Deep Image Prior for Sparse-sampling Photoacoustic Microscopy

Tri Vu, Anthony DiSpirito III, Daiwei Li, Zixuan Zhang, Xiaoyi Zhu, Maomao Chen, Dong Zhang, Jianwen Luo, Yu Shrike Zhang, Roarke Horstmeyer, and Junjie Yao

Abstract—Photoacoustic microscopy (PAM) is an emerging method for imaging both structural and functional information without the need for exogenous contrast agents. However, state-of-the-art PAM faces a tradeoff between imaging speed and spatial sampling density within the same field-of-view (FOV). Limited by the pulsed laser’s repetition rate, the imaging speed is inversely proportional to the total number of effective pixels. To cover the same FOV in a shorter amount of time with the same PAM hardware, there is currently no other option than to decrease spatial sampling density (i.e., sparse sampling). Deep learning methods have recently been used to improve sparsely sampled PAM images; however, these methods often require time-consuming pre-training and a large training dataset that has fully sampled, co-registered ground truth. In this paper, we propose using a method known as “deep image prior” to improve the image quality of sparsely sampled PAM images. The network does not need prior learning or fully sampled ground truth, making its implementation more flexible and much quicker. Our results show promising improvement in PA vasculature images with as few as 2% of the effective pixels. Our deep image prior approach produces results that outperform interpolation methods and can be readily translated to other high-speed, sparse-sampling imaging modalities.

Index Terms— convolutional neural network, deep image prior, deep learning, high-speed imaging, photoacoustic microscopy, raster scanning, undersampling

I. INTRODUCTION

Owing to its dependence on optical absorption, photoacoustic imaging (PAI) has seen increased use as a modality for functional imaging over the past decade [1, 2]. PAI combines the photoacoustic excitation from a laser source with the ultrasonic detection of the pressure waves. In detail, upon a short-pulsed laser illumination that satisfies thermal and stress confinement, thermo-elastic expansion of endogenous chromophores results in pressure waves, which propagate within the tissue and are subsequently detected by an ultrasonic sensing device. Photoacoustic microscopy (PAM) is a specific field of PAI that uses focused illumination or detection within a small FOV to capture high-resolution images. Compared to the more widely employed confocal and two-photon microscopy, PAM, especially optical-resolution PAM (OR-PAM), has the advantage of an imaging depth that extends beyond the diffusion limit. Accordingly, it has been a powerful microscopic tool for in vivo study of small animals, especially for imaging of mouse brain vasculature [3, 4].

Similar to other microscopic modalities, PAM relies on point-by-point scanning to acquire volumetric images over a large region-of-interest (ROI). This scanning strategy follows the Nyquist-Shannon sampling theorem (NSST), which demands that the step size be, at most, half the spatial resolution. This fundamental principle sets the required spatial sampling density and therefore the imaging rate of the system. In practice, depending upon the size and complexity of the desired targets, one can still technically violate NSST without losing too much useful information. For example, in OR-PAM, the targeted images are usually blood vessels with a minimal diameter of 10 μm [2, 5, 6]. Thus, a typical step size of 5 μm, larger than the spatial resolution of most traditional OR-PAM systems, is still satisfactory [7-9]. However, these systems are still limited by the pulse repetition rate (PRR) of the induced laser or, in older systems, the maximum mechanical scanning speed. Typical OR-PAM systems have a 4-5 Hz B-scan rate at 2-5 kHz PRR, which is extremely slow for certain functional imaging and clinical applications [8, 9].

Existing work has focused on modifying the mechanical or optical scanning schemes. For mechanical translation, previous groups have used voice-coil and piezo linear stage with modest scanning improvement [10, 11]. Other optical methods using Galvo, microelectromechanical, and polygon scanners achieved 10 to 100 times improvement in scanning speed by sacrificing the confocal alignment of illumination and detection [12-16]. These state-of-the-art approaches can reach up to 1 kHz B-scan rate. PAI is also spatially sparse [17-20], thus enabling compressed sensing (CS)-related studies to overcome NSST in traditional PAM [17, 21]. However, despite complex compressed sampling schemes, recent studies have demonstrated success for CS only up to a 20% sampling density [17, 21]. Despite their intrinsic advantages, all of these methods share the burden of extensive hardware modifications.

The advances of deep learning (DL) have been slowly

Manuscript received May 29, 2020; accepted MONTH DAY, YEAR. Date of publication MONTH DAY, YEAR; date of current version MONTH DAY, YEAR. This work was supported in part by the National Institutes of Health (R01 EB026143, R01 NS111039, R01 NS115581, R21 EB027304, R43 CA243822, R43 CA239830, R44 HL138185); Duke MEDx Basic Science Grant; Duke Center for Genomic and Computational Biology Faculty Research Grant; Duke Institute of Brain Science Incubator Award; American Heart Association Collaborative Sciences Award (18CSA4080277).

T. Vu, A. DiSpirito III, D. Li, X. Zhu, M. Chen, D. Zhang, and J. Yao are all with the Photoacoustic Imaging Lab, Duke University, Durham, NC 27708 USA (email: jy2111@duke.edu).

