Generative priors-constraint accelerated iterative reconstruction for extremely sparse photoacoustic tomography boosted by mean-reverting diffusion model: Towards 8 projections

Teng Lian [b,1], Yichen Lv [a,b,1], Kangjun Guo [a], Zilong Li [a], Jiahong Li [a], Guijun Wang [a], Jiabin Lin [a], Yiyang Cao [a], Qiegen Liu [a], Xianlin Song [a,c,d,*]

[a] School of Information Engineering, Nanchang University, Nanchang 330031, China
[b] Jiluan Academy, Nanchang University, Nanchang 330031, China
[c] Jiangxi Provincial Key Laboratory of Advanced Signal Processing and Intelligent Communications, Nanchang University, Nanchang 330031, China
[d] Jiangxi Provincial Engineering Research Center for Intelligent Medical Information Detection and Internet of Things, Nanchang University, Nanchang 330031, China

ABSTRACT

As a novel non-invasive hybrid biomedical imaging technology, photoacoustic tomography combines the advantages of high contrast of optical imaging and high penetration of acoustic imaging. However, the conventional standard reconstruction methods under sparse view may lead to low-quality image in photoacoustic tomography. To address this problem, an advanced sparse reconstruction method for photoacoustic tomography based on the mean-reverting diffusion model is proposed. By modeling the degradation process from a high-quality image under full-view scanning (512 projections) to a sparse image with stable Gaussian noise (i.e., mean state), a mean-reverting diffusion model is trained to learn prior information of the data distribution. Then the learned prior information is employed to generate a high-quality image from the sparse image by iteratively sampling the noisy state. Blood vessels simulation data and the animal *in vivo* experimental data were used to evaluate the performance of the proposed method. The results demonstrate that the proposed method achieves higher-quality sparse reconstruction compared with conventional reconstruction methods and U-Net method. In addition, the proposed method dramatically speeds up the sparse reconstruction and achieves better reconstruction results for extremely sparse images compared with the method based on conventional diffusion model. The proposed method achieves an improvement of 0.52 (~289 %) in structural similarity and 10.01 dB (~59 %) in peak signal-to-noise ratio for extremely sparse projections (8 projections), compared with the conventional delay-and-sum method. This method is expected to shorten the acquisition time and reduce the cost of photoacoustic tomography, thus further expanding the range of applications.

1. Introduction

Photoacoustic tomography (PAT) is a non-invasive hybrid biomedical imaging technique based on the photoacoustic effect, which is capable of providing high-resolution structural and functional information of tissue as well as the depth-penetration capability of optical imaging [1–4]. PAT has been widely used in medical imaging for tissue imaging [5,6], tumor diagnosis [7], vascular imaging [8], and image-guided surgery [9,10], and has gradually become an important tool in clinical experiments and clinical research [11–13]. In PAT, a

nanosecond-level pulsed laser beam is used in biological tissues, and absorbers located in the tissue body (e.g., tumors) absorb the pulsed light energy, which causes an instantaneous temperature rise and expansion, generating photoacoustic waves. At this time, ultrasonic transducer located on the surface of the tissue body can receive these outgoing photoacoustic waves and reconstruct an image of the light energy absorption distribution within the tissue based on the detected photoacoustic signals.

Conventional reconstruction methods are based on analytic algorithms, such as filtered back-projection, delay-and-sum (DAS) and time-

reversal methods [14–16]. However, although the above methods have wide applications in medical imaging, they still have some limitations. In an actual imaging system, conventional reconstruction methods are limited in resolution, resulting in the inability to accurately display fine structures. There are also problems such as sensitivity to data noise, excessive computational complexity, and severe artifacts. Therefore, high-quality and fast reconstruction under sparse sampling is a challenging task to be solved.

In response to the challenges faced by conventional photoacoustic tomography reconstruction methods, a series of hardware-based and algorithm-based improvements have been proposed to enhance imaging quality and efficiency. From the hardware perspective, the usual approach is to optimize the detector design, e.g., using ultrasonic transducer arrays to capture the photoacoustic signals from more views [17,18], thus improving the reconstruction quality. However, those hardware methods may increase system complexity and cost, thus limiting the feasibility of their widespread application. In contrast, improved algorithms may be a more feasible and economical solution to achieve high-quality reconstruction. Paltauf et al. designed iterative reconstruction algorithms to obtain better reconstruction through a model-based method [19]. Arridge et al. proposed compressed sensing reconstruction algorithms, which achieves better reconstruction by changing the method of signal acquisition and recovery [18,20]. In addition to this, variational methods are also common among model-based methods, which significantly improves the quality and usefulness of reconstructed images by solving an optimization problem with a regularization term and utilizing prior information of the image to guide the reconstruction [21–24]. However, these methods require complex data integration and accurate prior information, as well as extensive trial and tuning work, which are difficult to achieve in experiments.

In recent years, deep learning has been widely used in biomedical image processing, including photoacoustic tomographic image reconstruction, due to its powerful feature extraction and learning capabilities [25–28]. Madasamy et al. found that deep learning-based methods can compensate for the nonlinear luminous flux distributions more effectively and efficiently by comparing various network architecture [29]. Currently, most post-processing methods are based on U-Net network to eliminate artifacts. Chen et al. proposed a U-Net based method for photoacoustic imaging [30], and Davoudi et al. utilized the U-Net network to reduce artifacts in photoacoustic images and enhance anatomical contrast and image quantization [31]. Shahid et al. combined inverse compressed sensing with ResU-Net network to process photoacoustic data and restore high-quality images [32]. Guan et al. proposed a fully dense U-Net (FD-UNet) network for artifact removal [33]. Deng et al. designed the SE-UNet network architecture, which demonstrated better performance in artifact removal [34]. However, due to the need of learning unknown mappings, deep learning-based methods usually require a large amount of data and a large-scale network for training and testing, and they remain less practically effective than the conventional modified iterative methods.

With the development of generative models, a variety of generative models, including generative adversarial networks (GAN) [35], variational autoencoder (VAE) [36], denoising diffusion probabilistic model (DDPM) [37], and score-based generative model [38], have shown great advantages in the field of image processing and reconstruction. Among them, in the field of image restoration, the score-based generative model significantly improves the performance of the model by applying an efficient and stable sampling method, which can more effectively reduce the noise and errors when dealing with a variety of image data, thus achieving higher-quality restoration. On this basis, Song et al. further proposed a diffusion model based on stochastic differential equations (SDE). In this diffusion model, Stochastic differential equations are used in imaging to model the dynamics of image degradation and reconstruction by mathematically modeling the evolution of systems containing random noise. This SDE-based diffusion model can more

effectively handle complex image restoration with strong generative capabilities [39]. In the forward process, the SDE-based diffusion models gradually introduce Gaussian noise into the image until it is converted to a purely noisy image, and in the reverse process, it simulates a reverse SDE to reduce the purely noisy image to a high-quality image; however, modeling the diffusion process using conventional SDEs requires converting the original data distribution to a purely noisy prior distribution. This often requires more iterations and computational resources to obtain high-quality samples. Inspired by this, this paper proposes a novel reconstruction strategy based on the mean-reverting diffusion model, which significantly accelerates the reconstruction time as well as improves the reconstruction quality. The model used is a particular diffusion model based on mean-reverting stochastic differential equation (IR-SDE), which converts a high-quality image into a degraded counterpart with a fixed Gaussian noise in the forward process of reconstruction, and realizes the reconstruction of a high-quality image from noisy data without need of specific a prior information by simulating the corresponding reverse-time SDE [40]. The proposed mean-reverting SDE has a closed-form solution, allowing the computation of the ground truth time-dependent score and learning it with a neural network during the training process. In addition, the model uses maximum likelihood objective as the loss function and the learned prior information as the distributional constraints to guide the reconstruction process, and trains the neural network to find the optimal sparse image reconstruction trajectory. Compared to other methods, the proposed method shows significant improvement in reconstruction speed and image quality.

By applying the mean-reverting diffusion model to the sparse reconstruction of PAT, this study not only significantly improves the quality and speed of image reconstruction, but also verifies its superiority in both experimental data and practical applications. The successful realization of this model predicts that in future PAT applications, even in data-constrained situations, high-quality images required for clinical purposes can be efficiently reconstructed, thus providing more accurate diagnostic information for clinical medicine and promoting the further development and application of photoacoustic tomography in biomedical fields.

2. Principles and methods

2.1. The principle of photoacoustic tomography

In PAT, when a nanosecond-level pulsed laser beam is used to irradiate biological tissue, light absorbers (hemoglobin, melanin, etc.) within the biological tissue absorb the pulsed light energy and convert it to heat, and the thermal expansion and contraction of the absorber makes it a source of sound. Since soft tissue is a good medium for acoustic wave propagation, the photoacoustic signal at the absorber can effectively radiate to the surrounding medium and propagate in the tissue with low scattering and low loss. The ultrasonic transducer located around the tissue acquires the generated photoacoustic waves, and through signal processing and photoacoustic image reconstruction, a photoacoustic image reflecting the internal structure and function of the tissue can be formed. In the case of satisfying the thermal and pressure constraints, the initial acoustic pressure can be expressed by Eq. (1):

$$P_0 = \Gamma \eta_{th} \mu_a(r) F \tag{1}$$

where Γ is the Gruneisen coefficient, η_{th} represents the thermal conversion efficiency, μ_a is the absorption coefficient, and F represents the optical flux. The expression for the Gruneisen coefficient is expressed as Eq. (2):

$$\Gamma = \frac{\beta v_s^2}{C_P} \tag{2}$$

where, β is the coefficient of thermos-elastic expansion, v_s represents the speed of sound propagation in the tissue, and C_P is the heat capacity of the tissue at constant pressure. The initial sound pressure P_0 determines the value of the sound pressure at a certain moment and at a certain point, and the sound pressure at any moment and position is shown in Eq. (3):

$$\left(\nabla^2 - \frac{1}{v_S^2}\frac{\partial^2}{\partial t^2}\right)P\left(r, t\right) = -\frac{\beta}{k v_S^2}\frac{\partial^2 T(r, t)}{\partial t^2} \tag{3}$$

where T is the elevated temperature, $P(r, t)$ is the value of the sound pressure at point r and at moment t. The left side of the equal sign describes wave propagation, and the right side is the photoacoustic source term. Under the condition of satisfying thermal constraints, the heat equation can be expressed as Eq. (4):

$$\rho C_P \frac{\partial T(r, t)}{\partial t} = H(r, t) \tag{4}$$

where $H(r, t)$ is a heating function expressing the thermal energy conversion per unit time and per unit volume, and its relationship with the optical flux rate Φ can be expressed as $H = \eta_{th}\mu_a\Phi$. By bringing Eq. (4) into Eq. (3), the photoacoustic equation can be transformed as Eq. (5):

$$\left(\nabla^2 - \frac{1}{v_S^2}\frac{\partial^2}{\partial t^2}\right)P\left(r, t\right) = -\frac{\beta}{C_P}\frac{\partial H}{\partial t} \tag{5}$$

Eq. (5) represents the photoacoustic signal detected by the ultrasonic transducer placed on the surface of the biological tissue at any time and at any position after laser irradiation of the biological tissue. When an ideal laser beam is used to irradiate a light absorber with uniform acoustic properties, the amplitude of the photoacoustic signal generated by the absorber is proportional to the amplitude of the ideal pulsed laser, and the characteristics of the optical signal at this time are determined by the distribution of light absorption in the absorber. Therefore, it is possible to calculate the light absorption distribution of the absorber inside the biological tissue inversely based on the original photoacoustic signals detected by the ultrasonic transducer at each position, which is the reconstruction of the photoacoustic tomography.

