

Terabyte-scale supervised 3D training and benchmarking dataset of the mouse kidney

Journal Article**Author(s):**

Kuo, Willy; Rossinelli, Diego; Schulz, Georg; Wenger, Roland H.; Hieber, Simone; Müller, Bert; Kurtcuoglu, Vartan

Publication date:

2023-08

Permanent link:

<https://doi.org/https://doi.org/10.3929/ethz-b-000632024>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Scientific Data 10(1), <https://doi.org/10.1038/s41597-023-02407-5>



OPEN

DATA DESCRIPTOR

Terabyte-scale supervised 3D training and benchmarking dataset of the mouse kidney

Willy Kuo^{1,2,4}, Diego Rossinelli^{1,2,4}, Georg Schulz³, Roland H. Wenger^{1,2}, Simone Hieber³, Bert Müller^{3,5} & Vartan Kurtcuoglu^{1,2,5}✉

The performance of machine learning algorithms, when used for segmenting 3D biomedical images, does not reach the level expected based on results achieved with 2D photos. This may be explained by the comparative lack of high-volume, high-quality training datasets, which require state-of-the-art imaging facilities, domain experts for annotation and large computational and personal resources. The HR-Kidney dataset presented in this work bridges this gap by providing 1.7 TB of artefact-corrected synchrotron radiation-based X-ray phase-contrast microtomography images of whole mouse kidneys and validated segmentations of 33 729 glomeruli, which corresponds to a one to two orders of magnitude increase over currently available biomedical datasets. The image sets also contain the underlying raw data, threshold- and morphology-based semi-automatic segmentations of renal vasculature and uriniferous tubules, as well as true 3D manual annotations. We therewith provide a broad basis for the scientific community to build upon and expand in the fields of image processing, data augmentation and machine learning, in particular unsupervised and semi-supervised learning investigations, as well as transfer learning and generative adversarial networks.

Background & Summary

Supervised learning has been the main source of progress in the field of artificial intelligence/machine learning in the past decade¹. Impressive results have been obtained in the classification of two-dimensional (2D) color images, such as consumer photos or histological sections. Supervised learning requires high-volume, high-quality training datasets. The use of training data that do not fulfill these requirements may severely hamper performance: low-volume training data may result in poor classification of samples reasonably distant from any sample in the training set due to possible overfitting of the respective algorithm. Low quality may result in the algorithms learning the ‘wrong lessons’, as the algorithms typically do not include prior knowledge on which aspects of the images constitute artefacts. It is thus not surprising that the most used benchmarks in the field are based on large datasets of tiny 2D images, such as CIFAR², ImageNet³ and MNIST⁴.

In contrast to the semantic segmentation of 2D photos, shape detection within three-dimensional (3D) biomedical datasets arguably poses fewer technical challenges to machine learning. After all, 2D photos typically contain multiple color channels, represent the projection of a 3D object on a 2D plane, feature occlusions, are often poorly quantized and contain artefacts such as under- or overexposure and optical aberrations. Despite this, machine learning approaches have not reached the same performance in the analysis of 3D biomedical images as would be expected by their results achieved in 2D photos.

This discrepancy may be explained by the lack of sufficiently large-volume, high-quality training datasets, especially in the area of pre-clinical biomedical research data. The workload for manual annotation becomes excessive when a third dimension has to be considered. Also, domain experts capable of annotating biomedical data are in short supply compared to untrained personnel annotating recreational 2D photos^{5,6}. To reduce workload, 3D datasets are typically annotated only on a limited number of 2D slices⁷. This sparse annotation approach may, however, hamper machine learning, as the volume of training data may be insufficient to avoid overfitting. While this is in principle enough to train 3D segmentation models, the precise location of the 2D slices within

¹Institute of Physiology, University of Zurich, Zurich, Switzerland. ²National Centre of Competence in Research, Kidney.CH, Zurich, Switzerland. ³Biomaterials Science Center, Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland. ⁴These authors contributed equally: Willy Kuo, Diego Rossinelli. ⁵These authors jointly supervised this work: Bert Müller, Vartan Kurtcuoglu. ✉e-mail: vartan.kurtcuoglu@uzh.ch

the 3D grayscale signal has to be explicitly taken into consideration, possibly hampering straightforward training. Performance of supervised learning in the pre-clinical biomedical field may have reached a plateau, with potential improvements being held back by the lack of suitable training datasets, rather than by the intrinsic power of the underlying algorithms.

An example to demonstrate this issue can be found in a study by Pinto *et al.*⁸ who showed that a simple V1-like model was able to outperform state-of-the-art object recognition systems of the time, because the Caltech101 dataset⁹ used for benchmarking was inadequate for leveraging the advantages of the more advanced algorithms. The simple model's performance rapidly degraded for other images not included in the dataset, which confirmed that the limited variation in the benchmarking dataset was responsible for the observed performance ceiling. Had more varied datasets been used for benchmarking, the simple model would not have been able to outperform the state-of-the-art models and a larger performance gap would have been apparent.

Currently available large datasets, for example those published on grand-challenge.org, are based predominantly on clinical data. The major challenges for machine learning algorithms with this type of data are linked to widely varying intensity values, low signal-to-noise ratios, low resolution, and the presence of artefacts. These are, however, not the only challenges encountered in image processing, and they do not necessarily apply to other types of biomedical data. In pre-clinical imaging, continuous improvements of spatial resolution and acquisition speeds have led to ever larger data volumes, and with those to a proportional increase in human labor required for manual or semi-automatic segmentation. Machine learning algorithms have the potential to substantially reduce the amount of human labor needed, provided that they can work with low amounts of training data, where the annotation workload does not exceed the workload required for classical segmentation. Algorithms that perform well in such a low-data regime may greatly accelerate biomedical research, but promising candidates may be overlooked because of the current focus on clinical segmentation challenges.

In this work, we present HR-Kidney, a high-resolution kidney dataset. It is, to our knowledge, the largest supervised, fully 3D training dataset of biomedical research images to date, containing 33 729 renal glomeruli, viewed and validated by a domain expert. HR-Kidney is based on terabyte-scale synchrotron radiation-based X-ray phase-contrast microtomography (SR μ CT) acquisitions of three whole mouse kidneys at micrometer resolution. These training data, along with fully manually annotated 3D regions of interest, are available for download at the Image Data Resource (IDR, <https://idr.openmicroscopy.org/>)¹⁰ repository. Underlying raw data in the form of X-ray radiographs and reconstructed 3D volumes are supplied as well, as are reference segmentations of the vascular and tubular vessel trees.

