

DISS. ETH NO. 29001

# Development of Hybrid Optoacoustic and Ultrasound Imaging Systems

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES  
(Dr. sc. ETH Zurich)

presented by  
BERKAN LAFICI

M.Sc. in Biomedical Computing, Technical University of Munich  
born on 26.06.1992

accepted on the recommendation of  
Prof. Dr. Daniel Razansky  
Prof. Dr. Srirang Manohar  
Prof. Dr. Fernando Perez-Cruz

2023

## ABSTRACT

Optoacoustic (OA) imaging provides a unique contrast to visualize and quantify light-absorbing tissue chromophores in living organisms. The light excitation of tissues together with consecutive sound wave acquisition combines optical contrast with ultrasound (US) resolution for functional and molecular imaging applications. Furthermore, multispectral optoacoustic tomography (MSOT) can provide spectral information in multiwavelength illumination mode by extracting the absorption properties of underlying tissues. Due to the unique contrast generated by OA images, this biomedical imaging modality is well-established in preclinical settings and clinical applications are introduced recently. However, wider adaptation and application of OA imaging in research and clinical settings as a standard tool require further improvements in contrast, spatial and temporal resolution, standardization of image processing methods, and open-source data sharing. The rich contrast in OA images originating from absorption and scattering in the tissues lacks the information about anatomy and elastic properties. Reflection US is a well-established method in biomedical imaging to monitor anatomical structures. Transmission US can provide information about speed of sound (SoS) changes, acoustic attenuation, and elastic properties of the tissues. Hybridization of these modalities in one imaging setup will benefit from the unique contrast provided by each of them. However, the development of such a hybrid imaging system imposes different constraints on transducer array designs, data transfer rates, and signal acquisition methods. These constraints can be mitigated to some extent by the development of optimized hardware solutions, image acquisition methods, or signal processing methods. The scope of this thesis is the development of new data acquisition and image processing methods to optimize hybrid OA and US imaging systems. Therefore, automated methods to segment boundaries of the anatomical structures in hybrid optoacoustic ultrasound (OPUS) images are proposed. The limited view or sparse acquisition artifacts in the spatial domain are further reduced by data-driven image processing methods for OA imaging. Signal domain limited view artifact removal is proposed for custom designed detector array optimized for hybrid OPUS imaging. US image acquisition sequences and data processing methods are developed to increase spatial and temporal resolution in tomographic acquisition settings. In addition, the applications of the developed transmission reflection optoacoustic ultrasound (TROPUS) system are explored for quantitative multimodal assessment of mammary tumors and non-alcoholic fatty liver disease (NAFLD). The proposed methods are expected to expedite the adaptation of hybrid OA and US systems in research and clinical settings.

## ZUSAMMENFASSUNG

Die optoakustische Bildgebung (OA) bietet einen einzigartigen Kontrast zur Visualisierung und Quantifizierung von lichtabsorbierenden Gewebeschromophoren in lebenden Organismen. Die Licht-basierte Anregung von Gewebe, zusammen mit der konsekutiven Detektierung von Ultraschallwellen, kombiniert den optischen Absorptionskontrast mit der Auflösung von Ultraschall (US) für funktionelle und molekulare Bildgebungsanwendungen. Darüber hinaus kann die multispektrale optoakustische Tomographie (MSOT) durch Extraktion der Absorptionseigenschaften des Gewebes spektrale Informationen in einem Beleuchtungsmodus mit mehreren Wellenlängen liefern. Aufgrund des einzigartigen Kontrasts der OA-Bilder, ist diese biomedizinische Bildgebungsmodalität in präklinischen Umgebungen gut etabliert, und klinische Anwendungen wurden kürzlich eingeführt. Eine umfassendere Anpassung und Anwendung der OA-Bildgebung in Forschung und Kliniken als Standardinstrument erfordert jedoch weitere Verbesserungen des Kontrasts, der räumlichen und zeitlichen Auflösung, eine Standardisierung der Bildverarbeitungsmethoden und die gemeinsame Nutzung von Open-Source-Daten. Der vielfältige Kontrast in OA-Bildern, der auf Absorption und Streuung im Gewebe zurückzuführen ist, enthält keine Informationen über Anatomie und elastische Eigenschaften. Reflexions-US ist eine bewährte Methode in der biomedizinischen Bildgebung zur Überwachung anatomischer Strukturen. Transmissions-US kann Informationen über Änderungen der Schallgeschwindigkeit (engl. "speed of sound", SoS), akustische Dämpfung und elastische Eigenschaften des Gewebes liefern. Die Hybridisierung dieser Modalitäten in einer Bildgebungsplattform wird von dem einzigartigen Kontrast profitieren, den jede von ihnen bietet. Die Entwicklung eines solchen hybriden Bildgebungssystems bringt jedoch verschiedene Einschränkungen in Bezug auf das Design der US-Wandleranordnung, die Datenübertragungsraten und die Methoden der Signalerfassung mit sich. Diese Einschränkungen können bis zu einem gewissen Grad durch die Entwicklung optimierter Hardware-Lösungen, Bildfassungsmethoden oder Signalverarbeitungsmethoden verringert werden. Das Ziel dieser Arbeit ist die Entwicklung neuer Datenerfassungs- und Bildverarbeitungsmethoden zur Optimierung hybrider OA- und US-Bildgebungssysteme. Daher werden automatisierte Methoden zur Segmentierung der anatomischen Strukturen in hybriden optoakustischen Ultraschallbildern (engl. "optoacoustic ultrasound", OPUS) eingeführt. Die Artefakte der begrenzten Sicht oder der dünnbesetzter Erfassung im räumlichen Bereich werden durch datengesteuerte Bildverarbeitungsmethoden für die OA-Bildgebung weiter reduziert. Die Beseitigung von Artefakten bei eingeschränkter Sicht im Signalbereich wird für eine speziell entwickelte Detektoranordnung vorgeschlagen, die für die hybride OPUS-Bildgebung optimiert ist. Es werden US-Bildaufnahmesequenzen und Datenverarbeitungsmethoden entwickelt, um die räumliche und zeitliche Auflösung in tomographischen Aufnahmeeinstellungen zu erhöhen. Darüber hinaus werden die Anwendungen des entwickelten Transmissions Reflexion optoakustischen

Ultraschall-Systeme (engl. “transmission reflection optoacoustic ultrasound”, TROPUS) für die quantitative multimodale Bewertung von Mammatumoren und nichtalkoholischen Fettlebererkrankungen (engl. “non-alcoholic fatty liver disease”, NAFLD) untersucht. Es wird erwartet, dass die vorgeschlagenen Methoden die Anpassung von hybriden OA- und US-Systemen in der Forschung und im klinischen Umfeld beschleunigen werden.

## ACKNOWLEDGEMENTS

I would like to start this chapter by expressing my greatest appreciation for the support of the Lafci family members for their constant guidance and love all the time. I also want to give a special thanks to them for always being there for me and supporting me in good and bad times. Special thanks to my mother Ayten and my father Halil for growing me up and supporting me through this path. To my sisters Secil and Seda, for always setting good examples for me and broadening my horizon since my childhood. I would like to thank my cousins for sharing their memorable moments with me whenever I visit my hometown and for supporting me through this journey.

And then our amazing research group. Special thanks to my advisor Prof. Daniel Razansky for providing guidance and support for many projects. And Xose Luis Dean-Ben, thanks for always helping us with the experiments and guiding us through our research. I cannot skip the fun activities we were doing with my colleagues. The basketball crew and running team made this journey fun for me and showed me that we are not only colleagues but also friends.

I would like to thank the Swiss Data Science Center (SDSC) for supporting us with the academic collaboration grant. Special thanks to Prof. Fernando Perez-Cruz, Firat Ozdemir, and Anna Susmelj from SDSC for working with us and creating valuable outcomes together in the projects. I would like to thank Prof. Srirang Manohar for being the co-examiner of this PhD thesis.

Finally, I want to thank all my friends who helped me through this journey. We started this PhD journey together with Cagla and Kaan. We shared the same house, we shared the same office, and we shared our best times together. I would like to thank them for being great flat mates, office mates, and more. Special thanks to “The Sürveyan” family from Bogazici University for supporting me from Istanbul whenever I need and cheering me up even though we are far away. Thank you, Gizem and Onur for always being there for me in good and bad times since our days in the Bogazici University. To my friend Cem, thanks for making my Munich journey special and for sharing the most enjoyable moments with me during our years as students at Bogazici University and also in Munich. And special thanks to Alper for joining me for random basketball games in various cities around the world, for visiting Zurich many times, and for doing projects with me at Bogazici University. I would like to thank our project team members Alper and Nazlı for the joyful discussions and moments filled with laughter at Boğaziçi University and years after. Thank you, Naz, for the great discussions, Barcelona trips and more. Murat, thanks for always opening your home to me since the first time I started to travel between Munich and Zurich. Finally, I would like to

extend my gratitude to Fernando for sharing lots of great moments and memories with me during our life in Zurich and always cheering me up.

## PUBLICATIONS

### JOURNAL PUBLICATIONS

- **Lafci, B.**, Merčep, E., Morscher, S., Dean-Ben, X. L. and Razansky, D., 2020. Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 68(3), pp.688-696.
- **Lafci, B.**, Merčep, E., Herraiz, J. L., Dean-Ben, X. L. and Razansky, D., 2020. Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound. *Neoplasia*, 22(12), pp.770-777.
- **Lafci, B.**, Robin, J., Dean-Ben, X. L. and Razansky, D., 2022. Expediting image acquisition in reflection ultrasound computed tomography. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 69(10), pp.2837-2848.
- **Lafci, B.**, Hadjichambi, A., Konstantinou, C., Herraiz, J. L., Burton, N. C., Dean-Ben, X. L. and Razansky, D., 2022. Multimodal Assessment of Non-Alcoholic Fatty Liver Disease with Transmission-Reflection Optoacoustic Ultrasound. *bioRxiv* (under review).
- Ozdemir, F.<sup>†</sup>, **Lafci, B.**<sup>†</sup>, Dean-Ben, X.L., Razansky, D. and Perez-Cruz, F., 2023. OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing. *Transactions on machine learning research*.
- Davoudi, N.<sup>†</sup>, **Lafci, B.**<sup>†</sup>, Özbek, A., Dean-Ben, X.L. and Razansky, D., 2021. Deep learning of image-and time-domain data enhances the visibility of structures in optoacoustic tomography. *Optics letters*, 46(13), pp.3029-3032.
- Hu, Y.<sup>†</sup>, **Lafci, B.**<sup>†</sup>, Luzgin, A., Wang, H., Klohs, J., Dean-Ben, X.L., Ni, R., Razansky, D. and Ren, W., 2021. Deep learning facilitates fully automated brain image registration of optoacoustic tomography and magnetic resonance imaging. *Biomedical optics express*, 13(9), pp.4817-4833.
- Özsoy, Ç., **Lafci, B.**, Dean-Ben, X. L. and Razansky, D., 2022. Real-time assessment of high-intensity focused ultrasound (HIFU)-induced heating and cavitation with hybrid optoacoustic-ultrasound (OPUS) imaging. (under review).
- Özgürün, B., **Lafci, B.**, Dean-Ben, X. L. and Razansky, D., 2022. Sparse optoacoustic sensing with convolutional dictionary learning. (under review).

- Robin, J.<sup>†</sup>, Rau, R.<sup>†</sup>, **Lafci, B.**, Schroeter, A., Reiss, M., Dean-Ben, X.L., Goksel, O. and Razansky, D., 2021. Hemodynamic response to sensory stimulation in mice: Comparison between functional ultrasound and optoacoustic imaging. *NeuroImage*, 237, p.118111.
- Grünherz, L., Gousopoulos, E., Barbon, C., Uyulmaz, S., **Lafci, B.**, Razansky, D., Boss, A., Giovanoli, P., Lindenblatt N., 2022. Preoperative Mapping of Lymphatic Vessels by Multispectral Optoacoustic Tomography. *Lymphatic research and biology*, 20(6), pp.659-664.

#### CONFERENCE PUBLICATIONS

- **Lafci, B.**, Merčep, E., Morscher, S., Deán-Ben, X.L. and Razansky, D., 2020. Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks. In *Photons Plus Ultrasound: Imaging and Sensing 2020* (Vol. 11240, p. 112402N). International Society for Optics and Photonics.
- **Lafci, B.**, Merčep, E., Herraiz, J.L., Deán-Ben, X.L. and Razansky, D., 2021. Transmission-reflection optoacoustic ultrasound (TROPUS) imaging of mammary tumors. In *Photons Plus Ultrasound: Imaging and Sensing 2021* (Vol. 11642, p. 116422L). International Society for Optics and Photonics.
- Klimovskaia, A.<sup>†</sup>, **Lafci, B.**<sup>†</sup>, Ozdemir, F., Davoudi, N., Deán-Ben, X.L., Perez-Cruz, F., and Razansky, D., 2022. Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data. *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning 2022*, PMLR 172:1173-1191.
- Davoudi, N.<sup>†</sup>, **Lafci, B.**<sup>†</sup>, Özbek, A., Dean-Ben, X.L. and Razansky, D., 2022. Learning-based enhancement of limited-view optoacoustic tomography based on image- and time-domain data. In *Photons Plus Ultrasound: Imaging and Sensing 2022* (Vol. 11960, pp. 7-13). SPIE.
- Hu, Y.<sup>†</sup>, **Lafci, B.**<sup>†</sup>, Luzgin, A., Wang, H., Klohs, J., Dean-Ben, X.L., Ni, R., Razansky, D. and Ren, W., 2022. Automatic Image Registration of Optoacoustic Tomography and Magnetic Resonance Imaging Based on Deep Learning. *First Conference on Biomedical Photonics and Cross-Fusion (BPC) 2022*.
- Klug, H., **Lafci, B.**, Skachokova, Z., Rudin, M., Razansky, D. and Ioanas, H.I., 2021. Machine Learning Enabled Brain Segmentation for Small Animal Neuroimaging Registration. *European Molecular Imaging Meeting (EMIM) 2021*.

#### PRESENTATIONS/TALKS

- Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks. SPIE, BiOS, 2020, San Francisco, California, United States.



- Transmission-reflection optoacoustic ultrasound (TROPUS) imaging of mammary tumors. SPIE, BiOS, 2021, Virtual.
- Reflection Ultrasound Computed Tomography with Sparse Acquisition Sequences. International Ultrasonics Symposium (IUS), Coded Excitation and Tomography, 2021, Virtual.
- Multi-Parametric Characterization of Fatty Liver Disease with Transmission-Reflection Optoacoustic Ultrasound (TROPUS). European Molecular Imaging Meeting (EMIM), Ultrasound and Optoacoustic Technologies, 2022, Thessaloniki, Greece.
- Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data. Medical Imaging with Deep Learning (MIDL), 2022, Zurich, Switzerland.
- Non-alcoholic fatty liver disease assessment by transmission-reflection optoacoustic ultrasound (TROPUS) computed tomography. SPIE, BiOS, 2023, San Francisco, California, United States.
- Limited-view artifact removal in optoacoustic tomography by signal domain learning. SPIE, BiOS, 2023, San Francisco, California, United States.

† equal contribution

## INTRODUCTION

Biomedical imaging provides various tools to explore organs and tissues at different spatial and temporal resolutions. Even though biomedical imaging modalities showed great improvements in terms of technical capabilities and user experiences, there are still some challenges and burdens to reach an ideal technology. There are tradeoffs between spatial and temporal resolution or level of invasiveness. Another consideration regarding the wide availability of the specific systems is the cost of the required components, maintenance, and operation. User dependency and reproducibility combined with the standardization of data formats and processing algorithms play a key role in the wider adaptation of the proposed solutions. The physical constraints such as penetration depth, signal attenuation, and angular coverage of the detector arrays that can be placed around the organs are also the limiting factors in many biomedical imaging applications. The comfort of the subjects and possible side effects need to be considered as well when the new systems are developed. In this thesis, we focus on developments in optoacoustic (OA) and ultrasound (US) imaging modalities and combination of both modalities in hybrid imaging setups for preclinical research and clinical applications by accounting for abovementioned challenges and application dependent limitations.

OA or photoacoustic (PA) imaging is based on illumination of objects with nanosecond laser pulses and consecutive detection of US waves resulting from thermoelastic expansion of the light-absorbing materials. The label-free OA image contrast originating from absorption and scattering of light in the diffusing medium makes it suitable for biomedical image applications. In addition, the multispectral acquisition capabilities using various wavelengths sequentially in visible or near infrared (NIR) spectrum make OA imaging suitable for identification of chemical composition of biological samples such as oxy-hemoglobin ( $\text{HbO}_2$ ) and deoxy-hemoglobin (Hb), lipid, collagen, and melanin via spectroscopic analysis. Hence, the spectral readings based on multi-wavelength excitation render additional information not captured by other biomedical imaging modalities. The strong optical absorption of hemoglobin allows the visualization of vascular structures and hemodynamic responses, maintaining sub-millimeter resolutions at depths of several centimeters within highly scattering living tissues for visible and NIR wavelengths. This unique contrast provided by OA imaging modality at high imaging speed increased its applications in the last decades [1, 2]. Tomographic imaging capabilities and easy-to-use handheld operation mode also make this imaging modality attractive for preclinical and clinical imaging setups [3, 4]. However, there are still bottlenecks that need to be improved by novel hardware designs, image acquisition methods, and reconstruction algorithms. In addition, OA imaging lacks the information about anatomical structures and alternative validation methods to quantify tissue properties. These bottlenecks can be complemented with the inclusion of other imaging modalities such as US in the same

imaging setups or the development of advanced algorithms which accounts for the constraints.

US imaging is based on the transmission of sound waves to objects and the detection of reflected waves (pulse-echo US) from objects or transmitted waves through (transmission US) the imaged object [5]. Reflection US is a well-established method to provide anatomical information in biomedical applications. Transmission US can provide information about elasticity, acoustic attenuation, and speed of sound (SoS) changes in the tissues. OA and US are complementary imaging modalities that can provide functional, molecular, and anatomical information together. The nature of these two imaging modalities makes them suitable for hybrid imaging applications. The same imaging setup can be optimized to operate in hybrid mode to acquire different contrast and quantitative metrics [6, 7]. However, the combination of these imaging modalities in one imaging setup is not straightforward while both impose different constraints on transducer array designs, data transfer rates, and signal acquisition settings. The imaging sequences, hardware designs, and reconstruction algorithms need to be optimized to account for the different needs of both modalities. For example, while OA imaging can reach high frame rates, US imaging requires compounding multiple frames to increase contrast which reduces the frame rate. While quantitative OA imaging can be performed by multispectral acquisition, it needs a validation method such as SoS imaging.

## OBJECTIVES

This doctoral thesis intends to find efficient methods for hybridization of OA systems with US imaging. The overall project includes optimization of hardware setup such as faster and efficient data acquisition techniques and algorithms, improvement of OA and US reconstruction algorithms and implementation of deep learning methods to further improve image quality. We employed a large number of datasets from tissue phantoms, small animals, and human volunteers using dedicated hybrid optoacoustic ultrasound (OPUS) and transmission-reflection optoacoustic ultrasound (TROPUS) imaging systems to develop novel imaging methods. Finally, the applications of developed imaging methods are shown for mammary tumor progression and non-alcoholic fatty liver disease (NAFLD) characterization *in vivo*.

The objectives of this thesis can be summarized in four main categories:

- (a) Image quality improvement by deep learning-based processing algorithms is proposed for the reduction of limited view and sparse acquisition artifacts in signal and image domains.
- (b) Automated segmentation methods are developed to increase the quantification capabilities of imaging setups, to reduce inter user variability, and to enable registration of images from different modalities.
- (c) Optimized acquisition schemes are developed using OPUS and TROPUS imaging systems for increased contrast and resolution in spatial and temporal domains by considering the tradeoffs.
- (d) The developed approaches in OPUS and TROPUS systems are demonstrated for tumor progression quantification and NAFLD assessment applications.

---

In this thesis, deep learning-based image improvement algorithms are proposed to mitigate limited view and sparse acquisition artifacts to reach objective (a) [8, 9]. The accompanying datasets and benchmarks are released with a dataset paper which explains the details about acquisition, use cases and developed software packages [10]. An automated segmentation algorithm based on deep neural networks is developed to achieve the objective (b) [11]. The automated segmentation helps to delineate mouse boundaries in preclinical imaging systems which can be used to apply fluence correction and increase accuracy in quantitative analysis. The automation of this process also decreases the user dependency and increases the reproducibility. The new image acquisition sequences and accompanying image reconstruction algorithms are developed to reach objective (c) [12]. Specifically, US computed tomography acquisitions are optimized to reach higher temporal resolution which is a limiting factor in hybrid systems due to data transfer rates. Tumor progression monitoring [13] and NAFLD assessment [14] are demonstrated using hybrid imaging systems to reach objective (d).

There are tradeoffs in every imaging method that is aimed at increasing temporal or spatial resolution, contrast, and energy exposures. The proposed methods are optimized for specific imaging setups and applications. The details about imaging systems and specific applications are included in this thesis. The researchers can optimize the proposed methods for their own applications and hardware requirements in the future. We believe that proposed methods will accelerate the wider adaptation of hybrid OA and US systems in preclinical and clinical imaging settings. Moreover, they will provide available tools to explore new horizons for future researchers.

## THESIS OUTLINE

The thesis is organized as follows.

Introduction chapter gives a concise summary of the thesis to help readers navigate the document by introducing the OA and US imaging modalities and improvement points. It also gives a brief explanation regarding the objectives and the outline of the thesis.

Chapter 1 (Background) introduces general aspects of currently available biomedical imaging systems, mentions about the challenges to apply these systems for diagnosis and treatment of various diseases and explains the physical phenomena behind OA and US imaging and constraints imposed by both modalities in the process of development hybrid imaging systems. The hardware requirements are explained in detail including all components such as detector arrays, data acquisition systems (DAQs), light sources, power supply units, and workstations.

Chapter 2 (Deep Learning for Automatic Segmentation of Hybrid Optoacoustic Ultrasound (OPUS) Images) involves the paper published in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control in 2020 [11]. The paper proposes an automated segmentation algorithm using convolutional neural network (CNN) for OA and US images. The algorithm is proposed for preclinical imaging to segment outer boundaries of mouse cross sections. How-

ever, the same method can be extended for segmentation of other anatomical structures in hybrid OPUS images if the manual annotations become available for network training.

Chapter 3 (Expediting Image Acquisition in Reflection Ultrasound Computed Tomography) covers the paper that is published in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control in 2022 [12]. The paper proposes sparse acquisition methods for higher temporal resolution in reflection ultrasound computed tomography (RUCT) using synthetic transmit aperture (STA) image acquisition sequences. Specifically, two different approaches (undersampling and combination of transducer elements) are compared using contrast to noise ratio (CNR) and spatial resolution. The proposed method is expected to expedite image acquisition and reconstruction in RUCT by mitigating the problems originating from increasing amount of data acquired, transferred, and stored.

Chapter 4 (Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data) includes the paper that is presented and published in Medical Imaging with Deep Learning (MIDL) conference in 2022 [8]. The paper introduces a signal domain processing method to remove limited view artifacts in OA images. Specifically, two different networks are proposed to remove the domain gap between experimental and synthetic data (style network) and reduction of limited view artifacts originating from reduced angular coverage (side network) in OA imaging. The proposed methods improve the image quality and quantification accuracy in OA imaging.

Chapter 5 (OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing) contains the paper that is published in Transactions on Machine Learning Research (TMLR) in 2023 [10]. The paper introduces an open-source OA dataset including the raw signals and reconstructed images with different imaging system configurations from human forearm. The standard OA image processing algorithms and benchmarking for deep learning methods are also included with the dataset. The published dataset and benchmark algorithms are expected to help standardization and comparison of different methods in OA imaging.

Chapter 6 (Noninvasive Multiparametric Characterization of Mammary Tumors with Transmission-Reflection Optoacoustic Ultrasound) covers the paper that is published in Neoplasia in 2020 [13]. The paper shows the successful application of TROPUS system for assessment of mammary tumors in preclinical settings. Specifically, three different modalities (OA, RUCT and SoS) are used to evaluate mammary tumors. The tumors are localized with RUCT images using the anatomical information provided by pulse-echo US. Afterwards, OA images are used to assess the vessel characteristics around the tumor region. SoS maps from transmission ultrasound computed tomography (TUCT) are used for quantitative evaluation and validation of the same tumor regions.

Chapter 7 (Multimodal Assessment of Non-Alcoholic Fatty Liver Disease with Transmission-

Reflection Optoacoustic Ultrasound) contains the paper that is available online in bioRxiv and under review [14]. The paper shows the application of TROPUS imaging for characterization of NAFLD. Three co-registered modalities are used to acquire images from healthy (control) and fatty liver regions. RUCT images are used to locate liver region in the mouse cross-sections. Multispectral optoacoustic tomography (MSOT) images facilitated quantification of lipid accumulation in liver using the spectral analysis. SoS maps quantified the changes in the sound propagation speed due to fat accumulation in the liver tissues.

Chapter 8 (Discussion and Future Directions) discusses the results achieved by this thesis. The scientific conclusions are drawn based on the findings of experiments and published results. As well as providing perspectives about the research directions that can be pursued based on this thesis, the chapter also suggests improvement points for the proposed methodologies and discusses possible new application areas.

## BIBLIOGRAPHY

- [1] Lei Li and Lihong V Wang. “Recent advances in photoacoustic tomography”. In: *BME Frontiers* (2021).
- [2] Srirang Manohar and Daniel Razansky. “Photoacoustics: a historical review”. In: *Advances in optics and photonics* 8.4 (2016), pp. 586–617.
- [3] Adrian Taruttis and Vasilis Ntziachristos. “Advances in real-time multispectral optoacoustic imaging and its applications”. In: *Nature photonics* 9.4 (2015), pp. 219–227.
- [4] Amalina Binte Ebrahim Attia et al. “A review of clinical photoacoustic imaging: Current and future trends”. In: *Photoacoustics* 16 (2019), p. 100144.
- [5] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [6] Roy GM Kolkman et al. “Real-time in vivo photoacoustic and ultrasound imaging”. In: *Journal of biomedical optics* 13.5 (2008), pp. 050510–050510.
- [7] Geng-Shi Jeng et al. “Real-time interleaved spectroscopic photoacoustic and ultrasound (PAUS) scanning with simultaneous fluence compensation and motion correction”. In: *Nature communications* 12.1 (2021), p. 716.
- [8] Anna Klimovskaia Susmelj et al. “Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data”. In: *International Conference on Medical Imaging with Deep Learning*. PMLR. 2022, pp. 1173–1191.
- [9] Neda Davoudi et al. “Deep learning of image-and time-domain data enhances the visibility of structures in optoacoustic tomography”. In: *Optics letters* 46.13 (2021), pp. 3029–3032.
- [10] Firat Ozdemir et al. “OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing”. In: *Transactions on Machine Learning Research* (2023). ISSN: 2835-8856.
- [11] Berkan Lafci et al. “Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 68.3 (2020), pp. 688–696.
- [12] Berkan Lafci et al. “Expediting image acquisition in reflection ultrasound computed tomography”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 69.10 (2022), pp. 2837–2848.
- [13] Berkan Lafci et al. “Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound”. In: *Neoplasia* 22.12 (2020), pp. 770–777.

- 
- [14] Berkan Lafci et al. “Multimodal Assessment of Non-Alcoholic Fatty Liver Disease with Transmission-Reflection Optoacoustic Ultrasound”. In: *bioRxiv* (2022), pp. 2022–08.



# CONTENTS

<b>Abstract</b>	<b>i</b>
<b>Zusammenfassung</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Publications</b>	<b>vi</b>
Journal Publications . . . . .	vi
Conference Publications . . . . .	vii
Presentations/Talks . . . . .	vii
<b>Introduction</b>	<b>ix</b>
Objectives . . . . .	x
Thesis Outline . . . . .	xi
<b>1 BACKGROUND</b>	<b>1</b>
1.1 Overview . . . . .	2
1.1.1 Biomedical Imaging . . . . .	2
1.1.2 Challenges in Biomedical Imaging . . . . .	3
1.1.3 Optoacoustic and Ultrasound Imaging Overview . . . . .	4
1.2 Hardware . . . . .	7
1.2.1 Light Sources . . . . .	7
1.2.2 Power Supply Units . . . . .	8
1.2.3 Detector Arrays . . . . .	8
1.2.4 Data Acquisition Systems . . . . .	9
1.2.5 Work Stations . . . . .	10
1.3 Optoacoustic Imaging . . . . .	11
1.3.1 Signal Acquisition . . . . .	12
1.3.2 Simulations . . . . .	13
1.3.3 Image Reconstruction . . . . .	15
1.3.4 Image Enhancement . . . . .	17
1.3.5 Spectral Unmixing . . . . .	23
1.4 Ultrasound Imaging . . . . .	24
1.4.1 Signal Acquisition . . . . .	24
1.4.2 Reflection Ultrasound Computed Tomography . . . . .	24
1.4.3 Speed of Sound Imaging . . . . .	28
1.5 Hybrid Optoacoustic Ultrasound . . . . .	30
1.5.1 Signal Acquisition . . . . .	30
1.5.2 Image Reconstruction . . . . .	30
1.6 Transmission Reflection Optoacoustic Ultrasound . . . . .	32
1.6.1 Signal Acquisition . . . . .	32
1.6.2 Image Reconstruction . . . . .	33
<b>2 DEEP LEARNING FOR AUTOMATIC SEGMENTATION OF HYBRID OPTOACOUSTIC ULTRASOUND (OPUS) IMAGES</b>	<b>41</b>
2.1 Introduction . . . . .	43

2.2	Materials And Methods . . . . .	44
2.2.1	Convolutional Neural Network . . . . .	44
2.2.2	Ground Truth Manual Segmentation . . . . .	45
2.2.3	Active Contour by Edge Detection . . . . .	45
2.2.4	Experimental Data Sets . . . . .	47
2.2.5	Evaluation Criteria . . . . .	48
2.3	Results . . . . .	48
2.4	Conclusion and Discussion . . . . .	53
<b>3</b>	<b>EXPEDITING IMAGE ACQUISITION IN REFLECTION ULTRASOUND COMPUTED TOMOGRAPHY</b>	<b>61</b>
3.1	Introduction . . . . .	63
3.2	Materials And Methods . . . . .	64
3.2.1	Imaging System . . . . .	64
3.2.2	Image Formation . . . . .	65
3.2.3	Channel Grouping in Transmission . . . . .	66
3.2.4	Channel Grouping in Reception . . . . .	66
3.2.5	Channel Grouping in Both Transmission and Reception . . . . .	66
3.2.6	Sparse Data Acquisitions . . . . .	67
3.2.7	Phantom Experiments . . . . .	67
3.2.8	Animal Experiments . . . . .	67
3.3	Results . . . . .	68
3.4	Discussion . . . . .	75
3.5	Conclusion . . . . .	77
<b>4</b>	<b>SIGNAL DOMAIN LEARNING APPROACH FOR OPTOACOUSTIC IMAGE RECONSTRUCTION FROM LIMITED VIEW DATA</b>	<b>83</b>
4.1	Introduction . . . . .	85
4.2	Methodology . . . . .	87
4.2.1	Style Transfer Network . . . . .	87
4.2.2	Side Network . . . . .	89
4.3	Experiments and Results . . . . .	90
4.3.1	Datasets Description . . . . .	90
4.3.2	Results . . . . .	90
4.4	Conclusion . . . . .	92
	Appendix . . . . .	94
<b>5</b>	<b>OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing</b>	<b>103</b>
5.1	Introduction . . . . .	105
5.2	Background . . . . .	106
5.2.1	Transducer Arrays . . . . .	108
5.2.2	Reconstruction Method . . . . .	108
5.2.3	Sparse Sampling . . . . .	108
5.2.4	Limited View . . . . .	109
5.3	Datasets . . . . .	110
5.3.1	Multispectral Forearm Dataset . . . . .	110
5.3.2	Single Wavelength Forearm Dataset . . . . .	111
5.3.3	Simulated Cylinders Dataset . . . . .	111
5.3.4	OADAT-mini Dataset . . . . .	112
5.4	Tasks . . . . .	112
5.4.1	Image Translation . . . . .	112

5.4.2	Semantic Segmentation . . . . .	112
5.5	Experiments and Results . . . . .	113
5.5.1	Data Split and Preprocessing . . . . .	114
5.5.2	Results . . . . .	114
5.6	Discussion . . . . .	116
5.7	Conclusion . . . . .	118
	Appendix . . . . .	120
<b>6</b>	<b>NONINVASIVE MULTIPARAMETRIC CHARACTERIZATION OF MAMMARY TUMORS WITH TRANSMISSION-REFLECTION OPTOACOUSTIC ULTRASOUND</b>	<b>142</b>
6.1	Introduction . . . . .	144
6.2	Materials And Methods . . . . .	145
6.2.1	The Imaging Setup . . . . .	145
6.2.2	Optoacoustic Tomography . . . . .	145
6.2.3	Reflection (Pulse-Echo) Ultrasound Computed Tomography . . . . .	146
6.2.4	Speed of Sound Imaging . . . . .	147
6.2.5	Animal Handling . . . . .	147
6.3	Results . . . . .	148
6.4	Discussion and Conclusion . . . . .	151
	Supplementary Material . . . . .	155
<b>7</b>	<b>MULTIMODAL ASSESMENT OF NON-ALCOHOLIC FATTY LIVER DISEASE WITH TRANSMISSION-REFLECTION OPTOACOUSTIC ULTRASOUND</b>	<b>160</b>
7.1	Introduction . . . . .	162
7.2	Results . . . . .	163
7.3	Discussion . . . . .	168
7.4	Materials And Methods . . . . .	170
7.4.1	Imaging System . . . . .	170
7.4.2	Multispectral Optoacoustic Tomography (MSOT) Imaging . . . . .	170
7.4.3	Reflection Ultrasound Computed Tomography (RUCT) Imaging . . . . .	171
7.4.4	Speed of Sound (SoS) Imaging . . . . .	171
7.4.5	Ex Vivo Liver Imaging . . . . .	172
7.4.6	Animal Experiments . . . . .	172
<b>8</b>	<b>DISCUSSION AND FUTURE DIRECTIONS</b>	<b>180</b>
	<b>Abbreviations</b>	<b>182</b>
	<b>List of Figures</b>	<b>197</b>
	<b>List of Tables</b>	<b>199</b>
	<b>Curriculum Vitae</b>	<b>200</b>

# 1 BACKGROUND

This chapter focuses on the background information that is needed to understand projects/-papers presented in this thesis. The first section (Overview) gives general information about biomedical imaging, challenges in this field, and where OA and US imaging stand in this bigger picture. The second section (Hardware) introduces each component of the imaging systems used within the scope of this thesis in detail. The third section (Optoacoustic Imaging) introduces the general principles behind the OA phenomena and how data acquisition to image formation processes works and can be improved. The fourth section (Ultrasound Imaging) is about the principles of US imaging, data acquisition sequences, and image formation. The fifth section (Hybrid Optoacoustic Ultrasound) explains the combination of OA and reflection (pulse-echo) US modalities in the same imaging setup and the constraints and benefits in hybridization. The sixth section (Transmission-Reflection Optoacoustic Ultrasound) introduces working principles of triple modality imaging system combining OA tomography together with the transmission and reflection US.

## 1.1 OVERVIEW

Imaging starts with capturing different physical phenomena happening in the world with suitable sensors. There are several imaging methods proposed in numerous technical fields. The scope of the imaging applications and their importance are enormous for life. One of the main application areas of imaging is the biomedical field. In this thesis, we focus on the specific subfield of biomedical imaging, namely OA and US imaging, and we propose new methods to pave the way toward improved imaging systems.

### 1.1.1 BIOMEDICAL IMAGING

Biomedical imaging is an essential tool for the diagnosis and treatment of various diseases [1]. Many biomedical imaging technologies have evolved over time, each providing specific benefits but also limitations. Innovative ideas in biomedical imaging opened new avenues for capturing anatomical, functional, and molecular changes at different scales. For instance, x-ray imaging, US, magnetic resonance imaging (MRI), and positron emission tomography (PET) all provide a different level of specificity, sensitivity, spatial and temporal resolution for specific biomedical applications [2].

X-ray imaging is based on the excitation of tissues with ionizing radiation, and detection of absorbed energy through penetration [3]. X-ray imaging provides high spatial resolution with high penetration depths. The imaging technology also combined with tomographic acquisition designs by surrounding the imaged target and rotating the detector elements around it [4]. The x-ray computed tomography (CT) can provide slices of full body scan with high resolution. However, the imaging modality suffers from limited molecular specificity and sensitivity [5]. Moreover, ionizing radiation can harm living tissues, hence, the radiation dose and scan time of x-rays should be arranged accordingly by finding a balance between contrast and harmfulness. Specific contrast agents are proposed to increase contrast and functional information that can be captured with x-ray to overcome the limitations [6].

MRI is based on the generation of strong magnetic fields inside an isolated medium. The movements of protons under this magnetic field and the gradients are used to generate the images [7]. The images are acquired in tomographic slices which can be used to scan full body by the movement of detectors. MRI can also provide anatomical information at high spatial resolution. Also, MRI does not require exposure to any ionizing radiation which makes it safer for biomedical applications. However, MRI suffers from tradeoffs between low specificity, sensitivity, and temporal resolution [8]. The cost of required components and maintenance in MRI is relatively high compared to the other imaging modalities.

PET imaging is based on detection of radioactive gamma rays emitted by the radiotracer materials injected in the body [9]. The imaging system can provide high-sensitivity functional and molecular information [10]. However, PET imaging suffers from low spatial and temporal resolution. In addition, administration of radiotracers inside the body raise the

---

concerns about potential effects of ionizing radiation for the patient and people around the patient until the effects of radiotracers are eliminated [11].

The examples about biomedical imaging modalities can be augmented. However, each imaging modality comes with its advantages and the drawbacks or challenges about resolution, safety, costs and required resources.

### 1.1.2 CHALLENGES IN BIOMEDICAL IMAGING

One of the most important aspects of the biomedical imaging systems is the safety of the subjects and operators. The amount of injected contrast agents, dose and exposure time for excitation of the tissues should not exceed the safety limits to avoid any harmful effects [12]. However, the dose and time of exposure, and the amount of contrast agents limit the contrast (increased noise levels) that can be achieved via the biomedical imaging systems. Hence, the development of efficient hardware and software are crucial to maintain patient and operator safety and to reduce the harmful effects. Highly parallelized hardware components and processing algorithms constitute one of the main challenging points to achieve in most of the biomedical imaging modalities. The costs and the storage also need to be accounted while optimizing acquisition specific parameters.

There are also tradeoffs between achievable imaging parameters such as penetration depth, detected frequencies, spatial and temporal resolution. Spatial resolution is crucial to resolve the structures requiring detection to perform diagnosis or treatment [13]. The imaging modality properties should be chosen carefully based on the required spatial resolution. The spatial resolution is limited by several parameters originating from imaging system or imaged target such as the properties of the detectors, excitation source characteristics, scattering, and signal sources generated by the excited materials.

Temporal resolution of the imaging system should be enough to capture fast-changing functional or dynamic phenomena [13]. There are several factors limiting the achieved temporal resolution such as the number of detectors and data transfer rates. The number of detectors can be decreased to achieve higher frame rates (temporal resolution). However, this will result in arc-shaped undersampling artifacts and reduced contrast. The number of acquired sample points or sampling frequency of the signals can be decreased to achieve faster scans but they will lead to reduced field of view (FOV) and spatial resolution, respectively. Data transfer rates can be increased with the improved data acquisition hardware which in return requires specialized devices with higher costs. In addition, parallelized or compressed sensing signal processing algorithms need to be developed to fulfill the requirements originating from increased imaging speed [14].

Another challenge for non-invasive imaging is the physical constraints originating from access to different organs and tissues. The penetration depth of the employed imaging system is not always sufficient to capture information from deeper structures [15, 16]. It is not possible

to reach the organs or the tissues from all angles for specific penetration depths. Thus, the design of the detector arrays should be optimized to account for tradeoffs between physical constraints and imaging depths. The angular coverage that can be achievable based on the physical constraints affects the accuracy of the shapes of anatomical structures. For most of the imaging modalities, the optimal data acquisition can be performed by surrounding the objects with detectors. However, this is not possible due to above mentioned physical constraints and cost considerations. The costs increase with the number of detectors used, the processing power employed, parallelized processing operations and storage.

Yet, there is no ideal biomedical imaging system that can overcome all the challenges in the field. However, it is possible to make a choice of imaging system based on the required spatial and temporal resolution, allocated resources, and the area of the application. The new hardware and software designs can be developed by accounting for the tradeoffs between imaging parameters.

### 1.1.3 OPTOACOUSTIC AND ULTRASOUND IMAGING OVERVIEW

OA imaging combines the advantages of both optics and US as it uses pulsed nanosecond light radiation for the excitation of tissues and detects the US waves generated by the absorption and following thermoelastic expansion. The imaging modality uses safe levels of non-ionizing light radiation in the visible and NIR spectral ranges [17]. As a result, OA imaging generates contrast based on light absorption with molecular and functional detection capabilities. Compared to other optical methods, low scattering of US provides information from deeper structures by removing the limitations of optical diffusion. In addition, the high sensitivity, and specificity of OA imaging for functional and molecular applications makes it an attractive tool in the biomedical imaging field. Those properties can be combined in real-time, non-invasive, and handheld imaging setups for volumetric OA imaging. Sequential excitation with changing wavelengths of light is also used for spectral analysis of structures in OA images. MSOT exploits the advantages of spectral absorption differences among various chromophores to separate and quantify the existence of specific materials in the imaged objects [18].

OA imaging is rapidly finding its place as a preclinical and clinical imaging modality [19, 20]. Standalone applications of OA imaging for preclinical research showed promising results for the detection and assessment of several disease conditions. Specifically, tumor progression is monitored using the oxygenation levels and vessel structure changes at the core of the tumor region and surrounding tissues [21–23]. Cardiovascular investigation of heart functions and arrhythmias are performed with sparse acquisition sequences to reach ultrafast four-dimensional imaging [24]. Indocyanine green (ICG) is used for monitoring functional properties of the underlying tissues and organs [25, 26]. Functional neuroimaging is successfully shown using three-dimensional (3D) detector arrays [27]. Localization in OA tomography is performed to increase resolution in acquired images [28–30]. Magnetic microrobots at the nanometer scale are tracked in 3D using the real-time imaging capabilities of OA imaging [31]. OA microscopy

---

systems are proposed for monitoring skin cancer (melanoma) progression and wound healing processes [32, 33]. Further examples of OA imaging in preclinical research are available in the literature. Clinical translations of these applications are also an important aspect for the adaptation of this imaging modality in hospital settings.

Clinical applications of OA imaging are suggested using commercially available systems for patient studies or custom-engineered systems for volunteer studies. Crohn’s disease is characterized *in vivo* using multispectral properties of OA imaging [34, 35]. Carotid arteries are imaged with 3D handheld detector arrays [36] and carotid plaques are characterized in terms of lipid accumulation and oxygenation with two-dimensional (2D) detector arrays [37]. Brain imaging of humans with thin skulls is successfully performed by using a 3D detector array and rotating the acquisition elements around the brain [38]. Multispectral imaging of human breast cancer is performed with 2D concave detector arrays using 28 different wavelengths [39]. A 3D system is also developed for breast imaging by combining 12 arch-shaped detector arrays with 32 transducer elements [40, 41].

All these applications are made possible thanks to technical developments in the OA imaging field in terms of data acquisition and signal processing methods. Improved image reconstruction methods are developed to generate images with reduced noise levels and higher quantification accuracy [42, 43]. A Bayesian approach is employed to solve the reconstruction problem in spectral quantitative OA imaging [44]. The compressed sensing methods are proposed to reach higher imaging rates to monitor fast-changing dynamic properties [45, 46]. Deep learning methods are proposed to remove sparse acquisition artifacts [47, 48] and limited view artifacts resulting from reduced angular coverage [49–54]. Bandwidth enhancement is performed in the signal domain using end-to-end neural networks [55]. Dictionary learning is applied to OA images to recover vessel structures from undersampling acquisition scenarios [56] and to enhance the structures in a light emitting diode (LED) based imaging [57]. Deep learning methods enable semantic segmentation of structures in OA images [58, 59]. Multispectral OA images are unmixed to their components by learning-based methods [60, 61]. LEDs are used to replace laser sources in OA imaging, which resulted in lower contrast but reduced cost and higher repetition rates [62–64].

Despite the wide application areas of OA imaging as a standalone imaging modality in pre-clinical and clinical settings, it lacks the anatomical information to localize the structures. There is also a need to validate the quantitative findings in OA imaging with other modalities. To address those challenges, it is possible to combine OA imaging with US modality which uses similar detector arrays for image acquisition. This combination might help to complement the molecular and functional information provided by OA imaging with anatomical data from reflection US and stiffness and tissue density characteristics that can be extracted from transmission US imaging. However, integration of these modalities in one imaging device is not straightforward as they impose different constraints on detector array designs (i.e., pitch size, angular coverage), data acquisition rates, and image processing algorithms. For exam-



ple, a common approach employs linear US detector arrays combined with the light sources by fibers mounted on the probe [65, 66]. However, this geometry is better suited for reflection US imaging than OA imaging which requires increased angular coverage and transmission US imaging which requires the detection of transmitted waves through the imaged object.

In this thesis, we develop efficient hardware and software solutions for hybrid OA and US imaging systems. The imaging systems are optimized to capture the rich contrast provided by each modality. Novel algorithms are proposed to eliminate the constraints imposed by different modalities. Further, the applications of the proposed systems are shown in tumor and liver disease characterization. We hope that the proposed methods will help to mitigate the challenges in the biomedical imaging field.

## 1.2 HARDWARE

Hybrid OA and US imaging systems are composed of light sources, power supply units, detector arrays, DAQs, and workstations. The objects are excited by two different methods: light sources (for OA imaging) or US waves (for reflection or transmission US imaging). The nanosecond pulsed light sources are used to excite objects by sequentially switching wavelengths in OA imaging. The thermoelastic expansion caused by the light-absorbing materials creates US waves which are captured by detector arrays and converted into OA images by algorithms running on the workstations. The US waves can be also generated and transmitted by the transducer elements on the detector arrays for reflection and/or transmission US imaging. Reflection US can be performed by collecting the US waves reflected by the objects. Transmission US imaging is performed by attenuation of US waves during the transmission through the objects. So, the detector arrays are used as source and receiver in US imaging. Transducer elements on the detector arrays convert the US pressure waves into electrical signals. DAQ digitizes the acquired signals simultaneously for each channel of the detector array. The digitized signals are transferred to workstations for real time visualization using graphics processing unit (GPU) accelerated image reconstruction algorithms. The raw signals are stored by the workstations for offline processing. Each of these components is explained in detail in the following subsections.

### 1.2.1 LIGHT SOURCES

Short-duration pulsed light sources are utilized for OA imaging. The pulsed excitation of the objects with ultrafast (nanosecond) light sources causes heating by absorption and scattering. The sudden increase in the temperature results in the expansion and vibration of the structures in the excited objects. This effect is called thermoelastic expansion which generates the emission of US waves as a point source. The excitation light is delivered with the optical coupling by fiber bundles designed for specific illumination methodologies. The illumination pattern in this thesis is arranged to deliver planar focusing (multisegment array) or cylindrical focusing (circular arrays) depending on the detector geometry.

***Lasers:*** Nanosecond pulsed laser sources are used to excite the imaged objects by sequentially switching the illumination wavelengths. Specifically, tunable optical parametric oscillator (OPO) lasers in visible and NIR wavelength ranges are utilized to generate OA signals by thermoelastic expansion. The repetition rate of the lasers has a direct impact on the imaging speed and the laser energy. The laser repetition frequency can be optimized for the targeted applications. The lasers between 10 Hz and 100 Hz repetition rates are used in this thesis (InnoLas Laser GmbH, Krailling, Germany).

***Light Emitting Diodes:*** LEDs are used in OA imaging setups as an alternative to lasers. This choice of LED light source reduces the cost of the overall system and increases safety [62, 67]. The main disadvantage of LED-based excitation is the reduced energy delivery on the objects. The reduced energy comes with the cost of decreased contrast in OA images. The

problem of reduced energy delivery can be mitigated by increasing the repetition rate of LED pulses. Then, multiple image frames can be combined to create final high-contrast images. However, this method cannot be applied to fast-moving objects and dynamic processes. The use of LEDs increases the safety of the OA imaging systems by decreasing the risk of accidental laser damage by extreme energy level exposure and wrong usage of eye protection goggles.

### 1.2.2 POWER SUPPLY UNITS

US waves are generated by transmitting bipolar pulses in different combinations with individually assigned delay parameters to create desired wave shapes. The power supply units provide the required energy for DAQ to form US pulses by activating transducer elements. The voltage and the current of the power supply units are limited to 40 V<sub>pp</sub> and 20 Ampere, respectively, to avoid damaging the transducer elements and imaged objects. Principally, the main sources of excitation are the light sources and power supply units in OA and US imaging, respectively.

### 1.2.3 DETECTOR ARRAYS

The detector arrays are used to collect pressure waves generated by the OA effect, reflection, and transmission of US waves. The collected pressure signals are converted into electrical signals by the piezoelectric transducer elements in the detector arrays. The transducer elements are combined in various geometrical shapes for optimized OA and US imaging designs for specific applications.

**Circle Array:** The custom-engineered circular transducer array (Imasonic Sas, Voray, France) was designed for full-body preclinical tomographic imaging (Fig. 1.1a). It contains 512 transducer elements that are distributed on the two half circles (174°) equidistantly. Each element on the detector array can transmit and receive US signals sequentially. The two half circles of the circular detector array are combined in a 2D plane to create 348° tomographic coverage around the imaged object. The combined full-circle array has 40 mm of the radius of curvature. The transducer elements are cylindrically focused at a 38 mm distance. Each element has dimensions of 0.37 mm x 15 mm. The interelement pitch size is 0.47 mm. The array has a 5 MHz peak central frequency with a bandwidth of 60% at -6 dB.

**Semi-Circle Array:** The circle array can be separated into two semi-circles to perform handheld imaging in clinical applications (Fig. 1.1b). The semi-circle array has the same physical properties as the circle array except for the angular coverage and number of transducer elements. The semi-circle array can reach 174° angular coverage with 256 transducer elements. This specific array geometry in which transducer elements are distributed on an arc equidistantly helps to reduce artifacts generated by the discontinuity in the transducer element organization i.e., multisegment array. Hence, the semi-circle array can be used as an alternative to a multisegment array to reduce artifacts in OA images for handheld acquisitions. However, semi-circle array design is not optimized for planar pulse-echo US imaging.

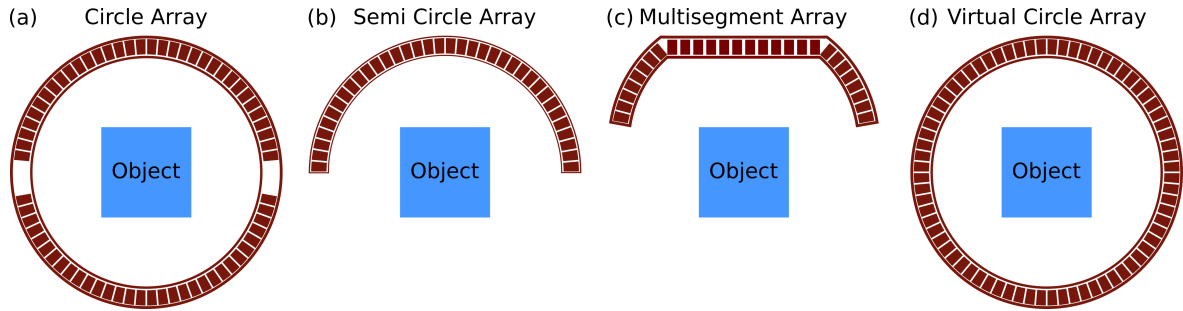


FIGURE 1.1: Representative drawings for detector arrays. (a) Circle array with  $348^\circ$  tomographic coverage including 512 transducer elements. (b) Semi-circle array formed by one half of the circle array including 256 transducer elements. (c) Multisegment array designed as a combination of linear and concave parts including 128 transducer elements on the linear part and 64 transducer elements on the concave parts (256 elements in total). (d) Virtual circle array to mimic  $360^\circ$  tomographic coverage with 1024 transducer elements.

**Multisegment Array:** The custom-engineered multisegment detector array (Imasonic Sas, Voray, France) is formed by the combination of two different geometries, namely, linear, and concave parts (Fig. 1.1c). The linear part is located at the center of the detector array with 128 transducer elements aligned on a straight line. The interelement pitch size of the linear part is 0.25mm. The linear geometry is optimized for reflection US data acquisition, but it is also used for OA imaging. There are two concave parts merged into the linear part from both sides with 64 transducer elements. The interelement pitch size of concave parts is 0.6 mm. The overall detector array is composed of 256 transducer elements focusing at 40 mm through the center of the imaging plane. The array has a 7.5 MHz peak central frequency with 70% detection bandwidth at -6 dB.  $170^\circ$  angular coverage is achieved with the specified array design.

**Virtual Circle Array:** The virtual circle detector array is simulated by mimicking 1024 elements distributed on a circle equidistantly (Fig. 1.1d). The virtual circle has the same radius (40 mm) with a circle array manufactured for hybrid OPUS and TROPUS acquisitions. In contrast to the circle detector array, there is no gap between the two semi-circles of virtual detector array. The number of transducer elements (1024) is doubled compared to the circle detector array (512). The virtual circle array was used in simulations to mimic the cases with increased angular coverage and spatial sampling.

#### 1.2.4 DATA ACQUISITION SYSTEMS

The detected pressure (US) waves are converted into electrical signals by piezoelectric transducer elements. The continuous electrical signals transferred from detector arrays need to be digitized by a DAQ. A custom-engineered DAQ (Falkenstein GmbH, Munich, Germany) is used in this thesis for the digitization of the acquired signals. The DAQ can digitize a maximum of 512 channels simultaneously. The transfer rate of the digitized signals is limited to 1 Gbit/s by DAQ. The sampling frequency and the number of samples can be optimized for required imaging frame rates (i.e., 24 or 40 mega samples per second (MSPS)). The DAQ is also responsible for the transmission of US waves with specified delays to perform US imaging applications with custom waveforms.

### 1.2.5 WORK STATIONS

Data transfer, synchronization, storage, and image processing are performed by the workstations with the GPUs. DAQs are connected to workstations with 1 Gbit/s ethernet connection to transfer the transmission and reception commands and the acquired signals. A Matlab package is used to synchronize the trigger events between light sources or function generators and DAQs. The same Matlab package with a custom-designed graphical user interface (GUI) reconstructs the images in real-time with GPU support using the OpenCL library. The acquired signals are saved in Matlab-compatible files for offline processing. Matlab and Python libraries are used for the offline processing of the acquired signals.

### 1.3 OPTOACOUSTIC IMAGING

OA imaging relies on the principle of excitation of objects with nanosecond light pulses and acquisition of generated US waves by transducer arrays that surround the imaged object. The main components used in OA imaging include nanosecond pulsed laser or LED arrays, US transducer array, DAQ, and processing workstation (see chapter 1.2 for details). The light excitation creates US (pressure) waves that propagate in a 3D imaging medium. OA wave equation for a short duration (less than stress and thermal relaxation time) pulsed excitation can be written as follows

$$\frac{\partial^2 p(r, t)}{\partial t^2} - c^2 \nabla^2 p(r, t) = \Gamma H(r, t) \frac{\partial \delta(t)}{\partial t} \quad (1.1)$$

where  $p(r, t)$  is the pressure wave that depends on spatial ( $r$ ) and temporal ( $t$ ) variables.  $c$  is the SoS in the imaging medium.  $\Gamma$  stands for the Grüneisen constant [68].  $H(r, t)$  is the absorbed energy field relative to the same spatial ( $r$ ) and temporal ( $t$ ) changes.  $\delta(t)$  represents the temporal ( $t$ ) laser light intensity change based on illumination. Initial value of the above equation can be written as

$$\frac{\partial^2 p(r, t)}{\partial t^2} - c^2 \nabla^2 p(r, t) = 0 \quad (1.2)$$

with the initial conditions

$$p(r, t)|_{t=0} = \Gamma H(r) \quad (1.3)$$

and

$$\frac{\partial p(r, t)}{\partial t} |_{t=0} = 0. \quad (1.4)$$

The Poisson solution of the above OA pressure wave equation can be written as

$$p(r, t) = \frac{\Gamma}{4\pi c} \frac{\partial}{\partial t} \int_{S'} \frac{H(r')}{|r - r'|} dS'(r, t) \quad (1.5)$$

where  $S'$  is the temporal spherical surface defined by  $|r - r'| = ct$  in 3D. The constant terms outside of the integral can be neglected. After dropping the constant terms, the pressure, in arbitrary units, can be expressed as

$$p(r, t) = \frac{\partial}{\partial t} \int_{S'} \frac{H(r')}{|r - r'|} dS'(r, t) \quad (1.6)$$

for 3D volumetric model of pressure waves. The same equation can be defined on a curve  $L'$  for 2D imaging medium as

$$p(r, t) = \frac{\partial}{\partial t} \int_{L'} \frac{H(r')}{|r - r'|} dL'(r, t). \quad (1.7)$$

These equations represent the OA forward model in 3D and 2D space. The inverse problem of given equations can be solved to reconstruct OA images using backprojection [69] (see chapter 1.3.3 for details). Alternatively, the inversion of the discretized version of forward model can be used for model-based reconstruction [43] (see chapter 1.3.3 for details).

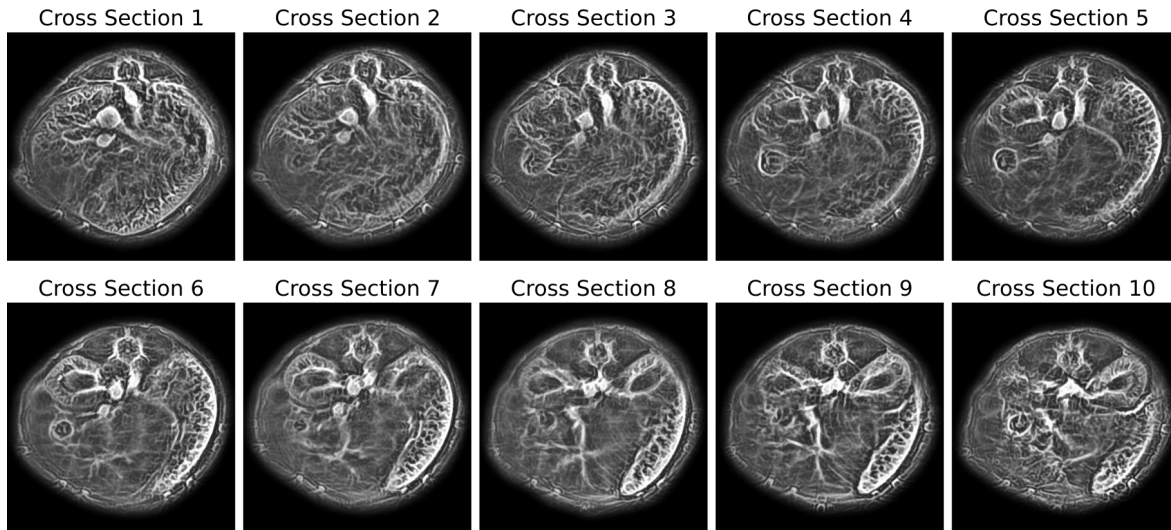


FIGURE 1.2: Example optoacoustic (OA) images from healthy mouse cross-sections using circle detector array with 1 mm step size in elevational direction. The images starting from the liver region (cross-section 1) to the kidney and spleen region (cross-section 10) are shown in the examples. Conventional image enhancement methods (see chapter 1.3.4 for details) are applied on all cross sections.

The example images from OA imaging systems after the reconstructions are shown above. Fig. 1.2 shows cross-sectional images acquired with the circler transducer array by moving the detector in vertical direction to acquire different anatomical regions of the mouse body. The images are generated by using two different SoS values for mouse body and surrounding medium (water). The conventional image enhancement methods, namely fluence correction, adaptive histogram equalization, Frangi filtering, and background suppression by masking, (see chapter 1.3.4 for details) are applied to increase the visibility of anatomical structures.

Handheld transducer arrays are used in clinical OA applications to acquire cross-sectional images from humans similar to clinically available US transducers. Fig. 1.3 shows example OA images from the human forearm acquired by semi-circle, multisegment, and linear detector arrays (see chapter 1.2.3 for details). The images are reconstructed by a standard backprojection algorithm using a single SoS algorithm. No image enhancement method is applied to the images after the reconstruction.

### 1.3.1 SIGNAL ACQUISITION

Time domain signals of US waves generated by thermoelastic expansion are acquired by the custom- engineered detector arrays. The design (interelement pitch size, focusing, element size, and sensitivity) and angular coverage of the transducer arrays vary depending on the application (see chapter 1.2.3 for details). The piezoelectric transducer elements on the detector arrays convert the mechanical pressure waves into electrical signals. Then, acquired electrical signals are digitized by DAQ. The signal acquisition rate is limited by a 1 Gbit/s ethernet interface between the workstation and DAQ. For example, 494 samples from 512 transducer elements can be acquired with a maximum frame rate of 100 Hz using this interface [45]. The

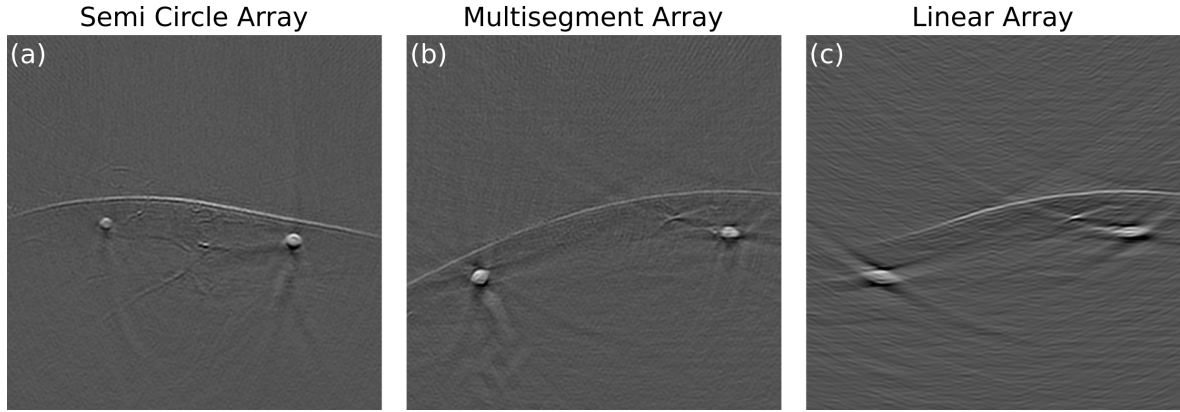


FIGURE 1.3: Example optoacoustic (OA) images from human forearms using semi-circle, multisegment, and linear (central part of multisegment) detector arrays. (a) Semi-circle array acquisition with 256 transducer elements and  $174^\circ$  angular coverage. (b) Multisegment detector array acquisition with 256 transducer elements and  $170^\circ$  angular coverage. (c) Linear (central part of multisegment) detector array acquisition with 128 transducer elements.

sampling frequency of DAQ, number of acquired samples, and activated transducer elements can be adjusted to specific values based on the desired sampling/discretization resolution.

Signal domain data provided with the projects presented in this thesis is stored in the following format in terms of dimensions: time points (number of samples)  $\times$  detector array channels  $\times$  frames (repetitions). The number of samples is restricted by the presets of the DAQ. Also, the sampling frequency of the time points can be adjusted to specific values predefined by the DAQ. The number of channels equals to transducer elements on the detector array, but the unused channels can be deactivated by DAQ for desired application (compressed sensing, sparse sampling) to speed up the signal acquisition. Frames (repetitions) are dependent on the duration of the acquisition and repetition frequency of the trigger events. MSOT acquisitions contain the frames acquired at different wavelengths for a defined number of repetitions. The raw signals from DAQ are saved without any processing. However, bandpass filtering and normalization around 0 mean are applied for each channel separately before feeding them into image reconstruction algorithms.

### 1.3.2 SIMULATIONS

Simulations of OA images are performed to test the implemented algorithms and to generate training datasets to apply learning-based image processing methods. The following steps are applied to generate simulated OA images: i) The geometrical shapes that mimic vessels and skin lines are simulated in 2D. ii) Forward OA model [43] is applied to generate corresponding raw signals for different detector array designs. iii) Backprojection or model-based image reconstruction algorithms are applied to create final OA images using the transducer element positions.

The simulations used in this thesis assume that the absorbed energy in the simulated structures is already known at single-wavelength excitation. The modelling of light illumination



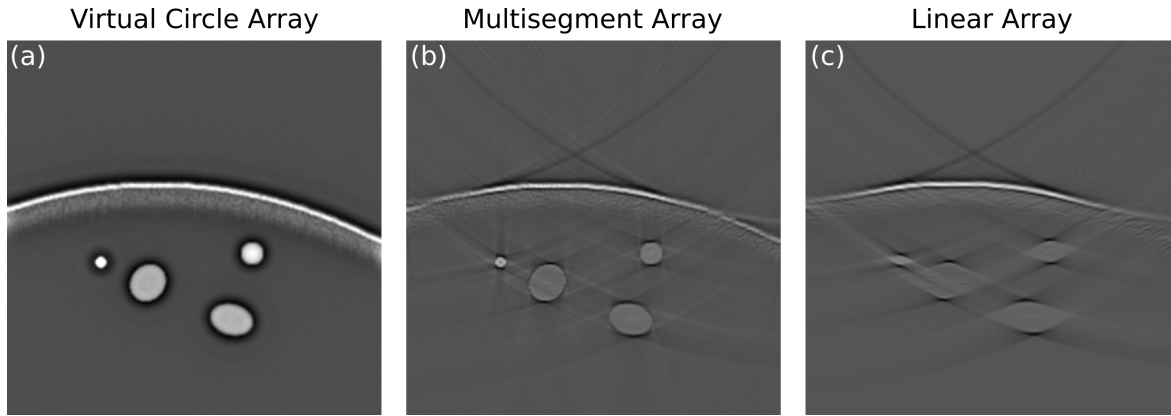


FIGURE 1.4: Simulated human forearm optoacoustic (OA) images using virtual circle, multisegment, and linear detector arrays. The circular high-intensity regions represent the vessels and parabolic line represents the skin line. (a) Simulation by virtual circle array with  $360^\circ$  angular coverage. (b) Simulation by multisegment array including linear and concave parts with  $170^\circ$  angular coverage. (c) Simulation by linear (central part of multisegment) detector array including 128 transducer elements.

and assignment of structure-specific absorption coefficients are not employed for the simulations. The simulations do not involve the multispectral acquisition capabilities of OA imaging. So, there is no MSOT data simulated in the scope of this thesis. Vessel structures are simulated using elongated and round shapes to mimic parallel and perpendicular vessels to the imaging plane. Specifically, a random number of ellipses with different intensity profiles are generated. The skin line is generated by fitting a second-degree polynomial function to 3 points that are defined in the specific height range from the top and the bottom of the images based on observations from experimental data. The generated curve after polynomial fitting is smoothed by Gaussian filtering. Finally, non-structured uniform noise is included under the skin line with exponential decay. Fig. 1.4 shows one cross-section of a simulated human forearm image using different transducer arrays, namely, virtual circle, multisegment, and linear (central part of multisegment) detector arrays. The simulations show artifacts similar to experimental data which originate from a limited view (Fig. 1.4c) and sparse acquisitions. The simulation of artifacts and noise structures is as important as simulating anatomical structures to test the efficiency of the proposed algorithms or to train learning-based methods.

Human forearm dataset is generated by assuming that vessels are positioned perpendicular to the imaging plane which is the case for experimental data acquired with multisegment and semi-circle array (Fig. 1.3). However, scanning with different orientation of detector array can generate structures that are not available in the presented simulation and experimental data. The scripts to generate simulations, to read the experimental and simulated data, and to try image processing algorithms are available online with the accompanying paper [70]. The simulations should be further modified if different vessel and skin geometries are preferred.

### 1.3.3 IMAGE RECONSTRUCTION

Image reconstruction refers to the algorithms that are utilized for the formation of images from acquired raw electrical signals by US detectors in the scope of this thesis. There are two main image reconstruction algorithms used in the OA imaging studies mentioned here. These are backprojection and model-based image reconstruction methods. Different versions of these image reconstruction algorithms are utilized based on the needs of specific applications. Two different signal preprocessing steps are applied before any type of image reconstruction algorithm. First, raw signals are bandpass filtered for the specific central frequency of the detector array used in the data acquisition. Then, the filtered signals are normalized around 0 mean for each individual channel. After these preprocessing steps following image reconstruction algorithms are utilized to generate OA images:

**Backprojection:** Standard backprojection algorithm creates a discrete grid in the FOV based on the desired resolution. Then, the time of flight (TOF) between each point on the grid and each transducer element on the detector array is calculated based on the defined SoS value. So, the electrical signals in the time domain are summed up based on their TOF from the discretized grid to US detectors to generate the final image. Theoretically, OA images are reconstructed by the absorbed energy field  $H(r')$  at a specific location depending on the measured pressure waves.  $H(r')$  at surface S for 3D or at curve L for 2D is calculated by the following equation (see chapter 1.3 for details)

$$H(r') = \frac{1}{\Gamma} \int_{\Omega} \frac{d\Omega}{\Omega} [2p(r, t) - 2t \frac{\partial p(r, t)}{\partial t}]_{t=\frac{r-r'}{c}} \quad (1.8)$$

where the constants in the equation can be omitted. After removing the constants, the equation can be discretized as

$$H(r'_j) = \sum [p(r_i, t_{ij}) - t_{ij} \frac{\partial p(r_i, t_{ij})}{\partial t}] \quad (1.9)$$

where  $r_i$  is the position of i-th transducer on the detector array,  $r'_j$  is the j-th point on the reconstruction grid and  $t_{ij} = |r_i - r'_j|/c$ . Overall, the backprojection equations summarize the summation of the signals for the calculated TOF between the specific location of the grid and transducer element.

In a standard backprojection algorithm, the medium is considered to have homogenous characteristics with the same SoS everywhere. However, the homogenous SoS assumption for imaging medium and different tissue types do not yield accurate image reconstructions. The SoS is changing based on the medium that the US waves travel in. The quality of the reconstructed image by backprojection algorithm depends on the accurate assignment of SoS for changing mediums. To account for heterogenous SoS in the medium, we developed backprojection algorithm that uses two different values for tissues inside the body and the background imaging medium. This algorithm improves the quality of OA images, but it requires access to the segmentation of the outer boundaries of the imaged object. Thus,, accurate segmen-

tation should be obtained to assign SoS values. The segmentation can be performed by manual annotations or automated methods [58, 59]. The same algorithm can be extended for the assignment of specific SoS values for each tissue type. However, this will require the segmentation of each organ or tissue type separately which increases the processing time substantially. Another approach to increase the performance of backprojection is to use the weighting of signals based on the distance from the surface. The signals can be weighted with linear or exponential models to account for light attenuation during the penetration.

**Model-Based:** The model-based reconstruction algorithm first discretizes the forward OA equation in finite space. The point sources on a 2D or 3D grid can be expressed as sparse matrix by this discretization. Then, the reconstruction is performed by minimizing the distance (loss or error function) between calculated values and experimental measurements for the same points. The calculation of discrete forward model has computational complexity. However, once the model matrix is calculated for a given detector array geometry and the size of the grid defined by users for specific FOV, the minimization takes relatively shorter time, especially when performed on GPU.

The 2D pressure wave function described in OA forward model (see chapter 1.3 for details) can be discretized as

$$p(r, t) \approx \frac{I(r, t + \Delta t) + I(r, t - \Delta t)}{2\Delta t} \quad (1.10)$$

for specific location ( $r$ ) on the grid for time ( $t$ ) where  $I(t)$  is

$$I(r, t) = \int_{L'} \frac{H(r')}{|r - r'|} dL'(t). \quad (1.11)$$

Then, the integral can be solved for  $M$  points on the curve  $L'(t)$  for positions  $r'_l$  as follows

$$I(r, t) = \frac{1}{2} \sum_{l=1}^{M-1} \left[ \frac{H(r'_l)}{|r - r'_l|} - \frac{H(r'_{l+1})}{r - r'_{l+1}} \right] d_{l,l+1} \quad (1.12)$$

or the same equation can be expressed as

$$I(r, t) \approx \frac{1}{2} \sum_{l=1}^{M-1} \frac{H(r'_l)}{|r - r'_l|} (d_{l-1,l} + d_{l,l+1}) \quad (1.13)$$

where  $d_{0,1} = d_{M,M+1} = 0$ . After the discretization, the pressure wave can be written for position  $r_i$  and time point  $t_j$  as

$$p(r_i, t_j) = \sum_{k=1}^N a_k^{ij} H(r'_k) \quad (1.14)$$

where  $N$  is the number of pixels in the image. The equation describes the linear combination of absorbed energy for spatial location ( $r_i$ ) defined by the pixels in the FOV. Then, the pressure wave equation is calculated for each transducer position on the detector array and for each discretization point on the mesh grid. The overall equation for the combination of

all discrete points and transducer elements can be expressed as

$$p = AH \quad (1.15)$$

where  $p$  is the theoretical pressure values,  $A$  is the model matrix and  $H$  is the reconstructed matrix. Then, the solution is achieved by minimizing the difference between measured values and theoretical values as

$$H_{solution} = \operatorname{argmin}_H \|p_{measured} - AH\|_2. \quad (1.16)$$

Different optimization techniques and distance functions can be employed for the minimization of above-defined function. Regularization parameters can be added to the equation for imposing image intensity priors.

#### 1.3.4 IMAGE ENHANCEMENT

Conventional post processing methods are applied to improve image quality in OA imaging. The image enhancement steps with conventional methods include fluence correction [71], adaptive histogram equalization [72], Frangi or vesseness filtering [73] and masking for background suppression [74]. The fluence correction maps fit a mathematical model (i.e., linear, or exponential) from the surface to the deeper structures of the object by assuming the light attenuation during the propagation. Then, the images are multiplied with this map to increase the contrast in deep structures (Fig. 1.5b). Histogram equalization stretches the contrast in the defined range to achieve better visualization (Fig. 1.5c). Frangi filter detects the vessel structures and boundaries of the tissues. The contrast from vessels and anatomical structures can be enhanced by applying Frangi filter after histogram equalization (Fig 1.5d). After the application of these steps output of the Frangi filter are combined with histogram equalized image (Fig. 1.5e). The segmentation map is applied on the combined image to suppress background which includes only noise from imaging medium (Fig. 1.5e). Fig. 1.5 visualizes the changes in every step applied on the OA images.

