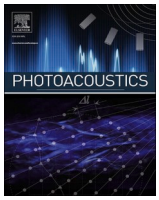
Photoacoustics 38 (2024) 100632

Contents lists available at ScienceDirect

Photoacoustics

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



Unsupervised deep learning enables real-time image registration of fast-scanning optical-resolution photoacoustic microscopy

Xiaobin Hong [a], Furong Tang [a], Lidai Wang [b,*], Jiangbo Chen [a,*]

[a] School of Mechanical & Automotive Engineering, South China University of Technology, Guangzhou, Guangdong, PR China
[b] Department of Biomedical Engineering, City University of Hong Kong, 83 Tat Chee Ave, Kowloon, Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:
Photoacoustic microscopy
Unsupervised deep learning
Image registration

ABSTRACT

A fast scanner of optical-resolution photoacoustic microscopy is inherently vulnerable to perturbation, leading to severe image distortion and significant misalignment among multiple 2D or 3D images. Restoration and registration of these images is critical for accurately quantifying dynamic information in long-term imaging. However, traditional registration algorithms face a great challenge in computational throughput. Here, we develop an unsupervised deep learning based registration network to achieve real-time image restoration and registration. This method can correct artifacts from B-scan distortion and remove misalignment among adjacent and repetitive images in real time. Compared with conventional intensity based registration algorithms, the throughput of the developed algorithm is improved by 50 times. After training, the new deep learning method performs better than conventional feature based image registration algorithms. The results show that the proposed method can accurately restore and register the images of fast-scanning photoacoustic microscopy in real time, offering a powerful tool to extract dynamic vascular structural and functional information.

1. Introduction

Optical-resolution photoacoustic microscopy (OR-PAM), characterized by subcellular resolution, rich optical contrasts, and label-free imaging capability, has shown inspiring prospects in anatomical, functional, and histological studies [1–7]. A fast scanning speed plays a crucial role in enhancing throughput and enabling the study of dynamic physiological or pathological processes in vivo [8,9]. Several fast scanners have been developed to achieve fast scanning imaging, such as water-immersible micro-electro-mechanical systems (MEMS) scanners and polygon scanners [10–14]. In fast-scanning OR-PAM systems, various factors such as manufacturing precision, installation error, and material fatigue of the scanners can impact the scanning trajectory uniformity. Under severe working conditions, for example, in high-speed water-immersible scanning with fiber based photoacoustic probe or in a handheld imaging system, the scanner suffers from constant random disturbances [15–19]. These factors may result in distortions of individual images and the misalignments among repetitive images, thereby affecting subsequent signal enhancement, feature extraction, and quantitative image analysis et al. Therefore, it is of great importance to restore and register distorted images for fast-scanning OR-PAM.

The scale-invariant feature transform (SIFT) and speeded-up robust features (SURF) methods are commonly used to extract coordinates of feature points from image pairs [20,21]. Then the deformed images can be corrected based on the coordinates. However, these feature point based registration algorithms may not be suitable for images with blurred feature points or a limited number of features [22]. Schwarz et al. proposed a method to address the displacement between adjacent B-scanning layers in acoustic-resolution photoacoustic microscopy (AR-PAM). However, because the limited penetration depth of OR-PAM offers an insufficient number of reference objects in depth, the dynamic reference surface required by this method is challenging to realize in high-resolution OR-PAM [23]. Huang et al. introduced a multi-scale vascular feature-matching algorithm based on the Demons transform to correct motion artifacts in mice vasculature [24]. Unfortunately, it suffered from slow data processing speed. In recent years, deep learning has played an increasingly vital role in medical image processing [25]. Some studies have attempted to remove artifacts from photoacoustic images using deep learning techniques [26–29]. Chen et al. proposed a deep learning based motion correction algorithm for OR-PAM that effectively corrected distortions in arbitrary directions [30].

* Corresponding authors.
E-mail addresses: lidawang@cityu.edu.hk (L. Wang), cjiangbo@scut.edu.cn (J. Chen).

https://doi.org/10.1016/j.pacs.2024.100632
Received 6 May 2024; Received in revised form 16 June 2024; Accepted 2 July 2024
Available online 5 July 2024

Nevertheless, the network required ground-truth data as input, which is not easily available in photoacoustic imaging. Sun et al. proposed a method for correcting motion artifacts in continuous sequences of intravascular photoacoustic images [31]. However, the application of deep learning methods to address the misalignment among multiple 2D or 3D images of OR-PAM has not been conducted.

Herein, we report an unsupervised deep learning based method that overcomes the limitations of previously mentioned techniques and achieves real-time image registration for fast-scanning OR-PAM. Our approach utilizes mutual information as the similarity metric between image sequences, approximating the nonlinear mapping from a distorted image to its undistorted counterpart. Importantly, this method eliminates the need for ground-truth input. We demonstrate the accuracy of our registration algorithm and its real-time processing capability using image frames obtained from a water-immersible resonant mirror based OR-PAM system. The results highlight the significant improvement achieved in correcting intra-image artifacts and inter-image misalignment, effectively addressing the challenge of real-time image registration encountered by high-speed photoacoustic microscopy during long-term imaging.

2. Materials and methods

2.1. Imaging system

The water-immersible resonant mirror based OR-PAM system has been described in previous work [12]. The optical beams were reflected onto the sample by a single-axis water-immersible resonant mirror coated with aluminum. The generated ultrasound waves were reflected by the resonant mirror, collimated by a planoconcave acoustic lens (#48–267-INK, Edmund Optics Inc), then transmitted through two prisms, and finally detected by a piezoelectric transducer (with a 50-MHz center frequency and 78 % bandwidth, V214-BC-RM, Olympus). A linear translation stage (PLS-85, Physik Instrumente GmbH & Co) was used to drive the fiber-based PA probe to scan in the slow axis direction. The A-line rate was 3.2 MHz, the B-scan rate was 2036 Hz, and the C-scan rate reached 1.7 Hz over an area of 2.5 × 6.7 mm^2. 5 micrograms of epinephrine was injected into the muscles of the hind legs of mice. It can induce changes in hemoglobin concentration and oxygen saturation (sO_2) of the blood vessels in the mouse ear.

2.2. Registration network

The network structure, as depicted in Fig. 1, comprises a convolutional neural network (CNN) and a spatial transformation module [32]. In this framework, an image pair of fixed I_f and moving I_m images (defined in a two-dimensional spatial domain $\Omega \subset \mathbb{R}^2$) is input into the network. The CNN, denoted as $g_\theta(I_f, I_m)$, then generates a deformation field ϕ. θ represents the parameter of the CNN. Subsequently, this deformation field ϕ and the moving image I_m are fed into the spatial transformation module, where the moving image undergoes warping. The final output is the predicted registration image ($I_m \circ \phi$). The spatial transformation module includes a grid generator and a sampler. For each pixel p, a sub-pixel position $p' = p + g_\theta(p)$ is computed within the moving image. Here, $g_\theta(p)$ represents the deformation of pixel p. The values of eight neighboring pixels are interpolated using the bilinear interpolation method:

$$I_m \circ \phi(p) = \sum_{q \in Z(p')} I_m(q) \prod_{d \in [x,y]} \left(1 - |p'_d - q_d|\right) \quad (1)$$

where $Z(p')$ is the pixel neighbors of p', and d iterates over dimensions of Ω.

