Resolution Enhancement of Under-Sampled Photoacoustic Microscopy Images Using Neural Representation

Youshen Xiao, Zhengyuan Zhang, Ruixi Sun, Yiling Shi, Sheng Liao, Fan Zhang, Yunhui Jiang, Xiyu Chen, Arunima Sharma, *Member, IEEE*, Manojit Pramanik, *Senior Member, IEEE*, Yuyao Zhang, and Fei Gao

Abstract—**Acoustic-Resolution Photoacoustic Microscopy (AR-PAM) has demonstrated great potential in subcutaneous vascular imaging. However, its spatial resolution is limited by the system's Point Spread Function (PSF). To enhance resolution, various deconvolution-based methods can be employed. Traditional deconvolution methods, such as Richardson-Lucy deconvolution and model-based deconvolution, typically use the PSF as prior knowledge to improve spatial resolution. However, accurately measuring the system's PSF is challenging, leading to the widespread adoption of low vision deconvolution methods, which often suffer from inaccurate deconvolution. Another major challenge of AR-PAM is the long scanning time. To accelerate image acquisition, downsampling can be applied to reduce scanning time. Subsequently, interpolation methods are commonly used to recover high-resolution images from the downsampled measurements. However, conventional interpolation methods struggle to achieve high-fidelity image recovery, particularly under high downsampling conditions. In this study, we propose a method based on Implicit Neural Representations (INR) to simultaneously address the challenges of unknown PSF and under-sampled image recovery. By leveraging INR, we learn a continuous mapping from spatial positions to initial acoustic pressure, effectively compensating for the discretization of the image space and enhancing the resolution of AR-PAM. Specifically, we treat the PSF as a learnable parameter to mitigate inaccuracies in PSF measurement. We qualitatively and quantitatively evaluated the proposed method on leaf vein data, mouse brain data, and real in vivo AR-PAM data, demonstrating superior performance compared to existing methods in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).**

Index Terms—**Photoacoustic microscopy (PAM), sparse image, implicit neural representation, unsupervised, deconvolution.**

I. INTRODUCTION

PHOTOACOUSTIC Imaging (PAI) is a non-invasive biomedical imaging technique that combines the advantages of optical absorption contrast with the deep penetration capabilities of ultrasound [1], [2]. In PAI, acoustic waves are generated by the thermoelastic expansion caused by transient light absorption, which are subsequently captured by ultrasonic sensors to construct PA images, representing the original distribution of initial pressure [3], [4]. With multi-wavelength light illumination, PAI exhibits high sensitivity to optical absorption contrast. Theoretically, any changes in the optical absorption coefficient directly reflects in the intensity variations of the detected PA signals, forming the basis for its functional and molecular imaging capability, such as the measurement of hemoglobin oxygen saturation [5]. Depending on the imaging system configuration, PAI is primarily categorized into three forms: Photoacoustic Microscopy (PAM), Photoacoustic Computed Tomography (PACT), and Photoacoustic Endoscopy (PAE).

PAM offers high-resolution imaging with subcutaneous depth, showcasing significant optical absorption contrast and revealing structural, functional, and molecular information of the tissue [6]. Optical-resolution PAM (OR-PAM), a variant of PAM, achieves near-diffraction-limited resolution through the use of tightly focused laser beams for excitation. Over the past decade, OR-PAM has garnered significant attention due to its superior imaging capabilities and has been widely applied in neuroscience research, tumor angiogenesis monitoring, histology examinations, dermatological diagnostics, and other preclinical and clinical studies [7], [8], [9]. On the other

Received 2 December 2024; revised 14 March 2025; accepted 13 April 2025. Date of current version 21 May 2025. The associate editor coordinating the review of this article and approving it for publication was Prof. David Lindell. *(Corresponding authors: Yuyao Zhang; Fei Gao.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Animal Care and Use Committee (IACUC) of Nanyang Technological University under Application No. SBS/NIEA0331.

Youshen Xiao, Ruixi Sun, Sheng Liao, Fan Zhang, Yunhui Jiang, Xiyu Chen, and Yuyao Zhang are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: xiaoysh2023@shanghaitech.edu.cn; sunrx2023@shanghaitech.edu.cn; liaosheng2022@shanghaitech.edu.cn; zhangfan2022@shanghaitech.edu.cn; jiangyh2022@shanghaitech.edu.cn; chenxy2022@shanghaitech.edu.cn; zhangyy8@shanghaitech.edu.cn).

Zhengyuan Zhang is with the School of Electrical and Electronic Engineering, Nanyang Technological University,, Singapore 639798 (e-mail: zhengyua002@e.ntu.edu.sg).

Yiling Shi is with the School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China (e-mail: shiyl2023@shanghaitech.edu.cn).

Arunima Sharma is with Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: arunima0111@gmail.com).

Manojit Pramanik is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: mano@iastate.edu).

Fei Gao is with the School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China, also with the School of Engineering Science, University of Science and Technology of China, Hefei 230026, China, and also with the Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China (e-mail: fgao@ustc.edu.cn).

The code for this work is availble at https://github.com/zhaowumian7/AR2OR-INR.

Digital Object Identifier 10.1109/TCI.2025.3565129

hand, acoustic-resolution PAM (AR-PAM), leveraging the good penetration of scattered light and sound waves in deeper tissues, demonstrates greater advantages in deep-tissue imaging compared to OR-PAM [10], which has proven its value in microvascular imaging [11].

The lateral resolution of AR-PAM is determined by the center frequency and numerical aperture (NA) of the focused acoustic transducer used. To achieve higher lateral resolution, AR-PAM can employ transducers with higher central frequency and NA. However, this approach comes with challenges, such as requiring the system to detect high-frequency acoustic waves, which experience significant attenuation in biological tissues, thus affecting their ability to penetrate deeply. Additionally, a higher NA can also limit the depth of focus, making it more difficult to maintain good focus over a certain depth range. To enhance the lateral resolution of AR-PAM while overcoming the issues mentioned above, deconvolution algorithms can be employed.

In recent years deconvolution algorithms have been widely applied in the field of photoacoustic imaging, including PACT [12], OR-PAM [13], [14], and AR-PAM [15]. For AR-PAM, various deconvolution techniques have been proposed, notably including Richardson-Lucy (RL) deconvolution and Malvar-Burton (MB) deconvolution. From the perspective of an image degradation model, restoring a clear image fundamentally involves a deconvolution process, which relies on the estimation of the Point Spread Function (PSF). In AR-PAM, the PSF is mainly determined by the characteristics of the acoustic transducer's focal region.

Another notable aspect of AR-PAM is that its performance is primarily constrained by two interrelated key factors: imaging speed and spatial resolution. According to the Nyquist sampling theorem, to ensure that PAM provides high spatial resolution images, the scanning step must be less than half of the lateral resolution to prevent aliasing, which would otherwise degrade image quality. Therefore, traditional PAM requires a large amount of sampling data, which increases acquisition and processing time, raises system memory requirements, and thus the cost. To achieve fast PAM systems without significantly reducing resolution or increasing costs, researchers have developed advanced data acquisition schemes and innovative signal processing methods [16], [17]. For AR-PAM, different scanning mechanisms can be adopted depending on the required imaging speed [18]. For instance, when video-rate three-dimensional imaging is needed, a raster scan of the excitation laser beam can be performed with the acoustic focus (typically with a diameter of about 50 micrometers), although this limits the field of view. To further increase imaging speed, a digital micromirror device can be used for random-access scanning of specific features with the region of interest, effectively skipping background areas. Nevertheless, even with these approaches, data acquisition speeds are still limited by point-by-point optical scanning. Recently, hybrid scanning using water-immersed microelectromechanical systems mirrors has achieved three-dimensional imaging rates of approximately 1 Hz with a moderate field of view (about 3x4 square millimeters) while maintaining high detection sensitivity. This method enhances both imaging speed and quality by simultaneously focusing the excitation laser beam and the received acoustic waves. At the algorithmic level, a sparse matrix recovery method based on the alternating direction method of multipliers has been proposed for sparse optical scanning PAM systems to enable rapid vascular imaging.

