

RESEARCH

Open Access



Surmounting photon limits and motion artifacts for biological dynamics imaging via dual-perspective self-supervised learning

Binglin Shen¹, Chenggui Luo¹, Wen Pang², Yajing Jiang³, Wenbo Wu³, Rui Hu¹, Junle Qu¹, Bobo Gu² and Liwei Liu^{1*}

*Correspondence:
liulw@szu.edu.cn

¹ Key Laboratory of Optoelectronic Devices and Systems of Guangdong Province and Ministry of Education, College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

² Med-X Research Institute and School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China

³ Department of Chemistry, Institute of Molecular Aggregation Science, Tianjin University, Tianjin 300072, China

Abstract

Visualizing rapid biological dynamics like neuronal signaling and microvascular flow is crucial yet challenging due to photon noise and motion artifacts. Here we present a deep learning framework for enhancing the spatiotemporal relations of optical microscopy data. Our approach leverages correlations of mirrored perspectives from conjugated scan paths, training a model to suppress noise and motion blur by restoring degraded spatial features. Quantitative validation on vibrational calcium imaging validates significant gains in spatiotemporal correlation (2.2×), signal-to-noise ratio (9–12 dB), structural similarity (6.6×), and motion tolerance compared to raw data. We further apply the framework to diverse *in vivo* experiments from mouse cerebral hemodynamics to zebrafish cardiac dynamics. This approach enables the clear visualization of the rapid nutrient flow (30 mm/s) in microcirculation and the systolic and diastolic processes of heartbeat (2.7 cycle/s), as well as cellular and vascular structure in deep cortex. Unlike techniques relying on temporal correlations, learning inherent spatial priors avoids motion-induced artifacts. This self-supervised strategy flexibly enhances live microscopy under photon-limited and motion-prone regimes.

Introduction

Living systems exhibit complex behaviors spanning spatial and temporal scales, from cardiac pulsation to neuronal spiking. Capturing and quantifying these fast biological dynamics can provide fundamental insights into physiology, development, neural function, biomechanics and more [1–4]. Commonly used two-photon laser-scanning microscopy (TPLSM) furnishes biologists with a practical tool for deep interrogation of biological structures and functions with high optical resolution and penetration depth [4–7]. Nevertheless, achieving clear visualization of rapid biological processes with TPLSM is fundamentally challenging. For example, imaging the rapidly beating heart requires an exposure time of less than 3 ms with enough detected photons to

minimize motion blur [8]. This exceeds the typical frame rate (30 Hz) of a resonant-scanning TPLSM. Photon limits, motion artifacts, and signal loss during acquisition degrade fidelity [9–12], while hardware constraints impose inherent trade-offs between imaging speed, field of view (FOV), resolution, and signal-to-noise ratio (SNR) [13–15]. Surmounting these barriers could drive transformative discoveries across the life sciences.

Advanced computational processing can extract subtle features difficult to resolve optically [16–18]. Data-driven deep learning has shown early promise in overcoming limitations of optical devices through software enhancements [19] and honing insights from microscopy data [20]. This approach provides a more flexible framework that can bypass explicit noise modeling [21, 22] and directly learn image features, enabling reliable mapping from corrupted, low-contrast data to high-quality approximations [23–28]. However, most techniques rely on registered raw input and ground truth image pairs for supervised training, presenting bottlenecks in scalability. Self-supervised learning is emerging as a promising method to circumvent this requirement by exploiting structure within the data itself to train neural networks [29–32]. Such built-in redundancy manifests in numerous forms, e.g. multimodal correlations, neighbor interpolation, noise consistency, temporal continuity, etc. The self-supervised learning overcomes the need for paired data exemplars, recovering useful information from signals obscured by photon noise. Nevertheless, for high-speed imaging of dynamically evolving biological processes, consistent temporal relationships for self-supervised training could be inaccessible, limiting their applications in denoising and deblurring of rapid biodynamics.

Here, we demonstrate DeepBID, a self-supervised paradigm for biodynamics imaging denoising and deblurring under challenging *in vivo* conditions. We focus specifically on harnessing the spatiotemporal relationships of microscopy data constructed by the bidirectional scan lines in TPLSM, and adapted a lightweight and efficient 3D model [32] to restore degraded spatial correlations by mapping between the conjugated scan lines. TPLSM sequentially samples the same structures with different noise distributions for mirrored perspective. This avoids potential spatiotemporal artifacts when utilizing temporal correlations across frames, which may confuse neural network mappings for dynamic imaging targets. Importantly, existing microscopes can readily provide suitable training data, facilitating adoption. We show the effectiveness of the approach in mitigating noise and motion artifacts in vibrational neuronal and astrocytic imaging, as well as reinforcing visualization of rapid nutrient flow in microcirculation and systolic and diastolic processes of cardiac dynamics.

