Neurocomputing 573 (2024) 127207

Contents lists available at ScienceDirect

Neurocomputing

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

Survey paper

Deep learning-powered biomedical photoacoustic imaging☆

Xiang Wei [a,b,c], Ting Feng [d], Qinghua Huang [e], Qian Chen [a,c], Chao Zuo [a,b,c], Haigang Ma [a,b,c,*]

[a] Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[b] Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210019, China
[c] Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing, Jiangsu Province 210094, China
[d] Academy for Engineering & Technology, Fudan University, Shanghai 200433, China
[e] School of Artificial Intelligence, Optics and Electronics(iOPEN), Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

ARTICLE INFO

Communicated by Zidong Wang

Keywords:
Biomedical imaging
Photoacoustic imaging
Deep learning
Convolutional networks

ABSTRACT

Photoacoustic Imaging (PAI) is an emerging hybrid imaging modality that combines optical imaging and ultrasound imaging, offering advantages such as high resolution, strong contrast, and safety. Despite demonstrating superior imaging capabilities, PAI still has certain limitations in its clinical application, such as the trade-off between imaging depth and spatial resolution, and the need for further improvement in imaging speed. Deep Learning, as a novel machine learning technique, has gained significant attention for its ability to improve medical image data and has been widely applied in PAI in recent years to overcome these limitations. In this review, we first introduce the principles of photoacoustic imaging, followed by the development and applications of popular deep neural network structures such as U-Net and GAN networks. Furthermore, we comprehensively discuss the recent advancements in the application of deep learning in photoacoustic imaging. Finally, a summary and discussion are provided.

1. Introduction

1.1. Photoacoustic imaging

Photoacoustic Imaging (PAI) is a novel non-invasive photon imaging technique used for disease detection, observing biological tissue structure, and assessing function. The physical basis of PAI is the photoacoustic effect in biological tissue. When a short-pulsed laser illuminates the imaged sample, the tissue or substance absorbs the light energy, resulting in thermal elastic expansion and causing instantaneous expansion and contraction of the surrounding medium, thereby generating ultrasound waves propagating towards the tissue surface and being received. By receiving the ultrasound signals and using acoustic inverse problems, the initial sound pressure signal map of the tissue surface can be reconstructed, enabling observation and diagnosis of biological tissue structure and function [1,2]. Due to the significant difference in scattering intensity between ultrasound waves and photons in biological tissue (approximately 2–3 orders of magnitude), ultrasound scattering is much lower than that of photons. Therefore, PAI can overcome the diffraction limit of optical imaging depth (i.e., 1 mm). Moreover, PAI combines the high imaging depth of ultrasound imaging with the high contrast and high resolution of optical imaging, thereby achieving high-depth, high-contrast, and high-resolution imaging of biological tissue by leveraging the advantages of both technologies.

The most common forms of photoacoustic imaging are photoacoustic tomography (PAT), photoacoustic microscopy (PAM), and photoacoustic endoscopy (PAE) [3,4]. PAT uses a non-focused large-diameter pulsed laser beam to achieve full-field illumination of the tissue surface and employs an array transducer to collect signals, which are then reconstructed into an image using inversion algorithms. Existing inversion algorithms include filtered back-projection (FBP), delay-and-sum (DAS) beamforming algorithm, Fourier-based algorithms, and time reversal (TR) algorithm. PAM, on the other hand, uses a focused short-pulsed laser to illuminate the target point and employs a focused transducer to collect the PA signal point-by-point, allowing for image reconstruction without the need for additional inversion algorithms. PAE is a type of endoscope-based photoacoustic imaging technology. Due to its unique imaging principles and the advantages of

* Corresponding author at: Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China.
E-mail address: mahaigang@njust.edu.cn (H. Ma).

https://doi.org/10.1016/j.neucom.2023.127207
Received 12 September 2023; Received in revised form 17 November 2023; Accepted 26 December 2023
Available online 29 December 2023

optics and acoustics, photoacoustic imaging has broad application prospects and has gradually played a role in clinical medicine, biomedical research, drug development, material science, and other biomedical fields. Its application areas are also constantly expanding and deepening.

1.2. Deep learning

With the advent of the big data era, computer computational capabilities have significantly improved, and continuously emerging open-source and user-friendly software frameworks have led to unprecedented development of artificial intelligence (AI) technology. Classical AI technology machine learning has attracted great interest in both industry and academia, especially data-driven artificial neural network technology, that is, deep learning [3–6]. The deep learning method aims to discover complex mappings from training data to achieve optimization of the existing parameter space problem. Unlike the lack of computing power in the past, today's graphic processing units allow neural networks to continuously improve their depth [7,8], width [9], computation speed and other aspects, gradually developing various basic network architectures. Deep learning has become an important method in computer vision, natural language processing, and AI fields. This article introduces various networks' applications and effects in the optical-acoustic field based on supervised learning perspective, focusing on classic deep neural network structures.

1.3. Deep learning-powered photoacoustic imaging

In the two previous sections, we introduced the advantages of photoacoustic imaging and deep learning and found that both have very good prospects in their respective fields. Especially in the medical imaging field, photoacoustic imaging has many advantages, such as combining acoustic depth, optical resolution, and non-invasiveness. However, photoacoustic imaging still faces many challenges, including image quality limited by sound and light diffraction, and various problems in the data acquisition, processing, and inversion processes. For example, in PAT, it is difficult to achieve low-cost equipment and high signal-to-noise ratio image reconstruction at the same time, and the widely used sparse detectors currently have difficulty obtaining good reconstruction results through conventional inversion methods. In PAM, there are also deficiencies in imaging speed. Although scanning speed can be improved by changing the repetition rate of the excitation light pulse and the scanning mechanism, these methods often have an unavoidable impact on image quality. In short, there is a certain contradiction between image quality, economic benefits, and time efficiency in photoacoustic imaging. Although many methods have been proposed to solve these problems, and these methods have achieved some effectiveness, further exploration and improvement are still needed.

The intervention of deep learning has had a huge impact in the field of photoacoustic imaging. We have found that a large number of photoacoustic imaging works based on deep learning have achieved imaging quality and efficiency that previous methods have difficulty achieving. This is also the reason why we want to write this review and organize and analyze recent related work. We want to organize and analyze our work in recent years from four important directions of photoacoustic imaging: PAT image reconstruction, PAM image reconstruction, photoacoustic image processing, and photoacoustic signal processing. Not only that, we also introduced the development and current status of common network structures such as U-Net and GAN networks in image processing. Finally, we summarized and prospected the review.

The first chapter of this article introduces the principle of photoacoustic imaging, the principle of deep learning network, and analyzes the current problems of photoacoustic imaging. Chapter 2 details the development of current popular deep learning networks, including U-Net, Residual Network (ResNet), and Super-Resolution Generative Adversarial Network (SRGAN). Chapter 3 lists and analyzes the application results of current deep learning technology in various fields of photoacoustic imaging. Chapter 4 summarizes the application results and problems of deep learning in the field of photoacoustics, and looks forward to future development directions. The following is the article flowchart and Chart of Recent works on Deep Learning-powered photoacoustic imaging.,.

2. The neural network structures based on photoacoustic imaging

In recent years, the combination of photoacoustic imaging and deep learning has brought significant improvements to photoacoustic imaging. Considering the effectiveness, real-time performance, and economy of the method, U-Net has emerged in various aspects of photoacoustic imaging in recent years due to its simple and efficient network structure. It has been applied to PAT reconstruction, PAM reconstruction, denoising, and image processing of photoacoustic images. Its superiority in image recognition and segmentation tasks was first discovered, and then it was applied to image denoising. It is worth noting that the skip connections in U-Net ensure the validity of the image, which effectively improves the signal-to-noise ratio of the image and greatly suppresses the possible artifacts produced during the processing. Subsequently, the U-Net network has also been widely applied to other aspects of various photoacoustic imaging methods, which also proves that network optimization for image recognition tasks is applicable to optimizing network performance for other image tasks. In addition, U-shaped deep neural networks also have certain robustness and generalization capabilities, can process different types and qualities of data, and can further improve model performance through techniques such as data augmentation.

Overall, U-Net has great advantages in the field of photoacoustic imaging. It can effectively process high-dimensional data, learn features in the data, and achieve accurate image segmentation and localization. This makes it a very promising tool in the field of photoacoustic imaging, which can help doctors and researchers make more accurate diagnoses and treatments. This chapter mainly introduces the U-Net and SRGAN network structures, details the birth of U-Net and the development process of its network architecture, shares the photoacoustic microscopy method we are working on based on SRGAN, and finally introduces the classic residual block structure.

2.1. U-Net

2.1.1. The proposal of U-Net network

Ronneberger et al. first proposed the U-Net network in 2015 [10],

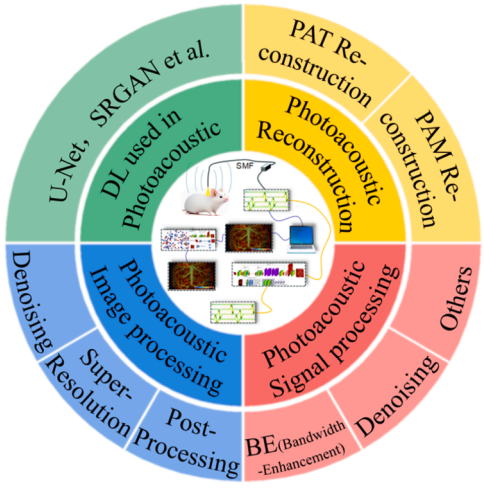

Fig. 1. The flowchart of this paper.

Recent works on Deep Learning-powered photoacoustic imaging				
Type	Name	Published time	Structure	Notes
PAT reconstruction	Deep learning for photoacoustic tomography from sparse data	2019	U-net with residual connection	It contains PAT filtered backprojection algorithm for the first layer.
	A New Deep Learning Network for Mitigating Limited-view and Under-sampling Artifacts in Ring-shaped Photoacoustic Tomography	2020	ring-array deep learning network (RADL-net)	It improves the quality of reconstructed images on a three-quarter ring transducer array.
	Limited-View and Sparse Photoacoustic Tomography for Neuroimaging with Deep Learning	2020	pixel-wise deep learning (Pixel-DL)	It contains pixel-wise interpolation governed by the physics of photoacoustic wave propagation.
	Deep-learning image reconstruction for real-time photoacoustic system	2020	a network using 3-D transformed arrays a multi-channel input 'upgUNET'	There is Reformatting raw channel data into a multi-channel array as a pre-processing step.
PAM reconstruction	Photoacoustic Microscopy with Sparse Data Enabled by Convolutional Neural Networks for Fast Imaging	2020(arxiv)	ResNet with Residual Block and SE Block	The CNN model utilized both squeeze and excitation blocks and residual blocks to achieve the enhancement.
	Deep Learning Enables Superior Photoacoustic Imaging at Ultralow Laser Dosages	2021	multitask residual dense network (MT-RDN)	It utilized an innovative strategy of integrating multi supervised learning, dual-channel sample collection, and a reasonable weight distribution.
PA image processing	De-Noising of Photoacoustic Microscopy Images by Attentive Generative Adversarial Network	2022	an attention enhanced generative adversarial network	It is an attention enhanced GAN that uses an improved U-net generator to remove noise from PAM images.
	High-resolution photoacoustic microscopy with deep penetration through learning	2022	Wasserstein distancegenerative adversarial Network (WGAN)	It uses Wasserstein distance to replace the Jensen–Shannon divergence as the objective to be opti mized.
	Photoacoustic Image Classification and Segmentation of Breast Cancer: A Feasibility Study	2018	AlexNet & GoogLeNet	It introduces the pre-trained AlexNet and GoogLeNet-based transfer learning for photoacoustic breast cancer classification.
PA signal processing	Deep Neural Network-Based Sinogram Super-Resolution and Bandwidth Enhancement for Limited-Data Photoacoustic Tomography	2020	U-Net (Hybrid) variant	It provides the generalization by effectively modeling the negative values in sinogram through the final layers while maintaining the advantages of ReLUs in U-Net architecture.
	Photoacoustic digital brain and deep-learning-assisted image reconstruction	2023	U-Net structure	It Uses a combination of simulation, MRA, and MRI to obtain prior ground truth images.
	Temporal and spectral unmixing of photoacoustic signals by deep learning	2021	conditional generative adversarial network (cGAN)	It automatically learns a loss that adapts to the data and distinguishes photoacoustic signals with phase differences

Chart 1. Chart of Recent works on Deep Learning-powered photoacoustic imaging.

which was initially applied to image segmentation and won the championships of the ISBI 2015 Cell Tracking Challenge and Caries Detection Challenge. To this day, U-Net has inspired the development of many network structures, and more and more deep learning strategies continue to extend based on U-Net. The U-Net network structure consists of a contracting path for capturing context and a symmetric expanding path for precise localization. U-Net can also be combined with data augmentation techniques to achieve end-to-end training with a small amount of data input [11–13]. Due to its network structure resembling the letter "U", it is named U-Net. The initial U-Net network structure is shown in Fig. 2(a).

In U-Net network structure, the vertical arrows form the processes of the encoder and decoder, while the horizontal arrows represent skip connections that jump across multiple layers. Its multi-layer encoder and decoder structure together constitute an overall layout resembling the letter "U". The left part of U-Net is the encoder, and the right part is the decoder. Let's discuss the encoder and decoder in detail.The encoder is responsible for extracting features from the input image. It gradually reduces the size of the feature map and increases the number of channels through multiple convolutional layers to extract more abstract features. Its structure consists of four blocks. Each block is composed of a 3×3 convolution (using the ReLU activation function) and a pooling layer with a stride of 2×2. After processing through the four blocks, the feature map is gradually reduced. The output of the encoder is passed to the decoder, and at the same time, skip connections are made between the output of each stage of the encoder and the symmetric stage of the decoder to preserve the detailed information of the feature map.

U-Net was initially applied to image segmentation, as shown in Fig. 2 (b) and (c). Fig. 2(b) demonstrates the U-Net's overlapping-tile strategy for seamless segmentation of images of arbitrary sizes by predicting the segmentation results of small selected areas through inputting the large selected frame image. Fig. 2(c) shows the process of observing HeLa cells

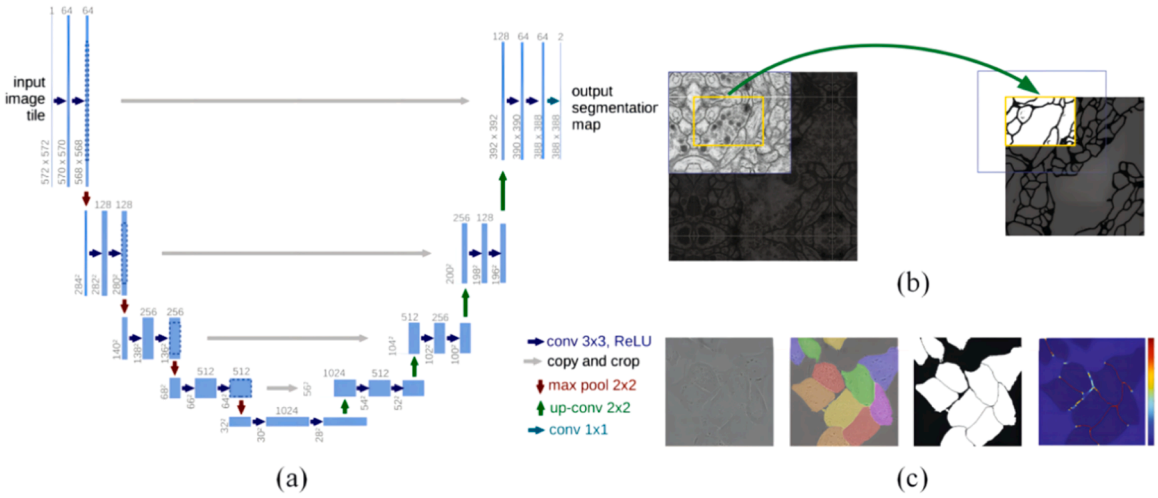

Fig. 2. Network structure and image segmentation; (a) Basic structure of U-Net network; (b) Seamless segmentation effect; (c) Progressive treatment of Hela cells, the four results are the original image, the image overlapped with the true value segmentation, the generated segmentation mask, and the result of using pixel loss weight mapping.

using differential interference contrast microscopy, where the four images represent the original image, the image overlaid with ground truth segmentation (different colors indicate different stages of HeLa cells), the generated segmentation mask (white represents foreground, black represents background), and the result using pixel loss weight mapping.

U-Net's overlapping-tile strategy has been widely used in medical image segmentation, effectively handling images of any size and achieving relatively accurate segmentation results. Meanwhile, U-Net can achieve end-to-end training with a small amount of data input by combining data augmentation and pixel loss weight mapping methods, making the network robust and capable of generalizing well.

2.1.2. The development of U-Net

U-Net is one of the currently popular network architectures, which was initially applied to image segmentation. With the continuous development and in-depth research of deep learning frameworks, the network structure of U-Net has also been continuously optimized and improved. More and more deep neural network structures have been discovered and combined with the U-Net architecture to further improve network efficiency. In addition, U-Net has been widely used in fields such as image reconstruction, image super-resolution, semantic segmentation, and signal processing, and has achieved good results.

In 2016, Cicek et al. proposed the 3D U-Net based on U-Net, which is used for 3D image segmentation [14]. Compared with U-Net, 3D U-Net only uses three downsampling operations and a normalization layer after each convolutional layer. It is worth noting that both 3D U-Net and U-Net do not use random dropout layers. In the 2018 MICCAI Brain Tumor Segmentation Challenge (BRATS), the team of the German Cancer Research Center used 3D U-Net and achieved the second place in the challenge with only a few modifications [15]. This indicates that compared to many new networks, 3D U-Net still has significant advantages.

