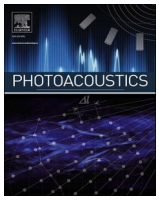
Photoacoustics 42 (2025) 100687

Contents lists available at ScienceDirect

Photoacoustics

journal homepage: www.elsevier.com/locate/pacs



Structure and oxygen saturation recovery of sparse photoacoustic microscopy images by deep learning

Shuyan Zhang [a,b], Jingtan Li [a,b], Lin Shen [a,b], Zhonghao Zhao [a,b], Minjun Lee [a,c], Kun Qian [a,b,*], Naidi Sun [a,b,*], Bin Hu [a,b,*]

[a] Key Laboratory of Brain Health Intelligent Evaluation and Intervention (Beijing Institute of Technology), Ministry of Education, China
[b] School of Medical Technology, Beijing Institute of Technology, China
[c] School of Computer Science and Technology, Beijing Institute of Technology, China

ARTICLE INFO

Keywords:
Deep learning
Photoacoustic microscopy
Sparse data
High-speed imaging
Image reconstruction

ABSTRACT

Photoacoustic microscopy (PAM) leverages the photoacoustic effect to provide high-resolution structural and functional imaging. However, achieving high-speed imaging with high spatial resolution remains challenging. To address this, undersampling and deep learning have emerged as common techniques to enhance imaging speed. Yet, existing methods rarely achieve effective recovery of functional images. In this study, we propose Mask-enhanced U-net (MeU-net) for recovering sparsely sampled PAM structural and functional images. The model utilizes dual-channel input, processing photoacoustic data from 532 nm and 558 nm wavelengths. Additionally, we introduce an adaptive vascular attention mask module that focuses on vascular information recovery and design a vessel-specific loss function to enhance restoration accuracy. We simulate data from mouse brain and ear imaging under various levels of sparsity (4 ×, 8 ×, 12 ×) and conduct extensive experiments. The results demonstrate that MeU-net significantly outperforms traditional interpolation methods and other representative models in structural information and oxygen saturation recovery.

1. Introduction

As an emerging hybrid imaging technology, photoacoustic microscopy (PAM) which constructs images based on the detection of acoustic waves generated by laser excitation through the photoacoustic effect, has been made a powerful tool with extensive applications in the biomedical and clinical fields, including morphological, functional, and molecular imaging for early diagnosis and treatment monitoring [1,2]. Based on differences in optical and acoustic focusing, PAM is further classified into optical resolution PAM (OR-PAM) and acoustic resolution PAM (AR-PAM) [3]. OR-PAM, in particular, has seen rapid development in various biomedical applications due to its high spatial resolution in both anatomical and functional imaging, such as tumor and brain diseases monitoring [4–7]. However, OR-PAM typically employs point-by-point scanning, which requires smaller step sizes to achieve high-resolution images, which creates a fundamental trade-off between imaging speed and spatial resolution [8]. Researchers usually need to reduce the spatial sampling density to increase the imaging speed for larger fields of view, which compromises resolution [9–11].

In recent years, several strategies, including high-repetition-rate lasers, supercontinuum sources, Raman crystals, and advanced scanning devices, such as high-speed voice coil stage, galvanometer, and polygon mirror scanners, have been employed to enhance PAM imaging speed and efficiency [8,11–14]. However, due to the Nyquist sampling theorem, which dictates that the scanning step size should not exceed half of the expected spatial resolution [15], the B-scan speed of high-speed PAM systems remains constrained in practical applications. It is noteworthy that the aforementioned strategies rely on complex and expensive hardware systems [16]. Consequently, an increasing number of studies have focused on reconstructing photoacoustic images from limited-view and undersampled data to achieve imaging quality comparable to full spatial sampling [17–19]. For instance, some studies have employed dictionary learning methods to recover signals from undersampled PAM data, achieving notable recovery performance [20]. With the rapid advancement of deep learning, an increasing number of deep learning techniques have been applied to photoacoustic imaging, including the reconstruction of degraded photoacoustic images and the acceleration of image reconstruction processes [21–28].

* Correspondence to: School of Medical Technology, Beijing Institute of Technology, No. 5 Zhongguancun South Street, Haidian District, Beijing 100081, China.
E-mail addresses: qian@bit.edu.cn (K. Qian), sunnaidi@bit.edu.cn (N. Sun), bh@bit.edu.cn (B. Hu).

https://doi.org/10.1016/j.pacs.2025.100687
Received 30 October 2024; Received in revised form 24 December 2024; Accepted 8 January 2025
Available online 9 January 2025

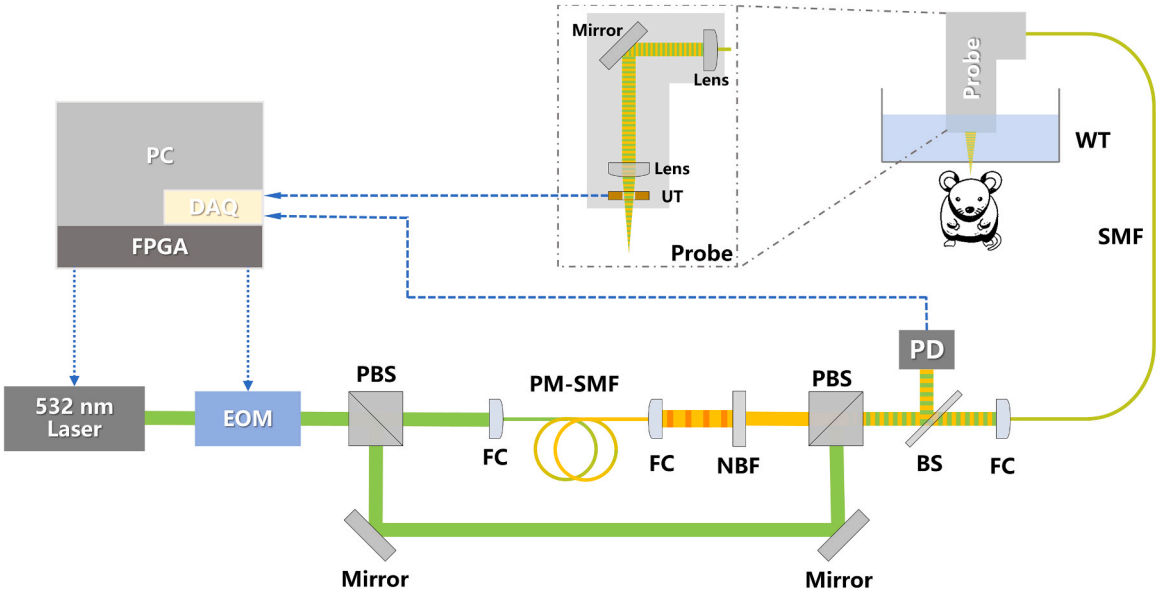


Fig. 1. Schematic of the multi-parametric PAM system. EOM, electro-optical modulator; PBS, polarizing beam splitter; PM-SMF, polarization-maintaining single-mode fiber; PD, photodiode; NBF, narrow bandpass filter; BS, beam sampler; FC, fiber coupler; SMF, single-mode fiber; UT, ultrasonic transducer; WT, water tank; PC, personal computer; DAQ, data acquisition card; FPGA, field-programmable gate array.

Moreover, numerous studies have concentrated on the recovery of undersampled PAM images, aiming to enhance imaging efficiency while maintaining high-quality outputs [29–34]. For example, DiSpirito *et al.* proposed an enhanced supervised Fully Dense U-net (FD U-net) model for recovering undersampled PAM images, demonstrating that their approach outperformed the standard U-net and other related models using only 2 % of the original pixel data [29]. Zhou *et al.* developed a novel convolutional neural network based approach for handling 1/4 and 1/16 sparse-sampling images, incorporating residual blocks, squeeze-and-excitation (SE) blocks, and perceptual loss functions. Their methods showed significant improvements in both quantitative and qualitative evaluations [30]. Besides, Liu et al. introduced Unified PhotoAcoustic Microscopy image reconstruction Network (UPAMNet) for PAM image super-resolution and denoising, demonstrating superior performance in sparse image reconstruction and noise reduction across various datasets [33]. Additionally, Vu et al. employed a deep image prior (DIP) approach to enhance the reconstruction of undersampled PAM images in an unsupervised manner, without the need for pre-training [31]. Similarly, Loc et al. introduced the DiffPam model, an unsupervised diffusion model that achieved performance comparable to U-net for undersampled PAM images [34]. Although existing studies demonstrate high accuracy in structural metrics such as structural similarity index (SSIM) and Peak Signal-to-Noise Ratio (PSNR), they have yet to address the calculation of functional parameters, such as hemoglobin oxygen saturation (sO_2) , which is crucial for cancer screening, diagnosis, and research into brain function and vascular diseases [6,35].

In view of this, this study proposes Mask-enhanced U-net (MeU-net), an improved U-net-based model, designed to improve the recovery of the structural and functional parameter sO_2 from sparse undersampled PAM data. It is important to note that unless otherwise specified, PAM refers to OR-PAM in the following text. Specifically, MeU-net leverages dual-channel input data consisting of photoacoustic signals acquired at 532 nm and 558 nm wavelengths. Additionally, an adaptive vascular attention mask module is incorporated to enhance the model's focus on vascular features. To further refine the reconstruction quality in vascular regions, we design a specialized loss function targeting these areas. Other improvements and detailed descriptions of the model will be further elaborated in the Methods section. To evaluate the performance of our model, we applied MeU-net to the recovery of simulated undersampled PAM images of mouse brain and ear vasculature and compared the results with those obtained using classical methods and models such as interpolation, Res U-net [36], and FD U-net. The findings revealed that MeU-net not only exhibited superior capabilities in recovering undersampled PAM images but also demonstrated exceptional performance in sO_2 estimation. These results underscore MeU-net's potential as a robust tool for enhancing the imaging speed of PAM in biomedical applications, particularly in scenarios involving sparse data acquisition.