D. Zhang and J. Luo are with the Department of Biomedical Engineering, Tsinghua University, Beijing, China, 100084.

Y. Zhang and Z. Zhang are with the Division of Engineering in Medicine, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Cambridge, MA 02139 USA

R. Horstmeyer is with the Computational Optics Lab, Duke University, Durham, NC 27708 USA

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier DOI HERE
incorporating into PAI. Over the past few years, researchers have primarily focused on artifact removal, object identification, and sparse sampling in photoacoustic computed tomography (PACT) [22-30]. DL, on the other hand, has only recently been implemented to address the undersampling problem in PAM [31, 32]. These prior applications of deep learning in PAM have impressive performance on in vivo data, with a structural similarity index (SSIM) of up to 0.92. However, all of these methods, including those that belong to PACT, depend on training convolutional neural networks (CNN) with large datasets that have ground truth. Generating such datasets is both time- and resource-intensive, especially for clinical cases. Additionally, pre-trained models are inclined to be biased toward learned features of their targeted training datasets [33, 34]. Thus, there is a high probability that they will fail to generalize when faced with unfamiliar input images [29]. Furthermore, with every new sampling pattern or sample added to the training dataset, the models will need to be retrained if this new information is to be integrated into the model weights. Therefore, a DL model that does not require prior training is preferable for producing a well-generalized solution.

One such DL approach that requires no prior training and no labeled ground truth data and has been shown to provide robust results is the emerging deep image prior (DIP) method [35]. The DIP network acts as a parametrization, which is solved by an iterative optimization scheme [35]. The output of the model, given a noise input, is the expected result of the restoration. Then, after distorted by a known operator, the output is compared to the downgraded image in the objective function for backpropagation. Thus, although traditional supervised DL methods require prior pairwise training with ground truth, the DIP model only needs the downgraded image itself, and the distorting operator. Speculatively, the optimized DIP network learns features within the distorted images from applied convolutional filters and imposes these implicit traits (i.e., prior) on the restored output [35].

The DIP approach shows success with natural images in various medical imaging recovery tasks, such as inpainting, denoising, and artifact removal [35]. In [36, 37], Gong et al. utilized DIP for PET reconstruction with MR images. Zhou et al. succeeded in 3D reconstruction of a refractive index in diffraction tomography [33]. Regarding the sparse-sampling scheme, Trampert et al. applied DIP to scanning electron microscopy (SEM) [38]. This work set a promising starting point for these technologies. However, this DIP method for SEM currently lacks hard statistical results and a thorough comparison with other methods, especially with pre-trained CNNs. Most importantly, previous works have not demonstrated the feasibility of DIP for PAI or sparse-sampling PAM.

In this work, inspired by DIP, we propose a new method using an untrained CNN for correcting undersampled PAM images (DIP-PAM), thereby providing an avenue for safe high-speed imaging. In terms of model training, the model iteratively finds an optimized, fully sampled image that approximates the known, sparsely sampled image given a known downsampling mask. Unlike CS, DIP-PAM does not need a randomized scanning pattern that complicates the hardware setup. Our approach also eliminates the need for a pre-captured pairwise training dataset, like those used in traditional DL methods, ensuring generalization with any given input. This model generalization is tested on mouse vasculature and a bioprinted phantom data. The in vivo experiments are carried out using two different PAM system platforms: a traditional mechanical scanning system and a state-of-the-art high-speed PAM system, for further proving our method’s universality to fast-scanning scheme. Subsequently, the performance of our method was compared with interpolation and pre-trained models. To the best of our knowledge, this is the first study of DIP for PAM and for sparse-sampling PAM in particular.

II. METHODS

A. Deep Image Prior (DIP) for Sparse-sampling Problem

To formulate a DIP solution, we first frame the problem in terms of a typical image restoration task solved by an optimization scheme. With a given distorted image \( x_0 \), one tries to find the original image \( x \) of width \( W \) and height \( H \) through an energy minimization problem:

\[
    x^* = \arg\min_x \{ L(d(x); x_0) + R(x) \} \tag{1}
\]

where \( L(x; x_0) \) is our objective function, \( d(\cdot) \) is the distorting operator, and \( R(x) \) is the regularization term. This regularizer serves as the data prior, embedding prior information about the image. For example, total variation is often used for \( R(x) \) to enforce smoothness. The choice of \( R(x) \) is critical to find the optimal solution. Besides existing denoising priors, such as Gaussian mixture models and block-matching-and-3D filtering (BM3D), various groups have investigated the use of CNN as \( R(x) \) [39-41], especially for PACT artifact removal [27]. However, these methods do not account for the data term \( L(x; x_0) \) and thus need training pairs. To get rid of the \( R(x) \), instead of searching for \( x \) explicitly, one can find \( \theta \) on the surjective function \( g \), such that Eq. (1) becomes:

\[
    \theta^* = \arg\min_\theta \{ L(d(g(\theta)); x_0) + R(g(\theta)) \} \tag{2}
\]

Indeed, a correct parametrization of \( x \) in place of \( g(\theta) \) can express the explicit prior itself, thus eliminating the prior term. Such parametrization can be represented by a neural network \( f_\theta(x) \) with random noise input \( x \) and weights \( \theta \):

\[
    \theta^* = \arg\min_\theta L(d(f_\theta(x)); x_0) \tag{3}
\]

which can be rewritten as a non-linear least squares solution as:

\[
    \theta^* = \arg\min_\theta \| d(f_\theta(x)) - x_0 \|^2 \tag{4}
\]

