nature machine intelligence

SUPPLEMENTARY INFORMATION

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

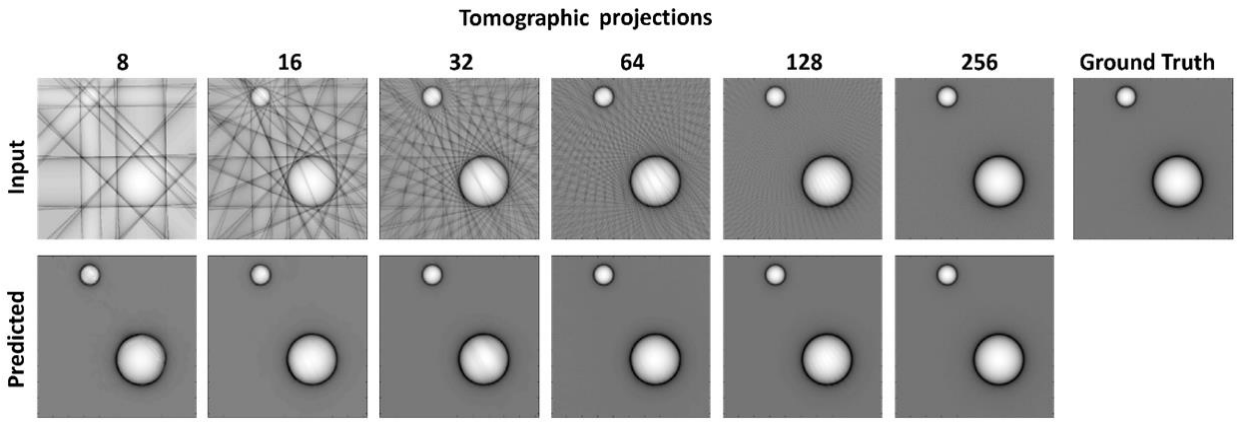
In the format provided by the authors and unedited.

Deep learning optoacoustic tomography with sparse data

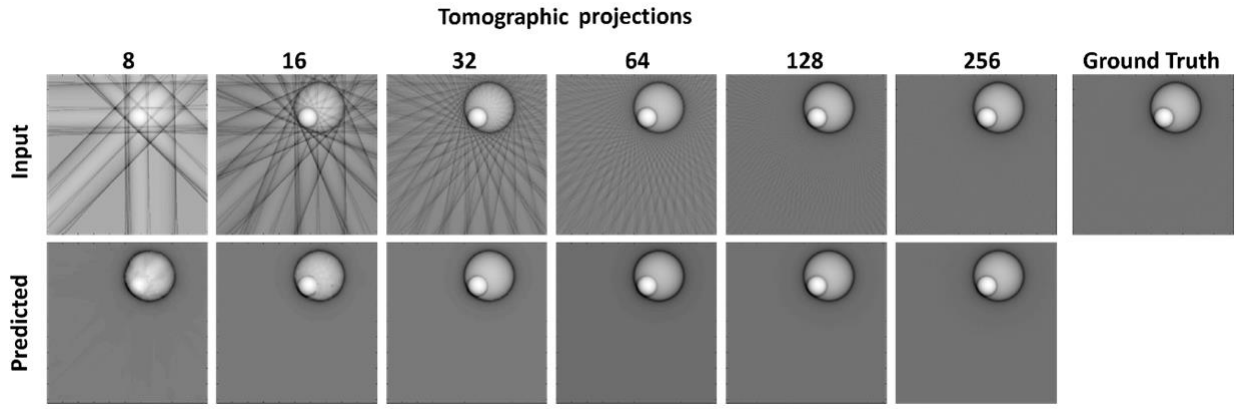
Neda Davoudi[1,2], Xosé Luís Deán-Ben[1,3] and Daniel Razansky[1,3]*