2. Methods

2.1. PAM system and data acquirement

The schematic of our used multi-parametric PAM system is shown in Fig. 1. The 532 nm laser beam from the nano-second pulsed laser (BX40-2-G, EdgeWave) passes through an electro-optical modulator (EOM; 350-80, Conoptics), where it is modulated into different polarization states. They are then separated into two distinct paths through a polarizing beam splitter (PBS; PBS121, Thorlabs). In the first path, the beam is coupled into a 10-meter-long single-mode polarization-maintaining fiber (PM-SMF; F-SPA, Newport), where Raman scattering induces the generation of new wavelengths. The resulting light exits through another fiber coupler and is directed into a narrow bandpass filter (NBF; FB560-10, Thorlabs) to isolate the 558 nm or 559 nm wavelength. And the second path beam is reflected by two mirrors and then recombined with the first beam through a PBS. To compensate for the fluctuation in laser intensity, the beam is sampled by a beam sampler (BS; BSF10-A, Thorlabs), with approximately 5 % energy being detected by a photodiode (PD; FDS100, Thorlabs) and recorded by a data acquisition card (DAQ). And the remaining 95 % is coupled into a single mode fiber (SMF; P1-460B-FC-2, Thorlabs) via a fiber coupler and directed to the scanning probe, which consists of a mirror, a pair of convex lenses (AC127-025-A, Thorlabs), and a ring-shaped ultrasonic transducer (UT; inner diameter: 2.2 mm; outer diameter: 4.0 mm; center frequency: 35 MHz; 6-dB bandwidth: 70 %). It is mounted on an X-Y stage driven by two high-precision stepper motors to enable scanning motion. The resulting photoacoustic signals are received by the UST and stored by the DAQ at a sampling rate of 500 Ms/s. In this system, the stepper motors, DAQ, laser, and EOM are all controlled by a computer and field-programmable gate array (FPGA).

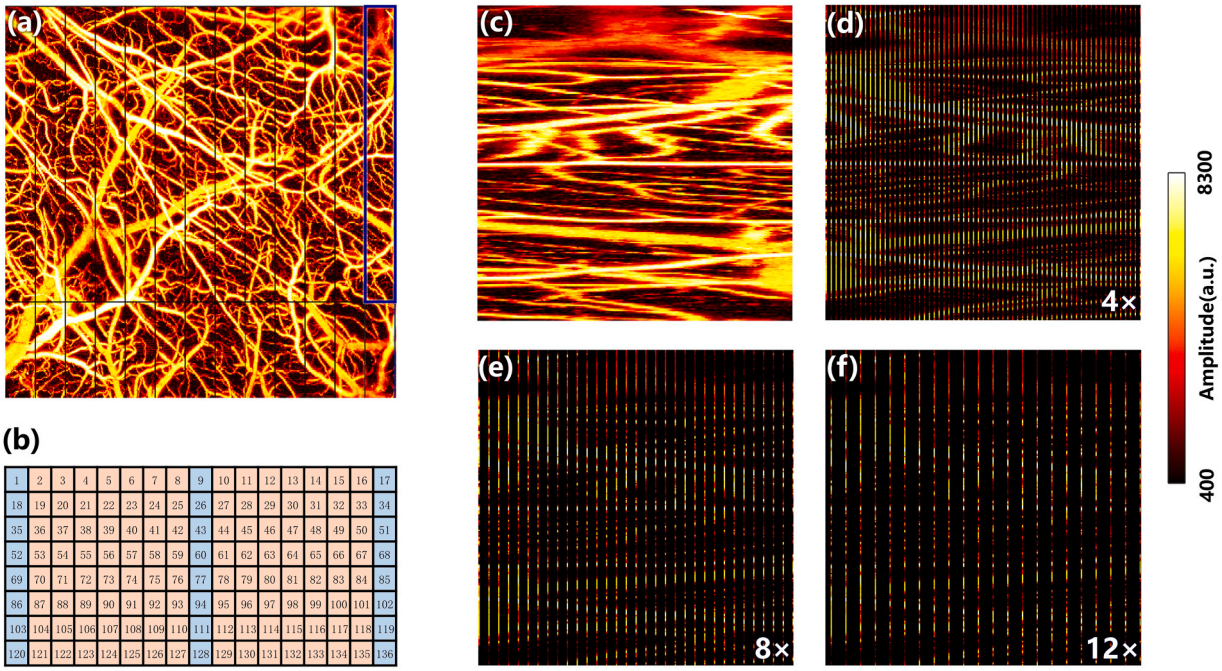


Fig. 2. Schematic diagram of data preparation and preprocessing. (a) Corrected original image and cropped example. (b) Distribution of data points after raster scanning correction and zero filling. The blue blocks represent the remaining points after sparseness, and the orange blocks represent the points filled with zeros at 8 × sparsity. (c) The blue box in (a) is the original cropped data. (d, e, f) are 4 × , 8 × , and 12 × undersampling images, respectively.

2.2. Animal preparation

To validate the accuracy of our model in reconstructing structural and functional images for practical applications, we intravitally imaged mouse brains and ears using the PAM system to build our dataset. The experiments were performed on CD-1 mice (male, 6–10 weeks old, Charles River Laboratories) under general anesthesia (isoflurane: 2 % for induction, 1.5 % for maintenance). And body temperature was maintained at 37°C using a heating pad (model DCT-15, Kent Scientific). The details of mouse cranial window preparation, brain imaging, and ear imaging followed the same protocol described in previous studies [37–39]. All experimental procedures were conducted in accordance with the protocols approved by the Animal Care and Use Committee at the University of Virginia.

2.3. Data preparation

In this work, we acquire photoacoustic images corresponding to both 532 nm and 558 nm wavelengths using the aforementioned system. In order to optimize the data for input our training model, preprocessing is required. As shown in Fig. 2, the necessary steps are outlined, taking the 532 nm photoacoustic image as an example. Specifically, the original intensity data in the acquired image exhibits inversions in the even-numbered columns, which is a result of the raster scanning pattern employed in this study. Therefore, prior to further processing, the raw data must be corrected by flipping the inverted columns. Following this correction, a complete PAM image is obtained. The changes in the peak intensity map and the layout of the raster scanning points after this correction are illustrated in Fig. 2(a). However, the image is evidently too large for direct use in the network model. To address this, we apply a sliding window of 256 × 256 pixels, partitioning the large image into multiple standard sub-images, as shown in Figs. 2(a)-1. For the excess part of the edge, we use the same amount of data from the other end to fill it so that images of different sizes can be completely divided into 256 × 256 pixels. Finally, these standard sub-images, derived from various full-sampled images, are randomly shuffled to create the dataset for model.

Through the preceding steps, we obtain full-sampled standard sub-images at 532 nm and 558 nm, upon which undersampling is performed to generate sparse data. According to previous studies, in deep learning, zero-padding is known to outperform interpolation methods [29]. Therefore, we employ zero-padding throughout this work to adjust the size of the undersampled input images. Specifically, we performed undersampling at scales of 4 × , 8 × , and 12 × , ensuring that the undersampled images retained the same dimensions as the original images by applying zero-padding, as illustrated in Figs. 2(a)-2. This process involved selecting one pixel for every 4, 8, or 12 pixels, with the blue blocks in the figure representing the retained pixel values during undersampling, while the orange blocks indicate regions where zeros were used for padding. Additionally, given the significant disparity between the lateral and axial resolutions (1.042 μm and 8.333 μm) in our data, undersampling is applied only along the later dimension. Fig. 2(b) illustrates examples of fully sampled images, along with three undersampled versions. Since ensuring the numerical accuracy of the images is critical to our work, all operations are conducted directly on the raw data without any enhancement of the zero-padded images. At the same time, we use the float 64 digital format in the entire process of normalization, model calculation and inference. Additionally, unlike other studies focused on undersampling, our model utilizes a dual-channel input, incorporating both the input data and the target simultaneously. Specifically, we train the model using photoacoustic data at two wavelengths, 532 nm and 558 nm. The undersampled PAM images serve as the input, while the corresponding fully sampled images are used as the target during training.

2.4. Network and settings

The proposed MeU-net is an enhanced model based on U-net, specifically designed for the reconstruction of undersampled images. The architecture of MeU-net is illustrated in Fig. 3.

In detail, two-channel undersampled PAM images are fed into the model, where they first undergo threshold segmentation to obtain corresponding binary images (i.e. two-channel vascular masks). These masks are then processed through separate encoders. In the encoder part

