Photoacoustics 32 (2023) 100536

Contents lists available at ScienceDirect

Photoacoustics

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

Learning spatially variant degradation for unsupervised blind photoacoustic tomography image restoration

Kaiyi Tang [a,b,c], Shuangyang Zhang [a,b,c], Yang Wang [a,b,c], Xiaoming Zhang [a,b,c], Zhenyang Liu [a,b,c], Zhichao Liang [a,b,c], Huafeng Wang [d], Lingjian Chen [d], Wufan Chen [a,b,c,*], Li Qi [a,b,c,*]

[a] School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong, China
[b] Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, Guangdong, China
[c] Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou, Guangdong, China
[d] Research Center of Narrative Medicine, Shunde Hospital, Southern Medical University, Foshan, Guangdong, China

ARTICLE INFO

Keywords:
Deep priors
Image restoration
Photoacoustic tomography
Spatially variant degradation
Unsupervised learning

ABSTRACT

Photoacoustic tomography (PAT) images contain inherent distortions due to the imaging system and heterogeneous tissue properties. Improving image quality requires the removal of these system distortions. While model-based approaches and data-driven techniques have been proposed for PAT image restoration, achieving accurate and robust image recovery remains challenging. Recently, deep-learning-based image deconvolution approaches have shown promise for image recovery. However, PAT imaging presents unique challenges, including spatially varying resolution and the absence of ground truth data. Consequently, there is a pressing need for a novel learning strategy specifically tailored for PAT imaging. Herein, we propose a configurable network model named Deep hybrid Image-PSF Prior (DIPP) that builds upon the physical image degradation model of PAT. DIPP is an unsupervised and deeply learned network model that aims to extract the ideal PAT image from complex system degradation. Our DIPP framework captures the degraded information solely from the acquired PAT image, without relying on ground truth or labeled data for network training. Additionally, we can incorporate the experimentally measured Point Spread Functions (PSFs) of the specific PAT system as a reference to further enhance performance. To evaluate the algorithm's effectiveness in addressing multiple degradations in PAT, we conduct extensive experiments using simulation images, publicly available datasets, phantom images, and in vivo small animal imaging data. Comparative analyses with classical analytical methods and state-of-the-art deep learning models demonstrate that our DIPP approach achieves significantly improved restoration results in terms of image details and contrast.

1. Introduction

Photoacoustic tomography (PAT) is a hybrid biomedical optical imaging technique that has been used in clinical and pre-clinical researches [1–4]. It utilizes wide field optical excitation to exploit the high optical contrast of different tissues, and uses broadband ultrasonic detection to enhance imaging depth. The PAT imaging process is as shown in Fig. 1. A pulsed laser exerts dense energy onto the object and excites acoustic waves. The transducer array intercepts the excited waves and converts them into electrical signal. Then, the acquired raw signal is reconstructed into the PAT image using computational algorithms such as delay-and-sum [5].

The image quality of PAT is subject to multiple degradations introduced in the above imaging process. The sources of degradation include: 1) spatial impulse response (SIR) of the finite size transducer element [6]; 2) limited and sparse spatial sampling of the transducer array [7,8]; 3) electrical impulse response (EIR) produced by the electronics [9]; and 4) errors induced by image reconstruction algorithm that uses an overly simplified imaging model and mismatched speed of sound (SOS). These accumulated system degradation results in low quality PAT images, leading to the loss of contrast and resolution.

Generally, the above forward imaging process can be described as a linear model using the point spread function (PSF) [10],

* Corresponding authors at: School of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong, China
E-mail addresses: chenwf@smu.edu.cn (W. Chen), qili@smu.edu.cn (L. Qi).

https://doi.org/10.1016/j.pacs.2023.100536
Received 5 May 2023; Received in revised form 18 July 2023; Accepted 19 July 2023
Available online 20 July 2023

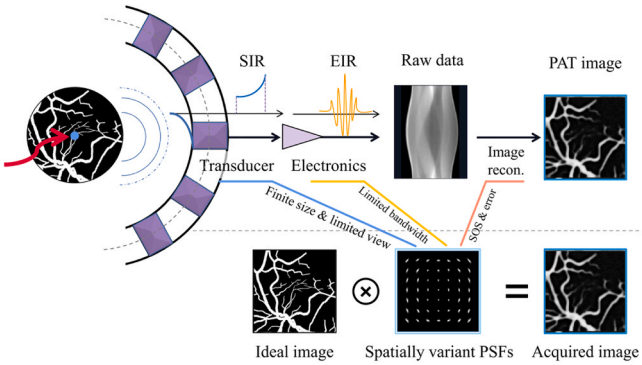


Fig. 1. PAT image formation process and system degradation. SIR: spatial impulse response, EIR: electrical impulse response, SOS: speed of sound. Image degradation can be represented by the spatially variant PSFs.

$$y(s) = PSF(s) * x(s) + n, \tag{1}$$

where s is the spatial coordinate in the image, $*$is the convolution operator in two-dimensional plane, y and x are the degraded image and the clean image, and n is the noise term. For simplicity, the PSF of an image is usually considered spatially uniformed. However, due to the aforementioned degradation sources, PAT image degradation should be systematically characterized by spatially variant PSFs $PSF(s)$ of the system, which have different shapes in different spatial locations [11,12].

In order to deal with the system degradation, restoration methods for both the signal and image domains have been proposed. Corrections in the signal domain account for the impulse response of the ultrasonic transducer [6,9,13–15]. However, these methods did not consider the detection geometry and algorithm error. Also, given a practical and highly integrated PAT system, the EIR and SIR of individual detection elements are not typically provided.

Image domain methods try to recover the clean image from the degraded one. For this purpose, many traditional image deconvolution techniques have been used, for example, Chaigne et al. [16] show the analysis of second order fluctuations of the photoacoustic images combined with image deconvolution enables resolving optically absorbing structures beyond the acoustic diffraction limit. In addition, the knowledge of the PSF of the system can be used during deconvolution. For example, Qi et al. [17] design a rigorous PSF measurement procedure to acquire a dense set of the spatially variant PSFs, and then restore sharp images based on a regularized optimization model (Fig. 2 (a)). Other than these methods, blind deconvolution approaches have also been proposed [18–20]. These traditional deconvolution methods are mostly model-based methods. They require accurate description of the imaging system or handcrafted prior to characterize the complex degradation in real PAT images.

With the rapid development of the deep learning (DL) technology, various methods have been proposed to solve the image restoration problem based on deep neural networks (DNN) [21–24]. Early DL-based deconvolution methods involved supervised learning techniques. For example, Cai et al. [25] propose an end-to-end DNN, ResU-net, for quantitative photoacoustic imaging. Hauptmann et al. [26] design a DNN to provide high-resolution images from restricted photoacoustic measurements. These end-to-end supervised DL methods may be able to correct for spatially variant degradation. However, they rely on training datasets with paired degraded-clean images, whereas in PAT imaging, these clean ground truth images are usually not available. One way to solve this is to use synthetic data to train a neural network, and then apply the trained model to a real PAT image (Fig. 2(b)). However, a model trained from simulated images cannot fully represent the real degradation and thus its performance is usually limited.

Recently, unsupervised DL methods emerge as a promising solution to the image restoration problem where there is no ground truth. For example, Lu et al. [27] introduce an image domain transformation method based on trained cyclic generative adversarial network to remove artifacts in PAT images. Li et al. [28] develop a dual-path quantitative photoacoustic tomography network with unsupervised data translation from simulation to experiment. Among these methods, the Deep Image Prior (DIP) [29] is a plug-and-play image processing framework that requires absolutely no image pairs for pre-training or labeled images as the target, either simulated or real data. The idea of DIP is to access a single input image and impose on the restored output the statistical prior information learned from the distorted image (Fig. 2 (c)) by using DNN. Using this property, DIP can perform various image processing tasks such as image super-resolution and deblurring [30]. DIP has been introduced to photoacoustic imaging by Tri et al. [31], in which they use DIP to improve image resolution of sparsely sampled photoacoustic microscopy.