Quantitative analysis demonstrates significant boosts in trace correlations, SNR, structural similarity, motion robustness, and segmentation fidelity over raw data. We highlight diverse *in vivo* contexts from neural activity, hemodynamics with cellular resolution to cardiac dynamics during rapid beating. Thereby, this work establishes a flexible computational imaging platform to strengthen live microscopy under challenging photon-limited and motion-prone regimes. By learning to extract maximally useful information from the signals directly registered by the sensor, our approach can relax hardware constraints to open up new imaging capabilities for morphological

and functional interrogation of biodynamics. The data-efficient training paradigm is readily scalable, while the framework is sufficiently generalizable for enhancing data from a variety of modalities.

Results and discussion

Dual-perspective self-supervised learning and testing

Microscopic imaging typically involves the sequential raster scanning of lines in both forward and backward trajectories. Bidirectional scanning demands meticulous alignment between the starting and ending positions across both scan directions to avoid undesirable jagged effects. This alignment challenge often leads to the discarding of information from one scanning direction, particularly noticeable in kHz resonant scanning processes, to attain high-quality images. To address this drawback, we introduce DeepBID with mirrored perspective self-supervised learning (MP-SSL) that leverages the information acquired from both forward and backward scan paths to construct a noise-diverse yet content-consistent dataset (Fig. 1a). Notably, MP-SSL allows full utilization of the bidirectional data for enhancement, overcoming the need to discard one scan direction. This strategy effectively bridges the temporal differences between adjacent frames, offering a stark contrast to the time-lapse perspective self-supervised learning (TP-SSL), which primarily learns the similarity between adjacent frames and shows excellent denoising performance in calcium imaging [31, 32]. The ensuing lightweight 3D network then harnesses this dataset to learn and restore the intricate spatiotemporal relationships inherent within images, thus enhancing clarity and quality. The essential components of the TPLSM system, the architecture of the 3D network, and the construction of the training dataset are elucidated in Fig. S1. The example low-SNR, corrupted images of astrocytes situated at different depths within the mouse brain were restored using MP-SSL, yielding significantly enhanced visibility and quality (Fig. 1b, Visualization 1). This network inference remained untrained, relying solely on the utilization of the pretrained synthetic data model that follows.

To quantitatively demonstrate the denoising capabilities of the model and its ability to remove motion artifacts, we initiated our exploration by applying the model to oscillatory calcium imaging data. This endeavor involved the creation of a comprehensive biological model incorporating various components, including blood vessels, neurons, and background dendrites/axons (Fig. S2). To enhance the realism of the synthetic recordings, we meticulously factored in optical propagation considerations, encompassing the point spread function (PSF), as well as the scanning dynamics intrinsic to the TPLSM system (Fig. 1c). This intricate process yielded synthetic noise-free recordings with hyperrealistic pixel distribution [32–34]. For the purpose of fostering a symmetric learning paradigm, we simulated a bidirectional scanning methodology. This entailed having the backward scan precisely mirror the trajectory of the corresponding forward scan, thus inducing the emergence of dual perspective within the resultant image (Fig. 1d). To facilitate subsequent MP-SSL training, we strategically allocated twice the number of pixels in the y -direction as compared to the x -direction.

The next stage involved the injection of mixed Poisson-Gaussian noise, taking into account the three primary sources of noise (dark noise, shot noise and readout noise) [35]. This process was meticulously executed on both the fluorescent neurons and the