We opt for the UNet [33] as the CNN to generate the deformation field, with subsequent improvement incorporated. Specifically, we introduce a simplified attention module after the downsampling layer. This module comprises a convolutional layer with a kernel size of 1, followed by a sigmoid layer. The attention mechanism assigns weights to the input feature map through convolutional operations and the sigmoid function, as expressed in the following Eq. (2). This allows the network to allocate varying weights to different pixels during feature map processing, enhancing its performance by focusing more on useful features.

$$\hat{x} = x \cdot Sigmoid(Conv(x)) \quad (2)$$

In this study, we focus on two types of image distortion. One is intra-image artifacts, which refer to the misalignment between adjacent odd and even columns. The other type of distortion is the deformation between image sequences. It specifically manifests as misalignment and certain local deformations between the current odd frames and their preceding odd frames and the subsequent even frames.

To address various distortion types, we adopt network structures with varying depths and loss functions tailored to each distortion scale. For the first type of distortion with smaller scales, two additional downsampling and upsampling layers are added to the UNet architecture shown in Fig. 1. This enhancement is designed to capture deformations at finer spatial resolutions. Each layer in the encoder has convolutional kernels of size 4, a stride of 2, and padding of 1. Following each convolutional layer, a LayerNorm layer and ReLU activation layer with a parameter of 0.2 are added. With each layer, the input size is halved. After 6 layers, the final feature map size output by the encoder is $(1/64)^2$ of the input image size. During the decoding phase, the output size is progressively increased through bilinear interpolation, with a sampling factor set to 2. The upsampling process culminates in a deformation field that is the same size as the input image. Skip connections integrate features learned during the encoding phase directly with corresponding layers in the decoding phase, thereby aiding the network in better preserving low-level features and spatial information.

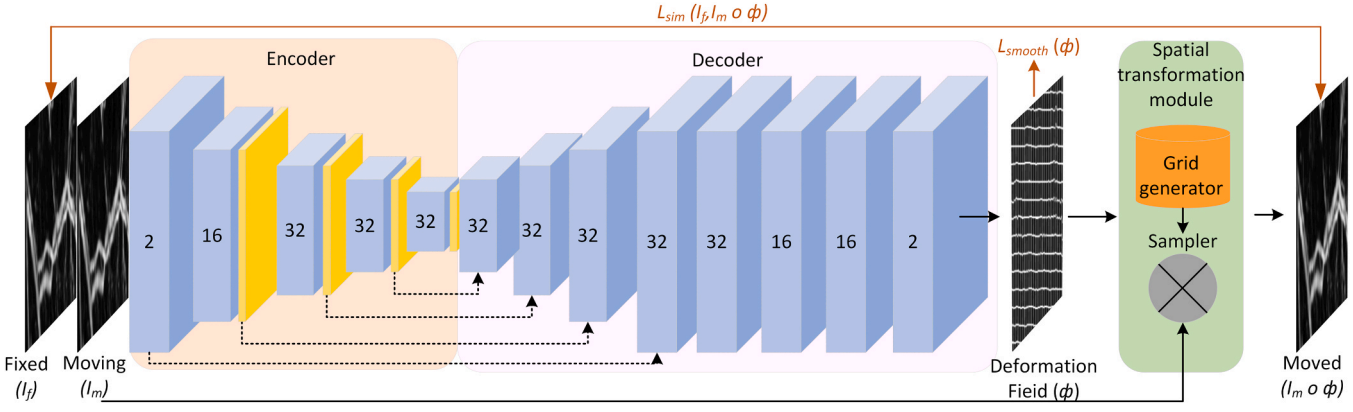


Fig. 1. The registration network diagram.

For scenarios where fixed and moving images exhibit similar image intensity distributions, the loss function is defined as follows:

$$L_{sim}\left(I_f, I_m \circ \phi\right) = \frac{1}{|\Omega|}\sum_{p\in\Omega}\left[I_f(p) - \left[I_m \circ \phi\right](p)\right]^2 \tag{3}$$

$$L_{smooth}(\phi) = \sum_{p\in\Omega}\left\|\nabla^2 u(p)\right\|^2 \tag{4}$$

$$\nabla^2 u(p) = \left(\left(\frac{\partial^2 u}{\partial x \partial x}\right), \left(\frac{\partial^2 u}{\partial y \partial y}\right)\right) \tag{5}$$

$$\frac{\partial u}{\partial x} \approx u\left(\left(p_x + 1, p_y\right)\right) - u\left(\left(p_x, p_y\right)\right) \tag{6}$$

$$\frac{\partial^2 u}{\partial x \partial x} \approx \frac{\partial u}{\partial x}\left(\left(p_x + 1, p_y\right)\right) - \frac{\partial u}{\partial x}\left(\left(p_x, p_y\right)\right) \tag{7}$$

$$\frac{\partial u}{\partial y} \approx u\left(\left(p_x, p_y + 1\right)\right) - u\left(\left(p_x, p_y\right)\right) \tag{8}$$

$$\frac{\partial^2 u}{\partial y \partial y} \approx \frac{\partial u}{\partial y}\left(\left(p_x, p_y + 1\right)\right) - \frac{\partial u}{\partial y}\left(\left(p_x, p_y\right)\right) \tag{9}$$

$$L = L_{sim}\left(I_f, I_m \circ \phi\right) + \lambda L_{smooth}(\phi) \tag{10}$$

Where λ is a hyperparameter. We choose the mean squared error (MSE) as the similarity loss, which capitalizes on the disparities between adjacent pixels to approximate spatial gradients. The reason for introducing the smoothing loss is that during the registration process, discontinuous deformation fields are often generated to maximize the similarity metric of the images, whereas ideal deformation fields should be diffeomorphic to ensure that the topological properties are not altered. Therefore, we introduce regularization penalties to enforce the continuity and overlap of the deformation field [34].

For the second type of deformation between image sequences with larger scales, the structure of the UNet is as shown in Fig. 1. The configurations of the convolutional layers are similar to the basic ones mentioned above. After 4 downsampling operations, the encoder outputs a feature map that is $(1/16)^2$ the size of the input image. The decoder remains the same configuration as above. However, we modify the loss function to better measure similarity. Given that MSE is more suitable for scenarios with comparable fixed and moving image intensities, it falls short when tracking dynamic photoacoustic image changes over time, such as sO_2 fluctuations, where image intensities also shift. As the MSE similarity loss function for registration can lead to inaccuracies, we opt for mutual information (MI) as the similarity loss function. MI facilitates a more accurate assessment of the similarity between photoacoustic images captured at different time points by effectively modeling the probabilistic relationships between pixel intensities [35]. The expression for mutual information is as follows:

$$L_{sim}\left(I_f, I_m \circ \phi\right) = -\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{M}\sum_{j=1}^{M}\exp\left(-\frac{1}{2}\left(\frac{\left|\left(I_m \circ \phi\right)_i - \left(I_f\right)_j\right|}{\sigma}\right)^2\right)\right) \tag{11}$$

We compute the mutual information using the Parzen window formula for the Gaussian function since the histogram based probability computation method is not differentiable and thus not applicable in deep learning [36]. In the Eq. (11), n represents the number of samples, M denotes the absolute value of the distance between the predicted output and the target. $\exp\left(-\frac{1}{2}\left(\frac{\left|(I_m \circ \phi) - I_f\right|}{\sigma}\right)^2\right)$ corresponds to the Gaussian kernel function, and σ signifies the standard deviation of the Gaussian kernel, serving as a hyperparameter.