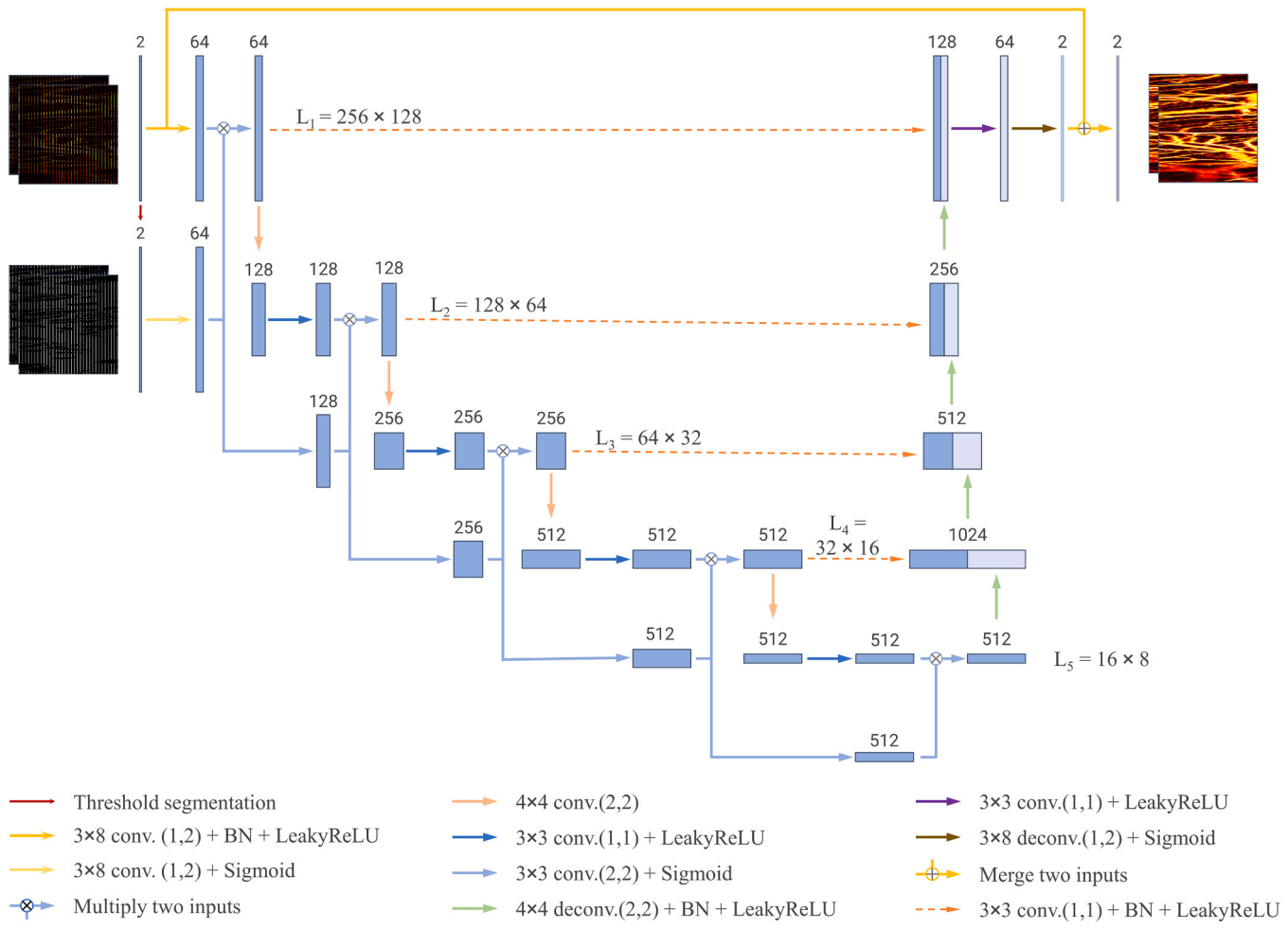


Fig. 3. The Architecture of MeU-net. On the left side of the figure, the two pictures represent, from top to bottom, the original undersampled PAM image and the binary image generated through threshold segmentation of the undersampled image, respectively. The rightmost picture represents the image generated by the network.

of the U-net architecture, we replaced the original max-pooling layers used for undersampling with convolutional layers. This modification allows for feature extraction while performing undersampling. The binary masks are also passed through convolutional layers to maintain the same number of channels and size as the feature maps. And then these masks are treated as adaptive weights and multiplied with the feature maps for integration, whose result serves as the input to the next undersampling convolution module, which constitutes the mask module we designed. This approach implements adaptive vascular attention, enhancing the expressiveness of feature representation. And We have visualized this module, as shown in Supplementary Fig. S1. Additionally, due to the zero-padding used for the missing values in the undersampled images, using conventional 3×3 or 4×4 convolution kernels will cause the model to generate a large number of zero values in the convolution operation and fail to extract valid information. To address this, we replaced the first layer of convolution kernels in the encoder with 3×8 kernels, enabling the model to capture more information along the x-axis of the image.

During decoding, the model employs transposed convolution for upsampling, gradually restoring the feature maps to higher resolutions. After each upsampling step, the model integrates the previously extracted high-resolution feature maps from the encoder with the upsampled feature maps through skip connections, which include additional convolution operations. Compared to simple skip connections, the convolutional operations further refine or compress the features, highlighting critical information and suppressing redundant features. They also balance the weight information from preceding modules, preventing excessive interference from the mask module on the features. This enhances gradient propagation in deep networks, facilitates convergence, and improves overall performance. These skip connections ensure that high-resolution information is not lost during the encoding process, thereby allowing the model to restore fine image details effectively. Moreover, given that the raw undersampled PAM images still retain some ground truth values, we merged the raw undersampled PAM images with the model's output. These improvements enable the model to accurately recover the structural and functional parameters of the input images while maintaining high resolution.

In our experiments, the proposed MeU-net model was implemented using Python 3.8.10 in PyTorch 2.0.0. During training, the Adam optimizer with a batch size of 16 was employed. The operating system was Ubuntu 20.04, and the workstation was equipped with an Intel Xeon(R) Platinum 8474 C CPU and an NVIDIA RTX 4090D GPU.

2.5. Loss function and evaluation indicators

For the loss function in our model, we employed a combination of mean absolute error (MAE) and SSIM between the ground truth image $\boldsymbol{x}$ and the generated image $\boldsymbol{y}$. The formulas for these loss functions are as follows:

$$L_{MAE} = \frac{1}{mn} \sum_{i=0}^{m} \sum_{j=0}^{n} |\boldsymbol{x}(i,j) - \boldsymbol{y}(i,j)| \tag{1}$$

Table 1
Experiment results of mouse brain data.

Sparsity	Image type	*Evaluation indicators*	Cubic interpolation	Res U-net	FD U-net	UPAMNet	SEResnet	MeU-net w/o L_{vessel}	MeU-net w/o mask module	MeU-net
4 ×	MAP	***SSIM***	0.8498	0.9614	0.9620	0.9625	0.9648	0.9643	**0.9651**	0.9630
		SSIM$_v$	0.9087	0.9351	0.9342	0.9353	0.9383	0.9369	0.9401	**0.9409**
		PSNR	28.1177	33.8489	33.5510	33.5709	33.7577	33.5394	33.6874	**33.7596**
		PSNR$_v$	24.7297	**29.4038**	29.0993	29.0774	29.3039	29.0270	29.2154	29.3475
		MSE	0.00161	**0.00044**	0.00047	0.00047	0.00046	0.00049	0.00047	0.00046
		MSE$_v$	0.00360	**0.00126**	0.00132	0.00134	0.00131	0.00140	0.00134	0.00130
	sO_2	***SSIM***	0.8584	0.8464	0.8364	0.8763	0.8616	0.8670	0.8654	**0.8800**
		SSIM$_v$	0.8525	0.8616	0.8331	0.8760	0.8606	0.8595	0.8735	**0.8823**
		PSNR	**26.4198**	23.3679	21.2511	25.5380	24.3366	25.1145	25.4008	26.3024
		PSNR$_v$	26.5097	24.5518	21.1973	25.1194	23.4818	25.1941	25.9330	**26.6561**
		MSE	**0.00233**	0.00500	0.00812	0.00295	0.00387	0.00318	0.00297	0.00240
		MSE$_v$	0.00230	0.00384	0.00855	0.00339	0.00496	0.00318	0.00273	**0.00228**
8 ×	MAP	***SSIM***	0.8210	0.9094	0.9321	0.9170	0.9478	0.9485	**0.9497**	0.9472
		SSIM$_v$	0.8697	0.8980	0.8907	0.9143	0.9113	0.9151	**0.9170**	0.9159
		PSNR	26.0589	30.8110	31.4679	31.2188	31.2700	**31.9446**	31.8736	31.8715
		PSNR$_v$	22.2264	26.9952	26.9755	27.4043	26.7507	**27.5326**	27.4806	27.5059
		MSE	0.00261	0.00088	0.00074	0.00079	0.00082	**0.00069**	0.00070	0.00070
		MSE$_v$	0.00644	0.00213	0.00210	0.00197	0.00237	**0.00194**	0.00197	0.00195
	sO_2	***SSIM***	0.7560	0.7735	0.6975	0.7966	0.7905	0.7936	0.8012	**0.8139**
		SSIM$_v$	0.7529	0.7838	0.6939	0.8217	0.7971	0.8129	0.8178	**0.8244**
		PSNR	23.1734	22.2772	18.7805	21.9916	21.7597	22.5667	22.6627	**23.6622**
		PSNR$_v$	23.3069	22.1184	17.5452	22.5285	20.9593	23.1216	23.4075	**23.9577**
		MSE	0.00489	0.00614	0.01413	0.00705	0.00712	0.00573	0.00555	**0.00444**
		MSE$_v$	0.00480	0.00657	0.01964	0.00626	0.00878	0.00503	0.00479	**0.00421**
12 ×	MAP	***SSIM***	0.7404	0.9181	0.9232	0.9054	0.9256	0.9233	**0.9337**	0.9303
		SSIM$_v$	0.7796	0.8856	0.8744	0.8656	0.8821	0.8797	**0.8958**	0.8944
		PSNR	23.2103	30.3884	30.5178	29.6225	29.5278	30.1750	**31.1834**	30.5548
		PSNR$_v$	19.5246	26.2917	26.2602	25.4910	25.2421	25.7860	**26.8376**	26.3527
		MSE	0.00502	0.00097	0.00092	0.00114	0.00123	0.00104	**0.00079**	0.00095
		MSE$_v$	0.01182	0.00248	0.00245	0.00300	0.00332	0.00286	**0.00214**	0.00250
	sO_2	***SSIM***	0.6778	0.7085	0.5494	0.6712	0.7021	0.7258	0.7367	**0.7435**
		SSIM$_v$	0.6802	0.7350	0.5657	0.7142	0.7179	0.7289	0.7536	**0.7586**
		PSNR	21.3625	19.9245	15.2125	17.6024	19.1368	19.6526	21.2778	**21.6534**
		PSNR$_v$	21.6189	20.0388	14.8981	18.0880	17.9004	19.2184	21.2867	**21.8405**
		MSE	0.00738	0.01089	0.03225	0.01865	0.01312	0.01146	0.00772	**0.00701**
		MSE$_v$	0.00701	0.01102	0.03715	0.01747	0.01813	0.01304	0.00807	**0.00685**

$$L_{SSIM} = 1 - \frac{1}{N}\sum_{q=1}^{N} SSIM\left(\boldsymbol{x}_q, \boldsymbol{y}_q\right) \tag{2}$$

where N is the total number of pixels in the image, and $\boldsymbol{x}_q, \boldsymbol{y}_q$ represent the patches of fixed size centered at the q-th pixel of the ground truth and generated images, respectively. To further enhance the model's ability to focus on vascular regions, we introduced an additional vessel-specific loss function, L_{vessel}, which is defined as:

$$L_{vessel} = \frac{1}{P}\sum_{p=1}^{P} \frac{1}{\left\|\boldsymbol{mask}_p\right\|_0} \sum_{\boldsymbol{mask}_p=1} \left|\boldsymbol{mask}_p - \boldsymbol{SSIM}(\boldsymbol{x}(p), \boldsymbol{y}(p))\right| \tag{3}$$

where P is the number of channels in the image, and $\boldsymbol{x}(p), \boldsymbol{y}(p)$ denote the ground truth and generated images in the p-th channel, respectively. The term $\boldsymbol{mask}_p$ is used to emphasize the vascular regions, and $\left\|\boldsymbol{mask}_p\right\|_0$ represents the number of non-zero elements in the mask. The total loss function for the model is expressed as:

$$L_{total} = \alpha L_{MAE} + \beta L_{SSIM} + \gamma L_{vessel} \tag{4}$$

In this study, we set the weights for each loss component to $\alpha = \beta = \gamma = 1$. This balanced weighting ensures that the model simultaneously focuses on minimizing the pixel-wise error, enhancing structural similarity, and accurately reconstructing the vascular structures.