In 2018, residual U-shaped network (Res-UNet) and dense U-shaped network (Dense-UNet) were born based on the U-Net architecture. Res-UNet and Dense-UNet are inspired by residual connections and dense connections, respectively, replacing each sub-module of U-Net with a form of residual connection or dense connection. Among them, dense connection means that the output of a layer in the sub-module is used as part of the input of subsequent layers, while any layer's input comes from the combination of outputs of previous layers. Res-UNet has been applied to the segmentation of retinal images [16], while Dense-UNet has been used to remove artifacts in images [17], which is the first case of using the U-Net architecture for image processing. The authors pointed out that U-Net is the most widely used CNN architecture for applying deep learning and post-processing methods to sparse tomographic image reconstruction [18]. It has many characteristics suitable for artifact removal, such as multi-level decomposition and multi-channel filtering. Moreover, on both synthetic data and experimental data [19], it shows better performance in removing sparse PAT image artifacts than iterative methods.The core idea of DenseNet is the Dense Block. In a Dense Block, the input of each layer is a concatenation of the outputs from all previous layers. Due to the direct connections between each layer and all preceding layers, DenseNet can effectively utilize parameters, resulting in a model with fewer parameters and reduced risk of overfitting. With the dense connectivity design, every layer in DenseNet has direct access to the feature maps from previous layers, facilitating feature propagation and reuse, which helps in learning richer feature representations.

Lin et al. proposed a robust deep learning network for ultrasonic photoacoustic microscopy with two modes dense network [20](US-PAM DenseNet), aimed at improving the performance of the model in distinguishing malignant from non-cancerous tissues based on co-registration of dual-mode ultrasound (US) and PAM images, as well as individualized normal reference images, as training. In Fig. 3, the US-PAM DenseNet similarly classifies the entire US-PAM B scan by ROI grade and computes the ROI heat map, highlighting the rectal cancer region.In Fig. 3(a), Five channels are generated from the selected ROI as the model input, which has dimensions of $128 \times 64 \times 5$. Solid arrows indicate data flows and connections inside the model: different colors correspond to different data origins. Connections are made between every pair of layers in the DenseNet architecture. The model has three layers, with 64 initial kernels in the first layer, a kernel growth rate of 12 from one layer to the next, and block repetition numbers of 4, 8, and 6 respectively for the three layers. The size of each model layer is marked under the layer icons.

In the same year, U-Net began to be applied to direct PAT reconstruction of sparse data from raw sensors.

Guan et al. proposed a new deep learning method called Pixel-DL (Pixel-wise Deep Learning) [21]. It first utilizes pixel-wise interpolation controlled by the physical propagation of photoacoustic waves, and then employs convolutional neural networks (CNNs) to reconstruct images. Synthetic phantom data from mouse brain, lung, and retinal vascular system were used for training and testing. The results show that Pixel-DL achieves comparable or better performance compared to iterative methods, making it suitable for real-time photoacoustic tomography (PAT) rendering and improving image reconstruction quality in

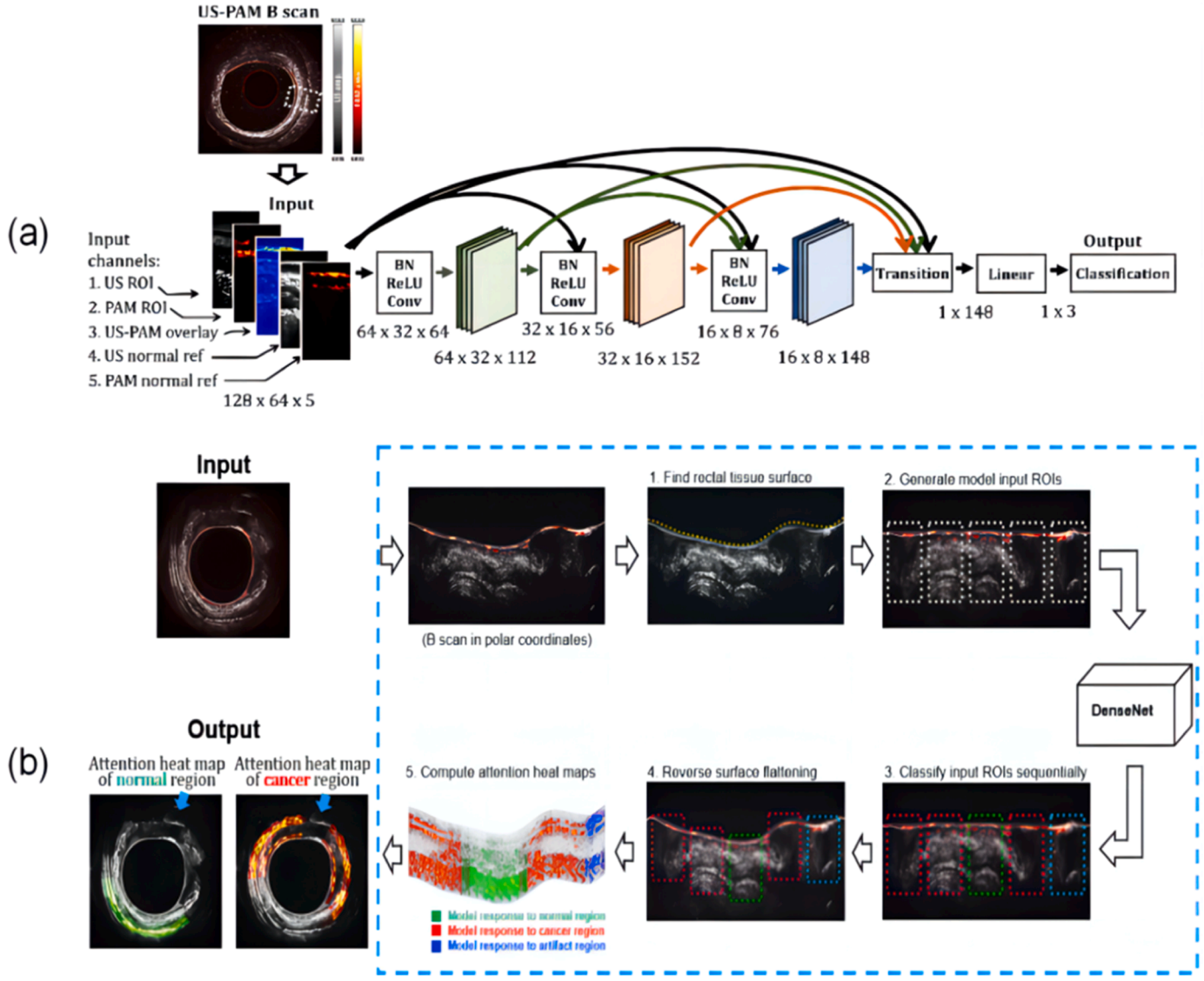

Fig. 3. The architecture of US-PAM DenseNet and the application of US-PAM DenseNet in generating thermal maps of suspicious tumor regions; (a) The white dotted box shows an example ROI selected from a co-registered US-PAM B scan. (b) The white dotted box shows an example ROI selected from a co-registered US-PAM B scan. Five channels are generated from the selected ROI as the model input, which has dimensions of 128 × 64 × 5. Solid arrows indicate data flows and connections inside the model: different colors correspond to different data origins. Connections are made between every pair of layers in the DenseNet architecture. The model has three layers, with 64 initial kernels in the first layer, a kernel growth rate of 12 from one layer to the next, and block repetition numbers of 4, 8, and 6 respectively for the three layers. The size of each model layer is marked under the layer icons.Pipeline for applying US-PAM DenseNet to diagnose a whole US-PAM B scan and generate an attention heat map of suspicious cancer regions to facilitate surgeons decision making. The processing pipeline is illustrated in the blue box. In steps 3, 4, and 5, green dotted boxes show the ROIs that US-PAM DenseNet classifies as normal, red shows the cancer ROIs, and cyan shows artifacts. In step 5, guided backpropagation is computed for all three potential classification outcomes, i.e., normal, cancer and artifact, weighted with their respective prediction scores.

limited-view and sparse PAT scenarios.Fig. 4(c) shows the improvement of Pixel-DL.Comparable to iterative reconstruction, Pixel-DL had similar performance for the fundus vasculature and outperformed it for the lung vasculature dataset.In this work, three different CNN-based deep learning methods for limited-view and sparse PAT image reconstruction were used, as shown in Fig. 4(a). Fig. 4(a)(I) shows inputs into the CNN for each deep learning approach. The Post-DL CNN implementation used residual learning which included a skip connection between the input and final addition operation. The initial Pixel-DL input contains "N" feature-maps corresponding to the number of sensors in the imaging system; (II) The FD-UNet is comprised of a contracting and expanding path with concatenation connections; (III) The output of the CNN is the desired PAT image. In Post-DL, residual learning is used to acquire the final PAT image. In Post-DL, the sensor data is reconstructed into an image with artifacts using time reversal, and then CNN is utilized as a post-processing step to remove the artifacts and enhance the image. In Pixel-DL, window-correlated information in the sensor data is interpolated on a pixel-by-pixel basis and mapped to the image space. In the improved Direct-DL implementation (mDirect-DL), a combination of linear interpolation and downsampling is used to ensure that the interpolated sensor data has the same dimensions as the final PAT image.In Fig. 5(b)(IV) The red semi-circle represents the sensor array, and the gray grid represents the defined reconstruction grid. In Fig. 5(b)(V)Color represents the time at which a pressure measurement was taken and is included to highlight the use of time-of-flight to map the sensor data to the reconstruction grid.

In 2019, Lan et al. proposed a Y-Net network based on the U-Net idea [22]. Unlike the general U-Net, Y-Net has two inputs and one output, i.e., two encoders and one decoder. By using the measured raw data and the beamformed image as inputs, Y-Net solves the PAT image

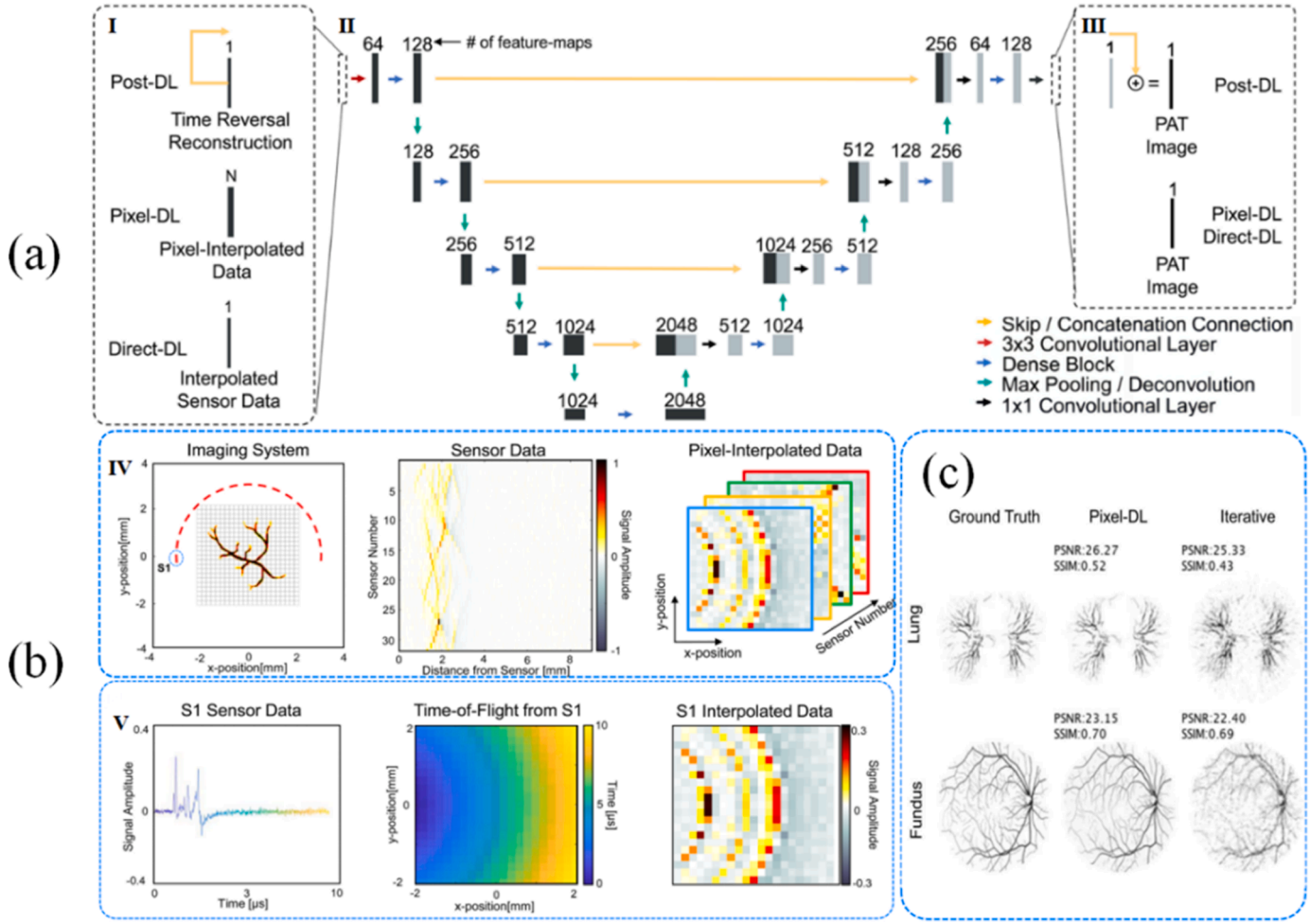

Fig. 4. The proposed FD-UNet network architecture, the introduced pixel interpolation process and PAT sensor data acquired with 32 sensors and a semi-circle view; (a)(I) Inputs into the CNN for each deep learning approach. The Post-DL CNN implementation used residual learning which included a skip connection between the input and final addition operation. The initial Pixel-DL input contains "N" feature-maps corresponding to the number of sensors in the imaging system; (II) The FD-UNet is comprised of a contracting and expanding path with concatenation connections; (III) The output of the CNN is the desired PAT image. In Post-DL, residual learning is used to acquire the final PAT image; (b)(IV) There are Schematic of the PAT system for imaging the vasculature phantom. The first sensor (S1) is circled and used as an example for applying pixel-wise interpolation to a single sensor; The PAT time series pressure sensor data measured by the sensor array; Resulting pixel-interpolated data after applying pixel-wise interpolation to each sensor based on the reconstruction grid; (V) There are Sensor data for S1; Calculated time-of-flight for a signal originating at each pixel position and traveling to S1; Pressure measurements are mapped from the S1 sensor data to the reconstruction grid based on the calculate time-of-flight for each pixel.(c)Data were acquired respectively on images of lung and fundus vasculature.

reconstruction problem, which can also be called hybrid processing. Inspired by the Y-Net network idea, in 2022, Guo et al. proposed an attention-guided network based on multi-feature fusion (AS-Net) for PA reconstruction, aiming to solve the PA reconstruction problem under sparse conditions of ultrasonic transducers in photoacoustic tomography [23].

In Fig. 5, Firstly, 2-D PA raw data is transformed into a 3-D square matrix by Folded Transformation (FT). Then AS-Net produces the multi-feature fusion base on the attention mechanism for PA reconstruction. ASKF-Net architecture consists of a basic PA reconstruction (BPR) module, semantic feature extraction (SFE) module, and feature fusion (FF) module. BPR module is a modified Auto-Encoder architecture used to reconstruct images from the PA signal, while the SFE module aims to extract semantic features from the DAS image. FF module is used to fuse the semantic feature into the output of the BPR module and generate the final reconstructed image.

In 2022, MENG et al. proposed a deep tissue acoustic-resolution photoacoustic microscopy technique based on a two-stage deep learning network [24]. This technique can adaptively restore high-resolution photoacoustic images at different defocusing depths, thereby partially solving the problem of poor imaging quality of off-focus plane targets. Specifically, the network structure consists of two stages. The first stage of the deep learning network is used to reconstruct the region far away from the focus, and the second stage reconstructs the region near the focus. In order to achieve image reconstruction, a residual U-shaped network with attention gates (Res-UNet_AG) is also designed in this study.

2.2. *Generative adversarial network*

Super-Resolution Generative Adversarial Network (SRGAN) is a network proposed by Christian Ledig et al.in 2017 in their paper [25]. This paper presents a super-resolution method based on generative adversarial networks, which can convert low-resolution images into high-resolution and realistic images. The appearance of SRGAN has attracted wide attention in the field of image processing and has achieved good results in practical applications.

The main body of the SRGAN network consists of two independent and combinable training network structures, namely the generator and discriminator. The network loss function consists of a perceptual loss

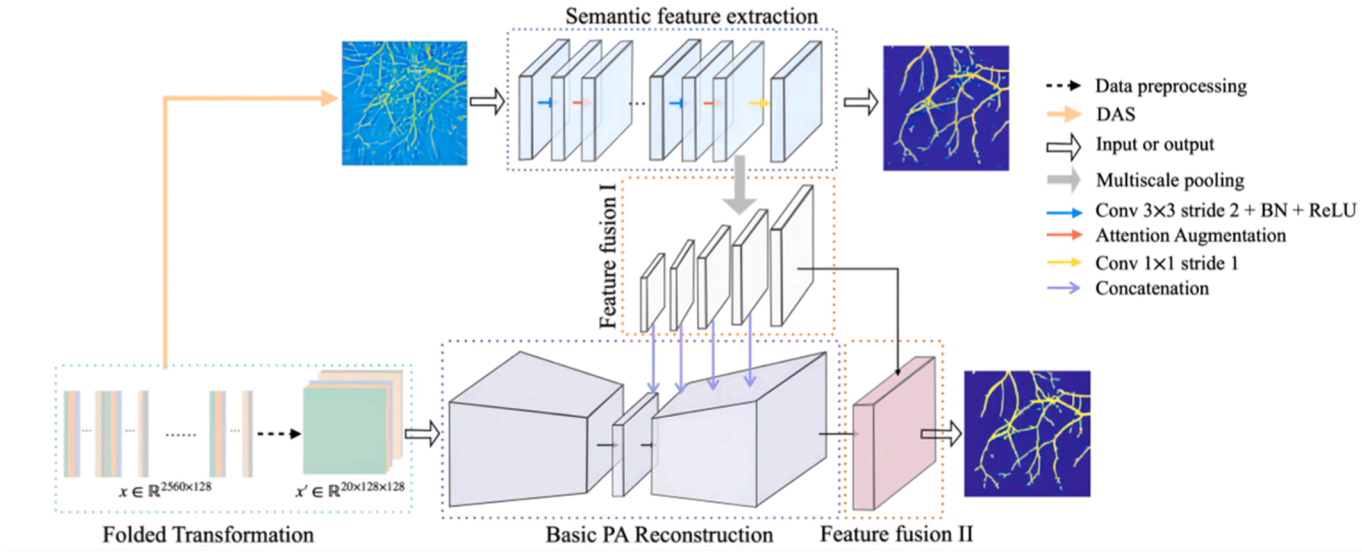

Fig. 5. Illustration of the reconstruction framework from Guo et al., which includes PA raw data preprocessing and AS-Net reconstruction network.

function, which corresponds to the content loss of the generator, and an adversarial loss function of the discriminator weighted by the two. The training idea of the network is to input low-resolution images into the generator network to obtain super-resolution images, and then input the super-resolution images into the discriminator network. The discriminator network will output a binary result representing the "truthfulness" of the image. If the discriminator outputs that the image is fake, the loss value will be returned to the generator network for further training; if it is true, the discriminator will continue to train. In summary, the generator and discriminator mutually constrain each other and train with a related loss function. The goal of the generator can be understood as "deceiving" the discriminator, while the goal of the discriminator is to optimize the authenticity of the generator.The network iteration stops when the minimum error is reached, and the generator at this time is taken as the final result of the network training. The content loss function used in this paper is different from the spatial loss and is based on the feature space Mean Squared Error (MSE) loss of a certain layer weight of the VGG19 model [26,27]. This loss function can improve the semantic recognition and readability of the image. The paper compares the reconstruction effects of three methods, interpolation, ResNet, and SRGAN (the generator is ResNet). The results show that SRGAN has a better effect in extracting image features. Fig. 6 shows the SRGAN network structure.