3. Experimental procedures

3.1. Data preparation

The dataset comprises 399 photoacoustic images of mouse ears along with their corresponding images of sO_2. During data preprocessing, the intensity values of the images were clamped based on their 1st and 99th percentiles to remove outliers. Next, we normalized these values to fall within the [0,1] range. Subsequently, a 3×3 median filter was applied to all data. Then an initial affine alignment was performed to roughly align the image position. The input image size of the network was 1024×320.

3.2. Network training

3.2.1. Intra-image registration

The workflow for intra-image even-odd column misalignment registration is illustrated in Fig. 2(a). Odd columns were extracted from the preprocessed data, halving the image length. Bilinear interpolation was then applied between columns, resulting in interpolated images of the same size as the input images. Although information from even columns was lost, this process ensured no misalignment between columns. These interpolated images served as fixed images, while the original unextracted and uninterpolated images served as moving images inputted into the registration network. This setup allowed the moving images to align with the fixed images while retaining information from even columns. In this scenario, the interpolated images provided an approximate baseline that was easily obtained, enabling the moving images to align with them without losing information. Ultimately, this process produced prediction moved images that were complete and free from misalignment.

The dataset was randomly divided into training and test sets in a ratio of 8:2. The Adam algorithm was chosen to optimize the parameters in the CNN, with an initial learning rate of 0.001, a batch size of 1, an iteration count of 6380, and λ set to 0.01.

3.2.2. Inter-image registration

The process of inter-image registration is shown in Fig. 2(b). Considering both distortions between adjacent odd and even frames and deformations between odd frames, we adopted a custom two-stage training strategy.

In Stage 1, an atlas based registration method was employed. We first extracted all odd frames from the training set. Assuming the first collected image had no distortion. Then, the first frame served as the fixed image, while the rest frames served as moving images to be aligned with the first frame. In Stage 2, we extracted all even frames from the training set. To preserve dynamic information between adjacent photoacoustic images, even frames were treated as moving images, while the preceding odd frame was used as the fixed image inputted into the network to learn how to align even frames with their preceding odd frames. During the testing phase, all odd frames in the testing set were first registered to the first frame to obtain moved predicted odd frames. Then, even frames in the testing set were treated as moving images, with adjacent moved predicted odd frames as fixed images, resulting in all moved predicted even frames.

80 % of the images were allocated to the training set and the remaining 20 % was reserved for the test set. The Gaussian kernel standard deviation σ for mutual information loss was set to 0.6 and the λ was set to 10. The total number of iterations for the training stages was 6380. Similarly, Adam was selected as the optimization algorithm with an initial learning rate of 0.001 and a batch size of 1. The network was implemented in Python 3.8 using the PyTorch framework. The workstation setup included a 13th Gen Intel(R) Core(TM) i7–13700KF 3.40 GHz CPU, 32 GB RAM, and an NVIDIA GeForce RTX 4080.

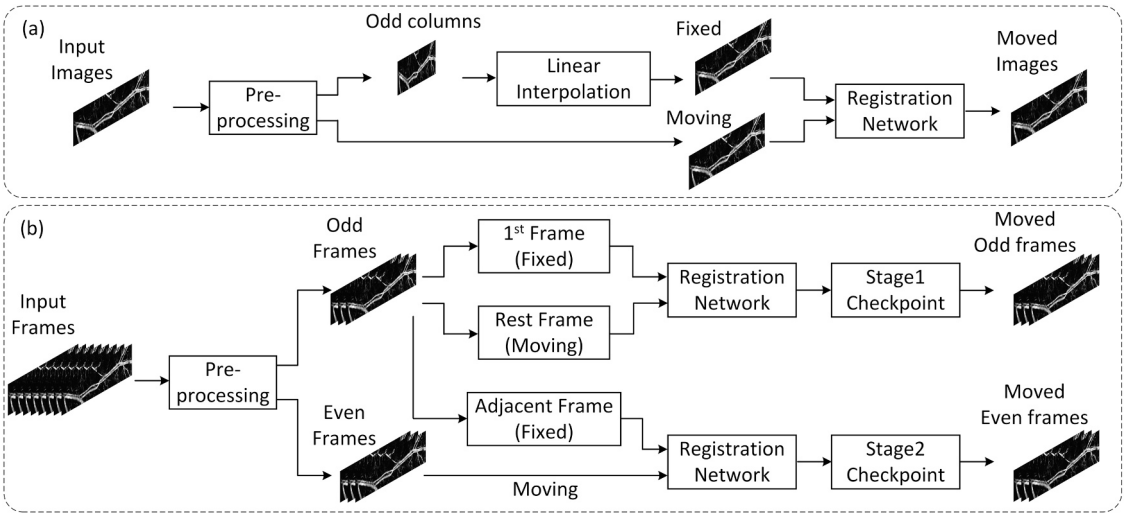


Fig. 2. (a) Intra-image registration workflow diagram. (b) Inter-image registration workflow diagram.

4. Results

4.1. Intra-image registration

The misalignment between even and odd columns and the corresponding registration results in the mouse ear region are illustrated in Fig. 3. Fig. 3(a) and (b) depict the maximum amplitude projection (MAP) images of the blood vessel structure before and after registration, respectively. To better demonstrate the performance of the proposed method, two regions marked with green boxes in Fig. 3(a) were enlarged and shown in Fig. 3(c) and (d). There are obvious jagged artifact structures caused by B-scan misalignment. Fig. 3(e) and (f) show the corresponding registration results. While the use of bilinear interpolation in the spatial transformation module causes a slight reduction in the spatial resolution which calculates new pixel values by weighted averaging of neighboring pixels, it does not hinder our ability to compare the images. We can still observe that our method effectively eliminates misalignment between the columns.

To fully validate the effectiveness of the proposed method, we extracted odd-column and even-column images from both the pre-corrected and post-corrected images, and quantitatively analyzed their similarity using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). A higher PSNR value indicates a higher similarity while a value of SSIM approaching 1 indicates a higher image similarity. If there is no misalignment between the odd and even columns, the odd-column and even-column images should exhibit high similarity. Additionally, to ensure that no image details were lost during processing, we calculated the ratio H of high-frequency information in the corrected image to that in the pre-corrected image. The closer this ratio H is to 1, the more high-frequency information is preserved. We also converted the images to the HSV color space and calculated the histogram intersection IS of the hue channel before and after correction. By comparing the histograms of the hue channel, we can measure the similarity of the color distribution without being affected by brightness and saturation. The closer the intersection IS is to 1, the more similar the color distributions of the images are, and the more information processed images retain. The results of the analysis are shown in Table 1.

It can be observed that our method effectively corrects the misalignment between odd and even columns, preserving the vast majority of the image information with less than 0.5 % loss of image detail.

