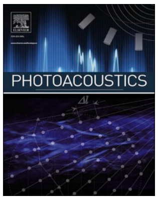
Photoacoustics 43 (2025) 100720

Contents lists available at ScienceDirect

Photoacoustics

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



High resolution photoacoustic vascular image reconstruction through the fast residual dense generative adversarial network

Yameng Zhang[a,b,1], Hua Tian[a,1], Min Wan[b], Shihao Tang[b], Ziyun Ding[c], Wei Huang[a], Yamin Yang[b], Weitao Li[b,*]

[a] School of Computer Engineering, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China
[b] Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China
[c] School of Engineering, University of Birmingham, Birmingham B15 2TT, UK

ARTICLE INFO

Keywords:
Photoacoustic microscopy
Generative adversarial network
Vascular imaging
Super-resolution imaging
Residual dense module

ABSTRACT

Photoacoustic imaging is a powerful technique that provides high-resolution, deep tissue imaging. However, the time-intensive nature of photoacoustic microscopy (PAM) poses a significant challenge, especially when high-resolution images are required for real-time applications. In this study, we proposed an optimized Fast Residual Dense Generative Adversarial Network (FRDGAN) for high-quality PAM reconstruction. Through dataset validation on mouse ear vasculature, FRDGAN demonstrated superior performance in image quality, background noise suppression, and computational efficiency across multiple down-sampling scales (×4, ×8) compared to classical methods. Furthermore, in the in vivo experiments of mouse cerebral vasculature, FRDGAN achieves the improvement of 2.24 dB and 0.0255 in peak signal-to-noise ratio and structural similarity metrics in contrast to SRGAN, respectively. Our FRDGAN method provides a promising solution for fast, high-quality PAM microvascular imaging in biomedical research.

1. Introduction

Photoacoustic imaging (PAI) is an advanced biomedical optical technique that detects photoacoustic signals generated when tissues absorb pulsed laser light and emit ultrasound waves. PAI offers several advantages, including high spatial resolution, substantial imaging depth, non-invasiveness, and the ability to perform imaging without the need for labeling or contrast agents [1,2]. The photoacoustic signals closely reflect the optical absorption properties of biological tissues, enabling the visualization of a wide range of physiological and pathological information [3,4]. Therefore, PAI has gained considerable attention as a promising modality for both preclinical and clinical imaging applications [5,6].

In contrast to Photoacoustic Tomography (PAT), which provides volumetric imaging by using diffuse light and detecting photoacoustic signals at a larger scale, Photoacoustic Microscopy (PAM) employs focused optical illumination and ultrasound detection. This configuration enables PAM to achieve superior spatial resolution, ranging from sub-micrometer to sub-millimeter, while maintaining imaging depths from several hundred micrometers to a few millimeters [7,8]. PAM offers distinct advantages in high-resolution (HR) imaging, particularly in providing detailed visualization of fine structures such as microvasculature, and tissue morphology [9,10]. In PAM, the axial resolution is determined by the bandwidth of the ultrasound detector, while the lateral resolution is governed by the size of the optical or acoustic focus. To ensure optimal spatial fidelity and image quality, following the Nyquist sampling criterion, the scanning step size in PAM must be no larger than half of the anticipated spatial resolution [11].

Imaging speed remains a significant challenge in PAM due to the complexities of simultaneously scanning the excitation beam and detecting the resultant ultrasonic waves in an aqueous environment. Significant advances have been made in enhancing the speed of PAM, primarily through the development of hardware technologies such as water-immersible Micro-Electro-Mechanical Systems (MEMS), galvanometer scanners, and microlens arrays [12–14]. However, rather than the speed of the scanning motors, the laser's pulse repetition rate remains a limiting factor for the overall imaging speed. This creates a fundamental trade-off between imaging speed and spatial resolution in

---

* Corresponding author.
E-mail address: liweitao@nuaa.edu.cn (W. Li).
[1] *co-first author*

https://doi.org/10.1016/j.pacs.2025.100720
Received 20 January 2025; Received in revised form 17 March 2025; Accepted 31 March 2025

PAM systems. Achieving higher resolution often necessitates slowing down the scanning process to acquire more detailed information, which in turn limits the speed. Conversely, Increasing the imaging speed typically requires reducing the number of pulses per unit time, which may reduce the spatial resolution. For instance, down-sampling is often used to accelerate the scanning process during PAM, but they come at the cost of reduced resolution. Improving the image quality of these down-sampled PAM images through advanced algorithms is therefore essential for real-time applications.

Recently, deep learning (DL) techniques have progressively been employed in biomedical imaging reconstruction and denoising, offering new opportunities for achieving fast, high-quality PAM imaging [15–18]. For example, Sharma *et al.* proposed a fully dense network to enhance resolution and reduce noise and artifacts of PAM images [19]. The DiSpirito team employed a fully dense U-Net for reconstructing down-sampled PAM images, which significantly improved image quality [20]. Liu *et al.* introduced an attention-steered network (AS-Net) for photoacoustic reconstruction, utilizing multi-feature fusion to remove artifacts and accelerate processing speed [21]. Zhou *et al.* developed a convolutional neural network (CNN) model incorporating residual blocks and channel-wise attention, which enhanced the performance of sparser image reconstruction [22]. Additionally, Yang *et al.* presented a unified PAM image reconstruction network (UPAMNet), designed to simultaneously achieve both super-resolution and denoising of PAM images [23].

Currently, a key focus in PAM research is the development of methods to adapt DL algorithms specifically for PAM, addressing issues such as image reconstruction, denoising, and improving the efficiency of sparse scanning. However, PAM presents several unique challenges in image reconstruction and denoising when processed with traditional DL techniques. Firstly, the scarcity of large, labeled training datasets for PAM images cause many traditional DL models with a high number of parameters to be prone to overfitting when trained on limited data. Secondly, PAM images typically lack the rich textures and distinct features that are commonly found in natural biological images, which are essential for many DL models. GANs can learn from smaller datasets and produce detailed, realistic outputs even when training data is scarce. Their adversarial training framework allows them to capture the underlying distribution of PAM images, effectively enhancing resolution and reducing noise without overfitting [24,25]. Zhang *et al.* proposed a domain transfer learning technique under GAN to enable the enhancement target with limited amount of matched in vivo AR to OR imaging data [26]. He *et al.* proposed an attention enhanced GAN to extract image features and adaptively remove various levels of Gaussian, Poisson, and Rayleigh noise for SR image reconstruction [27].

Traditional pixel-level metrics, such as Mean Squared Error (MSE), widely used in biomedical image reconstruction, often lead to over-smoothing when applied to PAM images [28]. As a result, directly transferring super-resolution networks designed for natural images to PAM tasks frequently leads to suboptimal image quality. Besides, although PAM images reconstructed with the traditional dense residual connection module exhibit good quality, their large model capacity and high computational demands significantly limit their application in real-world scenarios that require efficiency or real-time implementation [29,30].

To address the aforementioned challenges, we have proposed the Fast Residual Dense Generative Adversarial Network (FRDGAN) for two essential PAM image reconstruction tasks: super-resolution and computational efficiency. The main contributions of this paper include: (i) We innovatively integrated the Fast Residual Dense Block (FRDB) module and depth-wise separable convolutions (DSConv) into the SRGAN for PAM image reconstruction for the first time; (ii) We successfully acquired an experimental dataset consisting of 188 high-quality optical-resolution photoacoustic microscopy (OR-PAM) images (6000 ×500 ×256 pixels) of mouse ear vasculature to train the new model performance; (iii) We comprehensively demonstrated the performance of our new model using the acquired dataset in terms of image quality metrics (e.g., PSNR and SSIM) and computational efficiency(e.g., Inference Time and GPU Memory usage). (iv) Beyond synthetic downsampled data, we validated FRDGAN's performance on real low-speed PAM systems using a human hair phantom, a microfluidic chip, porcine liver tissue, and in vivo mouse cerebral vasculature. The results demonstrated that FRDGAN effectively reconstructed complex vascular structures with high quality. These findings highlight the feasibility of applying this approach in biomedical research and its potential for clinical applications.

2. Methodology

2.1. Classic SRGAN architecture

Following Cunningham [31], the SRGAN network consists of a discriminator network$D_{\theta_D}$and a generator network G_{θ_G}, and it solves the adversarial minimax problem through alternating optimization, as expressed in Eq.1:

$$\min_{\vartheta_G}\max_{\vartheta_D}\mathbb{E}_{I^{HR}\sim p_{train}}(I^{HR})[\log G_{\theta_D}(I^{HR})]+\mathbb{E}_{I^{LR}\sim p_G}(I^{LR})[\log(1-D_{\theta_D}G_{\theta_G}(I^{LR}))] \tag{1}$$

where $I^{LR}$is the LR version of its HR counterpart I^{HR}. $\theta_D$and θ_G represent the weights and biases of the discriminator network$D_{\theta_D}$and the generator network G_{θ_G}, respectively. The primary concept of this formulation is to train a generative model G that aims to deceive a differentiable discriminator D to differentiate between super-resolved images and real images. The ultimate objective is to develop a generative model capable of reconstructing the corresponding I^{HR} from a given LR input image I^{LR}.

The structure of the proposed network module, which builds upon the SRGAN framework, is illustrated in Fig. 1. To address the challenges of excessive model parameters and overly sharpened high-frequency details present in the original SRGAN design, we introduced the FRDB module as a replacement for the conventional SRResNet module [32]. Additionally, DSConv was integrated into the standard convolutional processes to reduce computational overhead. Detailed descriptions of these modifications are provided in Sections 2.2 and 2.4.

2.2. Generator network architecture

As shown in Fig. 1, our generator network is composed of four main components: a shallow convolutional net, FRDBs, a dense feature fusion net, and an up-sampling layer. This network takes a LR image (I^{LR}) as input and generates a super-resolution (SR) image (I^{SR}) that the discriminator cannot distinguish from the real HR image. The key improvement in our generator network is the replacement of the traditional SRResNet with the proposed FRDB. The detailed network architecture is depicted in Fig. 2.