In 2018, Wang et al. proposed an enhanced super-resolution generative adversarial network to solve the artifacts generated by SRGAN in image super-resolution [28]. The super-resolution generative adversarial network (SRGAN) can generate realistic textures during the single image super-resolution process. However, the details of the reconstructed image are often accompanied by artifacts. In order to further improve the visual quality of SRGAN, the authors conducted in-depth research on SRGAN and improved three key components: network structure, adversarial loss, and perceptual loss, to obtain an enhanced SRGAN (ESRGAN). The residual-in-residual dense block (RRDB) without batch normalization is introduced as the basic network building unit. In addition, the authors used the idea of relative error to let the discriminator predict the relative realism instead of the absolute value. Finally, the activation before the feature is used to improve the perceptual loss and provide stronger supervision for brightness consistency and texture

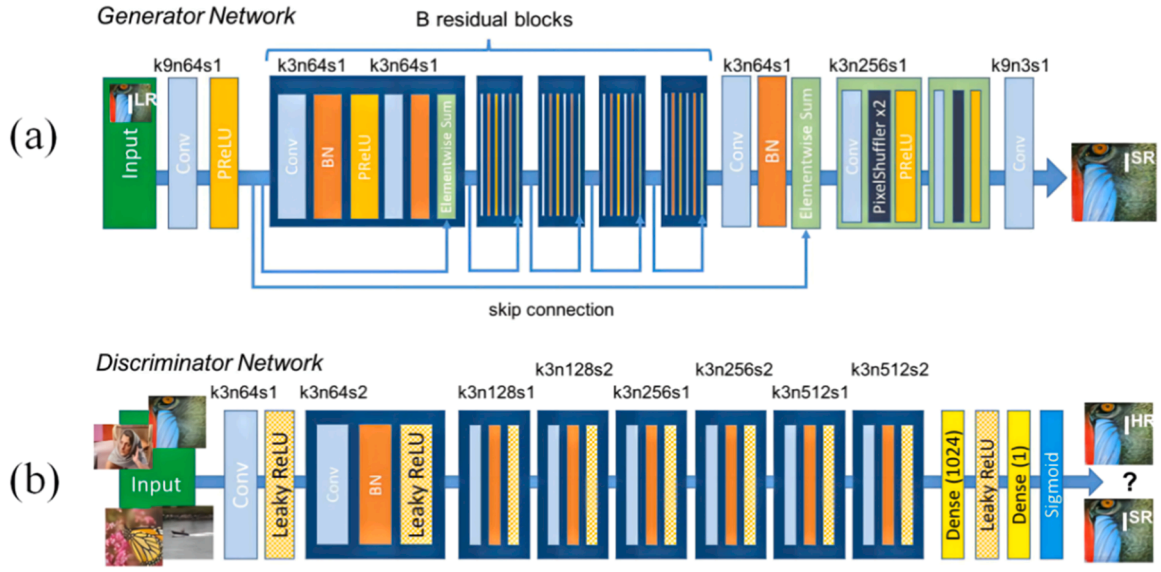

Fig. 6. The SRGAN structure. (a) The Generator Network of SRGAN structure.The main function is to generate parameters. (b) The Discriminator Network of SRGAN structure.The main function is to cooperate and build generate corresponding parameters which could be used to train Generator Network for improvement.

restoration. The network improvements are shown in Fig. 7.

Fig. 7(a) shows the network flow of ESRGAN, where Residual Blocks in the SRGAN generator are treated as a type of feature extraction layer called Basic Block. Fig. 7(b) shows that ESRGAN removes the normalization layer (BN, Batch Normalization) in the Residual Blocks and adds Dense Blocks as feature extraction layers after Residual Blocks. In each Residual Block, the BN layer appears twice to ensure that the network gradient does not explode, but it significantly slows down the network training speed and reduces the amount of feature information, which leads to a significant impact on the reconstruction effect. Dense Blocks can also solve the problem of gradient diffusion and explosion, and will not reduce the image reconstruction effect. The authors also proposed optimizing the loss function of the discriminator, which is different from the standard discriminator in SRGAN. It estimates the real probability of the input image, that is, trying to predict the probability that the real image is relatively more realistic than the fake image. This modification helps to learn sharper edges and more detailed textures.

2.3. Residual blocks and residual networks

U-Net is constantly developing in terms of width and depth in the field of photoacoustic imaging, especially in the areas of reconstruction and image processing in PAM. With the continuous improvement of imaging speed, image effectiveness, and the requirements for biaxial resolution, the ordinary U-Net network structure is difficult to meet the situation of deepening the width and depth. The introduction of Residual Blocks into U-Net has enabled the network to reach unprecedented depths, and Residual Blocks have been fully applied not only in U-Net but also in other network structures such as the generator network of SRGAN. Szegedy et al. summarized the impact of network structures including Residual Blocks on image recognition tasks [29]. The authors analyzed the inherent importance of residual connections for training very deep neural networks. High-performance networks are often very deep, and deep neural networks are difficult to train compared to shallow neural networks because of the problems of gradient vanishing and exploding, as well as the increased computational complexity that increases the hardware requirements for network training. Skip connections are an important component structure of Residual Blocks, which can obtain weights from a certain layer of the network layer and quickly feedback to another layer, usually skipping connections to deeper layers. This structure can reflect the weights of the lower layers of the network in the next layers of the network, thereby avoiding gradient vanishing and exploding problems, and improving the efficiency and stability of network training. Currently, it is common to use residual connections to replace filter cascading stages.

The authors also pointed out that optimizing convolutional neural networks with recognition performance as the goal can also be transformed into performance improvements in other tasks. Using Residual Blocks to construct residual networks (ResNet) that can train deep networks not only has good results in image segmentation but can also be further extended to other fields such as medical imaging. He et al. first proposed ResNet to solve the problems of gradient vanishing and exploding in deep networks while successfully increasing the number of network layers to the order of 10^3 while ensuring the constraint of the loss function. ResNet is composed of Residual Blocks [30,31]. The proposal of ResNet and Residual Blocks ensures the effectiveness of training deep neural networks. Even if the network depth reaches the level of 10^3, the loss function can be optimized to ensure a reduction in training error.

3. Application of deep learning in photoacoustic imaging

3.1. Photoacoustic image reconstruction

Deep learning methods, as a new information mining method, have a wide range of applications in multidimensional information processing, such as reconstruction, denoising, super-resolution, etc., and have achieved many good results.Currently, there are also some non-iterative reconstruction schemes proposed, such as direct estimation, PA signal model reconstruction, and PA signal or image enhancement through deep learning.

The direct reconstruction method solves the PA wave equation, which captures the mapping from signal to image with the PA signal as input. Waibel et al. established a direct estimation from light and sound signal detector data to PA imaging [32], input the synthetic data of the 128-element linear detector into an improved U-Net, and reconstructed the final initial PA pressure signal. Schwab et al. used deep learning to learn the weights of reflected data on different channels and trained neural networks for vessel phantom. Meanwhile, the model used Shepp-Logan phantom to verify. They also proposed a data-driven regularization method [33], which significantly suppresses noise by applying truncated singular value decomposition (SVD) [34] and then restoring truncated SVD coordinate coefficients. Lan proposed using three different sensor data (2.25 MHz, 5 MHz, 7.5 MHz) as input and using U-net for direct reconstruction. Feng et al. improved Res-Unet for direct reconstruction of simple phantoms and compared it with some U-net models [35]. Tong Tong [36] trained a feature pyramid network (FPnet) as post-processing using in vivo data. Mohammad Abu Anas et al. proposed a deep CNN network structure for beamforming PA data [37], which consists of five dense blocks consisting of convolutional layers with different sizes. The article discusses the influence of variable sound speed on this method and verifies its robustness under variable sound speed.

In particular, in the PAI system, due to the existence of optical

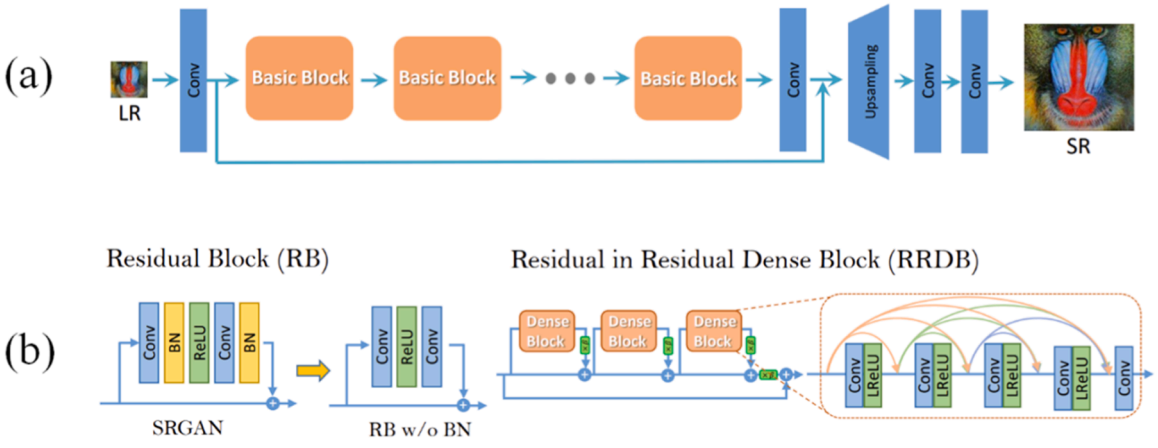

Fig. 7. ESRGAN Network; (a) ESRGAN flow; (b) Modified Residual Block and RRDB.

scattering, the effective excitation degree of deep targets is affected to a certain extent, which is a problem that cannot be ignored. To solve this problem, Johnstonbaugh et al. designed an encoder-decoder network for predicting objects in deep tissue [38]. This study introduced acoustic and optical attenuation in simulation and compared it with actual situations. Allman et al. used VGG16 beam to detect and eliminate reflections from point sources [39], and all experiments used simple (point) phantoms. This method uses neural networks to eliminate artifacts caused by reflections, which greatly improves imaging speed.

In summary, there is currently widespread research on non-iterative reconstruction methods, indicating the urgent need for real-time photoacoustic imaging. Not only in PAT, but also in PAM and PAE, there are various problems with real-time imaging. Factors affecting imaging speed in traditional PAM algorithms include the repetition rate of the excitation light pulse, scanning mechanism, signal preprocessing, and image post-processing. Common solutions such as increasing pulse repetition rate, sampling scanning method, and pixel stacking have been verified, but at the same time, they also face the lack of imaging quality. Considering the current research status of PAE, there are not many studies on PAE reconstruction using deep learning, so it will not be further described here.

3.1.1. Photoacoustic tomography image reconstruction

Image reconstruction is an important part of photoacoustic computed tomography (PAT), which is responsible for converting the raw signals received by the ultrasound transducer into an initial pressure distribution image. Due to the ill-posed nature of photoacoustic imaging and the lack of an accurate inverse model in practical situations (limited field of view and sparse sampling), photoacoustic tomography reconstruction is still challenging. In PAT, the purpose of image reconstruction is to reconstruct the initial PA pressure distribution, which is positively correlated with the optical absorption intensity of biological tissues. The sensor array receives PA signals P ($P(r, t) \mid r, t$ represents 3D position and time) excited by short-pulsed laser at different ionization levels, and based on these PA signals, the acoustic-thermal information H (r, t) is reconstructed through some inverse reconstruction methods, and then A (r) is further reconstructed, i.e., the distribution of tissue optical absorption intensity. Currently, the most commonly used inverse reconstruction methods include model-based methods such as back-projection (BP) and time reversal (TR); sparse data-based reconstruction methods such as compressed sensing (CS), wavelet transform (WT), and discrete cosine transform (DCT); data mining methods such as deep learning; and model-based iterative methods. Among them, back-projection method is the most widely used, while BP and its derived algorithms such as filtered backprojection (FBP) are considered the most famous PAT reconstruction algorithm due to their simple implementation [40,41].

If experimental conditions are sufficient, i.e., a sufficiently large and dense ultrasound transducer array is distributed on the inner radius of a circular or elliptical detector, the photoacoustic inverse problem of the backprojection method can be expressed as follows [40]:

$$A(r) \propto \int d\theta \frac{1}{t} \frac{\partial p(r_0, t)}{\partial t}\Big|_{t=(|r_0-r|/c)} \tag{1}$$

Here, r is the position of the acoustic pressure; c is the speed of sound; θ is the angle between the ultrasound transducer and the acoustic pressure signal; r_0 is any position of the ultrasound transducer on the inner radius of the circular or elliptical detector; $p(r_0, t)$ is the known condition for the inverse operation, i.e., the acoustic pressure signal received by the ultrasound array at that position; $A(r)$ is the spatial distribution of tissue optical absorption intensity.

Kim et al. proposed to modify 2D raw data (with time and detector dimensions) into a 3D array (with two spatial dimensions and one channel dimension), where the channel data packages correspond to the propagation delay distribution at a spatial point and serves as the input to the neural network [42]. Traditional popular machine learning methods train on incomplete images obtained under ill-posed conditions through standard reconstruction methods [43–46]. Due to the loss of previously captured weak information that is difficult to reconstruct, the fine structure of the reconstructed image is often unsatisfactory. Kim's method trains on the basis of the first step of most traditional reconstruction methods, greatly simplifying the learning process. The expansion of the channel dimension preserves more information and improves learning accuracy.

Fig. 8(a) shows the input data of the neural network. Using simple acoustic propagation physics rules in r (x, z) and the linear array transducer system, 3D transformed data is obtained by the propagation delay distribution of specific image points at different depths, which are used as inputs to the network.

Fig. 8(b) shows the CNN network architecture used in the study. Prior to data input into the network, pre-processing was performed by looking up a priori LUT tables on the original signal (2048×128 obtained by adding noise to real images), converting it into a $512 \times 128 \times 128$ data array containing delay information. Reformatting the original channel data into a multi-channel array as a preprocessing step improves learning efficiency for highly complex network structures. This neural network uses U-net as a basis and decomposes the signal through multi-scale feature mapping. By combining trainable networks with transformation methods, the structure of vascular networks was simulated in simulations and experiments. Overall, this method significantly improves image quality compared to traditional methods for reconstructing PA data, but loses a little complex absorption body geometry and may produce small artifacts.

Antholzer et al. proposed a direct and efficient reconstruction algorithm based on deep learning for the sparse data problem in reconstructions [47]. The first step uses the PAT filtered backprojection algorithm, followed by optimizing the reconstruction results using the U-net architecture. It not only solves the time-consuming forward and adjoint problems, but also has better imaging effects than direct filtered backprojection algorithms, and performs similarly to existing iterative methods for sparse data PAT.Because iterative algorithms have their own limitations. For example, the reconstruction quality strongly depends on the used a-priori model about the objects to be recovered. For example, TV minimization assumes sparsity of the gradient of the image to be reconstructed. Such assumptions are often not strictly satisfied in real world scenarios which again limits the theoretically achievable reconstruction quality.On the other hand, iterative reconstruction algorithms tend to be slower as they require repeated application of the PAT forward operator and its adjoint.Antholzer further proposed another three-layered S-net network for direct reconstruction for the sparse data problem, where the input is an image with artifacts and a real ground image obtained through a priori method. In simulation experiments, S-net can effectively eliminate artifacts caused by sparse data and greatly improve reconstruction efficiency compared to traditional image reconstruction methods [48–51]. The author also summarizes a deep network generally used for image enhancement after PAT image reconstruction. In the first step, the FBP algorithm (or another standard linear reconstruction method, using FBP as an example here) is applied to sparse data. In the second step, a deep CNN is applied to intermediate reconstruction, which outputs an image with almost no artificial artifacts.This can be explained as a deep network with FBP in the first layer and CNN in the remaining layers.

