OPEN ACCESS
Journal of Innovative Optical Health Sciences
Vol. 17, No. 3 (2024) 2350028 (18 pages)

DOI: 10.1142/S1793545823500281

World Scientific
www.worldscientific.com

Triple-path feature transform network for ring-array photoacoustic tomography image reconstruction

Lingyu Ma, Zezheng Qin, Yiming Ma and Mingjian Sun*
School of Astronautics, Harbin Institute of Technology
Harbin, Heilongjiang 150000, P. R. China
**sunmingjian@hit.edu.cn*

Received 7 July 2023
Revised 20 September 2023
Accepted 25 September 2023
Published 11 November 2023

Photoacoustic imaging (PAI) is a noninvasive emerging imaging method based on the photoacoustic effect, which provides necessary assistance for medical diagnosis. It has the characteristics of large imaging depth and high contrast. However, limited by the equipment cost and reconstruction time requirements, the existing PAI systems distributed with annular array transducers are difficult to take into account both the image quality and the imaging speed. In this paper, a triple-path feature transform network (TFT-Net) for ring-array photoacoustic tomography is proposed to enhance the imaging quality from limited-view and sparse measurement data. Specifically, the network combines the raw photoacoustic pressure signals and conventional linear reconstruction images as input data, and takes the photoacoustic physical model as a prior information to guide the reconstruction process. In addition, to enhance the ability of extracting signal features, the residual block and squeeze and excitation block are introduced into the TFT-Net. For further efficient reconstruction, the final output of photoacoustic signals uses 'filter-then-upsample' operation with a pixel-shuffle multiplexer and a max out module. Experiment results on simulated and *in-vivo* data demonstrate that the constructed TFT-Net can restore the target boundary clearly, reduce background noise, and realize fast and high-quality photoacoustic image reconstruction of limited view with sparse sampling.

Keywords: Deep learning; feature transformation; image reconstruction; limited-view measurement; photoacoustic tomography.

1. Introduction

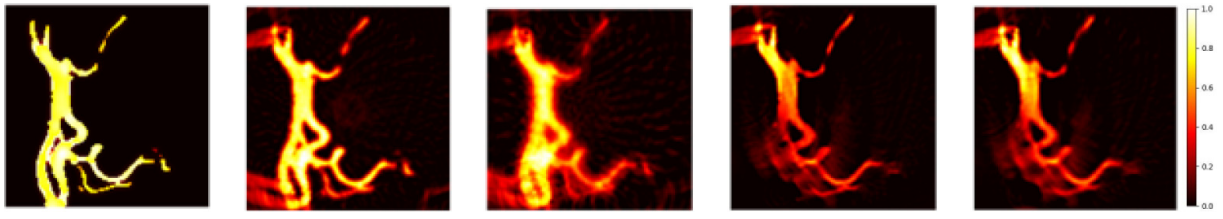
The ring-array photoacoustic tomography (PAT) system is often used to image the whole body of small animals or human organs and tissues. As a noninvasive biomedical imaging method, photoacoustic imaging (PAI) can reveal the optical absorption properties of biological tissue, and combined with molecular probes, the functional properties of tissue can be obtained, which can be used for medical diagnosis and treatment assistance.[1,2] For PAT, the fully sampled photoacoustic signals involve large amount of data which will lead to a

sharp increase in the cost of the acquisition device and the image reconstruction time.[3] Therefore, some annular array systems reduce the cost and improve the reconstruction speed by reducing the number of transducer elements and controlling the sampling angle of view. However, in this setting, it is difficult to meet the requirements of full sampling, which reduces the quality of the reconstructed image. As shown in Fig. 1, when the sampling angle of view is 180° and the number of transducer elements is reduced by 50%, the image reconstruction quality is greatly reduced. Finding a fast and high-quality photoacoustic image reconstruction algorithm to improve the quality of reconstructed images for under-sampled data with limited-view is of great significance for promoting the clinical transformation and application of PAT technology.

Conventional PAT reconstruction algorithms, e.g., filtered back-projection (FBP) and time reversal (TR) are widely used in photoacoustic image reconstruction. However, these reconstruction algorithms will result in distorted images with many artifacts in limited-view configuration.[4] Recently, the development of deep learning (DL) algorithm provides a new research perspective for the enhancement of PAT image quality, which is divided into image denoising based on low-quality reconstructed images, reconstruction process compensation based on traditional reconstruction algorithms, and direct image reconstruction.[5] Lan *et al.* established a generative adversarial network-based approach, Ki-GAN, which, in addition to time-series data, uses traditional delayed-sum reconstructed photoacoustic images as additional information to regularize neural network, the method is applied to a system distributed in a ring array.[6,7] Min *et al.* proposed a method to enrich time series data using a lookup table-based image transformation before reconstructing the image using U-Net.[8] Lu *et al.* proposed a hybrid data-driven deep learning method LV-GAN based on a generative adversarial network to recover high-quality images from sampled signals of a ring-array photoacoustic system with a limited-view angle. Experiments show that LV-GAN can achieve high recovery accuracy even with a limited detection angle of less than 60°.[9]

Also recently, some studies implemented flexible unsupervised DL strategies for photoacoustic image reconstruction. Lu *et al.* designed a PA-GAN based on CycleGAN to improve the limited-view image quality in PAT.[10] Li *et al.* proposed a SEED-Net to generate data-label pairs through unsupervised 'simulation-to-experiment' data translation and presented a QOAT-Net to estimate absorption coefficients.[11]

Although the above-mentioned DL algorithms have achieved advanced research results, the reconstruction process focuses on the post-processing of low-quality images obtained from traditional reconstruction. These reconstructions are highly dependent on traditional reconstruction methods. In the case of incomplete signals, the reconstruction results will also degrade due to the poor reconstruction quality of traditional methods. Waibel *et al.* used a modified U-Net architecture to estimate the initial pressure distribution directly from time-series pressure data, the first attempt to directly reconstruct images using the convolutional neural network (CNN).[12] Lan *et al.* developed a hybrid processing framework Y-Net to complete reconstruction by optimizing raw data and beamforming images, the network connects two encoders through a decoding path, which is more efficient than traditional algorithms.[13] The texture structure and high-dimensional features of the original signal are encoded, and the feasibility and robustness of



(a) (b) (c) (d) (e)

Fig. 1. Reconstruction images corresponding to the change in sampling angle and the number of transducer elements. (a) The ground truth. (b) and (c) represents reconstructed images of 128 and 64 detector elements with full-ring coverage, respectively. (d) and (e) represents reconstructed images of 128 and 64 detector elements with 180° limited-view coverage, respectively.

the method are verified by *in vitro* experiments. Tong *et al.* first attempted to take the physical model as the prior information of the network structure, designed a feature projection network FPnet connecting the signal domain and the image domain, and used U-Net as the subsequent image processing network.[14] Therefore, in the process of directly reconstructing the raw photoacoustic signal, the reconstruction performance can be further improved by adding the prior information of the physical model and the texture information of the low-quality images.

In this paper, we present the study on a multi-input parallel processing DL model triple-path feature transform network (TFT-Net), which is used to realize signal-to-image conversion and image-to-image feature extraction. TFT-Net combines the raw photoacoustic pressure signals and traditional linear reconstruction images as input, and takes the photoacoustic physical model (i.e., first-order partial derivative of signal to time) as a prior information to guide the reconstruction process. The residual block and squeeze and excitation block are introduced into TFT-Net to enhance the ability of extracting the limited-view signal features. For further efficient reconstruction, the final output of photoacoustic signals uses 'filter-then-upsample' operation with a pixel-shuffle multiplexer and a nonlinear max out module to quickly choose the maximum pixel value of all the channels. To verify the performance of the network, experiments on simulated data and *in-vivo* data were implemented.

2. Materials and Methods

2.1. *Photoacoustic tomography*

PAT uses a fixed-frequency short-pulsed laser to irradiate the biological tissue in the imaging area.[15] Instantaneous thermoelastic expansion of the tissue induces a rise in pressure, generating ultrasonic signals that propagate to the surface of the tissue.[16] The signal is collected by the ultrasonic transducer around the imaging area. The acquisition card realizes the setting of the acquisition frequency, amplifies the sound pressure signal generated by the tissue, and obtains a photoacoustic image that can reflect the structural characteristics after analysis by the reconstruction algorithm.

The photoacoustic signal collected by the ultrasonic transducer is related to the change in tissue temperature. The photoacoustic signal in the time and spatial domain $p(r,t)$ satisfies the following equation:

$$\left(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2}\right)p(r,t) = -\frac{\beta}{kv_s^2}\frac{\partial^2 T(r,t)}{\partial t^2}, \quad (1)$$

where v_s represents the speed of the ultrasonic wave, usually 1540 m/s in the experiment; β is the thermal expansion coefficient; k is the isothermal compressibility coefficient, which is related to the inherent properties of the heated tissue.