Furthermore, we compared the registration results with different loss

Table 1
Quantitative analysis of image similarity before and after registration.

	PSNR	SSIM	H	IS
Before registration	21.2694	0.9398	-	-
After registration	23.8099	0.9595	0.9975	0.9967

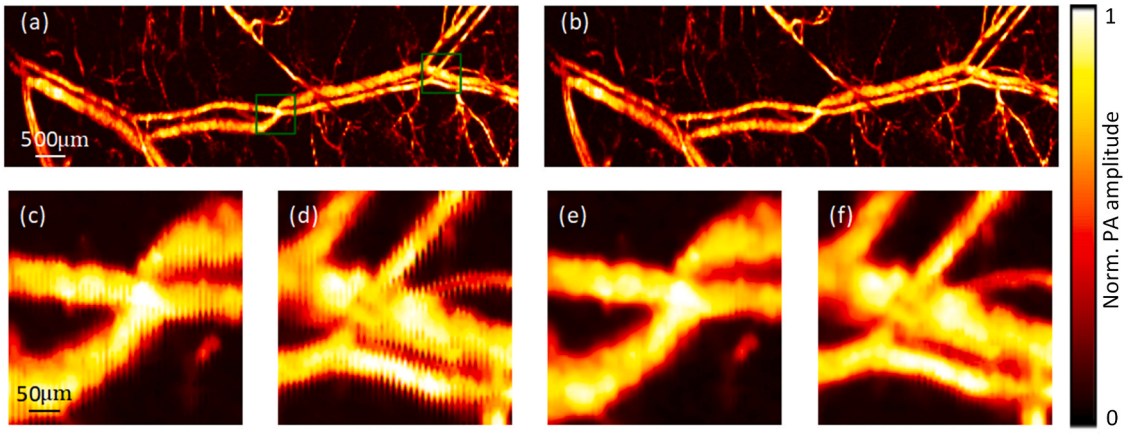


Fig. 3. (a) The MAP image of the blood vessel structure in the mouse ear before registration. (b) The MAP image after registration. (c) and (d) The enlarged images of the two green boxes in (a). (e) and (f) The enlarged images of corresponding areas in (b).

functions. For example, using SSIM as the similarity loss function, and comparing its results with those of MSE. We discuss the differences between the two loss functions from two perspectives.

First is computational complexity. The formula for calculating SSIM is shown below:

$$SSIM\left(x,y\right) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{12}$$

Where μ_x and μ_y are the means of images x and y respectively, σ_x^2 and σ_y^2 are their variances, σ_{xy} is the covariance between the two images, and C_1 and C_2 are constants used for stability to avoid division by zero. It's clear that SSIM has higher computational complexity, which can increase training time. In fact, we found that with the same number of training iterations, SSIM loss took 1.4 times longer than MSE during the experiments. Additionally, the SSIM loss involves more parameters than MSE loss, making it more susceptible to improper parameter choices that can affect the effectiveness of the loss function.

The second perspective is the registration effectiveness. We performed a quantitative analysis of the registration results for different loss functions in terms of structural similarity and information retention. The results are shown in Table 2.

Table 2 shows that when using SSIM as the loss function, although its structural similarity is slightly higher, it exhibits lower similarity in color distribution. This is because MSE directly measures the differences in pixel values between images, minimizing MSE loss aims to make images as close as possible at the pixel level. In contrast, SSIM takes multiple perspectives on brightness, contrast, and structure, emphasizing overall perceptual similarity rather than exact pixel-level matching. This may cause SSIM to overlook some details in certain cases.

Compared to the MSE loss, the SSIM loss only slightly improves structural similarity but requires longer training time and results in more loss of image information. Therefore, we believe that choosing MSE as the loss function is more reasonable.

4.2. Inter-image registration

Fig. 4 displays two pairs of unregistered MAP images collected at different times. We enlarged the distortions between adjacent frames and differentiated them with different colored box pairs. The area selected by the single blue box in (a) is shown enlarged in Fig. 6. It can also be seen that the amplitude of the signal in the blood vessels is changing due to the physiological response to adrenaline.

For comparison, we applied traditional non-learning based registration algorithms Demons [37] and the SIFT method to register the test set comprising a total of 80 images. We evaluated the results using common quantitative metrics, including PSNR, mutual information matrix (MIM), normalized cross-correlation (NCC), and the runtime on both the central processing unit (CPU) and graphics processing unit (GPU). The MIM was computed using mutual information and presented in grayscale form, where the higher overall brightness of the image indicates a higher correlation between the images. Similarly, NCC measures image similarity, with values closer to 1 indicating greater similarity.

Fig. 5 shows grayscale images representing the MIM for three different methods. The brightness of the MIM for the images corrected by SIFT, Demons, and the proposed method is significantly higher than that of the original images. This suggests effective suppression of deformations. Besides, our method appears to be the most optimal among the three methods. Fig. 6 shows the enlarged effect of overlapping adjacent odd frames within the blue box area of Fig. 4(a). Two images were overlaid in green and magenta colors, respectively. If the vessel positions in these two images perfectly align, the overlapped vessels should appear gray; otherwise, misalignment exists. Before registration, the misalignment is evident due to the presence of green and magenta on either side of the vessels. After correction by the registration network, these color disappears, indicating successful alignment between adjacent odd frames. Notably, while Demons also alignes the images, it sacrifices effective image information, as indicated by the yellow arrows in Fig. 6(i). This is attributed to the non-uniform grayscale of photoacoustic images, to which the Demons algorithm is highly sensitive. When the vascular morphology is changing, SIFT struggles to extract effective feature points for matching, leading to unsatisfactory registration results.

Table 2
Quantitative analysis results of different loss functions.

Loss function	SSIM	IS
MSE	0.9595	0.9967
SSIM	0.9627	0.9943

Table 3 provides a quantitative analysis of the PSNR, NCC, and runtime for different methods. It can be seen that after registration using the three methods, there is a great improvement in the PSNR and NCC of the image sequences, with our proposed method demonstrating superior performance. When executed on the same CPU, the Demons takes over 50 times longer than our approach. Moreover, when utilizing a GPU, our method achieves a doubled speed improvement. Overall, our proposed method achieves comprehensive performance advantages.

Fig. 7(a) - (d) show the corrected adjacent image pairs. The colored boxes indicate the areas where distortion was originally present. Compared to Fig. 4(a) - (d), Fig. 7(a) - (d) demonstrates that the proposed method can effectively mitigate edge distortion and deformation in even frames in contrast to their adjacent odd frames. The same network was applied to the sO_2 data. However, since the sO_2 images are composed of three color channels: red, green, and blue, we first split each image containing three channels into three images, each containing only one channel. Subsequently, each channel image was registered separately following the same process used for MAP images. Finally, the three images of different color channels were merged at the output to obtain the final corrected sO_2 image. In Fig. 7(e) - (h), we present the corresponding corrected images of sO_2. With the aligned image sequences, we analyzed changes in vessel diameter for two cross sections annotated in Fig. 7(a). Specifically, cross section A was extracted from the artery and cross section B was from the vein, as shown in Fig. 7(i). Notably, the diameter of the artery gradually decreases due to the effect of epinephrine and stabilizes around 180 seconds, while the diameter of the vein can be regarded as essentially unchanged. Similarly, we examined changes in sO_2 for two areas annotated in Fig. 7(e). Area 1 was extracted from the artery and area 2 was from the vein, as depicted in Fig. 7(j). It can be found that sO_2 in the artery remains essentially unchanged, whereas sO_2 in the vein exhibits a tendency to rise and then fall as a result of epinephrine, ultimately stabilizing around 120 seconds.