In recent years, deep learning methods have emerged in the field of photoacoustic imaging [19]. By training neural network models, these methods can automatically extract features from raw data and generate high-quality images without the need for traditional signal processing steps. Deep learning has shown remarkable performance in addressing issues such as sparse-view, limited-view, and artifact removal in PACT; it has also been applied to deconvolution and undersampling problems in PAM. Chen et al. [20] proposed using deep learning to transform blurred in vivo mouse vasculature images acquired by AR-PAM systems to achieve deep-penetration OR-PAM performance. Feng et al. [21] aiming to enhance the lateral resolution of AR-PAM, developed a multi-scale feature high-fidelity restoration algorithm based on deep convolutional neural networks. To accelerate PAM imaging, M. Burcin Unlu et al. introduced a novel and flexible algorithm called DiffPam [16], based on diffusion models, to speed up the photoacoustic imaging process. Although deep learning methods have shown high effectiveness in AR-PAM deconvolution, improving algorithm performance or accelerating PAM imaging speeds still relies on a large amount of training data. However, the high cost of acquiring PAM data and the difficulty in obtaining paired datasets, coupled with the increased ill-posedness of upsampling and deconvolution of sparsely sampled discrete points, remain significant challenges. Therefore, there is an urgent need to transition from discretized representations to continuous representations suitable for unsupervised deep learning, to support high-resolution deconvolution in AR-PAM imaging.

Implicit Neural Representations (INR) represent a new approach that parameterizes signals using Multi-Layer Perceptron (MLP) [22]. Unlike traditional explicit representations that use discrete elements such as pixels or voxels, INR represent the object itself as a continuous function of spatial coordinates. That is, the values at any spatial positions of the object can be retrieved by querying the corresponding coordinates of a trained MLP. This provides a universal solution for various applications in object reconstruction. With the application of MLP and appropriate encoding functions that map input coordinates into high-dimensional spaces [23], INR have achieved superior performance in multiple computer vision tasks [23], [24]. Previous studies have also shown that INR can solve inverse problems in the medical imaging domain in an unsupervised manner, such as CT image reconstruction [25], [26] and undersampled MRI [27]. Our previous work utilized INR to reconstruct photoacoustic images under sparse and limited-view conditions [28].

In this work, we explored the use of INR to accelerate the image reconstruction process in PAM and improve the resolution of AR-PAM. The effectiveness of INR-based sparse sampling deconvolution was first validated through simulated vascular images. The results showed that undersampled PAM data could be accurately restored, and the resolution was enhanced. Subsequently, we further verified the INR model using images of leaf veins and microvasculature of a mouse brain. The findings demonstrated that INR not only achieved high-fidelity restoration of feature sizes in undersampled AR-PAM data but also maintained good continuity in the images. These discoveries indicate that our method can significantly increase the speed of AR-PAM scanning while improving resolution, presenting broad prospects for application in biomedical imaging. The contributions of this paper are summarized as follows:

1) For the first time, we consider the use of INR to simultaneously perform sparse image reconstruction and deconvolution for AR-PAM. This novel approach aims to enhance both the imaging speed and resolution of AR-PAM.
2) In our method, the PSF of AR-PAM is learnable, which reduces the difficulty of accurately measuring the system's PSF. This feature makes our approach more flexible and adaptable, allowing it to automatically adjust and optimize the PSF under different imaging conditions. Consequently, this further enhances the quality and robustness of the image reconstruction.

II. PRELIMINARY

A. Problem Formulation

The image formation process of AR-PAM can be represented by the following linear model:

$$\mathbf{y}=D_k(\kappa\otimes\mathbf{x}+\mathbf{n}), \tag{1}$$

The convolution process is represented by $\otimes$, where κ denotes the PSF, and $D_k(\cdot)$ represents the k-fold uniform downsampling operation. Here, $\mathbf{n}\in\mathbb{R}^{m\times m}$ represents system noise. The inverse problem aims to recover the unknown OR-PAM image $\mathbf{x}\in\mathbb{R}^{m\times m}$ from the real AR-PAM sparse measurement data $\mathbf{y}\in\mathbb{R}^{(m/k)\times(m/k)}$. In the AR-PAM acquisition process, to improve acquisition speed, the measurement data $\mathbf{y}$ undergoes downsampling, making its dimension significantly smaller than that of $\mathbf{x}$ (i.e., $m<n$). This downsampling operation renders the AR-PAM deconvolution problem ill-posed. A common approach to address this challenge is to formulate it as a regularized inversion problem, expressed as follows:

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}\in\mathbb{R}^{m\times m}}\|\mathbf{y}-D_k(\kappa\otimes\mathbf{x})\|^2+\epsilon\cdot\mathcal{R}(\mathbf{x}), \tag{2}$$

The term $\|\mathbf{y}-D_k(\kappa\otimes\mathbf{x})\|^2$ represents the data fidelity term, which ensures that $\mathbf{x}$ remains consistent with $\mathbf{y}$. The regularization term $\mathcal{R}(\mathbf{x})$ introduces explicit constraints on $\mathbf{x}$ to limit the solution space, such as the total variation (TV) regularization used for image smoothing. This has been shown to be useful in the regularization of deconvolution algorithms [2]. The hyperparameter $\epsilon>0$ controls the strength of the regularization.

B. AR-PAM Imaging Model

According to [21], the AR-PAM system can be modeled as a linear shift-invariant system around the focal region, and its imaging quality is determined by ultrasonic beamforming principles, including lateral and axial resolution, and depth of field. The field pattern of a plane circular transducer can be derived using the Rayleigh-Sommerfeld diffraction formula, which calculates the field at a spatial point $\overrightarrow{r_t}$ as:

$$\widetilde{\mathrm{E}}(\overrightarrow{r_t},\omega)=\frac{1}{i\lambda}\int_0^R\int_{-\pi}^{\pi}\widetilde{\epsilon}(r_1,\omega)e^{i\omega r_{01}/c}\frac{z}{r_{01}^2}r_1d\varphi_1dr_1, \tag{3}$$

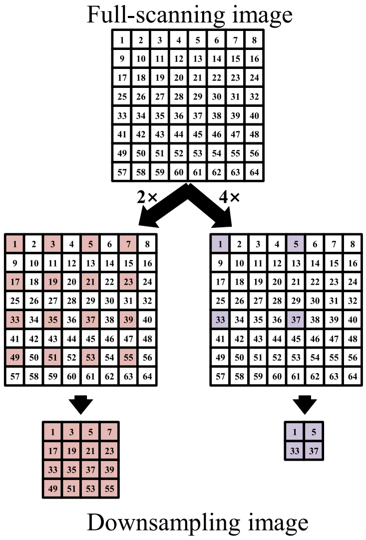
where r_{01} is the distance from the source to the observation point, λ is the wavelength, and c is the speed of sound.

For a spherically focused transducer, the field pattern can be obtained by adding a spherical compensation function:

$$\tilde{\epsilon}(r_1,\omega)=e^{-i\omega(\sqrt{L^2+r_1^2}-L)/c} \tag{4}$$

The PSF kernel κ is computed as the integral of the transducer's spectrum $T(\omega)$ weighted by the field distribution of the spherically focused transducer:

$$\kappa=\int T(\omega)\tilde{\mathbb{E}}(\vec{r_t},\omega)d\omega, \tag{5}$$



1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64

1	3	5	7
17	19	21	23
33	35	37	39
49	51	53	55

1	5
33	37

Fig. 1. The process of generating low-sampling images from the full-scanning ones.