Image reconstruction is also an important part of functional imaging, including blood oxygen detection and various molecular detections. Due to the fact that hemoglobin is the main substance absorbed by human cells below 1000 nm, PAT can quantitatively detect hemoglobin (HbO_2) and deoxyhemoglobin (HbR). Since the oxygen saturation (sO_2) of hemoglobin in normal tissue is higher than that in malignant tissue, sO_2 is an important physiological index of the body [52–54]. sO_2 is defined as the fraction of HbO_2 relative to the total hemoglobin concentration in the blood:

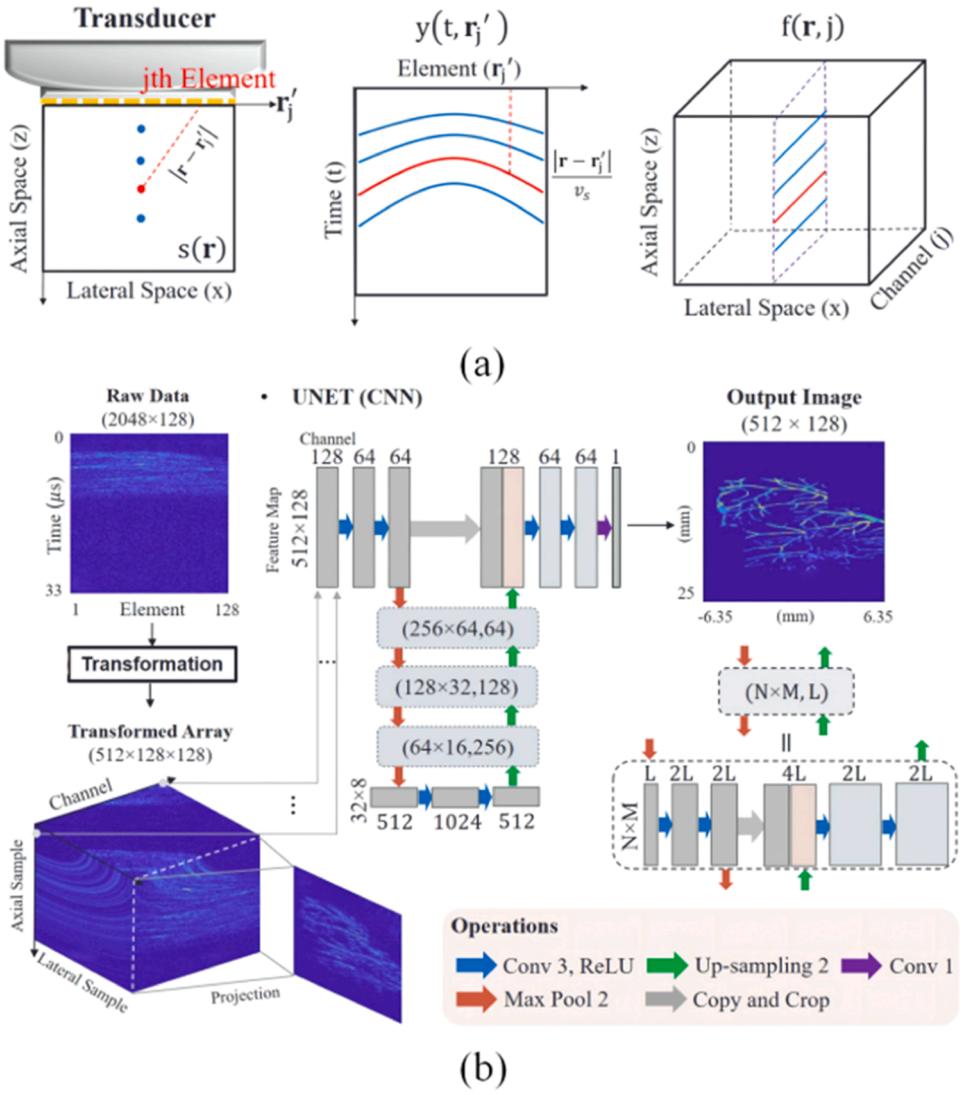

Fig. 8. System diagram. (a) Schematic of photoacoustic data acquisition; (b) CNN-Net.

$$sO_2(x,y) = \frac{C_{HbO_2}(x,y)}{C_{HbO_2}(x,y) + C_{HbR}(x,y)} \times 100\% \quad (2)$$

Here C_{HbO2} and C_{HbR} represent the concentration of oxygenated and deoxygenated hemoglobin, respectively, while x and y denote the spatial position. According to the above formula, the basic principle of using photoacoustic tomography (PAT) for quantitative blood oxygen synthesis imaging is that *HbO₂* and *HbR* have significant absorption differences at different wavelengths of light. Similarly, quantitative spectroscopic photoacoustic imaging (QS-PAI) is an imaging technique that measures at multiple wavelengths of light to provide information related to molecular composition [55]. The aim is to convert multi-wavelength PA images into a final image that mainly highlights the quantitative and accurate estimation of chromophore spatial concentration changes in scattering media. The main problem with QS-PAI is essentially an inversion problem of light propagation operators. The current common two-stage inversion strategy can be summarized as follows: 1. determining the absorption coefficient; 2. determining the chromophore concentration. Due to the characteristics of scattering media, accurate nonlinear inversion of spatially structured light flux is difficult to achieve, and it is unrealistic to rely on strict conditions, such as known scattering coefficients and homogeneous background optical properties. In the inversion process, linear substitution instead of nonlinear inversion is used to determine the absorption coefficient by using multi-wavelength PA images and light flux related to the absorption coefficient and scattering coefficient, which may result in large errors.

Cai et al. proposed the first deep learning framework Res-Unet for quantitative PA imaging [56]. Res-Unet takes the entire initial pressure image distributed at different wavelengths as input, so that reconstruction can best utilize all measurement signals. To prevent the degradation of deep networks, residual learning mechanism is adopted. In Res-Unet, comprehensive contextual information is extracted from multispectral initial pressure images to quantitatively estimate chromophore concentration or sO2. The CNN architecture implemented using U-net is used to measure object contours, perform optical inversion, estimate the main absorbing chromophores and their absorption spectra, and perform linear decomposition.

Yang et al. proposed a deep residual and recursive neural network (DR2U-net) for quantitative estimation of hemoglobin oxygenation in photoacoustic imaging [57]. The proposed DR2U-net can extract flux distribution information from the optical absorption image using only two wavelengths of light in Monte Carlo simulations, and then generate quantitative sO2 images. Through testing on simulated biological tissues, the measured sO2 results have high accuracy, with an error as low as 1.27 %, compared to traditional linear mixing methods (48.76 %). In the network structure, deep networks can enrich feature information, so

the article uses residual connections mentioned above to solve possible gradient explosion and improve training accuracy [58]. Batch normalization is also used to accelerate convergence speed and reduce covariate shift. This approach effectively reduces the nonlinear effect of scattered light flux while increasing system robustness and reducing noise interference.

Rajendran and Pramanik proposed a novel deep learning architecture for tangential resolution in circular-scan photoacoustic tomography (PAT) imaging system [59]. The article uses a U-Net-based convolutional neural network combined with 9 residual blocks to improve the tangential resolution of PAT images. This is the first study to use a U-Net structure neural network for tangential resolution of PAT images. In general, in photoacoustic tomography, axial resolution does not change and is influenced by the detection bandwidth. However, tangential resolution will change with the size of the detector aperture. Especially when the aperture size is smaller, the tangential resolution is higher. However, if a small-aperture detector is used, the sensitivity of the sensor will decrease. Therefore, a large-aperture detector is the main choice for circular-scan PAT imaging systems. The proposed TARES network was implemented using Python 3.7 and TensorFlow v2.3 deep learning library [60]. The model was trained using simulated PA data and validated using experimental model data and human PA images [61–64]. The training model can detect data well and simulate body images of humans and animals.

Gao et al. proposed a U-Net-based convolutional neural network to extract effective photoacoustic information hidden in speckle patterns in a vascular network image dataset under porous media [65]. As shown in Fig. 9, human skull belongs to a typical multi-scattering medium, and traditional ultrasound imaging has many challenges in imaging deep and fine structures due to significant scattering of sound signals during excitation and reception. The article uses photoacoustic imaging principles and deep neural networks to solve the issues of frequency-domain wideband scattering in transcranial photoacoustic microvascular imaging and superposition of spatial domain main lobe and side lobe signals [66,67]. In short, the neural network can effectively extract valid information from highly blurred speckle patterns for rapid reconstruction of target images, providing broad application prospects in transcranial photoacoustic imaging [68,69].

3.1.2. Photoacoustic microscopic imaging reconstruction

Zhou et al. proposed a method using ResNet to improve the quality of sparse PAM images [70], which can simultaneously maintain good image quality and accelerate image acquisition speed [71–73]. In this work, the dataset used was PAM images of oak and magnolia leaf veins. Immerse the leaves in a container with black ink for more than 7 h, then place them on a glass slide and seal them with silicone gel (GE sealant). For each PAM image, use an OR-PAM probe with a resolution of 2 μ m, consisting of a beam profiler and 10 × Beam expander measurement at 256 × Scan leaf samples at 256 scanning points with a scanning step of 8 μ m. Finally, a real image dataset of 268 original fully sampled PAM images was obtained. Corresponding low pixel images pass through 2 × And 4 × Downsampling acquisition.

The proposed ResNet structure is shown in Fig. 10(a). The authors used 16 residual blocks and 8 squeeze and excitation (SE) blocks as the key part of feature extraction. Inspired by SRGAN [74], the residual block shown in Fig. 10(b) can extract features well in the SR task. The SE block with channel attention mechanism (as shown in Fig. 10(c)) helps network convergence and performance. The "Upconv" block consists of 2× upsampling layers and standard convolution layers (kernel size 3, filter number 256, stride 1). The Tanh activation function is used after the final output layer.

Zhao et al. proposed a multi-task residual dense network (MT-RDN) deep learning system and method [75]. The MT-RDN network adopts an innovative strategy combining multi-supervised learning, dual-channel sample collection, and reasonable weight allocation. The proposed deep learning method is combined with an improved OR-PAM system for application. This study obtained good images for the first time under ultra-low laser dose (reduced by 32 times). The network method aims to solve the challenges of image quality deterioration caused by low single-pulse laser energy and undersampling during high-speed imaging.

In the proposed system method, the original images (i.e., undersampled images obtained under low excitation laser energy) are collected at 532 and 560 nanometer wavelengths and assigned to two different network input channels input1 and input2 respectively. The

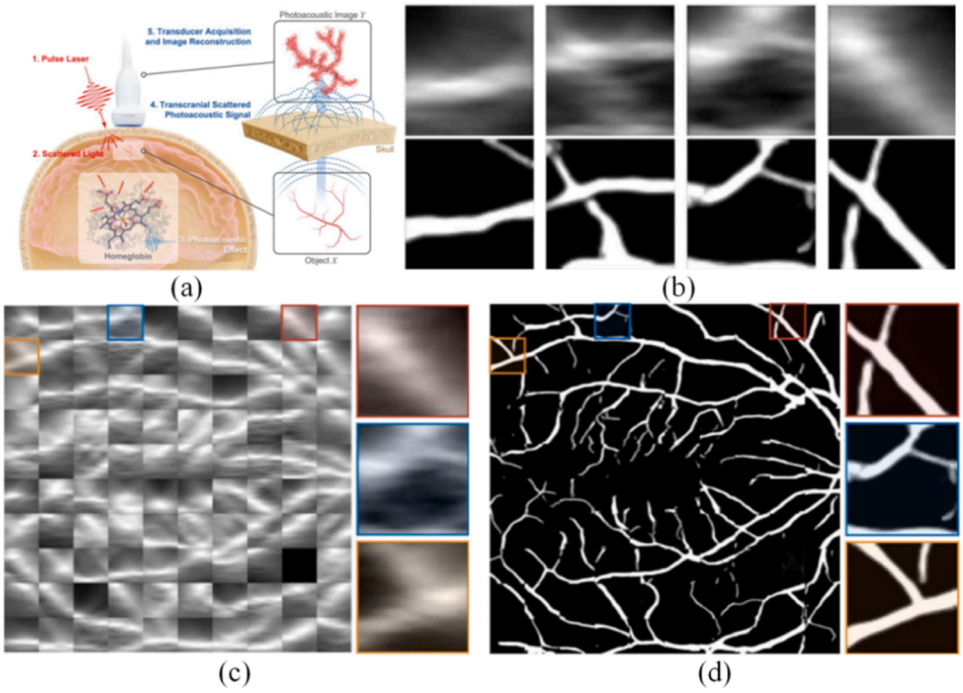

Fig. 9. Schematic diagram and method comparison diagram. (a) the schematic diagram of transcranial photoacoustic imaging; (b) the reconstruction effect of DAS and this network on plaque respectively; (c) and (d) the reconstruction effect of DAS and this network on whole image respectively.

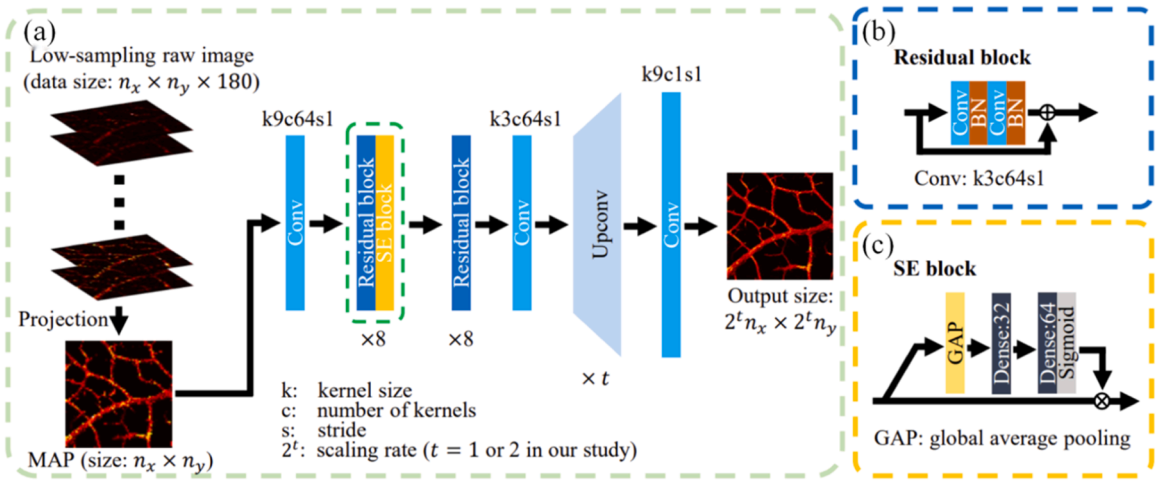

Fig. 10. Resnet Network diagram; (a) ResNet structure; (b) Residual Block structure; (c) Squeeze and Excitation (SE) Block.

low pixel input of the network is obtained by down-sampling the original image using 2x and 4x undersampling images at half of the single-pulse laser energy (i.e., ANSI limit of single-pulse laser energy), and then segmenting the original image as input to the MT-RDN network. The 2x undersampled image is cut into slices of 100 × 100 pixels, and the 4x undersampled image is cut into slices of 50 × 50 pixels. MT-RDN has three sub-networks. The first sub-network is used to process data input 1 (i.e., 532 nm data) to obtain output 1, and the second sub-network is used to process data input 2 (i.e., 560 nm data) to obtain output 2. Outputs 1 and 2 are further combined and processed by Sub-network 3 to obtain output 3. Ground truth images 1–3 are obtained from full-sampled images obtained at 532 nm and 560 nm ANSI limits of single-pulse laser energy, and ground truth images obtained using the Paivef method [76].

3.2. Photoacoustic image processing

The goal of image processing is to improve the quality and clarity of important details or targets in an image for specific applications by manipulating the image. Often, image enhancement is closely associated with the subsequent steps of photoacoustic image reconstruction. Image processing techniques such as noise reduction, smoothing, contrast stretching, sharpening, edge enhancement, and super resolution are commonly used to increase imaging readability and efficiency. These operations all belong to image processing, with the aim of improving the interpretability and effectiveness of the resulting image. Image processing often follows image reconstruction algorithms.

3.2.1. Improvement of signal-to-noise ratio of photoacoustic images

In photoacoustic (PA) signals, the initially acquired PA signal and image often suffer from low signal-to-noise ratio (SNR) due to the weak amplitude of the PA signal and strong random noise from external instruments and the environment. In practice, the PA waves generated by low-cost, low-energy laser diodes are very weak and almost buried by noise. Additionally, deep tissue imaging is accompanied by severe attenuation, such as scattering, leading to the problem of low SNR in PA signals [77]. Consequently, the reconstructed PA images have poor quality with noise. Therefore, effective denoising techniques are required for reconstructing artifact-free PA images from measurements containing noise signals [78–80]. Although traditional Kalman filters (KF) [81,82] can remove Gaussian noise in the time domain [83–86], they lack adaptability under real-time estimation conditions due to their fixed model. The effectiveness of the traditional KF relies on the proper definition of two key parameters: the system noise matrix (Q) and the measurement noise matrix (R). However, it is often challenging to obtain accurate statistical data for these parameters in practical situations. To overcome this challenge, there are existing methods for eliminating white noise. The most common one is data averaging, which has been used in PAI. However, it requires additional storage space for data and imposes high requirements on time [87]. In addition to white noise, electrical noise generated by the photoacoustic imaging system [88,89] and interference in the acquired photoacoustic signals can significantly degrade image contrast in multispectral photoacoustic tomography (MSOT).

He et al. proposed an attention enhanced GAN that uses an improved U-net generator to remove noise from PAM images [90]. The network does not need to manually select settings for different noisy images, but instead uses an attention enhanced generative adversarial network to extract image features and adaptively remove varying degrees of Gaussian, Poisson, and Rayleigh noise. The proposed method has been validated on both synthetic and real datasets, including phantom (leaf vein) and in vivo (mouse ear blood vessels and zebrafish pigment) experiments. The network structure diagram and denoising effect are shown in Fig. 12. To effectively capture features and distinguishing information with varying importance, an attention mechanism is applied in their network Different from regular CNNs which may treat all information equally, attention blocks additionally introduce attention weights for different feature channels or spatial positions Specifically, this method utilizes the attention block, ie., the GC block, to enhance the attention to long range dependencies and that better handle unexpected noise instance of focusing on signal pixels The detailed structure of the GC attention block includes 1 × 1 conversions and layer normalization GC blocks are placed after each standard unit block of the encoder in the generator.Fig. 11(a) shows the GAN network structure diagram. The network structure includes a generator and a discriminator. Fig. 11(b) displays a comparison of the results of neural network imaging and other methods in the mouse ear vascular region. (Scale bar: 250 μm. All images, excluding zoom images, share the same scale bar. The values in the colorbar indicate relative PA intensity) On the left side of Fig. 11(c) is the sample image before denoising, which includes mouse ear blood vessels, zebrafish pigment, and enlarged color box areas in the above samples. On the right is the denoised image, which includes mouse ear blood vessels, zebrafish pigment, and enlarged color box areas in the above image. (Scale: 500 μm).

