

PHOTONICS Research

Self-supervised denoising for enhanced volumetric reconstruction and signal interpretation in two-photon microscopy

JIE LI,^{1,2}  LIANGPENG WEI,³ AND XIN ZHAO^{1,2,*} 

¹National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, Engineering Research Center of Trusted Behavior Intelligence, Ministry of Education, Tianjin Key Laboratory of Intelligent Robotic (tikLIR), Institute of Robotics & Automatic Information System (IRAS), Nankai University, Tianjin 300350, China

²Shenzhen Research Institute of Nankai University, Shenzhen 518083, China

³School of Biomedical Engineering and Technology, Tianjin Medical University, Tianjin 300070, China

*Corresponding author: zhaoxin@nankai.edu.cn

Received 31 March 2025; revised 7 May 2025; accepted 22 May 2025; posted 27 May 2025 (Doc. ID 563812); published 1 August 2025

Volumetric imaging is increasingly in demand for its precision in statistically visualizing and analyzing the intricacies of biological phenomena. To visualize the intricate details of these minute structures and facilitate the analysis in biomedical research, high-signal-to-noise ratio (SNR) images are indispensable. However, the inevitable noise presents a significant barrier to imaging qualities. Here, we propose SelfMirror, a self-supervised deep-learning denoising method for volumetric image reconstruction. SelfMirror is developed based on the insight that the variation of biological structure is continuous and smooth; when the sampling interval in volumetric imaging is sufficiently small, the similarity of neighboring slices in terms of the spatial structure becomes apparent. Such similarity can be used to train our proposed network to revive the signals and suppress the noise accurately. The denoising performance of SelfMirror exhibits remarkable robustness and fidelity even in extremely low-SNR conditions. We demonstrate the broad applicability of SelfMirror on multiple imaging modalities, including two-photon microscopy, confocal microscopy, expansion microscopy, computed tomography, and 3D electron microscopy. This versatility extends from single neuron cells to tissues and organs, highlighting SelfMirror's potential for integration into diverse imaging and analysis pipelines. © 2025

Chinese Laser Press

<https://doi.org/10.1364/PRJ.563812>

1. INTRODUCTION

The advancements in two-photon fluorescent microscopy technology and fluorescent probes have enabled scientists to delve from two-dimensional imaging to more complex volumetric visualization. This progression has allowed researchers to capture detailed three-dimensional representations of biological samples, providing unparalleled insights into cellular and subcellular structures and processes [1–3]. To capture these intricate detail structures for subsequent biological analysis, a high-signal-to-noise ratio (SNR) image containing sufficient information about the sample is necessary. However, poor SNR imaging caused by the paucity of a fluorescent photon budget is the main challenge of fluorescent microscopy. That is, fluorescent imaging suffers from the dilemma among image SNR, illumination intensity, and imaging speed.

Although higher excitation power can directly improve fluorescent photon harvest, this comes with adverse effects such as concurrent photobleaching, phototoxicity, and tissue heating,

which can be detrimental to the health of the sample, skew experimental outcomes, and hinder the understanding of biological phenomena [4–6]. Another plausible strategy is to utilize more advanced fluorescent probes and detection techniques with higher quantum yields and improved photostability. Still, the usage conditions for these techniques are more demanding [7,8]. Apart from the above, fluorescent imaging still suffers from photon shot noise [9–11], which comes from the inevitable stochasticity of photon detection and electronic noise, leading to increased measurement uncertainty and imposing limitations on imaging resolution, speed, and sensitivity.

Extensive research efforts have been investigated to remove noise from fluorescent imaging; a “hard” approach is the development of high-performance fluorophores [7,12,13], as well as advancements in excitation and detection physics [10,14]. A softer way is with the assistance of a post-hoc computation aid algorithm. The conventional denoising approach attempts to use linear filters to model the noises and split them from signals [15,16]; however, these methods often fall short in

performance and applicability due to the inherent challenge of accurately characterizing the signal and noise distributions, which are rarely known in practice. Recently, the data-driven denoising algorithm, especially a deep-learning-based method, demonstrated promising performance in reconstructing signals from noisy images [17–23], and has been applied in fluorescent imaging [24–27]. These non-linear supervised learning methods have great potential to learn any arbitrary distribution of signals and noises, resulting in promising denoising outcomes with reduced variance. This potential is rooted in the utilization of extensive training data consisting of noisy-clean pairs of the same samples and events, allowing the network to learn the intricate mappings from noisy to clean images. Yet, the inherent presence of noise in fluorescent imaging makes the acquisition of corresponding noisy-clean image pairs impractical for denoising fluorescent imaging data. This limitation necessitates the development of alternative strategies for denoising that do not rely on the noisy-clean image pairs datasets.

Lately, self-supervised or unsupervised learning techniques [28–32] have emerged as a promising solution to remove the obstacle. Unlike the supervised learning method, which heavily relies on ground truth labels for supervision, the learning process in self-supervised methods only uses the noise data for denoising. One strategy is to extend denoising into the temporal domain by imaging the same signal at two different times and to train the denoiser by mapping these noisy pairs of images [28,31]. However, these methods are not practical for denoising fluorescent imaging data, especially for volumetric data, because data collection is expensive and time-consuming. Besides, these methods exhibit reduced generalizability to data that deviate significantly from the training distribution. An alternative denoising strategy involves training the data solely on themselves, eschewing any external information or knowledge about the noise's distribution or variance within the images [29,30,33–36]. These stem from that single images contain substantial internal data redundancy and provide useful priors [37]. Noise2Void [29] and Noise2Self [30] are two of these denoisers by training a neural network to predict the value of a pixel from its surrounding neighborhood. Further, Neighbor2Neighbor [34] and Noise2Fast [35] introduced downsampling strategies to generate synthetic training pairs from single noisy images for denoising. However, the execution time required for these methods, even for processing 2D images, can be substantial. The high-dimensional nature of 3D data significantly increases the computational complexity and memory requirements, making them impractical for use in fluorescent volumetric imaging data.

To this end, we propose SelfMirror, a deep self-supervised denoising method for fluorescent volumetric imaging data without any high-SNR observation for training. SelfMirror aims to capture the inherent patterns and structures from the low-SNR data and effectively denoise it. The inner assumption of our method is that the continuous and smooth characteristics of biological signals imply that neighboring slices within the volumetric data share common spatial structures. This indicates that the patterns and structures observed in one slice are likely to be consistent and recognizable in adjacent

slices. Based on the above assumption, we proposed a deep-learning network to fully utilize the intrinsic spatial information within volumetric imaging data. The intrinsic characteristics of SelfMirror make it particularly well-suited for denoising volumetric imaging using only spatial information, without the need to acquire temporal redundancy information as in Noise2Noise, and avoiding the necessity of masking information and computational resources demanding as in Noise2Void. Additionally, we introduce a feature maps learning mechanism to further optimize the denoising performance. By jointly learning the feature maps derived from the dual-network architecture, this approach effectively reduces the predictive variance of SelfMirror, thereby significantly enhancing the overall denoising efficiency.

Furthermore, our approach involves minimal assumptions regarding the noise statistics of the volumetric data and does not necessitate an explicit noise model. Consequently, SelfMirror is adaptable to various noise levels and different imaging modalities. We quantitatively evaluated our method on both simulated and diverse experimental data. The results demonstrate that SelfMirror exhibits great robustness and fidelity across varied noise level data, with only a minimal performance drop in extremely low-SNR data. The signals and structures that were previously obscured by noise become clearly visible after denoising with SelfMirror, with the noise being significantly suppressed. The versatility of SelfMirror is comprehensively evaluated using a diverse range of fluorescent volumetric imaging data acquired with various microscopy techniques. This includes two-photon microscopy, expansion light-sheet microscopy, and confocal microscopy. Besides, the biological samples employed demonstrate the broad applicability of SelfMirror, including dendrites, neurons, and cerebrovasculature of mice, as well as *Penicillium* and the intestines of mouse embryos. Beyond its effectiveness in fluorescent microscopy, we further demonstrate the generality of SelfMirror for denoising volumetric data from other modalities. This includes thoracoabdominal computed tomography (CT) from humans and neuron 3D electron microscopy (3D-EM) from a mouse cortex. These results highlight the exceptional capability of SelfMirror for denoising a wide range of volumetric imaging data. Its versatility and generality across diverse modalities and biological samples make it a valuable asset for researchers investigating various subcellular and macro phenomena.

2. RESULTS

A. Principle of SelfMirror

Fluorescent images are susceptible to noise, particularly from the dominant photon shot noise and electronic noise, resulting in low-SNR images. Theoretically, the low-SNR image X can be decomposed into the sample signals G , representing ground truth (GT), and noise N :

$$X = G + N. \quad (1)$$

G represents the structural information of the biological sample, which is not independently distributed, whereas N is normally assumed as the zero-mean Poisson-Gaussian

additive noise from the photon shot noise and electronic noise, in which each pixel is independent. Then the task of denoising aims to recover the underlying clean signal G from the observed noisy data X by finding a function: $f(X) = G$. Normally, we can train a network to approximate f by using pairs of noisy-clean ($X - G$) images:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X(i)), G(i)\}, \quad (2)$$

where i denotes the pixel index in an image, L is the loss function, θ represents the parameter of the network, and $\hat{\theta}$ is the optimized parameter of the network. This scheme is impractical for fluorescent imaging denoising because it inevitably suffers from noise degradation. Strategies like Noise2Noise extended it to the temporal domain, that is, imaging signal G twice: $X' = G + N'$ and $X'' = G + N''$. It proposed to train the network only using pairs of noisy-noisy images without ground truth supervision:

$$\begin{aligned} \hat{\theta} &= \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X'(i)), X''(i)\} \\ &= \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X'(i)), G(i) + N''(i)\} \\ &\approx \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X'(i)), G(i)\}. \end{aligned} \quad (3)$$

However, capturing a pair of images from the same scene in fluorescent imaging is also impractical, especially in scenarios when the imaging procedure itself is time-consuming, as is the case in structure light microscopy. In other kind of strategies like Noise2Void, these methods take adjacent pixels of pixel i in two-dimensional images as the receptive field $\text{RF}(i)$ to predict pixel i :

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X(\text{RF}(i) \setminus i)), X(i)\}. \quad (4)$$

By excluding the center pixel from the receptive field, the network is prevented from simply learning the identity mapping. However, the network may struggle to assign appropriate weights to the excluded pixel when calculating the output. This issue becomes particularly evident when denoising isolated pixels or dealing with high irregularities within images.

