electronics

MDPI

Article

Unsupervised Learning for Enhanced Computed Photoacoustic Microscopy

Lulin Yang [1], Wenjing Chen [1], Tingdong Kou [1], Chenyang Li [2], Meng You [2,*] and Junfei Shen [1,*]

[1] College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China; yanglulin@stu.scu.edu.cn (L.Y.); chenwj0409@scu.edu.cn (W.C.); koutingdong@stu.scu.edu.cn (T.K.)
[2] State Key Laboratory of Oral Diseases & National Center for Stomatology & National Clinical Research Center for Oral Diseases & Department of Oral Medical Imaging, West China Hospital of Stomatology, Sichuan University, Chengdu 610065, China; hxkq_lichenyang@stu.scu.edu.cn
* Correspondence: youmeng@scu.edu.cn (M.Y.); shenjunfei@scu.edu.cn (J.S.)

Abstract: Photoacoustic microscopy (PAM) is a medical-imaging technique with the merits of high contrast and resolution. Nevertheless, conventional PAM scans specimens in a diameter-by-diameter fashion, resulting in a time-consuming process. Furthermore, deep-learning-based PAM image enhancement necessitates acquiring ground-truth data for training purposes. In this paper, we built an optical-resolution photoacoustic microscopy system and introduced an innovative unsupervised-learning algorithm. First, we enhanced the rotational-scanning method, transitioning from a diameter-by-diameter approach to a sector-by-sector one, significantly reducing imaging time (from 280 s to 109 s). Second, by establishing a metric for unsupervised learning, we eliminated the need for collecting reliable and high-quality ground truth, which is a challenging task in photoacoustic microscopy. A total of 324 pairs of datasets (mouse ears) were collected for unsupervised learning, with 274 for training and 50 for testing. Additionally, carbon-fiber data were sampled for lateral resolution and contrast evaluation, as well as the effective rate evaluation of the algorithm. The enhanced images demonstrated superior performance compared with that of maximum projection, both subjectively and objectively. A 76% improvement in the lateral resolution was observed. The effective rate of the algorithm was measured to be 100%, which was tested on 50 random samples. The technique presented in this paper holds substantial potential for image postprocessing and opens new avenues for unsupervised learning in photoacoustic microscopy.

Keywords: photoacoustic imaging; photoacoustic microscopy; unsupervised learning

check for updates

Citation: Yang, L.; Chen, W.; Kou, T.; Li, C.; You, M.; Shen, J. Unsupervised Learning for Enhanced Computed Photoacoustic Microscopy. *Electronics* **2024**, *13*, 693. https://doi.org/10.3390/electronics13040693

Academic Editor: Elias Stathatos

Received: 24 November 2023
Revised: 24 January 2024
Accepted: 26 January 2024
Published: 8 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Photoacoustic imaging (PAI) is a medical-imaging technique that combines exceptional contrast with high resolution [1–4]. PAI offers superior imaging depth and higher contrast compared with optical and ultrasonic imaging, respectively. In contrast to positron-emission tomography (PET), X-ray imaging, and computed tomography (CT), PAI provides structural, molecular, and functional information about biological tissue without radiation damage. In comparison to magnetic resonance imaging (MRI), PAI boasts higher resolution, faster imaging speed, and lower costs. Additionally, it is more applicable to patients with implanted devices. With these advantages, PAI has garnered increasing attention in clinical and preclinical applications, including neuroscience [5], tumor angiogenesis [6], histology [7,8], dermatology [9], and more.

PAI relies on the photoacoustic effect [10], a physical phenomenon discovered by Bell in 1880. To put it simply, biological tissue undergoes thermal elastic expansion (resulting in an ultrasonic wave) after absorbing energy from a pulsed laser or intensity-modulated continuous laser. After that, the ultrasonic wave can be converted into an analog signal by the ultrasonic transducer and sampled by a data-acquisition device for subsequent image

reconstruction. Sound scatters about 1000 times less than light [11,12], which explains why PAI has a higher imaging depth than that of optical imaging.

PAI can be categorized into three types [13]: photoacoustic tomography (PAT), photoacoustic microscopy (PAM), and photoacoustic endoscopy (PAE). PAM is further divided into two types: optical-resolution photoacoustic microscopy (OR-PAM) and acoustic-resolution photoacoustic microscopy (AR-PAM). Our focus is on OR-PAM, which can capture micrometer-ordered capillaries. OR-PAM, first proposed by Maslov et al. in 2008 [14], has emerged as a crucial research direction in the PAI field, offering the most exceptional lateral resolution among PAIs, down to several micrometers [15]. However, it faces the inevitable tradeoff between lateral resolution and imaging depth [16–18], typically achieving a depth of up to 1 mm in most cases. Simultaneously enhancing lateral resolution and imaging depth proves challenging. Additionally, spatial resolution may be constrained by imaging speed. Implementing hardware-level solutions presents challenges due to factors such as cost considerations, device accuracy, and component quality. Liang et al. addressed this challenge by employing a new-type photoacoustic sensor based on a small-sized fiber laser [19]. They achieved a lateral resolution of 3.3 μm with a field of view (FOV) of up to 1.57 mm^2. Subsequently, they enhanced their photoacoustic sensor [20] and conducted high-speed imaging over 2 × 2 mm^2 at a frame rate of 2 Hz. Chen et al. proposed a functional OR-PAM relying on polygon scanning to achieve wide-field and fast imaging [21]. The A-line rate, B-line rate (in the 12 mm range), and volumetric imaging rate (over a scanning area of 12 × 5 mm^2) of their system were reported to be 1 MH, 477.5 Hz, and ~1 Hz, respectively.

Some teams took a different approach, addressing the issue from the aspect of software, particularly leveraging deep learning. Fortunately, deep learning (DL) has been developing rapidly since 2015, which was applied to PAI for fasting imaging and improving image quality, especially in PAM [22]. Chen et al. proposed a method to tackle with motion artifacts of OR-PAM based on deep learning [23]. It is a convolutional network that established an end-to-end map from images of motion artifacts to the corrected, proving effective for both large blood vessels and blood capillaries. Dispirito et al. utilized a deep-learning technique to overcome the tradeoff between spatial resolution and imaging speed [24]. Only 2% of the original pixel data is needed to reconstruct the PAM image by applying the FD U-net they trained. Zhou et al. also trained a CNN with a squeeze-and-excitation (SE) block for super-resolution tasks [25]. It is noteworthy that Cheng et al. employed a generative adversarial network (GAN) to process low-resolution PAM images [26], achieving an improvement of lateral resolution from 54.0 μm to 5.1 μm. However, it is important to note that all these techniques belong to supervised learning, requiring ground-truth data.

Obtaining reliable and high-quality ground truth for PAM is challenging. First, limited to the performance of the PAM system, it is difficult to obtain reliable and high-quality experimental data for training. Over half of the deep-learning papers about PAI rely on simulated data for supervised learning [27]. Second, in terms of PAM, not only do we need to build a high-precision system to get high-resolution images as ground truth but also low-resolution images in the corresponding area as input for the network. A complex and integrated system might need to be built.

In order to enhance image quality without ground truth, we proposed a technique based on unsupervised learning. The network is specifically designed by us to enhance the contrast and resolution of the source images. In the network, we add a continuous convolution to each step of the encoder and a decoder to strengthen the capacity of the model. No activation function is employed on the last layer in order to improve the forward propagation of the network. Drawing inspiration from U2fusion [28], we implemented two adaptive weights to retain structural information and reduce background noise in source images. By using this method, we get significantly higher image contrast and resolution compared to the conventional max projection approach. The novelty and contributions of our work are summarized as follows:

We proposed a framework and metric of unsupervised learning for improving image contrast and resolution. It eliminated the need for collecting reliable and high-quality ground truth, which is a challenging task in photoacoustic microscopy. We built an OR-PAM system and enhanced the rotary-scanning method from a diameter-by-diameter approach to a sector-by-sector one, greatly saving imaging time. The total imaging time was reduced from 280 s to 109 s. Two adoptive weights, named ω_1 and ω_2, were calculated from the pretrained VGG-16 network and applied in the loss function to preserve structural information in the source images during the training process. A total of 324 pairs of datasets (mouse ears) were collected for unsupervised learning, with 274 for training and 50 for testing. The datasets can be used for other unsupervised learning such as unsupervised imaging fusion. Except for that, the effective rate of our algorithm was tested to be 100% with 50 pairs of images of carbon fiber.