Databases of photographs such as CIFAR² and ImageNet³ are currently the principle sources of data for testing and benchmarking machine learning algorithms, which can be attributed to the very high volume of annotation data available. As a result, these databases feature a high performance ceiling for benchmarking. However, the workload for creating these datasets is tremendous. For example, annotations for the ImageNet database were carried out over the period of three years by a labor force of 49 000 people hired via Amazon Mechanical Turk³.

Biomedical databases cannot be created by similar use of crowdsourced, untrained personnel, as annotations require domain expertise to avoid misclassification caused by a lack of familiarity with the different classes. Furthermore, quantitative assessment of biomedical markers such as blood vessel density requires segmentation rather than classification, which takes considerably more time to complete. This is exemplified by the PASCAL VOC dataset¹¹, which contains four times less segmentations than classifications. Due to these additional challenges, data volumes in biomedical databases lag behind those of photographic databases. With HR-Kidney, we are providing 1701 GB of artefact-corrected image data and 33 729 segmented glomeruli, which represent a one to two orders of magnitude increase over currently available biomedical 3D databases^{12–16}, reaching a data volume comparable to the photographic datasets popular for machine learning benchmarks.

This substantial increase in size compared to the state-of-the-art may enable disruptive developments in machine learning, as most research groups in the field do not have the combination of sample preparation expertise, access to synchrotron radiation facilities, high performance computing resources and domain knowledge to create training datasets at this scale and quality.

Terabyte-scale X-ray microtomography images of the renal vascular network. We acquired SR μ CT images of whole mouse kidneys with 1.6 μ m voxel size, ensuring sufficient sampling of functional capillaries, the smallest of which are 4 μ m in diameter¹⁷. High image quality in terms of signal-to-noise ratio was achieved by employing the ID19 micro tomography beamline of the European Synchrotron Radiation Facility (ESRF), which provides several orders of magnitude higher brightness than conventional laboratory source microtomography devices. This allows for propagation-based phase contrast imaging, which leverages sample refractive index-dependent edge-enhancement for improved contrast¹⁸. Fig. 1 provides an overview of the data acquisition and processing pipeline. Therein, the raw SR μ CT radiographs are marked as **D1**.

Vascular signal-to-noise ratio was further improved by the application of a custom-developed mixture of contrast agent in a vascular casting resin capable of entering the smallest capillaries and filling the entire vascular bed. There was sufficient contrast to extract an initial blood vessel segment using curvelet-based denoising and hard thresholding. Connectivity analysis ensured and confirmed that the blood vessel segment, including capillary bed, was fully connected. Microscopic gas bubbles caused by external perfusion of the kidney, which is part of the organ preparation procedure, were excluded from the image set by applying connected component analysis to the background. This allowed for the elimination of all gas bubbles not in contact with the vessel boundary (**D5** in Fig. 1).

The contrast agent, 1,3-diiodobenzene, also diffused into the lipophilic white adipose tissue, resulting in high X-ray absorption in the perirenal fat, which is, therefore, included in the vascular segment (Fig. 2a). To visualize the segment without fat, a machine learning approach based on invariant scattering convolution networks¹⁹ was