C. Downsampling Images

The above operation can be considered as data obtained under dense full-scanning conditions. To speed up the acquisition of AR-PAM data, we can perform down-sampling to obtain sparsely sampled data [29]. This process is illustrated in Fig. 1 Specifically, using stride scaling, only half of the pixels along one lateral dimension are selected and used in the down-sampled image. This means that the down-sampled image (in terms of pixels) contains only one-quarter of the pixels of the fully scanned image. Consequently, the image acquisition time for the first approximation can be theoretically reduced to one-fourth. Similarly, when a four-times stride scaling is applied, the down-sampled image contains only one-sixteenth of the pixels compared to the full scan situation, thus the image acquisition time is expected to be reduced to one-sixteenth of the full scan scenario. The down-sampling method used here will generate a good approximation of the down-sampled images obtained experimentally.

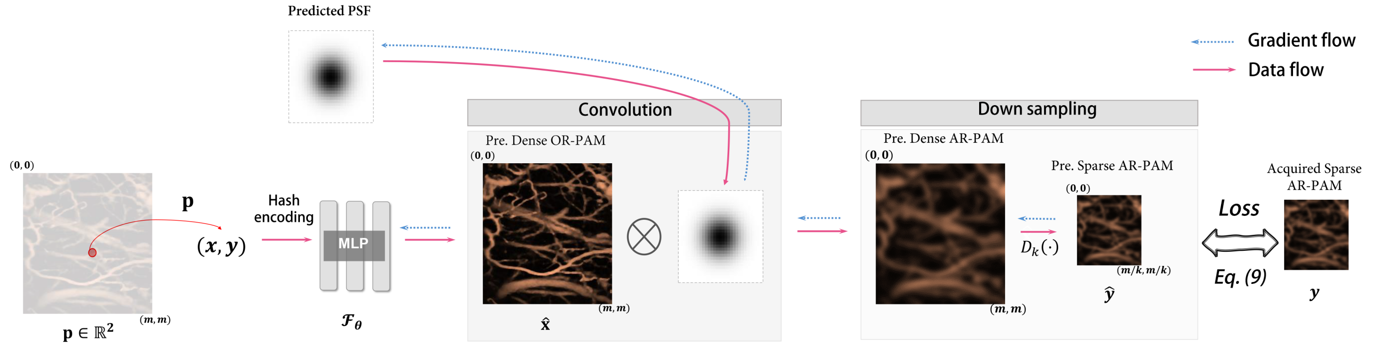


Fig. 2. The pipeline of the proposed sparse deconvolution reconstruction method is as follows: The network takes the densely sampled coordinates $\mathbf{p}$ of high-resolution images as input and predicts the image intensities I at these locations. Subsequently, the predicted results undergo PSF convolution and down-sampling operations. Finally, the network is optimized by minimizing the loss between the predicted sparsely sampled AR-PAM images and the actual acquired AR-PAM images. Here, $D_k(\cdot)$ represents the k-fold down-sampling operation.

D. Implicit Neural Representation for AR-PAM

INR is an emerging unsupervised paradigm that has shown significant potential in reconstruction tasks across various medical modalities, including CT and PAI [25], [26], [28]. In the INR framework, the desired OR-PAM image $\mathbf{x}$ is represented as a continuous function of pixel space positions:

$$f : \mathbf{p} = (x, y) \in \mathbb{R}^2 \to I(\mathbf{p}) \in \mathbb{R}, \tag{6}$$

where $I(\mathbf{p})$ denotes the intensity of the OR-PAM image $\mathbf{x}$ at the spatial coordinate $\mathbf{p}$. This function f is parameterized as a MLP network $\mathcal{F}_\Theta$, where Θ represents the learnable network parameters. Given the sparse measurements $\mathbf{y}$ from AR-PAM, the MLP $\mathcal{F}_\Theta$ is optimized using the following objective function:

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}(D_k(\kappa \otimes \mathbf{x}), \mathbf{y}), \tag{7}$$

$$I(\mathbf{p}) = \mathcal{F}_\Theta(\gamma(\mathbf{p})), \tag{8}$$

where $\gamma(\cdot)$ is an encoding module, such as positional encoding [30], Fourier encoding [31] for high-frequency components, or hash encoding [32] for fast rendering. $\mathcal{L}$ measures the discrepancy between the predicted measurements $D_k(\kappa \otimes \mathbf{x})$ and the true real AR-PAM sparse measurement $\mathbf{y}$, using either L_1 or L_2 norms. After optimization, the network $\mathcal{F}^*_\Theta$ generates the reconstructed image $\hat{\mathbf{x}}$, which closely approximates $\mathbf{x}$. This leverages the inherent inductive bias of neural networks on image manifolds [33].

III. METHOD

A. Proposed Framework

Our framework consists of four main components, as described in Fig. 2. Firstly, we predict the high resolution initial acoustic pressure for dense sampling in OR-PAM image $\mathbf{x}$ at given spatial positions $\mathbf{p}$. Secondly, we convolve the predicted image with the PSF to obtain a prediction of the image that would be acquired under dense sampling conditions in real AR-PAM. Thirdly, we uniformly down-sample the dense sampling image to simulate an image $\hat{\mathbf{y}}$ acquired under sparse sampling conditions in real AR-PAM. Finally, backpropagation of the computed loss function is used to update the parameters with the INR.

1) *Hash Encoding:* In the first step, the input space is divided into an $m \times m$ grid, with each grid cell assigned a two-dimensional coordinate (x, y). However, due to the spectral bias issue [33], fitting high-frequency signals using only two-dimensional coordinates is extremely challenging. To mitigate this problem, various encoding strategies [26], [31], [32] have been proposed to map low-dimensional inputs to high-dimensional feature vectors, which enables subsequent MLP to easily capture high-frequency components and thereby reduce approximation errors. In our method, we employ hash encoding. Hash encoding assigns a trainable feature to each input coordinate, transforming low-dimensional coordinates into high-dimensional features. This adaptive encoding strategy is task-specific and enables powerful fitting capabilities even with a relatively shallow MLP.

2) *Three-Layers MLP:* After hash encoding, the two-dimensional input coordinate $\mathbf{p} \in \mathbb{R}^2$ is encoded into a high-dimensional feature vector. Subsequently, a 3-layer MLP is used to transform the feature vector into image intensity I. The MLP consists of two hidden layers, each with 64 neurons and using the ReLU activation function, and an output layer with a Sigmoid activation function. Based on this encoding, the MLP predicts the initial acoustic pressure values with high resolution for dense sampling in AR-PAM.

3) *Learnable PSF:* The initial acoustic pressure image is then convolved with the learnable PSF to generate data that simulates real AR-PAM acquisition conditions. The PSF is initially estimated based on the measurements from the AR-PAM system. This initial PSF serves as a starting point. Then, in each iteration, the PSF is refined through a learning process.

To simulate the undersampling imaging process, we uniformly down-sample the densely sampled image. To constrain the INR, we include three different loss terms, a least-squares, a SSIM term and a total-variation term.

$$\mathcal{L} = (1-\lambda)\frac{1}{(m/k)^2}\sum_{\mathbf{r}\in\mathbf{y}}(\hat{\mathbf{y}}(\mathbf{r})-\mathbf{y}(\mathbf{r}))^2 + \lambda\mathrm{SSIM}(\hat{\mathbf{y}}(\mathbf{r}),\mathbf{y}(\mathbf{r})) + \epsilon TV(\hat{\mathbf{x}}), \tag{9}$$

We use L2 and SSIM loss as similarity metrics between the predicted image $\hat{\mathbf{y}}$ and the acquired corresponding image $\mathbf{y}$. Additionally, we employ TV loss to regularize the predicted OR-PAM image. TV loss has been shown to be effective in the regularization of deconvolution algorithms [34], and it also provides enhanced results in our case.

B. Implementation Details

The proposed framework is entirely implemented in PyTorch. Using hash encoding [32], the MLP can remain relatively small; therefore, we set it to have 3 layers with 64 neurons per layer. We employed the Adam optimizer [35] to minimize the loss function, with hyperparameters set as follows: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\lambda = 0.3$ and $\epsilon = 1e^{-5}$. The initial learning rate was set to 10^{-3}, with a decay factor of 0.5 every 500 epochs. The total number of training epochs was 5000, and the entire training process took approximately 2.3 minutes on a single NVIDIA 1080 Ti GPU. Notably, all the training parameters remained the same for different input views.