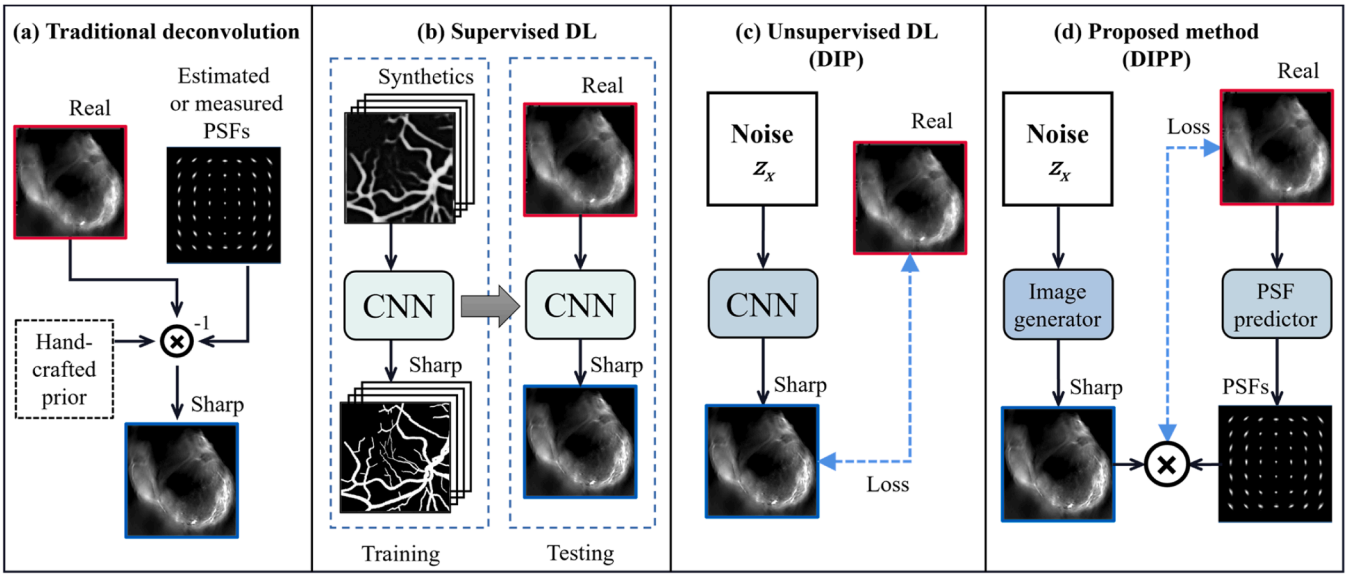


Fig. 2. Traditional and deep-learning-based methods for PAT image restoration. The red boxed image represents the degraded PAT image, and the blue boxed image represents the restored image. DIP: deep image prior, CNN: convolutional neural network.

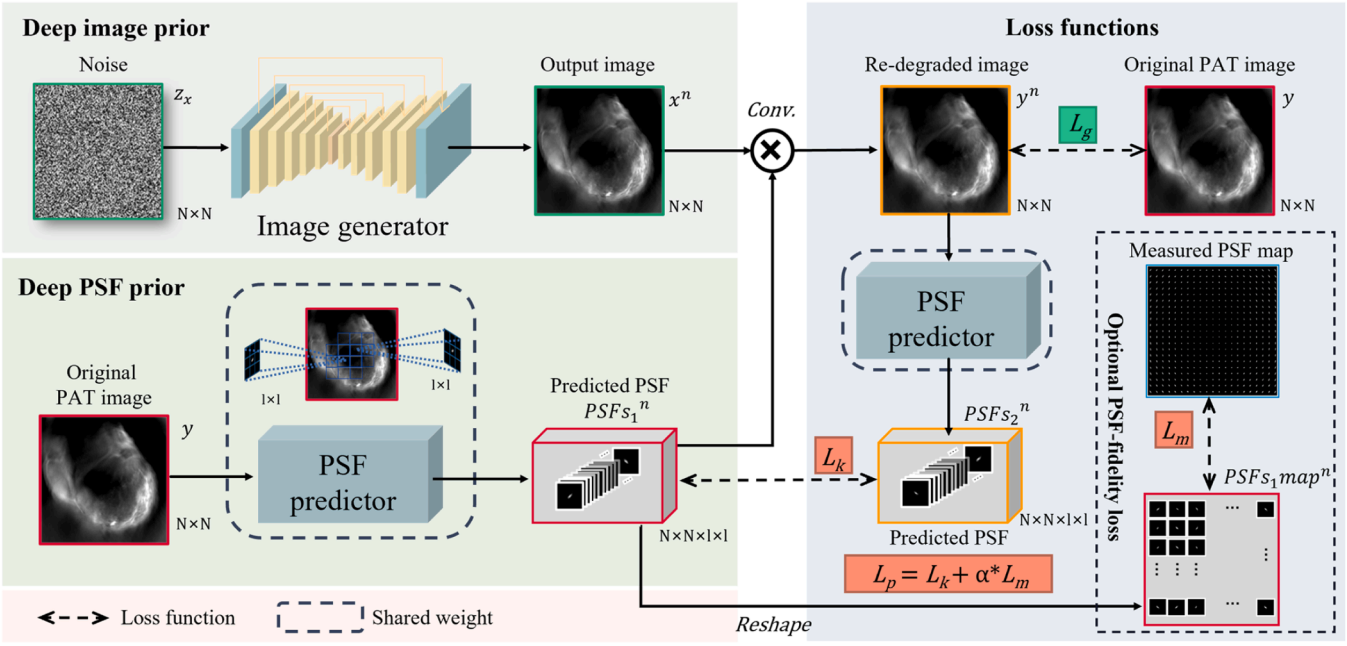


Fig. 3. Illustration of the DIPP framework for PAT image restoration. Deep image prior: an image generator takes noise z_x as input and output the recovery results x^n. Deep PSF prior: a PSF predictor takes the original degraded PAT image y as input, estimates the corresponding PSF for each image patch, and outputs pixel-wise PSF kernel $PSFs_1{}^n$. Loss function: Re-degraded image y^n is synthesized using x^n and $PSFs_1{}^n$. y^n is fed into the PSF predictor again and the output is $PSFs_2{}^n$. $PSFs_{1map}{}^n$ is obtained by reshaping $PSFs_1{}^n$. The loss function of the image generator L_g consists of y^n and $\mathbf{y}$. The loss function of the PSF predictor L_p consists of two parts: L_k and the optional PSF-fidelity term L_m. L_k is composed of $PSFs_1{}^n$ and $PSFs_2{}^n$, and L_m is composed of $PSFs_{1map}{}^n$ and the measured PSF map.

Given its unsupervised learning advantage, the DIP technique may be a suitable solution to our PAT image restoration problem. However, the DIP method is shown to handle spatially variant degradation poorly [30]. In addition, its performance on a single image fluctuates with different network initialization and the input random noise [32]. Without proper regularization, it cannot achieve stable and accurate image restoration performance for complex degradation. Based on the above considerations, herein we propose a new image restoration framework, Deep hybrid Image-PSF Priors, or DIPP, to account for the spatial-variant degradation in PAT. Embedding into a MAP-based deconvolution framework, our DIPP consists of an image generator network as the deep image prior to generate sharp image iteratively, and a PSF predictor network that served as the deep PSF prior to estimate degradation from the input PAT image (Fig. 2(d)). These two components are optimized by re-degrading the recovered result of the image generator in each iteration to match the original image. In addition, the experimentally measured PSFs of the PAT system can be integrated as an optional reference to guide PSF estimation. Extensive experiments are performed to prove our contributions on simulated images, publicly available datasets, phantom images as well as in vivo small animal PAT imaging data. Compared to traditional and state-of-the-art deconvolution methods, our DIPP shows superior performance in improving PAT image quality.

2. Methods

Our DIPP framework for PAT image restoration is illustrated in Fig. 3. Our model can be broadly divided into three parts. Specifically, the first part is a deep image prior for recovering the sharp image; the second part is a deep PSF prior for predicting individual PSFs from overlapping degraded image patches; and the third part is loss computation by using image re-degrading. In the following, we first introduce the formulation of the spatially varying deconvolution, and then propose our DIPP model for hybrid neural regularization. We present the restoration algorithm in the end.

2.1. Formulation of spatially variant deconvolution

Based on the physical degradation model in (1), we adopt the Maximum a Posterior (MAP) framework to obtain the optimal solution for deconvolution:

$$(x, PSFs) = arg\max_{(x,\ PSFs)} P(x)P(x|y), = arg\max_{(x,\ PSFs)} P(y|x)P(x)P(PSFs),$$
$$= arg\min_{(x,\ PSFs)} \|PSF(s) * x(s) - y(s)\|^2, \quad (2)$$

where $PSFs$ is composed of the spatially variant PSFs corresponding to each pixel s.

To obtain accurate image recovery, a regularization term $\phi(x)$ for the clean image x can be added, such that

$$(x, PSFs) = arg\min_{(x,\ PSFs)} \|PSF(s) * x(s) - y(s)\|^2 + \lambda \phi(x). \quad (3)$$

To account for spatially variant degradation, we can further incorporate the prior knowledge of the PSFs into the deconvolution problem by adding another regularization term $\phi(PSFs)$. Then (3) becomes:

$$(x, PSFs) = arg\min_{(x,\ PSFs)} \|PSF(s) * x(s) - y(s)\|^2 + \lambda \phi(x) + \tau\phi(PSFs), \quad (4)$$

where λ and $\tau$are trade-off parameters that control the strength of the regularization terms. In the following, we show that the hybrid image-PSF regularization in (4) can be modelled with deep neural networks.