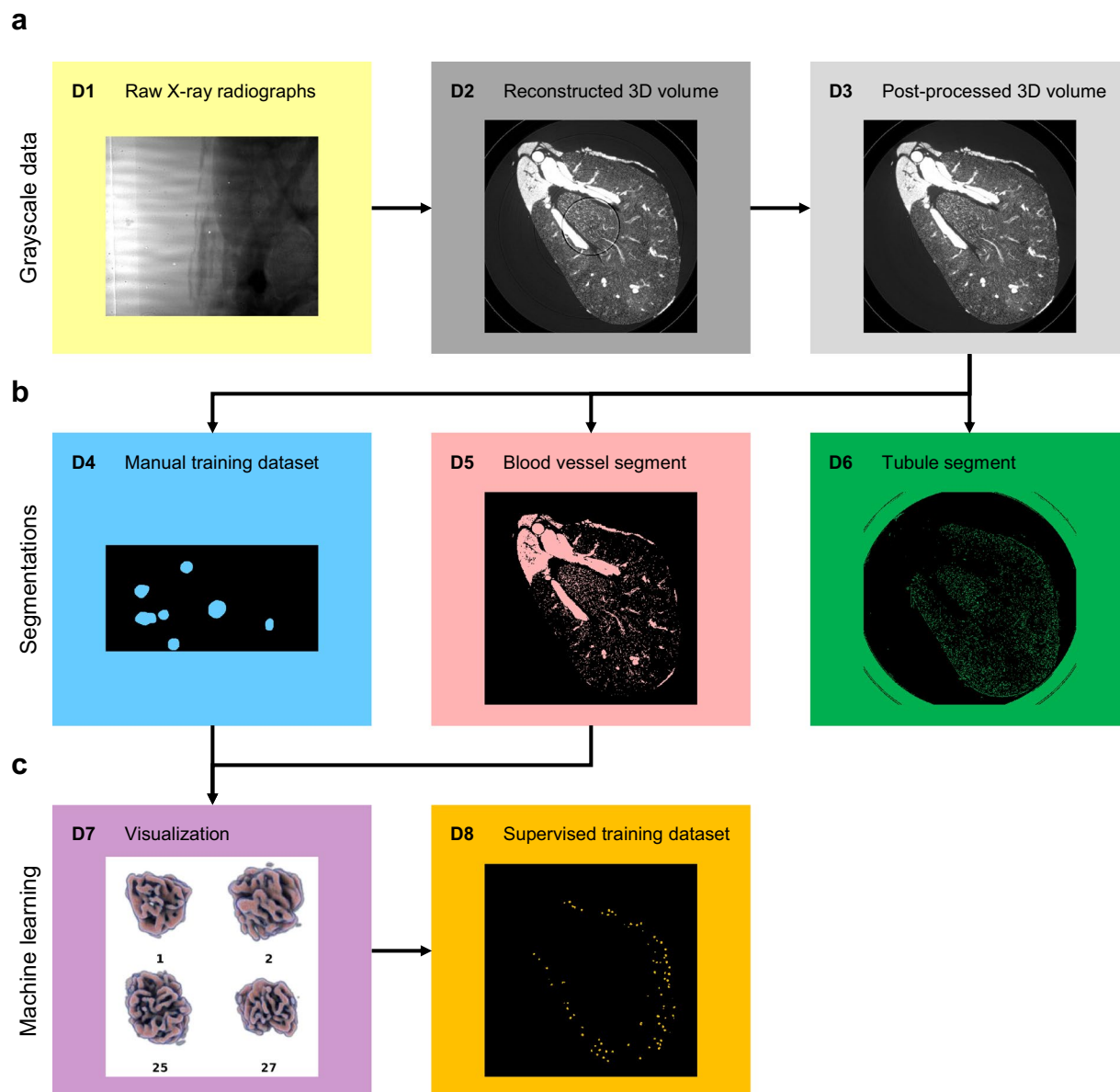


Fig. 1 Overview of the data made available in the repository. (a) Grayscale data: X-ray radiographs (D1) were reconstructed into 3D volumes (D2) and post-processed for ring artefact removal (D3). (b) Classical segmentations: Glomeruli were manually contoured in small regions of interests (D4). Blood vessels (D5) and tubules (D6) were segmented by noise removal, thresholding and connectivity analysis. (c) Machine learning segmentations: Glomeruli identified via machine learning were combined with the blood vessel segment for visualization (D7), which was viewed by a domain expert to validate the final glomerulus mask (D8). A list of dataset contents and file formats is provided in Table 1.

applied, removing most of the blob-shaped fat globules and the more dense fatty tissue surrounding the collecting duct (D5b). Using the resulting segment, vessel thickness was calculated according to the concept of largest inscribed sphere (Fig. 2b,d).

Post-processed dataset and reference segmentations. Ring artefacts compromise segmentation, as the corresponding areas may be erroneously attributed to the blood vessel or tubule segments, depending on their intensity (Fig. 3a). To avoid this, kidneys were scanned with overlapping height steps, and the overlapping regions of the neighboring height steps were employed to produce a post-processed 3D volume with greatly reduced ring artefacts (D3 in Fig. 1), on which blood vessel and tubular segmentations were performed. Connectivity analysis yielded the fully connected tree of the blood vessel segment (D5). Insufficient contrast and resolution within the inner medulla prevented similar connectivity analysis on the tubular segment (D6).

High quality annotations of glomeruli in full 3D. Glomeruli are the primary filtration units of the kidney. They feature characteristic ball-shaped vascular structures. As they present with the same gray values and

	Filenames (Kidney_2)	Data provided	Precision	Data format
D1	kidney2_D1_projections_height1.ome.tiff kidney2_D1_projections_height2.ome.tiff kidney2_D1_projections_height3.ome.tiff kidney2_D1_projections_height4.ome.tiff kidney2_D1_projections_height5.ome.tiff kidney2_D1_projections_height6.ome.tiff	X-ray radiographs	Grayscale, 16-bit unsigned integer	OME-TIFF (.tiff) In attachment: ESRF data format (.edf, raw binary with 1024 byte header describing image dimensions)
D2	kidney2_D2_reco_height1.ome.tiff kidney2_D2_reco_height2.ome.tiff kidney2_D2_reco_height3.ome.tiff kidney2_D2_reco_height4.ome.tiff kidney2_D2_reco_height5.ome.tiff kidney2_D2_reco_height6.ome.tiff	Paganin filtered, reconstructed 3D volume	Grayscale, 32-bit floating point	OME-TIFF (.tiff)
D3	kidney2_D3_inpainted.ome.tiff	Artifact-corrected, inpainted 3D volume	Grayscale, 32-bit floating point	OME-TIFF (.tiff)
D4	kidney2_D4_h3_glomeruli_roi1_annotation.ome.tiff kidney2_D4_h3_glomeruli_roi1_raw_abs.ome.tiff kidney2_D4_h3_glomeruli_roi1_raw_pag.ome.tiff kidney2_D4_h4_glomeruli_roi2_annotation.ome.tiff kidney2_D4_h4_glomeruli_roi2_raw_abs.ome.tiff kidney2_D4_h4_glomeruli_roi2_raw_pag.ome.tiff kidney2_D4_h5_glomeruli_roi3_annotation.ome.tiff kidney2_D4_h5_glomeruli_roi3_raw_abs.ome.tiff kidney2_D4_h5_glomeruli_roi3_raw_pag.ome.tiff	Glomeruli annotation, 3 regions of interests (512 × 256 × 256)	Binary, 8-bit integer	OME-TIFF (.tiff)
D5	kidney2_D5_blood_vessels.ome.tiff	Blood vessel segment	Binary, 8-bit integer	OME-TIFF (.tiff)
D5b	kidney2_D5b_blood_vessels_defatted.ome.tiff	Defatted blood vessel segment	Binary, 8-bit integer	OME-TIFF (.tiff)
D6	kidney2_D6_tubules.ome.tiff	Tubule segment	Binary, 8-bit integer	OME-TIFF (.tiff)
D7	kidney2_D7_glogallery_00.ome.tiff kidney2_D7_glogallery_01.ome.tiff kidney2_D7_glogallery_02.ome.tiff kidney2_D7_glogallery_03.ome.tiff kidney2_D7_glogallery_04.ome.tiff kidney2_D7_glogallery_05.ome.tiff kidney2_D7_glogallery_06.ome.tiff kidney2_D7_glogallery_07.ome.tiff kidney2_D7_glogallery_08.ome.tiff kidney2_D7_glogallery_09.ome.tiff kidney2_D7_glogallery_10.ome.tiff kidney2_D7_glogallery_11.ome.tiff	Gallery of visualized glomeruli	RGB, 8-bit integer per channel	OME-TIFF (.tiff)
D8	kidney2_D8_glomeruli_segment.ome.tiff	Glomeruli segment	Binary, 8-bit integer	OME-TIFF (.tiff)
D9	kidney2_D9_h2_fat_roi1_annotation.ome.tiff kidney2_D9_h2_fat_roi1_raw_abs.ome.tiff kidney2_D9_h2_fat_roi1_raw_pag.ome.tiff kidney2_D9_h2_fat_roi2_annotation.ome.tiff kidney2_D9_h2_fat_roi2_raw_abs.ome.tiff kidney2_D9_h2_fat_roi2_raw_pag.ome.tiff kidney2_D9_h4_fat_roi3_annotation.ome.tiff kidney2_D9_h4_fat_roi3_raw_abs.ome.tiff kidney2_D9_h4_fat_roi3_raw_pag.ome.tiff	Extrarenal fat annotation, 3 regions of interests (512 × 256 × 256)	Binary, 8-bit integer	OME-TIFF (.tiff)