2. Materials and Methods

2.1. OR-PAM System

A transmissive OR-PAM system was built to collect photoacoustic data, as depicted in Figures 1 and 2. The light source is a pulsed laser of 532 nm wavelength, 6.8 ns pulse width, and 20 μJ single-pulse energy. The beam from the laser source is delivered into a spatial filter for spatial filtering which is then coupled into the single-mode fiber (SMF) with an optical cap layer at both ends. The spatial filter comprises two 4× microscope objectives and a pinhole with a diameter of 50 μm. The SMF achieves a fiber-coupling efficiency of 44%. The usage of the SMF is to increase the flexibility of the OR-PAM system and enhance the quality of the laser beam. Raster scanning is achieved using the scanning galvanometer. The water tank is custom-designed with a hole of 2 cm diameter at the bottom to allow the laser beam to pass through. The water-immersed ultrasonic transducer (central frequency, 15 MHz; fractional bandwidth, 35%; acoustic focal length, 1.50 inches; element diameter, 0.63 inches; focus type, line focused) was fixed on the rotary table by a customized duralumin connector. The 3D adjustment mount facilitates the mounting of the rotary table. The ultrasonic waves induced by the laser are detected by the transducer and converted into electronic signals, subsequently amplified by a low-noise amplifier. The photoacoustic signal then passes through the customized bandpass filter (BPF, 3 MHz–30 MHz) to minimize noise interference. Finally, the signal is sampled by the data-acquisition card at a frequency of 250 MHz for subsequent image reconstruction. To minimize the noise brought by the rotary table, the connector that connects the transducer and rotary table is made of insulated material, and the transducer is grounded. The control card manages the scanning galvanometer and pulsed laser for raster-scanning operations. Detailed information about the devices is provided in Table 1.

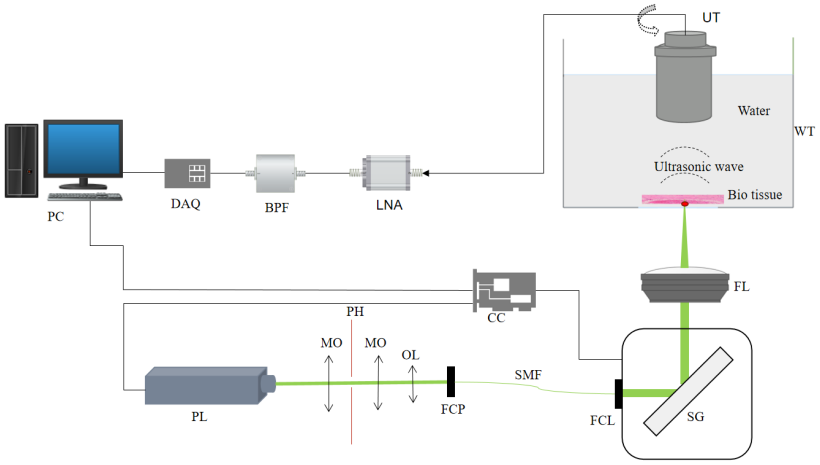


Figure 1. Block diagram of the PAM system; PC, personal computer; DAQ, data-acquisition card; BPF, band-pass filter; LNA, low noise amplifier; UT, ultrasonic transducer; WT, water tank; PL, pulsed

laser source; ML, microscope objective; PH, pinhole; OL, optical lens; FCP, fiber coupler; SMF, single-mode fiber; FCL, fiber collimator; SG, scanning galvanometer; FL, field lens; CC, control card.

Table 1. Devices used in the PAM system.

Modulation	Devices	Model
Optics	Laser	AO-Mini-532; Changchun New Industries, Changchun, China
	Spatial filter	Pinhole: PCI-50H; LBTEK, Changsha, China 2 Microscope objectives: GCO-2101; 4×; Daheng Optics, Beijing, China Adjustment support: GCM-SF03M (Daheng), OMHS-WJ (Zolix, China)
	Fiber coupling	Single-mode fiber: 460HP single-mode fiber with optical end caps; Pawo Optics, China Fiber-coupling adjustment support: GCX-C6PC-A; Daheng Optics, China Fiber collimator: GCX-LF6PC-532; Daheng Optics
Raster scanning	Galvanometer	SCANCUBE III 14; SCANLAB, Germany
	Field lens	F-theta-017700-209-26-ft-100-532-90; SCANLAB, Germany
	Control card	RTC4; SCANLAB, Germany
	Rotary table	GCD-012100M, GCD-0401M; Daheng Optics, China
Acquisition	Ultrasonic transducer	V319-SU-CF1.50IN-PTF; Olympus, Japan
	Amplifier	LNA-650; RF-Bay, US.
	Filter	Customized; Zhonghua Antenna, China
	Data-acquisition card	PCIe-8914; Art Beijing Technology, China
Other	Water tank	Customized; Yongtuo, China
	3D adjustment mount	LD-125-2-N; Shunrong, China

* The filter used is a passive band-pass filter of 3 MHz to 30 MHz. The water tank's dimensions are 15 cm × 15 cm × 5 cm, constructed from PMMA. Fiber end caps are employed to safeguard the optical fiber from potential damage caused by the pulsed laser at both ends. Fiber-coupling modulation is implemented to enhance the quality of the laser beam, a critical factor for achieving high system resolution.

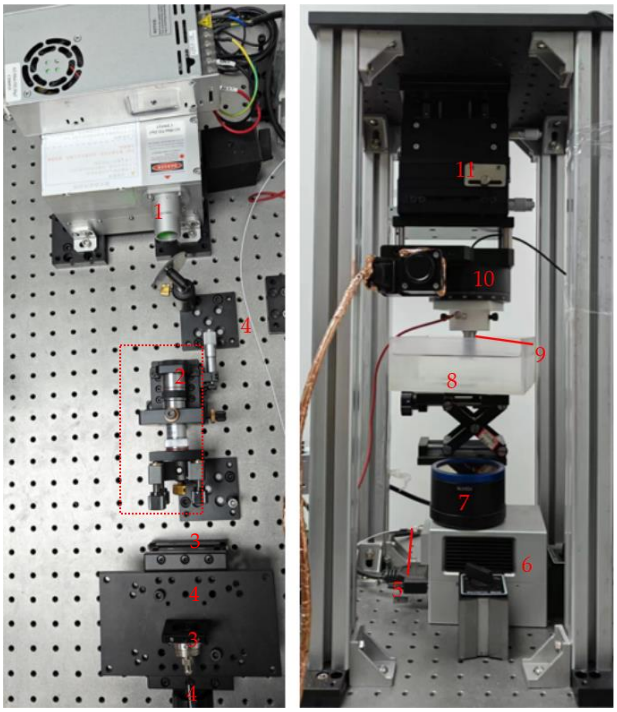


Figure 2. Physical map of the PAM system. 1, PL; 2, spatial filter; 3, FCP; 4, SMF; 5, FCL; 6, SG; 7, FL; 8, WT; 9, UT; 10, rotary table; 11, 3D adjustment mount.

2.2. System Working Sequence

The working sequence of the PAM system is shown in Figure 3. A pulse signal with a frequency of 50 kHz and a duty cycle of 1% is emitted from the control card to

control the pulsed laser source and the data-acquisition card to conduct raster scanning and data sampling, respectively, as long as the galvanometer works. The laser source and the data-acquisition card use the same signal sent out from the control card. The laser was designed to be low-level triggered to receive the externally triggered pulses. A triggered pulse corresponds to a laser pulse. The data-acquisition card is in trigger-sampling mode, which would conduct a sampling operation when triggered pulses come. For each triggered pulse, 1024 points are set to be sampled. The rotary table can be controlled by serial communication. Four main commands have been designed to precisely control it to work, namely degrees, speed, direction, and execution. It is important to note that the last command sends no information but serves to initiate the execution of the preceding commands.