2.2. Hybrid neural regularization modeling

Previously, the DIP framework [29,33] has proved that DNNs were capable of learning high-level texture priori information before fitting low-level features, such as degradation and noise. Given this powerful a priori acquisition capability, we suggest using deep priors G_x, G_p to model $x(s)$ and $PSF(s)$ in (4), respectively. The deep priors also implicitly include the regularization terms $\phi(x)$ and $\phi(PSFs)$. Therefore, we aim to

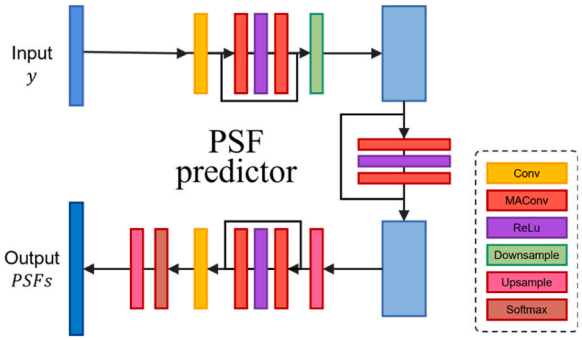


Fig. 4. The structure of the PSF predictor network. Given a degraded image input **y**, the network outputs PSFs estimated from the corre-sponding degraded image patches. The PSF predictor is consisted of repeated application of re-sidual blocks. Each residual module contains two mutual affine convolution layers with a ReLu layer between them. Downsample and upsample layers are implemented by 2 × 2 convolution (stride of 2) and 2 × 2 transpose convolu-tion (stride of 2).

solve:

$$\min_{(G_p,\ G_x)} \left\| G_p(y) * G_x(z_x) - y \right\|^2, \text{s.t.} 0 \le (G_x(z_x))_i \le 1, 0 \le i \le m \bullet n, (G_p(y))_j \ge 0, \sum_{0 \le j \le k^2} (G_p(y))_j = 1, \tag{5}$$

where z_x is a random input image sampled from the uniform distribution, y is the raw PAT image, which has size m × n. The size of a single PSF is k × k. $(\cdot)i$ and $(\cdot)j$ denote the i-th and j-th elements. G_p is the PSF predictor and G_x is the image generator. Thus, $G_p(y)$is the estimated PSF map and $G_x(z_x)$ is the restored image. Compared to DIP, our framework captures the priors of both the clean image and PSFs, and therefore is dubbed Deep hybrid Image-PSF Priors (DIPP).

In the following, we elaborate our neural prior model around the above formulation.

2.2.1. Image generator

We use the original DIP network to serve as the clean image generator G_x. As shown in Fig. 3, the generator takes random noise z_x of the same size as the original image y, and outputs a sharp restored result x^n. G_x is structured as a U-net framework [34] containing five encoder units and five decoder units with skip connections. Sigmoid nonlinearity is applied to the output layer to meet the intensity range constraint.

2.2.2. PSF predictor

Generally, we can assume that a PSF only has impacts on a local image patch of a fixed size. Motivated by [35], we propose using a neural network to serve as the PSF predictor G_p. Generally, the PSF predictor is expected to estimate the pixel-wise PSF from the adjustable receptive field. As a result, each pixel of the input original image corresponds to a specific PSF. The structure of G_p is shown in Fig. 4.

First, y is input to the 3 × 3 convolutional layer to extract image features, and then goes through a residual block. The residual block consists of two mutual affine convolution (MAConv) layers with a ReLU layer between them. After a downsampling layer (stride of 2), it passes through the same residual block as described above, and then feeds to an upsampling layer (stride of 2). Then the residual block is used again and finally the output PSFs are estimated using a 3 × 3 convolutional layer and a softmax layer. Unlike the image generator, which simply uses the normal convolutional layers, we use the MAConv layers [35] in the PSF predictor to reduce model parameter and computation cost for large-scale PSF estimation. The MAConv layer was originally used in autoencoder [34]. It enhances inter-channel dependence by mutual affine transformation instead of simply connecting different channels. It is also able to make connections between feature maps such that richer information is obtained.

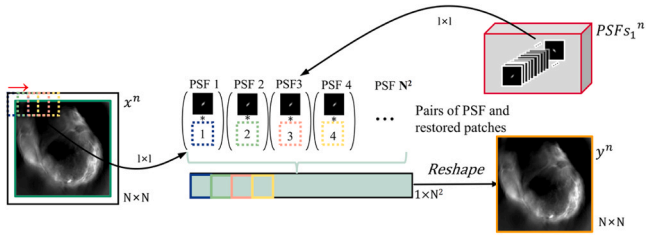


Fig. 5. Patch-wise convolution for image re-degradation.

Our DIPP adaptively estimates the PSFs from a single degraded image by using the PSF predictor. The PSF predictor network can either be obtained by labeled data training and then fine-tuning, as in natural image deblurring [35], or be initialized by random initialization. In our experiments, we found that for PAT images, random initialization performs better than pre-training, the details of which can be found in the Results section.

2.2.3. Image re-degrading and loss function

During each iteration, we generate the re-degraded image y^n using the outputs of the above two sub-networks to construct the loss functions (as shown in Fig. 3). This is achieved by patch-wise convolution, as shown in Fig. 5. Its specific procedure is as follows: first, pad around the restored result x^n and shift one window over it using a stride of 1; then, convolute the image patch with the respective PSF for each slide and fill the corresponding image pixel to obtain the re-degraded image y^n. The weight of each window is determined by the PSF corresponding to the center pixel of the restored image patch.

Once the re-graded image is acquired, y^n is fed into the PSF predictor to output $PSFs_2{}^n$. Then, we adopt the following loss functions,

$$L_i = \|y^n - y\|^2, \tag{6}$$

$$L_k = \left\| G_p(y) - G_p(y^n) \right\|^2 = \|PSFs_1 - PSFs_2\|^2, \tag{7}$$

$$L_m = \left\| G_p(y) - PSFs_m \right\|^2 = \|PSFs_1 - PSFs_m\|^2, \tag{8}$$

where y^n is the degraded image obtained in the n-th iteration, y is the raw PAT image. L_i is the image loss which directly compares the original image and the degraded image, L_k is the kernel loss which compares the individual PSF in order to maintain patch consistency, L_m is the measured PSF loss, such that the measured PSFs $PSFs_m$ can be incorporated into the model. $PSFs_1$ and $PSFs_2$ are pixel-wise PSFs predicted from y and y^n by the PSF predictor G_p, respectively. Then, we let

$$L_g = L_i \tag{9}$$

be the loss function of the image generator, and

$$L_p = L_k + \alpha * L_m \tag{10}$$

be the loss function of the PSF predictor. For the predictor loss L_p, the kernel loss L_k is a fidelity item, and L_m is control by the adjustable a parameter α and thus is optional.

According to this, in the training of the image generator G_x, we aim to reduce the distance between the original input image and the re-degraded image, whereas in the training of the PSF predictor G_p, we aim to reduce 1) the distance between the PSFs predicted from the original image and the re-degraded image and 2) the distance between the measured PSFs and the PSFs predicted from the re-degraded image.

Therefore, in the loss function L_g of the image generator, we simply use the image loss L_i to satisfy the constraint. For the PSF predictor loss L_p, we use y and y^n as inputs to perform $G_p(y)$ and $G_p(y^n)$, respectively.

Then, the kernel loss L_k is imposed to increase the consistency of the two PSF estimations. Noteworthy, in the predictor loss L_p, we choose to use the kernel loss L_k rather than the image loss L_i, because L_k is capable of comparing the differences of individual image patch instead of the whole image.

To improve the accuracy of the PSF predictor, we add an optional regularization term L_m to incorporate the measured PSFs of the specific PAT system. L_m is able to force the estimated PSFs to match the actually measured PSFs. However, as a strong fixed reference, L_m may also have a risk of reducing image-dependent PSF adaptation. In other words, the unknown PSFs should be derived mostly from the observed image, rather than only depending on the measured PSFs. This is controlled by the parameters α in (10).

2.3. Optimization algorithm

Our DIPP is a blind spatially variant deconvolution framework; we aim to search for the latent sharp image only from the acquired raw input image itself. Under the MAP-based deconvolution framework, the image generator and PSF predictor are trained iteratively when the stopping condition is reached, i.e., $T = 1000$. The iteration process is listed in Algorithm 1. In each iteration, we use G_p as the deep PSF prior to output the estimated PSFs k, and G_x as the deep image prior to obtain the recovered image result x. The loss functions are constructed by performing patch-wise convolution, and G_p and G_x are updated using the ADAM algorithm [36]. When the T^{th} iteration is reached, the output x in this iteration is the final restored result.