Table 1. List of filenames, data contents, data types, bit depths and data formats contained within the HR-Kidney dataset available at the Image Data Resource (IDR) repository. Filenames are indicated for "Kidney_2", which is the only dataset for which manual annotation data (**D4**, **D9**) and defatted blood vessel segment (**D5b**) are available.

vessel diameters as other blood vessels in the kidney, differ only in their morphology, possess diverse sizes and shapes and are present in large, discrete numbers in the kidney, they pose a difficult segmentation challenge with a large sample size.

When contouring features of interest in 3D, however, the required workload is multiplied by the number of slices in the third dimension. In our work, this corresponds to a factor of 256 for a region of interest and 7168 for an entire kidney. Due to this extremely high contouring workload, 3D data are generally not annotated in full 3D, but rather partially annotated by selecting and contouring only a few sparse 2D slices out of the whole dataset⁷. This approach decreases the accuracy of the segmentation in the third dimension, which may be acceptable for clinical images, where standardized body positions or anisotropic resolutions reduce the necessity of rotational invariance in both algorithms and segmentation. In more generalized segmentation problems encountered in biomedical imaging on the other hand, this approach prevents these invariances from being leveraged for data augmentation techniques and may reduce performance of corresponding algorithms²⁰. For this reason, the annotator manually contoured all slices in three full 3D regions of interest of 512 × 256 × 256 voxels in size (**D4** in Fig. 1, Fig. 3b).

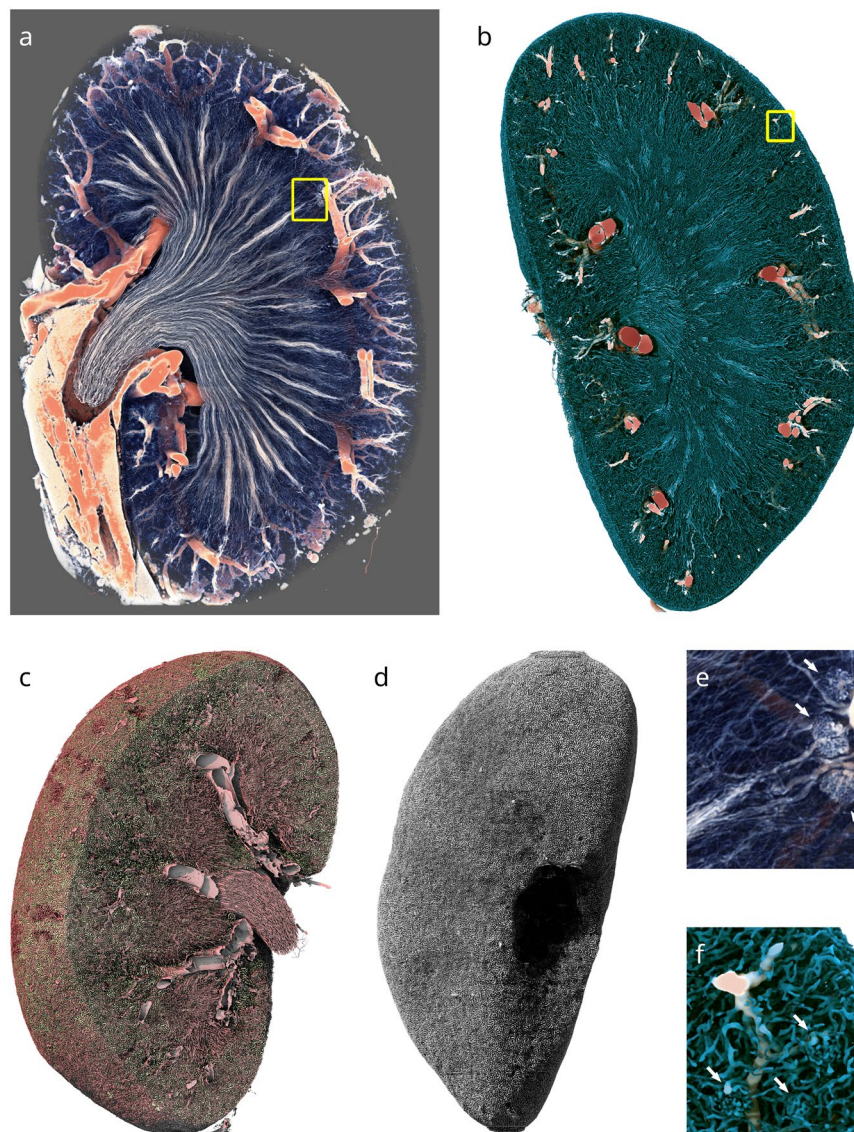


Fig. 2 Computer graphics renderings of the vascular and tubular structure of a mouse kidney. **(a)** Volume rendering of a Paganin-filtered kidney dataset (**D2**), clipped for visibility. Red colors represent higher intensity. Only the higher intensity values, which are due to the contrast agent in larger blood vessels and perirenal fat, are shown. **(b)** Opaque rendering of the thickness transform of the defatted blood vessel binary mask (**D2b**). Colors correspond to largest inscribed sphere radius. **(c)** Surface rendering of the post-processed segmented vascular (red) and tubular (green) lumina in the cortex and inner medulla. **(d)** Surface rendering of the segmented tubular lumina only (**D6**). **(e)** Magnified view of the region highlighted by the yellow square in **a**. Three juxtamedullary glomeruli are visible in the right part of the image (white arrows). **(f)** Magnified view of the region highlighted by the yellow square in **b**. Three cortical glomeruli are visible in the bottom half of the image (white arrows). Height of the kidney: 10 mm.