Our method takes a related but different approach. Instead of blinding the input image, we propose exploring the data's spatial excessiveness to denoise the images. The central principle of SelfMirror is to perform denoising based on the structural continuity of biological samples with minute bias by exploiting 3D information only in spatial domains (Fig. 1A). Structural continuity of biological samples results in the shape and size of the structure change gradually. When capturing images, this smoothness and consistency are observed as a gradual gradient in the images, as long as the sampling rate of two adjacent pixels is sufficiently close in all three axis dimensions. In other words, when we analyze the 3D stack images by slicing them in the depth dimension as the way it was imaged, each pair of consecutive frames can be treated as independent images with their independent noise and distinct biological signals. However, despite these differences, the structures within these consecutive frames exhibit remarkable

similarity due to the continuity and smoothness of the underlying biological sample:

$$\begin{aligned} X_s &= G_s + N', \\ X_{s+1} &= G_{s+1} + N'', \\ G_s &= G_{s+1} + e, \end{aligned} \quad (5)$$

where e is the error between G_s and G_{s+1} . When the imaging interval along the z -axis is small, G_s and G_{s+1} share great similarity, that is, e is near to zero (supplementary note and supplementary Figs. S4–S6 in Ref. [38]). Then the network can be trained as analogous to Eq. (3), and we have

$$\begin{aligned} \hat{\theta} &= \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X_s(i)), X_{s+1}(i)\} \\ &\approx \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X_s(i)), G_{s+1}(i)\} \\ &\approx \operatorname{argmin}_{\theta} \sum_i L\{f_{\theta}(X_s(i)), G_s(i)\}. \end{aligned} \quad (6)$$

Note that SelfMirror is particularly designed for denoising volumetric imaging by using frame-level adjacent spatial information, in which the time domain is not present. That is, three-dimensional spatial sequences typically contain substantially fewer images compared to temporal image sequences. Consequently, learning only spatial information between slices for denoising is insufficient for volumetric data (Section 2.B). To address this limitation, we propose leveraging full three-dimensional spatial information for denoising learning. As shown in Figs. 1A, the original raw image stack is sampled to generate pairs of sub-stacks (Fig. 1A and Section 4), which are then fed into our proposed network for training. For the implementation of the network f_{θ} , we proposed a sharing network for capturing this spatial information. We first adopted the 3D U-Net architecture as the backbones and turned them into a self-supervised network (Fig. 1A and Section 4). The parameters of the two networks in our SelfMirror are shared, in this way, by being encoded into the shared learnable network; the separated spatial information can interact and make the network more efficient. In addition, we propose a self-constrained learning consisting of feature map loss and image loss (Fig. 1A and Section 4). The feature map loss supervises the distance between the two sub-stacks, while the image loss supervises the distance between the output image and the sub-stacks. In the inference stage, the whole noisy stack would be fed into the trained SelfMirror model and output the denoised stack (Fig. 1B).

B. Performance Validation on Simulated Datasets

For the quantitative evaluation of SelfMirror's performance, we first validate it on simulated volumetric fluorescent imaging data, in which noise-free images were generated using NAOMi [39] and corresponding noisy images with different SNRs were synthesized by applying different levels of mixed Poisson-Gaussian noise (see supplementary Fig. S1 in Ref. [38]). Poisson distribution (P) simulates the photon shot received by the image; higher P means higher SNR. These noises significantly obscured the spatial information of neurons in the original data stack. After applying our method, SelfMirror effectively suppressed these noises, revealing previously hidden details (Figs. 1C and 2A, 2B). Notably, both neuronal somas

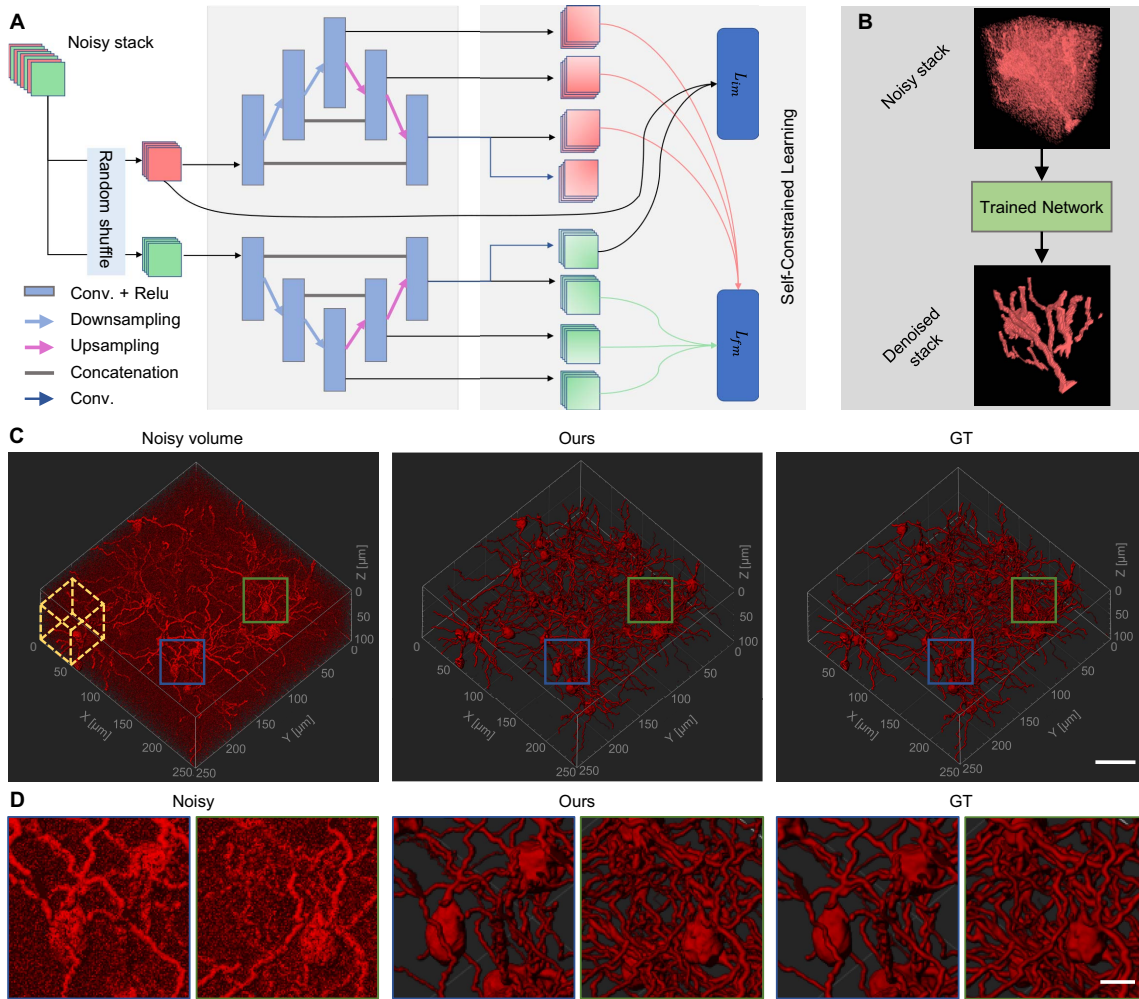


Fig. 1. Principle of SelfMirror and visualization of denoising three-dimensional imaging data. (A) Self-supervised learning strategy of SelfMirror. An imaging z stack was obtained by volumetric imaging. The original noisy image sequence is split into two sub-sequences. Each sub-stack is fed into a separate but identical network for training. The parameters of the two networks are shared. (B) The inference scheme of the trained SelfMirror; the whole noisy stacks would be fed into the network. (C) An example of 3D volume rendering of the neuronal population. The left shows the raw low-SNR volume, the middle shows the SelfMirror denoised volume, and the right shows the GT reference volume. Scale bars, 50 μm . (D) Magnified views of neuronal soma and spine structures in the color-boxed regions in (B). Scale bars, 15 μm .

and, more importantly, dendrites became clearly recognizable (see [Visualization 1](#)).

In our investigation of denoising for simulated datasets, we compared our proposed SelfMirror approach with eight established techniques: DIP [40], N2N [28], N2F [35], N2S [30], N2V [29], Ne2Ne [34], SUPPORT [36], and SN2N [41]. These methods were applied to eight noisy stacks with varying SNRs (varying P). To quantify the performance of each denoising method, PSNRs and Pearson correlation coefficients were evaluated along the three axes (x , y , z) of the image stacks (Table 1, supplementary Tables S1–S3 and Fig. S17 in Ref. [38]). As we can see, our SelfMirror achieves overall the best performance over the previous state-of-the-art results on all datasets. Although N2F also achieves a higher score in the Pearson correlation coefficient, its PSNR is not as competitive. N2F utilizes adjacent pixels within a 2D plane to generate training pairs but fails to exploit spatial redundancy, a limitation shared by N2N, N2S, and Ne2Ne. SUPPORT is

specifically designed for functional imaging denoising, requiring a significant number of slices to be sacrificed as feature channels during training. While this may be reasonable for functional imaging, it reduces data efficiency for structural imaging, where data availability is often limited. SN2N employs a 2D diagonal resampling strategy to create training pairs and uses Fourier rescaling to restore scale. However, Fourier rescaling can introduce smoothing artifacts [42], which may degrade the quality of the denoised results.

