Opto-Electronic Advances

CN 51–1781/TN ISSN 2096–4579 (Print) ISSN 2097–3993 (Online)

Enhanced photoacoustic microscopy with physics-embedded degeneration learning

Haigang Ma, Shili Ren, Xiang Wei, Yinshi Yu, Jiaming Qian, Qian Chen and Chao Zuo

Citation: Ma HG, Ren SL, Wei X, et al. Enhanced photoacoustic microscopy with physics-embedded degeneration learning. *Opto-Electron Adv* **8**, 240189(2025).

https://doi.org/10.29026/oea.2025.240189

Received: 16 August 2024; Accepted: 20 January 2025; Published online: 28 March 2025

Opto-Electronic Advances

Article

March 2025, Vol. 8, No. 3

DOI: 10.29026/oea.2025.240189 CSTR: 32247.14.oea.2025.240189

Enhanced photoacoustic microscopy with physics-embedded degeneration learning

Haigang Ma[1,2,3]*, Shili Ren[1,2,3], Xiang Wei[1,2,3], Yinshi Yu[1,2,3], Jiaming Qian[1,2,3], Qian Chen[1,3]* and Chao Zuo[1,2,3]*

Deep learning (DL) is making significant inroads into biomedical imaging as it provides novel and powerful ways of accurately and efficiently improving the image quality of photoacoustic microscopy (PAM). Off-the-shelf DL models, however, do not necessarily obey the fundamental governing laws of PAM physical systems, nor do they generalize well to scenarios on which they have not been trained. In this work, a physics-embedded degeneration learning (PEDL) approach is proposed to enhance the image quality of PAM with a self-attention enhanced U-Net network, which obtains greater physical consistency, improves data efficiency, and higher adaptability. The proposed method is demonstrated on both synthetic and real datasets, including animal experiments in vivo (blood vessels of mouse's ear and brain). And the results show that compared with previous DL methods, the PEDL algorithm exhibits good performance in recovering PAM images qualitatively and quantitatively. It overcomes the challenges related to training data, accuracy, and robustness which a typical data-driven approach encounters, whose exemplary application envisions to provide a new perspective for existing DL tools of enhanced PAM.

Keywords: photoacoustic microscopy; deep learning; high quality imaging; physical model

Ma HG, Ren SL, Wei X et al. Enhanced photoacoustic microscopy with physics-embedded degeneration learning. *Opto-Electron Adv* **8**, 240189 (2025).

Introduction

In recent years, photoacoustic imaging (PAI) has been drawing increasing attention from various research fields, including imaging, physics, chemistry, and biomedicine[1–3]. Briefly, in PAI, as photons travel in target, some of them are absorbed by intrinsic chromophores, and their energy is completely or partially converted into thermal energy. The energy then induces an initial pressure rise of target, which propagates as an acoustic wave. A piezoelectric ultrasonic transducer[4] or optical-acoustic transducer[5] detects the acoustic wave to yield an image, which maps the initial pressure rise distribution in the target[6]. By harnessing the high optical absorption contrast and low acoustic attenuation in the target, PAI is one of the rapidly growing biomedical imaging modalities, and has many interesting advantages, such as non-invasive, high resolution, and deep penetration to intrinsic and extrinsic chromophores, enabling morphological, functional, and molecular imaging of living tissues[7].

Photoacoustic microscopy (PAM) is a unique implementation of PAI where the spatial resolution is as fine as a micrometer or even nanometer, depending on the system configuration and imaging depth[8]. In PAM,

[1]Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; [2]Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing 210019, China; [3]Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing 210094, China.

*Correspondence: HG Ma, E-mail: mahaigang@njust.edu.cn; Q Chen, E-mail: chenqian@njust.edu.cn; C Zuo, E-mail: zuochao@njust.edu.cn

Received: 16 August 2024; Accepted: 20 January 2025; Published online: 28 March 2025

which inherits the characteristics of PAI, millimeter penetration depth can be obtained and is used for in vivo imaging of biological tissues[9]. The PAM systems have been developed in several forms and are used in many applications such as vascular biology[10], histology[11], oncology[12], dermatology[13], neuroscience[14], etc. However, PAM still faces several challenges. First, the resolution of PAM systems remains high within the depth of focus but deteriorates significantly at greater depths, primarily influenced by optical and acoustic scattering in biological tissues. Second, the detection sensitivity of PAM systems is determined by several critical parameters, such as the incident laser fluence, the target's optical absorption coefficient, and the detection efficiency of the ultrasonic transducer. Third, the Signal to Noise Ratio (SNR) of PAM images degrades at deep depths because of optical and acoustic attenuation in tissues[9,13]. The low SNR at deep depth can limit full understanding of the underlying conditions in large tissues. Thus, new methods are needed to improve the image quality of PAM while maintaining the safety of biological tissues and obtaining finer details, in order to expand the scope of PAM applications in preclinical studies.

Several methods have been shown to enhance the quality of photoacoustic (PA) images[15,16]. For PAM, image processing may be conducted for PA A-line signals[17] or for PA images directly[18]. Image processing of in vivo images was also tested in some of these works[19]. Among these image processing techniques, PA signal averaging is the most commonly used one[20], but acquisition of multiple signals is needed, which is time-consuming and leads to low temporal resolution. Further, researchers have tried to improve image quality using various algorithms to deal with the raw A-line signals, such as empirical mode decomposition[21], wavelet-based algorithms[22], and sparsity-based methods[23]. Alternatively, algorithms used in the image domain, such as synthetic aperture focusing technique (SAFT)[24], blind deconvolution technique (BDT)[25], and singular value decomposition (SVD)[26] can be used for PAM image processing directly. However, these methods still face some challenges: (i) prior information about the PAM image property is needed, which is a key bottleneck; (ii) time-consuming calculations are required, especially for those based on iterative and mutual optimization; (iii) some parameters such as the noise level or the attenuation compensation are needed to accommodate the manual settings of different input PA signals and images.

Deep learning (DL) is a data-driven approach that uses multilayered artificial neural networks for the automated processing of PA signals or images[27–31]. There are many applications to PAM such as artifact removal, super-resolution, denoising, and information enhancement, etc.[32]. For example, Allman et al. focused on using DL technology instead of traditional models based on iterative back-projection and compressive sensing algorithms, employing it as a post processing method to remove artifacts in reconstructed PAM images[33]. Song et al. utilized the DL method to improve lateral resolution of PAM at defocusing position, and remove background noise[34]. Liu et al. applied the DL models to achieve high-quality PAM images at low pulse laser energy for the purpose of high-speed PAM imaging[35]. Yao et al. applied modified U-Net structures for reconstructing the undersampled PAM images, and improving the PAM systems' imaging speed[36]. He et al. proposed a deep learning-based method to eliminate noise in PAM images without the need for manually selecting settings for different noisy images[37]. Cheng et al. trained a generative adversarial network (GAN) to enhance the lateral resolution of AR-PAM imaging[38]. Besides, data-driven and learning-based algorithms have also prevailed in PAM applications for image enhancement and reconstruction. For instance, Zheng et al. proposed an algorithm integrating both learning-based and model-based approaches, allowing a single framework to adaptively handle various distortion functions[39]. Furthermore, in ref.[40] the authors also displayed how the datasets obtained from a 4D spectral-spatial computational model could be utilized for neural network training to break the physical hardware boundary and biological constraints to improve the quality imaging capability of PAM. Although DL has many promising advantages and strengths compared with traditional processing approaches and physical modelling, there are also several challenges in making it robust and trustworthy for a wide range of PAM applications so that it can be reliably employed for policy- and decision-making. One of the greatest challenges of DL is that it does not always follow the underlying physical principles of the PAM systems it is applied to. While DL models can capture the underlying relationships within PAM data, their predictions do not always adhere to these principles, especially when applied to conditions outside their training scope, highlighting their limited ability to generalize to new scenarios. Moreover, achieving good results with DL models often relies on large, high-quality

datasets, which can be costly to obtain.

To address these challenges, here we developed a novel and effective physics-embedded degeneration learning (PEDL) approach to incorporate physical principles, governing laws, and domain knowledge of PAM into DL models. It is better suited to model the physical process of PAM by using tissue optics and acoustics to construct the physics-embedded degeneration mechanism and design a data degeneration model for PA with domain expertise. The adopted neural network of PEDL is based on a U-Net structure, incorporating multiple groups of residual blocks and a global context (GC) self-attention enhanced module. More complex features of PAM image can be extracted by using multiple residuals, and the gradient dispersion can be eliminated by residuals joining. Meanwhile, the global context information of the feature graph can be captured by GC self-attention mechanism, which can help the PEDL to better understand the overall structure of the PAM image. And, multiple sets of different weighted loss functions are combined to improve the network's effectiveness, enabling it to match PAM images under different degeneration situations. So that, the PEDL method can adaptively enhance PAM images at different depths, and recover finer structures. In addition, the PEDL does not require the time-consuming data collection tuning process compared with previous learning-based PAM enhancement algorithms, and by utilizing a lightweight GC module and batch normalization (BN) layer, which not only reduce the learning complexity but also accelerate convergence and help correct learning bias. Furthermore, the PEDL improves the generalizability of DL models to make reliable predictions for unknown scenes, including applicability to both acoustic resolution and optical resolution PAM systems, e.g. multiple types of noise interference, scattering changing biological tissue, and change of the laser energy, etc.

In the next few subsections, firstly, the proposed PEDL approach used to produce synthetic PAM images is discussed. Secondly, the architecture of self-attention enhanced U-Net network in the PEDL approach is detailed, including the residual block, the GC self-attention enhanced module, and the weighted loss functions. Thirdly, the PA dataset, feature network implementation, and evaluation metrics are presented in this work. Finally, other DL imaging methods for comparison with our proposed PEDL method are introduced. The results demonstrate that the proposed PEDL method can adaptively enhance the PAM images obtained across different noises and different imaging depths, scattering from different tissues, and different laser energy density, which can extend the application scenarios of the PEDL approach.

Physics-embedded degeneration learning method

In DL applications, a high-quality, large-scale and comprehensive dataset plays a vital role in determining the performance of the network. However, it is impractical to obtain a large number of in vivo image data for training for PAM in different situations. In this work, the physical model of PAM process is explored by adding a compensation function, and multiple groups of noises to simulate the imaging results under different conditions in the real environment, so as to characterize the degeneration of PAM, and establish a network to enhance the image information of PAM under different degeneration conditions. The following will introduce the degraded physical model and network structure, as well as the characterization functions Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM), and the combined loss function.

Degeneration model based on physical mechanism

This work utilizes high-quality images provided by Duke University's Photoacoustic Imaging Laboratory, which have been publicly announced to be available[41]. This dataset consists of in vivo microvascular images of the mouse brain obtained using an OR-PAM system with a lateral resolution of 5 micrometers. However, images under different imaging conditions for training are not provided. Therefore, for synthetic training dataset generation, a degeneration simulation model of photoacoustic microscopy under different imaging conditions is proposed, incorporating compensation functions and various noise sets to synthesize matching image pairs.