For laser pulses that satisfy thermal and stress constraints, the thermodynamic equation can be expressed as follows:

$$H(r,t) = \rho C_V \frac{\partial T(r,t)}{\partial t}. \quad (2)$$

$H(r,t)$ represents the thermal energy absorbed by the organization per unit time, which is affected by the equal volume specific heat capacity C_V and density ρ of biological tissue. Thus, by combining Eqs. (1) and (2), the photoacoustic wave equation expressing the photoacoustic propagation process can be obtained, as shown in the following equation:

$$\frac{\partial^2 p(r,t)}{\partial t^2} - v_s^2\nabla^2 p(r,t) = \Gamma H(r)\frac{\partial \delta(t)}{\partial t}, \quad (3)$$

where Γ is the Grüneisen coefficient, which represents the efficiency of tissue light energy to sound wave conversion. Then, the main idea of PAT reconstruction is to recover an accurate initial acoustic pressure distribution $H(r)$ from the detected raw photoacoustic signals.

2.2. *PAT image reconstruction*

At present, mature conventional photoacoustic image reconstruction algorithms are mainly divided into analytical reconstruction algorithm and iterative reconstruction algorithm.

Analytical reconstruction is the older approach to reconstruction, and consequently, has seen more development in PAI. Reconstruction is accomplished by deriving the analytical expression for photoacoustic images from the physical equation. Typical analytic reconstruction algorithms include delay-and-sum (DAS), FBP.[17–19] In the most common universal back-projection algorithm, $H(r)$ is determined by the weighted summation of the

back-projection term $b(d_i, t)$ and the coefficient ω_i.

$$H(r) = \sum_{i=1}^{N} b\left(d_i, t = \frac{|r - d_i|}{v_s}\right)\omega_i, \tag{4}$$

where i represents the index of the ultrasonic transducer element, ω is related to the centroid angle and position of the ultrasonic transducer, $b(d_i, t)$ is determined by the sound pressure signal and its derivative at the corresponding position and time, as shown in Eq. (5). Analytical reconstruction has high computational efficiency and fast imaging speed, but is highly data-dependent and prone to reconstruction artifacts.

$$b(d_i, t) = 2p(d_i, t) - 2t\frac{\partial p(d_i, t)}{\partial t}. \tag{5}$$

The iterative reconstruction algorithm is a kind of relatively flexible and variable reconstruction algorithm. The iterative algorithm is used to solve Eq. (3), which is written in matrix form:

$$Ax = y, \tag{6}$$

where x is the initial sound pressure distribution, A is the system matrix, and y is the measured photoacoustic signal. The objective of iterative reconstruction is to minimize the relationship between the measured signal y and the theoretical signal x predicted by the forward photoacoustic model, as shown in the following equation:

$$\underset{x}{\operatorname{argmin}} \frac{1}{2}\|Ax - y\|_2^2 + \lambda R(x), \tag{7}$$

where $\frac{1}{2}\|Ax - y\|_2^2$ represents the data item, the regularization factor $R(x)$ corresponds to the image prior, and λ is the weight of regularization.

Common iterative reconstruction algorithms include compressed sensing,[20,21] TR,[22,23] the alternating direction multiplier method,[24] etc. which are computationally expensive due to forward operation of each iteration.

Additionally, the data-driven artificial neural network, also known as deep learning,[25] has been developed to solve the inverse problem for imaging. DL is good at discovering complex patterns from massive amounts of data to determine the best model parameters to minimize the cost function. Generally, DL-based approaches follow the conventional reconstruction scheme mentioned above and can also be divided into iterative and noniterative reconstruction.

The DL-based iterative reconstruction scheme, that is, model-based learning, usually unrolls out the iterative process and simulates the process of multiple iterations through a series of network modules.[26] Establishing a model-based DL method and incorporating the physical forward model into the network can make full use of the theoretical basis of images, which has been studied in Refs. 27 and 28. However, due to the limitation of repeated simulation of the physical model, these improvements in reconstruction quality are usually time-consuming.

Noniterative reconstruction schemes, such as post-process reconstruction, using conventional reconstruction algorithms to obtain low-quality images, and then constructing a CNN to remove artifacts and noise, have been widely studied.[29–31] Compared with the deep learning-enhanced reconstruction method, the direct DL reconstruction does not depend on the image generated by the conventional reconstruction method and is easier to train. However, the data-driven DL direct reconstruction method does not include the physical model, and the quality of the reconstructed image is lower than other methods.[12] Then, some related research works[13,14] have attempted to introduce physical model prior information into the network to improve the reconstruction quality of signal-to-image domain transformation, and achieved good reconstruction results. However, Y-Net[13] is applied to image reconstruction of linear array transducers, and FPnet + Unet[14] has enhanced image quality based on signal domain transformation with slow reconstruction speed and numerous model parameters.

2.3. *Triple-path feature transform network*

In this paper, we integrate the post-processing image quality enhancement reconstruction method with the signal image domain reconstruction using prior information from physical models, and construct a TFT-Net to achieve parallel processing of multi-features while fusing rich prior information, as shown in Fig. 2. According to Eqs. (4) and (5), the raw photoacoustic signal $p(d_i, t)$ and the first-order partial derivative of the signal to time $\frac{\partial p(d_i,t)}{\partial t}$ have an important influence on the initial pressure reconstruction.[14] At the same time, taking the images obtained by the traditional

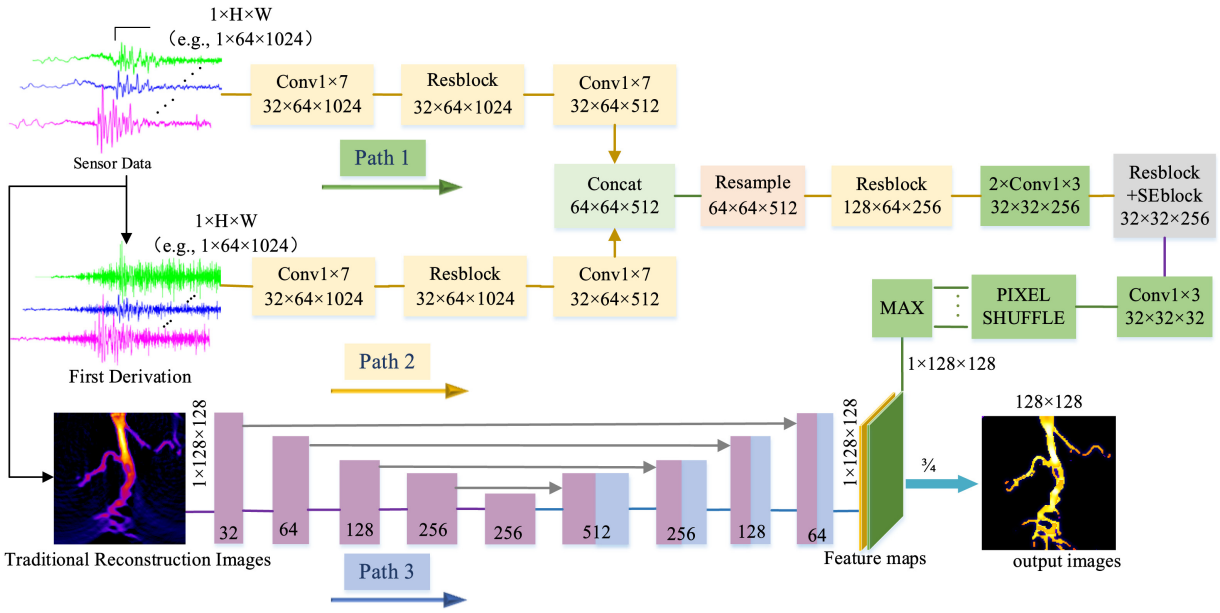


Fig. 2. Architecture of the proposed TFT-Net. Path 1, Path 2, and Path 3 are used to extract multiple features representing the raw photoacoustic signal, the first derivative of the photoacoustic signal and the traditional reconstruction images, respectively. Paths 1 and 2 present the output dimensions of each component in C × H × W. And the number in path 3 represents the number of filters.

reconstruction algorithm as the auxiliary input of network can provide additional texture information for the reconstruction results in addition to the sound pressure signal. TFT-Net realizes the hybrid processing of the raw sound pressure signal, the first derivative of the photoacoustic signal and the analytical reconstruction images. On the one hand, CNN is employed to extract and fuse the features of the original sound pressure signal and its first-derivative information, which is similar to the process of universal back projection (UBP) signal mapping pixels, so as to realize the feature transformation to photoacoustic images. On the other hand, CNN is used as a post-processing network, as shown in Fig. 2 using a typical encoder–decoder structure U-Net[32,33] to enhance the low-quality reconstructed images obtained by traditional reconstruction algorithms. Finally, the resulting feature maps are accumulated by summation to obtain qualified reconstruction images.

