ETH zürich

ETH Library

Deep learning optoacoustic tomography with sparse data

Journal Article

Author(s):
Davoudi, Neda; Dean-Ben, Xose Luis; Razansky, Daniel

Publication date:
2019-10

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000395989

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Nature Machine Intelligence 1(10), https://doi.org/10.1038/s42256-019-0095-3

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

nature machine intelligence

ARTICLES

https://doi.org/10.1038/s42256-019-0095-3

Deep learning optoacoustic tomography with sparse data

Neda Davoudi[1,2], Xosé Luís Deán-Ben[1,3] and Daniel Razansky[1,3]*

The rapidly evolving field of optoacoustic (photoacoustic) imaging and tomography is driven by a constant need for better imaging performance in terms of resolution, speed, sensitivity, depth and contrast. In practice, data acquisition strategies commonly involve sub-optimal sampling of the tomographic data, resulting in inevitable performance trade-offs and diminished image quality. We propose a new framework for efficient recovery of image quality from sparse optoacoustic data based on a deep convolutional neural network and demonstrate its performance with whole body mouse imaging in vivo. To generate accurate high-resolution reference images for optimal training, a full-view tomographic scanner capable of attaining superior cross-sectional image quality from living mice was devised. When provided with images reconstructed from substantially undersampled data or limited-view scans, the trained network was capable of enhancing the visibility of arbitrarily oriented structures and restoring the expected image quality. Notably, the network also eliminated some reconstruction artefacts present in reference images rendered from densely sampled data. No comparable gains were achieved when the training was performed with synthetic or phantom data, underlining the importance of training with high-quality in vivo images acquired by full-view scanners. The new method can benefit numerous optoacoustic imaging applications by mitigating common image artefacts, enhancing anatomical contrast and image quantification capacities, accelerating data acquisition and image reconstruction approaches, while also facilitating the development of practical and affordable imaging systems. The suggested approach operates solely on image-domain data and thus can be seamlessly applied to artefactual images reconstructed with other modalities.

Optoacoustic (OA) imaging and tomography has seen substantial growth across numerous preclinical research disciplines[1–3], while recently gaining maturity as a powerful new clinical diagnostic tool[4–7]. The growing interest in this hybrid imaging modality is chiefly driven by its unique label-free contrast, the myriad of possibilities for contrast enhancement and its high spatial resolution at centimetre-scale penetration into living tissues[8–10]. Recent OA technology development efforts have focused on increasing the temporal resolution to further complement its unique capabilities by enabling high-frame-rate imaging in two and three dimensions[11,12]. Given that entire tissue volumes can be simultaneously excited with nanosecond pulses of light, the imaging speed in OA tomography is ultimately limited by the unidirectional propagation of ultrasonic waves in biological tissues. Kilohertz volumetric imaging rates greatly exceeding the performance of other bio-imaging modalities have recently been demonstrated[13], enabling new insights into otherwise invisible biological phenomena[10].

In practice, the actual image acquisition speed is further conditioned by the data throughput and storage capacity of the imaging system. As a result, superior temporal resolution is commonly achieved by undersampling or by integrating the tomographic image data in spatial or spectral domains[14–17], but this results in limited-view reconstruction artefacts, image blurring or diminished spectral sensitivity[15,18]. Acoustic heterogeneities may further hinder the achievable resolution and overall image quality if not properly accounted for with dedicated reconstruction algorithms[19]. To this end, iterative image reconstruction algorithms based on total variation (TV) regularization have been proposed for rendering OA images from limited or ill-posed data[16]. Although such approaches were capable of mitigating streak-type artefacts in the reconstructed images, the imposed sparsity conditions unavoidably resulted in a loss of information or the appearance of new features, thus inhibiting correct image interpretation and quantification.

Machine and deep learning methods have led to a paradigm shift in the reconstruction and processing of biomedical images[20–23]. In the field of OA imaging, those methods have been applied for the identification of cancer lesions by k-means clustering and feature learning of texture patches[24], as well as support vector machine classification of local features extracted with a scale-invariant feature transform[25]. Convolutional neural networks (CNNs) based on the U-Net architecture have been proposed for the removal of streak-type artefacts and noise from OA images[26–28], reconstruction of the initial OA pressure[29] or extraction of oxygen saturation maps and chromophore distribution from multi-spectral OA images afflicted by wavelength-dependent light fluence attenuation[30]. Chromophore quantification has been attempted via fluence estimation on a per-voxel basis with a random forest regressor trained with voxel-based context images (CIs)[31]. CNNs based on the AlexNet and VGG-16 architectures have been used to localize point-like OA targets and remove reflection artefacts from the images[32,33], and the K-SVD dictionary learning algorithm has been used to correct for image artefacts generated by multiple reflections in mouse skull[34]. Dedicated deep neural network-based algorithms have also been suggested for the processing of OA signals. For example, a fully connected five-layer network has been used for signal bandwidth enhancement[35]. Furthermore, an iterative inversion algorithm based on a relatively small CNN architecture has recently been developed to learn the effect of the regularization term in each reconstruction step[28].

Previous attempts at OA image quality enhancement have primarily been based on methodologies seeking to recover images

[1]Institute for Biomedical Engineering and Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland. [2]Department of Informatics, Technical University of Munich, Munich, Germany. [3]Faculty of Medicine and Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland. *e-mail: daniel.razansky@uzh.ch