For further comparison, we demonstrated the denoising quality from randomly selected sub-stacks with max projection (Fig. 2A and supplementary Figs. S2, S3 in Ref. [38]). Maximum projections along three image axes revealed that SelfMirror and N2F can effectively reconstruct the signal, whereas the other methods struggled with signal reconstruction, particularly for features with only a few pixels in width. Beyond signal reconstruction, the analysis of maximum projections also highlighted the effectiveness of our proposed method in noise

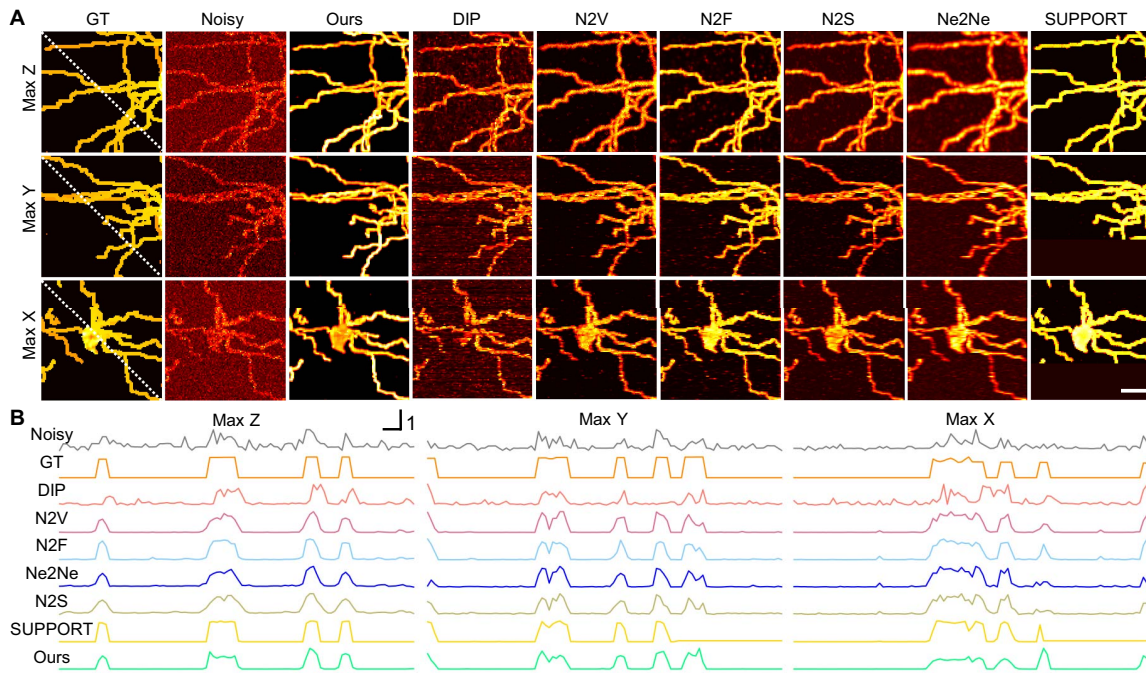


Fig. 2. Performance validation on simulated data. (A) Max projection along three (z , y , x) axes of blue-boxed region in Fig. 1A. From left to right are the GT, noisy, SelfMirror (ours), DIP, N2V, N2F, N2S, Ne2Ne, and SUPPORT denoised data. The images are colored with red hot. Scale bars, 20 μm . (B) From left to right are the pixel intensity profiles along the white dashed lines of max- z , max- y , and max- x in (A), respectively. The intensity is normalized to 0–1.

Table 1. PSNR and Pearson Correlation Coefficient (R) along the xy Plane of the Whole Volume from Eight Datasets (P , from 1 to 128)^a

P	PSNR (dB)/ R							
	RAW	N2N	N2F	N2S	Ne2Ne	SUPPORT	SN2N	SelfMirror
1	10.95/0.30	9.77/0.29	19.77/ 0.85	19.79/0.82	19.55/0.8	25.61 /0.64	24.54/0.82	26.35 / 0.90
2	13.43/0.40	8.60/0.48	20.12/ 0.90	20.07/0.88	19.74/0.84	24.10 /0.65	20.46/0.87	23.26 / 0.89
4	15.53/0.51	8.55/0.37	20.38/ 0.94	20.27/0.91	19.97/0.86	21.17 /0.66	20.44/0.88	23.97 / 0.92
8	17.16/0.62	11.86/0.53	20.56/ 0.95	20.57 /0.92	20.14/0.90	20.16/0.67	20.51/0.88	20.99 / 0.95
16	18.27/0.70	9.37/0.75	20.69/ 0.96	20.58/0.93	20.28/0.90	20.22/0.67	21.97 /0.86	21.14 / 0.96
32	18.96/0.77	8.95/0.65	20.77/ 0.96	20.65/0.93	20.34/0.91	20.20/0.67	21.97 /0.88	21.14 / 0.96
64	21.16/0.85	11.05/0.74	22.61/ 0.97	22.07/0.95	22.16/0.92	21.69/0.67	22.97 /0.89	22.83 / 0.96
128	26.95/0.95	6.33/0.78	27.69/ 0.97	27.18/0.95	27.73 / 0.97	25.62/0.67	25.22/0.80	27.75 /0.96

^aA full comparison is in the supplementary Tables S1–S3 in Ref. [38]. The first and second rank values are set in bold formatting.

suppression. While all methods offered some degree of noise reduction, SelfMirror stands out for its exceptional ability to substantially suppress noise, making a clear distinction between signal and background. The pixel trace (Fig. 2B) quantified the advantage of SelfMirror in both signal reconstruction and noise suppression, which verified the importance of fully exploiting spatial information.

C. Denoising Two-Photon Volumetric Imaging Data of Single Neurons

To verify the effectiveness of SelfMirror on experimentally obtained data, we first employed SelfMirror for noise removal and signal reconstruction for volume data of a single neuron in an anesthetized mouse. The results are shown in Fig. 3. We first captured an image stack of the single-neuron structure

in cortical layer 1 of an anesthetized mouse expressing dsRes2 under low laser power. The images were fraught with noise, to the extent that the signal from certain neuronal features was barely discernible. To validate our method, we also obtained a high-SNR image stack from the identical region but with an elevated laser power. The high-SNR image stack can serve as the reference for our denoising results. The application of SelfMirror to the low-SNR image stack yielded remarkable improvements (Fig. 3A). The original low-SNR image stack, which was initially filled with a significant amount of noise, became nearly noise-free after applying the SelfMirror denoising method. The background was rendered extremely clean, creating a very clear distinction between the background and the signal, even surpassing the quality of the high-SNR reference images. The crucial neuronal components were previously