2.2. Mean-reverting diffusion model

In order to solve the problem of low imaging quality under sparse projections in photoacoustic tomography, a diffusion model based on mean-reverting SDE is used in this study. The model converts a high-quality image into a degraded version with a fixed Gaussian noise in the forward process. And the high-quality image is then restored by simulating the corresponding reverse-time SDE in the reverse process. Song et al. pioneered the idea of using SDEs in diffusion modeling to simulate diffusion processes [39]. In forward SDE, Gaussian noise is added to perturb the data distribution to obtain the prior distribution. And in the reverse SDE, the prior distribution is transformed into data by slowly removing the noise to sample spare-view image from the data distribution. Compared with the conventional diffusion model, SDE-based diffusion models can model the entire diffusion process in the continuous time domain, which allows the models to simulate nonlinear or non-decomposable degradation processes. Meanwhile, since SDE simulates a continuous Gaussian process (Brownian motion), it makes the process of reconstructing the image smoother and more natural and maintains better detail fidelity. Assuming that $\{x(t)\}_{t=0}^{T}$ is a continuous diffusion process, such that p_0 denotes the initial distribution of the data and $t \in [0, T]$ denotes the continuous time variable, the forward diffusion of this diffusion process can be described by Eq. (6):

$$dx = f(x, t)\,dt + g(t)dw, \; x(0) \sim p_0(x) \tag{6}$$

where f and g are the drift and dispersion functions, respectively, w is a standard Wiener process, and $x(0) \in R^d$ is the initial condition. Usu-

ally, the end state $x(T)$ obeys a Gaussian distribution with fixed mean and variance. The designed SDE can gradually convert the data distribution to a fixed Gaussian noise. The process of sampling data from noise is then flipped by simulating the SDE backwards, and the reverse process of forward diffusion is also a diffusion process, i.e., reverse diffusion, described by Eq. (7):

$$dx = [f(x, t) - g(t)^2 \nabla_x \log p_t(x)]dt + g(t)d\widehat{w} \tag{7}$$

where $x(T) \sim p_T(x)$. $\widehat{w}$ is a reverse-time Wiener process, d_t denotes an infinitesimal negative time step, and $p_t(x)$ represents the marginal probability density function of $x(t)$ at time t. The score function $\nabla_x \log p_t(x)$ of the conventional SDE diffusion model is difficult to handle, and we usually train the relevant neural network $S_\theta(x, t)$ to approximate it under the score matching objective.

However, modeling the diffusion process using traditional SDEs requires converting the original data distribution to a purely noisy prior distribution. This often requires more iterations and computational resources to obtain high quality samples. On this basis, Luo et al. proposed image restoration with mean-reverting SDE (IR-SDE). Compared to the conventional SDE diffusion model, IR-SDE introduces parameters in the forward diffusion process and constructs a special case of the SDE, which makes the score function $\nabla_x \log p_t(x)$ analytically easy to deal with, and the new forward process is shown in Eq. (8):

$$dx = \theta_t(\mu - x)\,dt + \sigma_t\,dw \tag{8}$$

where μ denotes the state mean, θ_t and σ_t are time-dependent positive parameters that denote the mean reversion speed and stochastic volatility, respectively. There is a large degree of freedom in the choice of θ_t and σ_t. The choice can have a significant impact on the final recovery performance. μ and $x(0)$ are set to be a pair of high and low quality images. In order to make Eq. (8) have a closed-form solution, set $\frac{\sigma_t^2}{\theta_t} = 2\lambda^2$, where λ^2 is the stationary variance, then for any starting state $x(t)$, the solution of the SDE is shown in Eq. (9):

$$x(t) = \mu + (x(s) - \mu)e^{-\overline{\theta}_{s:t}} + \int_s^t \sigma_z e^{-\overline{\theta}_{z:t}}\,dw\,(z) \tag{9}$$

where $\overline{\theta}_{s:t} := \int_S^t \theta_t dz$ is known and the transition kernel $P(x(t)\,|x(s)\,) = N(x(t)\,|m_{s:t}(x(s)), v_{s:t}\,)$ is a Gaussian part with mean $m_{s:t}$ and variance $v_{s:t}$. Simplification from the above equation yields a distribution conditional on the initial state $x(0)$ at any time t, as shown in Eq. (10):

$$\begin{cases} p_t(x) = N(x(t)|m_t(x), v_t) \\ m_t(x) := \mu + (x(0) - \mu)e^{-\overline{\theta}_t} \\ v_t := \lambda^2(1 - e^{-2\overline{\theta}_t}) \end{cases} \tag{10}$$

when $t\to\infty$, the mean m_t converges to the low-quality image μ and the variance v_t converges to the stationary variance λ^2. In essence, the forward process of IR-SDE is to diffuse a high-quality image into a low-quality image with fixed Gaussian noise. In order to reconstruct the high-quality image from the end state $x(T)$, the reverse SDE for image restoration can be obtained by inverting Eq. (7) according to Eq. (8), The forward and reverse iterative reconstruction process is shown in Fig. 1:

$$dx = [\theta_t(\mu - x) - \sigma_t^2 \nabla_x \log p_t(x)]dt + \sigma_t d\widehat{w} \tag{11}$$

Eventually we can obtain the forward and reverse SDE of the mean-reverting diffusion model. The only unknown part of the SDE of the above process is the score function $\nabla_x \log p_t(x)$ of the marginal distribution at time t. However, during the training period, the high-quality image $x(0)$ is known, and therefore the network can be trained to estimate the conditional score function $\nabla_x \log p_t(x|x(0))$. Combined with the Eq. (10), the conditional score can be computed as shown in Eq. (12):

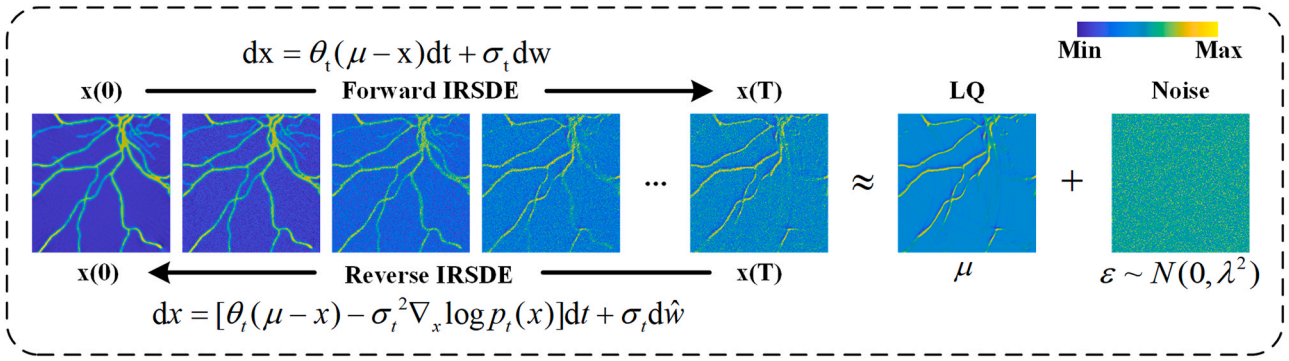$$\nabla_x \log p_t(x|x(0)) = -\frac{x(t) - m_t(x)}{v_t} \tag{12}$$

Fig. 1. Forward and reverse processes of IR-SDE.

when confronted with the sparse-view image, the training process usually becomes unstable, resulting in poor reconstruction results, this problem mainly stems from trying to learn the instantaneous noise at a given time. To solve this problem, the mean-reverting diffusion model adopts a maximum likelihood objective to allow the model to estimate along the optimal image reconstruction direction ($x_{1:T}$), which in turn leads to stabilized training and reconstruction of more accurate images. Specifically, this objective considers the entire trajectory from x_T to x_0, instead of a particular time step, and the model is guided towards the optimal path at each time step, avoiding local optimality and inconsistency in training. Estimating the optimal reconstruction path can be transformed into an optimization problem based on the maximum likelihood objective, i.e., maximizing the likelihood, which can be expressed as Eq. (13):

$$p(x_{1:T}|x_0) = p(x_T|x_0) \prod_{i=2}^{T} p(x_{i-1}|x_i, x_0) \qquad (13)$$

where $p(x_T|x_0) = N(x_T; m_T(x_0), v_T)$ is the distribution of the low-quality image. It can be derived from Bayes' theorem and Eq. (13), as described in Eq. (14):

$$p(x_{i-1}|x_i, x_0) = \frac{p(x_i|x_{i-1}, x_0)p(x_{i-1}|x_0)}{p(x_i|x_0)} \qquad (14)$$

Since all distributions are Gaussian distributions that can be computed according to the principles of the conventional diffusion model, it is natural to directly find the best reverse state that minimizes the negative log-likelihood, as shown in Eq. (15):

$$x_{i-1}^* = \mathrm{argmin}_{x_{i-1}}[-\log p(x_{i-1}|x_i, x_0)] \qquad (15)$$

where x_{i-1}^* denotes the ideal state reversed from x_i, i.e., the optimal reduction path. To simplify the notation, let $\theta_t' := \int_{i-1}^{i} \theta_t dt$. By solving the above task, the optimal inverse solution x_{i-1}^* of the IR-SDE can be obtained for any state x_i at discrete time $i > 0$ given the initial state x_0:

$$x_{i-1}^* = \frac{1-e^{-2\bar\theta_{i-1}}}{1-e^{-2\bar\theta_i}}e^{-\theta_i'}(x_i-\mu) + \frac{1-e^{-2\theta_i'}}{1-e^{-2\bar\theta_i}}e^{-\bar\theta_{i-1}}(x_0-\mu)+\mu \qquad (16)$$

After that, based on the possibility of maximizing the retrograde trajectory, we train the neural network $\widetilde{\varepsilon_\varnothing}(x(t), u, t)$ using an alternative loss function. The specific loss function is described in Eq. (17):

$$L_\gamma(\phi) := \sum_{i=1}^{T} \gamma_i E[||x_i - (dx_i)_{\tilde\varepsilon_\phi} - x_{i-1}^*||] \qquad (17)$$

where $\gamma_1, \cdots, \gamma_T$ are the positive weight and $\{x_i\}_i^T = 0$ denotes the discretization of the diffusion process. This loss function is shown to stabilize the training and consistently improve the image restoration performance compared to a common score matching objective. After training, the final reconstructed image can be obtained by sampling the

noise state x_T using network $\widetilde{\varepsilon_\varnothing}$ and solving the reverse SDE iteratively using a numerical scheme. The optimization objective of this loss function is to maximize the likelihood probability of the reverse trajectory, which reduces fluctuations in training, stabilizes training and improves recovery by explicitly modeling the optimal path for state transfer. During training, With the goal of convergence, the maximum likelihood loss is achieved by forcing the reverse trajectory $x_T \rightarrow x_0$ so that the predicted x_{i-1} is gradually approximated by x_{i-1}^*, and the above process is repeated until the loss function converges significantly.