In our sparse-sampling problem, the distorting operator is simply the pixel-wise multiplication between the fully sampled image and a sparse-sampling mask. We eventually have the following operator:

\[
    d(f_\theta(z)) = f_\theta(z) \circ m \tag{5}
\]
where ⊙ is the Hadamard product and \( m \in \{0, 1\}^{H \times W} \) is the binary mask of the sampling pattern, in which ‘1’ denotes the sampled pixels and ‘0’ denotes the skipped pixels. Substituting \( d(\cdot) \) in Eq. (5) to Eq. (4), we have the final formulation of DIP optimization for sparse-sampling restoration:

\[
\theta^* = \underset{\theta}{\arg\min}\|x' - x_0\|^2 \quad \text{with} \quad x' = f_0(z) \circ m 
\] (6)

Generally, instead of directly optimizing the fully sampled image \( x \), which requires a large number of training pairs, we iteratively update the weight (\( \theta \)) of the CNN, which outputs \( x = f_0(z) \). This output, multiplied by the sampling mask, is expected to give us the sparsely sampled image from the high-speed imaging system. A depiction of DIP optimization is shown in Fig. 1.

In this study, we use the noise-based regularization in the input as recommended by [35]. Instead of using a fixed uniform noise input, we add to \( z_0 \) an additional noise \( z' \) following a Gaussian distribution, with zero mean and a standard deviation \( \sigma_{z'} = 0.07 \). Instead of choosing the default value of 0.03 and 0.05 from [35], our \( \sigma_{z'} \) is experimentally determined by varying \( \sigma_{z'} \) from 0.01 to 0.21 (Fig. 2(d)). This evaluation gives us the local maxima at 0.06-0.08. Thus, we chose \( \sigma_{z'} = 0.07 \) for this study. This perturbing noise helps prevent overfitting and leads the optimization to a more generalized solution [42, 43]:

\[
z = z_0 + z' 
\] (7)

![Fig. 1. Schematics of the iterative DIP optimization. At each iteration \( k \) of \( N \) iterations, gradient of \( x_k \) and \( x_0 \) w.r.t. \( \theta \) is calculated and back-propagated to the encoder-decoder CNN model containing \( \theta \).](image)

### B. Network Architecture and Training

Our CNN for the deep prior takes the form of the U-Net with some modifications [44]. U-Net is a well-known network architecture for DL, especially in medical imaging applications. Its encoder-decoder framework effectively extracts important features in the compression (encoder/downsampling) path and reconstructs the target from those features in the decompression (decoder/upsampling) path. There are two reasons for using a U-Net architecture in DIP. First, even with a random weight and random noise input, this model can still reconstruct spatial structures with self-similarities, regardless of the stochastic inner hyperparameters [35]. Thus, it is possible that, after being exposed to the deteriorated image, the model will learn to apply certain similarities of the existing structures to the rest of the recovered image. The second reason lies in the inter-layer concatenations, or skip connections. By concatenating channels from deep layers, U-Net can pass high-level features to the upsampling scheme and create self-similarities on the output at various different scales [35].

The architecture of our modified U-Net is shown in Fig. 2(a). In general, it contains the main characteristics of U-Net. The modifications mainly reside in the inner blocks of our network. Instead of increasing the number of filters in the downsampling path and decreasing the number of filters in the upsampling path, we keep this number constant (64) for all the layers to decrease the number of model parameters. In addition to this channel reduction, we also incorporate a large kernel size (11 × 11) for the network. We hypothesize that a large receptive field can more readily draw from the information of nearby pixels in the sparse-sampled image, which are spatially separated due to the sampling pattern. In addition, we avoid using transposed (sub-pixel) convolutional layers in the decoder path because they result in severe checkerboard artifacts, as shown in Fig. 2(b). These artifacts come from the deconvolution overlap and random initialization [45-47]. Instead, as suggested from [45, 47], we split this upsampling process into standard 2D spatial convolution and bilinear interpolation.

![Fig. 2. (a) Network architecture of the modified U-Net. The input is 32-channel 300 × 300 random noise, and the output is single-channel recovered PAM image. All other layers have 64 filters. (b) Checkerboard artifact from the network with transposed convolution. Our model, with standard convolution and upsampling, eliminates this problem. Scale bar: 0.5 mm. (c) Training stability improved by AMSGrad, while Adam has exploding gradient at ~4,000th iteration. (d) Comparison of different \( \sigma' \) values with SSIM to choose \( \sigma' \) for our study. All SSIMs are averages of 15 samples, with three from each sparse sampling pattern. Error bar: standard deviation.](image)
C. Model Training and Evaluation

The modified U-Net is trained with the following general settings. The optimizer is AMSGrad with a clip value of $\tau = 10$, which has significant improvement in training stability compared to Adam (Fig. 2(c)). We train the model for 5,000 iterations and use mean-squared errors (MSE) loss for all experiments. All convolutional layers are followed by Leaky ReLU activation ($\alpha = 0.2$) instead of ReLU to help combat the vanishing gradient problem [48]. The model is trained on a 64-bit workstation with a NVIDIA RTX 2080 Ti GPU and an Intel Core i9-9900K, using Python v3.6.1 and Keras 2.2.5 with Tensorflow backend. Training time for each 300 × 300 pixel image takes approximately 15 minutes. More information regarding training parameters can be found in the code (https://github.com/trivu169/deep-prior-pam) and Fig. 2(a).