In order to enhance the network's ability to extract signal features, TFT-Net introduces typical feature extraction and transformation modules,[34,35] e.g., residual block (Resblock), squeeze and excitation block (SEblock) in Fig. 3. Based on the Resblock architecture, SEblock uses global average pooling,

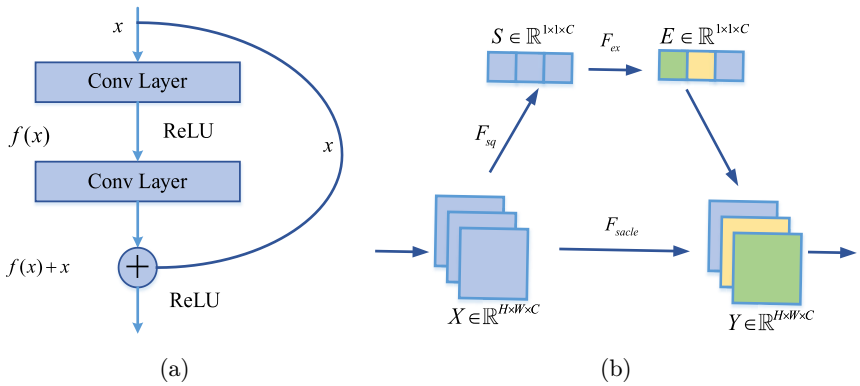


Fig. 3. Illustration of the feature extraction and transformation modules: (a) Resblock and (b) SEblock.

fully connected layer and objective function to redistribute the weight for each channel of the input features. Simultaneously, to increase the universality of the network and adapt to diverse input signal dimensions, a resample module of bilinear interpolation is adopted to adjust the processed dimension of (H, W) to a fixed (64, 512), where H represents the number of array elements and W is the signal length. For efficient reconstruction, the final output of path 1 and path 2 uses 'filter-then-upsample' operation with a pixel-shuffle multiplexer and a max out module, i.e., eSR-MAX,[36] instead of a fully connected layer. The learning parameters of the filtered convolution layer enable SEblock to explicitly model the weight relationship of each channel feature map in the training process, and finally achieve the goal of emphasizing important features and suppressing noise.

In addition, according to the difference between the characteristics represented by the input data and the location of the context, two normalization methods are used in the network, i.e., batch normalization (BN) and instance normalization (IN). Although the calculation methods of BN and IN are similar, each calculation of BN is oriented to the normalization of small batch samples, while IN normalizes a dimension of different channels of the feature map. Compared with BN, IN pays more attention to the independent features of each channel, which is often used in image stylization.[37] Therefore, in signal-to-image reconstruction, IN is used for feature extraction of photoacoustic pressure signal, combined with anisotropic convolution, which can preserve the independence of information collected by each transducer.

2.4. *Implementation*

2.4.1. *Performance measures*

The classic loss function mean square error (MSE) is used in the reconstructed model to measure the difference between the network prediction output and the label image in pixels, which is defined as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{8}$$

where n represents the number of pixels in the image, and the label image is consistent with the network output in size. y_i is the pixel value of label image, $\hat{y}_i$ is the pixel value of the network output image. The square processing of MSE makes the loss function sensitive to abnormal outliers, and its easy derivation makes it easier for the network model to approach the optimal solution.[38] In the training of the compared CycleGAN, the learned mapping function contains adversarial loss, cycle-consistency loss and identity loss, which is consistent with the original model.[39]

In order to quantitatively evaluate the performance of the network, mean absolute error (MAE), peak-signal-to-noise ratio (PSNR), and structural similarity index (SSIM)[40] are introduced as the evaluation metrics for image quality.

2.4.2. *Implementation details*

In network training, the batch size was set to 4. The model training epoch depends on the specific task and was usually set to 200. The neural network parameters were optimized using Adam optimizer with an initial learning rate of 1e-4 which is multiplied by 0.5 every 30 epochs. The photoacoustic signal data and traditional reconstructed low-quality input images are normalized to [0, 1]. The ground truth and final output of the model are 128×128 images. Based on the DL framework PyTorch,[41] the experiments are trained and verified on NVIDIA GTX1080Ti GPU with 11 GB memory.

3. Results

In order to validate the performance of the constructed reconstruction model, numerical simulation and *in-vivo* experiments were conducted. The reconstruction results of the proposed TFT-Net are compared with the widely used conventional reconstruction algorithm TR[23] and post-processing DL-based U-Net,[33] hybrid processing Y-Net,[13] domain transformation FPnet,[14] and unsupervised CycleGAN,[39] which are analyzed qualitatively and quantitatively on the simulation data set and the *in-vivo* data set, respectively. In experiments, the compared network models were adaptively adjusted based on the input dimension. In simulation and public *in-vivo* data, we vary the comparison model according to the input two path signal with dimension of (64, 1024), e.g., adjusting the last downsample layer Conv20 × 3 of Y-Net to Conv8 × 3 and changing the stride of the convolutional layer

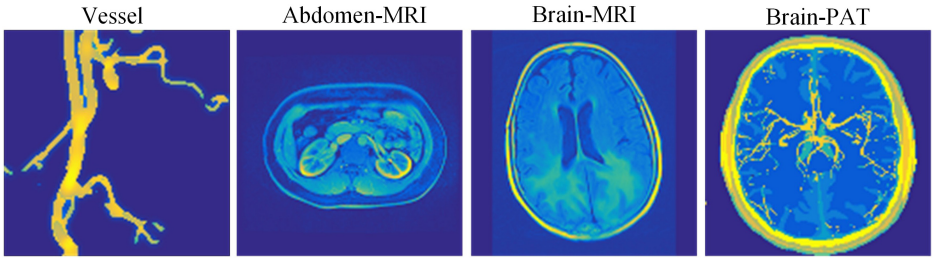


Fig. 4. Image examples for the four reconstructed datasets.

connected to the fully connected layer in FPnet from (1, 1) to (2, 1).

3.1. *Simulated data*

The MATLAB toolbox *k*-Wave[23] is used to generate the training data. As a supplement to the realistic dataset, the simulation dataset expands the data scale and increases the diversity of data. In simulation, we employed full-view 360° transducer ring-array with a radius of 55 mm, containing 128 detection elements. The center frequency of the transducer is set as 5.5 MHz with 80% bandwidth, and the speed of sound is set to 1540 m/s. Considering the need of different types of data for network generalization, four datasets with different structural characteristics are used to evaluate the performance of TFT-Net and other DL-based methods. For visual display, Fig. 4 shows example of images used in the four reconstructed datasets. The vessel dataset selected 1000 intricate and irregularly shaped cerebral angiography images to validate the convergence speed and stability of the model. The structure of abdomen medical images is relatively complex, and the distribution of tissue light absorption coefficient is diverse. Then based on the published 8270 abdomen MRI images data,[14] the model reconstruction performance is further verified. The brain dataset consists of public MRI images which can be acquired from the website of The Cancer Imaging Archive (TCIA).[42] Then, 1927 human brain MRI images were used for comprehensive comparison of reconstruction performance and human brain PAT transfer learning. Based on the optimal pre-trained model of human brain MRI, 10 human brain PAT images from the public photoacoustic ring-array tomography data set[43] were used for transfer learning of a few samples to validate the robustness of the proposed TFT-Net.

Due to the limited availability of the sample size, the human brain PAT images are divided into training set and testing set. Other datasets are randomly divided into train set, validation set, and test set according to the ratio of 70%, 10%, and 20%. Table 1 briefly describes each data set and lists the number of training, validation, and test samples. The training set is used for network training, the test set is used for model test, and the verification set is used to verify the training process and save the best model parameters. Further, based on the test set, three evaluation metrics are used to analyze the reconstruction performance. In the numerical simulation, we properly tested the reconstruction performance and robustness of the proposed TFT-Net.

3.1.1. *Model-fitting evaluation*

The fitting degree and reconstruction performance of the proposed TFT-Net was first evaluated based on a small sample vessel dataset. Table 2 lists the reconstruction performance of TFT-Net and baseline methods in terms of the evaluation metrics, with the

Table 1. Description of each simulated dataset.

Dataset	Description	Train set	Validation set	Test set
Vessel	Cerebral angiography images	700	100	200
Abdomen	Human abdomen MRI images	5789	827	1654
Brain	Human brain MRI images	1927	275	552
	Human brain PAT images	6	/	4

Table 2. Quantitative evaluation of proposed TFT-Net for vessel.

Method	PSNR (dB)	MAE ($\times 10^{-3}$)	SSIM
TR	19.7829	51.7811	0.3532
U-Net	23.0042	22.1096	0.8384
TFT-Net	**23.9837**	**16.1817**	**0.8440**

optimal values presented in bold. We can see that compared with the baseline method, TFT-Net with rich prior information fusion gains remarkable performance, with PSNR of 23.98 dB and SSIM of 0.84.

The loss and PSNR variation curves in the training based on U-Net and TFT-Net frameworks are shown in Fig. 5. It can be seen that there is no overfitting phenomenon, although the vessel training samples are insufficient. Due to the fusion of multiple features in TFT-Net to learn the reconstruction process, the convergence curve of the validation dataset based on TFT-Net is smoother and more stable compared to the baseline model U-Net.