3.2.2. Improvement of photoacoustic image resolution

Deep learning methods can also be applied to improve the resolution of photoacoustic (PA) images. Traditional acoustic-resolution PA imaging systems are often limited to imaging resolutions on the order of 100 micrometers due to the optical diffraction limit and the acoustic

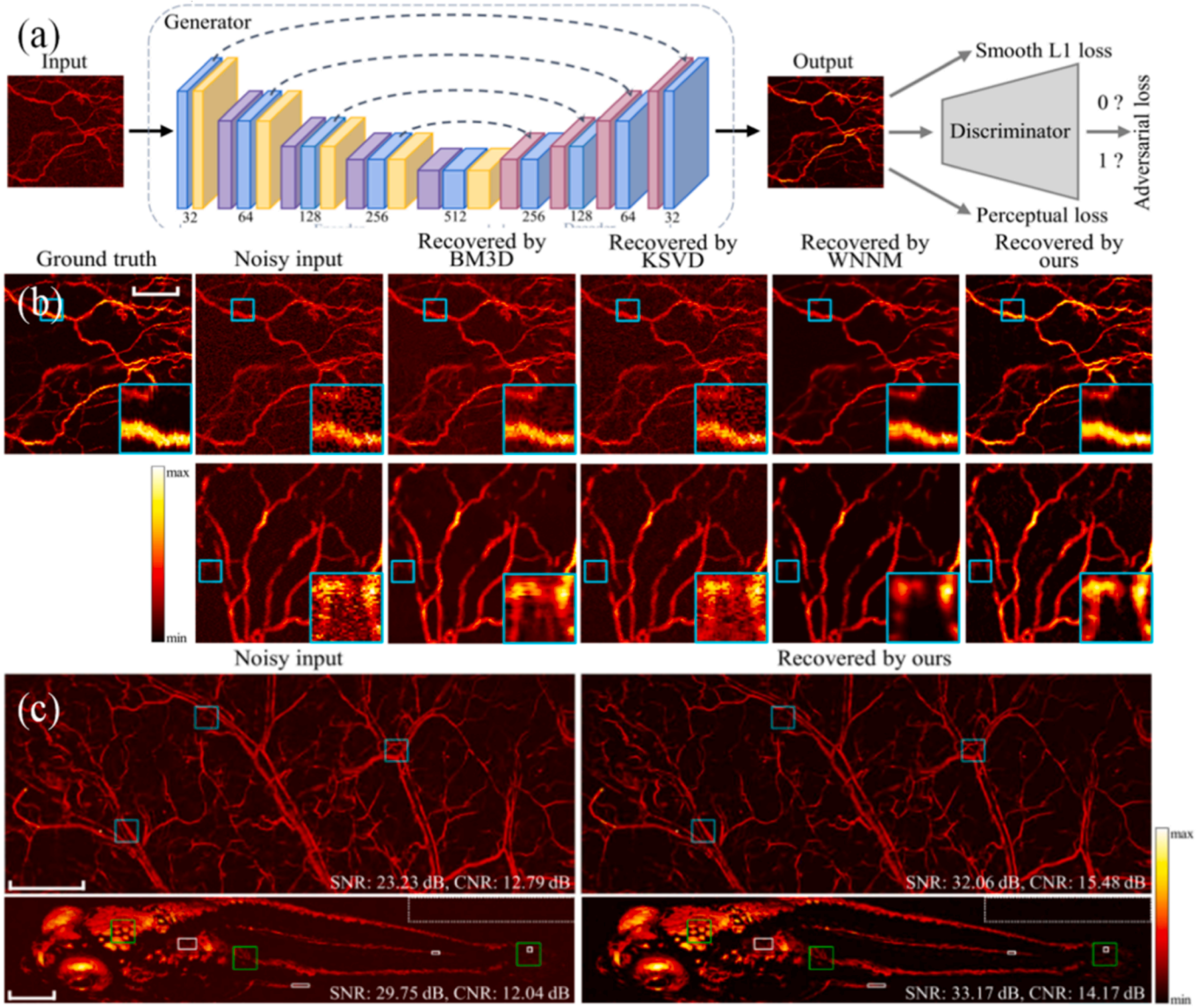

Fig. 11. Network structure diagram and denoising effect illustration. (a) Network structure diagram; (b) Representative results of the mouse ear blood vessel dataset acquired by in vivo experiment. Top raw: a representative sample from the synthetic noisy dataset; bottom row: a representative sample from the real noisy dataset; (c) Demonstration of denoising effects on mouse ear vasculature and zebrafish pigment.

diffraction limit. On the other hand, optical-resolution imaging systems can achieve spatial depths of around 1 mm due to the optical diffraction limit but have limited applicability in clinical medicine. Similar to the post-processing methods for sparse data or array-angle-limited problems in PA tomography (PAT), deep learning has been widely used in super-resolution reconstruction of photoacoustic images by implementing end-to-end image optimization.

Cheng et al. proposed a deep-penetration high-resolution photoacoustic microscopy technique based on deep learning generative adversarial network (GAN) architecture [91]. This method employed Wasserstein GAN (WGAN) as the training network to learn from low-resolution absorption-reconstruction photoacoustic microscopy (AR-PAM) images towards high-resolution optical-resolution photoacoustic microscopy (OR-PAM) images at the same depth. In this WGAN network, the generator takes AR images as input and generates high-resolution images, which are then passed to the discriminator to determine their similarity to ground truth and high-resolution images. As mentioned earlier, this generative adversarial network involves an adversarial cooperative training between the generator (G) and the discriminator (D): G generates an image that closely resembles the target image or its label to deceive D, while D provides feedback by discerning between real and generated images. In general, the network aims to minimize the mutual information difference (also known as Jensen-Shannon divergence) between the produced data and the real data. The article employed Wasserstein distance as the selected objective instead of Jensen-Shannon divergence to address the issues of vanishing gradients and model collapse in the generator [92–94]. The imaging results are shown in Fig. 12(a)–(c).

The degradation model of AR-PAM imaging is influenced by the imaging depth and the center frequency of the ultrasonic transducer, which may vary under different imaging conditions and cannot be processed using a single neural network model. To address this limitation, Zhang et al. proposed a supplementary framework that combines the advantages of model based and learning based methods and avoids their limitations, which can be used to enhance the image quality of AR-PAM images [95]. Firstly, a deep convolutional neural network is used to implicitly capture the image statistical and structural information of the target vascular image, thereby obtaining a Plug and Play (PnP) prior,

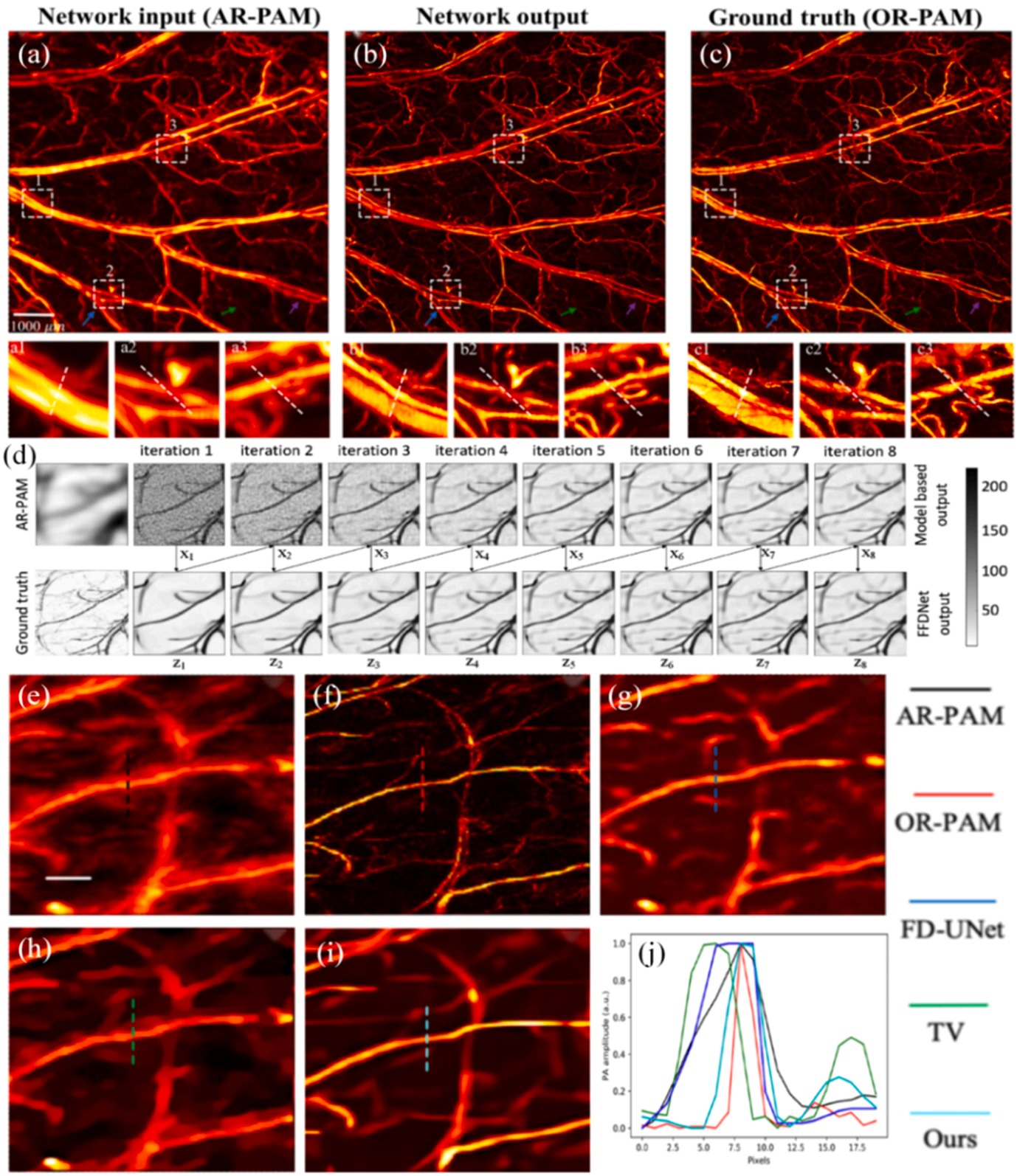

Fig. 12. WGAN network results for mouse ear vasculature and results of an adaptive enhancement method with a deep CNN prior. (a-c) WAGAN network for mouse ear vasculature: (a) Network input AR-PAM image, (a1, a2, a3) enlarged regions selected by white dashed boxes; (b) Network output image, (b1, b2, b3) enlarged regions selected by white dashed boxes; (c) Ground truth OR-PAM image, (c1, c2, c3) enlarged regions selected by white dashed boxes; (d-i) Adaptive enhancement method with a deep CNN prior: (d) Example AR-PAM image enhancement in different iterations by model based equation (upper row) and FFDNet (bottom row); (e) AR-PAM imaging result; (f) OR-PAM imaging result; (g) Result enhanced using the FDU-Net on (e); (h) Enhancement result using the total variation algorithm on (e); (i) Result enhanced using the proposed algorithm on (e); (j) Signal intensity distribution along the vertical dashed line. (Scale bar: 1 millimeter).

while avoiding the process of designing complex manual regularization terms. Subsequently, this PnP prior is further inserted into the model based framework so that it can adaptively handle the variational degradation process. The proposed framework was first applied to simulation research and demonstrated its excellent performance and adaptability (with a huge dynamic range). In addition, some in vivo experiments were conducted to test the framework's ability to handle various real imaging scenes. The results show that this method can adaptively enhance AR-PAM images obtained across different imaging systems and depths, expanding the application scenarios of this method. This work adopted a combination of network optimization and degradation models, iterating eight times each other. The advantage of this approach is to use neural networks to correct artifacts generated by degradation methods, and model methods to correct images when the optimization effect of network images decreases, achieving the effect of improving image effectiveness and resolution.The iteration model and imaging results are shown in Fig. 12(d)–(j). In the three simulation scenarios created, the proposed algorithm achieved optimal performance in terms of PSNR and SSIM values; In vivo testing results using this algorithm showed significant increases in SNR and CNR values from 6.34 and 5.79, respectively, to 35.37 and 29.66, as shown in Fig. 12(j).

3.2.3. Photoacoustic image segmentation and recognition processing

Image detection and recognition involve the task of identifying specific elements in medical images [96–98]. In many cases, the images are three-dimensional, making efficient analysis crucial. The ability to differentiate and classify different elements is fundamental in medical image analysis, and image segmentation is a necessary method for processing medical images. Image segmentation has greatly benefited from the latest developments in deep learning. In image segmentation, the goal is to accurately delineate the contours of organs or anatomical structures, and methods based on convolutional neural networks (CNNs) have gradually become dominant in this field. Deep learning not only helps in selecting and extracting features but also aids in constructing new features[99–101]. Moreover, it can provide predictive models that not only diagnose diseases but also measure and predict targets, offering actionable insights to improve efficiency for medical professionals. There have been numerous successful examples of deep learning-assisted image processing in photoacoustic imaging, and the segmentation and recognition methods used in photoacoustic imaging can be applied to other medical imaging modalities as well.

Zhang et al. proposed an emerging deep learning-based method for breast cancer diagnosis in photoacoustic tomography (PAT) [96]. This method employed a preprocessing algorithm to enhance the quality and uniformity of input breast cancer images. Additionally, a transfer learning algorithm was utilized to address the issue of insufficient training data, resulting in improved classification performance. The network categorized existing breast cancer datasets into six classes based on the BI-RADS level, helping doctors better diagnose and treat cancer based on breast imaging reports and data system levels.

In magnetic resonance imaging (MRI) field, Wu et al. proposed an oriented novel attention-based glioma grading network (AGGN) [102]. By applying the dual-domain attention mechanism, both channel and spatial information can be considered to assign weights, which benefits highlighting the key modalities and locations in the feature maps. Multi-branch convolution and pooling operations are applied in a multi-scale feature extraction module to separately obtain shallow and deep features on each modality, and a multi-modal information fusion module is adopted to sufficiently merge low-level detailed and high-level semantic features, which promotes the synergistic interaction among different modality information. The results have demonstrated the effectiveness and superiority of the proposed AGGN in comparison to other advanced models, which also presents high generalization ability and strong robustness.

Li et al. proposed a feature learning enhanced convolutional neural network (FLE-CNN) for cancer detection from histopathology images [103]. They built a highly generalized computer-aided diagnosis (CAD) system. The FLE-CNN included an information refinement unit employing depth- and point-wise convolutions is meticulously designed, where a dual-domain attention mechanism is adopted to focus primarily on the important areas. Experimental results demonstrate the merits of the proposed FLE-CNN in terms of feature extraction, which has achieved average sensitivity, specificity, precision, accuracy and F1 score of 0.9992, 0.9998, 0.9992, 0.9997 and 0.9992 in a five-class cancer detection task, and in comparison to some other advanced deep learning models, above indicators have been improved by 1.23 %, 0.31 %, 1.24 %, 0.5 % and 1.26 %, respectively.

3.3. Photoacoustic signal processing

Awasthi et al. proposed a deep learning-based photoacoustic (PA) sinogram super-resolution denoising model [104]. The loss function of this model is scaled root mean square error, which is used for super-resolution, denoising, and bandwidth (BW) enhancement of PA signals acquired at region boundaries [105]. The network and method presented in the paper have the following characteristics:It is the first single network that performs super-resolution, denoising, and BW enhancement of PA data in the sinogram domain. Most deep learning networks are proposed in the image space to improve reconstructed images. This network exhibits inherent robustness and generalization abilities. It also demonstrates robustness when trained on numerical models.The improved structure can be used to enhance raw data (sinogram) acquired experimentally, improving the results of inverse problems and inherently reducing biases introduced by image reconstruction methods.The introduction of scaled root mean square loss function to train the network on sinogram data containing extremely low values can be extended to other applications with similar properties as PA data. Fig. 13(a) and (b) show the network flowchart and structure diagram.

Similarly, Zhang et al. also proposed using sinogram data as input to remove artifacts produced by photoacoustic tomography imaging [106]. In their work, a two-dimensional brain PA numerical phantom dataset was generated based on magnetic resonance angiography (MRA) and T1-weighted images from the ixi dataset. The dataset was then used as input to a U-net network for training. The simulated artifact images were corrected against prior high-resolution images, resulting in a trained network that effectively corrects the acoustic aberration caused by the skull. Fig. 13(c) illustrates the experimental workflow, (d) presents comparisons of three sets of simulated brain imaging sinogram maps before and after artifact removal, and (e) shows a comparison of normalized signals from one of the models.