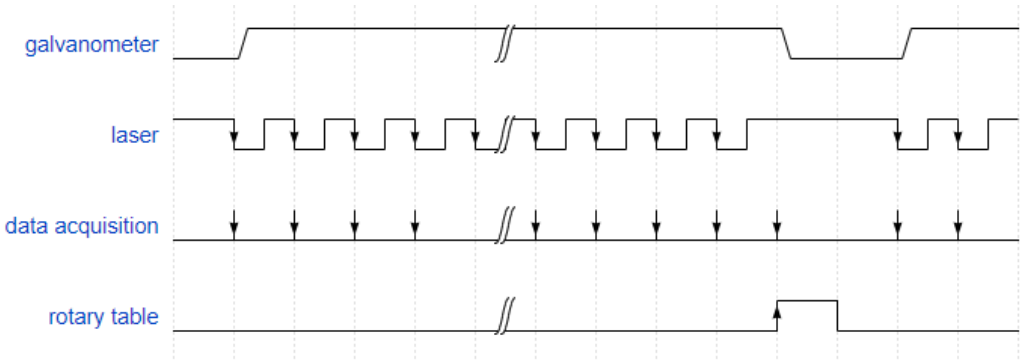


Figure 3. Working sequence of the PAM system.

2.3. Scanning Method

Through the experimental process, we adopted raster scanning of the rotary mode. We got 1000 points from each diameter after maximum projection. Considering imaging speed, we improved the scanning mode, sampling from a diameter-by-diameter approach to a sector-by-sector one, as Figure 4 shows. At first, we scan from $-18°$ to $18°$, which contains data of 360 diameters, each scanning diameter is separated by $0.1°$. Then, the transducer rotates $36°$ for the next raster scanning and sampling, and so forth. For the sector-by-sector approach, the UT keeps still until 360 diameters have been scanned in a sector. In other words, the rotation of the UT and the scan of the 2D galvanometer execute at different times. Repeatedly starting and stopping the rotary table is time-consuming. Therefore, the sector-by-sector approach saves time. Considering the response region of the line-focused UT, the central angle of a sector cannot be too large (36 degrees is appropriate). The transducer needs to rotate $144°$ to finish the total sampling of five sectors. Therefore, a total of $1024 \times 1000 \times 360 \times 5$ points are collected during an imaging process.

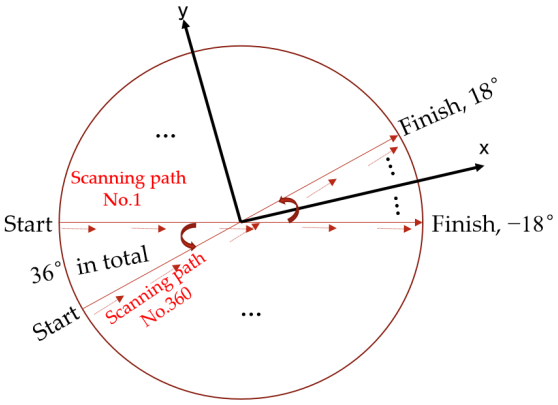


Figure 4. Sketch map of the first scanning sector.

3. Algorithm

3.1. U-Net Structure

To achieve PAM image transformation, we adopted U-Net as the framework for unsupervised learning. U-Net consists of two parts, called encoder and decoder. The U-Net architecture we used for image enhancement is illustrated in Figure 5. The encoder contains five downsampling convolutional blocks. Each convolutional block comprises a 3×3 convolution, with a Leaky Rectify Linear Unit (LeakyReLU, slope = 0.01) layer. There is a max pooling layer at the end of each step. The downsampling operation of the encoder is to extract high-level features of images. The decoder also comprises five steps. Each step consists of an upsampling transposed convolution (to halve the number of feature channels) and a LeakyReLU layer. In addition, a 3×3 convolution and a LeakyReLU layer are added to each step of the encoder and decoder as continuous convolution to strengthen the capacity of the model. The jump concatenating enables the decoder to get the high-level features extracted by the encoder. On the last layer of the decoder, a 1×1 convolution is employed to map each 32-component feature vector to the desired number of classes, without any activation function followed.

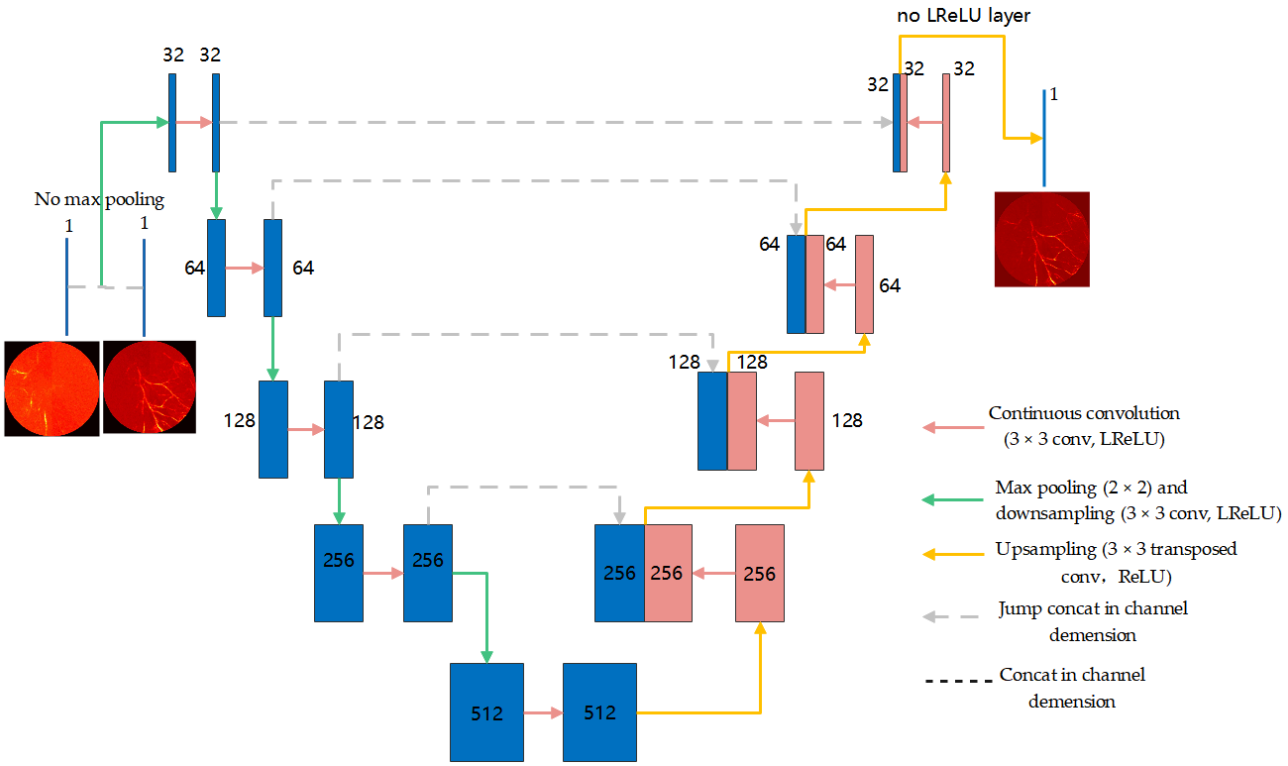


Figure 5. The structure of the U-Net model.

3.2. Loss Function

We utilized the pretrained VGG-16 network to extract feature information, as depicted in Figure 6. The inputs I_1 and I_2 are images of three channels, whereas photoacoustic images are single channel. Thus, we duplicated photoacoustic images into three channels as I_1 and I_2. The outputs of convolution before max pooling are feature maps of subsequent information, named from $\phi c_1(I)$ to $\phi c_5(I)$. The $\phi c_1(I)$ and $\phi c_2(I)$ contain shallow features, such as texture and shape. $\phi c_4(I)$ and $\phi c_5(I)$ can preserve deep-level features, such as content and spatial structures. With these features from shallow to deep, the pretrained VGG-16 model can present a comprehensive information extraction from an image.

