

# Physics-Guided Lateral and Axial Restoration for Fast Two-Photon Fluorescence Imaging

Zhengyuan Pan

School of Biomedical Engineering  
Shanghai Jiao Tong University  
Shanghai, China  
[panzhengyuan@sjtu.edu.cn](mailto:panzhengyuan@sjtu.edu.cn)

Man Lei

School of Biomedical Engineering  
Shanghai Jiao Tong University  
Shanghai, China  
[leiman@sjtu.edu.cn](mailto:leiman@sjtu.edu.cn)

Bobo Gu\*

School of Biomedical Engineering  
Shanghai Jiao Tong University  
Shanghai, China  
[bobogu@sjtu.edu.cn](mailto:bobogu@sjtu.edu.cn)

**Abstract**—Two-photon fluorescence (TPF) imaging opens a new avenue to achieve high resolution at extended penetration depths, particularly significant for intravital imaging. However, the conventional point-scanning TPF imaging systems face challenges in balancing imaging resolution and speed, making it extremely difficult to simultaneously achieve high resolution and speed. In this work, we develop an innovative deep learning framework of Lateral and Axial Restoration (LAR-Net) to break the limitation posed by trade-off between imaging quality and speed. LAR-Net employs a self-supervised training scheme to computationally restore the under-sampled TPF images to the level of fully-sampled images with axial isotropy, which has not been achieved before, by 4-fold axial and 16-fold lateral scanning speed increase, respectively. By exploiting the image formation model of point-scanning TPF and characteristics of vascular samples, we integrate physics-guided constraints for LAR-Net to restore image content with high fidelity. The simulation studies and experimental results demonstrate the excellent performance of LAR-Net to preserve fine structural features with improved signal-to-noise ratio and structure similarity index. Moreover, the TPF imaging system with the LAR-Net is able to achieve 60-fold improved imaging speed and comparable resolution and SNR as that of fully-sampled TPF system. The outstanding performance makes LAR-Net a potential tool for fast TPF imaging with high resolution.

**Keywords**—two-photon fluorescence microscopy, fast imaging, physics-guided, image restoration

## I. INTRODUCTION

Two-photon fluorescence (TPF) microscopy, which could provide optical sectioning and superior imaging performance in deep turbid tissues with a high resolution [1], has been an indispensable tool for biological structural and functional imaging [2, 3]. However, TPF images are usually captured by point-scanning strategy, which suffers from an innate trade-off problem among spatial resolution, imaging speed, field of view, photo-toxicity, and signal-to-noise-ratio, etc., especially when imaging 3D volumes [4]. Recently, point-scanning speeds have progressively improved, but various in vivo biological dynamics have remained out of reach due to the limited imaging speed, making it urgent and meaningful to further improve TPF imaging speed. Within the past few years, various novel high-speed TPF imaging strategies, including multiplane imaging [5-

7], temporal focusing [8-10], multifocal imaging [11, 12], etc., that overcome the speed limitations of classical TPF imaging modalities, have been proposed and demonstrated. Speed improvement techniques of other imaging modalities also provides insights for TPF imaging [13, 14]. However, these emerging imaging strategies required additional components or complicated optical systems, making them unfeasible for most research labs.

In recent years, rapidly developed deep learning opens a new avenue to boost the performance of optical microscopy, and is widely adopted in numerous optical imaging processing and analysis tasks, such as denoising [15], low-light enhancement [16], domain transformation [17], segmentation [18], and so on. By inferring high-resolution fully sampled images from under-sampled ones, deep-learning super-resolution method saves scanning time with no hardware requirements. Raw images and high-resolution images, which are captured using better optics [19], higher scanning rate [20], or super-resolved microscopy techniques (SIM, STED, etc.) [21, 22], were used to train supervised models by low-resolution/high-resolution pairs. However, it is almost impossible to collect perfectly registered training pairs of moving or living biological samples. To overcome these drawbacks, various emerging supervised models were proposed. A two-stage super-resolution and denoising pipeline based on conditional generative adversarial network (cGAN) was developed for two-photon 2D fiberscopy [23]. Ex vivo and in vivo low-resolution images were collected for the two stages respectively, and high-resolution training targets were generated computationally. ResNet CNN was proposed to restore under-sampled 2D point-scanning electron microscopy and TPF images [24, 25]. Low-resolution training images were generated by down-sampling and adding noise in high-resolution images. Modified neural network architecture was used to process 3D volumes acquired by scanning confocal microscopy and TPF, achieving a 64-fold theoretical improvement of image resolution [26, 27]. However, these methods mentioned above did not address the inherent issues of resolution anisotropy in TPF and other point-scanning microscopies. Resolution anisotropy is induced by axially elongated point spread function (PSF), resulting in lower axial resolution than lateral resolution [28]. A U-Net based supervised solution CARE was proposed to improve axial resolution by

about 10-fold, but the process for collecting paired images for training is cumbersome [29]. Self-supervised frameworks based on CycleGAN were proposed to restore 3D resolution isotropy [30, 31]. However, to our best knowledge, there is still a gap in combining 3D volume super-sampling and isotropic resolution restoration for speeding up TPF imaging.