In addition, we visually analyze the reconstruction performance on the test dataset, as shown in Fig. 6. The TR reconstructed result shows lower image quality and more artifacts. Based on the low-quality reconstructed image, post-processing U-Net can effectively reduce artifacts and enhance image quality. The TFT-Net model with multiple prior information fusion proposed on the baseline methods can reconstruct more accurate and clear vascular structures.

3.1.2. *Reconstruction performance on abdomen MRI*

The reconstruction performance was further compared and verified on the sufficient abdomen MRI dataset with relatively complex structures. The proposed TFT-Net is compared with multiple advanced DL-based photoacoustic image reconstruction algorithms, including U-Net, Y-Net, FPNet, and the unsupervised CycleGAN. The quantitative evaluation of the computed abdomen MRI test dataset is shown in Table 3. TFT-Net shows optimal reconstruction performance, achieving 30.70 dB (PSNR) and 0.89 (SSIM).

The visual effect comparison of different algorithms in the test set of abdomen MRI images is shown in Fig. 7. It can be seen that partial structural information can be restored by using post-processing U-Net or hybrid processing Y-Net. However, due to the relatively smooth filtering process, the obtained reconstructed image is blurry. FPNet using the measured raw data and its first derivative information as input, can reconstruct clearer image details. The unsupervised CycleGAN can improve image quality to some extent, but the generated reconstructed images contain significant background noise. The reconstruction performance of TFT-Net not only further improves the image details, but also greatly reduces the image with background noise and improves the analyzability of the reconstructed image.

3.1.3. *Reconstruction performance on human brain images*

Human brain MRI dataset with distinguishing features from abdomen structures was used to validate the reconstruction results of our method and the compared methods. In the reconstruction of human brain images with complicated structures, TFT-Net has also achieved excellent performance compared to other methods, as shown in Table 4, with PSNR of 25.71 dB and SSIM of 0.77. Parameters and

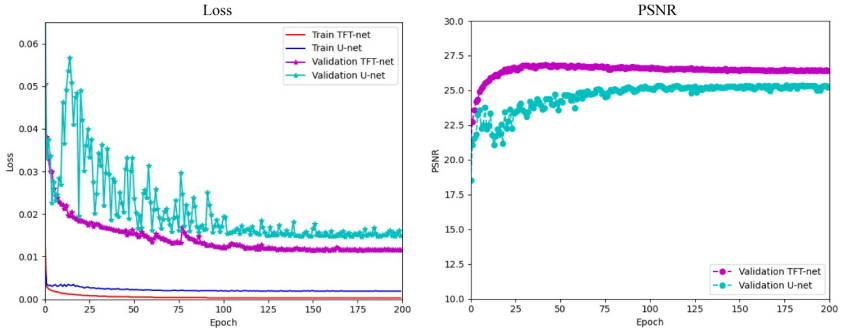


Fig. 5. Loss and PSNR evaluation metric change curves in the vessel dataset.

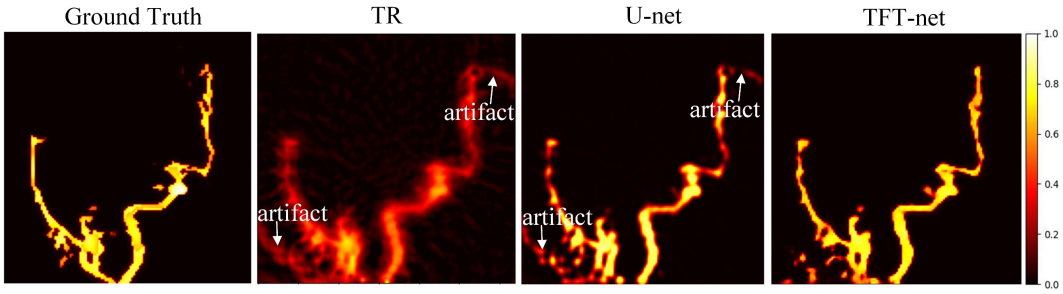


Fig. 6. Comparison of reconstruction performance based on different algorithms in vessel dataset.

computation amount are also calculated to measure the computational complexity and consumption of the model. As seen from Table 4, the quantity of parameters of TFT-Net is about 1.44 MB, achieving a better balance between image reconstruction performance and model complexity. And the floating-point operations (FLOPs) of TFT-Net is 100.78 GB, making it possible to realize high-quality reconstruction without intensive computation compared to other models e.g., CycleGAN and FPnet. As displayed in Table 4, the effectiveness of SEblock and eSR-MAX module is further verified. The SEblock following the last Resblock has a small improvement of 0.29 dB and 0.01 in PSNR and SSIM, respectively. In the eSR-MAX test experiment, a fully connected layer was used to replace the pixel-shuffle multiplexer and a max out module behind Con1 $\times$ 3 layer. The experimental results show that the parameter number of TFT-Net without eSR-MAX reaches 269.87 MB, which is nearly 188 times of the original model. In addition to significantly reducing model complexity, eSR-MAX also performs well in reconstruction performance, increasing PSNR by 1.87 dB and SSIM by 0.18.

Table 3. Quantitative evaluation comparison of different models for abdomen MRI.

Method	PSNR (dB)	MAE ($\times 10^{-3}$)	SSIM
U-Net	27.2310	23.8514	0.7692
Y-Net	26.2941	29.8859	0.5666
FPNet	25.7315	32.8905	0.6547
CycleGAN	22.5035	71.4363	0.4452
TFT-Net	**30.6961**	**15.6603**	**0.8931**

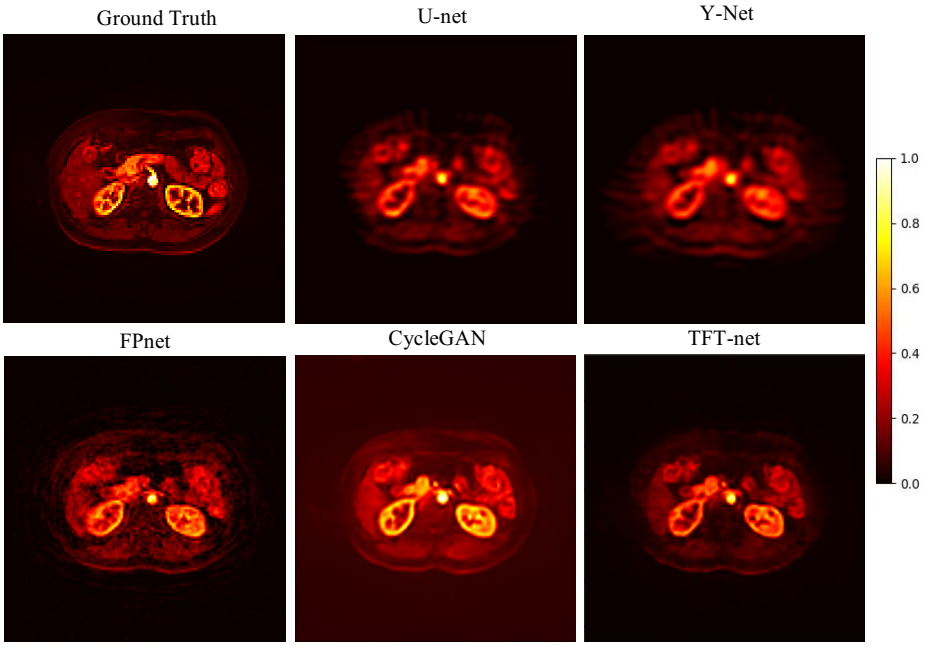


Fig. 7. Comparison of visual effect based on different algorithms in abdomen MRI.

Table 4. Quantitative evaluation comparison of different models for brain MRI.