Identification of all individual glomeruli. Combining these manual training data with the scattering transform approach, we were able to identify 10 031, 11 238 and 12 460 glomeruli, respectively, in the three kidney datasets. Each glomerulus was 3D visualized in a gallery (Fig. 4b), viewed by a single domain expert and classified by its shape as false positive, true positive with shape distortion, or true positive without shape distortion. Only 15, 4 and 4 glomeruli, respectively, were identified as false positives by the rater, and were typically the result of poor contrast-to-noise in the specific region of the underlying raw images. Such localized areas of poor contrast appear to be caused by limited diffusion of the radiopaque 1,3-diiodobenzene from the vascular cast into the surrounding tissue, increasing background gray values (Fig. 3c). False negatives were estimated to be approximately 2603, 2306 and 1420 using unbiased stereological counting on selected virtual sections, corresponding to miss rates of 20%, 17% and 10%, respectively. It should be noted that both the manual segmentation of training data and validation were performed by the same, single annotator, meaning that annotator-specific bias cannot

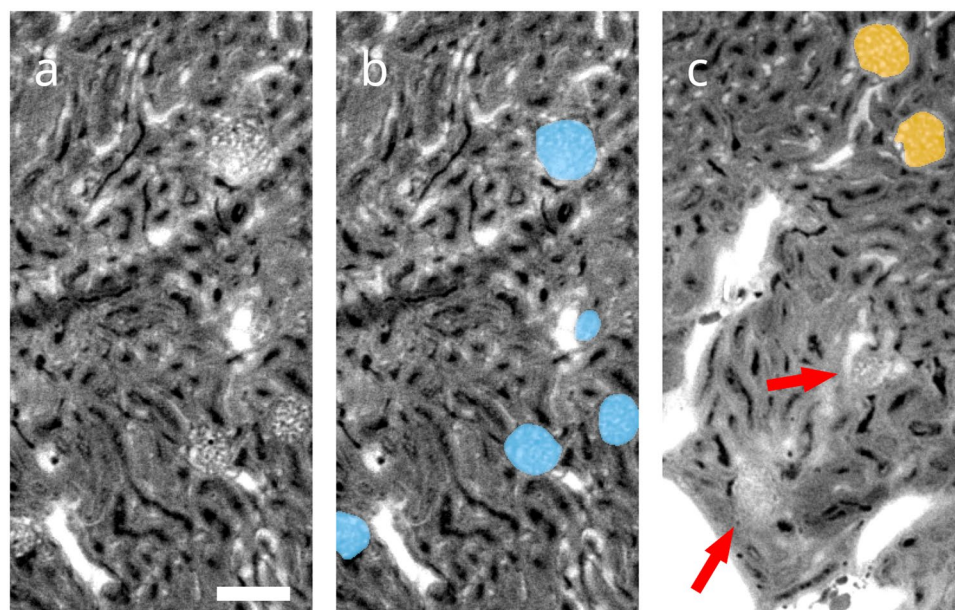


Fig. 3 Local views of the reconstructed 3D volume prior to artefact correction (D2). (a) A single virtual section of one of the regions of interest selected for manual contouring. Glomerular blood vessels feature the same gray values and size scales as other blood vessels in the kidney and differ only in their morphology. A prominent ring artefact can be observed touching the top glomerulus. Scale bar: 100 μm (b) Manual annotation (D4) of glomeruli overlaid in blue over raw data. (c) Different region of interest containing glomeruli identified by machine learning (D8) overlaid in orange over raw data, as well as false negatives denoted by red arrows. The majority of missed glomeruli are in similar regions of poor contrast-to-noise, which are characterized by elevated tissue background intensity. These are likely caused by limited diffusion of 1,3-diiodobenzene from the cast into surrounding tissue. The few false positives identified are also located mainly in such regions.

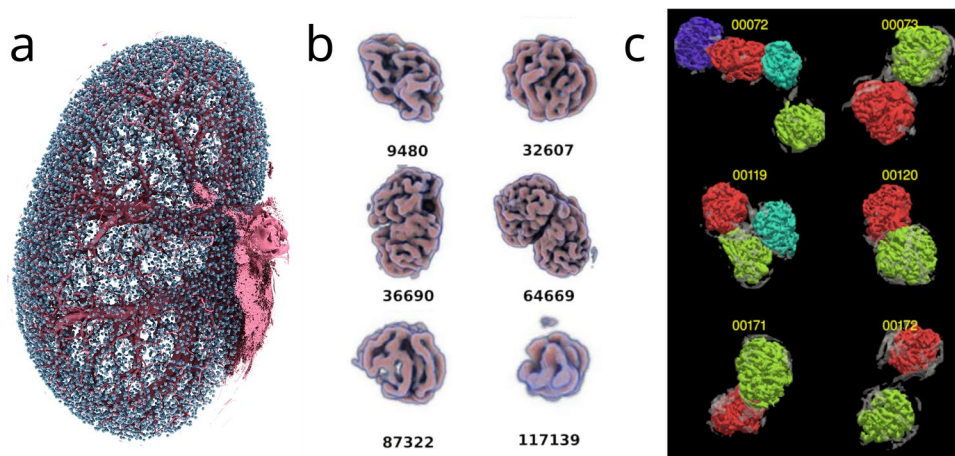


Fig. 4 Computer generated images of glomeruli. (a) All identified glomeruli (D8) of one kidney are shown in cyan in their original spatial location. Large pre-glomerular vessels are rendered in magenta, for orientation. (b) Excerpt from the gallery of volume-rendered glomeruli, exhibiting a selection of different glomerulus sizes and shapes. Each glomerulus or cluster of glomeruli was assigned an identification number and viewed by an expert. (c) Results of the morphometric analysis to separate clustered glomeruli.

be excluded. On the other hand, the lack of inter-rater variability removes a source of label noise, which may be beneficial in some applications, such as allowing for a higher performance ceiling in benchmarking²¹.

The binary masks containing these glomeruli represent the supervised and validated training dataset (D8 in Fig. 1) and are available in the repository, along with the visualizations and expert classifications. Morphometric analysis was employed to separate individual glomeruli within clusters in which they appeared fused due to shared vessels (Fig. 4c).