[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

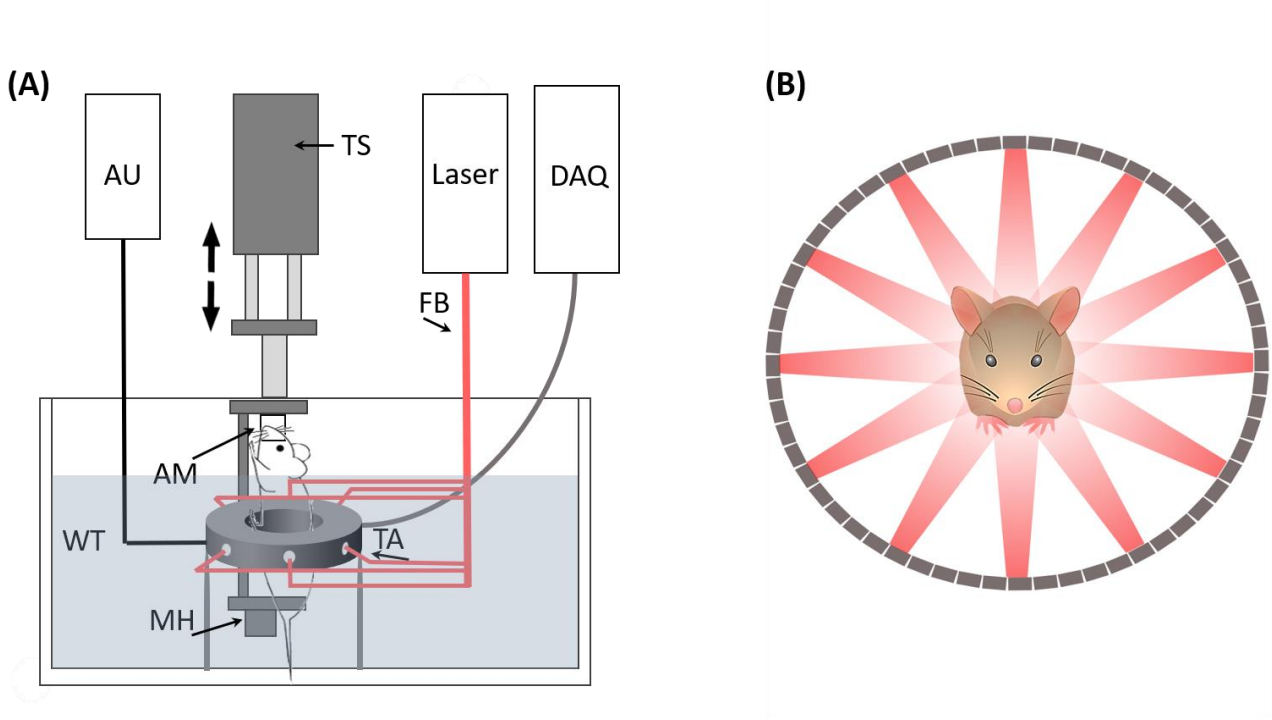
SUPPLEMENTARY FIGURES



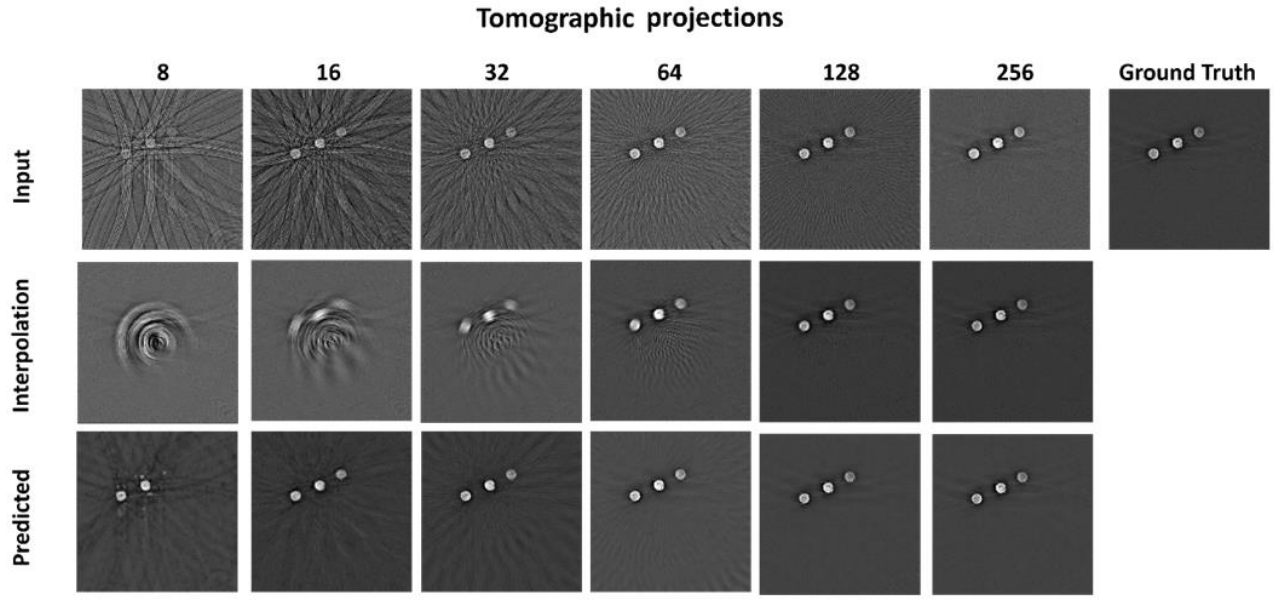
Supplementary Figure 1 | Deep learning network performance on simulated data with two separated absorbers – recovery of the original image quality from under-sampled tomographic OA data.



