Sparse-view reconstruction for photoacoustic tomography combining diffusion model with model-based iteration

Xianlin Song [1], Guijun Wang [1], Wenhua Zhong, Kangjun Guo, Zilong Li, Xuan Liu, Jiaqing Dong, Qiegen Liu [*]

School of Information Engineering, Nanchang University, Nanchang 330031, China

ARTICLE INFO

ABSTRACT

As a non-invasive hybrid biomedical imaging technology, photoacoustic tomography combines high contrast of optical imaging and high penetration of acoustic imaging. However, the conventional standard reconstruction under sparse view could result in low-quality image in photoacoustic tomography. Here, a novel model-based sparse reconstruction method for photoacoustic tomography via diffusion model was proposed. A score-based diffusion model is designed for learning the prior information of the data distribution. The learned prior information is utilized as a constraint for the data consistency term of an optimization problem based on the least-square method in the model-based iterative reconstruction, aiming to achieve the optimal solution. 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. In particular, under the extreme sparse projection (e.g., 32 projections), the proposed method achieves an improvement of ~ 260 % in structural similarity and ~ 30 % in peak signal-to-noise ratio for in vivo data, compared with the conventional delay-and-sum method. This method has the potential to reduce the acquisition time and cost of photoacoustic tomography, which will further expand the application range.

1. Introduction

Photoacoustic tomography (PAT) is a hybrid non-invasive imaging technique based on the photoacoustic effect, which combines the high contrast of optical imaging with the high penetration depth of ultrasound imaging [1–4]. PAT has been widely applied in tissue imaging [5,6], cancer detection [7,8], cardiovascular disease detection [9], image-guided surgery [10,11] and other fields. It has gradually become an important tool in preclinical and clinical research [12]. In PAT, a nanosecond-level pulsed laser beam is used to irradiate the tissue, and the tissue absorbs the light energy. An instantaneous temperature rise occurs after the tissue absorbs light energy, which results in a time-varying thermal expansion–relaxation process. Thus, the initial pressure (photoacoustic signal) is generated and propagated around. The photoacoustic waves are then collected by the ultrasonic transducer placed around the tissue. The absorption distribution of the tissue is ultimately reconstructed using a reconstruction algorithm [13]. The

conventional reconstruction methods are based on analytical algorithms, such as filtered back-projection, delay-and-sum (DAS), and time-reversal methods [14–16]. However, in an actual imaging system, it is difficult to obtain a complete photoacoustic signal due to limitations in the bandwidth and number of ultrasonic transducers. Sparse-view data can lead to severe artifacts in reconstructed images. Hence, high-quality reconstruction under sparse sampling is a challenging task to be solved.

Hardware-based and algorithm-based methods have been proposed to address this issue. From the hardware perspective, it is a common practice to capture photoacoustic signals from more views by utilizing ultrasonic transducer array, such as full-ring 512-element ultrasonic transducer array [17,18]. Arrays of other shapes have also been used to improve the reconstruction quality, such as spherical [19,20], cylindrical [21,22], and planar [23] detector array. However, these methods can lead to increased complexity and cost of the system, making it challenging to achieve widespread application. Improved algorithms

* Corresponding author.
E-mail address: liuqiegen@ncu.edu.cn (Q. Liu).
[1] contributed equally to this work.

can be a potential solution for achieving high-quality reconstruction. Since sparse sampling can make the reconstruction an ill-posed problem, a series of model-based improvement methods have been proposed [24–28]. The model-based method is to build a forward model from the initial sound pressure to the measured signal, and iteratively reconstruct the image under some optimization conditions [24–26]. Paltauf et al. designed a model-based iterative reconstruction algorithm to minimize the difference between the detected photoacoustic signal and the photoacoustic signal calculated from the image to obtain a better reconstruction [24]. Variational methods are also commonly used in model-based methods [27–30]. This approach involves solving an optimization problem with a regularization term to facilitate sparse reconstruction, utilizing prior information of the image to accelerate and constrain the reconstruction process. Compressed sensing [28,31,32] and iterative back-projection [33] were also employed to achieve high-quality reconstruction. However, these improved methods require accurate prior information, which is difficult to obtain in experiments.

In recent years, deep learning-based methods have gained widespread adoption in biomedical image processing [34,35], including PAT reconstruction. PAT reconstruction based on deep learning mainly includes post-processing methods, which use a network to eliminate artifacts in the images reconstructed by conventional analytical methods. Currently, most of the post-processing methods are based on the U-Net network for artifact removal [36–43]. Davoudi et al. used U-Net network to reduce artifacts in photoacoustic images and enhance their anatomical contrast and image quantification capacities [36]. Shahid et al. combined inverse compressed sensing and ResU-Net network to process photoacoustic data and restore high-quality images [38]. Guan et al. proposed a fully dense U-Net (FD-UNet) network for removing artifacts [39]. Direct reconstruction [44–46] is also a commonly employed method, which directly maps the detected signal to photoacoustic image through network. Tong et al. proposed the FPU-Net and utilized data-driven learning for signal-to-image transformation [44]. Kim et al. converted the raw channel data into a multi-channel array in preprocessing, and proposed the upgU-Net network to reconstruct photoacoustic images [46]. The above work is based on end-to-end network, which can achieve effective denoising and imaging. However, the networks are data-driven and require a significant amount of labeled and ground truth data for training, and obtaining such paired data in practice can be challenging. The third common method is the model-based learning method, which combines the physical model of PAT with the deep learning network, the prior information learned by deep network for constraints is added to the model-based iteration. Wang et al. applied a learned regularization to model-based deep learning networks to speed up the iterative reconstruction [47]. Hauptmann et al. proposed an iterative learning strategy called deep gradient descent (DGD) for PAT reconstruction based on a variational approach [48].

With the development of generative models, various deep generative models, such as generative adversarial network (GAN) [49,50], autoregressive model [51,52], flow [53], variational autoencoder (VAE) [54, 55], denoising diffusion probabilistic model (DDPM) [56] and score-based generative model [57] show great advantages in high-quality sample generation. Among them, the score-based generative model adopts a more efficient sampling method to further expand the generative ability. The probability model is obtained by learning the probability distribution of the given sample, and the target image is generated by fitting the sample through the model. The generative model can learn prior information during the learning process of fitting parameters. Inspired by this, this study proposed a sparse reconstruction strategy that combines score-based generative model and model-based iterative reconstruction method. A score-based generative model (diffusion model) is designed for learning the prior information of the data distribution. The learned prior information is utilized as a constraint for the data consistency term of an optimization problem based on the least-square method in the model-based iterative reconstruction, aiming to achieve the optimal solution. 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 can perform high-quality reconstruction under extreme sparse conditions.

2. Principles and methods

2.1. Photoacoustic tomography

In PAT, ultrasonic transducers placed around the tissue are used for the detection of the ultrasonic waves [2]. The amplitude of the photoacoustic signal is related to the absorption coefficient of biological tissue and the optical flux density. The initial pressure distribution p_0 can be reconstructed with the detected photoacoustic signal, which can be expressed by Eq. (1):

$$p_0 = \Gamma \eta_{th} \mu_a \phi, \tag{1}$$

where Γ is the Gruneisen coefficient, η_{th} represents the thermal conversion efficiency, μ_a is the absorption coefficient, and ϕ represents the optical flux. In PAT, if laser excitation satisfies both thermal and stress confinements, the heat conduction and fractional volume expansion are negligible [2]. The initial pressure will propagate in the medium, as expressed by the spatiotemporal equation Eq. (2):

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) p\left(r, t \right) = - \frac{\beta}{c_p} \frac{\partial H(r,t)}{\partial t}, \tag{2}$$

Where c represents the speed of sound, β represents the coefficient of thermos-elastic expansion, and c_p is the specific heat capacity, $p(r,t)$ represents the pressure field at position r and time t, $p(r,t)$ can be solved by the Green's function method, as shown in Eq. (3) :

$$p\left(r, t \right) = \frac{1}{4\pi c^2} \frac{\partial}{\partial t} \left[\frac{1}{ct} \int dr' p_0 \left(r' \right) \delta \left(t - \frac{|r - r'|}{c} \right) \right], \tag{3}$$

Eq. (3) describes the forward process of photoacoustic imaging, which can be equivalent to the linear process shown in Eq. (4):

$$y = Ax \tag{4}$$

where y represents the sparse-view time series photoacoustic signal measured by the ultrasonic transducer $p(r, t)$, x represents the initial pressure p_0. The linear operator A represents the forward process, which can be realized through k-Wave toolbox [58]. The reconstruction problem (inverse problem) in PAT is to reconstruct the initial pressure x from the measured time series photoacoustic signal y through optimization algorithms.