Methods

Materials list, in-depth surgery guide and more detailed description of the data processing are provided in the Supplementary Information.

Mouse husbandry. C57BL/6J mice were purchased from Janvier Labs (Le Genest-Saint-Isle, France) and kept in individually ventilated cages with *ad libitum* access to water and standard diet (Kliba Nafag 3436, Kaiseraugst, Switzerland) in 12 h light/dark cycles. Dataset 1 derives from the left kidney of a male mouse, 15 weeks of age with a body weight of 28.0 g. Dataset 2 is the right kidney of the same mouse. Dataset 3 derives from the right kidney of a female mouse, 15 weeks of age with a body weight of 22.5 g. All animal experiments were approved by the cantonal veterinary office of Zurich, Switzerland, in accordance with the Swiss federal animal welfare regulations (license numbers ZH177/13 and ZH233/15).

Perfusion surgery. Mice were anaesthetized with ketamine/xylazine. A blunted 21 G butterfly needle was inserted retrogradely into the abdominal aorta and fixed with a ligation (Figures 6, 7)²². The abdominal aorta and superior mesenteric artery above the renal arteries were ligated, the vena cava opened as an outlet and the kidneys were flushed with 10 ml, 37 °C phosphate-buffered saline (PBS) to remove the blood, then fixed with 50 ml 37 °C 4% paraformaldehyde in PBS (PFA) solution at 150 mmHg hydrostatic pressure.

Vascular casting. 2.4 g of 1,3-diodobenzene (Sigma-Aldrich, Schnellendorf, Germany) were dissolved in 7.5 g of 2-butanone (Sigma-Aldrich) and mixed with 7.5 g PU4ii resin (vasQtec, Zurich, Switzerland) and 1.3 g PU4ii hardener. The mixture was filtered through a paper filter and degassed extensively in a vacuum chamber to minimize bubble formation during polymerization, and perfused at a constant pressure of no more than 200 mmHg until the resin mixture solidified. Kidneys were excised and stored in 15 ml 4% PFA. For scanning, they were embedded in 2% agar in PBS in 0.5 ml polypropylene centrifugation tubes. Kidneys were quality-checked with a nanotom[®] m (phoenix|x-ray, GE Sensing & Inspection Technologies GmbH, Wunstorf, Germany). Samples showing insufficient perfusion or bleeding of resin into the renal capsule or sinuses were excluded.

ESRF ID19 micro-CT measurements. Ten kidneys were scanned at the ID19 tomography beamline of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using pink beam with a mean photon energy of 19 keV. Radiographs were recorded at a sample-detector distance of 28 cm with a 100 µm Ce:LuAG scintillator, 4 × magnification lens and a pco.edge 5.5 camera with a 2560 × 2160 pixel array and 6.5 µm pixel size, resulting in an effective pixel size of 1.625 µm. Radiographs were acquired with a half-acquisition scheme²³ in order to extend the field of view to 8 mm. Six height steps were recorded for each kidney, with half of the vertical field of view overlapping between each height step, resulting in fully redundant acquisition of the inner height steps.

5125 radiographs were recorded for each height step with 0.1 s exposure time, resulting in a scan time of 1 h for a whole kidney. 100 flat-field images were taken before and after each height step for flat-field correction. Images were reconstructed using the beamline's in-house PyHST2 software, using a Paganin-filter with a low δ/β ratio of 50 to limit loss in resolution and appearance of gradients close to large vessels^{18,24,25}. Registration for stitching two half-acquisition radiographs to the full field of view was performed manually with 1 pixel accuracy. Data size for the reconstructed datasets was 1158 GB per kidney.

Image stitching and inpainting. Outliers in intensity in the recorded flat fields were segmented by noise reduction with 2D continuous curvelets, followed by thresholding to calculate radius and coordinates of the ring artefacts. The redundant acquisition of the central four height steps allowed us to replace corrupted data with a weighted average during stitching. The signals of the individual slices were zeroed in the presence of the rings, summed up and normalized by counting the number of uncorrupted signals. In the outer slices, where no redundant data was available, and in locations where rings coincided in both height steps, we employed a discrete cosine transform-based inpainting technique with a simple iterative approach, where we picked smoothing kernels progressively smaller in size and reconstructed the signal in the target areas by smoothing the signal everywhere at each iteration. The smoothed signal in the target areas was then combined with the original signal elsewhere to form a new image. In the next iteration, in turn, the new image was then smoothed to rewrite the signal at the target regions. The final inpainted signal exhibits multiple scales since different kernel widths are considered at different iterations.

The alignment for stitching the six stacks was determined by carrying out manual 3D registration and double checking against pairwise stack-stack phase-correlation analysis²⁶. The stitching process reduced the dataset dimensions per kidney to 4608 × 4608 × 7168 pixels, totaling 567 GB.

Semi-automatic segmentation of the vascular and tubular trees. We performed image enhancement based on 3D discretized continuous curvelets²⁷, in a similar fashion as Starck *et al.*²⁸, but with second generation curvelets (i.e., no Radon transform) in 3D. The enhancement was carried out globally by leveraging the Fast Fourier Transform with MPI-FFTW²⁹, considering about 100 curvelets. The “wedges” (curvelets in the spectrum) have a conical shape and cover the unit sphere in an approximately uniform fashion. For a given curvelet, a per-pixel coefficient is obtained by computing an inverse Fourier transform of its wedge and the image spectrum. We then truncated these coefficients in the image domain against a hard threshold, and forward-transformed the curvelet again into the Fourier space, modulated the curvelets with the truncated coefficients and superposed them. As a result, the pixel intensities were compressed to a substantially smaller range of values, thus helping to avoid over- and under-segmentation of large and small vessels, respectively. A threshold-based segmentation followed the image enhancement. The enhancement parameters and threshold were manually chosen by examining six randomly chosen regions of interest. Spurious islands were removed by 26-connected component analysis, and cavities were removed by 6-connected component analysis.