In this work, we proposed a self-supervised framework of lateral and axial resolution restoration (LAR-Net) to restore the resolution of under-sampled TPF images and consequently enable high-speed TPF imaging. The traditional TPF imaging captures high-resolution images by fully-sampled point-scanning but sacrifices imaging speed. LAR-Net only requires non-paired under-sampled TPF images to restore the high-resolution images with significantly expedited imaging processes. In order to maintain a high image fidelity and minimize artifacts, the cascaded training scheme with a pixel substitution mechanism and CycleGAN model with saliency constraints are used in LAR-Net, respectively. In addition, to further preserve tubular-like structures, we impose a tubular constraint based-on gradient index and tubular content index during training. The superior restoration performance is demonstrated by synthetic simulation and experimental studies. After restoring the under-sampled TPF images with LAR-Net, the restored TPF images with improved resolution, structure similarity and less artifacts are achieved. Moreover, the theoretical TPF imaging speed is increased by over 60-fold.

## II. METHODS

### A. Animal and Surgery

The mouse was pre-anesthetized with isoflurane (5% inhalation, mixed with fresh air, 0.5 L min<sup>-1</sup>) and placed in a stereotactic frame. Meanwhile isoflurane was maintained at 1-2% throughout the surgical procedure, and body temperature was maintained at 37 °C with a heating pad. Fascia and connective tissue on the skull were gently removed with forceps and sterile wet cotton tips to avoid any internal bleeding inside the brain. A dental drill exposed a craniotomy (about 4 mm in diameter) in the right sensory cortex. The dura was carefully removed and artificial cerebrospinal fluid (ACSF) filled the exposed cortex. A coverslip with a diameter of 5 mm was pressed closely, and medical glue was used to seal its edge. Dental cement was used to create a water-immersed environment. After injecting the fluorescent dyes (5 mg kg<sup>-1</sup>) to label cerebral blood vessels through the tail vein, the mouse was prepared for the TPF imaging.

### B. TPF Microscope Setup and Image Acquisition

A commercial laser scanning microscopy system (OLYMPUS FLUOVIEW FVMPE-RS, Olympus Inc.) with a 20X objective lens (XLUMPLFLN20XW, 20X, 1.0-NA, water immersion, Olympus Inc.) was used for TPF imaging. A tunable Ti:Sapphire laser (Chameleon Ultra II, Coherent Inc.) was used for two-photon excitation of fluorescence dye. A continuously variable neutral density filter (NDC-50C-4M-B, Thorlabs Inc.) was used to adjust the excitation power for biological samples. A beam expander composed of a pair of convex lenses (LB1757-B, f = 30 mm; LB1676-B, f = 100 mm; Thorlabs Inc.) was used to expand the laser beam from 1 mm to 3.3 mm. Then, the laser beam was directly introduced into the microscope. Emitted

fluorescence was collected by the GaAsP photomultiplier tubes (PMTs) (FV30-AGAPD, Olympus Inc.) with a filter for red channel (610/35 nm) imaging. The excitation power was measured after the objective lens using a power meter. The sampling rate in lateral (X and Y) direction is 1.243  $\mu\text{m}/\text{px}$ , and in axial (Z) direction is 2  $\mu\text{m}/\text{slice}$ .

### C. Image Pre-Processing and Post-Processing

All volumes acquired from the microscopy system were in TIF format, which size was 512\*512\*N<sub>Z</sub> px (N<sub>Z</sub> is the number of slices). It indicated that the physical volume of the sample was 636.4\*636.4\*(N<sub>Z</sub>\*2)  $\mu\text{m}^3$ . The bit depth of each lateral slice was 24 bits with 8 bits per channel. They were normalized using percentile-based saturation with the bottom and top 0.25% for the simulated volumes and 0.03% for the experimental volumes.

Each lateral 2D slice was under-sampled by a factor of 4 using bilinear interpolation, i.e., every fourth pixel along both X and Y dimensions was reserved and reassigned. The corresponding 512\*512 under-sampling masks were also generated by setting every fourth pixel along both X and Y dimensions to 0 and the rest pixels to 256. Each input image (512\*512\*2) to the later restoration model for both training and inference was created by concatenating an under-sampled slice and an under-sampling mask. Before training, the input and ground truth images were divided 64\*64 laterally.

Image visualization was performed using Imaris software and ImageJ.