2.3. Sparse-view reconstruction for photoacoustic tomography based on diffusion model

In this study, Sparse-view reconstruction for PAT is realized using the IR-SDE diffusion model, which converts a high-quality image into a degraded version with a fixed Gaussian noise in the forward process, and reconstructs a high-quality image by simulating the corresponding reverse-time SDE in the reverse process.

In the training phase, a pair of low-quality and high-quality images is fed into the network, and the IR-SDE used by the model adjusts the forward process so that it models the image degradation itself, and thus the simulation of the degradation process can be achieved without the need to know any prior information of the degradation or the relevant parameters. In the reconstruction phase, the gradual removal of noise from the intermediate mean state and thus the reconstruction of a high-quality image is achieved by modeling the reverse diffusion process. In this reverse process, the reverse-time SDE is solved using numerical methods, usually the Euler-Maruyama method or other higher order methods such as the Milstein method [41,42]. The core step of the Euler-Maruyama method solution is to first divide continuous time intervals into discrete small-time steps, and then utilize the current estimated state at each time step and the gradient predicted by the model to update, and finally repeating the process until the original time point is reached. The specific training and reconstruction process are shown in Fig. 2(a).

In this process, the NAFNet network architecture plays a crucial role, which omits the linear activation layer and is a nonlinear activation network [43]. The mean-reverting diffusion model employs NAFNet instead of U-Net used in the conventional diffusion models as the backbone network for prediction noise, which enables more stable training and reconstruction, and ultimately achieves higher image reconstruction quality and computational efficiency. In the mean-reverting diffusion model, the simplified NAFNet structure is shown in Fig. 2(b), where some details are purposely omitted for simplicity, including the downsampling and upsampling layers, the feature fusion module, the inputs/outputs, etc. NAFNet adopts an innovative design concept called "nonlinear free activation", introducing "SimpleGate" instead of the conventional activation function. "SimpleGate" is an element-wise operation that divides the feature channel into two parts. This design reduces the computational burden
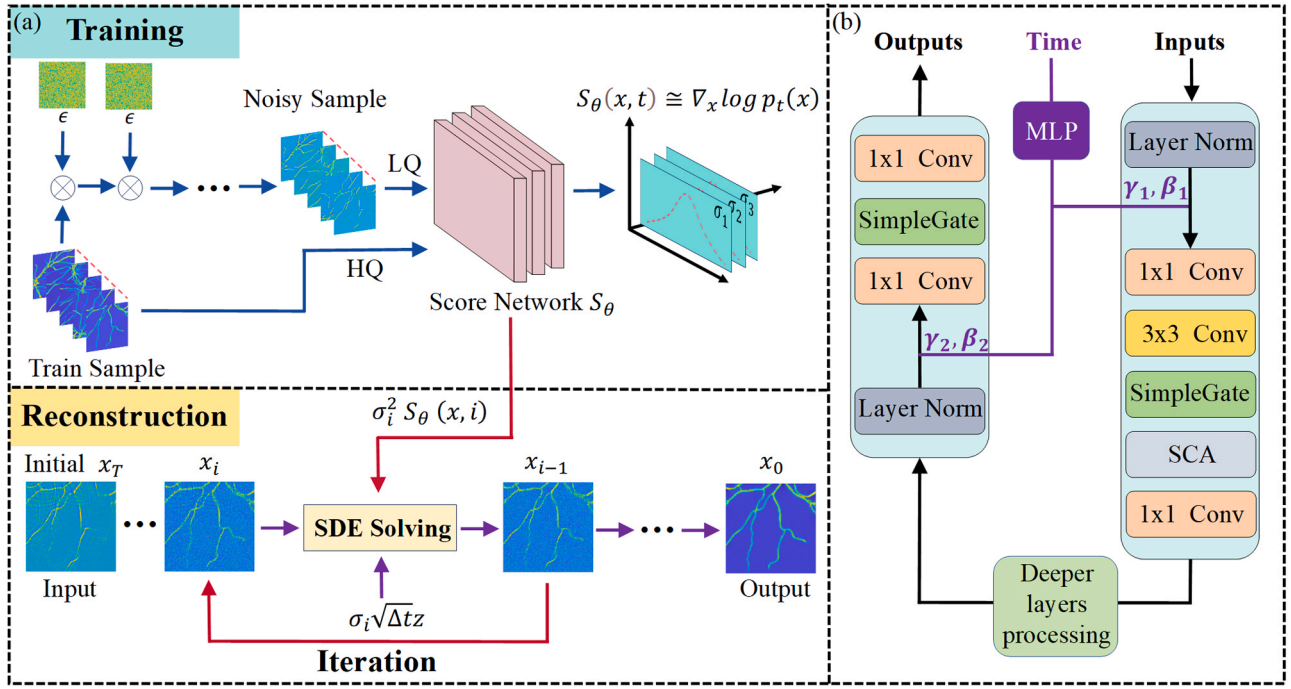
Fig. 2. Network Architecture and the Process of Training and Reconstruction. (a)Schematic diagram of the mean-reverting diffusion model during training and reconstruction. (b) Architecture of the ConditionalNAFNet. MLP, multi-layer perceptron; Conv, convolution; SCA, simplified channel attention; Layer Norm, layer normalization.

and enables high-speed computation, especially in hardware resource-constrained environments. In addition, without the extra parameters and computational overhead of the activation layer, NAFNet can be more easily scaled to deeper network levels without unduly increasing the computational burden. This allows the model to capture more complex features and perform better in handling highly complex image reconstruction [44]. In order to further optimize the quality of image reconstruction, the mean-reverting diffusion model used in this study employs an optimized NAFNet called ConditionalNAFNet. This special network introduces an improved nonlinear activation-free block on top of NAFNet thereby increasing the capability of conditional generation. As shown in Fig. 2(b), a multi-layer perceptron (MLP) is added to the network before the attention and feedforward layers. The design maps the time-step embedding of the diffusion process into channel scaling factor (α) and offset factor (β) to dynamically adjust the strength of the feature response. This design enables the network to adapt to different degradation stages (e.g., early deblurring and late denoising) and significantly improves the iterative stability. Specifically, the time-conditional dynamic modulation enhances the network's ability to perceive the degradation process by interacting the time-step information with the feature channel, thus exhibiting stronger robustness in sparse-view reconstruction. In addition, this improvement not only further enhances the adaptability of the model to temporal dynamics,

but also enables the model to provide more accurate outputs in image restoration, and maintains high performance and flexibility in image reconstruction under different conditions.

2.4. Simulation dataset acquisition

In order to obtain a sufficient number of photoacoustic images under full-view sampling, a virtual PAT is constructed based on k-Wave [45]. The overall flowchart for generating simulation datasets using the virtual PAT is shown in Fig. 3. The platform realizes the function of reconstructing the image under arbitrary sparse projections based on the full-view image. The size of the whole computational area set to 50 mm × 50 mm with a grid of 440 × 440 pixels. The ultrasonic transducers with a BW of 66 % and a center frequency set to 2.25 MHz are placed equidistantly at different viewpoints (views) in a circle with a radius of 21.6 mm, so that the images with different sparse projections can be reconstructed by reducing the number of ultrasonic transducers. The speed of sound is set to 1500 m/s and the surrounding medium is water with a density of 1000 kg/m³ .

2.5. Source of datasets

The datasets include simulation dataset and experimental dataset.
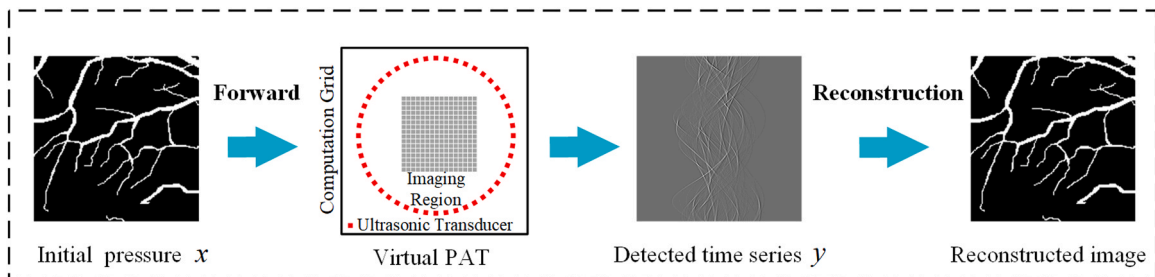
Fig. 3. Overall flow chart for the generation of simulation dataset using a virtual PAT.

The ground truth of the simulation dataset is obtained from the publicly available retinal vascularization datasets RAVIR and DRIVE [46], and 1500 images are finally obtained by cropping and transforming after selecting representative images. And then by constructing a virtual PAT based on k-Wave, a sufficient number of photoacoustic under full-view sampling is obtained (8, 16, 32, 64 and 128 projections). The training set consists of 1200 images and the test set consists of 300 images. The experimental dataset consists of a phantom dataset and animal *in vivo* data [31]. The phantom dataset contains 496 images under full-view scanning (512 projections) and an equal number of images under other sparse projections (8, 16, 32, 64, and 128 projections). The *in vivo* experimental data abdomen dataset contains 274 images under full-view scanning (512 projections) and an equal number of images under other sparse projections (8, 16, 32, 64, and 128 projections).

2.6. Network parameter settings

In the training stage, the proposed method uses the Lion optimizer combined with CosineAnnealingLR, the optimizer's exponential decay rate of the first-order moments is set to 0.9, the exponential decay rate of the second-order moments is 0.99, and the initial learning rate is 4×10^{-5}. The noise intensity in the SDE is fixed at 50, with the batch size of 4 and the image size of 256×256. A total of 400,000 iterations are performed, and the learning rate is self-multiplied by 0.5 (not less than 1×10^{-7}) after every 100,000 iterations. In the reconstruction stage, the image size is 256×256, and the number of iterations is 100 times. The proposed method is implemented based on the PyTorch framework, which mainly runs in the Python environment. A total of 6000 epochs were trained in the training phase, one checkpoint was saved for every 100 epochs, and the time spent for a single checkpoint was about 60 minutes, and an average model was saved synchronously after the final training was completed, from which the best training model was selected. In this study, the computational tasks were performed in a graphics processing unit (GPU; NVIDIA RTX A5000) with 24 GB of memory.

3. Results

3.1. Blood vessels simulation data

In the test phase, the simulated dataset was imported into the constructed virtual PAT and the photoacoustic signals from different sparse views are used as the input of the network. The reconstruction results of the conventional DAS method, the U-Net method, the VE-SDE method and the proposed method were compared. Fig. 4(a) shows the reconstruction of simulated blood vessels using the proposed method of this paper under 128, 64, 32, 16 and 8 projections, respectively. Among these methods, the VE-SDE method is a sparse-view reconstruction method integrating the diffusion model and model-based iterative reconstruction, which utilizes the variance exploding stochastic differential equation (VE-SDE) to model the diffusion process as a dynamic system affected by stochastic noise [47], whereas the IR-SDE method is a fast sparse-view reconstruction method based on the mean-reverting diffusion model proposed in this paper. The model accelerates the image reconstruction by introducing mean-reverting SDE to simulate the diffusion process, which in turn improves the efficiency and stability of image recovery.