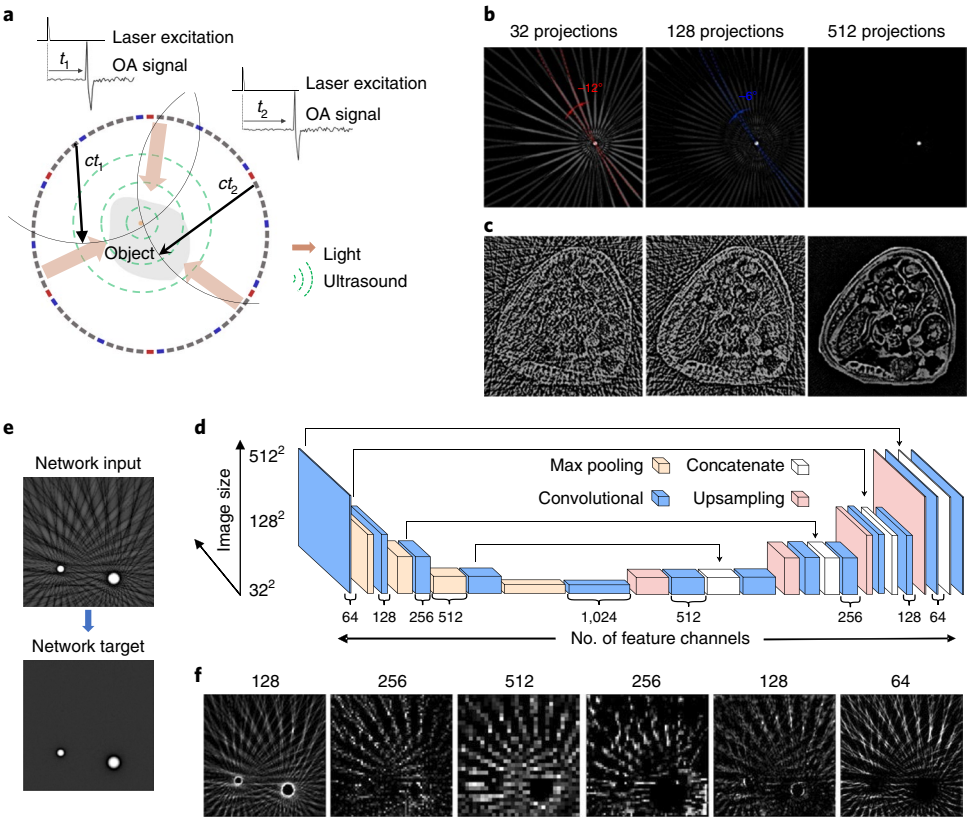


Fig. 1 | The deep convolutional neural network approach based on the U-Net architecture for artefact removal in OA imaging with undersampled data. **a**, Illustration of the optoacoustic tomographic image construction principle. **b**, Numerical simulation of a tomographic reconstruction of a single point absorber with different numbers of detection positions (projections) using a 2D filtered back-projection algorithm. **c**, Corresponding reconstruction of a more complex numerical phantom. **d**, The U-Net network architecture, consisting of contracting (downsampling) and expansive (upsampling) paths. **e**, Example pair of artefactual (network input) and artefact-free (network output) images, rendered after training with similarly simulated datasets. **f**, The corresponding intermediate outputs of the network for the example shown in **e**.

from sub-optimally acquired data by training networks on numerical phantoms or images from other modalities. Here, we propose a conceptually different approach that underlines the critical importance of training with high-quality OA images from real in vivo experiments for efficient recovery of image quality from undersampled or limited-view data.

The basic principle of OA tomographic image formation is depicted schematically in Fig. 1a. In short, tiny ultrasonic vibrations are induced by the absorption of pulsed laser radiation by tissue biochromes. The pressure (ultrasound) profiles are subsequently collected at multiple positions around the imaged object. The actual locations of the absorbers (OA sources) are encoded by the ultrasound time of flight; that is, the signal detected at time t_i has been generated on an arc-shaped surface of radius ct_i (where c is the speed of sound) centred at the detector position. Combining the time-resolved information from multiple detectors enables trilaterating the OA sources, thus building an image. This forms the basis of the common algorithms used for OA image reconstruction, including back-projection (radon-transform-based), iterative model-based as well as time-reversal approaches[36].

Results

Images reconstructed from undersampled data are commonly afflicted by so-called streak-type artefacts, which can shadow real anatomical structures. As shown in simulated OA reconstructions rendered from a single absorber (see Fig. 1b and Methods for details of the image reconstruction algorithm), arc-shaped structures separated by the angle subtended by adjacent detector positions are clearly visible when reconstructing the images using OA waveforms recorded from a relatively low number (32) of ultrasound sensors. The arc intensity diminishes when increasing the number of OA signals (tomographic projections) and eventually disappears (averages out) when operating with a dense distribution of 512 sensors. Streak-type artefacts are also generated in real samples containing a more heterogeneous distribution of absorbing structures. Even though individual arcs are more difficult to discern in this case, the streak artefacts similarly affect the visibility of structures when using a small number of sensors, as demonstrated by the simulated reconstructions in Fig. 1c.

The U-Net deep learning network[37] (Fig. 1d) was trained to recognize streak-type artefacts in OA images. The training was based on minimizing the cost function defined as the average L1-norm of the difference between the resulting predicted images of the network, I_p, and the reference artefact-free images, I_{GT}, which provided a ground truth (see Methods for details). Because learning of the residual is simpler for the network, the proposed model was trained on artefactual inputs with the corresponding target being an image of the streak artefacts, that is, the difference between the under- and fully-sampled reconstructions. The final artefact-free image was obtained by subtracting the learned residual from the artefactual image, as demonstrated in the numerical example in Fig. 1e, which contains two absorbers with different sizes. Note that Fig. 1e represents an illustrative example of the performance of the method, where training and testing were performed with simulated data.

When observing representative images rendered by the different layers of the network, one may notice that the superficial layers,