The model performance is evaluated using the average SSIM and peak signal-to-noise ratio (PSNR) on the testing datasets. These metrics represent both global and local information of the output of the reconstructed fully sampled images [29]. The testing datasets contain 37 vascular images with 300 × 300 pixels for each sampling pattern. SSIM and PSNR are also obtained for bilinear, bicubic, and lanczos (8 × 8 kernel) interpolation as well as for a pre-trained FD U-Net in [32] for comparison. Statistical evaluation of p-values is calculated using ANOVA with post-hoc Tukey’s HSD test. To minimize the reduction of statistical power [49], we only compare bicubic interpolation with DIP-PAM and FD U-Net. To apply DIP-PAM to larger images, we use a model patchwork algorithm to avoid border artifacts [32]. It is also possible to input the whole image to the model, if there is enough GPU and RAM memory.

D. Data Acquisition and Sparsely Sampled Mask Construction

Two types of optical-resolution PAM (OR-PAM) imaging systems are used in this study. The first one is the traditional OR-PAM system used in [9, 50] with lateral resolution of 3.5 μm, axial resolution of 15 μm, and step size of 5 μm. The testing datasets for evaluation are from this system due to its ability to acquire large test images. This system has been widely used in functional OR-PAM due to its high spatial resolution. However, its B-scan rate is only 5 Hz. In contrast, the second platform, a high-speed optical scanning OR-PAM using a 12-faced rotating polygon mirror, has an impressive B-scan rate of 1 kHz. Its spatial resolution is 10 μm. The system is downsampled 12 times along the slow axis and 6 times along the fast axis with no ground truth (fully sampled image).

All experimental protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Duke University. We used mouse brain and mouse ear vascular images, which are the most often imaged targets in PAM [2, 6, 51-53], especially for studying neuronal activity and hemodynamics [53, 54]. All vasculature images from the first system are acquired using a 532-nm laser and follow the post-processing steps mentioned in [32]. Even though the main imaged objects are blood vessels, we will reiterate that DIP-PAM is not target-specific. We expect that the DIP-PAM performance is generalizable to most sample types. To test this hypothesis, we applied DIP-PAM to phantom data of bioprinted hydrogels [55], acquired using both 532- and 590-nm lasers. All images in this work are maximum amplitude projections (MAP) of the volumetric raw data.

We follow a similar downsampling procedure as shown in [32] for the traditional PAM system images to artificially generate sparse-sampled images with effective areas that differ from fully sampled ones. Here, we also denote $[S_x, S_y]$ as our downsampling factor in $x$ (fast axis) and $y$ (slow axis), $\frac{1}{S_x}$ and $\frac{1}{S_y}$ are downsampling ratios along these axes. These ratios are then used to construct the binary downsampling mask. This 8-bit unsigned integer mask contains white pixels positioned $S_x - 1$ and $S_y - 1$ away from each other with black pixels in between (Fig. 3). The six sampling patterns used in this study are: [10, 5], [7, 3], [6, 1], [5, 1], and [4, 1] for mechanical-scanning system, and [6, 12] for high-speed optical-scanning PAM system.

III. Results

A. Qualitative DIP-PAM Results on in vivo Data

Representative outputs from all methods are shown in Fig. 5. The most striking improvements from DIP-PAM are seen with the high sparsity downsampling patterns (<20% pixel density). Particularly, at the [7, 3] and [10, 5] patterns, vessels in DIP-PAM outputs remain relatively continuous, as compared to the disjointed and aliased vessels of the interpolation-based methods, as denoted in Fig. 5. This is notable since our method could not learn this trait of vascular continuity from ground truth. Rather, this improvement arises from the generalized prior captured by the neural network at high noise.

![Fig. 3](imageurl) Fig. 3. (a) Picture of the bioprinted sample. (b) Sampling mask of the [7, 3] scanning pattern with the zoomed-in details. Yellow arrows denote the raster-scanning direction, and green arrows indicate the spaces between sampling (white) points. Here, $S_x - 1 = 6$ and $S_y - 1 = 2$. Scale bar: 0.5 mm.

![Fig. 4](imageurl) Fig. 4. DIP-PAM output after 5000 iterations under different noise regularization $\sigma_z$. Scale bar: 0.5 mm.
regularization, as depicted in Fig. 4. Low $\sigma_z' \leq 0.03$ leads to results similar to those of regular interpolation. In addition, considering the profile in Fig. 6(b), DIP-PAM does not over-smooth the textures and is able to accurately maintain the correct pixel intensities.

As expected, the pre-trained FD U-Net, at high sparsity, outperforms DIP-PAM as expected, with sharp and continuous vessels similar to the fully sampled image (Fig. 6(a)). Intensity profiles in Fig. 6(c) also depict improved contrast from FD U-Net. However, at low pixel density ([6, 1] or smaller $S_x$), the

![Image](image_url)

**Fig. 5.** Representative output from DIP-PAM and other methods for traditional PAM system. Green arrows represent disjointed and jagged vessels successfully corrected by DIP-PAM. At highly sparse patterns, DIP-PAM shows significant improvement over untrained methods, approximating the performance of FD U-Net. Scale bar: 0.5 mm.
difference between DIP and FD U-Net is less clear. The pre-trained method, despite improved contrast, has stripe-like artifacts (Fig. 6(c)), representing the leftover effects of the scanning pattern, which has not been completely removed. Up to 25% pixel density ([4, 1] pattern), outputs from all methods are visually indistinguishable.