To evaluate the quality of the reconstructed images, we compute three key metrics: *SSIM*, *PSNR*, and Mean Squared Error (*MSE*) [40,41]. Additionally, since our primary goal is to accurately recover the numerical values within the vascular structures, we have also designed and computed vessel specific metrics: $SSIM_v$, $PSNR_v$, MSE_v. These metrics are defined as follows:

$$SSIM_v = \frac{1}{N}\sum_{q=1}^{N} \frac{1}{\left\|\boldsymbol{mask}_q\right\|_0} \sum_{\boldsymbol{mask}_q=1} \boldsymbol{SSIM}\left(\boldsymbol{x}_q, \boldsymbol{y}_q\right) \tag{5}$$

$$PSNR_v = 10 \bullet \log_{10}\left(\frac{L^2}{MSE_v}\right) \tag{6}$$

$$MSE_v = \frac{1}{mn}\sum_{i=0}^{m}\sum_{j=0}^{n} (\boldsymbol{x}(i,j) - \boldsymbol{y}(i,j))^2 \bullet \boldsymbol{mask}(i,j) \tag{7}$$

where N is the number of image patches into which the full image is divided, L represents a constant that denotes the maximum dynamic range of the full image data type, and *mask* is a binary mask of the same size as the original image, generated using an empirical threshold k, defined as follows:

$$\boldsymbol{mask}(i,j) = \begin{cases} 1, & \text{if } \boldsymbol{x}(i,j) > k \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

In this work, beyond calculating the aforementioned evaluation metrics for structural results, we also compute the corresponding metrics for the functional parameter sO_2 images, and the detailed method of sO_2 quantification can be found in *Supplementary Section 2*.

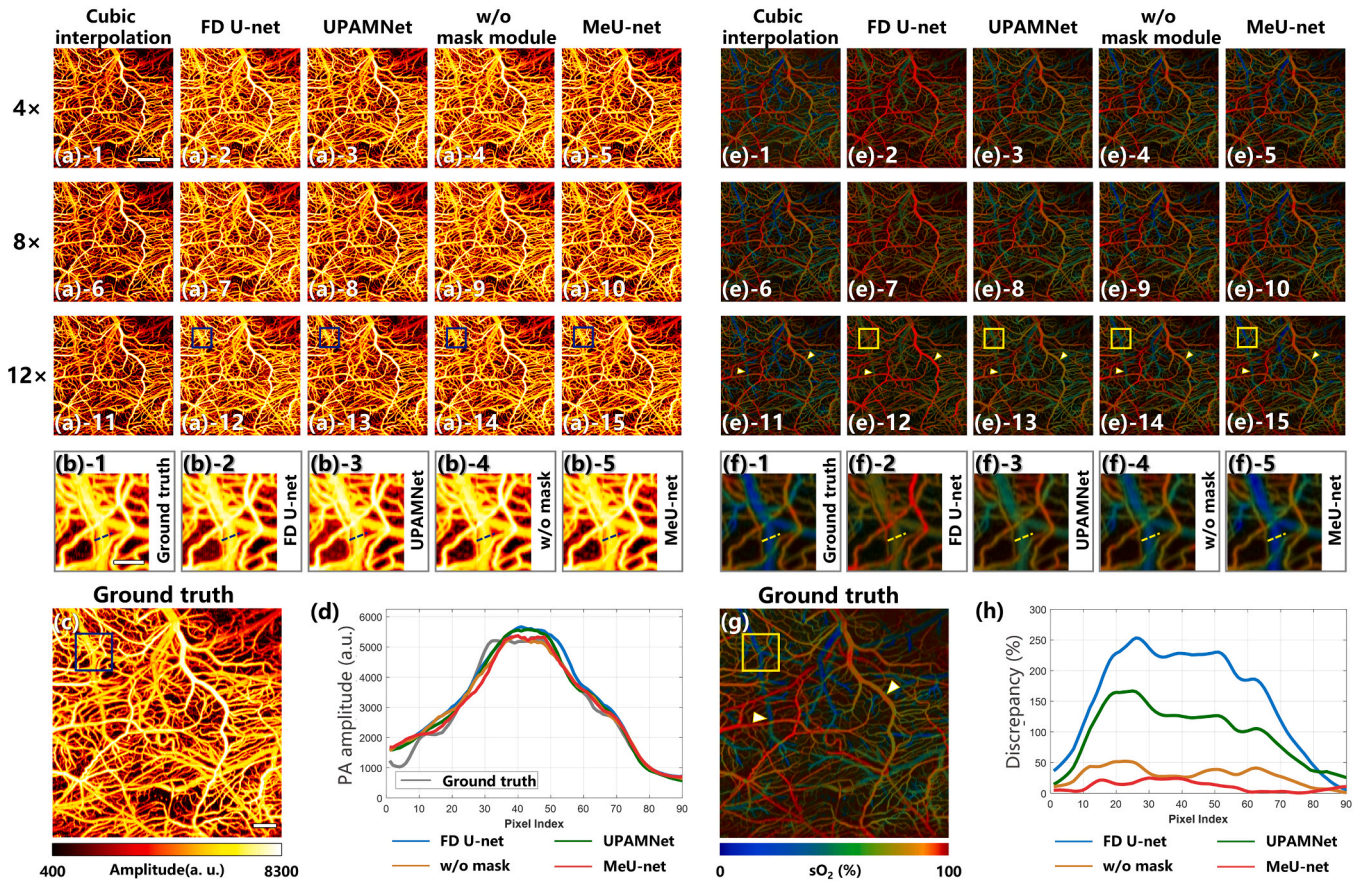


Fig. 4. Mouse brain data undersampling model recovery results. (a, b, c, d) are the results of mouse brain structure MAP recovery; (e, f, g, h) are the results of mouse brain sO_2 recovery. (a, e) are the results of the whole picture, from left to right, they are interpolation, FD U-net, UPAMNet, MeU-net without mask module, MeU-net; from top to bottom, they are 4 × , 8 × , 12 × . Scale bar: 1 mm; (b, f) are the regions of interest, from left to right, they are ground truth, FD U-net, UPAMNet, MeU-net without mask module, MeU-net. Scale bar: 250 μm; (c, g) are ground truth pictures. Scale bar: 0.5 mm; (d, h) are the comparisons of the values and sO_2 discrepancies of different models on selected line segments (dashed lines) in (b, f).

3. Results

3.1. Mouse brain results

We acquired full-brain vascular images from 11 mice and the corresponding scanning area was 5.5 mm × 4.2 mm. After cropping the effective region, removing unnecessary skull portions from the images, and segmenting the data, 920 images of 256 × 256 size were generated for the dataset. Among these, 734 images from 9 mice were used as the training data, while the remaining 186 images from 2 mice were allocated to the validation sets.

To demonstrate the superiority of our model, we conducted tests using newly acquired mouse brain data and trained four additional representative models for comparison, namely Res U-net, FD U-net, UPAMNet [33] and SEResnet [30]. The architectures of Res U-net and FD U-net models are illustrated in Supplementary Fig. S2 and Fig. S3, respectively. It is worth noting that for SEResnet, since the input data to our model underwent zero-padding, we removed the final upsampling layer from its original architecture to ensure compatibility. At the same time, to verify the effectiveness of our model improvement, we conducted ablation experiments on the mask module and L_{vessel}, respectively. Notably, to minimize the impact of different parameters on the model's results, we performed multiple trainings using learning rate of 0.001, and the random seed of 42. Following the aforementioned procedure, we trained multiple models under three different undersampling rates (4 ×, 8 × and 12 ×) to conduct experiments. The quantitative results of the construction on the test set are summarized in Table 1, while representative reconstructed PAM images from the models are shown in Fig. 4. It is important to note that the PAM images we present are the reassembled full-size images, rather than individual sub-images.

From the results presented above, it is evident that our model consistently outperforms others. Specifically, bicubic interpolation introduces substantial distortion in the data, which not only affects the structural performance but also the quantitative values. For the reconstructed images, the brightness of pixels indicate the amplitude value of photoacoustic signals. It can be clearly observed that the interpolation results significantly differ from the ground truth. For the sO_2 images of interpolation, a noticeable stair-step pattern in the red regions is visible, which is inconsistent with the actual physiological structures. In the case of the FD U-net, although it performs well in terms of structural metrics, there is a significant deviation in the sO_2 reconstruction. Although the results of the other models are relatively good, our model still outperforms them overall. For the two ablation models, the results on Table 1 show that our model's loss function design is superior to the model trained without L_{vessel} in the vast majority of cases. On the other hand, the model without the mask module does not show a significant difference compared to our model, and in certain cases, its structural metrics even surpass those of our model. However, in the sO_2 evaluation, both in terms of quantitative metrics and visual performance, the model without the mask module underperforms compared to our model, which illustrates the importance of the mask module for focusing on the internal features of the vessel.

To provide a clearer comparison of the performance differences between various methods and models, we zoomed in on specific regions of the reconstructed images under 12 × undersampling, along with the corresponding ground truth regions, as shown in Fig. 4(b, f).

Table 2
Experiment results of mouse ear data.