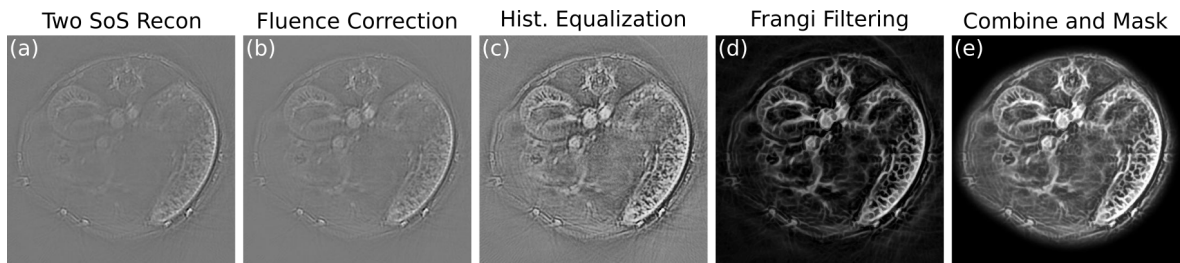


FIGURE 1.5: Image enhancement steps for optoacoustic (OA) images. (a) Image reconstruction with heterogenous backprojection algorithm by assigning different speed of sound (SoS) values for the background (water) and the mouse body. (b) Fluence correction using modified Bessel function approximation to account for light attenuation. (c) Adaptive histogram equalization. (d) Frangi filter to emphasize vessels. (e) Application of manual segmentation mask and the combination of histogram equalized image with Frangi filtered image.

In addition to conventional image enhancement methods, data-driven methods are also em-

ployed to improve image quality and temporal resolution in OA imaging. These methods can be grouped under four main categories, namely, removal of undersampling artifacts, segmentation for fluence correction, removal of limited view artifacts, and spatial resolution improvement. The following subsections focus on the presented papers in this thesis and already available methods in the literature for data-driven image enhancement methods in OA imaging.

***Removal of Undersampling Artifacts:*** The temporal resolution in OA imaging can be improved by reducing the number of channels in the detector array used for data acquisition. However, a reduced number of channels (increased temporal resolution) results in inferior image quality by causing reduced CNR, insufficient angular coverage, and undersampling artifacts (Fig. 1.6a). Data-driven methods were proposed to map the images reconstructed with channel undersampling to images with full sampling [47, 48]. At its core, CNN takes the spatially sparse sampled images as input. The network extracts the features in the images by passing them through the convolutional layers. The last layer of the network generates the same size of images as the input. After the last layer of the networks, the modified output images by layers are compared with the ground truth or target images. The distance between the output images of the networks and target images is calculated by loss functions (error metrics). The loss function is minimized iteratively using optimization methods and back-propagating the gradients of the losses. By following these steps, the layers of the networks are trained to adjust the weights of the features extracted from sparse sampling artifacts and anatomical structures. Finally, the trained network adjusts the weights to keep signals coming from tissue structures and to remove the artifacts from sparse sampling. The residual connections are also proposed to learn only the differences between input and target images by adding the input image to the output of the network before the calculation of loss [75]. We also performed a supervised training of CNN to remove sparse sampling artifacts in a semi-circle and multisegment detector array acquisitions. Fig. 1.6 shows examples of OA images with arc-shaped sparse sampling artifacts (Fig. 1.6a), predictions of CNN trained to remove artifacts (Fig. 1.6b), and target (ground truth) images used to train, validate, and test the network (Fig. 1.6c). The trained network can remove arc-shaped artifacts originating from undersampling and keep the artifacts originating from discontinuity of linear and concave parts of the multisegment array (thick arcs around the vessels). Also, the skin lines and vessel shapes are preserved by the network regardless of the thickness of the skin line and the size of the vessels. The dataset, trained network weights, training, and test scripts can be accessed with the accompanying paper presented in chapter 5 [70].

***Segmentation for Fluence Correction:*** Light attenuation or light fluence map of the imaged objects is not uniform in OA excitation while the material or tissue characteristics affect the amount of light that is absorbed and scattered. In addition, the amount of light energy reaching the structures decreases with the depth. The fluence correction helps to improve OA images qualitatively and quantitatively by accounting for light attenuation through the propagation path [76]. However, the application of fluence correction methods requires the

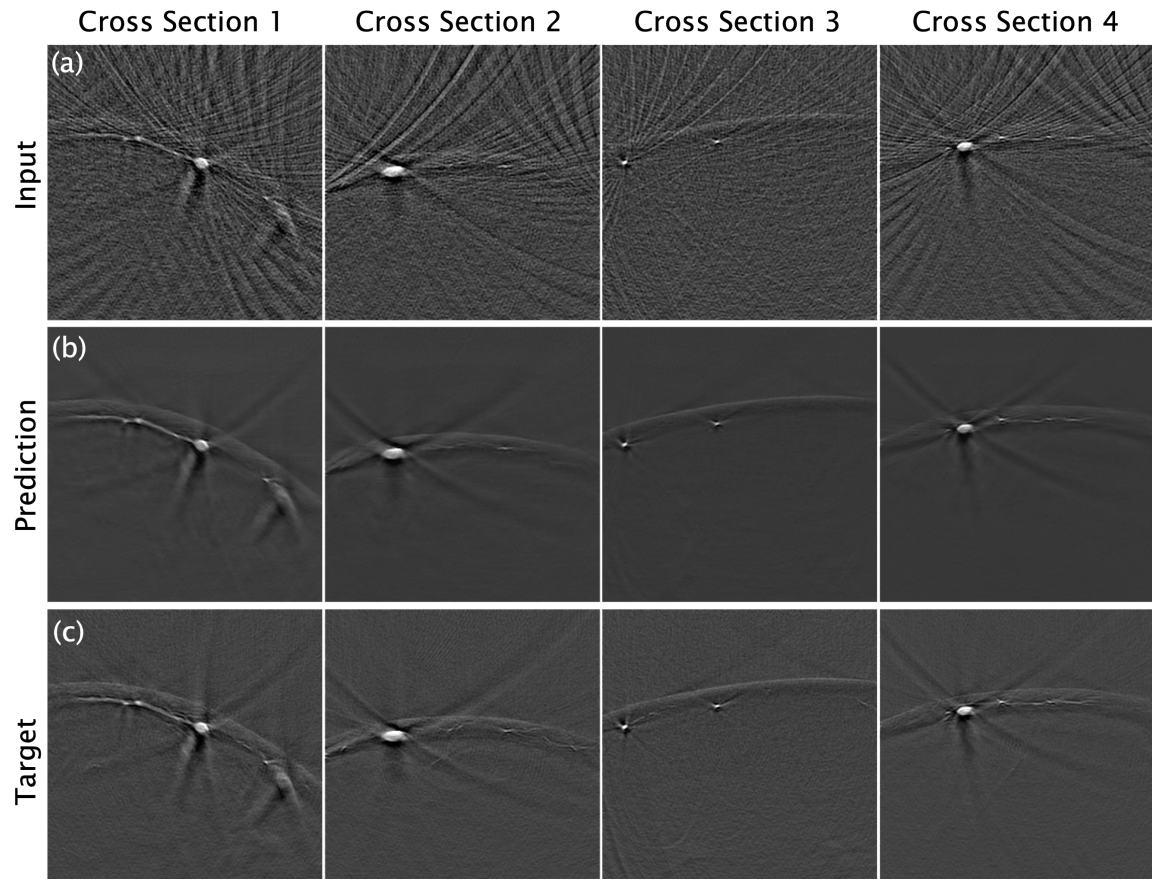


FIGURE 1.6: Removal of undersampling artifacts resulting from the sparsity of transducer elements with data-driven image enhancement algorithm. (a) Input images given to convolutional neural network (CNN) with sparse acquisition artifacts. (b) Output images (predictions) by CNN with removed sparse acquisition artifacts. (c) Target (ground truth) images acquired without sparsity in transducer elements.

segmentation of structures to assign tissue-specific characteristics. After the delineation of the tissue borders, specific attenuation values can be assigned to each region separately [71]. The segmentation can be performed by conventional methods such as the active contour method [77]. However, these methods require manual interaction or initialization points. With the wide use of data-driven methods in imaging, segmentation methods are also proposed to delineate the structures in biomedical images [78, 79]. In this thesis, we performed segmentation of mouse boundaries in preclinical hybrid OPUS images using CNNs and compared the results with manual and active contour segmentation methods (Fig. 1.7) [59]. Fig. 1.7a and 1.7d show manual annotation of mouse boundaries in OA and US images, respectively, which is user-dependent and time-demanding. active contour edge detection (ACED) is also performed by the circle contour initialization (Fig. 1.7b and 1.7e). Fig. 1.7c, and Fig. 1.7f show an example of mouse boundary segmentation using a CNN architecture which successfully delineates the mouse boundaries even in low contrast images. The vessel and the skin line segmentations are also performed in simulated OA images using data-driven methods [70].

After the segmentation of structures in OA images, fluence maps can be created using sev-

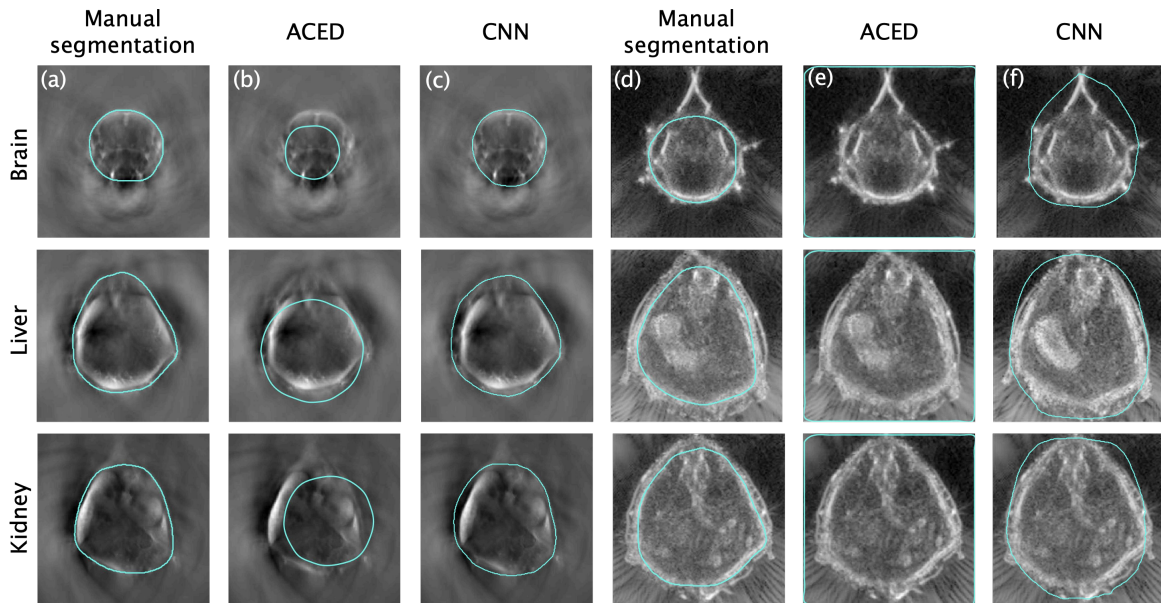


FIGURE 1.7: Segmentation of mouse boundaries using manual annotation, active contour edge detection (ACED), and convolutional neural network (CNN) for hybrid optoacoustic (OPUS) images. (a) Manual segmentation of mouse boundary in optoacoustic (OA) images. (b) ACED segmentation method for OA images. (c) CNN segmentation method for automated delineation of mouse boundaries in OA images. (d) Manual segmentation of mouse boundary in ultrasound (US) images. (e) ACED segmentation method for US images. (f) CNN segmentation method for automated delineation of mouse boundaries in US images. [80]

eral theoretical methods including linear and exponential decay [76]. The more sophisticated fluence maps require the segmentation of all organs and different tissue types which is a more challenging task than segmenting vessels and skin lines. However, generating manual annotations for each structure in biomedical images including OA imaging modality is a challenging or time-demanding task. The availability of such datasets might help to create more accurate segmentation results and fluence models for OA imaging.

**Removal of Limited View Artifacts:** OA is a tomographic imaging modality by its nature. Generated US waves by the OA effect (thermoelastic expansion of light absorbing materials) propagate in the 3D imaging medium in every direction. Detection of these US waves from different angles with wide angular coverage increases the image quality in OA imaging [81, 82]. However, full tomographic coverage of the tissues is not always possible due to physical constraints (tissue/organ is not accessible from every angle) and detector array design considerations (cost and design parameter requirements for different applications). The reduced angular coverage generates so-called limited view artifacts [83] which result in the elongation of round vessels (Fig. 1.8a). These limited view artifacts can be removed by supervised or unsupervised learning-based methods as proposed in the literature [49–54]. We also proposed to remove limited view artifacts using domain adaptation between simulated and experimental data in signal domain [84]. Since the limited view artifacts can be mitigated with simulations, the networks are trained to learn to remove these artifacts from the simulated images. However, the application of networks that are trained with simulated data on

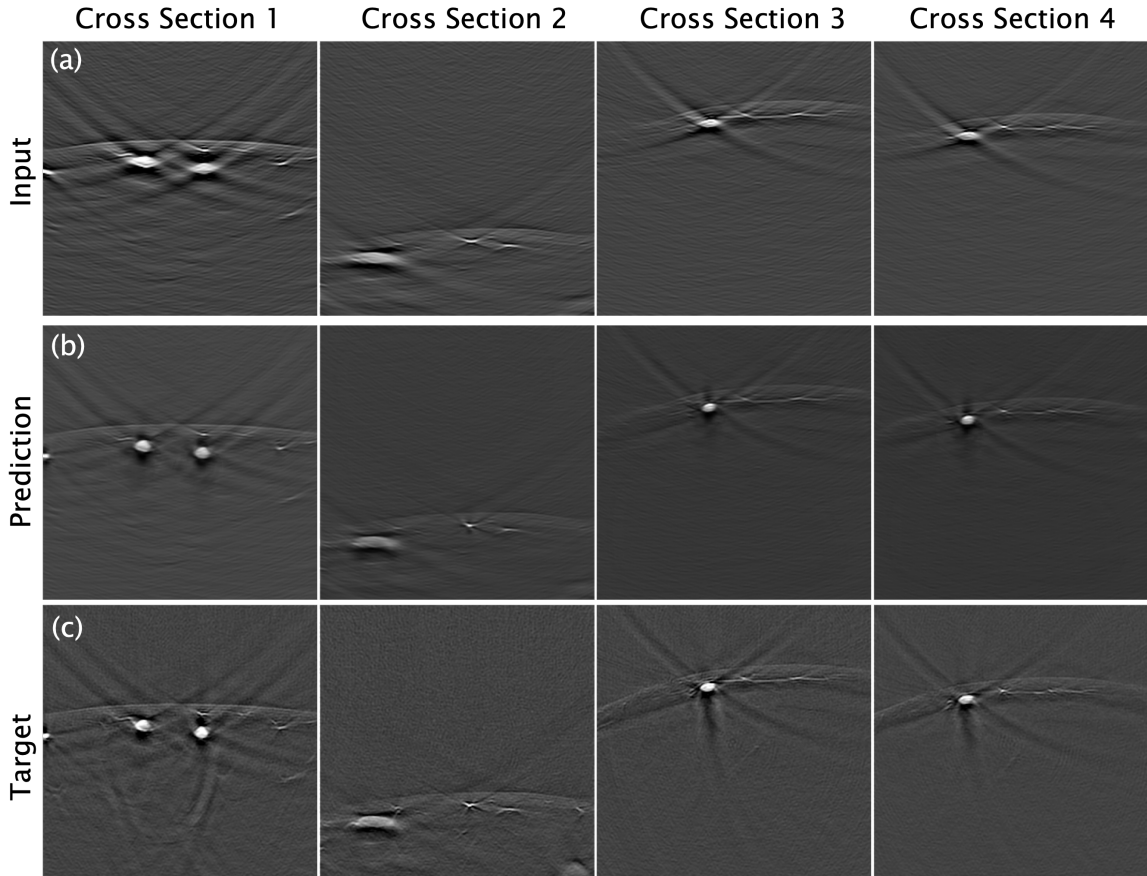


FIGURE 1.8: Removal of limited view artifacts resulting from reduced angular coverage of detector array by data-driven image enhancement algorithm. (a) Input images are given to convolutional neural network (CNN) with limited view artifacts. (b) Output images (predictions) by CNN with removed limited view artifacts. (c) Target (ground truth) images acquired with increased angular coverage.

the experimental data is not straightforward because the intensity distributions vary between these two types of datasets. Also, the simulations cannot replicate the same noise patterns originating from the background and electrical (intrinsic) noise of the system accurately [85]. In our proposed method, two different neural networks are trained to remove limited view artifacts. First, a style transfer network is used to reduce the gap between simulated and experimental data. Then, the second network which is trained on simulated data is applied to the experimental data to complete limited view projections in experimental data [84]. We also trained a network in a supervised method to remove limited view artifacts for linear array acquisitions from the central part of the multisegment detector array [70]. The input images that are acquired with a linear detector array clearly show the elongation of round vessel structures (Fig. 1.8a). The trained CNN with multisegment detector array acquisitions as ground truth images (Fig. 1.8c) can remove limited view artifacts and correct the vessel shapes (Fig. 1.8b).

***Spatial Resolution Improvement:*** Spatial resolution in OA imaging is limited by the diffraction limit of the acoustic waves. The US transducer arrays with higher bandwidths



can reach the diffraction limit and generate high-resolution OA images, but they suffer from the loss of FOV. Specifically, there is a trade-off between the bandwidth of the US transducer array and FOV that can be imaged with the same detector array. CNN architectures are proposed to enhance the bandwidth of transducers in the signal domain instead of the image domain for OA imaging [55]. We proposed a data-driven method to map low-resolution images acquired by the transducers with narrow bandwidths (Fig. 1.9a) to high-resolution images acquired by the wide bandwidth transducers (Fig. 1.9c). The CNN designed for the resolution improvement task receives low-resolution images as input and generates high-resolution images as the output of the network (Fig. 1.9b). Three different networks (SRCNN [86], UNet [78], and Attention UNet [87]) are trained and compared in performance for this specific task. The results from UNet are shown in Fig. 1.9b.

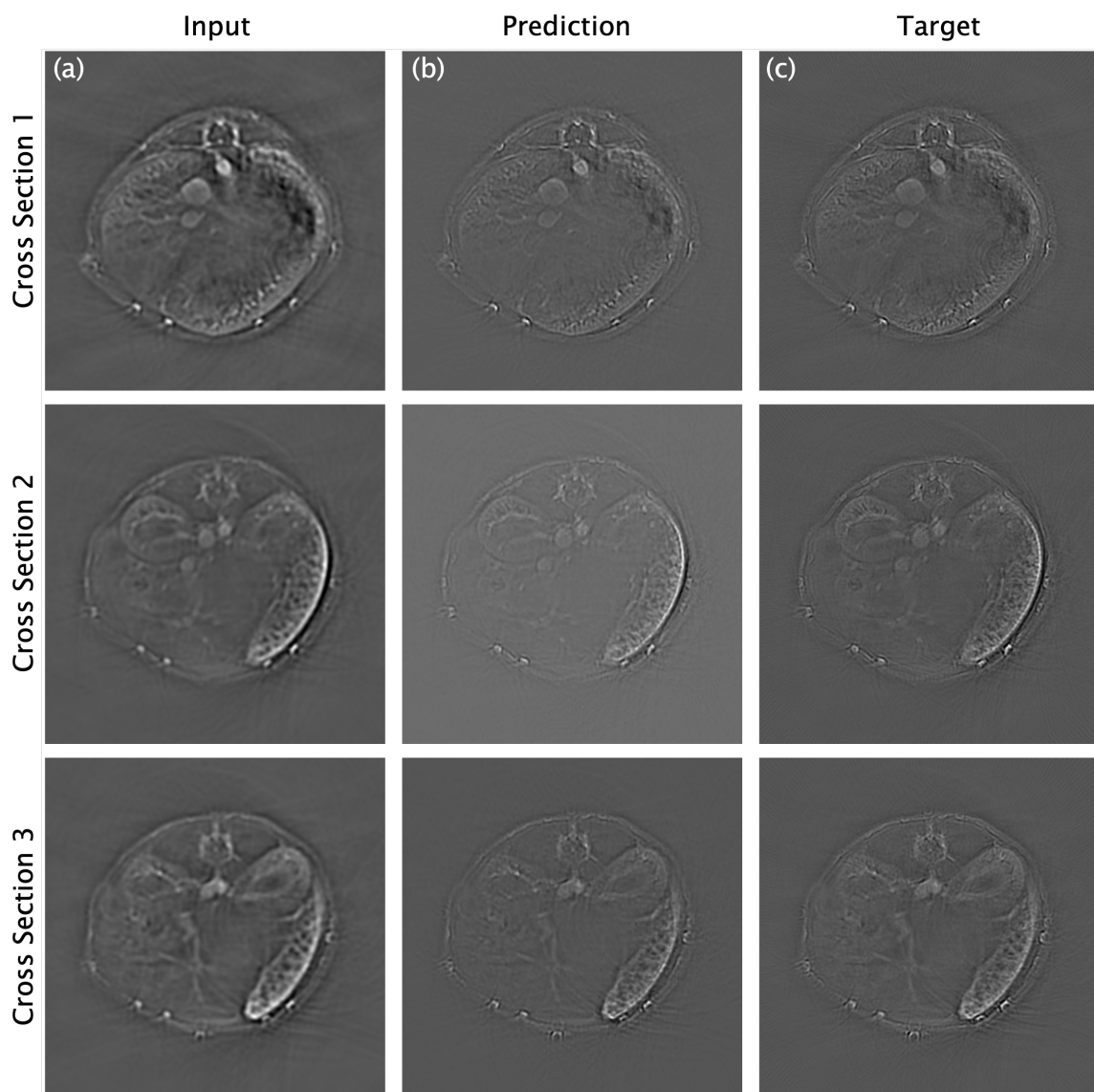


FIGURE 1.9: Low-resolution to high-resolution image enhancement using a convolutional neural network (CNN). (a) Input images given to CNN with low-resolution. (b) Output images (predictions) by CNN with increased resolution. (c) Target (ground truth) images with high spatial resolution.

### 1.3.5 SPECTRAL UNMIXING

MSOT is an efficient tool for spectral analysis that uses visible and NIR light excitation ranges with consecutive pulses of tunable lasers. Spectral unmixing is the method to identify the tissue content (concentrations of absorbers) using the varying absorption coefficients of chromophores at different illumination wavelengths [88]. The generated wavelength dependent pressure waves are expressed by the following equation

$$p(r, \lambda) = \Gamma H(r, \lambda) \quad (1.17)$$

where  $\Gamma$  stands for the Grüneisen constant [68] and  $H(r, \lambda)$  represents the absorbed energy at spatial location ( $r$ ) for the illumination wavelength ( $\lambda$ ).  $H(r, \lambda)$  can be expressed as the multiplication of light fluence with the absorption coefficient  $\mu(\lambda)$  and concentration  $C(r)$  of each absorber

$$H(r, \lambda) = \Phi(r, \lambda)[\mu_1(\lambda)C_1(r) + \mu_2(\lambda)C_2(r) + \dots + \mu_i(\lambda)C_i(r) + \dots + \mu_m(\lambda)C_m(r)] \quad (1.18)$$

where  $\Phi(r, \lambda)$  is the light fluence, that depends on spatial ( $r$ ) and wavelength ( $\lambda$ ) variables.  $\mu_i(\lambda)$  is the absorption coefficient of the  $i$ -th element dependent on the wavelength ( $\lambda$ ).  $C_i(r)$  is the spatially dependent ( $r$ ) concentration of the  $i$ -th element. The Grüneisen constant  $\Gamma$  can be neglected and light fluence is considered as constant for the fixed spatial location ( $r$ ) and wavelength ( $\lambda$ ) at specific pixel of the image for each absorber. Then, one linear equation can be written for each pixel as

$$H^k = \mu^k C^k \quad (1.19)$$

where  $H^k$  is the column vector containing pixel intensities for each illumination wavelength.  $\mu^k$  contains absorption coefficients for each absorber and wavelength.  $C^k$  is the unknown quantity of each absorber that needs to be calculated for spectral unmixing. For instance, Hb, HbO<sub>2</sub>, lipid, and melanin content of the different tissues can be extracted by solving the above equations for the concentration of each component by using the image intensities from MSOT.

## 1.4 ULTRASOUND IMAGING

US imaging is performed by the circle detector array in the scope of this thesis. The square US waves from single or multiple transducer elements are emitted to perform US imaging. The reflected waves are used for the reconstruction of reflection (pulse-echo) US images. Reflection US tomography is mentioned as RUCT in the remaining of this thesis. In addition, the transmitted signals through the imaged object are acquired at the opposite side of the transducer array to reconstruct transmission US images. TUCT imaging can be used for the generation of SoS or acoustic attenuation maps [74]. We only use and explore SoS imaging capabilities of TUCT in this thesis.

### 1.4.1 SIGNAL ACQUISITION

The emission of US waves and acquisition of reflected and/or transmitted waves are performed with the circle detector array. The circle detector array requires the STA [89] technique to complete full tomographic acquisition and to perform RUCT and SoS imaging [74]. The STA technique uses emission of signals with a single or group of elements to create an emission window. Then, reflected, or transmitted waves are acquired with a group of neighboring elements (for RUCT) or elements at the opposite side of the transducer array (for SoS imaging). The circle detector array has 512 transducer elements that can transmit US waves sequentially or simultaneously. The sequential transmission with each individual transducer element results in 512 transmission events to complete a full tomographic acquisition. Then, all the transducer elements can be used in reception mode simultaneously to receive reflected and transmitted waves at different locations. The transmitted US waveform is adjusted by assigning the specific delay for each transducer element by DAQ. The sampling frequency and the number of acquired samples are also adjusted by DAQ based on the required FOV and frame rate for specific applications.

### 1.4.2 REFLECTION ULTRASOUND COMPUTED TOMOGRAPHY

The image formation in RUCT is based on delay and sum (DAS) beamforming after the acquisition of signals using the STA method. The DAS method calculates the TOF between the emission of US wave by transducer and detection of the reflected signals with the receiving elements. The image (mesh grid) with the initial value of 0 is created before assigning the acquired values to each pixel. Then, the signal values are assigned to pixels by finding the corresponding time point on the signal based on the calculated TOF for each individual pixel and reception channel pair. After repeating this process for each pixel on the image, the DAS algorithm generates RUCT image from the single transmission event. Single transmission event results in low contrast RUCT images after the DAS beamforming. The STA includes transmission and reception of US signals from different angles with sequential transmission events to complete a full tomographic acquisition. The individual images resulting from single transmission events from different angles are then compounded (added up) to form final high-contrast RUCT images.

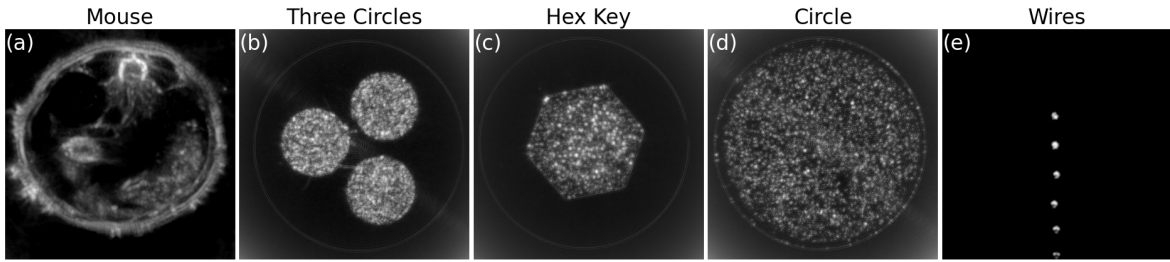


FIGURE 1.10: Example images from the reflection ultrasound computed tomography (RUCT). (a) Cross-sectional image of healthy nude mouse lower abdominal. (b) Three 6 mm circular insertions of 50  $\mu\text{m}$  reflective microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ) in 20 mm agarose phantom. (c) Hex key-shaped insertion of 50  $\mu\text{m}$  reflective microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ) in 20 mm agarose phantom. (d) 20 mm diameter circular insertion of 50  $\mu\text{m}$  reflective microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ). (e) Copper wires with 0.1 mm aligned perpendicular to the imaging plane.

Cross-sectional RUCT images of a healthy mouse and phantoms with different geometries are shown in Fig. 1.10. The raw signal data and DAS beamforming algorithms implemented in Python are available in the accompanying paper (see chapter 3 for details) [90]. The raw signal data from the healthy mouse is available with consecutive 1, 2, 3, and 4 transducer element transmissions to test *in vivo* imaging capabilities of the system. The same raw signals are also available for three circles, hex key, and circle phantoms to test wave focusing, FOV changes, and contrast. The raw signals from copper wires that are aligned perpendicular to the imaging plane are also included in the dataset to test the resolution changes based on different acquisition sequences or image reconstruction techniques.

**Consecutive Element Data Acquisition Methods:** Consecutive transducer elements are combined in transmit and receive modes to acquire RUCT images with the STA acquisition method. The goal of combining consecutive transducer elements was to mimic US detector arrays with a bigger element size and a reduced number of channels. This geometry increases the frame rate by reducing the required number of transmission events to complete full tomographic acquisition and data load resulting from the simultaneous acquisition in reception mode.

First, the effects of consecutive element transmission are investigated separately. The circle phantom is used to visualize the beam focusing for consecutive element transmission. Transmitting US waves with consecutive elements of the circle detector array results in focused beam directivity (Fig. 1.11). The directivity of the emitted US waves increases with the increasing number of combined consecutive transducer elements (Fig. 1.11a to 1.11d).

Then, the consecutive element transmission and reception were combined for *in vivo* acquisitions of the mouse lower abdominal region (Fig. 1.12). The same number of transducer elements (1, 2, 3, 4) are combined in the array for the emission and reception of US waves. This imaging sequence results in the preservation of contrast at the center of the image while causing the blurring at the peripheral regions (Fig. 1.12a to 1.12d). The consecutive element acquisition sequence can be preferred for applications with small FOV (i.e., 10 mm x 10 mm).

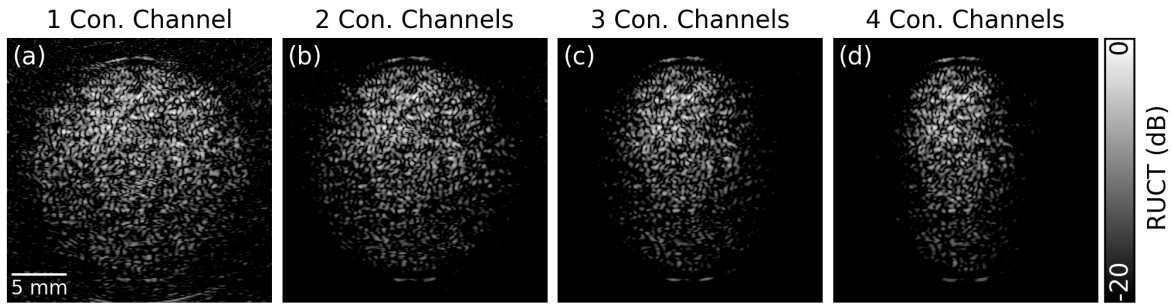


FIGURE 1.11: The effects of consecutive element transmission on beam focusing and field of view (FOV). (a) Transmission with 1 consecutive channel which is equal to synthetic transmit aperture (STA) with single elements. (b) Transmission with 2 consecutive transducer elements simultaneously. (c) Transmission with 3 consecutive transducer elements simultaneously. (d) Transmission with 4 consecutive transducer elements simultaneously.

**Sparse Data Acquisition Methods:** Subsampling of emission and reception channels were performed uniformly and randomly in RUCT imaging. Subsampling in transducer elements corresponds to skipping a fixed number of emission and reception channels of the detector array. This imaging sequence results in an increased frame rate by reducing the number of transmission events and the transferred data by decreasing the number of reception channels. However, an increased frame rate comes with the cost of reduced contrast and increased noise levels originating from reduced angular coverage in transmission events and decreased spatial sampling in reception.

First, uniform sparsity for transmission and reception is tested *in vivo* (Fig. 1.13). Uniform sparsity means skipping a fixed number of transducer elements that are distributed uniformly among emitting and receiving elements. For example, uniform sparsity 1 corresponds to using every element among emitting and receiving channels which is equal to full tomographic acquisition (Fig. 1.13a). Similarly, uniform sparsity 2 means performing the acquisition by emitting and receiving with every second element among the channels (Fig. 1.13b). The increased noise levels and reduced contrast are validated qualitatively and quantitatively.

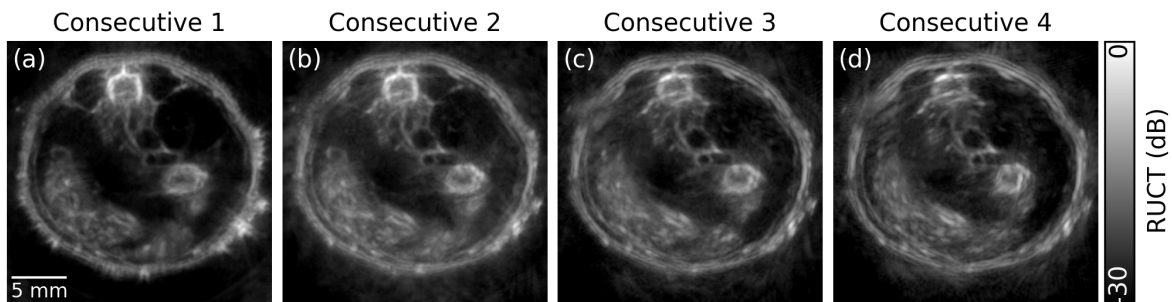


FIGURE 1.12: Combined effects of consecutive element transmission and reception were assessed *in vivo*. (a) *In vivo* mouse cross-section without consecutive element transmission or reception (individual element transmission and reception). (b) *In vivo* mouse cross-section with 2 consecutive element transmission and reception. (c) *In vivo* mouse cross-section with 3 consecutive element transmission and reception. (d) *In vivo* mouse cross-section with 4 consecutive element transmission and reception.

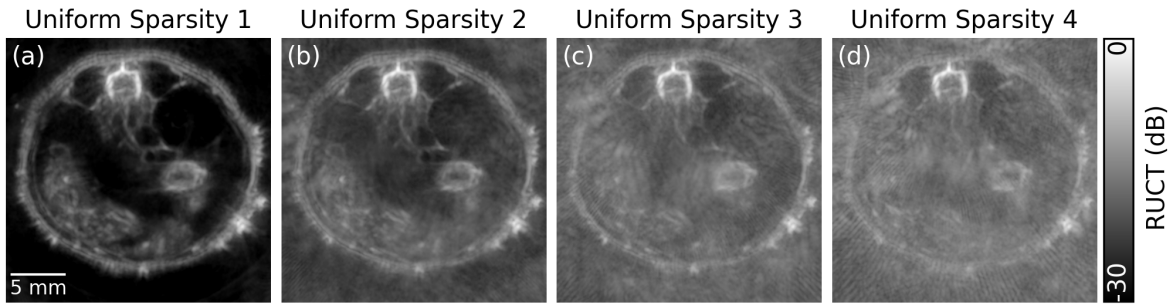


FIGURE 1.13: Combined effects of uniform sparsity in transmission and reception were assessed *in vivo*. (a) *In vivo* mouse cross-section without uniform sparsity in transmission or reception (all individual elements are used in transmission and reception) (b) *In vivo* mouse cross-section with uniform sparsity level of 2 in transmission and reception. (c) *In vivo* mouse cross-section with uniform sparsity level of 3 in transmission and reception. (d) *In vivo* mouse cross-section with uniform sparsity level of 4 in transmission and reception.

Then, random sparsity in transmission and reception is tested *in vivo* (Fig. 1.14). Random sparsity means undersampling of the transducer elements randomly for emitting and receiving channels by a fixed sparsity factor. For example, random sparsity 1 corresponds to using every element among emitting and receiving channels which is equal to full tomographic acquisition (Fig. 1.14a). Random sparsity 2 means performing the acquisition by emitting and receiving with half of the transducer elements in full acquisition (Fig. 1.14b). The increased noise levels and reduced contrast are validated qualitatively and quantitatively.

The effects of consecutive element and sparse data acquisitions are quantified and discussed in chapter 3 in the accompanying paper [90]. The performance analysis includes RUCT image quality validation based on contrast and resolution quantification with phantom images and *in vivo* mouse cross-sections.

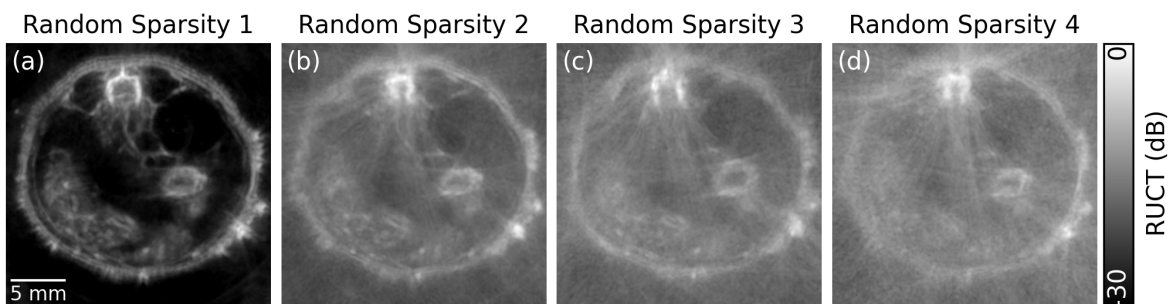


FIGURE 1.14: Combined effects of random sparsity in transmission and reception were assessed *in vivo*. (a) *In vivo* mouse cross-section without random sparsity in transmission or reception (all individual elements are used in transmission and reception) (b) *In vivo* mouse cross-section with random sparsity level of 2 in transmission and reception. (c) *In vivo* mouse cross-section with random sparsity level of 3 in transmission and reception. (d) *In vivo* mouse cross-section with random sparsity level of 4 in transmission and reception.

### 1.4.3 SPEED OF SOUND IMAGING

TUCT is based on the transmission of US waves from one side of the detector array and the collection of the transmitted waves through the object on the opposite side. In the scope of this thesis, TUCT is implemented with the circle detector array that is used to reconstruct SoS images of phantoms and mouse cross-sections. The STA imaging sequence with individual elements is performed for the acquisition of US signals (see chapter 1.4.1 for details). The consecutive element and sparse data acquisition methods are not evaluated for SoS imaging modality. Three different algorithms are tested for image reconstruction in SoS imaging, namely, straight ray, Bezier curves, and full wave inversion (FWI). The signals from 171 receiving transducer elements at the opposite side of the detector array are considered for each transmitting element for three different reconstruction techniques.

***Straight ray:*** Straight ray reconstruction algorithm assumes a linear path between emitting and receiving transducer elements which is an oversimplified assumption of the US wave propagation [91] (Fig. 1.15a). The shortest linear paths are simulated between each emitting and receiving pair at the opposite side of the transducer array based on the predefined initial SoS in the propagation medium. After the paths are simulated between the pairs, the cost function between simulated TOF and measurements is minimized using convex optimization methods. The challenging part for calculating the TOF between the transducer pairs is the detection of US wave arrival at the receiving elements. TOF picker algorithms that are already available in the literature are used to estimate US wave arrival [92, 93]. Median filtering and reciprocal pair comparison methods are used to remove outliers from TOF picker algorithm [94].

***Bezier curves:*** Bezier curve reconstruction assumes bent rays between emitting and receiving transducer element pairs [74, 95]. This method uses a similar approach with straight rays to calculate SoS in the medium. The path between transducer pairs is discretized and the TOF is calculated by assuming that US waves follow a curved path while traveling in the medium. The cost function between measured and calculated values is minimized by optimizing for SoS. The same TOF picker algorithm and postprocessing methods to remove outliers in the straight ray reconstruction are used for this method. Bezier curve assumption between transducer pairs is more accurate than the straight ray approach but it is less accurate than modeling the complete wave propagation (i.e., FWI) (Fig. 1.15).

***Full wave inversion:*** FWI method simulates the wave propagation in the imaging medium [96, 97]. Multiple paths are simulated between each emitter and receiver pair using GPU. The initial waveforms are extracted from the water measurements without placing any object between the transducer pairs. In addition, initial SoS values are estimated based on the arrival times. The waveforms and estimated SoS values are convolved for each simulated path. Then, the distance between calculated values from simulations and experimental values from measurements is minimized similar to the other SoS image reconstruction methods. A gradient descent algorithm is used to iteratively minimize the mean squared error between

simulations and measurements by optimizing for SoS values.

The complexity of the US propagation model increases from straight ray to FWI, while the contrast and resolution of the SoS predictions also increase (Fig. 1.15). The increased image quality (contrast and resolution) enables accurate quantification of SoS changes *in vivo*. However, increased complexity in reconstruction algorithms and accurate modeling bring the accompanying computing costs. FWI algorithm can reconstruct one cross-section in 5 minutes using GPU. While the increased reconstruction time is not a problem for offline processing, the trade-off between temporal resolution and quantification accuracy should be evaluated for desired real-time applications. The results are validated *in vivo* by reconstructing the same mouse cross-section with three different proposed algorithms (Fig. 1.15).

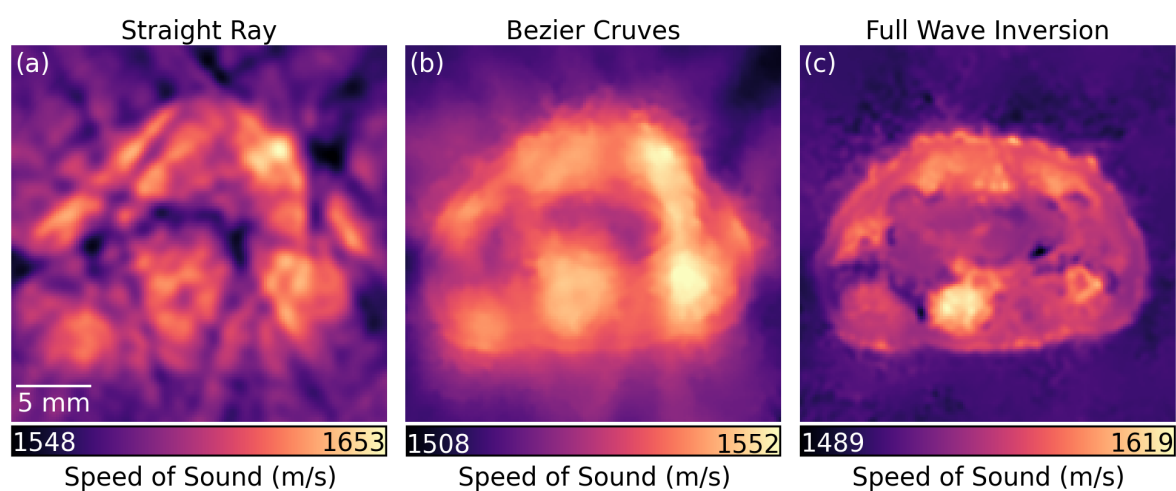


FIGURE 1.15: Speed of sound (SoS) maps of one mouse cross-section *in vivo* using different image reconstruction techniques. (a) Straight ray reconstruction method. (b) Bezier curves reconstruction method. (c) Full wave inversion (FWI) reconstruction method.



## 1.5 HYBRID OPTOACOUSTIC ULTRASOUND

The different contrasts provided by OA and US imaging can be complemented using the simultaneous acquisition of both modalities. The missing information on the anatomical properties of the imaged object in OA images can be obtained by US imaging. The molecular and functional properties obtained by OA imaging can be overlaid on US images for the quantification of localized responses. To this end, a hybrid OPUS system is implemented in this thesis for preclinical imaging using the circle detector array. As described in the OA and US imaging chapters (see chapters 1.3 and 1.4 for details), there are different considerations for the implementation of these systems. Specifically, different constraints are imposed by both imaging modalities on data transfer rate, probe designs, and image reconstruction algorithms. So, the combination of these two modalities for simultaneous acquisition is not straightforward. The image enhancement methods (see chapter 1.3.4 for details) for OA imaging and image acquisition sequences (see chapter 1.4.2 for details) for US imaging are proposed to increase image quality and acquisition speed.

### 1.5.1 SIGNAL ACQUISITION

The signal acquisition in OPUS imaging can be performed sequentially using both OA and US modalities or a combination in one signal acquisition instance. Sequential acquisition can be performed separately as defined by OA (see chapter 1.3.1 for details) and US (see chapter 1.4.1 for details) imaging. When the acquisition is performed in the same imaging instance/trigger, the delays of US transmission, laser trigger, US wave reception, and OA signal reception should be synchronized. In our simultaneous acquisitions, we triggered US transmission and laser pulses in a close time point window. So, the OA signals can arrive before the reflected US waves to avoid interference as light propagates faster than US waves in the imaging medium.

### 1.5.2 IMAGE RECONSTRUCTION

A hybrid OPUS imaging system requires the reconstruction of OA and US images separately as the governing physics behind each modality is different. The image reconstruction for OA images is performed by backprojection or model-based algorithms as explained in chapter 1.3.3. RUCT images acquired by the STA method are reconstructed by the DAS algorithm as explained in chapter 1.4.2. The OA image reconstruction is faster than RUCT images since the latter requires the compounding of multiple frames. Still, the RUCT image reconstruction can also be performed in real-time using GPUs. The final images after reconstructions can be overlaid without the need for any registration algorithms (Fig. 1.16).

Fig. 1.16 shows co-registered hybrid OPUS images from different cross-sections of a healthy mouse that were acquired simultaneously using the circle detector array. The applications of hybrid OPUS imaging can be augmented for the assessment and characterization of several diseases. For example, localized accumulation of lipid in the tissues can be obtained from OA images after spectral unmixing (see chapter 1.3.5 for details) and overlaid on US images (see

chapter 7 for details). The vessel structures and oxygenation in tumor areas can be investigated by OA images with the help of anatomical information from US images (see chapter 6 for details).

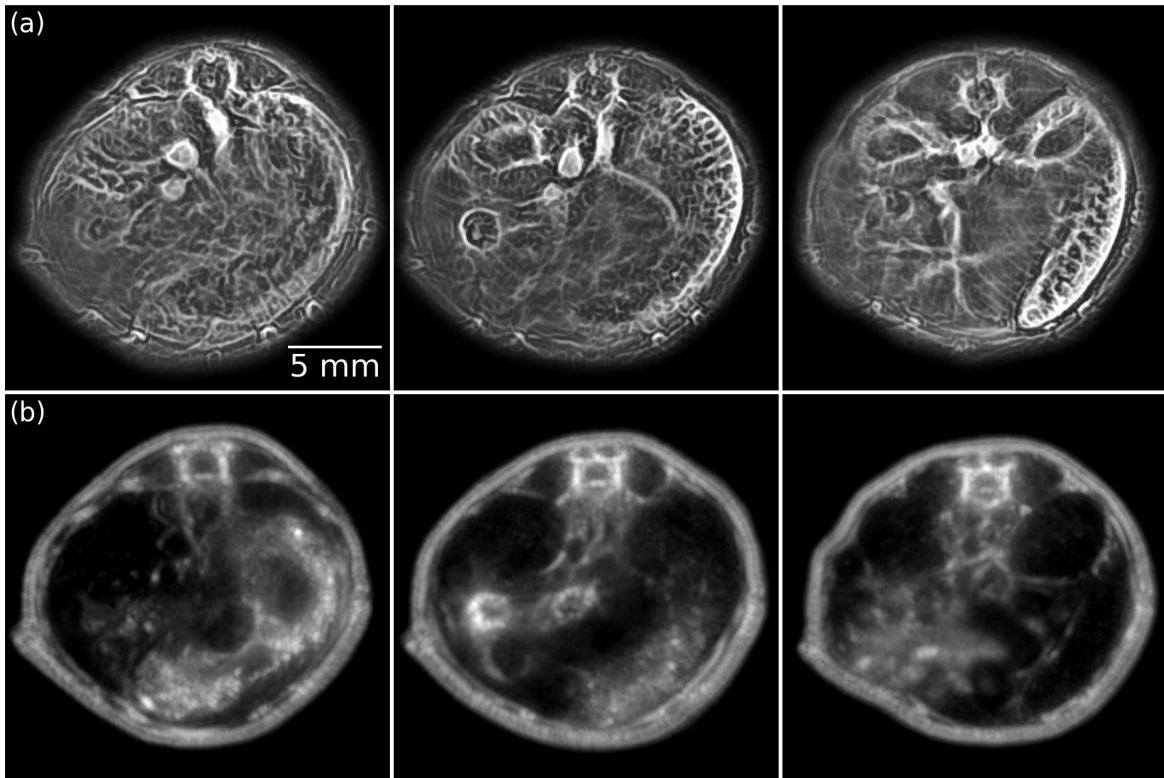


FIGURE 1.16: Optoacoustic ultrasound (OPUS) images of mouse cross-sections *in vivo*. (a) Optoacoustic (OA) images from different mouse cross-sections *in vivo*. (b) Corresponding reflection ultrasound computed tomography (RUCT) images from the same mouse cross-sections *in vivo*.

## 1.6 TRANSMISSION REFLECTION OPTOACOUSTIC ULTRASOUND

TUCT can provide complementary information on top of the hybrid OPUS imaging systems. SoS and acoustic attenuation maps can be generated using the signals acquired at the opposite side of the transmitting element by TUCT. To this end, we devised the TROPUS preclinical imaging system that combines OA imaging with both reflection and transmission US [74]. Only SoS imaging capabilities of TUCT are used, validated, and quantified in this thesis. The simultaneous or consecutive acquisition of OA and US signals using TROPUS system results in co-registered images from three modalities (OA, RUCT and SoS). *In vivo* mouse imaging experiments revealed fine details on vascularization, tissue reflectivity, density, and stiffness. The constraints imposed by the TUCT on the data acquisition are similar to RUCT system as they use the same US imaging sequences. However, their image reconstruction algorithms differ from each other as they quantify different physical phenomena.

### 1.6.1 SIGNAL ACQUISITION

TROPUS imaging is performed using only the circle detector array as it requires collection of transmitted US waves at the opposite side of the imaged object. The signal acquisition for OA imaging is performed as described in chapter 1.3.1 using all transducer elements on the detector array. US imaging is performed by the STA method (see chapter 1.4.1 and 1.5.1 for details) using single or group of elements for transmission events. The signals reflected from and transmitted through the imaged object are recorded at two opposite sides of the transducer array. Effects of consecutive element or sparse acquisitions are not quantified for SoS images in the scope of this thesis.

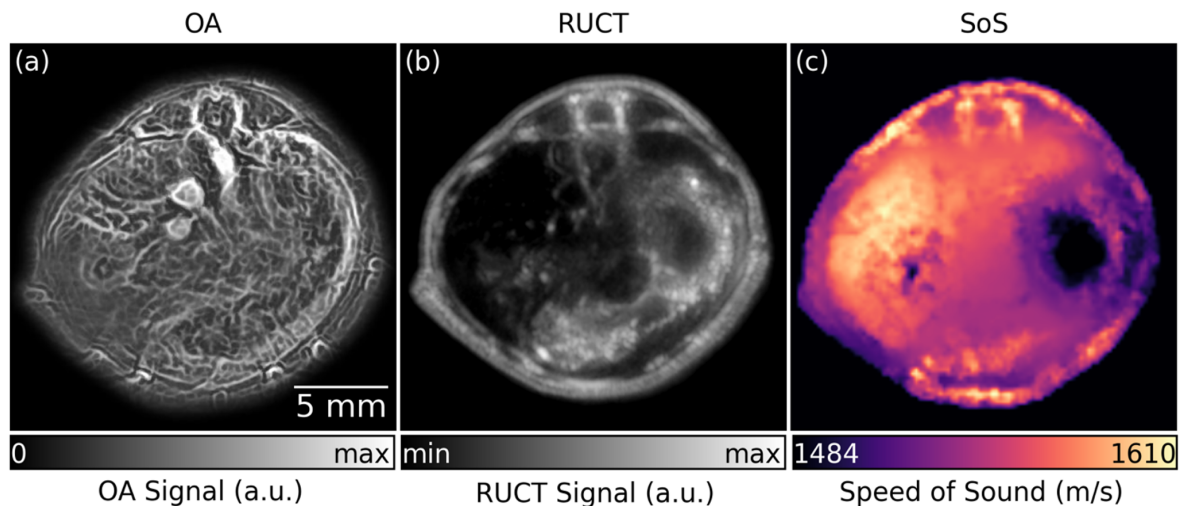


FIGURE 1.17: Simultaneous acquisition of one co-registered mouse cross-section with transmission-reflection optoacoustic ultrasound (TROPUS) system *in vivo*. (a) Optoacoustic (OA) image from a cross-section of the mouse *in vivo*. (b) Reflection ultrasound computed tomography (RUCT) image from the same cross-section of the mouse *in vivo*. (c) Full wave inversion (FWI) reconstruction of the speed of sound (SoS) image from the same cross-section of the mouse *in vivo*.

### 1.6.2 IMAGE RECONSTRUCTION

TROPUS system requires the reconstruction of images from each modality separately. OA images are reconstructed using backprojection or model-based algorithms as explained in chapter 1.3.3. RUCT images are reconstructed using the DAS algorithm as explained in 1.4.2. SoS maps are created by reconstruction algorithms that are explained in chapter 1.4.3. The resulting co-registered images from three different modalities (OA, RUCT, SoS) are shown in Fig. 1.17. Specifically, FWI reconstruction method is used for the reconstruction of SoS image (Fig. 1.17c) as it generates more accurate maps compared to the other methods.

## BIBLIOGRAPHY

- [1] Justine Wallyn et al. “Biomedical Imaging: Principles, Technologies, Clinical Aspects, Contrast Agents, Limitations and Future Trends in Nanomedicines”. In: *Pharmaceutical Research* 36 (2019), pp. 1–31.
- [2] Grit Gesine Ruth Schramek et al. “Imaging in anatomy: a comparison of imaging techniques in embalmed human cadavers”. In: *BMC medical education* 13 (2013), pp. 1–7.
- [3] Carl A Carlsson and Gudrun Alm Alm Carlsson. *Basic physics of X-ray imaging*. Linköping University Electronic Press, 1973.
- [4] Philip J Withers et al. “X-ray computed tomography”. In: *Nature Reviews Methods Primers* 1.1 (2021), p. 18.
- [5] Anita Kovács et al. “The sensitivity and specificity of chest CT in the diagnosis of COVID-19”. In: *European Radiology* 31 (2021), pp. 2819–2824.
- [6] Benjamin M Yeh et al. “Opportunities for new CT contrast agents to maximize the diagnostic potential of emerging spectral CT technologies”. In: *Advanced drug delivery reviews* 113 (2017), pp. 201–222.
- [7] Vijay PB Grover et al. “Magnetic resonance imaging: principles and techniques: lessons for clinicians”. In: *Journal of clinical and experimental hepatology* 5.3 (2015), pp. 246–255.
- [8] Riham H El Khouli et al. “Relationship of temporal resolution to diagnostic performance for dynamic contrast enhanced MRI of the breast”. In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 30.5 (2009), pp. 999–1004.
- [9] Arman Rahmim et al. “Dynamic whole-body PET imaging: principles, potentials and applications”. In: *European journal of nuclear medicine and molecular imaging* 46 (2019), pp. 501–518.
- [10] Simon R Cherry et al. “Total-body PET: maximizing sensitivity to create new opportunities for clinical research and patient care”. In: *Journal of Nuclear Medicine* 59.1 (2018), pp. 3–12.
- [11] Joanne C Wen et al. “Radiation-related cancer risk associated with surveillance imaging for metastasis from choroidal melanoma”. In: *JAMA ophthalmology* 131.1 (2013), pp. 56–61.
- [12] European Society of Radiology (ESR et al. “Patient safety in medical imaging: A joint paper of the European Society of Radiology (ESR) and the European Federation of Radiographer Societies (EFRS)”. In: *Radiography* 25.2 (2019), e26–e38.

- [13] Eugene Lin and Adam Alessio. “What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT?” In: *Journal of cardiovascular computed tomography* 3.6 (2009), pp. 403–408.
- [14] Jeffrey Tsao and Sebastian Kozerke. “MRI temporal acceleration techniques”. In: *Journal of Magnetic Resonance Imaging* 36.3 (2012), pp. 543–560.
- [15] Hairong Zhang et al. “Penetration depth of photons in biological tissues from hyperspectral imaging in shortwave infrared in transmission and reflection geometries”. In: *Journal of biomedical optics* 21.12 (2016), pp. 126006–126006.
- [16] Christian B Scheele et al. “Accuracy of a non-invasive CT-based measuring technique for cement penetration depth in human tibial UKA”. In: *BMC Medical Imaging* 19 (2019), pp. 1–8.
- [17] Jun Xia, Junjie Yao, and Lihong V Wang. “Photoacoustic tomography: principles and advances”. In: *Progress in electromagnetics research* 147 (2014), pp. 1–22.
- [18] Vasilis Ntziachristos and Daniel Razansky. “Molecular imaging by means of multispectral optoacoustic tomography (MSOT)”. In: *Chemical reviews* 110.5 (2010), pp. 2783–2794.
- [19] Volker Neuschmelting et al. “Performance of a multispectral optoacoustic tomography (MSOT) system equipped with 2D vs. 3D handheld probes for potential clinical translation”. In: *Photoacoustics* 4.1 (2016), pp. 1–10.
- [20] Lacey R McNally et al. “Current and Emerging Clinical Applications of Multispectral Optoacoustic Tomography (MSOT) in Oncology Clinical Uses of Multispectral Optoacoustic Tomography”. In: *Clinical cancer research* 22.14 (2016), pp. 3432–3439.
- [21] Michal R Tomaszewski et al. “Oxygen enhanced optoacoustic tomography (OE-OT) reveals vascular dynamics in murine models of prostate cancer”. In: *Theranostics* 7.11 (2017), p. 2900.
- [22] Emma Brown, Joanna Brunker, and Sarah E Bohndiek. “Photoacoustic imaging as a tool to probe the tumour microenvironment”. In: *Disease models & mechanisms* 12.7 (2019), p. dmm039636.
- [23] Sarah E Bohndiek et al. “Photoacoustic tomography detects early vessel regression and normalization during ovarian tumor response to the antiangiogenic therapy trebananib”. In: *Journal of Nuclear Medicine* 56.12 (2015), pp. 1942–1947.
- [24] Çağla Özsoy et al. “Ultrafast four-dimensional imaging of cardiac mechanical wave propagation with sparse optoacoustic sensing”. In: *Proceedings of the National Academy of Sciences* 118.45 (2021), e2103979118.
- [25] Nicolas Beziere et al. “Dynamic imaging of PEGylated indocyanine green (ICG) liposomes within the tumor microenvironment using multi-spectral optoacoustic tomography (MSOT)”. In: *Biomaterials* 37 (2015), pp. 415–424.
- [26] Jing Lv et al. “Quantitative functional evaluation of liver fibrosis in mice with dynamic contrast-enhanced photoacoustic imaging”. In: *Radiology* 300.1 (2021), pp. 89–97.

- [27] Sven Gottschalk et al. “Rapid volumetric optoacoustic imaging of neural dynamics across the mouse brain”. In: *Nature biomedical engineering* 3.5 (2019), pp. 392–401.
- [28] X Luís Dean-Ben and Daniel Razansky. “Localization optoacoustic tomography”. In: *Light: Science & Applications* 7.4 (2018), pp. 18004–18004.
- [29] Xosé Luís Deán-Ben et al. “Noninvasive three-dimensional optoacoustic localization microangiography of deep tissues”. In: *arXiv preprint arXiv:2007.00372* (2020).
- [30] Pengfei Zhang et al. “In vivo superresolution photoacoustic computed tomography by localization of single dyed droplets”. In: *Light: Science & Applications* 8.1 (2019), p. 36.
- [31] Paul Wrede et al. “Real-time 3D optoacoustic tracking of cell-sized magnetic micro-robots circulating in the mouse brain vasculature”. In: *Science Advances* 8.19 (2022), eabm9132.
- [32] Xosé Luís Deán-Ben and Daniel Razansky. “Optoacoustic imaging of the skin”. In: *Experimental Dermatology* 30.11 (2021), pp. 1598–1609.
- [33] Johannes Rebling et al. “Long-term imaging of wound angiogenesis with large scale optoacoustic microscopy”. In: *Advanced Science* 8.13 (2021), p. 2004226.
- [34] Ferdinand Knieling et al. “Multispectral optoacoustic tomography for assessment of Crohn’s disease activity”. In: *New England Journal of Medicine* 376.13 (2017), pp. 1292–1294.
- [35] Maximilian J Waldner et al. “Multispectral optoacoustic tomography in Crohn’s disease: noninvasive imaging of disease activity”. In: *Gastroenterology* 151.2 (2016), pp. 238–240.
- [36] Ivana Ivankovic et al. “Real-time volumetric assessment of the human carotid artery: handheld multispectral optoacoustic tomography”. In: *Radiology* 291.1 (2019), pp. 45–50.
- [37] Angelos Karlas et al. “Multispectral optoacoustic tomography of lipid and hemoglobin contrast in human carotid atherosclerosis”. In: *Photoacoustics* 23 (2021), p. 100283.
- [38] Shuai Na et al. “Massively parallel functional photoacoustic computed tomography of the human brain”. In: *Nature Biomedical Engineering* 6.5 (2022), pp. 584–592.
- [39] Gael Diot et al. “Multispectral Optoacoustic Tomography (Msot) of Human Breast Cancermsot Signatures of Human Breast Cancer”. In: *Clinical Cancer Research* 23.22 (2017), pp. 6912–6922.
- [40] Srirang Manohar et al. “The Twente Photoacoustic Mammoscope: system overview and performance”. In: *Physics in Medicine & Biology* 50.11 (2005), p. 2543.
- [41] Sjoukje M Schoustra et al. “Twente Photoacoustic Mammoscope 2: system overview and three-dimensional vascular network images in healthy breasts”. In: *Journal of biomedical optics* 24.12 (2019), pp. 121909–121909.
- [42] Seungwan Jeon et al. “Real-time delay-multiply-and-sum beamforming with coherence factor for in vivo clinical photoacoustic imaging of humans”. In: *Photoacoustics* 15 (2019), p. 100136.

- [43] X Luís Dean-Ben et al. “Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography”. In: *IEEE transactions on medical imaging* 31.10 (2012), pp. 1922–1928.
- [44] Aki Pulkkinen et al. “A Bayesian approach to spectral quantitative photoacoustic tomography”. In: *Inverse Problems* 30.6 (2014), p. 065012.
- [45] Ali Özbek, Xosé Luís Deán-Ben, and Daniel Razansky. “Optoacoustic imaging at kilohertz volumetric frame rates”. In: *Optica* 5.7 (2018), pp. 857–863.
- [46] Simon Arridge et al. “Accelerated high-resolution photoacoustic tomography via compressed sensing”. In: *Physics in Medicine & Biology* 61.24 (2016), p. 8908.
- [47] Stephan Antholzer, Markus Haltmeier, and Johannes Schwab. “Deep learning for photoacoustic tomography from sparse data”. In: *Inverse problems in science and engineering* 27.7 (2019), pp. 987–1005.
- [48] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky. “Deep learning optoacoustic tomography with sparse data”. In: *Nature Machine Intelligence* 1.10 (2019), pp. 453–460.
- [49] Steven Guan et al. “Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning”. In: *Scientific reports* 10.1 (2020), p. 8510.
- [50] Hengrong Lan, Jiali Gong, and Fei Gao. “Deep learning adapted acceleration for limited-view photoacoustic image reconstruction”. In: *Optics Letters* 47.7 (2022), pp. 1911–1914.
- [51] Dominik Waibel et al. “Reconstruction of initial pressure from limited view photoacoustic images using deep learning”. In: *Photons Plus Ultrasound: Imaging and Sensing 2018*. Vol. 10494. SPIE. 2018, pp. 196–203.
- [52] Andreas Hauptmann et al. “Model-based learning for accelerated, limited-view 3-D photoacoustic tomography”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1382–1393.
- [53] Tong Lu et al. “LV-GAN: A deep learning approach for limited-view optoacoustic imaging based on hybrid datasets”. In: *Journal of biophotonics* 14.2 (2021), e202000325.
- [54] Neda Davoudi et al. “Deep learning of image-and time-domain data enhances the visibility of structures in optoacoustic tomography”. In: *Optics letters* 46.13 (2021), pp. 3029–3032.
- [55] Navchetan Awasthi et al. “Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 67.12 (2020), pp. 2660–2673.
- [56] Fangyan Liu et al. “Dictionary learning sparse-sampling reconstruction method for in-vivo 3D photoacoustic computed tomography”. In: *Biomedical Optics Express* 10.4 (2019), pp. 1660–1677.



- [57] Parastoo Farnia et al. “Dictionary learning technique enhances signal in LED-based photoacoustic imaging”. In: *Biomedical optics express* 11.5 (2020), pp. 2533–2547.
- [58] Melanie Schellenberg et al. “Semantic segmentation of multispectral photoacoustic images using deep learning”. In: *Photoacoustics* 26 (2022), p. 100341.
- [59] Berkan Lafci et al. “Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 68.3 (2020), pp. 688–696.
- [60] Janek Gröhl et al. “Learned spectral decoloring enables photoacoustic oximetry”. In: *Scientific reports* 11.1 (2021), p. 6565.
- [61] Ivan Olefir et al. “Deep learning-based spectral unmixing for optoacoustic imaging of tissue oxygen saturation”. In: *IEEE transactions on medical imaging* 39.11 (2020), pp. 3643–3654.
- [62] Yunhao Zhu et al. “Light emitting diodes based photoacoustic imaging and potential clinical applications”. In: *Scientific reports* 8.1 (2018), pp. 1–12.
- [63] Wenfeng Xia et al. “Handheld real-time LED-based photoacoustic and ultrasound imaging system for accurate visualization of clinical metal needles and superficial vasculature to guide minimally invasive procedures”. In: *Sensors* 18.5 (2018), p. 1394.
- [64] Yunhao Zhu et al. “Towards clinical translation of LED-based photoacoustic imaging: a review”. In: *Sensors* 20.9 (2020), p. 2484.
- [65] Yuehang Wang et al. “Review of methods to improve the performance of linear array-based photoacoustic tomography”. In: *Journal of Innovative Optical Health Sciences* 13.02 (2020), p. 2030003.
- [66] Mucong Li et al. “Linear array-based real-time photoacoustic imaging system with a compact coaxial excitation handheld probe for noninvasive sentinel lymph node mapping”. In: *Biomedical optics express* 9.4 (2018), pp. 1408–1422.
- [67] Ali Hariri et al. “The characterization of an economic and portable LED-based photoacoustic imaging system to facilitate molecular imaging”. In: *Photoacoustics* 9 (2018), pp. 10–20.
- [68] Eduard Grüneisen. “Theorie des festen Zustandes einatomiger Elemente”. In: *Annalen der Physik* 344.12 (1912), pp. 257–306.
- [69] Ali Ozbek, XL Deán-Ben, and Daniel Razansky. “Realtime parallel back-projection algorithm for three-dimensional optoacoustic imaging devices”. In: *European conference on biomedical optics*. Optica Publishing Group. 2013, p. 88000I.
- [70] Firat Ozdemir et al. “OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing”. In: *Transactions on Machine Learning Research* (2023). ISSN: 2835-8856.
- [71] Frederic M Brochu et al. “Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography”. In: *IEEE transactions on medical imaging* 36.1 (2016), pp. 322–331.

- [72] Stephen M Pizer et al. “Adaptive histogram equalization and its variations”. In: *Computer vision, graphics, and image processing* 39.3 (1987), pp. 355–368.
- [73] Alejandro F Frangi et al. “Multiscale vessel enhancement filtering”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI’98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*. Springer. 1998, pp. 130–137.
- [74] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [75] Kyong Hwan Jin et al. “Deep convolutional neural network for inverse problems in imaging”. In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522.
- [76] Xuewen Zhou et al. “Evaluation of fluence correction algorithms in multispectral photoacoustic imaging”. In: *Photoacoustics* 19 (2020), p. 100181.
- [77] Subhamoy Mandal, Xosé Luís Deán-Ben, and Daniel Razansky. “Visual quality enhancement in optoacoustic tomography using active contour segmentation priors”. In: *IEEE transactions on medical imaging* 35.10 (2016), pp. 2209–2217.
- [78] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.
- [79] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. “V-net: Fully convolutional neural networks for volumetric medical image segmentation”. In: *2016 fourth international conference on 3D vision (3DV)*. Ieee. 2016, pp. 565–571.
- [80] Berkan Lafci et al. “Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks”. In: *Photons Plus Ultrasound: Imaging and Sensing 2020*. Vol. 11240. Spie. 2020, pp. 123–128.
- [81] Elena Merčep et al. “Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.9 (2015), pp. 1651–1661.
- [82] Kalloor Joseph Francis et al. “Tomographic imaging with an ultrasound and LED-based photoacoustic system”. In: *Biomedical optics express* 11.4 (2020), pp. 2152–2165.
- [83] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. “Combined pulse-echo ultrasound and multispectral optoacoustic tomography with a multi-segment detector array”. In: *IEEE transactions on medical imaging* 36.10 (2017), pp. 2129–2137.
- [84] Anna Klimovskaia Susmelj et al. “Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data”. In: *International Conference on Medical Imaging with Deep Learning*. PMLR. 2022, pp. 1173–1191.
- [85] Tanja Tarvainen et al. “Image reconstruction with noise and error modelling in quantitative photoacoustic tomography”. In: *Photons Plus Ultrasound: Imaging and Sensing 2016*. Vol. 9708. SPIE. 2016, pp. 670–677.

- 
- [86] Chao Dong et al. “Image super-resolution using deep convolutional networks”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015), pp. 295–307.
- [87] Ozan Oktay et al. “Attention u-net: Learning where to look for the pancreas”. In: *arXiv preprint arXiv:1804.03999* (2018).
- [88] X Luís Deán-Ben et al. “Fast unmixing of multispectral optoacoustic data with vertex component analysis”. In: *Optics and Lasers in Engineering* 58 (2014), pp. 119–125.
- [89] Jørgen Arendt Jensen et al. “Synthetic aperture ultrasound imaging”. In: *Ultrasonics* 44 (2006), e5–e15.
- [90] Berkan Lafci et al. “Expediting image acquisition in reflection ultrasound computed tomography”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 69.10 (2022), pp. 2837–2848.
- [91] Shengying Li et al. “Fast marching method to correct for refraction in ultrasound computed tomography”. In: *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*. IEEE. 2006, pp. 896–899.
- [92] Erol Kalkan. “An automatic P-phase arrival-time picker”. In: *Bulletin of the Seismological Society of America* 106.3 (2016), pp. 971–986.
- [93] Manfred Baer and Urs Kradolfer. “An automatic phase picker for local and teleseismic events”. In: *Bulletin of the Seismological Society of America* 77.4 (1987), pp. 1437–1445.
- [94] Cuiping Li, Neb Duric, and Lianjie Huang. “Comparison of ultrasound attenuation tomography methods for breast imaging”. In: *Medical Imaging 2008: Ultrasonic Imaging and Signal Processing*. Vol. 6920. SPIE. 2008, pp. 338–346.
- [95] Mailyn Perez-Liva et al. “Speed of sound ultrasound transmission tomography image reconstruction based on Bézier curves”. In: *Ultrasonics* 103 (2020), p. 106097.
- [96] Berkan Lafci et al. “Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound”. In: *Neoplasia* 22.12 (2020), pp. 770–777.
- [97] M Pérez-Liva et al. “Time domain reconstruction of sound speed and attenuation in ultrasound computed tomography using full wave inversion”. In: *The Journal of the Acoustical Society of America* 141.3 (2017), pp. 1595–1604.

## 2 DEEP LEARNING FOR AUTOMATIC SEGMENTATION OF HYBRID OPTOACOUSTIC ULTRASOUND (OPUS) IMAGES

This chapter includes the following publication:

IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 68, No. 3,  
Pages. 688-696 (2021)

Authors:

**Berkan Lafci**<sup>1,2</sup>, Elena Merčep<sup>3,4</sup>, Stefan Morscher<sup>3</sup>, Xosé Luís Deán-Ben<sup>1,2</sup>, Daniel Razansky<sup>1,2</sup>

<sup>1</sup> Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>2</sup> Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>3</sup> iThera Medical GmbH, Munich, Germany

<sup>4</sup> Faculty of Medicine, Technical University of Munich, Germany

**ABSTRACT**

The highly complementary information provided by multispectral optoacoustics and pulse-echo ultrasound have recently prompted development of hybrid imaging instruments bringing together the unique contrast advantages of both modalities. In the hybrid optoacoustic ultrasound (OPUS) combination, images retrieved by one modality may further be used to improve the reconstruction accuracy of the other. In this regard, image segmentation plays a major role as it can aid improving the image quality and quantification abilities by facilitating modeling of light and sound propagation through the imaged tissues and surrounding coupling medium. Here we propose an automated approach for surface segmentation in whole-body mouse OPUS imaging using a deep convolutional neural network (CNN). The method has shown robust performance, attaining accurate segmentation of the animal boundary in both optoacoustic and pulse-echo ultrasound images, as evinced by quantitative performance evaluation using Dice coefficient metrics.

---

## 2.1 INTRODUCTION

Optoacoustic (OA) is gaining maturity as a high performance biomedical imaging technique. Both clinical and small animal multispectral optoacoustic tomography (MSOT) scanners are being employed by a steadily growing number of groups in various fields including cancer research [1–3], cardio-vascular and inflammatory diseases [4, 5], neuroscience [6, 7] and drug development [8, 9]. The unique advantages of MSOT for biological investigations stem from its inherent capacity for spectroscopic differentiation of light-absorbing substances (chromophores) with high resolution deep in scattering mammalian tissues [10, 11]. The functional and molecular imaging capabilities of MSOT can greatly benefit from hybridization with pulse-echo ultrasound (US), which can provide additional information on tissue anatomy and blood flow [12, 13]. Efficient hybrid optoacoustic ultrasound (OPUS) imaging was recently implemented with concave arrays of cylindrically-focused transducers, which allowed mitigating limited-view and side-lobe effects in both MSOT and pulse-echo US reconstructions [14]. By implementing a full-ring array tomographic imaging geometry, it was further possible to render speed of sound and attenuation maps from transmission US measurements [15].

Image formation in both US and MSOT is commonly based on coherent summation of the detected acoustic waveforms. Sequential transmission from each transducer element and subsequent image formation based on the synthetic transmit aperture (STA) method has been shown suitable for pulse-echo US beamforming with the concave array systems used in MSOT [14]. Speed of sound variations in the imaged object may lead to misplacement of the compounded images and subsequent loss of spatial resolution. Due to its shared acoustic propagation path, MSOT reconstructions are challenged by a similar problem since assumption of uniform speed of sound in the imaged tissue and the surrounding medium has been shown to result in significant distortions, blurring and loss of contrast in the images [16]. Thereby, proper segmentation of the tissue surface and assignment of correct acoustic properties are essential for rendering high quality MSOT reconstructions [17, 18]. Knowledge of the object’s boundaries can also aid proper modelling of the excitation light distribution [19, 20], thus further contribute to easier interpretation and quantification of the reconstructed images.