IV. EXPERIMENTS

A. Data

1) Leaf Vein Data: To further validate the feasibility of our method, we used experimentally acquired sparse PAM images to make the scenario closer to practical applications. In this dataset [29], AR-PAM scanned the same region of interest (ROI) with scanning strides of 8 μm and 16 μm. In this demonstration, we obtained 256× 256pixel fully sampled PAM images and corresponding 128×128 (or 64× 64pixel) low-sampled PAM images for the same ROI. The low-sampled PAM images were used as input, while the 256× 256pixel images served as references.

2) Mouse Brain Microvasculature Data: We utilized in vivo mouse brain microvascular system data obtained by the Duke University Photoacoustic Imaging Laboratory, which were acquired using the PAM system previously published in [36]. The PAM system has a lateral resolution of 5 μm and an axial resolution of 15 μm. Since the data were acquired using OR-PAM, according to the study by Zhang et al. [37], the data can be degraded from OR-PAM to AR-PAM resolution. Therefore, for this dataset, we consider the data obtained by OR-PAM as the ground truth.

3) In vivo Data: To validate the proposed algorithm, in vivo experiments were conducted using the switchable AR-OR-PAM imaging system described in [38]. The system employs a diode-pumped Nd:YAG laser (INNOSLAB, Edgewave) and a dye laser (Credo-DYE-N, Sirah, Spectra Physics) to emit 570 nm pulses at a maximum repetition rate of 5000 Hz. A motorized prism (CR1/M-Z7, Thorlabs) directs the laser either to a multimode fiber (M29L01, Thorlabs) for AR-PAM or, via a fiber coupler (F-91-C1, Newport), to a single-mode fiber (P1-460B-FC-1, Thorlabs) for OR-PAM. In AR-PAM mode, the beam is weakly focused using a conical lens and a custom optical condenser, whereas in OR-PAM mode, it is collimated and then strongly focused with achromatic lenses. Both modes use confocally aligned 50 MHz ultrasound transducers to acquire photoacoustic signals, which are amplified (ZFL-500LN, Mini Circuits) and digitized at 250 Ms/s (M4i.4420, Spectrum). The system is mounted on a three-axis motorized stage (X/Y: PLS 85; Z: VT 80, PI—Physik Instrumente) for raster scanning. The scanning step size is 3 μm (x) and 2 μm (y) for OR-PAM, and 15 μm for AR-PAM. Acoustic coupling is achieved using a water tank with a polyethylene film at the bottom. To obtain high-quality images, in vivo imaging was performed on the mouse ear, where blood vessels are relatively superficial.

B. Compared Methods & Evaluation Metrics

1) Compared Methods: Comparative Methods: In this work, we compare the proposed method with four deconvolution techniques: 1) low vision Deconvolution [39], a classical analytical reconstruction algorithm. Due to the downsampling issue inherent in AR-PAM, we apply bilinear interpolation for methods requiring resampling before performing the deconvolution operation; 2) RL Deconvolution [40], a maximum likelihood estimation method, which performs deconvolution given a known PSF; 3) DIP [41], a deep learning-based approach that leverages the inherent prior of the network to perform interpolation and deconvolution; 4) DPIR [42], a plug-and-play image restoration method. DPIR employs a highly flexible and efficient CNN-based denoiser trained to set a benchmark for deep denoising.

2) Evaluation Metrics: To quantitatively assess the performance of the comparative methods, we employ three metrics to measure the difference between the ground truth image and the deconvolved image. Specifically, we consider peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [43], and learned perceptual image patch similarity (LPIPS) [44]. PSNR and SSIM capture pixel-level differences, while LPIPS quantifies perceptual similarity.

C. Effectiveness of PSF Prediction

To validate the effectiveness of the proposed method in PSF prediction, we conducted an experimental study on the mouse brain dataset [36] to assess the accuracy of the predicted PSF. Fig. 3 illustrates the deconvolution results at different training epochs, along with the corresponding predicted PSF.

In this experiment, the initial PSF was randomly generated following a Gaussian distribution and was iteratively optimized during training. As shown in Fig. 3, the predicted PSF progressively converges to the ground truth PSF over the training process. Concurrently, the quality of the deconvolved images is significantly improved. From a qualitative perspective, the reconstructed images exhibit enhanced spatial resolution, clearer vascular structures, and richer details. Additionally, from a quantitative standpoint, we evaluated the peak PSNR at different epochs. The results indicate that after 5000 epochs, the PSNR

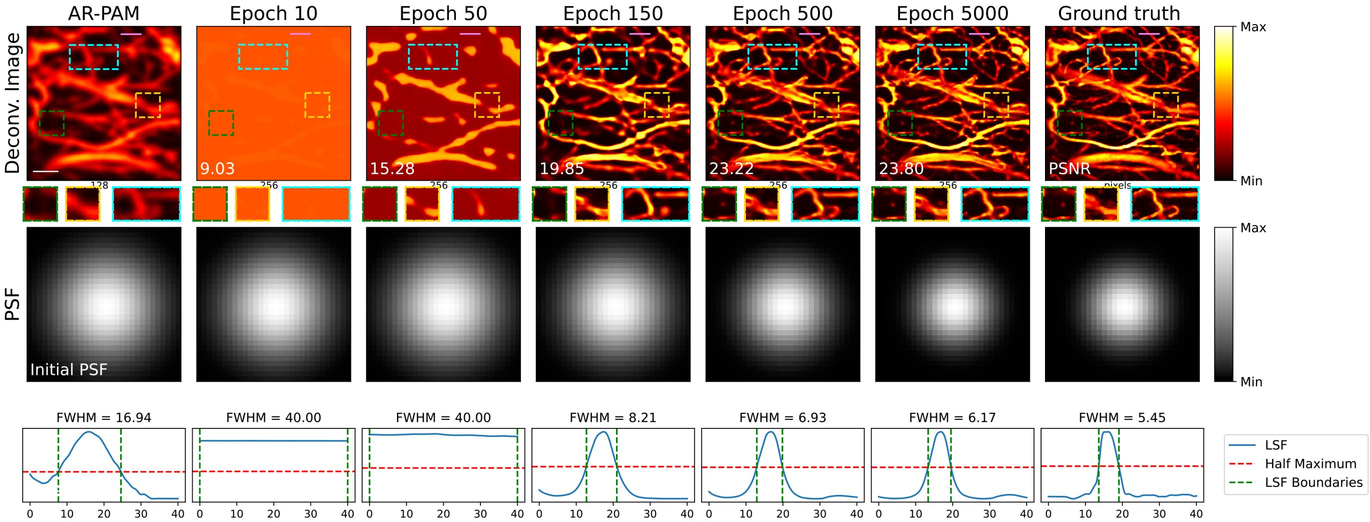


Fig. 3. The deconvolution results at different epochs, the predicted PSF of our method, and the FWHM of fine microvessels (purple solid line, LSF: line spread function) on the mouse brain data at a 2× downsampling rate. (Scale bar: 2mm) .

reached 23.8 dB, further verifying the accuracy of our PSF prediction.

To further describe the improvement in spatial resolution after network transformation, the impulse response of microvessels was measured using both OR-PAM and AR-PAM setups. As shown in Fig. 3, the normalized experimental PA data provides the line spread function (LSF, blue line). The full width at half maximum (FWHM) of the LSF is used to represent the system's lateral resolution. We observe that, as learning progresses in our method, the in situ response curve becomes steeper, indicating that previously blurred edges in AR-PAM imaging are now clearly distinguishable. Accordingly, the lateral resolution improves from approximately 16.94 in AR-PAM to 6.17, which is comparable to the ground truth value of 5.45 in OR-PAM, achieving nearly a threefold resolution enhancement. Although measurements may vary slightly at different positions along the microvessel edge, this result demonstrates the significant resolution improvement achieved through our method.