Algorithm 1 DIPP spatially variant deconvolution

Input: Original degraded image y, optional measured $PSFs$
Output: PSFs map K and a clean PAT image x
1: Sample z_x from uniform distribution with seed 0.
2: **for** t=1 to T **do**
3: $k = G_p(y)$
4: $x = G_x(z_x)$
5: Perform patch-wise convolution and compute the gradients, w.r.t. G_p and G_x
6: Update G_p and G_x using the ADAM algorithm[36]
7: **end for**
8: $k = G_p^T(y)$, $x = G_x^T(z_x)$

2.4. Experimental setups

2.4.1. Cross-sectional PAT imaging system

A commercial multispectral photoacoustic tomography system (MSOT inVision128, iThera Medical, Germany) is used for small animal imaging., as shown in Fig. 6.

The system employs a pulse laser that provides 360°ring illumination to excite the sample. The laser is tunable from 670 nm to 980 nm and the maximum pulse energy is 60 mJ at 760 nm. The tomographic ultrasound detection array is ring-shaped and contains 128 cylindrically focused elements. The detector covers a 270° angle and forms a 40.05 mm radius imaging area. The center frequency and bandwidth of the transducer are 5 MHz and 60%, respectively.

2.4.2. PSF map measurement

The PSF measurement procedure has been previously described in [17]. Briefly, it consists of the following steps: perform z-scanning of a black microsphere placed at the center of the FOV to determine the cross-sectional imaging plane; determine the speed of sound visually when the microsphere shape is smallest in the imaging plane; scan the microsphere within the x-y detection plane to get a sparse PSF map. The microsphere is positioned on a 21 mm × 21 mm Cartesian grid with a grid size of 1 mm, thus the obtained PSF map consists of 441 PSFs in the cross-sectional plane.

2.4.3. Simulation experiment

We use a synthetic image of capillaries to mimic the vascular network system within a tumor (Fig. 7). Spatially variant convolution is performed on the ground truth (GT) image using the measured PSF data to obtain the degraded image. We evaluate the performance of different PSF predictor settings, and perform an image restoration experiment on uneven SOS distribution to observe the performance of DIPP when local PSFs are mismatched.

2.4.4. Publicly available datasets

We use two public datasets to evaluate the performance of our DIPP with respect to state-of-the-art DL methods. The datasets are: 1) Levin dataset [37]. It contains 4 sharp images and 8 convolution kernels with different sizes. In total 32 images are obtained by performing spatially invariant convolution on the sharp images using the 8 kernels. We perform spatially invariant deconvolution experiment on this dataset to illustrate the feasibility of the proposed deep PSF prior. The results are presented in Supplementary Fig. S1. 2) Set 5 [38]. It consists of 5 sharp images (“baby”,“bird”,“butterfly”, “head”,“woman”). We use the spatially variant PSFs measured on our PAT system to obtain the non-uniformly degraded images. We perform spatially variant deconvolution using DIPP and other DL methods. The results can be found in Supplementary Fig. S2.

2.4.5. Phantom experiment

We perform a phantom experiment to observe the resolution improvement of the DIPP method. We use a colored dried leaf as the imaging target. We embed the leaf in a 2.6 cm diameter agarose cylinder (solution by adding 1.2 g agar to 100 ml distilled water) to obtain the phantom. We perform PAT imaging at 680 nm and reconstruct the images using a speed of sound of 1536 m/s.

2.4.6. In vivo animal experiment

Animal experiments have been approved by Southern Medical University and are performed in compliance with institutional guidelines. Two cancerous mice (mouse 1 and mouse 2 in Fig. 11) carrying 4T1 mammary carcinoma implanted on the lower flank and developed for 12 days, and a healthy female mouse (mouse 3 in Fig. 12) at the age of 5 weeks are used in the experiment. PAT imaging is performed as described in [17]. The imaging wavelength is 700 nm and the speed of sound is set to 1536 m/s. In both the phantom and animal experiments, PAT image reconstruction is done using the built-in algorithm of the MSOT inVision128 system.

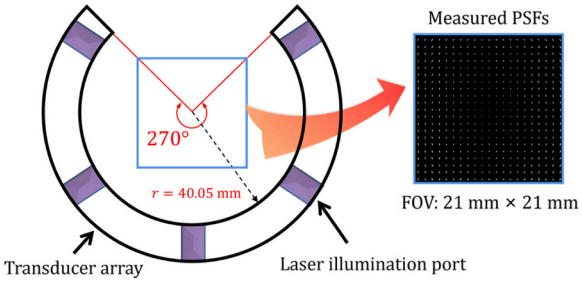


Fig. 6. The illumination and detection setup of our PAT system and the measured PSF map.

2.4.7. Image quality metrics

The metrics used to evaluate the performance of our deconvolution algorithm include the peak signal to noise ratio (PSNR), signal to noise ratio (SNR), and structural similarity (SSIM), which are defined as:

$$PSNR = 10\cdot\log_{10}\left(\frac{P^2}{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{ij}-x_{ij})^2}\right), \tag{11}$$

$$SNR = 10\cdot\log_{10}\left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{ij}-x_{ij})^2}\right), \tag{12}$$

$$SSIM = \frac{(2\mu_I\mu_x + C_1)(2\sigma_{I\cdot x} + C_2)}{(\mu_I^2 + \mu_x^2 + C_1)(\sigma_I^2 + \sigma_x^2 + C_2)}, \tag{13}$$

where, P is the dynamic range of the pixel value ($P = 255$ for 8 bit grayscale images). I and x are the sharp GT image and the restored image, which have size m × n.$(\cdot)_{ij}$ denotes the i-th row and j-th column element. μ_I and μ_x are the mean of I and x, respectively, σ_I^2 and σ_x^2 are the corresponding variance. $\sigma_{I\cdot x}$ is the covariance of I and x. The C_1 and C_2 are given by:

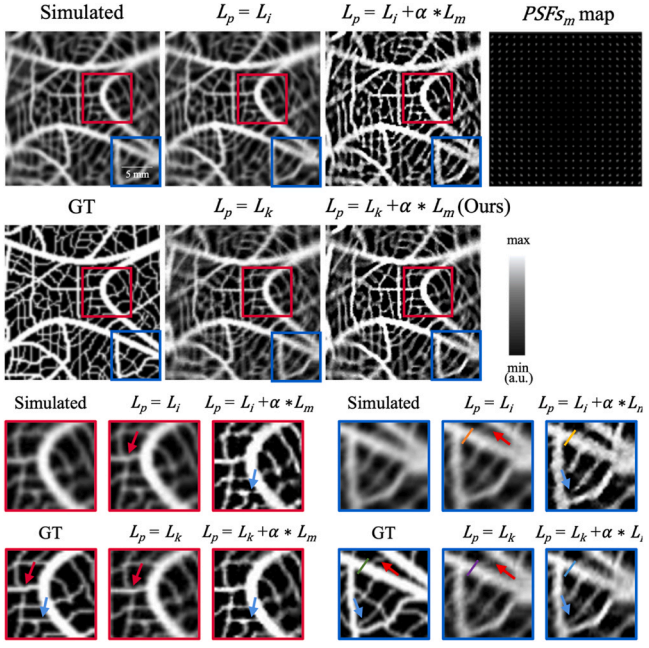


Fig. 7. Comparison of different loss function settings for the PSF predictor in the proposed DIPP framework. "GT" is a clean ideal image, and "Simulated" is the result of the GT image convolved with the measured PSFs map of our PAT system. $PSFs_m$ map: measured PSFs.

$$C_1 = (K_1P)^2, C_2 = (K_2P)^2 \tag{14}$$

where, K_1 and K_2 are set as $K_1 = 0.01$ and $K_2 = 0.03$, respectively.

3. Results

In this section, we show the experimental results. We first compare the effect of different configurations of the PSF predictor and test the performance using mismatched PSFs in the simulation experiments. Then we compare the proposed DIPP with a comprehensive set of other methods on public datasets. Finally, we test our method on phantom and real-world animal PAT images. In all the experiments, we set the maximum iteration to 1000 and α to 0.6. The learning rates of the image generator and the PSF predictor are 0.01 and $1e^{-4}$, respectively.

3.1. Comparison of different PSF predictor settings

3.1.1. Loss function for the PSF predictor

As mentioned previously, both the kernel loss L_k and the image loss L_i can be used in the loss function L_p of the PSF predictor [39]. In fact, in natural image processing, the image loss L_i is more frequently adopted [30,32,39,40]. To evaluate their performance, we conduct simulation experiments by using the following settings of the predictor loss L_p: 1) L_i only, 2) L_k only, 3) $L_i + \alpha * L_m$, and 4) $L_k + \alpha * L_m$. The results are shown in Fig. 7.

As can be seen, when L_m is not used, the result of using L_i only is an average degraded solution, which has uniform blurring over the whole image. In contrast, the result of using L_k only achieves a much better deconvolution effect. The red and blue boxes in Fig. 7 are regions with different degrees of degradation. As shown in their enlarged images in the bottom panel of Fig. 7, the result by using L_k has clearer detail and sharper edges than using L_i only (blue arrows). This indicates that the kernel loss L_k is more suitable than the image loss L_i for the PSF predictor.