Image segmentation is often challenged by low contrast, noise, and other experimental factors. In pulse-echo US images, common artifacts are related to attenuation, speckle noise or shadowing, which may result in missing boundaries [21]. Efficient segmentation of MSOT images is similarly hampered by the relatively low intrinsic contrast of large anatomical structures and tissue boundaries [22]. One potential solution consists in the application of active contour models for the segmentation of MSOT images, which was further used for enhancing imaging performance by accounting for differences in speed of sound and light fluence attenuation [20]. Methods based on deep learning (DL) have been increasingly utilized for medical image processing over the last two decades owing to their good performance in classification and function approximation, particularly, in the case of pre-processing (e.g. construction and

restoration), segmentation, registration and recognition [23]. The advent of DL methods consisting of more hidden layers and providing higher level of abstraction has further enhanced the performance of convolutional neural networks (CNNs) for medical image analysis.

Herein, we propose a CNN based on the U-Net architecture [24] for segmenting images acquired with a hybrid OPUS imaging scanner. A large number of manually-segmented OA images was employed for efficient training of the network, whose performance was then evaluated on images rendered with both modalities by calculating Dice coefficient metrics [25] of the difference between the automatically segmented and ground truth (manually segmented) images. The network performance was further compared with those of active contour segmentation and of a CNN trained with both OA and US images.

## 2.2 MATERIALS AND METHODS

The key system components and the general workflow are schematically depicted in Fig. 2.1. In short, pulse-echo US and OA images were acquired from living mice non-invasively with a small-animal OPUS scanner. The outer boundaries were manually segmented (labelled) from the reconstructed images and served as ground truth for training a CNN based on the U-Net architecture. The performance of the CNN was subsequently compared with that of the active contour segmentation method by quantifying differences in the segmentation results with the ground truth approach using Dice coefficient metric. A more detailed description of the methods is provided below.

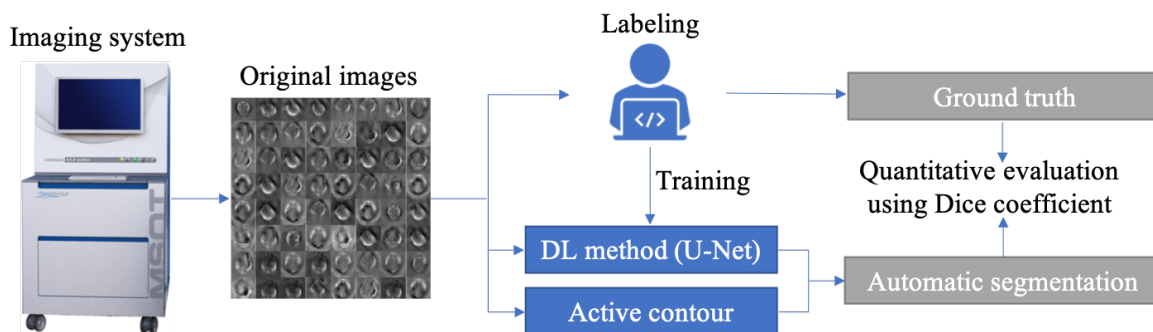


FIGURE 2.1: Key components and workflow of the conducted study. Images were acquired by multispectral optoacoustic tomography (MSOT) small animal scanners (Models inVision 256-TF and inVision 512-echo, iThera Medical GmbH, Munich, Germany) and manually segmented to create ground truth data. Deep learning (DL)-based segmentation method or active contour method are applied on the acquired images. The results are evaluated by comparing the output of the segmentation algorithms with manually segmented images using Dice coefficient.

### 2.2.1 CONVOLUTIONAL NEURAL NETWORK

CNNs use image kernels to extract image features and learn the optimal kernels by updating them during training for a specific task. Kernel combinations using different sizes and interconnections lay the foundation of the CNN architectures [26]. The CNN architecture

used in this work was first proposed in [24] and referred to as the U-Net architecture. U-Net architecture features a combination of encoders and decoders, where every encoder layer contains convolutional kernels followed by steps of batch normalization and rectified linear unit (ReLU) [27]. In our case, the network used 256x256 pixel image with one channel as input, increased the channel size of the image and downsampled the image in the encoder part. In the decoder part, the network up-sampled and concatenated the image while the channel size was reduced. The network was further adapted to the input image sizes and layers using the PyTorch framework, as shown in Fig. 2.2. After the application of two convolutional layers having 3x3 kernels with stride 1, a max pooling layer having 2x2 kernels with stride 2 was applied. The decoder layers first applied transposed convolution with kernel size 3x3, stride 1 and up-sampled the input with a factor of 2 using bilinear interpolation. Then, the same sized encoder layer was concatenated with the decoder layer. After concatenation, the convolutional layer with 3x3 kernels on inputs that have the stride value of 1 was applied followed by the batch normalization and ReLU.

Augmentation methods were further applied in order to increase the amount of training data and ensure variability in the dataset. Specifically, this was done by random scaling, rotation and shifting of the images. The network used a loss function combined in different ratios (0.25-0.75, 0.5-0.5, 0.75-0.25) of binary cross entropy and soft Dice loss. The loss function was optimized by stochastic gradient descent (SGD) with learning rate 0.01 and momentum 0.99. The training was done with batch size 5 over 100 epochs. The initialization of the weights was done by the Xavier uniform initializer [28].

### 2.2.2 GROUND TRUTH MANUAL SEGMENTATION

Manual segmentation still remains a common practice in image analysis applications to obtain ground truth data for evaluating the performance of image segmentation approaches [29, 30] or providing data for CNN training [31]. In this work, manual segmentation of all OA and US images was done under supervision of an experienced biologist well trained in the imaging field. It was performed by drawing polygons around the structures followed by spline interpolation of the polygon for smooth contour appearance. The final ground truth data contained binary 300x300 pixels images, where “1” and “0” corresponds to pixels located inside and outside the mouse, respectively.

### 2.2.3 ACTIVE CONTOUR BY EDGE DETECTION

The CNN-based segmentation performance was evaluated and compared with the active contour edge detection (ACED) method, an automatic segmentation approach widely employed in medical image analysis [32–34]. Specifically, we used the implementation of active contour model introduced by Li et al. [35] combined with an automatic initialization method by edge detection and circle fitting algorithm. Automatic initialization facilitated the application of active contour segmentation as an automatic method not affected by user-dependent inputs. This ensured a fair comparison with the CNN, which is also an automatic segmentation method. The algorithm steps as well as intermediate outputs are shown in Fig. 2.3.



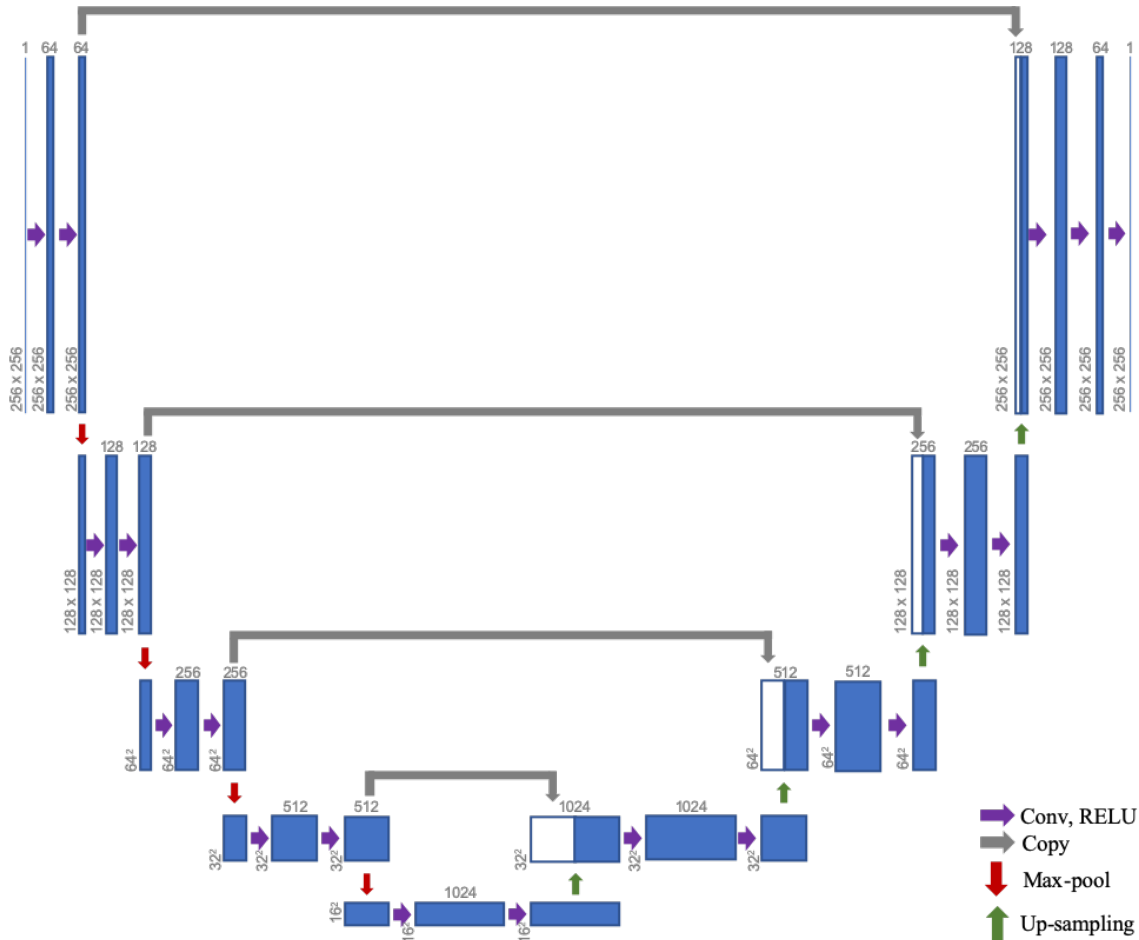


FIGURE 2.2: The U-Net convolutional neural network (CNN) architecture used for image segmentation. The network uses 256x256 pixel image with one channel as an input, increases the channel size of the image and downscales the image in the encoder part (left). In the decoder part, the network up-samples and concatenates the image while the channel size is reduced (right).

The ACED algorithm received as input OA or US images as a square array of 256x256 pixels (Fig. 2.3a). The images were first downscaled to 150x150 pixels to reduce the computation time whereas the pixel intensities were converted to 8-bit range between 0 and 255. Edge detection was implemented to overcome any dependency of the initial guess upon the user. For this, the Canny edge detector [36] was applied after smoothing the image using Gaussian filter with kernel size 3 and sigma 0.5. A binary image was then created with the pixels on the edges labelled as “1” and the background labelled as “0” (Fig. 2.3b). The outliers and the non-connected components in the pixels erroneously detected as edges were removed by applying morphological operations of dilation and erosion with a disc-shaped structuring element of 3-pixel size (Fig. 2.3c). A circle that has the minimum radius of all possible circles enclosing all the edge pixels was then selected (Fig. 2.3d). This circle was subsequently fed into the active contour function along with the number of iterations (20) and step size (1 pixel). The optimal number of iterations and step size that maximize Dice coefficient for the entire dataset was selected heuristically. The deformable spline contour evolves based on the minimization of an energy function with a distance regularization term and an external

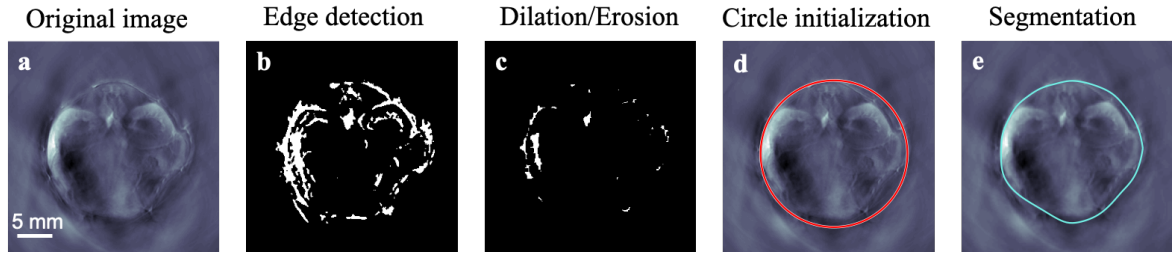


FIGURE 2.3: Algorithmic steps of the active contour edge detection (ACED) segmentation approach. (a) The originally reconstructed optoacoustic (OA) image. (b) Output of Canny edge detector. (c) Image after morphological operations of dilation/erosion. (d) Fitted circle on pixels detected in the previous step. (e) The final result of the active contour segmentation.

energy that drives the motion of the contour toward desired locations (Fig. 2.3e) [35].

#### 2.2.4 EXPERIMENTAL DATA SETS

All images were acquired with the commercial small animal imaging systems inVision 256-TF and inVision 512-echo (iThera Medical GmbH, Munich, Germany). All procedures involving animal care and experimentation were conducted in full compliance with the animal handling guidelines and with approval from the institutional review board. The detected OA signals for 800 nm excitation wavelength were band-pass filtered with 50 kHz and 6.5 MHz cut-off frequencies and deconvolved with the electrical impulse response of the transducer to account for its limited bandwidth and phase distortions [10]. A two-dimensional filtered back-projection algorithm [16] was subsequently used to reconstruct cross-sectional images from the OA pressure signals with a pixel size and field of view (FOV) set to 100  $\mu\text{m}$  and 30x30  $\text{mm}^2$ , respectively.

The US signals were acquired using STA by repeating transmission events for every single element of the ring-shaped detection array operating at 5 MHz central frequency. Signals from every transmission event were then beamformed using delay and sum (DAS) method to generate low-resolution images. The final high-resolution US images were then rendered by compounding low-resolution images that were generated by the individual transmission events. Overall, STA with sub-aperture size of 128 elements in combination with spatial compounding were applied, as described elsewhere [15, 37]. The final US images were generated in a 30x30  $\text{mm}^2$  FOV with a pixel size of 100  $\mu\text{m}$ . Details on the training and test datasets for OA and US images are provided below.

**Training Data Sets:** The dataset was carefully selected by assigning all acquired images to either training or test sets based on the mouse ID. This represents a standard procedure that avoids accidental use of images from the same mouse for both training and testing, thus prevents network from overfitting the data. Training was first performed using exclusively OA images. Subsequently, a combination of OA and US images was used for training. Note that the number of available US images was lower. In total, the OA training data contained 174 images (12 mice) for the brain region, 97 images (13 mice) for the kidney region and

108 images (14 mice) for the liver region. The US training data contained 98 images (13 mice) and 78 images (13 mice) from liver and kidney regions, respectively. The brain, kidney and liver data sets were subsequently combined to generate training dataset, i.e., training was eventually performed using 379 OA and 176 US images in total for brain, kidney and liver regions. A fragment of the training data was further separated for validation purposes during the training process. The ratio between training and validation sets was chosen as 0.80.

**Test Data Sets:** The datasets used for evaluation of segmentation methods included both OA and US images, as summarized in Table 2.1.

Anatomical Region	Number of Images	
	OA Mode	US Mode
Brain	28 (4 mice)	9 (2 mice)
Liver	38 (5 mice)	16 (5 mice)
Kidney	33 (5 mice)	30 (5 mice)
Total	99	55

TABLE 2.1: Summary of the test data sets.

Test datasets further included 5 brain, 4 kidney and 4 liver images with artifacts and 19 liver images acquired with mouse positioned off-center in the imaging chamber that are included in Table 2.1. Those data sets were included in order to test the performance of different segmentation algorithms under sub-optimal data acquisition conditions. Presence of prominent reflection artifacts in the OA images of the brain region can be ascribed to air-filled cavities or bones [38], while the off-center positioning of the animal in the imaging chamber may occur due to operation by unexperienced users.

### 2.2.5 EVALUATION CRITERIA

Final evaluation was done using the Dice coefficient, also called the overlap index [25]. The Dice coefficient can be calculated by pixel-wise comparison based on the true positive (TP), false positive (FP) and false positive (FN) ratios as follows

$$Dice = \frac{2TP}{2TP + FP + FN}.$$

## 2.3 RESULTS

The segmentation performance of the CNN trained exclusively with OA data and the ACED algorithm for exemplary test datasets of the brain (Fig. 2.4a-c), liver (Fig. 2.4d-f), and kidney (Fig. 2.4g-i) cross-sections is shown. It can readily be observed that the CNN method outperformed the ACED approach in all representative anatomical regions. It was further

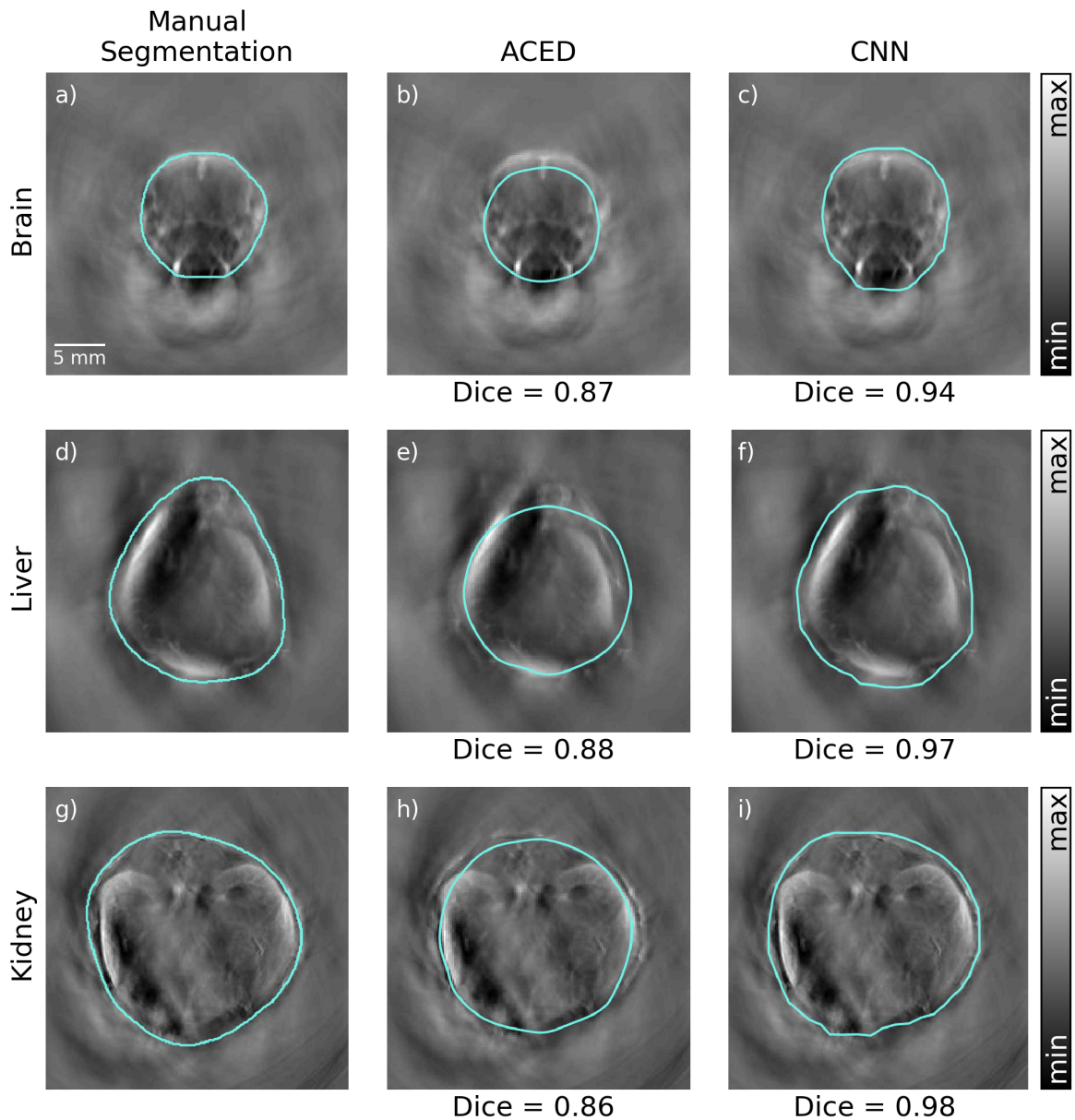


FIGURE 2.4: Segmentation results for optoacoustic (OA) images obtained with manual segmentation (left), active contour edge detection (ACED) (middle) and convolutional neural network (CNN) (right) for cross-sections from the brain (a-c), liver (d-f), and kidney (g-i) regions. The corresponding Dice coefficients are also shown.

possible to accurately segment the mouse surface in images containing indistinct boundaries (Fig. 2.4c, i) or non-circular shapes (Fig. 2.4f). This observation is quantitatively confirmed by the higher Dice coefficients obtained using CNN. The poorer performance of the ACED algorithm may be attributed to the step of circle initialization, when high-intensity regions may erroneously be classified as edges (Fig. 2.4b), or to the step of the active contour evolution, subject to over-fitting caused by low-contrast regions in the image (Figs. 2.4e,h). Manual initialization and parameter optimization on a per-image basis may yet recover some of the ACED performance to the detriment of the loss of an automated approach. A robust analysis was ensured by empirically optimizing the parameters to provide an optimal performance for the entire dataset rather than for individual images.

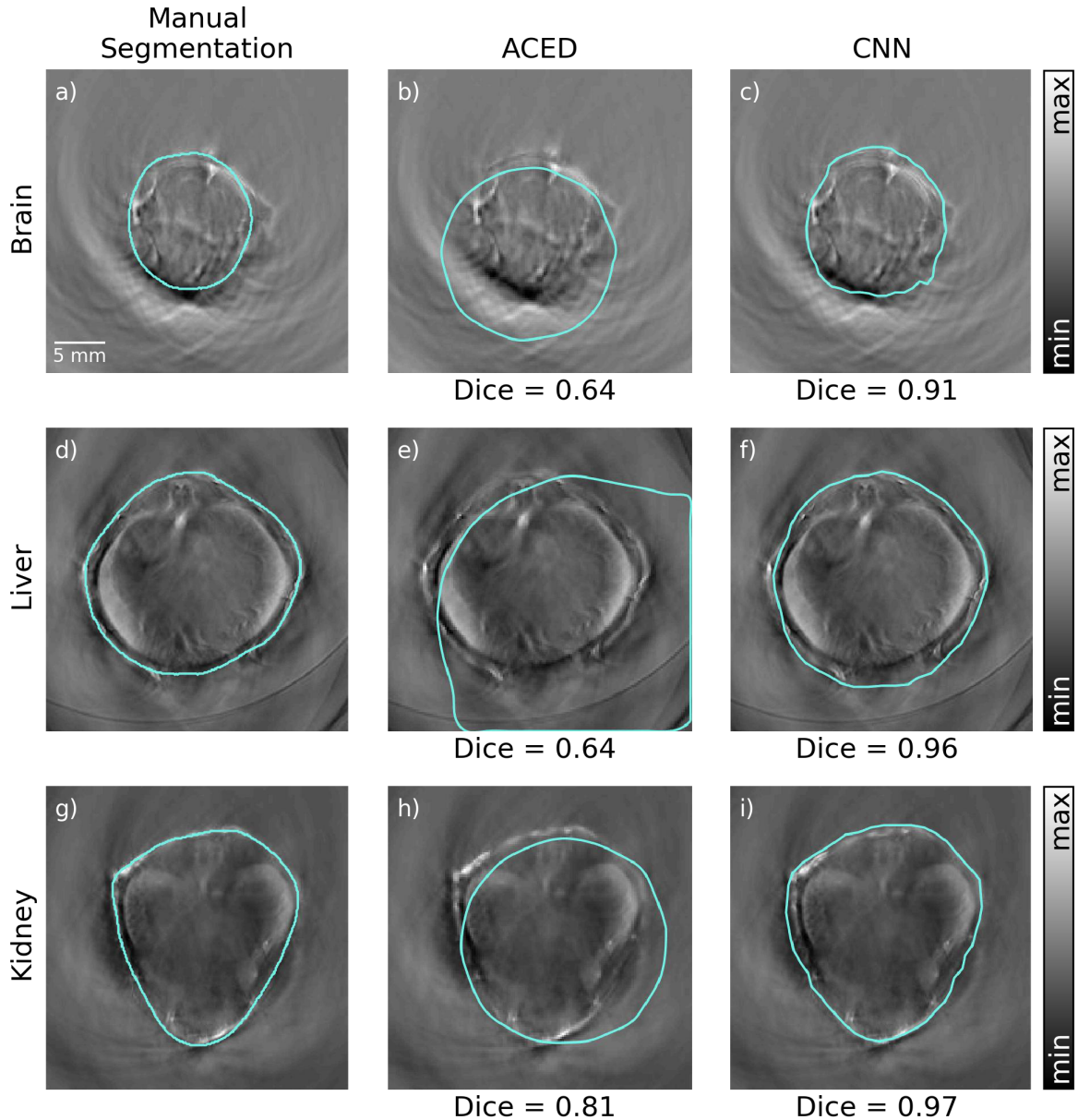


FIGURE 2.5: Segmentation results for optoacoustic (OA) images obtained with manual segmentation, active contour edge detection (ACED) and convolutional neural network (CNN) for mouse cross-sections through brain (a-c), liver (d-f), and kidney (g-i) regions that showcase sub-optimal (non-centered) animal positioning (kidney) and artifacts (liver, brain). The corresponding Dice coefficients are also shown.

The superior performance of the CNN trained exclusively with OA data over ACED was corroborated with sub-sets of images whose quality was compromised by sub-optimal data acquisition conditions, which resulted in image artifacts and/or diminished resolution with the animal positioned off-center in the imaging chamber (Fig. 2.5). Specifically, images of the brain and liver regions (first and second row in Fig. 2.5) with arc-shaped artifacts as well as an image of the kidney region (third row of Fig. 2.5) with sub-optimal positioning of animal (off-center) are shown. The performance of the CNN was barely affected in those cases (Figs. 2.5c, f, and i), whereas the performance of the ACED algorithm drastically deteriorated. For two out of the three analyzed anatomical regions, circle initialization failed as

the “arc”-shape reflection artefacts at the periphery of the FOV were classified as edges. This has led to selecting a larger initial circle and erroneous inclusion of pixels lying outside the mouse body into the segmented region (Fig. 2.5b, e). The CNN-based algorithm performed equally well also when the mouse was positioned off-center (Fig. 2.5i), with yielded Dice score of 0.98 versus 0.81 for the ACED method (Fig. 2.5h).

The applicability of the CNN trained exclusively on OA images for US image segmentation is exemplified in Fig. 2.6. Good segmentation performance was observed in the liver and kidney cross-sections (Fig. 2.6f, i), attaining Dice coefficients of 0.95 and 0.95, respectively. However, the CNN was less efficient when segmenting brain cross-sections (Fig. 2.6c). This

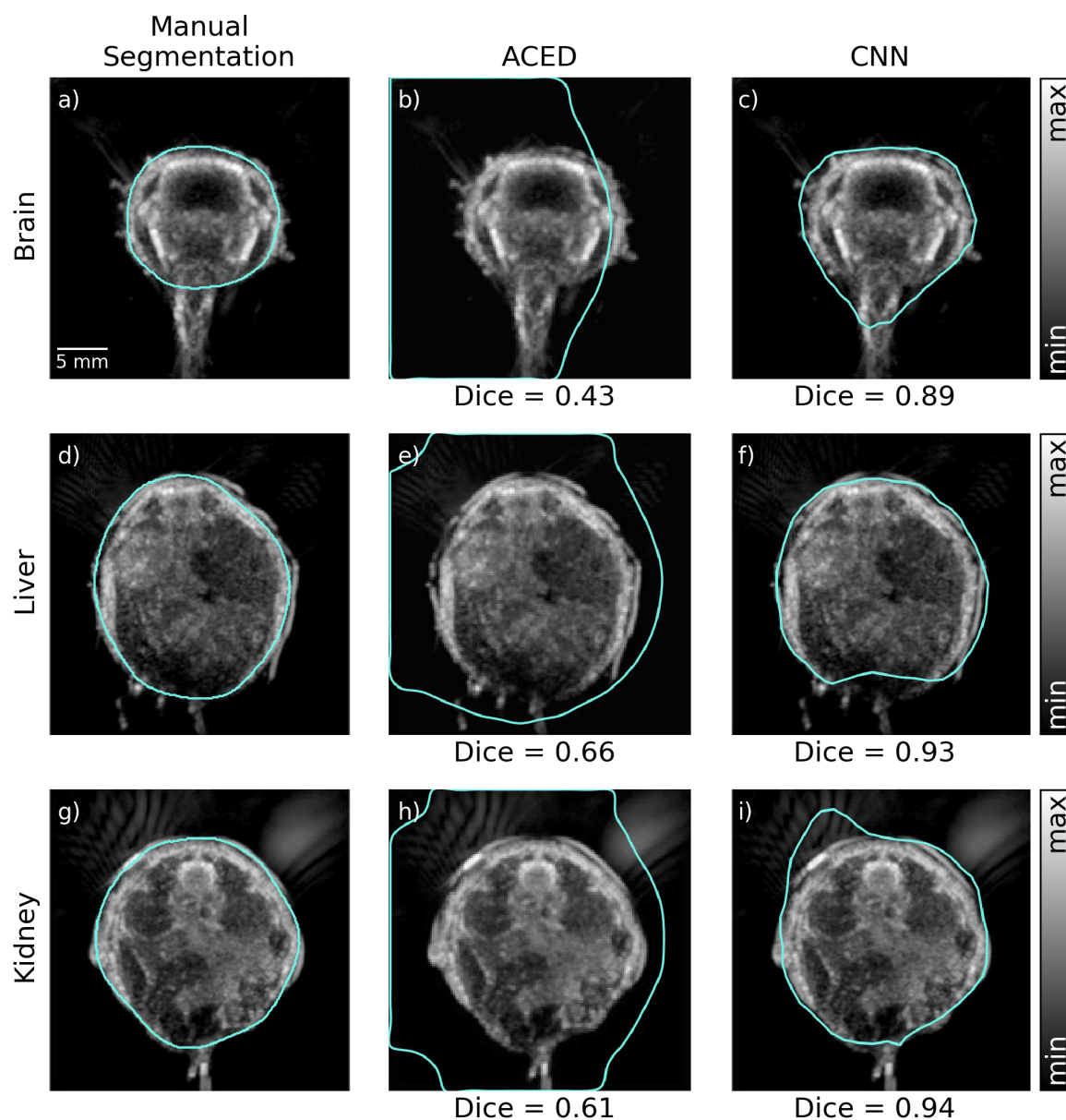


FIGURE 2.6: Segmentation results for ultrasound (US) images obtained with manual segmentation, active contour edge detection (ACED) and convolutional neural network (CNN) (trained on optoacoustic (OA) images) for mouse cross-sections through brain (a-c), liver (d-f), and kidney (g-i) regions with the corresponding Dice coefficients.

may be attributed to the presence of high-contrast features from the acoustically reflecting but optically transparent protective waterproof membrane (Fig. 2.6a), not present in the OA training images. In all cases, the ACED algorithm performed poorly owing to the failed initialization step in the presence of reflections outside the mouse body, erroneously classified as edges.

The overall performance of both algorithms for all the test datasets is summarized in Fig. 2.7. The CNN algorithm that was trained with a loss function ratio with 0.5-0.5 of binary cross entropy and soft Dice loss outperformed ACED for segmentation of OA images with a corresponding Dice score of  $0.95 \pm 0.04$  and a very narrow range of outliers, versus  $0.78 \pm 0.11$  for the ACED method. Both methods showed inferior performance for segmentation of US images, in particularly with respect to the number of outliers, whereas CNN still yielded better results than ACED with Dice score of  $0.80 \pm 0.08$  versus  $0.55 \pm 0.10$ .

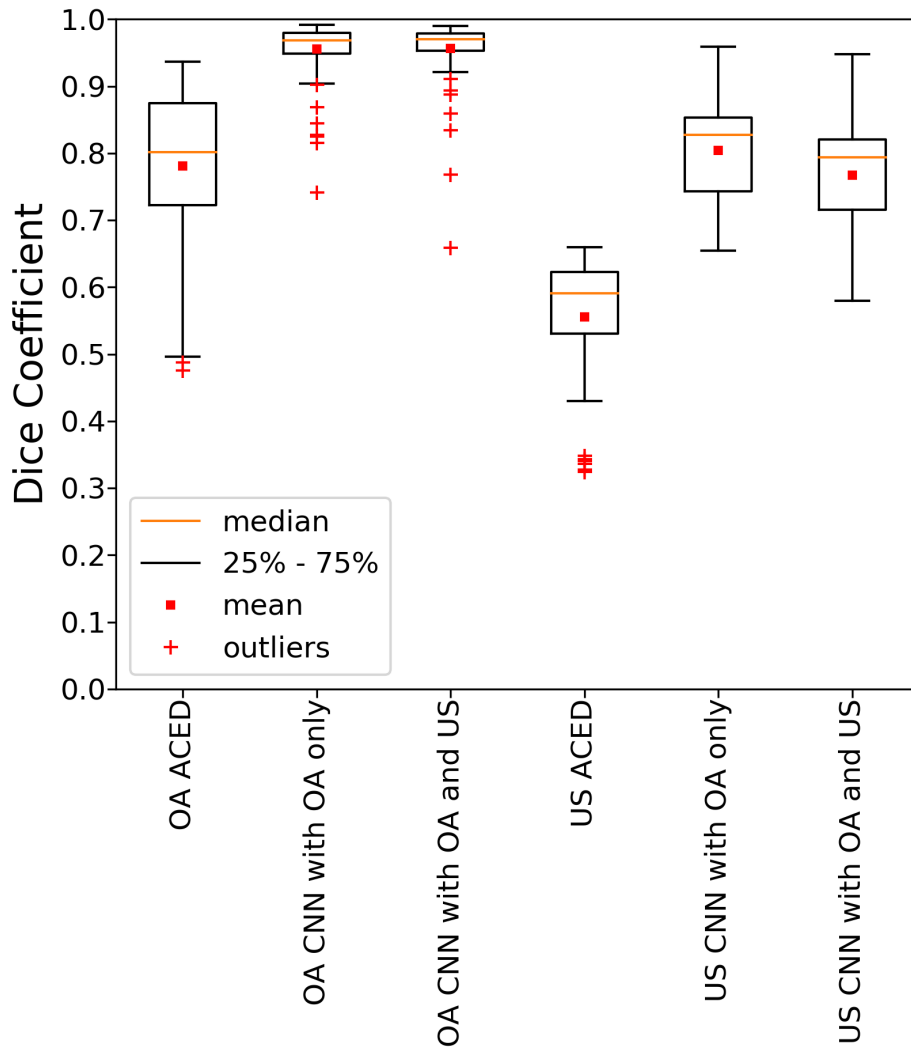


FIGURE 2.7: Boxplots illustrating the Dice coefficient variability for the segmented optoacoustic (OA) and ultrasound (US) images obtained with active contour edge detection (ACED), convolutional neural network (CNN) trained with only OA images and CNN trained with combination of OA and US images.

Different types of noise (Gaussian, salt & pepper and speckle) was further added to OA images to demonstrate robustness of the proposed algorithm. The network achieved Dice scores of  $0.94\pm 0.03$  in the presence of Gaussian noise (0 mean, 0.1 variance) for OA images and  $0.78\pm 0.10$  for US images. The salt and pepper noise (0.004 noise density) decreased the mean Dice scores to  $0.94\pm 0.02$  for OA images and  $0.76\pm 0.09$  for US images compared to noise free images. The network showed an inferior performance for OA images in the presence of speckle noise, i.e when randomly generated numbers between 0 and 255 were added. Specifically, it achieved Dice scores of  $0.93\pm 0.03$  for OA images and  $0.81\pm 0.06$  for US images.

We further analyzed the effects of loss functions on training. For this, the network was trained with different combinations of binary cross entropy and soft Dice loss. It achieved Dice score of  $0.96\pm 0.02$  for OA images and  $0.83\pm 0.08$  for US images for a 0.25-0.75 ratio of binary cross entropy and soft Dice loss ratio. On the other hand, it achieved Dice score of  $0.95\pm 0.02$  for OA images and  $0.80\pm 0.11$  for US images when considering a ratio 0.75-0.25 of binary cross entropy and soft Dice loss ratio. As expected, it appears that a higher weight in the soft Dice loss results in an increased Dice score on the test images.

The CNN was further trained by additionally including the available 176 US images from the kidney and liver cross sections. In this case, no improvement in the overall performance for OA image segmentation was observed. In fact, the measured Dice coefficient was  $0.95\pm 0.4$ , approximately the same as previously measured when training exclusively with the OA images (Fig. 2.7). In addition, the number of outliers increased in OA test images. On the other hand, the performance of US image segmentation degraded when including US images into training. The network trained with both OA and US images reached a Dice coefficient of  $0.76\pm 0.09$ , i.e. lower than the network trained only with OA images. We ascribe the inferior performance of the CNN trained with a combination of OA and US data to the relatively low number of available US images. We believe that the fact that the network trained with OA data can additionally accurately segment US images represents an important advantage of the suggested processing approach for hybrid OPUS imaging.

## 2.4 CONCLUSION AND DISCUSSION

This work reports on a new methodology for automatic segmentation of the dual-modality images acquired with hybrid OPUS imaging systems by means of the U-Net CNN architecture. It has shown robust performance in accurate segmentation of mouse boundary in both OA and pulse-echo US images, as evinced by quantitative evaluation of the CNN performance using Dice coefficient metrics based on the difference between the automatically segmented and ground truth images. In contrast, performance of the active contour based segmentation approach is severely compromised by image artifacts, contrast variations as well as off-center positioning of the mouse. However, the ultimate goal of medical image segmentation is elimination of user dependency without sacrificing segmentation accuracy. In this respect, the proposed CNN approach has shown great promise in accurate segmentation in cross-sections corresponding to the mouse brain, liver and kidney regions. Better performance of the net-



work may potentially be achieved by separating brain, kidney and liver datasets for every training process [39]. However, the data from three different anatomical regions was mixed to demonstrate the robustness of the network for different contrast and shapes. As accurate labeling of large amount of data is generally a time-consuming process, in this work we concentrated on showcasing performance in three selected anatomical regions. However, the CNN approach could be seamlessly applied to whole-body mouse segmentation as well as segmentation of distinct organs.

We further demonstrated applicability of the CNN trained on OA images for segmentation of pulse-echo US images, although this approach may have limitations due to different contrast exhibited by the two modalities. However, our results indicate that additional training with the US data did not improve the segmentation accuracy, although larger amount of datasets may need to be taken into account to draw unequivocal conclusions. In future studies, we will explore the possibility of feeding matching pairs of OA and US in the same network to increase the training performance. For this, a hybrid imaging system enabling simultaneous image acquisition in both modalities is required. The suggested approach can potentially be applied to hybrid images acquired by other types of OPUS imaging systems. For instance, multi-segment arrays consisting of concave and linear parts have been recently shown to provide an outstanding performance for multi-modal imaging [12, 40]. This particular embodiment employs different parts of the array for OA and pulse-echo US imaging, which may imply separate CNN training. The OPUS modality has also been extended to mapping the speed of sound and acoustic attenuation in tissues by using a transmission US approach [15], which may equally be used to facilitate more efficient segmentation. Three-dimensional systems providing both OA and US images have also been suggested for both small animal tomography and hand-held scanning in humans [41, 42]. Much like in the two-dimensional case, accurate segmentation of the tissue boundary can be used to improve spatial resolution and facilitate image co-registration.

The advent of DL methods represents a major shift from traditional machine learning approaches [43]. In addition to showing unprecedented performance in computer vision tasks, DL networks have been applied for processing of images from different modalities in the field of medical imaging, including computed tomography (CT) [44], magnetic resonance imaging (MRI) [45], positron emission tomography (PET) [46] and US [47–50]. The proposed CNN approach for automatic OPUS image segmentation can be used to more accurately compensate for image artifacts originating from light attenuation in deep tissues as well as speed of sound heterogeneities. Apart from image segmentation, DL methods have been shown to significantly enhance other signal and image processing steps, including tomographic reconstructions from sparse OA data [51–53], image artifact removal [54, 55] and quantitative analysis of OA images [56]. Considering the generally high complexity of the OA image reconstruction pipeline, in particular for multispectral methods like MSOT, the DL approaches are expected to play a major role in facilitating accurate image reconstruction and analysis.

#### ACKNOWLEDGEMENT

The authors acknowledge funding support from the Swiss Data Science Center (project C19-04) and the European Research Council under Consolidator Grant ERC-2015-CoG-682379.

#### DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors declare that they have no competing financial interests.

## BIBLIOGRAPHY

- [1] Adrian Taruttis, Gooitzen M van Dam, and Vasilis Ntziachristos. “Mesoscopic and macroscopic optoacoustic imaging of cancer”. In: *Cancer research* 75.8 (2015), pp. 1548–1559.
- [2] Keerthi S Valluru and Juergen K Willmann. “Clinical photoacoustic imaging of cancer”. In: *Ultrasonography* 35.4 (2016), p. 267.
- [3] Mohammad Mehrmohammadi et al. “Photoacoustic imaging for cancer detection and staging”. In: *Current Molecular Imaging (Discontinued)* 2.1 (2013), pp. 89–105.
- [4] Xosé Luís Deán-Ben, Steven James Ford, and Daniel Razansky. “High-frame rate four dimensional optoacoustic tomography enables visualization of cardiovascular dynamics and mouse heart perfusion”. In: *Scientific reports* 5.1 (2015), p. 10133.
- [5] Sophinese Iskander-Rizk, Antonius FW Van Der Steen, and Gijs Van Soest. “Photoacoustic imaging for guidance of interventions in cardiovascular medicine”. In: *Physics in Medicine & Biology* 64.16 (2019), 16TR01.
- [6] X Luís Deán-Ben et al. “Functional optoacoustic neuro-tomography for scalable whole-brain monitoring of calcium indicators”. In: *Light: Science & Applications* 5.12 (2016), e16201–e16201.
- [7] Junjie Yao and Lihong V Wang. “Photoacoustic brain imaging: from microscopic to macroscopic scales”. In: *Neurophotonics* 1.1 (2014), pp. 011003–011003.
- [8] Daniel Razansky et al. “Deep tissue optical and optoacoustic molecular imaging technologies for pre-clinical research and drug discovery”. In: *Current pharmaceutical biotechnology* 13.4 (2012), pp. 504–522.
- [9] Yuqi Zhang et al. “Photoacoustic drug delivery”. In: *Sensors* 17.6 (2017), p. 1400.
- [10] Rui Ma et al. “Multispectral optoacoustic tomography (MSOT) scanner for whole-body small animal imaging”. In: *Optics express* 17.24 (2009), pp. 21414–21426.
- [11] Martin P Mienkina et al. “Multispectral photoacoustic coded excitation imaging using unipolar orthogonal Golay codes”. In: *Optics Express* 18.9 (2010), pp. 9076–9087.
- [12] Xosé Luis Deán-Ben, E Merčep, and Daniel Razansky. “Hybrid-array-based optoacoustic and ultrasound (OPUS) imaging of biological tissues”. In: *Applied Physics Letters* 110.20 (2017), p. 203703.
- [13] Robert Nuster et al. *Hybrid photoacoustic and ultrasound section imaging with optical ultrasound detection*. 2013.
- [14] Elena Merčep et al. “Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.9 (2015), pp. 1651–1661.

- [15] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [16] Christian Lutzweiler and Daniel Razansky. “Optoacoustic imaging and tomography: reconstruction approaches and outstanding challenges in image performance and quantification”. In: *Sensors* 13.6 (2013), pp. 7345–7384.
- [17] Jithin Jose et al. “Speed-of-sound compensated photoacoustic tomography for accurate imaging”. In: *Medical physics* 39.12 (2012), pp. 7262–7271.
- [18] X Luís Deán-Ben, Vasilis Ntziachristos, and Daniel Razansky. “Effects of small variations of speed of sound in optoacoustic tomographic imaging”. In: *Medical physics* 41.7 (2014), p. 073301.
- [19] Frederic M Brochu et al. “Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography”. In: *IEEE transactions on medical imaging* 36.1 (2016), pp. 322–331.
- [20] Subhamoy Mandal, Xosé Luís Deán-Ben, and Daniel Razansky. “Visual quality enhancement in optoacoustic tomography using active contour segmentation priors”. In: *IEEE transactions on medical imaging* 35.10 (2016), pp. 2209–2217.
- [21] J Alison Noble and Djamel Boukerroui. “Ultrasound image segmentation: a survey”. In: *IEEE Transactions on medical imaging* 25.8 (2006), pp. 987–1010.
- [22] Christian Lutzweiler, Reinhard Meier, and Daniel Razansky. “Optoacoustic image segmentation based on signal domain analysis”. In: *Photoacoustics* 3.4 (2015), pp. 151–158.
- [23] Zhenghao Shi et al. “Survey on neural networks used for medical image processing”. In: *International journal of computational science* 3.1 (2009), p. 86.
- [24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.
- [25] Lee R Dice. “Measures of the amount of ecologic association between species”. In: *Ecology* 26.3 (1945), pp. 297–302.
- [26] Yann LeCun et al. “Object Recognition with Gradient-Based Learning”. In: *Shape, Contour and Grouping in Computer Vision* (1999), p. 319.
- [27] Vinod Nair and Geoffrey E Hinton. “Rectified linear units improve restricted boltzmann machines”. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010, pp. 807–814.
- [28] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feed-forward neural networks”. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 249–256.

- [29] Kathrin Tingelhoff et al. “Analysis of manual segmentation in paranasal CT images”. In: *European archives of oto-rhino-laryngology* 265 (2008), pp. 1061–1070.
- [30] Xiaoyi Jiang et al. “Distance measures for image segmentation evaluation”. In: *EURASIP Journal on Advances in Signal Processing* 2006 (2006), pp. 1–10.
- [31] Fernando C Monteiro and Aurélio C Campilho. “Performance evaluation of image segmentation”. In: *Image Analysis and Recognition: Third International Conference, ICIAR 2006, Póvoa de Varzim, Portugal, September 18-20, 2006, Proceedings, Part I 3*. Springer. 2006, pp. 248–259.
- [32] Foued Derraz, Mohamed Beladgham, and M’hamed Khelif. “Application of active contour models in medical image segmentation”. In: *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004*. Vol. 2. IEEE. 2004, pp. 675–681.
- [33] Riccardo Boscolo, Matthew S Brown, and Michael F McNitt-Gray. “Medical image segmentation with knowledge-guided robust active contours”. In: *Radiographics* 22.2 (2002), pp. 437–448.
- [34] Xiaohua Qian et al. “An active contour model for medical image segmentation with application to brain CT image”. In: *Medical physics* 40.2 (2013), p. 021911.
- [35] Chunming Li et al. “Distance regularized level set evolution and its application to image segmentation”. In: *IEEE transactions on image processing* 19.12 (2010), pp. 3243–3254.
- [36] John Canny. “A computational approach to edge detection”. In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), pp. 679–698.
- [37] Elena Merčep et al. “Whole-body live mouse imaging by hybrid reflection-mode ultrasound and optoacoustic tomography”. In: *Optics letters* 40.20 (2015), pp. 4643–4646.
- [38] X Luis Dean-Ben et al. “Statistical approach for optoacoustic image reconstruction in the presence of strong acoustic heterogeneities”. In: *IEEE Transactions on Medical Imaging* 30.2 (2010), pp. 401–408.
- [39] Kamal Jnawali et al. “Automatic cancer tissue detection using multispectral photoacoustic imaging”. In: *International Journal of Computer Assisted Radiology and Surgery* 15 (2020), pp. 309–320.
- [40] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. “Combined pulse-echo ultrasound and multispectral optoacoustic tomography with a multi-segment detector array”. In: *IEEE transactions on medical imaging* 36.10 (2017), pp. 2129–2137.
- [41] Volker Neuschmelting et al. “Performance of a multispectral optoacoustic tomography (MSOT) system equipped with 2D vs. 3D handheld probes for potential clinical translation”. In: *Photoacoustics* 4.1 (2016), pp. 1–10.
- [42] Thomas Felix Fehm, Xosé Luís Deán-Ben, and Daniel Razansky. “Four dimensional hybrid ultrasound and optoacoustic imaging via passive element optical excitation in a hand-held probe”. In: *Applied Physics Letters* 105.17 (2014), p. 173505.

- [43] Dan Claudiu Ciresan et al. “Flexible, high performance convolutional neural networks for image classification”. In: *Twenty-second international joint conference on artificial intelligence*. 2011.
- [44] Patrick Ferdinand Christ et al. “Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks”. In: *arXiv preprint arXiv:1702.05970* (2017).
- [45] Zeynettin Akkus et al. “Deep learning for brain MRI segmentation: state of the art and future directions”. In: *Journal of digital imaging* 30 (2017), pp. 449–459.
- [46] Paul Blanc-Durand et al. “Automatic lesion detection and segmentation of 18F-FET PET in gliomas: a full 3D U-Net convolutional neural network study”. In: *PLoS One* 13.4 (2018), e0195798.
- [47] Moi Hoon Yap et al. “Automated breast ultrasound lesions detection using convolutional neural networks”. In: *IEEE journal of biomedical and health informatics* 22.4 (2017), pp. 1218–1226.
- [48] Pádraig Looney et al. “Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning”. In: *JCI insight* 3.11 (2018).
- [49] Mina Amiri, Rupert Brooks, and Hassan Rivaz. “Fine tuning u-net for ultrasound image segmentation: which layers?” In: *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data: First MICCAI Workshop, DART 2019, and First International Workshop, MIL3ID 2019, Shenzhen, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 1*. Springer. 2019, pp. 235–242.
- [50] Manna E Philip et al. “Convolutional neural networks for automated fetal cardiac assessment using 4d b-mode ultrasound”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 824–828.
- [51] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky. “Deep learning optoacoustic tomography with sparse data”. In: *Nature Machine Intelligence* 1.10 (2019), pp. 453–460.
- [52] Stephan Antholzer, Markus Haltmeier, and Johannes Schwab. “Deep learning for photoacoustic tomography from sparse data”. In: *Inverse problems in science and engineering* 27.7 (2019), pp. 987–1005.
- [53] Andreas Hauptmann et al. “Model-based learning for accelerated, limited-view 3-D photoacoustic tomography”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1382–1393.
- [54] Derek Allman, Austin Reiter, and Muyinatu A Lediju Bell. “A machine learning method to identify and remove reflection artifacts in photoacoustic channel data”. In: *2017 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2017, pp. 1–4.
- [55] Steven Guan et al. “Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal”. In: *IEEE journal of biomedical and health informatics* 24.2 (2019), pp. 568–576.

- [56] Changchun Yang et al. “Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 741–744.

### 3 EXPEDITING IMAGE ACQUISITION IN REFLECTION ULTRASOUND COMPUTED TOMOGRAPHY

This chapter includes the following publication:

IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 69, No. 10,  
Pages. 2837 - 2848 (2022)

Authors:

**Berkan Lafci**<sup>1,2</sup>, Justine Robin<sup>1,2,3</sup>, Xosé Luís Deán-Ben<sup>1,2</sup>, Daniel Razansky<sup>1,2</sup>

<sup>1</sup> Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>2</sup> Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>3</sup> Present affiliation: Physics for Medicine Paris, INSERM U1273, ESPCI Paris, PSL University, CNRS UMR 8063, Paris, France



## ABSTRACT

Reflection ultrasound computed tomography (RUCT) attains optimal image quality from objects that can be fully accessed from multiple directions, such as the human breast or small animals. Owing to the full-view tomography approach based on compounding of images taken from multiple angles, RUCT effectively mitigates several deficiencies afflicting conventional pulse-echo ultrasound (US) systems, such as speckle patterns and inter-user variability. On the other hand, the small inter-element pitch required to fulfill the spatial sampling criterion in the circular transducer configuration used in RUCT typically implies the use of an excessive number of independent array elements. This increases the system's complexity and costs and limits the achievable imaging speed. Here we explore acquisition schemes that enable RUCT imaging with reduced number of transmit/receive elements. We investigated the influence of the element size in transmission and reception in a ring array geometry. Performance of a sparse acquisition approach based on partial acquisition from a sub-set of the elements has been further assessed. A larger element size is shown to preserve contrast and resolution at the center of the field of view (FOV), while a reduced number of elements is shown to cause uniform loss of contrast and resolution across the entire FOV. The trade-offs of achievable FOV, contrast to noise ratio, temporal and spatial resolution are assessed in phantoms and *in vivo* mouse experiments. The experimental analysis is expected to aid the development of optimized hardware and image acquisition strategies for RUCT, thus result in more affordable imaging systems facilitating wider adoption.

### 3.1 INTRODUCTION

Biomedical pulse-echo ultrasound (US) relies on acoustic impedance variations, giving rise to back-scattered or reflected acoustic waves, in order to visualize deeply embedded structures in biological tissues [1]. The portable and affordable nature of pulse-echo US based on linear or convex array transducers has led to its widespread use in clinical practice as a hand-held diagnostic imaging tool [2, 3]. The inter-element spacing (pitch) of US imaging arrays is typically designed to be around half the acoustic wavelength. In this manner, the emitted US beam can be efficiently shaped along arbitrary directions [4, 5]. Traditional pulse-echo US imaging is based on steering the transmitted beam generated by constructive interference of the waves emitted by individual elements [6], while more recent ultrafast US schemes are based on beamforming with unfocused beams (e.g. plane waves) along different directions [7]. In both cases, the aperture (size) of the array along with its central frequency (and thus the pitch) defines the lateral resolution, so that the number of elements is determined accordingly.

US-based imaging methods can further provide additional information beyond the conventional pulse-echo readings. For example, transmission ultrasound computed tomography (TUCT) maps the speed of sound and acoustic attenuation within the imaged sample enclosed by an US array, which have been shown to provide valuable information for diagnosis of breast cancer in patients [8, 9]. TUCT imaging can be performed by placing several linear transducer arrays around the imaged object [10]. Optoacoustic tomography (OAT) capitalizes instead on US generation via absorption of light in biological tissues, thus synergistically combines rich optical contrast and acoustic resolution unaffected by light scattering and renders high-resolution functional and molecular information complementing anatomical pulse-echo US readings [11–13]. Both TUCT and OAT are inherently tomographic imaging modalities that commonly rely on inverse scattering [14–16] or model-based [17–19] reconstruction algorithms to render accurate and quantitative results. For those methods, US signals must then be collected at different locations surrounding the imaged sample, so that the angular coverage provided by the US array is of key importance for rendering accurate tomographic reconstructions. Pulse-echo US has also been implemented with circular arrays enclosing the object. This approach, termed reflection ultrasound computed tomography (RUCT), has recently been combined with TUCT and OAT to result in a multimodal transmission-reflection optoacoustic ultrasound (TROPUS) imaging platform [20, 21]. Pulse-echo US with concave arrays have also been suggested e.g. for ophthalmic imaging [22], or for monitoring high-intensity focused ultrasound (HIFU) [23]. Combinations of linear and concave arrays have also been used for hybrid OAT and pulse-echo US [24, 25]. Typically, a very large number of elements is required for conforming with an optimal inter-element pitch when using concave arrays, particularly when employing high-frequency circular arrays having a large diameter implied by the design constraints. In practice, the number of elements of the array and associated electronic channels are constrained by the design complexity and cost of the system as well as the required imaging frame rate per given data throughput capacity. These trade-offs could be somewhat mitigated by applying a compressed sensing approach, which was recently explored with synthetically generated data for TUCT [26].

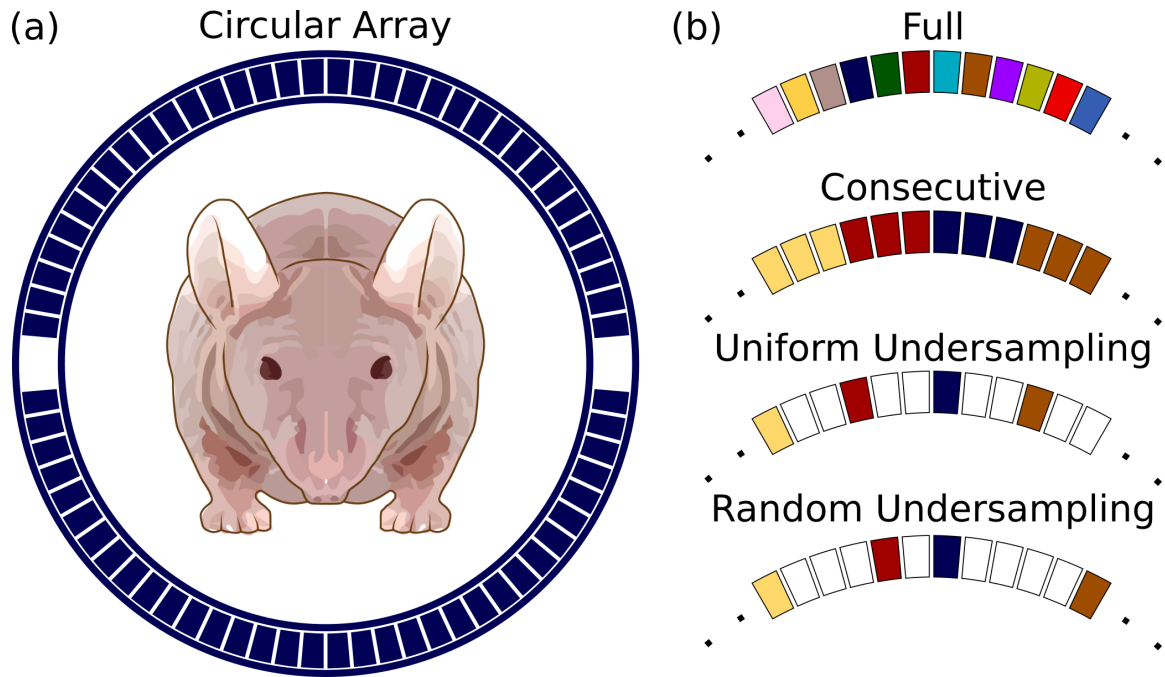


FIGURE 3.1: Imaging setup and transmit/receive schemes. (a) Schematic diagram of the circular array (top view). The animal position is indicated. (b) Active array elements for transmission and/or reception. Different colors indicate independent signals. Top row – all array elements are driven independently. Second row – consecutive elements are driven simultaneously in transmission or grouped (combined) in reception. Third row – Uniform undersampling acquisition scheme with only a few equidistant elements activated in transmission and reception. Bottom row – Sparse undersampling with randomly chosen transmission and reception elements.

In this work, we investigated practical strategies for effective reduction of the number of transmitting/receiving channels in RUCT without compromising the resulting image quality. Specifically, three different approaches are considered for reducing the number of required transmit and receive channels: i) Simultaneous transmission and reception using consecutive transducer elements, ii) Uniform undersampling of transmission and reception channels, iii) Random undersampling of transmission and reception channels. The effects of reducing transmission and reception channels were first investigated separately. Then, the combination of transmission and reception channel reduction was further attempted. Performance of the different acquisition schemes was experimentally validated with tissue phantoms and *in vivo* mouse RUCT data recorded by a 512-channel full-ring US array system. The ultimate goal is to define the optimal approach for reducing the complexity of the array and the number of acquired channels, which in turn will result in less expensive systems and also enable achieving higher frame rates.

## 3.2 MATERIALS AND METHODS

### 3.2.1 IMAGING SYSTEM

RUCT data acquisition was performed with a 40 mm radius full-ring custom-engineered piezocomposite array transducer (Imasonic SAS, Voray, France) (dimensions: 200 mm x 190

mm x 25 mm). It consists of 512 elements distributed with 0.47 mm inter-element pitch along two identical sub-arrays, each covering 174° around the imaged object (Fig. 3.1a). The individual 0.37 mm × 15 mm elements lie on a concave surface, resulting in cylindrical (toroidal) focusing at 38 mm distance, i.e. close to the center of the ring array geometry. The array has central frequency of 5 MHz and 60% transmission/reception bandwidth (at -6 dB) with each element having 0.76 mm beam width at the focus and 13.5 mm focal zone. The sensitivity field and acoustic beam profile of single element was characterized in previous works for an equivalent array [27, 28]. The array is driven by a custom-made data transmission-acquisition system (DAQ, Falkenstein Mikrosysteme GmbH, Taufkirchen, Germany). The DAQ can independently drive all the elements with defined digital waveforms and arbitrary transmission delays. In receive mode, it can perform simultaneous parallel digitization of all the signals collected by the array elements at 12, 24, or 40 mega samples per second (MSPS), which are then transmitted via Ethernet at a maximum transmission rate of 1 Gbit/s to a personal computer for storage and post-processing. In each transmission event, a bipolar 38 V<sub>pp</sub> voltage pulse centered at 5 MHz was transmitted to the individual elements. For all experiments, the transducer and the imaged objects (phantoms and mice) were immersed in a water tank to ensure acoustic coupling. The array was aligned by means of a stage system with four degrees of freedom (x, y, z translation and rotation) to ensure the sample is positioned around the center of the full-ring array (Fig. 3.1a).

### 3.2.2 IMAGE FORMATION

The synthetic transmit aperture (STA) image acquisition method was used to achieve high contrast and resolution [29, 30]. This approach is based on driving each element of the array successively in transmission while receiving simultaneously with neighboring elements. Images from every transmission event were reconstructed individually using the standard delay-and-sum beamforming algorithm [31], rendering low contrast to noise ratio (CNR) images with specular appearance. Pulse-echo US images are generated with the delay-and-sum algorithm by calculating the total travel time (time of flight (TOF)) it takes for each transmitted wave to return back to the transducer after being reflected by the sample. The individual images are then compounded (added up) to form a final RUCT image with high CNR and reduced speckle patterns, whereas the number of compounded images determines the achievable frame rate. Prior to reconstruction, the acquired signals were band-pass filtered between 2.5 and 6 MHz cut-off frequencies. The compounded images are displayed in a logarithmic (dB) scale, namely,  $20\log_{10}(I)$ , being  $I$  the final RUCT image. The image CNR is subsequently calculated via

$$\text{CNR} = \frac{\bar{I}_{\text{signal}} - \bar{I}_{\text{noise}}}{\text{std}(\bar{I}_{\text{noise}})}$$

where  $\bar{I}_{\text{signal}}$  is the mean value of the image intensity for a region within the sample,  $\bar{I}_{\text{noise}}$  is the mean value of the image intensity in a region with background noise and  $\text{std}(\bar{I}_{\text{signal}})$  is the standard deviation of the image intensity in a region with background noise.

### 3.2.3 CHANNEL GROUPING IN TRANSMISSION

We first investigated on the effects of simultaneously driving consecutive elements of the array in transmission, which effectively corresponds to using larger (wider) elements (Fig. 3.1b). This enables increasing the transmitted wave energy hence enabling higher CNR of the beamformed images for each transmission, which in turn results in a reduced number of compounded transmissions and (potentially) a higher effective frame rate. For instance, transmission with  $N_{CT} = 1, 2, 3, 4$  consecutive elements implies 512, 256, 171, 128 transmission events, respectively, in order to complete the full ( $360^\circ$ ) tomographic acquisition, i.e., the acquisition time is reduced by a factor of  $N_{CT}$ . However, the diffraction-associated directivity of the US beam increases with  $N_{CT}$ , consequently reducing the effective field of view (FOV) within the sample.

### 3.2.4 CHANNEL GROUPING IN RECEPTION

We also investigated on the effects of using larger (wider) elements for detecting US echoes. For this, an increasing number of transducer elements were grouped in reception. Specifically,  $N_{CR} = 1$  (full reception, Fig. 3.1b, top) and, 2, 3 and 4 consecutive elements (Fig. 3.1b, second row) were considered. The reception aperture was fixed, corresponding to an angle of  $45^\circ$  in all cases. A  $45^\circ$  reception aperture was used in all cases – since pulse-echo US is mainly based on back-scattered signals, larger reception angles have not resulted in substantial image quality improvement. The received signals from  $N_{CR}$  consecutive elements were averaged to emulate a single integrated waveform detected by a larger element to be used for delay-and-sum beamforming. To analyze the effects resulting only from changes in the size of the receiving transducer (i.e. number of consecutive reception channels  $N_{CR}$ ), a maximum number of transmission events (512) and a minimum number of consecutive transmission elements ( $N_{CT} = 1$ ) were considered. The frame rate before compounding is ultimately limited by the data transfer rate. Therefore, by grouping  $N_{CR}$  channels in reception, it can be increased by a factor of  $N_{CR}$  for a given data throughput capacity of the DAQ. However, much like for transmission, element’s directivity is affected by diffraction, thus the effective FOV is expected to decrease with increasing  $N_{CR}$ .

### 3.2.5 CHANNEL GROUPING IN BOTH TRANSMISSION AND RECEPTION

To further accelerate the data acquisition, we also investigated a configuration where the same number of consecutive transducer elements ( $N_{CT} = N_{CR} = N_{CTR}$ ) were used both at transmission and reception. Specifically,  $N_{CTR} = 1, 2, 3$  and 4 were considered. This approach combines the effects resulting from consecutive element transmission and reception. Therefore, considering full tomographic coverage ( $360^\circ$ ) in transmission and an amount of data in reception limited by the data throughput capacity of the DAQ, the frame rate can be increased by a factor of  $N_{CTR}^2$ , i.e. by 1, 4, 9 and 16 accordingly.

### 3.2.6 SPARSE DATA ACQUISITIONS

The feasibility to form RUCT images with a reduced number of elements was also analyzed by considering sparse (either uniform or random) distribution of elements. For the uniform undersampling case, a certain number of consecutive elements was skipped in transmission and reception (Fig. 3.1b, third row), i.e., only one out of  $N_{\text{STR}} = 1, 2, 3$  or 4 was used. In contrast, random undersampling assigns random single elements for transmission events (Fig. 3.1b, fourth row). Then, reception elements around the transmitting channel were assigned randomly by keeping the reception aperture fixed at  $45^\circ$ . The elements for transmission and reception were selected based on uniform random distributions. The sparse acquisition schemes result in the number of channels reduced by a factor of  $N_{\text{STR}}$ , both for transmissions and receptions. Therefore, much like in the alternative approach based on consecutive elements, the frame rate can be increased by a factor of  $N_{\text{STR}}^2 = 1, 4, 9$  or 16 per given data throughput capacity of the DAQ. The sparse acquisition schemes preserve the small size of individual elements and hence are not expected to be affected by diffraction. Instead, the resulting images are afflicted by spatial undersampling artifacts. Furthermore, the reduced intensity of the emitted signals and the reduced sensitivity in reception is expected to lead to a lower signal-to-noise ratio (SNR) of the acquired signals.

### 3.2.7 PHANTOM EXPERIMENTS

Phantoms were used to evaluate the effects of the element size in transmission and reception as well as to test performance of the proposed RUCT imaging approaches with a reduced number of elements. The first phantom, used to test the effects of US emission with consecutive elements, consisted of a 20 mm cylinder made of agar (1.3% w/v agar powder).  $50\mu\text{m}$  microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ) were embedded uniformly within the phantom, thus mimicking an approximately constant back-scattering background. The second phantom, used to test the effects of signal acquisition with consecutive and sparse acquisition schemes, was formed with parallelly oriented 0.1 mm diameter copper wires lying perpendicular to the imaging plane. The wires are distributed across 20 mm x 20 mm FOV whose dimensions are comparable with the intended *in vivo* experiments. They were attached to two rigid rods located above and below the array and fully immersed in water. The third phantom, used to test the imaging performance, consisted of a 20 mm cylinder made with agar (1.3% w/v agar powder) in which three 6 mm cylindrical insertions were included.  $50\mu\text{m}$  microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ) were embedded uniformly within these three insertions.

### 3.2.8 ANIMAL EXPERIMENTS

A female athymic nude mouse (8 weeks old, Janvier Lab, France) was used for testing the imaging performance *in vivo*. The animal was housed in a ventilated cage inside a temperature-controlled room under a 12-hour dark/light cycle. Pelleted food and water were provided ad-libitum. The mouse was anesthetized with isoflurane (4% v/v for induction and 1.5% for maintenance, Abbott, Cham, Switzerland) in an oxygen/air mixture (100/400

mL/min) during the experiment. The animal was placed in up-right position in the water tank using a custom-made animal holder. The head of the mouse was kept outside the water, and anesthesia was maintained throughout the experiment via a breathing mask. The water temperature was kept stable at 34°C during the experiment with an electrical heater. The imaging array was moved by an electronically-controlled stage in four-dimensional (4D) (x, y, z, and rotation axis) so the mouse was positioned around the center of the circular array geometry. Cross-sectional RUCT images were acquired from the intestinal region of the mouse. Animal experimentation was conducted in accordance with the Swiss Federal Act on Animal Protection and was approved by the Cantonal Veterinary Office Zürich.

### 3.3 RESULTS

We first explored the effects of using consecutive elements for transmission and/or reception, which is equivalent to increasing the size (width) of the elements. RUCT images reconstructed using a single transmission event (no compounding) with increasing numbers of emitting channels ( $N_{CT} = 1, 2, 3, 4$ ) and with  $N_{CR} = 1$  for a 45° angular aperture in reception are presented in Fig. 3.2a. Considering the uniform distribution of acoustic back-scatterers within the phantom, these images basically resemble the emitted US beam shape, whose diffraction-associated directivity increases with the number of consecutive transmission elements. Directivity of the emitted beam affects the maximum FOV that can be acquired without losing the information in the region of interest. The number of consecutively transmitting elements ( $N_{CT}$ ) was limited to 4 as larger number resulted in an inadequately small FOV. The enhanced directivity along with the larger sub-aperture contributes to higher acoustic energy density within a region surrounding the center of the phantom. This is manifested by higher CNR of the RUCT images achieved with a higher number of consecutive transmitting elements (Fig. 3.2b). The directivity of the US beam was further quantified by considering a section of the RUCT images (indicated in green in the fourth column of Fig. 3.2a). The averaged 20 horizontal profiles within this region are shown in Fig. 3.2c. The width of this profile is clearly reduced with increased number of consecutive transmitting channels, which results in a contracted FOV.

The enhanced directional sensitivity of larger elements further affects the detected signals. The effects of using consecutive elements for reception were analyzed with a phantom consisting of wires aligned perpendicularly to the imaging plane. Fig. 3.3a shows the reconstructed RUCT images with a group of  $N_{CR}$  consecutive channels (indicated by the same color) compounded from 512 transmission events with reduced number of reception events. The reduction in reception events results in a lower amount of data being transferred, which in turn increases the achievable frame rate by the factor of  $N_{CR}$ . It is shown that the CNR remains constant for the central wire (circular data points in Fig. 3.3b) when using consecutive reception channels. However, for a wire distant from the central part (triangular data points in Fig. 3.3b), the CNR decreases from 61 (for  $N_{CR} = 1$ ) to 48 (for  $N_{CR} = 4$ ). Horizontal image profiles for the selected wires are also shown in Figs. 3.3c and 3.3d. The profile corresponding to the central line does not change significantly when increasing the

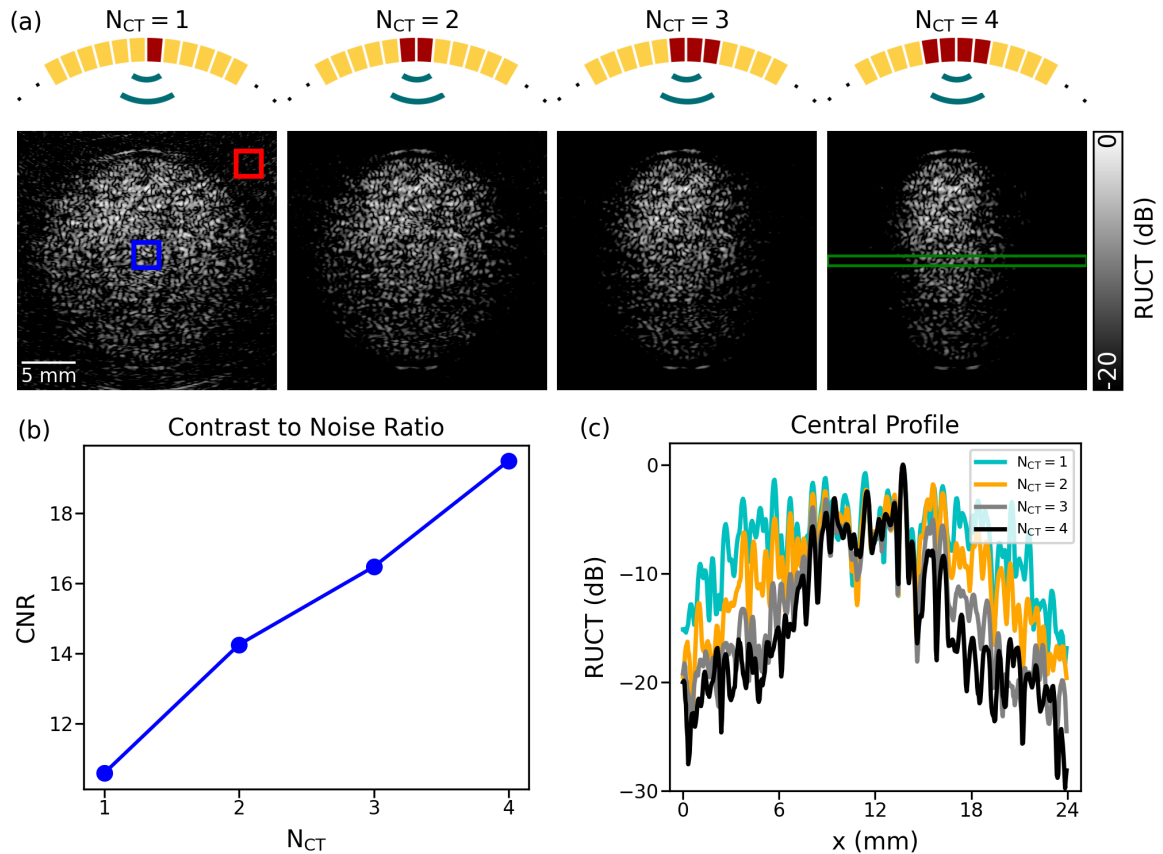


FIGURE 3.2: Effects of element size in transmission analyzed by varying the number of simultaneously transmitting consecutive elements ( $N_{CT}$ ). (a) Schematic representation of the transmission configuration with the positions of transmitting elements of the circular array shown in red. These elements ( $N_{CT} = 1, 2, 3, 4$ ) are driven simultaneously in one transmission event. Reflection ultrasound computed tomography (RUCT) images from a circular phantom with uniformly distributed microspheres show the effects of increasing the element size on the directivity of the transmitted ultrasound beam. (b) Contrast to noise ratio (CNR) of the RUCT images as a function of the number of consecutive transmitting elements ( $N_{CT}$ ). The signal and background regions for the CNR calculation are indicated in (a) with blue and red squares, respectively. (c) Averaged 20 horizontal profiles within the rectangular region indicated in green in (a) as a function of lateral distance for  $N_{CT} = 1, 2, 3, 4$ .

number of consecutive reception elements. However, the profile corresponding to a laterally shifted location shows a significant increase in the side lobe intensities and in the full width at half maximum (FWHM) of the main lobe for a higher number of consecutive elements in reception. This leads to a loss of resolution and contrast in peripheral regions of the FOV.