A typical Residual Dense Block (RDB), originally proposed by Zhang [33] consists of densely connected layers, local feature fusion layers, and a local residual learning layer, forming a continuous memory mechanism. In an RDB, each convolutional layer can access all subsequent layers and pass on the information that needs to be retained. However, this dense connectivity also results in information redundancy. To address this, we optimized the dense connection module and proposed the FRDB specifically for photoacoustic imaging. The improved connection in the FRDB is mathematically defined as follows:

$$F_{d,2^k}=\sigma(W_{d,2^k}[F_{d-1},F_{d,2^k-2^0},\cdots F_{d,2^k-2^n},\cdots,F_{d,2^k-2^{k-1}}]) \tag{2}$$

where $F_{d-1}$and F_d denote the input and output of the d-th FRDB, respectively, both containing G_0 feature maps. σ denotes the leaky ReLU activation function, and $W_{d,2^k}$represents the weights of the 2^k-th convolution layer, with n satisfying the condition that $n\leq k-1$.We assume $F_{d,2^k-2^n}$consists of G feature-maps, so $[F_{d-1},F_{d,2^k-2^0},\cdots F_{d,2^k-2^n},\cdots,$

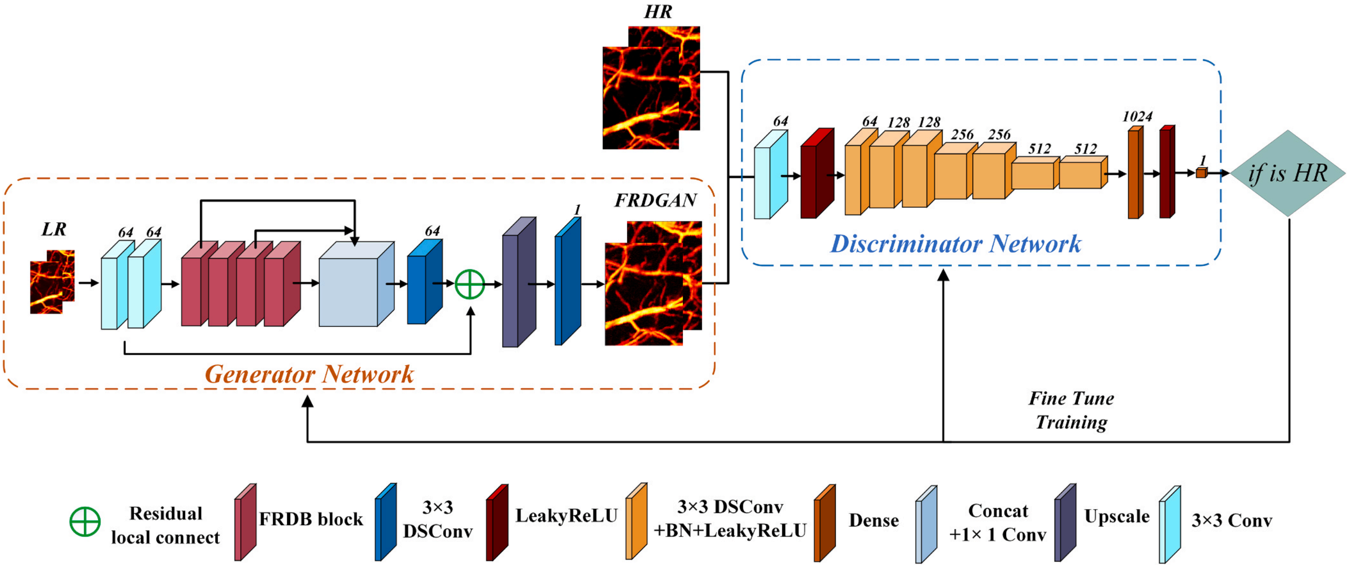


Fig. 1. Schematic representation of the FRDGAN architecture. The Generator Network takes LR input images and processes them through a sequence of layers, including 3 × 3 convolutions, fast residual dense blocks FRDBs, concatenation followed by a 1 × 1 convolution, a 3 × 3 DSConv, residual local connections, an upscaling layer, and a final 3 × 3 DSConv to generate HR outputs. The Discriminator Network evaluates the generated HR images to determine their authenticity by comparing them with real HR images. It includes a 3 × 3 standard convolution, a Leaky ReLU activation, a series of 3 × 3 DSConv with batch normalization (BN) and Leaky ReLU, a dense layer, a Leaky ReLU, and a final dense layer. Fine-tune training is employed to optimize the generator and discriminator collaboratively, enhancing the fidelity of super-resolved images.

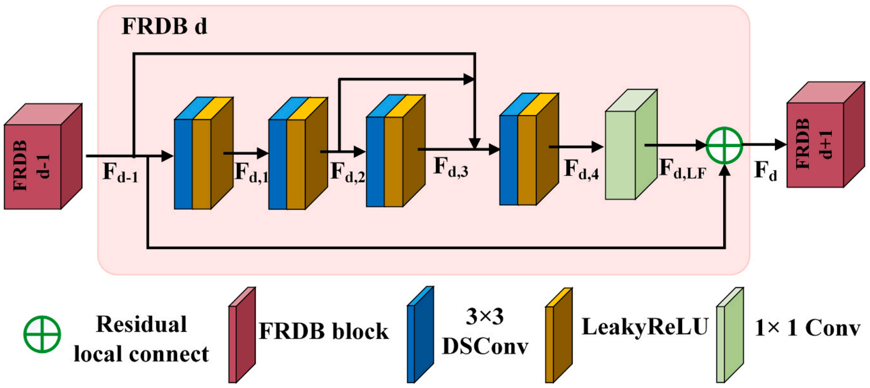


Fig. 2. Detailed structure of the Fast Residual Dense Block (FRDB) within the FRDGAN architecture. The FRDB consists of multiple sub-blocks ($F_{d,1}$ to $F_{d,4}$), each comprising a 3 × 3 DSConv and a ReLU activation. The features ($F_{d,LF}$) are passed through a 1 × 1 convolution to merge information before being added back to the input via a residual local connection, generating the output (F_d).

$F_{d,2^{k}-2^{k-1}}]$refers to the concatenation of the feature-maps produced by the$(d-1)$-*th* FRDB and by k convolution layers in the *d-th* FRDB, resulting in $G_0 + kG$ feature-maps. This improved connection strategy allows rapid and dense connectivity for each output feature, retaining sequential memory while simplifying the network structure. To reduce the dimensionality of the feature maps for subsequent residual learning, a 1 × 1 convolutional layer is introduced into $F_{d,2^{k}}$, yielding $F_{d,LF}$.

Local residual learning is further integrated into the FRDB to enhance information flow. The final output of the *d-th* FRDB can be expressed as:

$$F_d = F_{d,LF} + F_{d-1} \tag{3}$$

The F_d incorporates the residual structure into the dense network, connecting the state of the preceding layer $F_{d-1}$with subsequent ones. This continuous memory mechanism effectively integrates low-frequency and high-frequency features, enhancing the reconstruction quality of the network.

2.3. *Discriminator network architecture*

The discriminator network adopts the structure of SRGAN model [34], which consists of 8 convolutional layers. Each convolutional layer, except the initial layer for image input, is connected to Batch Normalization (BN) layers and utilizes Leaky ReLU as the activation function. After the convolutional processes, 512 feature maps are passed through two dense layers. Finally, a sigmoid activation function is applied to obtain the classification probability for determining whether the input is real or generated.

2.4. *Depth-wise separable convolution*

DSConv is utilized for all convolutional processes in FRDGAN. Unlike standard convolution, it typically executes depth-wise convolution first, followed by pointwise convolution [35], as illustrated in Fig. 3.

Assuming the input feature map has n channels with a size of $A \times B$, and a convolution is performed using a kernel of size C×C in case of stride length of 1, the output will be a feature map with m channels, also

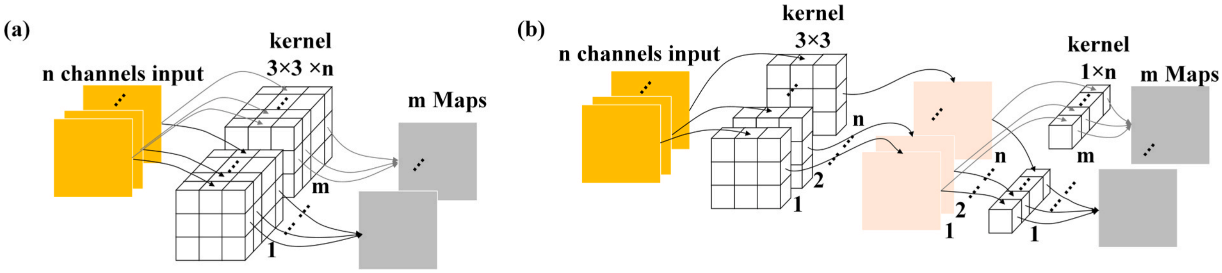


Fig. 3. Comparison between standard convolution and DSConv. (A) In standard convolution, a 3 × 3 kernel operates across all n input channels simultaneously to produce m output feature maps. (B) DSConv splits the operation into two steps: depth-wise convolution and pointwise convolution. First, a 3 × 3 depth-wise convolution is applied to each input channel individually, followed by a 1 × 1 pointwise convolution that combines these intermediate features across channels to generate the m output feature maps.

sized $A \times B$. The computational cost of the standard convolution is as follows:

$$O_{Conv} = n \cdot A \cdot B \cdot C \cdot C \cdot m \tag{4}$$

The computational cost of the DSConv is as follows:

$$O_{DSConv} = n \cdot A \cdot B \cdot C \cdot C + A \cdot B \cdot n \cdot m \tag{5}$$

The ratio of computational cost between the two convolutional structures is given by:

$$\alpha = \frac{O_{DSConv}}{O_{Conv}} = \frac{1}{m} + \frac{1}{C^2} \tag{6}$$

From Eq.6, it can be observed that when the number of output channels and the size of the convolution kernel are significantly greater than 1, replacing standard convolution with DSConv can greatly reduce the calculation amount of the same convolution layer, thereby saving the network's computation time.

2.5. Loss function

The SRGAN framework emphasizes the importance of high-frequency content in images and proposes a perceptual loss function based on perceptual similarity. Specifically, the content loss is defined on the high-frequency feature maps before the 4th convolutional layer, prior to the 5th pooling layer, in the Visual Geometry Group (VGG)-5.4 model (part of the 19-layer VGG network) [36]. The definition of typical SRGAN perceptual loss function $l^{SR}$is as follows:

$$l^{SR} = l^{SR}_{VGG} + 10^{-3} l^{SR}_{Gen} \tag{7}$$

where l^{SR}_{VGG} denotes the content loss and l^{SR}_{Gen} represents the adversarial loss.

The VGG loss is calculated as the Euclidean distance between the feature map representations of a reconstructed image $G_{\theta_D}(I^{LR})$and the reference image I^{HR}:

$$l^{SR}_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y}\right)^2 \tag{8}$$

where $\phi_{i,j}$ denotes the feature map extracted from the j-th convolution before the i-th max-pooling layer in the VGG network, and $W_{i,j}$ and $H_{i,j}$ are the dimensions of the respective feature maps.