Table 4 quantitatively evaluates the similarity of the sO_2 dataset before and after registration with different algorithms. It is evident that the similarity metrics of the images have significantly improved after registration. Among the three methods, our method demonstrates the most impressive effect. Well-registered images enable more accurate observation of changes in blood sO_2 at specified locations. The dynamic demonstration of vascular structure and sO_2 are shown in Supplementary video 1 and video 2 respectively.

Supplementary material related to this article can be found online at doi:10.1016/j.pacs.2024.100632.

Additionally, we discuss the impact of these two hyperparameters λ and σ on the experimental results during the process of registering image sequences. Fig. 8(a) depicts the pre-registration photoacoustic image of the mouse ear at a certain moment. Fig. 8(b) and (c) respectively depict the trends of SSIM with the sequence interval when setting λ and σ to different values. Fig. 8(d) and (j) represent the enlarged views of the green box regions in Fig. 8(a). Fig. 8(e) - (i) represent the registration results with different values of parameter λ. Fig. 8(k) - (o) represent the

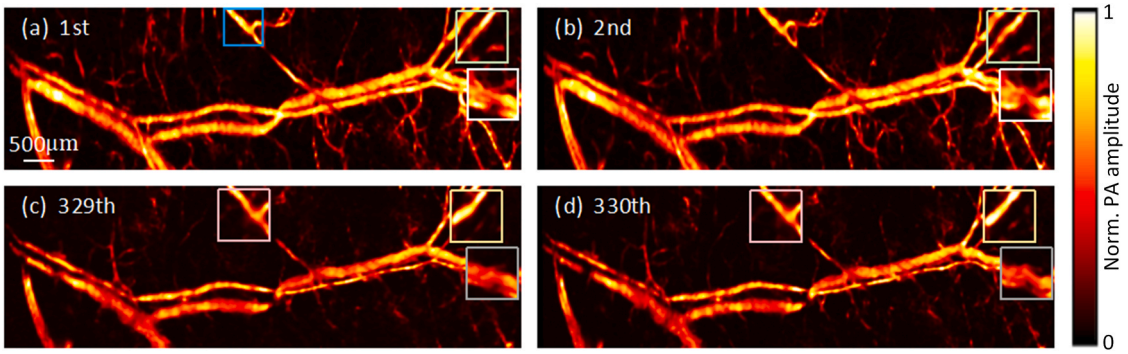


Fig. 4. Two pairs of unregistered MAP images were collected at different times. The area selected by the blue box in (a) is shown enlarged in Fig. 6. Other different colored box pairs highlight and magnify the distortion that occurs in even frames compared to adjacent odd frames.

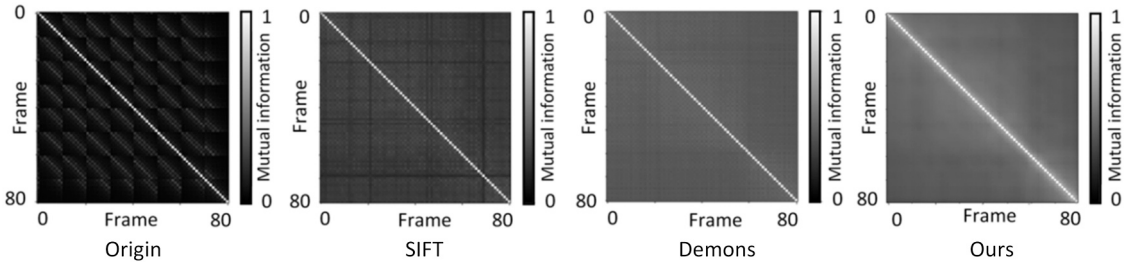


Fig. 5. Grayscale images representing the MIM of different methods.

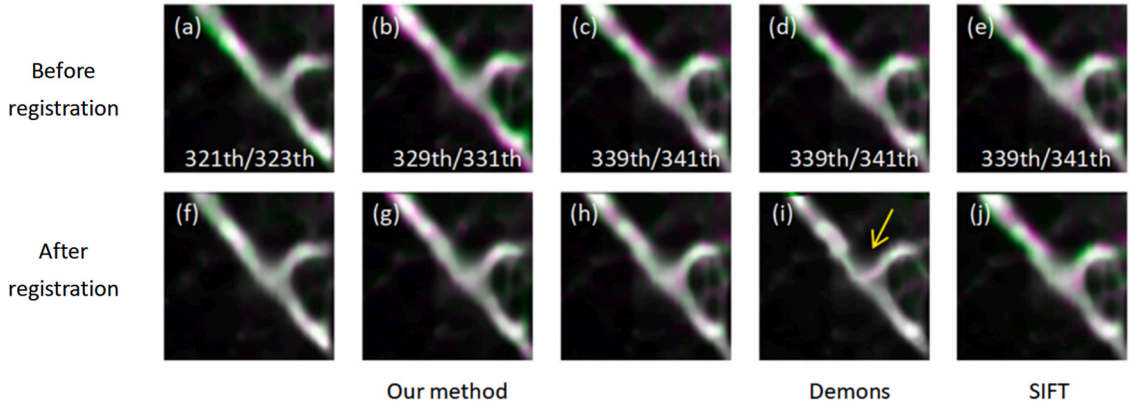


Fig. 6. (a) - (e) The enlarged images of adjacent odd frames overlapped at different times within the blue box area of Fig. 4(a). (f) - (h) Corresponding images of (a) - (c) after registration using our method. (i) Corresponding image of (d) after registration using Demons algorithm. The yellow arrow points out where vascular information is missing. (j) Corresponding image of (e) after registration using SIFT algorithm.

Table 3
Quantitative evaluation results of the similarity before and after the registration of MAP images with different algorithms.