Next, we evaluated the possibility of speeding up acquisitions by uniform undersampling with one out of  $N_{SR}$  equally spaced reception channels. The images reconstructed with different levels of sparsity in reception ( $N_{SR}$ ), compounded from 512 transmission events, are displayed in Fig. 3.4a. The reduction in number of reception channels results in a  $N_{SR}$ -fold increase in frame rate due to the reduced amount of data acquired for each transmission event. The CNR at central and peripheral regions of the reconstructed images is shown to clearly decrease with increased sparsity level (Fig. 3.4b). The CNR at the central region diminishes with respect



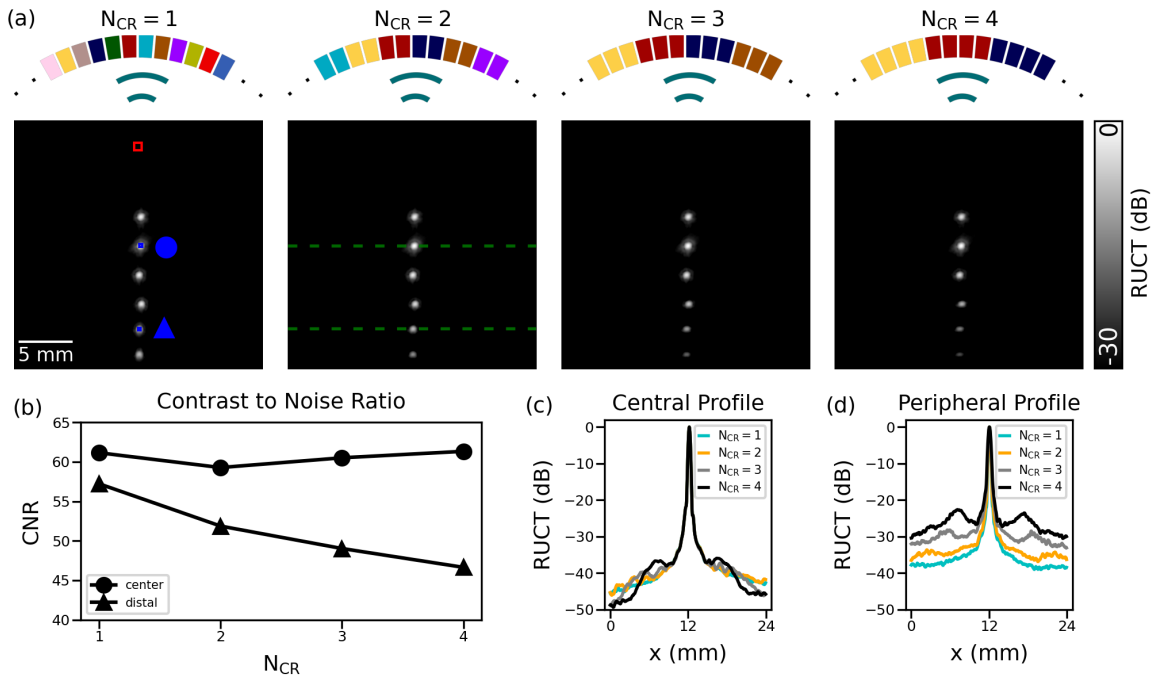


FIGURE 3.3: Effects of varying the number of consecutive reception elements ( $N_{CR}$ ). (a) Groups of reception channels ( $N_{CR} = 1, 2, 3, 4$ ) are indicated with different colors. Compounded (512 transmission events) reflection ultrasound computed tomography (RUCT) images of wires aligned perpendicular to the imaging plane. (b) Contrast to noise ratio (CNR) dependence on  $N_{CR}$ . (c) Horizontal image profile through the central wire as a function of  $N_{CR}$ . (d) Horizontal image profile through a peripheral wire as a function of  $N_{CR}$ .

to what has previously been achieved with consecutive channels in reception. The increased sparsity level in reception ( $N_{SR}$ ) also causes side-lobe artifacts with a defined pattern to appear in the images. Horizontal image profiles at the center and peripheral data points (green dashed lines in Fig. 3.4a) were analyzed to quantify the side lobes corresponding to individual wires. Approximately the same level of side lobe intensity increase was observed in central and peripheral profiles (Figs. 3.4c and 3.4d, respectively). The fact that both the CNR and image profiles remain nearly identical at central and peripheral regions indicates that the imaging performance achieved with this sparse acquisition approach is uniform across the entire FOV.

The effects of reducing the number of reception channels via randomized undersampling were further analyzed (Fig. 3.5). For this, images from the wire phantoms were formed by using the reception sequence corresponding to the random reception element assignments (Fig. 3.5a). Similar to the other acquisitions, the compounded transmission event number was kept constant at 512 by only reducing the number of reception channels. This acquisition scheme resulted in  $N_{SR}$ -fold increase in frame rate by reducing the transferred data volume. The CNR decreased with increasing sparsity level ( $N_{SR}$ ) in both central and peripheral regions (Fig. 3.5b), albeit at a slower pace as compared to the uniform undersampling case. We ascribe this slight increase in CNR with respect to the uniform undersampling case to the lack of pattern noise in the background. Side-lobe artifacts are still present in the images, but random sampling appears to more uniformly distribute these across the background. As a result, no significant increase in side lobes corresponding to individual wires was observed in

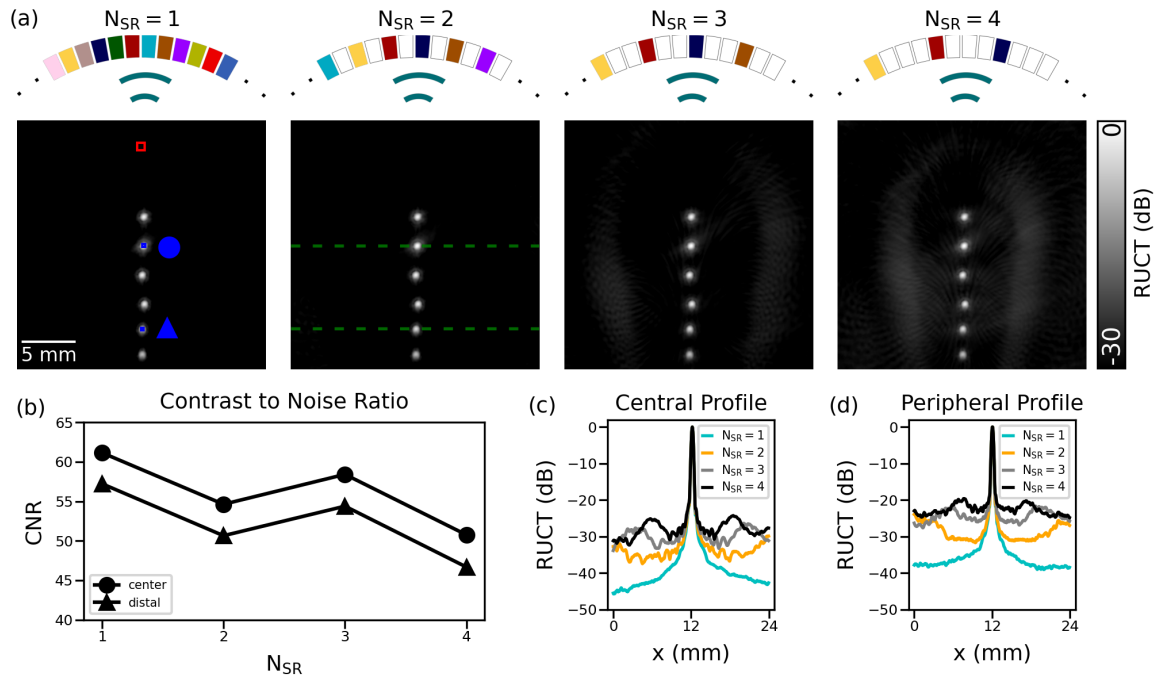


FIGURE 3.4: Effects of uniform undersampling in reception. (a) Different acquisition schemes were evaluated with increasing level of sparsity, i.e.,  $N_{SR} = 1, 2, 3, 4$ . The corresponding compounded reflection ultrasound computed tomography (RUCT) images of wires aligned perpendicular to the imaging plane are shown. (b) Contrast to noise ratio (CNR) as a function of  $N_{SR}$ . (c) Horizontal image profile through the central wire as a function of  $N_{SR}$ . (d) Horizontal image profile through the peripheral wire as a function of  $N_{SR}$ .

the selected horizontal profiles for central and peripheral wires (Figs. 3.5c and 3.5d). On the other hand, much like for the case of uniform undersampling, a similar performance across the entire FOV was observed in terms of CNR and resolution.

To explain the difference between the uniform and random undersampling approaches, one should distinguish between two effects: a global CNR decrease when the number of elements decreases (occurs for both undersampling methods) and the appearance of grading lobes, which increase in size and move towards the center in the case of uniform undersampling only. The latter effect is particularly visible in Figs. 3.4c and 3.4d: small humps progressively appear in the profiles; and in Fig 3.4a, for  $N_{SR} = 3$  and 4: high intensity regions appear and move towards the center, increasing in intensity for higher  $N_{SR}$ . This effect is not visible for random sampling - only an increase of the general background level can be seen without appearance of grading lobes.

The RUCT imaging performance with a reduced number of elements was then tested in a more realistic tissue phantom configuration consisting of three circular insertions with embedded microspheres to mimic various levels of scattering. The RUCT images reconstructed with increasing number of consecutively grouped elements ( $N_{CTR} = 1, 2, 3, 4$ ) enable increasing the frame rate by 1, 4, 9 and 16 fold, respectively, for a given data throughput capacity of the DAQ (Fig. 3.6a). Also, the CNR inside the phantom is preserved with increased number of grouped elements due to increased energy delivered with one transmission event and the

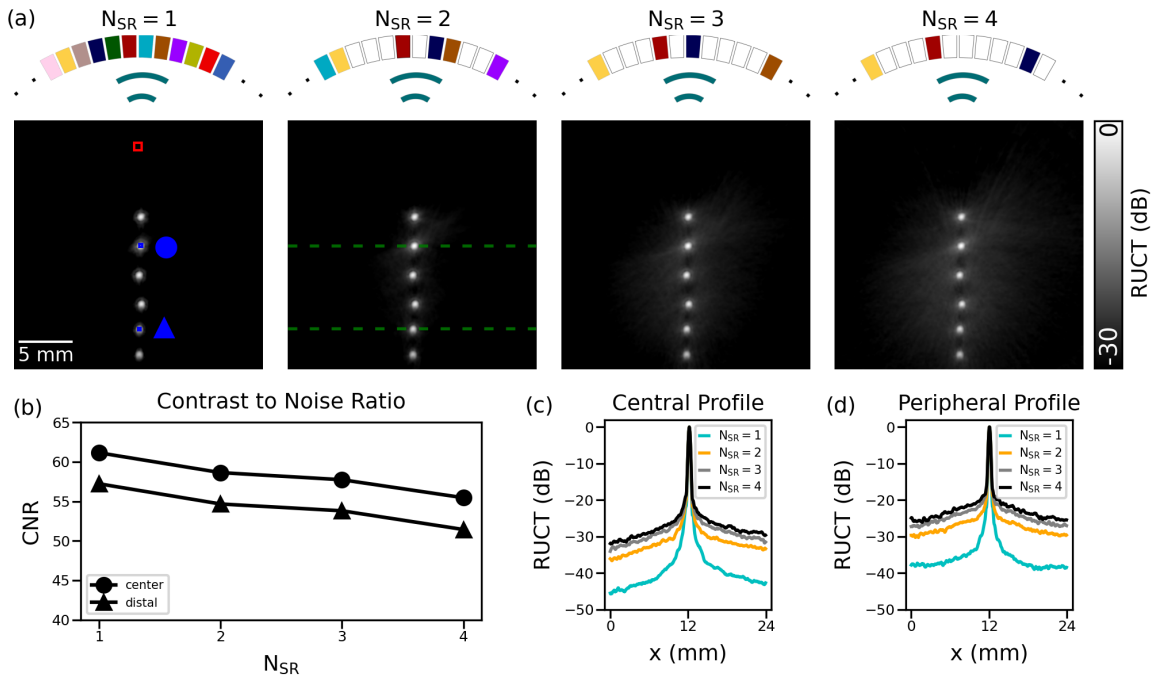


FIGURE 3.5: Effects of random undersampling in reception. (a) Illustration of the various random undersampling schemes having  $N_{SR} = 1, 2, 3, 4$  with the corresponding compounded reflection ultrasound computed tomography (RUCT) images of wires aligned perpendicular to the imaging plane. (b) Contrast to noise ratio (CNR) as a function of  $N_{SR}$ . (c) Horizontal image profile through the central wire as a function of  $N_{SR}$ . (d) Horizontal image profile through the peripheral wire as a function of  $N_{SR}$ .

enhanced detection sensitivity (circular data points in Fig. 3.6d). Note, however, that the image contrast diminishes in the peripheral regions of the phantom when grouping several consecutive elements. For the sparse acquisition methods (Figs. 3.6b and 3.6c), the increase in sparsity level  $N_{STR}$  resulted in diminished CNR (triangular and asterisk data points in Fig. 3.6d), which was more pronounced in the uniform undersampling case. However, contrary to the consecutive element grouping approach that results in non-uniform CNR across the imaged FOV, good image contrast uniformity was observed throughout the entire phantom volume with the sparse acquisition approach (Fig. 3.6e). The speckle patterns in the images obtained with random undersampling appears to be lower as compared to uniform undersampling, thus resulting in an enhanced CNR. This indicates that speckle patterns are minimized with random undersampling similarly to the side-lobe pattern artifacts corresponding to isolated wires.

The RUCT imaging performance with a reduced number of elements was then tested *in vivo* by acquiring cross-sectional images of the intestinal region of a healthy female athymic nude mouse. The images reconstructed with consecutively grouped elements (Fig. 3.7a) and sparse acquisition sequences (Fig. 3.7b and 3.7c) were compared per given frame rate (data throughput). CNR values were calculated by measuring the signal around the transverse colon area (blue square in Fig. 3.7a) with the background noise calculated in a region outside the mouse (red square in Fig. 3.7a). Much like for the phantom images, the CNR values remained relatively stable when using consecutive elements, dropping from 28 for  $N_{CTR} = 1$ ,

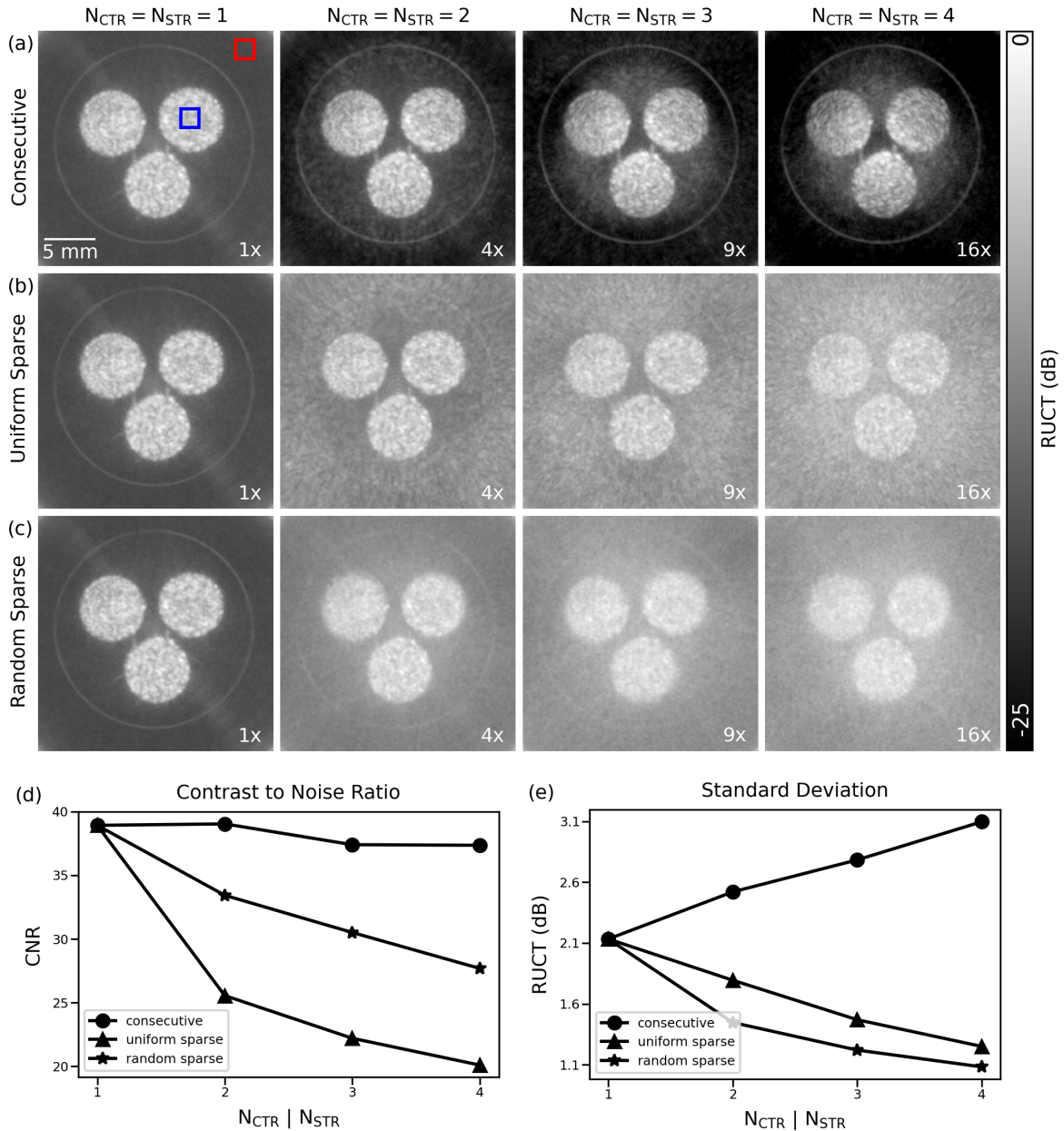


FIGURE 3.6: Reflection ultrasound computed tomography (RUCT) phantom images acquired with consecutive elements and with sparse acquisitions as a function of the number of consecutive elements grouped in transmission/reception ( $N_{CTR}$ ) and the sparsity level ( $N_{STR}$ ). The test phantom consisted of a 20 mm diameter cylinder made of agar (1.3% w/v agar powder) embedded with three 6 mm diameter cylindrical insertions filled with a uniform distribution of 50 $\mu$ m-sized microparticles (Cospheric BKPMS-1.2 45-53 $\mu$ m). The increase in frame rate for a given data throughput capacity is shown in the bottom right corner of the images. (a) Compounded images as a function of the number of consecutive elements ( $N_{CTR} = 1, 2, 3, 4$ ). (b) Compounded images from uniform undersampling as a function of the sparsity level ( $N_{STR} = 1, 2, 3, 4$ ). (c) Compounded images from random undersampling as a function of the sparsity level ( $N_{STR} = 1, 2, 3, 4$ ). (d) Contrast to noise ratio (CNR) of the images as a function of  $N_{CTR}$  and  $N_{STR}$ . The signal and background regions for the CNR calculations are indicated in panel a with blue and red squares, respectively. (e) Standard deviation of image intensity across the FOV as a function of  $N_{CTR}$  and  $N_{STR}$ .

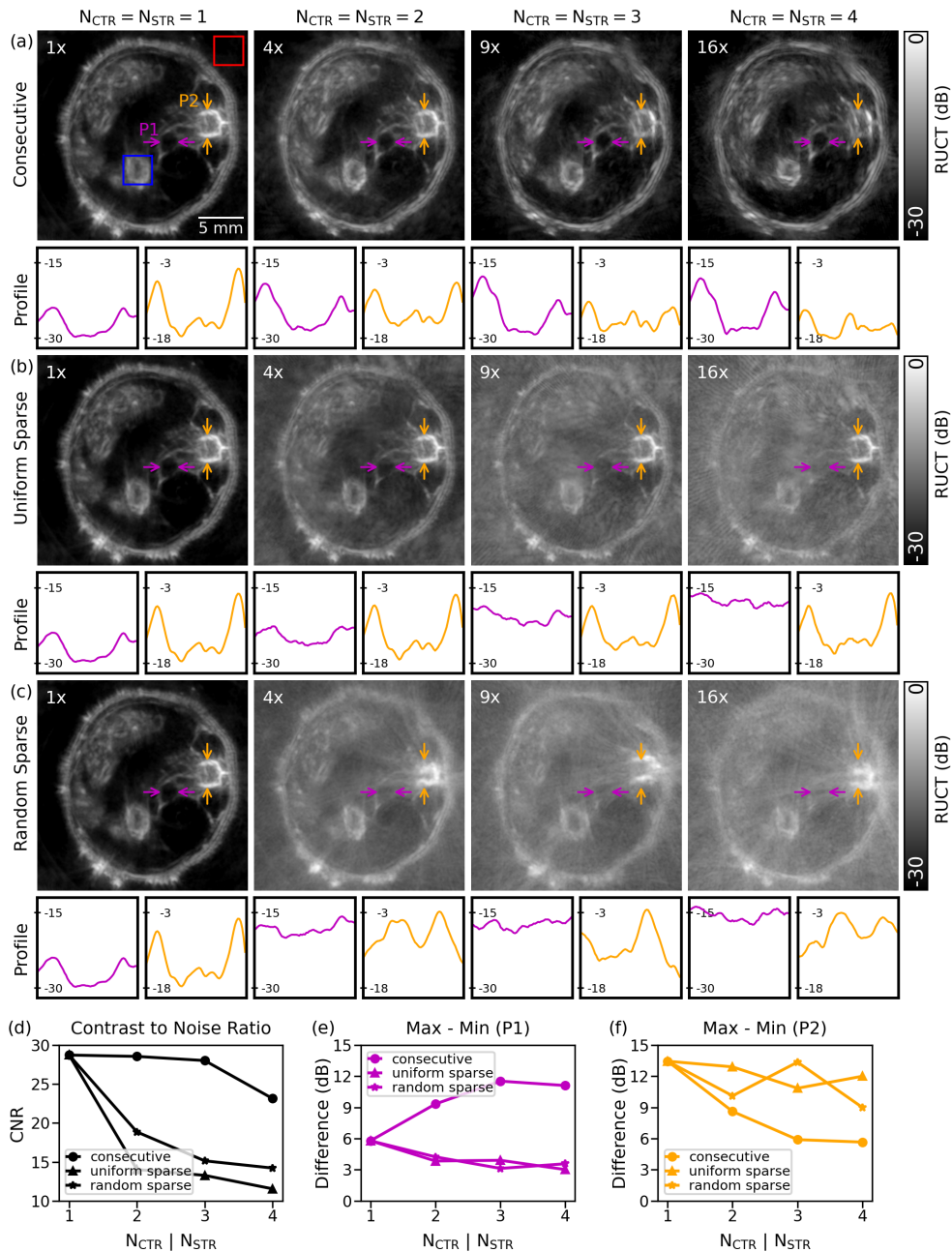


FIGURE 3.7: Comparison between *in vivo* reflection ultrasound computed tomography (RUCT) images acquired with the consecutive and undersampling schemes. A female athymic nude mouse (8 weeks old) was imaged *in vivo*. The achievable increase in frame rate for a given data throughput capacity of the data transmission-acquisition system (DAQ) is shown in upper left corner of the images. (a) RUCT images acquired with simultaneously driven consecutive elements ( $N_{CTR} = 1, 2, 3, 4$ ). (b) RUCT images acquired with uniform sparse acquisition as a function of sparsity levels ( $N_{STR} = 1, 2, 3, 4$ , respectively). (c) RUCT images acquired with random sparse acquisition as a function of sparsity levels ( $N_{STR} = 1, 2, 3, 4$ , respectively). (d) Change in contrast to noise ratio (CNR) calculated on the transverse colon using consecutive element and sparse acquisition methods. (e) Difference between minimum and maximum values of line profile of descending colon (P1) for the consecutive and undersampling acquisition schemes. (f) Difference between minimum and maximum values of line profile of vertebral column (P2) for the consecutive and undersampling acquisition schemes. Signal profiles through the descending colon (P1) and vertebral column (P2) are shown under the images in panels a-c.

2, 3 to 23 for  $N_{\text{CTR}} = 4$  (circular data points in Fig. 3.7d). On the contrary, increasing the undersampling levels has seen the CNR decreasing substantially, from 28 for  $N_{\text{STR}} = 1$  to 12 for  $N_{\text{STR}} = 4$  (Fig. 3.7d – triangular data points). Likewise, the CNR dropped from 28 to 14 for the random undersampling approach (Fig. 3.7d – asterisk data points). The observed CNR changes further lead to loss of resolution, as evinced from the one-dimensional (1D) signal profiles plotted under the images in Fig. 3.7. The difference between maximum and minimum values of signal profiles P1 and P2 are shown in Fig. 3.7e and 3.7f, respectively, to better visualize the change in contrast. The shape of the descending colon as well as contrast of its boundaries (P1 in Figs. 3.7a and 3.7e) remain well preserved when imaging with increasing number of consecutive elements ( $N_{\text{CTR}}$ ). On the contrary, uniform and random sparse acquisitions result in loss of resolving capacity with increased sparsity level ( $N_{\text{STR}}$ ). The image quality is affected differently when analyzing signal profile through the vertebral column located peripheral from the center (P2 in Fig. 3.7a). Here consecutive element acquisition results in substantial loss of resolving power (P2 in Figs. 3.7a and 3.7f), which remains less affected for the undersampling acquisition schemes (Figs. 3.7b, 3.7c and 3.7f). In the *in vivo* example presented here, the two structures P1 and P2 remain visible and undistorted for undersampling levels as high as  $N_{\text{CTR}} = 3$  or  $N_{\text{STR}} = 2$ , which respectively correspond to 170 or 256 independent channels. These numbers are compatible with commercially available preclinical scanners, such as the Vantage-256<sup>TM</sup> from Verasonics. Note that the vertebral column (P2) appears to be distorted when considering random undersampling. This appears to corroborate that random undersampling contributes to removing speckle patterns, which is undesirable in the present case as these may correspond to actual structures in the images.

### 3.4 DISCUSSION

In this work, different strategies to reduce the number of transmit/receive elements in RUCT were investigated. Optimal trade-off between number and size of array elements and RUCT imaging performance for the given target size and resolution have been determined. Two main performance metrics were considered, namely, the image contrast, primarily driven by the total active area of the transducer, and the FOV, which is chiefly influenced by the directivity of individual elements. Note that the spatial resolution is mainly determined by the detection bandwidth rather than by the number of active elements. Overall, it was found that the CNR is maintained in a region close to the center of the array when consecutive elements are grouped together mimicking a dense array with a lower number of elements. Indeed, a lower number of emission and reception events is compensated with a higher acoustic energy density due to the enhanced directivity and increased sub-aperture size. The resolution in peripheral regions can be restored by employing advanced algorithms such as pulse compression and Golay approaches [32–35]. On the other hand, channel undersampling (both uniform and random) resulted in a reduced CNR but a more uniform performance in central and peripheral regions, which contributes to a larger effective FOV. It was observed that random undersampling appears to contribute to removing speckle patterns in the images. This scheme was thus shown to enhance CNR by minimizing side-lobe pattern artifacts and speckle patterns but also resulted in distortion of actual structures in the images. Channel

undersampling can be further optimized by employing advanced randomization schemes [36].

The optimal approach for expediting image acquisition in RUCT must then be defined according to the application of interest. For instance, tomographic imaging of the human breast requires a large FOV attained with a dense circular array of detectors located in close proximity to the skin surface. In this case, an optimal imaging speed acceleration is achieved with a randomized sparse undersampling method. On the contrary, a typical mouse imaging RUCT scanner features a much larger ratio between the array diameter and the effective FOV, thus an acquisition scheme with a reduced number of large-sized elements is arguably preferred to maximize the CNR. Note that the FOV also scales with the frequency bandwidth of the elements, which is generally lower for arrays designed to image larger samples. The final number of independently driven channels must then be selected by maximizing the various imaging performance aspects pertaining contrast, resolution and FOV but also an overall system's cost and its anticipated image acquisition time (imaging speed).

A low number of independent channels contributes to reducing the complexity of both the US array and the accompanying data transmission/acquisition electronics, making the system more portable and affordable thus facilitating its dissemination among biologists and clinicians. For a given limit of the data throughput capacity, reducing the number of active transducer elements and of compounded transmission events in each frame further enhances the temporal resolution of the system, thus facilitating visualization of dynamic events such as blood flow and contrast agent perfusion [37]. High-frame-rate imaging is also essential in super-resolution imaging where tracking of individual microbubbles has been shown to massively enhance the spatial resolution of US-based angiography [38, 39]. The tracking accuracy and maximum admissible velocity of a given microbubble (e.g. to measure blood flow) is directly related to the frame rate. A higher imaging rate would thus allow the propagation of RUCT to many more promising applications, especially if the loss in image quality can be compensated for with advanced reconstruction methods.

Finally, the full-ring array configuration used in RUCT is particularly attractive for TROPUS imaging combining RUCT, TUCT and OAT in a single system [20, 21]. Such multimodal integration provides unique capabilities for anatomical, functional and molecular imaging of various disease models, e.g. in the context of multi-parametric tumor detection and characterization [21, 40]. Previous studies have shown that the overall considerations pertaining the size and number of transducer elements in OAT are analogous to what was observed herein for RUCT. For instance, the element directivity similarly confines the FOV in OAT while decreasing the contrast and resolution in peripheral regions [17, 41]. The reduction in the effective FOV is slightly lower in OAT due to the uni-directional propagation of US waves. On the other hand, the tomographic inversion in TUCT is very similar to the extensively studied X-Ray computed tomography (CT) problem where spatial undersampling is known to result in so-called streak artifacts reducing the quality and resolution of the images [42]. Much like for X-Ray CT, the use of model-based image reconstruction methods based

---

on total variation (TV) regularization has facilitated rendering accurate TUCT images with randomized transmissions [26].

The results presented in this work were based on delay-and-sum beamforming. Yet, the derived conclusions are expected to hold for other image formation methods, such as more advanced model-based reconstruction methods commonly employed for reducing undersampling artifacts in OAT [17–19]. Recently, model-based methods have also been suggested for improving image quality in pulse-echo US [43, 44], hence their deployment for RUCT may potentially further enhance the achievable FOV, resolution and contrast. Such methods are for instance expected to perform particularly well when exploiting the image-domain sparsity for implementing accurate compressed-sensing-based methods [45].

### 3.5 CONCLUSION

In this work, we have experimentally analyzed and established the trade-offs between number, distribution and size of array elements and performance of RUCT imaging for the given target size and resolution. Specifically, it was shown that a reduced number of large elements facilitates preserving imaging performance at the central part of the image to the detriment of resolution and contrast at peripheral regions. On the other hand, a sparse (either uniform or random) distribution of small elements results in a more uniform performance across the field of view with reduced contrast. The experimental analysis is expected to aid in the development of optimized hardware and image acquisition strategies for RUCT, thus result in more affordable imaging systems and facilitate a wider adoption of RUCT as an imaging tool.



## ACKNOWLEDGEMENT

This work was supported by Swiss Data Science Center grant C19-04.

## CODE AVAILABILITY

Delay and sum beamformer algorithm can be found here:

<https://github.com/berkanlafci/pyruct>

Detailed documentation of the code can be found here:

<https://berkanlafci.github.io/pyruct/>

## DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors declare that they have no competing financial interests.

## AUTHORS' CONTRIBUTION

Conception and design: B. Lafci, J. Robin, X. L. Deán-Ben, D. Razansky

Development of methodology: B. Lafci, J. Robin, X. L. Deán-Ben, D. Razansky

Acquisition of data: B. Lafci, J. Robin

Analysis of data: B. Lafci

Writing, review, and revision of the manuscript: B. Lafci, J. Robin, X. L. Deán-Ben, D. Razansky

Study supervision: D. Razansky

## BIBLIOGRAPHY

- [1] Jørgen Arendt Jensen. “Medical ultrasound imaging.” In: *Progress in biophysics and molecular biology* 93 1-3 (2007), pp. 153–65.
- [2] Alessandro Ramalli et al. “Design, implementation, and medical applications of 2-D ultrasound sparse arrays”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* (2022), pp. 1–1.
- [3] Ramona De Luca et al. “Diagnostic ultrasound probes: a typology and overview of technologies”. In: *Current Directions in Biomedical Engineering* 4.1 (2018), pp. 49–53. DOI: doi:10.1515/cdbme-2018-0013.
- [4] Paul D. Wilcox and Jie Zhang. “Quantification of the Effect of Array Element Pitch on Imaging Performance”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 65.4 (2018), pp. 600–616.
- [5] Hideyuki Hasegawa and Chris L. de Korte. “Impact of element pitch on synthetic aperture ultrasound imaging”. In: *Journal of Medical Ultrasonics* 43.3 (2016), pp. 317–325.
- [6] Jeff Powers and Frederick Kremkau. “Medical ultrasound systems”. In: *Interface Focus* 1.4 (2011), pp. 477–489.
- [7] Gabriel Montaldo et al. “Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 56.3 (2009), pp. 489–506.
- [8] Cuiping Li et al. “In vivo Breast Sound-Speed Imaging with Ultrasound Tomography”. In: *Ultrasound in Medicine & Biology* 35.10 (2009), pp. 1615–1628.
- [9] Nicolas Vinard et al. “Optimized transducer configuration for ultrasound waveform tomography in breast cancer detection”. In: *Proc. SPIE 10580, Medical Imaging 2018: Ultrasonic Imaging and Tomography*. Mar. 2018, p. 13.
- [10] Yijun Cheng et al. “Ultrasound tomography reconstruction algorithm using two opposite rotatable linear arrays for breast cancer diagnosis”. In: *AIP Advances* 11.5 (2021), p. 055124.
- [11] Elena Merčep et al. “Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.9 (2015), pp. 1651–1661.
- [12] Geng-Shi Jeng et al. “Real-time interleaved spectroscopic photoacoustic and ultrasound (PAUS) scanning with simultaneous fluence compensation and motion correction”. In: *Nature Communications* 12.1 (2021), p. 716.
- [13] Kalloor Joseph Francis et al. “Tomographic imaging with an ultrasound and LED-based photoacoustic system”. In: *Biomedical optics express* 11.4 (2020), pp. 2152–2165.

- 
- [14] Alexander Velichko, Eduardo Lopez Villaverde, and Anthony J. Croxford. “Local scattering ultrasound imaging”. In: *Scientific Reports* 11.1 (2021), p. 993.
- [15] J. Wiskin et al. “Non-linear inverse scattering: high resolution quantitative breast tissue tomography”. In: *The Journal of the Acoustical Society of America* 131.5 (2012), pp. 3802–3813.
- [16] James Wiskin et al. “Full wave 3D inverse scattering transmission ultrasound tomography in the presence of high contrast”. In: *Scientific reports* 10.1 (2020), pp. 20166–20166.
- [17] Christian Lutzweiler, Xosé Luís Deán-Ben, and Daniel Razansky. “Expediting model-based optoacoustic reconstructions with tomographic symmetries”. In: *Medical Physics* 41.1 (2014), p. 013302.
- [18] Daniel Queirós et al. “Modeling the shape of cylindrically focused transducers in three-dimensional optoacoustic tomography”. In: *Journal of Biomedical Optics* 18.7 (2013), pp. 1–13.
- [19] Lu Ding, Xosé Luís Deán-Ben, and Daniel Razansky. “Efficient 3-D Model-Based Reconstruction Scheme for Arbitrary Optoacoustic Acquisition Geometries”. In: *IEEE Transactions on Medical Imaging* 36.9 (2017), pp. 1858–1867.
- [20] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [21] Berkan Lafci et al. “Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound”. In: *Neoplasia* 22.12 (2020), pp. 770–777.
- [22] Jung Hyui Cha et al. “A 15-MHz 1-3 piezocomposite concave array transducer for ophthalmic imaging”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 62.11 (2015), pp. 1994–2004.
- [23] Yayun Wan and Emad S. Ebbini. “Imaging with concave large-aperture therapeutic ultrasound arrays using conventional synthetic-aperture beamforming”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 55.8 (2008), pp. 1705–1718.
- [24] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. “Combined pulse-echo ultrasound and multispectral optoacoustic tomography with a multi-segment detector array”. In: *IEEE transactions on medical imaging* 36.10 (2017), pp. 2129–2137.
- [25] Xue Wang et al. “Integrated thermoacoustic and ultrasound imaging based on the combination of a hollow concave transducer array and a linear transducer array”. In: *Physics in Medicine & Biology* 66.11 (2021), p. 115011.
- [26] Ruud van Sloun et al. “Compressed Sensing for Ultrasound Computed Tomography”. In: *IEEE Transactions on Biomedical Engineering* 62.6 (2015), pp. 1660–1664.

- [27] Daniel Razansky, Andreas Buehler, and Vasilis Ntziachristos. “Volumetric real-time multispectral optoacoustic tomography of biomarkers”. In: *Nature Protocols* 6.8 (2011), pp. 1121–1129.
- [28] Anton Bychkov et al. “Toroidally focused sensor array for real-time laser-ultrasonic imaging: The first experimental study”. In: *Photoacoustics* 17 (2020), pp. 100–160.
- [29] Jørgen Arendt Jensen et al. “Synthetic aperture ultrasound imaging”. In: *Ultrasonics* 44 (2006), e5–e15.
- [30] J.T. Ylitalo and H. Ermert. “Ultrasound synthetic aperture imaging: monostatic approach”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 41.3 (1994), pp. 333–339.
- [31] Jørgen Grythe. “Beamforming algorithms - beamformers”. In: *Online* (2016).
- [32] Jenho Tsao and Ming-Huang Chen. “An Adaptive Pulse Compression Filter for Ultrasound Contrast Harmonic Imaging”. In: *Engineering* 5 (2013), pp. 118–122.
- [33] Vera Behar and Dan Adam. “Parameter optimization of pulse compression in ultrasound imaging systems with coded excitation”. In: *Ultrasonics* 42.10 (2004), pp. 1101–1109.
- [34] Cheng Jin et al. “A new scheme of coded ultrasound using Golay codes”. In: *Journal of Zhejiang University SCIENCE C* 11.6 (2010), pp. 476–480.
- [35] Xiaochun Wang et al. “Research on Golay-coded excitation in real-time imaging of high frequency ultrasound biomicroscopy”. In: *Scientific Reports* 11.1 (2021), p. 1848.
- [36] Miguel Bernal, Daniel Rohrbach, and Ron Daigle. “Random aperture optimization for SRAC in high frame rate volume imaging”. In: *2021 IEEE International Ultrasonics Symposium (IUS)*. 2021, pp. 1–4.
- [37] Xosé Luís Deán-Ben et al. *Noninvasive three-dimensional optoacoustic localization microangiography of deep tissues*. 2020. arXiv: 2007.00372 [physics.bio-ph].
- [38] Claudia Errico et al. “Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging”. In: *Nature* 527.7579 (2015), pp. 499–502.
- [39] Charlie Demené et al. “Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients”. In: *Nature Biomedical Engineering* 5.3 (2021), pp. 219–228.
- [40] R. Gerhard Pratt et al. “Sound-speed and attenuation imaging of breast tissue using waveform tomography of transmission ultrasound data”. In: *Medical Imaging 2007: Physics of Medical Imaging*. Vol. 6510. International Society for Optics and Photonics. SPIE, 2007, pp. 1523–1534.
- [41] Lu Ding, Daniel Razansky, and Xosé Luís Deán-Ben. “Model-Based Reconstruction of Large Three-Dimensional Optoacoustic Datasets”. In: *IEEE Transactions on Medical Imaging* 39.9 (2020), pp. 2931–2940.
- [42] Jiang Hsieh. “Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise”. In: *Medical Physics* 25.11 (1998), pp. 2139–2147.

- [43] Beatrice Berthon et al. “An integrated and highly sensitive ultrafast acoustoelectric imaging system for biomedical applications”. In: *Physics in Medicine & Biology* 62.14 (2017), pp. 5808–5822.
- [44] Justine Robin et al. “Dual-Mode Volumetric Photoacoustic and Contrast Enhanced Ultrasound Imaging with Spherical Matrix Arrays”. In: *IEEE Transactions on Medical Imaging* (2021), pp. 1–1.
- [45] Hervé Liebgott et al. “Compressive sensing in medical ultrasound”. In: *2012 IEEE International Ultrasonics Symposium*. 2012, pp. 1–6.

## 4 SIGNAL DOMAIN LEARNING APPROACH FOR OPTOACOUSTIC IMAGE RECONSTRUCTION FROM LIMITED VIEW DATA

This chapter includes the following publication:

Proceedings of The 5th International Conference on Medical Imaging with Deep Learning, PMLR 172, Pages. 1173-1191 (2022).

Authors:

Anna Klimovskaia Susmelj<sup>1,†</sup>, **Berkan Lafci**<sup>2,3,†</sup>, Firat Ozdemir<sup>1</sup>, Neda Davoudi<sup>2,3</sup>, Xosé Luís Deán-Ben<sup>2,3</sup>, Fernando Perez-Cruz<sup>1,4</sup>, Daniel Razansky<sup>2,3</sup>

<sup>1</sup> Swiss Data Science Center, ETH Zürich and EPFL, Zürich, Switzerland

<sup>2</sup> Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>3</sup> Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>4</sup> Institute for Machine Learning, Department of Computer Science, ETH Zurich, Switzerland

† equal contribution

## ABSTRACT

Multispectral optoacoustic tomography (MSOT) relies on optical excitation of tissues with subsequent detection of the generated ultrasound waves. Optimal image quality in MSOT is achieved by detection of signals from a broad tomographic view. However, due to physical constraints and other cost-related considerations, most imaging systems are implemented with probes having limited tomographic coverage around the imaged object, such as linear array transducers often employed for clinical ultrasound (US) imaging. MSOT image reconstruction from limited-view data results in arc-shaped image artifacts and disrupted shape of the vascular structures. Deep learning methods have previously been used to recover MSOT images from incomplete tomographic data, albeit poor performance was attained when training with data from simulations or other imaging modalities. We propose a two-step method consisting of i) style transfer for domain adaptation between simulated and experimental MSOT signals, and ii) supervised training on simulated data to recover missing tomographic signals in realistic clinical data. The method is shown capable of correcting images reconstructed from sub-optimal probe geometries using only signal domain data without the need for training with ground truth (GT) full-view images.

---

## 4.1 INTRODUCTION

Multispectral optoacoustic tomography (MSOT) is a hybrid biomedical imaging modality based on optical excitation (thermal expansion) of biological tissues followed by detection of the generated ultrasound (US) waves. Interpretation and quantification of the MSOT data is often hampered by poor tomographic coverage provided by the common clinical imaging system implementations. Optimal tomographic inversion in MSOT implies recording sufficient information from the generated ultrasonic wave field with broad angular tomographic coverage [1]. Yet, full tomographic coverage is generally not possible due to physical constraints and other cost-related considerations.

The potential value of MSOT as a clinical imaging tool can be further enhanced when combined with pulse-echo (reflection) US imaging, which provides important reference anatomical information [2, 3]. However, such combination is not straightforward cause MSOT and US imaging impose different, often contradictory, constraints on the transducer array design, such as the pitch size or element directivity in transmission and reception. Linear arrays are frequently used in commercial hand-held US scanners and have further been suggested for implementation of hybrid optoacoustic ultrasound (OPUS) scanners [4, 5]. Linear arrays are easy to manufacture and well-established guidelines exist for the interpretation of the generated US images, offering clear advantages in the clinical setting. On the other hand, acquisition of MSOT images with linear arrays typically results in elongated vessel structures and arc-shaped limited-view artifacts, thus making image interpretation and quantification difficult [6].

Recently, specialized transducer geometries have been suggested for optimal implementation of hybrid OPUS scanners, such as multisegment transducer arrays incorporating both linear and concave array segments. In this way, the limited-view MSOT problem can be partially mitigated with image quality restored to a certain degree (Fig. 4.1A) [7]. In the multisegment configuration, the linear part renders standard US images while all elements contribute to an increased angular coverage for MSOT image reconstruction [8]. While providing optimal image quality in both MSOT and US modes, the manufacturing process of an array of this type is relatively complex. Furthermore, hand-held scanning of certain parts of the human body is also hampered with this array geometry due to the need for a customized water coupling approach [9].

A possible approach to resolve these unavoidable trade-offs is to make use of the optimized multisegment array configuration for the development of learning-based methods toward improving image quality in both MSOT and US imaging modes. In this work, we approach the hybrid image reconstruction problem by considering the raw signal domain data instead of dealing directly with the reconstructed images. With this, we aim to capture spatial and temporal correlations between the transducer array elements in latent space. A learning-based method trained on simulated data may be used to complete the missing view angles in the linear array recordings. To this end, deep learning methods have been used to par-



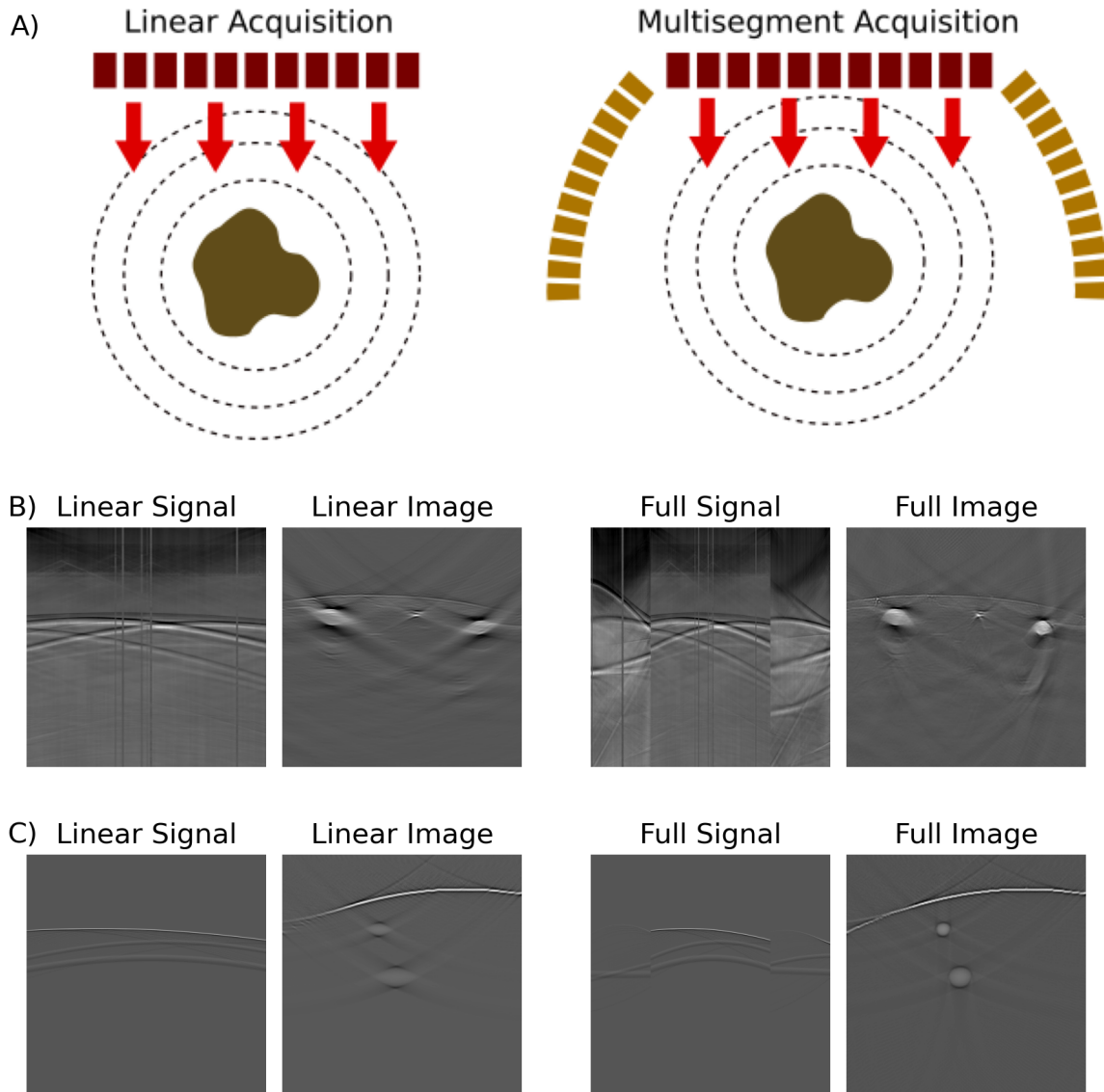


FIGURE 4.1: Handheld multispectral optoacoustic tomography (MSOT) imaging with linear and multisegment array configurations. A) Schematic diagram of the array geometries. The excitation light beam and generated ultrasound (US) waves are represented with red arrows and dashed circles, respectively. B) Raw data (time-resolved signals) along with the reconstructed MSOT images corresponding to a hand-held scan of the human arm at 1064 nm excitation wavelength with the linear and multisegment arrays, respectively. C) Simulated signals along with the reconstructed images for the linear and multisegment arrays, respectively.

tially restore quality of the MSOT images reconstructed from incomplete tomographic data. However, poor performance was achieved when training with data from simulations or other imaging modalities, which appears to be a result of the large domain gap between simulated and experimental data [10]. Thereby, we propose a two-step approach consisting of i) style transfer for domain adaptation between simulated and experimental MSOT signals, and ii) semi-supervised training on simulations from multisegment array geometry and experimental data from linear array to recover missing signals in experimental multisegment array data. We formulate the domain adaptation problem as an unpaired image translation between simulated and experimental signals. Signals detected by elements of the concave segments

are then estimated from the data provided by the linear part of the array using only simulated data after domain adaptation. Once the missing parts have been recovered, the MSOT image reconstruction can be performed with standard methods such as back-projection or model-based algorithms [11–13]. The main contributions of this paper with respect to prior art are summarized as follows: i) the MSOT limited-view problem is tackled in the signal domain instead of the image domain, ii) a method to reduce domain gap between simulated and experimental data is proposed, and iii) a learning-based method is used for estimating the signals from missing detection elements without the need for ground truth (GT) data from sophisticated and costly array configurations (e.g. multisegment array). To the best of our knowledge, this work constitutes the first attempt to address the limited-view MSOT problem in the raw signal domain using simulated data.

## 4.2 METHODOLOGY

### 4.2.1 STYLE TRANSFER NETWORK

The first step of the proposed method represents the key component to reduce the domain gap between simulated and experimental data. The main differences between simulated and experimental domains result from difficulties in mimicking complicated acoustic pressure wavefields corresponding to actual anatomical structures and from the lack of realistic noise components in simulated data. Herein, we use a style transfer network to reduce the distance between simulated and experimental datasets. The simulated and experimental domains are denoted as X and Y, respectively. The corresponding training samples are denoted as  $\{x_i\}_{i=1}^N$ ,  $\{y_i\}_{i=1}^M$  where N and M are the number of samples from simulated and experimental data, respectively. The signals from both domains are combined into one dataset alongside their labels, so that each signal is represented by a pair  $\mathcal{D} = \{(s_i, l_i)\}_{i=1}^{N+M}$ , where  $s_i$  is the signal itself and  $l_i$  is a label indicating if the signal belongs to the simulated ( $l_i = 0$ ) or experimental ( $l_i = 1$ ) domains.

An encoder-decoder architecture was used with domain-adversarial training on the latent space from fader networks [14] after minor modifications. As demonstrated in Fig. 4.2A, the model consists of two convolutional neural networks; an encoder  $E_{\theta_{\text{enc}}}$  and a decoder  $D_{\theta_{\text{dec}}}$ , a latent space discriminator and an additional fully connected discriminator network  $D_{\theta_{\text{disc}}}$  for the adversarial training to ensure domain alignment. The encoder  $E_{\theta_{\text{enc}}}$  takes as input the signal two-dimensional (2D) representation  $s_i$  and produces the latent representation  $z_i = E_{\theta_{\text{enc}}}(s_i)$ . The decoder  $D_{\theta_{\text{dec}}}$  takes as input the latent invariant representation  $z_i$  and corresponding label of the input domain  $l_i$  to produce the reconstructed signal  $\hat{s}_i$ . We use mean absolute error (MAE) as the reconstruction error as it is more suitable for medical imaging problems since it produces sharper images:

$$\ell_{\text{MAE}} = \frac{1}{N + M} \sum_{(s,l) \in \mathcal{D}} \|D_{\theta_{\text{dec}}}(E_{\theta_{\text{enc}}}(s), l) - s\|_1. \quad (4.1)$$

The decoder takes as an input latent representation and a label. If the label is “experimental”,

a latent vector is sampled from a Gaussian distribution with mean and variance parametrized by embedding layers in the decoder, and then a convolutional network is applied to it. If the label is “simulated”, another convolutional network is applied directly to the latent representation without sampling.

The latent discriminator  $D_{\theta_{\text{disc}}}$  is trained to make this representation  $z_i$  invariant to the domain via an adversarial loss as in generative adversarial networks (GANs) [15]. In particular, this is achieved by a two-player game; the discriminator is trained between two domains with a classification loss:

$$\ell_{\text{latent\_disc}} = -E_{x \sim P_{\text{exp}}}[\log D_{\theta_{\text{disc}}}(E_{\theta_{\text{enc}}}(x))] - E_{y \sim P_{\text{sim}}}[\log(1 - D_{\theta_{\text{disc}}}(E_{\theta_{\text{enc}}}(y)))] \quad (4.2)$$

and the parameters of the encoder are optimized through an additional adversarial loss:

$$\ell_{\text{adv\_latent}} = -E_{s \sim P_s}[\log D_{\theta_{\text{disc}}}(E_{\theta_{\text{enc}}}(s))], \quad (4.3)$$

where  $P_{\text{exp}}$  and  $P_{\text{sim}}$  are the experimental and simulated data distributions, and  $P_s$  is the joint distribution of both experimental and simulated data.

Since adversarial training can be unstable [16–18], we additionally add feature matching to stabilize adversarial training [17]:

$$\ell_{\text{FM}} = \frac{1}{2} \|E_{x \sim P_{\text{exp}}}[E_{\theta_{\text{enc}}}(x)] - E_{y \sim P_{\text{sim}}}[E_{\theta_{\text{enc}}}(y)]\|_2^2. \quad (4.4)$$

We opt for patchGAN discriminator [19] as proposed by Lample et al. [14] to improve the quality and sharpness of the reconstructions where the discriminators are trained by updating following two loss functions:

$$\ell_{\text{sim}} = -E_{s \sim P_{\text{exp}}}[\log D_{\theta_{\text{disc\_sim}}}(D_{\theta_{\text{dec}}}(E_{\theta_{\text{enc}}}(s)), l = \text{sim})], \quad (4.5)$$

$$\ell_{\text{exp}} = -E_{s \sim P_{\text{sim}}}[\log D_{\theta_{\text{disc\_exp}}}(D_{\theta_{\text{dec}}}(E_{\theta_{\text{enc}}}(s)), l = \text{exp})]. \quad (4.6)$$

Patch discriminators  $D_{\theta_{\text{disc\_sim}}}$  and  $D_{\theta_{\text{disc\_exp}}}$  are trained by minimizing the following losses respectively:

$$\ell_{\text{sim\_disc}} = -E_{x \sim P_{\text{sim}}}[\log D_{\theta_{\text{disc\_sim}}}(x)] - E_{y \sim P_{\text{exp}}}[\log(1 - D_{\theta_{\text{disc\_sim}}}(D_{\theta_{\text{dec}}}(E_{\theta_{\text{enc}}}(y)), l = \text{sim}))], \quad (4.7)$$

$$\ell_{\text{exp\_disc}} = -E_{x \sim P_{\text{exp}}}[\log D_{\theta_{\text{disc\_exp}}}(x)] - E_{y \sim P_{\text{sim}}}[\log(1 - D_{\theta_{\text{disc\_exp}}}(D_{\theta_{\text{dec}}}(E_{\theta_{\text{enc}}}(y)), l = \text{exp}))]. \quad (4.8)$$

The overall training procedure consists in iteratively updating the generator networks via minimization of the total loss

$$\ell_{\text{total}} = \ell_{\text{MAE}} + \ell_{\text{adv\_latent}} + \ell_{\text{Cycle}} + \ell_{\text{FM}} + \ell_{\text{sim}} + \ell_{\text{exp}}, \quad (4.9)$$

where  $\ell_{\text{Cycle}}$  is cycle consistency loss [20] and the discriminators are updated every  $n$  epochs

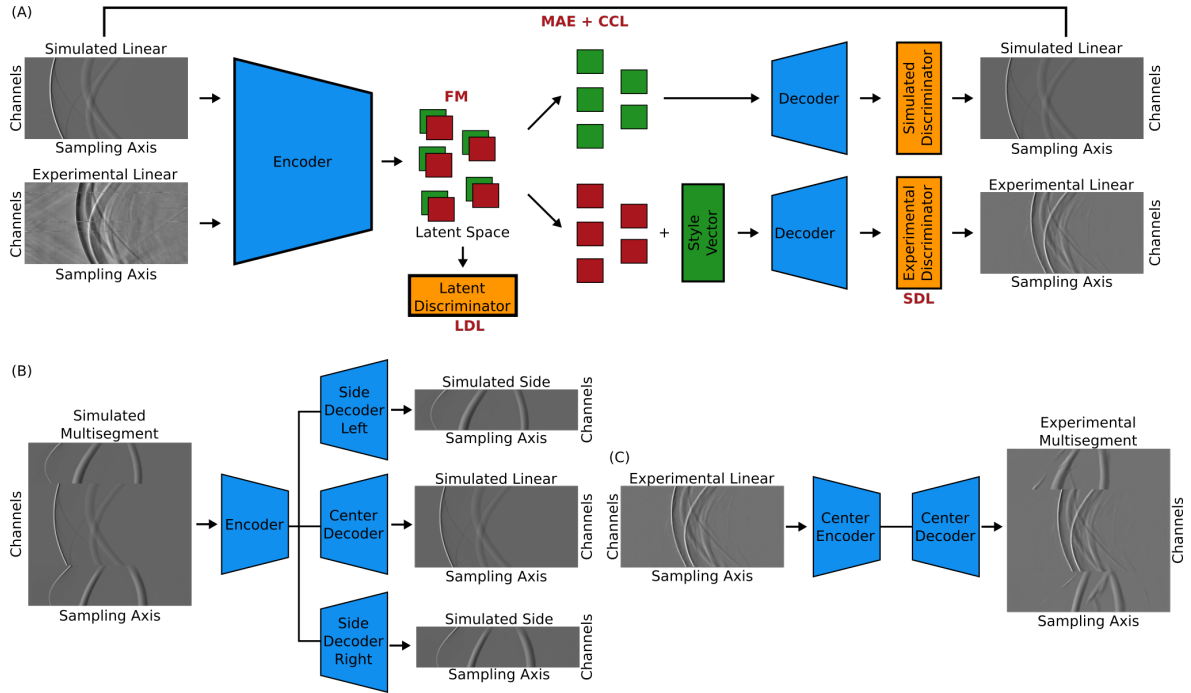


FIGURE 4.2: Summary of the proposed network architectures. A) Style network architecture with loss functions defined by red labels. Mean absolute error (MAE), cycle consistency loss (CCL), feature matching loss (FM), latent discriminator loss (LDL), style discriminator loss (SDL). B) Training of side network with simulated data. C) Training of side network using only experimental data from the linear part of the array.

via minimizing losses from equations 4.3, 4.7, 4.8 separately. This combination of losses was originally proposed in Lample et al. [14], except the cycle loss, which was proposed for domain adaptation in Hoffman et al. [21]. Such a particular combination of losses is important for a good quality of predictions for domain adaptation task. We will refer to this network as style network in the rest of this manuscript. The fully trained style network will be denoted as  $S_{\theta_{Style}}$  with  $s_{i,lin}$  linear parts of a signal  $i$  and  $s_{i,mul}$  multisegment array detection of this signal.

#### 4.2.2 SIDE NETWORK

After reducing the domain gap between simulated and experimental data, a second (side) network of auto-encoders is suggested to overcome limited-view-associated problems and yield geometrically corrected images. The main goal of this side network is to impute virtual signals at the concave parts of the multisegment array (Fig. 4.1A) using side decoders. Specifically, one encoder and three decoders are used in the auto-encoder architecture for simulated signals Fig. 4.2B. It is important to note that only the central part of the network (encoder and center decoder) is used in the training phase when the observation corresponds to an experimental signal. By training the network on simulated data, this is optimized to complete the concave sides of the imaging array. Accordingly, experimental signals are included in the dataset for optimization of the encoder and center decoder. Training on linear experimental signals helps to adapt network for experimental data distribution.

In a similar manner as in the style network, we propose an encoder-decoder architecture. As in  $S_{\theta_{Style}}$ , the encoder takes the signal matrix corresponding to linear array from either simulated or experimental images. The decoder, in contrast to  $S_{\theta_{Style}}$ , consists of three convolutional networks, where the goals of each networks are to produce signals from i) linear array (“center decoder”), ii) left concave segment, and iii) right concave segment. During training, the network only sees linear parts of experimental signals and whole multisegment array signals for simulated data. We use MAE loss to train the networks:

$$\ell_{sides} = \frac{1}{N} \sum_{(s_{lin}, s_{mul}) \in \mathcal{D}_{sim}} \|D_{\theta_{dec}}(E_{\theta_{enc}}(S_{\theta_{Style}}(s_{lin}))) - s_{mul}\|_1 + \frac{1}{M} \sum_{s_{lin} \in \mathcal{D}_{exp}} \|D_{\theta_{dec}}(E_{\theta_{enc}}(S_{\theta_{Style}}(s_{lin}))) - s_{lin}\|_1. \quad (4.10)$$

### 4.3 EXPERIMENTS AND RESULTS

#### 4.3.1 DATASETS DESCRIPTION

The datasets used in this study include two main parts. The simulated dataset contains a curved structure mimicking the skin surface and circular shapes similar to round vessel structures in the human forearm. It has 5500 cross-sectional images with different positioning of structures and number of vessels. The simulations were drawn as acoustic pressure maps in the spatial domain. Then, the corresponding signals were generated using the MSOT forward model for the multisegment array geometry [22]. 32 images were hold out for testing. The rest of the images were split into training and validation via 70/30 split. The second dataset was acquired from volunteers using a multisegment ultrasound transducer array as shown in Fig. 4.1A and described in detail in [8]. The corresponding simulated and experimental signals and images from the linear and all parts of the multisegment arrays are shown in Fig. 4.1B and 4.1C, respectively. In total, 5565 cross-sectional experimental images were collected from 22 forearms. 5501 images were allocated for training, 32 images for validation and 32 images for testing.

#### 4.3.2 RESULTS

The experimental signals acquired with the linear part were first processed with the style network to generate their virtual simulated counterparts in order to reduce the domain gap with actual simulated data. Then, these simulated signals were fed into the side network to impute the missing concave parts. The signals from the side network and the GT multisegment acquisitions were reconstructed using filtered back-projection and elastic-net algorithms [11, 22, 23] to generate the corresponding images. The comparison was made in the image domain because some of the samples from signal domain are not used in reconstruction as they stay out of the field of view (FOV) or otherwise modified/filtered by pre-processing algorithms before reconstruction. Another reason to evaluate the results in the image domain is that the main goal of this work is to enhance MSOT images by eliminating limited-view.

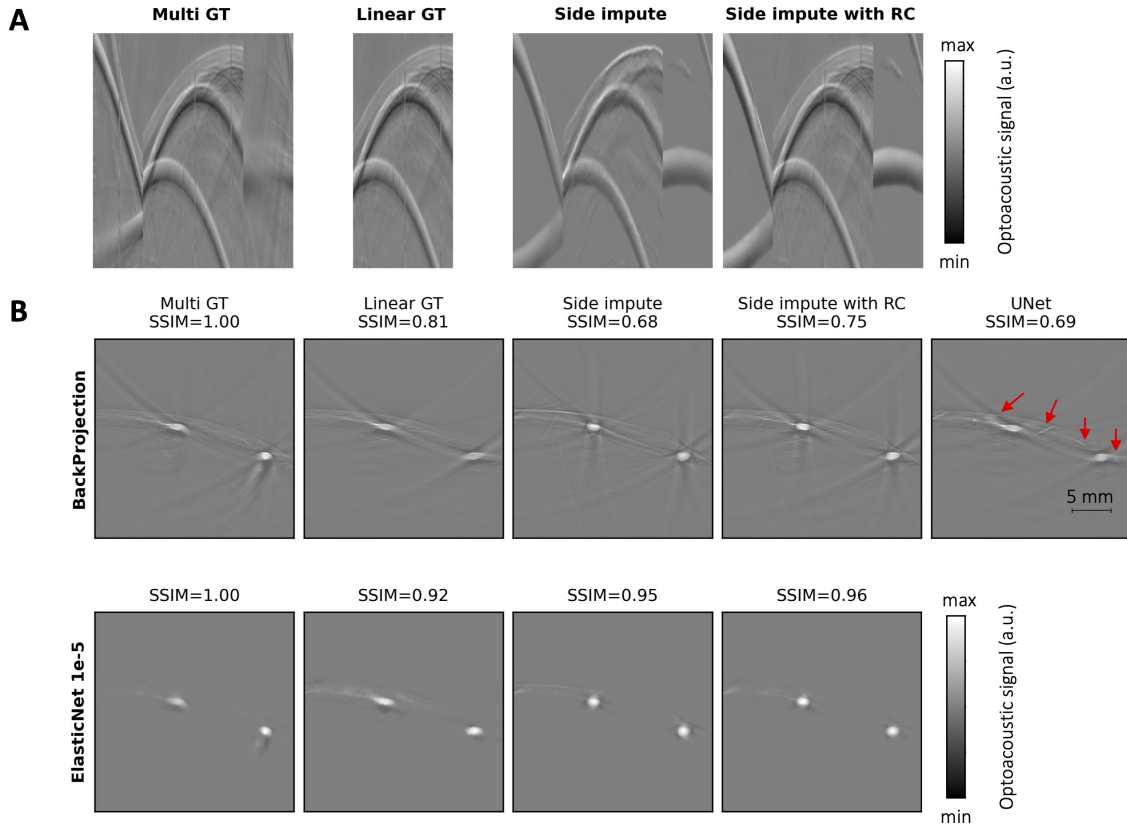


FIGURE 4.3: Results in the signal and image domains. A) Signal domain representation of a test image. Left to right: multisegment ground truth (GT) signals, linear GT signals, multisegment signals after style and side networks, multisegment signals after style and side network with real center (RC). B) Images reconstructed with back-projection and model-based elastic-net approach. Benchmark UNet result is added for comparison.

One example from test set is shown in Fig. 4.3. The third column in Fig. 4.3 shows the output of both networks after style network and side network. We refer to this output as “Sides impute” further in the paper. The linear part of this “Sides impute” output is further replaced with experimental linear part (Fig. 4.3A - 4th column). We refer to it further in the text as “Sides impute with RC” (RC stands for real center). The corresponding reconstructed images for each signal are displayed in the same column. The network results (side impute and side impute with RC) show clear improvement in round vessel structures which are the most important features visualized in MSOT images. In contrast, benchmark UNet is not able to correct the vessel shapes while further creating additional vessel-like structures.

The quantitative results calculated on the test set are summarized in Table 4.1 and a complete ablation study in Supplementary Tables 4.2, 4.3. Four different metrics were used to evaluate the proposed networks, namely structural similarity index (SSIM), mean squared error (MSE), Pearson correlation coefficient (Pearson) and peak signal to noise ratio (PSNR). The linear GT represents the images reconstructed from cropped multisegment signals; i.e., excluding concave parts. Hence, the linear GT may outperform some of the metrics. However, when the results are compared with the benchmark UNet, the proposed method becomes superior in each evaluation metric.

TABLE 4.1: Reconstruction scores with respect to ground truth (GT) multisegment reconstructions for elastic net (EN) with  $\alpha = 1e^{-5}$  and backprojection (BP) methods. Best score is in bold.

	SSIM		MSE		Pearson		PSNR	
	EN	BP	EN	BP	EN	BP	EN	BP
Linear GT	0.85	<b>0.67</b>	0.0019	<b>0.0013</b>	33.75	<b>35.37</b>	75.86	<b>77.48</b>
Ours	<b>0.90</b>	0.64	<b>0.0013</b>	0.0014	<b>35.28</b>	35.11	<b>77.39</b>	77.22
Unet	0.57	0.49	0.0026	0.0036	30.92	32.14	74.25	73.03

#### 4.4 CONCLUSION

This work is the first to pursue a signal domain solution to overcome limited-view reconstruction artifacts in MSOT imaging. Style transfer network was shown to reduce the domain gap between simulations and experimental signals, thus significantly improving quality of the reconstructed images versus conventional learning-free methods, such as back-projection or regularized model-based reconstruction. Previously-suggested methods trained exclusively with simulated data in the image domain (e.g. benchmark UNet) have shown inferior performance as compared to the domain adaptation networks proposed here. It was additionally demonstrated that, once the domain gap is reduced, training with simulated data can be used for imputing missing signals over a broader tomographic angle, thus leading to reduction of limited-view artifacts in backprojection-based reconstructions. However, the proposed method yields slight improvements in backprojection reconstructions. The results can be improved by using different architectures for the similar signal domain approach. Future work will make use of different training invariant architectures or other reconstruction methods.

#### ACKNOWLEDGEMENT

This work was supported by Swiss Data Science Center grant C19-04.



## APPENDIX A. RELATED WORK

### A.1. DOMAIN GENERALIZATION AND ADAPTATION

Domain adaptation and generalization address the problem of domain shift for settings where the distribution of observations for training and testing differ substantially. A popular approach to address the co-variate shift problem is to minimize the domain gap between two distributions in the latent space by minimizing maximum mean discrepancy [24–27], and adversarial feature alignment [28–32]. Recently, several methods based on image-level translation for domain adaptation and reduction of style bias have been proposed [21, 33, 34]. In particular, Hoffman et al. [21] proposed using core concepts from CycleGan [19]. Nam et al. [34] used style-agnostic networks to reduce domain shift by disentangling style encoding from class categories. Lample et al. [14] originally proposed using adversarial auto-encoder to swap attributes in the images, and later Lotfollahi et al. [35] demonstrated how a similar idea based on disentangling information from different domains could be applied for out-of-distribution predictions of unseen drug combinations. The method proposed herein combines aspects of these works in order to find a fast and easy way to train networks on simulated data that can be subsequently applied to experimental data.

### A.2. DEEP LEARNING IN OPTOACOUSTIC IMAGING

Several deep-learning-based methods have been used to enhance the MSOT imaging performance. For example, densely-sampled data was recovered from sparse signal acquisitions using supervised learning [10, 36, 37]. The inverse reconstruction problem has been tackled by learning the optimal regularization in iterative methods [38]. A semantic segmentation network was applied to hybrid OPUS images for delineating the mouse boundaries in pre-clinical data [39]. MSOT images are also segmented using convolutional neural networks [40, 41]. Multi-modal images from MSOT and magnetic resonance imaging (MRI) systems were registered using segmentation and spatial transformer networks [42]. Spectral unmixing between different wavelengths in MSOT images was performed using deep learning methods [43, 44]. Speed of sound values that are used in MSOT image reconstruction was corrected by a learning based method [45]. Noise caused by electromagnetic interference in MSOT imaging setups could be removed by means of signal domain learning approach [46]. Bandwidth enhancement was also proposed in the signal domain using supervised learning methods [47]. MSOT image and signal domain data were combined in hybrid networks to reduce limited-view artifacts [48, 49], although signal data was only used as complementary information for the image domain learning. The two-step method proposed herein solely operates in the signal domain to solve limited-view-associated problems.

## APPENDIX B. IMAGE RECONSTRUCTION METHODS

Two different methods were used in this study to generate images from time domain signals, namely back-projection and elastic-net [23]. Back-projection is a simple and widely used method based on delaying and summing the US signals according to their time of flight from

the pixels on the reconstruction grid [11]. It is applied as follows: i) the signals are first band-pass filtered between 0.1 and 6 MHz, ii) the filtered signals are normalized around zero mean, iii) the pre-processed signals are summed up based on their time of flights. The FOV was set to 25.6 mm (256x256 pixels). The elastic-net method was also used for reconstructing the images [23]. It is based on regularized inversion considering conventional model-based reconstruction [50]. This regularization approach was chosen as it allows removing some artifacts from the non-regularized version of the model-based algorithm and further enables computationally comparing the effect of geometry correction in the signal domain. We used a fixed parameter for the elastic net,  $\alpha = 10^{-5}$ . This value was empirically established for providing an optimal quality of the reconstructed images with simulated and experimental multisegment array data.

### APPENDIX C. BENCHMARK UNET

For comparison purposes, we trained a UNet model on the simulated image domain that learns how to map from MSOT reconstructions obtained with the linear part of the array to those achieved with the full multisegment geometry. For this, an Adam optimizer and MAE loss was used. As opposed to the original UNet [51], we use batch normalization [52] with zero padded convolutional layers in order to keep resolution constant within a convolutional block. In addition, benchmark UNet has a single residual connection from input to final model output [53] such that only residual mapping needs to be learned, and start with 32 convolutional kernels at full resolution instead of 64, going up to 512 instead of 1024 convolutional kernels at the coarsest level, consisting of a total of four max pooling layers. Below, we show that this baseline does not generalize well to experimental data.

### APPENDIX D. TRAINING

Two different network architectures were trained to reduce limited-view artifacts in the signal domain. The networks were implemented in pyTorch (v1.9) using CUDA (v11) and cuDNN (v8) libraries. Two NVIDIA Titan X GPUs were used in parallel for training. Both networks were trained for 200 epochs using batch size of 16. Adam optimizer with learning rate of 0.001 and weight decay of 10 were used for loss minimization. Style network was trained by leveraging the combination of six different losses as described in methodology section. The loss functions were weighted heuristically to arrange the effects of each loss. Specifically, weights  $\ell_{adv\_Latent}$ ,  $\ell_{patch\_sim}$  and  $\ell_{patch\_exp}$  were set to 0.001.  $\ell_{FM}$  and  $\ell_{Cycle}$  were weighted with 0.1.  $\ell_{MAE}$  was directly added with weight of 1. All discriminators are trained with gradient clipping penalty proposed in Arjovsky, Chintala, and Bottou [18] to mitigate the potentially too strong of discriminator and poor convergence of the adversarial losses.