While using VGG-5.4 (the 4th convolutional layer within the 5th block of the VGG network) in the perceptual loss function helps preserve high-frequency features, it often results in overly sharpened images and lower objective evaluation metrics. This study incorporates both VGG-2.2 (the 2nd convolutional layer within the 2nd block of the VGG network) and VGG-5.4 into the perceptual loss function, assigning different weight values to balance the representation of high-frequency and low-frequency features. The improved content loss function is shown as follows:

$$l^{SR}_{VGG} = \alpha l_{VGG/5.4} + \beta l_{VGG/2.2} \tag{9}$$

where α and β are hyperparameters. After validation, α is set to 1, and β is set to 0.01.

The generative loss l^{SR}_{Gen} is defined by the following formula:

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_D}(I^{LR})) \tag{10}$$

where $D_{\theta_D}(G_{\theta_D}(I^{LR}))$ represents the discriminator network that evaluates whether an image is real or generated $G_{\theta_D}(I^{LR})$ by the generator G_{θ_D}.

3. Experiments

3.1. PAM system set and data acquisition

The PAM system comprises a pulsed laser (VPFL-G-20, Spectra-Physics, USA), a high-frequency contact ultrasonic transducer (V214-BB-RM, Olympus, Japan), and a dual-channel data acquisition (DAQ) system (CSE1422, GAGE, USA). As shown in Fig. 4(a), the laser emits pulse at a wavelength of 532 nm. The laser beam is transmitted through the fiber collimator (F220FC-532, Thorlabs, USA) and passes through the fiber to the fiber coupler (PAF2P-11A, Thorlabs, USA). The outgoing laser is expanded by a pair of lenses and filtered through a band-pass filter (FB530–10, Thorlabs, USA) to remove the stray light. The laser beam is then reflected by a steering prism (CCM1-P01/M, Thorlabs, USA), focused by a plano-convex lens (LA1509-A, Thorlabs, USA) and an achromatic lens (GCL-010654, Daheng, China), and finally directed onto the target object using a combination of prisms. The acoustic signal induced by photon stimulation is reflected back and captured by a high-frequency ultrasonic transducer with a focal length of 6.7 mm and a diameter of 0.3 mm. Subsequently, the dual-channel DAQ system captures the instantaneous acoustic signals for processing.

The PAM system operates in a scanning mode, as illustrated in Fig. 4(b), using a two-dimensional motorized platform. When the pulse laser is focused on the target object, an A-line signal is generated at each scanning point. The system performs point-by-point scanning along the x-axis, organizing the A-line signals into a B-scan matrix. Sequential operation of two stepper motors enables an S-shaped scanning pattern. Using the optical path design and scanning path shown in Fig. 4, the system captures photoacoustic signals in a lattice configuration, reconstructing the photoacoustic structural image using the Maximum Amplitude Projection (MAP) algorithm [42]. The sampling frequency of the stepper motor during scanning determines the resolution of the reconstructed PAM images. The PAM system has a maximum scanning area of 20 mm× 20 mm, with the step sizes adjustable between 5–20 μm, and the depth sampling for each A-line consists of 256 units.

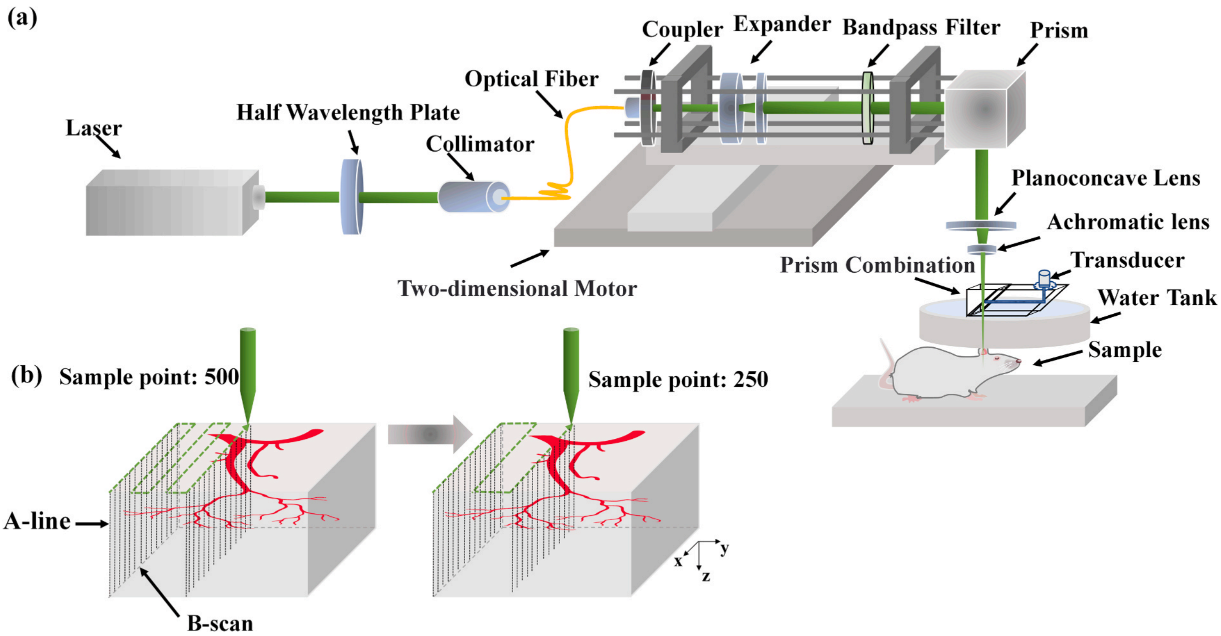


Fig. 4. Schematic representation of the Photoacoustic Microscopy (PAM) system. (a) Overview of the PAM system configuration, including the laser source, beam collimation, prism combination, and focusing optics for HR imaging. The system also incorporates a two-dimensional motor for precise spatial scanning of the imaging area. (b) Schematic diagram illustrating the system's scanning mode. An A-line scan represents the depth-resolved detection of photoacoustic signals at a fixed lateral position, capturing one-dimensional information along the z-axis. A B-scan is obtained by performing A-line scans sequentially along one lateral direction (x-axis), creating a two-dimensional cross-sectional image of the sample. From left to right down-sampling (×2) is applied during data acquisition or processing in microvascular imaging.

3.2. Dataset preparation

In this work, a dataset of PAM images of mouse ear vasculature was used to train and validate the FRDGAN model. First, ICR mice (20–25 g) were anesthetized with isoflurane before imaging. Depilatory cream was applied to expose the mouse ear vasculature. Ultrasound gel (Aquasonic 100) was applied to the mouse ear to enhance acoustic coupling, and the ear was fixed on a platform positioned beneath a water tank. Second, the PAM probe was placed above the target and immersed in the water tank to ensure proper acoustic coupling. The overall scanning area was 5 mm × 5 mm, with a pulse repetition rate of 5 kHz. The motor's longitudinal step size was 5 μm, and the x-axis scan consisted of 6000 steps. The signal depth was configured to 256 units, and the total scanning process costed approximately 20 minutes. A total of 188 PAM images of mouse ear vasculature were acquired. All animal experiments for PAI were conducted in compliance with the guidelines of the Institutional Animal Care and Use Committee at Nanjing University of Aeronautics and Astronautics.

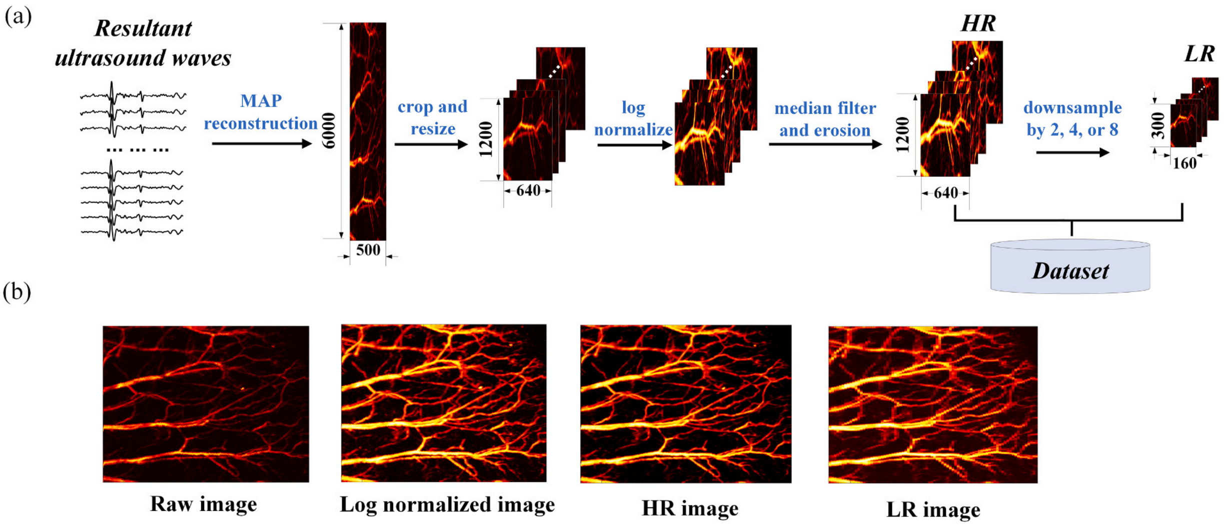


Fig. 5. Workflow for preprocessing Photoacoustic Microscopy (PAM) images. (a) Step-by-step process to generate HR and LR image datasets from raw photoacoustic data. The resultant ultrasound waves are reconstructed into Maximum Amplitude Projection (MAP) images, which are then cropped and resized to focus on the region of interest. The images are log-normalized to enhance contrast and detail, followed by median filtering and erosion to reduce noise and artifacts. The HR dataset is generated at full resolution (1200 ×640 pixels), while the LR dataset is created by systematically down-sampling the HR images by factors of 2, 4, or 8, resulting in resolutions such as 300 × 160 pixels. (b) Representative images obtained during preprocessing, including the raw image, log-normalized image, HR image, and LR image.

The dataset primarily consisted of mouse ear microvascular images obtained from the PAM system. The PAM image data preprocessing workflow is illustrated in Fig. 5. Due to the misalignment between the laser pulse trigger frequency and the motor scanning frequency, the raw PAM images had the X-axis accuracy of approximately 1 μm/pixel and the Y-axis accuracy of 10 μm/pixel, with an overall scanning size of 6000 × 500 × 256. The lateral resolution of this PA system, validated through experiments using ferrous microparticles, is approximately 24 μm [37].

To meet the input dimension requirements for the DL model, the images were cropped and resized, as demonstrated in Fig. 5(a). Additional preprocessing steps, including normalization and filtering, were applied to produce HR images suitable for training and evaluation. Fig. 5 (b) shows vascular images obtained after various preprocessing steps. Due to random motor movement noise and environmental interference during scanning, the collected PAM images contained numerous discrete bright spots considered as noise artifacts. To address this, logarithmic normalization was applied to compress the amplitude of the bright artifacts, aligning the data towards a normal distribution.