Comparing the case with and without the addition of the measured PSF loss L_m in Fig. 7, the inclusion of L_m achieves a significant improvement in image resolution. As can be seen in the enlarged images, the result using $L_i + \alpha * L_m$ has lost part of the microstructure compared to using $L_k + \alpha * L_m$ (red arrows). Fig. 8 shows the image profiles along the solid lines in Fig. 7. It can be seen that the profiles are proximity to GT when adding L_m regardless of which loss is used, and the profile by using L_k and L_m together is closest to the GT profile. Fig. 8 also shows the full-width at half maximum (FWHM) of the image profiles. The "$L_k + \alpha * L_m$" setting obtains the smallest FWHM. Overall, using a combination of L_k and L_m achieves the best performance over its counterparts. As shown in Fig. S3 in Supplementary Material, the difference maps between the GT image and the restored results using different loss function settings further confirm the above finding.

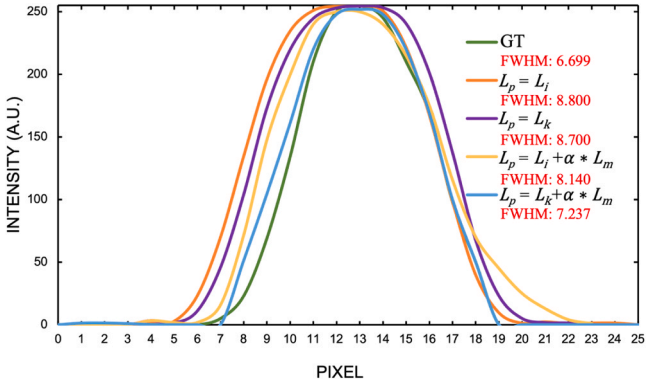


Fig. 8. The image profiles along the solid lines in Fig. 7.

Table 1
PSNR and SSIM of different initialization and loss function settings on the Levin et al. [37].

Method	PSF size	PSNR	SSIM
Pre-training / only L_k	21 × 21	23.98	0.7357
Pre-training / only $L_k + \alpha * L_m$	21 × 21	24.85	0.7575
Random initialization / only L_k	21 × 21	33.18	0.9448
Random initialization / only $L_k + \alpha * L_m$	21 × 21	**34.13**	**0.9533**

3.1.2. Pre-training vs. random initialization

As mentioned in the Methods section, the PSF predictor in DIPP can be either randomly initialized or pre-trained using simulated example images. To evaluate these two strategies, we test randomly initialized and pre-trained PSF predictor networks with and without the loss L_m of the measured PSFs.

For network pre-training, we follow all the settings in [35] except for replacing the training set DIV2K [41] with BSD68 [42], which contains 68 grayscale images. This is because PAT images are grayscale images. Since the pre-trained network can only estimate fixed size degradation, we only use 4 degraded images as the testing set to evaluate the performance. These images are synthesized with one uniform PSF and four clear images from the dataset of Levin et al. [37] such that the dataset used for pre-training is different from the testing data. This is to mimic the situation in PAT imaging where training datasets can only be obtained from simulation.

Table 1 lists the obtained PSNR and SSIM results of different initialization and loss function settings. As can be seen, when only the kernel loss L_k is used, random initialization has 38.4% and 28.4% improvement over pre-training in PSNR and SSIM, respectively. When L_k and L_m are used together, PSNR and SSIM are further improved by 37.3% and 35.6%, respectively, for random initialization. In both cases, the randomly initialized model performs better than the pre-training network. Moreover, the addition of L_m improves the performance for both pre-training and random initialization. The best performance is achieved by the combination of random initialization and using both the L_k and L_m loss functions. Notice that the ground truth PSFs used for the pre-training set and the testing set are of the same size, but with different shapes. Therefore, for cases where the exact size and types of degradation cannot be predicted (as in PAT imaging), it is suggested to use a PSF predictor network with randomly initialized parameters to avoid misleading of optimization direction.

3.2. Experiment on mismatched PSFs

Next, we test the image restoration performance of DIPP when the measured PSFs do not match the real PSFs. As shown in Fig. 9, the measured PSFs $PSFs_a$ (Fig. 9(a)) is obtained under a uniform SOS of 1533 m/s. We create a mismatched PSF map $PSFs_b$ by using 1536 m/s on one fourth of the PSF map (blue box in Fig. 9(b)) and keep the rest to 1533 m/s. We then used $PSFs_b$ to synthesize the unevenly degraded "vessel" image used for deconvolution (Fig. 9(d)).

As can be seen from the deconvolution results in Fig. 9(e) and (f), the traditional deconvolution method [17], which is also based on spatially variant PSFs, fails to recover image detail in both SOS-matched and -mismatched regions. In contrast, our DIPP corrects for the spatially variant degradation caused by uneven SOS distribution. The mean and variance of the absolute error between DIPP and GT are smaller than that of the traditional method [17] in either the whole image or the SOS-mismatched area (red box). Although the degradation in the SOS-mismatched region has not been fully restored, this experiment shows that the PSF predictor in DIPP is able to avoid certain errors in heterogeneous SOS distribution.

3.3. Phantom image restoration results

In the following real-world PAT imaging experiments, we use the aforementioned measured PSFs to guide image restoration. We compare our DIPP with the SV-Gauss. and SV-Spar. methods, which are traditional model-based spatially variant deconvolution algorithm proposed by our group previously [17]. SV-Gauss. and SV-Spar. use measured PSFs [17] to guide SV image deconvolution with Gaussian and sparsity priors as regularization, respectively. We also perform image deconvolution using the DIP method [29], in which only a single U-Net is served as an image generator. For fair comparison, we modify the original DIP method to incorporate the measured PSFs. This is done by using $||PSFs_m * x^n - y||^2$ in [29].

Fig. 10 shows the experimental results of the leaf phantom. As can be seen from the enlarged red boxed areas, DIPP successfully enhances the contrast of the vein structure (blue arrows), whereas serious artifacts (red arrow) are presented by DIP. The image profiles along the white solid line (Fig. 10(c)) confirm that our DIPP achieves the highest signal contrast.

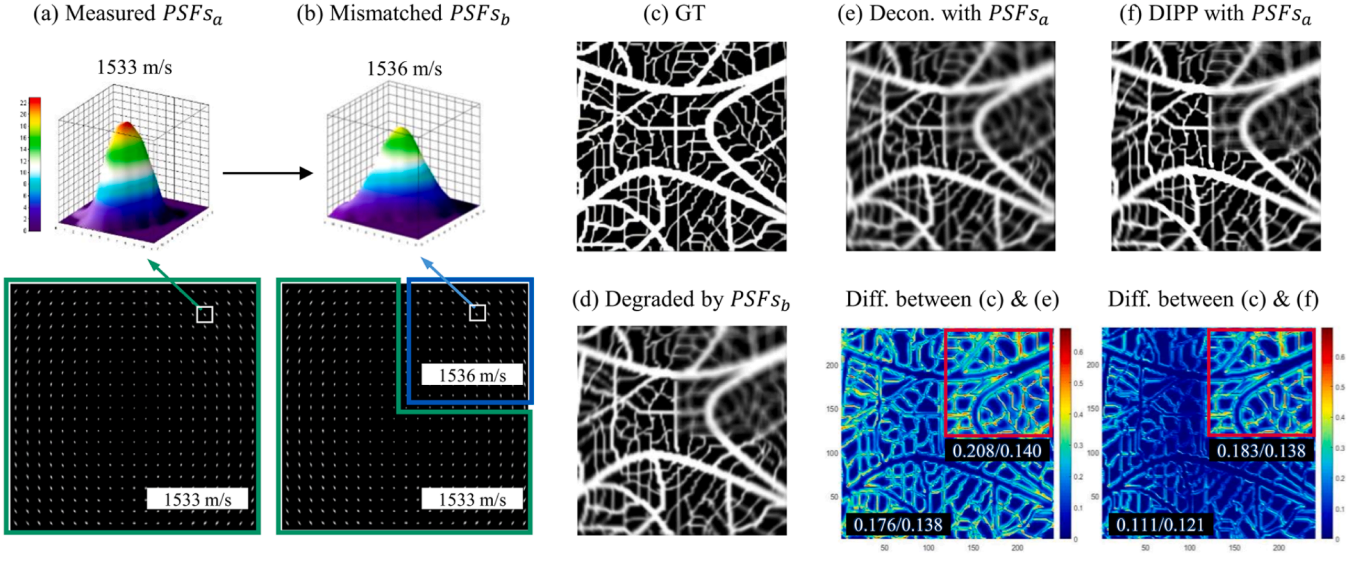


Fig. 9. Simulation experiment with mismatched SOS. (a) $PSFs_a$: measured PSFs reconstructed with SOS= 1533 m/s. (b) $PSFs_b$: mismatched PSFs where the upper right region is reconstructed using SOS= 1536 m/s. (c) The ground truth (GT) "vessel" image. (d) The blur "vessel" image degraded by using $PSFs_b$. It simulates the spatially variant degradation due to the mismatched SOS distribution. (e) SV-Gauss. deconvolution [17] of (d) by using $PSFs_a$. (f) DIPP deconvolution of (d) using $PSFs_a$. Below (e) and (f) are their difference map with the GT image (c). The numbers indicate mean/variance within either the red box or the global difference map.