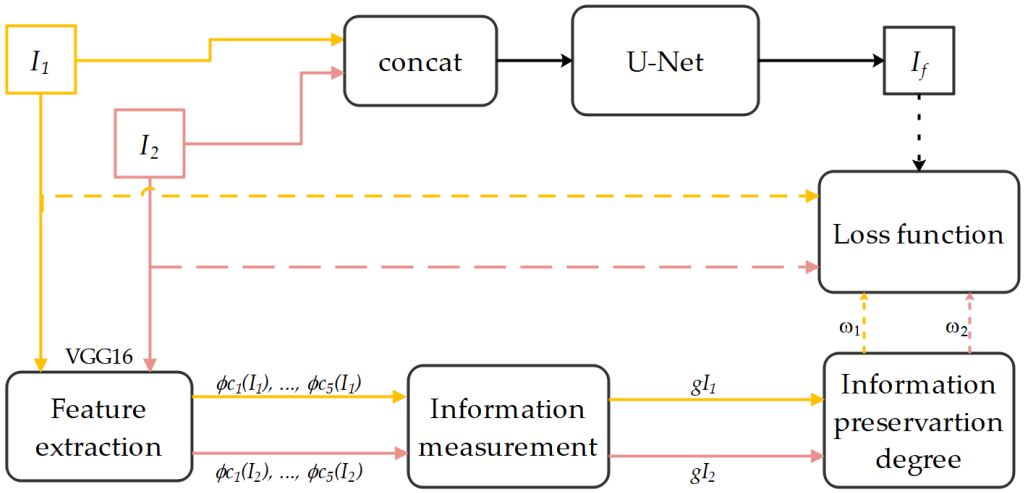


Figure 6. Training mechanism of U-Net.

We use the gradients of these five feature maps to evaluate the feature information, as follows:

$$gI = \frac{1}{5}\sum_{j=1}^{5} \frac{1}{H_j W_j D_j} \sum_{k=1}^{D_j} \left\| \nabla \Phi c_j^k(I) \right\|_F^2 \quad (1)$$

wherein $\phi c_j(I)$ denotes the feature map of convolutional layers before the j-th max pooling layer of the VGG-16 network. The k is the feature map in the k-th channel of D_j channels. $\|\bullet\|_F^2$ denotes the Frobenius norm, and ∇ is the Laplace operator.

To preserve structural information in each source image, two adaptive weights, named ω_1 and ω_2, are used to define the similarity between the source images and the fusion images. A higher weight corresponds to a greater similarity, indicating a higher degree of preservation for the corresponding source image. ω_1 and ω_2 are defined as follows:

$$[\omega_1 \omega_2] = softmax([\frac{gI_1}{c}, \frac{gI_2}{c}]) \quad (2)$$

A Softmax function is used to map the value of gI_1/c and gI_2/c between zero and one. The predefined positive constant c is applied to scale values for better assignment which guarantees that the sum of ω_1 and ω_2 is one. Then, ω_1 and ω_2 are exploited in the loss function to control the structure-preservation degree of the source images.

The loss function we use for model training is defined as follows:

$$L_{SSIM} = E[\omega_1 \bullet (1 - S_{I_1,I_f}) + \omega_2 \bullet (1 - S_{I_2,I_f})] \quad (3)$$

$$L_{MSE} = E[\omega_1 \bullet MSE_{I_1,I_f} + \omega_2 \bullet MSE_{I_2,I_f}] \quad (4)$$

$$L = L_{SSIM} + \alpha L_{MSE} \quad (5)$$

where I_1 and I_2 denote the source images, and I_f is the image reconstructed by the U-Net. E denotes the averaging operation (tf.reduce_mean, axis = -1 in tensorflow). $S_{x,y}$ is the value of structural similarity (*SSIM*) between image x and image y in Equation (3). $MSE_{x,y}$ is the value of the mean square error (*MSE*) between image x and image y in Equation (4). L is the total loss of the U-Net consisting of L_{SSIM} and L_{MSE}. L_{SSIM} focuses on the contrast and structure, while L_{MSE} is employed to constrain intensity distribution. α is a positive constant set to 0.5 to balance the tradeoff between L_{SSIM} and L_{MSE}. If α is too large, the I_f will get bright and blurry. Conversely, the I_f will be too dark.

4. Results

4.1. Training Details

All the raw data of the photoacoustic images were restored in the form of a matrix. Each of the images has a pixel size of 1000 × 1000, and no cropping operation was applied to them. A 3 × 3 median filtering was adopted to eliminate the extremely bright pixel point caused by salt and pepper noise. A total of 324 pairs of datasets were collected, with 274 designated for training and 50 for testing. The parameters of the U-Net model were optimized by RMSPropOptimizer at a learning rate of 0.0001. The number of the training epoch was set to 100, with 426 steps for each epoch. The training was executed on a Tesla A100 GPU and a 2.9 GHz Gold 6226R CPU.

4.2. Lateral Resolution and Contrast Evaluation

To assess the lateral resolution of our system, we employed carbon fibers for imaging, each with a diameter of 7 μm, affixed to a transparent glass sheet. The carbon fibers were fixed on a transparent plastic disc and immersed in the water tank to avoid displacement during the imaging process. As illustrated in Figure 7, we sampled 10 pixels for (a) and (b) along the white line and performed Gaussian fitting, respectively. The image reconstructed by max projection has higher pixel intensity while lower lateral resolution than that of deep learning in general. The FWHM of the Gaussian function and lateral resolution can be calculated by the following formulas:

$$\mathrm{FWHM} = 2\sigma\sqrt{2\ln 2} \tag{6}$$

$$\delta = 1/\mathrm{FWHM} \tag{7}$$

where σ is the standard deviation of the Gaussian function, and δ denotes the lateral resolution (pixels^{-1}). The image reconstructed by U-Net has a resolution of 0.37 pixels^{-1}, as opposed to the 0.21 pixels^{-1} of the max projection case, which indicates a substantial improvement in resolution, specifically around a 76% enhancement when transitioning from max projection to the deep-learning approach.

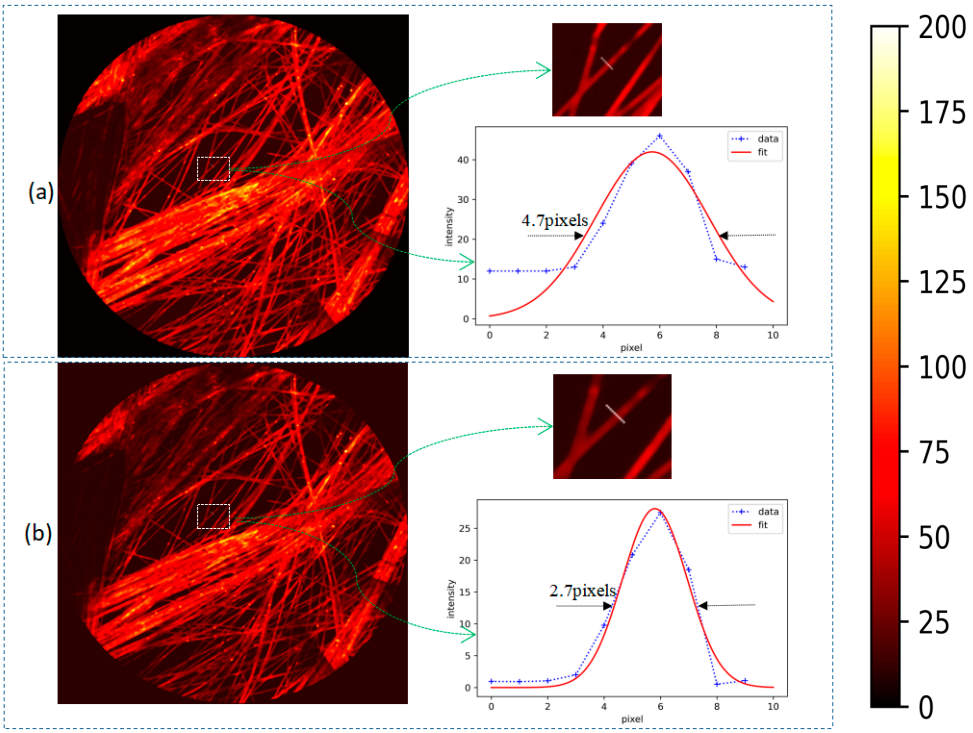


Figure 7. Comparison of lateral resolution. (**a**) Lateral resolution evaluation for max projection; (**b**) lateral resolution evaluation for deep learning. Note that we get an image of 1000 × 1000 with a

field of view of 5 × 5 mm^2. In other words, 1 unit in pixel corresponds to 5 μm in length. Thus, the FWHM length of (**a**) and (**b**) are 23.5 μm and 13.5 μm, respectively. The imaging depth of the system was measured to be about 600 μm. The limit of resolution is 12.98 μm (d = 1.22 λf/D; f = 100 mm, D = 5 mm, λ = 532 nm; f, focal length of the field lens; D, diameter of the laser beam; λ, wavelength of the laser).