Sparsity	Image type	*Evaluation indicators*	Cubic interpolation	Res U-net	FD U-net	UPAMNet	SEResnet	MeU-net w/o L_{vessel}	MeU-net w/o mask module	MeU-net
4 ×	MAP	***SSIM***	0.6619	0.8969	0.8970	0.8410	**0.9090**	0.8968	0.9048	0.8939
		SSIM$_v$	0.8186	0.8755	0.8725	0.8662	0.8847	0.8878	**0.8904**	0.8901
		PSNR	23.4886	29.9237	29.8761	28.5086	30.1536	30.1855	**30.3867**	29.9211
		PSNR$_v$	21.7538	25.0139	24.9292	24.5030	25.2043	25.4770	**25.5784**	25.1491
		MSE	0.00457	0.00108	0.00109	0.00153	0.00105	0.00105	**0.00100**	0.00108
		MSE$_v$	0.00684	0.00324	0.00329	0.00366	0.00313	0.00295	**0.00287**	0.00319
	sO_2	***SSIM***	0.7774	0.8019	0.7986	0.7950	0.8099	0.8108	0.8032	**0.8114**
		SSIM$_v$	0.7354	0.7657	0.7528	0.7562	0.7735	0.7730	0.7669	**0.7749**
		PSNR	22.1171	22.3290	22.0905	21.9354	22.2876	22.3911	22.2679	**22.7113**
		PSNR$_v$	21.1649	21.2381	20.8040	20.6389	21.2153	21.1958	**21.6522**	21.5594
		MSE	0.00646	0.00609	0.00650	0.00666	0.00613	0.00612	0.00624	**0.00552**
		MSE$_v$	0.00809	0.00792	0.00887	0.00924	0.00814	0.00825	**0.00723**	**0.00731**
8 ×	MAP	***SSIM***	0.5932	0.8198	0.8120	0.7521	0.8183	0.8096	0.8211	**0.8299**
		SSIM$_v$	0.7309	0.7944	0.7813	0.7866	0.8037	**0.8007**	0.7958	0.7969
		PSNR	22.1629	27.3235	27.3201	26.2713	27.2025	27.1139	27.1029	**27.5925**
		PSNR$_v$	19.7732	22.5425	22.3353	22.4873	22.4631	22.5781	22.3166	**22.6603**
		MSE	0.00623	0.00194	0.00192	0.00257	0.00201	0.00207	0.00207	**0.00182**
		MSE$_v$	0.01078	0.00567	0.00595	0.00575	0.00580	0.00565	0.00600	**0.00552**
	sO_2	***SSIM***	0.6464	0.6711	0.6220	0.6796	0.6404	0.6792	0.7013	**0.7027**
		SSIM$_v$	0.5940	0.6293	0.5837	0.6413	0.5821	0.6434	0.6575	**0.6579**
		PSNR	19.4167	17.9925	15.6181	19.2952	16.5172	18.3193	19.7484	**20.1312**
		PSNR$_v$	18.4200	16.6800	15.7607	18.4117	14.2181	18.1324	18.8686	**18.9053**
		MSE	0.01182	0.01675	0.02734	0.01232	0.02420	0.01557	0.01085	**0.00991**
		MSE$_v$	0.01499	0.02268	0.02889	0.01514	0.04092	0.01629	0.01332	**0.01324**
12 ×	MAP	***SSIM***	0.4903	0.7144	0.7642	0.6844	0.7593	0.7609	0.7441	**0.7668**
		SSIM$_v$	0.5868	0.7030	0.7266	0.7037	0.7253	0.7371	0.7392	**0.7450**
		PSNR	20.2654	25.2860	**25.9369**	24.9499	25.5311	25.2254	25.4992	25.7235
		PSNR$_v$	17.5563	20.7617	21.0477	20.8946	20.8316	20.5647	21.0369	**21.1363**
		MSE	0.00962	0.00307	0.00263	0.00334	0.00291	0.00317	0.00294	**0.00279**
		MSE$_v$	0.01799	0.00858	0.00802	0.00833	0.00847	0.00905	0.00807	**0.00788**
	sO_2	***SSIM***	0.5537	0.5989	0.3990	0.6038	0.6137	0.5785	0.6179	**0.6230**
		SSIM$_v$	0.5070	0.5593	0.3900	0.5566	0.5551	0.5400	0.5636	**0.5794**
		PSNR	17.8589	17.2099	12.5077	17.6207	17.6702	16.3522	18.1183	**18.2856**
		PSNR$_v$	16.9677	16.9870	15.1438	16.7774	16.2409	16.5387	17.0206	**17.3033**
		MSE	0.01681	0.01958	0.05846	0.01782	0.01748	0.02438	0.01585	**0.01537**
		MSE$_v$	0.02078	0.02061	0.03151	0.02180	0.02460	0.02316	0.02050	**0.01916**

Additionally, arrows were also used to highlight blood vessels with noticeable differences. Furthermore, the sO_2 values in major blood vessels are more representative of physiological changes. To demonstrate the effectiveness of the model, we selected the intensity distribution and the sO_2 differences between the reconstructed results and ground truth values within the cross-section of a major blood vessel in the region, as shown in Fig. 4(d, h).The sO_2 discrepancy is defined as $\Delta = \frac{sO_{2R} - sO_{2T}}{sO_{2T}} \times 100\%$, where sO_{2R} represents the reconstructed sO_2 values, and sO_{2T} represents the ground truth sO_2 values. It is evident that our model achieves the closest match to the ground truth in both structural intensity and sO_2 recovery. Therefore, our model not only achieves high-quality structural reconstruction from sparse data, but also ensures numerical accuracy, demonstrating superior performance in the calculation of the functional parameter sO_2.

3.2. Mouse Ear Results

In order to further validate the generalization capability of our model and assess its ability to recover data in fine vasculature, we acquired ear images from five mice using PAM, with a scanning area of 6 mm × 6 mm. After cropping the effective regions and segmenting the data, a dataset comprising 1266 images of 256 × 256 size was generated. Among these, 1048 images from the ears of four mice were used as the training set, while the remaining 218 images from one mouse were used for validation set. The models, training methods, parameter settings, and metric calculations are consistent with those described in the previous section. The quantitative results of the reconstructions on the test set are presented in Table 2, while Fig. 5 shows the visualized PAM images reconstructions from models.

Based on the evaluation metrics in the table and the visual appearance of the images, the overall performance is similar to that observed for the mouse brain vasculature in *Subsection 3.1*. Bicubic interpolation shows a significant discrepancy compared to the ground truth. While the FD U-net performs well in structural evaluation metrics, it still exhibits large deviations in the sO_2 calculation. The ablation experiments demonstrate that our model's designed loss function outperforms the results obtained by the model without L_{vessel}. Although removing the mask module improves some structural metrics, our model still outperforms it in sO_2 performance. It is important to note that in the visualized ear sO_2 images, models generated, including the ground truth, display a stair-step pattern in the red regions. This occurs due to minor damage to the outer skin during image acquisition, where small blood spots on the surface of the ear naturally create red spots. This phenomenon differs from the brain images, where the vasculature is internal. Thus, we believe that the presence of such artifacts is reasonable in this context.

As seen in the magnified regions in Fig. 5(b, f) and the intensity distribution and sO_2 discrepancies results of the vessel cross-sections in Fig. 5(d, h), our model also exhibits the closest match to the ground truth, maintaining superior performance. Our model consistently produces high-quality results across different samples, demonstrating strong generalization capabilities.