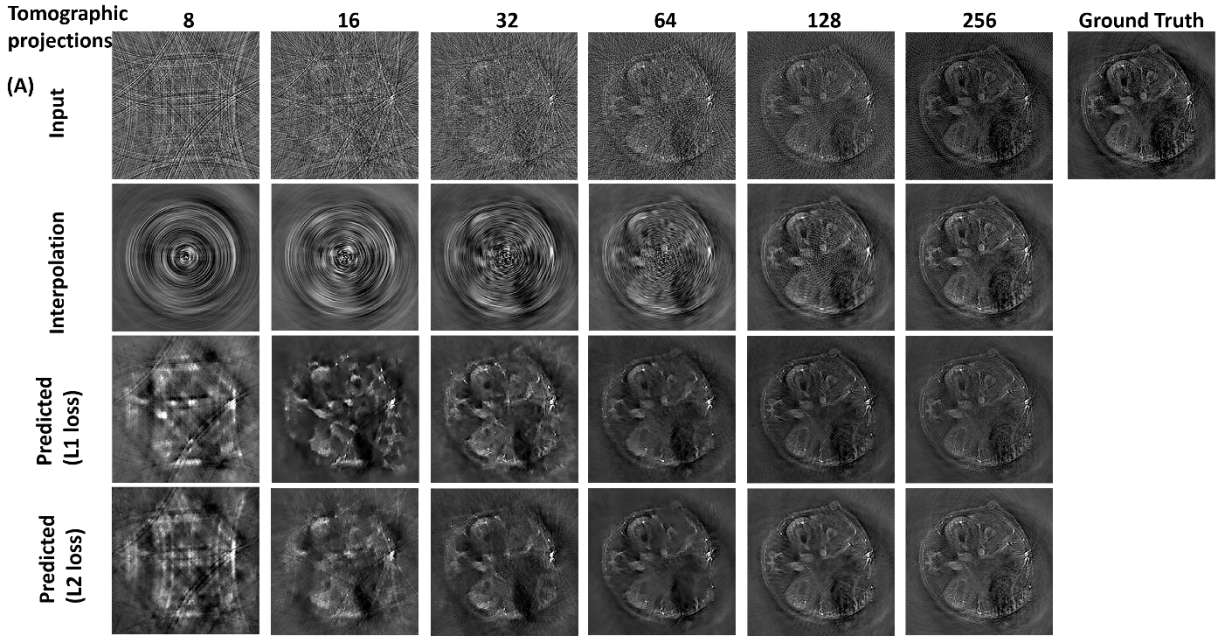
Supplementary Figure 2 | Deep learning network performance on simulated data with two overlapping absorbers – recovery of the original image quality from under-sampled tomographic optoacoustic data.