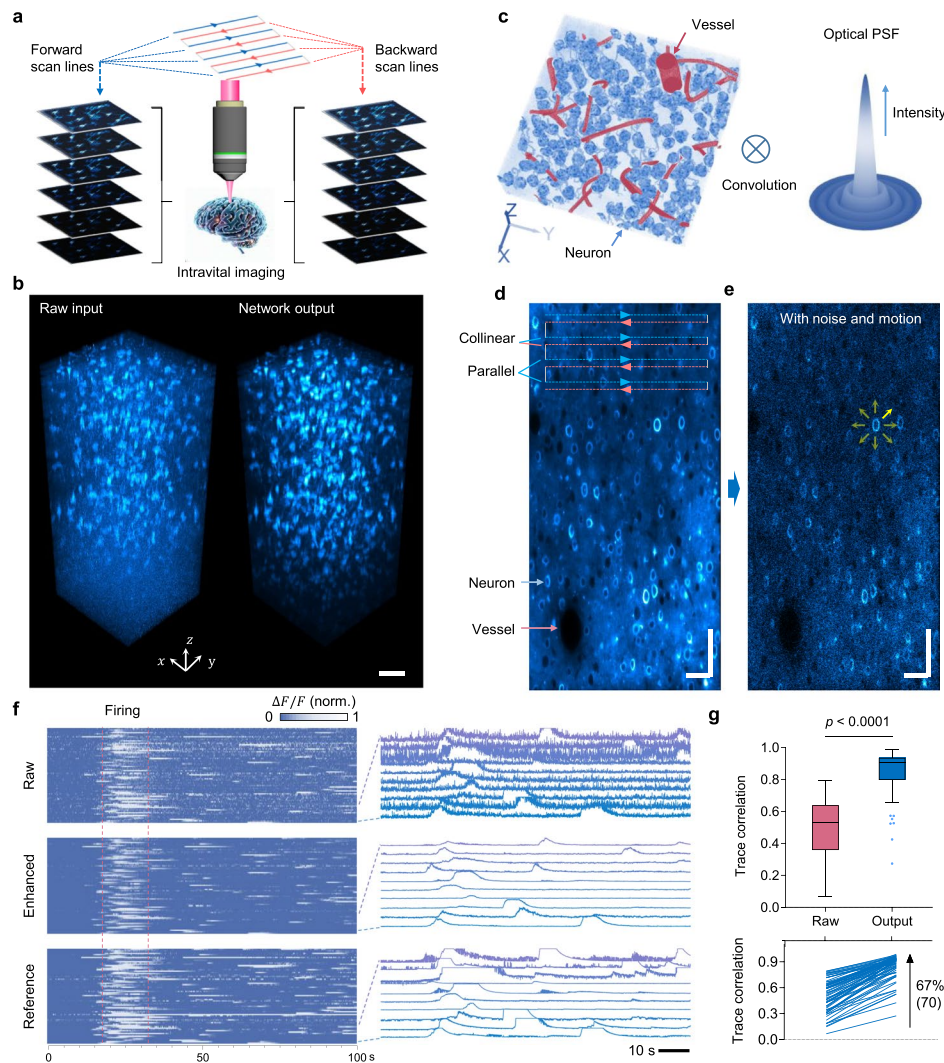


Fig. 1 Principle and performance validation of DeepBID. **a**, Diagram of the data construction. The raw stack from *in vivo* brain imaging, comprising forward and backward scan lines, is segmented into input and target sub-stacks for 3D network training, a process seamlessly integrated into the model. Post-training, the pretrained model enables direct testing of unidirectional or bidirectional scan images without division. **b**, Example test outcomes showcasing astrocyte images with vibrations. The noise in the volume was suppressed, rendering deeper structures more distinct. **c**, Left, a synthetic distribution of neurons (blue) and vessels (red). Right, convolution with the point spread function of the two-photon system. **d**, Generated reference image using bidirectional scanning for evaluating image quality metrics. Forward (blue dashed line) and backward (red dashed line) scanning paths are collinear, ensuring high semantic relevance for self-supervised learning. Scanning lines in the same direction remain parallel. These noise-free images serve as benchmarks for network performance assessment. **e**, Raw data constructed by introducing mixed Poisson-Gaussian noise and motion drifts (indicated by yellow arrows). **f**, Long-timescale calcium fluctuations evoked by 70 isolated neurons. All traces were normalized, with prominent fringes delineated between red dashed lines. Zoomed-in traces are featured in the right panel. **g**, Top, Tukey box-and-whisker plot illustrating Pearson correlations of calcium traces extracted from enhanced data versus raw noisy data with a two-tailed Wilcoxon matched-pairs signed rank test ($n = 70$). Bottom, Correlation augmentation post-denoising. Each line corresponds to a distinct recording. Scale bars, 50 μm in **b**, **d**, and **e**

non-fluorescent vasculature recordings, thus emulating the authentic complexities of a real microscopy scene (Fig. 1e). Unlike previous works focused on motionless calcium imaging, we deliberately introduced realistic motion artifacts into the data (Fig. 1e).

These artifacts stemmed from factors such as instrument vibrations, respiration, and cardiac activity [36–38]. This addition inherently induced a scenario where neurons exhibited random “jumps,” leading to misalignments between adjacent frames, especially at the high imaging speed of 30 Hz. Each frame underwent random motion-induced shifts, rendering it arduous to align algorithmically. Given this intricate blend of factors, we synthesized the highly realistic two-photon image data of neurons by amalgamating convolution acquisition, noise corruption, and motion blurring. We fed the generated synthetic records into the network to learn the mapping between two perspectives of the conjugated scan paths for self-supervised denoising and deblurring, without access to pristine data except for benchmarking and quantification.