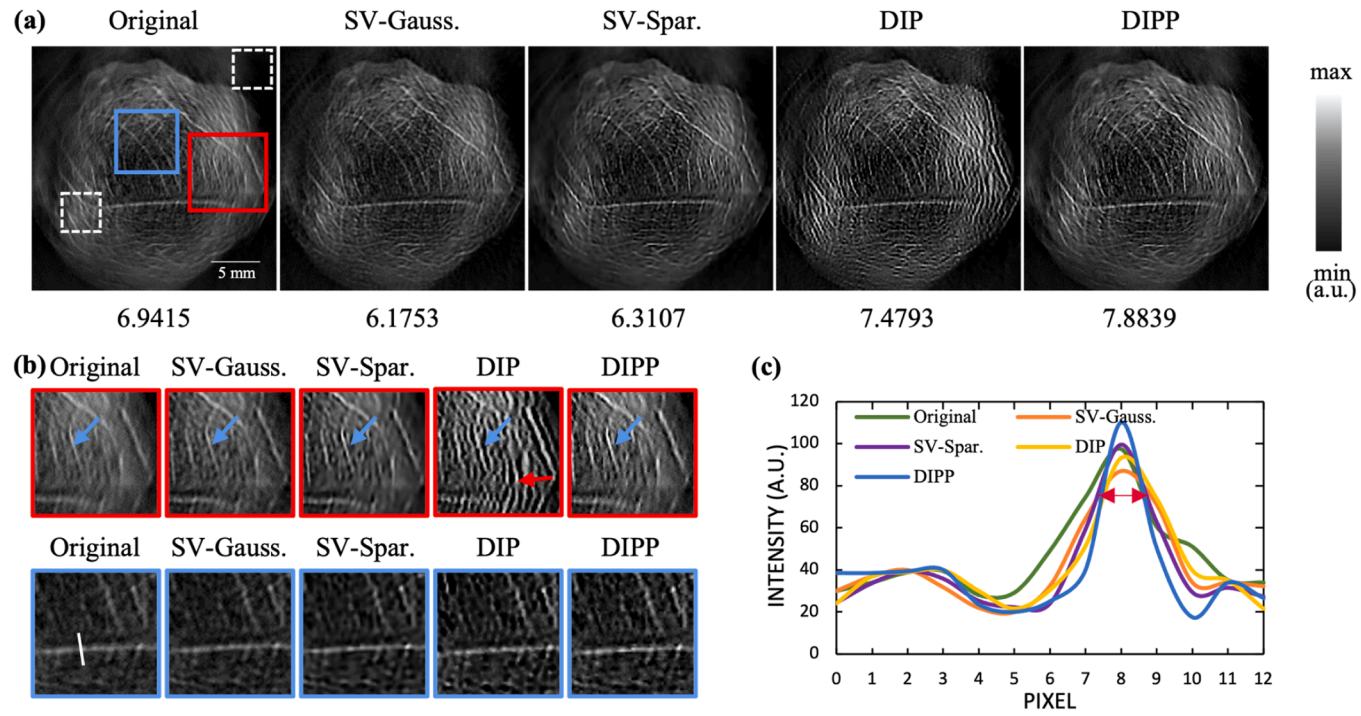


Fig. 10. PAT image restoration result of the tree leaf phantom. (a) The original PAT image and its deconvolution results with traditional methods (SV-Gauss. and SV-Spar.), DIP method [29] and ours DIPP method. The number below each image is the SNR calculated within the white boxes. (b) Enlarged areas corresponding to the red and blue box in (a). (c) Image profiles drown along the white solid line in (b).

3.4. Real-world animal PAT image restoration results

We further verify the performance of our method on real PAT images of cancerous and healthy mice. Fig. 11 shows the representative restoration results of the PAT images of the cancerous mice. The mouse 1 and mouse 2 images are taken from the tumor region where neovascular is presented. As can be seen, compared to traditional model-based deconvolution methods, DL-based methods (both DIP and DIPP) achieve better visualization of the vasculature. Compared to DIP, our DIPP framework further enhances the resolution and contrast of image details. This can be confirmed in the enlarged images shown in (b), where some fine structures have been successfully restored (arrows). In addition, as shown in the image profiles in (c), our DIPP can distinguish tiny structures at different locations, showing much higher peaks and lower valleys.

Fig. 12 shows the restoration results of the healthy mouse around the kidney region. It can be seen that our DIPP is able to recover clear vascular structure (blue arrows) compared to other methods. As shown in the enlarged blue box regions, DIPP also improves the resolution along tissue surface boundaries. Fig. 12(c) shows the signal intensity along the white solid line in (b), compared with other methods, the profile of the DIPP image has been improved in terms of contrast and resolution. Overall, our DIPP framework excels in restoring image contrast and detail in real-world PAT images.

4. Discussions

From the above results, compared to the state-of-the-art image restoration methods, the proposed DIPP framework is shown to handle the unknown and non-uniform degradation in PAT imaging nicely. To our best knowledge, our DIPP is the first attempt to apply the DIP concept to spatially variant image deconvolution. Our DIPP is an unsupervised learning strategy that requires no labeled data for network training. This increases its usability in real-world PAT imaging applications.

The major innovation of our DIPP is that it imposes hybrid image and PSF priors using deep neural networks. In contrast to the original DIP [29] framework, the addition of PSF predictor in our method stabilizes the iteration process and better adapts to the non-uniform degradation. In the loss function settings of the PSF predictor, we use kernel loss instead of image loss so that the predictor is forced to learn the spatially variant degradation rather than finding an average degraded solution for the whole image. In addition, our DIPP allows the incorporation of experimentally measured PSFs of the PAT system to further improve algorithm performance.

When compared with SelfDeblur [30], DIPP is designed to handle spatially variant image degradation, whereas SelfDeblur [30] only applies to spatially invariant degradation, i.e., the PSF is uniform. SelfDeblur uses a simple fully-connected network for deep PSF prior, while DIPP uses the mutual affine layer and residual connections in the PSF predictor. The mutual affine layer increases the connection between feature maps, and the residual connection deepens the network. This enables the PSF predictor to obtain richer feature information with fewer parameters, and to facilitate the estimation of degradation information for each image patch using kernel loss. In addition, the PSF predictor focuses on local image patches rather than the global image, making it less sensitive to PSF size settings, and therefore obtains consistent estimation results even though the PSF size is not known. This is useful for PAT image restoration, which has unknown spatially varying degradation. Therefore, we can see that the DIPP framework substantially improves the ability to predict the heterogeneous degradation information.

For network initialization strategy, random initialization does not impose restrictions on the network weights in advance. We found that by using random initialization rather than pre-training can help the PSF predictor to focus better on the degradation of the input image. This indicated that using simulated PAT datasets to pre-train the PSF predictor may lead to poor results because real and simulated PAT images have inconsistent degradation types.