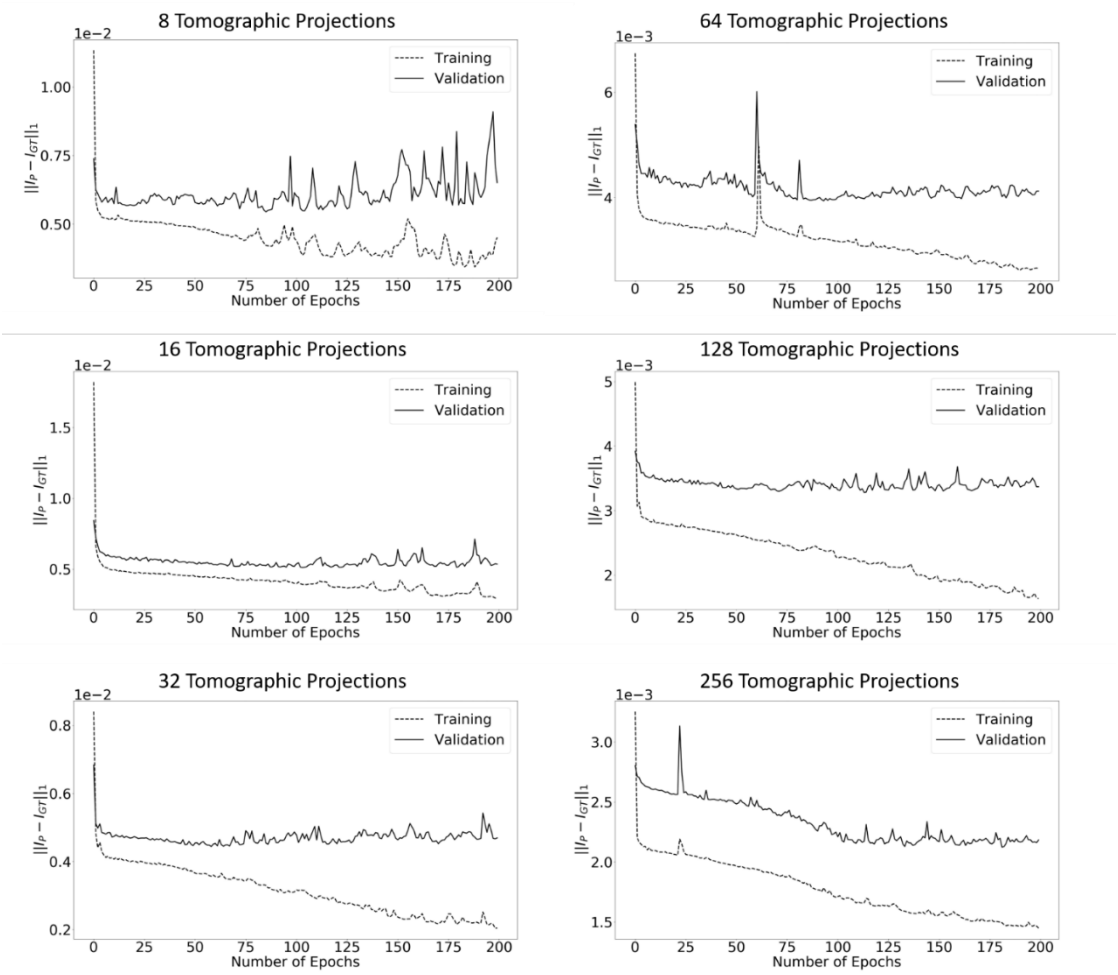
Supplementary Figure 3 | The newly developed whole-body mouse imaging scanner. (**A**) Lay-out of the complete full-ring array imaging setup. AU – Anesthesia unit, TS – Translation stage, DAQ- Data acquisition, FB – Fiber bundle, AM- Anesthesia mask, WT – Water tank, MH – Mouse holder, TA – Transducer array. (**B**) The cross-sectional (circumferential) light illumination configuration with 12-arm fiber bundle.



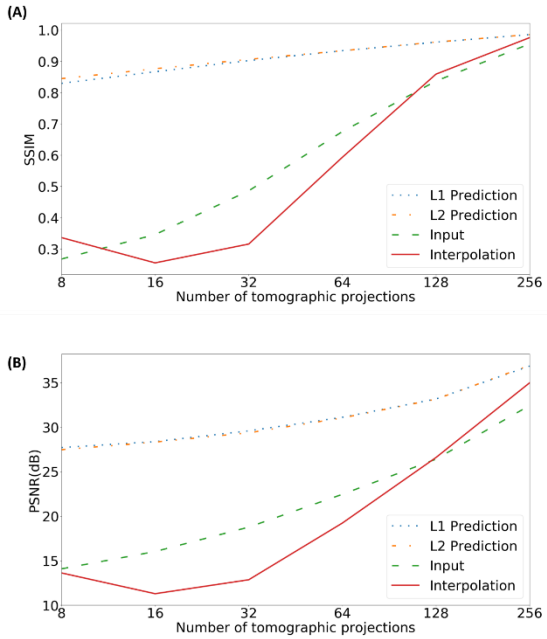
Supplementary Figure 4 | Deep learning network performance on experimental phantom data. The results of image quality recovery are further compared to images reconstructed with interpolated tomographic data.