In photoacoustic imaging, previous signal processing techniques have been found insufficient to eliminate the influence of electrical noise because they often rely on simplified models and fail to capture the complex characteristics of both the signal and the noise. Dehner et al. proposed a discriminative deep learning approach to separate electrical noise from the photoacoustic signal prior to image reconstruction as shown in Fig. 14 [107]. In Fig. 14(a), Data layout of a measured multispectral stack of sinograms. The depicted sinogram shows the recorded signals during a representative scan of a human breast lesion at 960 nm. Magnification of the marked signals, which were recorded prior to responses from tissue and thus are predominately comprised of electrical noise. Histogram and fitted Gaussian distribution ($R2 = 99.5$ %) for parts of the electrical noise with visually low amounts of parasitic noise (signals marked with the dashed rectangle) illustrating the characterization of the thermal noise of the system. In Fig. 14(b), there are Noisy sinogram from a representative scan of a human breast lesion.Electrical noise component inferred by the neural network on the left side. Denoised sinogram obtained by subtracting the above two. On the right side are Magnifications of the marked areas in the left charts. Quantitative evaluation of the denoising performance below. There are Comparison of the SNR distributions in simulated photoacoustic sinograms

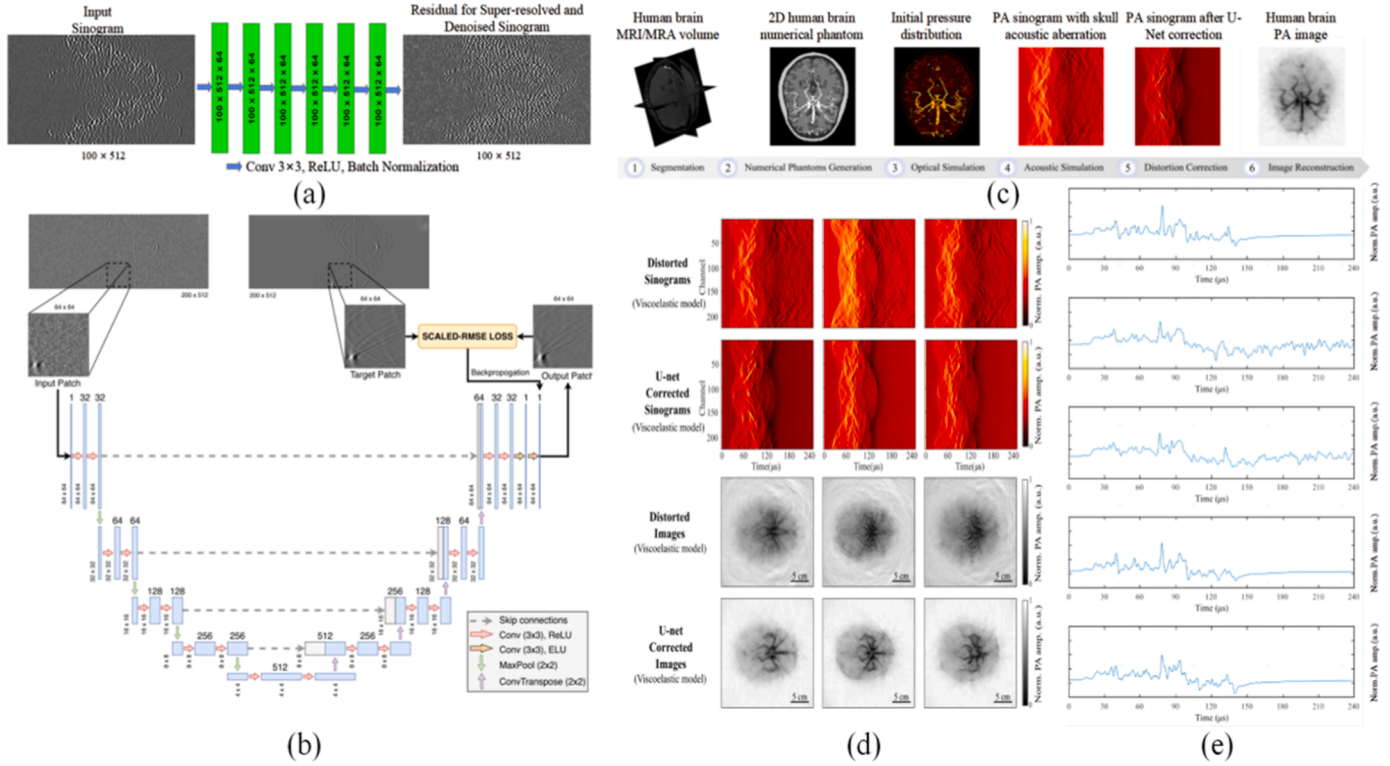

Fig. 13. Application of Sinogram Graph as Network Input in PAT. (a) Network diagram proposed by Awasthi et al.; (b) Network structure. (c) The experimental flow chart proposed by Zhang et al., whose flow includes prior image segmentation, making skull simulation structure, optical simulation, acoustic simulation, sinogram image training and finally obtaining the artifact free image; (d) Normalized PA sinograms and normalized DAS reconstructed human brain PAT images from viscoelastic media acoustic model. (e) Normalized PA signal is taken as the first channel of one of the skull simulation models. Signals are reference PA signal, PA signal with skull aberration obtained from fluid media acoustic model, PA signal with skull aberration obtained from viscoelastic media acoustic model, PA signal with skull aberration obtained from fluid media acoustic model after U-net Correction, PA signal with skull aberration obtained from viscoelastic media acoustic model after U-net correction.

that are distorted by electrical noise before and after denoising. The mean gain is 10.9 dB. Evaluation of in vivo scans of human breast lesions. Mean SNR (SNRmean) of individual time samples. The average increase is 20.8 dB. Individual SNRmean of all detectors. The average increase is 22.4 dB. Average SNR gains ("SNR after denoising - SNR before denoising") of the trained model for photoacoustic signals that were corrupted by a combination of measured electrical noise sinograms scaled with factor EN ∈{0, 0.5, 1,., 3}, and white Gaussian noise with standard deviations σGN ∈{0, 0.2, 0.4,., 2}. In Fig. 14(c), The first row shows the NMF spectra obtained from the original and denoised human breast lesion MSOT images from Dataset-BC, as well as the reference absorption spectra of the most prominent chromophores in breast tissue. The second and third rows show the before and after denoising comparison images, with the left column representing the pre-denoising image and the right column representing the post-denoising image. Visualizations of the NMF decomposition of a typical MSOT image are shown for pre-denoising and post-denoising at a depth of approximately 2 cm in malignant breast tumor. The contribution of the three spectra to the image is color-coded, with these spectra corresponding to the absorption spectra of hemoglobin (second row), fat, and water (third row). The tumor location determined from the ultrasound image is delineated by white contours. The proposed deep learning algorithm is based on two key features. Firstly, it learns the spatiotemporal correlation between the noise and the signal by using the entire photoacoustic sinogram as input. Secondly, it is trained on a large dataset consisting of experimentally acquired pure noise and synthetic photoacoustic signals.

The network utilizes a U-Net neural network architecture with 5 depths and 64 channel widths [108,109]. The basic expressive power of the network is reduced by estimating the interference signal. The L1 norm (L1 loss) is used as the loss function, and the ADAM optimizer [110] is employed with a learning rate of 0.0001, decayed linearly to 0. The ADAM optimizer has a batch size of 1, and the momentum parameters are set to $\beta 1 = 0.5$ and $\beta 2 = 0.999$. To speed up the learning process, Dehner et al. used a neural network input value of a constant 0.004, which brings the signal range to $[-1; 1]$. After passing through the artificial neural network, all signals were rescaled back to their original range. During training, the decomposition with the minimum loss over the data was validated, and the final model was selected.

Gutta et al. proposed a deep learning-based method for bandwidth enhancement of photoacoustic (PA) data [111]. During the process of photoacoustic tomography (PAT), the acquired PA signals from the surface of the tissue are always limited to a certain frequency band, while finer details of the image often reside in the high-frequency region of the PA signal. By utilizing a deep learning network, it is possible to effectively enhance the bandwidth of the PA signal without increasing computational complexity, thereby improving the contrast restoration and reconstruction quality of PA images.The network is trained with limited-bandwidth signals as input and full-bandwidth signals as output. The enhanced acoustic (PA) signal is then used as input to analysis reconstruction algorithms such as backprojection. This approach enables simple and efficient restoration of frequency band information but is limited by the constraints of prior algorithms and cannot achieve breakthroughs without real ground truth data.

Zhou et al. proposed a conditional generative adversarial network (cGAN) for distinguishing the photoacoustic (PA) signals generated by fiber-separated dual-wavelength excitation lasers [112]. The time delay between the signals is approximately 38 nm. Improving the imaging speed of multi-parameter photoacoustic microscopy (PAM) is a key focus in this direction. To avoid temporal overlap, the A-line rate is limited to within 3 MHz due to the speed of sound in biological tissues.In

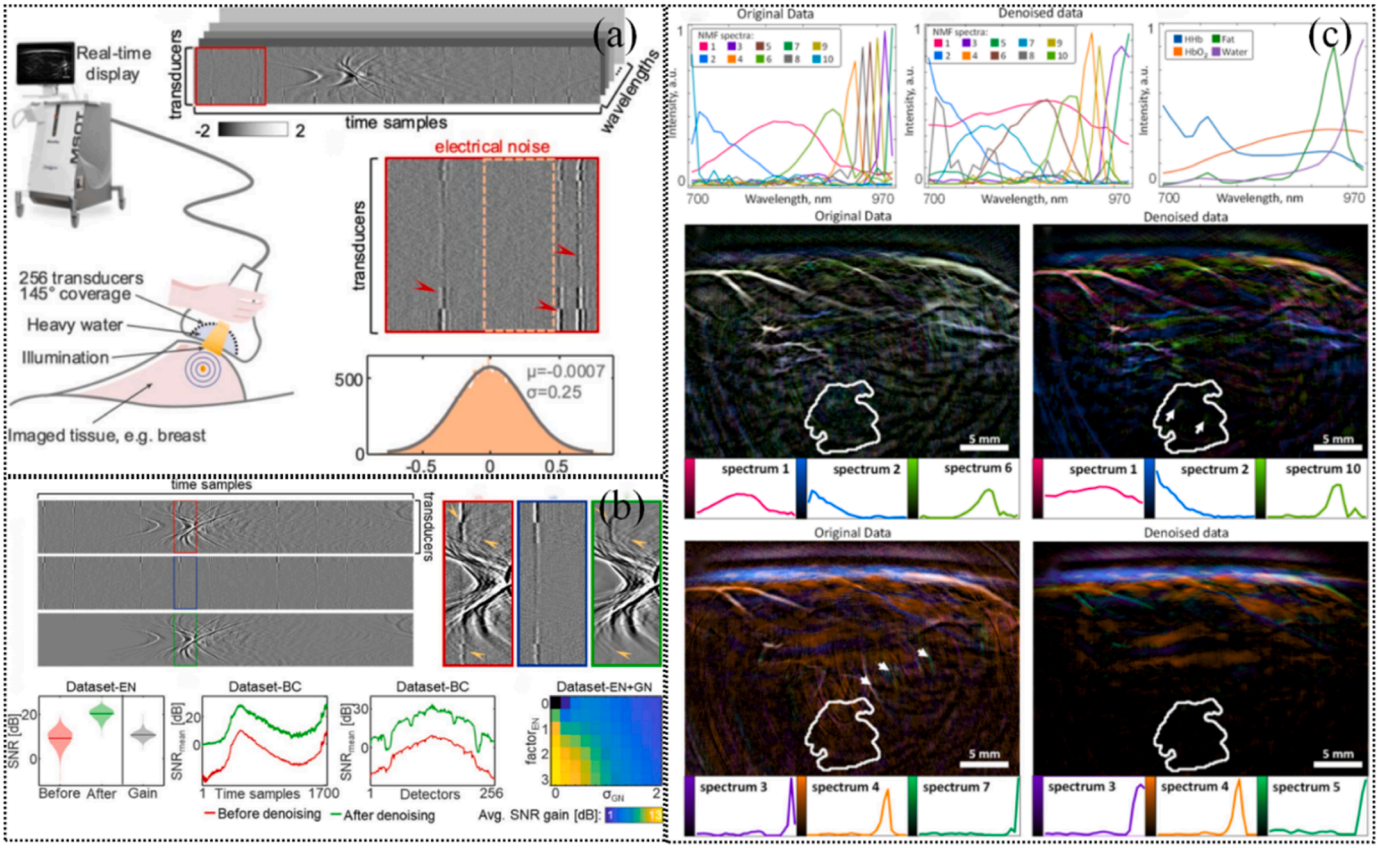

Fig. 14. The experiment is based on the schematic diagram of a handheld MSOT system, the evaluation diagram of signal signal-to-noise ratio, and the impact of denoising on the spectral content of photoacoustic images. (a) The experiment is based on the schematic diagram of the hand-held MSOT system Illustration of the scanning procedure using the handheld imaging probe of the test system. (b) Evaluation of the proposed denoising approach in the signal domain. (c) The impact of denoising on the spectral content of photoacoustic images.

order to achieve high-speed photoacoustic imaging of hemoglobin oxygen saturation, stimulated Raman scattering in optical fibers is widely used as a conventional method for generating dual-wavelength excitation at 558 nm from a commercially available 532 nm laser. However, the length of the fiber used for efficient wavelength conversion is typically short, resulting in only a small time delay being obtained, leading to significant overlap in the acquired A-line signals at the two wavelengths. The proposed cGAN network allows for PAM excitation using

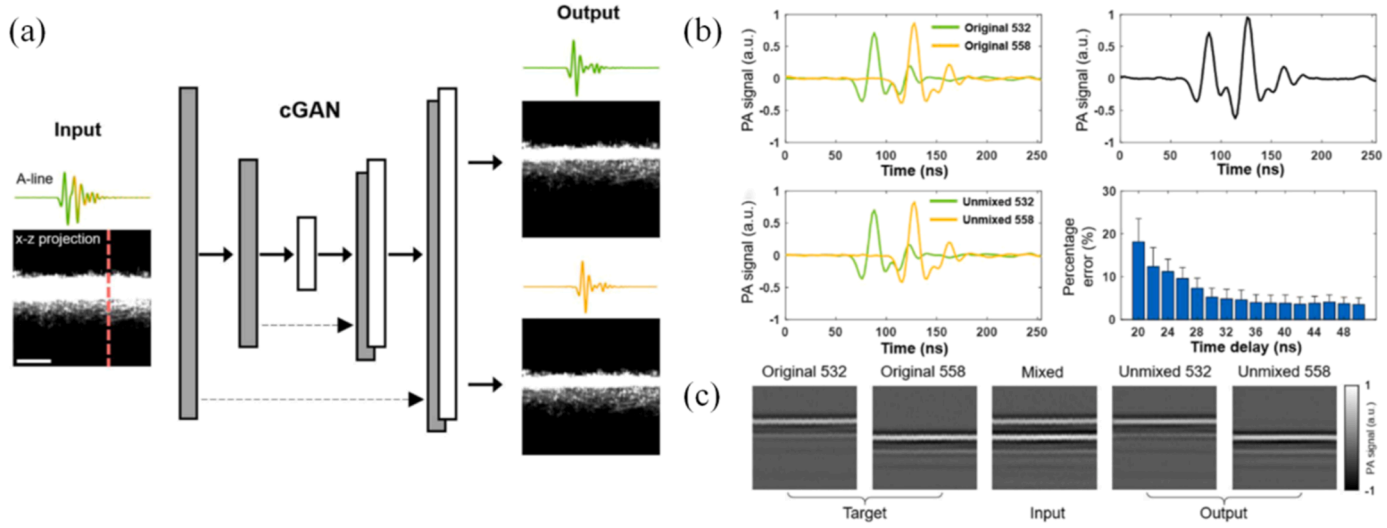

Fig. 15. Schematic of the cGAN and Dual wavelength A-scan signal graph. (a) The grayscale images are x – z projections of the three-dimensional dataset. A representative A-line, along the red dashed line, is shown above each ofthe x – z projection images. Scale bar: 300 μm. (b) 532-nm excited A-line (green) and digitally delayed 558-nm excited A-line (yellow). a.u., arbitrary units. Digital sum of the two A-lines. Non-overlapping A-lines (green: 532nmand yellow: 558 nm) generated by the cGAN. Percentage error of sO2 values as a function oftime delay. The error bars represent standard deviations. (c) Representative B-scans as the target, input, and output of the cGAN, consisting of 256 original, mixed, and unmixed A-lines, respectively.

multi-spectral laser pulses, addressing the issue of insufficient energy in single-color laser pulses as shown in Fig. 15. This technology presents an innovative approach towards achieving ultra-high-speed multi-parameter PAM.

4. Summary and outlook

4.1. Summary

Deep learning, as a cutting-edge data acquisition technology, has been widely used in various fields of photoacoustic imaging such as image reconstruction, image processing, and signal processing. It can adjust parameters according to different network requirements to achieve a balance between strong robustness, high imaging speed, and artifact removal.

Compared with the iterative reconstruction method, the overall error of the denoising process after back projection reconstruction is higher than that of the iterative reconstruction method, especially in the case of limited angle scanning, which will produce mechanical artifacts, and the error at the imaging boundary is also more obvious. However, it is known that both deep learning-based image processing algorithms and traditional iterative reconstruction algorithms can overcome these mechanical artifacts by using prior mapping relationships. Compared with the post-processing methods of sparse data reconstruction in PAM, the deep learning-based reconstruction method not only has faster reconstruction speed but also greatly improves reconstruction efficiency. In addition, in terms of photoacoustic image processing, deep learning also has great advantages, including: higher flexibility and accuracy in handling complex and variable photoacoustic signals; outstanding performance in many image processing tasks such as image segmentation, classification, and reconstruction; ability to handle large amounts of training data, thereby improving model generalization ability and prediction performance; good scalability, allowing the model complexity to be adjusted based on task requirements and computing resources. Furthermore, deep learning models can utilize knowledge learned from other domains for transfer learning, thereby accelerating model training and improving performance.

The application prospects of deep learning in the field of photoacoustic imaging are vast, with continuously emerging network architectures for reconstruction algorithms in photoacoustic tomography, as well as for subsequent processing and forward sinogram processing. There is also ample room for improvement in various aspects of photoacoustic microscopy (PAM), such as scanning mechanism enhancements, excitation mechanism improvements, and post-processing techniques.

4.2. Data acquisition

Big data is the core of deep learning, but there is currently no open dataset for photoacoustic image reconstruction. In current experiments, the test sets used to train and validate the CNN are generally obtained through three methods: real human photoacoustic imaging results, imaging results of phantoms, and computer-simulated images. Since photoacoustic imaging has not been widely used for clinical diagnosis and treatment of diseases, the available clinical case data is severely lacking. The flexibility of phantom images is low, and the cost of making phantoms is high. Furthermore, it takes a long time to construct the data set required for deep learning. Computer simulation involves forward numerical simulation of the optical forward problem (the propagation of pulsed laser in tissue) and the acoustic forward problem (the process of tissue absorbing light energy, expanding due to heating, then emitting ultrasonic waves and propagating towards the tissue surface), obtaining the simulated initial sound pressure distribution map as the expected output image. The low-quality images reconstructed from limited-angle photoacoustic measurement data using standard reconstruction algorithms are used as input images to form the training set of the CNN. The authenticity and effectiveness of the sample still need to be further discussed. In summary, there is currently a lack of large-scale open source training samples for photoacoustic imaging.