During the reconstruction process, the blood vessels keep gradually emerging from the noise with the increase of iterations. At the 25th iteration, the artifacts in the image under the original sparse projections are gradually eliminated. At the 50th iteration, most of the artifacts in the simulated vessel image have been removed and the main structure of
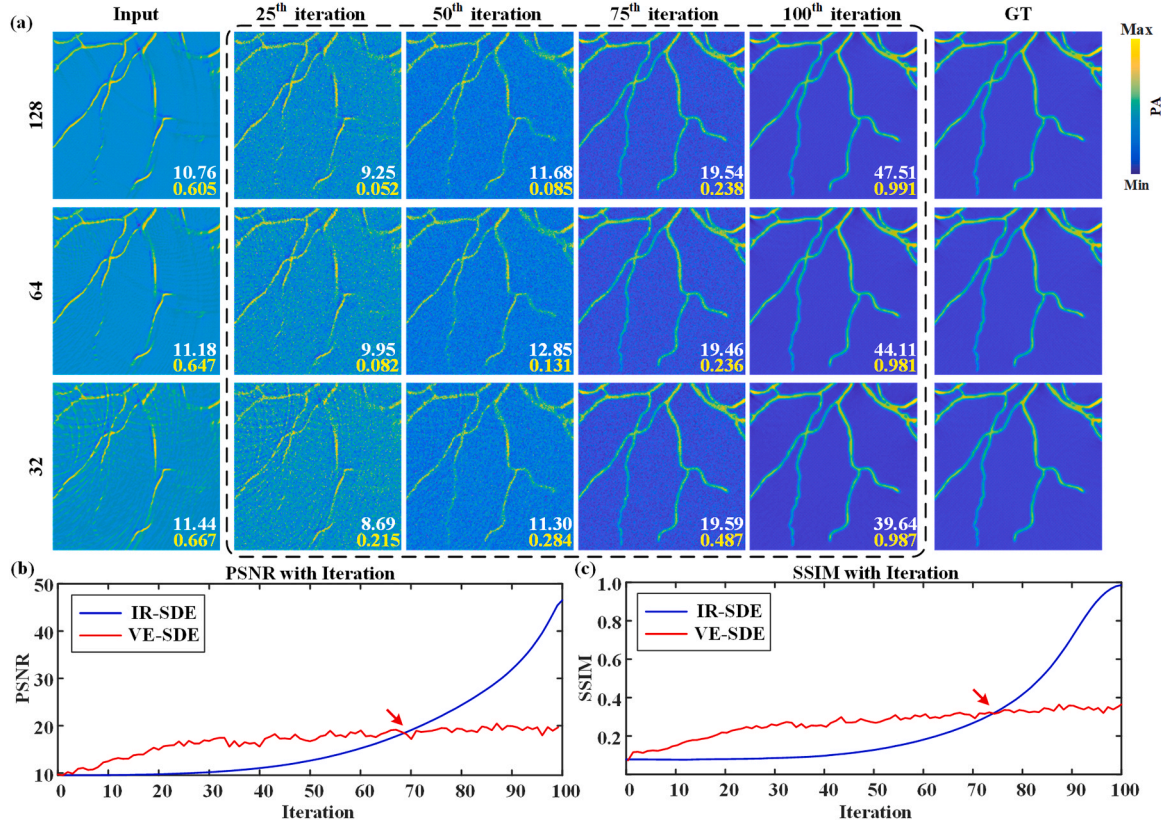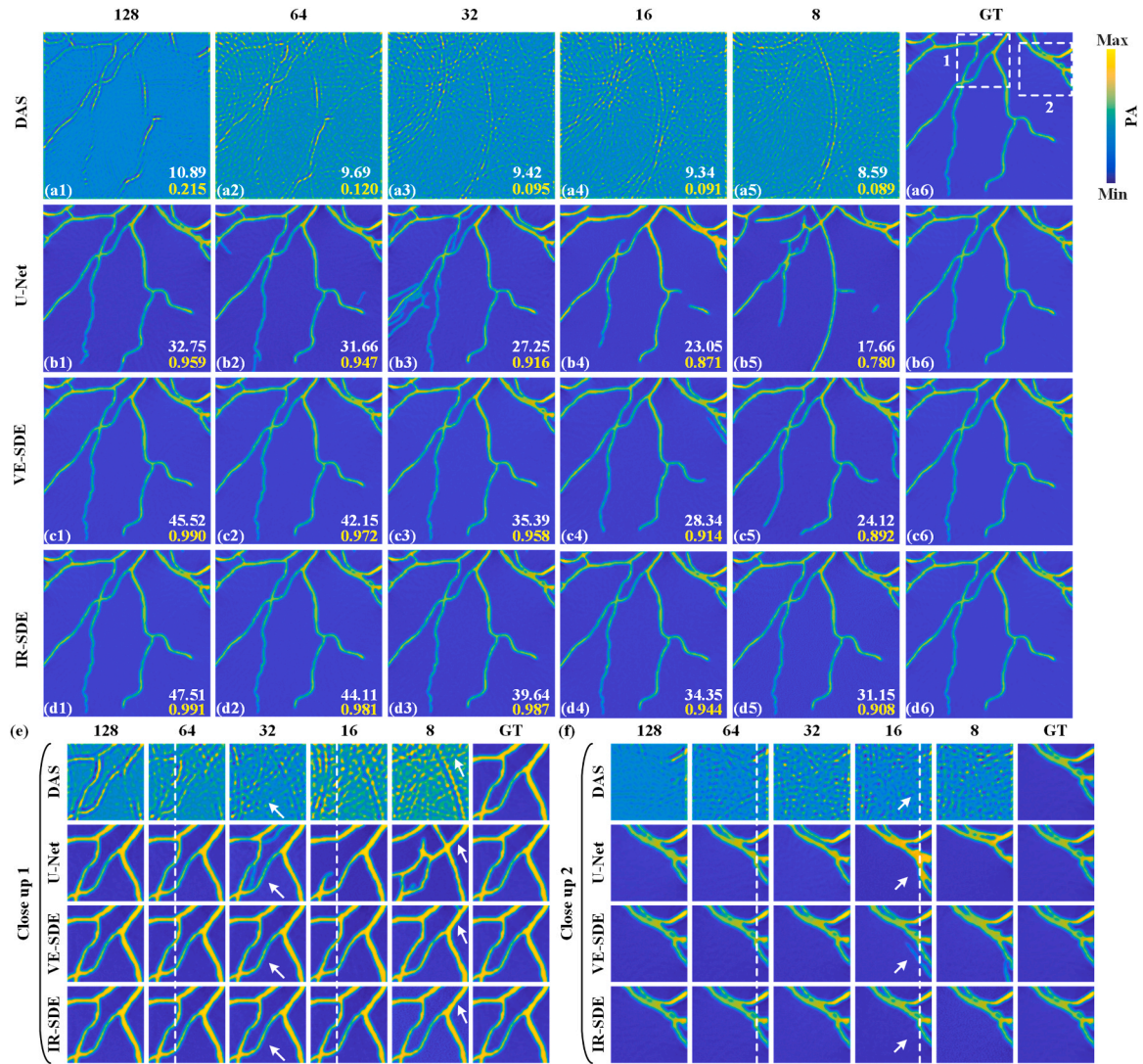

Fig. 4. The reconstruction of simulated blood vessels using the proposed method under 128, 64, and 32 projections, respectively. (a) shows the reconstruction results of the proposed method under different number of iterations. (b) and (c) show the changes of PSNR and SSIM within 100 iterations for the IR-SDE method and the VE-SDE method for the reconstruction process of 128-projections images, respectively.

the vessel is clearly visible. At the 75th iteration, the artifacts in the simulated blood vessel image are almost completely removed, and the noise in the image has begun to be gradually eliminated. At the 100th iteration, the simulated vessel image has been reconstructed and a high-quality reconstructed image is obtained. Fig. 4(b) and (c) quantitatively show the variation of PSNR and SSIM with the iteration in the sparse-view reconstruction using the VE-SDE method versus the proposed method for images under 128 projections, respectively. For the VE-SDE method, PSNR and SSIM slightly outperform the proposed method in the initial stage, however, as the number of iterations increases, PSNR and SSIM rise extremely slowly, both increasing at low values (at the 100th iteration, PSNR tends to be 19.85 dB and SSIM tends to be 0.39). At the same time, the VE-SDE method suffers from a large number of small fluctuations during the iterations and the reconstruction process is unstable. In VE-SDE method, the forward operator A and the adjoint operator A* need to be computed at each iteration by using the k-Wave toolbox. The calculation of two operators takes a lot of time [47]. The complete reconstruction using the VE-SDE method requires about 1000 iterations, and the reconstruction time is about 1 h (each iteration takes

about 4 s). For the proposed method, PSNR and SSIM are still lower than the VE-SDE method in the initial iterations before the position indicated by the red arrows in Fig. 4(b) and 4(c), due to the presence of considerable noise in the image. However, as the iteration exceeds about 70 iterations, the noise in the image is gradually removed, resulting in a rapid increase in PSNR and SSIM that greatly exceeds the results of the VE-SDE reconstruction. At the 100th iteration, the PSNR and SSIM are 47.51 dB and 0.991, respectively, and the complete iterative image reconstruction of the proposed method requires about 100 iterations, with a reconstruction time of about 5 s (~0.05 s/iteration), which is a great advantage over VE-SDE method in terms of time. Meanwhile, as shown in Fig. 4(b) and (c), the iterative curves of the proposed method are smoother, which demonstrates that the method proposed in this paper has the ability to be more stable in the iterative process. It should be noted that both performance curves for the methods presented in Fig. 4(b) and (c) rise very slowly at the beginning, and stabilize only in the last few iterations. This is due to the different inherent order in which model degradation is processed. The mean-reverting diffusion model tends to prioritize the main degradation (i.e., deblurring rings)

Fig. 5. The reconstruction results of different methods for simulated blood vessels under different projections. (a1)-(a5) are the reconstruction results of DAS method under 128, 64, 32, 16 and 8 projections, respectively, and (a6) is the ground truth (GT). (b1)-(b5) are the reconstruction results of U-Net method under 128, 64, 32, 16 and 8 projections, respectively, and (b6) is the ground truth. (c1)-(c5) are the reconstruction results of VE-SDE method under 128, 64, 32, 16 and 8 projections, respectively, and (c6) is the ground truth. (d1)-(d5) are the reconstruction results of IR-SDE method under 128, 64, 32, 16 and 8 projections, respectively, and (d6) is the ground truth. White and yellow numbers in the lower right of the image indicate PNSR and SSIM, respectively, and (e) and (f) are close-up images of rectangles 1 and 2, respectively.

and performs Gaussian denoising only in the last few iterations. As a result, the performance curve rises slowly in the early stages (the first 70 iterations). From Fig. 4(a), it can be seen that during the iterative reconstruction of the vessels, most of the blurring is removed in the intermediate stage (the first 50 iterations), however, at this point in time there is a large amount of noise in the image. Due to the presence of Gaussian noise, the improvement of the PSNR and SSIM is slow. In the last few iterations, the model starts to focus on denoising, which leads to a rapid improvement in image quality. This is reflected in a sharp increase in the performance curve and complete convergence at the 100th iteration. Taken together, these results show that the proposed method achieves significant advantages in sparse-view reconstruction compared to traditional generative models. The proposed method not only effectively improves the image quality, but also achieves significant image reconstruction quality by completing the image reconstruction in a faster time.