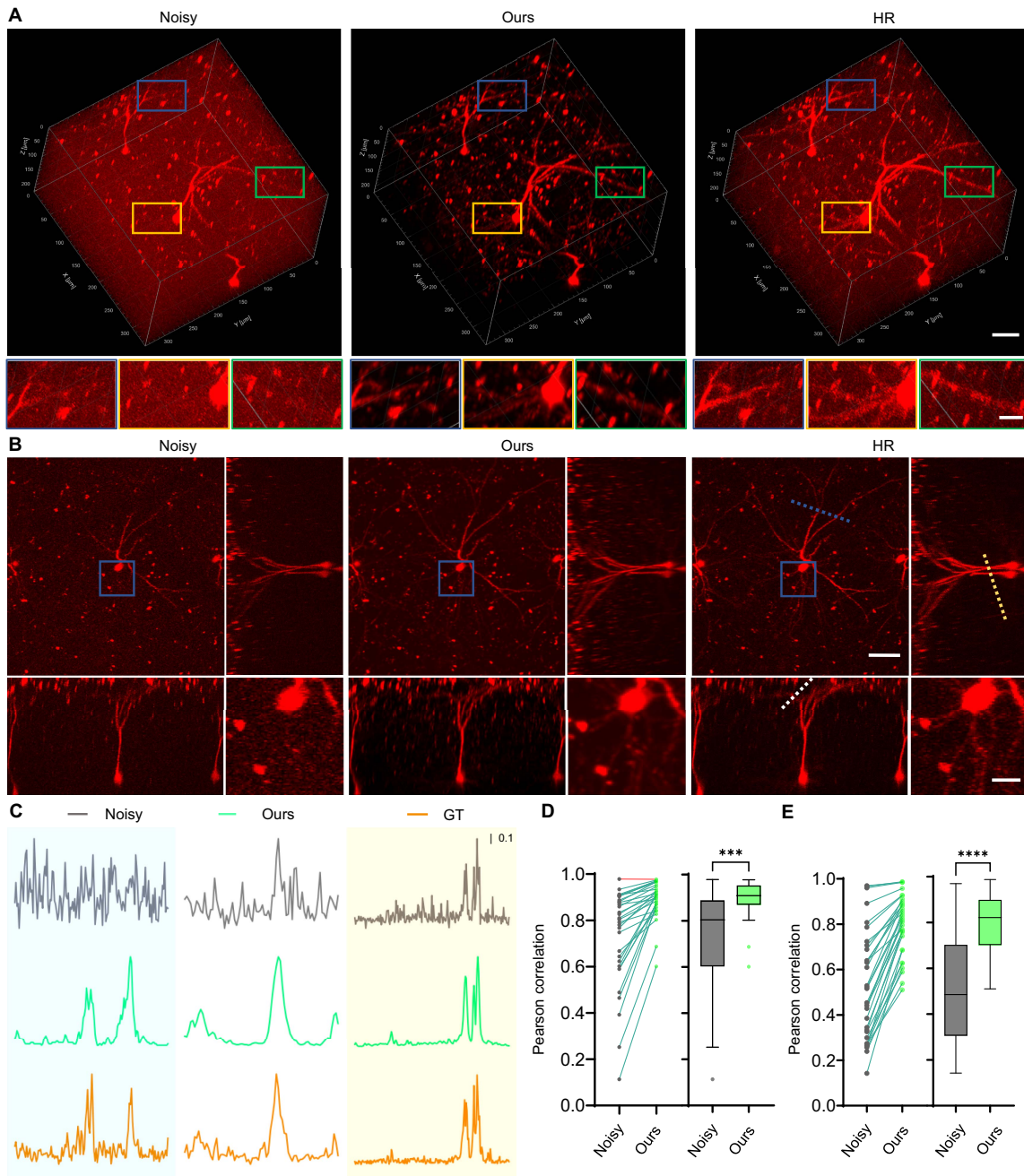


Fig. 3. SelfMirror denoises two-photon volumetric imaging data of single neurons. (A) Visualization of single neuronal morphological images with a $340\ \mu\text{m} \times 340\ \mu\text{m} \times 220\ \mu\text{m}$ volume (278 planes, $0.66\ \mu\text{m}/\text{pixel}$ in x and y , $0.8\ \mu\text{m}/\text{pixel}$ in z) in the mouse cortex. From left to right are the original noisy volume, the same volume denoised with SelfMirror, and the high-SNR reference volume. Color-boxed regions show the magnified views of the neuron structures with soma, dendrite, and spine. Scale bars, $50\ \mu\text{m}$ for the whole FOV and $20\ \mu\text{m}$ for magnified views. (B) Orthogonal views of the same imaging plane of the neuronal volumes. From left to right are the noisy data, SelfMirror denoised counterparts, and high-SNR reference data. Magnified views of blue-boxed regions are shown at the bottom right of the images. Scale bars, $50\ \mu\text{m}$ for the whole FOV and $15\ \mu\text{m}$ for magnified views. (C) From left to right are the pixel intensity profiles along the blue-, white-, and red-dashed lines in (B), respectively. The intensity is normalized to 0–1. (D) Left, the statistical evaluation of Pearson correlation coefficient increases of traces before and after denoising by SelfMirror, in which the trace is randomly selected in xy slices, $N = 32$. Each line represents one of 32 traces, and increased correlations are colored blue. Right, Tukey box-and-whisker plot of the fluorescent traces. p values calculated by two-sided one-way paired t -test are specified with asterisks, *** $p < 0.001$. (E) Left, the statistical evaluation of Pearson correlation coefficient increases of traces before and after denoising by SelfMirror, in which the trace is randomly selected in xz and yz slices, $N = 31$. Each line represents one of 31 traces, and increased correlations are colored blue, decreased correlations are colored red. Right, Tukey box-and-whisker plot of the fluorescent traces. p values calculated by two-sided one-way paired t -test are specified with asterisks, **** $p < 0.0001$.

obscured by noise, rendering them difficult or impossible to discern in the raw low-SNR image. However, after applying SelfMirror, the dendritic structures became visible, allowing for a detailed analysis of their morphology and distribution (see [Visualization 2](#)).

Maximum projection along image axes also further verified the effectiveness of our method (Fig. 3B). The maximum projection images of the low-SNR stack presented a significant challenge, particularly for visualizing faint structures. In our case, the low-SNR image contained small dendrites with weak fluorescence intensity, rendering them virtually indistinguishable from the background noise. These crucial structures were essentially “drowned out” by the noise, hindering a comprehensive analysis. However, the projection images of the SelfMirror enhanced stack demonstrated that the previously unrecognized dendritic structures became recognized. Additionally, fluorescent traces of dendritic pixels were accurately recovered, exhibiting high consistency with the high-SNR reference (Fig. 3C). To further quantify this improvement, we calculated the Pearson correlation coefficients of randomly selected fluorescent traces compared to the high-SNR reference in each maximum projection direction by giving z , y , or x steps (Figs. 3D, 3E and supplementary Fig. S7 in Ref. [38]). Pearson correlation was improved from 0.52 to 0.79 in the max- z projection (Fig. 3D, median value) and 0.73 to 0.89 in the max- x and max- y projection (Fig. 3E, median value).

D. Denoising Two-Photon Volume Imaging Data of Neuronal Population

Following the success of single-neuron data, we employed SelfMirror for noise removal in the imaging data of large neuronal populations. The results are demonstrated in Fig. 4. Similar to the denoising performed on single neurons, we first imaged low-SNR volume images of neuronal populations in cortical layer 1 of an anesthetized mouse expressing EGFP under low laser power, followed by high-SNR volume images of the same region with high laser power as a reference. Despite the high noise level of the low-SNR images, the neuronal structures can be revealed from the 3D rendering of the neural volume after SelfMirror denoising (Fig. 4A and [Visualization 3](#)). For a detailed comparison, we present snapshots of manually selected regions at different depths and an orthogonal view of the neural volume. With the enhancement of SelfMirror, the structure and distribution of the neurons became distinctly observable (Figs. 4B, 4C and supplementary Fig. S8 in Ref. [38]). Additionally, we extracted the fluorescent traces along a diagonal line. Consistent with the results from simulations and single-neuron structural imaging, the variance was significantly reduced, while the small dendrites with weak intensity were preserved (Fig. 4D). To quantify noise suppression, we statistically calculated the background variance in a selected non-signal region. The superior ability of SelfMirror can reveal high-fidelity noise suppression with extremely low variance (Fig. 4E). This high-fidelity noise suppression highlights SelfMirror’s effectiveness in enhancing the visibility and clarity of neuronal structures, making it a powerful tool for detailed and accurate imaging of large neuronal populations.

Moreover, to verify the reliability of SelfMirror on other experimental data, we demonstrated its performance on

publicly available multiphoton structural imaging data of deep cortical spines and (sub)cortical dendrites, a challenge for high-SNR imaging due to the depth of these structures in the brain (Fig. 4F). In these data, adaptive optics (AO) enhanced imaging in deep scattering tissues mitigates otherwise low SNR and serves as a high-SNR reference. Contaminated by detection noise, the spatial structures of spines and dendrites were severely corrupted in the imaging data without AO. After we applied SelfMirror to enhance these data, the structures became recognizable and free from noise. To further illustrate this improvement, we took snapshots of manually selected regions at three different depths for comparison (Fig. 4F and supplementary Fig. S9 in Ref. [38]). The previously barely perceptible spines and dendrites were now distinguishable and maintained their shapes, which would otherwise have been overwhelmed by noise (see [Visualization 4](#)).

E. Applying SelfMirror Denoising to Multiple Fluorescent Imaging Microscopy

Then, to demonstrate the versatility of SelfMirror, we evaluated it on denoising volumetric structural imaging data from different microscopy techniques and biological samples. SelfMirror was tested on three volumetric datasets that contained *Penicillium* imaged with confocal microscopy, mouse cerebrovasculature imaged with two-photon microscopy, and intestine of mouse embryos imaged with expansion microscopy. The qualitative assessment revealed that SelfMirror was able to enhance the signals and surpass the noise as illustrated in Fig. 5. For example, in the data with abundant fluorescent contents like *Penicillium* and intestine, SelfMirror successfully reconstructed the elaborate mycelium structures with exceptional clarity (Figs. 5A and 5B, supplementary Figs. S10, S12 in Ref. [38], [Visualization 5](#) and [Visualization 7](#)). For the data with sparse fluorescent contents like cerebrovasculature, SelfMirror surpassed the overwhelming noise (Fig. 5D, supplementary Fig. S11 in Ref. [38], and [Visualization 6](#)). These results collectively highlighted SelfMirror’s capability to learn and adapt to the statistical characteristics of a wide range of fluorescent imaging data. For the quantitative evaluation of SelfMirror with the *Penicillium* dataset, analysis of fluorescent pixel traces demonstrated SelfMirror denoising results exhibited high consistency with the high-SNR reference (Fig. 5A). Further measurements of the Pearson correlation coefficients and PSNR, using the high-SNR image as the ground truth for each plane along the z -axis, also reinforced the consistency (Fig. 5C). The average PSNR (27.35 ± 1.05 dB) of SelfMirror showed 5.76 dB increments compared to the low-SNR image (21.59 ± 0.90 dB) and the average Pearson correlation coefficient of SelfMirror (0.93 ± 0.001) showed 0.32 increment compared to the low-SNR image (0.61 ± 0.01). The successful denoising of diverse datasets across various microscopy techniques underscores the broad applicability of SelfMirror in biological research.