	PSNR	NCC	Runtime (CPU/GPU)
Origin	16.0671	0.9392	-
SIFT	24.6815	0.9465	4 s/-
Demons	31.4853	0.9882	212 s/-
Ours	40.4501	0.9989	4 s/2 s

registration results with different values of parameter σ. During the registration process of odd-numbered sequences, we set parameter λ to be 0.001, 0.01, 0.1, 1, and 10, while keeping other parameters and conditions unchanged. The results in Fig. 8(b) indicate that as λ decreases, the SSIM between different sequence intervals gradually increases, suggesting that the similarity between the images increases as λ decreases. However, smaller values of λ are not necessarily better. Compared to the pre-registration image (d), Fig. 8(e) - (h) show varying degrees of missing vascular information indicated by the arrows. The reason for this phenomenon is that a smaller bending energy penalty coefficient implies a lower requirement for the smoothness of the deformation field, allowing the network to perform larger deformations during registration to better match the details and structures of the image. However, excessive or irregular deformations can also damage image details and textures, leading to a loss of image information. Therefore, considering both the registration effectiveness and accurate preservation of image information, we choose to set λ to 10. Similarly, we set parameter σ to be 0.1, 0.3, 0.6, 0.9, and 1.2. Fig. 8(c) shows that the SSIM is significantly higher when σ is set to 0.1 and 0.3 compared to when σ is set to 0.6, 0.9, and 1.2. However, Fig. 8(k) and (l) also indicate that setting σ to 0.1 and 0.3 results in the loss of vascular information. Moreover, setting σ too large will increase the degree of smoothing,

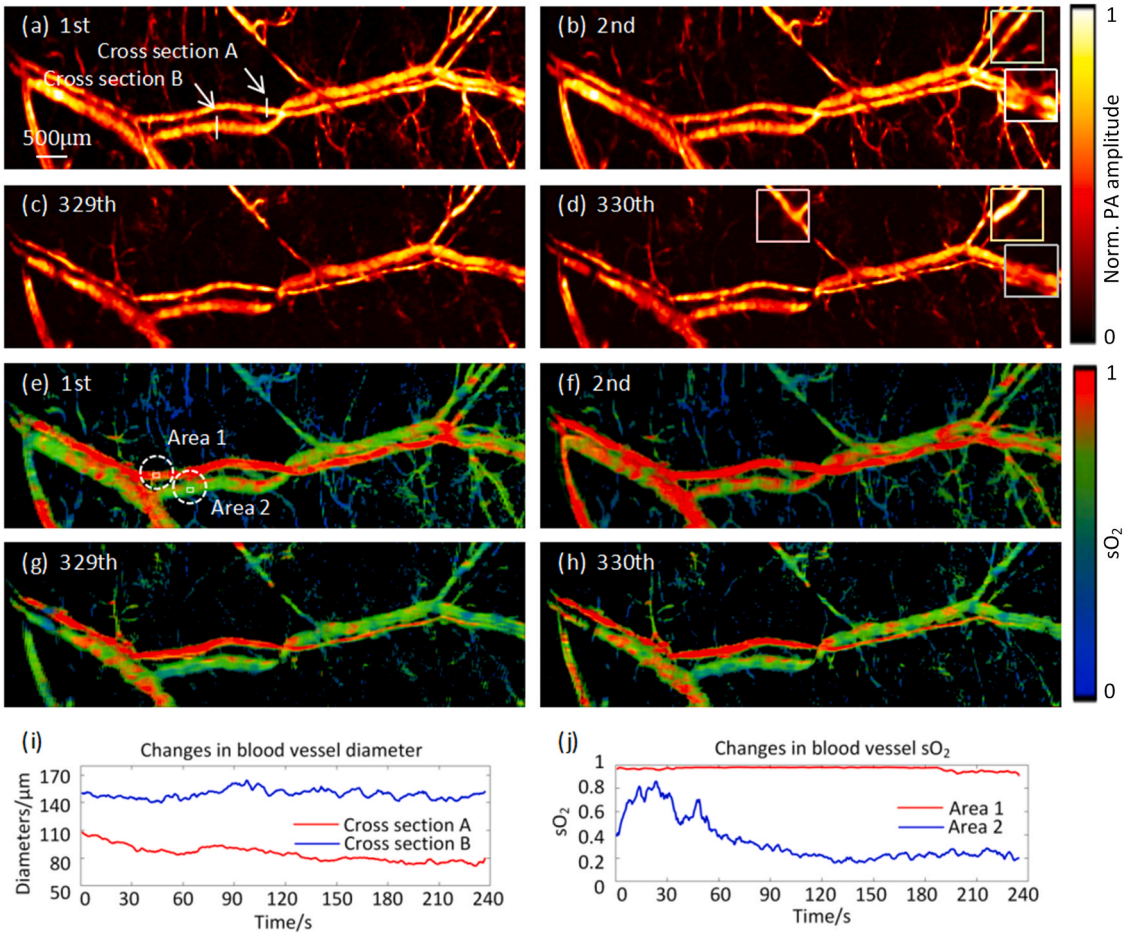


Fig. 7. (a)-(d) The aligned images of adjacent frames using our method. (e) - (h) The corresponding sO_2 images after registration. (i) The diameter changes in the specified cross sections of blood vessels. (j) The sO_2 changes in the specified areas of blood vessels.

Table 4
Quantitative evaluation results of the similarity. before and after the registration of sO_2 images with different algorithms.

	PSNR	NCC	SSIM
Origin	14.1879	0.8542	0.5212
SIFT	21.3737	0.9012	0.6839
Demons	22.8185	0.9551	0.8165
Ours	23.9173	0.9625	0.8317

which can adversely affect the registration effectiveness. To set the optimal σ with a more quantified reference, we incorporated a quantitative analysis based on IS and H. Specifically, we used IS to assess the degree of vascular information retention and H to evaluate the smoothness of the image. Since significant vascular information loss occurred at σ values of 0.1 and 0.3, we only conducted a quantitative analysis for σ values of 0.6, 0.9, and 1.2. The analysis results are demonstrated in Table 5 with maximum values highlighted in bold. It can be noted that when σ is equal to 0.6, the SSIM and H values are both the highest. Therefore, we believe that setting σ to 0.6 is preferable.

5. Discussion

We report an unsupervised deep learning based registration network for correcting image distortions caused by scanning distortion inherent to the scanner and dynamic perturbation from high-frequency scans of the OR-PAM system. By using the proposed method in this paper, both pseudo-artifacts between odd and even columns within a single image and misalignment among multiple images can be addressed. This method not only reduces expensive hardware costs [38] but also improves throughput. Compared to traditional registration algorithms, the proposed method achieves superior registration performance in a shorter time. In contrast to previously reported deep learning based correction methods, our method does not necessitate ground-truth input, which is challenging to acquire in large quantities. We incorporate mutual information into the loss function, adapt the network structure according to different needs, and design unique two-stage training strategies to better suit the application scenarios and requirements of images from OR-PAM. Experimental results show that the proposed method can effectively eliminate image artifacts and misalignment, facilitating observation and comparison of changes in the morphology and function of the vascular, as well as subsequent quantitative analysis. However, our method also has limitations. Its effectiveness is influenced by the size and quality of the dataset, as poor training results may occur with small or low-quality datasets. Therefore, it is necessary to continue to expand our dataset in the future. By incorporating labeled data, explore semi-supervised learning and contrastive learning to optimize label utilization efficiency and enhance the stability and validation of models. We will also investigate domain-adaptive methods, aiming to enable models to migrate efficiently between different domains and thus better adapt to different data distributions.