According to acoustic beamforming theory, the photoacoustic imaging mechanism at different depths can be characterized by the physical model of a spherical focusing transducer and its generated field pattern[42,43]. First, the ultrasound transducer is discussed, and from the Rayleigh-Sommerfeld diffraction formula, it is derived:

$$\tilde{\Phi}\left(\boldsymbol{r}_0,\omega\right)=\frac{1}{\mathrm{i}\lambda}\int_0^a\int_{-\pi}^{\pi}\left(\tilde{\Phi}\left(\boldsymbol{r}_1,\omega\right)\right)\mathrm{e}^{\mathrm{i}kr_{01}}\frac{z}{{r_{01}}^2}r_1\mathrm{d}r_1\mathrm{d}\phi_1 + \frac{1}{2\pi}\int_0^a\int_{-\pi}^{\pi}\left(\tilde{\Phi}\left(\boldsymbol{r}_1,\omega\right)\right)\mathrm{e}^{\mathrm{i}kr_{01}}\frac{z}{{r_{01}}^3}r_1\mathrm{d}r_1\mathrm{d}\phi_1\,. \quad (1)$$

In the above formula, the first term and the second term represent the high-frequency and low-frequency components, respectively. $\tilde{\Phi}(\boldsymbol{r}_0,\omega)$ is the Fourier transform of the wave field $\tilde{\Phi}(\boldsymbol{r}_0,t)$ at the spatial point, and $\tilde{\Phi}(\boldsymbol{r}_1,\omega)$ is the Fourier transform of the aperture weighting function $\tilde{\Phi}(\boldsymbol{r}_1,t)$. k represents the wave number. r_{01} is the distance between the field point r_0 and the source point r_1, given by the following formula:

$$r_{01}=\sqrt{(x_0-x_1)^2+(y_0-y_1)^2+z^2}\,. \tag{2}$$

Equation (1) involves two double integrals, making its computation time-consuming. Notably, in practical beamforming scenarios, $r_{01}\gg\lambda/2\pi$ is usually satisfied, allowing the second term in Eq. (1) to be neglected. Therefore, Eq. (1) can be simplified as follows:

$$\tilde{\Phi}(\boldsymbol{r}_0,\omega)=\frac{1}{\mathrm{i}\lambda}\int_0^a\int_{-\pi}^{\pi}\left(\tilde{\Phi}(\boldsymbol{r}_1,\omega)\right)\mathrm{e}^{\mathrm{i}kr_{01}}\frac{z}{{r_{01}}^2}r_1\mathrm{d}r_1\mathrm{d}\phi_1\,. \tag{3}$$

A planar circular transducer ($\tilde{\Phi}(\boldsymbol{r}_1,\omega)\equiv 1$) can be derived from Eq. (3) as follows:

$$\tilde{\Phi}(\boldsymbol{r}_0,\omega)=\frac{ka^2}{\mathrm{i}2z}\mathrm{e}^{\mathrm{i}k\left(z-r_0^2/2z\right)}\left[2\frac{\mathrm{J}_1(kar_0/z)}{kar_0/z}\right]\,. \tag{4}$$

where J_1 is the first-order Bessel function of the first kind, and a is the radius of the ultrasound transducer[44]. The field pattern of a spherical focusing ultrasound transducer can be obtained by combining the field pattern of a planar circular transducer with a spherical compensation function, which is given by:

$$\tilde{\Phi}(\boldsymbol{r}_1,\omega)=\mathrm{e}^{-\mathrm{i}k\left(\sqrt{F^2+r_1^2}-F\right)}\,, \tag{5}$$

where F is the focal length of the transducer, and from Eq. (3), it can be deduced that:

$$\tilde{\Phi}(\boldsymbol{r}_0,\omega)\propto 2\frac{\mathrm{J}_1(kar_0/F)}{kar_0/F}\,. \tag{6}$$

The spherical focusing transducer then integrates over the frequency $T(\omega)$ to the sensor. The integral is taken over the bandwidth of the ultrasound transducer. Thus, the PSF kernel can be obtained as follows:

$$h=\int T(\omega)\tilde{\Phi}(\boldsymbol{r}_0,\omega)\,\mathrm{d}\omega\,. \tag{7}$$

As the depth increases, the image information gradually weakens due to the increased scattering medium it passes through. However, using maximum intensity projection (MIP) images, which treats the image as a whole, is unreasonable because weaker signals experience more severe attenuation at greater depths.

Therefore, as shown in Fig. 1, this work introduces a compensation function to address the degeneration differences. Z represents depth, I denotes images from different processes, and FWHM stands for full width at half maximum. The signal gradually weakens, with weaker signals experiencing more severe attenuation. The specific compensation function is given by:

$$\alpha(x)=\mu x^{1+\xi z}\,, \tag{8}$$

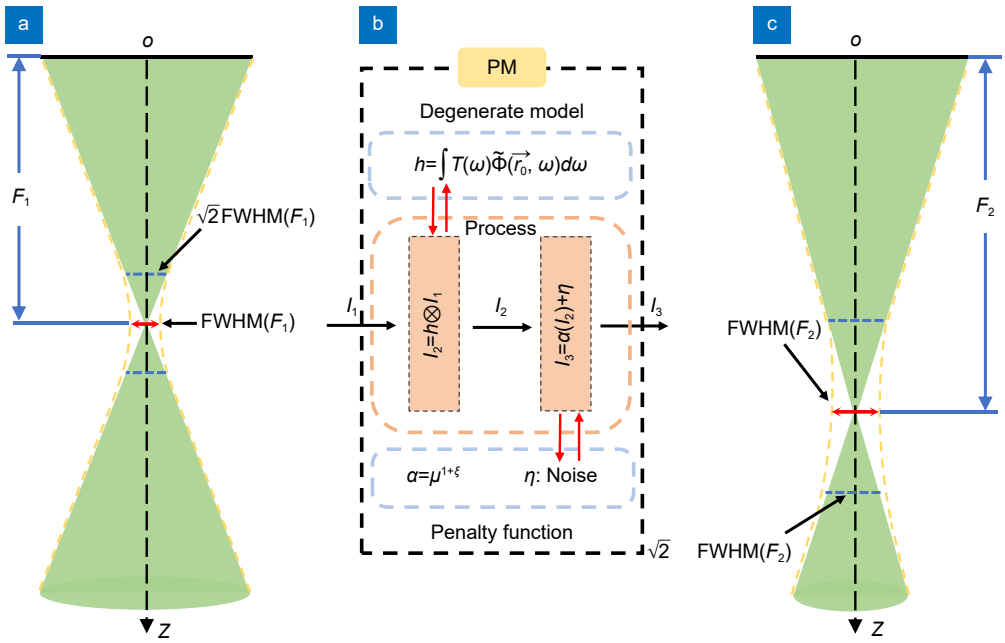

Fig. 1 | Schematic diagram of physical mechanism. (**a**) Field pattern of the transducer at a shallower focal plane. (**b**) A simple schematic of the physical model. (**c**) Field pattern of the transducer at a deeper focal plane.

where μ is the attenuation coefficient and ξ is the compensation coefficient. Simultaneously, during the attenuation process, the image information will inevitably be affected by noise. Therefore, it can be deduced that:

$$\eta = \eta_1 + \eta_2 + \eta_3 \,, \tag{9}$$

where η is the total noise. η_1, η_2, η_3 are Gaussian noise, Rayleigh noise, and Poisson noise, respectively, with Gaussian noise being the primary noise distribution. Finally, integrating the above equations, it can be deduced that:

$$I_m = h \otimes (\alpha(I_{m1})) + \eta \,, \tag{10}$$

where I_{m1} is the original image, and I_m is the reconstructed image at different depths. If η remains constant, it is unreasonable and unrealistic. Therefore, η will randomly vary within a range and will increase as the degree of degeneration intensifies.

Structure of the network

Deep learning has proven to be a promising technology for many visual recognition tasks. Deep learning-based medical imaging has been widely discussed and is considered to have great potential, especially in enhancing raw data. The proposed network in this work, PEDL, primarily uses convolutional neural networks to achieve adaptive enhancement of PAM imaging under different degeneration conditions.

Initially, the preliminary reconstruction network structure in this work is based on the U-net network, with several modifications to match the physical model's mechanism. These modifications are designed to improve the network's performance in image restoration and adaptive capacity. The physical model is integrated into the network. During each iteration, it generates dynamic data, satisfying physical laws while incorporating inherent randomness. Specifically, during network training, the convolution kernel (h), noise intensity (η), and compensation function (α) continuously vary and exhibit a certain degree of weak correlation. Each parameter varies within a defined range and is influenced by a common scaling factor. This weak correlation better reflects real-world scenarios, as different imaging systems have varying noise and resolution characteristics. Additionally, changing the physical model in each iteration enables a many-to-one training approach, consistent with the variation in noise distribution in actual imaging. The detailed structure of the network is shown in Fig. 2(b). The batch size for training is set to 12. The network's input uses a convolution kernel size of 9×9. Compared to the traditional 3×3 convolution kernel, the 9×9 convolution kernel provides a larger field of view, which helps in better extracting the global features of the image[45,46]. Following this, multiple Res blocks (detailed structure shown in Fig. 2(d)) are used to modify the original U-Net. A series of Res blocks with 3×3 convolution kernels are placed, and the stacking of multiple convolution blocks can extract more complex and abstract features layer by layer[47]. This allows for multi-resolution analysis and recovery of both major and micro vessels. Additionally, the introduction of residual parts enhances the network, addressing issues of gradient dispersion and accuracy degeneration in deep networks, allowing the network to go deeper while maintaining accuracy and speed. Due to the varying sizes of vessels, and since the U-Net structure is not very sensitive to very high spatial frequency information, this network incorporates BN methods in each down-sampling convolution layer and intermediate convolution layers to overcome this issue. The BN layers also improve the network's convergence speed.

Secondly, this network uses skip connections to link the encoder and decoder of the same layer, preserving spatial information for subsequent image reconstruction. A simple schematic of the physical model is shown in Fig. 2(c). During network training, the values of h, α, and η are changed with each iteration to improve the network's adaptability and enhance the image reconstruction performance under different degrees of degeneration. The GC block attention module is also introduced into the network of this work[48]. In Fig. 2(e), X, Q, and W are solely used as abstract symbols for modules, without specific meanings, and are only intended to represent different module concepts. The following formula provides a simple representation of the GC attention module.

$$z_i = x_i + W_{v2}\mathrm{ReLU} \cdot \left(LN\left(W_{v1} \sum\nolimits_{j=1}^{N_p} \frac{\exp(W_k x_j)}{\sum_{m=1}^{N_p} \exp(W_k x_m)} x_j \right) \right) . \tag{11}$$

In the above Eq. (11), $e^{W_k x_j} \Big/ \sum_{m=1}^{N_p} e^{W_k x_m}$ is the weight of the GC attention. $W_{v2}\mathrm{ReLU}\left(LN\left(W_{v1}(\cdot)\right)\right)$ is the transformation module. The GC attention module uses a self-attention mechanism to capture the global context information of the input feature map, which helps the model understand the overall structure of the image. Additionally, since the GC attention module is lightweight, it can