The normalization formula is given as follows:

$$\dot{I}_{i,j} = \log_{10}(I_{i,j})/\log_{10}(\max(I)), \quad \forall x, j \in N \tag{11}$$

Where $I_{i,j}$ represents the pixel value of the image I and $\max(I)$ represents the maximum pixel value of the image I.

After normalization, median filtering was applied using a 3 × 1 convolution kernel, followed by a 3 × 3 structural element erosion to eliminate remaining noise. The processed HR images, originally sized at 1200 × 640, were downsampled to generate LR images using the nearest neighbor (NN) interpolation method with down-sampling ratios of 2, 4, and 8.

Let I_{HR} denotes a HR image andI_{LR}denotes the downsampled LR image. The NN downsampling formula is defined as follows:

$$I_{LR}(i,j) = I_{LR}(i\cdot s_i, j\cdot s_j) \tag{12}$$

Where (i,j) represents coordinates in the LR image I_{LR}, s_i and s_j are down-sampling ratios. Then, the combined HR and LR images constituted the complete training dataset, as shown in Fig. 5(b).

For data augmentation, sub-images of 1200 × 640 pixels were randomly cropped from the original images. Additional augmentations included random rotation (up to 20°), random horizontal shift (up to 10 % of the image width), and random vertical shift (up to 10 % of the image height) [38]. These augmentations were applied during each batch loading process, ensuring robust training and evaluation of the DL model, in summary, the sample dataset consisted of 380 images, with 40 images allocated to the test set.

3.3. Network implementation details

To optimize training efficiency, the proposed FRDGAN network consists of four FRDB blocks, each containing four Conv+LeakyReLU layers with 64 channels. Each 3 × 3 DSConv layer in the FRDB block includes a depth-wise separable convolution (stride = 1, padding = 1, kernel size = 1 ×3) followed by a pointwise convolution (stride = 1, padding = 0, kernel size = 3 ×1). The training process involved 5400 iterations over 300 epochs. The Adam optimizer was used with β1 = 0.9 [39]. The discriminator's loss was based on the contrast between real images and the fake images generated by the generator, while the generator's loss was a weighted sum of adversarial loss, perception loss and MSE loss. The generator's loss served as the overall loss function for the network. During each epoch, the generator and discriminator were trained alternately. The training was conducted on an Ubuntu 20.04.3 LTS system using Python 3.9.18 and Pytorch 2.1.0. The FRDGAN network was deployed on a Linux machine equipped with eight 12 G NVIDIA GeForce RTX 2080 Ti GPUs. All of the networks were optimized using the Adam algorithm with a mini-batch size of 16 and initial learning rate of 0.001.

3.4. Evaluation metrics

To quantitatively assess the quality of reconstructed LR images, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were used [40]. Specifically, PSNR is defined by the following formula:

$$MSE = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\widehat{I_C}(i,j) - I_C(i,j)\right)^2 \tag{13}$$

Where $I_C(i,j)$和$\widehat{I_C}(i,j)$ refer to the pixel value of the ground-truth (HR) and reconstructed image, respectively.$H$and W are height and width of the ground-truth (HR) and reconstructed image.

$$PSNR = 10\log\left(\frac{MAX^2}{MSE}\right) \tag{14}$$

where MAX is the maximal pixel value of the ground-truth (HR) and reconstructed image.

$SSIM$ is defined by the following formula:

$$SSIM = \frac{(2\mu_{I_C}\mu_{\hat{I_C}} + \varepsilon_1)(2\sigma_{I_C\hat{I_C}} + \varepsilon_2)}{(\mu_{I_C}^2 + \mu_{\hat{I_C}}^2 + \varepsilon_1)(\sigma_{I_C}^2 + \sigma_{\hat{I_C}}^2 + \varepsilon_2)} \tag{15}$$

where μ_{I_C}, $\mu_{\hat{I_C}}$, σ_{I_C} and $\sigma_{\hat{I_C}}$ are the means and variances of the ground-truth (HR) and reconstructed image, and $\sigma_{I_C\hat{I_C}}$ is the covariance of reconstructed image and ground-truth (HR), respectively. The small constants ε_1 and ε_2 are applied to avoid a null denominator.

For real noisy datasets without ground truth, denoising performance was evaluated using signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) [41]. To calculate these metrics, representative signal regions and background regions from each sample were selected. SNR and CNR were calculated as follows, respectively:

$$SNR = 20\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu_i}{\sigma_b}\right) \tag{16}$$

$$CNR = 20\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{|\mu_i - \mu_b|}{\sqrt{\sigma_i^2 + \sigma_b^2}}\right) \tag{17}$$

where μ_i and σ_i denote the mean and standard deviation of the i-th signal region, respectively, μ_b and σ_b are the mean and standard deviation of the background region, and n is the number of signal regions, which was set as 3 in our research.

In this study, the mean option score (MOS) was used as a subjective evaluation metric, as described in [25]. We asked 20 raters, blinded to the image versions, to rate the image quality on a scale of 1 to 5. The final MOS was calculated as the average score across all raters.

3.5. Comparison method

For detailed performance assessment, we compared the image quality metrics and computational efficiency of bicubic interpolation [42], SRGAN, RDN [43], and FRDGAN across down-sampling ratios of × 2, × 4, and × 8. As FRDGAN represents an enhancement of the SRGAN framework, with its key innovation being the replacement of the standard residual module with FRDBs, we specifically focused on comparing the denoising and generalization performance of SRGAN and FRDGAN, detailed results are illustrated in Sections 4.2 and 4.3.

4. Results

4.1. Comparison of reconstructed PAM images

Fig. 6(a) and (f) show the LR and HR images of mouse ear vasculature

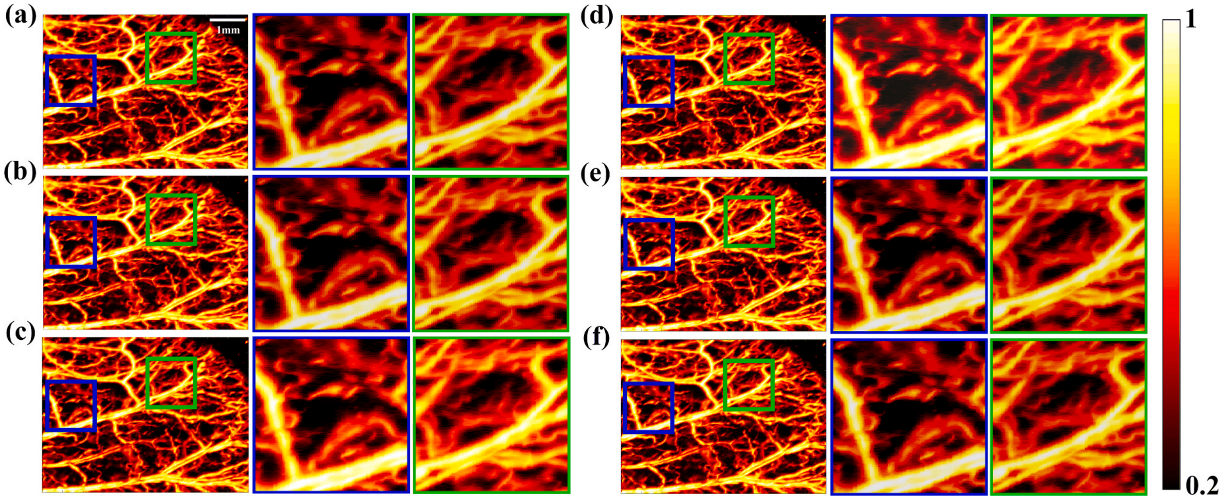


Fig. 6. Representative reconstructed PAM images of mouse ear vasculature (×2 down-sampling) obtained using different methods. (a) LR image, (b) bicubic interpolation, (c) RDN, (d) SRGAN, (e) FRDGAN, and (f) HR image. The images in the first column show the full field of view of the vascular network. The second and third columns present magnified close-up views of the regions indicated by the blue and green boxes, respectively.

obtained by our PAM system (Fig. 4), respectively. The HR image represents the vascular structure obtained by point-by-point scanning followed by the MAP process (Fig. 5). The LR image was generated by down-sampling the original MAP image with a ratio of 2. Based on the LR image, vascular structure images were reconstructed using various methods, including bicubic interpolation, RDN, SRGAN, and the proposed FRDGAN, as shown in Fig. 6(b)-(e), respectively. Close-up images of the selected region are presented sequentially. At low-to-moderate down-sampling ratio (×2), differences between the reconstruction methods are less apparent. However, upon closer inspection, the vascular structures reconstructed using bicubic interpolation (Fig. 6(b)) display slight jagged edges. In contrast, the vascular structures reconstructed by RDN, SRGAN, and FRDGAN (Fig. 6(c)-(e)) remain relatively smooth and continuous.

The differences in image quality among these methods become more significant at higher down-sampling ratios. For higher down-sampling ratios (×4 and ×8), the jagged edges in LR images become increasingly pronounced, as seen in Figs. 7(a) and 8(a). The close-up views of reconstructed images further emphasize these differences. Vascular structures reconstructed using bicubic interpolation (Figs. 7(b) and 8(b)) exhibit more noticeable jagged edges and poorer signal fidelity compared to HR images (Figs. 7(f) and 8(f)). In contrast, RDN, SRGAN, and FRDGAN achieve superior reconstruction of vascular edge details.

To quantitatively assess reconstruction performance, multiple objective and subjective metrics were employed, as summarized in Table 1. PSNR and SSIM have been used as widely accepted evaluation indexes for biomedical image restoration. PSNR emphasizes pixel-wise accuracy, while SSIM measures structural similarity between two images. Additionally, MOS was introduced as a subjective evaluation metric to account for visual effects and mitigate potential distortions caused by excessive optimization of objective metrics.

Table 1 shows a steady decline in SSIM, PSNR, and MOS values as the down-sampling ratio increases. At higher down-sampling ratios (×4 and ×8), the proposed FRDGAN method demonstrates significant improvements over other methods. These advantages become even more pronounced at × 8, where FRDGAN achieves a PSNR improvement of 1.75 dB and an SSIM improvement of 0.0254 over SRGAN, despite using less than 5 % of the pixels in the original HR image. This highlights FRDGAN's effectiveness in reconstructing high-quality images from severely down-sampled data.