Then, we extracted 30 pixels from the two images in the same position along the white lines in Figure 8. The pixel intensity distributions of the images reconstructed by max projection and deep learning are also depicted in Figure 8 for comparison. It is obvious that the image reconstructed by deep learning has a lower intensity compared to that of max projection. In other words, the former has a lower impact of background noise than the latter, indirectly implying that the image reconstructed by deep learning boasts higher contrast than that of the max projection.

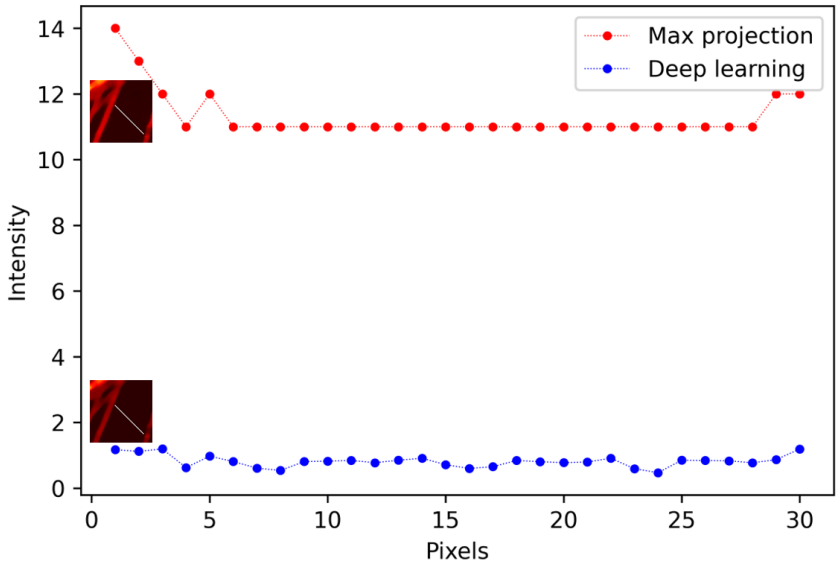


Figure 8. Comparison of the intensity between the image reconstructed by deep learning and that reconstructed by max projection.

4.3. Effective Rate Evaluation

In order to evaluate the effectiveness of our algorithm, we used 50 pairs (divided into 10 pairs, each containing five samples) of carbon-fiber images (300 × 300 $pixels^2$) for the resolution test. If the resolution of the image reconstructed by deep learning outperforms the one reconstructed by max projection, the algorithm of deep learning is considered to be effective. Effective rate means the ratio of the number of effective pairs to that of the total (50 pairs). The FWHM values of both the images reconstructed by max projection and those by deep learning are given in Table 2. The reciprocals of the data in Table 2 were calculated and are illustrated in Figure 9.

From Figure 9, it is evident that each of the systems using deep learning exhibits a higher resolution compared to that of max projection. The effective rate of our algorithm is measured to be 100%.

4.4. Imaging Accuracy Evaluation of the System

To evaluate the imaging accuracy of our system, a planar standard part made of carbon fiber with a fixed diameter of 7 μm was used for imaging. We imaged the standard part and compared the width of the carbon fiber (FWHM) in the image with the actual width of the carbon fiber (7 μm), which could reflect the imaging accuracy of our system to some extent. Fifteen groups of data were sampled from the image, as shown in Figure 10 and

Table 3. The average value of the measurement results (8.406 μm in Table 3) is 1.406 μm larger than the actual size (7 μm).

Table 2. FWHM (unit: pixels) of the images reconstructed by max projection (MP) and deep learning (DL).

Sample No. / Group No.	1		2		3		4		5	
	MP	**DL**	**MP**	**DL**	**MP**	**DL**	**MP**	**DL**	**MP**	**DL**
1	10.1	2.3	15.6	1.4	13.3	4.5	8.7	1.8	7.9	1.8
2	9.3	2.4	10.0	3.5	8.8	1.9	9.7	2.1	6.1	1.7
3	8.0	5.4	3.4	1.9	5.0	2.0	5.1	1.3	5.1	2.3
4	7.8	1.1	5.9	2.6	7.0	1.6	5.3	3.2	4.5	1.6
5	3.3	1.7	5.0	1.5	2.9	1.8	4.2	2.4	4.5	1.4
6	5.7	1.3	8.2	1.6	7.1	3.0	5.8	2.6	4.2	2.3
7	5.0	1.9	3.2	1.9	5.5	1.8	4.3	1.9	4.1	2.1
8	7.4	1.4	4.6	1.3	4.1	1.5	5.4	1.5	5.6	2.2
9	7.6	2.6	6.4	3.1	5.3	2.8	5.0	3.1	7.7	2.3
10	5.5	3.4	5.7	2.5	7.4	2.8	6.3	3.1	5.2	2.3
Effective rate	100%									

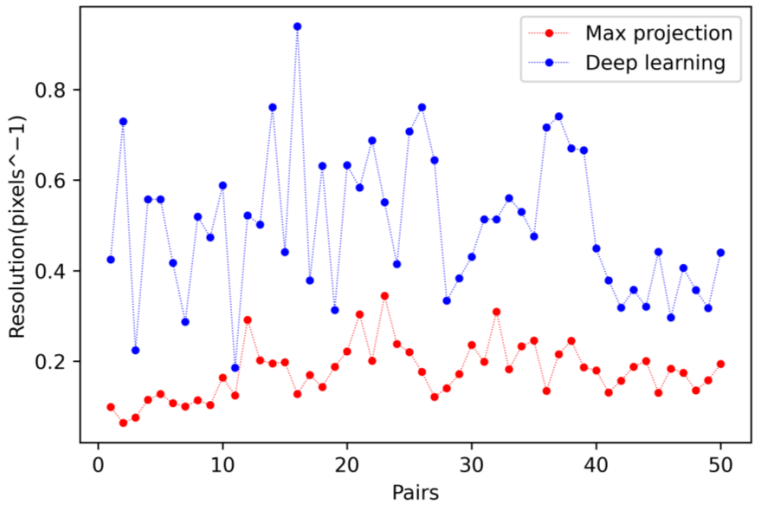


Figure 9. Comparison of the resolution between the images reconstructed by deep learning and those reconstructed by max projection.