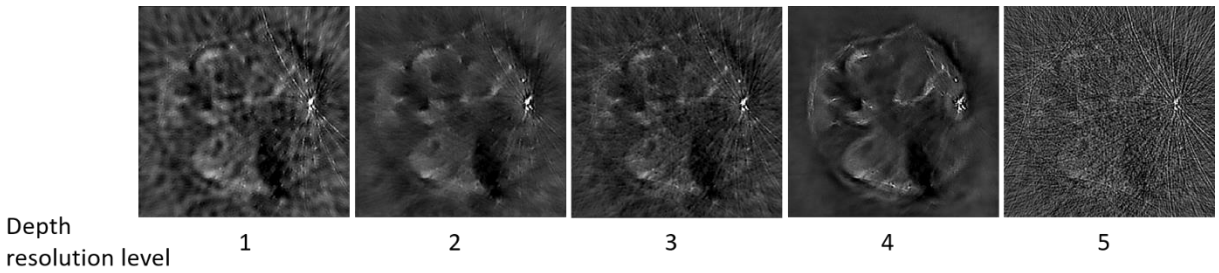
Supplementary Figure 5 | Qualitative comparison of network performance on the images recovered from under-sampled data using U-Net prediction with L1 and L2 loss function versus interpolation.



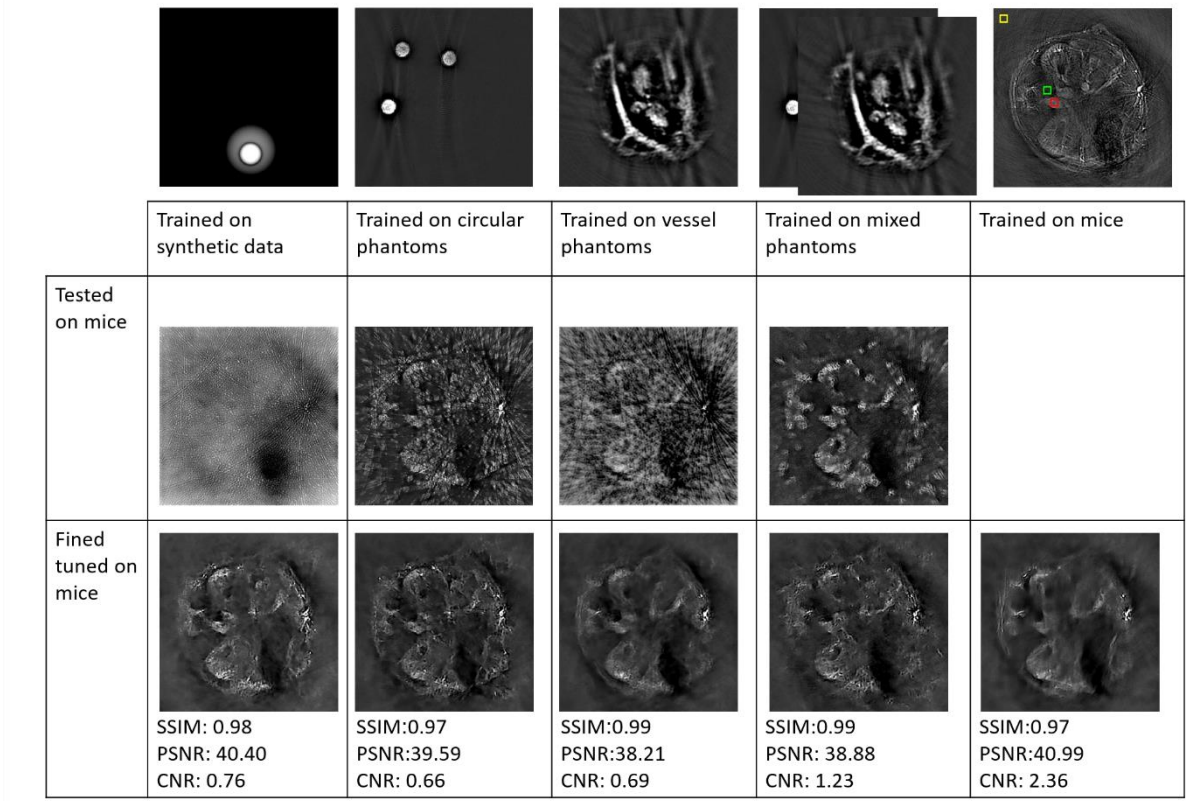
Supplementary Figure 6 | Quantification of network performance with the *in vivo* mouse images based on L1 norm of the error vs. number of epochs for different number of tomographic projections. I_P - predicted image; I_{GT} – ground truth image.