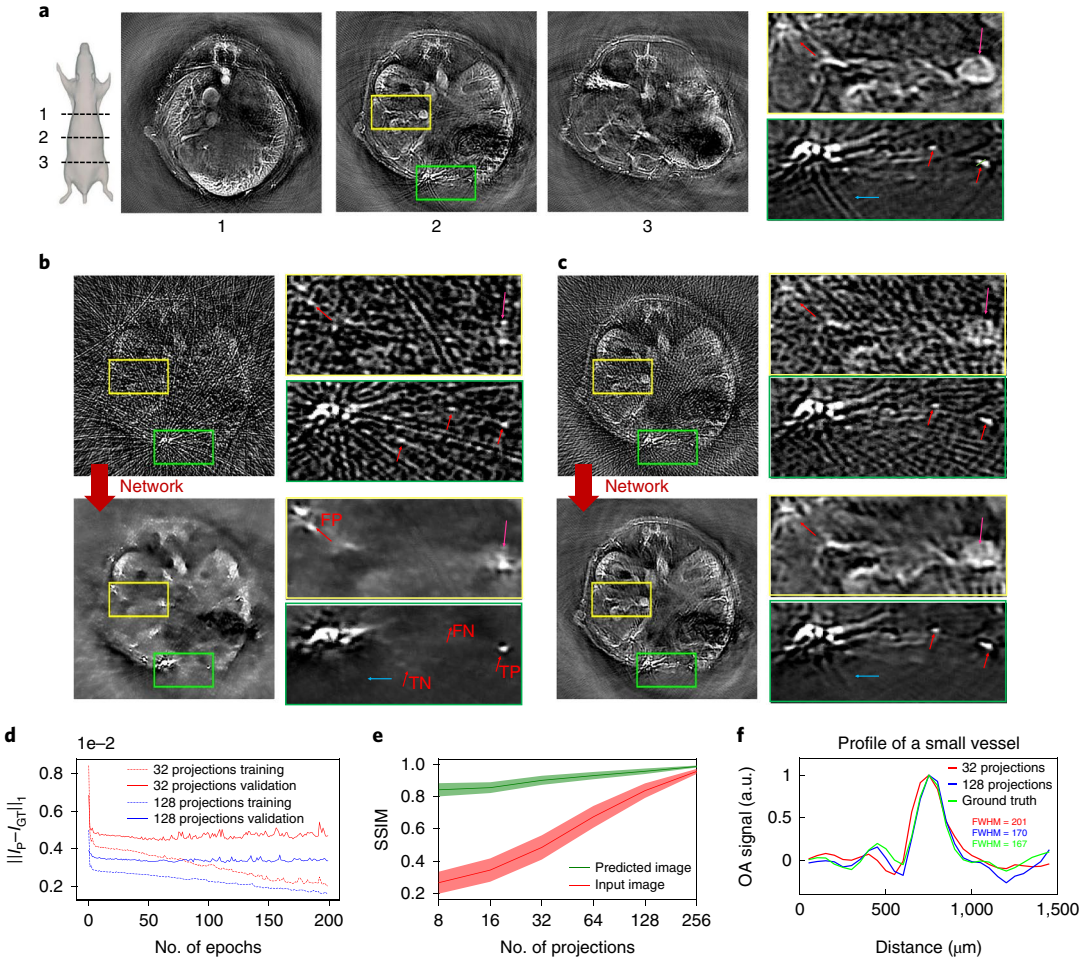


Fig. 2 | Restoring optoacoustic image quality in whole-body mouse tomography in vivo using a full-view ring-array transducer. a, Representative cross-sectional images from the abdomen region reconstructed using all 512 tomographic projections (detection elements), considered as the artefact-free targets (or ground truth). Enlarged views are shown on the right for the second image. **b**, Artefactual reconstructed image with undersampled (32 projections) data versus its artefact-free counterpart obtained with the trained network. TP, true positive; FP, false positive; TN, true negative; FN, false negative. **c**, The corresponding example for a reconstruction with 128 projections. **d**, Evolution of the L1-norm error estimate with the number of epochs throughout training and validation. l_p, predicted image; l_{GT}, ground truth image. **e**, Structural similarity index matrix (SSIM) of the ground truth image versus artefactual (red) and predicted artefact-free (green) images. **f**, Comparison among one-dimensional profiles extracted from images reconstructed with different numbers of tomographic projections (512, 128 and 32) in an area containing a small blood vessel. The calculated values of the full-width at half-maximum (FWHM) are indicated.

representing high-resolution information, feature arcs originating from both absorbers, while deeper layers mainly contain arcs stemming from the larger absorber (Fig. 1f). This is generally expected, as large structures and their associated artefacts extend over several scales in space representations. Similar observations were made when using a deep learning method for the removal of reflection artefacts from point-like targets[32,33]. As shown in numerical examples of images rendered from sparsely distributed absorbers, the multi-scale architecture of the proposed U-Net appears to efficiently detect and remove streak-type artefacts from OA images reconstructed from sixfold undersampled data (Supplementary Figs. 1 and 2).

The suggested approach was subsequently tested experimentally by tomographic OA imaging of tissue phantoms and mice in vivo. We used a dedicated whole-body small animal imaging scanner based on cross-sectional tomographic geometry, with detection performed by a ring-shaped transducer consisting of 512 cylindrically focused individual elements (see Methods and Supplementary Fig. 3 for details). The full-view geometry and dense distribution of the detection elements ensured efficient sampling of the tomographic data around the entire imaged object, leading to superior image quality when employing all 512 detected OA signals in the reconstructions (Fig. 2a). The image reconstruction algorithm also accounted for the speed of sound difference between the imaged object and the surrounding coupling medium (water) in order to achieve the best possible spatial resolution throughout the reconstructed cross-sections. Experimental validation with phantom data showed that the proposed deep learning algorithm previously trained on synthetic (numerical) data is capable of largely restoring the expected (reference) image quality, even when reducing the number of tomographic projections from 512 to as few as 16 (Supplementary Fig. 4).

We next acquired whole-body images from living mice with the newly developed small animal imaging scanner. Representative images from the abdomen region are shown in Fig. 2a, which confirm the unrivalled spatial resolution, penetration depth and overall OA image quality achieved. Datasets from $n = 6$ mice, each containing 100 different cross-sections, were then used for (re)training