As shown in our simulation experiments, some of the results are

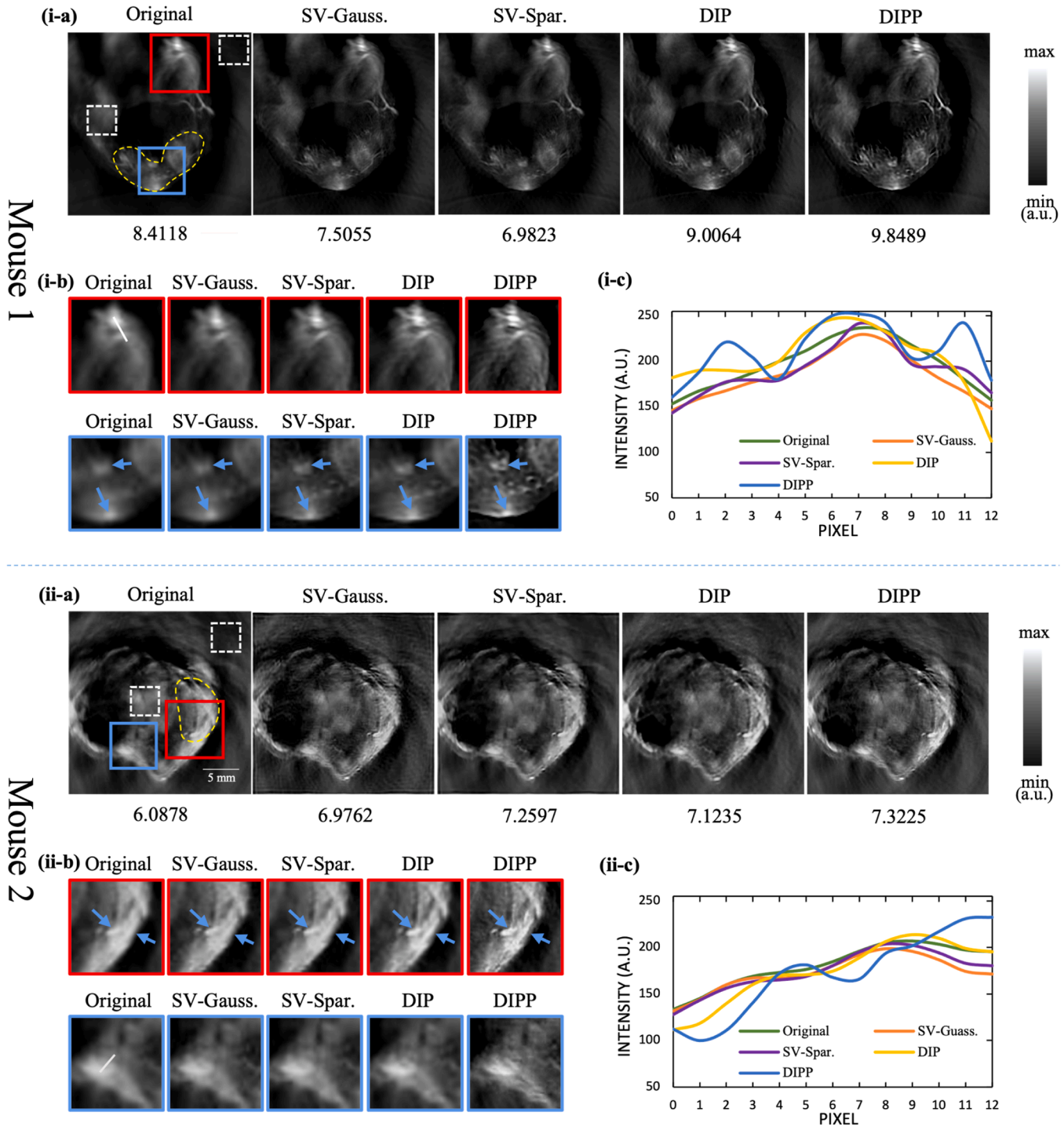


Fig. 11. PAT image restoration result of the cancerous mice images. (a) The original PAT image and its deconvolution results with traditional methods (SV-Gauss. and SV-Spar.), DIP method [29] and ours DIPP method. The tumors are outlined with the yellow dashed lines. The number below each image is the SNR calculated within the white boxes. (b) Enlarged areas corresponding to the red and blue box in (a). (c) Image profiles drown along the white solid line in (b) of each panel.

overfitted to the input and therefore may look better than the ground truth. This is one of the common shortcomings of unsupervised learning approaches such as DIP and etc. [29,30]. This overfitting problem may be mitigated by adding extra regularization on the model. Moreover, the two parameters, α and T, are tuned empirically in our DIPP. We have provided the restoration results of different α and T in the Supplementary Material. Finally, in our current work, we have only studied the feasibility of the proposed method for cross-sectional PAT images. However, 3D PAT imaging systems have been developed recently [43, 44]. In these systems, the PAT images are inherently 3D for each acquisition. Therefore, how to model the degradation in 3D space and how image restoration should be performed remain open questions for future research.

5. Conclusions

In this work, we present DIPP, a DL-based image restoration method for PAT imaging. Based on the classical MAP framework for image

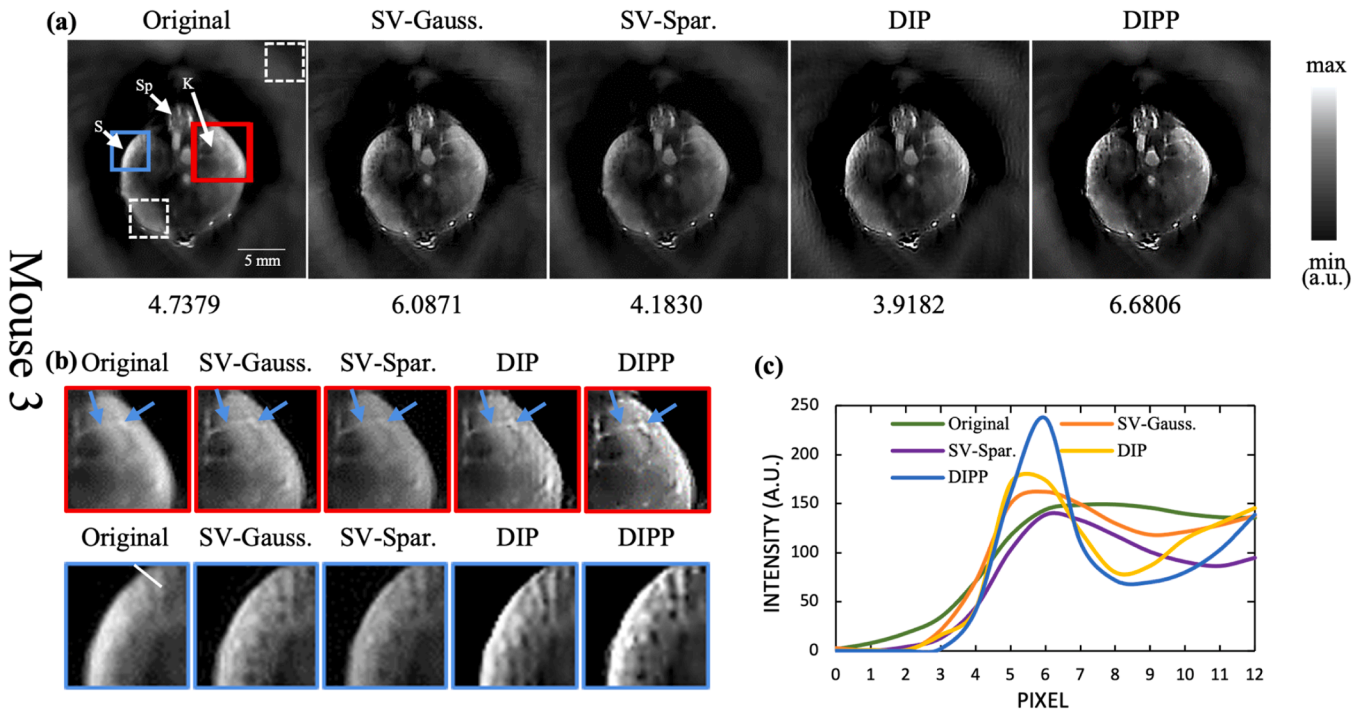


Fig. 12. PAT image restoration result of the healthy mice image. (a) The original PAT image and its deconvolution results with traditional methods (SV-Gauss. and SV-Spar.), DIP method [29] and ours DIPP method. The number below each image is the SNR calculated within the white boxes. (b) Enlarged areas corresponding to the red and blue box in (a). (c) Image profiles plotted along the solid white line across the abdominal cavity of the mouse for each panel in (b). K: kidney; S: spleen; Sp: spine.

deconvolution, our DIPP method employs a novel unsupervised learning model that takes one single input image for iterative optimization. We can also incorporate an experimentally measured PSF map to further improve image recovery. We conduct extensive experiments to demonstrate the performance of our DIPP framework, and the results show that DIPP offers superior image restoration capability compared with both classical approaches and state-of-the-art deep learning methods.

CRediT authorship contribution statement

Kaiyi Tang: Methodology, Software, Investigation, Data analysis, Writing – original draft. **Shuangyang Zhang**: Software. **Yang Wang**: Software. **Zhenyang LiuLingjian Chen**: Software. **Xiaoming Zhang**: PAT Experiments. **Zhichao Liang**: PAT Experiments. **Huafeng Wang**: PAT Experiments. **Wufan Chen**: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. **Li Qi**: Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Li Qi reports financial support was provided by Guangdong Provincial Natural Science Foundation. Li Qi reports financial support was provided by Guangdong Provincial Pearl River Talents Program. Wufan Chen reports financial support was provided by Special Project for Research and Development in Key Areas of Guangdong Province.

Data Availability

Data will be made available on request.

Acknowledgements

This work was supported in part by Guangdong Basic and Applied Basic Research Foundation (2021A1515012542, 2022A1515011748), Guangdong Pearl River Talented Young Scholar Program (2017GC010282), and Key-Area Research and Development Program of Guangdong Province (2018B030333001).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pacs.2023.100536.