F. Applying SelfMirror Denoising to Multiple Imaging Modalities

To further explore the generality of SelfMirror’s denoising capabilities, we ventured beyond fluorescence microscopy and applied it to different imaging modalities. This demonstrates

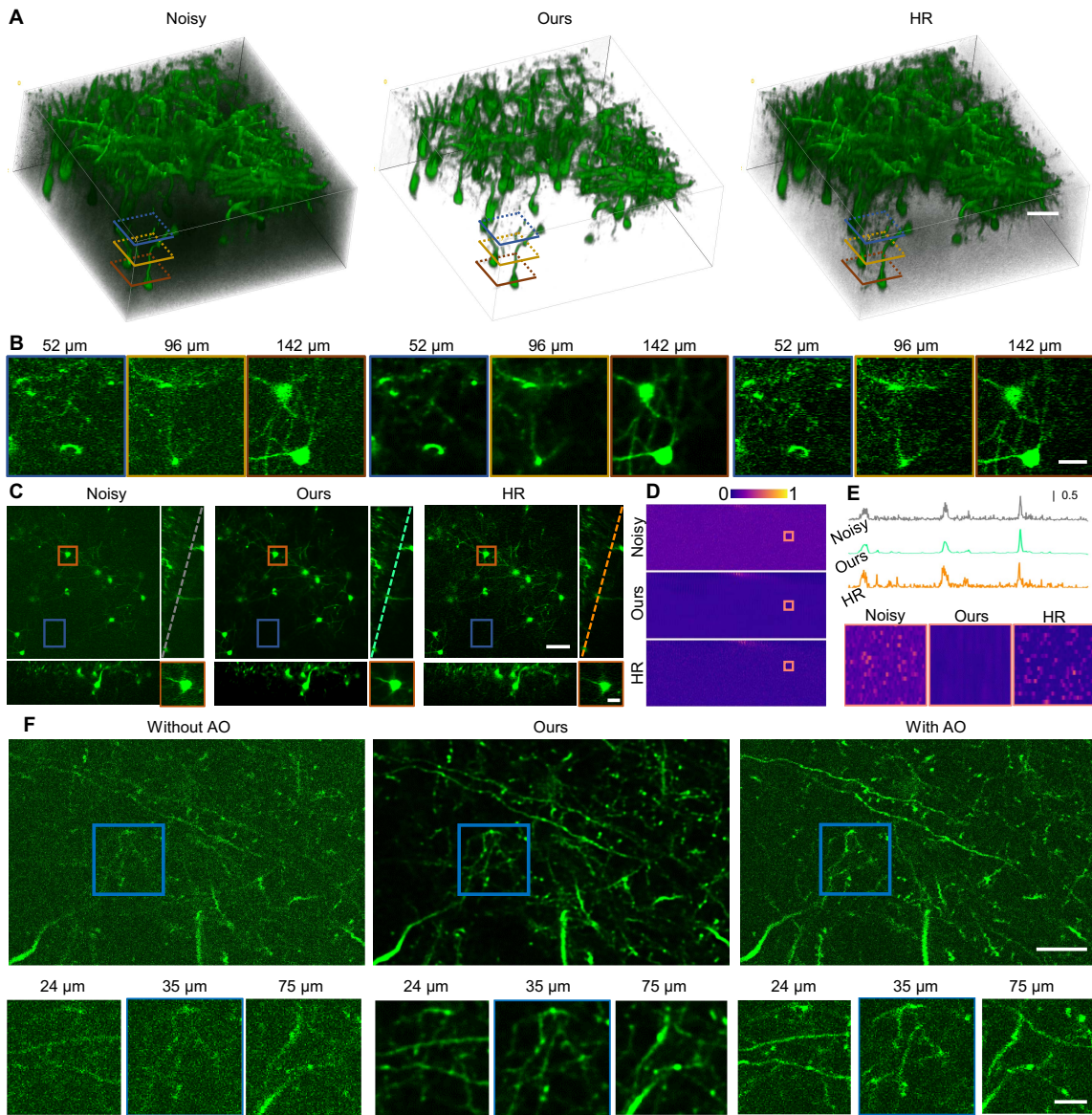


Fig. 4. SelfMirror denoises two-photon volumetric imaging data of neuronal population. (A) Visualization of morphological structure imaging of neuronal population with a $340 \mu\text{m} \times 340 \mu\text{m} \times 220 \mu\text{m}$ volume (144 planes, $0.66 \mu\text{m}/\text{pixel}$ in x and y , $1.5 \mu\text{m}/\text{pixel}$ in z) in the mouse cortex. From left to right are the original noisy volume, the same volume denoised with SelfMirror, and the high-SNR reference volume. Scale bars, $50 \mu\text{m}$. (B) Magnified views of the neuron structures with sliced soma, and dendrite in the three color-boxed regions at three different depths (52, 96, 142 μm) in (A). From left to right are the original noisy data, the same data denoised with SelfMirror, and high-SNR reference data. Scale bars, $20 \mu\text{m}$. (C) Example orthogonal frames of the neuronal volumes. Left, original noisy frames from three-dimensional views. Middle, SelfMirror denoised counterparts. Right, high-SNR reference frames. The imaging planes are the same on the left, middle, and right. A magnified view of the red-boxed region is shown at the bottom right of the images. Scale bars, $50 \mu\text{m}$ for the whole FOV and $10 \mu\text{m}$ for magnified views. (D) Pixel intensity profiles along the blue-dashed lines in (C). The intensity is normalized to 0–1. (E) Statistical spectrum plots of the intensity value of the blue-boxed signal-free region in (C), along the imaging z -axis. One row of the spectrum plot represents all pixels ($n_1 = 2800$) of the blue-boxed region in (C). The column direction of a spectrum plot represents the imaging planes ($n_2 = 144$). The equation for total pixels of a plot is $n = n_1 \times n_2 = 403,200$. All pixels were normalized with a plasma bar. Zoom-in blue-boxed region views are shown in the right panel. (F) Representative z -axis slice of structural imaging of dendrite and spine with a $180 \mu\text{m} \times 180 \mu\text{m} \times 150 \mu\text{m}$ volume (100 planes, $0.35 \mu\text{m}/\text{pixel}$ in x and y , $1.5 \mu\text{m}/\text{pixel}$ in z) in a mouse expressing GFP. Left, the original imaging data without AO as low-SNR image. Middle, the same image denoised with SelfMirror. Right, imaging data with AO as a high-SNR reference. Magnified views of the blue-boxed region at multiple axial locations are shown in the bottom panel. Axial location of $35 \mu\text{m}$ corresponds to the current frame. Scale bars, $25 \mu\text{m}$ for the whole FOV and $10 \mu\text{m}$ for magnified views.

that SelfMirror is not limited to a specific noise model but can effectively handle noise in various imaging techniques. Two datasets were employed for evaluation. (1) Human

thoracoabdominal CT images: this dataset presented a challenge due to the inherent noise limitations of CT scans, and SelfMirror was tested on a pair of low-SNR and high-SNR

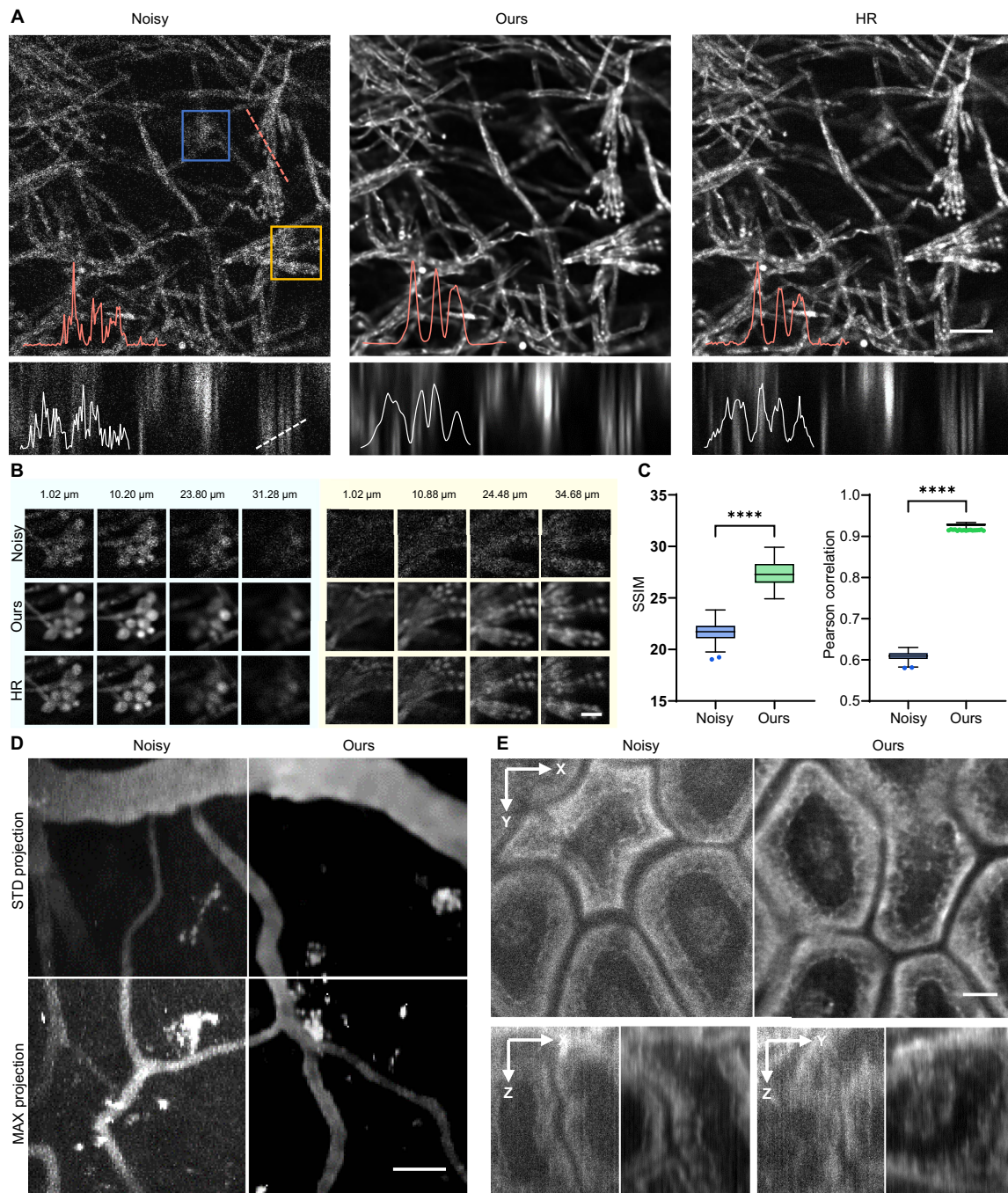


Fig. 5. Denoising structural imaging data from multiple fluorescent microscopies. (A) Visualization of representative z -axis slices (top) and y -axis slices (bottom) from left, noisy data, middle, corresponding SelfMirror denoised data, and right, high-SNR data of *Penicillium*. The pixel intensity profiles of the blue- and white-dashed lines are inserted on the left bottom of the images, respectively. Scale bars, 20 μm . (B) Magnified views of the blue- and yellow-boxed regions in (A) at multiple axial locations. Scale bars, 2 μm for all images. (C) Box-and-whisker plot showing PSNR and Pearson correlation coefficient for z -axial slices before and after SelfMirror denoising; high-SNR reference data were used as the ground truth for calculation. A two-sided paired-sample t -test is used, $N = 129$, which represents the number of planes along the z -axis ($****p < 0.0001$). (D) STD projection (top) and max projection (bottom) of vessels of a mouse cortex for the raw noisy data (left) and corresponding images using SelfMirror (right). Scale bars, 30 μm . (E) Example frames in three-axis planes (top, z -axis; bottom left, y -axis; bottom right, x -axis) of the intestine of a mouse embryo after expansion for the original noisy data (left) and corresponding denoised image using SelfMirror (right). Scale bars, 30 μm .