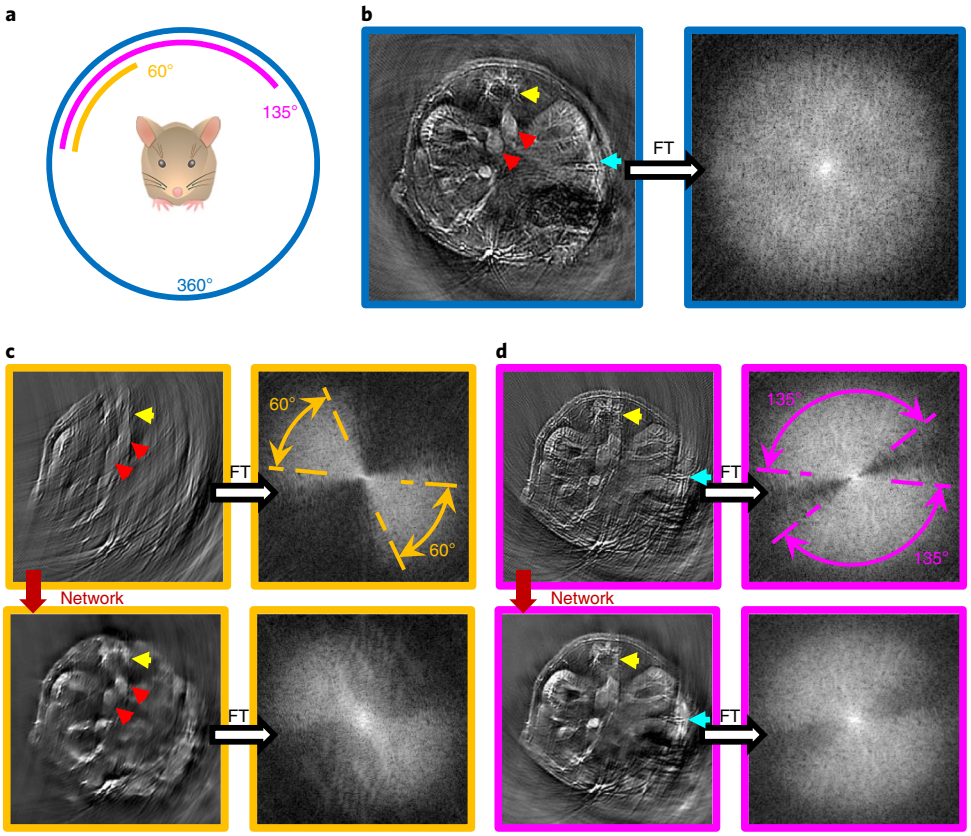


Fig. 3 | Network-based correction of limited-view effects induced by insufficient angular coverage of the optoacoustic imaging system. a, Schematic of the three considered imaging geometries of full 360° tomographic coverage (black circle) as well as representative sub-apertures of 60° (orange arc) and 135° (magenta arc). **b**, Image reconstructed with full angular coverage (taken as the ground truth for training the network) along with its amplitude spectrum plotted on a logarithmic scale. FT, Fourier transform. **c**, Top: image reconstructed with 60° angular coverage and its respective amplitude spectrum. Bottom: output image of the network and its corresponding amplitude spectrum. **d**, Top: image reconstructed with 135° angular coverage and its respective amplitude spectrum. Bottom: output image of the network and its corresponding amplitude spectrum.

(70% of the data), validating (20% of the data) and testing (10% of the data) the network. The ground truth images for training were taken as the reconstructed images with all 512 transducer elements. This high-quality experimental dataset effectively enabled proper learning of the effects of sparse acquisition in cross-sectional images of mice. Results of the artefact removal from images of the kidney region reconstructed with 32 and 128 projections are presented in Fig. 2b,c, respectively. Although the quality of the images predicted by the 128-channel data is comparable to the ground truth images, the network output naturally renders an inferior result for the 32-channel data. Nevertheless, the mouse outline and other internal structures, which can barely be revealed in the originally reconstructed 32-projection image due to the strong streak-type artefacts, have significantly better visibility after the network correction. Quantitatively, the algorithm is capable of differentiating the features corresponding to small blood vessels from the streak-type artefacts manifested as a speckle pattern of bright and dark spots (labelled by red arrows in the enlarged views in Fig. 2b,c as true positive (TP) and true negative (TN)). Some misclassified areas still exist in the recovered images—labelled by red arrows as false positive (FP) and false negative (FN) in the enlarged views in Fig. 2b,c. By manually comparing small vessels in 42 test slices, we estimated that 106 TPs, 4 FPs and 83 FNs were rendered. On the other hand, larger vessels are not correctly recovered when only using 32 projections and can only be visualized properly in the network output for 128 projections, where streak artefacts are less prominent (pink arrows in the enlarged view in Fig. 2b,c). This is consistent with the multi-scale nature of the artefacts generated by large-sized vessels, which seemingly require many layers of the multi-scale network for their accurate representation. Small vessels were also more faithfully identified with the network trained with images reconstructed with 128 projections. Specifically, we estimated 188 TPs, 0 FPs and 4 FNs in the test slices. Notably, the network was also capable of eliminating some of the minor artefacts present in the reference images obtained with fully sampled data. For reconstructions made from densely sampled data, the streak-type artefacts may still be present in the vicinity of strong absorbers[38]. For example, the blue arrow in the green-colour enlarged view in Fig. 2a shows the artefacts produced by a strongly absorbing superficial vessel, which are subsequently removed by the network (green-framed enlarged view in Fig. 2c). The results obtained for the L1- versus L2-norm cost functions as well as for a different number of array elements (projections) are presented in Supplementary Fig. 5 and are further compared with those obtained by interpolating the signals between adjacent projections. Additionally, two videos showing the performance of the method during a breathing cycle and for all slices covering the imaged region of the mouse are available in Supplementary Videos 1–4. Image reconstruction and processing with the trained network takes ~100 ms, comparable to the previously reported processing times for similar methods[30,39].

Note that the model was constantly saved throughout the iterative procedure. We applied an early stopping approach for training when the validation error was diverging for a certain consecutive number of epochs and we chose the best model mostly at around

100 epochs of training before the prediction became unstable, as determined by the corresponding L1-norm error behaviour (Fig. 2d and Supplementary Fig. 6). Hyper parameters and the loss function of the network were optimized during the validation phase of the network with regard to different evaluation metrics, including the L1- and L2-norm, peak signal-to-noise ratio (PSNR) and the SSIM (which combines local image structure, luminance and contrast into a single local quality score) (see Methods and Supplementary Fig. 7). Figure 2e presents the SSIM for the reference image (reconstructed with all 512 elements) and the input and output images of the network reconstructed with different numbers of tomographic projections. The test data (10% of the total number of images) were used to calculate these metrics. This figure confirms the considerable image quality enhancement achieved with the deep learning method, with SSIM values exceeding 0.9 even for a relatively low number of projections. The L2-norm performs slightly better (as a loss function) with a smaller number of projections, while a higher number of projections results in an approximately equivalent performance for the two norms (Supplementary Fig. 7). It appears that very shallow networks are too simple to achieve the desired performance, while very deep networks fail to recover the expected output, arguably due to overfitting (Supplementary Fig. 8). Importantly, images predicted by the network exhibit similar spatial resolution values as compared with the reference (ground truth) images, as evinced by observing the one-dimensional image profiles in a region containing a small vessel (Fig. 2f).