Fig. 5 shows the reconstruction results of simulated blood vessels by different methods with different projections. Fig. 5(a1)-5(a5) show the reconstruction results of the DAS method under 128, 64, 32, 16 and 8 projections, respectively. It can be observed that the reconstruction results of the DAS method under different projections have serious artifacts and the vessel images are blurred. As the number of projections decreases, the artifacts become more and more obvious, leading to a significant degradation of the reconstruction results. Fig. 5(a6) shows the ground truth (GT) obtained under full-view scanning (512 projections). Fig. 5(b1)-5(b5) shows the reconstruction results using the U-Net method, and the image quality of the reconstruction results of the U-Net method is greatly improved compared with that of the DAS method. As shown in Fig. 5(b1) and (b2), U-Net performs well in sparse-view reconstruction under 128 projections versus 64 projections, and the reconstructed image is almost free of artifacts, but there is still the problem of poor reconstruction of details in some parts of the blood vessel.

In addition, as shown in Fig. 5(b3)-(b5), using the U-Net method performs poorly in the sparse reconstruction under 32 projections and below, and there are still more artifacts, as shown in the close-up images Fig. 5(e) and (f), the reconstructed blood vessel images have poor details and more thin branches of blood vessels that should not be present. The white arrows in the close-up images point out the wrongly reconstructed vessels, and some of the branches that should be present are not reconstructed. Fig. 5(b6) is the ground truth (GT). Fig. 5(c1)-5(c5) shows the sparse reconstruction results of the VE-SDE method, and it can be observed that the reconstructed images based on the VE-SDE method are better in general, and are significantly better than those of the U-Net and the DAS methods, but not as good under extremely sparse projections (16 projections and 8 projections), with some artifacts and redundant blood vessels, as well as poorly reconstructed details, which does not allow the reconstruction of specific details of the vascular branches. As shown in the close-up images Fig. 5(e) and (f), where the white arrows point to in the corresponding images under 16 projections, it is obvious that the reconstruction of the VE-SDE method shows extra blood vessels with poor reconstruction details. Fig. 5(c6) is the ground truth (GT). Fig. 5(d1)- 5(d5) show the reconstruction results of the method based on the mean-reverting diffusion model under 128, 64, 32, 16 and 8 projections, respectively, and Fig. 5(d6) is the ground truth (GT). It can be observed that the proposed method shows significant reconstruction ability under different projections. Compared with the U-Net method, the IR-SDE-based method obtains less artifacts in the reconstructed images, and the vessel branches are clearer and more complete. Meanwhile, compared with the VE-SDE method, the reconstructed images obtained by the proposed method have better details, and this method is able to completely reconstruct clear and complete vessel images with higher quality even under extremely sparse projections. Fig. 5(d6) corresponds to the ground truth (GT). Quantitative analysis shows that the PSNR of the proposed method is 47.51 dB and SSIM is 0.991 under 128 projections, which improves PSNR by 14.76 dB

and SSIM by 0.032 compared to the U-Net method, and improves PSNR by 1.99 dB and SSIM by 0.001 compared to the VE-SDE method. Under 64 projections, the

PSNR and SSIM of the proposed method reach 44.11 dB and 0.981, which are 12.45 dB and 0.034 higher than that of U-Net, and 1.96 dB and 0.09 higher than that of VE-SDE, respectively. When the projections are further reduced to 32 projections, the proposed method still reaches 39.64 dB for PSNR and 0.987 for SSIM, which are higher than that of U-Net. Compared with the U-Net method, PSNR and SSIM are improved by 12.39 dB and 0.071, respectively, and compared with VE-SDE method, PSNR and SSIM are improved by 4.25 dB and 0.029, respectively. The above analyses show that the proposed method has better reconstruction effect under sparser projections, and the advantages of the proposed method are more significant under extremely sparse projections (16 and 8 projections). Under 16 projections, the proposed method improves PSNR and SSIM by 11.30 dB and 0.073 compared to U-Net, and 6.01 dB and 0.03 compared to VE-SDE, respectively. Under 8 projections, the reconstructed image of the proposed method improves PSNR and SSIM by 13.49 dB and 0.128 compared to U-Net, and PSNR and SSIM are improved by 7.03 dB and 0.016 compared to VE-SDE, respectively. The error maps corresponding to the reconstructed images of Fig. 5(a1)-5 (d5) are shown in Fig. 6(A1)-6(D5), respectively. The reconstructed images by the proposed method have fewer error and are closer to the ground truth, both under high projections and low projections. Fig. 6(g) and (h) show two sets of the signal distribution along the dashed lines, obtained under 64 projections and 16 projections, respectively. It can be found that the signals presented by the proposed method at all aspects are extremely close to the full-view images, and the reconstruction results are better, indicating the superiority of the method in sparse-view reconstruction for PAT. In the sparse-view reconstruction for PAT, the reconstruction under extremely sparse projections is usually difficult to optimize. The above results demonstrate that the proposed method can achieve high-quality reconstruction even under extremely sparse projections (e.g., 16, 8 projections), which is significantly better than the VE-SDE method and the U-Net method.

3.2. Phantom experiment

In order to verify the effectiveness of the proposed method in the real experimental data, the reconstruction performance tests based on DAS method, U-Net method, VE-SDE method and the proposed IR-SDE method on circular phantom were carried out and the results are shown in Fig. 7.

Fig. 7 (a1)-(a5) are the results of reconstructed images using DAS method under 128, 64, 32, 16 and 8 projections, respectively, and Fig. 7 (a6) is the ground truth. Circular phantom becomes clearer as the number of projections increases. However, the reconstructed images still have a lot of serious artifacts and there is a large difference from the ground truth image. Fig. 7(b1)-(b5) show the reconstructed images using the U-Net method under 128, 64, 32, 16 and 8 projections, respectively. The results show that the U-Net method can achieve the reconstruction of the main part of the body, but there are still some artifacts around the circular phantom, and the artifacts become more serious as the projection becomes sparser. Fig. 7(c1)-(c5) show the reconstructed images using the VE-SDE method under 128, 64, 32, 16 and 8 projections, respectively. The VE-SDE method eliminates most of the artifacts while achieving a partial reconstruction of the main body, but when the projection is extremely sparse (e.g., 16, 8 projections), the main part of the image reconstructed by the VE-SDE method appears to be partially unclear, and there is a large gap with the ground truth. Fig. 7 (d1)-(d5) shows the reconstructed images using the IR-SDE method under 128, 64, 32, 16 and 8 projections, respectively, and the proposed method achieves superior performance and produces high-quality image reconstructions compared with the DAS, U-Net, and VE-SDE methods, and the quality of the reconstructed images gradually improves with the increase of the number of projections. At the same time, better
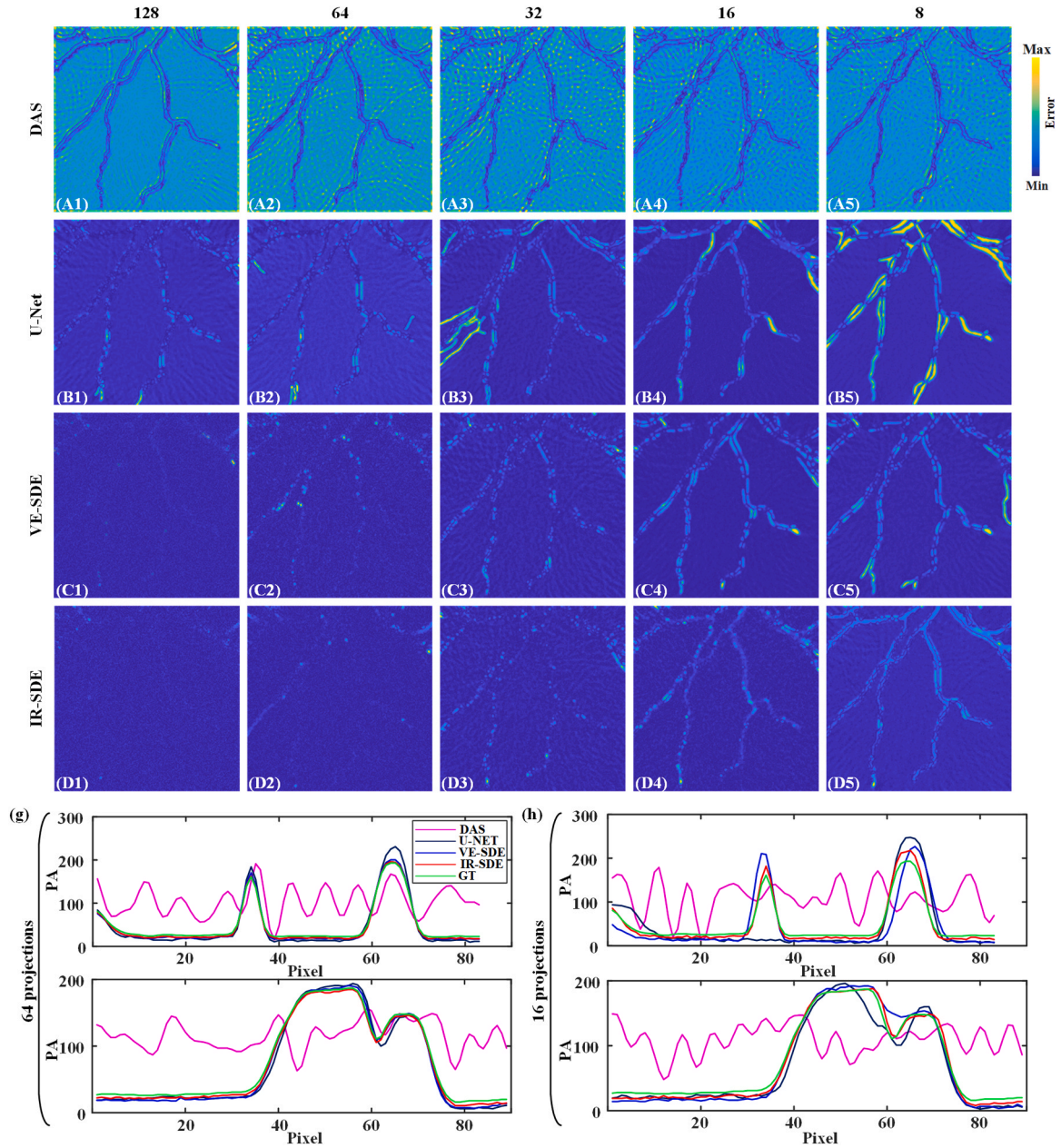
Fig. 6. Error map and the signal distribution. The top image is the error map corresponding to the reconstruction results for each image in Fig. 5(g) and (h) are the signal distribution along the dashed lines in the 64-projections and 16-projections images in the close-up image, respectively; the upper signal distribution in (g),(h) corresponds to the signal distribution along the white dashed line in the close-up image 1 in Fig. 5(e), and the lower signal distribution in (g),(h) corresponds to the signal distribution along the white dashed line in the close-up image 2 in Fig. 5(f).

reconstruction is also achieved on images under extremely sparse projections (16 and 8 projections), and the results of reconstructed images are closest to the ground truth. Quantitative analyses show the superiority of the proposed method. Taking Fig. 7 as an example, the PSNR and SSIM of the reconstructed image by the IR-SDE method are 39.55 dB and 0.96 under 128 projections, and the values of PSNR and SSIM are increased by 5.37 dB and 0.11, respectively, compared with U-Net method, and the values of PSNR is increased by 0.64 dB compared with VE-SDE method. The proposed method also achieves similar advantages under 64 projections and 32 projections. In addition, in the case of extremely sparse projections (16 and 8 projections), the advantage of the proposed method is even more significant. Under 16 projections, the PSNR and SSIM of the reconstructed image by the proposed method are 28.27 dB and 0.91, respectively, and the values of PSNR and SSIM are increased by 2.40 dB and 0.02 compared with the VE-SDE method,

by 5.58 dB and 0.56 compared with the U-Net method. Under 8 projections, the PSNR and SSIM of the images reconstructed by the proposed method are 26.64 dB and 0.86, respectively, and the values of PSNR and SSIM are improved by 2.72 dB and 0.04, respectively, compared with the VE-SDE method, and by 7.44 dB and 0.60, respectively, compared with the U-Net method.