Method	Parameters (MB)	FLOPs (GB)	PSNR (dB)	MAE ($\times 10^{-3}$)	SSIM
TR	/	/	15.5450	133.2564	0.4107
U-Net	**0.9268**	**8.9286**	22.8232	46.8272	0.6121
Y-Net	9.7593	22.8088	21.5640	55.6264	0.5177
FPNet	270.4076	300.9707	20.6364	62.4445	0.4567
CycleGAN	28.2956	118.0622	20.6310	77.9809	0.5544
TFT-Net w/o eSR-MAX	269.8719	101.8587	23.8336	43.5352	0.5885
TFT-Net w/o SEblock	1.4363	100.7839	25.4133	33.2374	0.7587
TFT-Net	1.4364	100.7849	**25.7069**	**32.3262**	**0.7708**

In addition to numerical results, the visual reconstruction results of different algorithms in brain MRI test images are shown in Fig. 8. To clearly compare the reconstruction effects of each model, the MSE ($\times 10^{-3}$) value between the recovered image and the ground truth is calculated. It can be intuitively seen that the supervised deep learning reconstruction methods can effectively eliminate background noise in low-quality images reconstructed by TR, but exhibit blurry and unclear details in image feature areas. CycleGAN with unsupervised style transfer enhances some details e.g., edges and textures, but shows unsatisfactory in removing background noise. The TFT-Net without SEblock module and the TFT-Net without eSR-MAX module both exhibit better reconstruction

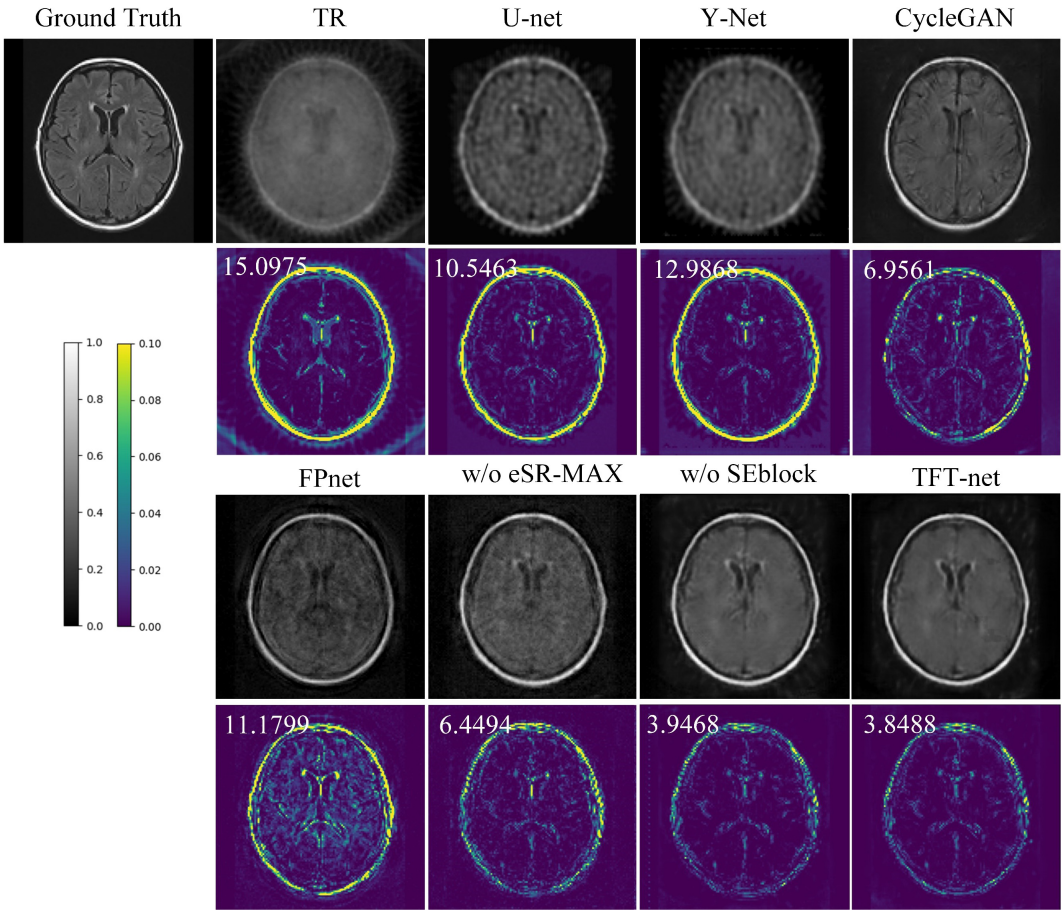


Fig. 8. Reconstruction results of different models for the human brain test images. Grayscale images represent the ground truth and corresponding generated images by each method. Viridis-color images represent the squared error maps between the recovered image and ground truth.

Table 5. Quantitative evaluation comparison for human brain PAT.

Method	PSNR (dB)	MAE ($\times 10^{-3}$)	SSIM
TR	14.9620	154.7668	0.5093
U-Net	12.9931	141.7165	0.3858
Y-Net	12.2102	145.2734	0.3957
FPNet	13.2545	141.7443	0.3972
CycleGAN	8.2101	192.7942	0.3915
TFT-Net	**18.1050**	**77.4122**	**0.6063**

performance than other DL-based models. At the same time, the reconstruction result of the TFT-Net without SEblock module is superior to those without eSR-MAX module, indicating that the eSR-MAX plays a more important role in improving the performance of the TFT-Net. TFT-Net outperforms other methods in reconstructing detail textures, resulting in clearer and more hierarchical images which are closer to the ground truth.

The proposed method was further experimentally validated on the public photoacoustic ring-array tomography dataset of human brain. Because of insufficient training samples for human brain PAT, we used pre-trained weights from human brain MRI simulation data as the network initial parameters. After transfer learning, the pre-processed PAT data was employed to optimize the constructed reconstruction models. In the photoacoustic test set, the quantitative reconstruction performance comparison of each model after 10 epochs of fine-tuning is shown in Table 5. It can be seen that in the transfer training with only a small number of samples, i.e., six human brain PAT images, only TFT-Net based on TR reconstructed low-quality images can effectively extract image feature information and improve performance metrics, demonstrating impressive model performance stability.

Figure 9 shows an example of visualization comparison results between the TFT-Net model and other methods. In human brain PAT, we can see that the reconstructed image obtained by TR has under-sampling artifacts, obvious background noise and blurred boundary of the imaging target. Although deep learning methods e.g., U-Net and CycleGAN have not shown performance improvements in terms of quantitative metrics, their subjective quality was relatively enhanced compared to low-quality images reconstructed by TR. The reconstruction of TFT-Net effectively eliminates the interference of background noise with a better visual effect. Although there are some differences in the intensity information of the target compared with the ground-truth, the reconstruction result is in line with the expectation and can clearly reflect the contour boundary of the target.

3.2. *In-vivo data*

After fully verifying the excellent reconstruction results in simulated data experiments, the proposed model is further applied to more complex and practical *in-vivo* data. Firstly, performance comparisons are conducted using mice's brain *in-vivo* data from public dataset MSOT-Brain.[14] According to the reconstruction settings, a transducer containing 64 channels' elements with a 180° view were used to rebuild the ground-truth image. Furthermore, we applied the model to finger cross-sectional

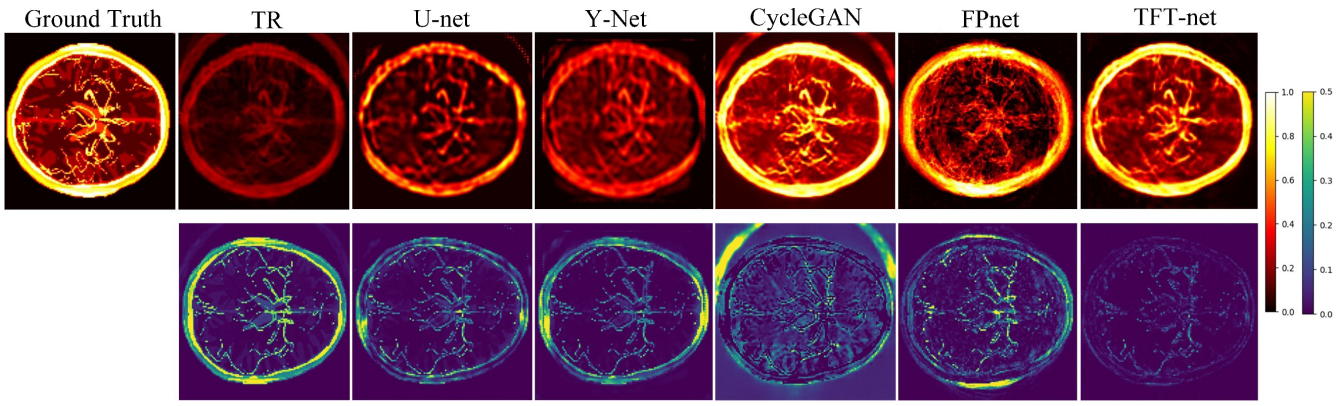


Fig. 9. Comparison of reconstruction performance for the human brain test sample. Row 1 represents the ground truth and corresponding generated images by each method. Row 2 represents the squared error maps between the recovered image and ground truth.

Table 6. Description of each *in-vivo* dataset.

Dataset	Description	Train set	Test set
Brain	Mouse brain PAT images	695	61
Finger	Human finger images in PAT system	800	200

images obtained from self-constructed PAT systems. The description of each photoacoustic dataset is shown in Table 6. Due to the limited sample size, the *in-vivo* images are divided into training set and testing set.

3.2.1. *Reconstruction performance on mouse brain PAT*

We implemented our experiment on public mouse brain *in-vivo* data to test the validity of TFT-Net. As shown in Table 7, TFT-Net still achieved optimal results in PSNR, SSIM, and MAE evaluation metrics, which is consistent with its performance on simulated data. Based on TR reconstruction results, TFT-Net increased PSNR by 17.78 dB and reduced MAE by 88.07%. In terms of reconstruction speed, TFT-Net using sound pressure signal and traditional reconstructed image as input, takes approximately 6.86 s to complete automatic reconstruction of testing images, which is about 4.35 s longer than the baseline U-Net. But compared with other popular reconstruction models e.g., Y-Net and CycleGAN, the time difference is within 1–2 s, and even exceeding FPNet.