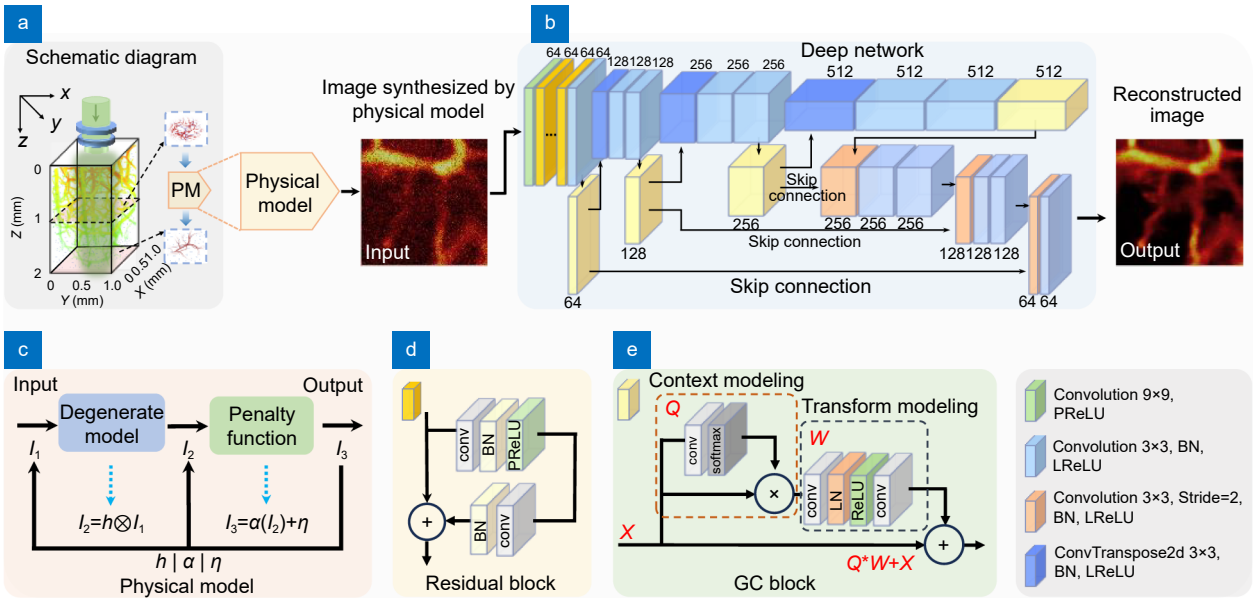

Fig. 2 | Network structure and flow chart. (**a**) Schematic diagram of the physical model. (**b**) Basic structure diagram of the proposed network. (**c**) Image generation process of the physical model. (**d**) Residual module. (**e**) GC attention module.

be applied to multiple different layers to better capture the global context information of the input feature map. At the same time, because GC pays attention to modular lightweight, it can improve the training speed very well. Using the GC attention module can enhance the analysis and recognition of microvasculature, providing possibilities for subsequent image reconstruction. Finally, the network structure uses a 9×9 convolution kernel for the output, aiming to integrate multi-scale information. This helps enhance feature representation capabilities, smooths the feature map, and preserves the overall structure of the image. By integrating these additional components, the network model demonstrates higher sensitivity to vascular features of different scales, stronger noise resistance, and excellent contrast enhancement effects. At the same time, it can match the physical model well.

Quantitative analysis

PSNR represents the peak signal-to-noise ratio. It is generally used as a ratio between the maximum signal and background noise, as follows:

$$PSNR = 10 \cdot \log_{10}\left(\frac{(2^n-1)^2}{MSE}\right) . \tag{12}$$

In the above formula, Mean Squared Error (MSE) is the pixel-level mean squared error between the two images. $(2^n-1)^2$ is the square of the maximum possible signal value, where n is the number of bits per sample. The higher the PSNR value, the better the quality of the reconstructed image.

The pixel similarity between images cannot fully and accurately reflect the quality of the reconstructed image. According to the SSIM theory, there is a strong correlation between pixels. SSIM evaluates image quality by measuring luminance, contrast, and structural similarity, and then combines these aspects for the final image quality assessment.

The final calculation formula for the SSIM index is as follows:

$$SSIM = l^{\alpha} c^{\beta} s^{\gamma} . \tag{13}$$

In Eq. (13), l, c and s represent adjustments for luminance, contrast, and structural similarity, respectively. α, β and γ are parameters indicating the weights of luminance, contrast, and structural similarity measures for the SSIM index. Typically, $\alpha = \beta = \gamma = 1$. The final SSIM value ranges from [0, 1], with higher values indicating a higher degree of structural similarity between the two images, thus implying better image quality of the reconstructed image.

Combined loss function

Throughout the entire network training process, the loss function is very important. To enhance the resolution of photoacoustic imaging at different depths while maintaining clarity and effective information, this work proposes a combined loss function to assist in network training. Our loss function mainly consists of two parts:

(1) MSE Loss[49]; (2) perceptual Loss[50].

MSE Loss: MSE is a loss function used to measure the difference between model predictions and actual observations. In deep learning, MSE is often used as a loss function for regression problems, especially in tasks such as image generation and numerical prediction. MSE is calculated by squaring the differences between predicted and true values and then averaging them. For a dataset with n samples, the MSE formula is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i')^2 , \tag{14}$$

where Y_i is the actual observed value, Y_i' is the model's reconstructed prediction, and n is the number of samples.

Perceptual Loss: Pixel-wise loss functions (e.g., MSE) directly compare differences based on all pixels, producing a single value for the entire image. This approach does not focus on the recovery of sharp edges and may lead to overly smooth patterns. In contrast, perceptual loss, utilizing convolutional filters (including edge detectors), is more likely to recover fine features. Therefore, to better restore images, this work adopts perceptual loss, which includes various convolutional filters to extract fine textures and focuses on high-level features, making the restoration performance more aligned with human visual perception. Specifically, high-level features are extracted from a pre-trained VGG-19 model.

$$Perceptual = \frac{1}{n}\sum_{i=1}^{n}(F(Y_i) - F(Y'_i))^2 , \tag{15}$$

where $F(Y_i)$ represents the i-th element of the actual image in this feature layer, $F(Y'_i)$ represents the i-th element of the generated image in this feature layer, and n represents the number of feature layers. By integrating both loss functions, the model benefits from precise pixel-level correction and enhanced perceptual quality, leading to more visually appealing and accurate reconstructions.

Experimental results and analysis

The size of the training dataset images is standardized to 128×128 sub-images and then 5000 images are selected for training. During network training, 96×96 images are randomly selected from the 128×128 sub-images for training at each iteration to avoid overfitting. We randomly crop 500 images of size 400×400 from large-scale images in the public dataset for testing. All input images are normalized to the range [0, 1] before training. Adam optimizer known for its robustness is used to optimize the model weights during training. The trained models are used as training generated models in the network framework. The neural network model is developed using PyTorch in Python 3.6 and trained on NVIDIA GeForce RTX 4060 laptop GPU platform. The model is trained for 80 epochs, with an initial learning rate of 5×10^{-5}, which is reduced to 60% of its original value at the 20 th and 40 th epochs. Additionally, model outputs and the loss function are monitored in real time using TensorBoard, and convergence is determined when these values stabilized. In this work, the PSNR and SSIM metrics are calculated based on the Y channel of the images. The Y channel, which represents the luminance component in the YCbCr color space, captures the brightness information of the image. It is also the aspect most sensitive to human perception of image quality. By adopting this approach, the quality assessment focuses on the details most relevant to human perception, such as edges and textures, which are particularly critical for the analysis of photoacoustic imaging.

Simulation experiments

To test the denoising performance of our network, this work first adjusts the physical model to generate various noise patterns and conducts network training. Then, the reconstruction effects under different noise levels are analyzed.

Different levels of noise are added to three types of images, and the network is used for reconstruction and comparison with the ground truth images. Figure 3(a–c) show three different ground truth images. Various levels of noise, including Gaussian noise, Rayleigh noise, Poisson noise, and background noise, are added to the three types of images, with Gaussian noise being the primary noise. Here we provide partial definitions of the noise distribution standards. For a "low" noise distribution, Gaussian noise with a standard deviation of 10 is added, and additional Gaussian noise is applied to regions with pixel intensities below 30 to enhance noise interference in small vessel signals. For a "medium" noise distribution, Gaussian noise with a standard deviation of 20 is added, with extra Gaussian noise introduced in regions where pixel intensities are below 30. For a "high" noise distribution, Gaussian noise with a standard deviation of 30 is added, along with additional Gaussian noise in areas with pixel intensities below 30. Different levels of noise are shown in Fig. 3(a1–c1, a3–c3, a5–c5). The reconstruction effects of different noise levels are evident,

as shown in Fig. 3(a2–c2, a4–c4, a6–c6). From these figures, it is clear that the network adaptively enhances images under different noise conditions. The proposed network effectively removes noise and significantly improves image contrast. In Fig. 3(d–f), the specific SSIM and PSNR characteristics of the three images are shown. The line charts indicate that as noise increases, the input PSNR and SSIM gradually decrease, and the difference between input and output PSNR and SSIM increases. The output PSNR and SSIM remain relatively stable, and in some cases, the restoration effect improves with increasing noise, as seen in Fig. 3(e). The bar chart represents the weighted product of PSNR and SSIM, demonstrating the significant improvement after network reconstruction, generally stabilizing at over 25. In addition to deep learning methods, there are many non-deep learning techniques for improving photoacoustic image quality, such as sparse coding and SLD-PAM[51,52]. However, sparse coding denoising techniques are typically not suitable for real-time processing and require a long processing time, while SLD-PAM has high hardware requirements and involves complex computations. Therefore, in this work, we select two conventional non-deep learning methods and compare them with our approach[53,54]. Quantitative analysis of the network reconstruction under different noise conditions is detailed in Table 1. From the SSIM and PSNR metrics, it is evident that the network output shows a significant improvement in these quantitative metrics, with SSIM values stabilizing around 0.9. This demonstrates the network's strong adaptability to different noise levels. Under high-noise conditions, the PEDL method can efficiently remove noise and significantly enhance the quality of photoacoustic microscopy images, outperforming the two conventional methods.

To further evaluate the denoising performance of the network, more detailed feature analysis is conducted on images under strong noise conditions.

As shown in Fig. 4, a significant improvement can be seen from Fig. 4(a) to Fig. 4(b). In the magnified regions, the network shows a strong recovery effect on signals masked by noise, as seen in Fig. 4(a1–a3, b1–b3, c1–c3).