## APPENDIX E. SUPPLEMENTARY RESULTS

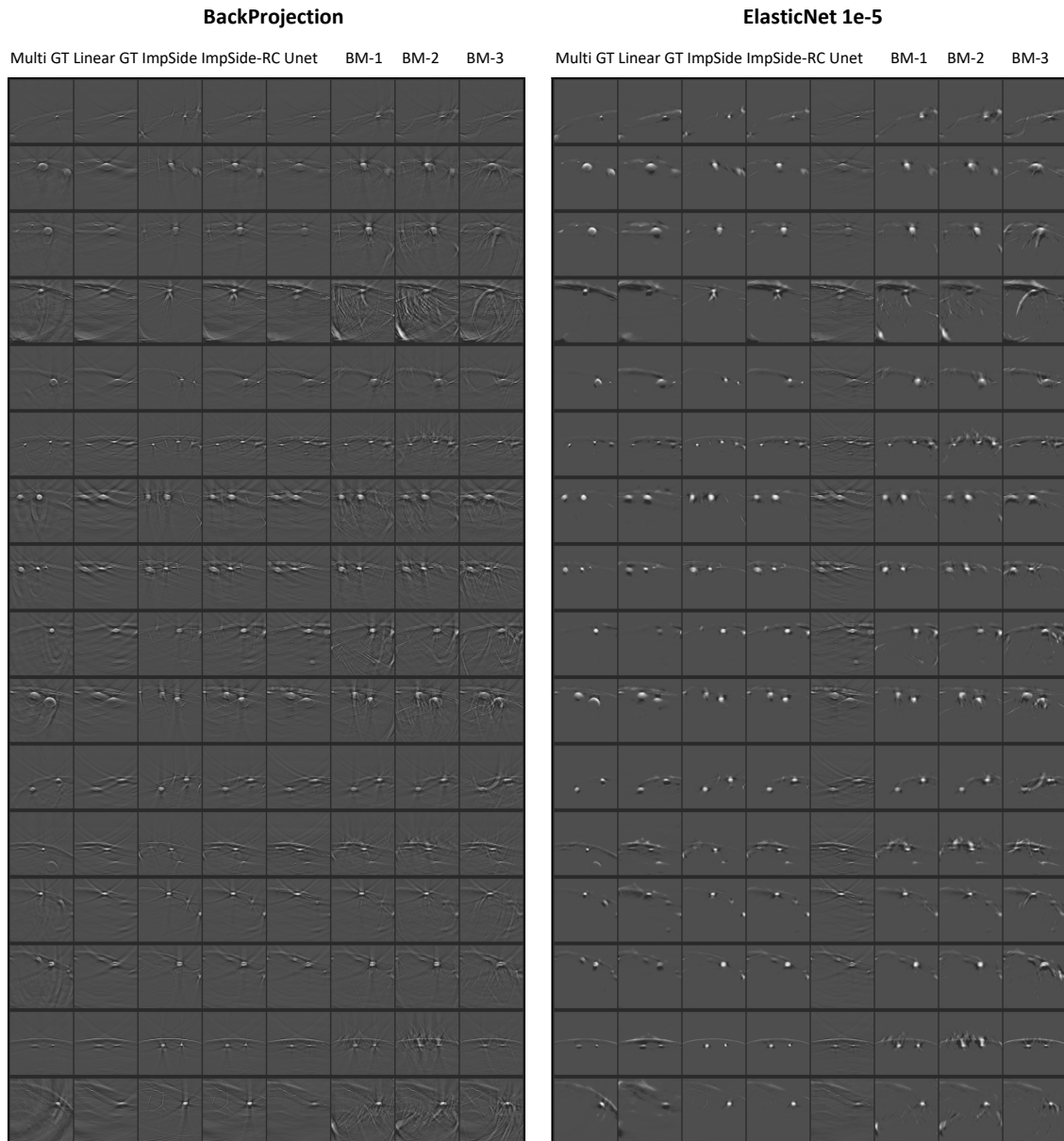


FIGURE 4.4: Example of reconstructions on the test set (sample number 1-16). Each row corresponds to a different input signal. Each column corresponds to a different method (best viewed digitally). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts.

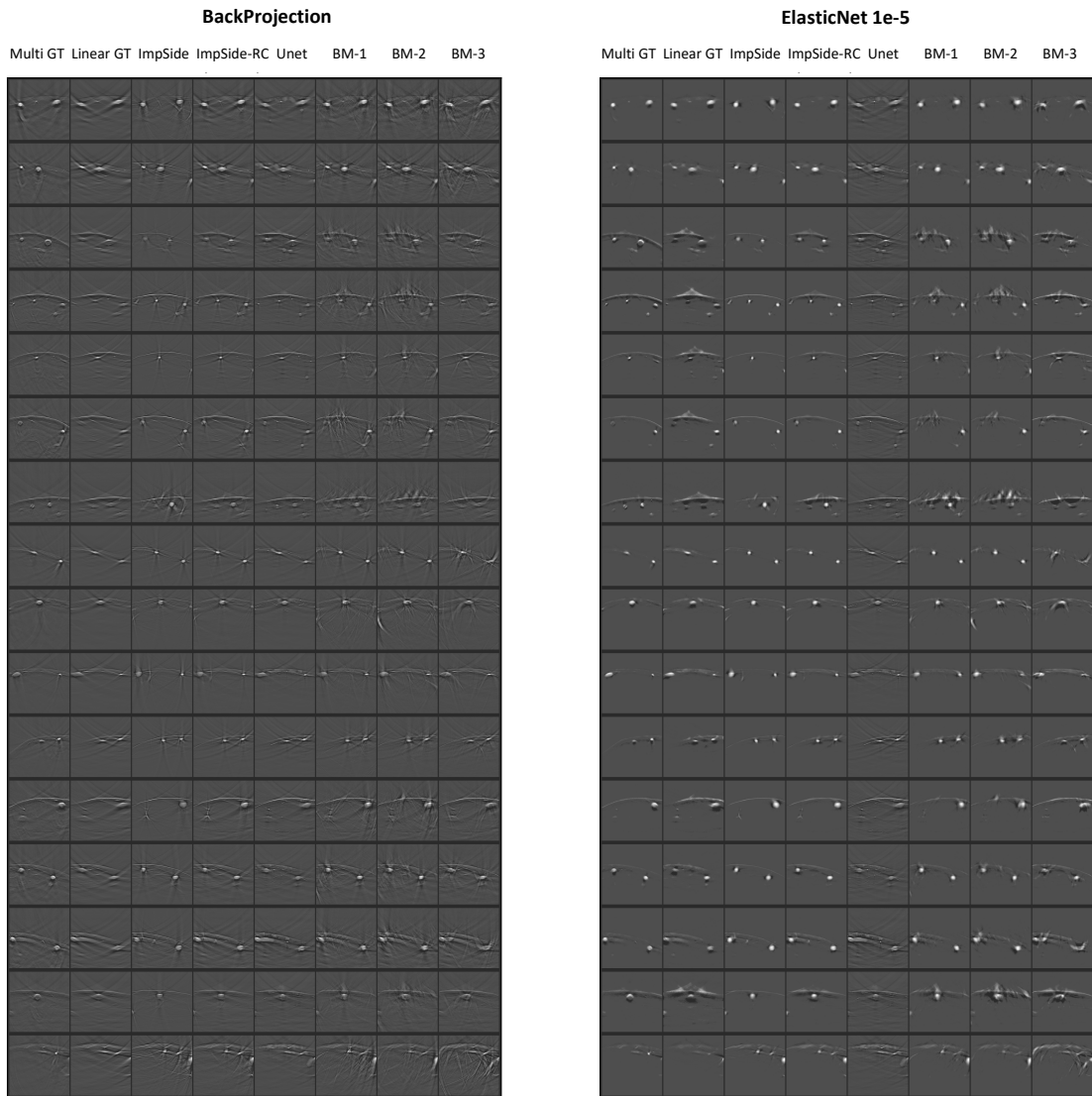


FIGURE 4.5: Example of reconstructions on the test set (sample number 17-32). Each row corresponds to a different input signal. Each column corresponds to a different method (best viewed digitally). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts

TABLE 4.2: Reconstruction scores with respect to ground truth (GT) multisegment reconstruction using Elastic Net ( $\alpha = 10^{-5}$ ). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts. Name of our main proposed method is in bold.

Dataset/Metrics	SSIM	MSE	Pearson	PSNR
Linear GT	$0.85 \pm 0.06$	$0.0019 \pm 0.0010$	$33.75 \pm 2.22$	$75.86 \pm 2.22$
Side impute	$0.89 \pm 0.04$	$0.0018 \pm 0.0008$	$33.89 \pm 1.93$	$76.00 \pm 1.93$
<b>Side impute RC</b>	$0.90 \pm 0.03$	$0.0013 \pm 0.0006$	$35.28 \pm 2.17$	$77.39 \pm 2.17$
Unet	$0.57 \pm 0.10$	$0.0026 \pm 0.0009$	$32.14 \pm 1.73$	$74.25 \pm 1.73$
BM-1	$0.86 \pm 0.05$	$0.0019 \pm 0.0011$	$33.92 \pm 2.26$	$76.03 \pm 2.26$
BM-2	$0.85 \pm 0.06$	$0.0025 \pm 0.0014$	$32.76 \pm 2.40$	$74.87 \pm 2.40$
BM-3	$0.84 \pm 0.05$	$0.0025 \pm 0.0011$	$32.50 \pm 2.02$	$74.61 \pm 2.02$

TABLE 4.3: Reconstruction scores with respect to ground truth (GT) multisegment reconstruction using backprojection (BP). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts. Name of our main proposed method is in bold.

Dataset/Metrics	SSIM	MSE	Pearson	PSNR
Linear GT	$0.67 \pm 0.11$	$0.0013 \pm 0.0007$	$35.37 \pm 2.39$	$77.48 \pm 2.39$
Side impute	$0.51 \pm 0.13$	$0.0025 \pm 0.0012$	$32.45 \pm 2.02$	$74.56 \pm 2.02$
<b>Side impute RC</b>	$0.64 \pm 0.11$	$0.0014 \pm 0.0008$	$35.11 \pm 2.52$	$77.22 \pm 2.52$
Unet	$0.49 \pm 0.13$	$0.0036 \pm 0.0018$	$30.92 \pm 2.15$	$73.03 \pm 2.15$
BM-1	$0.53 \pm 0.13$	$0.0022 \pm 0.0017$	$33.43 \pm 2.74$	$75.54 \pm 2.74$
BM-2	$0.50 \pm 0.12$	$0.0025 \pm 0.0024$	$32.93 \pm 2.67$	$75.04 \pm 2.67$
BM-3	$0.54 \pm 0.13$	$0.0022 \pm 0.0015$	$33.33 \pm 2.75$	$75.44 \pm 2.75$

## BIBLIOGRAPHY

- [1] Elena Merčep et al. “Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.9 (2015), pp. 1651–1661.
- [2] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [3] Berkan Lafci et al. “Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound”. In: *Neoplasia* 22.12 (2020), pp. 770–777.
- [4] Geng-Shi Jeng et al. “Real-time interleaved spectroscopic photoacoustic and ultrasound (PAUS) scanning with simultaneous fluence compensation and motion correction”. In: *Nature communications* 12.1 (2021), p. 716.
- [5] Leonardo G Montilla et al. “Real-time photoacoustic and ultrasound imaging: a simple solution for clinical ultrasound systems with linear arrays”. In: *Physics in Medicine & Biology* 58.1 (2012), N1.
- [6] Xosé Luís Deán-Ben and Daniel Razansky. “On the link between the speckle free nature of optoacoustics and visibility of structures in limited-view tomography”. In: *Photoacoustics* 4.4 (2016), pp. 133–140.
- [7] Xosé Luis Deán-Ben, E Merčep, and Daniel Razansky. “Hybrid-array-based optoacoustic and ultrasound (OPUS) imaging of biological tissues”. In: *Applied Physics Letters* 110.20 (2017), p. 203703.
- [8] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. “Combined pulse-echo ultrasound and multispectral optoacoustic tomography with a multi-segment detector array”. In: *IEEE transactions on medical imaging* 36.10 (2017), pp. 2129–2137.
- [9] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. “Imaging of blood flow and oxygen state with a multi-segment optoacoustic ultrasound array”. In: *Photoacoustics* 10 (2018), pp. 48–53.
- [10] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky. “Deep learning optoacoustic tomography with sparse data”. In: *Nature Machine Intelligence* 1.10 (2019), pp. 453–460.
- [11] Ali Ozbek, XL Deán-Ben, and Daniel Razansky. “Realtime parallel back-projection algorithm for three-dimensional optoacoustic imaging devices”. In: *European conference on biomedical optics*. Optica Publishing Group. 2013, p. 88000I.

- 
- [12] Minghua Xu and Lihong V Wang. “Universal back-projection algorithm for photoacoustic computed tomography”. In: *Physical Review E* 71.1 (2005), p. 016706.
- [13] Lu Ding, Daniel Razansky, and Xosé Luís Deán-Ben. “Model-Based Reconstruction of Large Three-Dimensional Optoacoustic Datasets”. In: *IEEE Transactions on Medical Imaging* 39.9 (2020), pp. 2931–2940.
- [14] Guillaume Lample et al. “Fader networks: Generating image variations by sliding attribute values”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 5963–5972.
- [15] Ian Goodfellow et al. “Generative adversarial nets”. In: *Advances in neural information processing systems* 27 (2014).
- [16] Martin Arjovsky and Léon Bottou. “Towards principled methods for training generative adversarial networks”. In: *arXiv preprint arXiv:1701.04862* (2017).
- [17] Jianmin Bao et al. “CVAE-GAN: fine-grained image generation through asymmetric training”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2745–2754.
- [18] Martin Arjovsky, Soumith Chintala, and Léon Bottou. “Wasserstein generative adversarial networks”. In: *International conference on machine learning*. PMLR. 2017, pp. 214–223.
- [19] Jun-Yan Zhu et al. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.
- [20] Phillip Isola et al. “Image-to-image translation with conditional adversarial networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.
- [21] Judy Hoffman et al. “Cycada: Cycle-consistent adversarial domain adaptation”. In: *International conference on machine learning*. PMLR. 2018, pp. 1989–1998.
- [22] X Luís Dean-Ben, Vasilis Ntziachristos, and Daniel Razansky. “Acceleration of optoacoustic model-based reconstruction using angular image discretization”. In: *IEEE Transactions on medical imaging* 31.5 (2012), pp. 1154–1162.
- [23] Hui Zou and Trevor Hastie. “Regularization and variable selection via the elastic net”. In: *Journal of the royal statistical society: series B (statistical methodology)* 67.2 (2005), pp. 301–320.
- [24] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. “Domain generalization via invariant feature representation”. In: *International Conference on Machine Learning*. PMLR. 2013, pp. 10–18.
- [25] Muhammad Ghifary et al. “Scatter component analysis: A unified framework for domain adaptation and domain generalization”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.7 (2016), pp. 1414–1430.

- [26] Muhammad Ghifary, W Bastiaan Kleijn, and Mengjie Zhang. “Domain adaptive neural networks for object recognition”. In: *Pacific Rim international conference on artificial intelligence*. Springer. 2014, pp. 898–904.
- [27] Mingsheng Long et al. “Deep transfer learning with joint adaptation networks”. In: *International conference on machine learning*. PMLR. 2017, pp. 2208–2217.
- [28] Ya Li et al. “Deep domain generalization via conditional invariant adversarial networks”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 624–639.
- [29] Mingsheng Long et al. “Learning transferable features with deep adaptation networks”. In: *International conference on machine learning*. PMLR. 2015, pp. 97–105.
- [30] Yaroslav Ganin et al. “Domain-adversarial training of neural networks”. In: *The journal of machine learning research* 17.1 (2016), pp. 2096–2030.
- [31] Mingsheng Long et al. “Conditional adversarial domain adaptation”. In: *arXiv preprint arXiv:1705.10667* (2017).
- [32] Eric Tzeng et al. “Adversarial discriminative domain adaptation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7167–7176.
- [33] Zak Murez et al. “Image to image translation for domain adaptation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 4500–4509.
- [34] Hyeonseob Nam et al. “Reducing Domain Gap by Reducing Style Bias”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 8690–8699.
- [35] Mohammad Lotfollahi et al. “Compositional perturbation autoencoder for single-cell response modeling”. In: *bioRxiv* (2021).
- [36] Stephan Antholzer, Markus Haltmeier, and Johannes Schwab. “Deep learning for photoacoustic tomography from sparse data”. In: *Inverse problems in science and engineering* 27.7 (2019), pp. 987–1005.
- [37] Steven Guan et al. “Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning”. In: *Scientific reports* 10.1 (2020), p. 8510.
- [38] Andreas Hauptmann et al. “Model-based learning for accelerated, limited-view 3-D photoacoustic tomography”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1382–1393.
- [39] Berkan Lafci et al. “Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 68.3 (2020), pp. 688–696.
- [40] Melanie Schellenberg et al. “Semantic segmentation of multispectral photoacoustic images using deep learning”. In: *Photoacoustics* 26 (2022), p. 100341.



- 
- [41] Nikolaos-Kosmas Chlis et al. “A sparse deep learning approach for automatic segmentation of human vasculature in multispectral optoacoustic tomography”. In: *Photoacoustics* 20 (2020), p. 100203.
- [42] Yexing Hu et al. “Deep learning facilitates fully automated brain image registration of optoacoustic tomography and magnetic resonance imaging”. In: *Biomedical Optics Express* 13.9 (2022), pp. 4817–4833.
- [43] Ivan Olefir et al. “Deep learning-based spectral unmixing for optoacoustic imaging of tissue oxygen saturation”. In: *IEEE transactions on medical imaging* 39.11 (2020), pp. 3643–3654.
- [44] Janek Gröhl et al. “Learned spectral decoloring enables photoacoustic oximetry”. In: *Scientific reports* 11.1 (2021), p. 6565.
- [45] Seungwan Jeon and Chulhong Kim. “Deep learning-based speed of sound aberration correction in photoacoustic images”. In: *Photons plus ultrasound: Imaging and sensing 2020*. Vol. 11240. SPIE. 2020, pp. 24–27.
- [46] Christoph Dehner et al. “Deep-learning-based electrical noise removal enables high spectral optoacoustic contrast in deep tissue”. In: *IEEE Transactions on Medical Imaging* 41.11 (2022), pp. 3182–3193.
- [47] Sreedevi Gutta et al. “Deep neural network-based bandwidth enhancement of photoacoustic data”. In: *Journal of biomedical optics* 22.11 (2017), pp. 116001–116001.
- [48] Neda Davoudi et al. “Deep learning of image-and time-domain data enhances the visibility of structures in optoacoustic tomography”. In: *Optics letters* 46.13 (2021), pp. 3029–3032.
- [49] Hengrong Lan et al. “Y-Net: a hybrid deep learning reconstruction framework for photoacoustic imaging in vivo”. In: *arXiv preprint arXiv:1908.00975* (2019).
- [50] X Luís Dean-Ben et al. “Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography”. In: *IEEE transactions on medical imaging* 31.10 (2012), pp. 1922–1928.
- [51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer. 2015, pp. 234–241.
- [52] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. pmlr. 2015, pp. 448–456.
- [53] Kyong Hwan Jin et al. “Deep convolutional neural network for inverse problems in imaging”. In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522.

## 5 OADAT: EXPERIMENTAL AND SYNTHETIC CLINICAL OPTOACOUSTIC DATA FOR STANDARDIZED IMAGE PROCESSING

This chapter includes the following publication:

Transactions on Machine Learning Research (2023).

Authors:

Firat Ozdemir<sup>1,†</sup>, **Berkan Lafci**<sup>2,3,†</sup>, Xosé Luís Deán-Ben<sup>2,3</sup>, Daniel Razansky<sup>2,3</sup>, Fernando Perez-Cruz<sup>1,4</sup>

<sup>1</sup> Swiss Data Science Center, ETH Zürich and EPFL, Zürich, Switzerland

<sup>2</sup> Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>3</sup> Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>4</sup> Institute for Machine Learning, Department of Computer Science, ETH Zurich, Switzerland

† equal contribution

## ABSTRACT

Optoacoustic (OA) imaging is based on excitation of biological tissues with nanosecond-duration laser pulses followed by subsequent detection of ultrasound waves generated via light-absorption-mediated thermoelastic expansion. OA imaging features a powerful combination between rich optical contrast and high resolution in deep tissues. This enabled the exploration of a number of attractive new applications both in clinical and laboratory settings. However, no standardized datasets generated with different types of experimental set-up and associated processing methods are available to facilitate advances in broader applications of OA in clinical settings. This complicates an objective comparison between new and established data processing methods, often leading to qualitative results and arbitrary interpretations of the data. In this paper, we provide both experimental and synthetic OA raw signals and reconstructed image domain datasets rendered with different experimental parameters and tomographic acquisition geometries. We further provide trained neural networks to tackle three important challenges related to OA image processing, namely accurate reconstruction under limited view tomographic conditions, removal of spatial undersampling artifacts and anatomical segmentation for improved image reconstruction. Specifically, we define 44 experiments corresponding to the aforementioned challenges as benchmarks to be used as a reference for the development of more advanced processing methods.

## 5.1 INTRODUCTION

Optoacoustic (OA) imaging is being established as a powerful method with increasing application areas in clinical [1, 2] and preclinical settings [3, 4]. Using nanosecond-duration pulsed lasers operating in the visible and near infrared (NIR) optical wavelength range, biological tissues are thermoelastically excited. This excitation yields ultrasound (US) waves, from which OA images are tomographically reconstructed (Fig. 5.1a). The rich optical contrast from endogenous tissue chromophores such as blood, melanin, lipids and others are combined with high US resolution, i.e., tens of micrometers. This unique feature makes OA particularly suitable for molecular and functional imaging. Other important advantages such as the feasibility of hand-held operation, the fast acquisition performance (real time feedback) and the non-invasive safe contrast (i.e., non-ionizing radiation) further foster the wide use of OA in multiple biomedical studies. OA imaging has been shown to provide unique capabilities in studies with disease models e.g., of breast cancer [5–7], as well as for the clinical assessment of Crohn’s disease [8], atherosclerotic carotid plaques [9] or skin cancer [10]. As the range of applications of OA imaging gets broader, the need for different data processing pipelines increases in parallel. Also, new methods are continuously being developed to provide an enhanced OA performance. Specific examples include increased temporal resolution with compressed/sparse data acquisitions [11], accurate image reconstruction algorithms [12], light fluence correction by segmenting the tissue boundaries [13] or enhanced spectral unmixing algorithms from multispectral data [14].

Three major challenges suitable for data-driven approaches in clinical OA imaging are summarized below:

**Sparse acquisition:** OA imaging provides a unique potential to monitor fast-changing events such as cardiac arrhythmias [15], neuronal activity [16] or indocyanine green clearance [17] *in vivo*. For this, ultra-fast imaging systems capable of capturing changes in living organisms occurring at up to millisecond temporal scales are required. The main limiting factor affecting the achievable frame rate is the data transfer capacity. This limitation can be eliminated by reducing the number of acquired channels (signals). Therefore, sparse or compressed sensing methods have been proposed both using conventional methods [11] and deep learning algorithms [18].

**Limited view reconstruction:** OA is inherently a tomographic imaging modality. Acquisition of pressure signals from different angles is essential to capture the information encoded in US waves traveling in a three-dimensional (3D) medium in order to render accurate tomographic reconstructions. This further increases the image contrast, resolution and quantitiveness. However, tomographic coverage of the samples is often hindered by physical restrictions. Thereby, new image processing pipelines have been suggested to improve limited-view-associated challenges in OA imaging by using data-driven algorithms in the image domain [19], signal domain [20] and combination of both domains [21, 22].

**Segmentation:** Optimal OA reconstruction algorithms need to account for different optical and acoustic properties in biological tissues and in the coupling medium (water). For example, the speed of sound (SoS) depends on the elastic properties of the medium. Proper assignment of SoS values in tissues and in water requires accurate delineation of the tissue boundaries. Thereby, segmentation of structures [13, 23] in OA images has been shown to enhance the image reconstruction performance. Additionally, the optical fluence (intensity) also varies with depth across different tissues. This issue remains as one of the main factors affecting quantification in OA images [24] and can also be corrected with tissue segmentation [25].

As an emerging method, OA imaging requires standardization, open source code publication and data sharing practices to expedite the development of new application areas and data processing pipelines. Particularly, the aforementioned challenges associated to high data throughput, limited angular coverage, SoS assignment and fluence corrections require coordinated efforts between experimental and data science communities. Initial efforts to standardize data storage formats and image reconstruction algorithms have been undertaken [26]. However, Gröhl et al. [26] focus on standard reconstruction algorithms and propose data storage formats for acquisition related metadata. Data-driven image and signal processing methods require additional initiatives on fast access to large bulks of OA image and signal data and benchmarks for learning-based methods.

Here, we provide experimental data and simulations of forearm datasets as well as benchmark networks aiming at facilitating the development of new image processing algorithms and benchmarking. These “Experimental and Synthetic Clinical Optoacoustic Data (OADAT)” include, (i) large and varied clinical and simulated forearm datasets with paired subsampled or limited view image reconstruction counterparts, (ii) raw signal acquisition data of each such image reconstruction, (iii) definition of 44 experiments with gold standards focusing on the aforementioned OA challenges, (iv) pretrained model weights of the networks used for each task, and (v) user-friendly scripts to load and evaluate the networks on our datasets. The presented datasets and algorithms will expedite the research in OA image processing.

## 5.2 BACKGROUND

For OA imaging, the objects are excited with the nanosecond-duration laser pulses in visible or NIR light wavelengths which result in thermoelastic expansion of the structures. This expansion generates pressure waves (US signals) that are detected by transducer arrays. Corresponding images are reconstructed by solving the OA inverse problem on the acquired signals. Below, we explain the transducer arrays used for data acquisition, the reconstruction algorithm used to generate images from acquired signals and the sampling/acquisition techniques. Detailed explanation about OA imaging and used tools can be found in Appendix 5.7.

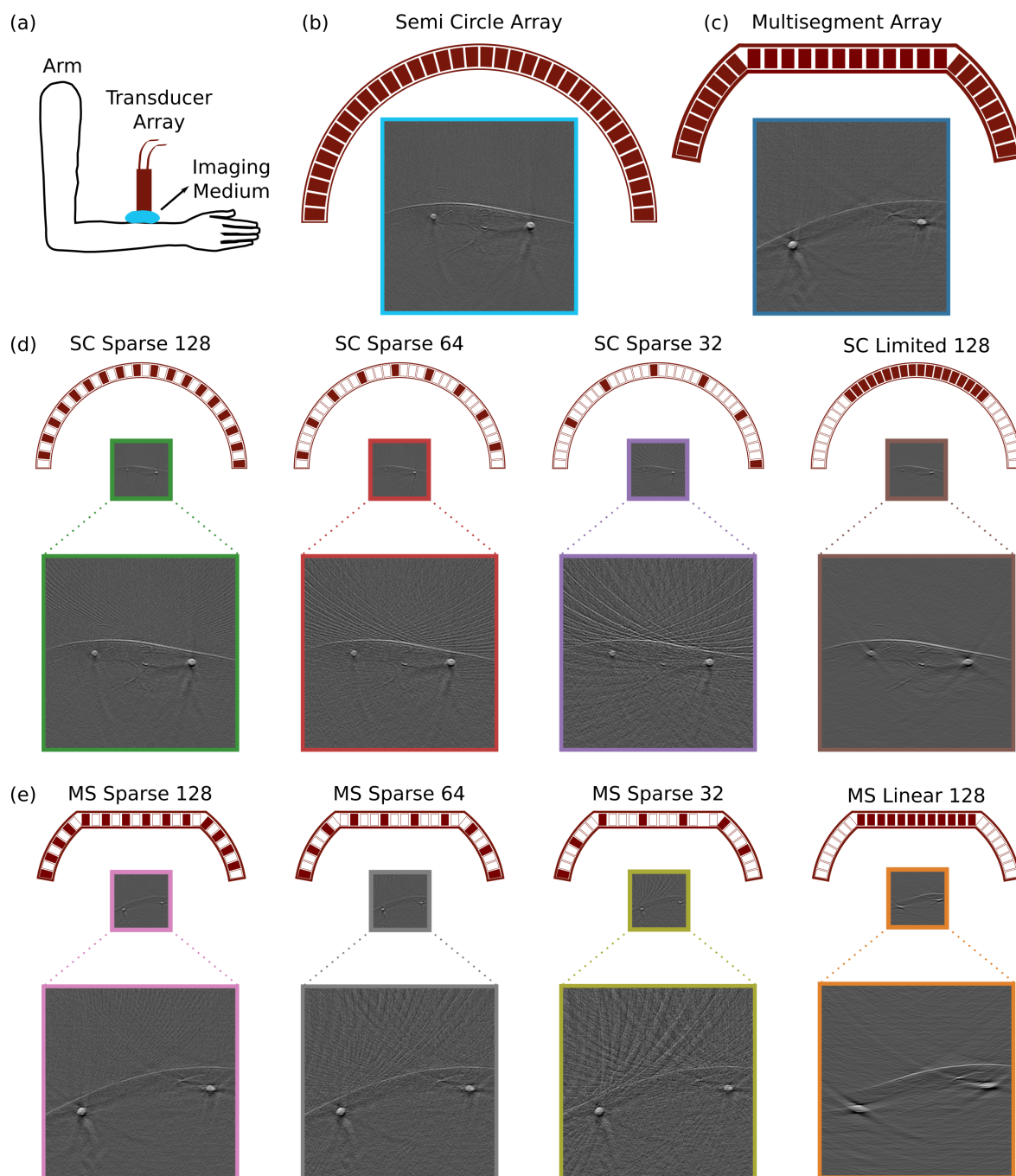


FIGURE 5.1: Experimental data acquisition, transducer arrays and resulting images. (a) Experimental setup for optoacoustic (OA) forearm imaging. (b) Semi circle array along with an example of acquired images. (c) Multisegment array along with an example of acquired images. (d) Uniform subsampling for the semi circle array (128, 64 and 32 elements) and limited view acquisition for the semi circle array with reduced angular coverage (128 elements). (e) Uniform subsampling for the multisegment array (128, 64 and 32 elements), and linear array acquisition for the multisegment array (128 elements). Transducer elements are shown as actively receiving (red) or off (white).

### 5.2.1 TRANSDUCER ARRAYS

**Semi circle** array contains 256 transducer elements distributed equidistantly over a semi circle (concave surface, Fig. 5.1b). **Multisegment** array is a combination of linear array in the center and concave parts on the right and left sides, designed to increase angular coverage, as shown in Fig. 5.1c. The linear part contains 128 elements and each of the concave parts consist of 64 elements, totaling to 256. **Linear** array is the central part of the multisegment array with 128 elements (Fig. 5.1c). The array geometry is optimized for US data acquisitions with planar waves. Hence, it produces OA images with limited view artifacts due to reduced angular coverage which is a limiting factor for OA image acquisitions. **Virtual circle** array is generated to simulate images with 360 degree angular coverage and yields artifact free reconstructions (Fig. 5.2a). It contains 1,024 transducer elements distributed over a full circle with equal distance. We also have a virtual multisegment array that correspond to its physical counterpart. Additional geometric and technical details are listed in Appendix 5.7.

### 5.2.2 RECONSTRUCTION METHOD

We use backprojection algorithm in this study to generate OA images from the acquired signals<sup>1</sup>. This algorithm is based on delay and sum beamforming approach [27] (see Appendix 5.7 for details). First, a mesh grid is created to represent the imaged field of view. Then, the distance between the points of the mesh grid and array elements are calculated based on the known locations of the transducers. Time of flight is obtained through dividing distance by the SoS values that are assigned based on the temperature of the imaging medium and tissue properties. The clinical and simulated data are reconstructed with SoS of 1,510 m/s in this study as the simulations and the experiments were done at the corresponding imaging medium temperature. Unlike more sophisticated model-based reconstruction approaches [12], backprojection is parameterized using only SoS, making it a stable choice across all imaged scenes.

### 5.2.3 SPARSE SAMPLING

Sparse sampling yields streak artifacts on reconstructed images due to large inter-element pitch size. For a given angular coverage, i.e., transducer array geometry, using less transducer elements for reconstruction causes stronger artifacts due to increased inter-element pitch size. We define sparse sampled semi circle array acquisitions **semi circle sparse 128**, **semi circle sparse 64** and **semi circle sparse 32** when using 128, 64 and 32 elements out of the 256 of semi circle array (Fig. 5.1d first three columns), respectively. Similarly, we define sparse sampled virtual circle array acquisitions **virtual circle sparse 128**, **virtual circle sparse 64** and **virtual circle sparse 32** when using 128, 64 and 32 elements out of 1,024 of virtual circle array (Fig. 5.2c first three rows), respectively. In addition, we also define sparse sampled multisegment array acquisitions **multisegment sparse 128**, **multisegment sparse 64** and **multisegment sparse 32** when using 128, 64 and 32 elements out of the 256 elements of the multisegment array (Figs. 5.1e and 5.2d first three columns), respectively. All items

<sup>1</sup>Python module for OA reconstruction: [github.com/berkanlafci/pyoat](https://github.com/berkanlafci/pyoat).

correspond to uniform and hence equidistant subsampling of the corresponding transducer array signals.

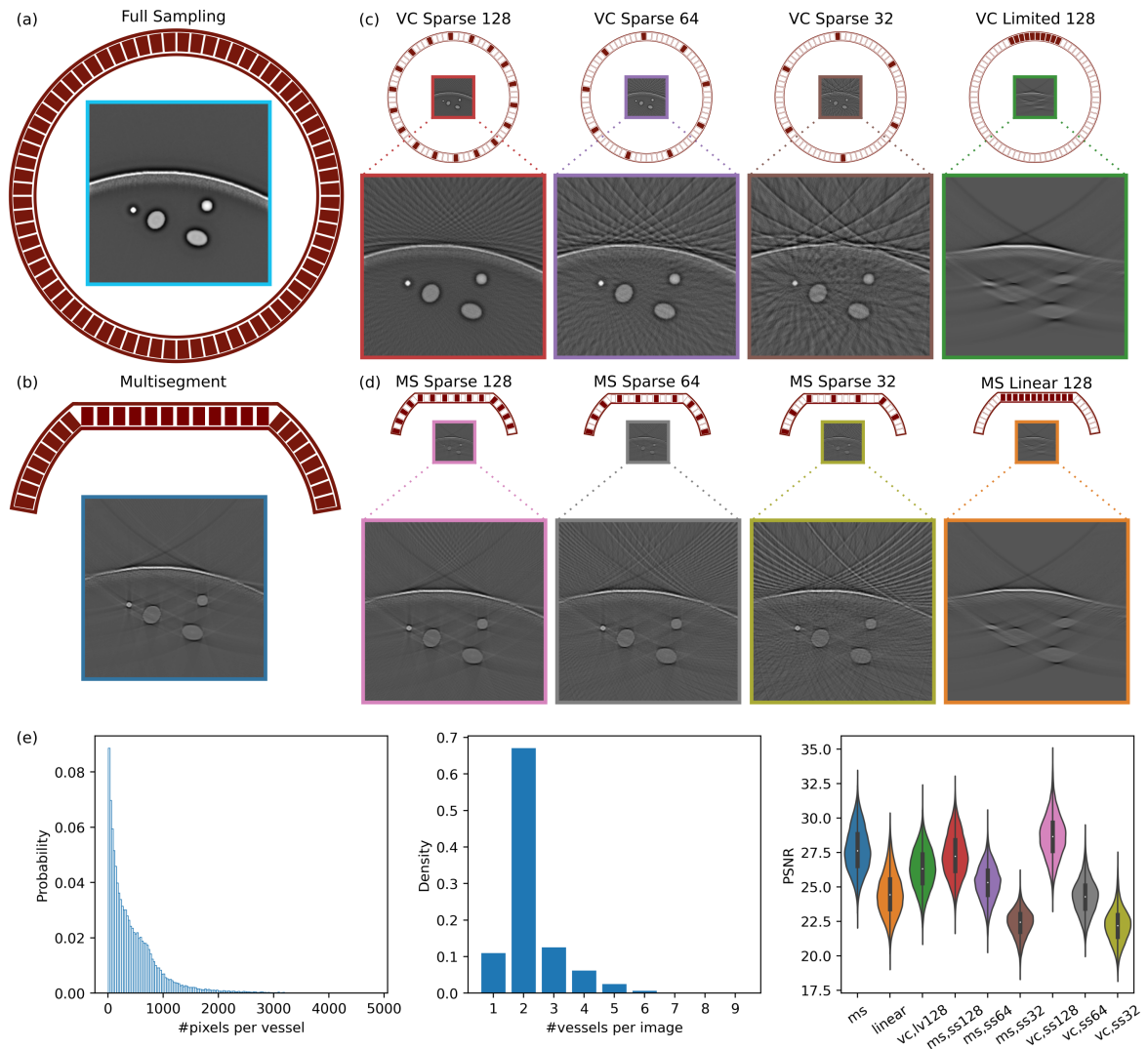


FIGURE 5.2: Overview of the simulated data. (a) Virtual circle array and an image reconstructed using 1,024 transducer elements. (b) Multisegment array and the corresponding image reconstructed using combined linear and concave parts of the transducer array. (c) Uniform subsampling of virtual circle array with 128, 64 and 32 elements and limited view acquisition with reduced angular coverage (128 elements). (d) Uniform subsampling of multisegment array with 128, 64 and 32 elements and linear array acquisition (128 elements). (e) Vessel size distribution (pixels per vessel), number of vessels per image, and peak signal-to-noise ratio of full sampling compared to other reconstructions (x axis naming conventions are explained in Sec. 5.3.3). Transducer elements are shown as actively receiving (red) or off (white).

#### 5.2.4 LIMITED VIEW

Limited view acquisitions lead to distorted geometry (e.g., elongated vessels) due to the reduced angular coverage (Figs. 5.1d,e & 5.2c,d last column, limited 128 and linear 128). To mimic commonly occurring limited view settings, we use a continuous subset of elements for a given transducer. This corresponds to retaining inter-element pitch size while reducing the



angular coverage. We define a limited view acquisition for each transducer array as follows: (i) **Linear array** is the common practice in clinical settings for US data acquisition [28, 29]. Typically, the same linear geometry is combined with OA imaging to provide complementary information [30]. To model this clinically realistic scenario, we use the linear part of the multisegment array for OA image reconstruction. (ii) **Semi circle limited 128** uses half of the semi circle array; 128 transducer elements, yielding a quarter circle. The differences between linear and semi circle limited view array acquisitions are the inter-element pitch size, focusing and geometry of the active area. (iii) **Virtual circle limited 128** uses 128 consecutive elements (one eighth of a circle) out of 1,024.

### 5.3 DATASETS

We present four datasets <sup>2 3</sup> (two experimental, one simulated, one fully annotated subset) where each has several subcategories for the purpose of tackling different challenges present in the domain. Raw signal acquisition data that is used to reconstruct all images are also provided with the datasets. Experimental datasets also include details about the volunteer Fitzpatrick skin phototype [31], which relates to the amount of melanin pigment in the skin (see Appendix 5.7 for distribution and further details). We also display a comparative overview of publicly available and our proposed OA datasets in Table 5.1. Please refer to Tables 5.5, 5.6, 5.7, 5.8, and 5.9 for summaries of the file contents of the datasets in the Appendix.

TABLE 5.1: Publicly available OA datasets, supported tasks, provided data format(s), size, and content. Davoudi et al. [18] contains 274 mice and 469 phantom slices. Huang et al. [32] has 10 mice with 10 frames (100 slices) at 27 different wavelengths and 20 phantom slices.

Dataset	tasks			image reconstruction	raw signal	size (>5k instances)	clinical data
	limited view	sparse sampling	pixel annotations				
Davoudi et al. [18]	✗	✓	✗	✓	✗	✗	✗
Huang et al. [32]	✗	✗	✗	✗	✓	✗	✗
MSFD (ours)	✓	✗	✗	✓	✓	✓	✓
SWFD (ours)	✓	✓	✗	✓	✓	✓	✓
SCD (ours)	✓	✓	✓	✓	✓	✓	✗
OADAT-mini (ours)	✓	✓	✓	✓	✓	✓	✗ ∪ ✓

#### 5.3.1 MULTISPECTRAL FOREARM DATASET

Multispectral forearm dataset (MSFD) is collected using multisegment array (Sec. 5.2.1) from nine volunteers at six different wavelengths (700, 730, 760, 780, 800, 850 nm) for both arms. Selected wavelengths are particularly aimed for spectral decomposition aiming to separate oxy- and deoxy-hemoglobin [33]. All wavelengths are acquired consecutively, yielding almost identical scene being captured for a given slice across different wavelengths with slight displacement errors. For each of the mentioned category 1,400 slices are captured, creating a

<sup>2</sup>Link to our datasets: [doi.org/10.3929/ethz-b-000551512](https://doi.org/10.3929/ethz-b-000551512)

<sup>3</sup>Repository for accessing and reading datasets: [github.com/berkanlafci/oadat](https://github.com/berkanlafci/oadat)

sum of  $9 \times 6 \times 2 \times 1,400 = 151,200$  unique signal matrices.

From this data, using backprojection algorithm, we reconstruct (i) linear array images  $\text{MSFD}_{\text{linear}}$ , (ii) multisegment array images  $\text{MSFD}_{\text{ms}}$ , (iii) multisegment sparse 128 images (Sec.5.2.3),  $\text{MSFD}_{\text{ms,ss128}}$ , (iv) multisegment sparse 64 images (Sec.5.2.3),  $\text{MSFD}_{\text{ms,ss64}}$ , and (v) multisegment sparse 32 images (Sec.5.2.3),  $\text{MSFD}_{\text{ms,ss32}}$ , where each dataset has 151,200 images of  $256 \times 256$  pixel resolution; totaling to 756,000 image instances.

### 5.3.2 SINGLE WAVELENGTH FOREARM DATASET

Single wavelength forearm dataset (SWFD) is collected using both multisegment and semi circle arrays (Sec. 5.2.1) from 14 volunteers at a single wavelength (1,064 nm) for both arms. The choice of the wavelength is based on maximizing penetration depth for excitation light source (laser) [34]. Out of the 14 volunteers, eight of them have also participated in the MSFD experiment and their unique identifiers match across the dataset files. For each array, volunteer, and arm, we acquired 1,400 slices, creating a sum of  $2 \times 14 \times 2 \times 1,400 = 78,400$  unique signal matrices. It is important to note that despite the data being acquired from the same volunteers, signals between multisegment array and semi circle array are not paired due to physical constraints.

From this data, using backprojection algorithm, we reconstruct (i) linear array images,  $\text{SWFD}_{\text{linear}}$ , (ii) multisegment array images,  $\text{SWFD}_{\text{ms}}$ , (iii) semi circle array images,  $\text{SWFD}_{\text{sc}}$ , (iv) semi circle array limited 128 images (Sec.5.2.4),  $\text{SWFD}_{\text{sc,lv128}}$ , (v) semi circle sparse 128 images (Sec.5.2.3),  $\text{SWFD}_{\text{sc,ss128}}$ , (vi) semi circle sparse 64 images (Sec.5.2.3),  $\text{SWFD}_{\text{sc,ss64}}$ , (vii) semi circle sparse 32 images (Sec.5.2.3),  $\text{SWFD}_{\text{sc,ss32}}$ , (viii) multisegment sparse 128 images (Sec.5.2.3),  $\text{SWFD}_{\text{ms,ss128}}$ , (ix) multisegment sparse 64 images (Sec.5.2.3),  $\text{SWFD}_{\text{ms,ss64}}$ , and (x) multisegment sparse 32 images (Sec.5.2.3),  $\text{SWFD}_{\text{ms,ss32}}$ , where each dataset has 39,200 images of  $256 \times 256$  pixel resolution; totaling to 392,000 image instances.

### 5.3.3 SIMULATED CYLINDERS DATASET

Simulated cylinders dataset (SCD) is a group of synthetically generated 20,000 forearm acoustic pressure maps that we heuristically produced based on a range of criteria we observed in experimental images. The acoustic pressure maps are generated with  $256 \times 256$  pixel resolution where skin curves and afterwards a random amount of ellipses with different intensity profiles are generated iteratively for a given image (see Fig. 5.2). We explain details for the simulation algorithm<sup>4</sup> for generating acoustic pressure map in Appendix 5.7.

Based on the acoustic pressure map, we generate its annotation map with three labels, corresponding to background, vessels, and skin curve. For each acoustic pressure map, we generate signal matrices for the geometries of linear, multisegment and virtual circle arrays. Using linear and multisegment array signals, we use backprojection algorithm to reconstruct (i) linear array images,  $\text{SCD}_{\text{linear}}$ , and (ii) multisegment array images,  $\text{SCD}_{\text{ms}}$ , (iii) multisegment

<sup>4</sup>Python module for acoustic map simulation: [renkulab.io/gitlab/firat.ozdemir/oa-armsim](https://github.com/renkulab/oa-armsim).

sparse 128 images (Sec.5.2.3),  $SCD_{ms,ss128}$ , (iv) multisegment sparse 64 images (Sec.5.2.3),  $SCD_{ms,ss64}$ , and (v) multisegment sparse 32 images (Sec.5.2.3),  $SCD_{ms,ss32}$ . From virtual circle array signals, we use backprojection algorithm to reconstruct (vi) virtual circle images,  $SCD_{vc}$ , (vii) virtual circle limited 128 images (Sec.5.2.4),  $SCD_{vc,lv128}$ , (viii) virtual circle sparse 128 images (Sec.5.2.3),  $SCD_{vc,ss128}$ , (ix) virtual circle sparse 64 images (Sec.5.2.3),  $SCD_{vc,ss64}$ , and (x) virtual circle sparse 32 images (Sec.5.2.3),  $SCD_{vc,ss32}$ , where each dataset has 20,000 images of  $256 \times 256$  pixel resolution; totaling to 200,000 image instances. All ten image reconstruction variations of SCD have corresponding pairs for each of the 20k image; i.e., produced from the same acoustic pressure map.

### 5.3.4 OADAT-MINI DATASET

Using a subset of 100 signal and corresponding reconstruction instances from each of the previously mentioned datasets, we also present OADAT-mini, which is a fragment of OADAT that is significantly smaller yet comprehensive for understanding the contents of OADAT. In addition, OADAT-mini contains manual annotation maps for vessels in the reconstructed images.

## 5.4 TASKS

Based on the datasets presented in Sec. 5.3, we define a list of experiments related to image translation to overcome (i) sparse sampling and (ii) limited view artifacts, and semantic segmentation of images.

### 5.4.1 IMAGE TRANSLATION

Through a list of permutations of our datasets, we can define several pairs of varying difficulty of image translation experiments where the target images are also available (see Table 5.2). We present sparse sampling and limited view reconstructions of SWFD, MSFD and SCD for all transducer arrays. Sparse sampling correction experiments learn mapping functions listed in Table 5.2, where the function notations denote the dataset used, the task of sparse sampling (ss) correction from the given number of elements used for image reconstruction and the array that is used to generate the input. Limited view correction experiments learn mapping functions, listed in Table 5.2, where the function notations denote the dataset used, the task of limited view (lv) correction from the given number of elements used for image reconstruction and the array that is used to generate the input.

### 5.4.2 SEMANTIC SEGMENTATION

SCD includes pixel annotations for skin curve, vessels and background. In addition to segmentation of these structures on the ideal reconstructions  $SCD_{vc}$ , we define this task on sparse sampling and limited view reconstructions that contain the relevant artifacts encountered in experimental data. Accordingly, we compose the nine segmentation experiments listed in Table 5.2, where the function notations denote the task segmentation (seg), type of the reconstructed input being used (virtual circle (vc), limited view (lv), and sparse sampling (ss))

TABLE 5.2: List of tasks and experiments we define on MSFD, SWFD, and SCD. Experiment names are built of (i) dataset name for translation tasks or seg for segmentation task, (ii) input data and corresponding number of active array elements; sparse sampling (ss), limited view (lv), virtual circle (vc), and (iii) input array type; semi circle (sc), virtual circle (vc), linear (li), multisegment (ms). Image data and annotation maps are represented with  $x$  and  $y$ , while predicted image and annotations are shown as  $x^*$  and  $y^*$ .

Limited view correction	Semantic segmentation
$f_{\text{MSFD\_lv128,li}}: x \sim \text{MSFD}_{\text{linear}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_vc,vc}}: x \sim \text{SCD}_{\text{vc}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_lv128,li}}: x \sim \text{SWFD}_{\text{linear}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_lv128,li}}: x \sim \text{SCD}_{\text{linear}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_lv128,sc}}: x \sim \text{SWFD}_{\text{sc,lv128}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_lv128,vc}}: x \sim \text{SCD}_{\text{vc,lv128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_lv128,li}}: x \sim \text{SCD}_{\text{linear}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_ss128,vc}}: x \sim \text{SCD}_{\text{vc,ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_lv128,vc}}: x \sim \text{SCD}_{\text{vc,lv128}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_ss64,vc}}: x \sim \text{SCD}_{\text{vc,ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_lv256,ms}}: x \sim \text{SCD}_{\text{ms}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_ss32,vc}}: x \sim \text{SCD}_{\text{vc,ss32}} \rightarrow y^* \sim \text{labels}$
<hr/>	
<b>Sparse sampling correction</b>	
$f_{\text{SWFD\_ss128,sc}}: x \sim \text{SWFD}_{\text{sc,ss128}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_ss128,ms}}: x \sim \text{SCD}_{\text{ms,ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss64,sc}}: x \sim \text{SWFD}_{\text{sc,ss64}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_ss64,ms}}: x \sim \text{SCD}_{\text{ms,ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss32,sc}}: x \sim \text{SWFD}_{\text{sc,ss32}} \rightarrow x^* \sim \text{SWFD}_{\text{sc}}$	$f_{\text{seg\_ss32,ms}}: x \sim \text{SCD}_{\text{ms,ss32}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss128,vc}}: x \sim \text{SCD}_{\text{vc,ss128}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_MSFD\_lv128,ms}}: x \sim \text{MSFD}_{\text{linear}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss64,vc}}: x \sim \text{SCD}_{\text{vc,ss64}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_MSFD\_ss128,ms}}: x \sim \text{MSFD}_{\text{ms,ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss32,vc}}: x \sim \text{SCD}_{\text{vc,ss32}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_MSFD\_ss64,ms}}: x \sim \text{MSFD}_{\text{ms,ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss128,ms}}: x \sim \text{SWFD}_{\text{ms,ss128}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_MSFD\_ss32,ms}}: x \sim \text{MSFD}_{\text{ms,ss32}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss64,ms}}: x \sim \text{SWFD}_{\text{ms,ss64}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_lv128,ms}}: x \sim \text{SWFD}_{\text{linear}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SWFD\_ss32,ms}}: x \sim \text{SWFD}_{\text{ms,ss32}} \rightarrow x^* \sim \text{SWFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_lv128,sc}}: x \sim \text{SWFD}_{\text{sc,lv128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss128,ms}}: x \sim \text{SCD}_{\text{ms,ss128}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_SWFD\_ms,ms}}: x \sim \text{SWFD}_{\text{ms}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss64,ms}}: x \sim \text{SCD}_{\text{ms,ss64}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_SWFD\_sc}}: x \sim \text{SWFD}_{\text{sc}} \rightarrow y^* \sim \text{labels}$
$f_{\text{SCD\_ss32,ms}}: x \sim \text{SCD}_{\text{ms,ss32}} \rightarrow x^* \sim \text{SCD}_{\text{vc}}$	$f_{\text{seg\_SWFD\_ss128,sc}}: x \sim \text{SWFD}_{\text{sc,ss128}} \rightarrow y^* \sim \text{labels}$
$f_{\text{MSFD\_ss128,ms}}: x \sim \text{MSFD}_{\text{ms,ss128}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_ss64,sc}}: x \sim \text{SWFD}_{\text{sc,ss64}} \rightarrow y^* \sim \text{labels}$
$f_{\text{MSFD\_ss64,ms}}: x \sim \text{MSFD}_{\text{ms,ss64}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_ss32,sc}}: x \sim \text{SWFD}_{\text{sc,ss32}} \rightarrow y^* \sim \text{labels}$
$f_{\text{MSFD\_ss32,ms}}: x \sim \text{MSFD}_{\text{ms,ss32}} \rightarrow x^* \sim \text{MSFD}_{\text{ms}}$	$f_{\text{seg\_SWFD\_ss128,ms}}: x \sim \text{SWFD}_{\text{ms,ss128}} \rightarrow y^* \sim \text{labels}$
	$f_{\text{seg\_SWFD\_ss64,ms}}: x \sim \text{SWFD}_{\text{ms,ss64}} \rightarrow y^* \sim \text{labels}$
	$f_{\text{seg\_SWFD\_ss32,ms}}: x \sim \text{SWFD}_{\text{ms,ss32}} \rightarrow y^* \sim \text{labels}$

and the array that is used to generate the input. All data is generated from SCD and the objective is to match the ground truth annotations of the acoustic pressure map. Different than SCD, experimental datasets under OADAT-mini include pixel annotations for vessels and consist of 14 segmentation experiments, also listed in Table 5.2.

## 5.5 EXPERIMENTS AND RESULTS

For all experiments we standardize the architecture that we based on UNet [35]. Specifically, we adopt the five spatial feature abstraction levels and use skip connections to concatenate with features of matching spatial dimension along the upsampling path. However, we make several distinct design choices that vary from vanilla UNet. First, we use attention gates [36] at the end of each skip connection. Second, we opt for residual convolutional blocks with batch normalization [37] at each level, shown in Fig. 5.3. Third, we use two-dimensional (2D) bilinear upsampling instead of deconvolutions. Finally, we use half

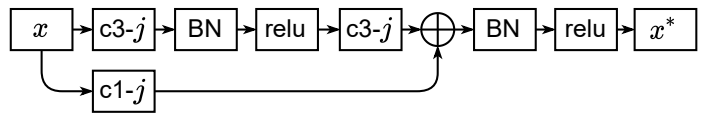


FIGURE 5.3: Residual convolutional block with batch normalization (BN).  $ci-j$  conv. layer have  $i \times i$  kernels and  $j$  filters.

we use two-dimensional (2D) bilinear upsampling instead of deconvolutions. Finally, we use half

the number of convolutional kernels at each layer; e.g., start with 32 convolutional filters as opposed to 64. Full schematic as well as other implementation details are discussed in Appendix 5.7. We refer to our modified UNet architecture as modUNet hereon.

### 5.5.1 DATA SPLIT AND PREPROCESSING

We standardize how we split each dataset into training and test sets regardless of the task in order to ensure consistency in our and future experiments. Out of the nine volunteers in MSFD, we use five for training (IDs: 2, 5, 6, 7, 9), one for validation (ID: 10) and three for testing (IDs: 11, 14, 15). Out of the 14 volunteers in SWFD, we use eight for training (IDs: 1, 2, 3, 4, 5, 6, 7, 8), one for validation (ID: 9) and five for testing (IDs: 10, 11, 12, 13, 14). Out of the 20k slices in SCD, we use the first 14k for training, following 1k for validation, and the last 5k for testing. For each experiment conducted on OADAT-mini, we use the first 75 samples for training, next 5 for validation and the last 20 for quantitative evaluation. This translates to six times the numbers for MSFD-mini, where there are 100 samples for each of the six wavelengths. As before, we conduct the MSFD-mini segmentation experiments as a single modUNet attempting segment any of the given six wavelength samples. As a preprocessing step, all data instances (except for annotation maps) are independently scaled by their maximum and then clipped at a minimum value of  $-0.2$  [38].

### 5.5.2 RESULTS

We evaluate modUNet performance on the test sets using standard metrics. Namely, we report mean absolute error (MAE), root mean squared error (RMSE), structural similarity index (SSIM), and peak signal to noise ratio (PSNR) for image translation experiments between modUNet predictions and targets. Segmentation task performance is reported using Dice coefficient (F1-score), intersection over union (IoU) (i.e., Jaccard index) and 95-percentile Hausdorff distance (HD95) metrics between modUNet predictions and annotation maps for vessels and skin curves. HD95 is calculated in several steps: First, the set of pixels along the contour for each predicted (set A) and annotated (set B) target structures are found. For each set point, the closest point from the other set is determined based on  $l_2$  distance. Different from the standard Hausdorff distance, the 95-percentile distance value is taken as the directional distance from set A to B (and vice versa) instead of the maximum distance. Then the maximum of these two values is calculated as the symmetric HD95 for a

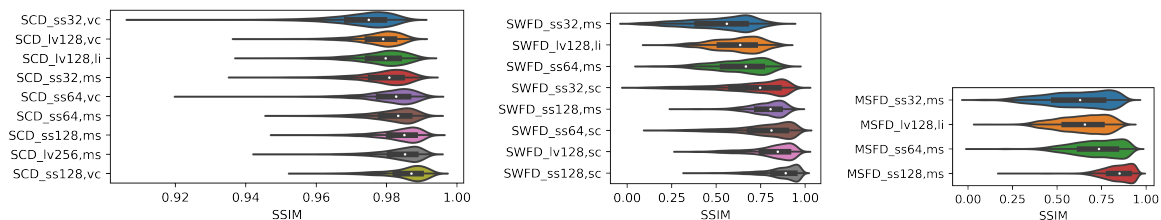


FIGURE 5.4: Distribution of modUNet structural similarity index (SSIM) performance on simulated cylinders dataset (SCD)(left), single wavelength forearm dataset (SWFD) (middle) and multispectral forearm dataset (MSFD) (right) image translation experiments, sorted in ascending median test sample performance.

TABLE 5.3: Image translation results of the proposed modUNet model reported as mean  $\pm$ std. Each row corresponds to the results of the experiment where input data is identified through (i) the input data and corresponding number of active transducer elements; sparse sampling (ss), limited view (lv) and (ii) the array type used for input; semi circle (sc), virtual circle (vc), linear (li), multisegment (ms).

		MAE	RMSE	SSIM	PSNR
<b>SCD</b>					
Limited view	lv128,li	0.005 $\pm$ 1.3e-3	0.012 $\pm$ 3.3e-3	0.978 $\pm$ 7.2e-3	39.909 $\pm$ 2.15
	lv128,vc	0.005 $\pm$ 1.3e-3	0.013 $\pm$ 3.4e-3	0.978 $\pm$ 6.2e-3	39.853 $\pm$ 2.18
	lv256,ms	0.005 $\pm$ 1.2e-3	0.009 $\pm$ 3.0e-3	0.984 $\pm$ 6.1e-3	42.459 $\pm$ 2.52
Sparse view	ss128,vc	0.004 $\pm$ 1.1e-3	0.008 $\pm$ 2.0e-3	0.985 $\pm$ 6.1e-3	44.068 $\pm$ 2.12
	ss64,vc	0.005 $\pm$ 1.3e-3	0.011 $\pm$ 2.9e-3	0.981 $\pm$ 6.9e-3	41.402 $\pm$ 2.25
	ss32,vc	0.006 $\pm$ 1.5e-3	0.013 $\pm$ 3.7e-3	0.973 $\pm$ 8.7e-3	39.309 $\pm$ 2.23
	ss128,ms	0.005 $\pm$ 1.1e-3	0.010 $\pm$ 3.0e-3	0.984 $\pm$ 6.1e-3	42.345 $\pm$ 2.39
	ss64,ms	0.005 $\pm$ 1.2e-3	0.010 $\pm$ 3.0e-3	0.982 $\pm$ 6.7e-3	41.525 $\pm$ 2.19
	ss32,ms	0.005 $\pm$ 1.3e-3	0.011 $\pm$ 3.0e-3	0.979 $\pm$ 7.3e-3	40.678 $\pm$ 2.10
<b>SWFD</b>					
Limited view	lv128,li	0.028 $\pm$ 1.6e-2	0.039 $\pm$ 2.0e-2	0.613 $\pm$ 1.4e-1	29.397 $\pm$ 4.75
	lv128,sc	0.016 $\pm$ 1.2e-2	0.021 $\pm$ 1.5e-2	0.811 $\pm$ 1.3e-1	36.791 $\pm$ 5.30
Sparse view	ss128,sc	0.015 $\pm$ 1.2e-2	0.019 $\pm$ 1.5e-2	0.863 $\pm$ 1.0e-1	38.233 $\pm$ 5.35
	ss64,sc	0.019 $\pm$ 1.5e-2	0.024 $\pm$ 1.8e-2	0.769 $\pm$ 1.6e-1	35.605 $\pm$ 5.22
	ss32,sc	0.021 $\pm$ 1.6e-2	0.028 $\pm$ 2.0e-2	0.693 $\pm$ 2.0e-1	33.852 $\pm$ 5.37
	ss128,ms	0.023 $\pm$ 1.5e-2	0.028 $\pm$ 1.9e-2	0.784 $\pm$ 9.7e-2	33.764 $\pm$ 4.53
	ss64,ms	0.029 $\pm$ 1.9e-2	0.037 $\pm$ 2.3e-2	0.636 $\pm$ 1.5e-1	31.311 $\pm$ 4.80
	ss32,ms	0.033 $\pm$ 2.1e-2	0.042 $\pm$ 2.5e-2	0.521 $\pm$ 1.8e-1	29.813 $\pm$ 5.11
<b>MSFD</b>					
limited view	lv128,li	0.023 $\pm$ 1.1e-2	0.035 $\pm$ 1.4e-2	0.640 $\pm$ 1.4e-1	29.731 $\pm$ 3.97
Sparse view	ss128,ms	0.017 $\pm$ 9.7e-3	0.022 $\pm$ 1.2e-2	0.839 $\pm$ 8.1e-2	35.798 $\pm$ 3.84
	ss64,ms	0.022 $\pm$ 1.2e-2	0.029 $\pm$ 1.5e-2	0.719 $\pm$ 1.4e-1	33.104 $\pm$ 4.02
	ss32,ms	0.026 $\pm$ 1.4e-2	0.036 $\pm$ 1.7e-2	0.608 $\pm$ 1.8e-1	30.873 $\pm$ 4.22

given image. In OADAT-mini experiments, only vessel annotation maps are available. Since OADAT-mini consist of subsets of the other three datasets, we do not repeat image translation tasks. In Tables 5.3 & 5.4 we report modUNet results for mean and standard deviations aggregated over the corresponding test set images.

Using SSIM, we show performance across all our datasets in Fig. 5.4. Upon exploring the reason behind the long tails, we notice that most of the lower scores occur when acquisition noise and/or artifacts are more pronounced. Depending on the sample, this can imply either modUNet reduced the noise present in the target, or both input/output pair in the test set had low signal-to-noise ratio (SNR). Nevertheless, modUNet successfully corrects geometric distortions for limited view experiments. Given the low mean and standard deviations in MAE and RMSE, we can comment that modUNet can generalize well to previously unseen volunteer data. This is further corroborated with the narrow SSIM interquartile range in Fig. 5.4 violin plots.

Similarly, we plot the segmentation performance for IoU across different experiments on SCD

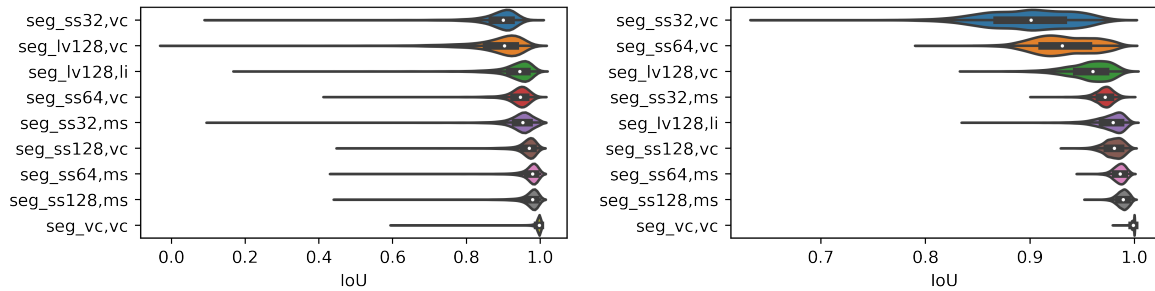


FIGURE 5.5: Distribution of modUNet intersection over union (IoU) performance on simulated cylinders dataset (SCD) semantic segmentation experiments for vessel (left) and skin curve (right) labels, sorted in ascending median test sample performance.

in Fig. 5.5 for vessel and skin curve labels. While skin curve segmentation performance almost never drops below IoU of 0.80 for SCD, one can see that IoU can be drastically lower for vessel segmentation. We observed that this only happens when the size of vessel is small. For example, there are examples with ground truth vessels having as low as four pixels while the prediction has six, leading to an IoU of 0.6. As for experiments with OADAT-mini, vessel segmentation performance in experimental datasets are worse. Specifically, the worst scores are observed for MSFD-mini experiments. This is due to the movement artifact between different wavelengths of a given slice. All MSFD-mini instances have expert annotations for 800 nm wavelength. Slight movement across different wavelengths can yield poor quantitative metrics, particularly exacerbated when the observed vessels are small. Given the limited training and test sizes of OADAT-mini experiments, we believe that the quantitative results should be taken as a reference. The qualitative results in the Appendix can be more informative for gaining insight for modUNet performance when trained with a very limited amount of data. We provide qualitative results as well as conduct further analysis for all tasks in Appendix 5.7.

## 5.6 DISCUSSION

Major differences exist between simulated and experimental datasets. Even if the content is different, training and test samples of the simulated dataset are inherently sampled from the same distribution. On the other hand, experimental datasets feature shifts due to different volunteers being imaged, inherent noise from data acquisition system, and difference in directional sensitivity resulting from transducer alignment and positioning of the hand-held probe. Furthermore, despite the efforts to avoid corrupted acquisitions during the data collection, experimental datasets still contain samples with relatively low signal-to-noise ratio. Such samples are expected to yield reduced performance metrics for image translation tasks due to significant mismatch between the predicted and noisy target images. In a clinical setting, a medical expert typically repeats an acquisition if they deem the signal quality is significantly lower than expected. However, beyond this subjective filtering step, one needs to make sure that even the worst results are either sufficiently good or their poor performance can be attributed to a cause. Accordingly, we further analyze some of the worst samples in the Appendix and believe that this should be a standard for future work. Provided dataset

TABLE 5.4: Segmentation results of our proposed modUNet model reported as mean  $\pm$ std. Each row corresponds to the results of the experiment where input data is identified through (i) the input data and corresponding number of active transducer elements; sparse sampling (ss), limited view (lv), virtual circle (vc), semi circle (sc), multisegment (ms) and (ii) the array type used for input; virtual circle (vc), semi circle (sc), multisegment (ms), and linear (li). single wavelength forearm dataset (SWFD)- and multispectral forearm dataset (MSFD)-mini correspond to experiment conducted on OADAT-mini dataset.

		Dice		IoU		HD95	
		vessels	skin curve	vessels	skin curve	vessels	skin curve
<b>SCD</b>							
Full view	vc,vc	0.996 $\pm$ 9.8e-3	0.999 $\pm$ 8.7e-4	0.993 $\pm$ 1.8e-2	0.999 $\pm$ 1.7e-3	0.792 $\pm$ 1.0e+0	0.164 $\pm$ 2.0e+0
Limited view	lv128,li	0.963 $\pm$ 3.4e-2	0.988 $\pm$ 7.3e-3	0.931 $\pm$ 5.5e-2	0.976 $\pm$ 1.4e-2	2.734 $\pm$ 2.9e+0	2.196 $\pm$ 7.4e+0
	lv128,vc	0.933 $\pm$ 5.8e-2	0.978 $\pm$ 1.1e-2	0.878 $\pm$ 8.6e-2	0.957 $\pm$ 2.1e-2	4.310 $\pm$ 8.9e+0	7.372 $\pm$ 1.4e+1
Sparse view	ss128,vc	0.980 $\pm$ 2.3e-2	0.990 $\pm$ 4.7e-3	0.961 $\pm$ 4.0e-2	0.980 $\pm$ 9.1e-3	2.499 $\pm$ 5.2e+0	3.897 $\pm$ 1.0e+1
	ss64,vc	0.967 $\pm$ 2.8e-2	0.965 $\pm$ 1.5e-2	0.937 $\pm$ 4.6e-2	0.933 $\pm$ 2.8e-2	2.738 $\pm$ 3.0e+0	9.634 $\pm$ 1.5e+1
	ss32,vc	0.938 $\pm$ 3.9e-2	0.946 $\pm$ 2.3e-2	0.886 $\pm$ 6.0e-2	0.899 $\pm$ 4.1e-2	3.074 $\pm$ 3.1e+0	12.220 $\pm$ 1.6e+1
	ss128,ms	0.983 $\pm$ 2.4e-2	0.994 $\pm$ 2.9e-3	0.968 $\pm$ 4.1e-2	0.989 $\pm$ 5.6e-3	2.289 $\pm$ 2.4e+0	5.311 $\pm$ 1.3e+1
	ss64,ms	0.984 $\pm$ 2.4e-2	0.993 $\pm$ 3.1e-3	0.969 $\pm$ 4.1e-2	0.985 $\pm$ 6.1e-3	2.772 $\pm$ 8.2e+0	2.668 $\pm$ 8.5e+0
	ss32,ms	0.971 $\pm$ 2.8e-2	0.985 $\pm$ 4.8e-3	0.945 $\pm$ 4.5e-2	0.971 $\pm$ 9.2e-3	3.026 $\pm$ 7.0e+0	6.795 $\pm$ 1.5e+1
<b>SWFD-mini</b>							
Full view	sc,sc	0.857 $\pm$ 5.7e-2	N/A	0.755 $\pm$ 8.5e-2	N/A	17.225 $\pm$ 2.8e+1	N/A
	ms,ms	0.842 $\pm$ 5.9e-2	N/A	0.732 $\pm$ 8.6e-2	N/A	17.218 $\pm$ 2.6e+1	N/A
Limited view	lv128,li	0.794 $\pm$ 6.4e-2	N/A	0.662 $\pm$ 8.3e-2	N/A	28.685 $\pm$ 3.9e+1	N/A
	lv128,sc	0.843 $\pm$ 5.5e-2	N/A	0.732 $\pm$ 7.8e-2	N/A	24.444 $\pm$ 2.6e+1	N/A
Sparse view	ss128,sc	0.864 $\pm$ 4.3e-2	N/A	0.763 $\pm$ 6.5e-2	N/A	23.146 $\pm$ 3.0e+1	N/A
	ss64,sc	0.841 $\pm$ 1.1e-1	N/A	0.737 $\pm$ 1.3e-1	N/A	20.216 $\pm$ 2.8e+1	N/A
	ss32,sc	0.864 $\pm$ 4.1e-2	N/A	0.762 $\pm$ 6.3e-2	N/A	26.832 $\pm$ 3.0e+1	N/A
	ss128,ms	0.836 $\pm$ 5.8e-2	N/A	0.722 $\pm$ 8.3e-2	N/A	17.909 $\pm$ 2.8e+1	N/A
	ss64,ms	0.837 $\pm$ 5.6e-2	N/A	0.723 $\pm$ 8.0e-2	N/A	15.811 $\pm$ 2.3e+1	N/A
	ss32,ms	0.809 $\pm$ 6.0e-2	N/A	0.684 $\pm$ 8.2e-2	N/A	20.603 $\pm$ 3.1e+1	N/A
<b>MSFD-mini</b>							
Limited view	lv128,li	0.474 $\pm$ 1.3e-1	N/A	0.320 $\pm$ 1.2e-1	N/A	20.134 $\pm$ 3.3e+1	N/A
Sparse view	ss128,ms	0.563 $\pm$ 1.0e-1	N/A	0.400 $\pm$ 1.1e-1	N/A	14.312 $\pm$ 1.9e+1	N/A
	ss64,ms	0.572 $\pm$ 1.2e-1	N/A	0.411 $\pm$ 1.3e-1	N/A	19.716 $\pm$ 2.1e+1	N/A
	ss32,ms	0.639 $\pm$ 1.4e-1	N/A	0.485 $\pm$ 1.5e-1	N/A	11.499 $\pm$ 1.7e+1	N/A

is limited to one body part of volunteers without any known health issues. The image reconstruction methods for OA imaging can be applied on any other body part or imaging setup as they solve the same physical inverse problem. The image translation algorithms for sparse acquisition and limited view problems can be adapted for different devices and acquisitions by another clinician/technician at another center, as the streak artifacts originating from sparse acquisition and limited view follow the same pattern.

We envision that future research will tackle additional challenges such as unsupervised or weakly supervised domain adaptation across the datasets provided in this work. There are initial studies to correct limited view artifacts in OA using transfer learning between simulated and experimental datasets after domain adaptation [20]. Similarly, transfer learning between simulated and experimental domains can enhance segmentation performance of the vessels and skin curve in clinical images. Using properties of the detected tissues for fluence correction and heterogeneous SoS image reconstructions would then yield more accurate and quantitative images [39]. We anticipate additional contributions in the field of representation learning using OADAT, e.g. through self-supervised learning, could allow overcoming bottlenecks for specialized downstream tasks with limited amount of task-specific available data.



The multispectral datasets with paired images across multiple wavelength acquisitions are expected to facilitate the investigation of generative modeling of multispectral signals from a given wavelength. We also anticipate future work to explore novel multispectral unmixing approaches using MSFD, enabling more accurate quantification of oxygenation, melanin and lipid content of the tissues. Finally, the provided raw signal data, not available in commercial devices, has a high value for data-driven research. For example, it can serve to benchmark methodologies in image reconstruction e.g., based on variational networks with loop unrolling [40], as well as ultra-fast imaging through adaptive channel sampling.

## 5.7 CONCLUSION

In this work, we provide experimental and synthetic clinical OA data covering a large variety of examples to be used as common ground to compare established and new data processing methods. The datasets correspond to samples from volunteers of varying skin types, different OA array geometries, and several illumination wavelengths. A subset of this experimental data is annotated by an expert. The dataset is supplemented with simulated samples containing ground truth acoustic pressure maps, annotations, and combine pairs of samples reconstructed with different OA array geometries. We define a set of 44 experiments tackling major challenges in the OA field and provide reconstructions of the images under these scenarios along with their corresponding ground truths. We propose and release 44<sup>5</sup> trained neural networks that achieve a good performance for all these examples which can be used as baselines for future improvements. Additional problems can further be defined with the data provided, such as the effects of random sparse sampling or the presence of noise in the signal matrices prior to reconstruction. We believe that these datasets and benchmarks will play a significant role in fostering coordinated efforts to solve major challenges in OA imaging.

---

<sup>5</sup>Pretrained model weights and various scripts to train and evaluate modUNet are available at <https://renkulab.io/gitlab/firat.ozdemir/oadat-evaluate>.

#### ACKNOWLEDGEMENT

This work was supported by Swiss Data Science Center grant C19-04.

#### BROADER IMPACT STATEMENT

The dataset is anonymized by randomly assigning identity numbers for each volunteer. The true identities will not be accessible by any third party now or in the future. The dataset should not be used to draw any medical conclusions. The purpose of the dataset is to help researchers to develop new image processing methods and provide benchmark scores.

#### AUTHOR CONTRIBUTIONS

B.L. setup and acquired raw datasets, B.L. and F.O. analyzed and parsed datasets, F.O. conducted numerical experiments based on the defined tasks. All authors reviewed the manuscript.