Gao et al. proposed a computing method of four-dimensional (4D) spectral-spatial imaging for PAD [113] , This method takes the optical and acoustic properties of heterogeneous skin tissues into account, which can be used to correct the optical field of excitation light, detectable ultrasonic field, and provide accurate single-spectrum analysis or multi-spectral imaging solutions of PAD for multilayered skin tissues. Simulation datasets obtained from the computational model were used to train neural networks to further improve the imaging quality of the PAD system.Most deep learning-based photoacoustic imaging needs thousands pairs of labeled input-output data to train the neural network, especially those applications in clinical skin imaging, which requires even larger amounts of data. However , in many cases the ground truth corresponding to the experimental data is inaccessible. This work as an efficient “learning from computational model” implemented an efficient method for obtaining simulation data.Considering human skin tissues are multilayered physiopathological structures with variability in optical absorption and acoustic impedance, this work further verify the simulation method from angles such as beam type, ultrasonic transducer performance, laser focusing position, and multi-spectral analysis.The article also proposes two neural networks trained on the dataset obtained through this method, namely the spread spectrum network and the enhanced imaging depth network.The feasibility of simulated datasets generated by computational modeling for neural network training was also demonstrated, helping to solve the major challenge of deep learning techniques in photoacoustic skin imaging that cannot obtain ground truth in many cases, with the potential to further improve the imaging quality of the PAD system through image reconstruction, information processing, and artificial intelligence methods as shown in Fig. 16(b–c). Fig. 16 (a) verifies the effectiveness of the simulation model under different luminous flux conditions of 6 mJ/cm^2, 12 mJ/cm^2, and 18 mJ/cm^2. Fig. 16(b) shows the results of palm skin imaging optimized by two networks.

4.3. Interpretability

In addition to the lack of widely available and specialized data, the reliability and interpretability of deep learning methods are also receiving increasing attention. Lisboa et al. published a review of interpretability discussions in machine learning in 2020 [114], in which they classified interpretability discussions corresponding to the development of machine learning.

The article further proposes that there is currently no complete consensus on how to evaluate the quality of interpretable or interpretable methods. The evaluation methods that can explain ML include the "Real Humans in Ieal Tasks" proposed by Doshi Velez and Kim and the "AI rationalization" proposed by Ehsan et al. The quality of a given explanation needs to be evaluated in the context of its task, measuring the extent to which the explanation promotes and improves decision-making.

Salahuddin et al. published a review of interpretable methods for deep neural networks in medical image analysis in 2022 [115]. The article pointed out that interpretable artificial intelligence (XAI) refers to an AI solution that can provide some details about its functionality in a way that end users can understand. At present, the interpretability of deep neural networks is widely defined as attempting to explain the decision-making process of a model in a way that can be understood by end users.

Dai et al. [116] used a conceptual alignment deep autoencoder to analyze tongue images that represent different body constituent types based on traditional Chinese medicine principles. Koh et al. introduced Concept Bottleneck Models for osteoarthritis grading and used 10 clinical concepts such as joint space narrowing, bone spurs, calcification, etc.In their study, Dai et al. employed a novel deep autoencoder with

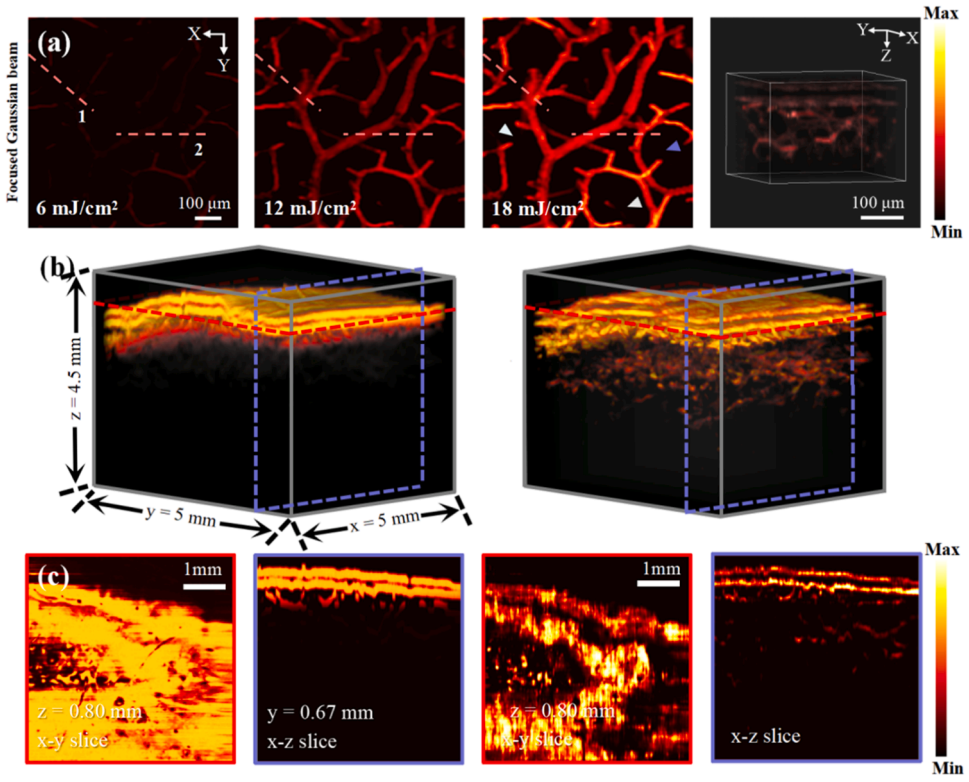

Fig. 16. The model adopts focused Gaussian beam images with different power densities of 532 nm wavelength incident beams and network generalization test results. (a) When the power densities are 6 mJ/cm^2, 12 mJ/cm^2, and 18 mJ/cm^2 and he three dimensional imaging result; (b) 3D PA image of palm skin and 3D PA image obtained after the spread spectrum network and the depth enhanced network processing (c) The corresponding color slices in the figure.

conceptual alignment to investigate tongue images, which are representative of diverse body constituent types according to the principles of traditional Chinese medicine. The utilization of this approach enabled a comprehensive analysis and interpretation of the underlying features associated with each body type.

Furthermore, Koh et al. [117] presented an innovative framework known as Concept Bottleneck Models for the purpose of osteoarthritis grading. This model incorporated ten important clinical concepts including joint space narrowing, bone spurs, calcification, among others. By leveraging these concepts, the researchers were able to establish a robust and informative grading system for the evaluation of osteoarthritis severity.

The above two are both based on the perspective of conceptual interpretability, and there are also more interpretable classifications, including: Case based models, Counter actual interpretation, Language description et al.

4.4. Conclusion and outlook

In the switchable optical and acoustic resolution photoacoustic endomicroscope proposed by Ma et al. in 2020 [118], high-resolution imaging of the surface and deep layers is achieved by switching between optical and acoustic resolution systems at different depths in the skin. We can contemplate by acquiring a dataset from this system, the high-resolution surface images and deep-layer images are combined and fed into a deep neural network, enabling the high-resolution images to learn depth information from the deep-layer images, and the deep-layer images to learn resolution from the high-resolution images. This approach may ultimately lead to a single system that combines the advantages of both types of photoacoustic microscopy.There is also the possibility of mutual learning between systems with different numerical apertures (NA) and corresponding scanning mechanisms, or between different excitation wavelengths. Can we achieve complementary effects between penetration depth and imaging resolution? These are all worth considering.Furthermore, in the field of photoacoustic endoscopy, the lack of corresponding datasets has limited the widespread use of deep learning methods. Therefore, exploring deep neural network-based approaches for photoacoustic endoscopy is also an important area of research.

In the previous article [94], the example of learning from low resolution AR-PAM images to high-resolution OR-PAM images can improve a certain imaging performance and expand the applicability of the system through learning between different imaging systems. In the field of photoacoustic imaging microscopy, AR-PAM and OR-PAM are complementary in imaging depth and resolution, OR-PAM can currently achieve an imaging depth of around 1.5 mm, with resolution at the micron or submicron level, while AR-PAM has an imaging depth of over 10 mm, but the corresponding resolution also has an order of magnitude attenuation. If a prior method can be used to obtain a prior of OR-PAM images at the same depth, the image features of this prior can be retained through a neural network method and applied in the corresponding AR-PAM system, that is, the results of the AR-PAM imaging system can be obtained through a neural network, and a high depth AR-PAM image with corresponding OR-PAM resolution can be obtained. Compared with general prior methods, the biggest advantage of OR-PAM prior is that it preserves the basic features of photoacoustic images, and its feedback signal composition is also ultrasound. This brings great convenience to the preservation of image features of tissue signal strength and phase. It is obvious that the combined imaging system can effectively improve imaging quality and obtain high-quality images at corresponding depths that were previously difficult to obtain. The corresponding potential mutual learning work can be envisioned. For example, OR-PAM learns imaging depth from AR-PAM images. Although the resolution of AR-PAM images is not as good as that of OR-PAM, it is possible to learn the intensity of AR-PAM signals from the perspective of photoacoustic signals by preserving the features of OR-PAM images, aiming to discover

weak photoacoustic signals in deep tissues. Due to the complementarity between AR-PAM and OR-PAM in the field of photoacoustic microscopy, it can be imagined that their mutual learning will become a reality in the near future. Photoacoustic tomography, on the other hand, has a higher imaging depth, and its imaging speed and imaging range are significantly different from microscopic systems. Its image features and system application scenarios are also inconsistent with microscopic systems. Therefore, the mutual learning between fault systems and microscopic systems still requires further development and integration of photoacoustic imaging. Not only is there mutual learning within photoacoustic imaging, but this approach can also be applied in bimodal imaging systems and in conditions of different system parameters. Bimodal imaging refers to the combination of two or more imaging techniques to obtain different types of information simultaneously or sequentially for image reconstruction and analysis. Mutual learning can be extended to, for example, mutual learning between ultrasound imaging and photoacoustic imaging, mutual learning between 1064 nm and 532 nm wavelength systems, and mutual learning between different NA systems.In summary, the ultimate goal of potential mutual learning currently lies in improving the system's penetration ability or imaging quality, provided that the two imaging systems are close or the imaging results can learn from each other.

The achievements of deep learning in photoacoustic imaging are undeniable, such as its applications in image reconstruction, signal-to-noise ratio improvement, and super-resolution. These achievements provide new ideas and methods for the development of photoacoustic imaging technology. However, there are still some challenges and limitations for deep learning models in photoacoustic imaging. For example, deep learning models require high training data demands, requiring a large amount of labeled data and computing resources. Although many network methods, such as U-Net structure and unsupervised learning, attempt to solve the data problem, there is still considerable room for improvement. Additionally, interpretability of deep learning models is also an issue that needs to be addressed.

In future research, we can try to further optimize the performance and interpretability of deep learning models to better meet the application requirements of photoacoustic imaging technology in clinical medicine and life sciences. At the same time, we can also explore combining deep learning with other technologies to discover more potential applications. As mentioned earlier, the work of Zhang et al. [95] is a good application and extension of neural network methods. For the interpretability of neural network methods, the author proposed a new approach that combines network and model methods. The model method corrected the image content forward, while the neural network corrects image artifacts. This alternating iteration method greatly improved the interpretability of neural networks. However, the drawback of this method is that the training difficulty and reconstruction time of the network have increased. Perhaps a more efficient network structure can be used as an alternative to iterative methods to improve imaging speed. In short, the idea of combining the principle of preserving models with neural networks is worth learning and continuing. Can deep learning also be better applied and explained from the perspective of photoacoustic signals? In photoacoustic tomography, the quality of the sine wave of the signal determines the quality of the reconstructed image. Unlike general image learning neural networks such as Awasthi [104] and Zhang [106], training photoacoustic signals to achieve signal amplification and denoising is also a way of applying deep learning. In future work, a collaborative learning approach can be envisioned. The signal learning network provides deep imaging signals and amplifies them, while the image learning network provides deep image features to achieve joint learning, discover deep structures, and efficiently image.

This review focuses on deep learning enabled photoacoustic imaging, and analyzes recent deep learning work from four perspectives: photoacoustic imaging PAT reconstruction, PAM reconstruction, image processing, and signal processing. The article also starts from neural network structures such as U-Net, GAN network, and Dense Block, organizing their early work in the field of biomedical imaging, and introducing readers to common neural network structures and their origins in the biomedical field. Finally, the article summarizes the analysis and summary of deep learning in improving imaging capabilities from four perspectives, proposes the current problems and difficulties of neural networks, and further provides ideas for solving the problems. In summary, the rapid development of neural networks has continuously empowered photoacoustic imaging and even biomedical imaging in recent years. They have made epoch-making contributions to the depth and quality of imaging results, as well as to the improvement of imaging system efficiency and imaging speed.

CRediT authorship contribution statement

Chen Qian: Supervision. **Zuo Chao:** Funding acquisition. **Ma Haigang:** Conceptualization, Data curation, Investigation, Supervision, Writing – review & editing. **Wei Xiang:** Investigation, Methodology, Writing – original draft, Writing – review & editing. **Huang Qinghua:** Resources. **Feng Ting:** Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by National Natural Science Foundation of China (62275121, 12204239, 12326609, 62071382, 62227818), Youth Foundation of Jiangsu Province (BK20220946), Fundamental Research Funds for the Central Universities (30923011024), Jiangsu Provincial Basic Research Program Frontier Leading Special Project (BK20192003).

References

[1] Yang C., Lan H., Gao F. Accelerated photoacoustic tomography reconstruction via recurrent inference machines[C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019: 6371–6374.
[2] P. Beard, Biomedical photoacoustic imaging, Interface Focus 1 (4) (2011) 602–631.
[3] C. Yang, H. Lan, F. Gao, et al., Review of deep learning for photoacoustic imaging, Photoacoustics 21 (2021) 100215.
[4] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press,, 2016.
[5] Q. Huang, H. Tian, L. Jia, et al., A review of deep learning segmentation methods for carotid artery ultrasound images, Neurocomputing (2023) 126298.
[6] Y. Guo, Y. Liu, A. Oerlemans, et al., Deep learning for visual understanding: a review, Neurocomputing 187 (2016) 27–48.
[7] A. Salehi, M. Balasubramanian, DDCNet: deep dilated convolutional neural network for dense prediction, Neurocomputing 523 (2023) 116–129.
[8] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[9] K. Kawaguchi, J. Huang, L.P. Kaelbling, Effect of depth and width on local minima in deep learning, Neural Comput. 31 (7) (2019) 1462–1498.
[10] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation[C]//medical image computing and computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015. Proceedings, Part III 18, Springer International Publishing,, 2015, pp. 234–241. October 5-9, 2015.
[11] Du G., Cao X., Liang J., et al. Medical image segmentation based on u-net: A review[J]. Journal of Imaging Science and Technology, 2020.
[12] N. Man, S. Guo, K.F.C. Yiu, et al., Multi-layer segmentation of retina OCT images via advanced U-net architecture, Neurocomputing 515 (2023) 185–200.
[13] Z. Huang, J. Miao, H. Song, et al., A novel tongue segmentation method based on improved U-Net, Neurocomputing 500 (2022) 73–89.
[14] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, et al., 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing