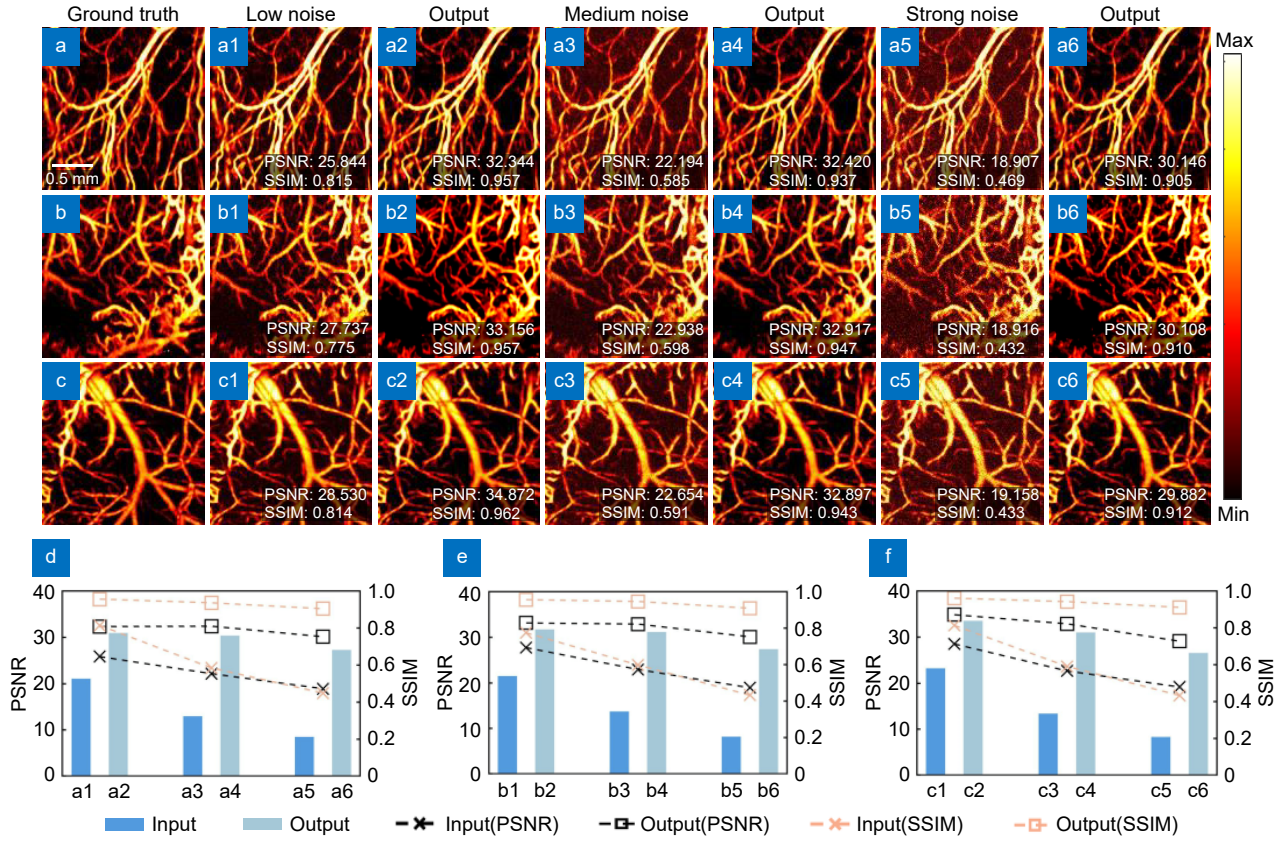

Fig. 3 | Processing effect diagram of different noises. (**a–c**) are the ground truth images of three different tests. (a1–c1, a3–c3, a5–a5) are the low noise, medium noise, and high noise images of the three different tests, respectively. (a2–c2, a4–c4, a6–c6) are the images output by the network for the three different test images. (**d–f**) are the data analyses of the three test images, with PSNR on the left and SSIM on the right, and the bar chart showing the product of PSNR and SSIM weights.

Table 1 | Quantitative analysis of different noises.

	Low noise		Medium noise		Strong noise	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Input	27.730	0.784	23.409	0.506	19.684	0.415
NLMD[53]	30.358	0.917	28.690	0.879	23.371	0.756
NAD[54]	27.376	0.866	26.151	0.834	23.970	0.798
PEDL	33.076	0.957	32.602	0.940	30.343	0.906

Figure 4(d–f) demonstrate that originally, the signals in the highlighted areas are covered by noise. However, after network training, small signals are significantly recovered, and their energy intensity is enhanced. Comparing Fig. 4(b3) and 4(c3), it is evident that the small signal intensity is strengthened after network output. Particularly in Fig. 4(d), analyzing the maximum intensity projection reveals that small signals, which are barely observable in the original image due to noise, are recovered after network reconstruction. From the above results, it can be concluded that the network effectively removes various noise levels and recovers small signals masked by noise, while also enhancing weaker signals. The comparison of PEDL, cGAN[55] and U-net[56] algorithms in denoising effect is also explored. The denoising performance of the above two networks and the PEDL method is studied using simulated data. A detailed comparison is shown in Fig. S1. The results indicate that the PEDL method outperforms the other two networks in terms of both PSNR and SSIM.

To comprehensively test the performance of our network in all aspects, the physical model is adjusted to degeneration mode and network training is carried out. Then, this work analyzes the reconstruction effect of photoacoustic microscopic images under different degeneration degrees.

In the degeneration model, the value of K represents the degree of degeneration, related to the sizes of a and F in Eq. (6). We use K to represent the overall image degradation, where an increase in K leads to a decrease in

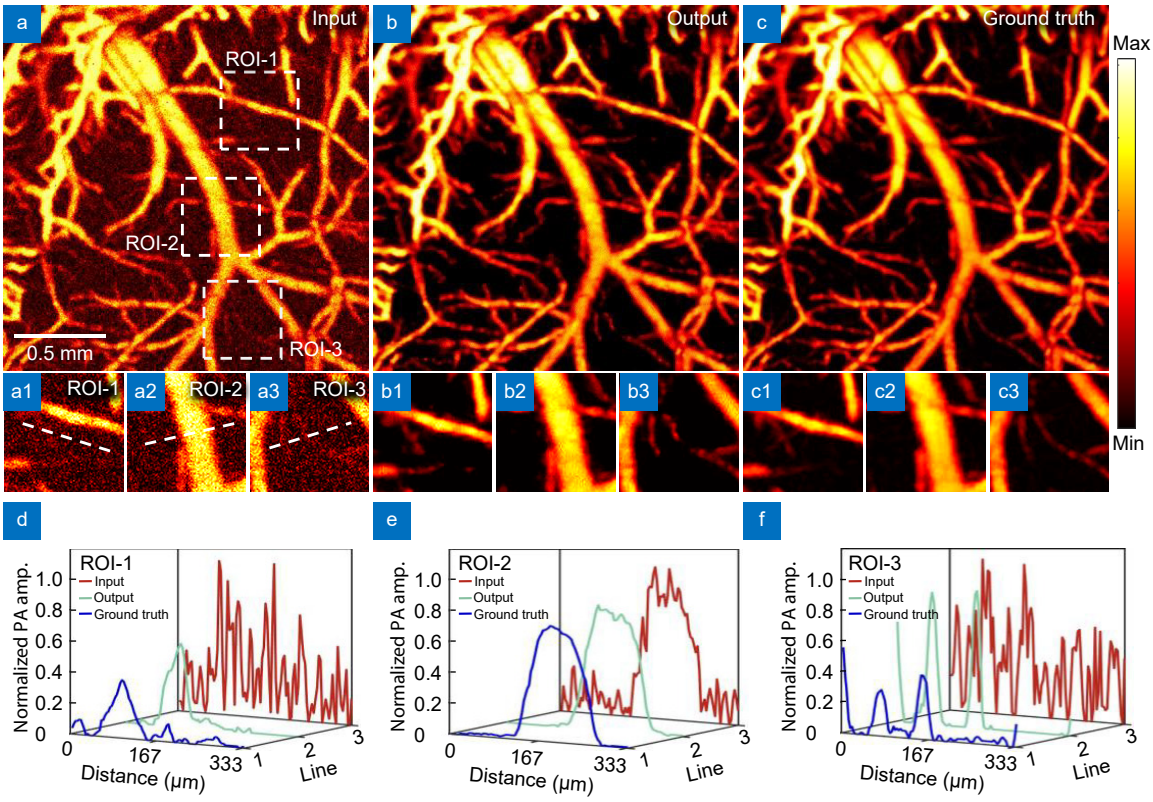

Fig. 4 | Effect diagram under strong noise. (**a**) Image input with strong noise. (**b**) Network reconstruction result. (**c**) Ground truth image. (a1−a3, b1−b3, c1−c3) are the magnified images of the corresponding ROI regions. (**d**) Maximum intensity projection analysis along the line in ROI-1. (**e**) Maximum intensity projection analysis along the line in ROI-2. (**f**) Maximum intensity projection analysis along the line in ROI-3.

resolution, accompanied by an increase in noise, including Gaussian noise, Rayleigh noise, Poisson noise, and background noise, with Gaussian noise being the predominant type. Figure 5 shows the resolution enhancement with K=15 but without noise. From Fig. 5(a1–a2) to Fig. 5(a3–a4), it can be seen that the resolution is significantly improved, and the weak signal is enhanced. Figure 5(b) and 5(c) show that, compared to the input images, the network output images have a smaller FWHM and that the network's output at small vessel locations is closer to the ground truth image. The FWHM of the fine blood vessels improve by 12.0% after network processing.

Figure 6(a) shows data under different degeneration conditions, where increasing K values lead to lower resolution and higher noise levels. Figure 6(b) shows network-reconstructed images under different degeneration conditions, where vascular structures are significantly improved. Figure 6(c, d) show PSNR and SSIM values between reconstructed and input images compared to ground truth images, as well as the differences between input and output images, with the specific differences determined by the coordinates on the right side of the graph. When the K value is 3, there is little difference between the reconstructed image and the input image. As K increases to 6, an enhancement in the signal intensity of small vessels can be observed in the reconstructed image compared to the input. With a further increase in K to 15, the input image resolution significantly decreases, and vessel structures become difficult to distinguish. However, after network reconstruction, the vessel structures are well restored. Overall, the network reconstruction demonstrates strong generalization, as it does not forcefully alter the vessel diameter when K is small. Instead, it adapts to the input image and adjusts the reconstruction strength accordingly. Table 2 provides quantitative analysis results for K values of 3, 6, 9, 12, and 15. Although PSNR of network-reconstructed images decreases with increasing K values, the difference between input and output PSNR gradually increases. SSIM improvement remains stable, with reconstructed image SSIM generally above 0.8. When K=15, the PEDL method can still effectively improve both PSNR and SSIM, keeping them within a valid range. It can also be seen from the above figure that the improvement effect of the network is adaptive under different degeneration conditions, and the reconstruction effect will change according to the different input data.

To further evaluate the performance of the network in enhancing resolution, the features of severely degraded images are analyzed in detail.

As shown in Fig. 7, there is a significant improvement

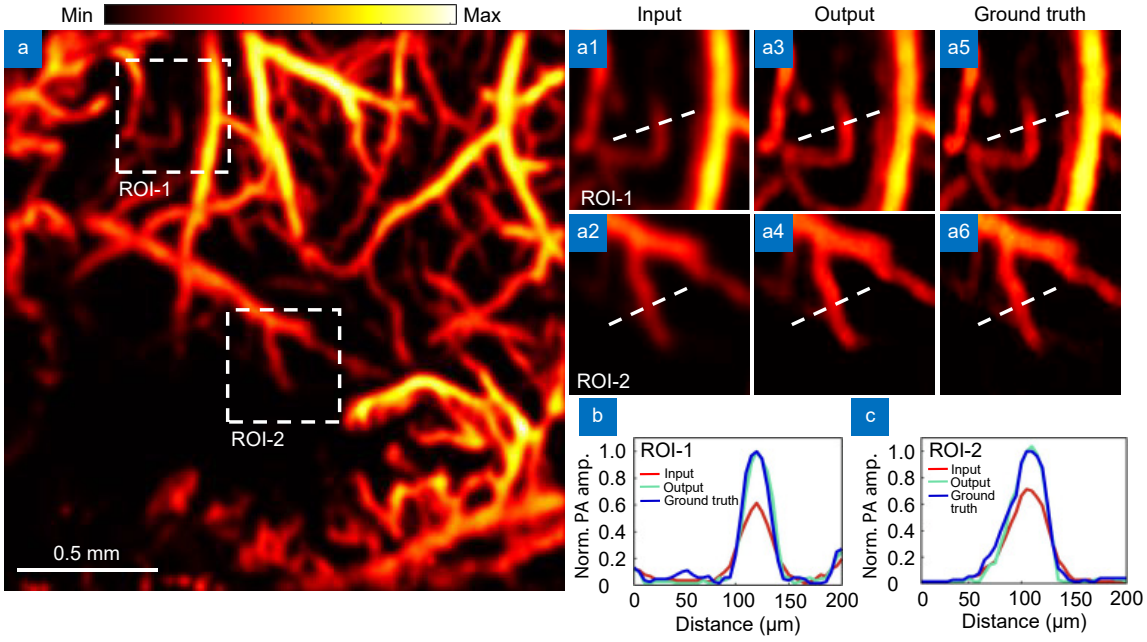

Fig. 5 | Resolution enhancement effect diagram. (**a**) Severely degraded, noise-free image input. (a1–a2) are magnified images of the ROI regions of the input images. (a3–a4) are magnified images of the ROI regions of the output images. (a5–a6) are magnified images of the ground truth ROI regions. (**b**) Maximum intensity projection analysis along the line in ROI-1. (**c**) Maximum intensity projection analysis along the line in ROI-2.