2.2. Principle of sparse reconstruction based on the diffusion model

2.2.1. Diffusion model

Generative models have achieved great success in generating realistic and diverse data. Common deep generative models can be divided into two categories: explicit generative models and implicit generative models [59]. GAN [60] is a well-known implicit probabilistic generative model that generates high-quality images by optimizing the generator and discriminator. However, GAN suffer from mode collapse when the network training is unstable. Explicit generative models, which use probability density functions and probability mass functions to describe the data distribution, provide an explicit parameter specification for the data distribution. Common explicit generative models include
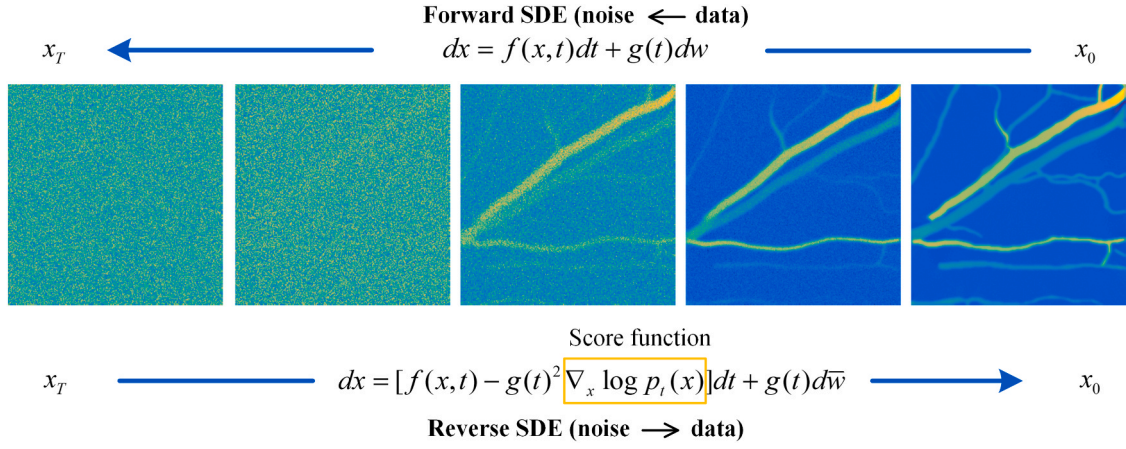
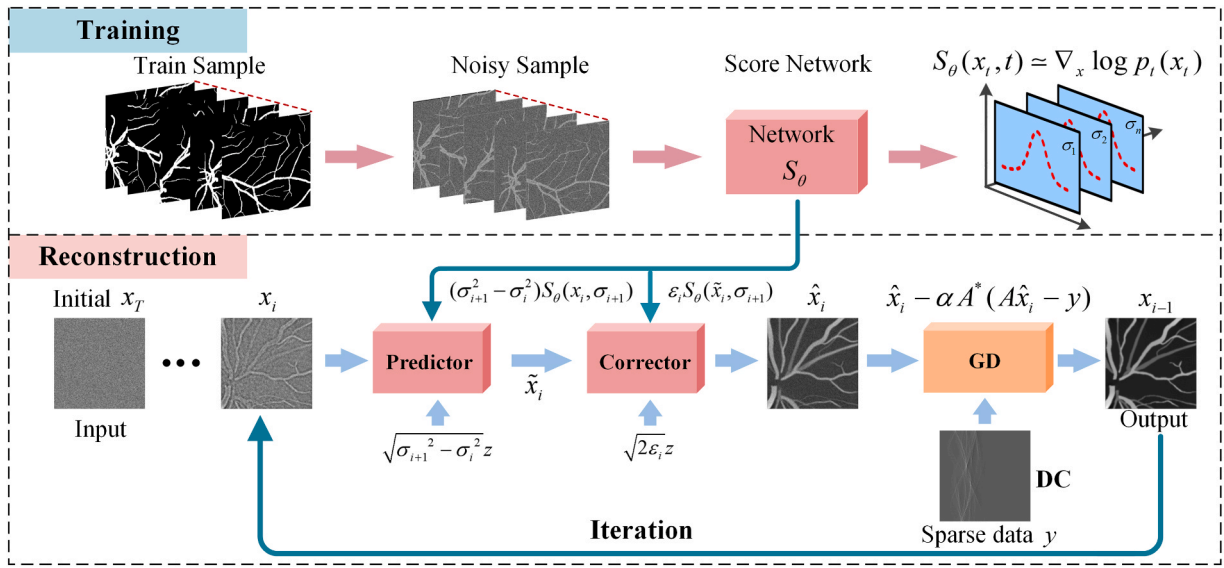**Fig.1.** Forward and reverse processes of diffusion model.

Fig. 2. Sparse reconstruction flowchart of PAT. Top: Training process to learn the gradient distribution using denoising score matching. Bottom: Reconstruction process, iteration between numerical SDE solver and gradient descent to achieve sparse reconstruction. GD, gradient descent; DC, data consistency.

autoencoders (AE) [61,62], flow-based generative models [53,63], and deep Boltzmann machines [64]. In generative models, each data in the training datasets is an independent and identically distributed random sample that conforming to the probability distribution [57]. The probability distribution is usually represented by a score function, where the score function is the gradient of the log-probability density function. Generative model based on score matching (diffusion model) is used to estimate the score function of the data distribution by optimizing a score network. The key to score matching-based generative models lies in perturbing the data with multi-scale noises.

The diffusion model includes a forward diffusion process (stochastic differential equation, SDE) and a corresponding reverse SDE, as shown in Fig. 1. The data distribution is perturbed by adding Gaussian noise in forward SDE to obtain prior distribution. And in reverse SDE, the prior distribution is transformed into the data by slowly removing the noise to achieve the purpose of sampling from the data distribution. $(x_t)_{t=0}^{T}$ is assumed to be a continuous diffusion process with $x_t \in \mathbb{R}$, where $t \in [0, T]$ is a continuous-time variable. $x_0 \sim p_{data}$, p_{data} is the data distribution of target image. $x_T \sim p_T$ (p_T is the prior distribution containing p_{data}

information) is the target-related prior information learned in the forward SDE process. The forward diffusion process is described by Eq. (5):

$$dx = f(x,t)dt + g(t)dw, \tag{5}$$

where $f(x,t) \in \mathbb{R}$ and $g(t) \in \mathbb{R}$ are the drift and diffusion coefficients of x_t, respectively, $w \in \mathbb{R}$ induces Brownian motion. To sample from a data distribution, a neural network can be trained to estimate the gradient of the logarithmic data distribution $\nabla_x \log p_t(x)$ (score function), and this gradient can be used to solve the reverse-time SDE to realize data generation from noise.

The above inverse process is also a diffusion process, which can be expressed as the reverse-time SDE shown in Eq. (6):

$$dx = \left[f(x,t) - g(t)^2 \nabla_x \log p_t(x) \right] dt + g(t) d\overline{w}, \tag{6}$$

where dt is an infinitesimal negative time step, $\overline{w}$ is the inverse Brownian motion. Different SDEs can be constructed by choosing different $f(x,t)$ and $g(t)$. Since variance exploding (VE) SDEs can generate higher-quality samples, VE-SDE was adopted in this work, as illustrated in
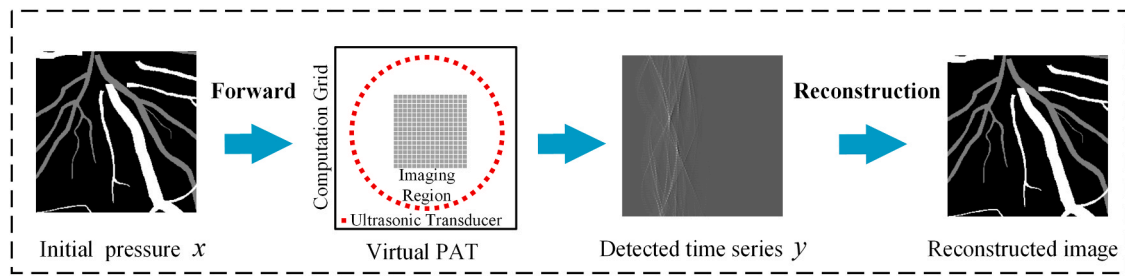
Fig.3. Overall flow chart for the generation of simulation dataset using a virtual PAT.