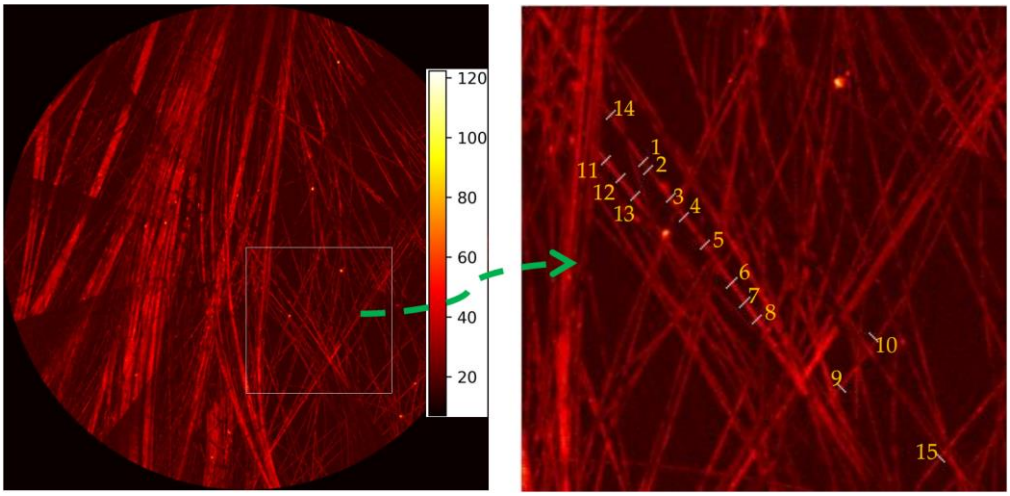


Figure 10. The images of the standard part and sampling regions (marked from 1 to 15).

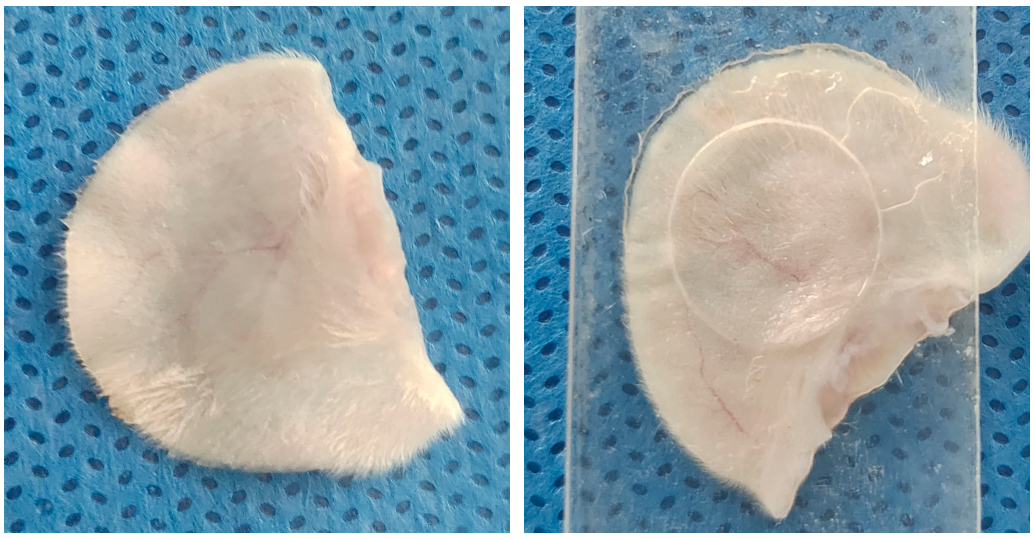
Table 3. The measurement results.

Group No.	FWHM (Pixels)	FWHM (μm)
1	1.714	8.570
2	1.449	7.245
3	1.994	9.970
4	2.076	10.380
5	1.904	9.520
6	1.814	9.070
7	1.304	6.520
8	1.883	9.415
9	1.887	9.435
10	1.518	7.590
11	1.584	7.920
12	2.096	10.480
13	1.617	8.085
14	1.340	6.700
15	1.037	5.185
Ave	1.681	8.406
Err	0.281	1.406

4.5. Biological Tissue Imaging Experiment

4.5.1. Sample Preparation

To obtain datasets for unsupervised learning, SD mouse (6–10 weeks) ears were supplied by West China Dental Hospital for in vitro imaging. The ears were positioned at the bottom of the water tank, secured in the middle by two 10 mm-diameter perforated plastic sheets to allow light in and ultrasonic waves out, as Figure 11 shows. This study was approved by the Ethics Committee of West China Hospital of Stomatology, Sichuan University (No. WCHSIRB-AT-2024-004) and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). An area of 5×5 mm^2 of the mouse ears was imaged at a step size of 5 μm. We got 324 pairs of mouse-ear images from 30 different mice, with 274 for training and 50 for testing. Five or six pairs of images were taken of different locations of a mouse ear. Notably, the hair on the mouse ears remained intact throughout the entire experiment.



(**a**) (**b**)

Figure 11. Images of a mouse ear in vitro. (**a**) The mouse ear with hair intact. (**b**) The mouse ear caught by two plastic sheets.

4.5.2. Imaging Experiment and Results

The length of an A-line is 1024. Normally, there is only a small part of the data that contains desired information (vessels and capillaries). We calibrated the distribution of the valid data and discarded the useless 1024 depths. Only valid data were chosen for image reconstruction, which saved computing resources and time.

As Figure 12 shows, I_1 and I_2 are source images derived from the results of the max projection of two distinct 100-length layers in 1024 depths, which cost 18 s. After that, they are concatenated in channel dimension and sent to the U-Net for reconstruction. I_f is the reconstructed image, which preserves the structural information of I_1 and I_2. It took 4.82 s from I_1 and I_2 to I_f.

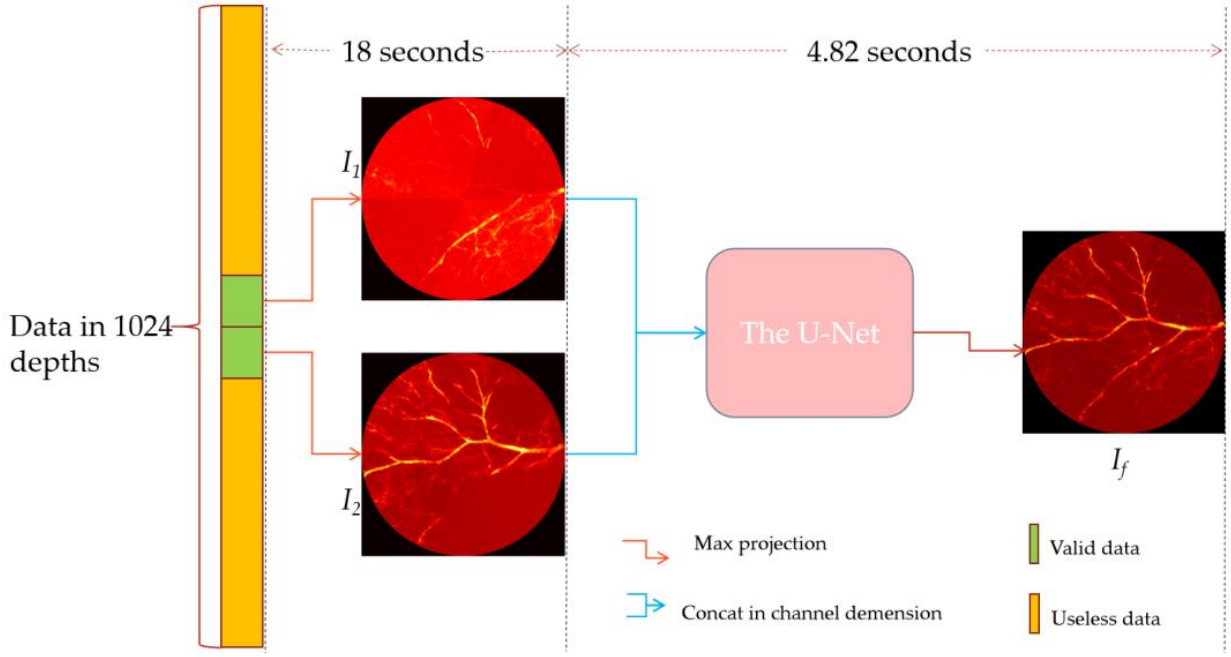


Figure 12. The process of image reconstruction with deep learning.

The images labeled column (a) and column (b) in Figure 13 are derived from two distinct 100-length layers containing vasculature information. Images in column (c) are reconstructed from the max projection of the data within a 200-length layer. Images in column (d) are obtained from the trained U-Net, utilizing the images in column (a) and column (b) as the input.