We manually segmented neurons to extract time-dependent calcium traces (Methods) on the raw data, MP-SSL inferred data, and pristine data. At long timescales, calcium traces extracted from the raw data exhibit a high noise level, leading to difficulty in signal extraction for the calcium spikes. However, the denoised data exhibit enhanced congruence with the noise-free ground truth, especially at low SNRs (Fig. 1f). Using MP-SSL, calcium traces are clearly visible, and smoother signals are extracted from neurons heavily affected by noise and motion artifacts, surpassing the corresponding high SNR reference in some cases. The calcium traces produced through the application of MP-SSL may exhibit greater levels of visibility and smoothness compared to those associated with the high SNR reference example. This is because any image noise of zero mean cannot be learned by the network, and the network can only learn the map of the clear data, so the self-supervised method may obtain smoother data than the data with a high SNR. After denoising, there is a substantial improvement in Pearson correlation when compared to the raw traces (depicted in Fig. 1g), showcasing an average increase of 67%. Remarkably, even calcium traces affected by motion blur could be effectively recovered from the initially noisy raw data, with concurrent attenuation of background noise. These observations collectively underscore the potential of self-supervised spatiotemporal enhancement to markedly enhance the precision of neural signal extraction, thereby facilitating the intricate analysis of neural circuits.

Denoising and deblurring motion-affected imaging data

We conducted a comparative analysis between MP-SSL and the chosen TP-SSL method (Fig. 2a–d, Visualization 2). Under a large image distortion caused by motion, the TP-SSL restoration of neurons exhibited noticeable blurring, while MP-SSL remained largely unaffected by motion artifacts due to its primary focus on learning intra-frame spatiotemporal correlations within the 16-kHz bi-scan paths (Methods). We extracted an individual neuron with spontaneous neural activity (zoom-in of Fig. 2a–d), which exhibited a drifting pattern indicated by the yellow arrow, attributed to vibrations. The TP-SSL approach suffered from spatiotemporal confusion and ghosting in the restored neurons, stemming from its learning of inter-frame dynamic correlations at a mediocre frame rate. In contrast, MP-SSL could clearly retrieve individual neurons and calcium signals, bypassing motion blurring and enriching fluorescence photons. Moreover, MP-SSL was able to resolve previously invisible neurons with significantly lower SNR, in contrast to the blurred visualization achieved with TP-SSL (Fig. S3a–d). This facilitated