Eq. (7):

$$f(x, t) = 0, g(t) = \sqrt{d[\sigma^2(t)]/dt}, \tag{7}$$

where $\sigma(t) > 0$ is a monotonically increasing function in noise scale.

The training process of the sparse reconstruction of PAT proposed in this paper is shown in the upper part of Fig. 2. In the training process, to solve Eq. (6), the score function $\nabla_x \log p_t(x)$ needs to be known for all time steps. However, the true $\nabla_x \log p_t(x)$ is unknown, it can be estimated by training a scoring network $S_\theta(x_t, t)$. The unknown $\nabla_x \log p_t(x)$ can be replaced with $\nabla_x \log p_t(x_t|x_0)$ using the denoising score matching [65], where $\nabla_x \log p_t(x_t|x_0)$ is the gradient of the Gaussian perturbation kernel centered at x_0. During the denoising score matching training, the parameters of the score network $S_\theta(x_t, t)$ are optimized according to Eq. (8):

$$\theta^* = \arg\min_\theta \mathbb{E}_t \left\{ \lambda(t) \mathbb{E}_{x_0} \mathbb{E}_{x_t|x_0} \left[\|S_\theta(x_t, t) - \nabla_{x_t} \log p_t(x_t|x_0)\|_2^2 \right] \right\}, \tag{8}$$

where $\mathbb{E}_t \{\lambda(t) \mathbb{E}_{x_0} \mathbb{E}_{x_t|x_0} [\|S_\theta(x_t, t) - \nabla_{x_t} \log p_t(x_t|x_0)\|_2^2]\}$ can be regarded as the loss function. Once the score network is trained by Eq. (8), sparse reconstruction in PAT can be achieved by solving the reverse SDE using approximate conditions $S_\theta(x_t, t) \simeq \nabla_x \log p_t(x_t)$. Therefore, the reverse SDE shown in Eq. (6) can be rewritten to Eq. (9):

$$dx = -d[\sigma^2(t)]S_\theta(x_t, t) + \sqrt{\frac{d[\sigma^2(t)]}{dt}} d\bar{w}, \tag{9}$$

2.2.2. Sparse reconstruction of PAT based on diffusion model

The forward process of photoacoustic imaging can be expressed by Eq. (4). The reconstruction of PAT is to use sparse-view PA data y to reconstruct the initial pressure x through an optimization algorithm. Hence, the reconstruction of PAT can be transformed into an optimization problem based on the least-square method shown in Eq. (10):

$$x = \arg\min_x \|Ax - y\|_2^2, \tag{10}$$

where $\|Ax - y\|_2^2$ is the data consistency term. The gradient descent (GD) method can be used for iterative solution, as shown in Eq. (11):

$$x_{i-1} = x_i - \alpha A^*(Ax_i - y), \tag{11}$$

where α is the iterative step size, A^* is the adjoint operator of A, and $A^*(Ax_i - y)$ is the gradient of the data consistency term. It is worth noting that as the number of iterations increases, the subscript i varies in the direction of decreasing since the step t in the reconstruction part of the diffusion model is from T to 0, as shown in Fig. 2. To reduce the ill-posed problem and achieve the optimal solution, a regularization term can be introduced as a constraint, then the optimization problem (Eq. (10)) can be transformed to Eq. (12):

$$x = \arg\min_x \|Ax - y\|_2^2 + \lambda R(x), \tag{12}$$

where $R(x)$ is the regularization term that contains the prior information

of the target image, λ is the regularization parameter. A well-designed regularization method can lead to high-quality reconstruction. Therefore, by incorporating a differentiable regularization term, the iterative solution for Eq. (12) can be expressed as Eq. (13).

$$x_{i-1} = x_i - \alpha \left[A^*\left(Ax_i - y\right) + \lambda \frac{\partial R(x_i)}{\partial x} \right], \tag{13}$$

Common regularization methods include Tikhonov regularization [27] and TV regularization [28]. Here, learned regularization was adopted in this work, the regularization term $R(x)$ can be expressed as the probability density function $\log p_t(x)$, and the new regularization part was included in the output of the diffusion model network (the regularization coefficient λ is implicit). For the reconstruction problem in PAT, an alternating optimization algorithm is used for the minimization of the decoupling of the prior information term and the data consistency term. The details for the alternating optimization algorithm are further given by Eq. (14), which is an approximate expression of Eq. (13), the reconstruction problem in PAT can be converted to an optimization problem that two sub-problems are alternately updated.

$$\begin{cases} \widehat{x}_i = x_i - \lambda\alpha S_\theta \\ x_{i-1} = \widehat{x}_i - \alpha A^*(A\widehat{x}_i - y) \end{cases} \tag{14}$$

The first equation of Eq. (14) is used for data generation based on reverse SDE ($S_\theta \simeq \nabla_x \log p_t(x)$) that includes the regularization part in the process of iterative generation. The second equation of Eq. (14) is an iterative equation, which is obtained by solving the data consistency term through gradient descent method. In the first term, the generative model is responsible for estimating the complex prior data distribution from the target image dataset. For the solution of the reverse SDE (Eq. (9)) in the first term, the Euler discretization method [57] can be employed. According to the prediction-correction (PC) concept, a PC sampler is introduced to rectify errors in the evolution of the reverse SDE. In the diffusion model, the preliminary prediction is obtained by numerically solving the reverse SDE, and then the direction of gradient ascent is corrected using the Langevin Markov Chain Monte Carlo algorithm [66]. After each PC sampling update, a data consistency is necessary. Further elaborate on the above process, the iterative reconstruction in PAT based on the diffusion model consists of two steps: prediction and correction, as depicted in the lower part of Fig. 2:

Step 1: In the prediction step, Eq. (15) is used for data prediction, the target image $\widetilde{x}_i$ is generated from the prior distribution learned.

$$\widetilde{x}_i = x_i + \left(\sigma_{i+1}^2 - \sigma_i^2 \right) S_\theta\left(x_i, \sigma_{i+1} \right) + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} z, \tag{15}$$

where σ_i is the noise scale, $i = N - 1, \cdots, 1, 0$ is the number of discretization steps for the reverse SDE (i.e., the number of iterations), and $z \sim \mathbb{N}(0, 1)$ is Gaussian white noise with zero mean (standard normal).

Step 2: In the correction step, the correction algorithm shown in Eq. (16) is used to correct the gradient rising direction.

$$\widehat{x}_i = \widetilde{x}_i + \varepsilon_i S_\theta\left(\widetilde{x}_i, \sigma_{i+1} \right) + \sqrt{2\varepsilon_i} z, \tag{16}$$

After the prediction and correction, the gradient descent method of Eq. (17) is used to constrain the result. Eq. (17) is called the data consistency step, which can effectively enhance the consistency of data.

$$x_{i-1} = \widehat{x}_i - \alpha A^*(A\widehat{x}_i - y) \tag{17}$$

During the reconstruction phase, high-quality sparse reconstruction of PAT is achieved by using the input noise and the sparse-view photoacoustic signal detected by the ultrasonic transducer, and iterative updates is performed under the constraints of the data consistency term.

The pseudo-code of the reconstruction algorithm in this paper is shown in Algorithm 1, and the algorithm contains two loops: (1) In the outer loop, the sparse-view PA data is fed into network to achieve a prediction of data distribution, The number of iterations N of the outer loop is determined by the number of discrete steps of the reverse SDE. (2) The inner loop is corrected by annealing Langevin iteration. During the whole iteration process, the data prior term (regularization term) of the PC process and data consistency term are updated.

Algorithm 1. Training for prior learning.

Fig. 3, the augmented data is imported into the computational region of the virtual PAT. The DAS algorithm is used to reconstruct artifact-free simulation data at full view required for training. The training set consists of 1200 images and the test set consists of 300 images. (2) Experimental dataset: The experimental dataset consists of a phantom dataset and animal in vivo data. 1) Phantom dataset: The phantom dataset is a circular phantom dataset [36], which contains 469 reconstructed images under full-view scanning (512 projections) and other projections (8, 16, 32, 64 and 128 projections), respectively. 2) The in vivo experimental data of mice's abdomens is from the public dataset MOST-Abdomen [36], which contains 274 reconstructed images under full-view scanning (512 projections) and other projections (8, 16, 32, 64 and 128 projections), respectively.