Table 7. Quantitative evaluation comparison for mouse brain PAT.

Method	PSNR (dB)	MAE ($\times 10^{-3}$)	SSIM	Time (s)
TR	14.6909	152.0110	0.5927	/
U-Net	28.1431	31.1618	0.8207	**2.5151**
Y-Net	22.5791	67.3953	0.7624	4.8379
FPNet	23.6841	60.3207	0.7234	7.4335
CycleGAN	25.4436	44.6794	0.7843	5.8443
TFT-Net	**32.4731**	**18.1349**	**0.8703**	6.8645

Figure 10 shows the imaging performance of different algorithms in mouse brain PAT test images. It can be seen that based on TR low-quality images with poor visibility of structures, TFT-Net can accurately reconstruct more image details with clear texture edges, which is closest to the ground-truth, as shown by the red arrow. Although CycleGAN exhibits strong feature generation capabilities, some texture information of the image generated by CycleGAN is inconsistent with the ground truth, as shown by the yellow arrow. In terms of visualization and objective indicators, the reconstruction results of TFT-Net in the *in-vivo* data of ring-array PAT meet expectations, achieving high-quality reconstruction of photoacoustic images while considering imaging time. Furthermore, we tested the reconstruction performance of TFT-Net on different spatial sampling sparsity using mouse brain PAT images. The raw sensor data were detected by 8, 16, 32, and 64 elements with 180° angular coverage, respectively. As shown in the last row of Fig. 10, TFT-Net can effectively

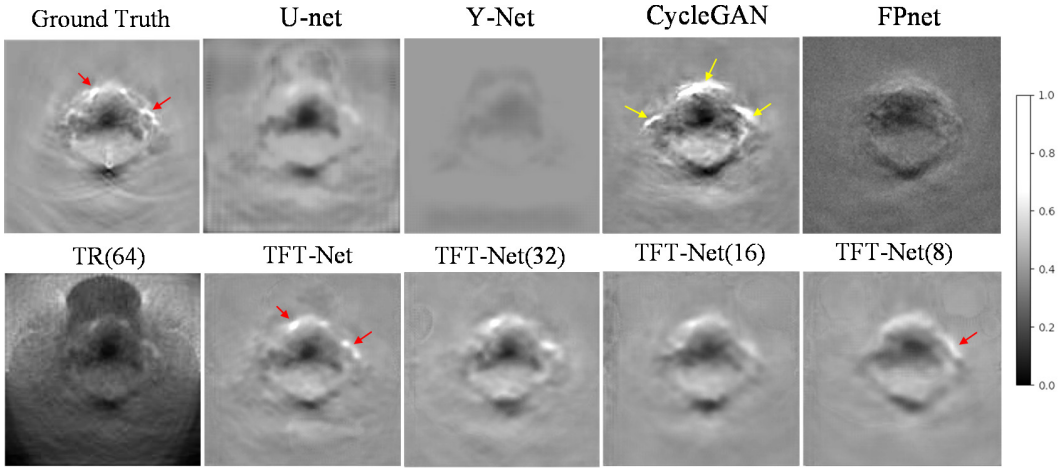


Fig. 10. Comparison of imaging performance of the mouse brain PAT. Row 1 represents the ground truth and corresponding generated images by contrastive DL-based methods. Row 2 represents the traditional reconstructed images and TFT-Net reconstructed images based on different sparse sampling data.

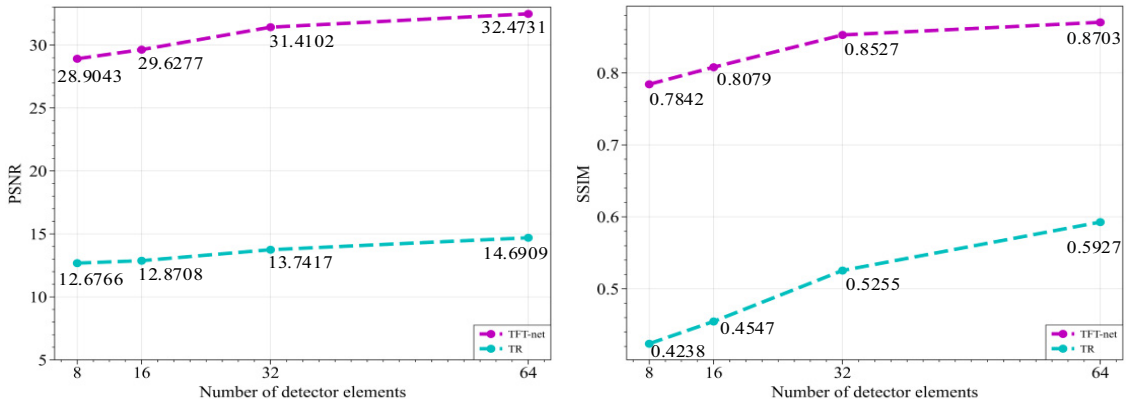


Fig. 11. The quantitative evaluation metric change curves of TR and TFT-Net with 8, 16, 32, and 64 elements. The left and right charts represent the PSNR and SSIM, respectively.

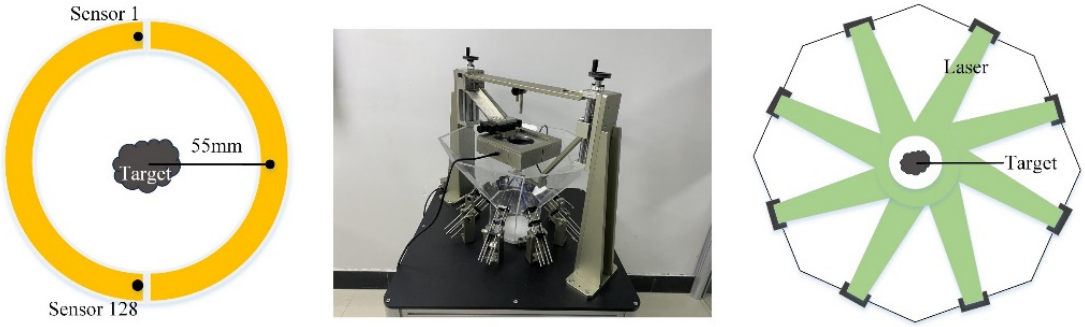
recover the missing information and improve the quality of PAT images suffering sparse sampling.

Moreover, from the comparison of quantitative metrics before and after TFT-Net reconstruction in Fig. 11, it can be seen that TFT-Net shows stable reconstruction performance for images with different levels of sampling sparsity, and can still show excellent performance in the reconstruction of low number (8) of ultrasound sensors, with a difference of only 3.57 dB in PSNR and 0.09 in SSIM compared to 64-channel reconstruction images.

3.2.2. *Validating on PAT system*

To further evaluate the reconstruction performance of TFT-Net, we used a full-ring tomographic scanner consisting of two semicircular ultrasound transducer arrays for finger imaging. The distribution of the two semi-circular ultrasound transducer arrays is shown in Fig. 12(a), each with 128 elements for ultrasound detection evenly distributed in the imaging area with a radius of 55 mm. The PA ring-array imaging system adopts multi-path annular illumination to realize full angle illumination of the imaging area, as shown in Fig. 12(b). The optical fiber is used to transmit the laser, the beam is collimated by the lens, and the tomography irradiation is formed in the imaging area after the light is transmitted through the water tank wall. The central frequency of the transducer is 5.5 MHz, and the signal acquisition frequency is set to 25.51 MHz. At the same time, 532 nm wavelength laser is used to excite the signal of the imaging target, and the laser repetition rate is set to 20 Hz. As shown in Fig. 12(c), the imaging target is sampled using the ring-array PA system, which is evenly covered by the illumination area. The photoacoustic signal generated by absorbing laser energy is finally received by the transducer.

In systematic experiments, the semi-ring128-channel signal is employed as the subsampled signal, and the original 256-channel signal is employed



(a) (b) (c)

Fig. 12. System experimental equipment: (a) Probe distribution and view setting, (b) Ring-array photoacoustic system, and (c) Circular lighting diagram.

Table 8. Quantitative evaluation metrics comparison for finger PAT.