While the quality of image reconstruction is critical, the computational efficiency and resource requirements of reconstruction algorithms

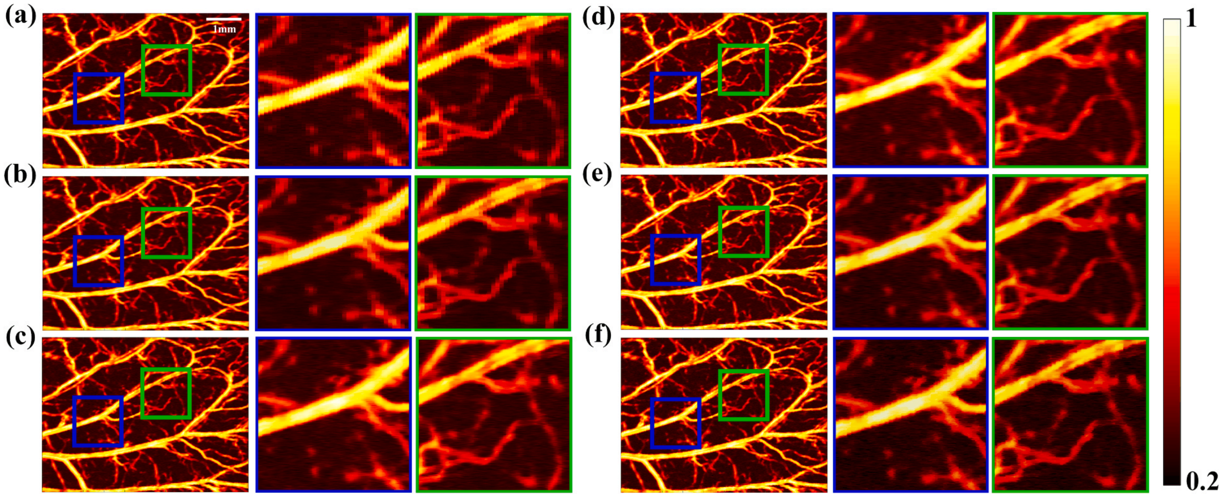


Fig. 7. Representative reconstructed PAM images of a mouse ear vasculature (×4 down-sampling) obtained using different methods. (a) LR image, (b) bicubic interpolation, (c) RDN, (d) SRGAN, (e) FRDGAN, and (f) HR image. The images in the first column show the full field of view of the vascular network. The second and third columns present magnified close-up views of the regions indicated by the blue and green boxes, respectively.

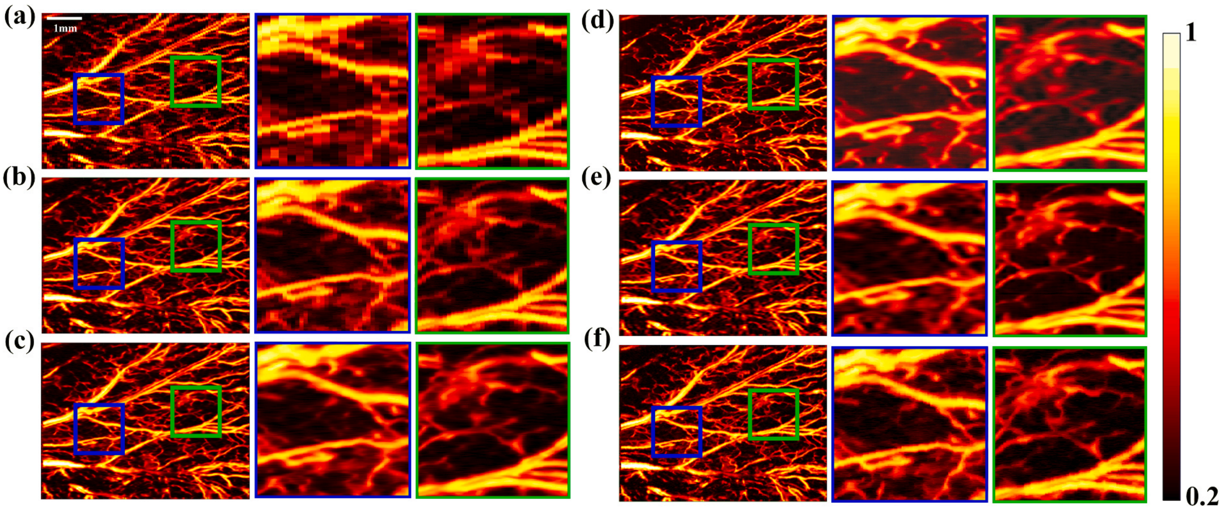


Fig. 8. Representative reconstructed PAM images of a mouse ear vasculature (×8 down-sampling) obtained using different methods. (a) LR image, (b) bicubic interpolation, (c) RDN, (d) SRGAN, (e) FRDGAN, and (f) HR image. The images in the first column show the full field of view of the vascular network. The second and third columns present magnified close-up views of the regions indicated by the blue and green boxes, respectively.

Table 1
Performance comparison of reconstruction methods across three down-sampling factors (×2, ×4, ×8). Quantitative metrics include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean Opinion Score (MOS), Inference Time (ms), Floating Point Operations (FLOPs), Parameters, and GPU Memory Usage.

Scale	Methods	PSNR (dB)	SSIM	MOS	Inference [ms]	FLOPs [G]	Parameters [M]	Memory [GB]
× 2	Bicubic	38.26	0.9658	3.38	-	-	-	-
	RDN	42.35	0.9822	4.35	648.30	4245.71	22.12	13.315
	SRGAN	41.55	0.9791	4.33	27.37	101.17	5.63	1.738
	FRDGAN	**42.08**	**0.9826**	**4.31**	**27.24**	**103.12**	**5.65**	**2.538**
× 4	Bicubic	31.10	0.8985	2.67	-	-	-	-
	RDN	37.81	0.9476	3.78	169.89	1090.07	22.27	3.706
	SRGAN	37.06	0.9378	3.52	15.55	58.89	5.91	1.086
	FRDGAN	**38.00**	**0.9523**	**3.81**	**15.40**	**58.66**	**5.79**	**1.012**
× 8	Bicubic	23.79	0.7911	2.07	-	-	-	-
	RDN	32.07	0.8742	3.40	48.03	301.16	22.42	1.303
	SRGAN	30.67	0.8516	3.27	13.82	45.52	6.03	0.814
	FRDGAN	**32.42**	**0.8770**	**3.44**	**9.78**	**43.32**	**5.94**	**0.630**

are equally important. Four metrics were used to evaluate computational efficiency and resource utilization, including Inference, Floating Point Operations (FLOPs), Parameters, and GPU Memory Usage. Reduced inference time and FLOPs indicate faster processing and higher computational efficiency, enabling rapid reconstruction of large datasets. Parameters and memory requirements reflect the computational footprint of the model during testing, making them important considerations for sparse scanned PAM imaging. Although RDN and FRDGAN yield comparable results in terms of PSNR, SSIM, and MOS image evaluation metrics, FRDGAN offers superior computational efficiency. Specifically, when the down-sampling factor is × 8, FRDGAN reduces inference time by 38.25 ms, decreases FLOPs by 257.84 G, and lowers both the model parameters and GPU Memory Usage by 16.48 M and 0.673 G, respectively. The tables show that FRDGAN demonstrates significant improvements across all metrics as the down-sampling ratio increases, particularly at down-sampling ratio of × 8. Taken together, FRDGAN achieves better PSNR, SSIM, and MOS values while significantly reducing computational and memory requirements. These enhancements make FRDGAN a more efficient and effective choice for rapid PAM imaging, especially under sparse scanning conditions.

4.2. Analysis of reconstructed signal strength and evaluation of SNR and CNR

Fig. 9 compares the reconstructed signal strength of different algorithms on small blood vessels. The first column shows the reconstructed PAM images at down-sampling ratio of × 2, × 4, and × 8, while columns 2 to 4 present the MAP images of the corresponding white-marked vessels. Upon comparison, the effective vessel signal widths reconstructed by LR and bicubic interpolation are wider than those of HR images, with signal strengths that are lower than HR.

In contrast, the vessel MAP signal widths reconstructed by SRGAN and FRDGAN closely match those of HR images, while SRGAN produces a higher background signal. As the down-sampling ratio increases, the deviation of the MAP signal peak from the HR becomes more pronounced for both bicubic interpolation and SRGAN. Conversely, the MAP signal peak reconstructed by FRDGAN exhibit much smaller deviations, maintaining a closer alignment with the original HR data.

To further evaluate the quality of the reconstructed vasculature signals, SNR and CNR were calculated for the white-marked vascular sites and background noise. SNR is calculated as the ratio of the mean signal intensity of the vascular region to the standard deviation of the background noise. CNR analyzes the difference between the mean signal intensity of the vascular region and that of the background noise. Both higher SNR and higher CNR correlate with better image quality. High SNR indicates smoother, more uniform backgrounds, and high CNR reflects better contrast and separability between the signal and background noise. As shown in Fig. 10, it is evident that FRDGAN achieves significant improvements over existing methods in terms of both SNR and CNR, particularly at higher down-sampling ratios. A statistical