The training of the model utilizes the Adaptive Moment Estimation (Adam) optimization method with a learning rate of 2×10^{-4}. During the reconstruction stage, the number of iterations is set to 1000. The image size is set to 256×256, and the pixel values are normalized. The data distribution is perturbed using Gaussian noise with a noise value in the range of 0.01–300. The implementation of proposed method is based

---

Algorithm 1

Training for prior learning

Dataset: Photoacoustic dataset: x

1: Training $S_\theta(x_t, t) \simeq \nabla_x \log p_t(x_t)$

2: Output: Trained $S_\theta(x_t, t)$

Iterative reconstruction

Setting: Score network S_θ, Noise iteration step ε_i, Noise scale σ_i, Number of discretization steps for the reverse SDE N, Number of corrector steps M

1: $x_N \sim \mathbb{N}(0, \sigma_{\max}^2 I)$

2: $z \sim \mathbb{N}(0,1)$

3: For $i = N\text{-}1$ to 0 **do (Outer loop)**

4: Update $\tilde{x}_i = x_i + (\sigma_{i+1}^2 - \sigma_i^2)S_\theta(x_i, \sigma_{i+1}) + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} z$ **(Prediction)**

5: **For** $j = 1$ to M **do (Inner loop)**

6: Update $\hat{x}_i \leftarrow \tilde{x}_{i,j} + \varepsilon_i S_\theta(\tilde{x}_{i,j}, \sigma_{i+1}) + \sqrt{2\varepsilon_i} z$ **(Correction)**

7: **End for**

8: Update $x_{i-1} = \hat{x}_i - \alpha A^*(A\hat{x}_i - y)$ **(Gradient descent)**

9: End for

Return x_0

2.3. Dataset acquisition and network parameter setting

The datasets include simulation dataset and experimental dataset. (1) Simulation dataset: To obtain a sufficient number of photoacoustic images under full-view sampling, a virtual PAT was constructed based on k-Wave [58], as illustrated in Fig. 3. The platform realizes the forward process and reverse reconstruction of PAT under any detection view. The size of the entire computational area is set to 50 mm × 50 mm with grid of 440 × 440 pixels. Ultrasonic transducers with a BW of 66 % having a central frequency of 2.25 MHz are placed at different views in a circularly equidistant manner at a radius of 21.6 mm (the number of ultrasonic transducers is determined by the experimental requirements, for the acquisition of ground truth, it is set to 512). The sound velocity is set to 1500 m/s, and the surrounding medium is water with a density of 1000 kg/m³. The public retinal vessel datasets RAVIR and DRIVE [67] are augmented by cropping and rotating transformation. As shown in

on the PyTorch framework and primarily implemented in the Python environment. In this work, the computation is performed on a graphical processing unit (GPU; GeForce RTX 2080Ti).

3. Results

3.1. Blood vessels simulation data

During the test phase, the simulated dataset is imported into the virtual PAT depicted in Fig. 3, and the sparse-view photoacoustic signal is used as the input (data fidelity) of the network. The reconstruction results using conventional DAS method, compressed sensing method, U-Net method, and the proposed diffusion model-based method were compared. Fig. 4(a) and (b) show the process of reconstruction of the simulated blood vessel using the proposed method and GD method, respectively, under 32 projections. The reconstruction iterates from a
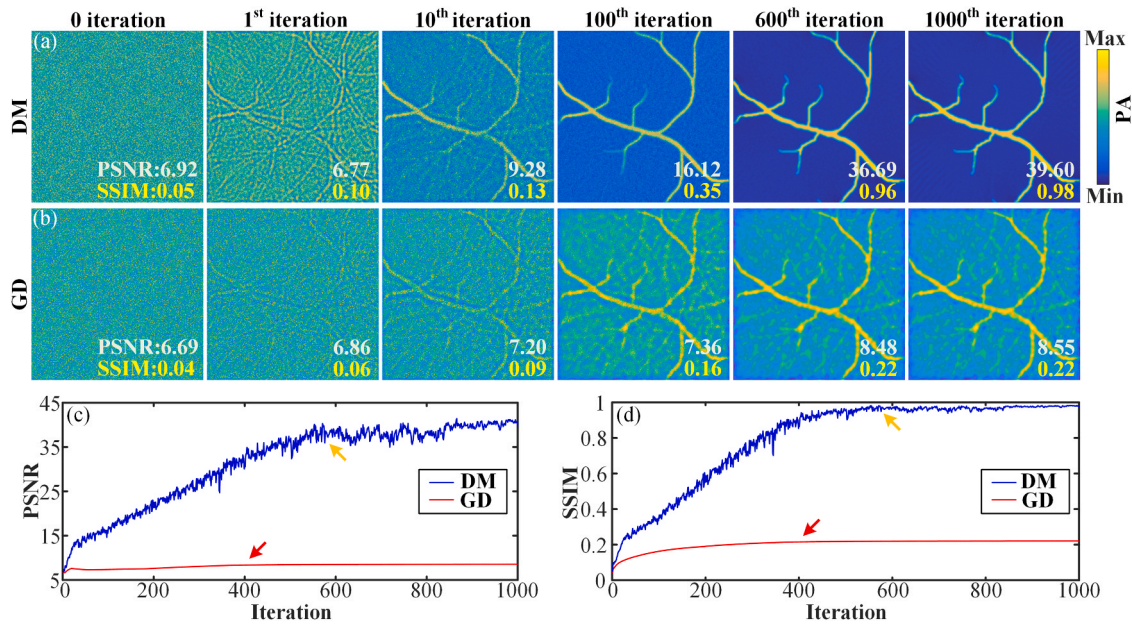
Fig.4. The process of reconstruction of the simulated blood vessels using the proposed method under 32 projections. (a) and (b) Reconstruction results under different iterations using the proposed method and GD method, respectively. (c) and (d) The variation of PSNR and SSIM with the iteration, respectively. DM, diffusion model; GD, gradient descent; PA, photoacoustic amplitude.

noise image, and the blood vessels begin to emerge from the noise at the first iteration, the general outline of the blood vessels becomes visible. With the iterations increase, the noise is gradually eliminated, and the blood vessels exhibit improved definition at the 10th iteration compared with earlier iterations. The result of the proposed method is slightly better that of the GD method. At the 600th iteration, it can be observed that the blood vessels have been essentially reconstructed using the diffusion model-based method, and the image quality has been further improved. However, the reconstruction result of the GD method still has severe artifacts. Fig. 4(c) and (d) show the variation of PSNR and SSIM with the iteration. For the proposed method, the PSNR and SSIM rise rapidly when the number of iterations is around 0–600. At the 600th iteration, the PSNR and SSIM are improved to 36.69 dB and 0.96, respectively (indicated by the yellow arrows in Fig. 4(c) and (d)). As the number of iterations increases further, the PSNR and SSIM tend to be stable (the PSNR tends to 39 dB, the SSIM tends to 0.98). However, for the GD method, the PSNR and SSIM increase slowly with the number of iterations, and remain at a low value (the PSNR tends to 8.36 dB, the SSIM tends to 0.22) after about 400 iterations indicated by the red arrows in Fig. 4(c) and (d). The results show that the proposed method based on diffusion model has good performance in sparse-view reconstruction.

Fig. 5 shows the reconstruction results of different methods for simulated blood vessels under different projection. Fig. 5(a)–(c) are the reconstruction results using DAS method under 128, 64 and 32 projections, respectively. It can be observed that the reconstruction results by the DAS method exhibit noticeable artifacts under different projections. As the number of projections decreases, the artifacts become more pronounced, resulting in a significant degradation of the reconstruction. Fig. 5(d) is the corresponding ground truth. Fig. 5(e)–(g) in the second row are the reconstruction results using compressed sensing method [68–70]. The compressed sensing method is achieved by k-Wave (the configuration of the compressed sensing simulation model is the same as the virtual PAT). The measurement matrix in the simulation was set as the Hadamard matrix [68], and the reconstruction algorithm is OMP algorithm [69,70]. The number of measurements (i.e., the number of projections) is set to 32, 64 and 128, respectively. It can be seen that under 128 measurements, the reconstruction quality of compressed