In summary, the images reconstructed by deep learning exhibit a lower intensity distribution but higher contrast. It is intuitively observable that the vasculature images in column (d) are clearer than those in column (c). Column (c) exhibits higher intensity, causing the background to appear brighter and the vasculature to appear blurred. Therefore, images in column (d) have a better performance compared to those in column (c).

To make a better comparison of the images of column (c) and column (d) in Figure 13, we sample 200 points along the white lines, as shown in Figure 14. It is obvious that the images reconstructed by deep learning have a much lower level of background noise than those reconstructed by max projection, which means the images of (d) outperform those of (c).

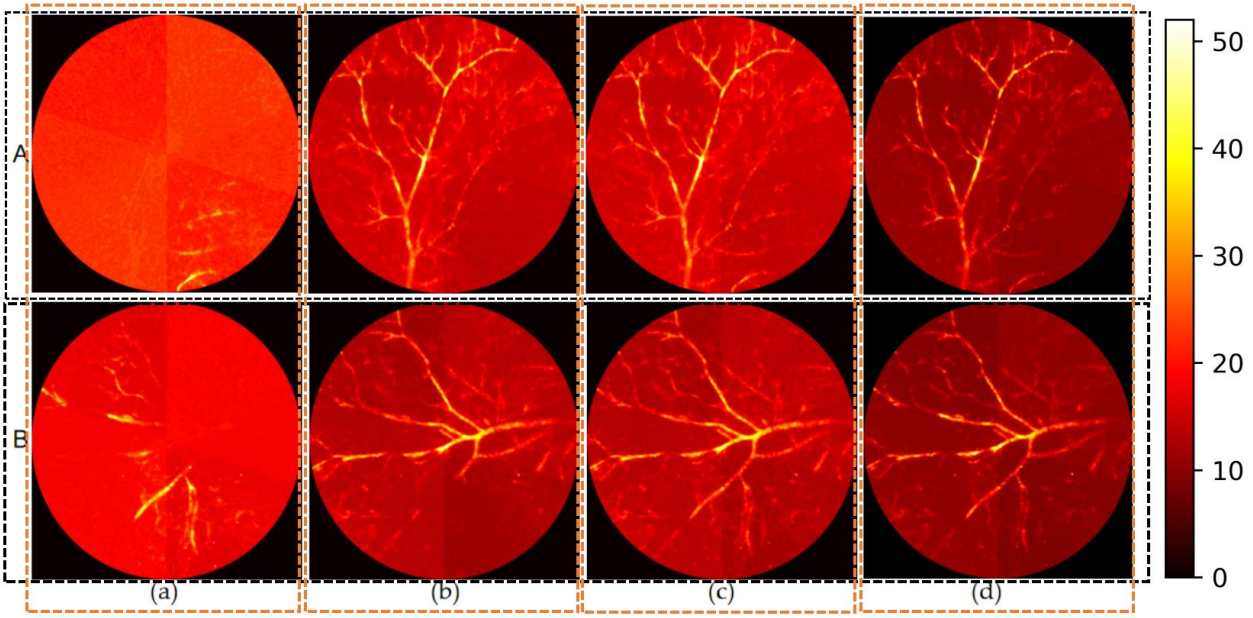


Figure 13. Two groups of images are marked as row A and row B, each surrounded by a black dashed box, respectively. Column (**a**), source images I_1; column (**b**), source images I_2; column (**c**), the images reconstructed by max projection; column (**d**), the images reconstructed by deep learning.

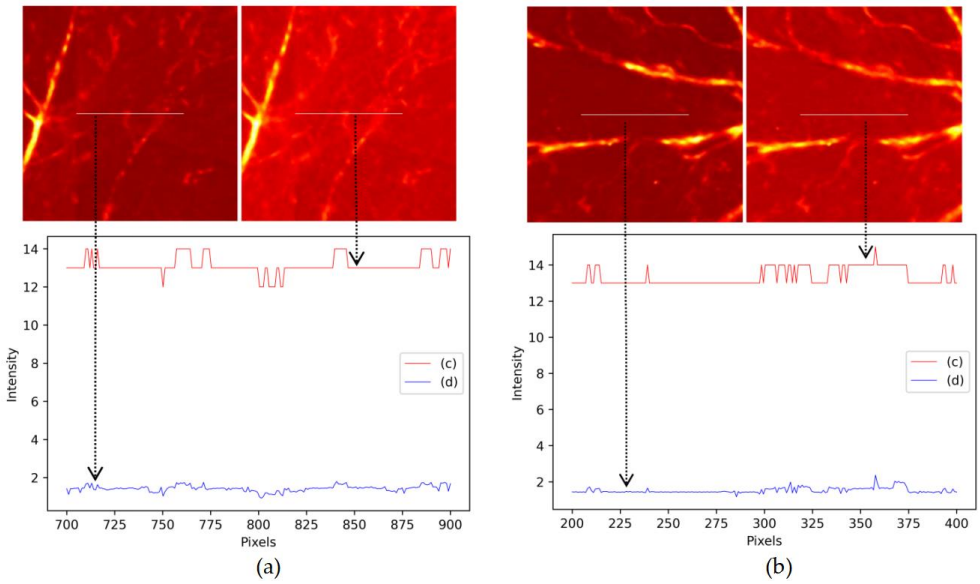


Figure 14. (**a**) The comparison of the background noise of the images in row A; the average intensities of the background noise of (c) and (d) are 13.12 and 1.43, respectively. (**b**) The comparison of the background noise of the images in row B; The average intensities of the background noise of (c) and (d) are 13.35 and 1.52, respectively.

5. Discussion and Conclusions

To acquire datasets for unsupervised learning, we constructed a transmissive OR-PAM system capable of imaging biological tissue in vitro. In order to accelerate the imaging speed, we improved the scanning method from a diameter-by-diameter approach to a sector-by-sector one. The data were preprocessed from polar coordinates to Cartesian coordinates by a specific max projection algorithm. The in vitro mouse ears were supplied by West China Dental Hospital. A total of 324 pairs of datasets were collected, with 274 designated for training and 50 for testing. No validation data were utilized in the training process. After a 100-epoch training, we evaluated the reconstructed images both

subjectively and objectively. Carbon fibers were also utilized for imaging to evaluate the lateral resolution of the PAM system and the effective rate of our algorithm.

It took 109 s to capture an image of 1000 × 1000 pixels^2 with a field of view (FOV) of 5 × 5 mm^2 by using our PAM system. Overall, the images reconstructed by the U-Net exhibit higher contrast and better lateral resolution compared to those reconstructed by max projection. Some blurred vasculature in the max projection images becomes clear in the images outputted by the trained U-Net model. The lateral resolution improves from the original 23.5 μm to 13.5 μm, an impressive improvement of 76% achieved. The imaging depth of our system is about 600 μm. The effectiveness rate of our algorithm was measured to be 100%. Our analysis also clearly indicates a significant improvement in image contrast.

Photoacoustic imaging is a promising medical-imaging modality that combines the advantages of high contrast and high resolution. It is difficult to improve the imaging depth and resolution simultaneously. Many research teams attempt to address the challenge of simultaneously improving the imaging depth and resolution in photoacoustic imaging using deep-learning methods. However, an obstacle they face is the requirement to obtain ground-truth data for training the deep-learning models, which can be time-consuming and difficult in photoacoustic microscopy. The approach outlined in this paper liberates researchers from the necessity of collecting ground-truth data for deep learning. Additionally, the training task demonstrated effectiveness, with just 274 pairs of training data. This work not only provides a solution for unsupervised learning in photoacoustic microscopy but also expands the possibilities for its application in medical imaging.

However, there still exist some shortcomings in our study. (1) We have to rotate the ultrasonic transducer because of its line-focus property, which has a higher SNR compared to a flat-field one but increases time cost. A more efficient ultrasonic transducer is needed to improve the image speed and lateral resolution. The new-type fiber ultrasonic transducer [19,20] brings hope to the PAM research field, which can improve SNR and imaging speed simultaneously due to its larger response area and higher sensitivity compared to a traditional piezoelectric ultrasonic transducer. (2) Slight information loss appeared during the image-fusion process due to limitations in the network model. To address these limitations, future efforts should focus on proposing more effective metrics and enhancing the network to preserve structural information to the maximum extent. Moreover, ongoing work is needed to improve the image quality of photoacoustic microscopy from both hardware and algorithmic perspectives.