### D. Stage#1 Network: Lateral and Axial Content Restoration

The basic block of Enhanced Deep Super-Resolution (EDSR) network is modified residual block. Compared to the original residual block, batch normalization and the ReLU activation after addition at the end of the block were removed to save memory and increase efficiency [32]. The number of residual blocks is 32, and the number of filters is 256. To process volumetric input, the dimensions of all layers are adapted for 3D volume data.

### E. Stage#2 Network: Axial Isotropy Restoration

The generator and discriminator of axial restoration network adopted the architecture of the original CycleGAN. The generator network G<sub>A</sub> and G<sub>B</sub> were constructed in 3D U-Net architecture, which consisted of a down-sampling path and a up-sampling path. The padded convolutional layers were used to keep the image size unchanged when passing through convolutional layers. All the down-sampling and up-sampling layers were implemented with stride convolution with trainable parameters. All 3D convolution layers had a kernel size of 3, stride size of 1, and padding size of 1. All transposed convolution layers had a kernel size of 2, stride size of 2, and no padding. The discriminators, which adopted the PatchGAN classifier, had convolution layers with a kernel size of 4, stride size of 2, and padding size of 1. The discriminators estimated the statistical distance between the distribution in isotropic resolution domain and that in anisotropic resolution domain. The statistical distance of the forward path was defined as the sum of the statistical distance in the XY, YZ, and XZ planes.

### F. Physics-Guided Constraints and Loss Design

Since the foreground structures differed from the background noise and defocused signal in fluorescence imaging, so the saliency constraint was applied to improve the restoration of meaningful structures. Using an adaptive thresholding segmentation technique, the saliency maps of semantic structures could be extracted. The saliency constraint maintained the consistency of the image content and avoided image distortions:

$$L_{SC}(X, Y) = E_{X \sim p_{data}}(X) \left[ \|T_\alpha(X) - T_\beta(G(X))\|_1 \right] + E_{Y \sim p_{data}}(Y) \left[ \|T_\beta(Y) - T_\alpha(F(Y))\|_1 \right] \quad (1)$$

where  $T_\alpha$  and  $T_\beta$  were segmentation operators with thresholds of  $\alpha$  and  $\beta$ , which were set with three steps: 1. select a background region and average all pixel values in that region; 2. select a foreground region and average all pixel values in that region; 3. set the threshold as the average of the average background and foreground intensities. Then Otsu thresholding was performed to produce saliency masks.

A major component of the image content are tubular structures like blood vessels and axons, so we introduce an additional constraint to better preserve tubular features. Specifically, two indices are calculated: gradient index (GI) and tubular content index (TCI). GI measures the magnitude and orientation consistency of intensity gradients along the radial around a tubular centerline. It reflects how sharply the image intensity drops off perpendicular to the tubular structure, defined as,

$$GI = \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M |\nabla I(p_i + r \cdot r_{i,j}) \cdot r_{i,j}| \quad (2)$$

where  $N$  is the number of sampled points along the centerline,  $M$  is the number of radial directions per point,  $\nabla I$  is 3D gradient vector,  $p_i$  is centerline point  $i$ ,  $r_{i,j}$  is unit radial direction  $j$  orthogonal to the tangent at  $p_i$ , and  $r$  is the radial sampling distance.

Tubular content index measures the tubular structure based on local gradient matrix, defined as,

$$Q = s_1 \cdot s_2 \cdot \frac{s_1 - s_3}{s_1 + s_3} \quad (3)$$

where  $s_1$ ,  $s_2$ , and  $s_3$  are the singular values of local gradient matrices, and  $s_1 \geq s_2 \geq s_3$ . Both GI and TCI are summed with weights as tubular structure maps for training input and target volumes to calculate tubular structure loss.

$$L_{TS}(X, Y) = \alpha L_{GI}(X, Y) + \beta L_{TCI}(X, Y) \quad (4)$$

where  $\alpha$  and  $\beta$  are weights,  $L_{GI}$  is the L1 difference between gradient maps of input  $X$  and target  $Y$ , and  $L_{TCI}$  is the L1 difference between tubular content maps of input  $X$  and target  $Y$ .

The total losses for stage#1 network and stage#2 network are defined as the following:

$$L_{stage\#1 total}(X, Y) = L_1(X, Y) + \lambda_1 L_{TS}(X, Y) + \lambda_2 L_{SC}(X, Y) \quad (5)$$

where  $\lambda_1$  and  $\lambda_2$  are weights for loss terms,  $L_1$  is the L1 loss between  $X$  and  $Y$ , and

$$L_{stage\#2 total}(X, Y) = L_{GAN}(X) + L_{GAN}(Y) + \lambda [L_{cycle}(X, Y) + \rho_1 L_{TS}(X, Y) + \rho_2 L_{SC}(X, Y)] \quad (6)$$

Where  $\lambda$ ,  $\rho_1$ , and  $\rho_2$  are weights of loss terms,  $L_{GAN}$  is the adversarial loss, and  $L_{cycle}$  is the cycle-consistency loss.