## APPENDIX A. OPTOACOUSTIC IMAGING BACKGROUND

OA imaging setups contain light source, transducer array, data acquisition system and workstation PC. Usually, nanosecond-duration pulsed lasers are used as light sources. Alternatively, light emitting diodes (LEDs) can be used with high repetition rates (shorter pulses) as cheaper alternatives to lasers. The excitation of objects/tissues with short, pulsed light sources results in heating and expansion (thermoelastic expansion) of the materials. This expansion generates OA signals (US waves) that propagate to various directions in a 3D imaging medium (see 5.7 for detailed explanation of pressure waves). Transducer arrays are used to capture the propagating waves in the imaging medium. Piezocomposite elements in these arrays convert mechanical pressure waves into electrical signals (see 5.7 for transducer array specifications). After the detection and conversion, the electrical signals are digitized by the data acquisition systems at the defined sampling rate. The signals are acquired and digitized by all the elements in a transducer array simultaneously. The OA acquisition generates time domain (raw) signals acquired from each transducer element. Image reconstruction is performed by solving the inverse problem of signal to image domain conversion. In this study, we used backprojection method to reconstruct OA images (see 5.7 for the detailed explanation of the method and used algorithms). Overall data acquisition system and experimental design are explained in 5.7.

### A.1. OPTOACOUSTIC WAVE EQUATIONS

OA imaging is based on thermoelastic expansion of the tissues which results in propagation of pressure waves in imaging medium depending on spatial and temporal changes. The OA wave equation can be written as follows

$$\frac{\partial^2 p(r,t)}{\partial t^2} - c^2 \nabla^2 p(r,t) = \Gamma H(r,t) \frac{\partial \delta(t)}{\partial t}$$

where  $r$  and  $t$  are the spatial and temporal variables, respectively.  $\Gamma$  is the Grüneisen constant [41].  $c$  stands for speed of sound.  $H(r,t)$  is the absorbed energy field based on the location and time of the sample.  $\delta(t)$  stands for temporal laser light intensity change based on the illumination.  $p(r,t)$  represents the pressure wave dependent on spatial and temporal variables. The Poisson solution of OA wave equation for pressure wave can be written as

$$p(r,t) = \frac{\Gamma}{4\pi c} \frac{\partial}{\partial t} \int_{S'} \frac{H(r')}{|r-r'|} dS'$$

where  $S'$  is the time dependent spherical surface defined by  $|r - r'| = ct$ . This equation represents OA forward model which inverse problem of reconstruction can be derived. OA images are reconstructed by absorbed energy field  $H(r')$  at specific location based on measured pressure waves.  $H(r')$  is calculated from detected pressure waves at the surface  $S$  as follows

$$H(r') = \frac{1}{\Gamma} \int_{\Omega} \frac{d\Omega}{\Omega} \left[ 2p(r,t) - 2t \frac{\partial p(r,t)}{\partial t} \right]_{t=\frac{r-r'}{c}}$$

The constants at this equation can be omitted. After omitting the constants in the formula, the equation is discretized as

$$H(r'_j) = \sum \left[ p(r_i, t_{ij}) - t_{ij} \frac{\partial p(r_i, t_{ij})}{\partial t} \right]$$

where  $r'_j$  is the  $j$ -th point on the defined reconstruction grid,  $r_i$  is the position of  $i$ -th transducer and  $t_{ij} = |r_i - r'_j|/c$ . In summary, the equation calculates the distance between a point on the defined grid and an element of the transducer array. Then, it finds the corresponding wave intensity in signal (or defined as surface) based on the time of flight calculated using speed of sound in the imaging medium.

## A.2. TRANSDUCER ARRAY DETAILS

**Semi circle:** The array contains 256 piezocomposite transducers distributed over a semi circle (concave surface) equidistantly with the radius of 40 mm (Fig. 5.1b, main manuscript). The single transducer elements have dimensions of 0.37 mm  $\times$  15 mm with inter-element distance of 0.10 mm. This configuration of transducer elements results in cylindrical (toroidal) focusing at 38 mm (close to the center of the array). The central peak frequency of array is 5 MHz with 60% bandwidth at -6 dB.

**Multisegment:** The array is formed by the combination of a linear detector array and concave parts on the right and left sides as shown in Fig. 5.1c (main manuscript). The linear part contains 128 elements distributed on a linear surface with inter-element pitch size of 0.25 mm. Both of the concave parts include 64 elements which make the total number of elements equal to 256 (128 linear + 128 concave). The inter-element pitch size of concave part is 0.6 mm with 40 mm radius of curvature. The height of all elements are equal to 10 mm. Concave parts are designed to increase angular coverage in OA imaging. This configuration results in a cylindrical focusing at 38 mm close to the center of the array. The array has 7.5 MHz central frequency with 70% bandwidth at -6 dB.

**Linear array:** The array is central part of the multisegment array with 128 transducer elements distributed over a line with pitch size of 0.25 mm (Fig. 5.1c, main manuscript). Similar to concave parts, the linear array has 7.5 MHz central frequency with 70% bandwidth at -6 dB. The linear array is optimized for US data acquisitions with planar waves. Hence, the array produces OA images with limited view artifacts due to reduced angular coverage which is a limiting factor for OA image acquisitions.

**Virtual circle:** The array is generated to simulate images with 360 degree angular coverage which results in artifact free reconstructions (Fig. 5.2a, main manuscript). It contains 1,024 transducer elements distributed over a full circle with equal distance. The radius of the transducer array is kept equal to semi circle array (40 mm) to allow comparison between simulations and experimental acquisitions.

## A.3. OPTOACOUSTIC IMAGE RECONSTRUCTION

A Python package “pyoat” is presented with the datasets to reconstruct images from raw signals. The library uses “pip” package manager to install and use the functions. The

package provides functions to bandpass filter and normalize raw signals as preprocessing step. Implementation of backprojection algorithm is also included in the package (Sec. 5.2.2, main manuscript). The forward model operator is implemented to simulate signals from acoustic pressure maps. Data readers and savers are integrated into the package for loading the raw data and saving the reconstructed images, respectively. The package provides examples to use different functions. The examples are easy to use Python scripts which require only raw signal data paths as input. Positions of elements in all transducer arrays used in this study are included in the library to enable reconstruction from raw signals.

#### A.4. EXPERIMENTAL DATA ACQUISITION SETUP

Signal acquisition is performed with OA imaging setup that combines four main components, namely, transducer arrays, nanosecond pulsed lasers, data acquisition system and workstation PC (Fig. 5.1, main manuscript). SWFD dataset was acquired with multisegment and semi circle transducer arrays (Sec. 5.2.1, main manuscript) using a nanosecond laser at 1,064 nm with repetition rate of 10 Hz. MSFD dataset was acquired with only multisegment array at six different wavelengths (700, 730, 760, 780, 800, 850 nm) using the laser with repetition rate of 50 Hz. The increased repetition rate guarantees that displacement within the scene between the frames of different wavelengths are minimal. Data acquisition system is used to digitize signals acquired by the transducer arrays. A sampling rate of 40 mega samples per second (MSPS) was used in all experiments. Then, digital signals are sent to workstation PC to store the data and display both raw signals and reconstructed images in real time with lower resolution. The high resolution images presented in this work were reconstructed after the acquisition (offline) using the backprojection method (Sec. 5.2.2, main manuscript). The real time feedback helps to position the transducer arrays. The workstation PC also synchronizes all the imaging setup components by setting delays for data acquisition system and triggers for laser pulses. Imaging medium was filled with water to increase coupling efficiency between the transducer arrays and the skin. Water has low attenuation at the wavelengths used in this study. Hence, the signal attenuation in the medium was low enough to be neglected. Transducer arrays are held orthogonal to the forearm surface throughout acquisition and swiftly moved from elbow towards wrist. All participants joined the experiments voluntarily and were informed about details of the experiments.

#### APPENDIX B. FITZPATRICK SKIN PHOTOTYPE IN EXPERIMENTAL DATASETS

Fitzpatrick skin phototype is a metric to quantify the amount of melanin pigment in the skin of a subject [31]. The metric ranges from one to six, going from pale white skin to black skin color. This is relevant for various OA applications due to different skin types (melanin concentration) lead to varying amount of contrast (absorption) at skin surfaces in the acquired images. Accordingly, the distribution of the skin types of the volunteers across SWFD and MSFD are shown in Fig. 5.6.

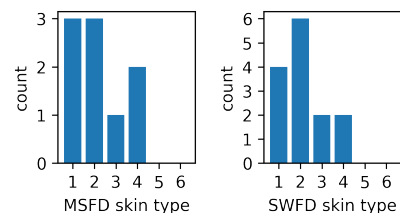


FIGURE 5.6: Fitzpatrick skin phototype [31] distribution of volunteers in datasets.

## APPENDIX C. ARCHITECTURE AND IMPLEMENTATION DETAILS

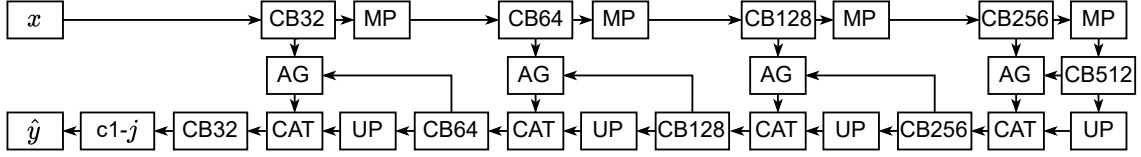


FIGURE 5.7: Schematic of the proposed modUNet architecture.  $CB_j$  represents the residual two-dimensional (2D) convolutional block with batch normalization shown in Fig. 5.3 (main manuscript) where each convolution has  $j$  filters. Other abbreviations correspond to 2D-maxpooling (MP) of poolsize 2, 2D bilinear upsampling (UP) by a factor of 2, concatenation (CAT), and attention gates (AGs) [36]. Finally,  $c1-j$  represents a convolutional layer of  $j$  filters (1 for image translation and 3 for semantic segmentation experiments) and  $1 \times 1$  kernels without activation.

We show the schematic of the proposed modUNet architecture in Fig. 5.7.

We use categorical cross entropy loss and mean squared error (MSE) loss for segmentation and image translation experiments, respectively. We add an additional  $l_1 = l_2 = 0.01$  regularization weight for each learned model parameter.

For all experiments, we use Adam optimizer and scale the learning rate with exponential decay (decay rate of 0.98 and decay steps of 1,000) from a peak of 0.0001 following a linear warmup of 10,000 steps. Except for experiments on OADAT-mini, all models are trained for 150 epochs with a mini-batch size of 25 and best validation set loss is used as the early stopping criteria for which test set performance metrics are presented. Experiments done on OADAT-mini are optimized for 10,000 epochs. For each experiment, we used an Nvidia P100 or Titan X GPU with 12 GB memory to train the model. We do not augment data during training. Experiments on MSFD, SWFD, and SCD took about, 61, 32 and 15 hours, respectively. The training time variation across datasets is due to the difference in size of the datasets.

## APPENDIX D. QUALITATIVE RESULTS

Despite the impressive performance of the modUNet, there is a long tail in almost all tasks for SSIM or IoU distributions as seen in Figs. 5.4 and 5.5 (main manuscript). Accordingly, we qualitatively examine some of the worst samples at the bottom of these tails for each task.

## D.1. IMAGE TRANSLATION

We showcase and analyze a selection of the worst performing samples from the most challenging limited view and sparse sampling reconstruction experiments on SWFD and report SSIM for these samples with respect to the target sample.

In Figs. 5.8 & 5.9 (experiments  $f_{\text{SWFD}_{\text{lv128,li}}}$  and  $f_{\text{SWFD}_{\text{lv128,sc}}}$ ) one can see that some of the worst samples based on SSIM correspond to images with little to no signal content. The

remaining sample from the 10th percentile in Fig. 5.8 suggests that low SSIM in multisegment array images can be deceiving due to different set of artifacts that appear on the target (multisegment) array. In Fig. 5.9, where the target is the semi circle array, the artifacts are not as intense. Nevertheless, given the accurate vessel geometry correction, low SSIMs in 1st, 5th and 10th percentiles can be attributed to other local noise patterns, such as the circular noise that exist on the target image. Thanks to the majority of samples in our datasets having high SNR, most of the undesirable artifacts that seldomly occur in test samples are removed. For example, in Figs. 5.9 & 5.10, columns with relatively higher signal content are either very close to the target or even more artifact-free.

In Figs. 5.11 & 5.12, we show some of the samples with the worst SSIM from the experiments  $f_{\text{SWFD\_ss32,sc}}$  and  $f_{\text{SWFD\_ss32,ms}}$ , respectively. The samples in the left two columns for both experiments have very poor SNR. For all other samples, modUNet predictions from the heavily subsampled image reconstructions indicate higher SNR than the target samples thanks to removing most of the noise pattern (especially visible within the water medium prior to/above the forearm) while keeping all soft tissue structures intact and faithful to their counterparts in the target image. Unfortunately, mismatch of the noise patterns causes low SSIM for such samples with low SNR target images.

For completeness, we show the worst performing samples for SSIM in image translation experiments using SCD in Fig. 5.13. One can see that the rounded SSIM already reaches 1.0 by the 1st- (right figure) or 5th- (left figure) percentile performing sample in the test set. In the left figure 2nd column, one can see that the strong artifact passing through the small vessel caused modUNet to separate the vessel into two pieces, shown with a red arrow.

In addition, we also showcase some of the best performing samples for SSIM in Figs. 5.14, 5.15, 5.16, 5.17 & 5.18. Beyond correcting distorted vessel geometry, it can be seen in some samples that certain vessels that are barely visible get accurately redrawn to match the target (e.g., Fig. 5.9 3rd and 4th columns, Fig. 5.11 2nd, 3rd and 4th column, Fig. 5.12 3rd and 4th columns, and Fig. 5.18 4th column).

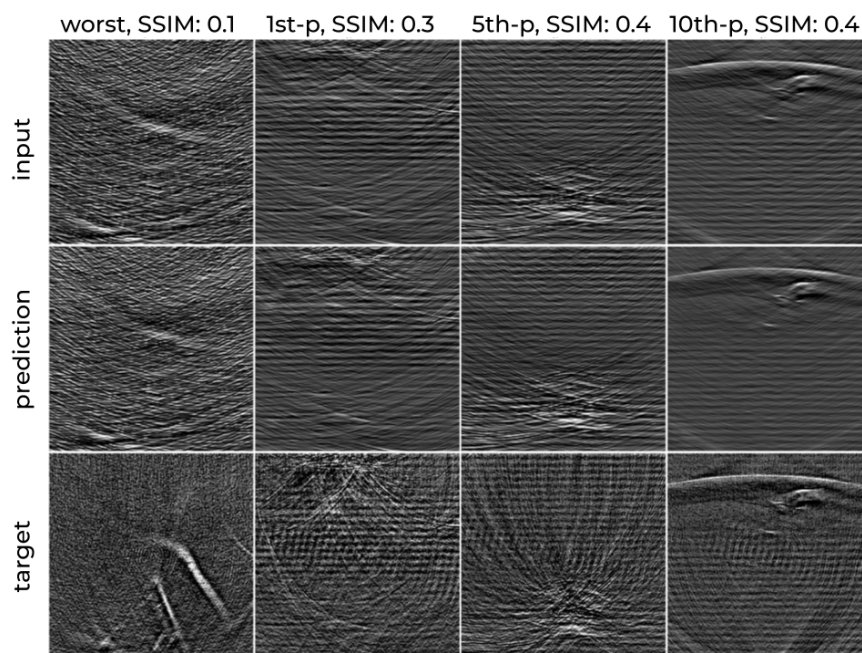


FIGURE 5.8: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{l_v128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

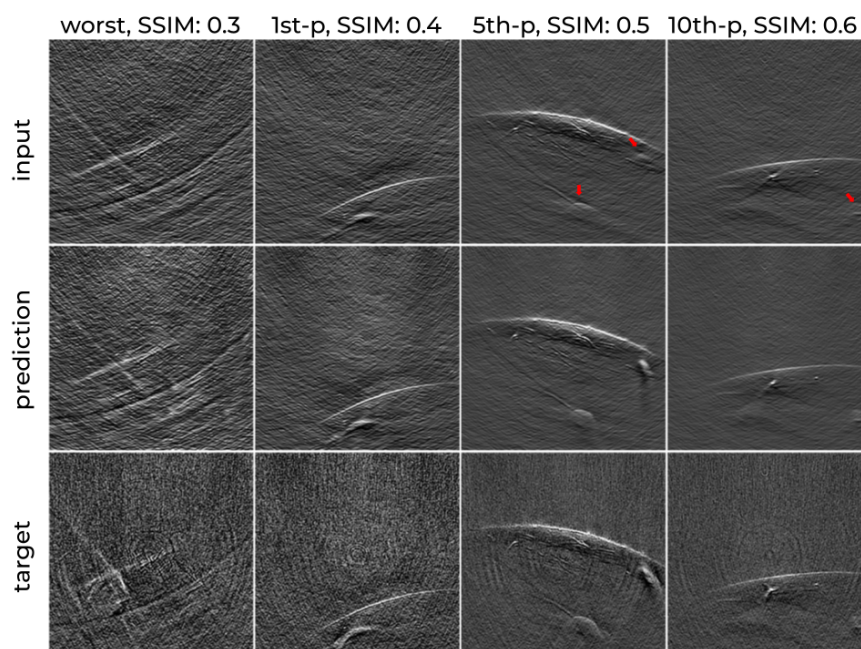


FIGURE 5.9: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{l_v128,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.



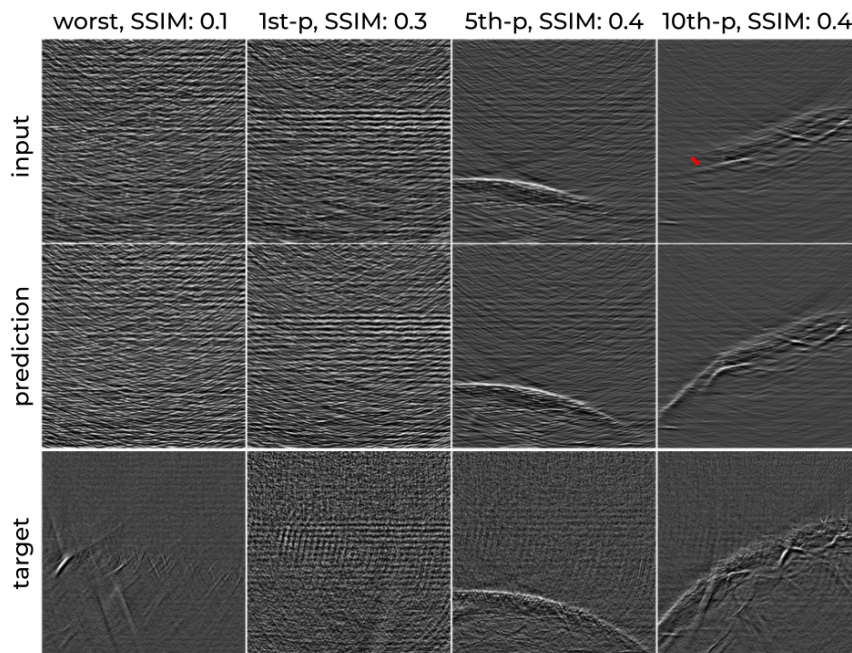


FIGURE 5.10: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{MSFD}_{1v128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

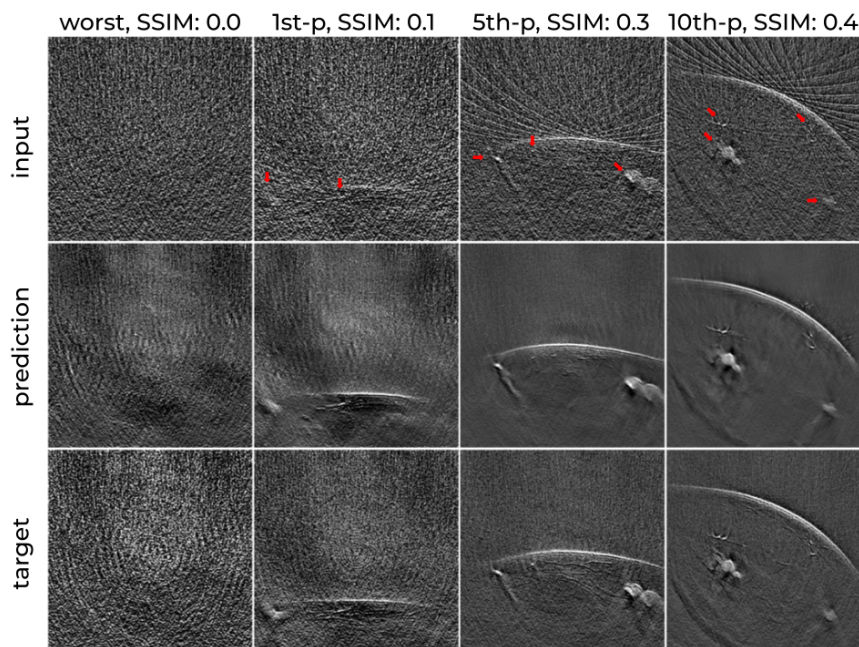


FIGURE 5.11: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{ss32,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

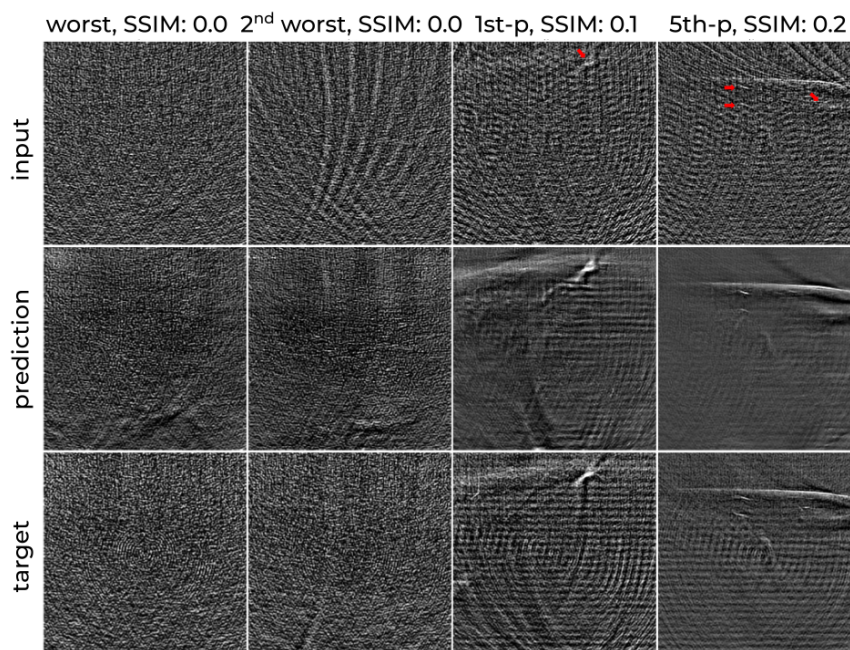


FIGURE 5.12: We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the nearly invisible vessels at input getting corrected at modUNet predictions.

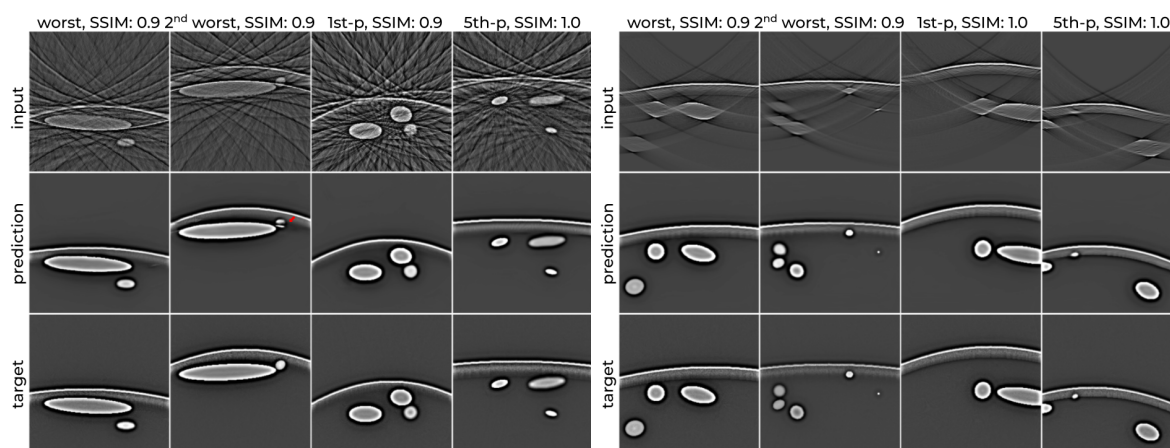


FIGURE 5.13: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiments  $f_{\text{SCD\_ss32,vc}}$  (left) and  $f_{\text{SCD\_lv128,li}}$  (right) with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

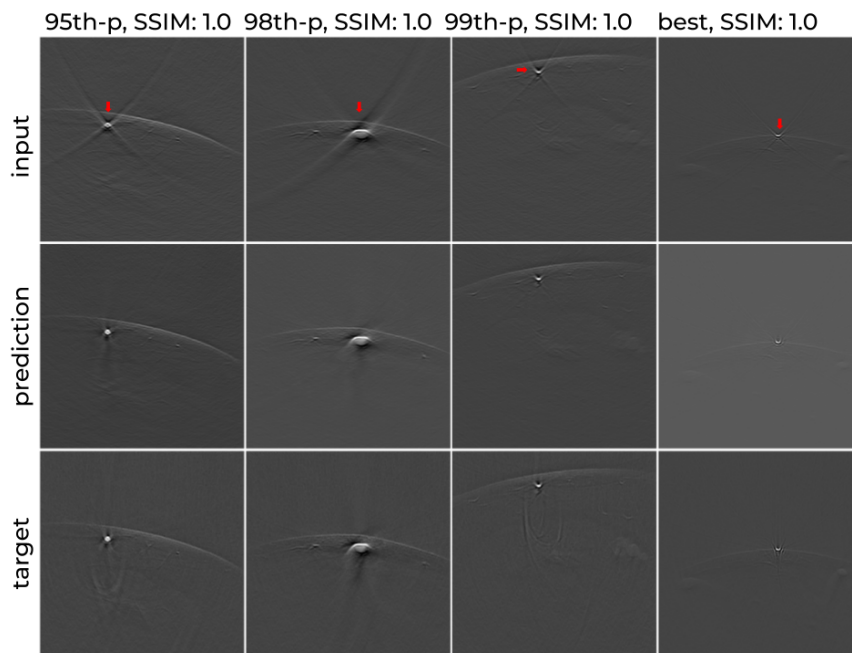


FIGURE 5.14: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{1v128,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

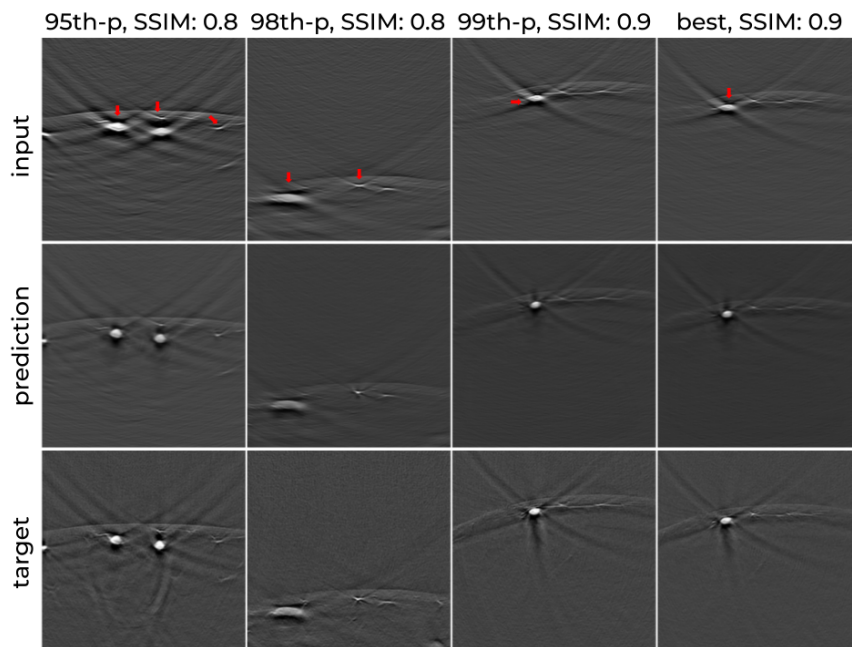


FIGURE 5.15: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{1v128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

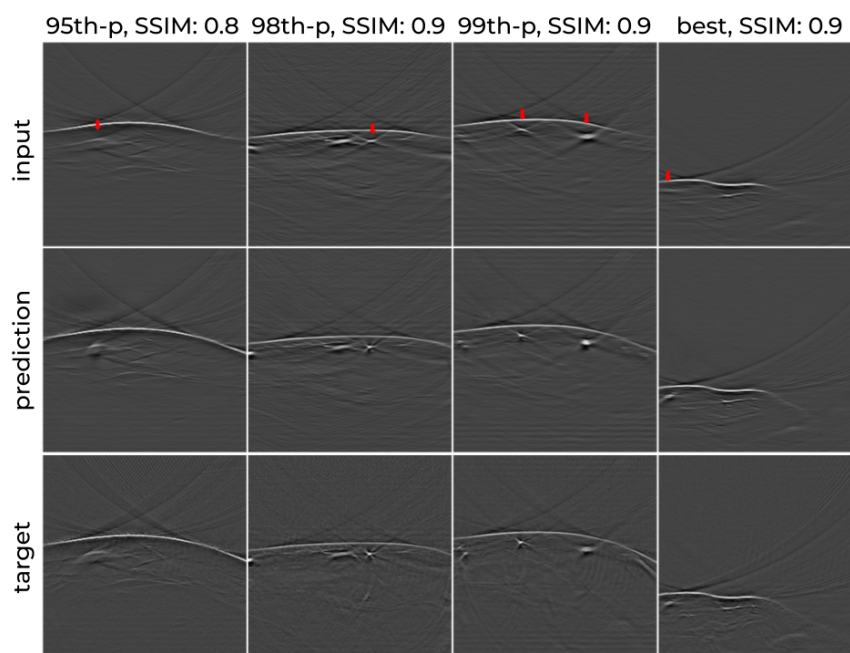


FIGURE 5.16: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{MSFD}_{lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions.

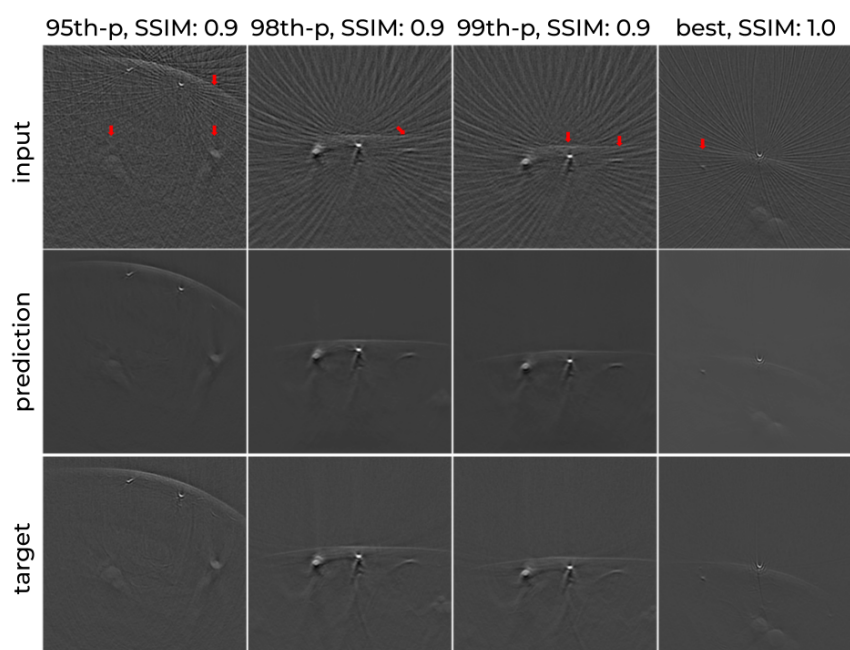


FIGURE 5.17: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{ss32,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

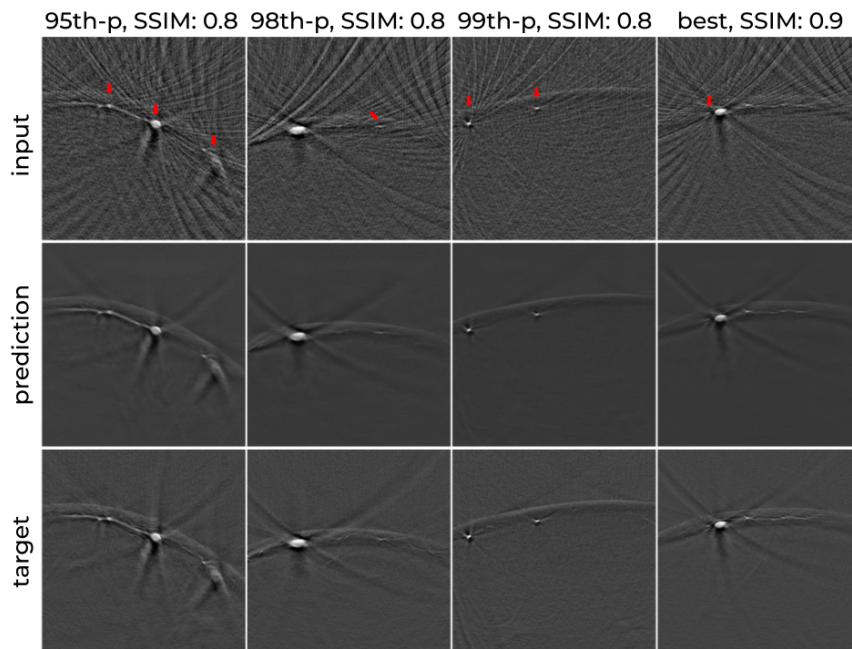


FIGURE 5.18: We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

## D.2. SEMANTIC SEGMENTATION

Similar to image translation task, we look at some of the samples from SCD with the worst vessel segmentation performance with respect to IoU metric in Figs. 5.19 & 5.20 (experiments  $f_{\text{seg\_lv128,li}}$  and  $f_{\text{seg\_ss32,vc}}$ ). As a result of SCD having flawless ground truth annotations, modUNet can achieve near perfect segmentation performance for all classes in all experiments of both tasks. The decrease of IoU seems to be exclusive for under- or over-segmenting little vessels by a handful of pixels. Despite most vessels being small (see Fig. 5.2e in the main manuscript, distribution of number of pixels per vessel), low IoU is limited to only a handful of extreme cases. This is also corroborated by the fact that skin curve segmentation distribution (see Fig. 5.5 right in the main manuscript) having a significantly shorter tail, barely falling below an IoU of 0.95 in most experiments. In order to understand the single experiment with a considerable performance drop, we look at experiment  $f_{\text{seg\_ss32,vc}}$  in Fig. 5.21. It can be observed that under heavy sparse sampling, the skin curve often gets ambiguous, rarely leading to suboptimal segmentation performance in these cases. Nevertheless, for most such subsampled reconstructions having intense duplicates of skin curve (Fig. 5.20 first row), modUNet successfully segments the right curve. Considering OADAT-mini segmentation experiments, in Fig. 5.22, we show the worst performing four samples from the test set of MSFD-mini. In Fig 5.23, it can be seen that such poor IoU is not observed in the worst performing samples for SWFD-mini experiments.

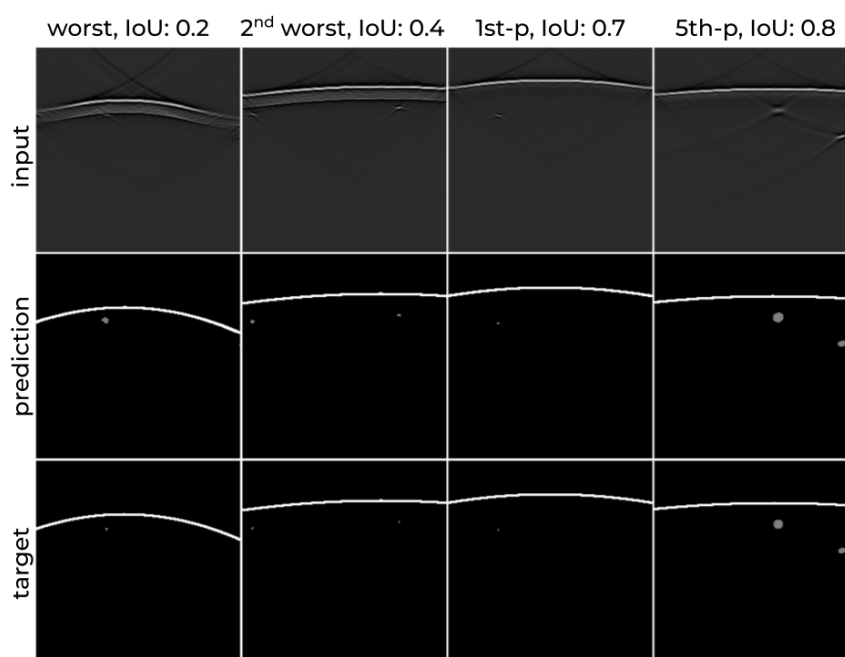


FIGURE 5.19: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) vessel intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs.

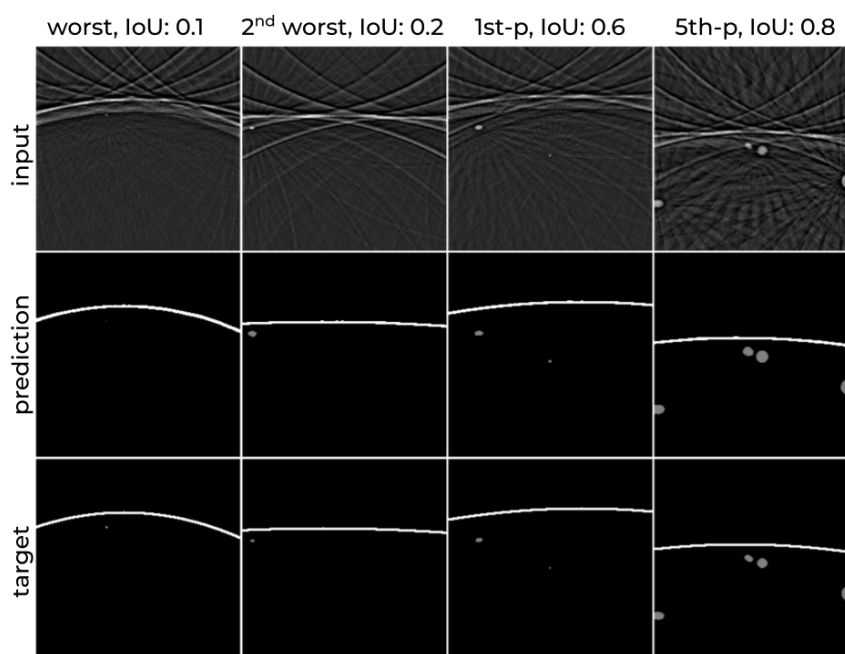


FIGURE 5.20: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) vessel intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_ss32,vc}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs.

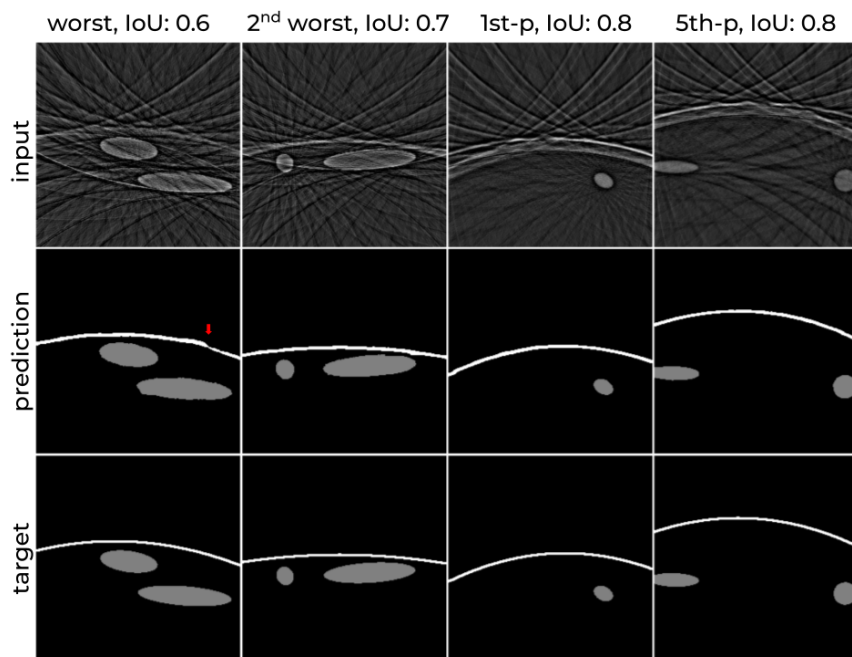


FIGURE 5.21: We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) skin curve intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_ss32,vc}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs.

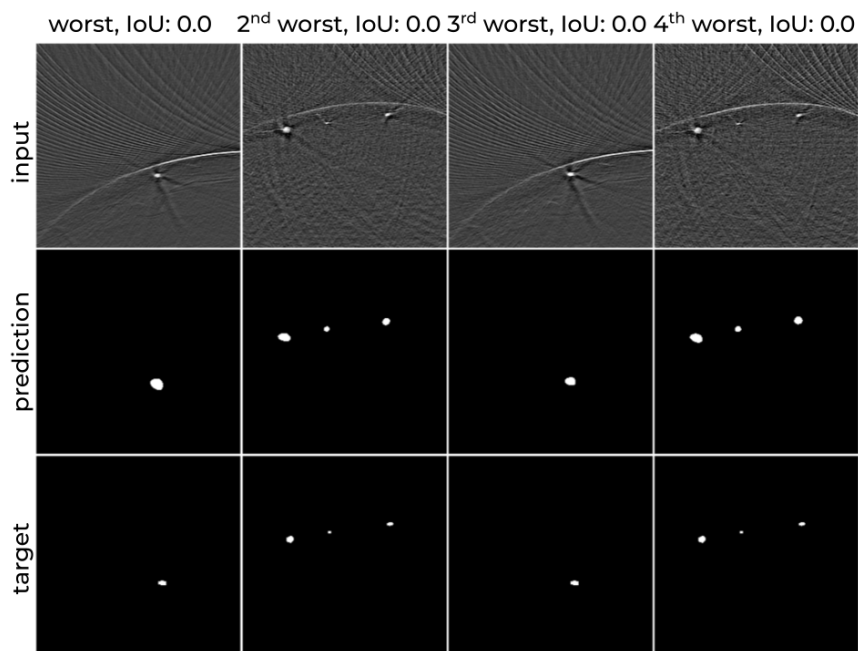


FIGURE 5.22: We showcase the worst (1st column), 2nd- (2nd column), 3rd- (3rd column), and 4th-worst (4th column) intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_MSFD\_ss64,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

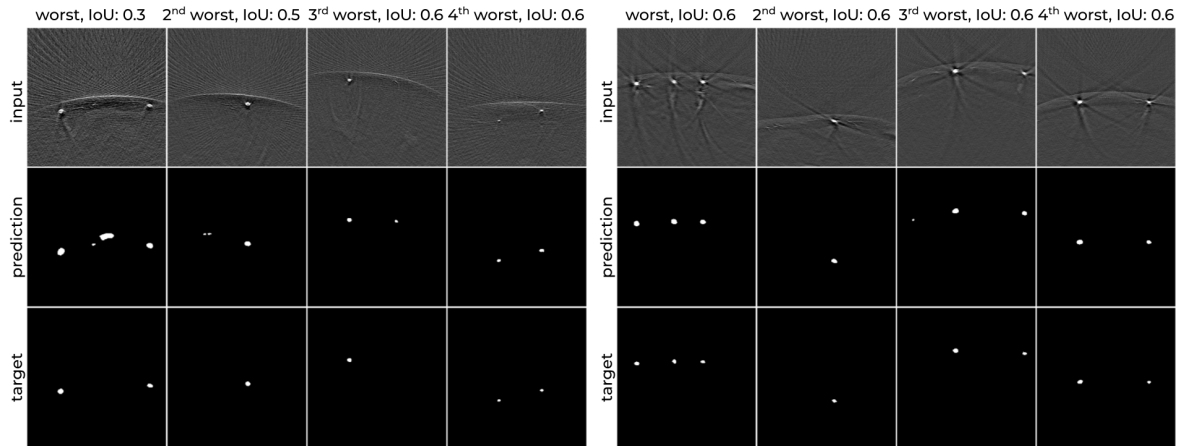


FIGURE 5.23: We showcase the worst (1st column), 2nd- (2nd column), 3rd- (3rd column), and 4th-worst (4th column) intersection over union (IoU) samples based on modUNet predictions for experiments  $f_{\text{seg\_SWFD\_ss64,sc}}$  (left) and  $f_{\text{seg\_SWFD\_ss128,ms}}$  (right) with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs.

## APPENDIX E. SIMULATED CYLINDER DATASET GENERATION

Our proposed simulated dataset, SCD, follows a group of heuristics we derived from observations on experimental images. For any given sample, we first generate an acoustic pressure map, from which we also generate ground truth annotations. We then apply certain post processing steps to imitate other phenomena such as patterns under the skin surface and different vessel textures. Given the geometry of the transducers we want to simulate, we then apply forward transform that gives the raw signals. Finally, we reconstruct these signals using backprojection algorithm to generate the images used in various SCD experiments in this manuscript.

The acoustic pressure map generation consists of initially drawing the curve that represents the laser pulse absorption on the skin surface mainly due to melanin. Given that experimental data is acquired with making sure that forearm is roughly at a certain distance range from the arrays, we also limit the drawn skin curve distance. We define the skin curve as a 2nd degree polynomial that is fitted to three points randomly sampled at the two horizontal edges and the center of the image at varying heights. As a post-processing step to mimic experimental data, the curve is first smoothed with a Gaussian filter. Then, an exponential decay of randomized length is applied under the curve along vertical axis. Finally, a non-structured uniform normal noise is multiplied with the aforementioned exponential decay region. For vessel generation, the number of cylinders to be drawn is sampled based on a coin flip. Based on the outcome, either two cylinders drawn or the number of cylinders is sampled from Poisson distribution (see Fig. 5.2e, distribution of number of vessels per image, main manuscript). Each vessel is initially represented by a cylinder orthogonal to the image plane ( $z$ -axis) with a randomly sampled radius. We then randomly rotate the cylinder around  $x$ - and  $y$ - axes. The vessel is determined as the cross-section of the cylinder at the imaging plane, yielding ellipses based on the final angle of the cylinders. As a post-processing step, we flip a coin to determine whether the vessel has a homogeneous intensity profile or has a linearly decreasing intensity



from its center. We then apply a Gaussian filter on the vessel to smooth its edges. Finally, based on a coin flip, we decide whether or not to multiply the intensity profile of the vessel with uniform normal noise. The same process is iteratively repeated until desired number of non-overlapping vessels are generated. All parameters used for the aforementioned steps are empirically selected based on our observations on the experimental datasets. We provide the script to simulate acoustic pressure maps that we used for SCD.

## APPENDIX F. ORGANIZATION OF DATASETS

The datasets and required algorithms to read raw signals and reconstructed images are summarized in Fig. 5.24 along with the codes to rerun and evaluate all the presented experiments. All datasets are stored as HDF5 files and corresponding indices across datasets within an HDF5 file have the same scene as raw signal or reconstructed image. Similarly, metadata such as *patientID*, *side* (i.e., left vs right arm), *skin type*, and *sliceID* contain the information for the corresponding indices. Some of these metadata such as *patientID*, which corresponds to the unique anonymized identifier of the volunteer, only exists in experimental datasets. Matching identifiers across experimental datasets (MSFD and SWFD) correspond to the same volunteer.

Each sample (raw signal or image reconstruction) is chunked in a single piece for optimal compression/decompression overhead when reading individual images and storing datasets. For example, for a given HDF5 dataset of shape (number of instances, height, width), each block of (1, height, width) is compressed individually. In our experiments, this ensured zero additional idle GPU time reading random training sample from a compressed HDF5 file when compared with its non-compressed HDF5 counterpart.

Raw signal datasets have the shape of (number of instances, temporal acquisition axis, receiving element axis). Temporal acquisition axis is fixed to 2,030 sampling points across all transducer arrays. Receiving element axis depends on the number of transducer elements that is used for signal acquisition. For sparse sampling and limited view reconstructions,

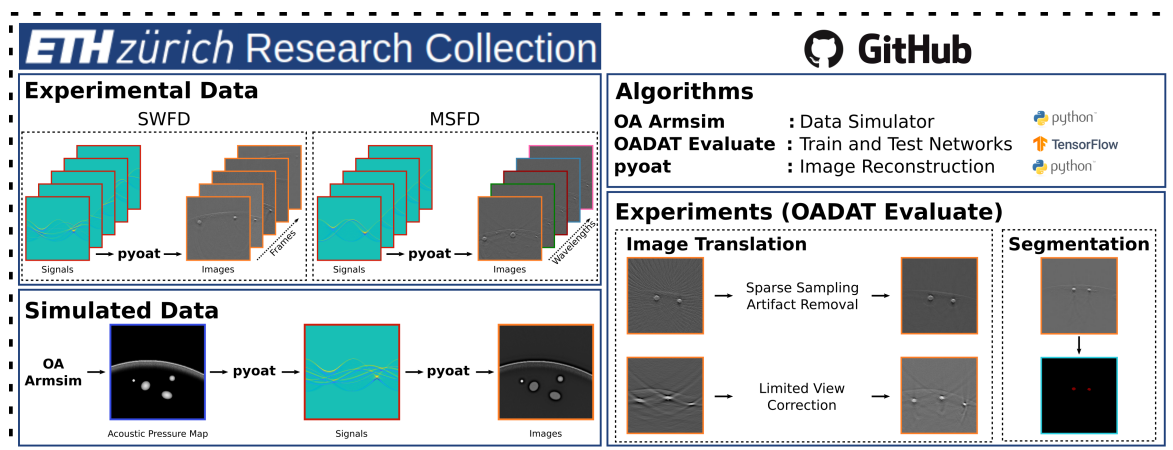


FIGURE 5.24: Pipeline figure to summarize data storage and algorithms.

TABLE 5.5: Contents of the MSFD file.

Name	Shape	Content	Name	Shape	Content
patientID	(25200,)	(2, 5, 6, 7, 9, 10, 11, 14, 15)	ms_ss64_BP_w700	(25200, 256, 256)	image reconstructions
side	(25200,)	(left, right)	ms_ss64_BP_w730	(25200, 256, 256)	image reconstructions
skin_type	(25200,)	(1, 2, 3, 4)	ms_ss64_BP_w760	(25200, 256, 256)	image reconstructions
sliceID	(25200,)	(1, ..., 1400)	ms_ss64_BP_w780	(25200, 256, 256)	image reconstructions
linear_BP_w700	(25200, 256, 256)	image reconstructions	ms_ss64_BP_w800	(25200, 256, 256)	image reconstructions
linear_BP_w730	(25200, 256, 256)	image reconstructions	ms_ss64_BP_w850	(25200, 256, 256)	image reconstructions
linear_BP_w760	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w700	(25200, 256, 256)	image reconstructions
linear_BP_w780	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w730	(25200, 256, 256)	image reconstructions
linear_BP_w800	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w760	(25200, 256, 256)	image reconstructions
linear_BP_w850	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w780	(25200, 256, 256)	image reconstructions
ms_BP_w700	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w800	(25200, 256, 256)	image reconstructions
ms_BP_w730	(25200, 256, 256)	image reconstructions	ms_ss128_BP_w850	(25200, 256, 256)	image reconstructions
ms_BP_w760	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w700	(25200, 2030, 256)	raw signals
ms_BP_w780	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w730	(25200, 2030, 256)	raw signals
ms_BP_w800	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w760	(25200, 2030, 256)	raw signals
ms_BP_w850	(25200, 256, 256)	image reconstructions	ms_ss32_raw_w780	(25200, 2030, 256)	raw signals
linear_raw_w700	(25200, 2030, 256)	raw signals	ms_ss32_raw_w800	(25200, 2030, 256)	raw signals
linear_raw_w730	(25200, 2030, 256)	raw signals	ms_ss32_raw_w850	(25200, 2030, 256)	raw signals
linear_raw_w760	(25200, 2030, 256)	raw signals	ms_ss64_raw_w700	(25200, 2030, 256)	raw signals
linear_raw_w780	(25200, 2030, 256)	raw signals	ms_ss64_raw_w730	(25200, 2030, 256)	raw signals
linear_raw_w800	(25200, 2030, 256)	raw signals	ms_ss64_raw_w760	(25200, 2030, 256)	raw signals
linear_raw_w850	(25200, 2030, 256)	raw signals	ms_ss64_raw_w780	(25200, 2030, 256)	raw signals
ms_raw_w700	(25200, 2030, 256)	raw signals	ms_ss64_raw_w800	(25200, 2030, 256)	raw signals
ms_raw_w730	(25200, 2030, 256)	raw signals	ms_ss64_raw_w850	(25200, 2030, 256)	raw signals
ms_raw_w760	(25200, 2030, 256)	raw signals	ms_ss128_raw_w700	(25200, 2030, 256)	raw signals
ms_raw_w780	(25200, 2030, 256)	raw signals	ms_ss128_raw_w730	(25200, 2030, 256)	raw signals
ms_raw_w800	(25200, 2030, 256)	raw signals	ms_ss128_raw_w760	(25200, 2030, 256)	raw signals
ms_raw_w850	(25200, 2030, 256)	raw signals	ms_ss128_raw_w780	(25200, 2030, 256)	raw signals
ms_ss32_BP_w700	(25200, 256, 256)	image reconstructions	ms_ss128_raw_w800	(25200, 2030, 256)	raw signals
ms_ss32_BP_w730	(25200, 256, 256)	image reconstructions	ms_ss128_raw_w850	(25200, 2030, 256)	raw signals
ms_ss32_BP_w760	(25200, 256, 256)	image reconstructions			
ms_ss32_BP_w780	(25200, 256, 256)	image reconstructions			
ms_ss32_BP_w800	(25200, 256, 256)	image reconstructions			
ms_ss32_BP_w850	(25200, 256, 256)	image reconstructions			

we retain the shape of the raw signal data, however, we switch off the receiving elements that are not active, hence they are padded with zeros. This convention also allows for direct compatibility with the *pyoat* reconstruction package.

We provide a header dataset file “OADAT.h5” as a convenient access point to each other HDF5 dataset file. For convenience, we also provide a header dataset file “OADAT\_v2.h5” that allows more intuitive access to addendums of MSFD, SWFD, and SCD with multi-segment sparse sampling datasets. Additionally, a header dataset file “OADAT-mini.h5” is provided as an access point to reach all relevant OADAT-mini HDF5 files. These header files also contain metadata module of the Dataset Nutrition Label, which should help making the dataset self explanatory.

### F.1. MSFD

Contents of MSFD are listed in Table 5.5. The contents of the dataset are stored in the files “MSFD\_multisegment\_RawBP.h5” and “MSFD\_multisegment\_ss\_RawBP.h5”. Provided *sliceID* corresponds to the time index of the slice acquired from a given volunteer for a given acquisition. For example, *sliceID*  $i + 1$  is recorded right after  $i$  for a given *patientID* and *side*. This information can be relevant for future work that does not treat each slice independently, but exploit correlations from consecutive slice acquisitions. Nevertheless, a *sliceID*  $i \in [1, \dots, 1400]$  does not necessarily correspond to the same position on the forearm across the volunteers.

### F.2. SWFD

Contents of SWFD are split into two separate HDF5 files with contents listed in Tables 5.6 & 5.7 (“SWFD\_semicircle\_RawBP.h5” and “SWFD\_multisegment\_RawBP.h5”). Despite having

TABLE 5.6: Contents of SWFD semi circle file.

Name	Shape	Content	Name	Shape	Content
patientID	(39228,)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14)	sc_lv128_raw	(39228, 2030, 256)	raw signals
side	(39228,)	(left, right)	sc_ss128_raw	(39228, 2030, 256)	raw signals
skin_type	(39228,)	(1, 2, 3, 4)	sc_ss32_raw	(39228, 2030, 256)	raw signals
sliceID	(39228,)	(100, ..., 1500)	sc_ss64_raw	(39228, 2030, 256)	raw signals
sc_lv128_BP	(39228, 256, 256)	image reconstructions	sc_raw	(39228, 2030, 256)	raw signals
sc_ss128_BP	(39228, 256, 256)	image reconstructions			
sc_ss32_BP	(39228, 256, 256)	image reconstructions			
sc_ss64_BP	(39228, 256, 256)	image reconstructions			
sc_BP	(39228, 256, 256)	image reconstructions			

TABLE 5.7: Contents of SWFD multisegment file.

Name	Shape	Content	Name	Shape	Content
patientID	(39228,)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14)	ms_ss128_BP	(39228, 256, 256)	image reconstructions
side	(39228,)	(left, right)	ms_ss32_BP	(39228, 256, 256)	image reconstructions
skin_type	(39228,)	(1, 2, 3, 4)	ms_ss64_BP	(39228, 256, 256)	image reconstructions
sliceID	(39228,)	(100, ..., 1500)	ms_ss128_raw	(39228, 2030, 256)	raw signals
linear_BP	(39228, 256, 256)	image reconstructions	ms_ss32_raw	(39228, 2030, 256)	raw signals
ms_BP	(39228, 256, 256)	image reconstructions	ms_ss64_raw	(39228, 2030, 256)	raw signals
linear_raw	(39228, 2030, 256)	raw signals			
ms_raw	(39228, 2030, 256)	raw signals			

the same number of instances per dataset category across the two files, the indices are not paired across semi circle and multisegment array. Hence we decided to split the dataset into two files. Furthermore, we include two additional HDF5 files “SWFD\_semicircle\_ss\_RawBP.h5” and “SWFD\_multisegment\_ss\_RawBP.h5” that contain sparse sampling image and raw data for multisegment array. As an artifact from discarding first few slices to filter out low SNR instances (both in MSFD and SWFD), SWFD acquisitions have 1,401 slices, having *sliceID* in the range  $[100, \dots, 1500]$ , yielding  $2 \times 14 \times 1,401 = 39,228$  samples per dataset category as opposed to 39,200.

### F.3. SCD

Contents of SCD are listed in Table 5.8 and stored in the file “SCD\_RawBP.h5”. Different than the experimental datasets, SCD also contains the ground truth acoustic pressure map and annotations (labels). Annotations are encoded as 0: background, 1: vessel, and 2: skin curve.

### F.4. OADAT-MINI

OADAT-mini consists of a subset of 100 instances of each category of OADAT along with their annotation maps for vessels. Accordingly, its contents are split into multiple files, one for each

TABLE 5.8: Contents of SCD files.

Name	Shape	Content	Name	Shape	Content
sliceID	(20000,)	(0, ..., 19999)	vc_lv128_raw	(20000, 2030, 1024)	raw signals
ground_truth	(20000, 256, 256)	acoustic pressure map	vc_ss128_raw	(20000, 2030, 1024)	raw signals
labels	(20000, 256, 256)	(0, 1, 2)	vc_ss32_raw	(20000, 2030, 1024)	raw signals
linear_BP	(20000, 256, 256)	image reconstructions	vc_ss64_raw	(20000, 2030, 1024)	raw signals
ms_BP	(20000, 256, 256)	image reconstructions	vc_raw	(20000, 2030, 1024)	raw signals
linear_raw	(20000, 2030, 256)	raw signals	ms_ss128_BP	(20000, 256, 256)	image reconstructions
ms_raw	(20000, 2030, 256)	raw signals	ms_ss32_BP	(20000, 256, 256)	image reconstructions
vc_lv128_BP	(20000, 256, 256)	image reconstructions	ms_ss64_BP	(20000, 256, 256)	image reconstructions
vc_ss128_BP	(20000, 256, 256)	image reconstructions	ms_ss128_raw	(20000, 2030, 256)	raw signals
vc_ss32_BP	(20000, 256, 256)	image reconstructions	ms_ss32_raw	(20000, 2030, 256)	raw signals
vc_ss64_BP	(20000, 256, 256)	image reconstructions	ms_ss64_raw	(20000, 2030, 256)	raw signals
vc_BP	(20000, 256, 256)	image reconstructions			

TABLE 5.9: Contents of OADAT-mini files.

Name	Shape	Content	Name	Shape	Content
<b>SCD - mini</b>					
sliceID	(100,)	C (0, ..., 19999)	vc,lv128_raw	(100, 2030, 1024)	raw signals
ground_truth	(100, 256, 256)	acoustic pressure map	vc,ss128_raw	(100, 2030, 1024)	raw signals
labels	(100, 256, 256)	(0, 1, 2)	vc,ss32_raw	(100, 2030, 1024)	raw signals
linear_BP	(100, 256, 256)	image reconstructions	vc,ss64_raw	(100, 2030, 1024)	raw signals
ms_BP	(100, 256, 256)	image reconstructions	vc_raw	(100, 2030, 1024)	raw signals
linear_raw	(100, 256, 256)	raw signals	ms,ss128_BP	(100, 256, 256)	image reconstructions
ms_raw	(100, 2030, 256)	raw signals	ms,ss32_BP	(100, 256, 256)	image reconstructions
vc,lv128_BP	(100, 256, 256)	image reconstructions	ms,ss64_BP	(100, 256, 256)	image reconstructions
vc,ss128_BP	(100, 256, 256)	image reconstructions	ms,ss128_raw	(100, 2030, 256)	raw signals
vc,ss32_BP	(100, 256, 256)	image reconstructions	ms,ss32_raw	(100, 2030, 256)	raw signals
vc,ss64_BP	(100, 256, 256)	image reconstructions	ms,ss64_raw	(100, 2030, 256)	raw signals
vc_BP	(100, 256, 256)	image reconstructions			
<b>SWFD semi circle - mini</b>					
patientID	(100,)	(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13)	sc,lv128_raw	(100, 2030, 256)	raw signals
side	(100,)	(left, right)	sc,ss128_raw	(100, 2030, 256)	raw signals
skin_type	(100,)	(1, 2, 3, 4)	sc,ss32_raw	(100, 2030, 256)	raw signals
sliceID	(100,)	C (100, ..., 1500)	sc,ss64_raw	(100, 2030, 256)	raw signals
labels	(100, 256, 256)	(0, 1)	sc_raw	(100, 2030, 256)	raw signals
sc,lv128_BP	(100, 256, 256)	image reconstructions			
sc,ss128_BP	(100, 256, 256)	image reconstructions			
sc,ss32_BP	(100, 256, 256)	image reconstructions			
sc,ss64_BP	(100, 256, 256)	image reconstructions			
sc_BP	(100, 256, 256)	image reconstructions			
<b>SWFD multisegment - mini</b>					
patientID	(100,)	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)	ms,ss128_BP	(100, 256, 256)	image reconstructions
side	(100,)	(left, right)	ms,ss32_BP	(100, 256, 256)	image reconstructions
skin_type	(100,)	(1, 2, 3, 4)	ms,ss64_BP	(100, 256, 256)	image reconstructions
sliceID	(100,)	C (100, ..., 1500)	ms,ss128_raw	(100, 2030, 256)	raw signals
labels	(100, 256, 256)	(0, 1)			
linear_BP	(100, 256, 256)	image reconstructions	ms,ss32_raw	(100, 2030, 256)	raw signals
ms_BP	(100, 256, 256)	image reconstructions	ms,ss64_raw	(100, 2030, 256)	raw signals
linear_raw	(100, 2030, 256)	raw signals			
ms_raw	(100, 2030, 256)	raw signals			
<b>MSFD multisegment - mini</b>					
patientID	(100,)	(2, 5, 6, 7, 9, 10, 11, 14, 15)	ms_ss64_BP_w700	(100, 256, 256)	image reconstructions
side	(100,)	(left, right)	ms_ss64_BP_w730	(100, 256, 256)	image reconstructions
skin_type	(100,)	(1, 2, 3, 4)	ms_ss64_BP_w760	(100, 256, 256)	image reconstructions
sliceID	(100,)	C (1, ..., 1400)	ms_ss64_BP_w780	(100, 256, 256)	image reconstructions
labels	(100, 256, 256)	(0, 1)			
linear_BP_w700	(100, 256, 256)	image reconstructions	ms_ss64_BP_w800	(100, 256, 256)	image reconstructions
linear_BP_w730	(100, 256, 256)	image reconstructions	ms_ss64_BP_w850	(100, 256, 256)	image reconstructions
linear_BP_w760	(100, 256, 256)	image reconstructions	ms_ss128_BP_w700	(100, 256, 256)	image reconstructions
linear_BP_w780	(100, 256, 256)	image reconstructions	ms_ss128_BP_w730	(100, 256, 256)	image reconstructions
linear_BP_w800	(100, 256, 256)	image reconstructions	ms_ss128_BP_w760	(100, 256, 256)	image reconstructions
linear_BP_w850	(100, 256, 256)	image reconstructions	ms_ss128_BP_w780	(100, 256, 256)	image reconstructions
ms_BP_w700	(100, 256, 256)	image reconstructions	ms_ss128_BP_w800	(100, 256, 256)	image reconstructions
ms_BP_w730	(100, 256, 256)	image reconstructions	ms_ss128_BP_w850	(100, 256, 256)	image reconstructions
ms_BP_w760	(100, 256, 256)	image reconstructions	ms_ss32_raw_w700	(100, 2030, 256)	raw signals
ms_BP_w780	(100, 256, 256)	image reconstructions	ms_ss32_raw_w730	(100, 2030, 256)	raw signals
ms_BP_w800	(100, 256, 256)	image reconstructions	ms_ss32_raw_w760	(100, 2030, 256)	raw signals
ms_BP_w850	(100, 256, 256)	image reconstructions	ms_ss32_raw_w780	(100, 2030, 256)	raw signals
linear_raw_w700	(100, 2030, 256)	raw signals	ms_ss32_raw_w800	(100, 2030, 256)	raw signals
linear_raw_w730	(100, 2030, 256)	raw signals	ms_ss32_raw_w850	(100, 2030, 256)	raw signals
linear_raw_w760	(100, 2030, 256)	raw signals	ms_ss64_raw_w700	(100, 2030, 256)	raw signals
linear_raw_w780	(100, 2030, 256)	raw signals	ms_ss64_raw_w730	(100, 2030, 256)	raw signals
linear_raw_w800	(100, 2030, 256)	raw signals	ms_ss64_raw_w760	(100, 2030, 256)	raw signals
linear_raw_w850	(100, 2030, 256)	raw signals	ms_ss64_raw_w780	(100, 2030, 256)	raw signals
ms_raw_w700	(100, 2030, 256)	raw signals	ms_ss64_raw_w730	(100, 2030, 256)	raw signals
ms_raw_w730	(100, 2030, 256)	raw signals	ms_ss64_raw_w760	(100, 2030, 256)	raw signals
ms_raw_w760	(100, 2030, 256)	raw signals	ms_ss64_raw_w780	(100, 2030, 256)	raw signals
ms_raw_w780	(100, 2030, 256)	raw signals	ms_ss64_raw_w800	(100, 2030, 256)	raw signals
ms_raw_w800	(100, 2030, 256)	raw signals	ms_ss64_raw_w850	(100, 2030, 256)	raw signals
ms_raw_w850	(100, 2030, 256)	raw signals	ms_ss128_raw_w700	(100, 2030, 256)	raw signals
ms_ss32_BP_w700	(100, 256, 256)	image reconstructions	ms_ss128_raw_w730	(100, 2030, 256)	raw signals
ms_ss32_BP_w730	(100, 256, 256)	image reconstructions	ms_ss128_raw_w760	(100, 2030, 256)	raw signals
ms_ss32_BP_w760	(100, 256, 256)	image reconstructions	ms_ss128_raw_w780	(100, 2030, 256)	raw signals
ms_ss32_BP_w780	(100, 256, 256)	image reconstructions	ms_ss128_raw_w800	(100, 2030, 256)	raw signals
ms_ss32_BP_w800	(100, 256, 256)	image reconstructions	ms_ss128_raw_w850	(100, 2030, 256)	raw signals
ms_ss32_BP_w850	(100, 256, 256)	image reconstructions			

transducer array and SCD: “SWFD\_semicircle\_RawBP-mini.h5”, “SWFD\_multisegment\_RawBP-mini.h5”, “MSFD\_multisegment\_RawBP-mini.h5”, “SCD\_RawBP-mini.h5”. We list the contents of each of these files in Table 5.9.

## APPENDIX G. CONSIDERATIONS WHEN USING OUR DATASETS

As the volunteers in experimental datasets are considered to be healthy, anonymization is done one way, and true identities of volunteers are not possible to trace. Furthermore, to the best of our knowledge, usage of our proposed datasets cannot pose threat to the volunteers, even with malicious intent. However, as with all clinical data, it should be acknowledged that our dataset would represent a particular subset of all potential forearm images collected at respective wavelengths and transducer arrays. Accordingly, any subsequent work that makes use of our datasets for validation purposes need to ensure that the diversity of our datasets is sufficient for their target application.

## APPENDIX H. PERSISTENCE OF PROPOSED DATASETS

All datasets are hosted on Libdrive at ETH Zurich Research Collection repository. Accordingly, the datasets have a landing page with digital object identifier (DOI) and metadata. Data will be freely accessible with no restriction. As per ETH Zurich Research Collection documentation, the dataset is guaranteed to have a retention period for 10 years. Based on the frequency of usage by the community, the Research Collection may continue to host the data beyond the 10 year period.

## APPENDIX I. AUTHOR STATEMENT ON DATA LICENSE

The dataset is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC-BY-NC).

## APPENDIX J. TRAINING AND EVALUATION CODE

In oadat-evaluate repository, along with saved model weights, we provide all necessary scripts to train modUNet from scratch for all 44 experiments as well as evaluating them over the whole test set. We provide standalone examples on the repository landing page to show how to;

- (i) load a pretrained model, for example, to do inference or finetuning it,
- (ii) train modUNet from scratch for either of the two tasks for a given experiment,
- (iii) evaluate a serialized model, whether it is one of the pretrained models we provide as is, or any other Tensorflow model that is already serialized, and
- (iv) use our provided data loader to read any sample from OADAT.

## BIBLIOGRAPHY

- [1] Idan Steinberg et al. “Photoacoustic clinical imaging”. In: *Photoacoustics* 14 (2019), pp. 77–98.
- [2] Jimmy L Su et al. “Advances in clinical and biomedical applications of photoacoustic imaging”. In: *Expert opinion on medical diagnostics* 4.6 (2010), pp. 497–510.
- [3] Berkan Lafci et al. “Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound”. In: *Neoplasia* 22.12 (2020), pp. 770–777.
- [4] Elena Merčep et al. “Whole-body live mouse imaging by hybrid reflection-mode ultrasound and optoacoustic tomography”. In: *Optics letters* 40.20 (2015), pp. 4643–4646.
- [5] Srirang Manohar and Maura Dantuma. “Current and future trends in photoacoustic breast imaging”. In: *Photoacoustics* 16 (2019), p. 100134.
- [6] Gael Diot et al. “Multispectral Optoacoustic Tomography (Msot) of Human Breast Cancermsot Signatures of Human Breast Cancer”. In: *Clinical Cancer Research* 23.22 (2017), pp. 6912–6922.
- [7] Reni Butler et al. “Optoacoustic breast imaging: imaging-pathology correlation of optoacoustic features in benign and malignant breast masses”. In: *American Journal of Roentgenology* (2018), pp. 1155–1170.
- [8] Ferdinand Knieling et al. “Multispectral optoacoustic tomography for assessment of Crohn’s disease activity”. In: *New England Journal of Medicine* 376.13 (2017), pp. 1292–1294.
- [9] Angelos Karlas et al. “Multispectral optoacoustic tomography of lipid and hemoglobin contrast in human carotid atherosclerosis”. In: *Photoacoustics* 23 (2021), p. 100283.
- [10] Xosé Luís Deán-Ben and Daniel Razansky. “Optoacoustic imaging of the skin”. In: *Experimental Dermatology* 30.11 (2021), pp. 1598–1609.
- [11] Ali Özbek, Xosé Luís Deán-Ben, and Daniel Razansky. “Optoacoustic imaging at kilohertz volumetric frame rates”. In: *Optica* 5.7 (2018), pp. 857–863.
- [12] X Luís Dean-Ben et al. “Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography”. In: *IEEE transactions on medical imaging* 31.10 (2012), pp. 1922–1928.
- [13] Berkan Lafci et al. “Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 68.3 (2020), pp. 688–696.

- [14] Stratis Tzoumas and Vasilis Ntziachristos. “Spectral unmixing techniques for optoacoustic imaging of tissue pathophysiology”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375.2107 (2017), p. 20170262.
- [15] Çağla Özsoy et al. “Ultrafast four-dimensional imaging of cardiac mechanical wave propagation with sparse optoacoustic sensing”. In: *Proceedings of the National Academy of Sciences* 118.45 (2021), e2103979118.
- [16] Justine Robin et al. “Hemodynamic response to sensory stimulation in mice: Comparison between functional ultrasound and optoacoustic imaging”. In: *NeuroImage* 237 (2021), p. 118111.
- [17] Lisanne Grünherz et al. “Preoperative mapping of lymphatic vessels by multispectral optoacoustic tomography”. In: *Lymphatic research and biology* 20.6 (2022), pp. 659–664.
- [18] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky. “Deep learning optoacoustic tomography with sparse data”. In: *Nature Machine Intelligence* 1.10 (2019), pp. 453–460.
- [19] Steven Guan et al. “Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning”. In: *Scientific reports* 10.1 (2020), p. 8510.
- [20] Anna Klimovskaia Susmelj et al. “Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data”. In: *International Conference on Medical Imaging with Deep Learning*. PMLR. 2022, pp. 1173–1191.
- [21] Neda Davoudi et al. “Deep learning of image-and time-domain data enhances the visibility of structures in optoacoustic tomography”. In: *Optics letters* 46.13 (2021), pp. 3029–3032.
- [22] Hengrong Lan et al. “Y-Net: a hybrid deep learning reconstruction framework for photoacoustic imaging in vivo”. In: *arXiv preprint arXiv:1908.00975* (2019).
- [23] Melanie Schellenberg et al. “Semantic segmentation of multispectral photoacoustic images using deep learning”. In: *Photoacoustics* 26 (2022), p. 100341.
- [24] Frederic M Brochu et al. “Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography”. In: *IEEE transactions on medical imaging* 36.1 (2016), pp. 322–331.
- [25] Subhamoy Mandal, Xosé Luís Deán-Ben, and Daniel Razansky. “Visual quality enhancement in optoacoustic tomography using active contour segmentation priors”. In: *IEEE transactions on medical imaging* 35.10 (2016), pp. 2209–2217.
- [26] Janek Gröhl et al. “The IPASC data format: A consensus data format for photoacoustic imaging”. In: *Photoacoustics* 26 (2022), p. 100339.
- [27] Ali Ozbek, XL Deán-Ben, and Daniel Razansky. “Realtime parallel back-projection algorithm for three-dimensional optoacoustic imaging devices”. In: *European conference on biomedical optics*. Optica Publishing Group. 2013, p. 88000I.