Author Contributions: Conceptualization, L.Y. and J.S.; methodology, L.Y. and W.C.; software, L.Y. and T.K.; validation, L.Y., T.K. and C.L.; formal analysis, L.Y. and M.Y.; investigation, M.Y. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2022YFC2410102), the National Natural Science Foundation of China (62105227), the Research and Develop Program, West China Hospital of Stomatology Sichuan University (RD-03-202408), the Jiangxi Science and Technology Program (20224AAC01011), and the Sichuan Science and Technology Program (2022YFS0113).

Data Availability Statement: Data are partly contained within the article. Due to confidentiality, raw data cannot be disclosed at this time but are available from the corresponding author upon a reasonable request.

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Ntziachristos, V.; Razansky, D. Molecular imaging by means of multispectral optoacoustic tomography (MSOT). *Chem. Rev.* **2010**, *110*, 2783–2794. [CrossRef] [PubMed]
2. Xu, M.; Wang, L.V. Photoacoustic imaging in biomedicine. *Rev. Sci. Instrum.* **2006**, *77*, 041101. [CrossRef]
3. Wang, L.V. Multiscale photoacoustic microscopy and computed tomography. *Nat. Photonics* **2009**, *3*, 503–509. [CrossRef] [PubMed]
4. Wang, L.V.; Yao, J. A practical guide to photoacoustic tomography in the life sciences. *Nat. Methods* **2016**, *13*, 627–638. [CrossRef] [PubMed]

5. Cao, R.; Li, J.; Ning, B.; Sun, N.; Wang, T.; Zuo, Z.; Hu, S. Functional and oxygenmetabolic photoacoustic microscopy of the awake mouse brain. *Neuroimage* **2017**, *150*, 77–87. [CrossRef] [PubMed]
6. Jin, T.; Guo, H.; Jiang, H.; Ke, B.; Xi, L. Portable optical resolution photoacoustic microscopy (pORPAM) for human oral imaging. *Opt. Lett.* **2017**, *42*, 4434–4437. [CrossRef] [PubMed]
7. Zabihian, B.; Weingast, J.; Liu, M.; Zhang, E.; Beard, P.; Pehamberger, H.; Drexler, W.; Hermann, B. In vivo dual-modality photoacoustic and optical coherence tomography imaging of human dermatological pathologies. *Biomed. Opt. Express* **2015**, *6*, 3163–3178. [CrossRef]
8. Wong, T.T.; Zhang, R.; Hai, P.; Zhang, C.; Pleitez, M.A.; Aft, R.L.; Novack, D.V.; Wang, L.V. Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy. *Sci. Adv.* **2017**, *3*, e1602168. [CrossRef]
9. Berezhnoi, A.; Schwarz, M.; Buehler, A.; Ovsepian, S.V.; Aguirre, J.; Ntziachristos, V. Assessing hyperthermia-induced vasodilation in human skin in vivo using optoacoustic mesoscopy. *J. Biophotonics* **2018**, *11*, e201700359. [CrossRef]
10. Bell, A.G. The photophone. *J. Frankl. Inst.* **1880**, *110*, 237–248. [CrossRef]
11. Lengenfelder, B.; Mehari, F.; Hohmann, M.; Heinlein, M.; Chelales, E.; Waldner, M.J.; Klämpfl, F.; Zalevsky, Z.; Schmidt, M. Remote photoacoustic sensing using speckleanalysis. *Sci. Rep.* **2019**, *9*, 1057. [CrossRef] [PubMed]
12. Wang, L.V.; Hu, S. Photoacoustic tomography: In vivo imaging from organelles toorgans. *Science* **2012**, *335*, 1458–1462. [CrossRef] [PubMed]
13. Beard, P. Biomedical photoacoustic imaging. *Interface Focus* **2011**, *1*, 602–631. [CrossRef] [PubMed]
14. Maslov, K.; Zhang, H.F.; Hu, S.; Wang, L.V. Optical-resolution photoacoustic microscopy for in vivo imaging of single capillaries. *Opt. Lett.* **2008**, *33*, 929–931. [CrossRef] [PubMed]
15. Hu, S.; Maslov, K.; Wang, L.V. Second-generation optical-resolution photoacoustic microscopy with improved sensitivity and speed. *Opt. Lett.* **2011**, *36*, 1134–1136. [CrossRef] [PubMed]
16. Maslov, K.; Stoica, G.; Wang, L.V. In vivo dark-field reflection-mode photoacoustic microscopy. *Opt. Lett.* **2005**, *30*, 625–627. [CrossRef]
17. Chuangsuwanich, T.; Moothanchery, M.; Yan, A.T.C.; Schmetterer, L.; Girard, M.J.A.; Pramanik, M. Photoacoustic imaging of lamina cribrosa microcapillaries in porcine eyes. *Appl. Opt.* **2018**, *57*, 4865–4871. [CrossRef]
18. Bi, R.; Balasundaram, G.; Jeon, S.; Tay, H.C.; Pu, Y.; Li, X.; Moothanchery, M.; Kim, C.; Olivo, M. Photoacoustic microscopy for evaluating combretastatin A4 phosphate induced vascular disruption in orthotopic glioma. *J. Biophotonics* **2018**, *11*, e201700327. [CrossRef]
19. Liang, Y.; Jin, L.; Wang, L.; Bai, X.; Cheng, L.; Guan, B.O. Fiber-laser-based ultrasound sensor for photoacoustic imaging. *Sci. Rep.* **2017**, *7*, 40849. [CrossRef]
20. Liang, Y.; Liu, J.W.; Jin, L.; Guan, B.O.; Wang, L. Fast-scanning photoacoustic microscopy with a side-looking fiber optic ultrasound sensor. *Biomed. Opt. Express* **2018**, *9*, 5809–5816. [CrossRef]
21. Chen, J.; Zhang, Y.; He, L.; Liang, Y.; Wang, L. Wide-field polygon-scanning photoacoustic microscopy of oxygen saturation at 1-MHz A-line rate. *Photoacoustics* **2020**, *20*, 100195. [CrossRef] [PubMed]
22. Zhao, H.; Ke, Z.; Yang, F.; Li, K.; Chen, N.; Song, L.; Zheng, C.; Liang, D.; Liu, C. Deep learning enables superior photoacoustic imaging at ultralow laser dosages. *Adv. Sci.* **2020**, *8*, 2003097. [CrossRef] [PubMed]
23. Chen, X.; Qi, W.; Xi, L. Deep-learning-based motion-correction algorithm in optical resolution photoacoustic microscopy. Visual Computing for Industry. *Biomed. Art* **2019**, *2*, 12. [CrossRef]
24. DiSpirito, A.; Li, D.; Vu, T.; Chen, M.; Zhang, D.; Luo, J.; Horstmeyer, R.; Yao, J. Reconstructing undersampled photoacoustic microscopy images using deep learning. *IEEE Trans. Med. Imaging* **2020**, *40*, 562–570. [CrossRef]
25. Zhou, J.; He, D.; Shang, X.; Guo, Z.; Chen, S.-L.; Luo, J. Photoacoustic microscopy with sparse data by convolutional neural networks. *Photoacoustics* **2021**, *22*, 100242. [CrossRef] [PubMed]
26. Cheng, S.; Zhou, Y.; Chen, J.; Li, H.; Wang, L.; Lai, P. High-resolution photoacoustic microscopy with deep penetration through learning. *Photoacoustics* **2021**, *25*, 100314. [CrossRef] [PubMed]
27. Gröhl, J.; Schellenberg, M.; Dreher, K.; Maier-Hein, L. Deep learning for biomedical photoacoustic imaging: A review. *Photoacoustics* **2021**, *22*, 100241. [CrossRef] [PubMed]
28. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 502–518. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.