sensing method is improved compared with DAS method, and the artifacts can be removed significantly. As the number of measurements decreases, the blood vessels cannot be completely reconstructed under 64 measurements (indicated by the yellow arrows in the Fig. 5). Under extreme sparse projection conditions (e.g., 32 measurements), the blood vessels can hardly be reconstructed. Fig. 5(i)–(k) in the third row are the reconstruction results using U-Net method under 128, 64 and 32 projections, respectively. Fig. 5(l) is the corresponding ground truth. The image quality of the reconstruction results using U-Net has been greatly improved compared with that of DAS and compressed sensing method. U-Net has a better performance under 128 projections and 64 projections, and the reconstruction results have almost no artifacts, as shown in Fig. 5(i) and (j). However, artifacts still exist under extremely sparse conditions (e.g., 32 projections), as shown in Fig. 5(k). Figs. 5 (m)–(o) show the reconstruction results of the diffusion model-based method under 128, 64 and 32 projections, respectively. Fig. 5(p) is the corresponding ground truth. It can be observed that the proposed method performs remarkably performance under different projections. Compared with the U-Net, the reconstructed image obtained by the diffusion model-based method has fewer artifacts and higher reconstruction quality. Fig. 5(q) and (r) are the close-up images indicated by the white dashed rectangles 1 and 2, respectively. It can be seen that the reconstruction results using the diffusion model method have no significant artifacts (indicated by the white arrows) compared with DAS. And more details are contained than the reconstruction results using U-Net and compressed sensing method. Fig. 5(s)–(x) show the signal distribution indicated by the white dashed lines in Fig. 5(q) and (r) under 128, 64 and 32 projections, respectively. It can be observed that the signal distribution of the method based on the diffusion model is closer to the ground truth, which verifies the superiority of the proposed method in the sparse reconstruction of PAT. Quantitatively, the proposed method achieves a PSNR of 45.09 dB and an SSIM of 0.99 under 128 projections, with an improvement of 9 dB in PSNR and 0.03 in SSIM compared with U-Net. Under 64 projections, the proposed method can reach 41.10 dB and 0.98 in PSNR and SSIM, respectively, with an improvement of 8.7 dB and 0.04 compared with U-Net, which suggests that the proposed method outperforms U-Net under sparse views. When the projection is further reduced to 32 projections, the proposed method
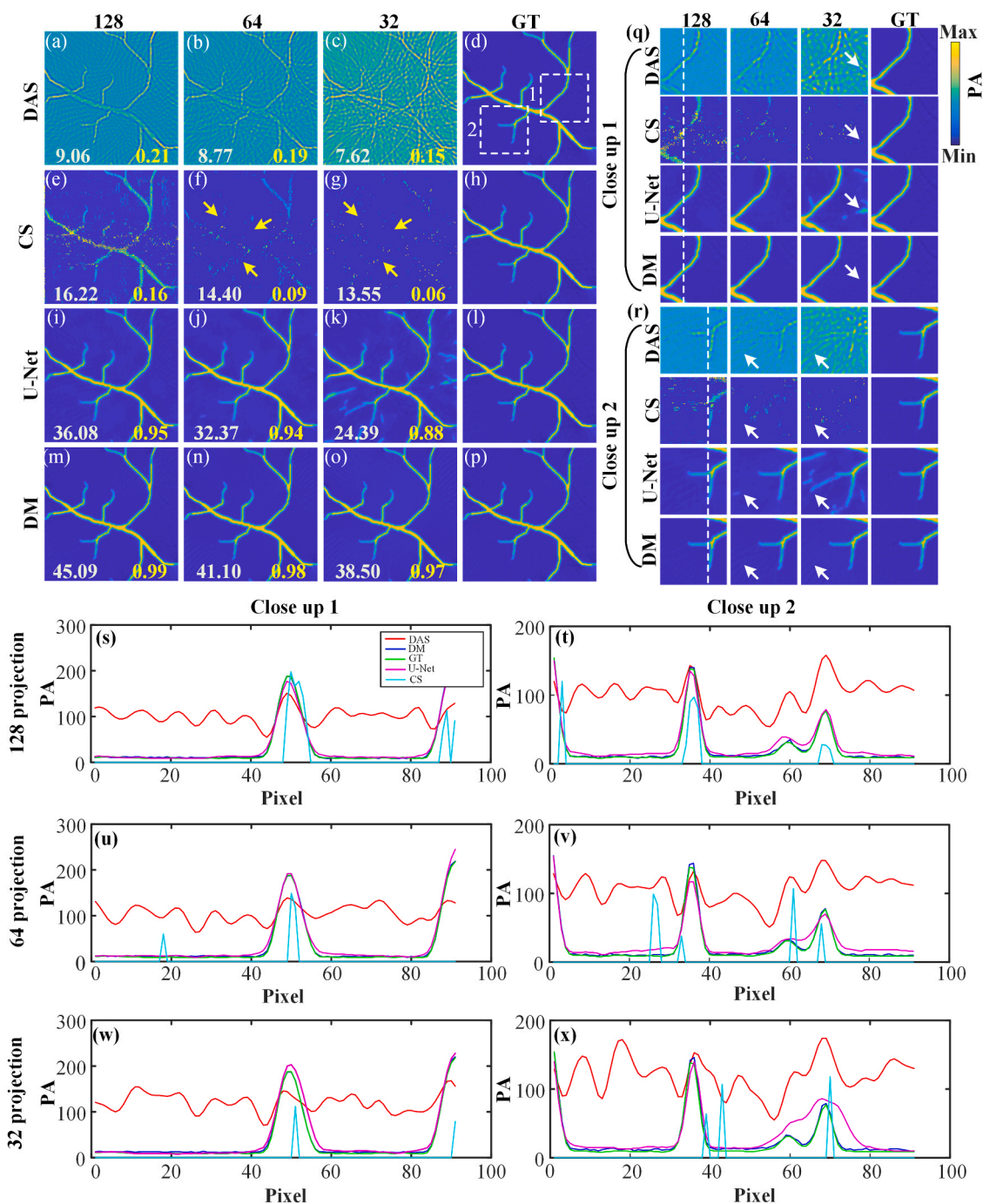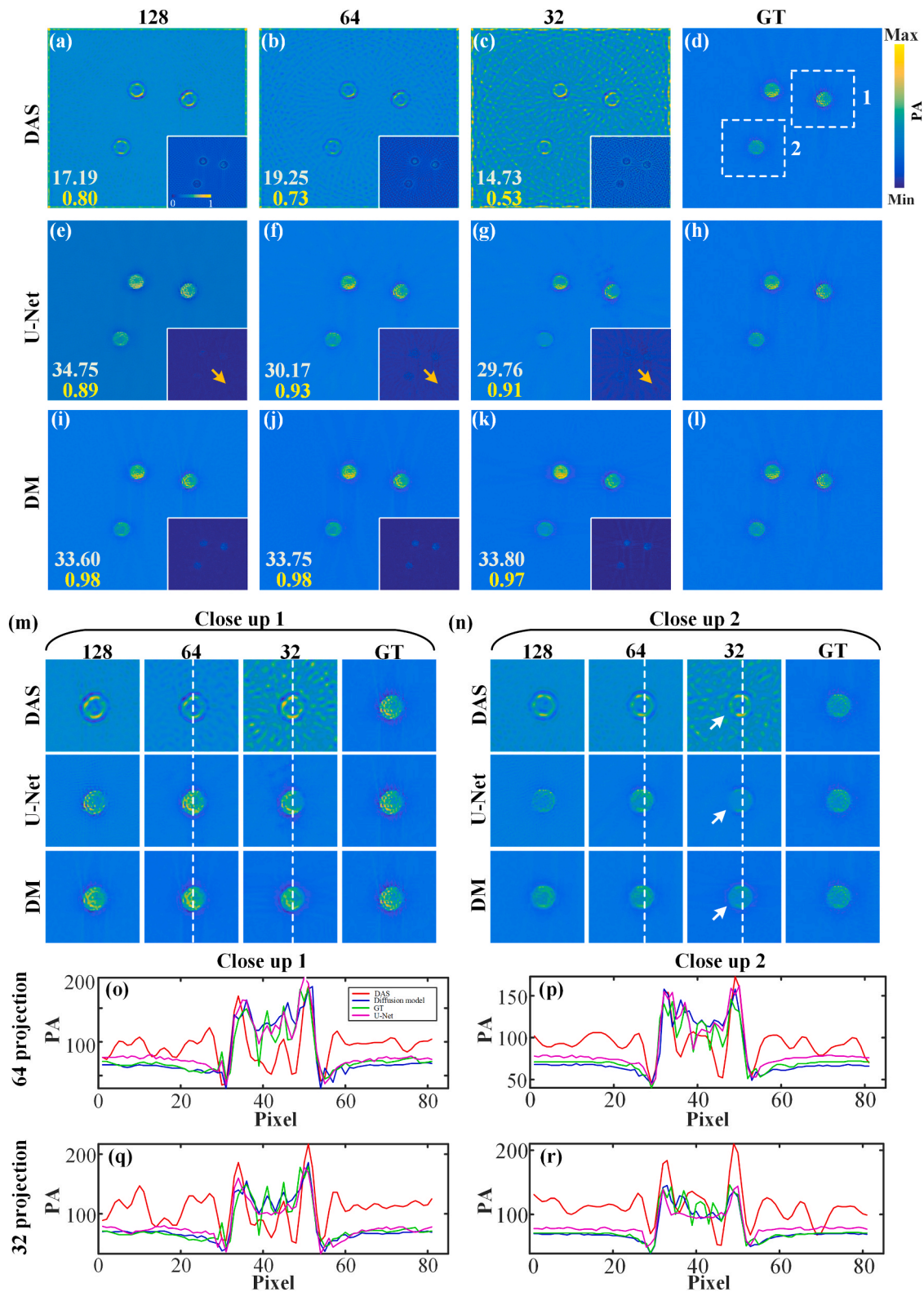
Fig.5. The reconstruction results of different methods for simulated blood vessels under different projection. (a)-(c) are the reconstruction results using DAS under 128, 64 and 32 projections, respectively. (d) is the ground truth. (e)-(g) the reconstruction results using compressed sensing method under 128, 64 and 32 measurements, respectively. (h) is the ground truth. The yellow arrows in (f) and (g) denote the vessels which cannot be completely reconstructed. (i)-(k) the reconstruction results using U-Net under 128, 64 and 32 projections, respectively. (l) is the ground truth. (m)-(o) are the reconstruction results using proposed method under 128, 64 and 32 projections, respectively. The white numbers below the images represent the PSNR, and the yellow numbers represent the SSIM. (l) is the ground truth. (q) and (r) are the close-up images indicated by the white dashed rectangles 1 and 2, respectively. (s)-(x) are the signal distribution indicated by the white dashed lines in the (q) and (r) under 128, 64 and 32 projections, respectively. CS, compressed sensing; DM, diffusion model; GT, ground truth; PA, photoacoustic amplitude.