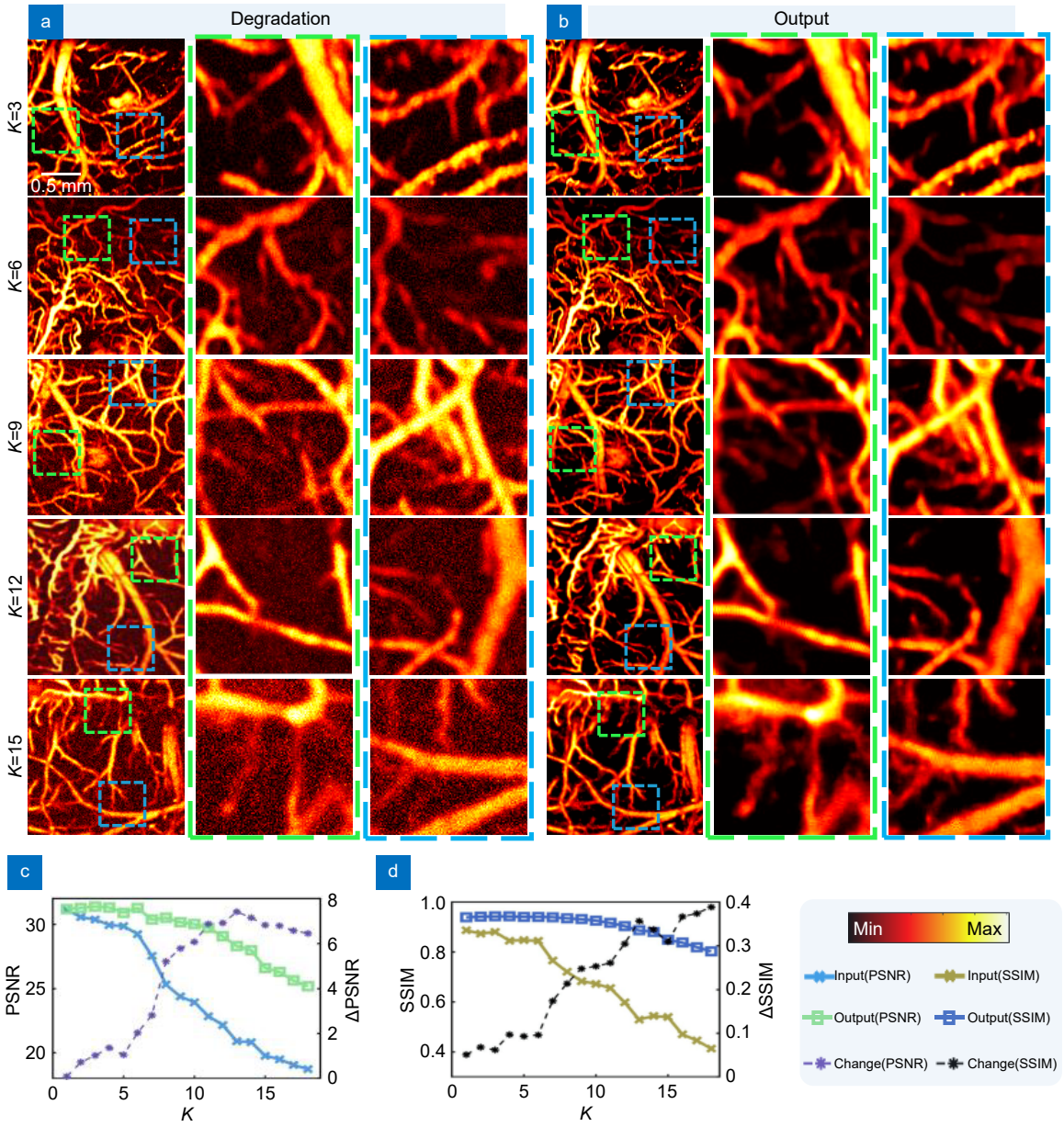

Fig. 6 | Reconstructed images with different degeneration effects. (**a**) Images under different degeneration models. (**b**) Results of different degeneration models after network output. (**c**) PSNR feature analysis for different degradation models, where ΔPSNR represents the difference in PSNR between the input and output. (**d**) SSIM feature analysis for different degradation models, where ΔSSIM represents the difference in SSIM between the input and output.

Table 2 | Quantitative analysis of different degeneration degrees.

	K=3		*K*=6		*K*=9		*K*=12		*K*=15	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Input	30.371	0.881	29.239	0.845	24.380	0.684	22.151	0.599	19.766	0.540
PEDL	31.392	0.943	31.271	0.941	30.163	0.932	29.064	0.904	26.597	0.849

from Fig. 7(a) to Fig. 7(b). In the magnified regions, the resolution enhancement is evident, as shown in Fig. 7(a1–a3, b1–b3, c1–c3). Figure 7(d–f) show that the originally noisy regions are effectively denoised after network reconstruction, improving the image resolution. In Fig. 7(a–c), the yellow arrows indicate blood vessels that are difficult to discern due to noise interference; however, after network reconstruction, the reconstructed

images are much closer to the ground truth. In Fig. 7(d–f), the areas marked by orange arrows show that the noise is effectively removed, and the FWHM of the vessels is well restored. Specifically, in Fig. 7(d), analyzing the maximum intensity projection reveals that the input image is noisy except for the vascular regions. After network reconstruction, the noise is clearly removed, and the FWHM of the vessels is reduced from 51 μm to 31 μm, close to the ground truth of 30 μm, enhancing the vessel signal. Analyzing these results, the conclusion is that the network can effectively restore the images under different degeneration conditions, significantly improve the resolution and reduce the noise. At the same time, this work investigates the performance of different algorithms in terms of resolution enhancement on simulated data. A detailed comparison can be found in Fig. S1. The results indicate that PEDL outperforms both the cGAN and U-Net algorithms across various metrics.

Phantom experiments

To validate the capability of the preliminary model in processing real imaging data, phantom experiments are conducted. Two experiments are done here to verify the processing ability and generalization ability of the PEDL. Because the most important medical application of PA is vascular system imaging, and our previous model is also trained with mouse vascular system data set, we adopt the phantom created with the imaging results of mouse vascular system. The experimental setup employs a nanosecond laser system operating at 532 nm with a pulse frequency of 10 kHz[57]. The laser beam is attenuated, spatially filtered and coupled to one optical fiber as the excitation source, and then it is cleaned by a spatial filter system, which passes through a 10 nm diameter pinhole for spatial filtering, and then coupled into a single-mode fiber by a fiber coupler. The single-mode fiber guides the laser beams into a PA probe with a microscope objective (MPlan Apo HL, 5X/0.13, Sigma), a ringed focused ultrasonic transducer, and a 2D scanner. The transducer has a central frequency of 30 MHz. The 2D scanning is performed using a bow-shaped scanning trajectory.

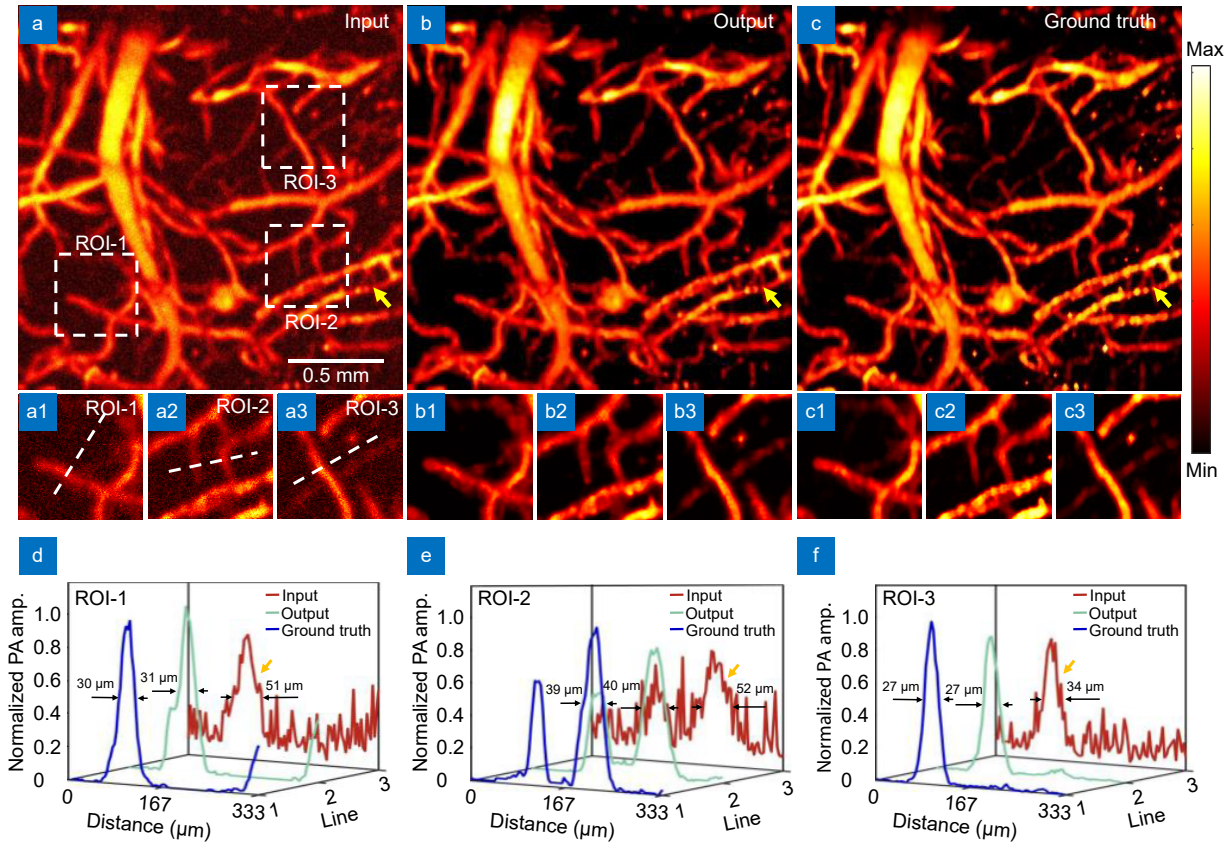

Fig. 7 | Reconstruction effect diagram under high degeneration. (**a**) Input images with severe degeneration. (**b**) Network reconstruction results. (**c**) Ground truth images. (a1–a3), (b1–b3), (c1–c3) are the magnified images of the corresponding ROI regions. (**d**) Maximum intensity projection analysis along the line in ROI-1. (**e**) Maximum intensity projection analysis along the line in ROI-2. (**f**) Maximum intensity projection analysis along the line in ROI-3.

The first experiment uses silicone to simulate skin by covering a mouse's ear with silicone of different thicknesses to mimic real imaging at different depths. Silicone and skin have similar absorption and scattering properties. The same area of mouse's ear is imaged with silica gel sheets with different thicknesses. The specific experimental results are shown in Fig. 8.