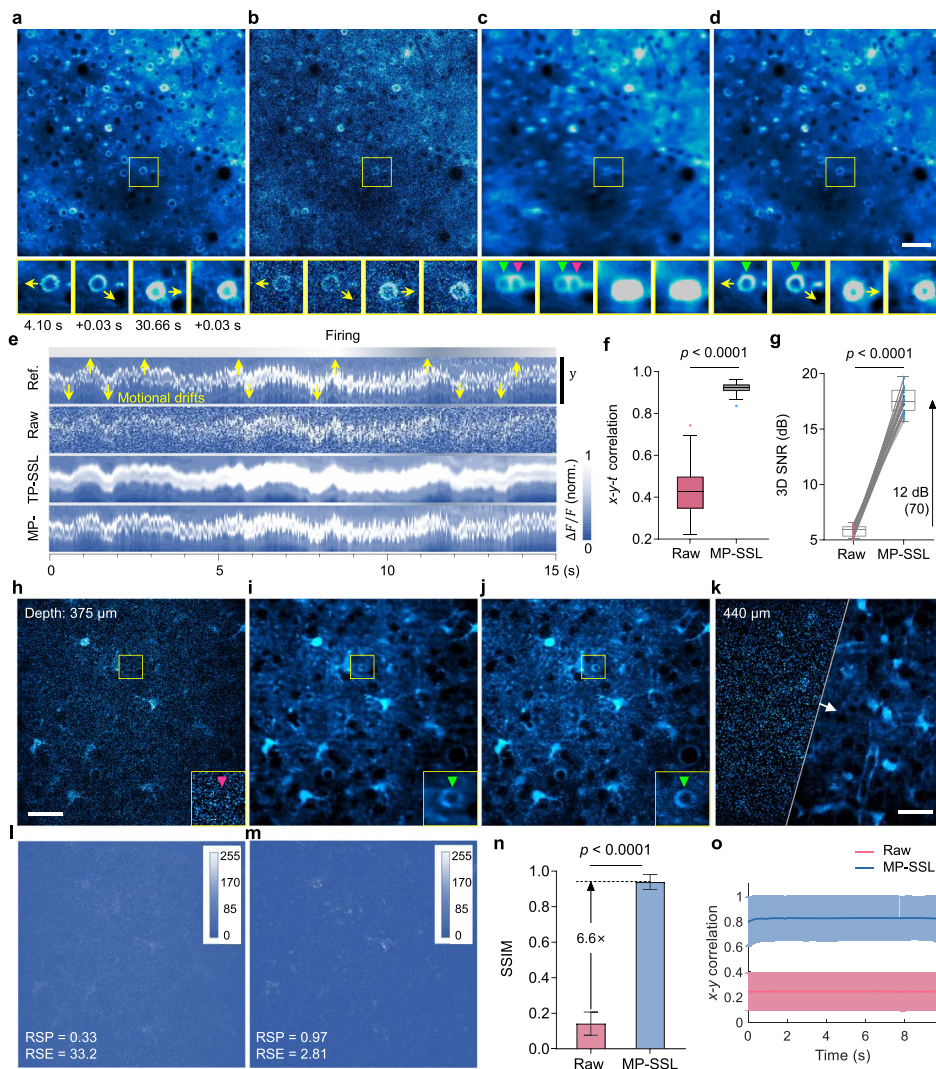


Fig. 2 Deep learning-enhanced motion-affected synthetic and experimental data. **a**, Synthetic portrayal of spontaneous calcium dynamics within the mouse cortex without noise. **b**, Raw data degraded from the clean image. TP-SSL restoration (**c**) and MP-SSL restoration (**d**) of the low-SNR recording. The magnified neuron in the lower sections, depicted at various time points, is highlighted within yellow boxes, with the moving direction indicated by yellow arrows. Correct restoration via MP-SSL is marked by green arrowheads, while ghosted neurons resulting from TP-SSL are denoted by magenta arrowheads. **e**, The **y-t** views of the neuron exhibiting calcium signals and temporal shifts within a 15-second window. Also, see the orthogonal **x-t** views in Fig. S4. **f**, Tukey box-and-whisker plot illustrating spatiotemporal correlation changes in calcium data pre and post-denoising ($n = 70$ **x-y-t** stacks). **g**, Improvement of the 3D SNR. Each line represents 1 of 70 spatiotemporal data. The overlay of a statistical Tukey box-and-whisker plot provides context. Correlation and SNR calculations reference clean stacks. **h**, Experimentally captured astrocyte image with a low SNR using the bidirectional resonant scan TPLSM. **i**, Image restored from the low-SNR image using the pretrained calcium MP-SSL network. **j**, Temporal average of the raw input frames ($n = 300$). Yellow boxes indicate the extracted neuron magnified in the insets. Magenta arrowhead indicates the vague neuron (**h**), while green arrowheads point to the clear neuron (**ij**). **k**, Astrocyte images at a larger cortical depth. The TPLSM input and MP-SSL result are shown in the left and right portion, respectively. Error maps, RSP, and RSE values for raw input (**l**) and network output (**m**) were calculated in relation to the temporal average image. Column bar graph of SSIM (**n**) and **x-y** spatial correlation (**o**) calculated between each frame and the 300-frame average ($n = 51$ stacks with 300 temporal frames per stack from a depth range of 300–550 μm). Two-tailed Wilcoxon matched-pairs signed rank tests were applied between the raw input and MP-SSL output in **f, g** and **n**, and mean \pm standard deviation (SD) was shown in **n** and **o**. Scale bars, 50 μm in **d** and 30 μm in the other images

a more precise extraction of spatial and temporal components of neurons (Fig. S3e–h) using constrained non-negative matrix factorization (CNMF [39, 40]).

Presented in Fig. 2e is the y - t orthogonal view of the neuron, centered around the signal firing instance. The TP-SSL method erased finer details of the neuron contour due to the pronounced oscillations in the y direction (see x direction in Fig. S4a), whereas MP-SSL exhibited an impressive alignment with the pristine data. We quantified x - y - t correlations by employing the neurons extracted from the data, as showcased in Fig. 2f. In comparison to the temporal (t) correlation of calcium traces, the x - y - t correlation simultaneously assessed spatial drifts of neurons and temporal dynamics of the calcium signal. MP-SSL significantly enhanced the overall correlation by 114% over the input (0.92 vs. 0.43), resulting in lower restoration variance. To quantify motion robustness, we computed the correlation enhancements of the x - t slices, the y - t slices, and the x - y slices to assess spatial structure preservation and temporal trace consistency before and after denoising (Fig. S4b). The improved correlations demonstrate the effectiveness of the method in mitigating motion-induced degradation. Specifically, x - y correlation shows maintenance of spatial alignment, while x - t and y - t correlation validates recovering the temporally consistent calcium dynamics. Moreover, the 3D SNR (x - y - t) of the MP-SSL restoration experienced a notable improvement of 12 dB (Fig. 2g), indicating remarkable noise suppression and a remarkable spatiotemporal recovery of neural signal correlations. Note S1 details that MP-SSL also achieved a significant enhancement in the 3D structural similarity index measure (SSIM).