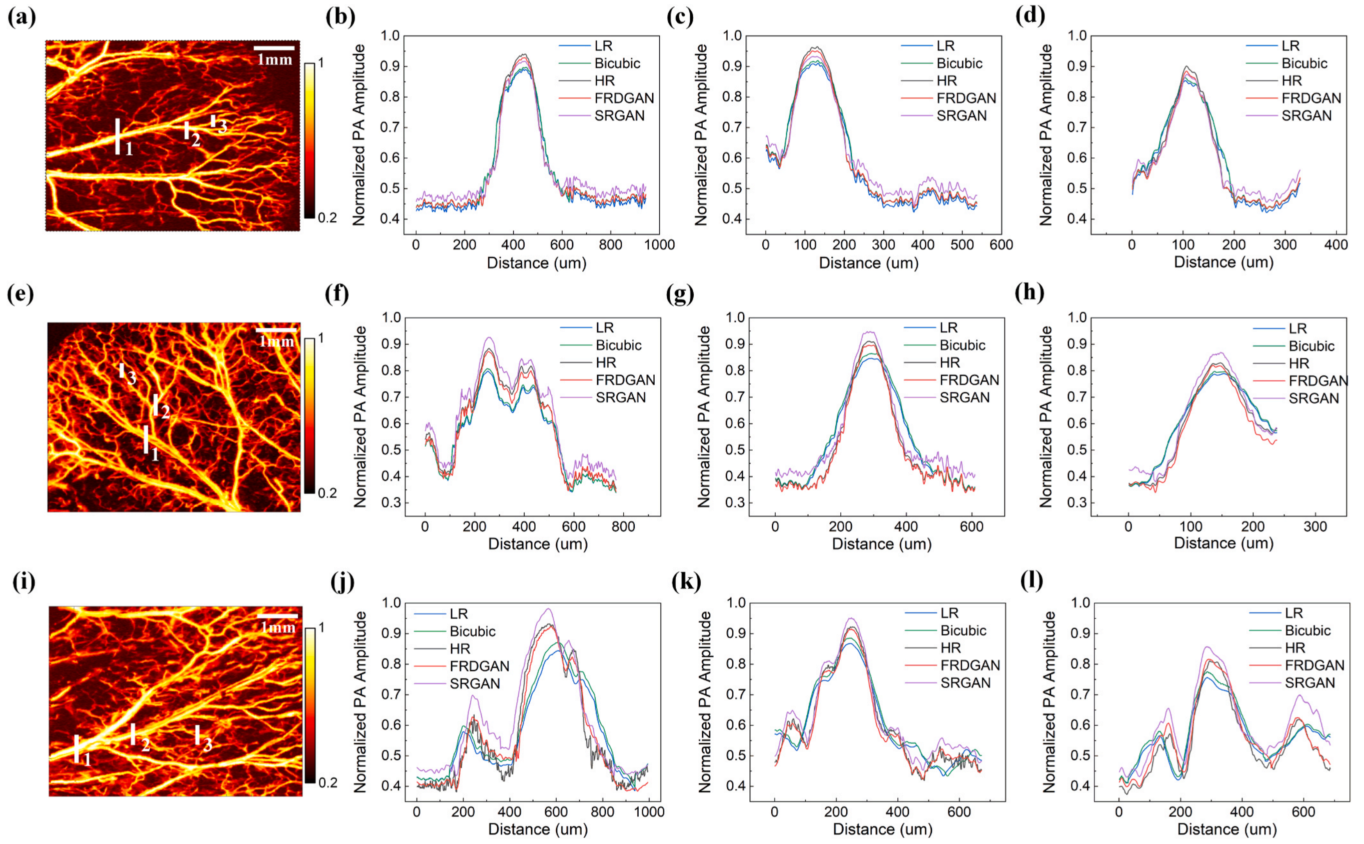


Fig. 9. Comparison of reconstructed PAM images obtained by FRDGAN with varying down-sampling ratios and corresponding pixel intensity profiles. (a, e, i) Reconstructed images for down-sampling ratios of 2, 4, and 8, respectively. White lines (labeled 1, 2, and 3) in each image indicate the locations where pixel intensity profiles are measured. (b–d) Normalized photoacoustic amplitude profiles along the white lines 1, 2, and 3 for the down-sampling ratio of 2, showing the performance of different reconstruction methods, including LR, bicubic interpolation, HR, FRDGAN, and SRGAN. (f–h) Corresponding intensity profiles for the down-sampling ratio of 4. (j–l) Corresponding intensity profiles for the down-sampling ratio of 8.

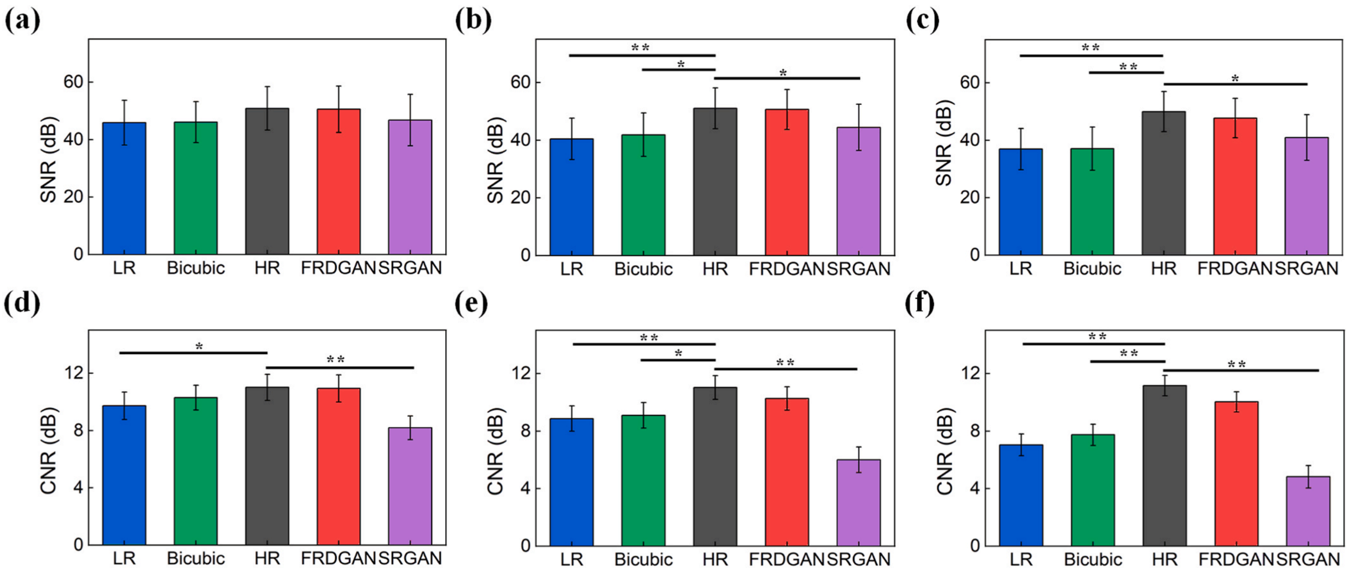


Fig. 10. Statistical analysis of the performance of FRDGAN compared with other methods for reconstructed PAM images. (a–c) Signal-to-noise ratio (SNR) of images reconstructed with down-sampling ratios of × 2, × 4, and × 8, respectively, for LR, bicubic interpolation, HR, FRDGAN, and SRGAN. (d–f) Contrast-to-noise ratio (CNR) of images reconstructed with down-sampling ratios of × 2, × 4, and × 8, respectively. Statistical comparisons are performed between FRDGAN and other methods, with $* *$ indicated $P < 0.01$ and $*$ indicated $P < 0.05$.

analysis shows that, the SNR values of SRGAN-reconstructed images show some differences compared to HR images ($P < 0.05$) while the CNR values of SRGAN-reconstructed images exhibit significant differences compared to HR images ($P < 0.01$). By contrast, despite increasing down-sampling ratios, the SNR and CNR values of images reconstructed using FRDGAN remain comparable to those of HR images, with no significant difference observed.

4.3. Comparison of ablation study

To evaluate the effectiveness of the proposed FRDGAN network, we conducted ablation studies by removing key modules, such as the DSconv and FRDB modules. FRD-Conv refers to the FRDGAN variant where the DSConv module is replaced with a standard convolution module, while RDN-GAN represents the FRDGAN variant where the FRDB module is replaced with a conventional RDB module. Supplementary Figs. S1-S3 show selected images of the same mouse ear vasculature with down-sampling ratios (×2, ×4, ×8), as shown in Figs. 6–8. The significance of each module in FRDGAN is further evaluated using quantitative metrics, including PSNR, SSIM, MOS, inference time, FLOPs, parameter count, and GPU memory usage, as summarized in Table S1.

The removal of the DSconv module led to computational efficiency degradation across all down-sampling ratios. While the reconstructed images by FRD-Conv exhibit slightly higher SSIM, PSNR, and MOS values compared to FRDGAN at a down-sampling ratio of × 2, the evaluation of computational efficiency reveals that metrics, including inference time, FLOPs, parameter count, and GPU memory usage, are markedly higher for FRD-Conv than for FRDGAN, as shown in Table S1. For down-sampling ratios of × 4 and above, FRDGAN outperforms FRD-Conv in both image quality (PSNR, SSIM, and MOS) and computational efficiency (inference time, FLOPs, parameter count, and GPU memory usage). This highlights the critical role of the DSConv module in balancing feature extraction efficiency and computational overhead. This emphasizes the critical role of the DSconv module in balancing feature extraction efficiency and computational overhead.

Table S1 further demonstrates that excluding the FRDB module results in more pronounced performance degradation across all down-sampling ratios. Specifically, for RDN-GAN at a × 2 down-sampling ratio, PSNR decreases to 42.07, while SSIM drops to 0.9827. Notably, the inference time increases sharply to 70.44 ms compared to 27.24 ms for FRDGAN, with FLOPs and GPU memory usage rising to 207.84 G and 5.145 G, respectively. These results indicate the indispensable role of the FRDB module in reducing network redundancy. Additionally, through a comparative analysis of the close-up images of Fig. S2 - S3, it can be observed that RDN-GAN introduces noise artifacts during reconstruction, whereas FRDGAN demonstrates effective noise suppression in terms of SNR and CNR. This suggests that the redundancy caused by dense residual connections can lead to instability in the network itself.

4.4. Generalization and validation of FRDGAN

To further validate the performance and generalization capability of FRDGAN method, we simulated its actual application under rapid scanning conditions using the PAM system. We acquired a dataset of PAM images, including mouse cerebral vasculature with an open cranial window, human hair phantoms, microfluidic chips, and porcine liver tissue. These included both low-speed and high-speed scans of the same region, resulting in raw 3D matrices with B-scaning points of 6000 × 500 and 1500 × 250, respectively. Real HR and LR with down-sampling ratios of [2,4] were reconstructed using MAP. The trained FRDGAN, SRGAN, and bicubic interpolation methods were tested on these datasets, and the corresponding parameter metrics are presented in Table 2 and Table. S2.

Table 2
Performance comparison of reconstruction methods for cerebral vasculature and hair phantom. Quantitative metrics include PSNR and SSIM.

Data	Methods	PSNR(dB)	SSIM
Cerebral vasculature	Bicubic	22.06	0.7434
	SRGAN	26.86	0.8320
	FRDGAN	**29.10**	**0.8575**
Hair phantom	Bicubic	39.54	0.9546
	SRGAN	40.54	0.9711
	FRDGAN	**41.30**	**0.9736**

Based on the close-up images marked with blue boxes in Fig. 11 and Fig. S4, tubular structures reconstructed using bicubic interpolation exhibit discontinuous edges and a jagged appearance, indicating poor reconstruction quality, especially in complex regions. While SRGAN performs adequately for images with simple structures, such as straight hair and microfluidic chips, its performance diminishes when reconstructing more complex structures like cerebral vasculature and porcine liver tissue. It often overlooks finer vascular details and fails to restore intricate features. Yet, our proposed FRDGAN method consistently demonstrates superior generalization ability and stability across both simple and complex structures.