still achieves a PSNR of 38.50 dB and an SSIM of 0.97, which shows a remarkable improvement of 14.1 dB and 0.09 in PSNR and SSIM, respectively, compared with U-Net. These results demonstrate the proposed method can achieve higher-quality reconstruction even under extreme sparse projection (e.g., 32 projections).

3.2. Phantom experiment

To validate the effectiveness of the proposed method in the experimental data, circular phantom was selected for the performance test. The results are shown in Fig. 6, and Fig. 6(a)–(c) are the reconstruction results using the DAS method under 128, 64 and 32 projections, respectively. As the number of projections increases, the outline of the

Fig.6. The reconstruction results of the circular phantom. (a)-(c) are the reconstruction results of DAS method under 128, 64 and 32 projections, respectively. (e)-(g) are the results of the U-Net method under 128, 64 and 32 projections, respectively. (i)-(k) are the reconstruction results of the proposed method under 128, 64 and 32 projections, respectively. The bottom right corner shows the error maps corresponding to each reconstruction result. The white numbers below the images represent the PSNR, and the yellow numbers represent the SSIM. (d), (h) and (l) represent the same ground truth. (m) and (n) are the close-up images indicated by the white dashed rectangles 1 and 2, respectively. (o)-(r) are the signal distribution along the dashed lines in (m) and (n), respectively.
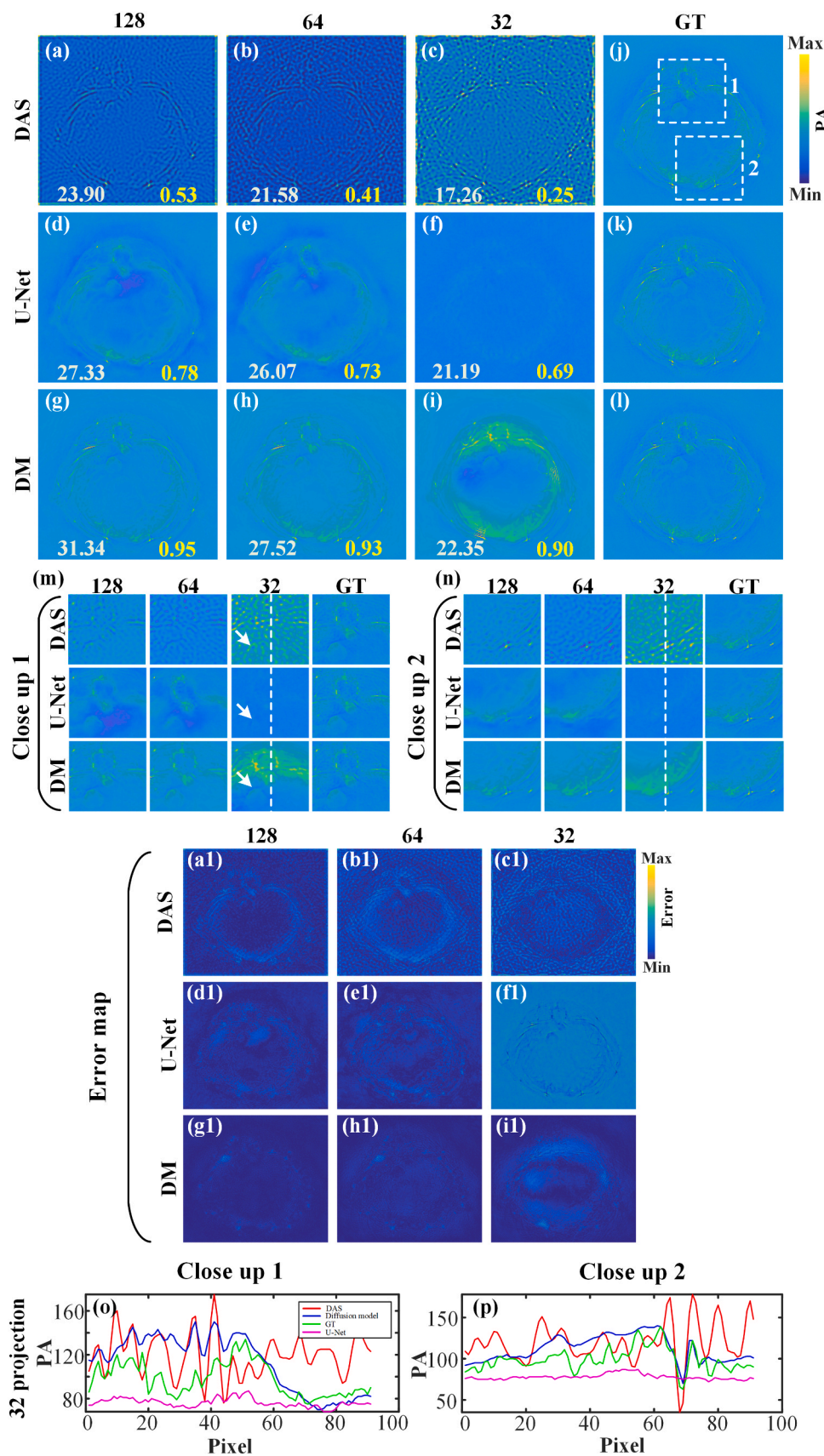
Fig.7. The reconstruction results of in vivo mouse abdomen. (a)-(c) are the reconstruction results of DAS under 128, 64 and 32 projections, respectively. (d)-(f) are the de-artifact images of U-Net under 128, 64 and 32 projections, respectively. (g)-(i) are the reconstructed images of the proposed method under 128, 64 and 32 projections, respectively. (j)-(l) are the same ground truths. The white numbers in the bottom left corner represent PSNR, and the yellow numbers represent SSIM. (m) and (n) are the close-up images indicated by the white dashed rectangles 1 and 2, respectively. (a1)-(i1) are the error maps corresponding to (a)-(i), respectively. (o) and (p) are the signal distribution along the white dashed lines in (m) and (n) under 32 projections, respectively.