In summary, this experiment demonstrates that our method can effectively approximate the true PSF through iterative optimization and significantly improve the deconvolution reconstruction quality of AR-PAM images, thereby validating its effectiveness in PSF prediction.

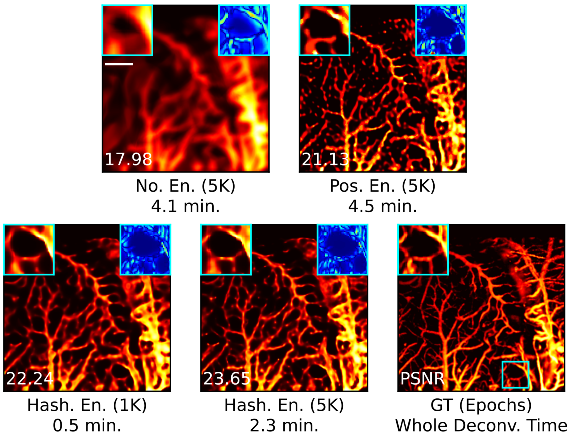


Fig. 4. The results of the model with three encoding modules on mouse brain data at a 2× downsampling rate. (Scale bar: 2mm).

D. *Effectiveness of Hash Encoding*

To validate the effectiveness of hash encoding [32], we conducted a comparative analysis of three different encoding modules based on the mouse brain dataset [36]: (1) No encoding, employing a pure 9-layer MLP; (2) Positional encoding, utilizing a 9-layer MLP with positional encoding [30]; and (3) Hash encoding, adopting a 3-layer MLP with hash encoding.

Fig. 4 presents both qualitative and quantitative PSNR results. The experimental findings indicate that, compared to the non-encoding approach, both positional encoding and hash encoding significantly enhance PSNR performance in all scenarios. For instance, in the deconvolution task with 2× downsampling, positional encoding improves PSNR by 3.15 dB (21.13 vs. 17.98), while hash encoding achieves an increase of 5.67 dB (23.65 vs. 17.98). This improvement is mainly attributed to the spectral bias problem [33], where a pure MLP inherently favors learning low-frequency signals during training. Therefore, incorporating encoding modules is crucial for enhancing the MLP's ability to learn high-frequency components.

Moreover, we observed that hash encoding slightly outperforms positional encoding. Overall, hash encoding benefits from its adaptive encoding strategy and shallower MLP architecture, achieving superior image quality and the fastest reconstruction speed. Numerical analysis shows that hash encoding requires only about 0.5 minutes to reach the same performance level as positional encoding, whereas positional encoding takes 4.5 minutes, approximately nine times longer.

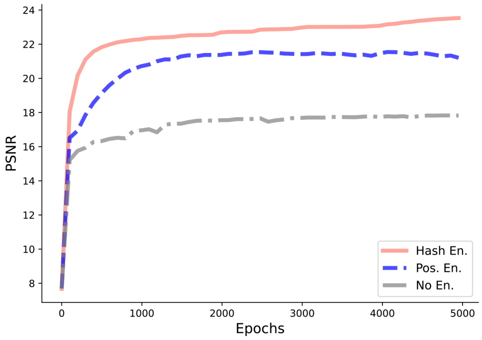


Fig. 5. The performance curves of the model with three encoding modules on mouse brain data at a 2× downsampling rate over training epochs.

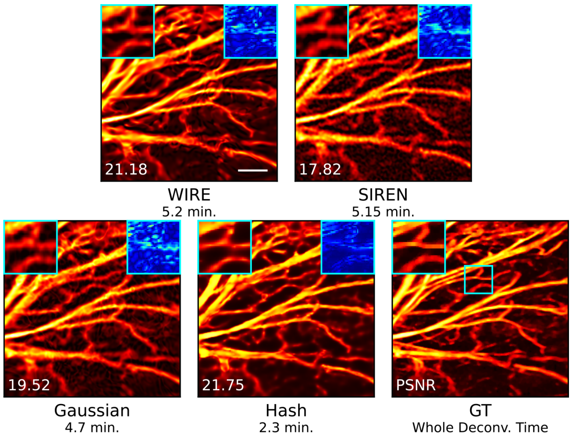


Fig. 6. The results of different nonlinear network architectures on mouse brain data at a 2× downsampling rate. (Scale bar: 2mm).

Fig. 5 further illustrates the performance curves of the three encoding modules during the training phase. The results clearly demonstrate that hash encoding yields the best performance.

E. Influence of Noises

In the AR-PAM acquisition process, noise is inevitably introduced due to various factors. To evaluate the performance of hash encoding in terms of reconstruction accuracy, reconstruction speed, and robustness to noise, we introduced speckle noise based on the mouse brain dataset [36], following the method in [45]. We then compared hash encoding with other nonlinear INR models, including WIRE [45], SIREN [22], and Gaussian [46]. The qualitative and quantitative results PSNR are shown in Figs. 6 and 7.

From the experimental results, we summarize the following key observations: 1) Hash encoding outperforms other nonlinear INR structures in both reconstruction accuracy and speed; 2) As noise is introduced, the performance of all models degrades, which aligns with expectations. The presence of noise exacerbates the ill-posed nature of the AR-PAM deconvolution problem, making models more prone to local minima. 3) Under high noise conditions, WIRE exhibits the best robustness, while hash encoding, which tends to fit high-frequency image content more effectively, is more susceptible to overfitting and thus performs relatively worse. However, by incorporating a TV-loss constraint, the robustness of hash encoding against noise can be significantly improved.

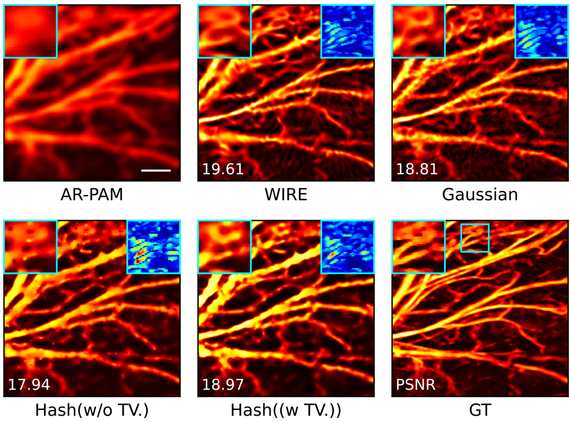


Fig. 7. The results of different nonlinear network architectures on mouse brain data with added shot noise at a 2× downsampling rate. (Scale bar: 2mm).

In summary, our method achieves superior performance in most cases in terms of reconstruction accuracy, computational efficiency, and robustness to noise.

F. Leaf Vein Experiment

To validate the effectiveness of our method, we first applied it, along with other comparative approaches, to AR-PAM images synthesized using the AR-PAM imaging model. As described in Section II-B, the PSF kernel at the focal plane can be derived from the AR-PAM imaging model, enabling the generation of synthetic AR-PAM images via convolution. We then applied several low vision deconvolution methods, including RL deconvolution, DIP, and DIPR, to perform sparse deconvolution on the AR-PAM images. The results were compared with those obtained using our proposed method. Fig. 8 presents the qualitative deconvolution results on the leaf vein dataset at different downsampling rates. The low vision deconvolution method failed to produce satisfactory results, recovering only partial details of the leaf veins. The RL deconvolution method demonstrated a noticeable improvement over low vision deconvolution, but the enhancement in resolution remained limited. DIP achieved better reconstruction, successfully recovering most of the leaf vein structures. However, discontinuities were still observed, particularly at 4× downsampling rate. DIPR, on the other hand, generated a smooth but overly blurred result, leading to a loss of fine details. In contrast, our proposed method consistently produced high-quality images that closely resembled the GT across all downsampling rates. The quantitative results, presented in Table I, further confirm the superiority of our approach. On the leaf vein dataset, our method outperformed the others in most cases. Notably, at 4× downsampling rate, our approach improved the PSNR by 0.08dB (23.72vs. 23.64) and by 0.29dB (23.72vs. 23.43). Interestingly, although DIPR is designed as a plug-and-play image restoration approach based on pre-trained