An illustration of the optimization is depicted in Fig. 8. Note that the model quickly goes from an arbitrary output in the beginning to a meaningful representation at 1000th iteration. The learning curve in Fig. 2(c) is in accordance with this behaviour, with a sharp reduction in loss after the first 1000 iterations. The remainder of the optimization focuses on recovering higher frequency details and smaller vessels. Furthermore, from both Fig. 2(c) and Fig. 8, one can observe that there is negligible improvement after the 3000th iteration. Thus, we can effectively reduce the training time by half (from 15 to 7.5 minutes) without sacrificing the image equality of the output.

We also applied DIP-PAM to the mouse vasculature data acquired by the high-speed polygon scanning system, which is [6, 12] downsampled (Fig. 7). For this significantly sparse sampling pattern, we further appreciate the superiority of DL-based methods over interpolation. The severity of the vasculature aliasing, denoted in Fig. 7, invalidates the use of interpolation for functional quantification of vessel profiles at these sparsities. DIP-PAM and FD U-Net maintain the vessel shape continuity and edge smoothness. FD U-Net, however, appears to have some remaining artifacts, as seen by horizontal traces similar to those from the traditional PAM system. DIP-PAM, conversely, is free of upsampling artifacts, maintains the

Fig. 6. Pixel intensity profiles of line 1-3 (a, b, and c, respectively) from Fig. 5. In all cases, DIP-PAM displays consistent agreement with the ground truth (fully sampled images). In line 1, taken from the [10,5] pattern, traditional interpolation shifts the position of the vessel due to stair-case artifact, while FD U-Net results in unwanted vertical stripes (line 2 from [5,1]). At low sparsity (line 3 from [4,1]), all methods exhibit minimal difference in performance, apart from the increased contrast of FD U-Net. Scale bar: 0.5 mm.

Fig. 7. Representative output from DIP-PAM and other methods for high-speed polygon scanning PAM system. Top rows: whole-field mouse brain vasculature images. Bottom rows: close-up region marked by red square. Red arrows denote jagged and continuous vessels from interpolation and DIP-PAM, respectively. Yellow arrows indicate broken vessels recovered successfully by DIP-PAM. Yellow circle emphasizes the artificial artifacts from FD U-Net. Scale bar: 125 μm.

Fig. 8. Output after every 1000 iterations of the DIP-PAM optimization. The image is downsampled by [7,3] scanning pattern. Scale bar: 0.5 mm.
connectivity of vasculatures after sparse-sampling reconstruction, and avoids jagged vessel shapes after upsampling. For certain vessels, our method can even retain the continuity that both interpolation and FD U-Net fail to capture (Fig. 7). The success at this level of sparsity corresponds to a 72-fold improvement in scanning time, increasing the imaging rate of the high-speed system from 1 kHz to 72 kHz for laser PRR limited systems.

B. Quantification of DIP-PAM Performance

The previous qualitative observations are supported by our quantification with SSIM and PSNR (Fig. 9 and Table I). FD U-Net remains at the top for most of the quantitative comparisons, followed by DIP-PAM. At the [10,5] downsampling pattern, the SSIM difference between FD U-Net and DIP-PAM is relatively large (0.756 vs. 0.693, respectively). However, that performance gap reduces as the pixel density of sparse-sampled images increases (0.910 vs. 0.903 with the [6,1] pattern, respectively). PSNR differences between DIP-PAM and FD U-Net are similar across all scanning patterns with \( p < 0.05 \) (Fig. 9), potentially because PSNR and the loss function of our model both depend on MSE. This indicates that our approach and FD U-Net are on a par for both SSIM and PSNR at [10,5], [7,3], [6,1], and [5,1] downsampling patterns. These patterns correspond to 50-, 21-, 6- and 5-times respective improvements in modern PAM system image acquisition time. DIP-PAM outperforms all other methods at [6,1] sampling pattern with the highest PSNR of 29.71 dB.

Statistical tests also determine that both DL-based methods outperform conventional interpolation at most of the cases, except the [4,1] pattern (Fig. 9). Interestingly, at this sampling pattern, both bicubic and FD U-Net outperform DIP-PAM significantly for SSIM \( (p < 0.05 \text{ and } p = 0.01, \text{ respectively}) \), but there are no differences for PSNR. This quantification indicates that our method performs better at high sparsity (< 25%) and suggests that, with higher pixel densities, DIP-PAM performance saturates as compared to other methods.

### Table I: Comparison of SSIM and PSNR Between DIP-PAM and Other Methods (Mean ± SD)

|         | [4,1]            | [5,1]            | [6,1]            | [7,3]            | [10,5]           |
|---------|------------------|------------------|------------------|------------------|------------------|
| SSIM    |                  |                  |                  |                  |                  |
| Bilinear| 0.951 ± 0.020    | 0.874 ± 0.042    | 0.840 ± 0.054    | 0.786 ± 0.072    | 0.546 ± 0.096    |
| Bicubic | 0.957 ± 0.019    | 0.877 ± 0.045    | 0.842 ± 0.055    | 0.801 ± 0.073    | 0.542 ± 0.095    |
| Lanczos | 0.957 ± 0.020    | 0.873 ± 0.046    | 0.836 ± 0.059    | 0.794 ± 0.077    | 0.529 ± 0.096    |
| DIP     | 0.940 ± 0.040    | 0.928 ± 0.036    | 0.903 ± 0.047    | 0.853 ± 0.054    | 0.693 ± 0.109    |
| FD U-Net| **0.959 ± 0.018** | **0.944 ± 0.026** | **0.910 ± 0.039** | **0.881 ± 0.055** | **0.756 ± 0.096** |