### G. Evaluation Metrics

The restoration accuracy of the trained networks was evaluated using peak signal-to-noise ratio (PSNR) and structure similarity index (SSIM). For analysis of resolution improvement, ImageJ is used to plot the pixel intensity profiles, apply Gaussian fitting, and full-width half-maximum (FWHM) calculation.

### H. Hardware for Computation

The graphics hardware used for training and inference is NVIDIA GeForce RTX 3080. The processing unit is Intel® Xeon® CPU E5-2673 v4 @ 2.30GHz, with a RAM of 128 GB.

## III. RESULTS

### A. Restoration Framework of Under-Sampled Volumes

In the lateral and axial restoration stage (Fig.1), all under-sampled volumes are transformed into fully-sampled slices by the network. It is trained in a cascaded manner. For cascade stage  $T$ , the network takes under-sampled volume  $L$  ( $T=1$ ) or the combined volume  $\hat{B}$  from stage  $T-1$  ( $T>1$ ) as input, and generates a restored fully-sampled volume. When performing training, the fully sampled volume  $I$  is convolved with the 3D point spread function (PSF) of the imaging system to generate a blurred volume  $B$ . Pixels, which are both predicted by the model (at locations  $x \neq ds$ ) and in the digitally up-sampled volume  $D^T L$  (at locations  $x = ds$ ), are preserved in the combined volume  $\hat{B}$ . In the end, the volume  $\hat{B}$  served as the input of the next cascade stage to the network.

In the axial isotropy restoration stage (Fig.2), the volume is transformed by a CycleGAN network. As shown in CycleGAN scheme, two generator networks  $G_A$  and  $G_B$  learn the resolution enhancement and resolution degradation, respectively. The resolution enhancement generator  $G_A$  transformed the 3D

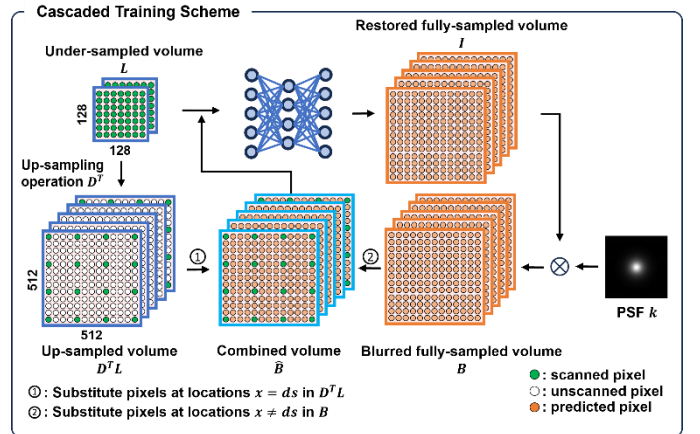


Fig. 1. Cascade training of lateral and axial content restoration network.

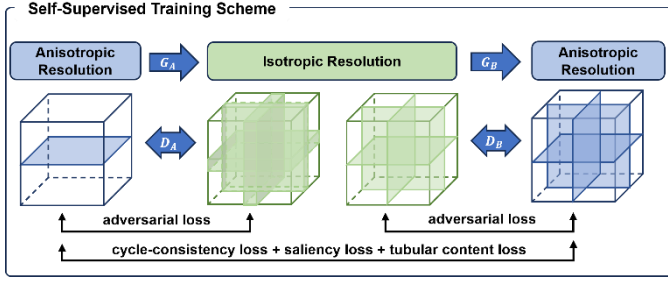


Fig. 2. Self-supervised training of axial isotropy restoration network. volume from anisotropic domain to isotropic domain, and the discriminative network  $D_A$  compared the 2D maximum intensity projection images from the generated 3D volume with 2D lateral images from the real 3D volume. On the other hand, the generator network  $G_B$  transforms the 3D volume from isotropic domain to anisotropic domain, and the discriminative network  $D_B$  compares 2D images from the generated 3D volume with 2D images from the real 3D volume. The cycle-consistency loss guarantees that the resolution enhancement and resolution degradation processes are reciprocal mappings of each other, while the saliency loss and tubular content loss further reduce artifacts and preserve content.