volume images acquired with different CT scan settings. (2) 3D-EM images of the mouse cortex: this dataset involved high-resolution 3D-EM images of the mouse cortex. While

3D-EM offers exceptional detail, it can also be susceptible to noise artifacts. The qualitative visualization of the thoracoabdominal CT (Fig. 6A, supplementary Fig. S13 in

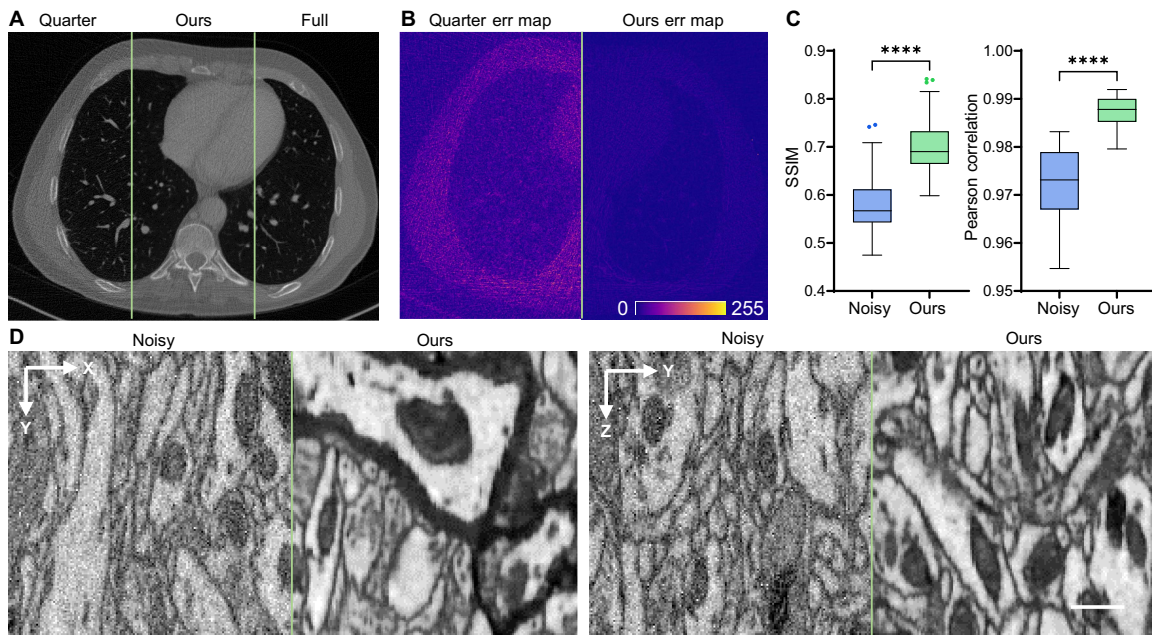


Fig. 6. Denoising multiple volumes from multiple imaging modalities. (A) Representative slice from low-SNR (left), SelfMirror denoised (middle), high-SNR (right) CT volumes of human thoracoabdominal body. (B) Low-SNR (left) and SelfMirror denoised (right) error maps with high-SNR image in (A). Error maps are colored with a plasma bar. (C) Box-and-whisker plot showing Pearson correlation coefficient (left) and structural similarity index (right) of axial slices. A paired-sample t -test is used, $N = 560$, which represents the number of planes along the z -axis (**** $p < 0.0001$). (D) Example xy plane (left) and yz plane (right) frames of a mouse somatosensory cortex layer 4 from a 3D-EM volume $93\ \mu\text{m} \times 60\ \mu\text{m} \times 93\ \mu\text{m}$. Scale bars, $10\ \mu\text{m}$.

Ref. [38], and Visualization 8) and cortex 3D-EM (Fig. 6D, Figs. S14, S15 in Ref. [38], and Visualization 9) revealed a significant reduction in noise and an overall enhancement of signal intensity in the volumetric data processed by SelfMirror. The absolute error maps to the high-SNR reference showed that our SelfMirror denoised image has a smaller difference to the high-SNR reference (Fig. 6B and supplementary Fig. S13 in Ref. [38]). For the quantitative evaluation of SelfMirror with the CT dataset, the SSIM and Pearson correlation coefficients for each plane along the z -axis were calculated (Fig. 6C). The average SSIM (0.71 ± 0.043) of SelfMirror shows more than 22% increment compared to low-SNR images (0.58 ± 0.050), and the average Pearson correlation coefficients increase to 0.99 ± 0.002 by SelfMirror from 0.97 ± 0.007 of low-SNR images. The successful denoising of CT and 3D-EM images, alongside the previous demonstrations with fluorescence microscopy, underscores the broad applicability of SelfMirror across diverse imaging modalities. Its ability to function independently of specific noise model assumptions makes it a valuable tool for researchers working with various imaging techniques in biological research.

3. DISCUSSION

In summary, SelfMirror is based on deep self-supervised learning for spatial enhancement of diverse imaging datasets. It capitalizes on the inherent property of biological tissues: their structures generally change smoothly and continuously across space. This core principle empowers SelfMirror to be readily

adaptable to a wide range of biological samples and imaging modalities. Examples include multiphoton imaging, confocal microscopy, and even expansion microscopy. The limitation of SelfMirror lies in the basic assumption that the adjacent slices in the z -axis of the volume image have high similarity. If the z interval between captured slices is too large to maintain this similarity, SelfMirror would produce bias and over-smoothed denoising results, and its performance may degrade (supplementary Figs. S4–S6 in Ref. [38]). Therefore, when determining the spacing between z -slices, the step size should be made as small as possible to ensure a sufficiently high level of similarity between adjacent slices. In our experimental validation of the SelfMirror principle, it was demonstrated that datasets employing z intervals below $6\ \mu\text{m}$ achieved satisfactory denoising performance across specimen scales ranging from 2 to $30\ \mu\text{m}$. Notably, exceptional noise reduction efficacy was observed in datasets with z intervals under $4\ \mu\text{m}$, with SNR improvements exceeding threefold compared to original noisy stacks (supplementary Fig. S4 in Ref. [38]).

Since SelfMirror does not make specific assumptions about the noise model present in different imaging modalities, retraining the model might be necessary to achieve optimal results. This becomes particularly important when factors like imaging parameters, sample types, or modalities change. Additionally, it is also worth noting that SelfMirror is specifically tailored to address zero-mean stochastic noise, which is a common type of noise in fluorescent imaging data. However, the method may not be as effective in dealing with deterministic artifacts, such as those caused by motion, photobleaching, or fixed-pattern noise. The presence of significant drift artifacts

in imaging datasets primarily induced by physiological motions like respiration and cardiac activity may compromise inter-slice continuity. This disruption compromises the denoising efficacy of SelfMirror, resulting in bias and oversmoothing artifacts in drifted regions. To mitigate this limitation, we recommend prerequisite non-rigid registration using the ImageJ [43] prior to applying our denoising framework (supplementary Fig. S16 in Ref. [38]). And the computational reconstruction involved in techniques such as CT may introduce artifacts and modify the noise distribution, potentially deviating from the ideal zero-mean stochastic noise assumption. These factors do indeed pose challenges to the performance of our SelfMirror method, as we can see that certain structured noise patterns, such as radial noise artifacts, remain in the denoised images while full-dose CT suffers less of these patterns. These artifacts often require more specialized approaches or additional preprocessing steps to be effectively mitigated. Despite this, our comparative studies indicate that SelfMirror consistently outperforms other methods in CT denoising.

The successful application of SelfMirror for denoising CT and 3D-EM images demonstrates its versatility beyond fluorescent imaging data. This is a significant advancement, as these imaging modalities often involve biological samples susceptible to radiation exposure or electron beam damage. By effectively denoising such data with SelfMirror, researchers can potentially reduce the required radiation or electron beam dose, minimizing potential harm to the samples while maintaining image quality. Encouragingly, training SelfMirror can be completed within half an hour, and subsequent predictions for a $512 \times 512 \times 200$ pixel volume can be made within a minute on an NVIDIA RTX 3090 GPU, which facilitates its general usage in many laboratories with common desktop settings. Overall, SelfMirror's self-supervised learning scheme, coupled with its robustness in noise suppression and signal reconstruction, positions it as a versatile tool for a wide range of image data processing tasks. We expect that its core strategy, learning the statistical relationships between neighboring similarities, will not only enhance image denoising but also be adapted to process a broader spectrum of biological data.