|         | [4,1]            | [5,1]            | [6,1]            | [7,3]            | [10,5]           |
|---------|------------------|------------------|------------------|------------------|------------------|
| PSNR (dB) |                  |                  |                  |                  |                  |
| Bilinear| 32.41 ± 3.70     | 27.38 ± 3.34     | 26.27 ± 3.51     | 25.17 ± 3.67     | 20.46 ± 3.03     |
| Bicubic | 33.19 ± 3.80     | 27.44 ± 3.42     | 26.31 ± 3.67     | 25.41 ± 3.88     | 20.05 ± 3.05     |
| Lanczos | 33.25 ± 3.84     | 27.38 ± 3.43     | 26.23 ± 3.69     | 25.37 ± 3.93     | 19.91 ± 3.07     |
| DIP     | 32.67 ± 4.86     | 31.42 ± 4.30     | **29.71 ± 4.23** | 27.35 ± 3.79     | 23.04 ± 3.90     |
| FD U-Net| **34.37 ± 4.63** | **32.06 ± 4.30** | 29.43 ± 4.09     | **28.17 ± 4.09** | **24.05 ± 4.03** |

Fig. 9. Boxplots of DIP-PAM and other methods. Top and bottom show the result of SSIM and PSNR, while left to right represent decreasing pixel density. Gray line: median; black dashed line: mean; * \( p = 0.01 \), ** \( p < 0.05 \), *** \( p > 0.05 \).
C. DIP-PAM on Non-vascular Data

All methods are applied to a PAM image of a bioprinted sample to test the robustness and universality of DIP-PAM outside vasculature data. The representative outputs of the [7, 3] pattern are displayed in Fig. 10. Qualitatively, the close-up regions agree well with the in vivo results. All outputs from interpolation methods appear blurry and jagged. The FD U-Net image suffers from the aforementioned checkerboard artifacts caused by its use of transposed convolutions. The DIP-PAM image, conversely, has clear edges and is artifact-free. This phantom result further eliminates the concern that our DL-based method can introduce artificial features, especially in a novel input.

Nonetheless, quantification (Table II) reveals the broader performance of DIP-PAM. Although it still dominates traditional interpolation at high sparsity of [10, 5] downsampling, DIP-PAM gains diminished improvement with increasing pixel density. This is due to the simplicity of the phantom shape that contains mostly low-frequency spatial features, which are easily reconstructed by interpolation approaches. At the [4, 1] downsampling ratios, DIP-PAM falls short of all other methods for both SSIM and PSNR, which confirms the sparsity-related performance hypothesis that arose from the vascular data.

IV. DISCUSSION

One of the central missions in improving PAM technology is to increase imaging speed, which serves as a fundamental roadblock to the application of PAM to time-sensitive tasks. Unlike traditional PAM systems, which rely on slow mechanical scanning, the imaging speed of state-of-the-art, high-speed PAM systems is often limited by the laser’s repetition rate. Sparse sampling has become a necessary compromise when imaging speed is increased. In many cases with fast optical scanning, the imaging speed is inversely proportional to the number of effective pixels. Therefore, reducing the number of pixels through sparse sampling will lead to the same percentage increase in imaging speed.

---

**TABLE II**

COMPARISON OF SSIM AND PSNR BETWEEN METHODS FOR PHANTOM DATA. ALL INDICES ARE THE AVERAGE OF 532- AND 590-NM IMAGES.

| Method     | [4,1]  | [5,1]  | [6,1]  | [7,3]  | [10,5] |
|------------|--------|--------|--------|--------|--------|
| Bilinear   | 0.935  | 0.843  | 0.779  | 0.678  | 0.556  |
| Bicubic    | 0.937  | 0.835  | 0.765  | 0.659  | 0.516  |
| Lanczos    | 0.937  | 0.831  | 0.757  | 0.647  | 0.500  |
| DIP-PAM    | 0.762  | 0.745  | 0.775  | 0.662  | 0.595  |
| FD-UNet    | 0.910  | 0.859  | 0.850  | 0.699  | 0.640  |

| Method     | [4,1] dB | [5,1] dB | [6,1] dB | [7,3] dB | [10,5] dB |
|------------|----------|----------|----------|----------|-----------|
| Bilinear   | 30.68    | 25.63    | 23.58    | 22.95    | 19.49     |
| Bicubic    | 30.98    | 25.43    | 23.28    | 22.65    | 19.06     |
| Lanczos    | 31.01    | 25.35    | 23.18    | 22.50    | 18.93     |
| DIP-PAM    | 26.05    | 25.51    | 25.41    | 23.60    | 22.39     |
| FD-UNet    | 24.73    | 27.98    | 28.02    | 28.20    | 22.93     |

---

**Fig. 10.** Phantom upsampling outputs from DIP-PAM and other methods with [7,3] sampling pattern. Each output is the summation of images acquired with 532- and 590-nm lasers. Middle and bottom rows are close-up images of region 1 and 2, respectively. Red arrows denote jagged and blurry edges from interpolation. Yellow circles indicate checkerboard artifacts from FD U-Net. Scale bar: 1 mm.
Despite not being trained with any fully sampled ground truth data, DIP-PAM manages to correct sparsely sampled images with promising and competitive results. When the targets are vessels, the reconstructed images from DIP-PAM maintain a high level of connectivity and smoothness—key physiological features of vasculature. DIP-PAM prevails over interpolation-based methods that introduce disjointed and blurry textures into vessels, which may jeopardize functional assessments. Our method’s improvement is still preserved when the target is completely different from the typical vessel image input (i.e., our bioprinted phantom). Although artifacts are a concern for not only FD U-Net, but also other pre-trained DL-methods, primarily due to an insufficient amount of training data, DIP-PAM does not result in any spurious artifacts in either in vivo or phantom data.