To further validate the capability of our model in reconstructing

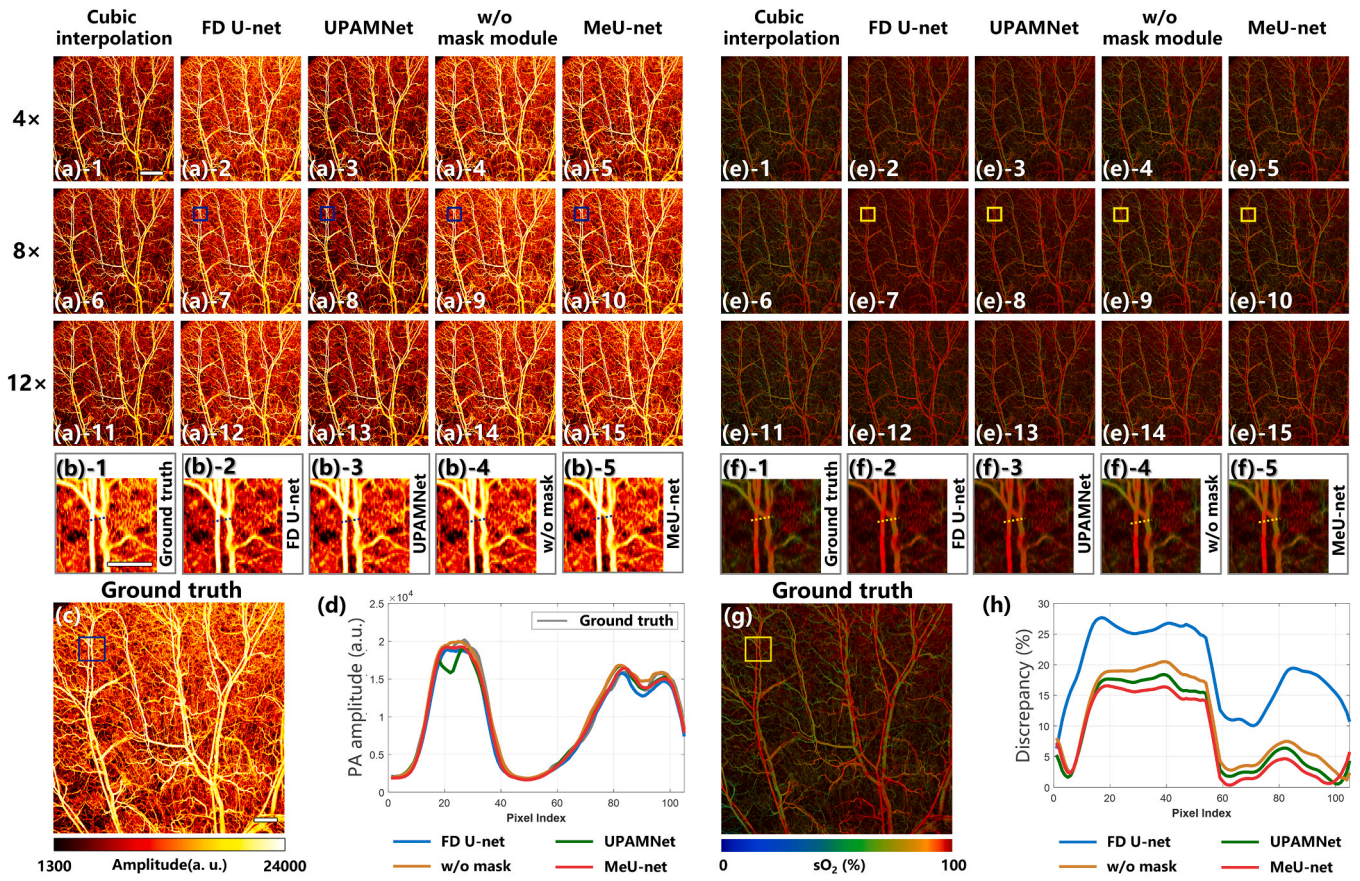


Fig. 5. Mouse ear data undersampling model recovery results. (a, b, c, d) are the results of mouse ear structure MAP recovery; (e, f, g, h) are the results of mouse ear sO_2 recovery. (a, e) are the results of the whole picture, from left to right, they are interpolation, FD U-net, UPAMNet, MeU-net without mask module, MeU-net; from top to bottom, they are 4 ×, 8 ×, 12 ×. Scale bar: 1 mm; (b, f) are the regions of interest, from left to right, they are ground truth, interpolation, FD U-net, UPAMNet, MeU-net without mask module, MeU-net. Scale bar: 250 μm. (c, g) are ground truth pictures. Scale bar: 0.5 mm. (d, h) are the comparisons of the values and sO_2 discrepancies of different models on selected line segments (dashed lines) in (b, f).

structural and sO_2 images on unseen data, we employed a transfer learning approach. Specifically, the model trained on mouse brain data (*Subsection 3.1*) was directly applied to the reconstruction of mouse ear data at three sparsity levels without additional training. The reconstruction results are presented in Table 3. Although the performance of the transferred model decreased slightly compared to the model trained specifically on mouse ear data, our model still outperformed other representative models in the majority of cases. This demonstrates the strong generalization capability and robustness of our model, further validating its superiority in reconstructing structural and functional information under varying conditions.

4. Discussion

The primary objective of this work is to address the challenge of enhancing imaging speed in PAM systems under hardware constraints while maintaining the accuracy of structural features and functional parameters, specifically sO_2, at high resolution. For improving PAM imaging speed, both hardware advancements and algorithmic compression-and-reconstruction approaches can be employed [42,43]. However, hardware-based improvements often come with significantly increased system costs. On the algorithmic front, most existing deep learning approaches for undersampled image recovery focus primarily on structural aspects, with few studies addressing the recovery of functional parameters such as sO_2. To make up for this shortage, we developed the MeU-net model, which incorporates key innovations such as an adaptive vascular attention mask module and a vessel-specific loss function. By comparing our model with bicubic interpolation and other representatives, we demonstrate that MeU-net not only achieves high-quality structural recovery from sparse images but also ensures accurate sO_2 parameter's recovery. Using only 8.33–25 % of the original data, our method theoretically enables PAM systems to increase imaging speed by 4–12 times, while preserving the accuracy of the sO_2 parameter.

In our study, we conducted experiments using PAM to image both the brain and ear regions of mice. Two key aspects need to be discussed in this context. First, the uneven lateral and axial resolution used during data acquisition stems from practical considerations. In high-resolution imaging, achieving the same level of axial resolution as lateral resolution would significantly increase acquisition time, which is impractical in most real-use scenarios. Moreover, the high lateral resolution is sufficient to ensure accurate calculation of functional parameters such as sO_2. Secondly, the selection of the mouse brain and ear regions was intentional. The brain contains a large number of major blood vessels with well-defined gaps between them, providing a high signal-to-noise ratio (SNR) environment, which highlights the model's performance under low-noise conditions. In contrast, the mouse ear features a dense network of micro vessels and, during the imaging process, presents challenges due to the presence of blood clots. This creates a low SNR scenario that is ideal for evaluating the model's ability to recover sO_2 in more difficult conditions involving small vessels. Our experiments demonstrate that MeU-net outperforms other models in both settings, consistently delivering superior results under both high SNR and low SNR conditions. At the same time, we also tested the recovery of data after undersampling in both lateral and axial directions ((4 ×, 2 ×) and (4 ×, 4 ×)), as well as larger lateral sparsity (16 ×), as shown in

Table 3
Transfer learning results of mouse ear data.

Sparsity	Image type	*Evaluation indicators*	FD U-net	UPAMNet	SEResnet	MeU-net
4 ×	MAP	***SSIM***	0.8671	0.8548	0.8713	**0.8791**
		$SSIM_v$	0.8532	0.8681	0.8705	**0.8741**
		PSNR	29.2333	29.0296	29.2799	**29.4432**
		$PSNR_v$	24.4497	24.8680	24.8532	**24.9067**
		MSE	0.00125	0.00138	0.00127	**0.00122**
		MSE_v	0.00366	0.00336	0.00338	**0.00333**
	sO_2	***SSIM***	0.5518	**0.7876**	0.7836	0.7872
		$SSIM_v$	0.5448	0.7514	0.7480	**0.7616**
		PSNR	13.5145	21.8508	21.3607	**21.7318**
		$PSNR_v$	16.4105	20.6900	19.7539	**21.0021**
		MSE	0.04646	**0.00689**	0.00766	0.00709
		MSE_v	0.02398	0.00915	0.01152	**0.00852**
8 ×	MAP	***SSIM***	0.7981	0.7012	0.7934	**0.8072**
		$SSIM_v$	0.7907	0.7799	0.7905	**0.7981**
		PSNR	27.0216	25.6778	26.7991	**27.0701**
		$PSNR_v$	22.6708	22.5922	22.3142	**22.6848**
		MSE	0.00208	0.00281	0.00222	**0.00208**
		MSE_v	0.00550	0.00561	0.00601	**0.00550**
	sO_2	***SSIM***	0.5941	0.6524	0.6439	**0.6548**
		$SSIM_v$	0.5586	0.6279	0.6277	**0.6401**
		PSNR	16.4635	18.8523	18.8063	**19.3590**
		$PSNR_v$	15.9170	18.3041	17.9331	**18.5773**
		MSE	0.02411	0.01357	0.01365	**0.01202**
		MSE_v	0.02767	0.01561	0.01714	**0.01465**
12 ×	MAP	***SSIM***	0.7241	0.6615	0.7067	**0.7241**
		$SSIM_v$	0.7233	0.6910	0.7194	**0.7265**
		PSNR	25.2217	24.7198	24.8524	**25.3728**
		$PSNR_v$	21.0833	20.9546	20.7746	**21.1978**
		MSE	0.00312	0.00348	0.00342	**0.00302**
		MSE_v	0.00798	0.00818	0.00862	**0.00780**
	sO_2	***SSIM***	0.3620	0.4396	0.5619	**0.5748**
		$SSIM_v$	0.3162	0.4661	0.5331	**0.5597**
		PSNR	12.8586	14.5163	16.9301	**17.2025**
		$PSNR_v$	13.9525	15.4876	15.6387	**16.9226**
		MSE	0.05309	0.03803	0.02101	**0.01972**
		MSE_v	0.04143	0.03021	0.02884	**0.02128**

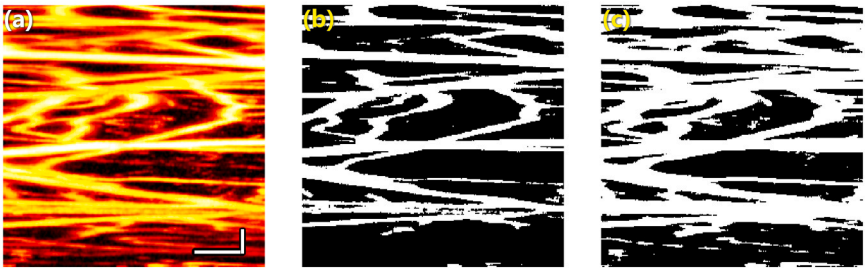


Fig. 6. Schematic diagram of vascular mask segment. (a) Example image of ground truth; (b) The vascular mask made with the threshold obtained by graythresh method; (c) The vascular mask made with 80 % of the threshold obtained by graythresh method. Scale bars: vertical: 200 μm, horizontal: 50 μm.

Supplementary Section 4, and our model also showed good results. However, we also acknowledge that our model has several areas in need of improvement. For instance, in the 12 × undersampling results for the ear images, as shown in Fig. 5(a)-11–15 and Fig. 5(e)-11–15, it is undeniable that both our model and other models exhibit suboptimal performance. We attribute this to the presence of blood spots and capillaries in the ear, which introduce significant challenges to accurate reconstruction.

It is evident that the choice of threshold values for vascular mask significantly influences the performance of the model, the loss function, and the evaluation metrics. To address this, we adopted a systematic approach for selecting mask thresholds. For evaluation metrics, the mask threshold was determined using Otsu's method for global image thresholding [44]. The schematic diagram of the vascular mask result is shown in Fig. 6. We computed thresholds for all acquired data and used their average as the final value for the evaluation indicators' mask. In contrast, for the loss function and the mask values used within the model, we considered the potential influence of features surrounding the vasculature on the recovery of vascular edges. To account for this, we reduced the mask threshold slightly. Specifically, we set the mask threshold for the loss function and model at 80 % of the average value calculated earlier. This adjustment ensures a balance between accurately capturing vascular details and mitigating the influence of peripheral features.