We further demonstrated the transfer learning capability of MP-SSL by applying the pretrained denoising model to the raw images of experimentally captured astrocytes (Fig. 2h–j). Despite the real noise profiles differing from the simulated training data, the deep network significantly reduced noise after transfer learning, enhancing image quality. The motion-affected dendritic spines and branches (Fig. 2h) were clearer upon restoration (Fig. 2i), aligning well with the average over the raw 300 frame input (Fig. 2j). This significant reduction in noise improved image quality and greatly reduced the number of scans required to obtain high-SNR images. At a larger cortical depth (Fig. 2k), our approach effectively resolved the invisible structural details and textures that were originally submerged in noise, thereby providing a clearer view of cellular morphology. We also denoised the $2\times$ zoom-in images (all astrocyte images were restored using the pretrained calcium MP-SSL network). The achieved sharpness in the 3D and optical-section restoration at different depths (Fig. S5) showed that the structural and dynamic information of nerve and glial cells were clearly resolved at the network output, agreeing well with the high-SNR cumulative images.

Additionally, we presented the intensity profiles along the terminal branch of the astrocyte in Fig. S6. The irregular profile for the input image shows a low distinguishability of noise and informative signals, which results in difficulty in structure visualization for the weak-signal regions. Nevertheless, these unwanted fluctuations were effectively removed during the restoration process, preserving tissue structure features well. For verification, we quantified the error mapping of the raw input image (Fig. 2h) in comparison with that of MP-SSL output image (Fig. 2i) concerning the temporally average reference (Fig. 2j), as shown in Fig. 2l, m. The calculated error maps, resolution scaled error (RSE), and resolution scaled Pearson coefficient (RSP) [41, 42] (Methods) reveal

that MP-SSL did not introduce noticeable restoration artifacts or blurring, as evidenced by the significantly high RSP of 0.97, compared to the original 0.33. The network output results had a much lower level of spatiotemporal mismatch error and high correlation with the high-SNR reference, even when considering large depths over 400 μm (Fig. S7). Note that the 300-frame average was used for the metric calculations since pristine reference was unattainable and the selected images for evaluating the metrics were motionless and therefore highly similar to the average frame. The network inference achieved a 6.6-fold enhancement in SSIM (0.14 for raw input and 0.94 for MP-SSL, Fig. 2n), as well as a 3.4-fold spatial correlation between each output frame and the temporal average (Fig. 2o) at the cortical depths from 300 to 550 μm . Thereby, MP-SSL can highly suppress noise fluctuations affecting the visualization of astrocyte and neuronal network without “freezing” their spatiotemporal dynamics, and avoid restoration artifacts and blurring. This highlights the generalizability of the MP-SSL approach to experimental imaging data beyond the synthesized training distribution.

Visualizing and measuring rapid hemodynamics

Observing hemodynamics and understanding microcirculation provides crucial insights into cerebral vascular health and disease. Abnormal hemodynamics have been implicated in conditions such as atherosclerosis, hypertension, and aneurysms [43–45]. Conventional hemodynamic visualization is susceptible to motion artifacts attributed to swift flow, resulting in significant inter-frame hemodynamic disparities constrained by inadequate photon availability. To demonstrate the efficacy and adaptability of our approach to experimentally acquired hemodynamic data, we conducted high-speed imaging of mouse brain vessels utilizing a custom TPLSM setup in conjunction with a synthesized contrast agent (Methods). Utilizing the low-SNR data as the input for the MP-SSL network learning, we accounted for the $<1\text{-}\mu\text{m}$ mismatch between forward and backward scan paths. Synchronized high-SNR data, on the other hand, were utilized for a quantitative assessment of denoising efficacy. Post-network inference, both TP-SSL and MP-SSL proficiently eliminated mixed noise from bidirectionally scanned time-lapse stacks (Fig. 3a–d, Visualization 3). Although both methods (Fig. 3c–d) delineated the outlines of low-SNR vessels (Fig. 3b), TP-SSL, due to substantial inter-frame shifts, struggled to distinguish intricate hemodynamic dynamics such as erythrocyte motility and nutrient flow in the microcirculation [45]. In contrast, MP-SSL precisely resolved these processes, effectively capturing particle size and position in excellent agreement with the pristine reference (Fig. 3a). By leveraging the distance l traversed by the particle over the time interval t , the flow velocity was computed to be approximately 0.6 mm/s, based on the denoised frames. This insight provides a quantitative measure of the resolved microcirculation through our approach. Furthermore, the intensity profiles across the cross-section of the small vessel in Fig. S8 highlighted the diminished visibility of the raw data, which underwent substantial enhancement after the network inference, strongly aligning with the averaged data. However, it is important to note that intravascular transport within the averaged image remains concealed.