The corresponding error maps are shown in the upper right corner of each reconstructed image in Fig. 7. In the error maps, the proposed method reconstructs images with fewer error both under higher and lower projections, and is closest to the ground truth. Fig. 7(e) and (f) show the close up images indicated by the white dashed rectangles 1 and 2, respectively. The reconstructed images using the proposed method exhibit more accurate details compared to the other three methods. Fig. 7(g) and (h) show the signal distribution along the dashed lines in Figs. 7(e) and 7(f). Similar to the processing in the simulated vessel
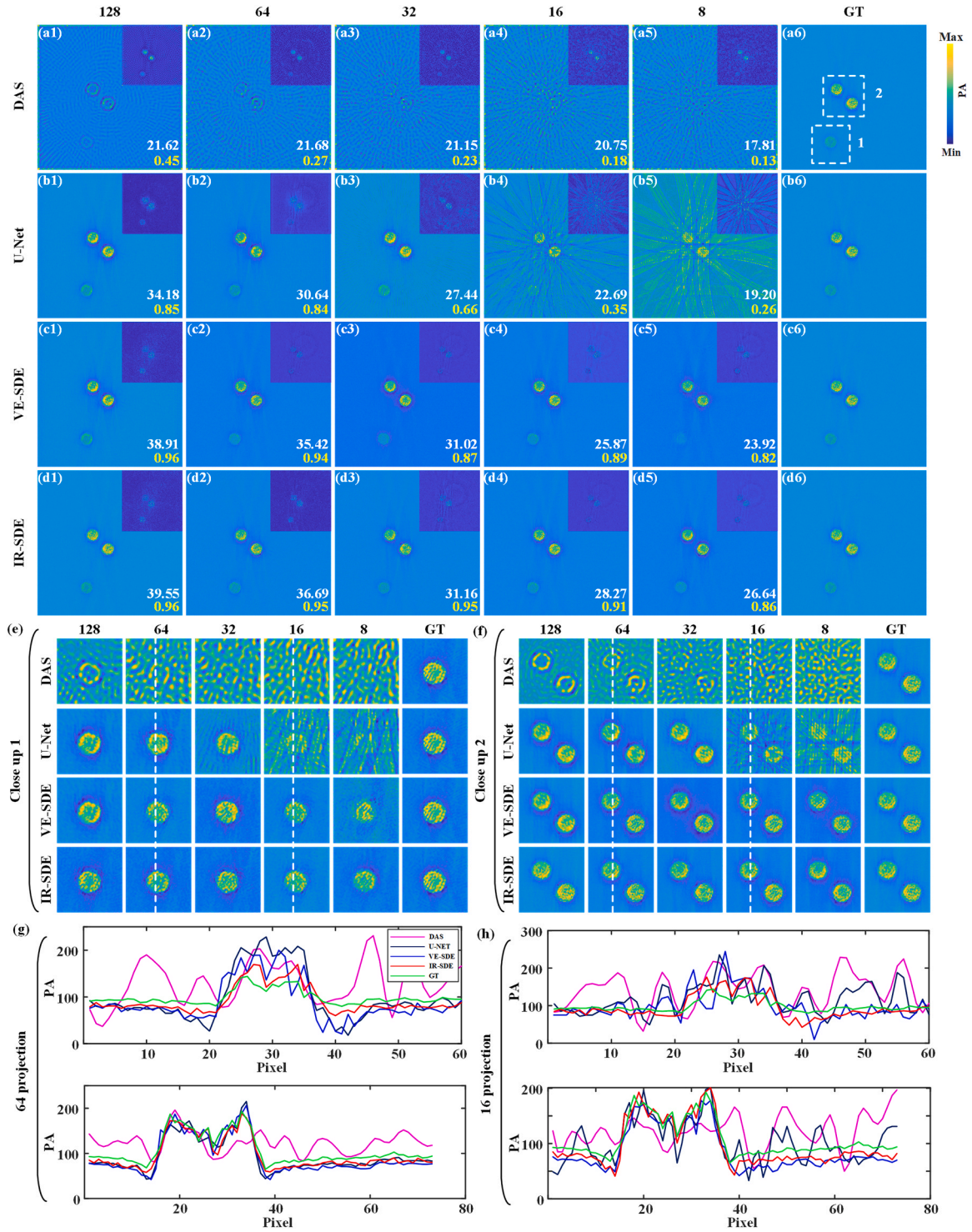
Fig. 7. The reconstruction results of different methods for circular phantom under different projections. (a1)-(a5) are the reconstruction results of DAS method under 128, 64, 32, 16 and 8 projections, respectively, and (a6) is the ground truth (GT). (b1)-(b5) are the reconstruction results of U-Net method under 128, 64, 32, 16 and 8 projections, respectively, and (b6) is the ground truth. (c1)-(c5) are the reconstruction results of VE-SDE method under 128, 64, 32, 16 and 8 projections, respectively, and (c6) is the ground truth. (d1)-(d5) are the reconstruction results of the proposed method under 128, 64, 32, 16 and 8 projections, respectively, and (d6) is the ground truth. The white and yellow numbers in the lower right of the image denote PNSR and SSIM, respectively, and the upper right image of the image is the error map corresponding to the reconstruction results, (e),(f) are the close-up images of rectangles 1 and 2, respectively. (g),(h) are the signal distribution along the dashed line of the 64-projections and 16-projections images in the close-up image, respectively; the upper signal distribution in (g),(h) corresponds to the signal distribution along the dashed line in (e) close-up image 1, and the lower signal distribution in (g),(h) corresponds to the signal distribution along the dashed line in (f) close-up image 2.

dataset, the signal distribution along the dashed line is selected for the two sets of images under 64 projections and 16 projections. Fig. 7(g) and (h) demonstrate that the signal distribution of the reconstruction results of the proposed method is much closer to the ground truth. Especially, the advantage of the proposed method is more obvious under extremely sparse projections.

To better showcase the superiority of the proposed method, ten images from the circular phantom test set are selected for the reconstruction experiments. The mean values of PSNR and SSIM of the reconstruction results are given in Table 1 and Table 2, respectively. Under extremely sparse projections (8 projections), The average PSNR of the proposed method is 27.10 dB and the average SSIM is 0.87, which are 9.56 dB and 0.69 better than the DAS method, respectively. Compared with the VE-SDE method, the results of the proposed method are on average improved by 3.33 dB and 0.05, respectively. The above results demonstrate that the reconstruction effect of the proposed method under extremely sparse projections is significantly better than that of the VE-SDE method and the U-Net method, and the experimental results verify the authenticity and superiority of the proposed method in sparse reconstruction of experimental data.

3.3. In vivo experimental data

In order to further demonstrate the effectiveness of the proposed method *in vivo* data, the reconstruction of mice's abdomens using the DAS method, the U-Net method, the VE-SDE method and the proposed method were compared, respectively, as shown in Fig. 8.

Fig. 8(a1)-(a5) are the reconstructed images of DAS method under 128, 64, 32, 16 and 8 projections respectively. Fig. 8(b1)-(b5) are the reconstructed images of U-Net method under 128, 64, 32, 16 and 8 projections respectively. Fig. 8(c1)-(c5) are the reconstructed images of VE-SDE method under 128, 64, 32, 16 and 8 projections, and Fig. 8(d1)-(d5) are the reconstructed images of the proposed method under 128, 64, 32, 16 and 8 projections, respectively. It can be clearly observed that all the images reconstructed by the DAS method under different projections have extremely severe artifacts. The U-Net method reconstructs images under sparser projections (32 projections or the sparser projections) with severe loss of detail and some artifacts, which becomes more obvious in the reconstruction of images under sparser projections. Although the VE-SDE method is more effective compared to the U-Net method, its reconstruction is still poor under extremely sparse projections (16 and 8 projections). In the reconstruction under 16 projections, the subject part is extremely blurred with unclear details, and in the reconstruction under 8 projections, the subject is reconstructed incorrectly, and even the round abdomen is reconstructed as a polygon, which seriously lacks details. In contrast, the proposed method achieves better reconstruction under sparse projections, especially in the reconstruction of images under sparse projections, the reconstruction is more superior and stable. Compared with DAS, U-Net and VE-SDE method, the proposed method has significant advantages in sparse image reconstruction of experimental data.

Fig. 8 (A1)-8(D1) in Fig. 8(g) are the error maps of Fig. 8(a1)-(d1), respectively, in which it can be found that the errors of DAS are more obvious and the reconstruction effect is poorer, and the reconstruction effect is poorer when the projections are sparser. Although the U-Net and the VE-SDE method are more effective in the reconstruction of

Table 1
The mean for the PSNR of circular phantom data.

Methods	128 projections	64 projections	32 projections	16 projections	8 projections
DAS	22.43 dB	21.74 dB	20.98 dB	19.87 dB	17.54 dB
U-Net	34.12 dB	30.02 dB	27.29 dB	23.74 dB	19.70 dB
VE-SDE	38.52 dB	35.21 dB	31.64 dB	26.05 dB	23.77 dB
IR-SDE	39.14 dB	36.59 dB	32.12 dB	29.27 dB	27.10 dB

Table 2
The mean for the SSIM of circular phantom data.