It should be emphasized that the high-quality images shown in Fig. 2 and Supplementary Videos 1–4 could only be attained when training the network with actual in vivo mouse scans. No comparable gains were achieved when the training was performed with synthetic or phantom data of different complexity, unless also performing fine tuning with real mouse data—see the qualitative and quantitative (SSIM, PSNR and contrast to noise ratio (CNR)) comparison of the results of the different strategies shown in Supplementary Fig. 9. When attempting different training strategies (Supplementary Fig. 10), the network performance is consistently enhanced whenever training data better resembles the complex mouse anatomy. In particular, when the network was trained with simulations or phantom experiments it yielded sub-optimal results and was only able to arrive at reasonable mouse image quality after being fine-tuned with sufficient experimental data. This further highlights the importance of training with high-quality experimental data.

Finally, the acquired whole-body images of mice were used to evaluate the performance of the same network architecture to enhance the visibility of images reconstructed under limited-view conditions[40]. Much like in other tomographic imaging modalities, accurate images can only be rendered when signals are collected from a surface enclosing the entire sample[41]. This becomes particularly challenging when imaging large objects like the human body, where tomographic access is limited. We thus evaluated the network's ability to correct for the so-called limited-view effects often produced by imaging systems with limited angular coverage (Fig. 3). The image reconstruction was performed by considering 85, 192 and 512 array elements covering 60°, 135° and 360° around a transverse section of the mouse (Fig. 3a). The image reconstructed by considering all the array elements (Fig. 3b) was again taken as the ground truth for training the network. Note that the amplitude spectrum of this image (Fig. 3b) is uniformly distributed across all angles in 2D Fourier space, as expected when full angular coverage is provided[40]. However, the Fourier domain information is lacking at angles not covered by the measured locations (Fig. 3c,d, top), which is manifested as a lack of visibility of boundaries, blood vessels and other elongated structures in the spatial image domain. Specifically, in spite of the dense spatial sampling of signals, vertical structures, such as the spinal cord (indicated by a yellow arrow) and the spleen (indicated by a blue arrow), are barely visible, even with an angular coverage of 135°. Concurrently, major blood vessels (indicated by red arrows) are elongated and distorted in the image rendered with an arc array covering only 60°, which is typical for linear handheld array transducers. These and other distortions are effectively corrected by the network (Fig. 3c,d, bottom), with the visibility of structures clearly enhanced and most information in the Fourier domain recovered, particularly at relatively low frequencies.

Discussion

Recent progress in OA system design and data-processing approaches has enabled imaging with unprecedented frame rates that are ultimately limited by the unidirectional propagation of ultrasound waves in biological tissues. Sparse data acquisition schemes may further boost the achievable frame rates[13], while reducing the number of simultaneously acquired signals can also reduce the complexity and associated costs of OA tomographic imaging systems. Yet, inevitable trade-offs exist between spatial resolution, imaging speed, penetration depth and detection sensitivity, commonly resulting in sub-optimal sampling of tomographic data and diminished image quality. The framework we suggest here is a fast and efficient approach for the acquisition and reconstruction of high-quality OA data. Our approach is conceptually different from the previously suggested machine and deep learning methods that are based on network training on numerical phantoms or images from other modalities. Instead, we have devised a dedicated full-view tomographic system that can provide the most relevant training and ground truth data for improving image quality in various OA image acquisition scenarios. The excellent image quality of the reference images used to train the network was achieved by (1) ensuring sufficient angular coverage to avoid limited-view effects[42], (2) providing a sufficient number of sensing elements to prevent the appearance of streak artefacts[43] and (3) having sufficiently small element size to avoid lateral blurring at points away from the centre[44]. Accounting for the heterogeneous speed of sound distribution in the reconstruction procedure further enhanced the resolution and contrast of the reference images[43,45], while real-time image rendering with simultaneously acquired projection data prevented other artefacts due to motion-related blurring[46]. The high quality of the images used for training is arguably the main reason behind the outstanding network performance with heavily undersampled data, as clearly demonstrated by the quantitative metrics for the different training strategies. The method could significantly enhance structural visibility, even with strong limited-view imaging scenarios and high sparsity of the acquired data. Significant image quality deterioration was only observed with a very low number of detection channels (<16) and/or angular coverage below 60°.

The proposed algorithmic and experimental framework can potentially be used for the efficient recovery of image data in a variety of OA imaging applications and system configurations. Given the divergence of results obtained with different training datasets, it is expected that the network needs to be retrained for optimal performance for a given imaging paradigm. The suggested methodology can further facilitate the development of more efficient data acquisition approaches, thus making technology more affordable without significantly impairing the imaging performance. Alternatively, or additionally, the approach can be employed to boost imaging frame rates by reducing the effective amount of data required for reconstructing each image frame. It was shown that accurate images could still be recovered in the presence of motion, which anticipates the general applicability of the method for visualizing rapid biological dynamics. This may benefit diverse applications, such as monitoring neural activity[2,12], cardiac function[3] or the free behaviour of model organisms[13,47]. Other deep learning networks could also be attempted to facilitate the efficient processing of multi-scale OA images. Of note is that the incorporation of neural

networks within iterative (model-based) reconstruction algorithms has been shown previously to outperform U-Net-based image post-processing in terms of robustness to perturbations in simulated data[28]. Combining more sophisticated network architectures with such model-based approaches can potentially greatly enhance image reconstruction from sparse data if properly trained with high-quality experimental OA images. Given the high performance of the network under limited-view conditions, the herein described image correction pipeline can readily be applied to other OA imaging configurations, for example for boosting image quality rendered with linear-array-based hand-held clinical imaging probes[29,48] or concave probe geometries with higher tomographic coverage[49]. Because the learning approach solely operates with image-domain data, it can be adapted in a manner similar to other imaging modalities.