As shown in Table 2 and Table S2, due to the structural simplicity of the hair phantom and microfluidic chip, reconstruction results using FRDGAN, SRGAN, and bicubic interpolation methods exhibit minimal differences in PSNR and SSIM values. All methods perform relatively well on this dataset. On the other hand, both mouse cerebral vasculature and porcine liver tissue contain fine capillary networks and exhibit complex overall tissue background structures. It is these common characteristics that highlight the superiority of the FRDGAN method. Images reconstructed by FRDGAN exhibit significantly higher PSNR and SSIM values compared to other methods, demonstrating its ability to preserve finer details and achieve higher image quality. It is worth mentioning that the these datasets were acquired through low-speed scanning to generate LR images, which differ from the LR images produced by HR down-sampling. Considering that the process of low-speed scanning would possibly introduce additional noise and artifacts, to our interests, under real-world scanning conditions, FRDGAN outperforms other methods, showcasing its robustness in handling adverse noise and artifacts while maintaining high-quality reconstructions.

5. Discussion

Our work builds on recent advancements in DL and applies these innovations to the rapidly growing field of PAM. By training on an expanding dataset of microvascular images, our FRDGAN model demonstrated the ability to reconstruct images at down-sampling ratios of up to × 8. Using only 1.5 % to 25 % of the original pixels, our FRDGAN model can potentially accelerate the imaging speeds of PAM systems by 4 to 64 times, assuming the imaging speed is mainly limited by the laser's pulse repetition rate. This approach avoids the need for costly hardware upgrades and creates an avenue for further performance improvements through continuous dataset expansion and model refinement. The evaluation across three representative down-sampling ratios demonstrates the superior performance of FRDGAN in improving the resolution of down-sampled PAM images, as evidenced by its ability to effectively reconstruct vasculature with minimal loss of detail.

Adversarial training, particularly with GANs, has been a key approach for achieving photo-realistic image reconstruction [44]. One of the milestones in this area is SRGAN, which uses residual blocks and perceptual loss within a GAN framework to significantly enhance visual quality compared to traditional PSNR-oriented methods. However, SRGAN heavily emphasizes perceptual loss based on VGG-5.4, which excessively prioritizes high-frequency details while introducing distortions in low-frequency regions, resulting in skewed reconstructions of noisy areas. In contrast, our FRDGAN approach introduces a joint

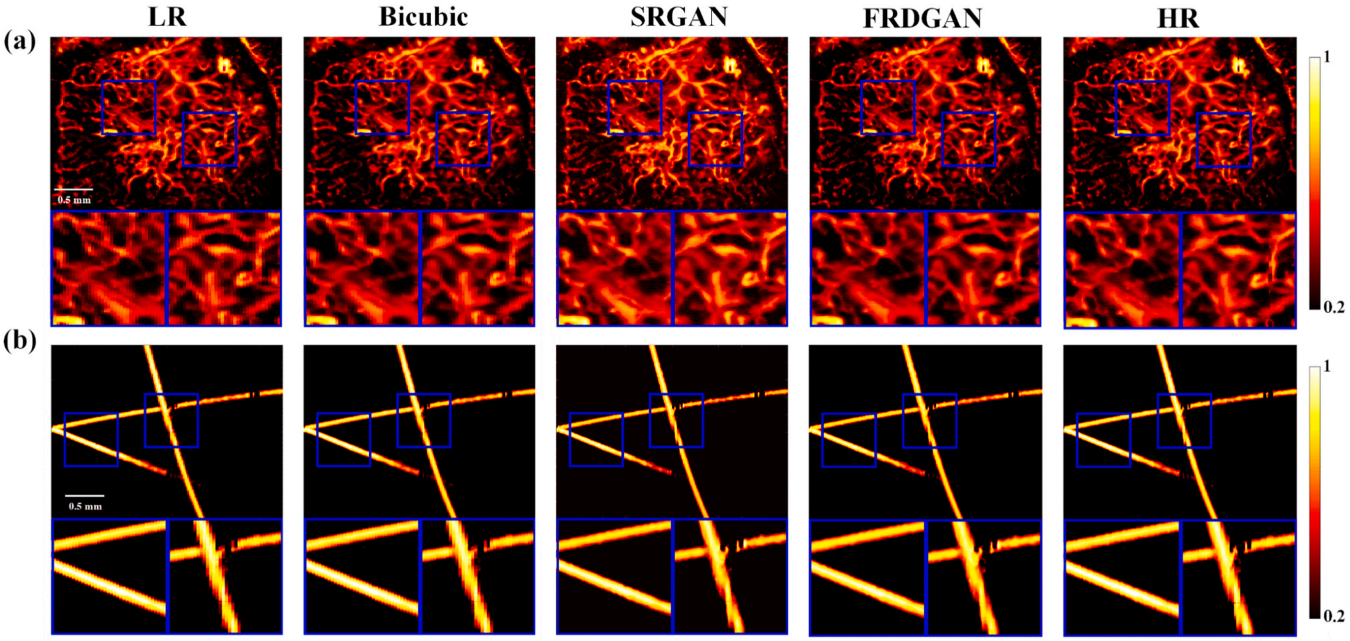


Fig. 11. Validation of PAM image reconstruction using LR images obtained from actual scanning ratios of [2,4]. (a) Reconstructed images of mouse cerebral vasculature using LR, Bicubic interpolation, SRGAN, FRDGAN, and HR. (b) Reconstructed images of a hair phantom under the same conditions, with similar comparisons across reconstruction methods. The bottom row shows magnified views of selected regions to assess the preservation of structural details.

perceptual loss utilizing both VGG-2.2 and VGG-5.4, enabling the reconstructed images to retain smoother details while suppressing noise, As shown in Fig. 10, FRDGAN's reconstructed vascular signal peaks align closely with HR ground truth, with SNR and CNR surpassing SRGAN by 14.1 % and 51.8 %, respectively (×8 down-sampling). This improvement results in superior performance in terms of CNR and SNR, providing a balanced reconstruction that maintains structural and textural fidelity. Furthermore, by comparing the model generalization experiments in Table 2 and Table S2, compared to simpler tubular structures (such as human hair phantoms or microfluidic chips), FRDGAN achieves higher-quality image reconstruction and exhibits stronger noise suppression capabilities in complex vascular structures or noisy conditions.

Compared to conventional RDN frameworks, the FRDGAN generator demonstrates enhanced reconstruction capabilities while utilizing fewer computational resources. While the standard RDN network uses L1 loss to achieve higher PSNR values, it often fails to preserve high-frequency details, resulting in visually unsatisfying reconstructions [45]. As shown in Table 1, both RDN and FRDGAN achieve similar SSIM values, but FRDGAN outperforms RDN in terms of image fidelity. The densely connected structure of RDN can lead to shallow feature over-fusion, inefficient computation, redundant parameters, and an increased risk of gradient explosion. In contrast, FRDGAN's sparse connection design avoids these issues, achieving reconstructions closer to HR images while significantly reducing training time during alternating generator-discriminator training cycles. Besides, the ablation studies confirm that both the DSConv module and FRDB module are indispensable for FRDGAN's performance. Their removal leads to significant declines in image quality, increased computational costs, and reduced efficiency. FRDGAN, with both modules intact, consistently outperforms its ablated variants across all down-sampling factors, achieving state-of-the-art results in both reconstruction fidelity and computational efficiency.

A major challenge in deploying deep learning (DL) models for photoacoustic microscopy (PAM) lies in their reliance on extensive fully sampled datasets, which poses a critical bottleneck in biomedical imaging due to the high costs and ethical constraints of acquiring large-scale, annotated PAM data. Traditional GANs, by optimizing the generator-discriminator framework, can effectively enhance image texture realism and preserve high-frequency details, even with limited training data. FRDGAN effectively inherits the advantages of GAN-based networks while mitigating the overfitting issues observed during SRGAN training by incorporating FRDBs and DSConv modules, exhibiting strong generalization performance, as shown by its ability to reconstruct LR images without manual parameter tuning after effective training on a well-prepared initial dataset. Unlike other methods, our FRDGAN-based approach does not require pretraining and can reconstruct high-quality HR images of mouse brain vasculature and hair phantoms while retaining fine details. This proves that FRDGAN is well-suited for transfer learning on small sample datasets, achieving detailed and high-quality reconstructions without requiring a large number of computational resources.

While focuses on the OR-PAM reconstruction algorithm, our study lays the groundwork for improving the resolution of acoustic-resolution photoacoustic microscopy (AR-PAM) by leveraging the strengths of OR-PAM. Previous studies have demonstrated techniques that combine AR-PAM and OR-PAM data, or utilize advanced methods like Wasserstein GANs and DL models to enhance image quality at greater depths [39, 46]. In the future, we could integrate AR-PAM and OR-PAM data in our model to improve image reconstruction further. Another important area for future application is functional imaging. As high-definition PAM typically requires long scanning times, it can be problematic for *in vivo* imaging that requires fast temporal resolution. Approaches like conditional GANs (cGANs) have shown promise in speeding up multi-parameter PAM imaging by separating signals from dual-wavelength lasers [47]. By using FRDGAN, we may further accelerate the functional imaging process, making it feasible to capture high-quality functional images in real-time with minimal computational overhead. Additionally, improvements can be made to PAM's scanning mechanisms, excitation methods, and post-processing techniques [48]. Future work should focus on improving the performance and interpretability of DL models by developing more efficient algorithms, refining data acquisition and labeling processes, and enhancing the transparency of model predictions to ensure their reliability and applicability in clinical settings.

6. Conclusion

To address the challenges of HR PAM imaging, we propose an optimized FRDGAN that enhances resolution while reducing scanning times. By incorporating a sparse-dense connection mechanism and replacing standard convolutions with DSConv, FRDGAN improves image detail and reconstruction efficiency. These advancements enable superior performance in reconstructing microvascular details across multiple down-sampling scales, with a lighter network load. Validated on external datasets, the model demonstrates robust generalizability, offering a solution to balance imaging speed and resolution in PAM.

CRediT authorship contribution statement

Zhang Yameng: Writing - review & editing, Writing - original draft, Visualization, Validation. Methodology, Conceptualization. **Tian Hua:** Visualization, Software, Formal analysis. **Wan Min:** Writing - review & editing, Investigation. **Tang Shihao:** Writing - review & editing, Investigation. **Ding Ziyun:** Writing - review & editing, Conceptualization. **Huang Wei:** Writing - review & editing, Funding acquisition. **Yang Yamin:** Writing - review & editing. **Li Weitao:** Writing - review & editing, Validation, Supervision, Funding acquisition.

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.

Acknowledgement

This research was supported by National Natural Science Foundation of China (81727804, 61875085), Nanjing Health Science and Technology Development Foundation (ZKX24043), Fundamental Research Funds for the Central Universities (NJ2024029), and Research Project of Jiangsu Province (22KJB460024).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pacs.2025.100720.

Data availability

Data will be made available on request.