### B. Restoration Performance on in vivo TPF Volumes of Mouse Cerebrovasculature

To test the effectiveness of the proposed restoration network, in vivo TPF volumes of mouse cerebrovascular are captured. The lateral dimension of the captured volumes is  $512 \times 512$  px with a sampling rate of  $1.243 \mu\text{m}/\text{px}$ , and the axial dimension varies between 300 to 600 px with a sampling rate of  $2 \mu\text{m}/\text{px}$ . Fifteen 3D volumes are collected from different mice with the same capture setting. The training data for the lateral restoration model is generated by extracting and applying 4-fold under-sampling to lateral 2D slices of each 3D volume. The sampling rate of the under-sampled images is  $4.972 \mu\text{m}/\text{px}$ . Meanwhile, each 3D volume is divided into smaller patches of  $64 \times 64 \times 64$ . After discarding the low-contrast and noisy patches, the remaining high-quality patches are used for cascaded training of the lateral and axial content restoration model. And the number of cascaded stages is set to 3 to balance the effectiveness and the speed.

A combined 3D volume consisting of under-sampled input, LAR-Net output, and GT is displayed in Fig.3a. A comparison of a 2D XY and XZ MIP is shown in Fig.3b-c. The image restoration quality and axial resolution improvement is quantified with PSNR, SSIM, FWHM (Fig.3d-i). The fine structures, which are indistinguishable in the under-sampled volume, can be recognized after axial resolution enhancement. The SSIM of restored images is not improved as compared with that of under-sampled input, it could be ascribed to the fact that the axial resolution of the original 3D volume is lower than that of the restored 3D volume, resulting in thicker vascular structure and different appearance in terms of SSIM. The Gaussian-fitted pixel intensity profiles along the orange lines in Fig.3i exhibit 3 times axial resolution improvement by LAR-Net. The enhanced resolution does not mean the network altered the vascular structure, and Fig.3e-f suggest that structures in XY planes maintain a high similarity in LAR-Net output and GT. Moreover, the proposed algorithm has much higher PSNR and SSIM than

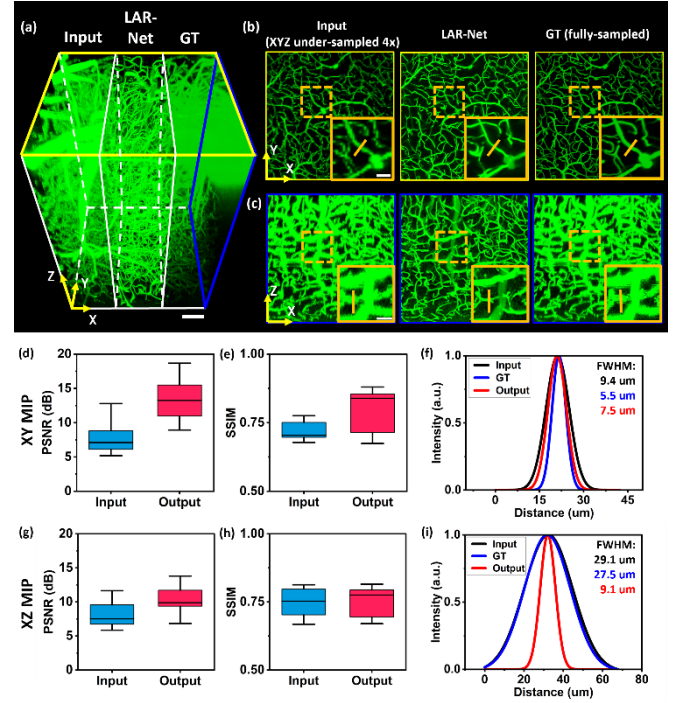


Fig. 3. Restoration of in vivo TPF images of mouse cerebral blood vessels with enhanced resolution. a) Combined 3D volume of 64x under-sampled input, LAR-Net output, and ground truth. Scale bar:  $100 \mu\text{m}$ . b) Comparison of 2D XY MIP. Scale bar:  $30 \mu\text{m}$ . c) Comparison of 2D XZ MIP. Scale bar:  $30 \mu\text{m}$ . d) PSNR and e) SSIM of 2D XY MIPs,  $n=32$ . f) Gaussian fitting of orange line intensity profile in b), and the FWHM values of the corresponding curve. g) PSNR and h) SSIM of 2D XZ MIPs,  $n=32$ . i) Gaussian fitting of orange line intensity profile in c), and the FWHM values of the corresponding curve. Excitation:  $960 \text{ nm}$ , Power =  $40 \text{ mW}$ . the classic Bicubic, and less PSNR and SSIM than CARE and other axial resolution restoration methods (Table I). It was worth noting that CARE requires supervised training and strictly paired and registered data, making it significantly difficult to deploy.