Besides, it is an important point for discussion that during the evaluation of network performance, the optimal results for structural metrics

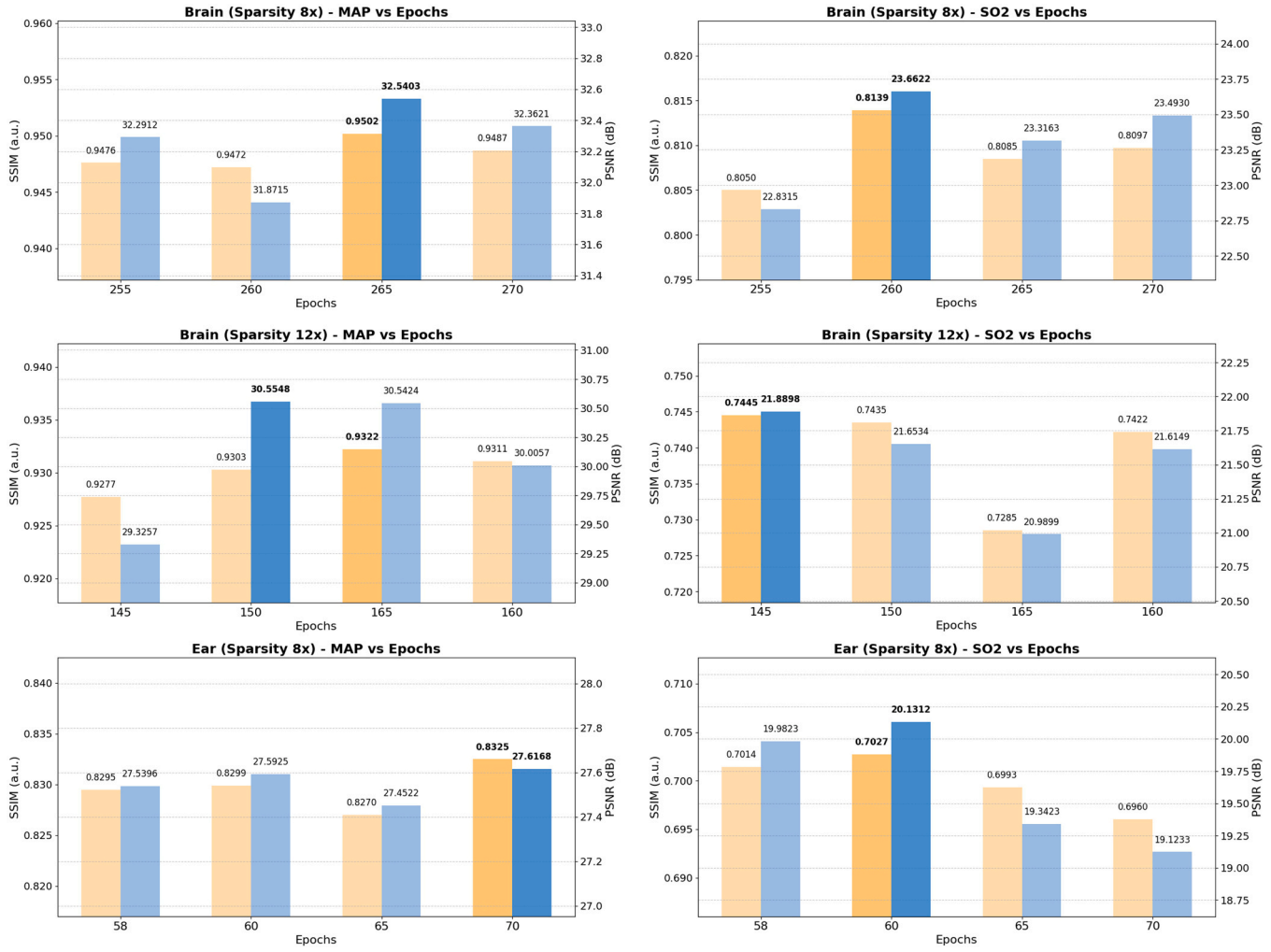


Fig. 7. Different conditions training results with different epochs.

and sO_2 metrics are not always achieved simultaneously. This suggests that a trade-off may be necessary in practical applications. From extensive observations, we found that sO_2 tends to reach its optimal value within approximately 10 epochs before or after the point at which structural results are at their peak. As shown in Fig. 7, we plotted the SSIM and PSNR of the MAP graph and sO_2 graph under different conditions training. This is because sO_2 us derived from data that must first be reconstructed by the model, and the current loss function cannot fully guarantee an explicit optimization for sO_2 during training. This represents both a limitation of the current work and a direction for future research. Nonetheless, the loss function we designed in this work has played a significant role in optimizing sO_2, as evidenced by the results in Table 1 and Table 2. Additionally, we observed another phenomenon in which, under different levels of sparsity, the model without the mask module performed very similarly to our model across all evaluation metrics, but in the actual visual effect of Figs. 4(e)-4 and Figs. 5(e)-4, it is far from the ground truth. And although our model slightly outperforms the model without mask module in terms of the quantitative metrics, the disparity between the two models' metrics is not as large as the observed difference in visual quality. We attribute this discrepancy to fact that when calculating the final evaluation metrics, we used the raw data format without converting it into compressed image formats such as PNG or JPEG, which has affected the metric calculations. However, in terms of physiological parameter recovery, ensuring numerical accuracy is critical. Thus, developing evaluation metrics that are more aligned with this application is another direction we aim to explore in the future.

5. Conclusion

In summary, to achieve an increase in PAM acquisition speed while ensuring the accuracy of structural and sO_2 parameters, we designed an adaptive attention module focused on vascular recovery and developed a vessel-specific loss function L_{vessel}, resulting in the MeU-net algorithm. This algorithm has demonstrated outstanding performance in the recovery of images from both the mouse brain and ear. We validated the accuracy of our model by comparing it with representative methods such as bicubic interpolation, Res U-net, and FD U-net. Furthermore, we discussed the rationale behind our design choices and acknowledged the limitations present in our work. We provided a more in-depth discussion of the superior numerical and visual recovery capabilities of our model and outlined potential directions for future improvements. Our work offers valuable insights for accelerating the scanning speed of PAM systems and other systems employing similar scanning methodologies, while also enhancing the recovery of undersampled data. Additionally, we explored the potential of deep learning to recover physiological functional parameters. However, we recognize that we have yet to achieve high-accuracy recovery from the signal level and the simultaneous recovery of flow and the metabolic rate of oxygen functional parameters alongside structural images. This remains a key focus for our future research endeavors.

Funding sources

This work was supported in part by the Natural Science Foundation of Beijing Municipality under Grants 4244108 and the Key Technologies Research and Development Program under Grants 2023YFC2506804.

CRediT authorship contribution statement

Shuyan Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jingtan Li:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Lin Shen:** Writing – review & editing, Software, Methodology. **Zhonghao Zhao:** Writing – review & editing, Software, Methodology. **Minjun Lee:** Methodology, Software. **Kun Qian:** Resources, Funding acquisition. **Naidi Sun:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Bin Hu:** Supervision, Resources, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (version GPT-4o, OpenAI) in order to improve the readability and language. After using ChatGPT, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Mengkai Sun, Zhigang Li and Gang Luo for assistance on the method of model analysis and visualization.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pacs.2025.100687.

Data availability

Data will be made available on request.