Methods

Network training. U-Net[37] is a neural network architecture that has performed well for medical image segmentation and reconstruction, in particular when trained with a limited amount of data. The network consists of two main components (Fig. 1d). The encoder part is responsible for analysing and capturing the features from the input data. The dimensionally symmetric decoding component enables precise localization for generating results relying on the learned and forwarded features extracted from the encoder part. Both parts are composed of consecutive convolutional layers. In the encoder part, after each of the two convolution layers the output is passed through a pooling layer that halves the image size. As the data propagate through different stages along the compression path, the resolution is curtailed. The image contraction is reversed in a symmetric fashion by the upsampling layers in the decoder section, which restore the images to their size in the corresponding encoder layers. This procedure is repeated for the output of each convolutional layer until the output of the layers in the expanding path (decoder) reaches the same dimension as the first layer of the network. As outlined in Fig. 1d, concatenated copies are established between layers of equal resolution in the encoder and decoder to compensate for the loss in spatial resolution that results from the multiple downsampling operations. This allows the network to use the original image features in addition to the features appearing after upsampling. Therefore, data are not only recovered from a compressed latent representation, but all the relevant information coming from previous layers in the encoder is further generalized to prevent any loss of fine details when pooled in the encoder.

We fed pairs of training datasets $\{x_i, y_i\}$ to the network to render the model function f that maps from x to y, that is, $y=f(x)$, where x_i and y_i represent artefactual and artefact-free images, respectively, and y is an estimate of the target image by our model. The different types of dataset used for training and the different tested training strategies are depicted in Supplementary Figs. 9 and 10. Specifically, training was performed with simulated data consisting of spherical absorbers of different sizes, experimentally acquired data from paper-printed phantoms consisting of several absorbing dots or vascular networks, as well as real cross-sectional images of mice acquired by the full ring scanner. The network estimates the parameters in each iteration to minimize the predefined loss functions that are selected to be the L1-norm ($||x||_1 = \sum_i |x|_i$) and the L2-norm $\left(||x||_2 = \sqrt{\sum_i x_i^2}\right)$ of the difference between the predicted I_p and ground truth I_{GT} images. Weight matrices are trained by back-propagating the derivative of the loss with respect to the parameters throughout the network and updating the parameters via stochastic gradient descent. During training we used the Adam optimization algorithm[50] to minimize the loss. The initial value of the learning rate was 0.0005 and had a time-based decay with a momentum of 0.8. The mini-batch size equalled to 1. We considered the reconstruction with 512 projections as an artefact-free ground truth and reconstruction with sub-sampled data with 8, 16, 32, 64, 128 and 256 projections employed as artefactual inputs. Alternatively, we also considered sub-sampled data with 85 and 192 (out of 512) projections covering angles of 60° and 135°. Time spent for training each epoch was 50 s, which was considered an offline process. Testing and recovering each image lasted ~0.1 s, which sped up image reconstruction extremely with conventional iterative methods using signals from all of the 512 available projections. The models were implemented in Python using the Theano framework[51] and Lasagne library[52] running on a PC with a GTX 1080Ti GPU.

Cross-sectional OA tomography system. To generate accurate high-resolution reference images for the optimal network training, we devised a full-view tomographic scanner capable of attaining superior cross-sectional image quality from living mice (Supplementary Fig. 3). The scanner consisted of a 80 mm diameter ring detection array with 512 individual detection elements (5 MHz central frequency, >80% detection bandwidth, 0.37 mm width and 15 mm height along the elevational direction)[53]. Each element was cylindrically focused at a distance of 38 mm to selectively capture signals from the imaged cross-section. Manufacturing of the array was subcontracted to transducer array (Imasonics SaS). OA signal excitation was performed with a short-pulsed (<10 ns) Nd:YAG laser (Spectra Physics) with a repetition rate of 15 Hz and 1,064 nm optical wavelength. The laser beam was guided with a custom-made fibre bundle (CeramOptec) having 12 output ferrules (arms) fixed to both sides (six on each side) of the full-ring transducer with an angular separation of 60° between the arms. The total per-pulse laser energy at the output of the bundle was ~100 mJ, distributed over an ~6 cm^2 ring-shaped area around the mouse. Acoustic coupling between the imaged objects and the array was guaranteed by immersing the samples in water at a controlled temperature of 34 °C. The signals recorded from all of the 512 array elements were simultaneously digitized at 40 megasamples per second using a custom-made data acquisition system (Falkenstein Mikrosysteme). The acquisition was triggered with the Q-switch output of the laser and the data were transmitted to a PC via Ethernet for further processing and image reconstruction.

Image reconstruction. Reconstruction of 2D (cross-sectional) OA images was performed with a universal back-projection algorithm[54]. Before reconstruction, the signals were bandpass-filtered between 0.5 and 8 MHz to reduce noise and cancel out low-frequency offsets. The reconstruction algorithm was further adapted to account for a heterogeneous distribution of the speed of sound in the mouse versus the surrounding coupling medium[45]. For this, an initial image was rendered by considering a uniform speed of sound corresponding to the speed of sound in water for the measured temperature. The outline of the animal was then manually segmented and the reconstruction was redone by assigning a different speed of sound within the segmented area to optimize image quality.

Animal imaging experiments. A total of six athymic nude-Fox1^{nu} mice (Harlan Laboratories) were imaged in vivo. The mice were placed in a specifically designed holder attached to the water tank (Supplementary Fig. 3). The fore and hind paws were attached to the holder to fix the mice in a stationary position during the experiment. Anaesthesia was induced in a chamber containing 3% vol/vol isoflurane (in air) before quickly transferring and fixing the animals in the specifically designed holder. Once in place, anaesthesia was maintained using isoflurane (2% in air) via a coupling mask attached to the holder. A mouth clamp ensured coupling of the mask to the mouse. The water tank was kept at a constant temperature of 34 °C during data acquisition. For each mouse, the ring array was vertically scanned over 50 mm in 0.5 mm steps, thus acquiring 100 cross-sectional images covering the full torso from the shoulders down to the lower abdomen. All experiments involving animals were conducted in full compliance with the institutional guidelines of the Helmholtz Center Munich and with approval from the Government of Upper Bavaria.

Phantom imaging experiments. Phantom training data were generated by imaging paper-printed absorbing targets embedded in a 16 mm diameter agar cylinder. The agar matrix was prepared by diluting 1.3% (wt/wt) agar powder in water. A total of 28 phantoms were measured, from which we collected 450 images through data augmentation with random rotation and translation. For each phantom, the printed paper contained three randomly distributed 2.5 mm diameter black circles. Thirty-three phantoms consisting of printed vascular structures adapted from images acquired with an OA microscopy set-up[55] were also measured, from which we collected 100 images through data augmentation with random rotation. The speed of sound was assumed to be uniform for the reconstruction in all cases.