References

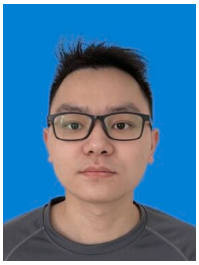
[1] V. Ntziachristos, D. Razansky, Molecular imaging by means of multispectral optoacoustic tomography (MSOT), Chemical reviews 110 (5) (2010) 2783–2794.
[2] L.V. Wang, S. Hu, Photoacoustic tomography: in vivo imaging from organelles to organs, Science 335 (6075) (2012) 1458–1462.
[3] S. Zhang, et al., Photoacoustic imaging of living mice enhanced with a low-cost contrast agent, Biomed. Opt. Express 10 (11) (2019) 5744–5754.
[4] S. Zhang, et al., Pixel-wise reconstruction of tissue absorption coefficients in photoacoustic tomography using a non-segmentation iterative method, Photoacoustics 28 (2022), 100390.
[5] K.E. Thomenius. Evolution of ultrasound beamformers, IEEE, 1996, 1996 IEEE Ultrasonics Symposium. Proceedings.
[6] Q. Sheng, et al., A constrained variable projection reconstruction method for photoacoustic computed tomography without accurate knowledge of transducer responses, IEEE Trans. Med Imaging 34 (12) (2015) 2443–2458.
[7] X. Li, et al., Multispectral interlaced sparse sampling photoacoustic tomography based on directional total variation, Comput. Methods Prog. Biomed. 214 (2022), 106562.
[8] X. Li, et al., Multispectral interlaced sparse sampling photoacoustic tomography, IEEE Trans. Med Imaging 39 (11) (2020) 3463–3474.
[9] M.L. Li, Y.C. Tseng, C.C. Cheng, Model-based correction of finite aperture effect in photoacoustic tomography, Opt. Express 18 (25) (2010) 26285–26292.
[10] M. Xu, Photoacoustic imaging in biomedicine, Review of scientific instruments 77 (4) (2006).
[11] Lauer, T.R., 2002. Deconvolution With a Spatially-Variant PSF. 2002, arXiv. p. 167–173.

[12] J.G. Nagy, D.P. O'Leary, Restoring images degraded by spatially variant blur, SIAM Journal on Scientific Computing 19 (4) (1998) 1063–1082.
[13] M. Haltmeier, Spatial resolution in photoacoustic tomography: effects of detector size and detector bandwidth, Inverse Problems 26 (12) (2010), 125002.
[14] Y. Wang, et al., Photoacoustic imaging with deconvolution algorithm, Phys. Med. Biol. 49 (14) (2004) 3117–3124.
[15] L. Qi, et al., Cross-sectional photoacoustic tomography image reconstruction with a multi-curve integration model, Comput. Methods Prog. Biomed. 197 (2020), 105731.
[16] T. Chaigne, et al., Super-resolution photoacoustic fluctuation imaging with multiple speckle illumination, Optica 3 (1) (2016) 54–57.
[17] L. Qi, et al., Photoacoustic tomography image restoration with measured spatially variant point spread functions, IEEE Trans. Med. Imaging 40 (9) (2021) 2318–2328.
[18] J. Chen, et al., Blind-deconvolution optical-resolution photoacoustic microscopy in vivo, Opt. Express 21 (6) (2013) 7316–7327.
[19] T. Jetzfellner, V. Ntziachristos, Performance of blind deconvolution in optoacoustic tomography, Journal of innovative optical health sciences 4 (4) (2011) 385–393.
[20] S. Zhang, et al., MRI information-based correction and restoration of photoacoustic tomography, IEEE Trans. Med Imaging 41 (9) (2022) 2543–2555.
[21] C.J. Schuler, et al., Learning to Deblur, IEEE Trans. Pattern Anal. Mach. Intell. 38 (7) (2016) 1439–1451.
[22] Wieschollek, P., et al., 2017. End-to-end learning for image burst deblurring. in Computer Vision–ACCV 2016: *13th Asian C*onference on Computer Vision, Taipei, Taiwan, November 20–24, 2016, Revised Selected Papers, Part IV 13. 2017. Springer.
[23] Nah, S., T. Hyun Kim, and K. Mu Lee, 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[24] Hradiš, M., 2015. Convolutional Neural Networks for Direct Text Deblurring. in British Machine Vision Conference. 2015.
[25] C. Cai, et al., End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging, Optics letters 43 (12) (2018) 2752–2755.
[26] A. Hauptmann, et al., Model-based learning for accelerated, limited-view 3-D photoacoustic tomography, IEEE transactions on medical imaging 37 (6) (2018) 1382–1393.
[27] M. Lu, et al., Artifact removal in photoacoustic tomography with an unsupervised method, Biomed. Opt. Express 12 (10) (2021) 6284–6299.
[28] Agrawal, S., et al., 2021. Learning Optical Scattering Through Symmetrical Orthogonality Enforced Independent Components for Unmixing Deep Tissue Photoacoustic Signals. 2021(5–5).
[29] Ulyanov, D., A. Vedaldi, and V. Lempitsky, 2017. Deep Image Prior. in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2017.
[30] Ren, D., et al., 2020. Neural Blind Deconvolution Using Deep Priors. in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
[31] T. Vu, et al., Deep image prior for undersampling high-speed photoacoustic microscopy, Photoacoustics 22 (2021), 100266.
[32] G. Bredell, et al., Wiener Guided DIP for Unsupervised Blind Image Deconvolution, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (2023).
[33] Gandelsman, Y., A. Shocher, and M. Irani, 2019. "Double-DIP": Unsupervised Image Decomposition via Coupled Deep-Image-Priors. in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.
[34] Ronneberger, O., P. Fischer, and T. Brox, 2015. U-net: Convolutional networks for biomedical image segmentation. in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. 2015. Springer.
[35] J. Liang, et al., Mutual affine network for spatially variant kernel estimation in blind image super-resolution, Proceedings of the IEEE/CVF International Conference on Computer Vision (2021).
[36] D.P. Kingma, J. Ba Adam, A method for stochastic optimization, arXiv Prepr. arXiv 1412 (2014) 6980.
[37] A. Levin, et al., Understanding and evaluating blind deconvolution algorithms, in: *IEEE conference on computer vision and pattern recognition*, IEEE, 2009.
[38] Bevilacqua, M., et al., 2012. Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding. in British Machine Vision Conference. 2012.
[39] J. Pan, et al., Deblurring images via dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 40 (10) (2018) 2315–2328.
[40] Tao, X., et al., 2018. Scale-recurrent Network for Deep Image Deblurring. in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.
[41] Agustsson, E. and R. Timofte, 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017.
[42] Martin, D., et al., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. in Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. 2001. IEEE.
[43] R. Ni, et al., Multiscale optical and optoacoustic imaging of amyloid-β deposits in mice, Nat. Biomed. Eng. 6 (9) (2022) 1031–1044.
[44] L. Lin, et al., High-speed three-dimensional photoacoustic computed tomography for preclinical research and clinical translation, Nat. Commun. 12 (1) (2021) 882.

Kaiyi Tang is pursuing an academic master's degree in Biomedical Engineering from Southern Medical University. Her research interests include photoacoustic tomography and image restoration.

Shuangyang Zhang received his B. Eng. degree in 2017 from Southern Medical University. He is currently pursuing a Ph.D. degree in Engineering at the School of Biomedical Engineering, Southern Medical University. His research interests include multimodal image registration and information fusion.

Yang Wang is currently pursuing a master's degree at Southern Medical University. His research focuses on high-quality photoacoustic tomography imaging methods based on sound speed correction.



Xiaoming Zhang received his bachelor's degree from Southern Medical University in 2022. He is currently pursuing the master's degree in the school of Biomedical Engineering, Southern Medical University. His research interests include image restoration based on photoacoustic Tomography and relevant system based on photoacoustic imaging and MRI.

Zhenyang Liu received his B.S. degree in Biomedical Engineering in 2021 from the school of Biomedical Engineering of Southern Medical University. He is currently a master student at the Institute of Medical Information, School of Biomedical Engineering, Southern Medical University. His research interests include photoacoustic imaging and optical coherence tomography.

Zhichao Liang received his bachelor's degree from Southern Medical University in 2021. He is currently pursuing the master's degree in the school of Biomedical Engineering, Southern Medical University. His research interests include photoacoustic tomography and image segmentation.

Wufan Chen received his B.S. degree in 1975 and M.Sc. degree in 1981, both from Beihang University. He is currently a full professor at the Guangdong Provincial Key Laboratory of Medial Image Processing, School of Biomedical Engineering, Southern Medical University. His research interests include biomedical imaging principle and image processing.

Li Qi received his Ph.D. degree in Optical Engineering in 2016 from Nanjing University. He is currently an associate professor at the Guangdong Provincial Key Laboratory of Medial Image Processing, School of Biomedical Engineering, Southern Medical University in Guangzhou, China. His research interests include photoacoustic imaging and optical coherence tomography.