Methods	128 projections	64 projections	32 projections	16 projections	8 projections
DAS	0.45	0.28	0.22	0.19	0.18
U-Net	0.87	0.83	0.62	0.46	0.29
VE-SDE	0.95	0.93	0.88	0.87	0.82
IR-SDE	0.97	0.95	0.94	0.91	0.87

images under high-projections image, there are still artifacts when the view is sparse, and the reconstruction is not effective. While the proposed method reconstructs images with fewer error under low projections, and is closest to the ground truth. Fig. 8(e) and (f) show the close up images indicated by the white dashed rectangles 1 and 2, respectively. The reconstructed images using the proposed method exhibit more accurate details compared to the other three method. Fig. 8 (h) and (i) are the signal distribution along the white dashed lines. The reconstructed signal by the proposed method is closer to the signal of the ground truth. Quantitative analysis shows the superiority of the method. As shown in Fig. 8, the proposed method achieves a PSNR of 32.27 dB and an SSIM of 0.94 under 128 projections, and the values of PSNR and SSIM are improved by 5.26 dB and 0.2 compared with the U-Net method, and 0.44 dB and 0.02 compared with the VE-SDE method, respectively. Under 64 and 32 projections, the PSNR and SSIM of the reconstructed image by the proposed method are 27.02 dB and 0.70, respectively, and the values of PSNR and SSIM are improved by 7.76 dB and 0.18 compared with the U-Net method, and by 3.33 dB and 0.09 compared with the VE-SDE method, respectively. The above experimental results fully verify the authenticity and superiority of the proposed method in extremely sparse image reconstruction of *in vivo* experimental data.
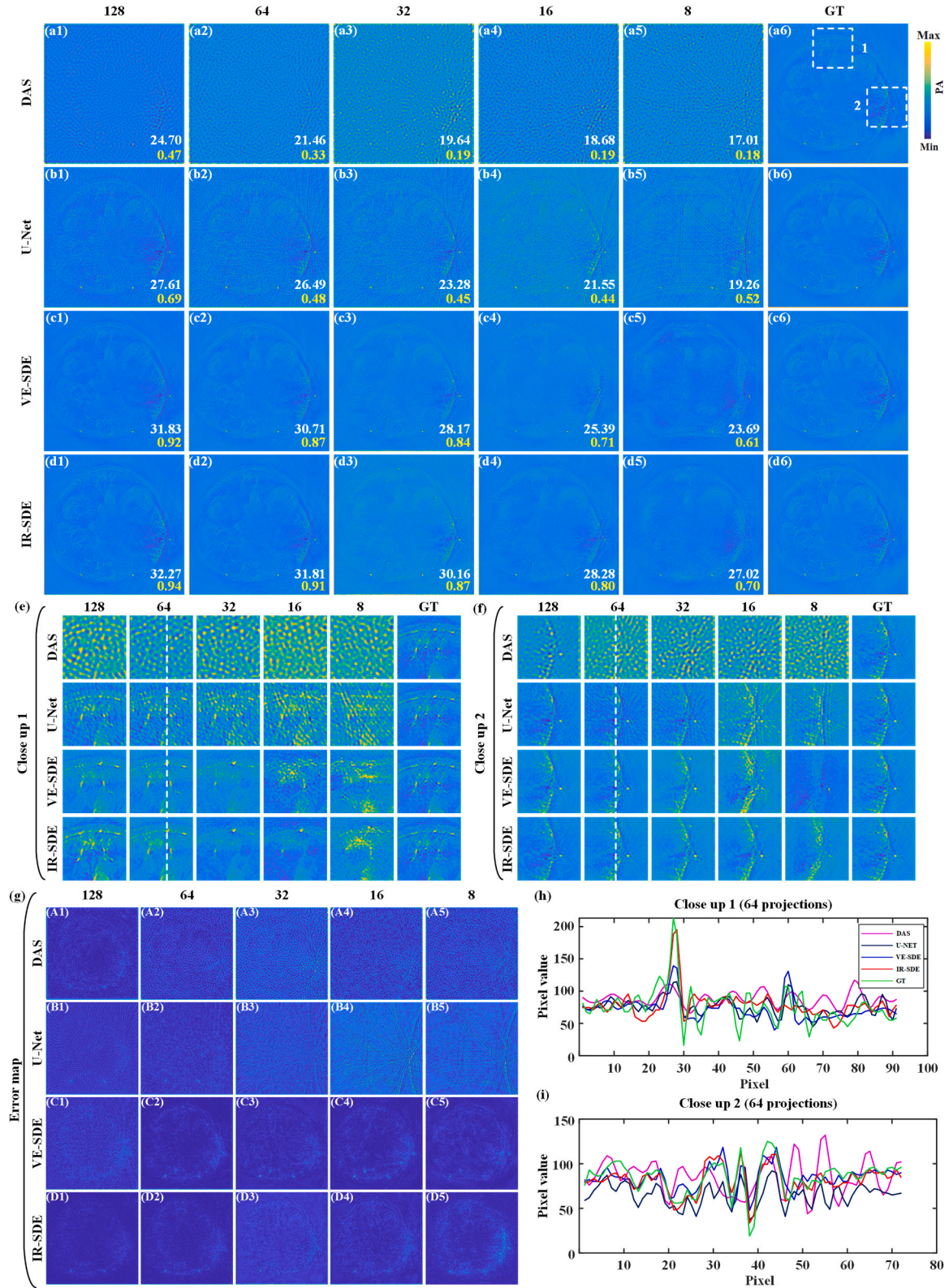
4. Discussion and conclusion

In conclusion, to address the problem that conventional standard reconstruction methods for PAT in sparse views may lead to low image quality in photoacoustic tomography, this study proposes a novel sparse view reconstruction method based on a mean-reverting diffusion model. This score-based diffusion model is used to learn prior information about the data distribution, and in the iterative reconstruction, the learned prior information is utilized as a distribution constraint to guide the reconstruction process, which is combined with the maximum likelihood objective to find the optimal sparse image reconstruction trajectory, aiming to stabilize the training and improves the restoration results. In the forward process, the proposed method converts the degradation of high-quality images into low-quality images with fixed Gaussian noise, and in the reverse process, since the mean-reverting SDE in the proposed method has a closed-form solution, the solution of the reverse SDE can be performed iteratively by numerical methods to realize the recovery of high-quality images from the mean state. Blood vessels simulation data and the animal *in vivo* experimental data were used to evaluate the performance of the proposed method. Quantitative analysis was carried out to verifies the superiority of the proposed method. For the *in vivo* experimental data, under 128 projections, the PSNR and SSIM of the reconstructed images of the proposed method are improved to 32.27 dB and 0.94, and the values of PSNR and SSIM are increased by 5.26 dB and 0.27, respectively, compared with U-Net method. In addition, the proposed method remains superior under extremely sparse conditions and significantly outperforms other conventional generative models. Under 8 projections, the PSNR and SSIM of the reconstructed image of the proposed method are improved to 27.02 dB and 0.70, respectively, and the values of PSNR and SSIM are increased by 7.76 dB and 0.18, respectively, compared with U-Net method, and by 3.33 dB and 0.09 respectively, compared with VE-SDE method. Compared with the conventional delay-and-sum (DAS), the

Fig. 8. The reconstruction results of different for *in vivo* mouse abdomen under different projections. (a1)-(a5) are the reconstruction results of DAS method under 128, 64, 32, 16 and 8 projections, respectively, and (a6) is the ground truth (GT). (b1)-(b5) are the reconstruction results of U-Net method under 128, 64, 32, 16 and 8 projections, respectively, and (b6) is the ground truth. (c1)-(c5) are the reconstruction results of VE-SDE method under 128, 64, 32, 16 and 8 projections, respectively, and (c6) is the ground truth. (c1)-(c5) are the reconstruction results of IR-SDE method under 128, 64, 32, 16 and 8 projections, respectively, and (a6) is the ground truth (GT). The white and yellow numbers in the lower right of the image denote PNSR and SSIM. (e) and (f) are the close-up images of rectangles 1 and 2, respectively. (g) is the corresponding error map of each reconstructed image. (h) is the signal distribution along the dashed line for the image under 64 projections in the close-up image 1. (i) is the signal distribution along the dashed line for the image under 64 projections in the close-up image 2.

proposed method achieves an improvement of 0.68 (~358 %) in SSIM and 10.52 dB (~54 %) in PSNR under sparser projection (32 projections), and achieves an improvement of 0.52 (~289 %) in SSIM and 10.01 dB (~59 %) in PSNR under extremely sparse projections (8 projections). These results demonstrate that the proposed method has a more stable and excellent reconstruction capability in sparse reconstruction for PAT, and it gives an effective solution to the problem of poor reconstruction for PAT under extremely sparse conditions by conventional generative models. At the same time, the faster imaging speed and lower cost of the proposed method will enable wider application of PAT in scenarios such as clinical diagnostics, *in vivo* vascular detection, and other scenarios that require real-time imaging feedback. This breakthrough may significantly reduce the mechanical complexity of scanning devices, enabling high-speed dynamic imaging (e.g., real-time cardiovascular imaging) while reducing the risk of motion artifacts. The proposed method is expected to reduce the time required for biological tissue reconstruction while improving the accuracy of tissue reconstruction due to incomplete data acquisition, which may further extend the practical application of PAT.

However, this study also has some limitations, on the one hand, pairs of scanned images are required (i.e., sparse-view images and full-view images), and due to physical limitations for PAT, it is still challenging for real imaging systems to acquire pairs of images [48]. PAT systems usually use circular scanning, but the equipment and data complexity are high, which leads to limited data acquisition efficiency and makes it difficult to quickly generate large-scale training datasets. On the other hand, real-time performance is an important consideration in practical clinical performance. Unlike traditional diffusion models that takes about 1 h for a single reconstruction, the proposed method only takes about 5 s. However, considering the clinical demand for real-time dynamic imaging [49], how to further reduce the model iteration time without affecting the reconstruction quality, so as to meet the needs of clinical dynamic imaging, is still a direction to be explored.

During the training of the mean-reverting diffusion model, the unknown score function needs to be estimated by training the score network, and the training process takes a lot of time due to the need to continuously transforms a high-quality image into a degraded counterpart as a mean state with fixed Gaussian noise, and to learn the data distribution simultaneously in the process. The exact time is related to the graphics processing unit configuration (GPU; NVIDIA RTX A5000 for this experiment), dataset size and parameter settings. A total of 6000 epochs were trained in the training phase, one checkpoint was saved for every 100 epochs, and a single checkpoint took about 60 minutes, and an average model was saved synchronously after finalizing the training, from which the best training model was selected. As shown in Fig. 4(b), at the 100th iteration, PSNR and SSIM reach a steady state. The time taken for single image reconstruction is about 5 s (~20 iterations/s). To summarize, the DAS method has the lowest computational complexity and the shortest single reconstruction time, but its image quality is poor. The U-Net method has a simple structure and its convolutional layers have limited receptive domains, which hinders its ability to learn global features when dealing with sparse-view images, and makes it not as effective as the diffusion model, but its reconstruction time is shorter. The VE-SDE method requires higher computational resources and takes longer to reconstruct, but the reconstructed images are of higher quality. The method proposed in this paper has a higher demand for computational resources, but it can achieve high-quality reconstruction in a shorter time, which is expected to shorten the acquisition time of PAT in the practical application and reduce the cost, thus further expanding the application scope of PAT.

Disclosures

The authors declare no conflicts of interest.

CRediT authorship contribution statement

Lin Jiabin: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Wang Guijun:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Li Jiahong:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Li Zilong:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Song Xianlin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Liu Qiegen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Cao Yiyang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lv Yichen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lian Teng:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Guo Kangjun:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

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.

Data Availability

Our code is publicly available at https://github.com/yqx7150/GAID-PAT.

References

[1] L.V. Wang, Multiscale photoacoustic microscopy and computed tomography, Nat. Photonics 3 (9) (2009) 503–509.

[2] M.H. Xu, L.V. Wang, Photoacoustic imaging in biomedicine, Rev. Sci. Instrum. 77 (4) (2006) 041101.

[3] A.P. Jathoul, J. Laufer, O. Ogunlade, B. Treeby, B. Cox, E. Zhang, P. Johnson, A. R. Pizzey, B. Philip, T. Marafioti, M.F. Lythgoe, R.B. Pedley, M.A. Pule, P. Beard, Deep *in vivo* photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter, Nat. Photonics 9 (2015) 239–246.