Statistical testing of SSIM and PSNR data indicate minimal differences in performance between DIP-PAM and FD U-Net. These evaluation matrices also suggest that our method can reconstruct sparsely sampled images at sparsities as low as 2%, corresponding to a 50-fold improvement in scanning time over modern PAM systems. Such a gain without the need for any hardware modification or large, fully sampled training data is significant for not only PAM, but also any other point-by-point scanning-based imaging modalities. In addition, DIP-PAM can easily adjust the binary masks to adapt to any sampling trajectories, such as spiral, Lissajous, or sine-wave patterns. This flexibility is not available for pre-trained methods because they have to retrain every time a new downsampling mask is used.

Nonetheless, our method is not without its downsides. Thorough examination shows that our method’s performance saturates with decreasing sparsity. This saturation worsens when the imaged target is relatively simple with smooth textures, similar to our phantom. This drawback, however, is not of grave concern as most PAM applications focus on high-frequency spatial targets, such as blood vessels and circulating tumor cells. The main disadvantage of this technique lies in the reconstruction time of the iterative optimization. A 300-by-300 (1.5 mm × 1.5 mm) pixel region takes approximately 7.5 minutes, which is infeasible for real-time processing. In addition, with a large number of B-scan images, the total processing time can extend to several hours.

Future work will concentrate on reducing the optimization time and improving the performance of DIP-PAM. Reconstruction time can be decreased by reducing the total number of model parameters, but potentially at the cost of output image quality. This trade-off can be circumvented with transfer learning. For example, we can apply certain upper layers of a pre-trained neural network, such as VGG-Net or FD U-Net, to the deep prior. During DIP-PAM optimization, these layers will be frozen, reducing the number of trained hyperparameters and the subsequent training time. Integrating these pre-learned features will reduce the workload the deep prior needs to perform in optimization and could potentially improve image quality.

Upcoming studies will also expand the applications of DIP-PAM. Depending on the reconstruction time problem, we can flexibly modify the degradation operator, as it is not constrained by the Hadamard product or a binary spatial mask. For example, an optically scanning system with non-uniform sensitivity across the FOV can be integrated into the degradation operator for deep prior restoration. We are also interested in using DIP for PACT to resolve sparse-sampling and limited-view artifacts. Unlike PAM, in which there is fully sampled in vivo data with high image quality, PACT often lacks ground truth due to wide-field imaging of the heterogeneous medium. PACT’s dependence on simulation data limits the capability of DL correction. DIP is a compelling candidate for such applications, in which a ground truth is difficult or nearly impossible to obtain. Thus, it is expected that the success and application of DIP will not be limited to sparsely sampled PAM reconstruction.

V. Conclusion

In this study, we introduce the application of a novel DL method called deep image prior to reconstruct sparsely sampled PAM data. Deep prior represents a DL-based approach that does not require training on a large and diverse set of imaging data with co-registered ground truth. Our work succeeds in recovering fully sampled in vivo data with key physiological features, such as vessel continuity and smoothness. DIP outperforms conventional interpolation and is equivalent to state-of-the-art pre-trained DL methods in terms of performance. Our method can achieve 5- to 72-fold improvements in scanning time over modern laser repetition rate-limited PAM systems and will serve as an important new asset for high-speed PAM functional imaging.

DATA AND CODE AVAILABILITY

All mouse brain microvasculature datasets used for this study are available upon request. The main code used to produce the results in this paper is available on https://github.com/trivul69/deep-prior-pam.

ACKNOWLEDGEMENTS

A portion of our code is adapted from: https://github.com/satoshi-kosugi/DeepImagePrior. The authors thank the code owners of this repository. We also thank Dr. Caroline Connor for editing the manuscript.

REFERENCES

[1] T. Vu, D. Razansky, and J. Yao, “Listening to tissues with new light: recent technological advances in photoacoustic imaging,” Journal of Optics, vol. 21, no. 10, p. 103001, 2019.
[2] S. Jeon, J. Kim, D. Lee, J. W. Baik, and C. Kim, “Review on practical photoacoustic microscopy,” Photoacoustics, vol. 15, p. 100141, 2019.
[3] J. Yao and L. V. Wang, “Photoacoustic microscopy,” Laser & photonics reviews, vol. 7, no. 5, pp. 758-778, 2013.
[4] H. F. Zhang, K. Maslov, G. Stoica, and L. V. Wang, “Functional photoacoustic microscopy for high-resolution and noninvasive in vivo imaging,” Nature biotechnology, vol. 24, no. 7, pp. 848-851, 2006.
[5] B. Xiong et al., “Precise cerebral vascular atlas in stereotactic coordinates of whole mouse brain,” Frontiers in neuroanatomy, vol. 11, p. 128, 2017.
[6] L. V. Wang and J. Yao, “A practical guide to photoacoustic tomography in the life sciences,” Nature methods, vol. 13, no. 8, p. 627, 2016.
[7] J. Kim, J. Y. Kim, S. Jeon, J. W. Baik, S. H. Cho, and C. Kim, “Super-resolution localization photoacoustic microscopy using intrinsic red blood cells as contrast absorbers,” Light: Science & Applications, vol. 8, no. 1, pp. 1-11, 2019.
[8] S. Hu, K. Maslov, and L. V. Wang, “Second-generation optical-resolution photoacoustic microscopy with improved sensitivity and speed,” Optics letters, vol. 36, no. 7, pp. 1134-1136, 2011.
M. Chen et al., "Simultaneous photoacoustic imaging of intravascular and tissue oxygenation," Optics letters, vol. 44, no. 15, pp. 3773-3776, 2019.