### C. Restoration Performance on in vivo TPF Volumes of Mouse Neurons

To demonstrate the generalization capacity of the model for different biological samples, the fluorescence images of the cortical region of the Thy 1-eYFP mouse brain are used to train and test the 3D restoration model. The mouse brain volumes are collected with  $3\text{-}\mu\text{m}$  interval along axial dimension [29]. The volume is first digitally under-sampled by a factor of 4 along X, Y, and Z dimensions, and then passed into LAR-Net. A combined 3D volume consisting of the under-sampled input and the restored output is shown in Fig.4a. Zoomed-in regions in 2D XY MIP and 2D XZ MIP show that axon and soma structures and anatomical shapes are maintained and better resolved with improved contrast across the whole volume using LAR-Net (Fig.4b-c). A comparison of segmented 2D MIPs is shown in Figure S6b in Supporting Information. And the restored images also improved the PSNR significantly and SSIM slightly as compared with the under-sampled images (Fig.4d-e). Soma and axon structures exhibit highly identical profile in XY plane (Fig.4f), while less blurry along axial direction in the restored volume, and exhibit around twice higher axial resolution



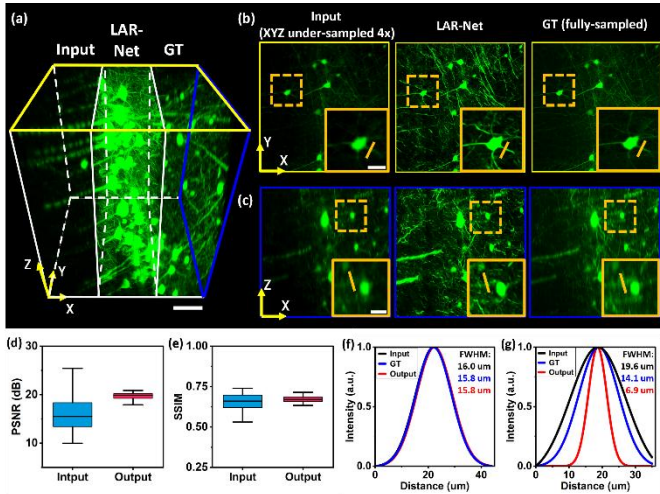


Fig. 4. Restoration performance of in vivo TPF images of mouse cortical neurons. a) Combined 3D volume of 64x under-sampled input, LARR output, and ground truth. Scale bar: 100  $\mu\text{m}$ . b) Comparison of 2D XY MIP. Scale bar: 30  $\mu\text{m}$ . c) Comparison of 2D XZ MIP. Scale bar: 30  $\mu\text{m}$ . d) PSNR and e) SSIM of 3D XY slices,  $n=600$ . f) Gaussian fitting of orange line profile in b), and the FWHM values of the corresponding curve. g) PSNR and h) SSIM of 2D XZ MIPs,  $n=32$ . i) Gaussian fitting of orange line profile in c), and the FWHM values of the corresponding curve. (Fig.4g). These results indicated that the model had an good generalization capacity for different types of biological tissue.

#### D. Imaging Speed Improvement

Considering that the under-sampled TPF images could be restored with excellent PSNR and SSIM, the proposed LARR could efficiently accelerate the point-scanning processes. The theoretical imaging speed could be improved by 64 times by LAR-Net. Due to mechanical actuation and imaging software processing time, the actual enhancement would be slightly lower than the theoretical one. The imaging time for the typical setting (251 slices with 2  $\mu\text{m}$  interval) is about 292 s, and could be shortened to 63 s with LAR-Net (Table II). More comparison with different settings also demonstrated the effectiveness of LAR-Net.

### IV. DISCUSSION

Despite the rapid development and huge success of deep learning for optical imaging, the high cost of computational resources and the demand for large amount of training data remain practical challenges. In this work, we develop a deep learning framework of LAR-Net to improve the lateral and axial resolution of under-sampled TPF images without ground truth data, resolving the contradiction between imaging resolution and speed. The proposed LAR-Net does not require perfectly registered high-resolution images, making it feasible in resource-limited settings e.g., pathological sample evaluation of rare diseases. Meanwhile, LAR-Net employs physics-guided priors, including image formation process, saliency information, and tubular structure characteristics, as constraints to limit the solution space, thereby reducing artifacts and preserve content structures. LAR-Net can be applied to any other commercial or custom-built scanning-based imaging techniques, such as confocal microscopy, without the need to modify the optical hardware. The design of two separate stages was adopted due

to the consideration of computational demands and task specialization. Training a single network to handle both 3D super-sampling and axial isotropic resolution restoration has to balance potentially competing demands, resulting in suboptimal performance.

LARR exhibits excellent performance in TPF imaging, but it has several limitations:

I), the deep-learning-based image restoration is an ill-posed problem under all circumstances. The output generated by a deep learning model is at best a statistical inference of the training data, which can infinitely approximate the ground truth but never 100% match it. The image formation process constraint in this work is just a way to mitigate the flaw instead of completely overcoming it.

II), the restoration capability of LAR-Net has a limitation. The maximum under-sampling rate is 4 for any dimension. If the under-sampling rate increases beyond 4, network hallucination increases drastically, and the fine structures with a width smaller than 3-5 pixels will be lost.