circular phantom becomes clearer, however, there are still serious arti-facts in the reconstructed image. The lower right corner of the figures show the error maps between the ground truth image and the recon-struction results of the DAS under different projection, which shows that the reconstruction results are quite different from the ground truth. Fig. 6(d) is the ground truth, which is obtained under the full view (512 projections). Fig. 6(e)–(g) show the artifact-reduced images using the U-Net method under 128, 64 and 32 projections, respectively. Fig. 6(h) is the ground truth. The results show that the U-Net method can effectively remove most of the artifacts. However, the corresponding error map shows that there are still some artifacts around the circular phantom (indicated by the yellow arrows in Fig. 6(e)–(g)). Fig. 6(i)–(k) are the reconstructed images using the proposed method under 128, 64 and 32 projections, respectively. The proposed method can effectively remove the artifacts in the image. Even under sparse conditions (e.g., 32 pro-jections), the proposed method achieves superior performance and produces higher-quality reconstructions compared with the other two methods. Moreover, as the number of projections increases, the imaging performance can be further improved. Figs. 6(m) and (n) show the close-up images indicated by the white dashed rectangles 1 and 2, respec-tively. The images reconstructed by the diffusion model exhibit more accurate details compared with the other two methods. It can be observed that in the region indicated by the white arrow in Fig. 6(n), the proposed method is able to reconstruct more details. Fig. 6(o)–(r) show the signal distribution along the dashed lines in Fig. 6(m) and (n), which shows that the signal distribution of the reconstructed result by the proposed method is closer to the ground truth. Under extremely sparse condition (32 projections), the signal distribution of the proposed method is still closer to the ground truth. Quantitative analysis was carried out to demonstrates the superiority of the proposed method over the U-Net. Under 64 projections, the PSNR and SSIM of the proposed method rise to 33.75 dB and 0.98, respectively, and the PSNR and SSIM are increased by 3.5 dB and 0.05, respectively, compared with U-Net. Additionally, the proposed method performed better even under extremely sparse conditions. Under 32 projections, the PSNR and SSIM are 33.80 dB and 0.97, respectively, with an improvement of 4 dB and 0.06 compared with U-Net. Experimental results demonstrate the effectiveness of the proposed method in sparse reconstruction of experimental data.

3.3. In vivo experimental data

To further verify the effectiveness of the proposed method in in vivo data, the reconstruction of mice's abdomens using the conventional method, the U-Net method and the proposed method were compared, respectively, as shown in Fig. 7. Fig. 7(a)–(c) are the reconstructed im-ages of DAS method under 128, 64 and 32 projections, respectively. Fig. 7(d)–(f) are the de-artifact images of U-Net under 128, 64 and 32 projections, respectively. Fig. 7(g)–(i) are the reconstructed images of the proposed method under 128, 64 and 32 projections, respectively. Fig. 7(j)–(l) are the same ground truths. It is evident that the image reconstructed by the DAS method contains significant artifacts, while the image reconstructed by the U-Net method under the sparse 32 projections suffers from substantial loss of intricate details (Fig. 7(f)). These results can also be observed in the close-up images indicated by the white dashed boxes 1 and 2 in Fig. 7(j), as shown in Fig. 7(m) and (n). The positions indicated by the white arrows in Fig. 7(m) show that the reconstruction using DAS has more serious artifacts, the details of the reconstruction results based on U-Net are lost to a certain extent, while the diffusion model method could reconstruct a completer and more accurate signal. Fig. 7(a1)–7(i1) are the error maps corresponding to Fig. 7(a)–7(i), which show that the reconstruction of the proposed method is closer to the ground truth. Particularly.

in the error maps obtained under 32 projections, compared with U-Net, the proposed method show less deviation from the ground truth. Fig. 7(o) and 7(p) are the signal distribution along the white dashed lines

in Fig. 7(m) and 7(n) under 32 projections, respectively. The signal reconstructed by the proposed method is closer to the signal of the ground truth. Quantitative analysis shows that the proposed method achieves a PSNR of 27.52 dB and an SSIM of 0.93 under 64 projections. Compared with U-Net, the PSNR and SSIM are increased by 1.5 dB and 0.21, respectively. Under the extreme sparse condition of 32 projections, the proposed method achieves a PSNR of 22.35 dB and an SSIM of 0.90 (while the PSNR and SSIM using the DAS method are 17.26 dB and 0.25, respectively), with an improvement of 1.17 dB and 0.21 over the U-Net. These results demonstrate that the proposed method can perform high-quality reconstruction under extreme sparse conditions, which further verify the validity and superiority of the reconstruction method based on the diffusion model.

4. Conclusion and discussion

In conclusion, a novel model-based sparse reconstruction method for PAT via diffusion model was proposed to address the low-quality recon-struction in conventional standard reconstruction of PAT under sparse view. The sparse reconstruction strategy combines score-based generative model and model-based iterative reconstruction method. A score-based diffusion model is designed for learning the prior information of the data distribution. The learned prior information is utilized as a constraint for the data consistency term of an optimization problem based on the least-square method in the model-based iterative reconstruction, aiming to achieve the optimal solution. 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 dem-onstrates the superiority of the proposed method. For the in vivo experi-mental data, under 64 projections, the SSIM and PSNR of the proposed method rise to 0.93 and 27.52 dB, respectively, and the SSIM and PSNR are increased by 0.2 and 1.5 dB, respectively compared with U-Net method. Additionally, the proposed method performed better under extremely sparse conditions. Under 32 projections, the SSIM and PSNR are 0.90 and 22.35 dB, respectively, with an improvement of 0.21 and 1.2 dB compared with U-Net. Compared with the conventional delay-and-sum method, the proposed method achieves an improvement of 0.65 ($\sim$ 260 %) in SSIM and 5.1 dB ($\sim$ 30 %) in PSNR. These results demonstrate the proposed method can achieve higher-quality reconstruction even under extreme sparse projection (e.g., 32 projections).

The training process of the diffusion model is to estimate the un-known score function by training the score network. In the training process, it is necessary to continuously add noise to the data in the training set and learn the data distribution, which takes a lot of time. The training time of the proposed method is mainly related to the configu-ration of the graphical processing unit used in the experiment, the number and size of the training dataset. During the training phase, one checkpoint is saved for every 10,000 epochs completed, which takes about 50 min. Ten checkpoints were obtained in the experiment and the best training model (i.e., the score network) was selected. Therefore, the required time for training is about 8 h. The reconstruction is an iterative process, the reconstruction time is associated with the number of iter-ations. It can be seen from Figs. 4(c) and 4(d), PSNR and SSIM reach a stable state at about the 600th iteration. Therefore, the reconstruction time is about 1 h (the time required for each iteration is about 6 s). In terms of efficiency, the speed and quality of reconstruction need to be considered. The DAS method is a conventional analytical algorithm without training and iteration. For the U-Net method, the network is a data-driven end-to-end network, which can directly use the pre-trained model to remove artifacts without iteration. Therefore, DAS and U-Net method have advantages in terms of reconstruction speed. In terms of reconstruction quality, the proposed method has obvious advantages over the other two methods, as shown in Figs. 4–7. For the DAS method, to achieve comparable reconstruction quality, more projection data need to be obtained, which will significantly limit the imaging speed and increase the system cost. U-Net method needs more paired datasets to

obtain higher-quality reconstruction results, which leads to further increase in training time.

The noise used in this diffusion model is Gaussian white noise with zero mean. For noise that does not satisfy the Gaussian distribution, it can be realized by other types of diffusion models [71,72]. For example, the cold diffusion [71] proposed by Bansal et al. can work on not only Gaussian noise, but also arbitrary noise or even noiseless/cold image transforms. Therefore, it's possible to work on noise that does not follow a Gaussian distribution. The implementation of this method has the potential to effectively decrease both the imaging time and cost of PAT, and holds the potential to significantly expand the application range of PAT in the acquisition of physiological and pathological process.

Funding

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

Data will be made available on request.

Acknowledgments

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] Y. Wu, J. Kang, W.G. Lesniak, A. Lisok, H.K. Zhang, R.H. Taylor, M.G. Pomper, E. M. Boctor, System-level optimization in spectroscopic photoacoustic imaging of prostate cancer, Photoacoustics 27 (2022), 100378.

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

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

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

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

[13] J. Xia, J. Yao, L.V. Wang, Photoacoustic tomography: principles and advances, Electro Waves (Camb.) 147 (2014) 1–22.

[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, Single-breath-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 small-animal whole-body dynamics at high spatiotemporal resolution, Nat. Biomed. Eng. 1 (5) (2017), 0071.

[19] J. Gamelin, A. Aguirre, A. Maurudis, F. Huang, D. Castillo, L.V. Wang, Q. Zhu, Curved array photoacoustic tomographic system for small animal imaging, J. Biomed. Opt. 13 (2) (2008), 024007.

[20] R.A. Kruger, C.M. Kuzmiak, R.B. Lam, D.R. Reinecke, S.P. Del Rio, D. Steed, Dedicated 3D photoacoustic breast imaging, Med. Phys. 40 (11) (2013), 113301.