[4] P. Beard, Biomedical photoacoustic imaging, Interface Focus 1 (4) (2011) 602–631.

[5] C. Tian, Z. Xie, M.L. Fabiilli, X. Wang, Imaging and sensing based on dual-pulse nonlinear photoacoustic contrast: a preliminary study on fatty liver, Opt. Lett. 40 (10) (2015) 2253–2256.

[6] C. Tian, W. Qian, X. Shao, Z. Xie, X. Cheng, S. Liu, Q. Cheng, B. Liu, X. Wang, Plasmonic nanoparticles with quantitatively controlled bioconjugation for photoacoustic imaging of live cancer cells, Adv. Sci. 3 (12) (2016) 1600237.

[7] M. Mehrmohammadi, S. Joon Yoon, D. Yeager, S. Y. Emelianov, Photoacoustic imaging for cancer detection and staging, Curr. Mol. Imaging 2 (1) (2013) 89–105.

[8] M. Wu, N. Awasthi, N.M. Rad, J.P. Pluim, R.G. Lopata, Advanced ultrasound and photoacoustic imaging in cardiology, Sensors 21 (2021) 7947.

[9] S.H. Han, Review of photoacoustic imaging for imaging-guided spinal surgery, Neurospine 15 (4) (2018) 306–322.

[10] M.A. Lediju Bell, Photoacoustic imaging for surgical guidance: principles, applications, and outlook, J. Appl. Phys. 128 (6) (2020) 060904.

[11] P.K. Upputuri, M. Pramanik, Recent advances toward preclinical and clinical translation of photoacoustic tomography: a review, J. Biomed. Opt. 22 (4) (2016) 041006.

[12] L. Lin, P. Hu, X. Tong, S. Na, R. Cao, X. Yuan, D.C. Garrett, J. Shi, K. Maslov, L. V. Wang, High-speed three-dimensional photoacoustic computed tomography for preclinical research and clinical translation, Nat. Commun. 12 (1) (2021) 882.

[13] G. Diot, S. Metz, A. Noske, E. Liapis, B. Schroeder, S.V. Ovsepian, R. Meier, E. Rummeny, V. Ntziachristos, Multispectral optoacoustic tomography (MSOT) of human breast cancer, Clin. Cancer Res 23 (22) (2017) 6912–6922.

[14] M.H. Xu, L.V. Wang, Pulsed-microwave-induced thermoacoustic tomography: Filtered backprojection in a circular measurement configuration, Med. Phys. 29 (8) (2002) 1661–1669.

[15] B.E. Treeby, E.Z. Zhang, B.T. Cox, Photoacoustic tomography in absorbing acoustic media using time reversal, Inverse Probl. 26 (11) (2010) 115003.

[16] G. Matrone, A.S. Savoia, G. Caliano, G. Magenes, The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging, IEEE Trans. Med. Imaging 34 (4) (2015) 940–949.

[17] L. Lin, P. Hu, J. Shi, C.M. Appleton, K. Maslov, L. Li, R. Zhang, L.V. Wang, Singlebreath-hold photoacoustic computed tomography of the breast, Nat. Commun. 9 (1) (2018) 2352.

[18] L. Li, L. Zhu, C. Ma, L. Lin, J. Yao, L. Wang, K. Maslov, R. Zhang, W. Chen, J. Shi, L. V. Wang, Single-impulse panoramic photoacoustic computed tomography of smallanimal whole-body dynamics at high spatiotemporal resolution, Nat. Biomed. Eng. 1 (5) (2017) 0071.

[19] G. Paltauf, J.A. Viator, S.A. Prahl, S.L. Jacques, Iterative reconstruction algorithm for optoacoustic imaging, J. Acoust. Soc. Am. 112 (4) (2002) 1536–1544.

[20] E.J. Candès, J. Romberg, T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, J. IEEE Trans. Inf. Theory 52 (2) (2006) 489–509.

[21] S. Gutta, S.K. Kalva, M. Pramanik, P.K. Yalavarthy, Accelerated image reconstruction using extrapolated Tikhonov filtering for photoacoustic tomography, Med. Phys. 45 (8) (2018) 3749–3767.

[22] S. Arridge, P. Beard, M. Betcke, B. Cox, N. Huynh, F. Lucka, O. Ogunlade, E. Zhang, Accelerated high-resolution photoacoustic tomography via compressed sensing, Phys. Med. Biol. 61 (2016) 8908.

[23] C.B. Shaw, J. Prakash, M. Pramanik, P.K. Yalavarthy, Least Squares QR-based decomposition provides an efficient way of computing optimal regularization parameter in photoacoustic tomography, J. Biomed. Opt. 18 (080501) (2013) 1–3.

[24] J. Wang, Y.Y. Wang, Photoacoustic imaging reconstruction using combined nonlocal patch and total-variation regularization for straight-line scanning, Biomed. Eng. Online 21 (1) (2022) 1–24.

[25] J. Gröhl, M. Schellenberg, K. Dreher, L. Maier-Hein, Deep learning for biomedical photoacoustic imaging: a review, Photoacoustics 22 (2021) 100241.

[26] M. Kim, G.S. Jeng, I. Pelivanov, M. O'Donnell, Deep-learning image reconstruction for real-time photoacoustic system, IEEE Trans. Med. Imaging 39 (11) (2020) 3379–3390.

[27] S. Antholzer, M. Haltmeier, J. Schwab, Deep learning for photoacoustic tomography from sparse data, Inverse Prob. Sci. Eng. 27 (7) (2019) 987–1005.

[28] A. Hauptmann, B. Cox, Deep learning in photoacoustic tomography: Current approaches and future directions, J. Biomed. Opt. 25 (11) (2020), 112903–112903.

[29] A. Madasamy, V. Gujrati, V. Ntziachristos, J. Prakash, Deep learning methods hold promise for light fluence compensation in three-dimensional optoacoustic imaging, J. Biomed. Opt. 27 (10) (2022).

[30] T. Chen, T. Lu, S. Song, S. Miao, F. Gao, J. Li, A deep learning method based on UNet for quantitative photoacoustic imaging, Photons Ultrasound.: Imaging Sens. (2020).

[31] N. Davoudi, X.L. Deán-Ben, D. Razansky, Deep learning optoacoustic tomography with sparse data, Nat. Mach. Intell. 1 (10) (2019) 453–460.

[32] H. Shahid, A. Khalid, X. Liu, M. Irfan, D. Ta, A deep learning approach for the photoacoustic tomography recovery from undersampled measurements, Front. Neurosci. 15 (2021) 598693.

[33] S. Guan, A.A. Khan, S. Sikdar, P.V. Chitnis, Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal, IEEE J. Biomed. Heal. Inform. 24 (2) (2020) 568–576.

[34] J. Deng, et al., Unet-based for photoacoustic imaging artifact removal, Proc. IEEE Int. Ultrason. Symp. (IUS) (2018) 1–4.

[35] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courvile, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[36] D. Rezende and S. Mohamed, Variational inference with normalizing flows, in International Conference on Machine Learning (ICML2015), pp. 1530–1538.

[37] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. 33 (2020) 6840–6851.

[38] Y. Song, S. Ermon, Improved techniques for training score-based generative models, arXiv, arXiv (2020), 2006.09011.

[39] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, arXiv, arXiv 2011 (2020) 13456.

[40] Z. Luo, F.K. Gustafsson, Z. Zhao, J. Sjölund, T.B. Schön, Image Restoration with Mean-Reverting Stochastic Differential Equations, arXiv,arXiv 2301 (2023) 11699.

[41] X. Mao, The truncated Euler–Maruyama method for stochastic differential equations, J. Comput. Appl. Math. 290 (2015) 370–384.

[42] G.N. Milstein, Approximate integration of stochastic differential equations. Theory Probab. Its Appl. 19 (3) (1975) 557–562.

[43] L. Chen, X. Chu, X. Zhang, J. Sun, Simple baselines for image restoration, arXiv, arXiv (2022), 2204.04676.

[44] Z. Luo, F.K. Gustafsson, Z. Zhao, J. Sjölund, T.B. Schön, Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. Computer Vision and Pattern Recognition (CVPR2023), IEEE Press, Vancouver, 2023, pp. 1680–1691.

[45] B.E. Treeby, B.T. Cox, k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, 021314-021314-12, J. Biomed. Opt. 15 (2) (2010), 021314-021314-12.

[46] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Ridgebased vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509.

[47] X. Song, G. Wang, W. Zhong, K. Guo, Z. Li, X. Liu, J. Dong, Q. Liu, Sparse-view reconstruction for photoacoustic tomography combining diffusion model with model-based iteration, Photoacoustics 33 (2023) 100558.

[48] F. Dreier, S. Pereverzyev Jr, M. Haltmeier, Operator learning approach for the limited view problem in photoacoustic tomography, Comput. Methods Appl. Math. 19 (4) (2019) 749–764.

[49] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, R. Salmelin, Dynamic imaging of coherent sources: studying neural interactions in the human brain, Proc. Natl. Acad. Sci. 98 (2) (2001) 694–699.

Teng Lian is currently studying for bachelor degree in Computer Science and Technology in Nanchang University, Nanchang, China. His research interests include image processing, artificial intelligence, deep learning and photoacoustic tomography.

Yichen Lv is currently studying for bachelor degree in Artificial Intelligence in Nanchang University, Nanchang, China. His research interests include artificial intelligence, optical imaging and photoacoustic tomography.

Kangjun Guo is currently studying for bachelor degree in Automation Science in Nanchang University, Nanchang, China. His research interests include image processing, artificial intelligence and photoacoustic tomography.

Zilong Li received the bachelor degree in Electronic Information Engineering from Guilin University of Electronic Technology, Guilin, China. He is currently studying in Nanchang University for master's degree in Electronic Information Engineering. His research interests include optical imaging, deep learning and photoacoustic tomography.

Yiyang Cao is currently studying for bachelor degree in Computer Science and Technology in Nanchang University, Nanchang, China. His research interests include optical imaging, deep learning and photoacoustic microscopy.

Jiahong Li received the bachelor degree in Electronic Information Engineering from Yanshan University, Qinhuangdao, China. He is currently studying in Nanchang University for master's degree in Electronic Information Engineering. His research interests include optical imaging, deep learning and photoacoustic tomography.

Qiegen Liu received his PhD degree in Biomedical Engineering from Shanghai Jiao Tong University, Shanghai, China in 2012. Currently, he is a professor at Nanchang University. He is the winner of Excellent Young Scientists Fund. He has published more than 50 publications and serves as committee member of several international and domestic academic organizations. His research interests include artificial intelligence, computational imaging and image processing.

Guijun Wang received the bachelor degree in Electronic Information Science and Technology from Nanchang Hangkong University, Nanchang, China. He is currently studying in Nanchang University for master's degree in Electronic Information Engineering. His research interests include image processing, deep learning and photoacoustic tomography.

Xianlin Song received his PhD degree in Optical Engineering from Huazhong University of Science and Technology, China in 2019. He joined School of Information Engineering, Nanchang University as an assistant professor in Nov. 2019. He has published more than 20 publications and given more than 15 invited presentations at international conferences. His research topics include optical imaging, biomedical imaging and photoacoustic imaging.

Jiabin Lin is currently studying for bachelor degree in Automation Science in Nanchang University, Nanchang, China. His research interests include optical imaging, deep learning and photoacoustic tomography.