4. METHODS

A. Network and Training Details

We adopted the 3D U-Net architecture as the backbones and turned them into a self-supervised network (Fig. 1A). The network is composed of two 3D encoders, a bottom block, and two 3D decoders with three skip connections from the encoder blocks to the decoder blocks. Each 3D encoder block consists of two $3 \times 3 \times 3$ convolutional layers, followed by a ReLU activation without batch normalization, with a final $2 \times 2 \times 2$ max pooling with strides of two to downsample the feature maps. The bottom block is similar to the encoder without downsampling. And each 3D decoder block contains two $3 \times 3 \times 3$ convolutional layers, followed by a ReLU activation function, with a final 3D nearest interpolation to upsample the feature maps. The 3D encoders form a contracting path to encode the input to high-dimensional space. Then the 3D decoders construct an expanding path to decode the high-

dimensional representation to output the same shape as the input. In the training process, as illustrated in Fig. 1A, the image stack is split into two sub-stacks: one consisting of odd-numbered slices and the other of even-numbered slices. Each sub-stack is fed into a separate but identical network for training. A random shuffle block—rotate, flip, or swap—is applied for data augmentation to reduce overfitting. The parameters of the two networks are shared, in this way, by being encoded into the shared learnable network; the separated spatial information can interact and make the network more efficient. The feature maps output from the bottom block and the decoder blocks of each network, together with the final output of the network, participate in the optimization of the entire training; thus, the loss function of SelfMirror contains two parts: feature map loss L_{fm} and image loss L_{im} :

$$L_{fm} = \frac{1}{n} \sum_{i=1}^n (\|d_1^i - d_2^i\|_2 + \|d_1^i - d_2^i\|_1),$$

$$L_{im} = \alpha \|f(X_1) - X_2\|_2 + (1 - \alpha) \|f(X_1) - X_2\|_1, \quad (7)$$

where d_1^i and d_2^i represent the feature maps output from the decoders of each network, $i \in (1, n)$; n is the number of the decoder blocks. f represents the SelfMirror network, X_1 and X_2 are the input and the target sub-stacks randomly shuffled from the noisy stack, respectively, and α denotes the hyperparameter and is set to 0.5. $\|\cdot\|_2$ and $\|\cdot\|_1$ are the L_2 norm and L_1 norm, respectively. During the training, the ratio between the feature map loss and the image loss was set to 1:2 to balance their contributions. The model was trained using 3D patches, each with a spatial resolution of $64 \times 64 \times 64$ pixels. Network optimization was conducted with the Adam optimizer, employing a fixed learning rate of 1×10^{-4} . The training process spanned 20 epochs, with each epoch encompassing a complete pass through all available patches. To facilitate detailed monitoring of the training dynamics, loss values were recorded after each gradient update. All training experiments were performed using PyTorch 2.0.1 and CUDA 11.8, executed on an NVIDIA RTX 3090 GPU. In the inference stage, the whole noisy stacks would be fed into the trained SelfMirror model, for inference, without the need to split them into two separate sequences (Fig. 1B).

B. 3D Imaging Data Simulation

We utilized simulated structural fluorescent imaging data to quantitatively evaluate the performance of SelfMirror and other SOTA methods for comparisons. Our simulation pipeline followed a two-step process to generate the datasets. First, we synthesized noise-free morphological imaging volumes, which served as the ground truth for our experiments. We adopted the *in silico* NAOMi [39] to obtain these noise-free volumes. The simulator was originally designed for simulating two-photon calcium imaging data. We modified the codes to create realistic morphological fluorescent imaging datasets with fluorescently labeled neuron somata, axons, and dendrites. Next, the simulated data were corrupted with different levels of mixed Poisson-Gaussian noise, which is dominant in fluorescent imaging [11], to obtain different imaging SNR volume datasets. Poisson sampling on noise-free images was first performed to

simulate the content-dependent Poisson noise, then Gaussian noise was added to these data, and Poisson noise was set as the dominant noise source. We created eight different SNR datasets from low to high SNR and used them in the paper for the performance evaluation of our SelfMirror and other methods.

C. *In vivo* Two-Photon 3D Imaging of Mouse Brain

All experiments were supported by the Animal Care and Use Committee of Nankai University, following the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

For single neuronal imaging, mice were obtained from Tianjin Medicinal University. *In utero* electroporation surgery was performed to transfect layer 2/3 progenitor neurons in embryonic day (E) 15.5 embryos. During the procedure, the pregnant mice were anesthetized, the uterine horns were carefully exposed, and approximately 1 μ L of a buffer solution was pressure-injected into the right lateral ventricle of each embryo using a pulled-glass pipette. This buffer solution contained 1.2 μ g/ μ L SEP-GluA1 and myc-GluA2 plasmid, 0.3 μ g/ μ L Dsred2 plasmid, and a trace amount of Fast Green (Sigma) to aid visualization. Then, the 3 mm tweezer electrode of the electroporator was positioned to target the somatosensory cortex, and five pulses of 35 V (50 ms on, 950 ms off, 1 Hz) were delivered to the embryos. Following the electroporation procedure, the embryos were carefully returned to the maternal abdominal cavity, and the abdominal wall muscles and skin were sutured. Once the pregnant mice recovered from anesthesia, they were placed back in their cages and closely monitored, with ample water and food provided. At 4 weeks after the birth of the offspring, males and females were separated into different cages. At 8 weeks of age, craniotomy surgery was performed for imaging.

For neuronal population imaging, mice were genetically modified to express EGFP using AAV transfection. During the procedure, mice were first anesthetized with an isoflurane-oxygen mixture (1.5% isoflurane/O₂, volume ratio) and given the analgesic buprenorphine (SC, 0.3 mg/kg). Next, virus injection was performed using a specialized glass pipette with a 45° beveled tip and a 15–20 μ m opening. This pipette was filled with mineral oil and loaded with the viral solution using a plunger controlled by a hydraulic manipulator; 20–30 nL of AAV-hSyn-EGFP-containing solution (2×10^{13} infectious units/mL) was carefully injected into a targeted region of the cortex. To prevent the viral solution from leaking back out after injection, the pipette was left in the brain for over 15 min and then the plunger was withdrawn (1 nL in volume) before the pipette was pulled up. After that, the skin was sutured and mice recovered from anesthesia. Following the procedure, the mice were placed back in their cages and closely monitored, with ample water and food provided. Craniotomy surgery was performed for imaging at 2–3 weeks after AAV transfection.

For cerebrovascular imaging, mice were injected with DsRed2 fluorescent protein via the tail vein. The procedure involved dilating the tail veins, disinfecting the injection site, and carefully injecting the protein solution into one of the lateral tail veins using a fine-gauge needle. Approximately

100–200 μ L of DsRed2 solution was administered per mouse. After injection, the mice were monitored to ensure proper recovery and the absence of any adverse effects.

For craniotomy surgery, mice were first anesthetized on a heating pad with eye ointment covered. The scalp was then carefully dissected to expose the skull. A custom-made recording chamber was secured onto the skull using glue and dental cement. Following a 20 min curing for the cement, a 3 mm diameter cranial window was drilled using a trephine drill. Next, the designated skull disc was removed with tweezers, and the dura mater was carefully excised. Finally, a cover glass was placed over the cranial window to protect the exposed brain and maintain a sterile environment. After the surgery, the mice were recovered from anesthetization, placed back in their cages, and closely monitored, with ample water and food provided. Post-surgery administration of anti-inflammatory agents continued for seven consecutive days for long-term imaging. Then, two-photon imaging could be performed after a week of recovery.

All imaging experiments were conducted using a two-photon microscope equipped with a Ti-sapphire laser (Mai-Tai Deep See Spectra Physics). The excitation beam was focused with a water-dipping objective (Nikon 40 $\times$, 0.80 NA). The resulting backward-scattered signal emanating from the brain tissue was gathered by the same objective lens and detected with photomultiplier tubes (PMTs, Hamamatsu GaAsP PMT). In single-neuron imaging experiments, SEP-GluA1 and dsRed2 were excited using a 910 nm wavelength with power output ranging from 15 to 100 mW directed to the back aperture of the objective lens, according to the fluorescent expression condition and imaging depth. The emitted fluorescence was separated by dichroic mirrors and filters (ET525/50m for green, ET629/56m for red). The laser power was adjusted to a ratio of 1:3 for images with low and high SNRs. In neuronal population imaging experiments, the excitation wavelength was set to 920 nm with 15 to 100 mW and a 1:2 laser power ratio for low-SNR versus high-SNR images. In cerebrovascular imaging experiments, the excitation wavelength was set to 920 nm with 15 to 80 mW. During high-SNR imaging, careful power adjustments were crucial to prevent photobleaching. All image stacks were acquired at 512 $\times$ 512 pixels with a horizontal size of 0.66 μ m/pixel in x and y directions.

D. Method Comparison

We conducted a comprehensive performance comparison of SelfMirror against eight established denoising methods: DIP [40], Noise2Noise (N2N) [11], Noise2Fast (N2F) [35], Noise2Self (N2S) [30], Noise2Void (N2V) [29], Neighbor2Neighbor (Ne2Ne) [34], SUPPORT [36], and SN2N [41]. These methods were implemented using open-source codes provided by the respective research publications. Each method's denoising model was trained and tested on the same datasets to ensure a fair comparison. For methods originally designed for two-dimensional images, we adapted the image stack by splitting it into a series of two-dimensional frames to match the required input dimensions. Both training and inference were performed on a frame-by-frame basis. We adhered to the default training settings, including network architectures and hyperparameters, for all methods to maintain consistency.

E. Evaluation Metrics

To assess the effectiveness of various denoising methods, we employed the following quantitative metrics. For an image (or an image stack) x and its corresponding ground truth (noise-free image) y , the metrics are defined as follows.