Method	PSNR (dB)	MAE ($\times 10^{-3}$)	SSIM
DAS	26.5690	29.3944	0.5443
U-Net	28.6303	21.7006	0.7658
CycleGAN	28.2305	29.4088	0.7069
WGAN-GP	31.3003	16.0932	0.8331
TFT-Net w/o path3	24.4316	31.8167	0.5950
TFT-Net w/o path2	31.2913	16.5161	0.8365
TFT-Net w/o path1	31.5761	15.8830	0.8399
TFT-Net	**31.6961**	**14.8912**	**0.8418**

as the fully sampled signal. Based on the effective length, the data is intercepted to obtain the dual path signal with dimension of (128, 1600) as input. The low-quality reconstructed images of 128×128 were obtained by traditional reconstruction algorithm DAS.[44]

A quantitative analysis of test samples from the realistic test set is displayed in Table 8. Due to the mismatch between the signal dimension and the FPNet and Y-Net models, we mainly compared the performance with the post-processing models, i.e., U-Net and CycleGAN. At the same time, considering the strong feature generation capability of GANs, we introduced the supervised learning-based Wasserstein generative adversarial network with gradient penalty (WGAN-GP)[45] in sparse and 180° limited-view PAT acquisitions, which has been successfully used in removing limited-view and limited-bandwidth artifacts in PAT images.[46] In the model construction, the U-Net framework shown in Fig. 2 is used as the generator, and a discriminator is the same structure as Ref. 46.

In the system experiment, the evaluation metrics of DAS reconstruction results is acceptable when employing the 128 sensors with 180° view, and the

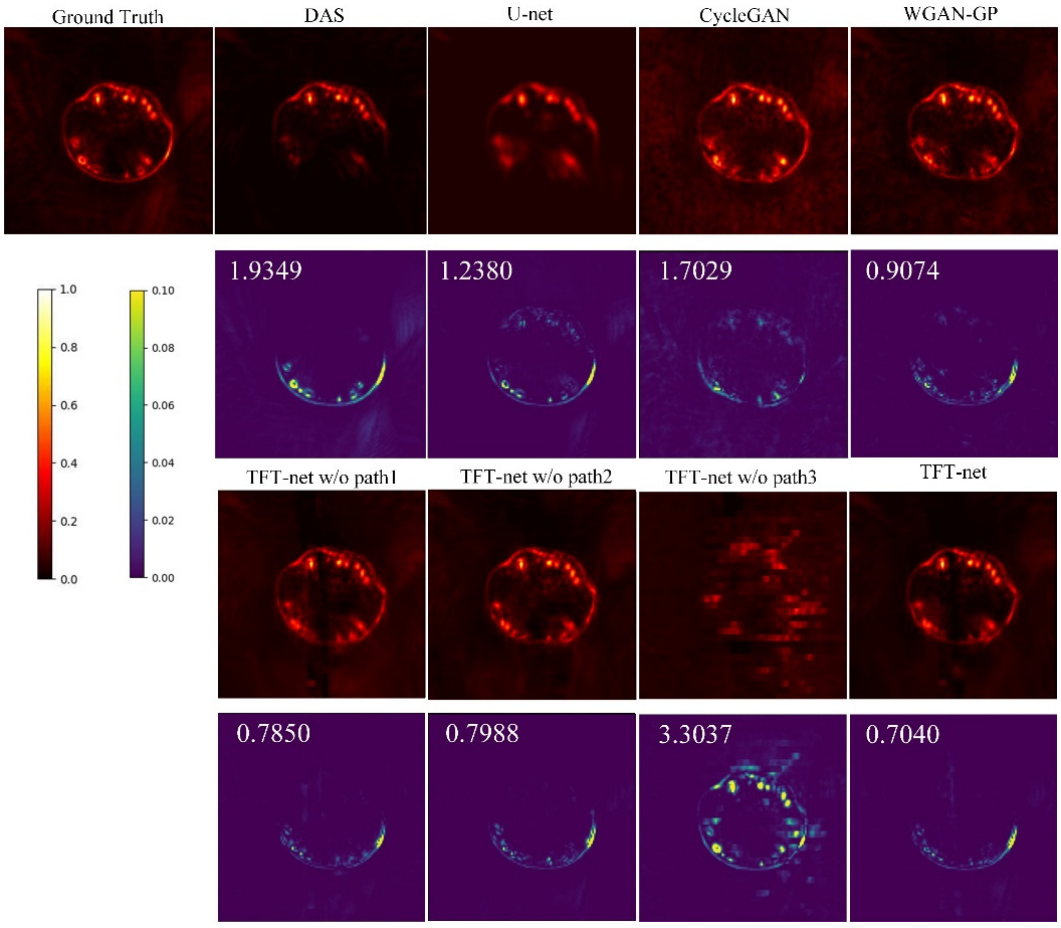


Fig. 13. Comparison of reconstruction performance for the finger PAT test sample. Hot-color images represent the ground truth and corresponding generated images by each method. Viridis-color images represent the squared error maps between the recovered images and ground truth, where the values represent the MSE ($\times 10^{-3}$) between the restored image and ground truth.

PSNR reaches 26.57 dB. On this basis, U-Net increased PSNR by 2.06 dB and SSIM by 0.22. Compared to U-Net, unsupervised reconstruction method CycleGAN showed a smaller performance increase, with PSNR by 1.66 dB and SSIM by 0.16. WGAN-GP improves the PSNR by 2.67 dB and 3.07 dB, and SSIM by 0.07 and 0.13 compared to U-Net and CycleGAN, demonstrating good reconstruction performance. The evaluation metrics of TFT-Net are slightly higher than WGAN-GP. It still shows excellent evaluation metrics in the finger imaging of the PAT system, achieving PSNR of 31.70 dB and SSIM of 0.84, which are 3.07 dB and 0.08 higher than the baseline U-Net, respectively.

The contributions of different paths in TFT-Net are compared in Table 8. It can be seen that TFT-Net without path 1 shows a slightly better performance than TFT-Net without path 2, both of which are significantly superior to other comparison models, e.g., U-Net and CycleGAN. The performance of TFT-Net without path 1 is slightly inferior to TFT-Net. However, TFT-Net without path 3 has unsatisfactory evaluation metrics, with PSNR of 24.43 dB and SSIM of 0.60, even lower than U-Net, demonstrating that path 3 in TFT-Net plays a more important role in high-quality image reconstruction.

Finger cross sections PAT examples of visualization performance between the TFT-Net model and other compared reconstruction methods are shown in Fig. 13. It can be seen that due to missing information, the DAS-based images acquired with limited-view 180° has incomplete finger structure. Although CycleGAN is inferior to U-Net in evaluation metrics, it demonstrates the image generation capability of GANs and can effectively recover missing structural information in visualization results. However, the background noise in the reconstructed image is still obvious, and the ability to remove artifacts and noise shows general performance. Compared with unsupervised CycleGAN, WGAN-GP can complete the reconstruction of missing information and effectively remove background noise, resulting in clearer reconstructed images. From Fig. 13, it can be seen that the reconstruction effect of TFT-Net without path 1 or path 2 is significantly better than that without path 3, indicating that the input of low-quality reconstructed images is the key to effectively guide the network to complete high-quality image reconstruction. On this basis, combined with the features of the original photoacoustic signal and its first-order derivative to time, the maximum extraction of reconstructed detail information can be achieved. The reconstructed image obtained by TFT-Net has better visibility, which not only accurately restores the missing structure, but also effectively minimized prominent artifacts, showing well effect in practice.

4. Discussion

Noniterative schemes based on deep learning can provide image reconstruction with low latency, high quality, and real-time performance, demonstrating wide potential application prospects e.g., early tumor screening or surgical guidance.[13] In this study, TFT-Net was used to remove artifacts and distortions caused by sparse sampling with limited-view. Model training was conducted on simulated datasets with different microstructures, and the results showed that TFT-Net can effectively restore high-quality, high-fidelity images and adapt well to various imaging targets. Moreover, the model integrates abundant prior information and adopts parallel processing, which is conducive to faster rate of convergence and more stable reconstruction performance. Due to the lack of widespread clinical application of PAI, real experimental photoacoustic data is insufficient.[47] Therefore, we validated the feasibility of transfer training the model to small-sample PAT based on reported human brain MRI. Through comparative experiments, TFT-Net achieves rapid feature extraction and can effectively improve the quality of reconstructed images. Based on *in-vivo* mouse brain data and finger experimental data, the effectiveness of the model was further verified. TFT-Net still shows good reconstruction performance in the sparse sampling with 8, 16, and 32 detection elements, and can recover sparsely sampled photoacoustic images under limited-view, indicating its great potentialities for transplantation or expansion to other scenarios.

This study still has some limitations and potential deviations. The model was only validated on sparse sampling with full-ring or 180° limited-view coverage, without verifying more severe imaging conditions. At the same time, the method of multi prior fusion can effectively improve reconstruction performance, while also accompanied by relatively inconvenient data preprocessing processes. Based on the experimental results, we can see that the GANs perform outstandingly in the recovery of

missing information, while the CNNs have a more obvious effect on denoising and artifact removal. Therefore, combining the advantages of both may further improve the reconstruction quality and speed in the limited view with sparse sampling. In the future work, we will seek the representation relationship between various prior information, further optimize TFT-Net method, and extend TFT-Net to 3D for real-time PAI.