At higher flow velocities, substances within the vessels exhibit motion blurring and trailing artifacts in the raw data due to their rapid displacement across scan lines. We employed residence time line scanning (RTLs [46]) to directly analyze flow velocity by

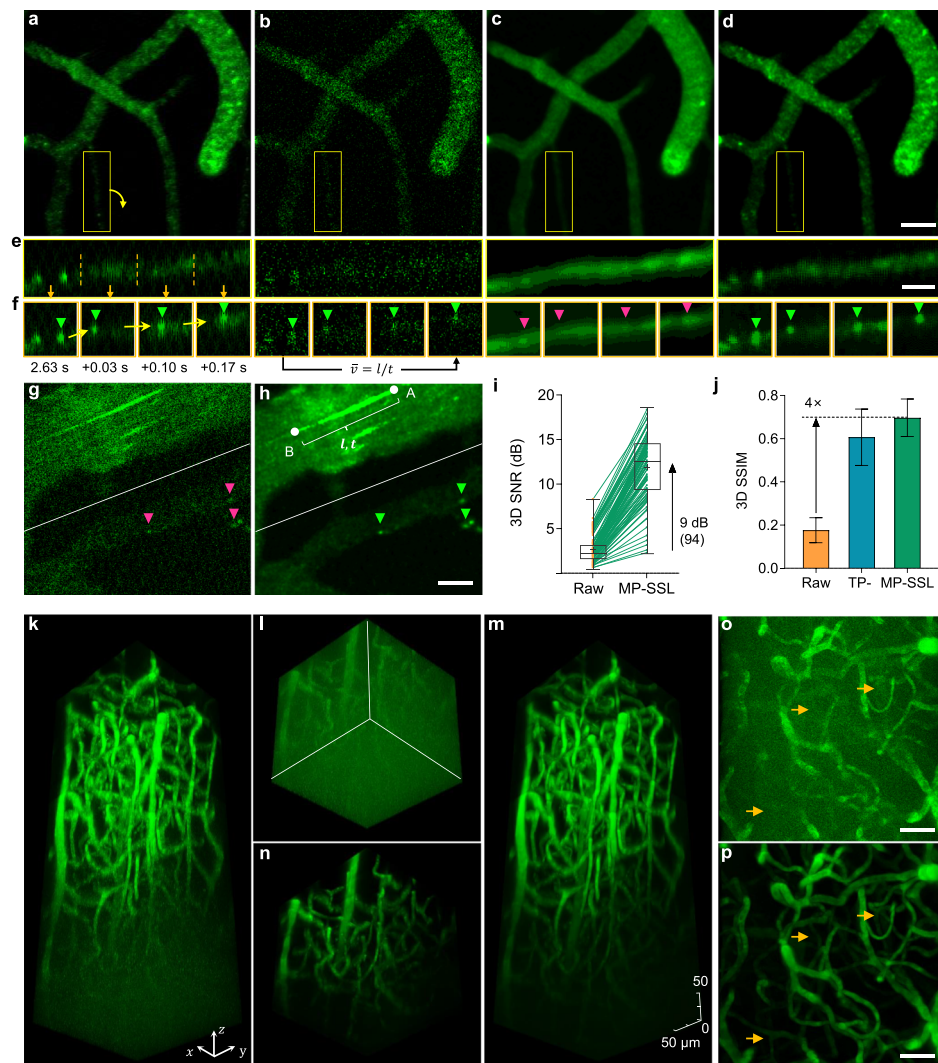


Fig. 3 Deep learning-enhanced high-speed hemodynamics imaging. **a**, Mouse cerebrovascular images captured by the TPLSM with a high SNR. **b**, Quality degradation with the depth-related mixed Poisson-Gaussian noise as raw data, which was restored using TP-SSL (**c**) and MP-SSL (**d**). Magnified views of the yellow boxed regions show an out-of-focus vessel (**e**), with its continuous segments at distinct time points displayed in **f**, with corresponding segments indicated by orange arrows and nutrient flow in microcirculation by yellow arrows. Notably, MP-SSL resolved the instantaneous positions of nutrient particles (green arrowheads), which remain indistinct (magenta arrowheads) in TP-SSL restoration. Flow velocity is derivable from travel distance (l) against travel time (t) calculations. **g**, Raw image depicting rapid hemodynamics within larger brain vessels. **h**, The restoration outcome using MP-SSL, allowing computation of high flow velocity via single-frame l and t values using the RTLS technique. Faint vessels (**g**) in the deeper layer (separated by the white lines) were restored clearly (**h**). MP-SSL distinguishes previously unclear substances (magenta arrowheads), now evident (green arrowheads). **i**, Improvement of the 3D SNR. Each line represents 1 of 92 spatiotemporal stacks, accompanied by an overlaid Tukey box-and-whisker plot for statistical context. **j**, Column bar graph of the 3D SSIM with mean \pm SD. **k**, Volumetric vasculature reconstruction using experimentally captured time-lapse series at 5 $\mu\text{m}/\text{stack}$. Brightened contrast in the lower section (deeper tissue) is revealed in **l**. **m**, The denoised brain volume, with the deep portion displayed in **n**. Volumes are reconstructed for maximum projection and are also projected to 2D (**o**, **p**) for dynamic temporal observation. Orange arrows point out instances of the initially obscured vessels (**o**) significantly influenced by noise, which were efficiently restored through network inference (**p**). Scale bars: 10 μm in **e** and 30 μm in the remaining images

scanning a line at an arbitrary angle to the vessel (Fig. 3g, h). Calculating the velocity from points AB yielded a remarkable 30 mm/s, which is 50 times faster than that seen in Fig. 3f. In the raw data, structural features are barely distinguishable and appear as faint shadows at the bottom of Fig. 3g, h. In contrast, the MP-SSL restoration dramatically enhances visibility, effectively recovering intricate vessel morphology and hemodynamics (Fig. 3h, Visualization 4), in harmony with the high-SNR sequence (Fig. S9a–c). Substances associated with nutrient flow that were faint in the raw data were now distinctly detected (Fig. 3g, h and Fig. S9e, f). Moreover, the