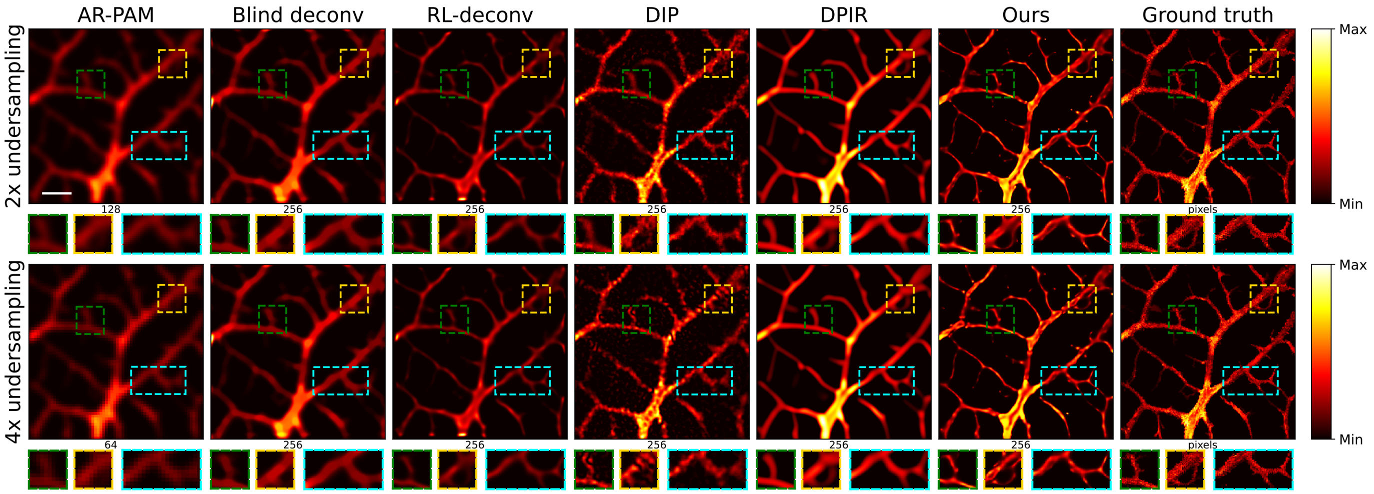


Fig. 8. Qualitative deconvolution results of all comparative methods on the leaf vein data at different downsampling rates. (Scale bar: 500 μm).

TABLE I
QUANTITATIVE COMPARISON OF DIFFERENT METHODS FOR LEAF DATA SPARSE DECONVOLUTION RECONSTRUCTION

Downsampling	Method	SSIM↑	PSNR↑	LPIPS↓
2×	Blind deconv. [39]	0.5860	18.74	0.4774
	RL deconv. [40]	0.6745	20.86	0.4001
	DIP [41]	0.6125	**24.32**	0.2871
	DPIR [42]	0.7731	23.43	0.4279
	Ours	**0.8166**	23.96	**0.2869**
4×	Blind deconv. [39]	0.4991	17.11	0.4805
	RL deconv. [40]	0.5884	19.71	0.4247
	DIP [41]	0.5118	23.64	0.3612
	DPIR [42]	0.7564	23.35	0.4427
	Ours	**0.8077**	**23.72**	**0.2886**

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT METHODS FOR MOUSE BRAIN DATA SPARSE DECONVOLUTION RECONSTRUCTION

Downsampling	Method	SSIM ↑	PSNR↑	LPIPS↓
2×	Blind deconv. [39]	0.3435	16.26	0.4053
	RL deconv. [40]	0.3890	17.26	0.3754
	DIP [41]	0.5166	20.31	0.2721
	DPIR [42]	0.5700	20.37	0.3504
	Ours	**0.7150**	**22.33**	**0.1399**
4×	Blind deconv. [39]	0.3031	15.70	0.4110
	RL deconv. [40]	0.3271	16.53	0.3816
	DIP [41]	0.4527	19.69	0.3079
	DPIR [42]	0.5517	20.16	0.3770
	Ours	**0.7112**	**22.00**	**0.1561**

models, its performance was inferior to unsupervised methods in our experiments. This can be attributed to the fact that DIPR is primarily trained on natural images and does not incorporate the PSF characteristics of AR-PAM. In contrast, our method is image-specific and demonstrates superior performance on the AR-PAM leaf vein dataset, achieving both high-resolution restoration and improved quantitative metrics.

G. Mouse Brain Data Experiment

To further validate the deconvolution capability of our method on more complex vascular structures, we conducted experiments on synthetic AR-PAM in vivo mouse brain data. Fig. 9 presents the qualitative deconvolution results at different downsampling rates. From the qualitative analysis, it can be observed that when processing images with more intricate structures, both low vision deconvolution and RL deconvolution fail to produce satisfactory reconstructions, leaving significant blurring in large regions. The DIP method maintains relatively good performance in this experiment. However, in vascular images with complex structures, the continuity issue becomes more pronounced, especially at higher downsampling rates. The DIPR method produces smoother results, but its over-smoothing effect leads to the loss of crucial vascular details. In contrast, our method demonstrates superior performance in AR-PAM images with intricate vascular structures. It not only accurately restores the vascular edge details but also maintains the integrity and continuity of internal structures, effectively reducing blurring and preserving information. The quantitative analysis further supports the superiority of our approach.

Table II presents the PSNR evaluation results for the mouse brain dataset. Our method consistently achieves the best performance across all experimental settings, demonstrating significant improvements compared to DIP and DIPR. Specifically, under the 4× downsampling condition, our method improves the PSNR by 2.31 dB compared to DIP (22.00 vs. 19.69) and by 1.84 dB compared to DIPR (22.00 vs. 20.16). These results further confirm the remarkable ability of our method to enhance AR-PAM resolution. Moreover, it remains robust to variations in data complexity, consistently performing well even in highly complex biological structures.

H. In Vivo Experiment

To further validate the effectiveness of the proposed method on real AR-PAM data, we conducted AR-PAM and OR-PAM imaging experiments on the same biological tissue region. Given that OR-PAM provides higher spatial resolution, its imaging

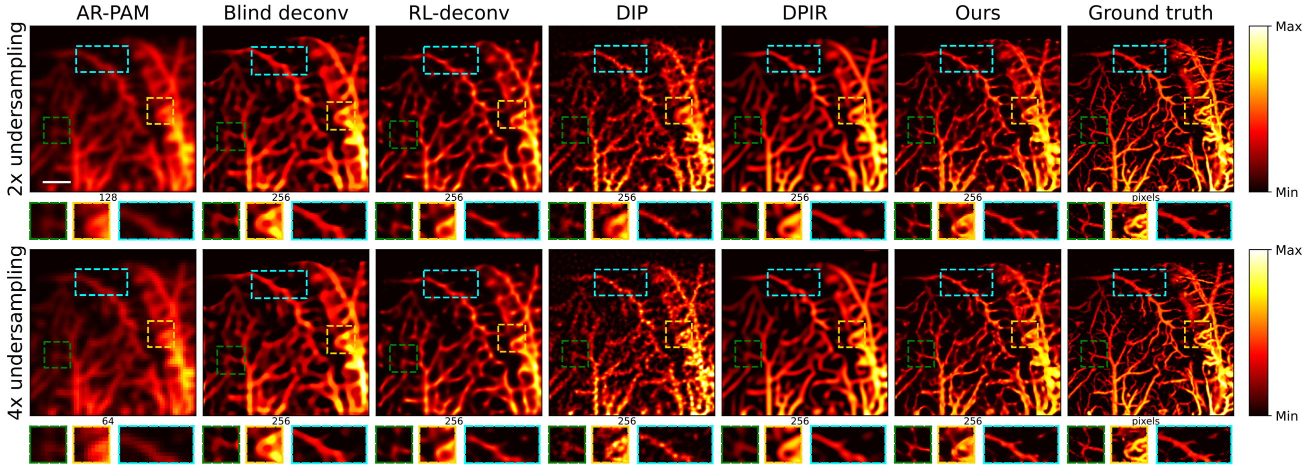


Fig. 9. Qualitative deconvolution results of all comparative methods on the mouse brain at different downsampling rates. (Scale bar: 2 mm).