In the experiment, it is evident from Fig. 8(a, c, e) that as the thickness increases, the overall brightness of the obtained images decreases, and the noise gradually increases. Figure 8(a, c, e, g) are photoacoustic microscopy images with four layers, two layers, one layer, and no silicone sheets, respectively, while Fig. 8(b, d, f) correspond to the network output images. After network processing, it can be seen from Fig. 8(a, c, e) to Fig. 8(b, d, f) that small signals are enhanced, noise is reduced, and overall image resolution is improved. Figure 8(h–m) show that compared to the input images, the network output images have a smaller FWHM. Figure 8(h, j) show that after network processing, the FWHM is smaller and background noise is effectively removed. Analyzing these results, the conclusion is that the network effectively improves the blood vessels with different depths, improves the overall resolution and effectively eliminates the background noise. Under the condition of 4 silicone sheets, the FWHM improves by an average of 60.4% after processing through the network. With 2 silicone sheets, the FWHM improves by an average of 19.9%, and with 1 silicone sheet, the improvement is 10.2%. These results indicate that the network possesses an adaptive capability, showing greater improvement as the initial FWHM of the imaging result deviates further from the ground-truth FWHM. This quantitative evidence further confirms that our network achieves resolution enhancement across different imaging thicknesses. The data at different depths are expressed in different ways, which proves that the method can improve the resolution and denoising. In the actual experiment, we can see that the network remains adaptive. The more significant resolution improvement and adaptability of PEDL can be seen in Fig. S2.

In the second experiment, the ears of mice are imaged by photoacoustic microscopy at different energy levels. The mouse's ears are irradiated with lasers with different power levels and image information is obtained. In this experiment, the mouse's ears are imaged using lasers with 100% power, 66% power, and 33% power. The specific experimental results are shown in Fig. 9.

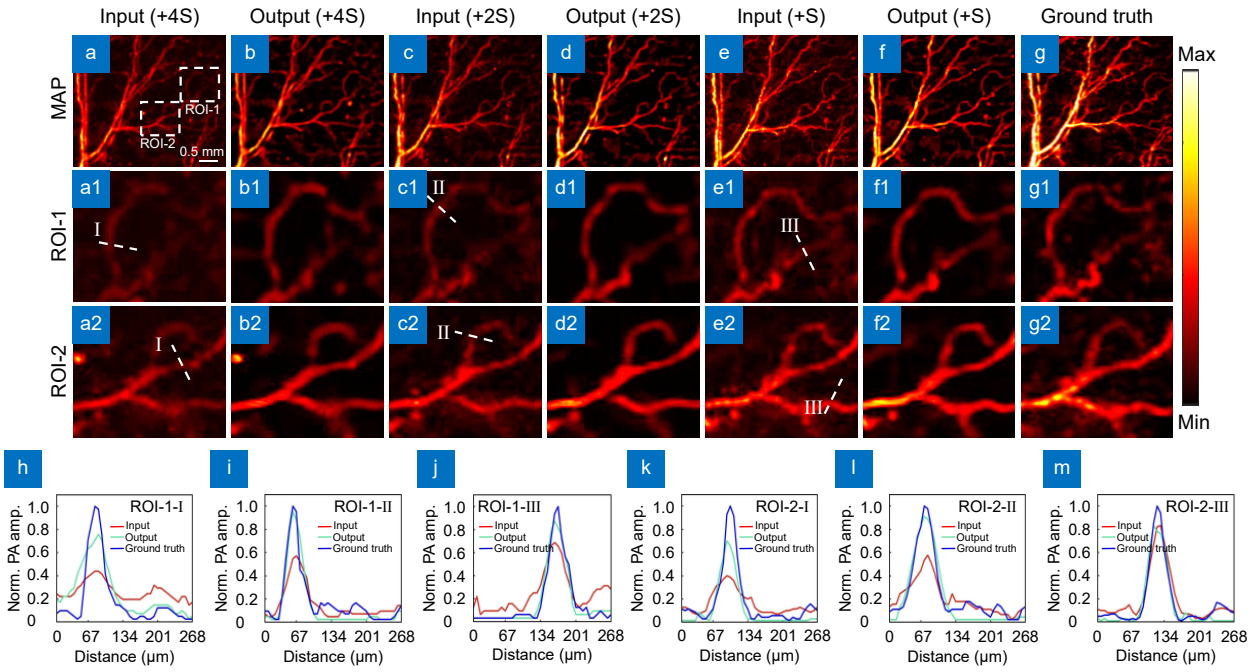

Fig. 8 | Imaging effect and reconstruction results at varying thickness levels. (**a**, **b**) Photoacoustic microscopy images and corresponding network output for four silicone layers. (**c**, **d**) Photoacoustic microscopy images and network output for two silicone layers. (**e**, **f**) Photoacoustic microscopy images and network output for one silicone layer. (**g**) Photoacoustic microscopy image with no silicone sheet added. (a1–g1, a2–g2) Magnified views of two ROI regions in the input images. (**h–j**) Maximum intensity projection analysis along the line in the magnified images of ROI-1. (**k–m**) Maximum intensity projection analysis along the line in the magnified images of ROI-2.

In the experiment, it is evident from Fig. 9(a–f) that as the laser power decreases, the overall brightness of the obtained images diminishes, and fine vessels gradually disappear. After processing with our network, from Fig. 9(a–f) to Fig. 9(a1–f1), small signals are enhanced, and overall image resolution is improved. Particularly from Fig. 9(f) to Fig. 9(f1), the structure of fine vessels is restored. Figure 9(a3–f3, a4–f4) show that compared to the input error maps, the output error maps have significant changes in fine vessels, and overall brightness in the output error maps is lower than in the input error maps. In Fig. 9, the arrows clearly highlight that the vascular structures in the input image are difficult to observe. However, after PEDL reconstruction, the vascular structures are effectively restored. Using the 100% power image as a reference, the restored vascular structures are verified to be reasonable. In Fig. 9(a5–f5), the images display histograms, where the *x*-axis represents intensity levels and the *y*-axis indicates the number of pixels corresponding to each value. Figure 9(a5–f5) show the histogram statistics of different datasets. Specifically, Fig. 9(a5–c5) and 9(d5–f5) represent the histograms of two different datasets. The low-energy region primarily represents background noise, while the high-energy region

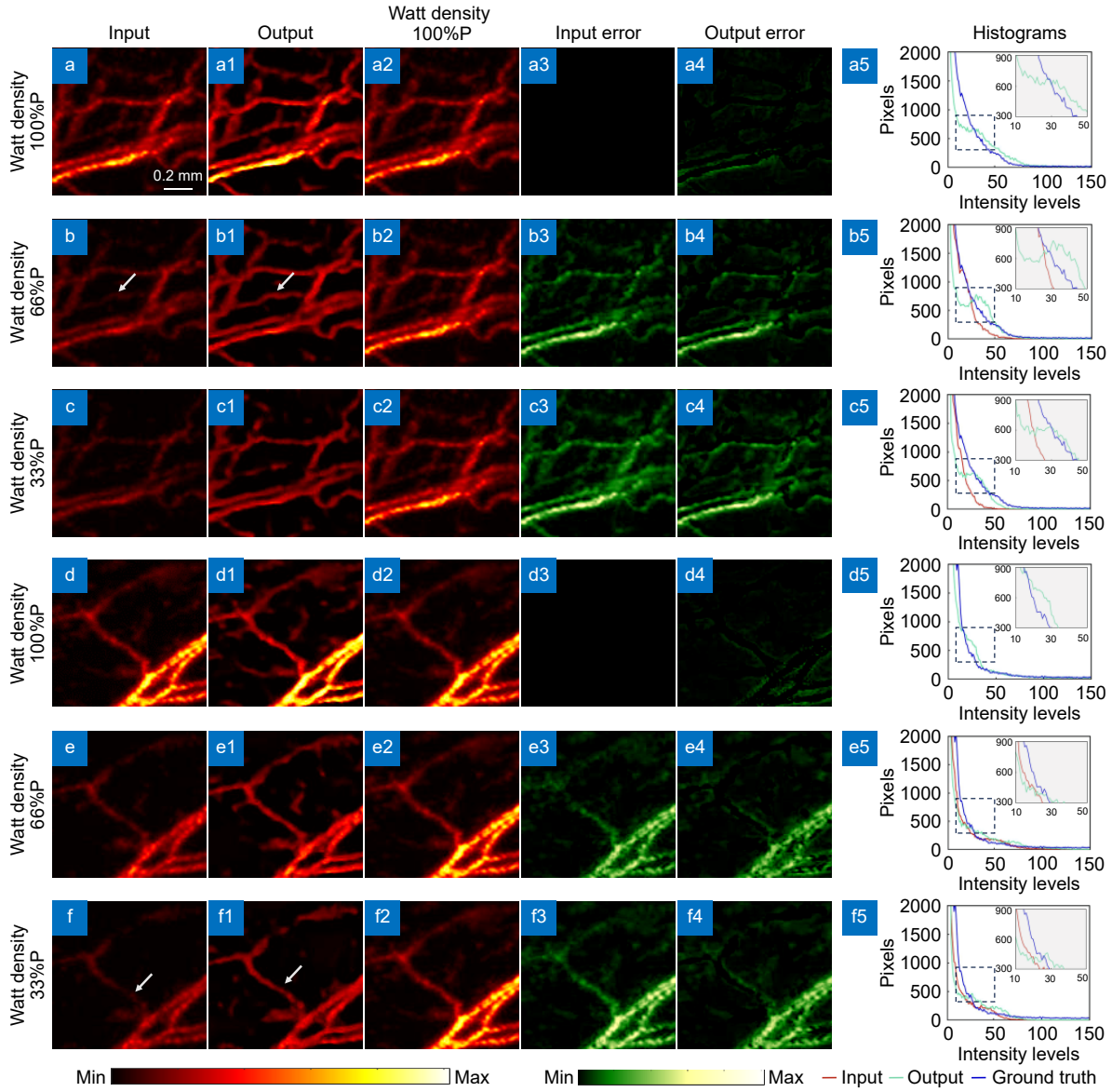

Fig. 9 | Effect diagram and reconstruction diagram under different laser energy. (**a–f**) Photoacoustic microscopy images at 100%, 66%, and 33% power, respectively. (a1–f1, a2–f2) Correspond to the network output images of the input images and the photoacoustic images obtained at 100% power. (a3–f3, a4–f4) Correspond to the error maps of the input and output images compared to the images obtained at 100% power. (a5–f5) Histograms of the input images, output images, and images obtained at 100% power.

reflects vessel signals. The zoomed-in area shows that the background noise is reduced in the low-energy region, and the vascular signals are enhanced in the high-energy region, indicating improved signal quality in specific energy ranges during processing. Analyzing these results demonstrates that the network can effectively restore image information at different energy levels, enhance weak signals, and effectively remove background noise. The effect of PEDL is not limited to the improvement of PAM at low energy, but also has obvious quality improvement for high power PAM imaging, and its adaptability has been fully reflected. To evaluate the denoising capability of PEDL on high-noise images, additional studies are conducted and presented in Fig. S3.

Mouse brain experiments

In order to study the output performance of the network under the conditions of deep absorption and scattering and over a large area, the mouse brain is imaged and analyzed by photoacoustic microscopy. The resulting structures are compared using cGAN, U-Net, and our network. The specific experimental results are shown in Fig. 10.