Numerical simulations. Two numerical simulations were performed to illustrate the effects of streak-type artefacts in the images (Fig. 1b). In both cases, OA signals arriving at a set of locations distributed along a 360° arc of 40 mm were calculated, thus simulating the experimental ring array transducer geometry. In the first simulation, a numerical phantom consisting of a planar parabolic absorber with a diameter of 300 mm was considered, for which the OA signals could be calculated analytically[56]. In the second case, a cross-sectional image of a mouse taken with a cryoslicer ex vivo was taken as the basis for simulating the absorbing target. The OA signals emanating from the numerical phantom were estimated with a discrete 2D forward model, as described elsewhere[57]. Reconstructions were performed with the back-projection formula assuming a uniform speed of sound. An additional simulation was performed to train the network and test its performance (Fig. 1c). For this, a sequence of OA signals from two randomly distributed parabolic absorbers were calculated. Images reconstructed by using 512 projections were taken as a reference (gold standard), whereas images reconstructed with a fraction of sensors were considered as artefactual inputs of the network. The network performance was further tested with a sequence of randomly distributed uniform volumetric absorbers having different sizes (Supplementary Figs. 1 and 2), for which the OA signals could also be calculated analytically[58]. In this case, the theoretical images were considered as a reference while the reconstructed images with a fraction of sensors were taken as the artefactual inputs.

Data availability

The datasets used for the current study were generated and analysed in our laboratory and are downloadable at https://doi.org/10.6084/m9.figshare.9250784.

The code to reproduce the results of the paper is available at https://github.com/ndavoudi/sparse_artefact_unet.

Received: 19 November 2018; Accepted: 9 August 2019;
Published online: 16 September 2019