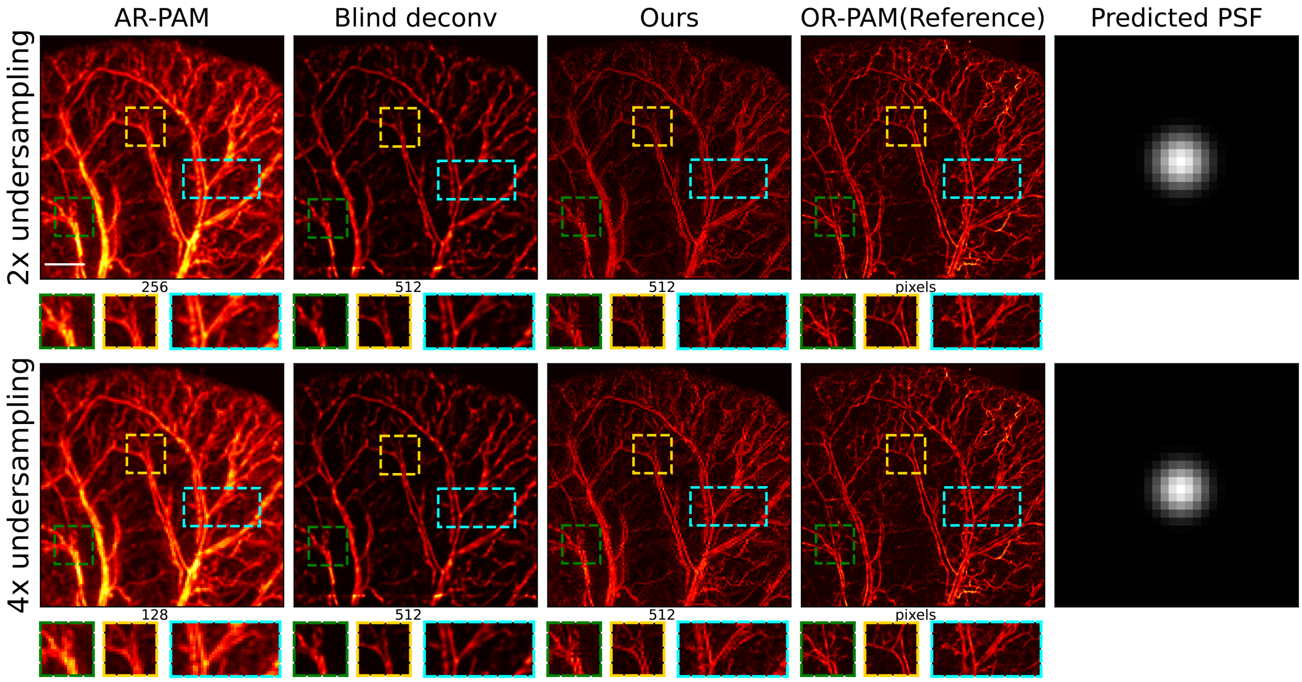


Fig. 10. Enhancement results of in vivo mouse ear AR-PAM images using different methods, along with our method's PSF predictions at various downsampling rates. (Scale bar: 5 mm).

results serve as a reference standard for performance evaluation. Fig. 10 presents the imaging results of AR-PAM and OR-PAM, where AR-PAM images under different downsampling rates are further enhanced using both low vision deconvolution methods and our proposed approach. Additionally, Fig. 10 shows the PSF predicted by our method.

The experimental results demonstrate that our method achieves superior perceptual resolution enhancement and structural detail preservation while effectively reducing artifacts. To further evaluate reconstruction fidelity, we analyzed the correspondence between the OR-PAM image and its enhanced version. The results indicate that our method significantly improves the quality of AR-PAM images. However, despite the noticeable improvements, the reconstructed images still exhibit lower resolution compared to the OR-PAM images.

This phenomenon can be attributed to the intrinsic differences between the two imaging modalities. AR-PAM offers deeper tissue penetration than OR-PAM, resulting in AR-PAM images containing more structural information from deeper tissue layers. Consequently, our method, while enhancing resolution, also amplifies these additional details, leading to differences between the reconstructed AR-PAM images and the OR-PAM reference images.

V. DISCUSSION

In this work, we explore sparse deconvolution techniques based on INR to enhance the resolution of AR-PAM images. The novelty of this work lies in the first application of INR methods to improve the resolution of focal region AR-PAM images. Compared to other deep learning methods, our approach does not require a large amount of paired data, thereby improving cost-effectiveness and reducing the difficulty of data acquisition. We evaluate our method against low vision deconvolution, RL deconvolution, DIP, and DIPR methods. Experimental results demonstrate that our approach excels in sparse deconvolution reconstruction, achieving sharper edges and finer details compared to these methods, while remaining robust to variations in data distribution and structural complexity. The sparse deconvolution achieved through the INR method can be explained as follows. First, each pixel in the high-resolution densely sampled image is assigned a two-dimensional coordinate, which is then hashed and passed as input to the INR network. Our network consists of two hidden layers, each containing 64 parameters. Finally, the network outputs the pixel intensity values of the OR-PAM image at each location. After obtaining the densely sampled OR-PAM image, convolution and down-sampling operations are performed, followed by loss computation and backpropagation optimization based on the actual acquired data. Notably, since it is challenging to obtain an exact PSF in practical applications, our network incorporates a learnable PSF, allowing for better PSF estimation. Additionally, due to the use of hashing, the entire reconstruction process is highly efficient, requiring only 2.3 minutes for 5000 iterations. Compared to supervised deep learning methods that demand a large amount of paired training data, our approach exhibits superior generalization, as supervised networks often struggle with unseen data types. Given that AR-PAM is primarily used for vascular imaging, we selected simulated vascular data for our experiments. The results indicate that, compared to other non-data-driven and pre-trained plug-and-play methods, our method achieves clearer sparse reconstruction and deconvolution, effectively reducing image blurring. To further validate the effectiveness of our approach, we also conducted experiments on real AR-PAM-acquired in vivo mouse ear data. The results demonstrate that our method maintains excellent reconstruction performance on real data, significantly enhancing image details and resolution.

VI. CONCLUSION

In this work, we explore sparse deconvolution techniques based on INR and propose a deep learning framework leveraging INR. We compare the proposed method with low vision deconvolution, RL deconvolution, DIP, and DPIR through both simulation and in vivo experiments. The results demonstrate that our method consistently achieves higher resolution and richer image details across structures of varying complexity, exhibiting significant advantages over existing deconvolution methods. In the in vivo AR-PAM imaging experiment of a mouse ear, our method successfully achieves high-fidelity deconvolution of sparse vasculature while maintaining excellent structural continuity. This work holds promise for enhancing the resolution of AR-PAM images while accelerating data acquisition. Furthermore, the proposed method demonstrates strong generalizability and can be extended to other imaging modalities to improve their resolution and image quality.

REFERENCES

[1] C. Paul Beard, “Biomedical photoacoustic imaging,” *Interface Focus*, vol. 1, pp. 602–631, 2011.

[2] C. Tian et al., “Dual-pulse nonlinear photoacoustic technique: A practical investigation,” *Biomed. Opt. Exp.*, vol. 6, no. 8, pp. 2923–2933, 2015.

[3] L. V. Wang, “Multiscale photoacoustic microscopy and computed tomography,” *Nature Photon.*, vol. 3, no. 9, pp. 503–509, 2009.