The bulk of the processing workload, required to transform data into an actionable training set, was carried out at the Zeus cluster of the Pawsey supercomputing centre. Zeus consisted of hundreds of computing nodes featuring Intel Xeon Phi (Knights Landing) many-core CPUs, together with 96 GB of “special” high-bandwidth memory (HBM/MCDRAM), as well as 128 GB of conventional DDR4 RAM. The final training and assessments were carried out at the Euler VI cluster of ETH Zurich, with two-socket nodes featuring AMD EPYC 7742 (Rome) CPUs and 512 GB of DDR4 RAM.

Identification of glomeruli via scattering transform. A machine learning-based approach relying on invariant scattering convolution networks was employed to segment the glomeruli and remove perirenal fat from the blood vessel segment¹⁹. For the glomerular training data, three selected regions of interest of $512 \times 256 \times 256$ voxels in size were selected from the cortical region of one kidney (dataset 2) and segmented by a single annotator by fully manual contouring in all slices. For the fat, manual work was reduced by providing an initial semiautomatic segmentation, which the manual annotation then corrected. The training data were supplemented by additional regions of interest that contained no glomeruli or fat at all, and thus did not require manual annotation. The manual annotations were then used to train a hybrid algorithm that relied on a 3D scattering transform convolutional network topped with a dense neural network. The scattering transform relied upon ad-hoc designed 3D kernels (Morlet’s wavelet with different sizes and orientations) that uniformly covered all directions at different scales. In the scattering convolutional network, filter nonlinearities were obtained by taking the magnitude of the filter responses and convolving them again with the kernels in a cascading fashion. These nonlinearities are designed to be robust against small Lipschitz-continuous deformations of the image¹⁹.

In contrast to our curvelet-based image enhancement approach, we decomposed the image into cubic tiles, then applied a windowed (thus local) Fourier transform on the tiles by considering regions about twice their size around them. While it would have been possible to use a convolutional network based upon a global scattering transform, this would have produced a very large number of features that would have had to be consumed at once, leading to an intermediate footprint in the petabyte-scale, exceeding the available memory of the cluster.

The scattering transform convolutional network produced a stack of a few hundred scalar feature maps per pixel. If considered as a “fiber bundle”^{30,31}, the feature map stack is equivariant under the symmetry group of rotations (i.e., the stack is a regular representation of the 3D rotation group $SO(3)$). This property can be exploited by further processing the feature maps with a dense neural network with increased parameter sharing across the hidden layers, making the output layer-invariant to rotations.

Data Records

The dataset is available at the Image Data Resource (IDR) repository at <https://doi.org/10.17867/10000188> under accession number idr0147³². As per the repository’s guidelines, all data is available in the OME-TIFF format, which features the ability to load downsampled versions of the image data, as well as viewing them on the repository’s online image viewer. Raw X-ray radiographs (**D1**) are further provided as an attachment in the original ESRF data format as well, which are raw binary image files with a 1024 bytes header describing the necessary metadata to open the images. All radiographs provided in this format feature image dimensions of 2560×2160 pixels, 16-bit unsigned integer bit depth and little-endian byte order.

File structure. The HR-Kidney datasets deposited in IDR are collected under a main folder named “idr0147-kuo-kidney3d”. The datasets of the three kidneys are collected in three subfolders “Kidney_1”, “Kidney_2” and “Kidney_3”. The filenames of all data provided in the folder “Kidney_2” are provided in Table 1. Datasets of the other kidneys follow the same naming scheme, differing only in the kidney number.

Raw X-ray radiographs (**D1**) are provided both in the OME-TIFF-format, and in the original ESRF data format (.edf) as attachments on IDR. Manual annotation data (**D4**, **D9**) and defatted blood vessel segment (**D5b**) are only available for “Kidney_2”. Annotation data are provided along with the excerpts (**D4**) from the reconstructed, Paganin-filtered 3D volumes (**D2**), denoted with a filename ending with “_pag”. Excerpts of the ROIs of the equivalent 3D volumes reconstructed without Paganin-filtering (“absorption images”) are provided with a filename ending with “_abs”, as they formed the basis of the manual annotation. They were not employed for training and are only provided for documentation. As the two versions are reconstructions of the same dataset, the manual annotations are equally valid for both. Coordinates and dimensions for extracting the ROIs are provided as metadata on IDR.

Technical Validation

Validation of glomeruli by a domain expert. To reduce the workload to the level required to make validation of each glomerulus feasible, a volumetric visualization of the overlap between the blood vessel mask and the glomerular mask was generated for rapid evaluation of easy to recognize glomeruli or artefacts. In a first round, glomeruli were classified by their shape into three categories: 1. certain true positive with shape distortion, 2. certain true positive without shape distortion and 3. uncertain. Candidates of the third category, which constituted about 3 to 5% of all machine-identified glomeruli, were then reviewed in a second round as an overlay over the original raw data on a slice-by-slice basis. These candidates were then assigned as false positives or as certain glomeruli of categories 1 or 2.

The numbers of false negative rates were assessed using stereological counting, to ensure proper unbiased sampling³³. Four pairs of virtual sections were selected at equidistant intervals in one half of each kidney. The distance between the pairs was 31 slices or $50 \mu\text{m}$. To extrapolate the number of counted glomeruli to the whole kidney, both the considered volume and the volume of the entire kidney were calculated based on a mask derived from morphological closing on the vascular segmentation (**D5**).

The vascular and tubular structures do not form individual, countable units. Therefore, the accuracy of their semi-automatic segmentations (**D5**, **D6**) cannot be assessed on a similar per-object-basis. An assessment would have to be performed on each individual voxel instead. This would require voxel-accurate ground truth data with higher precision than can be created, with reasonable workload, using manual annotation of the highly intricate and convoluted vascular and tubular trees. Accordingly, these segmentations have not been validated to be used as ground truth for machine-learning or benchmarking datasets, but are rather supplied as examples of segmentations for other purposes, such as testing image processing, segmentation or vascular analysis methods at the terabyte scale.

Code availability

The HR-Kidney dataset is freely available for download at the Image Data Resource under accession number idr0147: <https://doi.org/10.17867/10000188>.