SNR quantifies the pixel-level discrepancy between the two images on a logarithmic decibel scale; higher SNR values indicate that the reconstructed image is closer to the original, meaning better quality, which is formulated as

$$\text{SNR(dB)} = 10\log_{10} \frac{y^2}{(x-y)^2}. \quad (8)$$

Peak signal-to-noise ratio (PSNR) is a widely recognized metric for gauging the fidelity of image reconstruction. It quantifies the difference between the two images and is expressed in decibels; higher PSNR values indicate that the reconstructed image is closer to the original, meaning better quality, which is formulated as

$$\text{PSNR(dB)} = 10\log_{10} \left(\frac{\max(x)^2}{\text{mse}} \right), \quad (9)$$

where $\max(x)$ indicates the maximum intensity value of x ; mse measures the mean square error between x and y , which is defined as $\text{mse} = E(x-y)^2$, where E represents the arithmetic mean.

The Pearson correlation coefficient quantifies the linear correlation between the intensity values of two images; higher Pearson correlation coefficient values indicate that the reconstructed image is closer to the original in terms of pixel intensity and distribution, meaning better quality, which is formulated as

$$R_{xy} = \frac{E(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y}, \quad (10)$$

where μ_x , μ_y and σ_x , σ_y are the means and variances of x and y , respectively.

SSIM measures the perceptual similarity between two images on a perceptual level, including luminance, contrast, and structure; higher SSIM values indicate that the reconstructed image is closer to the original, meaning better quality. The definition is

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (11)$$

where σ_{xy} is the covariance of x and y . The two constants c_1 and c_2 are defined as $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ with $k_1 = 0.01$, $k_2 = 0.03$, and $L = 65,535$.

Funding. National Natural Science Foundation of China (62027812, 62333012).

Acknowledgment. We thank C. Wang for his support of proofreading.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Both simulated neuron volumetric data and two-photon volumetric imaging of single neurons, neuron

populations, as well as cerebrovasculature data used in this work are available in Ref. [38]. The dataset of two-photon volumetric imaging of dendrites is available in Ref. [44]. The dataset of confocal imaging of volumetric structural imaging of *Penicillium* is available in Ref. [36]. The dataset of expansion imaging of volumetric structural imaging of the intestine of mouse embryo is available in Ref. [36]. The dataset of thoracoabdominal computed tomography of humans is available in Ref. [45]. The dataset of neuron 3D electron microscopy of mouse cortex is available in Ref. [46]. The open-source Python code of SelfMirror is available at GitHub in Ref. [47]. See supplementary information in Ref. [38].

REFERENCES

1. Y. Guo, D. Li, S. Zhang, *et al.*, "Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond time-scales," *Cell* **175**, 1430–1442.e17 (2018).
2. R. Li, M. Zhu, J. Li, *et al.*, "Precise segmentation of densely interweaving neuron clusters using G-Cut," *Nat. Commun.* **10**, 1549 (2019).
3. D. M. Chow, D. Sinefeld, K. E. Kolkman, *et al.*, "Deep three-photon imaging of the brain in intact adult zebrafish," *Nat. Methods* **17**, 605–608 (2020).
4. K. Podgorski and G. Ranganathan, "Brain heating induced by near-infrared lasers during multiphoton microscopy," *J. Neurophysiol.* **116**, 1012–1023 (2016).
5. T. Wang, C. Wu, D. G. Ouzounov, *et al.*, "Quantitative analysis of 1300-nm three-photon calcium imaging in the mouse brain," *Elife* **9**, e53205 (2020).
6. P. P. Laissue, R. A. Alghamdi, P. Tomancak, *et al.*, "Assessing phototoxicity in live fluorescence imaging," *Nat. Methods* **14**, 657–661 (2017).
7. H. Dana, Y. Sun, B. Mohar, *et al.*, "High-performance calcium sensors for imaging activity in neuronal populations and microcompartments," *Nat. Methods* **16**, 649–657 (2019).
8. N. Samantaray, I. Ruvo-Berchera, A. Meda, *et al.*, "Realization of the first sub-shot-noise wide field microscope," *Light Sci. Appl.* **6**, e17005 (2017).
9. M. A. Taylor and W. P. Bowen, "Quantum metrology and its application in biology," *Phys. Rep.* **615**, 1–59 (2016).
10. C. A. Casacio, L. S. Madsen, A. Terrasson, *et al.*, "Quantum-enhanced nonlinear microscopy," *Nature* **594**, 201–206 (2021).
11. W. Meinel, J.-C. Olivo-Marin, and E. D. Angelini, "Denoising of microscopy images: a review of the state-of-the-art, and a new sparsity-based method," *IEEE Trans. Image Process.* **27**, 3842–3856 (2018).
12. S. R. Schultz, C. S. Copeland, A. J. Foust, *et al.*, "Advances in two-photon scanning and scanless microscopy technologies for functional neural circuit imaging," *Proc. IEEE* **105**, 139–157 (2016).
13. M. Jing, Y. Li, J. Zeng, *et al.*, "An optimized acetylcholine sensor for monitoring in vivo cholinergic activity," *Nat. Methods* **17**, 1139–1146 (2020).
14. C. Xu, "An adaptive excitation source for multiphoton deep and fast imaging," *Proc. SPIE* **11971**, 1197102 (2022).
15. J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.* **11**, 670–684 (2002).
16. K. Dabov, A. Foi, V. Katkovnik, *et al.*, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**, 2080–2095 (2007).
17. U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 2774–2781.
18. W. Aili, Z. Ye, M. Shaoliang, *et al.*, "Image denoising method based on curvelet transform," in *2008 3rd IEEE Conference on Industrial Electronics and Applications* (2008), pp. 571–574.

19. A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 1692–1700.
20. M. Aittala and F. Durand, "Burst image deblurring using permutation invariant convolutional neural networks," in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 731–747.
21. B. Du, Q. Wei, and R. Liu, "An improved quantum-behaved particle swarm optimization for endmember extraction," *IEEE Trans. Geosci. Remote Sens.* **57**, 6003–6017 (2019).
22. R. Jaroensri, C. Biscarrat, M. Aittala, *et al.*, "Generating training data for denoising real RGB images via camera pipeline simulation," *arXiv*, [arXiv:1904.08825](https://arxiv.org/abs/1904.08825) (2019).
23. X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proceedings of the 30th International Conference on Neural Information Processing Systems* (2016), pp. 2810–2818.
24. M. Weigert, U. Schmidt, T. Boothe, *et al.*, "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nat. Methods* **15**, 1090–1097 (2018).
25. C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," *Nat. Methods* **16**, 1215–1225 (2019).
26. L. Fang, F. Monroe, S. W. Novak, *et al.*, "Deep learning-based point-scanning super-resolution imaging," *Nat. Methods* **18**, 406–416 (2021).
27. H. Wang, Y. Rivenson, Y. Jin, *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods* **16**, 103–110 (2019).
28. J. Lehtinen, J. Munkberg, J. Hasselgren, *et al.*, "Noise2Noise: learning image restoration without clean data," *arXiv*, [arXiv:1803.04189](https://arxiv.org/abs/1803.04189) (2018).
29. A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 2129–2137.
30. J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," in *International Conference on Machine Learning* (2019), pp. 524–533.
31. J. Xu, Y. Huang, M.-M. Cheng, *et al.*, "Noisy-As-Clean: Learning self-supervised denoising from corrupted image," *IEEE Trans. Image Process.* **29**, 9316–9329 (2020).
32. T. Huang, S. Li, X. Jia, *et al.*, "Neighbor2Neighbor: self-supervised denoising from single noisy images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 14781–14790.
33. Y. Quan, M. Chen, T. Pang, *et al.*, "Self2Self with dropout: learning self-supervised denoising from single image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 1890–1898.
34. T. Huang, S. Li, X. Jia, *et al.*, "Neighbor2Neighbor: a self-supervised framework for deep image denoising," *IEEE Trans. Image Process.* **31**, 4023–4038 (2022).
35. J. Lequyer, R. Philip, A. Sharma, *et al.*, "A fast blind zero-shot denoiser," *Nat. Mach. Intell.* **4**, 953–963 (2022).
36. M. Eom, S. Han, P. Park, *et al.*, "Statistically unbiased prediction enables accurate denoising of voltage imaging data," *Nat. Methods* **20**, 1581–1592 (2023).
37. M. Zontak and M. Irani, "Internal statistics of a single natural image," in *Proceedings of the Conference on IEEE Computer Vision and Pattern Recognition* (2011), pp. 977–984.
38. https://github.com/utixx/SelfMirror/blob/main/Supplementary%20Information/Supplementary_Information_of_SelfMirror.pdf.
39. A. Song, J. L. Gauthier, J. W. Pillow, *et al.*, "Neural anatomy and optical microscopy (NAOMi) simulation for evaluating calcium imaging methods," *J. Neurosci. Methods* **358**, 109173 (2021).
40. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 9446–9454.
41. L. Qu, S. Zhao, Y. Huang, *et al.*, "Self-inspired learning for denoising live-cell super-resolution microscopy," *Nat. Methods* **21**, 1895–1908 (2024).
42. M. Zhou, H. Yu, J. Huang, *et al.*, "Deep Fourier up-sampling," in *Proceedings of the 36th International Conference on Neural Information Processing Systems* (2022), pp. 22995–23008.
43. J. Schindelin, I. Arganda-Carreras, E. Frise, *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nat. Methods* **9**, 676–682 (2012).
44. L. Streich, J. C. Boffi, L. Wang, *et al.*, "High-resolution structural and functional deep brain imaging using adaptive optics three-photon microscopy," *Nat. Methods* **18**, 1253–1258 (2021).
45. C. H. McCollough, A. C. Bartley, R. E. Carter, *et al.*, "Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge," *Med. Phys.* **44**, e339–e352 (2017).
46. M. Berning, K. M. Boergens, and M. Helmstaedter, "Segem: efficient image analysis for high-resolution connectomics," *Neuron* **87**, 1193–1206 (2015).
47. <https://github.com/utixx/SelfMirror>.