6. Conclusion

To tackle the challenges posed by the misalignment of inter-image

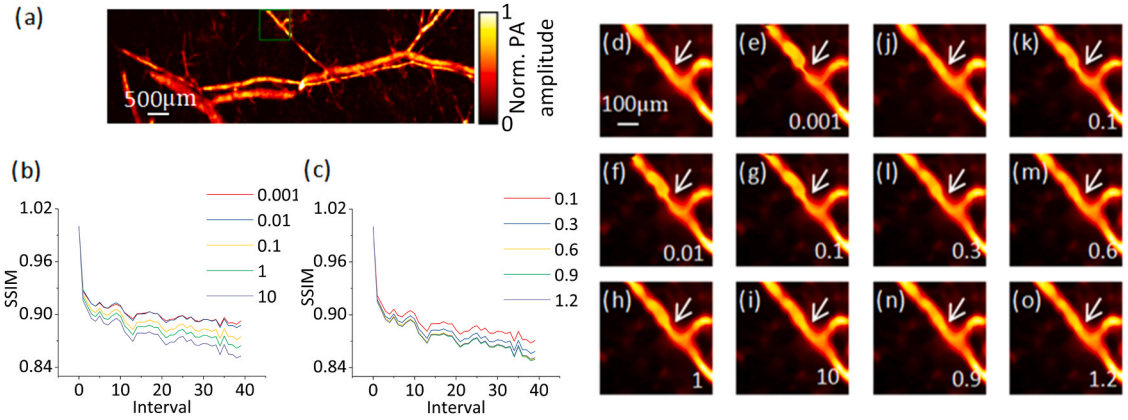


Fig. 8. (a) The MAP images before registration. (b) The SSIM between two frames with an interval of k frames for different values of λ. (c) The SSIM between two frames with an interval of k frames for different values of σ. (d) and (j) The enlarged images of the green boxes in (a) before registration. (e) - (i) The enlarged images of the green boxes in (a) after registration for different values of λ. (k) - (o) The enlarged images of the green boxes in (a) after registration for different values of σ.

Table 5
Quantitative analysis results with σ set to different values.

σ	SSIM	IS	H
0.6	**0.8779**	0.9972	**0.9995**
0.9	0.8770	**0.9975**	0.9985
1.2	0.8776	0.9974	0.9972

arising from B-scan trajectory distortion and dynamic perturbation in the fast scanning OR-PAM system, we propose a deep learning based registration network. Operating without the need for ground-truth inputs, our method approximates the distorted image to the undistorted image using mean square deviation and mutual information as similarity metrics. We integrated the proposed algorithm into a resonant mirror based OR-PAM system, enabling the correction of intra-image motion artifacts in the microvascular structure, as well as misalignment among images, thereby achieving stable video frame display. With the registered image frames, we can accurately quantify changes in microvessel diameter and sO_2 at specified locations, verifying that our approach can facilitate the observation of dynamic tissue structural changes and extraction of quantitative functional information. The results show that our method can achieve efficient alignment of 40 image frames per second, demonstrating the potential for real-time imaging processing capabilities in fast-scanning systems. It is expected to be a tool to promote the development of not only OR-PAM but also microscopic imaging systems that adopt fast scanners.

CRediT authorship contribution statement

Xiaobin Hong: Writing – review & editing. **Furong Tang:** Methodology, Writing – original draft, Software. **Lidai Wang:** Writing – review & editing. **Jiangbo Chen:** Conceptualization, Project administration, Writing – review & editing, Resources.

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

The data that has been used is confidential.

Acknowledgments

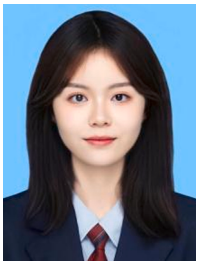
This work is supported by the Guangzhou 2024 basic and applied basic research theme project (SL2023A04J00875), the Guangdong Basic and Applied Basic Research Foundation (2023A1515110065), the Research Grants Council of the Hong Kong Special Administrative Region (11103320, 11101618), and the Natural Science Foundation of China (NSFC) (52375537, 81627805, 61805102).

References

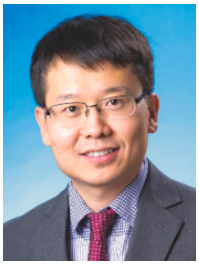
[1] R. Cao, A. Tran, J. Li, et al., Hemodynamic and oxygen-metabolic responses of the awake mouse brain to hypercapnia revealed by multi-parametric photoacoustic microscopy, J. Cereb. Blood Flow. Metab. 41 (10) (2021) 2628–2639, https://doi.org/10.1177/0271678X211010352.
[2] B. Paul, Biomedical photoacoustic imaging, Interface Focus 1 (4) (2011) 602–631, https://doi.org/10.1098/rsfs.2011.0028.
[3] X.Y. Zhu, Q. Huang, A. DiSpirito, et al., Real-time whole-brain imaging of hemodynamics and oxygenation at micro-vessel resolution with ultrafast wide-field photoacoustic microscopy, Light Sci. Appl. 11 (1) (2022) 138, https://doi.org/10.1038/s41377-022-00836-2.
[4] L.V. Wang, L. Gao, Photoacoustic microscopy and computed tomography: from bench to bedside, Annu. Rev. Biomed. Eng. 16 (2014) 155–185, https://doi.org/10.1146/annurev-bioeng-071813-104553.
[5] J. Shi, T.T.W. Wong, Y. He, et al., High-resolution, high-contrast mid-infrared imaging of fresh biological samples with ultraviolet-localized photoacoustic microscopy, Nat. Photon 13 (2019) 609–615, https://doi.org/10.1038/s41566-019-0441-3.
[6] X.Y. Zhu, Q. Huang, L.M. Jiang, et al., Longitudinal intravital imaging of mouse placenta, Sci. Adv. 10 (12) (2024) eadk1278, https://doi.org/10.1126/sciadv.adk1278.
[7] J.N. Zhang, D.L. Peng, W. Qin, et al., Organ-PAM: photoacoustic microscopy of whole-organ multiset vessel systems, Laser Photonics Rev. 17 (7) (2023) 2201031, https://doi.org/10.1002/lpor.202201031.
[8] C. Taboada, J. Delia, M.M. Chen, et al., Glassfrogs conceal blood in their liver to maintain transparency, Science 378 (6626) (2022) 1315–1320, https://doi.org/10.1126/science.abl6620.
[9] S.W. Cho, S.M. Park, B. Park, et al., High-speed photoacoustic microscopy: a review dedicated on light sources, Photoacoustics 24 (2021) 100291, https://doi.org/10.1016/j.pacs.2021.100291.
[10] K.Y. Wang, C.Y. Li, R.M. Chen, J.H. Shi, Recent advances in high-speed photoacoustic microscopy, Photoacoustics 24 (2021) 100294, https://doi.org/10.1016/j.pacs.2021.100294.
[11] J.J. Yao, L.D. Wang, J.M. Yang, et al., Wide-field fast-scanning photoacoustic microscopy based on a water-immersible MEMS scanning mirror, J. Biomed. Opt. 17 (8) (2012) 080505, https://doi.org/10.1117/1.JBO.17.8.080505.
[12] J.B. Chen, Y.C. Zhang, S.N. Bai, et al., Dual-foci fast-scanning photoacoustic microscopy with 3.2-MHz A-line rate, Photoacoustics 23 (2021) 100292, https://doi.org/10.1016/j.pacs.2021.100292.
[13] J.J. Yao, L.D. Wang, J.M. Yang, et al., High-speed label-free functional photoacoustic microscopy of mouse brain in action, Nat. Methods 12 (2015) 407–410, https://doi.org/10.1038/nmeth.3336.
[14] B.X. Lan, W. L, Y.C. Wang, et al., High-speed widefield photoacoustic microscopy of small-animal hemodynamics, Biomed. Opt. Express 9 (10) (2018) 4689–4701, https://doi.org/10.1364/BOE.9.004689.