References

[1] L.H.V. Wang, J.J. Yao, A practical guide to photoacoustic tomography in the life sciences, Nat. Methods 13 (8) (2016) 627–638.
[2] C.C. Yang, H.R. Lan, F. Gao, et al., Review of deep learning for photoacoustic imaging, Photoacoustics 21 (2021) 13.
[3] W.Y. Li, J. Lv, H.H. Li, et al., Quantification of vascular remodeling and sinusoidal capillarization to assess liver fibrosis with photoacoustic imaging, Radiology 314 (1) (2025).
[4] W. Choi, B. Park, S. Choi, et al., Recent advances in contrast-enhanced photoacoustic imaging: overcoming the physical and practical challenges, Chem. Rev. 123 (11) (2023) 7379–7419.
[5] J. Lv, H.R. Lan, A.J. Qin, T, et al., Dynamic synthetic-scanning photoacoustic tracking monitors hepatic and renal clearance pathway of exogeneous probes in vivo, Light-Sci. Appl. 13 (1) (2024).
[6] Y. Qiu, H.H. Li, K. Yu, et al., Collagen fibers quantification for liver fibrosis assessment using linear dichroism photoacoustic microscopy, Photoacoustics 42 (2025).
[7] J. Yao, L.V. Wang, Photoacoustic microscopy, Laser Photonics Rev. 7 (5) (2013) 758–778.
[8] E. Hysi, M.J. Moore, E.M. Strohm, et al., A tutorial in photoacoustic microscopy and tomography signal processing methods, J. Appl. Phys. 129 (14) (2021) 141102.
[9] C. Liu, L.D. Wang, Functional photoacoustic microscopy of hemodynamics: a review, Biomed. Eng. Lett. 12 (2) (2022) 97–124.
[10] R. Cao, S.D. Nelson, S. Davis, et al., Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy, Nat. Biomed. Eng. 7 (2) (2023) 124–134.
[11] S. Jeon, J. Kim, D. Lee, et al., Review on practical photoacoustic microscopy, Photoacoustics 15 (2019) 100141.
[12] J.Y. Kim, C. Lee, K. Park, et al., Fast optical-resolution photoacoustic microscopy using a 2-axis water-proofing MEMS scanner, SCI REP-UK 5 (1) (2015) 7932.
[13] Y. Yuan, S. Yang, D. Xing, Optical-resolution photoacoustic microscopy based on two-dimensional scanning galvanometer, Appl. Phys. Lett. 100 (2) (2012) 023702.
[14] J. Yao, Ultra-high-speed wide-field photoacoustic microscopy. High-Speed Biomedical Imaging and Spectroscopy VI, SPIE, 2021, p. 1165407.
[15] F. Zhang, J.D. Zhang, Y.T, et al., Photoacoustic digital brain and deep-learning-assisted image reconstruction, Photoacoustics 31 (2023) 100517.
[16] J. Yang, S. Choi, J. Kim, et al., Recent advances in deep-learning-enhanced photoacoustic imaging, Adv. Photonics Nexus 2 (5) (2023) 054001.
[17] P. Rajendran, A. Sharma, M. Pramanik, Photoacoustic imaging aided with deep learning: a review, Biomed. Eng. Lett. 12 (2) (2022) 155–173.
[18] H.X. Zhao, Z.W. Ke, F. Yang, et al., Deep Learning enables superior photoacoustic imaging at Ultralow Laser Dosages, Adv. Sci. 8 (3) (2021) 2003097.
[19] A. Sharma, M. Pramanik, Convolutional neural network for resolution enhancement and noise reduction in acoustic resolution photoacoustic microscopy, Biomed. Opt. Express 11 (12) (2020) 6826–6839.
[20] A. DiSpirito, D.W. Li, T. Vu, et al., Reconstructing udersampled potoacoustic mcroscopy Images Using Deep Learning, IEEE Trans. Med. Imaging 40 (2) (2021) 562–570.
[21] M.J. Guo, H.R. Lan, C.C. Yang, et al., AS-Net: fast photoacoustic reconstruction wth multi-feature fusion from sparse data, IEEE Trans. Comput. Imaging 8 (2022) 215–223.
[22] J.S. Zhou, D. He, X.Y. Shang, et al., Photoacoustic microscopy with sparse data by convolutional neural networks, Photoacoustics 22 (2021) 100242.
[23] Y.X. Liu, J.S. Zhou, Y.T. Luo, et al., UPAMNet: a unified network with deep knowledge priors for photoacoustic microscopy, Photoacoustics 38 (2024) 100608.
[24] S. Chen, X. Tian, Y. Wang, et al., DAEGAN: generative adversarial network based on dual-domain attention-enhanced encoder-decoder for low-dose PET imaging, Biomed. Singal. Proces. 86 (2023) 105197.
[25] W. Du, S. Tian, Transformer and GAN-based super-resolution reconstruction network for medical images, Tsinghua Sci. Technol. 29 (1) (2023) 197–206.
[26] Z. Zhang, H. Jin, Z. Zheng, et al., Deep and domain transfer learning aided photoacoustic microscopy: acoustic resolution to optical resolution, IEEE T Med Imaging 41 (12) (2022) 3636–3648.
[27] D. He, J. Zhou, X. Shang, et al., De-noising of photoacoustic microscopy images by attentive generative adversarial network, IEEE T Med Imaging 42 (5) (2022) 1349–1362.
[28] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, Comput. Vis. –ECCV 2016 14th Eur. Conf. (2016) 694–711.
[29] Y. Zhang, K. Li, K. Li, et al., Image super-resolution using very deep residual channel attention networks, Proc. Eur. Conf. Comput. Vis. (ECCV) (2018) 286–301.
[30] H. Zhao, Z. Ke, F. Yang, et al., Deep learning enables superior photoacoustic imaging at ultralow laser dosage, Adv. Sci. 8 (3) (2021) 2003097.
[31] C. Ledig, L. Theis, F. Huszár, et al., Photo-realistic single image super-resolution using a generative adversarial network, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2017) 4681–4690.
[32] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2017) 3147–3150.
[33] Y. Zhang, Y. Tian, Y. Kong, et al., Residual dense network for image super-resolution, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2018) 2472–2481.
[34] A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, CoRR abs/1511.06434, 2015.
[35] A.G. Howard, M. Zhu, B. Chen, et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017. arXiv:1704.04861.
[36] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014. arXiv: 1409.1556.
[37] M. Wan, Y. Zhang, J. Li, et al., Optical-resolution photoacoustic microelastography system for elasticity mapping: phantom study and practical application, J. Biophotonics 17 (8) (2024) e202400032.
[38] Z. Li, Y. Liu, X. Chen, et al., Blueprint separable residual network for efficient image super-resolution, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) (2022) 833–843.
[39] Z.Y. Zhang, H.R. Jin, W.W. Zhang, A. Sharma, M. Pramanik, Y.J. Zheng, et al., Adaptive enhancement of acoustic resolution photoacoustic microscopy imaging via deep CNN prior, Photoacoustics 30 (2023) 13.
[40] W. Zhou, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE T IMAGE PROCESS 13 (4) (2004) 600–612.
[41] B. Stephanian, M.T. Graham, H.Y. Hou, et al., Additive noise models for photoacoustic spatial coherence theory, Biomed. Opt. Express 9 (11) (2018) 5566–5582.
[42] D. Han, Comparison of commonly used image interpolation methods, Conf. 2nd Int. Conf. Comput. Sci. Electron. Eng. (ICCSEE) (2013) 1556–1559.
[43] D. Zhu, D. Qiu, Residual dense network for medical magnetic resonance images super-resolution, CMPB 209 (2021) 106330.
[44] X. Wang, K. Yu, S. Wu, et al., ESRGAN: enhanced super-resolution generative adversarial networks, Proc. Eur. Conf. Comput. Vis. (ECCV) (2018) 1–23.
[45] F. Kong, M. Li, S. Liu, et al., Residual Local Feature Network for Efficient Super-Resolution, 2022. arXiv:2205.07514.

[46] S.F. Cheng, Y.Y. Zhou, J.B. Chen, et al., High-resolution photoacoustic microscopy with deep penetration through learning, Photoacoustics 25 (2022) 12.
[47] Y.F. Zhou, F.H. Zhong, S. Hu, Temporal and spectral unmixing of photoacoustic signals by deep learning, Opt. Lett. 46 (11) (2021) 2690–2693.
[48] X. Wei, T. Feng, Q. Huang, et al., Deep learning-powered biomedical photoacoustic imaging, Neurocomputing 573 (2024) 127207.

Yameng Zhang received her Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China in 2019. She is a Lecturer at Nanjing Institute of Technology and working as a postdoctoral fellow at Nanjing University of Aeronautics and Astronautics, Nanjing, China, since 2020. Her research interests are biomedical image enhancement, biomedical signal processing and laser speckle imaging.

Hua Tian is currently an undergraduate student majoring in Data Science and Big Data at Nanjing Institute of Technology, Nanjing, China. His interests focus on medical image processing and modelling.

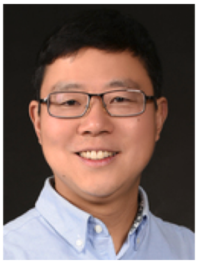
Min Wan received her B.S. degree in biomedical engineering from Hefei University of Technology, Hefei, China in 2022. She is currently a graduate student at the Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research project focuses on photoacoustic imaging, optical intrinsic signal imaging, and laser speckle imaging.

Shihao Tang received his B.S. degree from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2023. He is currently a graduate student in the department of biomedical engineering at the same university. His main research directions are photoacoustic imaging and traumatic brain injury.

Ziyun Ding received her B.Eng.and M.Eng degrees from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2006 and 2009, respectively, and Ph.D. degree from the University of Liverpool, Liverpool, UK, in 2013. She has been a Lecturer in Mechanical Engineering at the School of Engineering, University of Birmingham, UK, since 2019. Her research focuses on biomedical engineering, medical imaging, biomedical signal processing, modelling, and optimization.

Wei Huang is currently a Professor in the School of Computer Science at Nanjing Institute of Technology, Nanjing, China. Her research interests are big data security, big data Computing and cloud computing.

Yamin Yang is an Associate Professor in the Department of Biomedical Engineering at Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research focuses on biophotonics methods for cancer diagnosis and therapy.



Weitao Li is a Professor in the Department of Biomedical Engineering at Nanjing University of Aeronautics and Astronautics, Nanjing, China. He currently serves as the Deputy Secretary-General of the Jiangsu Society of Biomedical Engineering, and a member of the Third Biomedical Photonics Professional Committee of the Chinese Optical Society. His research interests are the development of biomedical imaging and detecting techniques, biomedical signal processing methods, and nanotechnology for early diagnosis of diseases.