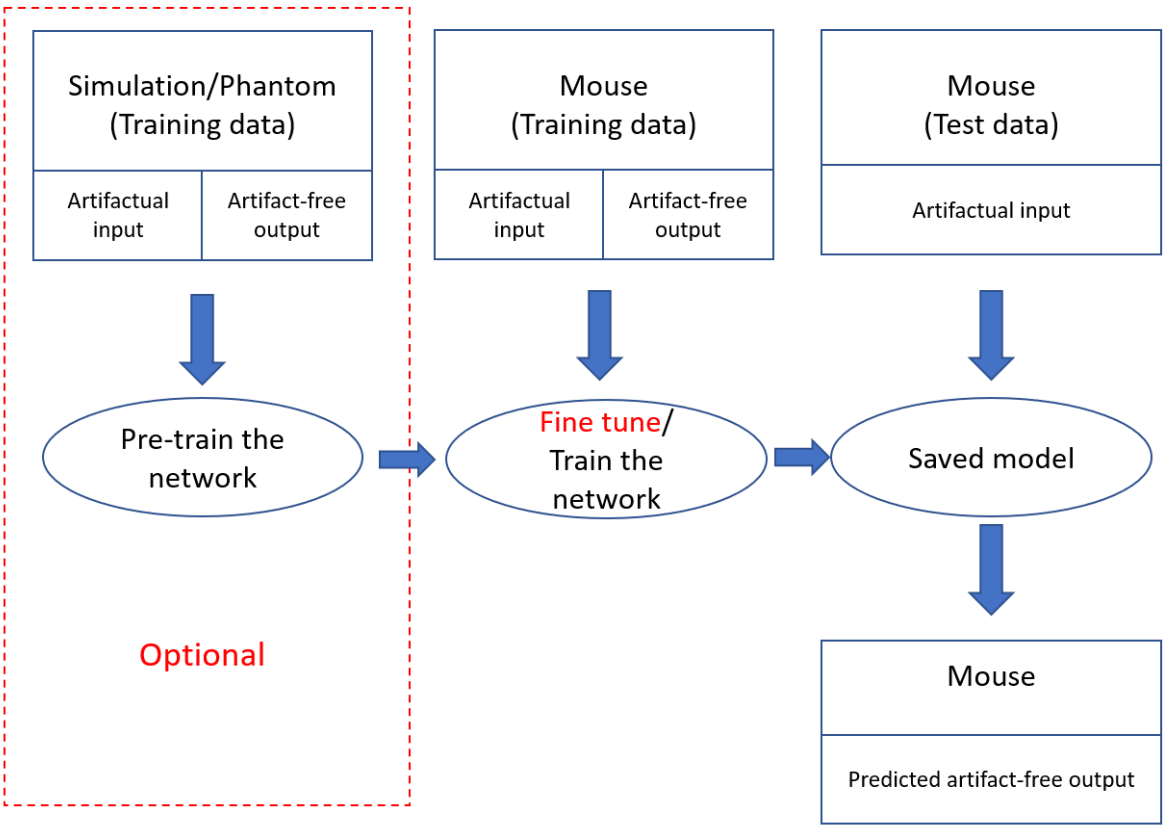
Supplementary Figure 7 | Numerical comparison of ground truth image reconstructed with full (512) transducers and sparsely reconstructed with fraction of transducers and recovered by signal interpolation and deep learning. (A) Comparison based on structural similarity index (SSIM). (B) Comparison based on peak signal-to-noise ratio (PSNR).



Supplementary Figure 8 | Example of network performance for different depth-resolution levels (defined as number of down-sampling and up-sampling layers in U-Net) for an image reconstructed with 32 projections considering optimal number of epochs for training.



Supplementary Figure 9 | Qualitative and quantitative (SSIM, PSNR, CNR) comparison of network performance with different training data and strategies. The red, green, and yellow rectangles in the reference mouse image refer to structure, background, and noise area respectively for calculating the CNR.



Supplementary Figure 10 | Flow diagram of different training strategies for the network.

Supplementary Video 1 | Fly-through of all cross-sectional images of the mouse reconstructed with 128 channels (artifactual input of the network) and with all 512 channels (reference image) along with the predicted output of the network.

Supplementary Video 2 | Fly-through of all cross-sectional images of the mouse reconstructed with 32 channels (artifactual input of the network) and with all 512 channels (reference image) along with the predicted output of the network.

Supplementary Video 3 | Cross-sectional images for 4 respiration cycles reconstructed with 128 channels (artifactual input of the network) and with all 512 channels (reference image) along with the predicted output of the network.

Supplementary Video 4 | Cross-sectional images for 4 respiration cycles reconstructed with 32 channels (artifactual input of the network) and with all 512 channels (reference image) along with the predicted output of the network.