[15] J.B. Kim, J.Y. Kim, S.W. Jeon, et al., Super-resolution localization photoacoustic microscopy using intrinsic red blood cells as contrast absorbers, Light Sci. Appl. 8 (2019) 103, https://doi.org/10.1038/s41377-019-0220-4.
[16] J.B. Chen, Y.C. Zhang, J.Y. Zhu, et al., Freehand scanning photoacoustic microscopy with simultaneous localization and mapping, Photoacoustics 28 (2022) 100411, https://doi.org/10.1016/j.pacs.2022.100411.
[17] P. Hajireza, W. Shi, R.J. Zemp, Real-time handheld optical-resolution photoacoustic microscopy, Opt. Express 19 (21) (2011) 20097–20102, https://doi.org/10.1364/OE.19.020097.
[18] Y. Zhou, W.X. Xing, K.I. Maslov, L.A. Cornelius, L.V. Wang, Handheld photoacoustic microscopy to detect melanoma depth in vivo, Opt. Lett. 39 (16) (2014) 4731–4734, https://doi.org/10.1364/OL.39.004731.
[19] W.Y. Zhang, H.G. Ma, Z.W. Cheng, Z.Y. Wang, L. Zhang, S.H. Yang, Miniaturized photoacoustic probe for in vivo imaging of subcutaneous microvessels within human skin, Quant. Imaging Med. Surg. 9 (5) (2019) 807–814, https://doi.org/10.21037/qims.2019.05.07.
[20] Q.R. Yu, Y.X. Liao, K.C. Liu, et al., Registration of photoacoustic tomography vascular images: comparison and analysis of automatic registration approaches, Front. Phys. 10 (2022), https://doi.org/10.3389/fphy.2022.1045192.
[21] H. Bay, A. Ess, T. Tuytelaars, et al., Speeded-up robust features (SURF, Comput. Vis. Image Underst. 110 (3) (2008) 346–359, https://doi.org/10.1016/j.cviu.2007.09.014.
[22] Z.H. Li, Z. Dong, A.X. Yu, et al., A robust image sequence registration algorithm for videosar combining surf with inter-frame processing, IGARSS (2019), https://doi.org/10.1109/IGARSS.2019.8899848.
[23] M. Schwarz, N. Garzorz-Stark, K. Eyerich, Motion correction in optoacoustic mesoscopy, Sci. Rep. 7 (2017) 10386, https://doi.org/10.1038/s41598-017-11277-y.
[24] Ningbo Chen Huangxuan Zhao, Tan Li, et al., Motion correction in optical resolution photoacoustic microscopy, IEEE Trans. Med. Imaging 24 (2019) 2139–2150, https://doi.org/10.1109/TMI.2019.2893021.
[25] C.C. Yang, H.R. Lan, F. Gao, et al., Review of deep learning for photoacoustic imaging, Photoacoustics 21 (2021) 100215, https://doi.org/10.1016/j.pacs.2020.100215.
[26] 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, https://doi.org/10.1109/TMI.2018.2829662.
[27] T. Vu, M. Li, H. Humayun, et al., A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer, Exp. Biol. Med. 245 (7) (2020) 597–605, https://doi.org/10.1177/1535370220914285.
[28] X.T. Zhang, F. Ma, Y.K. Zhang, et al., Sparse-sampling photoacoustic computed tomography: deep learning vs. compressed sensing, Biomed. Signal Proces. 71 (2022) 103233, https://doi.org/10.1016/j.bspc.2021.103233.
[29] A. Reiter, M.A.L. Bell, A machine learning approach to identifying point source locations in photoacoustic data (J), Proc. SPIE (2017) 100643, https://doi.org/10.1117/12.2255098.
[30] X.X. Chen, W.Z. Qi, L. Xi, Deep-learning-based motion-correction algorithm in optical resolution photoacoustic microscopy, Vis. Comput. Ind. Biomed. Art. 2 (2019) 12, https://doi.org/10.1186/s42492-019-0022-9.
[31] Z. Sun, J.J. Du, Y. Yao, et al., A deep learning method for motion artifact correction in intravascular photoacoustic image sequence, IEEE Trans. Med. Imaging 42 (2022) 66–78, https://doi.org/10.1109/TMI.2022.3202910.
[32] G. Balakrishnan, A. Zhao, M.R. Sabuncu, et al., VoxelMorph: a learning framework for deformable medical image registration, IEEE Trans. Med. Imaging 38 (2019) 1788–1800, https://doi.org/10.1109/TMI.2019.2897538.
[33] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, MICCAI, Lect. Notes Comput. Sci. 9351 (2015) 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
[34] B.D. de Vos, F.F. Berendsen, et al., A deep learning framework for unsupervised affine and deformable image registration, Med Image Anal. 52 (2019) 128–143, https://doi.org/10.1016/j.media.2018.11.010.
[35] B.D. de Vos, Bas H.M. van der Velden, J. Sander, et al., Mutual information for unsupervised deep learning image registration, Med. Imaging 2020 11313 (2020), https://doi.org/10.1117/12.2549729.
[36] H.O. Velesaca, G. Bastidas, M. Rouhani, et al., Multimodal image registration techniques: a comprehensive survey, Multimed. Tools Appl. (2024), https://doi.org/10.1007/s11042-023-17991-2.
[37] J.P. Thirion, Image matching as diffusion process: an analogy with maxwell's demons, Med. Image Anal. 2 (3) (1998) 243–260, https://doi.org/10.1016/S1361-8415(98)80022-4.
[38] W.H. Shu, M. Ai, T. Salcudean, et al., Image registration for limited-view photoacoustic imaging using two linear array transducers, Proc. SPIE 9323 Photons Ultrasound. Imaging Sens. (2015) 932348, https://doi.org/10.1117/12.2077740.

Xiaobin Hong received a Ph.D. degree in mechanical engineering from the South China University of Technology. He is currently a Professor at the School of Mechanical and Automotive Engineering, South China University of Technology. His research interests include instrumentation design, signal processing, and machine learning methods for acoustic testing



Furong Tang is a Master's student at the South China University of Technology. She received her Bachelor's degree from Southwest University of Science and Technology and was jointly trained at the University of Science and Technology of China during her undergraduate studies. Her research focuses on photoacoustic imaging and image processing.



Prof. Lidai Wang received the Bachelor and Master's degrees from the Tsinghua University, Beijing, and received the Ph.D. degree from the University of Toronto, Canada. After working as a postdoctoral research fellow in the Prof Lihong Wang's group, he joined the City University of Hong Kong in 2015. His research focuses on biophotonics, biomedical imaging, wavefront engineering, instrumentation and their biomedical applications.

Jiangbo Chen is an Associate professor at the South China University of Technology. He received his Ph.D. degree from the City University of Hong Kong and then worked as a postdoctoral fellow at the Polytechnic University of Hong Kong. He received a Bachelor's degree from Northeast Forestry University and a Master's degree from the Harbin Institute of Technology. His research focuses on biophotonics and biomedical imaging.