Figure 10(a–d) show the imaging experiments of mouse brain with different methods (cGAN, U-net, our network), cGAN and U-Net methods have limited enhancement of the mouse brain, where PEDL has a more significant effect on the recovery of fine blood vessels and resolution enhancement compared to these two methods. In particular, the resolution enhancement is more pronounced in the ROI indicated by the green arrows. In Fig. 10(c), the U-Net method enhances large blood vessels but has limited enhancement for small blood vessels, which might be due to the network's simple structure and the use of a simpler loss function, leading to unstable output. Figure 10(e) illustrates the vascular branch labeling of the input and output images using different methods. The output image of our network has more blood vessel branches, indicating the ability to restore more fine blood vessels. Figure 10(f) gives the error analysis of the vessel branch labeling. Figure 10(g) shows the projection analysis of the maximum intensity of the dashed line in the ROI area, from which it can be clearly observed that the full width at half maximum of the blood vessel decreases obviously after PEDL, which can effectively prove the improvement of resolution. Figure 10(h) shows the Contrast-to-Noise Ratio (CNR) and SNR of the three methods, and the weighted analysis of CNR and SNR, where the weighted curves give 50% of CNR and SNR. The formulas for CNR and SNR can be found in the supplementary materials under Parameter Definition. The input image processed by our network shows a significant improvement in CNR, SNR, and the weighted curves, outperforming the other methods. Figure 10(i) shows the input images and histogram analyses for the three methods. In the magnified region of the histogram, it can be observed that between 60 and 80, the output from PEDL is significantly higher than that of the other methods. In the overall chart, it can be seen that below 10, PEDL effectively suppresses background noise. It can be effectively verified that PEDL enhances weak signals while simultaneously eliminating background noise. To further investigate the authenticity of the vessels generated by PEDL, additional studies are conducted and are presented in Fig. S4. Based on the analysis of Fig. S4, it can be concluded that the vessels generated by PEDL exhibit high authenticity.

Discussion and conclusion

Discussion

The purpose of this work is to develop a physical model and a deep learning-based tool to enhance the quality of PAM imaging under various imaging conditions. To achieve this, the paper presents a mathematical derivation and constructs a synthetic dataset that incorporates diverse degeneration factors. Based on a physics-driven degeneration model, the PEDL network is applied to extract features from the synthetic dataset. Compared to other deep learning-based image reconstruction algorithms, the PEDL network integrates additional components, including residual blocks and GC attention modules, to enhance feature extraction and information processing. These components improve the model's capacity to restore vessel structures and enhance overall reconstruction performance across diverse conditions. The superiority of the PEDL network is validated through quantitative analysis in simulation and phantom experiments.

Specifically, the image simulation results under different noise environments have clearly shown that the network can effectively remove various levels of noise and restore small signals covered by noise, while also enhancing weaker signals. The image simulation results under different degeneration conditions show that the network can effectively restore images under various degeneration

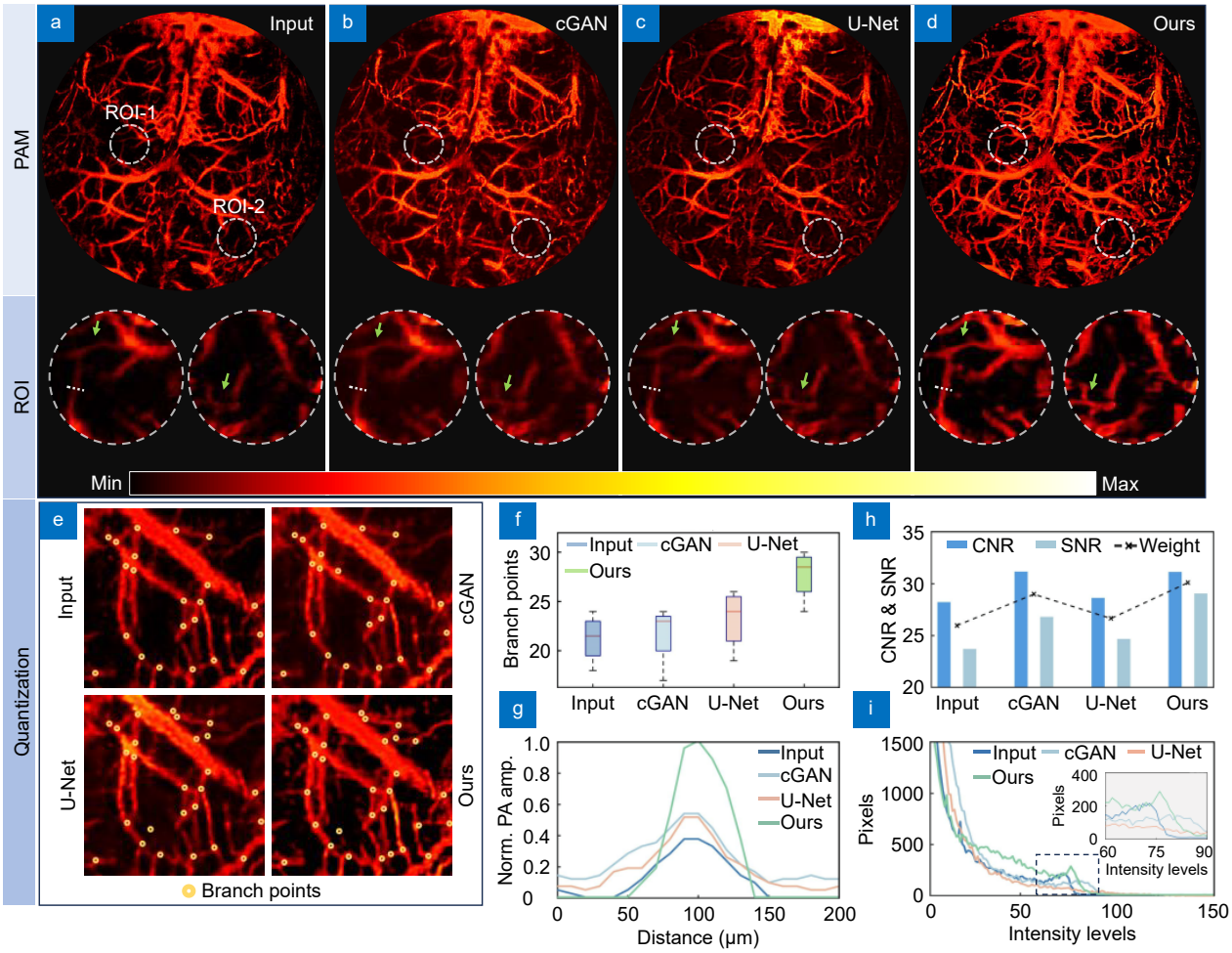

Fig. 10 | Different methods of brain reconstruction in mice. (**a**–**d**) Input images, photoacoustic microscopy images generated by cGAN, U-Net, and our network, respectively. (**e**) Blood vessel branch marking images of input and output images using cGAN, U-Net, and our network methods. (**f**) Error analysis of blood vessel branch markings. (**g**) Maximum intensity projection analysis of the dashed line in ROI area. (**h**) Comparison of Contrast-to-Noise Ratio (CNR), SNR, and their weights between input images and the three methods. (**i**) Histogram analysis of input images and the three methods.

conditions, significantly improving resolution. The performance of the network is also demonstrated through statistical measures such as PSNR and SSIM. Under strong noise conditions, both PSNR and SSIM are improved, achieving noise removal and signal recovery. Even under severe degeneration conditions, PSNR and SSIM can still be effectively enhanced. To further validate the network's ability to process real imaging data, this work conducted photoacoustic microscopy experiments at different depths and energy levels. In the depth experiments, translucent silicone sheets of various thicknesses are used for imaging. The results showed that the network significantly improved the resolution and contrast of blood vessels at different depths. In the energy experiments, the network effectively enhanced small signals and suppressed background noise. Finally, a mouse brain is imaged to explore the network's generalization capability and compared with other networks. The results demonstrated that the network maintained excellent performance in real photoacoustic microscopy imaging, significantly improving overall image resolution and contrast, and restoring and enhancing fine vascular structures. Additionally, the degeneration model based on physical mechanisms, combining mathematical derivation and practical data, has good generalizability and can be adjusted to different imaging conditions. However, this work has limitations because PAM imaging still requires a long time and cannot achieve real-time imaging.

The purpose of introducing the attention module in the network structure is to enhance the network's focus on key features, thereby improving its generalization

ability. Since embedding the physical model in the network causes the generated data to vary with each iteration, the attention module helps the network better adapt to these changes and respond effectively to different input conditions. This dynamic adjustment aligns closely with the imaging process in real-world scenarios, enabling the model to more accurately simulate the uncertainties encountered in practical applications. Additionally, the residual module is introduced in the network to capture deeper features. This design allows the network to retain and utilize more subtle feature layers when extracting information, further enhancing image quality and the resolution of complex structures. The residual module effectively mitigates the vanishing gradient problem in deep network training, allowing the network to maintain high learning efficiency even at greater depths. To evaluate the computational efficiency of the PEDL method, we compared it with cGAN and U-Net in reconstructing 100 images of 400×400 pixels under the same hardware conditions. The results show that the PEDL method's runtime is 10.930 seconds, with a peak memory usage of 1.46 MB. In contrast, cGAN's runtime is 8.050 seconds with 1.46 MB of memory usage, and U-Net's runtime is 6.467 seconds with the same memory usage of 1.46 MB. This indicates that the PEDL method has a certain disadvantage in computational efficiency, which may be due to the more complex structure and larger number of parameters in the PEDL model, thus increasing the computational load.

Under undersampling conditions, the decline in imaging resolution and the weakening of small vessel signals are common challenges. Our method can partially support imaging under undersampling conditions and mitigate the loss of information by enhancing resolution. However, due to the current limitations of the physical model, there is still room for improvement in integrating it with undersampling approaches. Additionally, when processing high-resolution PAM images or performing real-time imaging, the PEDL method may require significant computational resources, which could pose challenges in resource-limited settings such as portable devices or small laboratories. While PEDL performs well on specific tissue types, differences in optical and acoustic properties across various tissues may impact its applicability. Future research can be tested and popularized in various tissue types, and the function of this method can be increased to improve its popularization in clinical and research environments. What's more, integrating PEDL into existing PAM systems may encounter challenges with system compatibility and data processing adaptation, as different PAM systems have varying data formats and processing workflows, necessitating appropriate adjustments. The robustness and generalization ability of deep learning models depend on sufficient high-quality datasets. However, due to the difficulty of acquiring medical imaging data, future efforts could focus on improving imaging techniques to reduce data acquisition costs or on generating high-quality datasets using physical models and simulation techniques. At last, increasing dataset diversity and designing well-structured networks can further enhance the robustness and generalization ability of the model.

In the current PEDL method, the resolution enhancement effect is very apparent on simulated data, but for real-world data, especially images with more complex structures and noise, the performance is relatively poor. We believe the main reason for this difference is that, despite the degeneration process applied to the simulated data, there remains a significant gap between the simulated and real data. For example, the noise types and imaging conditions in real data are likely more complex, which reduces the effectiveness of our method in denoising and recovery on real-world images. Furthermore, while our method is capable of effective denoising, some smaller vessels still cannot be recovered after PEDL reconstruction, indicating that there is still room for improvement in the ability of PEDL to restore vessels post-denoising. In the future, we plan to further optimize the PEDL model to enhance its capacity for vessel structure restoration, especially in the presence of high noise and complex backgrounds. Additionally, we intend to extend the PEDL method to more application scenarios. Besides vessel structure restoration, we aim to incorporate the detection of biomarkers like sO2, enabling PEDL to handle a wider range of imaging tasks. Through these improvements, we hope to enhance the performance of PEDL on real-world data and increase its generalization ability across various imaging applications.