Received: 23 June 2022; Accepted: 24 July 2023;

Published online: 03 August 2023

References

- Benjamens, S., Dhunoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit. Med.* **3**, 118 (2020).
- Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. *University of Toronto* (2009).
- Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255, <https://doi.org/10.1109/CVPR.2009.5206848> (IEEE, 2009).
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
- Ørting, S. *et al.* A Survey of Crowdsourcing in Medical Image Analysis. *arXiv:1902.09159 [cs]* (2019).
- Tajbakhsh, N. *et al.* Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis* **63**, 101693 (2020).
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (eds. Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W.) vol. 9901 424–432 (Springer International Publishing, 2016).
- Pinto, N., Cox, D. D. & DiCarlo, J. J. Why is Real-World Visual Object Recognition Hard? *PLoS Comput Biol* **4**, e27 (2008).
- Li F-F, Fergus, R. & Perona, P. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. in *2004 Conference on Computer Vision and Pattern Recognition Workshop* 178–178, <https://doi.org/10.1109/CVPR.2004.383> (IEEE, 2004).
- Williams, E. *et al.* Image Data Resource: a bioimage data integration and publication platform. *Nat Methods* **14**, 775–781 (2017).
- Everingham, M. *et al.* The Pascal Visual Object Classes Challenge: A Retrospective. *Int J Comput Vis* **111**, 98–136 (2015).
- Litjens, G. *et al.* Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Medical Image Analysis* **18**, 359–373 (2014).
- Savva, M. *et al.* Large-Scale 3D Shape Retrieval from ShapeNet Core55. *Eurographics Workshop on 3D Object Retrieval* <https://doi.org/10.2312/3DOR.20171050> (2017).
- Allen, J. L. *et al.* HuBMAP ‘Hacking the Kidney’ 2020–2021 Kaggle Competition Dataset - Glomerulus Segmentation on Periodic Acid-Schiff Whole Slide Images. *The Human BioMolecular Atlas Program (HuBMAP)* <https://doi.org/10.35079/HBM925.SGXL.596> (2021).
- Sekuboyina, A. *et al.* VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images. *Medical Image Analysis* **73**, 102166 (2021).
- Baid, U. *et al.* The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. <https://doi.org/10.48550/ARXIV.2107.02314> (2021).
- Hall, C. N. *et al.* Capillary pericytes regulate cerebral blood flow in health and disease. *Nature* **508**, 55–60 (2014).
- Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R. & Wilkins, S. W. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J Microsc* **206**, 33–40 (2002).
- Bruna, J. & Mallat, S. Invariant Scattering Convolution Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1872–1886 (2013).
- Antoniou, A., Storkey, A. & Edwards, H. Data Augmentation Generative Adversarial Networks. *arXiv:1711.04340 [cs, stat]* (2018).
- Zhu, X. & Wu, X. Class Noise vs. Attribute Noise: A Quantitative Study. *Artificial Intelligence Review* **22**, 177–210 (2004).
- Czogalla, J., Schweda, F. & Loffing, J. The Mouse Isolated Perfused Kidney Technique. *Journal of Visualized Experiments* 54712 <https://doi.org/10.3791/54712> (2016).
- Kyrieleis, A., Ibison, M., Titarenko, V. & Withers, P. J. Image stitching strategies for tomographic imaging of large objects at high resolution at synchrotron sources. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **607**, 677–684 (2009).
- Mirone, A., Brun, E., Gouillart, E., Tafforeau, P. & Kieffer, J. The PyHST2 hybrid distributed code for high speed tomographic reconstruction with iterative reconstruction and a priori knowledge capabilities. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **324**, 41–48 (2014).
- Rodgers, G. *et al.* Optimizing contrast and spatial resolution in hard x-ray tomography of medically relevant tissues. *Appl. Phys. Lett.* **116**, 023702 (2020).
- Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156 (2008).
- Candès, E. J. & Donoho, D. L. Continuous curvelet transform. *Applied and Computational Harmonic Analysis* **19**, 198–222 (2005).
- Starck, J.-L., Murtagh, F., Candès, E. J. & Donoho, D. L. Gray and color image contrast enhancement by the curvelet transform. *IEEE Trans. on Image Process.* **12**, 706–717 (2003).
- Frigo, M. & Johnson, S. G. The Design and Implementation of FFTW3. *Proceedings of the IEEE* **93**, 216–231 (2005).
- Cohen, T. S. & Welling, M. Steerable CNNs. *arXiv:1612.08498 [cs, stat]* (2016).
- Weiler, M., Geiger, M., Welling, M., Boomsma, W. & Cohen, T. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. *arXiv:1807.02547 [cs, stat]* (2018).
- Kuo, W. *et al.* Terabyte-scale supervised 3D training and benchmarking dataset of the mouse kidney. *Image Data Resource (IDR)* <https://doi.org/10.17867/10000188> (2023).
- Gundersen, H. J. Stereology of arbitrary particles. A review of unbiased number and size estimators and the presentation of some new ones, in memory of William R. Thompson. *J Microsc* **143**, 3–45 (1986).

Acknowledgements

We gratefully acknowledge access to and support at the Swiss National Supercomputing Centre (Maria Grazia Giuffreda) and the Pawsey Supercomputing Centre (Ugo Varetto, David Schibeci). We acknowledge the European Synchrotron Radiation Facility (ESRF) for provision of synchrotron radiation facilities under proposal numbers MD-1014 and MD-1017 and we would like to thank Alexander Rack and Margie Olbinado for assistance in using beamline ID19. We wish to thank Axel Lang, Erich Meyer, Anastasios Marmaras and Virginia Meskenaite for help with vascular casting. Jan Czogalla for help with establishing the perfusion surgery. Gilles Fourestey for support in high performance computing, Lucid Concepts AG for collaboration in software development. Michael Bergdorf for help with scattering transform schemes. We are grateful for the financial support provided by the Swiss National Science Foundation (grant 205321_153523, to V.K., S.H. and B.M.) and NCCR Kidney.CH (grant 183774, to V.K. and R.H.W.).

Author contributions

W.K. developed the vascular resin mixture and surgery protocol, and performed SR μ CT imaging, reconstructions, manual annotation, validation, data curation and analysis. D.R. developed the pipeline and performed the processing for artefact-correction, segmentations, machine learning, path calculations, visualizations and data analysis. W.K. and D.R. drafted the manuscript. G.S. and S.H. performed SR μ CT imaging and designed scanning parameters. R.H.W., S.H., B.M. and V.K. conceived and designed the research, and edited and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02407-5>.

Correspondence and requests for materials should be addressed to V.K.

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023