[4] L. Nie and X. Chen, “Structural and functional photoacoustic molecular tomography aided by emerging contrast agents,” *Chem. Soc. Rev.*, vol. 43, no. 20, pp. 7132–7170, 2014.

[5] M. M. Menger et al., “Photoacoustic imaging for the study of oxygen saturation and total hemoglobin in bone healing and non-union formation,” *Photoacoustics*, vol. 28, 2022, Art. no. 100409.

[6] S. Jeon, J. Kim, D. Lee, J. W. Baik, and C. Kim, “Review on practical photoacoustic microscopy,” *Photoacoustics*, vol. 15, 2019, Art. no. 100141.

[7] R. Cao et al., “Optical-resolution photoacoustic microscopy with a needle-shaped beam,” *Nature Photon.*, vol. 17, pp. 89–95, 2022.

[8] Q. Zhao et al., “Quantitative analysis on in vivo tumor-microvascular images from optical-resolution photoacoustic microscopy,” *J. Biophotonics*, vol. 12, 2019, Art. no. e201800421.

[9] R. Bi, Q. Ma, H. Mo, M. Olivo, and Y. Pu, “Optical-resolution photoacoustic microscopy of brain vascular imaging in small animal tumor model using nanosecond solid-state laser,” *Neurophotonics Biomed. Spectrosc.*, pp. 159–187, 2019, doi: 10.1016/B978-0-323-48067-3.00008-1.

[10] J. Ahn et al., “Fully integrated photoacoustic microscopy and photoplethysmography of human in vivo,” *Photoacoustics*, vol. 27, 2022, Art. no. 100374.

[11] J. Brunker and P. C. Beard, “Acoustic resolution photoacoustic Doppler velocimetry in blood-mimicking fluids,” *Sci. Rep.*, vol. 6, 2016, Art. no. 20902.

[12] D. Egolf, K. W. R. Chee, and R. Z. James, “Sparsity-based reconstruction for super-resolved limited-view photoacoustic computed tomography deep in a scattering medium,” *Opt. Lett.*, vol. 43, no. 10, pp. 2221–2224, 2018.

[13] C. Zhang, K. Maslov, J. Yao, and L. V. Wang, “In vivo photoacoustic microscopy with 7.6-mathrmmum axial resolution using a commercial 125-mhz ultrasonic transducer,” *J. Biomed. Opt.*, vol. 17, pp. 116016–116016, 2012.

[14] J. Chen, R. Lin, H. Wang, J. Meng, H. Zheng, and L. Song, “Blind-deconvolution optical-resolution photoacoustic microscopy in vivo,” *Opt. Exp.*, vol. 21, no. 6, pp. 7316–7327, 2013.

[15] F. Feng, S. Liang, and S.-L. Chen, “Image enhancement in acoustic-resolution photoacoustic microscopy enabled by a novel directional algorithm,” *Biomed. Opt. Exp.*, vol. 13, no. 2, pp. 1026–1044, 2021.

[16] I. Loc and M. B. Unlu, “Accelerating photoacoustic microscopy by reconstructing undersampled images using diffusion models,” *Sci. Rep.*, vol. 14, 2024, Art. no. 16996.

[17] X. Wang et al., “Sparse sampling photoacoustic reconstruction with graph regularization group sparse dictionary,” *Appl. Opt.*, vol. 63, pp. 5292–5302, 2024.

[18] X. Zhu et al., “Real-time whole-brain imaging of hemodynamics and oxygenation at micro-vessel resolution with ultrafast wide-field photoacoustic microscopy,” *Light Sci. Appl.*, vol. 11, 2022, Art. no. 138.

[19] C. Yang, H. Lan, F. Gao, and F. Gao, “Review of deep learning for photoacoustic imaging,” *Photoacoustics*, vol. 21, 2020, Art. no. 100215.

[20] S.-W. Cheng, Y. Zhou, J. Chen, H. Li, L. Wang, and P. Lai, “High-resolution photoacoustic microscopy with deep penetration through learning,” *Photoacoustics*, vol. 25, 2021, Art. no. 100314.

[21] F. Feng, S. Liang, J. Luo, and S.-L. Chen, “High-fidelity deconvolution for acoustic-resolution photoacoustic microscopy enabled by convolutional neural networks,” *Photoacoustics*, vol. 26, 2022, Art. no. 100360.

[22] V. Sitzmann, N. P. Julien, A. W. Martel, D. B. Bergman, Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” *Adv. neural inf. process. syst.*, vol. 33, pp. 7462–7473, 2020.

[23] K. Zhang, G. Riegler, N. Snavely, and V. Koltun, “NERF : Analyzing and improving neural radiance fields,” 2020, *arXiv:2010.07492*.

[24] J. J. Park, P. R. Florence, J. Straub, R. A. Newcombe, and S. Lovegrove, “DeepSDF: Learning continuous signed distance functions for shape representation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 165–174.

[25] Q. Wu, R. Feng, H. Wei, J. Yu, and Y. Zhang, “Self-supervised coordinate projection network for sparse-view computed tomography,” *IEEE Trans. Comput. Imag.*, vol. 9, pp. 517–529, 2023.

[26] Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov, “CoIL: Coordinate-based internal learning for tomographic imaging,” *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1400–1412, 2021.

[27] L. Shen, J. M. Pauly, and L. Xing, “NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 770–782, Jan. 2024.

[28] Y. Xiao, Y. Shen, B. Yao, X. Cai, Y. Zhang, and F. Gao, “Limited-view photoacoustic imaging reconstruction via high-quality self-supervised neural representation, *Photoacoustics*, 2024, Art. no. 100685.

[29] J. Zhou, D. He, X. Shang, Z. Guo, S.-L. Chen, and J. Luo, “Photoacoustic microscopy with sparse data by convolutional neural networks,” *Photoacoustics*, vol. 22, 2021, Art. no. 100242.

[30] B. Mildenhall et al., “NeRF: Representing scenes as neural radiance fields for view synthesis,” *Commun. ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[31] M. Tancik et al., “Fourier features let networks learn high frequency functions in low dimensional domains, *Adv. neural inf. process. syst.*, vol. 33, pp. 7537–7547, 2020.

[32] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Trans. Graph.*, vol. 41, pp. 1–15, 2022.

[33] N. Rahaman et al., “On the spectral bias of neural networks,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 5301–5310.

[34] C. Dalitz, R. P. Frohlich, and T. Michalk, “Point spread functions and deconvolution of ultrasonic images,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, no. 3, pp. 531–544, Mar. 2015.

[35] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, 2014, *arXiv:1412.6980*.

[36] A. DiSpirito et al., “Reconstructing undersampled photoacoustic microscopy images using deep learning,” *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 562–570, Feb. 2021.

[37] Z. Zhang et al., “Deep and domain transfer learning aided photoacoustic microscopy: Acoustic resolution to optical resolution,” *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3636–3648, Dec. 2022.

[38] M. Moothanchery and M. Pramanik, “Performance characterization of a switchable acoustic resolution and optical resolution photoacoustic microscopy system,” *Sensors*, vol. 17, 2017, Art. no. 357.

[39] M. R. Chowdhury, J. Qin, and Y. Lou, “Non-blind and blind deconvolution under poisson noise using fractional-order total variation,” *J. Math. Imag. Vis.*, vol. 62, pp. 1238–1255, 2020.

[40] D. S. C. Biggs and M. Andrews, “Acceleration of iterative image restoration algorithms,” *Appl. Opt.*, vol. 36, no. 8, pp. 1766–1775, 1997.

[41] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, “Deep image prior,” *Int. J. Comput. Vis.*, vol. 128, pp. 1867–1888, 2017.

[42] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. V. Gool, and R. Timofte, “Plug-and-play image restoration with deep denoiser prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[44] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.

[45] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. Baraniuk, “Wire: Wavelet implicit neural representations,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 18507–18516.

[46] S. Ramasinghe and S. Lucey, “Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps,” in *Eur. Conf. Comput. Vis.*, 2021, pp. 142–158.