- [28] Jørgen Arendt Jensen. “Medical ultrasound imaging.” In: *Progress in biophysics and molecular biology* 93 1-3 (2007), pp. 153–65.
- [29] Ramona De Luca et al. “Diagnostic ultrasound probes: a typology and overview of technologies”. In: *Current Directions in Biomedical Engineering* 4.1 (2018), pp. 49–53. DOI: doi:10.1515/cdbme-2018-0013.
- [30] Elena Merčep, Xosé Luís Deán-Ben, and Daniel Razansky. “Combined Pulse-Echo Ultrasound and Multispectral Optoacoustic Tomography With a Multi-Segment Detector Array”. In: *IEEE Transactions on Medical Imaging* 36.10 (2017), pp. 2129–2137.
- [31] Vishal Gupta and Vinod Kumar Sharma. “Skin typing: Fitzpatrick grading and others”. In: *Clinics in dermatology* 37.5 (2019), pp. 430–436.
- [32] Shan Huang et al. “Functional multispectral optoacoustic tomography imaging of hepatic steatosis development in mice”. In: *EMBO Molecular Medicine* 13.9 (2021), e13490.
- [33] Valeriya Perekatova et al. “Optimal wavelengths for optoacoustic measurements of blood oxygen saturation in biological tissues”. In: *Biomedical optics express* 7.10 (2016), pp. 3979–3995.
- [34] Arunima Sharma, Vijitha Periyasamy, and Manojit Pramanik. “Photoacoustic imaging depth comparison at 532-, 800-, and 1064-nm wavelengths: Monte Carlo simulation and experimental validation”. In: *Journal of Biomedical Optics* 24.12 (2019), pp. 121904–121904.
- [35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.
- [36] Ozan Oktay et al. “Attention u-net: Learning where to look for the pancreas”. In: *arXiv preprint arXiv:1804.03999* (2018).
- [37] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. pmlr. 2015, pp. 448–456.
- [38] Lu Ding et al. “Efficient non-negative constrained model-based inversion in optoacoustic tomography”. In: *Physics in Medicine & Biology* 60.17 (2015), p. 6733.
- [39] X Luís Deán-Ben, Vasilis Ntziachristos, and Daniel Razansky. “Effects of small variations of speed of sound in optoacoustic tomographic imaging”. In: *Medical physics* 41.7 (2014), p. 073301.
- [40] Valery Vishnevskiy, Richard Rau, and Orcun Goksel. “Deep variational networks with exponential weighting for learning computed tomography”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*. Springer. 2019, pp. 310–318.
- [41] Eduard Grüneisen. “Theorie des festen Zustandes einatomiger Elemente”. In: *Annalen der Physik* 344.12 (1912), pp. 257–306.



## 6 NONINVASIVE MULTIPARAMETRIC CHARACTERIZATION OF MAMMARY TUMORS WITH TRANSMISSION-REFLECTION OPTOACOUSTIC ULTRASOUND

This chapter includes the following publication:

Neoplasia, Vol. 22, Issue 12, pp. 770-777 (2020).

Authors:

**Berkan Lafci**<sup>1,2</sup>, Elena Merčep<sup>3,4</sup>, Joaquin L. Herraiz<sup>5,6</sup>, Xosé Luís Deán-Ben<sup>1,2</sup>, Daniel Razansky<sup>1,2</sup>

<sup>1</sup> Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>2</sup> Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>3</sup> Institute for Biological and Medical Imaging, Technical University of Munich and Helmholtz Center Munich, Neuherberg, Germany

<sup>4</sup> iThera Medical GmbH, Munich, Germany

<sup>5</sup> Nuclear Physics Group and IPARCOS, Complutense University of Madrid, Madrid, Spain

<sup>6</sup> Health Research Institute of Hospital Clínico San Carlos (IdISSC), Madrid, Spain

## ABSTRACT

Development of imaging methods capable of furnishing tumor-specific morphological, functional and molecular information is paramount for early diagnosis, staging, and treatment of breast cancer. Ultrasound (US) and optoacoustic (OA) imaging methods exhibit excellent traits for tumor imaging in terms of fast imaging speed, ease of use, excellent contrast, and lack of ionizing radiation. Here we demonstrate simultaneous tomographic whole body imaging of optical absorption, US reflectivity and speed of sound (SoS) in living mice. *In vivo* studies of 4T1 breast cancer xenografts models revealed synergistic and complementary value of the hybrid imaging approach for characterizing mammary tumors. While neo-vasculature surrounding the tumor areas was observed based on the vascular anatomy contrast provided by the OA data, the tumor boundaries could be discerned by segmenting hypoechoic structures in pulse-echo US images. Tumor delineation was further facilitated by enhancing the contrast and spatial resolution of the SoS maps with a full wave inversion method. The malignant lesions could thus be distinguished from other hypoechoic regions based on the average SoS values. The reported findings corroborate the strong potential of the hybrid imaging approach for advancing cancer research in small animal models and fostering development of new clinical diagnostic approaches.

## 6.1 INTRODUCTION

Breast cancer is the most frequent non-cutaneous type of cancer in women and the second cause of cancer-related deaths in the female population [1]. Imaging-based mammography screening is considered to be a major factor leading to a 15-30% reduction of breast cancer mortality [2]. X-ray imaging of the breast remains the gold standard for breast screening in the clinical setting. Yet, this approach involves exposure to ionizing radiation and pain caused by breast compression. Moreover, false positives are produced e.g. due to the presence of cysts [3] and the sensitivity is low in women with radiographically dense breast [4]. Magnetic resonance imaging (MRI) and pulse-echo ultrasound (US) are then also routinely used in the clinics to complement the drawbacks of x-ray mammography. MRI provides high sensitivity for the detection of breast cancer, yet it attains low specificity and comes with high operational costs [5]. Pulse-echo US can distinguish between liquid-filled cysts from solid masses and even detect tumors not visible in x-ray images [6]. However, standard hand-held scans are operator dependent, which prevents the wide use of US as a standalone method.

In recent years, multispectral optoacoustic tomography (MSOT) imaging has been shown to significantly enhance the capabilities of pulse-echo (reflection) US for the detection of breast carcinomas [7]. Identification of tumors in MSOT images is facilitated by key bio-markers such as local increases in vessel density around the tumor region [8, 9], changes in oxygen saturation in the tumor microenvironment [10, 11] or alterations in the local distribution of fat, collagen and other intrinsic tissue chromophores [12]. The MSOT imaging depth is maximized for optical wavelengths around 1064 nm due to relatively low scattering and absorption of light by living tissues and the high energy of commonly available lasers at this wavelength [13]. However, optoacoustic (OA) imaging is generally incapable of accurate delineation of tumor shape and boundaries. Complementary anatomical information can be provided with pulse-echo US images rendered with hybrid systems [14–16]. Also, US imaging can be performed in transmission mode, in which case additional important mechanical and elastic tissue parameters can be extracted, such as maps of speed of sound (SoS) and acoustic attenuation (AA) [17]. Recent studies showed that SoS maps provide a powerful means to identify the tumor volume [18, 19], while AA maps can provide enhanced contrast for different tissue types [20].

In this work we employ a tri-modal transmission-reflection optoacoustic ultrasound (TROPUS) imaging platform for simultaneous characterization of solid tumors in mice. The imaging approach is based on a full ring of cylindrically-focused transducers that can provide high-resolution cross-sectional OA images in real time by exciting the tissue with a single laser pulse. Sequential excitation of the array elements and detection of the reflected and transmitted US waveforms further enables forming pulse-echo US as well as SoS images. We further employed a full wave inversion (FWI) method for reconstructing the transmission US data [17], which resulted in enhanced contrast and resolution as compared to the previously reported TROPUS implementation [21].

## 6.2 MATERIALS AND METHODS

### 6.2.1 THE IMAGING SETUP

The experimental set-up employed for image acquisition consists of four main modules, namely, a high-speed active transmission and data acquisition system (DAQ), an US ring-shaped transducer array, a pulsed laser source and a workstation computer (Fig. 6.1a). OA pressure waves are generated by illuminating the tissues with a nanosecond Nd:YAG pumped laser source (Spectra-Physics, Santa Clara, CA, USA) operating at 15 Hz pulse repetition rate. The full-ring-shaped transducer array was custom engineered (Imasonic Sas, Voray, France) for tomographic cross-sectional small animal imaging. It consists of 512 cylindrically-focused transducer elements with dimensions  $0.37 \times 15 \text{ mm}^2$  and interelement pitch of 0.47 mm. All the transducer elements are distributed equidistantly on two arcs, each covering a  $174^\circ$  angle. The radius of curvature is 40 mm, and every single transducer element is cylindrically-focused at 38 mm distance (x-z plane in Fig. 6.1a) to create a cross-sectional imaging geometry. The peak central frequency and transmission/reception bandwidth of the array elements at -6 dB are 5 MHz and 60%, respectively. The US array generates pressure waves used to interrogate the imaged sample in transmission US imaging mode while also detecting the pressure waves transmitted/reflected or generated within the imaged cross-section in the pulse-echo US and OA modes, respectively. In the US transmission mode, the excitation pulses are transmitted by the DAQ to each element of the array to generate US waves. When DAQ is switched to the receive mode, the detected pressure signals collected by the elements of the transducer array are digitized and transmitted over 1 Gbit/s ethernet connection to the host PC. Digitization sampling rate of 40 mega samples per second (MSPS) and vertical resolution of 12 bits were used for data acquisition. For collecting three-dimensional (3D) image data, the US array was translated in the vertical direction (y axis in Fig. 6.1a) with 1 mm step size using a motorized stage. For mouse imaging, the array was placed inside a temperature-controlled ( $34^\circ\text{C}$ ) water tank to increase the acoustic coupling efficiency between the imaged object and US sensors. The workstation computer having 128 GB of random access memory (RAM) and NVIDIA GeForce GTX 1060 graphics processing unit (GPU) synchronizes the DAQ and the laser by setting the transmission parameters and controlling reception events. It is also used to record and process the acquired signals to reconstruct images.

### 6.2.2 OPTOACOUSTIC TOMOGRAPHY

OA tomographic imaging of mice was performed at 1064 nm as this particular wavelength is known to have deep penetration into living mammalian tissues [13]. A fiber bundle (Light-GuideOptics GmbH, Rheinbach, Germany) separated into 12 output ferules on its distal end was used to deliver the light beam from the laser output to the imaging sample. For this, 6 output ferules were placed with 60 degrees separation (equidistantly) on each side of the transducer array (Fig. 6.1a) to facilitate uniform light delivery to the imaged mouse cross section. The output ferules of the bundle having  $0.21 \times 12.65 \text{ mm}^2$  dimensions were tilted  $24^\circ$  to attain an illumination ring with an area of  $6 \text{ cm}^2$  upon the mouse surface. The pressure waves excited within the sample were received with 512 elements after every laser pulse and

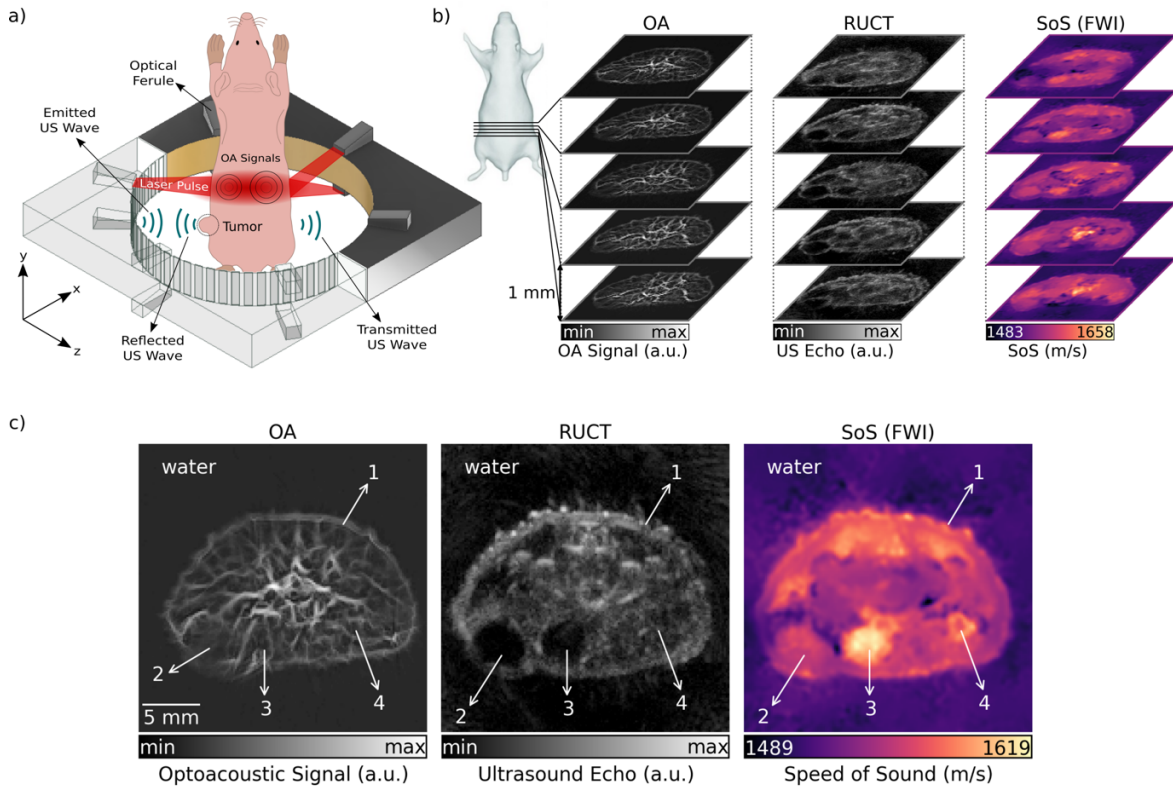


FIGURE 6.1: The tri-modal transmission-reflection optoacoustic ultrasound (TROPUS) imaging platform. a) Excitation and acquisition steps in the optoacoustic (OA) imaging mode, reflection ultrasound computed tomography (RUCT) mode and transmission ultrasound computed tomography (TUCT) speed of sound (SoS) imaging mode. b) Illustration of three-dimensional (3D) stacks of cross-sectional multimodal images acquired non-invasively from tumor-bearing mice. c) Representative TROPUS images of a cross-section of the tumor region in a mouse. From left to right, OA image, RUCT image, and SoS image acquired in the TUCT mode and reconstructed with the full wave inversion (FWI) method. 1 - Skin, 2 - Tumor, 3 - Urinary Bladder, 4 - Femur.

simultaneously digitized with the DAQ. OA images over a field of view (FOV) of  $25 \times 25 \text{ mm}^2$  were reconstructed using a back-projection algorithm after band-pass filtering the raw data in the 0.5-6 MHz frequency range [22]. The mouse boundary was manually segmented in the OA images to suppress the background. The images were subsequently normalized with a modified Bessel-function that was previously shown to approximate well the diffuse light distribution within a homogenous scattering and absorbing cylinder [23]. Finally, a vesselness (Frangi) filter was applied on the images to increase the vascular contrast [24].

### 6.2.3 REFLECTION (PULSE-ECHO) ULTRASOUND COMPUTED TOMOGRAPHY

US imaging was performed by sending a short pulse consisting of one cycle of bipolar signal ( $20 V_{pp}$ ) with duration of 0.16 microseconds to each element of the array in every transmission event. The transmission events were repeated to transmit pressure signals with all array elements. In one transmission event, the DAQ can transmit with a single element and receive reflected or transmitted signals from other 128 elements. Thereby, the pulse transmission events for each element were repeated 4 times so that the signals from all 512 elements ( $360^\circ$  full coverage) are acquired. The reflection ultrasound computed tomography (RUCT) images

were reconstructed with the synthetic transmit aperture (STA) technique. STA uses different single element in each transmission event and then coherently compounds the images from those transmission events to form the final image [25, 26]. For the beamforming process, 64 elements located to the left and 64 elements to the right from the transmitting element were included. In total 129 channels including the signals detected by the transmitting element were used for reconstruction of one sub-frame for every transmission event. The standard delay and sum (DAS) algorithm was used for reconstructing low-resolution sub-frames over  $25 \times 25 \text{ mm}^2$  FOV equivalent to that of the OA images. This process resulted in 512 low resolution images that were acquired by each transmission event, which were then coherently compounded to form the final high-resolution image.

#### 6.2.4 SPEED OF SOUND IMAGING

SoS tomography images were reconstructed from the US waves traversing the sample (mouse). For each transmitting element, the signals collected from 171 elements located on its the opposite side were considered (Fig. 6.1a). SoS maps were reconstructed with a FWI method [17]. A time of flight (TOF) picker algorithm was used to calculate the difference between the TOF of waves propagating in water and through the sample [27, 28]. This TOF picker algorithm was improved by weighting, median filtering and reciprocal pair comparison of the calculated TOF values, as previously described [29]. The wave propagation model was based on sampling the space between emitter and receiver along multiple paths using a family of Bezier curves. In short, the FWI method convolves a reference waveform with estimated TOF values from different paths corresponding to the defined curves [30]. Then, it minimizes the cost function between simulated waves and the measurements by changing SoS values in the defined image grid. This process is repeated iteratively until the cost function converges. The estimated SoS values provide the corresponding wave propagation speed for the defined cross-sectional reconstruction grid containing the mouse and background medium. Herein, the FWI method was employed to achieve improved resolution and contrast in the transmission US imaging mode as compared to the previously reported TROPUS implementation [21], which used the less accurate bent-rays approach. In the latter case, the wave propagation was significantly simplified and modeled as a narrow ray going through the path with the lowest TOF between the emitter and the receiver. Despite its merits, FWI has high computational complexity, which results in 5 minutes of reconstruction time per slice when using GPU. In this study, we further performed a comparison of the SoS maps reconstructed with all the three approaches. Transmission ultrasound further enables reconstructing AA maps. However, these were not considered in the current study due to lack of valuable tumor-related contrast.

#### 6.2.5 ANIMAL HANDLING

In total, 5 mice of the same age were imaged with the TROPUS system. All mice were anesthetized with 1.8% isoflurane in 100% oxygen flowing at a rate of 0.8 l/min. A custom-designed animal holder was used to keep the imaged mouse in vertical position inside the ring-shaped detector array. The head of the mouse was kept above the water level and

a mask was used to deliver the oxygen-anesthesia mixture. The water temperature was maintained at 34°C during the measurements. One of the mice (M1) was used as a control with no tumor. The other 4 mice had orthotopic tumors induced via inoculation of 4T1 mammary carcinoma cells in the mammary fat pad. Cell inoculation was performed at different days to characterize the ability of the TROPUS system to image tumors at different stages. Specifically, 2 orthotopic tumors in mice M3 and M4 were induced one month before the experiment, while the tumor cells were inoculated 3 weeks before the experiment in mice M2 and M5. All procedures involving animal care and experimentation were conducted in full compliance with the institutional guidelines of the Institute for Biological and Medical Imaging and with approval from the government of Upper Bavaria.

### 6.3 RESULTS

The representative three-dimensional (3D) image stacks acquired non-invasively from a tumor-bearing mouse are shown in Fig. 6.1b, corroborating the system's ability to simultaneously deliver whole body multi-modal OA and US data from mice. Representative cross-sections of the OA, RUCT, and SoS images covering 25x25 mm<sup>2</sup> FOV containing the tumor are further shown in Fig. 6.1c. The tumor location can be readily identified as a hypoechoic region in the RUCT image and as a region with lower vessel density in the OA image. In addition, SoS images enable the tissue assessment based on the distribution of its elastic modulus and density.

Note that the quality of the SoS images strongly depends on the inversion method. Indeed, the tumor mass is barely visible in the images reconstructed using straight ray approximation, which generally exhibit size-distortion due to refraction and poor contrast and resolution (first column in Fig. 6.2). When a more accurate modeling approach is attempted for the SoS reconstruction assuming Bezier-curve type of wave propagation, the reconstructed image quality readily improves (second column in Fig. 6.2). In our previous work the Bezier curve approximation was shown to be sufficiently accurate to enable the segmentation of outer boundaries and major anatomical structures [21]. However, quantitative analysis of smaller regions such as tumors cannot reliably be performed with this approach due to insufficient spatial resolution and contrast. While increased values of SoS are observed in the tumor region, similar values also appear in other regions, thus hampering unequivocal tumor differentiation. The state-of-the-art FWI reconstruction method significantly improves the resolution, contrast, and overall quality of the SoS maps (third column in Fig. 6.2), facilitating clear delineation of the tumor boundaries and other anatomical structures. This is also shown in Fig. 6.1c by comparing RUCT and SoS images. Further analysis was done to compare SoS values in different anatomical regions, namely tumor and urinary bladder, estimated with each reconstruction method. These anatomical structures were manually segmented in the cross sections rendered with the FWI method. The calculated mean and standard deviation of the SoS values in the tumor were  $1614 \pm 11.45$  m/s,  $1544 \pm 3.60$  m/s and  $1564 \pm 6.17$  m/s for the straight ray approximation, the Bezier curves method and the FWI method, respectively. The corresponding values for the urinary bladder were  $1611 \pm 7.84$  m/s,  $1541 \pm 2.14$  m/s and  $1584 \pm 11.02$  m/s. Thus, no clear distinction between the tumor and urinary bladder can be

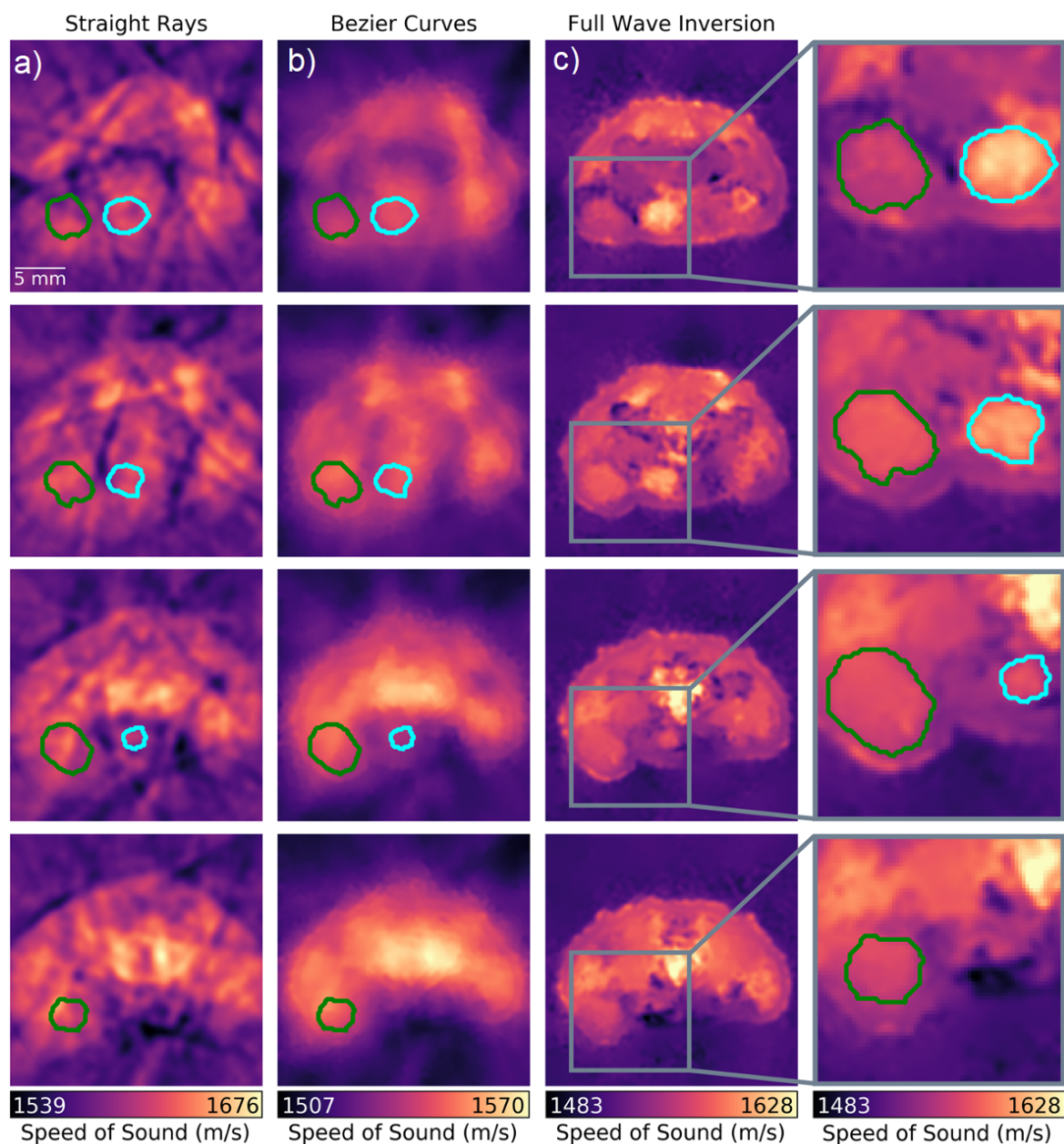


FIGURE 6.2: Comparison between different methods for rendering the speed of sound (SoS) images with transmission ultrasound computed tomography (TUCT). (a) Straight ray approximation. (b) Bezier curve reconstruction. (c) Full wave inversion (FWI). Reconstructions from different cross sections acquired from the same mouse with 1 mm steps are shown – see Fig. 6.1b. The SoS values were calculated over the tumor (green curve) and urinary bladder (cyan curve) areas manually segmented in the FWI images.

made by analyzing the SoS values rendered with the straight ray and Bezier curves methods, yet such differentiation is possible based on values extracted with the FWI methods. The volumes of the tumor and urinary bladder regions were further estimated by integrating the segmented regions over consecutive slices. This resulted in  $219 \text{ mm}^3$  and  $196 \text{ mm}^3$  estimates for the tumor and urinary bladder, respectively.

We subsequently analyzed *in vivo* data from  $n = 4$  tumor-bearing mice (M2-M5) and a



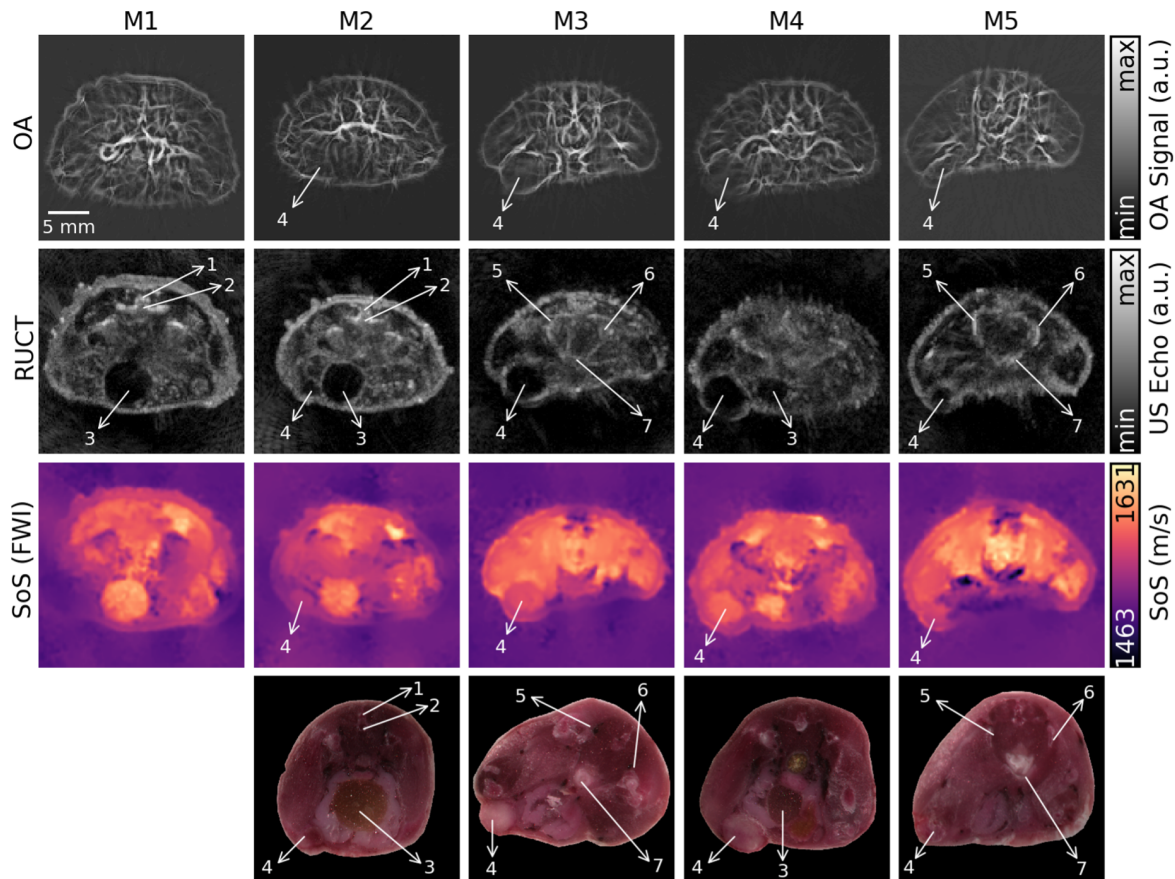


FIGURE 6.3: Cross-sectional multi-modal images of the tumor-bearing (M2-M5) and tumor free (M1) mice acquired from approximately the same abdominal region. Histological cryo-sections taken from approximately corresponding regions *ex vivo* are shown in the bottom row. The tumors and other anatomical structures are labeled: 1 - Vertebral Column, 2 - Caudal Vertebrae, 3 - Urinary Bladder, 4 - Tumor, 5 - Coxal Bone, 6 - Ischium, 7 - Female Urethra.

tumor-free mouse (M1). The results are shown in Fig. 6.3. While vascular density/size is clearly altered in the tumor regions according to the OA data, the lesion boundaries cannot be accurately discerned from those images. The RUCT data is used instead for anatomical guidance and segmentation of the tumor boundaries. Yet, anechoic or hypoechoic structures corresponding to malignant tumors may easily be confused with other mouse organs. For example, the urinary bladder, clearly distinguished as a low intensity region in the RUCT image of M1 (Fig. 6.3), exhibits similar characteristics to tumors regions shown for other mice. In this regard, the additional information provided by OA facilitates classifying this region as benign since no increase in vessel density or thickness occurs around expected tumor region. OA images from M3 and M4 show increased vessel density around the tumor region but not in the tumor core, which is consistent with previous studies [10, 31]. However, OA images are often corrupted with streak type artifacts [32] that can be observed in the urinary bladder area. Note that such artifacts are commonly amplified by the Frangi filtering (see image comparison in the supplementary information), which may lead to misinterpreting the artifacts for blood vessels. In general, the streak artifacts can be mitigated by using an imaging system with higher number of elements or employing more sophisticated reconstruction approaches

[33]. Note also that the RUCT images of M3 and M4 exhibit several regions with low intensity, which seem to be difficult to classify as benign or malignant even in conjunction with the vascularization information provided by the OA data. This turns even more challenging for M2 and M5, which have small-sized tumors. In those cases, the SoS data may serve as a complementary modality for increasing specificity of tumor detection and characterization. The tumor locations identified by the multi-modal *in vivo* TROPUS imaging were further confirmed by studying the histological cryo-sections taken from approximately corresponding regions *ex vivo* (last row in Fig. 6.3). Note that while the anatomical structures visible in the cryo-sections were generally matching the information obtained by TROPUS, the exact shape and size of the different structures might have changed due to compression and freezing of the samples.

To compare the information provided by reflection and transmission ultrasound, the tumors were manually segmented in the RUCT and SoS images following their identification and localization using the multi-modal data. The segmented regions based on the SoS (Fig. 6.4a) and RUCT (Fig. 6.4b) data were then superimposed on the OA images (Fig. 6.4c). Even though the US-based segmentations generally match the regions with low vessel density in the OA images, robust tumor differentiation based on vascular OA features seems challenging. We subsequently generated binary masks from the segmented region of interests (ROIs) and extracted the mean and standard deviation of the SoS values in the tumor regions (Fig. 6.4d). The measured mean SoS values in the tumors ranged from 1541 to 1572 m/s with the respective standard deviations ranging from 2.77 to 6.26 m/s. Also, the SoS values increased as a function of tumor size, though the limited sample size prevents establishing such a correlation unambiguously. In general, the reconstructed SoS values in the tumor regions are in the range of previously reported data for solid breast tumors [34], yet longitudinal study with an increased sample size is needed to validate quantification by the proposed methods.

Finally, the anatomical localization capabilities of the RUCT and SoS images were compared based on the segmented tumor areas from both modalities. While comparison between the extracted tumor areas yields similar values for both modalities (Fig. 6.4e), slightly larger area estimations were generally obtained when segmenting tumors based on the RUCT images in 3 out of 4 tumor-bearing mice. In one mouse having the smallest tumor size (M2), the segmented area was slightly larger in the SoS image, though the actual SoS values were smaller, which may just indicate an early stage of the tumor development.

#### 6.4 DISCUSSION AND CONCLUSION

The presented results indicate that the marriage between diverse OA and US contrasts in one single TROPUS platform has the potential to provide complementary information for characterizing mammary tumors in mice. Angiogenesis is a central hallmark of solid tumors, representing formation of new vascular network necessary to support tumor growth and metastasis. In our study, neo-vasculature was clearly observed in the areas surrounding the tumors in the OA images. However, tumor boundaries could not clearly be discerned

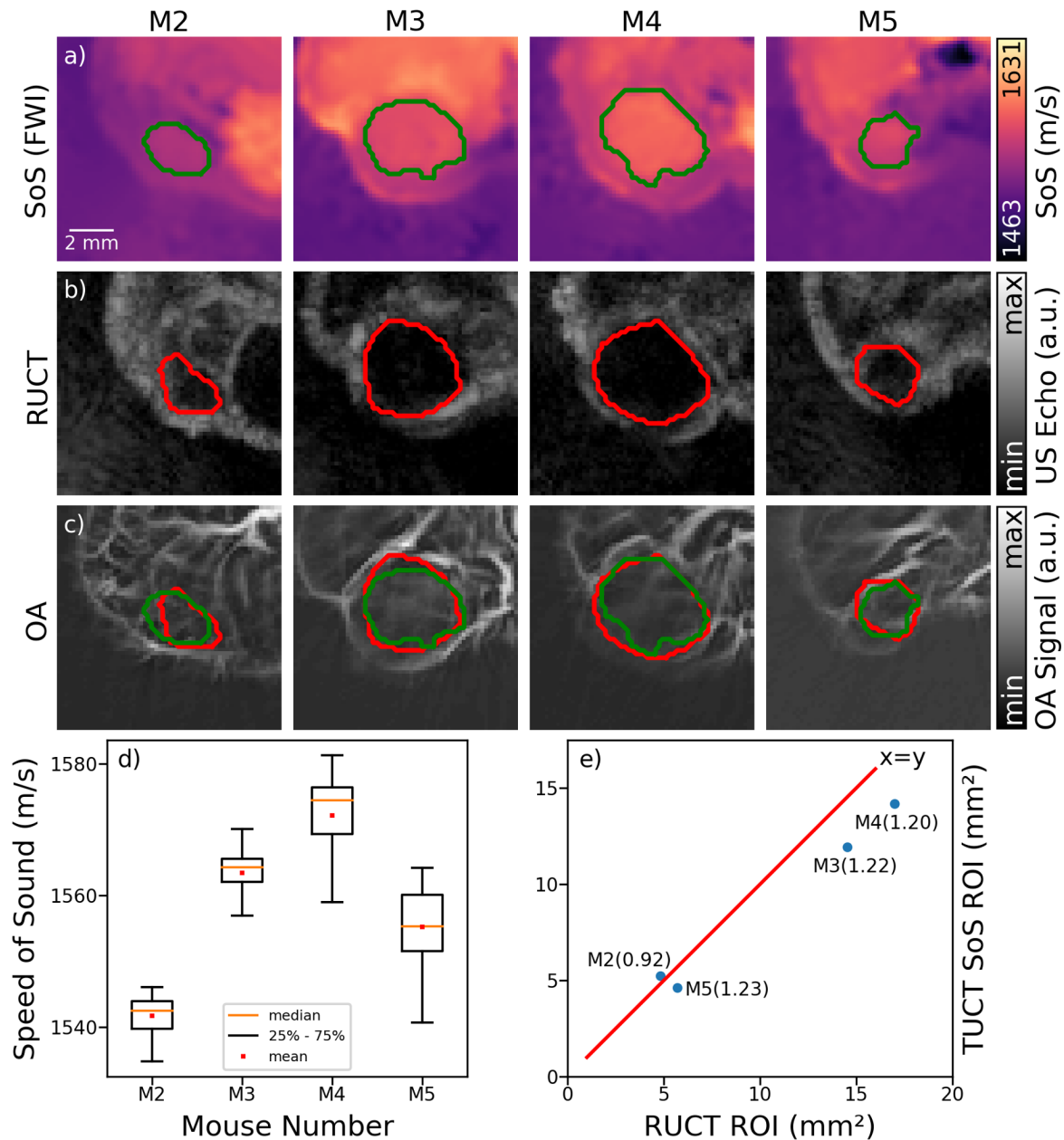


FIGURE 6.4: Tumor segmentation and characterization. (a) The manually segmented tumor areas in the speed of sound (SoS) images reconstructed with full wave inversion (FWI) method. (b) The corresponding segmentations based on the reflection ultrasound computed tomography (RUCT) images. (c) The segmented areas superimposed onto the optoacoustic (OA) images. (d) Measured mean and standard deviation of SoS values inside the segmented region of interests (ROIs) based on SoS-based tumor segmentations. (e) Comparison of the measured tumor areas based on the SoS- and RUCT-based tumor segmentations. The ratios between the calculated areas are presented for each imaged mouse.

based on the vascular anatomy contrast provided by the OA data. On the other hand, the tumors appeared as anechoic or hypoechoic structures in pulse-echo US, although some mouse organs like the urinary bladder may have a similar appearance in those images. The improved resolution and contrast of SoS maps reconstructed with the FWI method facilitated the delineation of the tumor mass. In this case, we were able to assign the tumors to areas having

---

sharp boundaries and a relatively uniform SoS different from the background. The extracted average values of SoS in the tumor regions could also potentially be used to distinguish malignant lesions from other regions with uniform SoS. Generally, reliable identification of tumors appears to be challenging in images from standalone modalities and the complementary information provided by OA, pulse-echo and SoS images aided an unambiguous identification. Yet, further work is required to strengthen the synergistic and complementary value of the suggested hybrid imaging methodology.

Even though the present study solely focused on the anatomical imaging capabilities of OA and US, these modalities are generally equipped with a range of additional functional and molecular imaging features that can aid tumor identification and characterization. Previous studies using MSOT approaches have demonstrated high resolution readings of tumor oxygenation gradients across tumors [35] as well as perfusion and targeted uptake of nanoparticles and other molecular agents by the tumor microenvironment [31, 36]. Similarly, Doppler and contrast-enhanced US imaging have been used for multi-parametric characterization of functional tumor parameters [37–39]. From the imaging point of view, hybridization of multiple modalities based on US and OA excitation can enhance the reconstruction accuracy of those methods by exploiting synergistic information on the underlying optical and acoustic tissue properties [32, 40]. The accuracy and quantification capabilities of the proposed system shall be validated in future longitudinal studies.

Clinical translation potential is another important aspect of the TROPUS platform that can be explored for identifying new diagnostics bio-markers of breast cancer. To this end, pulse-echo US is routinely used in the clinics for anatomical guidance and characterization of breast lesions. Transmission US has also shown promising diagnostic results in clinical trials [41]. Initial clinical studies aimed at early breast cancer detection have also been performed with OA imaging [8]. In one recent study, a full-ring array analogous to the one used in the TROPUS system has been tested for OA imaging of human breast [42], further supporting the clinical translation potential of our approach.

In conclusion, we demonstrated the potential of TROPUS imaging for detection and characterization of mammary tumors in mice. The reported findings corroborate the strong potential of the hybrid imaging approach for advancing cancer research in small animal models and fostering development of new clinical diagnostic approaches.

## ACKNOWLEDGEMENT

This work was supported by the German Research Foundation research grant RA1848/5-1.

## DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors declare that they have no competing financial interests.

## AUTHORS' CONTRIBUTION

Conception and design: E. Merčep, D. Razansky

Development of methodology: B. Lafci, E. Merčep, D. Razansky

Acquisition of data (provided animals, acquired, and managed patients, provided facilities, etc.): E. Merčep, D. Razansky

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B. Lafci, J. L. Herraiz, X. L. Deán-Ben

Writing, review, and/or revision of the manuscript: B. Lafci, J. L. Herraiz, X. L. Deán-Ben, D. Razansky

Study supervision: D. Razansky

## SUPPLEMENTARY FIGURE

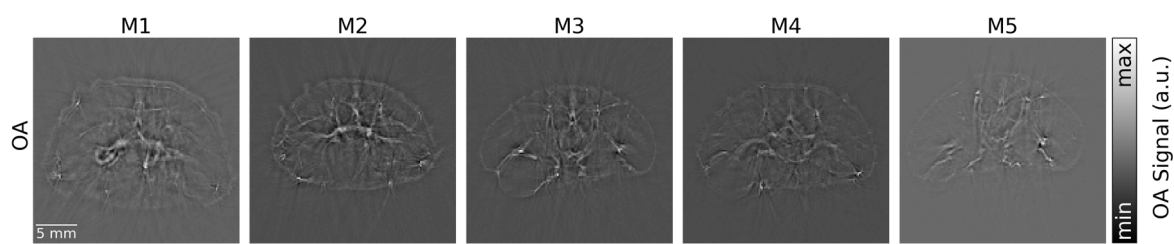


FIGURE 6.5: Cross-sectional optoacoustic (OA) images without Frangi filter. The same OA images with Frangi filter are shown in Fig. 6.3.

## BIBLIOGRAPHY

- [1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. “Cancer statistics, 2019”. In: *CA: a cancer journal for clinicians* 69.1 (2019), pp. 7–34.
- [2] Diana L Lam et al. “Imaging-based screening: understanding the controversies”. In: *AJR. American journal of roentgenology* 203.5 (2014), p. 952.
- [3] Wendie A Berg et al. “Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer”. In: *Jama* 299.18 (2008), pp. 2151–2163.
- [4] Valerie P Jackson et al. “Imaging of the radiographically dense breast.” In: *Radiology* 188.2 (1993), pp. 297–301.
- [5] Monica Morrow, Janet Waters, and Elizabeth Morris. “MRI for breast cancer screening, diagnosis, and treatment”. In: *The Lancet* 378.9805 (2011), pp. 1804–1811.
- [6] Qiuting Li et al. “Combining Ultrasound and X-Ray Imaging for Mammography: A Prototype Design”. In: *Bildverarbeitung für die Medizin 2019: Algorithmen-Systeme-Anwendungen. Proceedings des Workshops vom 17. bis 19. März 2019 in Lübeck*. Springer, 2019, pp. 245–250.
- [7] Srirang Manohar and Maura Dantuma. “Current and future trends in photoacoustic breast imaging”. In: *Photoacoustics* 16 (2019), p. 100134.
- [8] Mohammad Mehrmohammadi et al. “Photoacoustic imaging for cancer detection and staging”. In: *Current Molecular Imaging (Discontinued)* 2.1 (2013), pp. 89–105.
- [9] Isabel Quiros-Gonzalez et al. “Optoacoustics delineates murine breast cancer models displaying angiogenesis and vascular mimicry”. In: *British journal of cancer* 118.8 (2018), pp. 1098–1106.
- [10] Avihai Ron et al. “Volumetric Optoacoustic Imaging Unveils High-Resolution Patterns of Acute and Cyclic Hypoxia in a Murine Model of Breast Cancer Imaging High-Resolution Patterns of Acute and Cyclic Hypoxia”. In: *Cancer research* 79.18 (2019), pp. 4767–4775.
- [11] Emma Brown, Joanna Brunker, and Sarah E Bohndiek. “Photoacoustic imaging as a tool to probe the tumour microenvironment”. In: *Disease models & mechanisms* 12.7 (2019), p. dmm039636.
- [12] Gael Diot et al. “Multispectral Optoacoustic Tomography (Msot) of Human Breast Cancer msot Signatures of Human Breast Cancer”. In: *Clinical Cancer Research* 23.22 (2017), pp. 6912–6922.

- [13] Arunima Sharma, Vijitha Periyasamy, and Manojit Pramanik. “Photoacoustic imaging depth comparison at 532-, 800-, and 1064-nm wavelengths: Monte Carlo simulation and experimental validation”. In: *Journal of Biomedical Optics* 24.12 (2019), pp. 121904–121904.
- [14] Elena Merčep et al. “Whole-body live mouse imaging by hybrid reflection-mode ultrasound and optoacoustic tomography”. In: *Optics letters* 40.20 (2015), pp. 4643–4646.
- [15] Jeesu Kim et al. “Programmable real-time clinical photoacoustic and ultrasound imaging system”. In: *Scientific reports* 6.1 (2016), p. 35137.
- [16] Milan Oeri et al. “Hybrid photoacoustic/ultrasound tomograph for real-time finger imaging”. In: *Ultrasound in Medicine & Biology* 43.10 (2017), pp. 2200–2212.
- [17] M Pérez-Liva et al. “Time domain reconstruction of sound speed and attenuation in ultrasound computed tomography using full wave inversion”. In: *The Journal of the Acoustical Society of America* 141.3 (2017), pp. 1595–1604.
- [18] Adrian Taruttis, Gooitzen M van Dam, and Vasilis Ntziachristos. “Mesoscopic and macroscopic optoacoustic imaging of cancer”. In: *Cancer research* 75.8 (2015), pp. 1548–1559.
- [19] Xosé Luís Deán-Ben, Steven James Ford, and Daniel Razansky. “High-frame rate four dimensional optoacoustic tomography enables visualization of cardiovascular dynamics and mouse heart perfusion”. In: *Scientific reports* 5.1 (2015), p. 10133.
- [20] Hiroaki Okawai, Kazuto Kobayashi, and Shin’ichi Nitta. “An approach to acoustic properties of biological tissues using acoustic micrographs of attenuation constant and sound speed.” In: *Journal of Ultrasound in Medicine: Official Journal of the American Institute of Ultrasound in Medicine* 20.8 (2001), pp. 891–907.
- [21] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [22] X Luís Deán-Ben, Vasilis Ntziachristos, and Daniel Razansky. “Effects of small variations of speed of sound in optoacoustic tomographic imaging”. In: *Medical physics* 41.7 (2014), p. 073301.
- [23] Daniel Razansky and Vasilis Ntziachristos. “Hybrid photoacoustic fluorescence molecular tomography using finite-element-based inversion”. In: *Medical physics* 34.11 (2007), pp. 4293–4301.
- [24] Alejandro F Frangi et al. “Multiscale vessel enhancement filtering”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI’98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*. Springer. 1998, pp. 130–137.
- [25] Søren K Jespersen, Jens E Wilhjelm, and Henrik Sillesen. “Multi-angle compound imaging”. In: *Ultrasonic imaging* 20.2 (1998), pp. 81–102.
- [26] Ihor Trots et al. “Synthetic aperture method in ultrasound imaging”. In: *Ultrasound Imaging* (2011), pp. 37–56.



- [27] Cuiping Li et al. “An improved automatic time-of-flight picker for medical ultrasound tomography”. In: *Ultrasonics* 49.1 (2009), pp. 61–72.
- [28] Erol Kalkan. “An automatic P-phase arrival-time picker”. In: *Bulletin of the Seismological Society of America* 106.3 (2016), pp. 971–986.
- [29] Cuiping Li, Neb Duric, and Lianjie Huang. “Comparison of ultrasound attenuation tomography methods for breast imaging”. In: *Medical Imaging 2008: Ultrasonic Imaging and Signal Processing*. Vol. 6920. SPIE. 2008, pp. 338–346.
- [30] Mailyn Perez-Liva et al. “Speed of sound ultrasound transmission tomography image reconstruction based on Bézier curves”. In: *Ultrasonics* 103 (2020), p. 106097.
- [31] Eva Herzog et al. “Optical imaging of cancer heterogeneity with multispectral optoacoustic tomography”. In: *Radiology* 263.2 (2012), pp. 461–468.
- [32] Elena Merčep et al. “Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.9 (2015), pp. 1651–1661.
- [33] X Luís Dean-Ben et al. “Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography”. In: *IEEE transactions on medical imaging* 31.10 (2012), pp. 1922–1928.
- [34] Cuiping Li et al. “In vivo Breast Sound-Speed Imaging with Ultrasound Tomography”. In: *Ultrasound in Medicine & Biology* 35.10 (2009), pp. 1615–1628.
- [35] Michal R Tomaszewski et al. “Oxygen enhanced optoacoustic tomography (OE-OT) reveals vascular dynamics in murine models of prostate cancer”. In: *Theranostics* 7.11 (2017), p. 2900.
- [36] Katheryne E Wilson, Keerthi S Valluru, and Jürgen K Willmann. “Nanoparticles for Photoacoustic Imaging of Cancer”. In: *Design and Applications of Nanoparticles in Biomedical Imaging* (2017), pp. 315–335.
- [37] HHT Madsen and F Rasmussen. “Contrast-enhanced ultrasound in oncology”. In: *Cancer Imaging* 11.1A (2011), S167.
- [38] Hyun-Jung Jang, Hojun Yu, and Tae Kyoung Kim. “Contrast-enhanced ultrasound in the detection and characterization of liver tumors”. In: *Cancer Imaging* 9.1 (2009), p. 96.
- [39] L Alamo and U Fischer. “Contrast-enhanced color Doppler ultrasound characteristics in hypervascular breast tumors: comparison with MRI”. In: *European radiology* 11 (2001), pp. 970–977.
- [40] Xosé Luis Deán-Ben, E Merčep, and Daniel Razansky. “Hybrid-array-based optoacoustic and ultrasound (OPUS) imaging of biological tissues”. In: *Applied Physics Letters* 110.20 (2017), p. 203703.
- [41] Mark Sak et al. “Using speed of sound imaging to characterize breast density”. In: *Ultrasound in medicine & biology* 43.1 (2017), pp. 91–103.

- 
- [42] Sjoukje M Schoustra et al. “Twente Photoacoustic Mammoscope 2: system overview and three-dimensional vascular network images in healthy breasts”. In: *Journal of biomedical optics* 24.12 (2019), pp. 121909–121909.

## 7 MULTIMODAL ASSESMENT OF NON-ALCOHOLIC FATTY LIVER DISEASE WITH TRANSMISSION-REFLECTION OPTOACOUSTIC ULTRASOUND

This chapter includes the following publication:

Under Review (2023).

Authors:

**Berkan Lafci**<sup>1,2</sup>, Anna Hadjichambi<sup>3,4,5</sup>, Christos Konstantinou<sup>3,4,5</sup>, Joaquin L. Herraiz<sup>6,7</sup>,  
Luc Pellerin<sup>5,8</sup>, Neal C. Burton<sup>9</sup>, Xosé Luís Deán-Ben<sup>1,2</sup>, Daniel Razansky<sup>1,2</sup>

<sup>1</sup> Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, University of Zurich, Switzerland

<sup>2</sup> Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

<sup>3</sup> The Roger Williams Institute of Hepatology, Foundation for Liver Research, London, UK

<sup>4</sup> Faculty of Life Sciences and Medicine, King's College London

<sup>5</sup> Department of Biomedical Sciences, University of Lausanne, Lausanne, Switzerland

<sup>6</sup> Nuclear Physics Group and IPARCOS, Complutense University of Madrid, Madrid, Spain

<sup>7</sup> Health Research Institute of Hospital Clínico San Carlos (IdISSC), Madrid, Spain

<sup>8</sup> Inserm U1313, Université et CHU de Poitiers, Poitiers, France

<sup>9</sup> iThera Medical GmbH, Munich, Germany

## ABSTRACT

### Rationale

Non-alcoholic fatty liver disease (NAFLD) is an umbrella term referring to a group of conditions associated to fat deposition and damage of liver tissue. Early detection of fat accumulation is essential to avoid progression of NAFLD to serious pathological stages such as liver cirrhosis and hepatocellular carcinoma.

### Methods

We exploited the unique capabilities of transmission-reflection optoacoustic ultrasound (TROPUS), which combines the advantages of optical and acoustic contrasts, for an early-stage multiparametric assessment of NAFLD in mice.

### Results

The multispectral optoacoustic imaging allowed for spectroscopic differentiation of lipid content, as well as the bio-distributions of oxygenated and deoxygenated hemoglobin in liver tissues *in vivo*. The pulse-echo (reflection) ultrasound (US) imaging further provided a valuable anatomical reference whilst transmission US facilitated the mapping of speed of sound changes in lipid-rich regions, which was consistent with the presence of macrovesicular hepatic steatosis in the NAFLD livers examined with *ex vivo* histological staining.

### Conclusion

The proposed multimodal approach facilitates quantification of liver abnormalities at early stages using a variety of optical and acoustic contrasts, laying the ground for translating the TROPUS approach toward diagnosis and monitoring NAFLD in patients.

## 7.1 INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is a common disorder comprising a progressive spectrum of diseases, defined as an accumulation of fat in the liver (steatosis), in the absence of significant alcohol consumption [1]. NAFLD progresses to non-alcoholic steatohepatitis, characterized by inflammation and hepatocyte damage (which includes ballooning and cell death), together with deposition of collagen and fibrosis progression [2], results in enlarging and discoloring of the organ [3]. Further progression of fibrosis may lead to the irreversible stages of cirrhosis and, eventually, hepatocellular carcinoma [4]. At present, no approved interventions are available to treat liver fibrosis, which calls for the development of new research tools aimed at better understanding the underlying causes of NAFLD, as well as new methods capable of detecting this condition at the earliest reversible stage before it progresses to fibrosis [5, 6]. NAFLD and liver fibrosis have become a major health concern due to the growing prevalence of overweight and obese individuals in developed countries [7]. The worldwide mortality rate related to liver diseases follows an upward trend, reaching 2 million disease-related deaths annually in 2019 [8]. However, detection of early liver damage is challenged by the small size and sparsity of the scars formed before the appearance of fibrosis [9]. Currently, liver disease assessment is performed with biopsies and histopathology imaging [10]. Liver biopsy is however an invasive and user-dependent (sampling bias) procedure hindering a continuous monitoring of liver tissue abnormalities. Therefore, the development of non-invasive methods enabling the quantitative assessment of NAFLD is paramount both for preclinical studies aiming at advancing our knowledge of the disease, as well as for early diagnosis purposes in the clinical setting.

Whole-body clinical imaging methods have been shown to provide important advantages for liver disease diagnosis. Magnetic resonance imaging (MRI) achieves high specificity for fat accumulation by using the proton density fat fraction technique [11]. X-ray computed tomography (CT) has also been reported for the assessment of liver abnormalities with high resolution [12]. However, the use of these methods is associated with high installation and maintenance costs, exposure to ionizing radiation, and insufficient sensitivity to molecular (fat) contrast [13, 14]. Ultrasound (US) imaging is a more affordable and accessible bedside technology which has also been used for visualizing and assessing liver abnormalities [15, 16]. Linear array probes are typically used in clinics to provide a quick assessment of the liver with pulse-echo (B-mode) US. However, this approach does not provide sufficient tomographic (angular) coverage needed for accurate localization and quantitative characterization of the damaged liver areas, further lacking the necessary contrast for assessing fat accumulation. In response, tomographic US methods have been developed to provide enhanced tissue contrast. Reflection ultrasound computed tomography (RUCT) is based on tomographic pulse-echo US imaging with waves being sequentially emitted and detected at different angular positions around the sample. The broad angular coverage has been shown to increase the image contrast, resolution and field of view (FOV) with respect to those achieved with linear arrays [17]. Transmission ultrasound computed tomography (TUCT) further enables mapping the speed of sound (SoS) distribution in tissues by considering US waves transmitted

through the sample. SoS maps have been shown to provide improved specificity for detecting fatty and glandular tissue abnormalities and delineation of lesions [18].

Hybrid optoacoustic (OA) imaging combining light with sound has emerged as another powerful functional and molecular preclinical imaging approach. It is based on optical excitation of tissues at near infrared (NIR) wavelengths and tomographic detection of the thermoelastically-induced US waves, thus rendering rich optical contrast with high spatial resolution unaffected by photon scattering in deep tissues [19, 20]. In particular, multispectral optoacoustic tomography (MSOT) capitalizes on optical excitation at different wavelengths to spectroscopically differentiate between oxy-hemoglobin ( $\text{HbO}_2$ ) and deoxy-hemoglobin (Hb), melanin, lipids and other tissue bio-chromes as well as extrinsically administered contrast agents [21, 22]. However, unambiguous anatomical differentiation of lesions and organs is hindered with MSOT whose main contrast stems from hemoglobin-rich structures such as major blood vessels.

Recently, a multi-modal transmission-reflection optoacoustic ultrasound (TROPUS) imaging has been suggested as a versatile imaging approach for multi-parametric anatomical, functional and molecular characterization of murine disease models [23, 24]. The full tomographic coverage of the circular transducer array used in TROPUS results in an improved contrast and resolution with MSOT, RUCT and TUCT, while further providing real-time imaging capabilities for visualizing dynamic processes [25, 26]. Here, we employed TROPUS for assessing early-stage NAFLD in mice. The lipid accumulation in the liver was delineated and quantified with TUCT and MSOT while RUCT facilitated anatomical interpretation. The *in vivo* imaging results were validated with Hematoxylin and Eosin (H&E) staining of excised specimens.

## 7.2 RESULTS

The TROPUS imaging setup consists of a circular US transducer array, a nanosecond laser source, a data acquisition-transmission system DAQ and a workstation PC used for the system synchronization, data transfer, storage and processing (Fig. 7.1a, see Methods for details) [24]. Imaging in the MSOT mode was performed by quickly switching the optical wavelength of the nanosecond optical parametric oscillator (OPO) laser from 740 to 940 nm with a 20 nm step size at 25 Hz repetition rate in order to enable the separation of Hb,  $\text{HbO}_2$ , melanin and lipid components, the latter having a distinct peak in its absorption spectrum at 920 nm (Fig. 7.1b) [27]. Exemplary cross-sectional MSOT images acquired from living mice at 800nm excitation wavelength are shown in Fig. 7.1c. RUCT imaging was based on the synthetic transmit aperture (STA) image acquisition technique [25], which employs sequential transmission of US pulses with each array element followed by detection of the reflected signals. Image compounding was subsequently performed by adding up multiple low-contrast delay-and-sum images acquired from different views around the sample, resulting in a final high-contrast RUCT images (Fig. 7.1d). Quantitative TUCT images representing the SoS distribution in the mouse in m/s were reconstructed from the US waves that traversed

the imaged object using a full wave inversion (FWI) algorithm (Fig. 7.1e) [28].

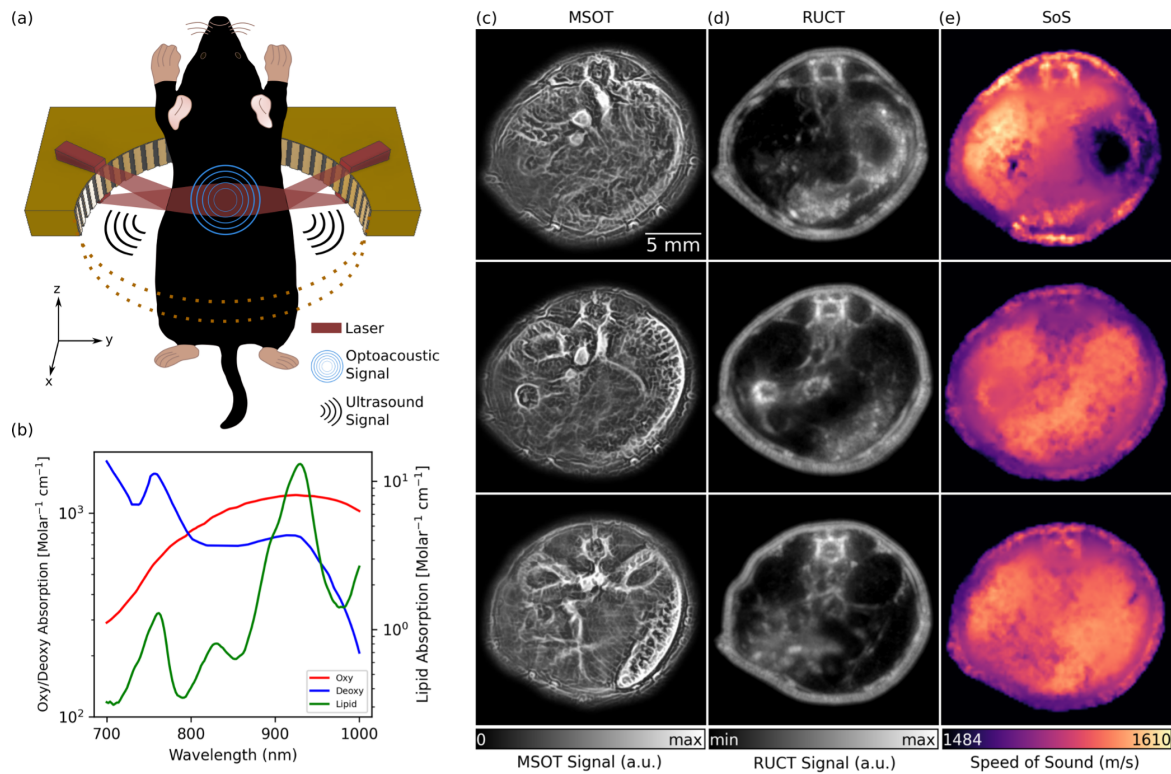


FIGURE 7.1: TROPUS imaging. (A) Lay-out of the imaging set-up combining three modalities, namely MSOT, RUCT and SoS imaging. Only half of the ring array is shown for better visualization. (B) Absorption spectrum of Hb, HbO<sub>2</sub> and lipid in 700 nm and 1000 nm wavelength range. (C) Exemplary MSOT images from different cross sections recorded at 1064 nm excitation wavelength. (D) The corresponding cross sections reconstructed with the RUCT modality. (E) The corresponding cross sections showing the SoS maps reconstructed with TUCT modality. Hb: deoxy-hemoglobin; HbO<sub>2</sub>: oxy-hemoglobin, MSOT: multispectral optoacoustic tomography; RUCT: reflection ultrasound computed tomography; SoS: speed of sound; TROPUS: transmission-reflection optoacoustic ultrasound; TUCT: transmission ultrasound computed tomography.

The basic ability of the multimodal TROPUS system to differentiate between NAFLD and control liver tissues was first evaluated with *ex vivo* samples (Fig. 7.2). Specifically, livers excised from 3 NAFLD and 3 control mice were imaged at two different vertical positions, resulting in 12 cross-sectional images. The MSOT images enable resolving the fat content by capitalizing on the distinctive optical absorption spectrum of lipids (Fig. 7.1b). The spectrally un-mixed bio-distributions of lipids (green color in Fig. 7.2a), overlaid onto the structural MSOT images rendered by averaging signals acquired at all the excitation wavelengths, clearly evince a higher fat content in the NAFLD liver tissue with respect to the control. RUCT images were further acquired for anatomical reference (Fig. 7.2b), which were used to delineate the borders of the excised livers in order to create binary masks to suppress background and conduct quantitative analysis. The corresponding SoS images acquired with TUCT manifest lower SoS values in the entire cross-sections of NAFLD livers as compared to the controls (Fig. 7.2c), which is generally expected considering a slower sound wave propagation in fat compared to liver tissue [29]. Histology images based on H&E staining were also acquired for validation (Fig. 7.2d). The spectra of the MSOT signals averaged over selected

region of interests (ROIs) revealed the presence of fat in the liver tissue from animals with NAFLD (Fig. 7.2e). Specifically, a distinctive peak at 920 nm was observed in the MSOT signal spectra matching well the known local maximum in the optical absorption of lipids [27]. This spectral peak was not present in the spectrum of the MSOT signals recorded from the control liver tissue. The lipid signals in *ex vivo* liver tissues were then averaged based on the pixel number after removing the non-distinct absorption background. The averaged lipid signal un-mixed from the MSOT images was 47% higher in NAFLD livers as compared to controls, exhibiting statistically significant differences for the 12 imaged cross-sections (Fig. 7.2f,  $p=0.001$ ). Note that a similar standard deviation (STD) of the lipid signal (12% of the average value of all images) was observed in both cases. Statistical analysis of the measured SoS values for the 12 imaged cross-sections further revealed significant differences between NAFLD liver tissues and controls (Fig. 7.1g,  $p=0.010$ ). The measured SoS mean and STD in NAFLD mice were 1495 m/s and 12 m/s, respectively, while these values were 1525 m/s and 15 m/s, respectively, in control mice. Histology images based on H&E staining further revealed macrovesicular hepatic steatosis in the NAFLD livers (Figs. 7.2d). The lipid accumulates in the hepatocytes as vacuoles detectable with H&E staining [30]. These intracytoplasmatic fat droplets were not observed in histology images of control tissues. The difference in fat content observed in histology images is thus consistent with the observations in the *in vivo* MSOT and SoS images.

TROPUS was then used to image NAFLD and control mice *in vivo*. The good anatomical contrast provided by RUCT facilitated identification of the liver cross-sections (Figs. 7.3a-b). The SoS images further provided sufficient contrast and resolution to differentiate the liver from other surrounding tissues (Figs. 7.3c-d). Segmentation of the liver was done by an experienced biologist considering both the RUCT and SoS images. This served to define binary masks to quantify differences in SoS between NAFLD and control mice. SoS values were averaged for the segmented binary masks for 4 NAFLD (20 cross-sections) and 4 control (20 cross-sections) mice. Statistical analysis revealed a significant drop in SoS in liver ROIs for the NAFLD (average: 1475 m/s, STD: 34 m/s) versus control (average: 1538 m/s, STD: 18 m/s) mice (Fig. 7.3e,  $p=0.007$ ), which is consistent with reduced SoS values in fat tissues versus healthy liver tissues [29]. A clear difference between the body weights of NAFLD and control mice was further observed (Fig. 7.3f,  $p=0.0005$ ), with mean values of 42g and 30g, respectively. The cross-sectional areas were further calculated by manually segmenting the outer boundaries of the mouse body in the RUCT images, where the skin surface was clearly distinguishable. Statistically significant differences in cross-sectional areas of NAFLD (average area: 562 mm<sup>2</sup>, STD: 29 mm<sup>2</sup>) and control (average area: 333 mm<sup>2</sup>, STD: 21 mm<sup>2</sup>) mice were also observed (Fig. 7.3g,  $p=7e-9$ ). Despite the increased body weight and cross-sectional area in NAFLD mice, RUCT manifested sufficient penetration depth to visualize structures in the central region of the mouse. Also, the transmitted US waves through mouse body were shown to have sufficient amplitude to enable reconstructing SoS images through the whole mouse body using the FWI reconstruction algorithm.

MSOT images were subsequently analyzed to visualize the distribution of different tissue chromophores. Specifically, linear un-mixing was performed by considering four components,



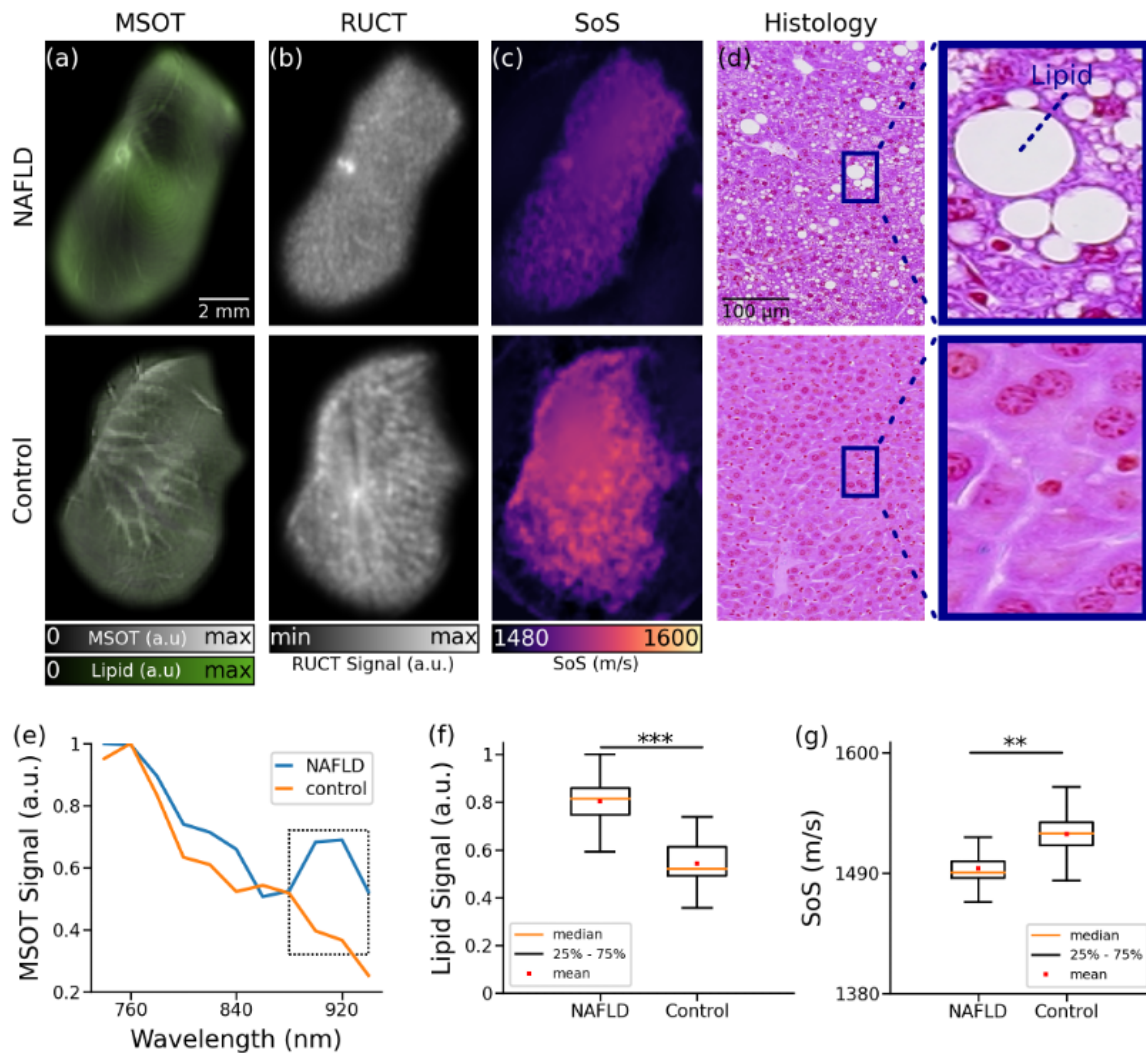


FIGURE 7.2: TROPUS imaging of liver tissues excised from NAFLD and control mice. (A) Un-mixed lipid distribution (green color) overlaid onto the anatomical MSOT images corresponding to averaged signals over all the acquired wavelengths for excised livers from mouse with NAFLD and control mouse. (B) RUCT images of excised livers from mouse with NAFLD and control mouse. (C) SoS images of excised livers from mouse with NAFLD and control mouse. (D) Histology images of excised livers from mouse with NAFLD and control mouse. (E) MSOT signal spectra of liver tissue shown in the panels a. (F) Average lipid signal intensities from 3 NAFLD and 3 control mice. (G) Average SoS values from 3 NAFLD and 3 control mice. MSOT: multispectral optoacoustic tomography; NAFLD: non-alcoholic fatty liver disease; RUCT: reflection ultrasound computed tomography; SoS: speed of sound; (p values are indicated by \* $\leq 0.05$ , \*\* $\leq 0.01$  and \*\*\* $\leq 0.001$ )

namely, Hb, HbO<sub>2</sub>, melanin and lipids. One NAFLD and one control animal were excluded from the MSOT data analysis due to the saturated signal intensity from the melanin channel as a result of skin pigmentation. MSOT images of NAFLD and control mice corresponding to averaged signals over all the wavelengths used for acquisition are shown in Figs. 7.4a-b. NAFLD mice clearly manifest an increased lipid content in the liver region (Fig. 7.4c), indicated by the yellow contour. On the contrary, a relatively low accumulation of fat in the liver was observed in control mice (Fig. 7.4d). Much like for the *ex vivo* samples, analysis of the MSOT signal spectra averaged over the liver areas enabled detection of lipids. While spectra from both NAFLD and control mice monotonically decrease with wavelength,

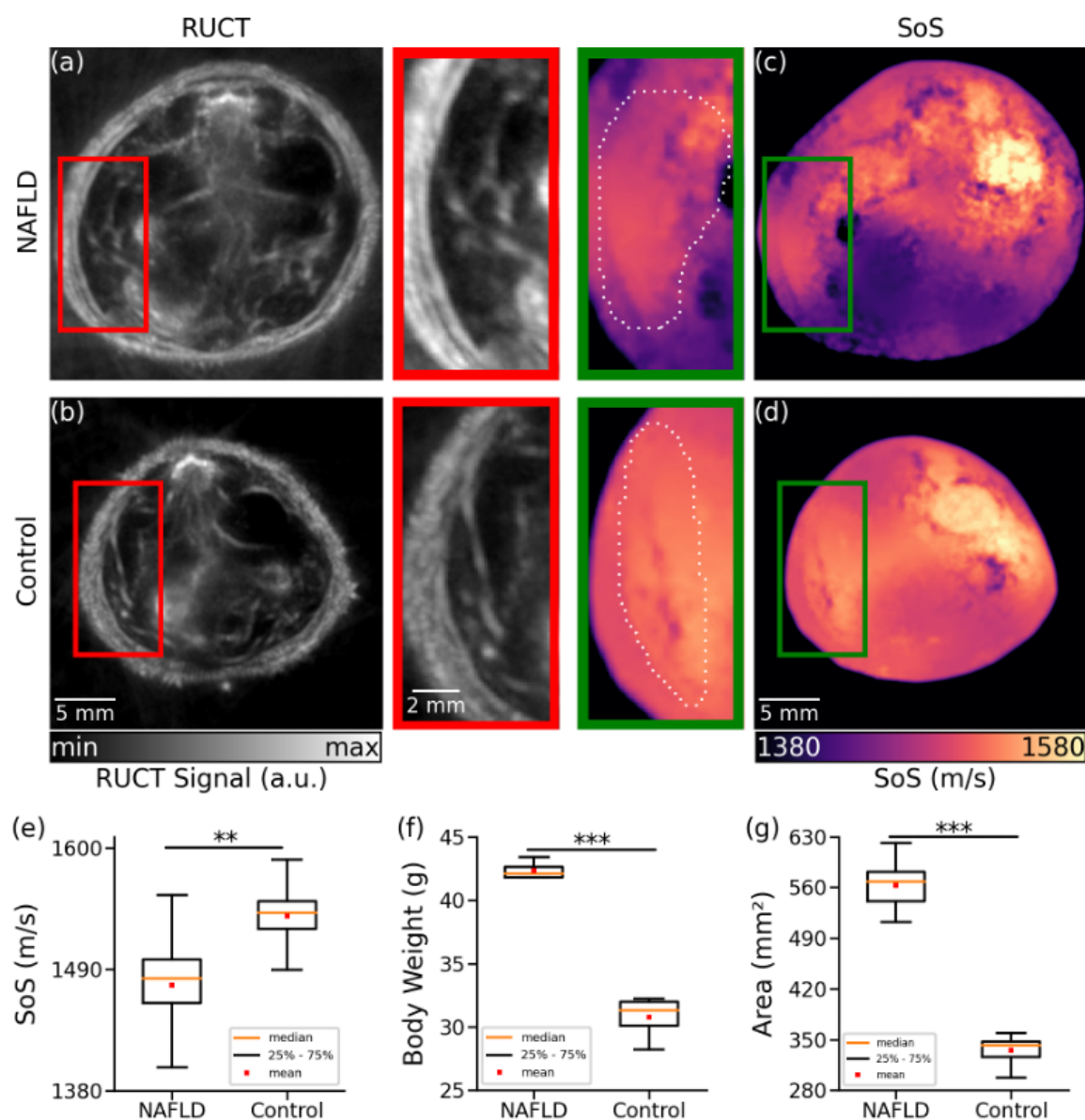


FIGURE 7.3: Cross-sectional RUCT and SoS images of NAFLD and control mice *in vivo*. (A) RUCT image of a NAFLD mouse cross section. Zoom-in of the liver region is shown. (B) RUCT image of a control mouse cross section. Zoom-in of the liver region is shown. (C) SoS image of a NAFLD mouse cross section. Zoom-in of the liver region is shown. (D) SoS image of a control mouse cross section. Zoom-in of the liver region is shown. (E) Boxplots of the measured SoS values in the segmented liver regions for NAFLD vs control mice cross sections. (F) Boxplots of the measured body weights for NAFLD and control mice. (G) Boxplots of the measured cross-sectional areas for NAFLD and control mice. NAFLD: non-alcoholic fatty liver disease; RUCT: reflection ultrasound computed tomography; SoS: speed of sound. (p values are indicated by  $*\leq 0.05$ ,  $**\leq 0.01$  and  $***\leq 0.001$ )

the lipid peak at 920 nm can only be detected in NAFLD mice (Fig. 7.4e). A statistically significant (19%) difference in lipid accumulation between NAFLD and control mice was found by calculating the averaged lipid signal values in the liver regions from all the measured cross sections (Fig. 7.4f,  $p=0.05$ ).