R. Ma, S. Söntges, S. Shobham, V. Ntzichristos, and D. Razansky, "Fast scanning coaxial optoacoustic microscopy," Biomedical optics express, vol. 3, no. 7, pp. 1724-1731, 2012.

L. Wang, K. Maslov, J. Yao, B. Rao, and L. V. Wang, "Fast voice-coil scanning optical-resolution photoacoustic microscopy," Optics letters, vol. 36, no. 2, pp. 139-141, 2011.

J. Yao et al., "Ultra-high-speed wide-field photoacoustic microscopy using a polygon mirror scanner (Conference Presentation)," in Photons Plus Ultrasound: Imaging and Sensing 2020, 2020, vol. 11240, p. 11240W-11. International Society for Optics and Photonics.

B. Lan, W. Liu, and J. Yao, "High-speed wide-field photoacoustic microscopy," in Clinical and Translational Biophotonics, 2018, p. JT3a. 68: Optical Society of America.

J. Y. Kim, C. Lee, K. Park, G. Lim, and C. Kim, "Fast optical-resolution photoacoustic microscopy using a 2-axis water-proofing MEMS scanner," Scientific reports, vol. 5, p. 7932, 2015.

J. Yao et al., "Wide-field fast-scanning photoacoustic microscopy based on a water-immersible MEMS scanning mirror," Journal of biomedical optics, vol. 17, no. 8, p. 080505, 2012.

Y. Yuan, S. Yang, and D. Xing, "Optical-resolution photoacoustic microscopy based on two-dimensional spectrometer," Applied Physics Letters, vol. 100, no. 2, p. 023702, 2012.

T. Liu, M. Sun, J. Meng, Z. Wu, Y. Shen, and N. Feng, "Compressive sampling photoacoustic microscope system based on low rank matrix completion," Biomedical Signal Processing and Control, vol. 26, pp. 58-63, 2016.

M. Sun et al., "Photoacoustic imaging method based on arc-direction compressed sensing and multi-antenna observation," Optics express, vol. 19, no. 16, pp. 14801-14806, 2011.

Z. Guo, C. Li, L. Song, and L. V. Wang, "Compressed sensing in photoacoustic tomography in vivo," Journal of biomedical optics, vol. 15, no. 2, p. 021311, 2010.

X. Liu, D. Peng, W. Guo, X. Ma, Y. Yang, and J. Tian, "Compressed sensing photoacoustic imaging based on fast alternating direction algorithm," International journal of biomedical imaging, vol. 2012, 2012.

R.-X. Huang et al., "Dual-wavelength OR-PAM with compressed sensing for cell tracking in a 3D cell culture system," in Photons Plus Ultrasound: Imaging and Sensing 2018, 2018, vol. 10494, p. 104946D. International Society for Optics and Photonics.

K. Jinwali, B. Chinni, V. Dogra, and N. Rao, "Automatic cancer tissue detection using multispectral photoacoustic imaging," International Journal of Computer Assisted Radiology and Surgery, pp. 1-12, 2019.

J. Zhang, B. Chen, M. Zhou, H. Lan, and F. Gao, "Photoacoustic image classification and segmentation of breast cancer: A feasibility study," IEEE Access, vol. 7, pp. 5457-5466, 2018.

A. R. Rajanna, R. Ptucha, S. Sinha, B. Chinni, V. Dogra, and N. A. Rao, "Prostate cancer detection using photoacoustic imaging and deep learning," Electronic Imaging, vol. 2016, no. 15, p. 1-6, 2016.

S. Guan, A. Khan, S. Sildar, and P. V. Chitnis, "Pixel-wise deep learning for improving image reconstruction in photoacoustic tomography," The Journal of the Acoustical Society of America, vol. 145, no. 3, pp. 1811-1811, 2019.

N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," Nature Machine Intelligence, vol. 1, no. 10, pp. 453-460, 2019.

A. Hauptmann et al., "Model-based learning for accelerated, limited-view 3-d photoacoustic tomography," IEEE transactions on medical imaging, vol. 37, no. 6, pp. 1382-1393, 2018.

S. Guan, A. Khan, S. Sildar, and P. Chitnis, "Fully dense UNet for 2D sparse photoacoustic tomography artifact removal," IEEE Journal of biomedical and Health Informatics, vol. 24, no. 2, pp. 568-576, 2019.

T. Yu, M. Li, H. Hunayun, Y. Zhou, and J. Yao, "A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer," Experimental Biology and Medicine, p. 153537020914285, 2020.

A. Reiter and M. A. L. Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in Photons Plus Ultrasound: Imaging and Sensing 2017, 2017, vol. 10064, p. 100643J. International Society for Optics and Photonics.