References

1. Jathoul, A. P. et al. Deep in vivo photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter. *Nat. Photon.* **9**, 239–246 (2015).
2. Yao, J. et al. High-speed label-free functional photoacoustic microscopy of mouse brain in action. *Nat. Methods* **12**, 407–410 (2015).
3. Lin, H.-C. A. et al. Characterization of cardiac dynamics in an acute myocardial infarction model by four-dimensional optoacoustic and magnetic resonance imaging. *Theranostics* **7**, 4470–4479 (2017).
4. Stoffels, I. et al. Metastatic status of sentinel lymph nodes in melanoma determined noninvasively with multispectral optoacoustic imaging. *Sci. Transl. Med.* **7**, 317ra199 (2015).
5. Diot, G. et al. Multi-spectral optoacoustic tomography (MSOT) of human breast cancer. *Clin. Cancer Res.* **23**, 6912–6922 (2017).
6. Knieling, F. et al. Multispectral optoacoustic tomography for assessment of Crohn's disease activity. *N. Engl. J. Med.* **376**, 1292–1294 (2017).
7. Lin, L. et al. Single-breath-hold photoacoustic computed tomography of the breast. *Nat. Commun.* **9**, 2352 (2018).
8. Wang, L. V. & Yao, J. A practical guide to photoacoustic tomography in the life sciences. *Nat. Methods* **13**, 627–638 (2016).
9. Weber, J., Beard, P. C. & Bohndiek, S. E. Contrast agents for molecular photoacoustic imaging. *Nat. Methods* **13**, 639–650 (2016).
10. Deán-Ben, X., Gottschalk, S., Mc Larney, B., Shoham, S. & Razansky, D. Advanced optoacoustic methods for multiscale imaging of in vivo dynamics. *Chem. Soc. Rev.* **46**, 2158–2198 (2017).
11. Taruttis, A. & Ntziachristos, V. Advances in real-time multispectral optoacoustic imaging and its applications. *Nat. Photon.* **9**, 219–227 (2015).
12. Deán-Ben, X. L. et al. Functional optoacoustic neuro-tomography for scalable whole-brain monitoring of calcium indicators. *Light Sci. Appl.* **5**, e16201 (2016).
13. Özbek, A., Deán-Ben, X. L. & Razansky, D. Optoacoustic imaging at kilohertz volumetric frame rates. *Optica* **5**, 857–863 (2018).
14. Guo, Z., Li, C., Song, L. & Wang, L. V. Compressed sensing in photoacoustic tomography in vivo. *J. Biomed. Opt.* **15**, 021311 (2010).
15. Sandbichler, M., Krahmer, F., Berer, T., Burgholzer, P. & Haltmeier, M. A novel compressed sensing scheme for photoacoustic tomography. *SIAM J. Appl. Math.* **75**, 2475–2494 (2015).
16. Arridge, S. et al. Accelerated high-resolution photoacoustic tomography via compressed sensing. *Phys. Med. Biol.* **61**, 8908–8940 (2016).
17. Meng, J. et al. High-speed, sparse-sampling three-dimensional photoacoustic computed tomography in vivo based on principal component analysis. *J. Biomed. Opt.* **21**, 076007 (2016).
18. Gratt, S., Nuster, R., Wurzinger, G., Bugl, M. & Paltauf, G. 64-line-sensor array: fast imaging system for photoactive tomography. *Proc. SPIE* **8943**, 894365 (2014).
19. Rosenthal, A., Ntziachristos, V. & Razansky, D. Acoustic inversion in optoacoustic tomography: a review. *Curr. Med. Imaging Rev.* **9**, 318–336 (2013).
20. Hyun, C. M., Kim, H. P., Lee, S. M., Lee, S. & Seo, J. K. Deep learning for undersampled MRI reconstruction. *Phys. Med. Biol.* **63**, 135007 (2018).
21. Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555**, 487–492 (2018).
22. Hammernik, K., Würfl, T., Pock, T. & Maier, A. in *Bildverarbeitung für die Medizin 2017* 92–97 (Springer, 2017).
23. Kelly, B., Matthews, T. P. & Anastasio, M. A. Deep learning-guided image reconstruction from incomplete data. Preprint at https://arxiv.org/abs/1709.00584 (2017).
24. Bungart, B. L. et al. Photoacoustic tomography of intact human prostates and vascular texture analysis identify prostate cancer biopsy targets. *Photoacoustics* **11**, 46–55 (2018).
25. Zhang, J., Chen, B., Zhou, M., Lan, H. & Gao, F. Photoacoustic image classification and segmentation of breast cancer: a feasibility study. *IEEE Access* **7**, 5457–5466 (2019).
26. Antholzer, S., Haltmeier, M. & Schwab, J. Deep learning for photoacoustic tomography from sparse data. *Inverse Probl. Sci. Eng.* **27**, 987–1005 (2018).
27. Guan, S., Khan, A., Sikdar, S. & Chitnis, P. V. Fully dense UNet for 2D sparse photoacoustic tomography artifact removal. *IEEE J. Biomed. Health Inform.* https://doi.org/10.1109/JBHI.2019.2912935 (2019).
28. Hauptmann, A. et al. Model-based learning for accelerated, limited-view 3-D photoacoustic tomography. *IEEE Trans. Med. Imaging* **37**, 1382–1393 (2018).
29. Waibel, D. et al. Reconstruction of initial pressure from limited view photoacoustic images using deep learning. *Proc. SPIE* **10494**, 104942S (2018).
30. Cai, C., Deng, K., Ma, C. & Luo, J. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Opt. Lett.* **43**, 2752–2755 (2018).
31. Kirchner, T., Gröhl, J. & Maier-Hein, L. Context encoding enables machine learning-based quantitative photoacoustics. *J. Biomed. Opt.* **23**, 056008 (2018).
32. Reiter, A. & Bell, M. A. L. A machine learning approach to identifying point source locations in photoacoustic data. *Proc. SPIE* **10064**, 100643J (2017).
33. Allman, D., Reiter, A. & Bell, M. A. L. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE Trans. Med. Imaging* **37**, 1464–1477 (2018).
34. Govinahallisathyanarayana, S., Ning, B., Cao, R., Hu, S. & Hossack, J. A. Dictionary learning-based reverberation removal enables depth-resolved photoacoustic microscopy of cortical microvasculature in the mouse brain. *Sci. Rep.* **8**, 985 (2018).
35. Gutta, S. et al. Deep neural network-based bandwidth enhancement of photoacoustic data. *J. Biomed. Opt.* **22**, 116001 (2017).
36. Lutzweiler, C. & Razansky, D. Optoacoustic imaging and tomography: reconstruction approaches and outstanding challenges in image performance and quantification. *Sensors* **13**, 7345–7384 (2013).
37. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer, 2015).
38. Dean-Ben, X. L., Buehler, A., Ntziachristos, V. & Razansky, D. Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography. *IEEE Trans. Med. Imaging* **31**, 1922–1928 (2012).
39. Schwab, J., Antholzer, S., Nuster, R. & Haltmeier, M. Real-time photoacoustic projection imaging using deep learning. Preprint at https://arxiv.org/abs/1801.06693 (2018).
40. Omar, M. et al. Optical imaging of post-embryonic zebrafish using multi orientation raster scan optoacoustic mesoscopy. *Light Sci. Appl.* **6**, e16186 (2017).
41. Xu, Y., Wang, L. V., Ambartsoumian, G. & Kuchment, P. Reconstructions in limited-view thermoacoustic tomography. *Med. Phys.* **31**, 724–733 (2004).
42. Deán-Ben, X. L. & Razansky, D. On the link between the speckle free nature of optoacoustics and visibility of structures in limited-view tomography. *Photoacoustics* **4**, 133–140 (2016).
43. Li, L. et al. Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution. *Nat. Biomed. Eng.* **1**, 0071 (2017).
44. Queiros, D. et al. Modeling the shape of cylindrically focused transducers in three-dimensional optoacoustic tomography. *J. Biomed. Opt.* **18**, 076014 (2013).
45. Dean-Ben, X. L., Ntziachristos, V. & Razansky, D. Effects of small variations of speed of sound in optoacoustic tomographic imaging. *Med. Phys.* **41**, 073301 (2014).
46. Fehm, T. F., Deán-Ben, X. L., Ford, S. J. & Razansky, D. In vivo whole-body optoacoustic scanner with real-time volumetric imaging capacity. *Optica* **3**, 1153–1159 (2016).
47. Tang, J., Dai, X. & Jiang, H. Wearable scanning photoacoustic brain imaging in behaving rats. *J. Biophoton.* **9**, 570–575 (2016).
48. Nair, A. A., Gubbi, M. R., Tran, T. D., Reiter, A. & Bell, M. A. L. A fully convolutional neural network for beamforming ultrasound images. In *2018 IEEE International Ultrasonics Symposium (IUS)* 1–4 (IEEE, 2018).
49. Neuschmelting, V. et al. Performance of a multispectral optoacoustic tomography (MSOT) system equipped with 2D vs. 3D handheld probes for potential clinical translation. *Photoacoustics* **4**, 1–10 (2016).
50. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
51. Bastien, F. et al. Theano: new features and speed improvements. Preprint at https://arxiv.org/abs/1211.5590 (2012).
52. Dieleman, S. et al. Lasagne: first release 3 (Zenodo, 2015).
53. Merčep, E., Herraiz, J. L., Deán-Ben, X. L. & Razansky, D. Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals. *Light Sci. Appl.* **8**, 18 (2019).
54. Xu, M. & Wang, L. V. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E* **71**, 016706 (2005).
55. Estrada, H. C. et al. Virtual craniotomy for high-resolution optoacoustic brain microscopy. *Sci. Rep.* **8**, 1459 (2018).
56. Rosenthal, A., Razansky, D. & Ntziachristos, V. Fast semi-analytical model-based acoustic inversion for quantitative optoacoustic tomography. *IEEE Trans. Med. Imaging* **29**, 1275–1285 (2010).
57. Dean-Ben, X. L., Ntziachristos, V. & Razansky, D. Acceleration of optoacoustic model-based reconstruction using angular image discretization. *IEEE Trans. Med. Imaging* **31**, 1154–1162 (2012).
58. Wang, L. V. & Wu, H.-i. *Biomedical Optics: Principles and Imaging* (Wiley, 2012).

Acknowledgements

This work was partially supported by the European Research Council under grant agreement ERC-2015-CoG-682379.

Author contributions

N.D., X.L.D.-B. and D.R. conceived the study. N.D. and X.L.D.-B. carried out the experiments. N.D. implemented the image reconstruction and processing algorithms and analysed the data. D.R. and X.L.D.-B. supervised the study and data analysis. All authors discussed the results and contributed to writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42256-019-0095-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.R.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019