III), the generalizability of LAR-Net is limited to samples with similar structures. However, this limitation can be overcome by transfer learning with data from the target domain or more parameters of the network.

Further increase of speed is a huge challenge using computational approach alone due to the limitation imposed by the Nyquist theorem. However, it is possible to break the speed limitation while maintaining a decent degree of ease of use by introducing prior information or small hardware changes.

LAR-Net has the potential of serving as part of the pre-processing step for various down-stream tasks, including vascular analysis and neuron tracing. While the performance for these tasks is not directly demonstrated in this work, due to the fact that most of the down-stream tasks take segmented images as input, the evaluation on the segmented images serves as an indirect proof that the output of LAR-Net achieves a similar result as the ground truth. It was worth noting that the image formation model and saliency constraints can also be integrated to other model architectures for imaging performance enhancement with content-preserving capability.

### V. CONCLUSION

In summary, this study presents LAR-Net, a physics-guided self-supervising super-sampling post-processing framework for TPF imaging, to increase the imaging speed by over 60-fold and maintain the resolution and image content as compared with that achieved using conventional system. The restoration using mouse vasculature and neuron volumes shows that the proposed method effectively maintains the fidelity of the restored image content. The distinguished features, i.e., no any requirement on ground truth data collection for training, hardware modification, and prior knowledge of the sample, make LAR-Net a promising post-processing tool for fast TPF imaging.

TABLE I QUANTITATIVE EVALUATIONS OF RESTORATION METHODS ON CEREBROVASCULATURE VOLUMES

Methods	PSNR	SSIM
<i>Bicubic</i>	6.8	0.53
<i>CARE</i>	13.1	0.75
<i>LAR-Net</i>	10.8	0.71

TABLE II. IMAGING TIME COMPARISON BETWEEN THE FULLY-SAMPLED IMAGES AND UNDER-SAMPLED IMAGES WITH LAR-NET

Number of Scanned Slices ( $z = 500 \mu\text{m}$ )	Lateral Image Size	Total Imaging Time	Theoretical Speed Increase
<i>63 slices (8-um interval)</i>	128*128 px	16 sec	64 times
	512*512 px	74 sec	4 times
<i>126 slices (4-um interval)</i>	128*128 px	32 sec	32 times
	512*512 px	147 sec	2 times
<i>251 slices (2-um interval)</i>	128*128 px	63 sec	16 times
	512*512 px	292 sec	1 times

#### ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Program of China (2021YFF0502900); the National Natural Science Foundation of China (62375169); Shanghai Jiao Tong University (YG2022ZD005, YG2024QNA06).