5. Conclusions

Based on the ring-array photoacoustic system, in this paper, a new DL imaging algorithm TFT-Net is proposed to improve the imaging quality while considering the imaging speed in the limited-view and sparse-sample signal acquisition situation. The proposed TFT-Net implements parallel processing of three inputs, integrating the texture structure of traditional algorithms, the high-dimensional features of the raw photoacoustic signal and the prior knowledge of the basic physical model. We train the model using photoacoustic simulated data generated by the k-wave simulation toolbox and evaluated it on the test set. Compared with other DL-based models and conventional reconstruction methods, TFT-Net shows better convergence stability and optimal performance, significantly enhancing the definition and contrast of images under sparse sampling. In addition, we verify the performance of TFT-Net on *in-vivo* data. TFT-Net still shows excellent image reconstruction performance in the finger imaging of the PAT system, achieving PSNR of 31.70 dB and SSIM of 0.84, respectively. Compared with the widely used methods, it exhibits superior performance and can obtain high-quality reconstruction results under sparse sampling with limited-view.

Acknowledgments

This work was supported by National Key R&D Program of China [2022YFC2402400], the National Natural Science Foundation of China [Grant No. 62275062] and Guangdong Provincial Key Laboratory of Biomedical Optical Imaging Technology [Grant No. 2020B121201010-4].

Conflicts of Interest

The authors declare that there are no conflicts of interest relevant to this paper.

ORCID

Lingyu Ma https://orcid.org/0000-0001-6050-5914
Zezheng Qin https://orcid.org/0009-0006-3169-2929
Yiming Ma https://orcid.org/0000-0001-7580-7288
Mingjian Sun https://orcid.org/0000-0001-8719-524X

References

1. V. Ntziachristos, D. Razansky, "Molecular imaging by means of multispectral optoacoustic tomography (MSOT)," *Chem. Rev.* **110**(5), 2783–2794 (2010).
2. J. Zhao *et al.*, "H_2O_2-sensitive nanoscale coordination polymers for photoacoustic tumors imaging via in vivo chromogenic assay," *J. Innov. Opt. Heal. Sci.* **15**(5), 2250026 (2022).
3. Z. Qin *et al.*, "The sparse array elements selection in sparse imaging of circular-array photoacoustic tomography," *J. Innov. Opt. Heal. Sci.* **15**(5), 2250030 (2022).
4. E. Mensah *et al.*, "Deep learning in the management of intracranial aneurysms and cerebrovascular diseases: A review of the current literature," *World Neurosurg.* **161**, 39–45 (2022).
5. C. Yang *et al.*, "Review of deep learning for photoacoustic imaging," *Photoacoustics* **21**, 100215 (2021).
6. H. Lan *et al.*, *Ki-GAN: Knowledge Infusion Generative Adversarial Network for Photoacoustic Image Reconstruction In Vivo*, Springer, Cham (2019).
7. H. Lan *et al.*, Hybrid neural network for photoacoustic imaging reconstruction, *2019 41st Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6367–6370 (2019).
8. W. K. Min *et al.*, "Deep-learning image reconstruction for real-time photoacoustic system," *IEEE Trans. Med. Imaging* **39**(11), 3379–3390 (2020).
9. T. Lu *et al.*, "LV-GAN: A deep learning approach for limited-view optoacoustic imaging based on hybrid datasets," *J. Biophoton.* **14**(2), e202000325 (2021).
10. M. Lu *et al.*, "Artifact removal in photoacoustic tomography with an unsupervised method," *Optica* **9**(1), 32–41 (2022).
11. J. Li *et al.*, "Deep learning-based quantitative opto acoustic tomography of deep tissues in the absence of labeled experimental data," *Biomed. Opt. EUBPxpress* **12**(10), 6284–6299 (2021).
12. D. Waibel *et al.*, "Reconstruction of initial pressure from limited view photoacoustic images using deep learning," *Photons Plus Ultrasound: Imaging Sens.* **10494**, 104942S1-8 (2018), doi: 10.1117/12.2288353.
13. H. Lan *et al.*, "Y-Net: Hybrid deep learning image reconstruction for photoacoustic tomography in vivo," *Photoacoustics* **20**, 100197 (2020).

14. T. Tong *et al.*, "Domain transform network for photoacoustic tomography from limited- view and sparsely sampled data," *Photoacoustics* **19**, 100190 (2020).
15. Y. Wang *et al.*, "Review of methods to improve the performance of linear array-based photoacoustic tomography," *J. Innov. Opt. Heal. Sci.* **13**(2), 2030003 (2020).
16. L. Li, L. V. Wang, "Recent advances in photoacoustic tomography," *BME Front.* **118**, 1–17 (2021).
17. M. Mozaffarzadeh *et al.*, "Double-stage delay multiply and sum beamforming algorithm: Application to linear-array photoacoustic imaging," *IEEE Trans. Biomed. Eng.* **65**(1), 31–42 (2018).
18. M. Xu, L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Phys. Rev. E* **71**, 016706 (2005).
19. L. Zeng *et al.*, "High antinoise photoacoustic tomography based on a modified filtered back-projection algorithm with combination wavelet," *Med. Phys.* **34**(2), 556–563 (2007).
20. S. Arridge *et al.*, "Accelerated high-resolution photoacoustic tomography via compressed sensing," *Phys. Med. Biol.* **61**(24), 8908 (2016).
21. M. M. Betcke *et al.*, "Acoustic wave field reconstruction from compressed measurements with application in photoacoustic tomography," *IEEE Trans. Comput. Imag.* **3**(4), 710–721 (2017).
22. T. D. Mast *et al.*, "A k-space method for large-scale models of wave propagation in tissue," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **48**(2), 341–354 (2001).
23. B. E. Treeby, B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.* **15**(2), 021314 (2010).
24. S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Le.* **3**(1), 1–122 (2010).
25. Y. Lecun, Y. Bengio, G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
26. S. Diamond *et al.*, "Unrolled optimization with deep priors," preprint, arXiv:1705.08041 (2017).
27. A. Hauptmann *et al.*, "Model based learning for accelerated, limited-view 3D photoacoustic tomo graphy," *IEEE Trans. Med. Imag.* **37**(6), 1382–1393 (2018).
28. Y. E. Boink, S. Manohar, C. A. Brune, "Partially learned algorithm for joint photoacoustic reconstruction and segmentation," *IEEE Trans. Med. Imag.* **39**(1), 129–139 (2019).
29. J. Schwab *et al.*, "Real-time photoacoustic projection imaging using deep learning," preprint, arXiv: 1801.06693 (2018).
30. H. Zhang *et al.*, "A new deep learning network for mitigating limited-view and under-sampling artifacts in ring-shaped photoacoustic tomography," *Comput. Med. Imag. Grap.* **84**, 101720 (2020).
31. T. Wang *et al.*, "Sparse view photoacoustic image quality enhancement based on a modified U-Net network." *Laser Optoelectron. P.* **59**(6), 0617022 (2022).
32. O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, Springer, New York (2015).
33. L. Ma *et al.*, "Deep learning for classification and localization of early gastric cancer in endoscopic images," *Biomed. Signal Process.* **79**, 104200 (2023).
34. K. He *et al.*, "Deep residual learning for image recognition," *2016 IEEE Conf. Computer Vision & Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778, IEEE (2016).
35. J. Hu *et al.*, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal.* **42**(8), 2011–2023 (2019).
36. P. N. Michelini *et al.*, "Edge–SR: Super–resolution for the masses," preprint, arXiv:2108.10335 (2021).
37. D. Ulyanov, A. Vedaldi, V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," preprint, arXiv:1607.08022 (2016).
38. H. Zhao *et al.*, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.* **3**(1), 47–57 (2017).
39. J. Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE Int. Conf. Computer Vision (ICCV)*, Venice, pp. 2242–2251 (2017).
40. Z. Wang *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
41. A. Paszke *et al.*, "Automatic differentiation in PyTorch," *31st Conf. Neural Information Processing Systems*, Long Beach, CA, USA (2017).
42. K. Clark *et al.*, "The cancer imaging archive (tcia): Maintaining andoperating a public information repository," *J. Digit. Imag.* **26**(6), 1045–1057 (2013).
43. T. Lyu *et al.*, "Photoacoustic digital brain: Numerical modelling and image reconstruction via deep learning," preprint, arXiv:2109.09127 (2021).
44. J. Xia, J. Yao, L. V. Wang, "Photoacoustic tomography: Principles and advances," *Electromagn. Waves (Camb)* **147**, 1–22 (2014).
45. I. Gulrajani *et al.*, "Improved Training of Wasserstein GANs," preprint, arXiv:1704.00028 (2017).

46. T. Vu *et al.*, "A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer," *Exp. Biol. Med. (Maywood)*, **245**(7), 597–605 (2020).

47. H. Deng *et al.*, "Deep learning in photoacoustic imaging: A review," *J. Biomed. Opt.* **26**(4), 040901 (2021).