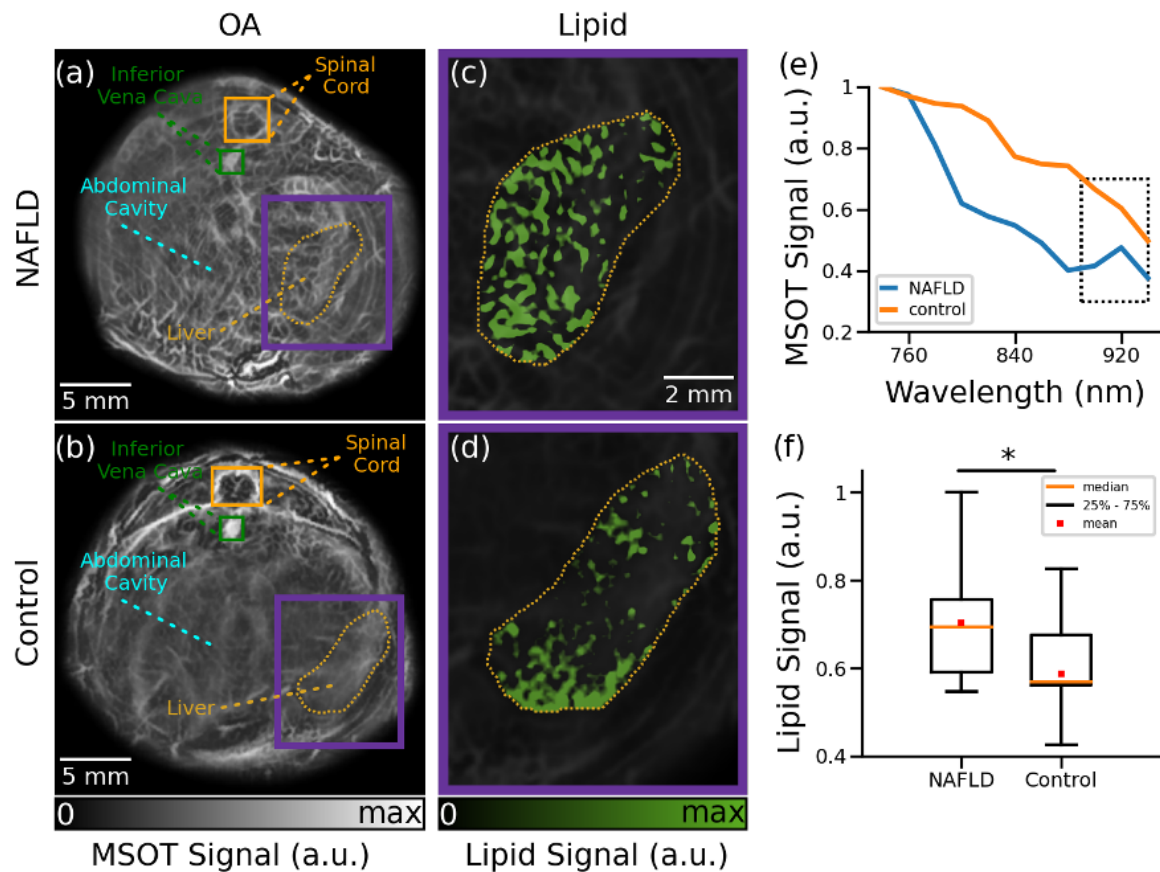


FIGURE 7.4: Cross-sectional MSOT images of NAFLD and control mice *in vivo*. (A) Cross section of a NAFLD mouse. (B) Cross section of a control mouse. (C) The un-mixed bio-distribution of lipids within selected liver region is shown for NAFLD mouse. (D) The un-mixed bio-distribution of lipids is shown for control mouse. (E) Spectrum of the MSOT signals in liver region indicated in panels (A-B). (F) Boxplots of the lipid signals in the liver cross sections for all mice. MSOT: multispectral optoacoustic tomography; NAFLD: non-alcoholic fatty liver disease. (p values are indicated by  $* \leq 0.05$ ,  $** \leq 0.01$  and  $*** \leq 0.001$ )

### 7.3 DISCUSSION

Early detection of NAFLD is essential for preventing progression of this condition to more advanced stages [31]. The process of fat accumulation in the liver is generally reversible before the onset of fibrosis by adjusting the daily diet and following a healthy lifestyle [32]. However, development of effective treatment strategies for NAFLD implies the *in vivo* validation of potential therapies in preclinical disease models. Histopathology imaging is conventionally used for this purpose [33], which however only allows measurements at single time points thus hampering longitudinal treatment follow-up studies. *In vivo* imaging modalities have thus been attempted for liver screening, predominantly MRI and pulse-echo US [34], which however suffer from low sensitivity and insufficient quantification accuracy. Multi-modal imaging with TROPUS represents a valuable alternative that can provide multi-parametric readings of the liver tissue condition. We have shown that the optical absorption peak of lipids at 920 nm facilitates quantification of fat accumulation in the MSOT images. The tomographic RUCT imaging was further shown to achieve improved resolution and contrast with respect

to standard pulse-echo US, thus enabling clear delineation of the outer boundary and internal structures in the cross-sectional images. The FWI reconstruction method further enabled the rendering of accurate SoS maps. Segmentation accuracy of the liver in the images benefits from the combination between SoS and RUCT images, which facilitated the observation of statistically significant differences in the measured parameters in NAFLD versus control liver tissues.

While TROPUS exploits the synergistic combination of three modalities for detecting and evaluating liver abnormalities, each modality is associated with certain limitations. For instance, MSOT imaging is affected by light attenuation in biological tissues, which effectively limits the achievable depth [35]. The so-called spectral-coloring effects may further hamper accurate quantification of chromophore distribution in deep tissues [36]. Model-based reconstruction algorithms can be used for rendering more quantitative results, although accurate modelling of all the factors affecting MSOT signals remains challenging [37–40]. On the other hand, due to the need for image compounding, RUCT and SoS imaging have an inferior imaging speed as compared to MSOT. While the achieved frame rates were generally sufficient for the experiments performed in this work, a high temporal resolution may be required to quantify dynamic biological processes, such as contrast-enhanced imaging for better assessment of NAFLD. The frame rate of all the TROPUS modalities can be increased by compressed sensing methods [26, 41] or optimized data acquisition strategies [25]. Deep learning-based methods could further enhance the performance of sparse acquisition strategies and provide more accurate segmentation of the liver boundaries [42, 43]. It is also important to take into account that three-dimensional (3D) imaging can only be achieved by vertically scanning the transducer array [24]. Alternatively, 3D imaging systems have been proposed for both MSOT and US imaging [44, 45].

Going forward, longitudinal studies starting from the early fat accumulation to fibrosis development, all the way to advanced pathological stages, such as liver cirrhosis or hepatocellular carcinoma, can also be performed with TROPUS, thus revealing new insights on the underlying mechanism of disease progression. Contrast agents may also be used for an enhanced TROPUS performance. For example, indocyanine green (ICG) is mainly cleared through the liver, and thus can be used to assess functional differences between healthy and NAFLD mice by comparing clearance time [22]. Different types of nanoparticles can also be used to boost imaging sensitivity and contrast [46–49]. Additional functional parameters such as blood flow can be extracted with contrast-enhanced MSOT imaging and Doppler US [50, 51].

In conclusion, we have demonstrated the capacity of the multi-modal TROPUS imaging for detecting and assessing NAFLD. The proposed approach facilitates quantification of liver abnormalities at early stages using a variety of optical and acoustic contrasts. It thus defines new *in vivo* imaging biomarkers to evaluate the efficacy of potential treatment strategies, providing a valuable alternative to conventional methods of assessing fat accumulation in the liver. All the three imaging modalities, namely, MSOT, pulse-echo, and transmission

US, have already been used in clinics [18, 52, 53] laying the groundwork for translating the TROPUS approach toward diagnosis and monitoring of NAFLD in humans.

## 7.4 MATERIALS AND METHODS

### 7.4.1 IMAGING SYSTEM

The TROPUS imaging setup contains four main components, namely a circular US transducer array, a nanosecond laser source, a DAQ and a workstation PC used for the system synchronization, data transfer, storage and processing [24]. The custom-engineered ring-shaped detector array (Imasonic Sas, Voray, France) consists of 512 individual elements operated in both transmit mode for US wave generation and in receive mode for the detection of OA, pulse-echo (reflection) and transmitted US signals (Fig. 7.1a). The array has 40 mm radius with the individual elements having 0.37 mm x 15 mm dimensions, interelement spacing of 0.1 mm, peak central frequency of 5 MHz and transmit/receive bandwidth of 60% at -6 dB. The array's active surface is shaped to provide cylindrical (toroidal) focusing in the imaged (two-dimensional (2D)) plane. During the experiments, the array was connected to an electronically controlled stage system with 4 degrees of freedom (x, y, z translations and azimuthal rotation) enabling accurate positioning of the imaged mouse at the center followed by volumetric scanning along the elevational (z) dimension. The mouse and the transducer array were placed in a temperature-controlled (36.5°C) water tank to ensure optimal physiological conditions and uninterrupted acoustic coupling. A tunable nanosecond OPO laser (SpitLight, InnoLas Laser GmbH, Krailling, Germany) was used for the OA excitation. The laser delivers ~20 mJ per pulse energy at 25 Hz repetition rate and optical wavelength between 680 and 1200 nm tunable on a per-pulse basis. The output beam was guided through an optical fiber separated into 12 output ferules with dimensions 0.21 mm x 12.65 mm to illuminate the object from different angles with uniform fluence (CeramOptec GmbH, Bonn, Germany) and optical energy density below safety limits [54]. The output ferules were equidistantly distributed on the top and bottom parts of the array with 24° tilt angle to optimize the uniformity of the illumination profile in the imaging plane. The MSOT and US signals collected by the array were digitized with a custom engineered DAQ (Falkenstein Mikrosysteme GmbH, Taufkirchen, Germany). The DAQ is connected to a workstation PC via 1 Gbit/s Ethernet to transfer the acquired signals. The workstation employed has 128 GB random access memory (RAM) and an NVIDIA GeForce GTX 1060 graphics processing unit (GPU) for real-time reconstruction of images. This PC was also used for synchronizing the delays between laser emission and US transmission, controlling the stages, and storing the acquired signals.

### 7.4.2 MULTISPECTRAL OPTOACOUSTIC TOMOGRAPHY (MSOT) IMAGING

Imaging in the MSOT mode was performed by quickly switching the optical wavelength of the nanosecond OPO laser from 740 to 940 nm with 20 nm step size at 25 Hz repetition rate. For each laser pulse, the OA signals recorded by all 512 elements were simultaneously sampled by the DAQ at 40 mega samples per second (MSPS) (2030 samples were acquired per

laser pulse from each element). The acquired signals were first bandpass filtered with cut-off frequencies 0.1 and 6 MHz. Then, MSOT images were reconstructed with a back-projection algorithm assigning different SoS values for the background (water) and the mouse body (Fig. 7.1c) [55]. 200 frames were averaged for each cross section to improve contrast to noise ratio (CNR). Frames affected by breathing motion were separated before averaging using an automatic detection algorithm based on the cross correlation between the frames [56]. The outer boundaries of each cross section were manually segmented in the reconstructed images using the combined information from RUCT and MSOT images. The segmented binary masks were also used to correct for light attenuation through the mouse using a simple modified Bessel function approximation [57]. Then, adaptive histogram equalization was applied on the MSOT images. After the histogram equalization, Frangi (vesselness) filter was used to detect vessels inside the mouse body [23]. As a final step, the segmented binary mask was applied for background suppression while combining reconstructed image and Frangi filtered image. The MSOT images acquired at the 11 wavelengths were used by a linear un-mixing algorithm [38] in order to separate Hb, HbO<sub>2</sub>, melanin and lipid components, the latter having a distinct peak in its absorption spectrum at 920 nm (Fig. 7.1b) [27].

#### 7.4.3 REFLECTION ULTRASOUND COMPUTED TOMOGRAPHY (RUCT) IMAGING

RUCT imaging was based on the STA image acquisition technique [25]. Data acquisition was performed by sequential transmission of a single-cycle bipolar square US wave (0.16  $\mu$ s, 38 V<sub>pp</sub>) with each array element. After each transmission event, all transducer elements were switched to receive the reflected and transmitted US waves. The acquisition scheme thus resulted in 512x512 time-resolved signals for the 512 transmission events. Cross-sectional images from single transmission events were reconstructed individually using delay and sum (DAS) algorithm [25]. It combines the information contained in 128 neighboring channels (90°) around the transmitting element to reconstruct a low-contrast RUCT image from each individual transmission event. Image compounding was subsequently performed by adding up the 512 low-contrast images, which resulted in a better image contrast owing to consolidation of the different views around the sample. The final (high contrast) RUCT images are presented on a logarithmic scale (Fig. 7.1d).

#### 7.4.4 SPEED OF SOUND (SOS) IMAGING

Data acquisition for SoS mapping was based on the same STA-based method described above. Quantitative images of the SoS distribution in m/s were reconstructed from the US waves that traversed the imaged object. Specifically, signals collected from 171 elements on the opposite side of each transmitting element were considered. A gradient-descent FWI algorithm was used to iteratively vary the estimation of the SoS values in the defined image grid to minimize the mean squared error (MSE) between the estimated waves and the actual measurements (Fig. 7.1e) [28]. In this work, 40 iterations were used in all cases.

FWI methods are able to improve resolution and contrast in the transmission US imaging mode when compared to the less precise bent-ray-based approach previously reported for

TROPUS reconstructions [23]. Nevertheless, FWI is more computationally complex, and typically requires large computational times, even when employing a GPU. Therefore, in this work, we used two approaches to speed-up the reconstruction time. On the one hand, the initial estimation of the SoS mapping was obtained from the time of flight (TOF) of each emitter-receiver pair. Reference waveforms for each emitter-receiver pair were obtained from acquisitions in water, i.e. no sample placed within the FOV, and the transmitted signals were decomposed as the sum of scaled and time-shifted versions of the reference waveforms. The TOF values were obtained as the minimum of the time shifts obtained from this decomposition. This approach is more robust and less sensitive to noise than the conventional TOF picker algorithms [24, 58, 59]. On the other hand, the estimation of the transmitted waves for a specific SoS mapping within the iterative algorithm was obtained by sampling the space between each emitter and receiver with multiple paths using parallel computations on a GPU, and convolving the reference waveform with the estimated TOF of each path [60]. This avoids the need of using slow acoustic solvers. With the proposed method, the total SoS reconstruction time was within 5 minutes per slice.

#### 7.4.5 EX VIVO LIVER IMAGING

*Ex vivo* imaging was performed to validate that the quantitative readings provided by TROPUS enable the differentiation between the diseased and normal liver. For this, *ex vivo* liver samples from 3 NAFLD and 3 control mice were imaged. The samples were embedded in a 20 mm cylindrical agarose phantom (1.3% w/v agarose powder). The same data acquisition protocol was executed as for the *in vivo* animal imaging experiments. Cross-sectional images from two different slices were acquired by vertically shifting the electronically controlled stage with a 1 mm step size.

Fat accumulation in liver tissues was further validated with tissue histological sections. Specifically, a Leica ASP300S tissue processor (Leica, Heerbrug, Switzerland) was used for paraffin embedding. Then, a microtome (Model: Microm HM 335 E, Thermo Scientific, Walldorf, Germany) was used to generate 3  $\mu\text{m}$  thick tissue samples. The samples were stained by H&E and examined by an experienced liver histopathologist using a Nikon Eclipse 80i microscope (Nikon AG, Egg, Switzerland). 10x and 20x magnification images were acquired using a brightfield microscope.

#### 7.4.6 ANIMAL EXPERIMENTS

Animal housing and experiments were performed in accordance with the Swiss animal welfare laws approved by the Committee on Animal Experimentation for the Canton de Vaud, Switzerland (VD 3401.c). Mice of C57BL/6 background were housed at the department of Biomedical Sciences, University of Lausanne, Switzerland for 24 weeks under a 12 h dark/-light cycle. The cages were ventilated and kept in a room with temperature and moisture controlled to 20-22°C and 50-60%, respectively. After the first 8 weeks, half of the mice continued having *ad libitum* access to normal chow (Granovit, Switzerland; 3242.PX.F12) and water, while the other half was given *ad libitum* access to high fat diet (Envigo, Harlan

Teklad, USA; Cat no TD.93075.PWD, Adjusted Calories Diet [55/fat]) with fructose and glucose included in the water (23.1 g/L d-fructose (Axonlab) + 18.9 g/L d-glucose (Axonlab)) for 16 weeks, to develop a diet-induced model of NAFLD [61]. Body weight was measured weekly with a digital balance.

4 mice with NAFLD and 4 control mice were imaged with TROPUS. Anaesthesia was induced with an initial dose of 4% isoflurane (Abbott, Cham, Switzerland) in an oxygen/air mixture (200:800 mL/min) and was maintained at 1.5% isoflurane supplied via a nose cone under normal air supply (oxygen/air 100:400 mL/min). The fur around the abdomen was shaved and depilated. The mice were vertically placed inside the water tank at the center of the ring array transducer with their head kept above the water surface by means of a custom-designed animal holder. Body temperature was maintained at 36.5°C by heating of the water using an electrical heater. For each mouse, 5 cross-sectional images from different sections corresponding to the liver region (40 images in total) were acquired by vertically shifting the US array with an electronically controlled stage.



## ACKNOWLEDGEMENT

This work was supported by Swiss Data Science Center grant C19-04.

## DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

N. C. Burton is an employee of iThera Medical.

## AUTHORS' CONTRIBUTION

Conception and design: B. Lafci, X. L. Deán-Ben, D. Razansky

Development of methodology: B. Lafci, X. L. Deán-Ben, D. Razansky

Acquisition of data: B. Lafci, A. Hadjihambi, C. Konstantinou, L. Pellerin, X. L. Deán-Ben

Analysis of data: B. Lafci, N. C. Burton, J. L. Herraiz, X. L. Deán-Ben

Writing, review, and revision of the manuscript: B. Lafci, A. Hadjihambi, J. L. Herraiz, L. Pellerin, N. C. Burton, X. L. Deán-Ben, D. Razansky

Study supervision: D. Razansky

## BIBLIOGRAPHY

- [1] Naga Chalasani et al. “The diagnosis and management of non-alcoholic fatty liver disease: Practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association”. In: *Hepatology* 55.6 (2012), pp. 2005–2023.
- [2] Mark Benedict and Xuchen Zhang. “Non-alcoholic fatty liver disease: An expanded review”. In: *World journal of hepatology* 9.16 (2017), p. 715.
- [3] Natascha Roehlen, Emilie Crouchet, and Thomas F Baumert. “Liver fibrosis: mechanistic concepts and therapeutic perspectives”. In: *Cells* 9.4 (2020), p. 875.
- [4] Jacopo Baglieri, David A Brenner, and Tatiana Kisseleva. “The role of fibrosis and liver-associated fibroblasts in the pathogenesis of hepatocellular carcinoma”. In: *International journal of molecular sciences* 20.7 (2019), p. 1723.
- [5] Alessandro Mantovani and Andrea Dalbeni. “Treatments for NAFLD: state of art”. In: *International journal of molecular sciences* 22.5 (2021), p. 2350.
- [6] Manuel Romero-Gómez, Shira Zelber-Sagi, and Michael Trenell. “Treatment of NAFLD with diet, physical activity and exercise”. In: *Journal of hepatology* 67.4 (2017), pp. 829–846.
- [7] Hiroki Nishikawa et al. “Sarcopenic obesity in liver cirrhosis: possible mechanism and clinical impact”. In: *International Journal of Molecular Sciences* 22.4 (2021), p. 1917.
- [8] Sumeet K Asrani et al. “Burden of liver diseases in the world”. In: *Journal of hepatology* 70.1 (2019), pp. 151–171.
- [9] Massimo Pinzani. “Pathophysiology of liver fibrosis”. In: *Digestive diseases* 33.4 (2015), pp. 492–497.
- [10] Alexander Boyd et al. “Medical liver biopsy: background, indications, procedure and histopathology”. In: *Frontline Gastroenterology* 11.1 (2020), pp. 40–47.
- [11] Scott B Reeder, Houchun H Hu, and Claude B Sirlin. “Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration”. In: *Journal of magnetic resonance imaging: JMRI* 36.5 (2012), p. 1011.
- [12] Carole Vitellius et al. “Liver fibrosis staging by computed tomography: prospective randomized multicentric evaluation of image analyses”. In: *Clinics and Research in Hepatology and Gastroenterology* 46.2 (2022), p. 101797.
- [13] Angelo Pietrobelli et al. “Dual-energy X-ray absorptiometry: fat estimation errors due to variation in soft tissue hydration”. In: *American Journal of Physiology-Endocrinology and Metabolism* 274.5 (1998), E808–E816.

- [14] ZiMian Wang et al. “Estimation of percentage body fat by dual-energy x-ray absorptiometry: evaluation by in vivo human elemental composition”. In: *Physics in Medicine & Biology* 55.9 (2010), p. 2619.
- [15] Arsenii V Telichko et al. “Noninvasive estimation of local speed of sound by pulse-echo ultrasound in a rat model of nonalcoholic fatty liver”. In: *Physics in Medicine & Biology* 67.1 (2022), p. 015007.
- [16] Juvenal Ormachea and Kevin J Parker. “A preliminary study of liver fat quantification using reported longitudinal ultrasound speed of sound and attenuation parameters”. In: *arXiv preprint arXiv:2109.04419* (2021).
- [17] Elena Merčep et al. “Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.9 (2015), pp. 1651–1661.
- [18] Cuiping Li et al. “In vivo Breast Sound-Speed Imaging with Ultrasound Tomography”. In: *Ultrasound in Medicine & Biology* 35.10 (2009), pp. 1615–1628.
- [19] XL Deán-Ben et al. “Advanced optoacoustic methods for multiscale imaging of in vivo dynamics”. In: *Chemical Society Reviews* 46.8 (2017), pp. 2158–2198.
- [20] Lihong V Wang and Song Hu. “Photoacoustic tomography: in vivo imaging from organelles to organs”. In: *science* 335.6075 (2012), pp. 1458–1462.
- [21] Vasilis Ntziachristos and Daniel Razansky. “Molecular imaging by means of multispectral optoacoustic tomography (MSOT)”. In: *Chemical reviews* 110.5 (2010), pp. 2783–2794.
- [22] Shan Huang et al. “Functional multispectral optoacoustic tomography imaging of hepatic steatosis development in mice”. In: *EMBO Molecular Medicine* 13.9 (2021), e13490.
- [23] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18.
- [24] Berkan Lafci et al. “Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound”. In: *Neoplasia* 22.12 (2020), pp. 770–777.
- [25] Berkan Lafci et al. “Expediting image acquisition in reflection ultrasound computed tomography”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 69.10 (2022), pp. 2837–2848.
- [26] Ali Özbek, Xosé Luís Deán-Ben, and Daniel Razansky. “Optoacoustic imaging at kilohertz volumetric frame rates”. In: *Optica* 5.7 (2018), pp. 857–863.
- [27] Robert H Wilson et al. “Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization”. In: *Journal of biomedical optics* 20.3 (2015), pp. 030901–030901.

- [28] M Pérez-Liva et al. “Time domain reconstruction of sound speed and attenuation in ultrasound computed tomography using full wave inversion”. In: *The Journal of the Acoustical Society of America* 141.3 (2017), pp. 1595–1604.
- [29] PA Hasgall et al. *Itis database for thermal and electromagnetic parameters of biological tissues, version 3.0*. 2015.
- [30] Christopher M Milroy. “Fatty liver and the forensic pathologist”. In: *Academic forensic pathology* 8.2 (2018), pp. 296–310.
- [31] Daryl Ramai et al. “Progressive liver fibrosis in non-alcoholic fatty liver disease”. In: *cells* 10.12 (2021), p. 3401.
- [32] Theresa J Hydes et al. “Evidence-based clinical advice for nutrition and dietary weight loss strategies for the management of NAFLD and NASH”. In: *Clinical and Molecular Hepatology* 26.4 (2020), p. 383.
- [33] Stefan G Hübscher. “Histological assessment of the liver”. In: *Medicine* 39.9 (2011), pp. 523–528.
- [34] Bernard E Van Beers, Jean-Luc Daire, and Philippe Garteiser. “New imaging techniques for liver diseases”. In: *Journal of hepatology* 62.3 (2015), pp. 690–700.
- [35] Arunima Sharma, Vijitha Periyasamy, and Manojit Pramanik. “Photoacoustic imaging depth comparison at 532-, 800-, and 1064-nm wavelengths: Monte Carlo simulation and experimental validation”. In: *Journal of Biomedical Optics* 24.12 (2019), pp. 121904–121904.
- [36] Hongzhi Zuo et al. “Spectral crosstalk in photoacoustic computed tomography”. In: *Photoacoustics* 26 (2022), p. 100356.
- [37] X Luís Dean-Ben et al. “Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography”. In: *IEEE transactions on medical imaging* 31.10 (2012), pp. 1922–1928.
- [38] Lu Ding et al. “Constrained inversion and spectral unmixing in multispectral optoacoustic tomography”. In: *IEEE transactions on medical imaging* 36.8 (2017), pp. 1676–1685.
- [39] Lu Ding, Daniel Razansky, and Xosé Luís Deán-Ben. “Model-Based Reconstruction of Large Three-Dimensional Optoacoustic Datasets”. In: *IEEE Transactions on Medical Imaging* 39.9 (2020), pp. 2931–2940.
- [40] Weiye Li et al. “Broadband model-based inversion enables optoacoustic microscopy beyond the acoustic diffraction limit”. In: *arXiv preprint arXiv:2009.10173* (2020).
- [41] Kathyayini Sivasubramanian and Manojit Pramanik. “High frame rate photoacoustic imaging using clinical ultrasound system”. In: *Photons Plus Ultrasound: Imaging and Sensing 2016*. Vol. 9708. SPIE. 2016, pp. 832–837.
- [42] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky. “Deep learning optoacoustic tomography with sparse data”. In: *Nature Machine Intelligence* 1.10 (2019), pp. 453–460.

- [43] Berkan Lafci et al. “Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 68.3 (2020), pp. 688–696.
- [44] Justine Robin et al. “Dual-Mode Volumetric Photoacoustic and Contrast Enhanced Ultrasound Imaging with Spherical Matrix Arrays”. In: *IEEE Transactions on Medical Imaging* (2021), pp. 1–1.
- [45] SA Ermilov et al. “Three-dimensional optoacoustic and laser-induced ultrasound tomography system for preclinical research in mice: design and phantom validation”. In: *Ultrasonic imaging* 38.1 (2016), pp. 77–95.
- [46] Donald A Fernandes and Michael C Kolios. “Intrinsically absorbing photoacoustic and ultrasound contrast agents for cancer therapy and imaging”. In: *Nanotechnology* 29.50 (2018), p. 505103.
- [47] Yu-Hsin Wang et al. “Photoacoustic/ultrasound dual-modality contrast agent and its application to thermotherapy”. In: *Journal of biomedical optics* 17.4 (2012), pp. 045001–045001.
- [48] X Luís Dean-Ben and Daniel Razansky. “Localization optoacoustic tomography”. In: *Light: Science & Applications* 7.4 (2018), pp. 18004–18004.
- [49] Claudia Errico et al. “Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging”. In: *Nature* 527.7579 (2015), pp. 499–502.
- [50] Xosé Luís Deán-Ben et al. “Noninvasive three-dimensional optoacoustic localization microangiography of deep tissues”. In: *arXiv preprint arXiv:2007.00372* (2020).
- [51] Ammar A Oglat et al. “A review of medical doppler ultrasonography of blood flow in general and especially in common carotid artery”. In: *Journal of medical ultrasound* 26.1 (2018), p. 3.
- [52] Lisanne Grünherz et al. “Preoperative mapping of lymphatic vessels by multispectral optoacoustic tomography”. In: *Lymphatic research and biology* 20.6 (2022), pp. 659–664.
- [53] Jørgen Arendt Jensen. “Medical ultrasound imaging.” In: *Progress in biophysics and molecular biology* 93 1-3 (2007), pp. 153–65.
- [54] Jiaqi Huang et al. “Empirical assessment of laser safety for photoacoustic-guided liver surgeries”. In: *Biomedical Optics Express* 12.3 (2021), pp. 1205–1216.
- [55] X Luís Deán-Ben, Vasilis Ntziachristos, and Daniel Razansky. “Effects of small variations of speed of sound in optoacoustic tomographic imaging”. In: *Medical physics* 41.7 (2014), p. 073301.
- [56] Thomas Felix Fehm et al. “In vivo whole-body optoacoustic scanner with real-time volumetric imaging capacity”. In: *Optica* 3.11 (2016), pp. 1153–1159.
- [57] Daniel Razansky and Vasilis Ntziachristos. “Hybrid photoacoustic fluorescence molecular tomography using finite-element-based inversion”. In: *Medical physics* 34.11 (2007), pp. 4293–4301.

- 
- [58] Cuiping Li et al. “An improved automatic time-of-flight picker for medical ultrasound tomography”. In: *Ultrasonics* 49.1 (2009), pp. 61–72.
- [59] Erol Kalkan. “An automatic P-phase arrival-time picker”. In: *Bulletin of the Seismological Society of America* 106.3 (2016), pp. 971–986.
- [60] Mailyn Perez-Liva et al. “Speed of sound ultrasound transmission tomography image reconstruction based on Bézier curves”. In: *Ultrasonics* 103 (2020), p. 106097.
- [61] James M Eng and Jennifer L Estall. “Diet-induced models of non-alcoholic fatty liver disease: food for thought on sugar, fat, and cholesterol”. In: *Cells* 10.7 (2021), p. 1805.

## 8 DISCUSSION AND FUTURE DIRECTIONS

This thesis focuses on the development of hybrid OA and US systems by proposing both hardware and software solutions. The constraints imposed by hybrid imaging modalities are investigated throughout the thesis. Several solutions including imaging setups, optimization of data acquisition schemes, and development of new image processing methods using conventional and state-of-the-art computer vision techniques are proposed to overcome these constraints. Each method comes with advantages and challenges. This section summarizes the findings and focuses on the improvement points for the proposed methods.

The project presented in “Deep Learning for Automatic Segmentation of Hybrid Optoacoustic Ultrasound (OPUS) Images” focuses on the development of automated segmentation methods for outer boundary of the mouse body from 2D cross-sectional images of OA and US. The study shows the successful implementation of deep learning-based semantic segmentation. The results are quantified by calculating the Dice metric between predictions of deep learning method and manual segmentation. The same metric is used to compare the proposed method with active contour. The proposed deep learning method outperforms the active contour method quantitatively and qualitatively. The study can be extended to 3D imaging modalities including the hybrid OA and US systems. The work does not involve the segmentation of inner structures such as kidney, liver, and spleen. However, the accurate light attenuation modeling requires information of all tissue structures in the cross-sections. The future works can include investigation of multiorgan segmentation and accurate light attenuation modeling methods.

US data acquisition sequences and possible array designs for tomographic imaging are proposed in “Expediting Image Acquisition in Reflection Ultrasound Computed Tomography”. Combination of consecutive transducer elements and spatial undersampling are investigated to increase frame rate in RUCT. The results are quantified by comparing CNR and spatial resolution for the same increase in frame rate (temporal resolution) for each method. The combination of multiple transducer elements for transmission and reception of US waves are implemented without changing the waveforms from each individual channel. Different waveforms (i.e., planar and diverging waves) can be generated by arranging the delays of each transducer elements separately. The investigation of alternative waveform can help to achieve higher frame rates or increase in contrast and/or resolution. Alternatively, the development of faster DAQ with higher data transfer rate can help to achieve higher frame rates.

In the work presented in “Signal Domain Learning Approach for Optoacoustic Image Reconstruction from Limited View Data” chapter, a method to reduce limited view artifacts

in signal domain are proposed. The results are validated by comparing the reconstructed OA images using several image processing performance metrics. In the presented project, the ground truth data is available in the test dataset. Thus, validation of the methods and calculation of the evaluation metrics are possible. However, the application of the proposed method on the images without ground truth data can result in the removal and addition of structures that are not present in the reality. The constraints on the networks should be adjusted to guarantee convergence of anatomically correct structures. The image reconstruction algorithms can be incorporated in the deep learning networks to learn and perform image generation directly in the inference time.

The dataset presented in “OADAT: Experimental and Synthetic Clinical Optoacoustic Data for Standardized Image Processing” chapter includes raw signals and reconstructed images with different sparse and angular coverage acquisitions from two different detector arrays. The benchmarks and training, validation, and test scripts for common deep learning algorithms are also included to give researchers a tool to explore and validate further image processing methods. In the presented data acquisitions, the orientation of the probe is perpendicular to the vessels in human forearm. However, the detector array can be used in different orientations in clinical settings. In this regard, the dataset and the benchmarks should be extended for different orientations. Different types of the detector arrays can be used to increase dataset variability, i.e., 3D probes. The extension of the dataset and benchmarks will be beneficial for the development and validation of new image processing methods.

The applications of TROPUS system for multimodal assessment of mammary tumor and liver disease progression are shown in chapters 6 and 7. While TROPUS imaging setup is suitable for achieving optimal imaging performance for small animal applications, it is not practical for human imaging where only one-sided access to the region of interest is available due to the large size of the human body. *In vivo* human imaging with OA imposes number of physical constraints and image reconstruction challenges related to the strong attenuation of the excitation light in living tissues, restricted tomographic access to the region of interest (limited-view problem) and acoustic heterogeneities. In addition, the full view tomographic coverage requires the use of extended number of detection elements to meet the strict inter-element pitch requirements. This results in excessive hardware complexity and implementation cost as well as excessive computational and memory requirements, hindering the real-time image rendering which is crucial in a bedside clinical imaging setting. In order to overcome these challenges, optimized detector array geometries with higher number of transducer elements and increased angular coverage are needed. In addition to considerations pertaining excessive instrumentation costs, it is not possible from a physical standpoint to achieve full tomographic coverage when imaging human tissues.

We believe that the methods proposed in this thesis will foster the applications of hybrid OA and US systems in preclinical and clinical settings.



## ABBREVIATIONS

<b>OA</b>	optoacoustic . . . . .	ix
<b>US</b>	ultrasound . . . . .	ix
<b>PA</b>	photoacoustic . . . . .	ix
<b>NIR</b>	near infrared . . . . .	ix
<b>HbO<sub>2</sub></b>	oxy-hemoglobin . . . . .	ix
<b>Hb</b>	deoxy-hemoglobin . . . . .	ix
<b>SoS</b>	speed of sound . . . . .	x
<b>OPUS</b>	optoacoustic ultrasound . . . . .	x
<b>TROPUS</b>	transmission-reflection optoacoustic ultrasound . . . . .	x
<b>NAFLD</b>	non-alcoholic fatty liver disease . . . . .	x
<b>DAQ</b>	data acquisition system . . . . .	xi
<b>CNN</b>	convolutional neural network . . . . .	xi
<b>RUCT</b>	reflection ultrasound computed tomography . . . . .	xii
<b>STA</b>	synthetic transmit aperture . . . . .	xii
<b>CNR</b>	contrast to noise ratio . . . . .	xii
<b>TUCT</b>	transmission ultrasound computed tomography . . . . .	xii
<b>MSOT</b>	multispectral optoacoustic tomography . . . . .	xiii
<b>MRI</b>	magnetic resonance imaging . . . . .	2
<b>PET</b>	positron emission tomography . . . . .	2
<b>CT</b>	computed tomography . . . . .	2
<b>FOV</b>	field of view . . . . .	3
<b>ICG</b>	indocyanine green . . . . .	4
<b>3D</b>	three-dimensional . . . . .	4
<b>2D</b>	two-dimensional . . . . .	5
<b>LED</b>	light emitting diode . . . . .	5
<b>GPU</b>	graphics processing unit . . . . .	7
<b>OPO</b>	optical parametric oscillator . . . . .	7
<b>MSPS</b>	mega samples per second . . . . .	9
<b>GUI</b>	graphical user interface . . . . .	10
<b>TOF</b>	time of flight . . . . .	15
<b>ACED</b>	active contour edge detection . . . . .	19
<b>DAS</b>	delay and sum . . . . .	24
<b>FWI</b>	full wave inversion . . . . .	28
<b>DL</b>	deep learning . . . . .	43

---

<b>ReLU</b>	rectified linear unit . . . . .	45
<b>SGD</b>	stochastic gradient descent . . . . .	45
<b>TP</b>	true positive . . . . .	48
<b>FP</b>	false positive . . . . .	48
<b>FN</b>	false positive . . . . .	48
<b>OAT</b>	optoacoustic tomography . . . . .	63
<b>HIFU</b>	high-intensity focused ultrasound . . . . .	63
<b>SNR</b>	signal-to-noise ratio . . . . .	67
<b>4D</b>	four-dimensional . . . . .	68
<b>FWHM</b>	full width at half maximum . . . . .	69
<b>1D</b>	one-dimensional . . . . .	75
<b>TV</b>	total variation . . . . .	77
<b>GT</b>	ground truth . . . . .	84
<b>MAE</b>	mean absolute error . . . . .	87
<b>CCL</b>	cycle consistency loss . . . . .	89
<b>FM</b>	feature matching loss . . . . .	89
<b>LDL</b>	latent discriminator loss . . . . .	89
<b>SDL</b>	style discriminator loss . . . . .	89
<b>GAN</b>	generative adversarial network . . . . .	88
<b>RC</b>	real center . . . . .	91
<b>SSIM</b>	structural similarity index . . . . .	91
<b>MSE</b>	mean squared error . . . . .	91
<b>Pearson</b>	Pearson correlation coefficient . . . . .	91
<b>PSNR</b>	peak signal to noise ratio . . . . .	91
<b>EN</b>	elastic net . . . . .	92
<b>BP</b>	backprojection . . . . .	92
<b>MSFD</b>	multispectral forearm dataset . . . . .	110
<b>SWFD</b>	single wavelength forearm dataset . . . . .	111
<b>SCD</b>	simulated cylinders dataset . . . . .	111
<b>ss</b>	sparse sampling . . . . .	112
<b>lv</b>	limited view . . . . .	112
<b>vc</b>	virtual circle . . . . .	112
<b>seg</b>	segmentation . . . . .	112
<b>sc</b>	semi circle . . . . .	113
<b>li</b>	linear . . . . .	113
<b>ms</b>	multisegment . . . . .	113
<b>BN</b>	batch normalization . . . . .	113
<b>RMSE</b>	root mean squared error . . . . .	114
<b>IoU</b>	intersection over union . . . . .	114

---

<b>HD95</b>	95-percentile Hausdorff distance . . . . .	114
<b>MP</b>	maxpooling . . . . .	123
<b>UP</b>	upsampling . . . . .	123
<b>CAT</b>	concatenation . . . . .	123
<b>AG</b>	attention gate . . . . .	123
<b>DOI</b>	digital object identifier . . . . .	138
<b>AA</b>	acoustic attenuation . . . . .	144
<b>RAM</b>	random access memory . . . . .	145
<b>ROI</b>	region of interest . . . . .	151
<b>H&amp;E</b>	Hematoxylin and Eosin . . . . .	163
<b>STD</b>	standard deviation . . . . .	165

## LIST OF FIGURES

1.1	Representative drawings for detector arrays. (a) Circle array with 348° tomographic coverage including 512 transducer elements. (b) Semi-circle array formed by one half of the circle array including 256 transducer elements. (c) Multisegment array designed as a combination of linear and concave parts including 128 transducer elements on the linear part and 64 transducer elements on the concave parts (256 elements in total). (d) Virtual circle array to mimic 360° tomographic coverage with 1024 transducer elements. . . . .	9
1.2	Example optoacoustic (OA) images from healthy mouse cross-sections using circle detector array with 1 mm step size in elevational direction. The images starting from the liver region (cross-section 1) to the kidney and spleen region (cross-section 10) are shown in the examples. Conventional image enhancement methods (see chapter 1.3.4 for details) are applied on all cross sections. . . . .	12
1.3	Example optoacoustic (OA) images from human forearms using semi-circle, multisegment, and linear (central part of multisegment) detector arrays. (a) Semi-circle array acquisition with 256 transducer elements and 174° angular coverage. (b) Multisegment detector array acquisition with 256 transducer elements and 170° angular coverage. (c) Linear (central part of multisegment) detector array acquisition with 128 transducer elements. . . . .	13
1.4	Simulated human forearm optoacoustic (OA) images using virtual circle, multisegment, and linear detector arrays. The circular high-intensity regions represent the vessels and parabolic line represents the skin line. (a) Simulation by virtual circle array with 360° angular coverage. (b) Simulation by multisegment array including linear and concave parts with 170° angular coverage. (c) Simulation by linear (central part of multisegment) detector array including 128 transducer elements. . . . .	14
1.5	Image enhancement steps for optoacoustic (OA) images. (a) Image reconstruction with heterogenous backprojection algorithm by assigning different speed of sound (SoS) values for the background (water) and the mouse body. (b) Fluence correction using modified Bessel function approximation to account for light attenuation. (c) Adaptive histogram equalization. (d) Frangi filter to emphasize vessels. (e) Application of manual segmentation mask and the combination of histogram equalized image with Frangi filtered image. . . . .	17

- 1.6 Removal of undersampling artifacts resulting from the sparsity of transducer elements with data-driven image enhancement algorithm. (a) Input images given to convolutional neural network (CNN) with sparse acquisition artifacts. (b) Output images (predictions) by CNN with removed sparse acquisition artifacts. (c) Target (ground truth) images acquired without sparsity in transducer elements. . . . . 19
- 1.7 Segmentation of mouse boundaries using manual annotation, active contour edge detection (ACED), and convolutional neural network (CNN) for hybrid optoacoustic ultrasound (OPUS) images. (a) Manual segmentation of mouse boundary in optoacoustic (OA) images. (b) ACED segmentation method for OA images. (c) CNN segmentation method for automated delineation of mouse boundaries in OA images. (d) Manual segmentation of mouse boundary in ultrasound (US) images. (e) ACED segmentation method for US images. (f) CNN segmentation method for automated delineation of mouse boundaries in US images. [80] . . . . . 20
- 1.8 Removal of limited view artifacts resulting from reduced angular coverage of detector array by data-driven image enhancement algorithm. (a) Input images are given to convolutional neural network (CNN) with limited view artifacts. (b) Output images (predictions) by CNN with removed limited view artifacts. (c) Target (ground truth) images acquired with increased angular coverage. . . . . 21
- 1.9 Low-resolution to high-resolution image enhancement using a convolutional neural network (CNN). (a) Input images given to CNN with low-resolution. (b) Output images (predictions) by CNN with increased resolution. (c) Target (ground truth) images with high spatial resolution. . . . . 22
- 1.10 Example images from the reflection ultrasound computed tomography (RUCT). (a) Cross- sectional image of healthy nude mouse lower abdominal. (b) Three 6 mm circler insertions of 50  $\mu\text{m}$  reflective microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ) in 20 mm agarose phantom. (c) Hex key-shaped insertion of 50  $\mu\text{m}$  reflective microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ) in 20 mm agarose phantom. (d) 20 mm diameter circular insertion of 50  $\mu\text{m}$  reflective microparticles (Cospheric BKPMS-1.2 45-53 $\mu\text{m}$ ). (e) Copper wires with 0.1 mm aligned perpendicular to the imaging plane. . . . . 25
- 1.11 The effects of consecutive element transmission on beam focusing and field of view (FOV). (a) Transmission with 1 consecutive channel which is equal to synthetic transmit aperture (STA) with single elements. (b) Transmission with 2 consecutive transducer elements simultaneously. (c) Transmission with 3 consecutive transducer elements simultaneously. (d) Transmission with 4 consecutive transducer elements simultaneously. . . . . 26

- 1.12 Combined effects of consecutive element transmission and reception were assessed *in vivo*. (a) *In vivo* mouse cross-section without consecutive element transmission or reception (individual element transmission and reception). (b) *In vivo* mouse cross-section with 2 consecutive element transmission and reception. (c) *In vivo* mouse cross-section with 3 consecutive element transmission and reception. (d) *In vivo* mouse cross-section with 4 consecutive element transmission and reception. . . . . 26
- 1.13 Combined effects of uniform sparsity in transmission and reception were assessed *in vivo*. (a) *In vivo* mouse cross-section without uniform sparsity in transmission or reception (all individual elements are used in transmission and reception) (b) *In vivo* mouse cross-section with uniform sparsity level of 2 in transmission and reception. (c) *In vivo* mouse cross-section with uniform sparsity level of 3 in transmission and reception. (d) *In vivo* mouse cross-section with uniform sparsity level of 4 in transmission and reception. . . . . 27
- 1.14 Combined effects of random sparsity in transmission and reception were assessed *in vivo*. (a) *In vivo* mouse cross-section without random sparsity in transmission or reception (all individual elements are used in transmission and reception) (b) *In vivo* mouse cross-section with random sparsity level of 2 in transmission and reception. (c) *In vivo* mouse cross-section with random sparsity level of 3 in transmission and reception. (d) *In vivo* mouse cross-section with random sparsity level of 4 in transmission and reception. . . . . 27
- 1.15 Speed of sound (SoS) maps of one mouse cross-section *in vivo* using different image reconstruction techniques. (a) Straight ray reconstruction method. (b) Bezier curves reconstruction method. (c) Full wave inversion (FWI) reconstruction method. . . . . 29
- 1.16 Optoacoustic ultrasound (OPUS) images of mouse cross-sections *in vivo*. (a) Optoacoustic (OA) images from different mouse cross-sections *in vivo*. (b) Corresponding reflection ultrasound computed tomography (RUCT) images from the same mouse cross-sections *in vivo*. . . . . 31
- 1.17 Simultaneous acquisition of one co-registered mouse cross-section with transmission-reflection optoacoustic ultrasound (TROPUS) system *in vivo*. (a) Optoacoustic (OA) image from a cross-section of the mouse *in vivo*. (b) Reflection ultrasound computed tomography (RUCT) image from the same cross-section of the mouse *in vivo*. (c) Full wave inversion (FWI) reconstruction of the speed of sound (SoS) image from the same cross-section of the mouse *in vivo*. . . . . 32

2.1	Key components and workflow of the conducted study. Images were acquired by multispectral optoacoustic tomography (MSOT) small animal scanners (Models inVision 256-TF and inVision 512-echo, iThera Medical GmbH, Munich, Germany) and manually segmented to create ground truth data. Deep learning (DL)-based segmentation method or active contour method are applied on the acquired images. The results are evaluated by comparing the output of the segmentation algorithms with manually segmented images using Dice coefficient. . . . .	44
2.2	The U-Net convolutional neural network (CNN) architecture used for image segmentation. The network uses 256x256 pixel image with one channel as an input, increases the channel size of the image and downscales the image in the encoder part (left). In the decoder part, the network up-samples and concatenates the image while the channel size is reduced (right). . . . .	46
2.3	Algorithmic steps of the active contour edge detection (ACED) segmentation approach. (a) The originally reconstructed optoacoustic (OA) image. (b) Output of Canny edge detector. (c) Image after morphological operations of dilation/erosion. (d) Fitted circle on pixels detected in the previous step. (e) The final result of the active contour segmentation. . . . .	47
2.4	Segmentation results for optoacoustic (OA) images obtained with manual segmentation (left), active contour edge detection (ACED) (middle) and convolutional neural network (CNN) (right) for cross-sections from the brain (a-c), liver (d-f), and kidney (g-i) regions. The corresponding Dice coefficients are also shown. . . . .	49
2.5	Segmentation results for optoacoustic (OA) images obtained with manual segmentation, active contour edge detection (ACED) and convolutional neural network (CNN) for mouse cross-sections through brain (a-c), liver (d-f), and kidney (g-i) regions that showcase sub-optimal (non-centered) animal positioning (kidney) and artifacts (liver, brain). The corresponding Dice coefficients are also shown. . . . .	50
2.6	Segmentation results for ultrasound (US) images obtained with manual segmentation, active contour edge detection (ACED) and convolutional neural network (CNN) (trained on optoacoustic (OA) images) for mouse cross-sections through brain (a-c), liver (d-f), and kidney (g-i) regions with the corresponding Dice coefficients. . . . .	51
2.7	Boxplots illustrating the Dice coefficient variability for the segmented optoacoustic (OA) and ultrasound (US) images obtained with active contour edge detection (ACED), convolutional neural network (CNN) trained with only OA images and CNN trained with combination of OA and US images. . . . .	52

- 3.1 Imaging setup and transmit/receive schemes. (a) Schematic diagram of the circular array (top view). The animal position is indicated. (b) Active array elements for transmission and/or reception. Different colors indicate independent signals. Top row – all array elements are driven independently. Second row – consecutive elements are driven simultaneously in transmission or grouped (combined) in reception. Third row – Uniform undersampling acquisition scheme with only a few equidistant elements activated in transmission and reception. Bottom row – Sparse undersampling with randomly chosen transmission and reception elements. . . . . 64
- 3.2 Effects of element size in transmission analyzed by varying the number of simultaneously transmitting consecutive elements ( $N_{CT}$ ). (a) Schematic representation of the transmission configuration with the positions of transmitting elements of the circular array shown in red. These elements ( $N_{CT} = 1, 2, 3, 4$ ) are driven simultaneously in one transmission event. Reflection ultrasound computed tomography (RUCT) images from a circular phantom with uniformly distributed microspheres show the effects of increasing the element size on the directivity of the transmitted ultrasound beam. (b) Contrast to noise ratio (CNR) of the RUCT images as a function of the number of consecutive transmitting elements ( $N_{CT}$ ). The signal and background regions for the CNR calculation are indicated in (a) with blue and red squares, respectively. (c) Averaged 20 horizontal profiles within the rectangular region indicated in green in (a) as a function of lateral distance for  $N_{CT} = 1, 2, 3, 4$ . . . . . 69
- 3.3 Effects of varying the number of consecutive reception elements ( $N_{CR}$ ). (a) Groups of reception channels ( $N_{CR} = 1, 2, 3, 4$ ) are indicated with different colors. Compounded (512 transmission events) reflection ultrasound computed tomography (RUCT) images of wires aligned perpendicular to the imaging plane. (b) Contrast to noise ratio (CNR) dependence on  $N_{CR}$ . (c) Horizontal image profile through the central wire as a function of  $N_{CR}$ . (d) Horizontal image profile through a peripheral wire as a function of  $N_{CR}$ . . . . . 70
- 3.4 Effects of uniform undersampling in reception. (a) Different acquisition schemes were evaluated with increasing level of sparsity, i.e.,  $N_{SR} = 1, 2, 3, 4$ . The corresponding compounded reflection ultrasound computed tomography (RUCT) images of wires aligned perpendicular to the imaging plane are shown. (b) Contrast to noise ratio (CNR) as a function of  $N_{SR}$ . (c) Horizontal image profile through the central wire as a function of  $N_{SR}$ . (d) Horizontal image profile through the peripheral wire as a function of  $N_{SR}$ . . . . . 71



- 3.5 Effects of random undersampling in reception. (a) Illustration of the various random undersampling schemes having  $N_{SR} = 1, 2, 3, 4$  with the corresponding compounded reflection ultrasound computed tomography (RUCT) images of wires aligned perpendicular to the imaging plane. (b) Contrast to noise ratio (CNR) as a function of  $N_{SR}$ . (c) Horizontal image profile through the central wire as a function of  $N_{SR}$ . d) Horizontal image profile through the peripheral wire as a function of  $N_{SR}$ . . . . . 72
- 3.6 Reflection ultrasound computed tomography (RUCT) phantom images acquired with consecutive elements and with sparse acquisitions as a function of the number of consecutive elements grouped in transmission/reception ( $N_{CTR}$ ) and the sparsity level ( $N_{STR}$ ). The test phantom consisted of a 20 mm diameter cylinder made of agar (1.3% w/v agar powder) embedded with three 6 mm diameter cylindrical insertions filled with a uniform distribution of 50 $\mu$ m-sized microparticles (Cospheric BKPMS-1.2 45-53 $\mu$ m). The increase in frame rate for a given data throughput capacity is shown in the bottom right corner of the images. (a) Compounded images as a function of the number of consecutive elements ( $N_{CTR} = 1, 2, 3, 4$ ). (b) Compounded images from uniform undersampling as a function of the sparsity level ( $N_{STR} = 1, 2, 3, 4$ ). (c) Compounded images from random undersampling as a function of the sparsity level ( $N_{STR} = 1, 2, 3, 4$ ). (d) Contrast to noise ratio (CNR) of the images as a function of  $N_{CTR}$  and  $N_{STR}$ . The signal and background regions for the CNR calculations are indicated in panel a with blue and red squares, respectively. (e) Standard deviation of image intensity across the FOV as a function of  $N_{CTR}$  and  $N_{STR}$ . . . . . 73
- 3.7 Comparison between *in vivo* reflection ultrasound computed tomography (RUCT) images acquired with the consecutive and undersampling schemes. A female athymic nude mouse (8 weeks old) was imaged *in vivo*. The achievable increase in frame rate for a given data throughput capacity of the data transmission-acquisition system (DAQ) is shown in upper left corner of the images. (a) RUCT images acquired with simultaneously driven consecutive elements ( $N_{CTR} = 1, 2, 3, 4$ ). (b) RUCT images acquired with uniform sparse acquisition as a function of sparsity levels ( $N_{STR} = 1, 2, 3, 4$ , respectively). (c) RUCT images acquired with random sparse acquisition as a function of sparsity levels ( $N_{STR} = 1, 2, 3, 4$ , respectively). (d) Change in contrast to noise ratio (CNR) calculated on the transverse colon using consecutive element and sparse acquisition methods. (e) Difference between minimum and maximum values of line profile of descending colon (P1) for the consecutive and undersampling acquisition schemes. (f) Difference between minimum and maximum values of line profile of vertebral column (P2) for the consecutive and undersampling acquisition schemes. Signal profiles through the descending colon (P1) and vertebral column (P2) are shown under the images in panels a-c. . . 74

- 4.1 Handheld multispectral optoacoustic tomography (MSOT) imaging with linear and multisegment array configurations. A) Schematic diagram of the array geometries. The excitation light beam and generated ultrasound (US) waves are represented with red arrows and dashed circles, respectively. B) Raw data (time-resolved signals) along with the reconstructed MSOT images corresponding to a hand-held scan of the human arm at 1064 nm excitation wavelength with the linear and multisegment arrays, respectively. C) Simulated signals along with the reconstructed images for the linear and multisegment arrays, respectively. . . . . 86
- 4.2 Summary of the proposed network architectures. A) Style network architecture with loss functions defined by red labels. Mean absolute error (MAE), cycle consistency loss (CCL), feature matching loss (FM), latent discriminator loss (LDL), style discriminator loss (SDL). B) Training of side network with simulated data. C) Training of side network using only experimental data from the linear part of the array. . . . . 89
- 4.3 Results in the signal and image domains. A) Signal domain representation of a test image. Left to right: multisegment ground truth (GT) signals, linear GT signals, multisegment signals after style and side networks, multisegment signals after style and side network with real center (RC). B) Images reconstructed with back-projection and model-based elastic-net approach. Benchmark UNet result is added for comparison. . . . . 91
- 4.4 Example of reconstructions on the test set (sample number 1-16). Each row corresponds to a different input signal. Each column corresponds to a different method (best viewed digitally). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts. . . . . 96
- 4.5 Example of reconstructions on the test set (sample number 17-32). Each row corresponds to a different input signal. Each column corresponds to a different method (best viewed digitally). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts . . . . . 97

5.1	Experimental data acquisition, transducer arrays and resulting images. (a) Experimental setup for optoacoustic (OA) forearm imaging. (b) Semi circle array along with an example of acquired images. (c) Multisegment array along with an example of acquired images. (d) Uniform subsampling for the semi circle array (128, 64 and 32 elements) and limited view acquisition for the semi circle array with reduced angular coverage (128 elements). (e) Uniform subsampling for the multisegment array (128, 64 and 32 elements), and linear array acquisition for the multisegment array (128 elements). Transducer elements are shown as actively receiving (red) or off (white). . . . .	107
5.2	Overview of the simulated data. (a) Virtual circle array and an image reconstructed using 1,024 transducer elements. (b) Multisegment array and the corresponding image reconstructed using combined linear and concave parts of the transducer array. (c) Uniform subsampling of virtual circle array with 128, 64 and 32 elements and limited view acquisition with reduced angular coverage (128 elements). (d) Uniform subsampling of multisegment array with 128, 64 and 32 elements and linear array acquisition (128 elements). (e) Vessel size distribution (pixels per vessel), number of vessels per image, and peak signal-to-noise ratio of full sampling compared to other reconstructions (x axis naming conventions are explained in Sec. 5.3.3). Transducer elements are shown as actively receiving (red) or off (white). . . . .	109
5.3	Residual convolutional block with batch normalization (BN). $ci-j$ conv.layer have $i \times i$ kernels and $j$ filters. . . . .	113
5.4	Distribution of modUNet structural similarity index (SSIM) performance on simulated cylinders dataset (SCD)(left), single wavelength forearm dataset (SWFD) (middle) and multispectral forearm dataset (MSFD) (right) image translation experiments, sorted in ascending median test sample performance. . . . .	114
5.5	Distribution of modUNet intersection over union (IoU) performance on simulated cylinders dataset (SCD) semantic segmentation experiments for vessel (left) and skin curve (right) labels, sorted in ascending median test sample performance. . . . .	116
5.6	Fitzpatrick skin phototype [1] distribution of volunteers in datasets. . . . .	122
5.7	Schematic of the proposed modUNet architecture. $CB_j$ represents the residual two-dimensional (2D) convolutional block with batch normalization shown in Fig. 5.3 (main manuscript) where each convolution has $j$ filters. Other abbreviations correspond to 2D-maxpooling (MP) of poolsize 2, 2D bilinear upsampling (UP) by a factor of 2, concatenation (CAT), and attention gates (AGs) [2]. Finally, $c1-j$ represents a convolutional layer of $j$ filters (1 for image translation and 3 for semantic segmentation experiments) and $1 \times 1$ kernels without activation. . . . .	123

- 5.8 We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions. . . . . 125
- 5.9 We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{lv128,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions. . . . . 125
- 5.10 We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{MSFD}_{lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions. . . . . 126
- 5.11 We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{ss32,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 126
- 5.12 We showcase the worst (1st column), 1st- (2nd column), 5th- (3rd column), and 10th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{ss32,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the nearly invisible vessels at input getting corrected at modUNet predictions. . . . . 127
- 5.13 We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiments  $f_{\text{SCD}_{ss32,vc}}$  (left) and  $f_{\text{SCD}_{lv128,li}}$  (right) with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 127
- 5.14 We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD}_{lv128,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 128

- 5.15 We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions. . . . . 128
- 5.16 We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{MSFD\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. Red arrows indicate some of the distorted vessel geometries at input getting corrected at modUNet predictions. . . . . 129
- 5.17 We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) structural similarity index (SSIM) samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,sc}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 129
- 5.18 We showcase 95th- (1st column), 98th- (2nd column), 99th-percentile (3rd column), and the best (4th column) SSIM samples based on modUNet predictions for experiment  $f_{\text{SWFD\_ss32,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 130
- 5.19 We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) vessel intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_lv128,li}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs. 131
- 5.20 We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) vessel intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_ss32,vc}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs. 131
- 5.21 We showcase the worst (1st column), 2nd worst (2nd column), 1st- (3rd column), and 5th-percentile (4th column) skin curve intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_ss32,vc}}$  with input (1st row), modUNet prediction (2nd row), and ground truth (3rd sample) pairs. 132
- 5.22 We showcase the worst (1st column), 2nd- (2nd column), 3rd- (3rd column), and 4th-worst (4th column) intersection over union (IoU) samples based on modUNet predictions for experiment  $f_{\text{seg\_MSFD\_ss64,ms}}$  with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 132
- 5.23 We showcase the worst (1st column), 2nd- (2nd column), 3rd- (3rd column), and 4th-worst (4th column) intersection over union (IoU) samples based on modUNet predictions for experiments  $f_{\text{seg\_SWFD\_ss64,sc}}$  (left) and  $f_{\text{seg\_SWFD\_ss128,ms}}$  (right) with input (1st row), modUNet prediction (2nd row), and target (3rd sample) pairs. . . . . 133
- 5.24 Pipeline figure to summarize data storage and algorithms. . . . . 134

- 6.1 The tri-modal transmission-reflection optoacoustic ultrasound (TROPUS) imaging platform. a) Excitation and acquisition steps in the optoacoustic (OA) imaging mode, reflection ultrasound computed tomography (RUCT) mode and transmission ultrasound computed tomography (TUCT) speed of sound (SoS) imaging mode. b) Illustration of three-dimensional (3D) stacks of cross-sectional multimodal images acquired non-invasively from tumor bearing mice. c) Representative TROPUS images of a cross-section of the tumor region in a mouse. From left to right, OA image, RUCT image, and SoS image acquired in the TUCT mode and reconstructed with the full wave inversion (FWI) method. 1 - Skin, 2 - Tumor, 3 - Urinary Bladder, 4 - Femur. . . . . 146
- 6.2 Comparison between different methods for rendering the speed of sound (SoS) images with transmission ultrasound computed tomography (TUCT). (a) Straight ray approximation. (b) Bezier curve reconstruction. (c) Full wave inversion (FWI). Reconstructions from different cross sections acquired from the same mouse with 1 mm steps are shown – see Fig. 6.1b. The SoS values were calculated over the tumor (green curve) and urinary bladder (cyan curve) areas manually segmented in the FWI images. . . . . 149
- 6.3 Cross-sectional multi-modal images of the tumor-bearing (M2-M5) and tumor free (M1) mice acquired from approximately the same abdominal region. Histological cryo-sections taken from approximately corresponding regions *ex vivo* are shown in the bottom row. The tumors and other anatomical structures are labeled: 1 - Vertebral Column, 2 - Caudal Vertebrae, 3 - Urinary Bladder, 4 - Tumor, 5 - Coxal Bone, 6 - Ischium, 7 - Female Urethra. . . . . 150
- 6.4 Tumor segmentation and characterization. (a) The manually segmented tumor areas in the speed of sound (SoS) images reconstructed with full wave inversion (FWI) method. (b) The corresponding segmentations based on the reflection ultrasound computed tomography (RUCT) images. (c) The segmented areas superimposed onto the optoacoustic (OA) images. d) Measured mean and standard deviation of SoS values inside the segmented region of interests (ROIs) based on SoS-based tumor segmentations. e) Comparison of the measured tumor areas based on the SoS- and RUCT-based tumor segmentations. The ratios between the calculated areas are presented for each imaged mouse. . . . . 152
- 6.5 Cross-sectional optoacoustic (OA) images without Frangi filter. The same OA images with Frangi filter are shown in Fig. 6.3. . . . . 155

- 7.1 TROPUS imaging. (A) Lay-out of the imaging set-up combining three modalities, namely MSOT, RUCT and SoS imaging. Only half of the ring array is shown for better visualization. (B) Absorption spectrum of Hb, HbO<sub>2</sub> and lipid in 700 nm and 1000 nm wavelength range. (C) Exemplary MSOT images from different cross sections recorded at 1064 nm excitation wavelength. (D) The corresponding cross sections reconstructed with the RUCT modality. (E) The corresponding cross sections showing the SoS maps reconstructed with TUCT modality. Hb: deoxy-hemoglobin; HbO<sub>2</sub>: oxy-hemoglobin, MSOT: multispectral optoacoustic tomography; RUCT: reflection ultrasound computed tomography; SoS: speed of sound; TROPUS: transmission-reflection optoacoustic ultrasound; TUCT: transmission ultrasound computed tomography. . . . . 164
- 7.2 TROPUS imaging of liver tissues excised from NAFLD and control mice. (A) Un-mixed lipid distribution (green color) overlaid onto the anatomical MSOT images corresponding to averaged signals over all the acquired wavelengths for excised livers from mouse with NAFLD and control mouse. (B) RUCT images of excised livers from mouse with NAFLD and control mouse. (C) SoS images of excised livers from mouse with NAFLD and control mouse. (D) Histology images of excised livers from mouse with NAFLD and control mouse. (E) MSOT signal spectra of liver tissue shown in the panels a. (F) Average lipid signal intensities from 3 NAFLD and 3 control mice. (G) Average SoS values from 3 NAFLD and 3 control mice. MSOT: multispectral optoacoustic tomography; NAFLD: non-alcoholic fatty liver disease; RUCT: reflection ultrasound computed tomography; SoS: speed of sound; (p values are indicated by \* $\leq 0.05$ , \*\* $\leq 0.01$  and \*\*\* $\leq 0.001$ ) . . . . . 166
- 7.3 Cross-sectional RUCT and SoS images of NAFLD and control mice *in vivo*. (A) RUCT image of a NAFLD mouse cross section. Zoom-in of the liver region is shown. (B) RUCT image of a control mouse cross section. Zoom-in of the liver region is shown. (C) SoS image of a NAFLD mouse cross section. Zoom-in of the liver region is shown. (D) SoS image of a control mouse cross section. Zoom-in of the liver region is shown. (E) Boxplots of the measured SoS values in the segmented liver regions for NAFLD vs control mice cross sections. (F) Boxplots of the measured body weights for NAFLD and control mice. (G) Boxplots of the measured cross-sectional areas for NAFLD and control mice. NAFLD: non-alcoholic fatty liver disease; RUCT: reflection ultrasound computed tomography; SoS: speed of sound. (p values are indicated by \* $\leq 0.05$ , \*\* $\leq 0.01$  and \*\*\* $\leq 0.001$ ) . . . . . 167

- 7.4 Cross-sectional MSOT images of NAFLD and control mice *in vivo*. (A) Cross section of a NAFLD mouse. (B) Cross section of a control mouse. (C) The un-mixed bio-distribution of lipids within selected liver region is shown for NAFLD mouse. (D) The un-mixed bio-distribution of lipids is shown for control mouse. (E) Spectrum of the MSOT signals in liver region indicated in panels (A-B). (F) Boxplots of the lipid signals in the liver cross sections for all mice. MSOT: multispectral optoacoustic tomography; NAFLD: non-alcoholic fatty liver disease. (p values are indicated by \* $\leq 0.05$ , \*\* $\leq 0.01$  and \*\*\* $\leq 0.001$ ) . . . 168



## LIST OF TABLES

2.1	Summary of the test data sets. . . . .	48
4.1	Reconstruction scores with respect to ground truth (GT) multisegment reconstructions for elastic net (EN) with $\alpha = 1e^{-5}$ and backprojection (BP) methods. Best score is in bold. . . . .	92
4.2	Reconstruction scores with respect to ground truth (GT) multisegment reconstruction using Elastic Net ( $\alpha = 10^{-5}$ ). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts. Name of our main proposed method is in bold. . . . .	98
4.3	Reconstruction scores with respect to ground truth (GT) multisegment reconstruction using backprojection (BP). “BM-1” - our sides network applied without prior style transfer network and trained with both synthetic data and linear parts of experimental data, “BM-2” - our sides network applied without prior style transfer network trained only on synthetic data, “BM-3” - supervised sides network which predicts from linear part array signal of concave parts. Name of our main proposed method is in bold. . . . .	98
5.1	Publicly available OA datasets, supported tasks, provided data format(s), size, and content. Davoudi et al. [1] contains 274 mice and 469 phantom slices. Huang et al. [2] has 10 mice with 10 frames (100 slices) at 27 different wavelengths and 20 phantom slices. . . . .	110
5.2	List of tasks and experiments we define on MSFD, SWFD, and SCD. Experiment names are built of (i) dataset name for translation tasks or seg for segmentation task, (ii) input data and corresponding number of active array elements; sparse sampling (ss), limited view (lv), virtual circle (vc), and (iii) input array type; semi circle (sc), virtual circle (vc), linear (li), multisegment (ms). Image data and annotation maps are represented with $x$ and $y$ , while predicted image and annotations are shown as $x^*$ and $y^*$ . . . . .	113
5.3	Image translation results of the proposed modUNet model reported as mean $\pm$ std. Each row corresponds to the results of the experiment where input data is identified through (i) the input data and corresponding number of active transducer elements; sparse sampling (ss), limited view (lv) and (ii) the array type used for input; semi circle (sc), virtual circle (vc), linear (li), multisegment (ms). . . . .	115

---

5.4	Segmentation results of our proposed modUNet model reported as mean $\pm$ std. Each row corresponds to the results of the experiment where input data is identified through (i) the input data and corresponding number of active transducer elements; sparse sampling (ss), limited view (lv), virtual circle (vc), semi circle (sc), multisegment (ms) and (ii) the array type used for input; virtual circle (vc), semi circle (sc), multisegment (ms), and linear (li). single wavelength forearm dataset (SWFD)- and multispectral forearm dataset (MSFD)-mini correspond to experiment conducted on OADAT-mini dataset. . . . .	117
5.5	Contents of the MSFD file. . . . .	135
5.6	Contents of SWFD semi circle file. . . . .	136
5.7	Contents of SWFD multisegment file. . . . .	136
5.8	Contents of SCD files. . . . .	136
5.9	Contents of OADAT-mini files. . . . .	137

# CURRICULUM VITAE

## EDUCATION

03/19 - 01/23      ETH Zurich, Zurich, Switzerland

*PhD. Information Technology and Electrical Engineering*

Thesis: Development of Hybrid Optoacoustic and Ultrasound Imaging Systems

Supervisor: Prof. Dr. Daniel Razansky, ETH and University of Zurich

10/16 - 12/18      Technical University of Munich, Munich, Germany

*M.Sc. Biomedical Computing*

Thesis: Temporal Prediction in Longitudinal Data Analysis

Supervisor: Prof. Dr. Ender Konukoglu, ETH Zurich and Prof. Dr. Björn Menze, University of Zurich

09/10 - 01/16      Bogazici University, Istanbul, Turkey

*B.Sc. Electrical and Electronics Engineering*

Thesis: Glucose sensors using split ring resonators and conductive polymer structures

Supervisor: Prof. Dr. Hamdi Torun, Northumbria University

03/14 - 09/14      Technical University of Munich, Munich, Germany

*Exchange Semester, Electrical and Computer Engineering*

## EXPERIENCE

02/23 - *pres.*      University of Zurich, Zurich, Switzerland

*Postdoctoral Researcher*

- Implemented 3D scanner hardware for small objects and multi-view reconstruction algorithms
- Developed automated segmentation and growth modelling pipelines

03/19 - 01/23      ETH Zurich, Zurich, Switzerland

*Researcher and Teaching Assistant*

- Developed image reconstruction algorithms for multimodality optoacoustic and ultrasound imaging
- Taught in “Molecular Imaging - Basic Principles and Biomedical Applications” course

11/16 - 01/19      iThera Medical GmbH, Munich, Germany

*R&D Engineer (Part Time)*

- Implemented deep learning-based optoacoustic image segmentation methods
- Developed artifact removal techniques by artificial intelligence algorithms

---

10/15 - 03/16 Siemens Healthineers, Innovation Lab, Istanbul, Turkey

*Sales Intern (Part Time)*

- Analyzed clinical workflow at multiple hospitals to improve Siemens products and services
- Performed comparative analysis of healthcare equipment features

07/15 - 09/15 Eindhoven University of Technology, Eindhoven, Netherlands

*Visiting Researcher (Summer Intern)*

- Worked on trajectory planning algorithms for drones
- Designed Simulink (circuit design program) structures for control mechanisms of drones

#### AWARDS AND SCHOLARSHIPS

'20 - '22 Swiss Data Science Center (SDSC) Project Grant

Funded for development of deep learning algorithms in medical imaging

'16 - '19 DAAD Master's Study Scholarship

Chosen by German Academic Exchange Service (DAAD) for 2 years of master's studies

'10 - '15 Republic of Turkey & Bogazici University High Merit Scholarship

Awarded for top 100 students in University Entrance Exam (OSS)

#### SKILLS

Programming: Python, Tensorflow, PyTorch, Matlab, C++, AWS, Cloud Computing, Git, HTML, Latex

Language: English (Fluent), German (Intermediate), Turkish (Native Language)

#### PERSONAL INTERESTS

Basketball, Mountain Biking, Running, Snowboarding