#### REFERENCES

- [1] F. Helmchen, and W. Denk, "Deep tissue two-photon microscopy," *Nat. Methods*, 2, 932–940, 2005.
- [2] L. Lin, J. Liu, Z. Pan, W. Pang, X. Jiang, M. Lei, et al., "General post-regulation strategy of AIEgens' photophysical properties for intravital two-photon fluorescence imaging," *Adv. Sci.* 2404792, 2024.
- [3] N. Song, and B. Gu, "In vivo and in situ robust quantitative optical biopsy of hepatocellular carcinoma and metastasis based on two-photon autofluorescence imaging," *Opt. Lett.*, 49, 4054–4057, 2024.
- [4] P. Luu, S. E. Fraser, and F. Schneider, "More than double the fun with two-photon excitation microscopy," *Commun. Biol.*, 7, 364, 2024.
- [5] D. R. Beaulieu, I. G. Davison, K. Kılıç, T. G. Bifano, J. Mertz, "Simultaneous multiplane imaging with reverberation two-photon microscopy," *Nat. Methods*, 17, 283–286, 2020.
- [6] T. D. Weber, M. V. Moya, K. Kılıç, J. Mertz, M. N. Economu, "High-speed multiplane confocal microscopy for voltage imaging in densely labeled neuronal populations," *Nat. Neurosci.*, 26, 1642–1650, 2023.
- [7] S. Geissbuehler, A. Sharipov, A. Godinat, N. L. Bocchio, P. A. Sandoz, A. Huss, et al., "Live-cell multiplane three-dimensional super-resolution optical fluctuation imaging," *Nat. Commun.*, 5, 5830, 2014.
- [8] E. Papagiakoumou, E. Ronzitti, and V. Emiliani, "Scanless two-photon excitation with temporal focusing," *Nat. Methods*, 17, 571–581, 2020.
- [9] C. Y. Chang, C. Y. Lin, Y. Y. Hu, S. F. Tsai, F. C. Hsu, and S. J. Chen, "Temporal focusing multiphoton microscopy with optimized parallel multiline scanning for fast biotissue imaging," *J. Biomed. Opt.*, 26, 016501, 2021.
- [10] Y. Hu, C. Hsu, Y. Tseng, C. Lin, H. Chiang, A. Chiang, et al., "Temporal focusing multiphoton microscopy with cross-modality multi-stage 3d unet for fast and clear bioimaging," *Biomed. Opt. Express*, 14, 2478–2491, 2023.
- [11] J. Bewersdorf, R. Pick, and S. W. Hell, "Multifocal multiphoton microscopy," *Opt. Lett.*, 23, 655–657, 1998.
- [12] J. N. Hansen, A. Gong, D. Wachten, R. Pascal, A. Turpin, J. F. Jikeli, et al., "Multifocal imaging for precise, label-free tracking of fast biological processes in 3D," *Nat. Commun.*, 12, 4574, 2021.
- [13] Z. Wang, T. Zhao, Y. Cai, J. Zhang, H. Hao, Y. Liang, et al., "Rapid, artifact-reduced, image reconstruction for super-resolution structured illumination microscopy," *The Innovation*, 4(3), 100425, 2023.
- [14] Z. Wang, T. Zhao, H. Hao, Y. Cai, K. Feng, X. Yun, et al., "High-speed image reconstruction for optically sectioned, super-resolution structured illumination microscopy," *Adv. Photon.*, 4(2), 026003, 2022.
- [15] X. Li, G. Zhang, J. Wu, Y. Zhang, Z. Zhao, X. Lin, et al., "Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising," *Nat. Methods*, 18, 1395–1400, 2021.
- [16] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, et al., "Zero-reference deep curve estimation for low-light image enhancement," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1780–1789, 2020.
- [17] S. Sun, H. Shi, and Y. Wu, "A survey of multi-source domain adaptation," *Inf. Fusion*, 24, 84–92, 2015.
- [18] E. Shelhamer, J. Long and T. Darrell. "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 39, 640–651, 2017.
- [19] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, 4, 1437–1443, 2017.
- [20] S. McAleer, A. Fast, Y. Xue, M. J. Seiler, W. C. Tang, M. Balu, et al., "Deep learning–assisted multiphoton microscopy to reduce light exposure and expedite imaging in tissues with high and low light sensitivity," *Trans. Vis. Sci. Tech.*, 10, 30, 2021.
- [21] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Gunaydin, et al., "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods*, 16, 103–110, 2019.
- [22] J. Chen, H. Sasaki, H. Lai, Y. Su, J. Liu, Y. Wu, et al., "Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes," *Nat. Methods*, 18, 678–687, 2021.
- [23] H. Guan, D. Li, H. Park, A. Li, Y. Yue, Y. A. Gau, et al., "Deep-learning two-photon fiberscopy for video-rate brain imaging in freely-behaving mice," *Nat. Commun.*, 13, 1534, 2022.
- [24] L. Fang, F. Monroe, S. W. Novak, L. Kirk, C. R. Schiavon, S. B. Yu, et al., "Deep learning-based point-scanning super-resolution imaging," *Nat. Methods*, 18, 406–416, 2021.
- [25] A. Zhou, S. A. Mihelcic, S. A. Engelmann, A. Tomar, A. K. Dunn, and V. M. Narasimhan, "A deep learning approach for improving two-photon vascular imaging speeds," *Bioengineering*, 11, 111, 2024.
- [26] H. Zhang, Y. Zhao, C. Fang, G. Li, M. Zhang, Y.-H. Zhang, and P. Fei, "Exceeding the limits of 3D fluorescence microscopy using a dual-stage-processing network," *Optica*, 7, 1627–1640, 2020.
- [27] Y. He, J. Yao, L. Liu, Y. Gao, J. Yu, S. Ye, et al., "Self-supervised deep-learning two-photon microscopy," *Photon. Res.*, 11, 1–11, 2023.
- [28] C. Ibáñez-López, G. Saavedra, G. Boyer, and M. Martínez-Corral, "Quasi-isotropic 3-D resolution in two-photon scanning microscopy," *Opt. Express*, 13, 6168–6174, 2005.
- [29] M. Weigert, U. Schmidt, T. Boothe, A. Muller, A. Dibrov, A. Jain, et al., "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nat. Methods*, 15, 1090–1097, 2018.
- [30] X. Li, G. Zhang, H. Qiao, F. Bao, Y. Deng, J. Wu, et al., "Unsupervised content-preserving transformation for optical microscopy," *Light Sci. Appl.*, 10, 44, 2021.
- [31] H. Park, M. Na, B. Kim, S. Park, K. H. Kim, S. Chang, et al., "Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy," *Nat. Commun.*, 13, 3297, 2022.
- [32] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," 2nd NTIRE: New Trends in Image Restoration and Enhancement Workshop and Challenge on Image Super-Resolution in Conjunction with CVPR, 2017.