Conclusions

In the process of photoacoustic microscopy imaging, various interfering factors can lead to low-quality imaging results. This work combines physical models with deep learning, introducing various degeneration factors in the physical model to match real imaging conditions, and designs a network structure that deeply integrates

the physical model. To address the challenge of obtaining large datasets, a method of synthesizing datasets using physical models is employed. The network structure includes GC attention modules, residual blocks, BN layers, and composite loss functions. Extensive results demonstrate that the proposed model outperforms others.

By implementing these optimization techniques, the PEDL network significantly improves the resolution and contrast of photoacoustic microscopy imaging. In experiments conducted at different depths, the PEDL network adaptively optimizes data, enhancing resolution and reducing noise. Additionally, the PEDL network effectively enhances photoacoustic microscopy imaging data under different energy conditions, showing reliable recovery of fine blood vessels. The results show that the PEDL network is superior to other networks in improving the imaging effect in the experiment with mouse brain, and its generalization ability on different datasets is verified. Future work will focus on developing a faster PAM imaging system and a network structure compatible with the PAM system.

References

1. Wang LV. Multiscale photoacoustic microscopy and computed tomography. *Nat Photonics* **3**, 503–509 (2009).
2. Yao JJ, Wang LV. Photoacoustic tomography: fundamentals, advances and prospects. *Contrast Media Mol. Imaging* **6**, 332–345 (2011).
3. Yu YS, Feng T, Qiu HX et al. Simultaneous photoacoustic and ultrasound imaging: a review. *Ultrasonics* **139**, 107277 (2024).
4. Abdullah A, Shahini M, Pak A. An approach to design a high power piezoelectric ultrasonic transducer. *J Electroceram* **22**, 369–382 (2009).
5. Zeng YG, Da X, Wang Y et al. Photoacoustic and ultrasonic coimage with a linear transducer array. *Opt Lett* **29**, 1760–1762 (2004).
6. Xu MH, Wang LV. Universal back-projection algorithm for photoacoustic computed tomography. *Phys Rev E* **71**, 016706 (2005).
7. Beard P. Biomedical photoacoustic imaging. *Interface Focus* **1**, 602–631 (2011).
8. Wang LV, Gao L. Photoacoustic microscopy and computed tomography: from bench to bedside. *Annu Rev Biomed Eng* **16**, 155–185 (2014).
9. Yao JJ, Wang LV. Photoacoustic microscopy. *Laser Photonics Rev* **7**, 758–778 (2013).
10. Cao R, Li J, Ning B et al. Functional and oxygen-metabolic photoacoustic microscopy of the awake mouse brain. *Neuroimage* **150**, 77–87 (2017).
11. Wong TTW, Zhang RY, Hai PF et al. Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy. *Sci Adv* **3**, e1602168 (2017).
12. Lin L, Wang LV. The emerging role of photoacoustic imaging in clinical oncology. *Nat Rev Clin Oncol* **19**, 365–384 (2022).
13. Favazza CP, Wang LV, Jassim OW et al. *In vivo* photoacoustic microscopy of human cutaneous microvasculature and a nevus. *J Biomed Opt* **16**, 016015 (2011).
14. Zhu XY, Huang Q, DiSpirito A et al. Real-time whole-brain imaging of hemodynamics and oxygenation at micro-vessel resolution with ultrafast wide-field photoacoustic microscopy. *Light Sci Appl* **11**, 138 (2022).
15. Jeon S, Kim J, Lee D et al. Review on practical photoacoustic microscopy. *Photoacoustics* **15**, 100141 (2019).
16. Wu JH, Feng T, Chen Q et al. Photoacoustic guided wavefront shaping using digital micromirror devices. *Opt Laser Technol* **174**, 110570 (2024).
17. Chen JB, Zhang YC, He LY et al. Wide-field polygon-scanning photoacoustic microscopy of oxygen saturation at 1-MHz A-line rate. *Photoacoustics* **20**, 100195 (2020).
18. Jeon S, Park J, Managuli R et al. A novel 2-D synthetic aperture focusing technique for acoustic-resolution photoacoustic microscopy. *IEEE Trans Med Imaging* **38**, 250–260 (2019).
19. Kim C, Favazza C, Wang LV. In vivo photoacoustic tomography of chemicals: high-resolution functional and molecular optical imaging at new depths. *Chem Rev* **110**, 2756–2782 (2010).
20. Telenkov SA, Alwi R, Mandelis A. Photoacoustic correlation signal-to-noise ratio enhancement by coherent averaging and optical waveform optimization. *Rev Sci Instrum* **84**, 104907 (2013).
21. Sun MJ, Feng NZ, Shen Y et al. Photoacoustic signals denoising based on empirical mode decomposition and energy-window method. *Adv Adapt Data Anal* **4**, 1250004 (2012).
22. Haq IU, Nagoaka R, Makino T et al. 3D Gabor wavelet based vessel filtering of photoacoustic images. In *Proceedings of 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 3883–3886 (IEEE, 2016); http://doi.org/10.1109/EMBC.2016.7591576.
23. Shang RB, Archibald R, Gelb A et al. Sparsity-based photoacoustic image reconstruction with a linear array transducer and direct measurement of the forward model. *J Biomed Opt* **24**, 031015 (2018).
24. Deng ZL, Yang XQ, Gong H et al. Two-dimensional synthetic-aperture focusing technique in photoacoustic microscopy. *J Appl Phys* **109**, 104701 (2011).
25. Chen JH, Lin RQ, Wang HN et al. Blind-deconvolution optical-resolution photoacoustic microscopy *in vivo*. *Opt Express* **21**, 7316–7327 (2013).
26. Hill ER, Xia WF, Clarkson MJ et al. Identification and removal of laser-induced noise in photoacoustic imaging using singular value decomposition. *Biomed Opt Express* **8**, 68–77 (2016).
27. Yang CC, Lan HR, Gao F et al. Review of deep learning for photoacoustic imaging. *Photoacoustics* **21**, 100215 (2021).
28. Li ZS, Sun JS, Fan Y et al. Deep learning assisted variational Hilbert quantitative phase imaging. *Opto-Electron Sci* **2**, 220023 (2023).
29. Saba A, Gigli C, Ayoub AB et al. Physics-informed neural networks for diffraction tomography. *Adv Photonics* **4**, 066001 (2022).
30. Wu ZL, Kang I, Yao YD et al. Three-dimensional nanoscale re-

duced-angle ptycho-tomographic imaging with deep learning (RAPID). *eLight* **3**, 7 (2023).
31. Wei X, Feng T, Huang QH et al. Deep learning-powered biomedical photoacoustic imaging. *Neurocomputing* **573**, 127207 (2024).
32. Gröhl J, Schellenberg M, Dreher K et al. Deep learning for biomedical photoacoustic imaging: a review. *Photoacoustics* **22**, 100241 (2021).
33. Allman D, Reiter A, Bell MAL. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE Trans Med Imaging* **37**, 1464–1477 (2018).
34. Wang R, Zhang ZP, Chen RY et al. Noise‐insensitive defocused signal and resolution enhancement for optical‐resolution photoacoustic microscopy via deep learning. *J Biophotonics* **16**, e202300149 (2023).
35. Zhao HX, Ke ZW, Yang F et al. Deep learning enables superior photoacoustic imaging at ultralow laser dosages. *Adv Sci* **8**, 2003097 (2021).
36. Vu T, DiSpirito III A, Li DW et al. Deep image prior for undersampling high-speed photoacoustic microscopy. *Photoacoustics* **22**, 100266 (2021).
37. He D, Zhou JS, Shang XY et al. De-noising of photoacoustic microscopy images by deep learning. arXiv: 2201.04302 (2022). https://doi.org/10.48550/arXiv.2201.04302
38. Cheng SF, Zhou YY, Chen JB et al. High-resolution photoacoustic microscopy with deep penetration through learning. *Photoacoustics* **25**, 100314 (2022).
39. Zhang ZY, Jin HR, Zhang WW et al. Adaptive enhancement of acoustic resolution photoacoustic microscopy imaging *via* deep CNN prior. *Photoacoustics* **30**, 100484 (2023).
40. Gao Y, Feng T, Qiu HX et al. 4D spectral-spatial computational photoacoustic dermoscopy. *Photoacoustics* **34**, 100572 (2023).
41. DiSpirito A, Li DW, Vu T et al. Reconstructing undersampled photoacoustic microscopy images using deep learning. *IEEE Trans Med Imaging* **40**, 562–570 (2021).
42. Lu JY, Zou HH, Greenleaf JF. Biomedical ultrasound beam forming. *Ultrasound Med Biol* **20**, 403–428 (1994).
43. Lin HN, Cheng JX. Computational coherent Raman scattering imaging: breaking physical barriers by fusion of advanced instrumentation and data science. *eLight* **3**, 6 (2023).
44. Zhang ZY, Jin HR, Zheng ZS et al. Deep and domain transfer learning aided photoacoustic microscopy: acoustic resolution to optical resolution. *IEEE Trans Med Imaging* **41**, 3636–3648 (2022).
45. Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. *arXiv preprint* arXiv:1603.07285, 2016
46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 26th International Conference on Neural Information Processing Systems* 1097–1105 (Curran Associates Inc. , 2012).
47. He KM, Zhang XY, Ren SQ et al. Deep residual learning for image recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016); http://doi.org/10.1109/CVPR.2016.90.
48. Cao Y, Xu JR, Lin S et al. GCNet: non-local networks meet squeeze-excitation networks and beyond. In *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshops* 1971–1980 (IEEE, 2019); http://doi.org/10.1109/ICCVW.2019.00246.
49. Kim T, Oh J, Kim N et al. Comparing Kullback-Leibler divergence and mean squared error loss in knowledge distillation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence* 2628–2635 (ijcai. org, 2021).
50. Yang QS, Yan PK, Zhang YB et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging* **37**, 1348–1357 (2018).
51. Wang ZY, Zhou YF, Hu S. Sparse coding-enabled low-fluence multi-parametric photoacoustic microscopy. *IEEE Trans Med Imaging* **41**, 805–814 (2022).
52. Zhang YC, Chen JB, Zhang J et al. Super‐low‐dose functional and molecular photoacoustic microscopy. *Adv Sci* **10**, 2302486 (2023).
53. Siregar S, Nagaoka R, Haq IU et al. Non local means denoising in photoacoustic imaging. *Jpn J Appl Phys* **57**, 07LB06 (2018).
54. Frangakis AS, Hegerl R. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *J Struct Biol* **135**, 239–250 (2001).
55. Gauthier J. Conditional generative adversarial nets forconvolutional face generation. Class project for Stanford CS231N: convolutional neural networks for visual recognition. *WinterSemester* **2014**, 2 (2014).
56. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, 2015); http://doi.org/10.1007/978-3-319-24574-4_28.
57. Ma HG, Yu YS, Zhu YH et al. Monitoring of microvascular calcification by time-resolved photoacoustic microscopy. *Photoacoustics* **41**, 100664 (2025).

Acknowledgements

This work was supported by National Natural Science Foundation of China (62227818, 12204239, 62275121), Youth Foundation of Jiangsu Province (BK20220946), Fundamental Research Funds for the Central Universities (30923011024), and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (JSGP202201).

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at https://doi.org/10.29026/oea.2025.240189