References

[1] W. Liu, J. Yao, Photoacoustic microscopy: principles and biomedical applications, Biomed. Eng. Lett. 8 (2018) 203–213.
[2] J. Park, S. Choi, F. Knieling, B. Clingman, S. Bohndiek, L.V. Wang, C. Kim, Clinical translation of photoacoustic imaging, Nat. Rev. Bioeng. (2024) 1–20.
[3] J. Xia, J. Yao, L.V. Wang, Photoacoustic tomography: principles and advances, Electromagn. Waves (Camb., Mass.) 147 (2014) 1.
[4] J. Yao, L.V. Wang, Sensitivity of photoacoustic microscopy, Photoacoustics 2 (2014) 87–101.
[5] P. Hai, J. Yao, K.I. Maslov, Y. Zhou, L.V. Wang, Near-infrared optical-resolution photoacoustic microscopy, Opt. Lett. 39 (2014) 5192–5195.
[6] M. Li, Y. Tang, J. Yao, Photoacoustic tomography of blood oxygenation: A mini review, Photoacoustics 10 (2018) 65–73, https://doi.org/10.1016/j.pacs.2018.05.001.
[7] S. Hu, K. Maslov, V. Tsytsarev, L.V. Wang, Functional transcranial brain imaging by optical-resolution photoacoustic microscopy, J. Biomed. Opt. 14 (2009), 040503–040503.
[8] K. Wang, C. Li, R. Chen, J. Shi, Recent advances in high-speed photoacoustic microscopy, Photoacoustics 24 (2021) 100294.
[9] J. Yao, C. Kim, M. Kolios, S. Hu, Breaking the speed limits in photoacoustic microscopy, Photoacoustics 32 (2023).
[10] Z. Zhang, H. Jin, Z. Zheng, W. Zhang, W. Lu, F. Qin, A. Sharma, M. Pramanik, Y. Zheng, HSD-PAM: high speed super resolution deep penetration photoacoustic microscopy imaging boosted by dual branch fusion, Netw., arXiv Prepr. arXiv 2308 (2023) 04922.
[11] S.-W. Cho, S.M. Park, B. Park, T.G. Lee, B.-M. Kim, C. Kim, J. Kim, S.-W. Lee, C.-S. Kim, others, High-speed photoacoustic microscopy: a review dedicated on light sources, Photoacoustics 24 (2021) 100291.
[12] W. Shi, P. Hajireza, P. Shao, A. Forbrich, R.J. Zemp, In vivo near-realtime volumetric optical-resolution photoacoustic microscopy using a high-repetition-rate nanosecond fiber-laser, Opt. Express 19 (2011) 17143–17150.
[13] J. Lee, S. Han, D. Seong, J. Lee, S. Park, R. Eranga Wijesinghe, M. Jeon, J. Kim, Fully waterproof two-axis galvanometer scanner for enhanced wide-field optical-resolution photoacoustic microscopy, Opt. Lett. 45 (2020) 865–868.
[14] J. Chen, Y. Zhang, L. He, Y. Liang, L. Wang, Wide-field polygon-scanning photoacoustic microscopy of oxygen saturation at 1-MHz A-line rate, Photoacoustics 20 (2020) 100195.
[15] P. Vaidyanathan, Generalizations of the sampling theorem: Seven decades after Nyquist, IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. 48 (2001) 1094–1109.
[16] J. Wang, B. Li, T. Zhou, C. Liu, M. Lu, W. Gu, X. Liu, D. Ta, Reconstructing cancellous bone from down-sampled optical-resolution photoacoustic microscopy images with deep learning, Ultrasound Med. Biol. 50 (2024) 1459–1471.
[17] R. Wang, J. Zhu, J. Xia, J. Yao, J. Shi, C. Li, Photoacoustic imaging with limited sampling: a review of machine learning approaches, Biomed. Opt. Express 14 (2023) 1777–1799.
[18] H. Deng, H. Qiao, Q. Dai, C. Ma, Deep learning in photoacoustic imaging: a review, J. Biomed. Opt. 26 (2021), 040901–040901.
[19] S. Choi, J. Yang, S.Y. Lee, J. Kim, J. Lee, W.J. Kim, S. Lee, C. Kim, Deep learning enhances multiparametric dynamic volumetric photoacoustic computed tomography in vivo (DL-PACT), Adv. Sci. 10 (2023) 2202089.
[20] S.G. Sathyanarayana, B. Ning, S. Hu, J.A. Hossack, Simultaneous dictionary learning and reconstruction from subsampled data in photoacoustic microscopy, in: 2019 IEEE International Ultrasonics Symposium (IUS), IEEE, 2019: pp. 483–486.
[21] H. Lan, D. Jiang, C. Yang, F. Gao, F. Gao, Y-Net: Hybrid deep learning image reconstruction for photoacoustic tomography in vivo, Photoacoustics 20 (2020) 100197, https://doi.org/10.1016/j.pacs.2020.100197.
[22] H. Lan, K. Zhou, C. Yang, J. Cheng, J. Liu, S. Gao, F. Gao, Ki-GAN: Knowledge Infusion Generative Adversarial Network for Photoacoustic Image Reconstruction In Vivo, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Springer International Publishing, Cham, 2019, pp. 273–281, https://doi.org/10.1007/978-3-030-32239-7_31.
[23] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, S. Arridge, Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography, IEEE Trans. Med. Imaging 37 (2018) 1382–1393, https://doi.org/10.1109/TMI.2018.2820382.
[24] A. Hauptmann, B. Cox, F. Lucka, N. Huynh, M. Betcke, P. Beard, S. Arridge, Approximate k-Space Models and Deep Learning for Fast Photoacoustic Reconstruction, in: F. Knoll, A. Maier, D. Rueckert (Eds.), Machine Learning for Medical Image Reconstruction, Springer International Publishing, Cham, 2018, pp. 103–111, https://doi.org/10.1007/978-3-030-00129-2_12.
[25] M.A. Anastasio, Deep learning and photoacoustic image formation: promises and challenges. Photons Plus Ultrasound: Imaging and Sensing 2023, SPIE, 2023, p. PC1237901.
[26] H. Lan, J. Gong, F. Gao, Deep learning adapted acceleration for limited-view photoacoustic image reconstruction, Opt. Lett. 47 (2022) 1911, https://doi.org/10.1364/OL.450860.
[27] H. Lan, L. Huang, X. Wei, Z. Li, J. Lv, C. Ma, L. Nie, J. Luo, Masked cross-domain self-supervised deep learning framework for photoacoustic computed tomography reconstruction, Neural Netw. 179 (2024) 106515, https://doi.org/10.1016/j.neunet.2024.106515.
[28] H. Lan, C. Yang, F. Gao, A jointed feature fusion framework for photoacoustic image reconstruction, Photoacoustics 29 (2023) 100442.
[29] A. DiSpirito, D. Li, T. Vu, M. Chen, D. Zhang, J. Luo, R. Horstmeyer, J. Yao, Reconstructing undersampled photoacoustic microscopy images using deep learning, IEEE Trans. Med. Imaging 40 (2020) 562–570.
[30] J. Zhou, D. He, X. Shang, Z. Guo, S.-L. Chen, J. Luo, Photoacoustic microscopy with sparse data by convolutional neural networks, Photoacoustics 22 (2021) 100242.
[31] T. Vu, A. DiSpirito III, D. Li, Z. Wang, X. Zhu, M. Chen, L. Jiang, D. Zhang, J. Luo, Y. S. Zhang, et al., Deep image prior for undersampling high-speed photoacoustic microscopy, Photoacoustics 22 (2021) 100266.
[32] J. Kim, G. Kim, L. Li, P. Zhang, J.Y. Kim, Y. Kim, H.H. Kim, L.V. Wang, S. Lee, C. Kim, Deep learning acceleration of multiscale superresolution localization photoacoustic imaging, Light.: Sci. Appl. 11 (2022) 131.
[33] Y. Liu, J. Zhou, Y. Luo, J. Li, S.-L. Chen, Y. Guo, G.-Z. Yang, UPAMNet: A unified network with deep knowledge priors for photoacoustic microscopy, Photoacoustics 38 (2024) 100608.
[34] I. Loc, M.B. Unlu, Accelerating photoacoustic microscopy by reconstructing undersampled images using diffusion models, Sci. Rep. 14 (2024) 16996.
[35] H. An, W. Lin, Quantitative measurements of cerebral blood oxygen saturation using magnetic resonance imaging, J. Cereb. Blood Flow. Metab. 20 (2000) 1225–1236, https://doi.org/10.1097/00004647-200008000-00008.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of, IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 770–778.

[37] R. Cao, A. Tran, J. Li, Z. Xu, N. Sun, Z. Zuo, S. Hu, Hemodynamic and oxygen-metabolic responses of the awake mouse brain to hypercapnia revealed by multi-parametric photoacoustic microscopy, J. Cereb. Blood Flow. Metab. 41 (2021) 2628–2639.
[38] R. Cao, J. Li, B. Ning, N. Sun, T. Wang, Z. Zuo, S. Hu, Functional and oxygen-metabolic photoacoustic microscopy of the awake mouse brain, Neuroimage 150 (2017) 77–87.
[39] B. Ning, M.J. Kennedy, A.J. Dixon, N. Sun, R. Cao, B.T. Soetikno, R. Chen, Q. Zhou, K. Kirk Shung, J.A. Hossack, others, Simultaneous photoacoustic microscopy of microvascular anatomy, oxygen saturation, and blood flow, Opt. Lett. 40 (2015) 910–913.
[40] A. Horé, D. Ziou, Image Quality Metrics: PSNR vs. SSIM, 20th Int. Conf. Pattern Recognit. 2010 (2010) 2366–2369, https://doi.org/10.1109/ICPR.2010.579.
[41] U. Sara, M. Akter, M.S. Uddin, Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study, J. Comput. Commun. (2019). ⟨https://api.semanticscholar.org/CorpusID:104425037⟩.
[42] J. Yao, L. Wang, J.-M. Yang, K.I. Maslov, T.T. Wong, L. Li, C.-H. Huang, J. Zou, L. V. Wang, High-speed label-free functional photoacoustic microscopy of mouse brain in action, Nat. Methods 12 (2015) 407–410.
[43] F. Zhong, Y. Bao, R. Chen, Q. Zhou, S. Hu, High-speed wide-field multi-parametric photoacoustic microscopy, Opt. Lett. 45 (2020) 2756–2759.
[44] T.Y. Goh, S.N. Basah, H. Yazid, M.J.A. Safar, F.S.A. Saad, Performance analysis of image thresholding: Otsu technique, Measurement 114 (2018) 298–307.

Shuyan Zhang received his B.S. degrees from the College of Physics and Optoelectronic, Taiyuan University of Technology, Taiyuan, China, in 2022. He is currently pursuing a Ph.D. degree with the School of Medical Technology, Beijing Institute of Technology, Beijing, China. His current research interests include photoacoustic microscopy and deep learning.

Jingtan Li received his B.S. degree in Computer Science and Technology from Zhengzhou University. In 2023, he began his M.S. studies in Artificial Intelligence at Beijing Institute of Technology. His research focuses on the application of deep learning in photoacoustic imaging, with recent work exploring super-resolution techniques for photoacoustic microscopy.

Lin Shen received his B.S. and M.Eng. degrees from the School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, China, in 2020 and 2023. He is currently pursuing a Ph.D. degree with the School of Medical Technology, Beijing Institute of Technology, Beijing, China. His current research interests include music therapy, deep learning, and generative artificial intelligence.

Zhonghao Zhao received the B.S. degree from Nanjing Agricultural University, Nanjing, China, in 2022. He is currently pursuing the M.S. degree with the School of Medical Technology, Beijing Institute of Technology, China. His research interests include Computer Audition, Sound Field Modelling, and Body Sound Perception.

Minjun Lee received his B.S. degree in Data Science and Big Data Engineering from the Beijing Institute of Technology in 2023. He is currently pursuing an M.S. degree in Computer Science Engineering at the same university. His research interests include photoacoustic imaging, deep learning, and medical artificial intelligence. His recent work focuses on enhancing photoacoustic image resolution through deep learning techniques, with applications in neural disease diagnosis and monitoring.

Kun Qian (Senior Member, IEEE) received his doctoral degree for his study on automatic general audio signal classification in 2018 in electrical engineering and information technology from Technische Universität München Germany. From 2021, he has been appointed (Full) Professor at the Beijing Institute of Technology, China. His research focuses on computer audition and affective computing. He (co-)authored more than 150 publications in peer reviewed journals, and conference proceedings having received more than 2.4k citations (h-index 30).

Naidi Sun received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 2010 and 2012, respectively, and the Ph.D. degree in biomedical engineering from the University of Virginia, Charlottesville, VA, USA, in 2020. He is currently an Assistant Professor with the School of Medical Technology, Beijing Institute of Technology, Beijing, China. His research focuses on photoacoustic microscopy and functional–metabolic photoacoustic imaging for biomedical applications. He (co-)authored more than 40 publications in peer reviewed journals, and conference proceedings having received more than 700 citations (h-index 12).

Bin Hu (Fellow, IEEE and IET) received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science in 1998. He is a (Full) Professor and the Dean of the School of Medical Technology at Beijing Institute of Technology, China. He is a National Distinguished Expert, Chief Scientist of 973 as well as National Advanced Worker in 2020. His research focuses on computational psychophysiology and affective computing. He (co-)authored more than 400 publications in peer reviewed books, journals, and conference proceedings leading to more than 12k citations (h-index 58).