and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016. Proceedings, Part II 19, Springer International Publishing,, 2016, pp. 424–432. October 17-21, 2016.
[15] F. Isensee, P. Kickingereder, W. Wick, et al., No new-net[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018. Revised Selected Papers, Part II 4, Springer International Publishing,, 2019, pp. 234–244. September 16, 2018.
[16] X. Xiao, S. Lian, Z. Luo, et al., Weighted res-unet for high-quality retina vessel segmentation[C]//2018 9th international conference on information technology in medicine and education (ITME), IEEE (2018) 327–331.
[17] S. Guan, A.A. Khan, S. Sikdar, et al., Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal, IEEE J. Biomed. Health Inform. 24 (2) (2019) 568–576.
[18] Han Y.S., Yoo J., Ye J.C. Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis[J]. arXiv preprint arXiv: 1611.06391, 2016.
[19] S. Hakakzadeh, Z. Kavehvash, M. Pramanik, Artifact removal factor for circular-view photoacoustic tomography[C]//2022 IEEE International Ultrasonics Symposium (IUS), IEEE (2022) 1–4.
[20] Y. Lin, S. Kou, H. Nie, et al., Deep learning based on co-registered ultrasound and photoacoustic imaging improves the assessment of rectal cancer treatment response, Biomed. Opt. Express 14 (5) (2023) 2015–2027.
[21] S. Guan, A.A. Khan, S. Sikdar, et al., Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning, Sci. Rep. 10 (1) (2020) 8510.
[22] H. Lan, K. Zhou, C. Yang, et al., Hybrid neural network for photoacoustic imaging reconstruction[C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE (2019) 6367–6370.
[23] M. Guo, H. Lan, C. Yang, J. Liu, et al., As-net: fast photoacoustic reconstruction with multi-feature fusion from sparse data, IEEE Trans. Comput. Imaging 8 (2022) 215–223.
[24] J. Meng, X. Zhang, L. Liu, et al., Depth-extended acoustic-resolution photoacoustic microscopy based on a two-stage deep learning network, Biomed. Opt. Express 13 (8) (2022) 4386–4397.
[25] Ledig C., Theis L., Huszár F., et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681–4690.
[26] T. Carvalho, E.R.S. De Rezende, M.T.P. Alves, et al., Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN[C]//2017 16th IEEE international conference on machine learning and applications (ICMLA), IEEE (2017) 866–870.
[27] V. Rajinikanth, A.N. Joseph Raj, K.P. Thanaraj, et al., A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection, Appl. Sci. 10 (10) (2020) 3429.
[28] Wang X., Yu K., Wu S., et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0–0.
[29] Szegedy C., Ioffe S., Vanhoucke V., et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).
[30] F. He, T. Liu, D. Tao, Why resnet works? residuals generalize, IEEE Trans. Neural Netw. Learn. Syst. 31 (12) (2020) 5349–5362.
[31] H. Lin, S. Jegelka, Resnet with one-neuron hidden layers is a universal approximator, Adv. Neural Inf. Process. Syst. (2018) 31.
[32] D. Waibel, J. Gröhl, F. Isensee, et al., Reconstruction of initial pressure from limited view photoacoustic images using deep learning[C]//Photons Plus Ultrasound: Imaging and Sensing 2018, SPIE 10494 (2018) 196–203.
[33] J. Schwab, S. Antholzer, R. Nuster, et al., Deep learning of truncated singular values for limited view photoacoustic tomography[C]//Photons Plus Ultrasound: Imaging and Sensing 2019, SPIE 10878 (2019) 254–262.
[34] H. Abdi, Singular value decomposition (SVD) and generalized singular value decomposition, Encycl. Meas. Stat. 907 (2007) 912.
[35] J. Feng, J. Deng, Z. Li, et al., End-to-end Res-Unet based reconstruction algorithm for photoacoustic imaging, Biomed. Opt. Express 11 (9) (2020) 5321–5340.
[36] T. Tong, W. Huang, K. Wang, et al., Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data, Photoacoustics 19 (2020) 100190.
[37] E.M.A. Anas, H.K. Zhang, C. Audigier, et al., Robust photoacoustic beamforming using dense convolutional neural networks[C]//Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation: International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16–20, 2018. Proceedings, Springer International Publishing,, 2018, pp. 3–11.
[38] K. Johnstonbaugh, S. Agrawal, D. Abhishek, et al., Novel deep learning architecture for optical fluence dependent photoacoustic target localization[C]//Photons Plus Ultrasound: Imaging and Sensing 2019, SPIE 10878 (2019) 95–102.
[39] D. Allman, A. Reiter, M.A.L. Bell, Photoacoustic source detection and reflection artifact removal enabled by deep learning, IEEE Trans. Med. Imaging 37 (6) (2018) 1464–1477.
[40] M. Xu, L.V. Wang, Universal back-projection algorithm for photoacoustic computed tomography, Phys. Rev. E 71 (1) (2005) 016706.
[41] J. Provost, F. Lesage, The application of compressed sensing for photo-acoustic tomography, IEEE Trans. Med. Imaging 28 (4) (2008) 585–594.
[42] M.W. Kim, G.S. Jeng, I. Pelivanov, et al., Deep-learning image reconstruction for real-time photoacoustic system, IEEE Trans. Med. Imaging 39 (11) (2020) 3379–3390.
[43] H. Lan, C. Yang, D. Jiang, et al., Reconstruct the photoacoustic image based on deep learning with multi-frequency ring-shape transducer array[C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE (2019) 7115–7118.
[44] M.T. McCann, K.H. Jin, M. Unser, Convolutional neural networks for inverse problems in imaging: a review, IEEE Signal Process. Mag. 34 (6) (2017) 85–95.
[45] C. Cai, K. Deng, C. Ma, et al., End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging, Opt. Lett. 43 (12) (2018) 2752–2755.
[46] J. Bell, What is machine learning? Mach. Learn. City.: Appl. Archit. Urban Des. (2022) 207–216.
[47] S. Antholzer, M. Haltmeier, J. Schwab, Deep learning for photoacoustic tomography from sparse data, Inverse Probl. Sci. Eng. 27 (7) (2019) 987–1005.
[48] L.V. Wang, Tutorial on photoacoustic microscopy and computed tomography, IEEE J. Sel. Top. Quantum Electron. 14 (1) (2008) 171–179.
[49] D. Allman, A. Reiter, M.A.L. Bell, Photoacoustic source detection and reflection artifact removal enabled by deep learning, IEEE Trans. Med. Imaging 37 (6) (2018) 1464–1477.
[50] 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.
[51] P.C.M. Van Zijl, S.M. Eleff, J.A. Ulatowski, et al., Quantitative assessment of blood flow, blood volume and blood oxygenation effects in functional magnetic resonance imaging, Nat. Med. 4 (1998) 159–167.
[52] B. Cox, J.G. Laufer, S.R. Arridge, et al., Quantitative spectroscopic photoacoustic imaging: a review, J. Biomed. Opt. 17 (6) (2012), 061202-061202.
[53] M. Li, Y. Tang, J. Yao, Photoacoustic tomography of blood oxygenation: a mini review, Photoacoustics 10 (2018) 65–73.
[54] L.J. Rich, M. Seshadri, Photoacoustic imaging of vascular hemodynamics: validation with blood oxygenation level–dependent MR imaging, Radiology 275 (1) (2015) 110.
[55] B.T. Cox, J.G. Laufer, P.C. Beard, et al., Quantitative spectroscopic photoacoustic imaging: a review, J. Biomed. Opt. 17 (6) (2012) 061202.
[56] C. Cai, K. Deng, C. Ma, et al., End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging, Opt. Lett. 43 (12) (2018) 2752–2755.
[57] C. Yang, H. Lan, H. Zhong, et al., Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network[C]//2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE (2019) 741–744.
[58] N. Sobahi, A. Sengur, R.S. Tan, et al., Attention-based 3D CNN with residual connections for efficient ECG-based COVID-19 detection, Comput. Biol. Med. 143 (2022) 105335.
[59] P. Rajendran, M. Pramanik, Deep learning approach to improve tangential resolution in photoacoustic tomography, Biomed. Opt. Express 11 (12) (2020) 7311–7323.
[60] Abadi M., Agarwal A., Barham P., et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems[J]. arXiv preprint arXiv:1603.04467, 2016.
[61] B.E. Treeby, B.T. Cox, k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, J. Biomed. Opt. 15 (2) (2010), 021314-021314-12.
[62] M. Schellenberg, K.K. Dreher, N. Holzwarth, et al., Semantic segmentation of multispectral photoacoustic images using deep learning, Photoacoustics 26 (2022) 100341.
[63] B.E. Treeby, J. Jaros, B.T. Cox, Advanced photoacoustic image reconstruction using the k-Wave toolbox[C]//Photons Plus Ultrasound: Imaging and Sensing 2016, SPIE 9708 (2016) 517–530.
[64] X. Song, X. Zhou, Photoacoustic microscopy simulation platform based on K-Wave simulation toolbox[C]//Photonics for Quantum 2021, SPIE 11844 (2021) 54–57.
[65] Y. Gao, W. Xu, Y. Chen, et al., Deep learning-based photoacoustic imaging of vascular network through thick porous media, IEEE Trans. Med. Imaging 41 (8) (2022) 2191–2204.
[66] M. Schmitt, C.M. Poffo, J.C. de Lima, et al., Application of photoacoustic spectroscopy to characterize thermal diffusivity and porosity of caprocks, Eng. Geol. 220 (2017) 183–195.
[67] R. Manwar, K. Kratkiewicz, K. Avanaki, Investigation of the effect of the skull in transcranial photoacoustic imaging: a preliminary ex vivo study, Sensors 20 (15) (2020) 4189.
[68] J.P. Monchalin, L. Bertrand, G. Rousset, et al., Photoacoustic spectroscopy of thick powdered or porous samples at low frequency, J. Appl. Phys. 56 (1) (1984) 190–210.
[69] C.F. Ramirez-Gutierrez, J.D. Castano-Yepes, M.E. Rodriguez-Garcia, In situ photoacoustic characterization for porous silicon growing: detection principles, J. Appl. Phys. 119 (18) (2016) 185103.
[70] Zhou J., He D., Shang X., et al. Photoacoustic Microscopy with Sparse Data Enabled by Convolutional Neural Networks for Fast Imaging[J]. arXiv preprint arXiv:2006.04368, 2020.
[71] S. Jeon, J. Kim, D. Lee, et al., Review on practical photoacoustic microscopy, Photoacoustics 15 (2019) 100141.
[72] L.V. Wang, J. Yao, A practical guide to photoacoustic tomography in the life sciences, Nat. Methods 13 (8) (2016) 627–638.
[73] Ledig C., Theis L., Huszár F., et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681–4690.
[74] Mathieu M., Couprie C., LeCun Y. Deep multi-scale video prediction beyond mean square error[J]. arXiv preprint arXiv:1511.05440, 2015.

[75] H. Zhao, Z. Ke, F. Yang, et al., Deep learning enables superior photoacoustic imaging at ultralow laser dosages, Adv. Sci. 8 (3) (2021) 2003097.
[76] H. Zhao, K. Li, N. Chen, et al., Multiscale vascular enhancement filter applied to in vivo morphologic and functional photoacoustic imaging of rat ocular vasculature, IEEE Photonics J. 11 (6) (2019) 3900912.
[77] Q. Yao, Y. Ding, G. Liu, et al., Low-cost photoacoustic imaging systems based on laser diode and light-emitting diode excitation, J. Innov. Opt. Health Sci. 10 (04) (2017) 1730003.
[78] R. Manwar, M. Hosseinzadeh, A. Hariri, et al., Photoacoustic signal enhancement: towards utilization of low energy laser diodes in real-time photoacoustic imaging, Sensors 18 (10) (2018) 3498.
[79] S. Wang, J. Lin, T. Wang, et al., Recent advances in photoacoustic imaging for deep-tissue biomedical applications, Theranostics 6 (13) (2016) 2394.
[80] A. Hariri, M. Hosseinzadeh, S. Noei, et al., Photoacoustic signal enhancement: towards utilization of very low-cost laser diodes in photoacoustic imaging[C]// Photons Plus Ultrasound: Imaging and Sensing 2017, SPIE 10064 (2017) 822–826.
[81] Y. Cao, R. Wang, J. Peng, et al., Humidity enhanced N2O photoacoustic sensor with a 4.53 μm quantum cascade laser and Kalman filter, Photoacoustics 24 (2021) 100303.
[82] M. Alaeian, H.R.B. Orlande, B. Lamien, Kalman filter temperature estimation with a photoacoustic observation model during the hyperthermia treatment of cancer, Comput. Math. Appl. 119 (2022) 193–207.
[83] S.C. Rutan, S.D. Brown, Pulsed photoacoustic spectroscopy and spectral deconvolution with the Kalman filter for determination of metal complexation parameters, Anal. Chem. 55 (11) (1983) 1707–1710.
[84] R. Manwar, M. Zafar, Q. Xu, Signal and image processing in biomedical photoacoustic imaging: a review, Optics 2 (1) (2020) 1–24.
[85] S. Telenkov, A. Mandelis, Signal-to-noise analysis of biomedical photoacoustic measurements in time and frequency domains, Rev. Sci. Instrum. 81 (12) (2010) 124901.
[86] C. Hide, T. Moore, M. Smith, Adaptive Kalman filtering for low-cost INS/GPS, J. Navig. 56 (1) (2003) 143–152.
[87] M. Zhou, H. Xia, H. Lan, et al., Wavelet de-noising method with adaptive threshold selection for photoacoustic tomography[C]//2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE (2018) 4796–4799.
[88] E.Y. Park, H. Lee, S. Han, et al., Photoacoustic imaging systems based on clinical ultrasound platform, Exp. Biol. Med. 247 (7) (2022) 551–560.
[89] M. Lassen, A. Brusch, D. Balslev-Harder, et al., Phase-sensitive noise suppression in a photoacoustic sensor based on acoustic circular membrane modes, Appl. Opt. 54 (13) (2015) D38–D42.
[90] D. He, J. Zhou, X. Shang, et al., De-noising of photoacoustic microscopy images by attentive generative adversarial network, IEEE Trans. Med. Imaging (2022).
[91] S. Cheng, Y. Zhou, J. Chen, et al., High-resolution photoacoustic microscopy with deep penetration through learning, Photoacoustics 25 (2022) 100314.
[92] Goodfellow I. Nips 2016 tutorial: Generative adversarial networks[J]. arXiv preprint arXiv:1701.00160, 2016.
[93] Arjovsky M., Bottou L. Towards principled methods for training generative adversarial networks[J]. arXiv preprint arXiv:1701.04862, 2017.
[94] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks [C]//International conference on machine learning, PMLR (2017).
[95] Z. Zhang, H. Jin, W. Zhang, et al., Adaptive enhancement of acoustic resolution photoacoustic microscopy imaging via deep CNN prior, Photoacoustics 30 (2023) 100484.
[96] J. Zhang, B. Chen, M. Zhou, et al., Photoacoustic image classification and segmentation of breast cancer: a feasibility study, IEEE Access 7 (2018) 5457–5466.
[97] Q. Huang, Z. Miao, J. Li, et al., Classification of breast ultrasound with human-rating BI-RADS scores using mined diagnostic patterns and optimized neuro-network, Neurocomputing 417 (2020) 536–542.
[98] Q. Huang, L. Ye, Multi-task/single-task joint learning of ultrasound BI-RADS features, IEEE Trans. Ultrason., Ferroelectr., Freq. Control 69 (2) (2021) 691–701.
[99] Y. Luo, Q. Huang, L. Liu, Classification of tumor in one single ultrasound image via a novel multi-view learning strategy, Pattern Recognit. (2023) 109776.
[100] Q. Huang, L. Zhao, G. Ren, et al., NAG-Net: nested attention-guided learning for segmentation of carotid lumen-intima interface and media-adventitia interface, Comput. Biol. Med. 156 (2023) 106718.
[101] Q. Huang, H. Luo, C. Yang, et al., Anatomical prior based vertebra modelling for reappearance of human spines, Neurocomputing 500 (2022) 750–760.
[102] P. Wu, Z. Wang, B. Zheng, et al., AGGN: attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion, Comput. Biol. Med. 152 (2023) 106457.
[103] H. Li, P. Wu, Z. Wang, et al., A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis, Comput. Biol. Med. 151 (2022) 106265.
[104] N. Awasthi, G. Jain, S.K. Kalva, et al., Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography, IEEE Trans. Ultrason., Ferroelectr., Freq. Control 67 (12) (2020) 2660–2673.
[105] Awasthi N., Pardasani R., Kalva S.K., et al. Sinogram super-resolution and denoising convolutional neural network (SRCN) for limited data photoacoustic tomography[J]. arXiv preprint arXiv:2001.06434, 2020.
[106] F. Zhang, J. Zhang, Y. Shen, et al., Photoacoustic digital brain and deep-learning-assisted image reconstruction, Photoacoustics (2023) 100517.
[107] C. Dehner, I. Olefir, K.B. Chowdhury, et al., Deep-learning-based electrical noise removal enables high spectral optoacoustic contrast in deep tissue, IEEE Trans. Med. Imaging 41 (11) (2022) 3182–3193.
[108] M. Liu, Z. Wang, H. Li, et al., AA-WGAN: attention augmented Wasserstein generative adversarial network with application to fundus retinal vessel segmentation, Comput. Biol. Med. 158 (2023) 106874.
[109] H. Li, N. Zeng, P. Wu, et al., Cov-Net: a computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision, Expert Syst. Appl. 207 (2022) 118029.
[110] Kingma D.P., Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
[111] S. Gutta, V.S. Kadimesetty, S.K. Kalva, et al., Deep neural network-based bandwidth enhancement of photoacoustic data, J. Biomed. Opt. 22 (11) (2017) 116001.
[112] Y. Zhou, F. Zhong, S. Hu, Temporal and spectral unmixing of photoacoustic signals by deep learning, Opt. Lett. 46 (11) (2021) 2690–2693.
[113] Y. Gao, T. Feng, H. Qiu, et al., 4D spectral-spatial computational photoacoustic dermoscopy, Photoacoustics (2023) 100572.
[114] P.J.G. Lisboa, S. Saralajew, A. Vellido, et al., The coming of age of interpretable and explainable machine learning models, Neurocomputing 535 (2023) 25–39.
[115] Z. Salahuddin, H.C. Woodruff, A. Chatterjee, et al., Transparency of deep neural networks for medical image analysis: a review of interpretability methods, Comput. Biol. Med. 140 (2022) 105111.
[116] Y. Dai, G. Wang, Analyzing tongue images using a conceptual alignment deep autoencoder, IEEE Access 6 (2018) 5962–5972.
[117] P.W. Koh, T. Nguyen, Y.S. Tang, et al., Concept bottleneck models[C]// International conference on machine learning, PMLR (2020) 5338–5348.
[118] H. Ma, Z. Cheng, Z. Wang, et al., Switchable optical and acoustic resolution photoacoustic dermoscope dedicated into in vivo biopsy-like of human skin, Appl. Phys. Lett. 116 (7) (2020).

Xiang Wei is a master student from Nanjing University of Science and Technology. He is now in the second year of his master's degree and his current research focuses on applications of photoacoustic imaging in biomedicine.

Ting Feng received her bachelor's degree, master's degree and Ph.D. degree from Nanjing University in 2010, 2012 and 2016, respectively. She is currently working at Fudan university in China. She was the visiting scholar at the University of Michigan in 2018 and 2019, and she was the joint-Ph.D. student at the University of Michigan in 2013–2015. Her current research interest includes photoacoustic imaging and measurements. A major part of her research is clinical application of photoacoustic techniques for bone health assessment.

Qinghua Huang received the Ph.D. degree in biomedical engineering from the Hong Kong Polytechnic University, Hong Kong, in 2007. Now he is a full professor in School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, China. His research interests include multi-dimensional u1trasonic imaging, medical image analysis, machine learning for medical data, and intelligent computation for various applications.

Qian Chen as a leading expert in the National Key Discipline of "Optical Engineering" at Nanjing University of Science and Technology. As the primary contributor, he has won a second-class State Technological Invention Award, a second-class State Scientific and Technological Progress Award and five first-class provincial and ministerial-level science and technology awards. As the first inventor, he has obtained 74 granted invention patents, 16 PCT international patents, and 6 U.S. patents. He has authored three books and 374 SCI papers, among which 27 have been featured on the cover. Currently, he serves as a Fellow and Executive Director of the Chinese Society of Optical Engineering and Executive Director of the Chinese Institute of Electronics.

Chao Zuo is a professor in optical engineering, Nanjing University of Science and Technology (NJUST), China. He leads the Smart Computational Imaging Laboratory (SCILab: www.scilaboratory.com) at the School of Electronic and Optical Engineering, NJUST. He has long been engaged in the development of novel Computational Optical Imaging and Measurement technologies, with a focus on Phase Measuring Imaging Metrology such as Holographic Interferometric Microscopy, Non-interferometic Quantitative Phase Imaging (QPI), Fringe Projection Profilometry (FPP), and Structured Illumination Microscopy (SIM). He has authored $>$ 200 peer-reviewed publications in prestigious journals with over 11,000 citations.

Haigang Ma received the Ph.D. degree in Optics from the South China Normal University, China, in 2020. Now he is an associate professor in School of Electronic and Optical Engineering, Nanjing University of Science and Technology, China. His research interests include photoacoustic imaging, ultrasonic imaging, photoelectric detection and processing, and photoacoustic imaging for various biomedical applications.