[21] J. Xia, M.R. Chatni, K. Maslov, Z. Guo, K. Wang, M. Anastasio, L.V. Wang, Whole-body ring-shaped confocal photoacoustic computed tomography of small animals in vivo, J. Biomed. Opt. 17 (5) (2012), 050506.

[22] C. Yeh, L. Li, L. Zhu, J. Xia, C. Li, W. Chen, A. Garcia-Uribe, K.I. Maslov, L.V. Wang, Dry coupling for whole-body small-animal photoacoustic computed tomography, J. Biomed. Opt. 22 (4) (2017), 041017.

[23] Y. Wang, Z. Guo, L.V. Wang, T.N. Erpelding, L. Jankovic, J.L. Robert, G. David, In vivo three-dimensional photoacoustic imaging based on a clinical matrix array ultrasound probe, J. Biomed. Opt. 17 (6) (2012), 061208.

[24] 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.

[25] C. Huang, K. Wang, L. Nie, L.V. Wang, M.A. Anastasio, Full-Wave Iterative Image Reconstruction in Photoacoustic Tomography with Acoustically Inhomogeneous Media, IEEE Trans. Med. Imaging 32 (6) (2013) 1097–1110.

[26] K. Wang, R. Su, A.A. Oraevsky, M.A. Anastasio, Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography, Phys. Med. Biol. 57 (2012) 5399–5423.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] H. Lan, J. Zhang, C. Yang, F. Gao, Compressed sensing for photoacoustic computed tomography based on an untrained neural network with a shape prior, Biomed. Opt. Express 12 (12) (2021) 7835–7848.

[32] X. Zhang, F. Ma, Y. Zhang, J. Wang, C. Liu, J. Meng, Sparse-sampling photoacoustic computed tomography: deep learning vs. compressed sensing, Biomed. Signal Process Control 71 (2022), 103233.

[33] X. Liu, D. Peng, X. Ma, W. Guo, Z. Liu, D. Han, X. Yang, J. Tian, Limited-view photoacoustic imaging based on an iterative adaptive weighted filtered backprojection approach, Appl. Opt. 52 (15) (2013) 3477–3483.

[34] C. Wang, H. Qin, G. Lai, G. Zheng, H. Xiang, J. Wang, D. Zhang, Automated classification of dual channel dental imaging of auto-fluorescence and white light by convolutional neural networks, J. Innov. Opt. Health Sci. 13 (04) (2020), 2050014.

[35] S. Xie, X. Zheng, Y. Chen, L. Xie, J. Liu, Y. Zhang, J. Yan, H. Zhu, Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction, Sci. Rep. 8 (1) (2018), 6700.

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

[37] H. Lan, C. Yang, D. Jiang, and F. Gao, Reconstruct the photoacoustic image based on deep learning with multi-frequency ring-shape transducer array, in International Engineering in Medicine and Biology Conference (EMBC2019), pp. 7115–7118.

[38] 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.

[39] 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.

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

[41] H. Shan, G. Wang, Y. Yang, Accelerated correction of reflection artifacts by deep neural networks in photo-acoustic tomography, Appl. Sci. 9 (13) (2019) 2615.

[42] S. Guan, K.T. Hsu, M. Eyassu, and P.V. Chitnis, Dense dilated UNet: deep learning for 3D photoacoustic tomography image reconstruction, arXiv, arXiv:2104.03130 (2021).

[43] Y.E. Boink, S. Manohar, C. Brune, A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation, IEEE Trans. Med. Imaging 39 (1) (2020) 129–139.

[44] T. Tong, W. Huang, K. Wang, Z. He, L. Yin, X. Yang, S. Zhang, J. Tian, Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data, Photoacoustics 19 (2020), 100190.

[45] D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, L. Maier-Hein, Reconstruction of initial pressure from limited view photoacoustic images using deep learning, Proc. SPIE 10494 (2018) 104942S.

[46] M.W. 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.

[47] T. Wang, M. He, K. Shen, W. Liu, C. Tian, Learned regularization for image reconstruction in sparse-view photoacoustic tomography, Biomed. Opt. Express 13 (11) (2022) 5721–5737.

[48] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, S. Arridge, Model-based learning for accelerated, limited-view 3-d photoacoustic tomography, IEEE Trans. Med. Imaging 37 (6) (2018) 1382–1393.

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

[50] T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive growing of gans for improved quality, stability, and variation, arXiv, arXiv:1710.10196 (2017).

[51] F. Moreno-Pino, P.M. Olmos, A. Artés-Rodríguez, Deep autoregressive models with spectral attention, Pattern Recognit. 133 (2023), 109014.

[52] S. Bond-Taylor, A. Leach, Y. Long, C.G. Willcocks, Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models, IEEE Trans. Pattern Anal. Mach. Intel. 44 (2021) 7327–7347.

[53] D.P. Kingma, P. Dhariwal, Glow: Generative flow with invertible 1×1 convolutions, J. Inf. Process. Syst. 31 (2018).

[54] C. Doersch, Tutorial on variational autoencoders. arXiv, arXiv: 1606.05908 (2016).

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

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

[57] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon, and B. Poole, Score-based generative modeling through stochastic differential equations, arXiv, arXiv: 2011.13456 (2020).

[58] 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.

[59] B. Guan, C. Yang, L. Zhang, S. Niu, M. Zhang, Y. Wang, W. Wu, and Q. Liu, Generative Modeling in Sinogram Domain for Sparse-view CT Reconstruction, arXiv, arXiv:2211.13926 (2022).

[60] A. Graves, Generating sequences with recurrent neural networks, arXiv, arXiv: 1308.0850 (2013).

[61] D.P. Kingma and M. Welling, Auto-encoding variational bayes, arXiv, arXiv: 1312.6114 (2013).

[62] K.C. Tezcan, C.F. Baumgartner, R. Luechinger, K.P. Pruessmann, E. Konukoglu, MR image reconstruction using deep density priors, IEEE Trans. Med. Imaging 38 (7) (2018) 1633–1642.

[63] T. Taskaya-Temizel, M.C. Casey, A comparative study of autoregressive neural network hybrids, Neural Netw. 18 (5–6) (2005) 781–789.

[64] R. Salakhutdinov and H. Larochelle, Efficient learning of deep Boltzmann machines, in Proceedings of the thirteenth international conference on artificial intelligence and statistics (AISTATS2010), pp. 693–700.

[65] P. Vincent, A connection between score matching and denoising autoencoders, Neural Comput. 23 (7) (2011) 1661–1674.

[66] G. Parisi, Correlation functions and computer simulations, Nucl. Phys. 180 (3) (1981) 378–384.

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

[68] S. Salari, F. Chan, Y.T. Chan, W. Read, TDOA estimation with compressive sensing measurements and Hadamard matrix, IEEE Aerosp. Electron. Syst. Mag. 54 (6) (2018) 3137–3142.

[69] A. Zhao, J. Liu, Q. He, B. Li, J. Zhang, J. Wei, L. Song, X. Song, Virtual compressed sensing photoacoustic tomography with high-detection efficiency using k-wave, Proc. SPIE 11590 (2021) 106–111.

[70] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.

[71] A. Bansal, E. Borgnia, H.M. Chu, J.S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein,Cold diffusion: Inverting arbitrary image transforms without noise, arXiv, arXiv:2208.09392 (2022).

[72] E. Hoogeboom and T. Salimans,Blurring diffusion models, arXiv, arXiv:2209.05557 (2022).

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.
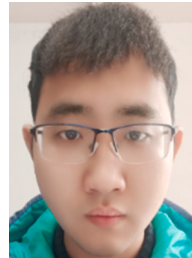
Wenhua Zhong received the bachelor degree in Electronic Information Engineering from Jiangxi Science and Technology Normal University, Nanchang, China. He is currently studying in Nanchang University for master's degree in Electronic Information. His research interests include computer vision, deep learning, photoacoustic microscopy 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.

Xuan Liu received the bachelor degree in Electronic Information Engineering from Pingxiang University, Pingxiang, China. He is currently studying in Nanchang University for master's degree in Information and Telecommunications Engineering. His research interests include single-pixel imaging, deep learning and image processing.

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.

Jiaqing Dong received the bachelor degree in Communication Engineering from Nanchang University, Nanchang, China. He is currently studying in Nanchang University for master's degree in Information and Communication Engineering. His research interests include holographic display and deep learning.

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.