ultrafast learning capability of MP-SSL significantly reduces persistent background artifacts, such as scanning fringe artifacts (SFA) stemming from periodic flickering ambient light and resonant scan coupling [15]. Comparing with the raw data, the SNR experienced an enhancement of approximately 9 dB (11.8 dB for MP-SSL compared with 2.7 dB for the raw data). The SSIM of MP-SSL restoration in relation to the high-SNR sequence reached 0.70, representing a four-fold improvement over the raw data. In contrast, TP-SSL achieved a lower SSIM of 0.60, accompanied by blurred details. Fig. S9d highlights a 1.9-fold improvement in the x - y - t correlation. These enhancements verify the robust visualization of vascular structures and blood flow in the microcirculation at high flow speeds.

Comparative evaluations were undertaken by presenting the denoising results obtained using different models [30–32, 47], which demonstrated restoration artifacts, stagnant dynamics, lower correlation, SNR and SSIM, against the MP-SSL outcomes (Fig. S10). This comparison demonstrates the higher-fidelity restoration achieved by our method for hemodynamic visualization, serving as valuable inputs for video tracking. Note S2 underscores the challenge posed by noise-affected vascular structures for automatic segmentation using the segment and track anything network (SAM-Track) [48, 49]. However, post-noise reduction, these dynamics can be effectively distinguished, thereby enabling real-time multi-object tracking and propagation.

We further conducted extensive vasculature restoration on time-lapse stacks at depths ranging from 310 to 750 μm (Fig. 3k–n, Visualization 5). After MP-SSL denoising, even low-contrast depths beyond 550 μm in the volumetric images (Fig. 3l) experienced a marked signal recovery, vividly demonstrated in Fig. 3n. As a result, vessel structures and real-time dynamic transports were both vividly visualized (Fig. 3o, p, Visualization 6). The cortical parenchyma is permeated by an intricate network of blood vessels, which run approximately parallel and delve into the deeper cortical layers, with the main vessels sending out smaller branches along their course. Remarkably, these invisible small branches were resolved without introducing artifacts using the network (Fig. 4). To quantify the performance of restoration, we calculated the x - y correlation between individual frames and the temporal average for nearly motionless vascular images displaying minimal temporal changes (Fig. S12a–f). In contrast to the original consecutive frames of the raw stack, which were temporally decorrelated and exhibited higher standard errors, the MP-SSL restoration showcased enhanced correlations that approached unity, along with a substantial reduction in errors (Fig. S12g). Thereby, the meticulous clarity in observing and describing the orientation and distribution of vessels and internal hemodynamics holds the potential to greatly enhance the understanding of the pathological mechanisms underlying vascular functional impairment in various brain disorders [50].

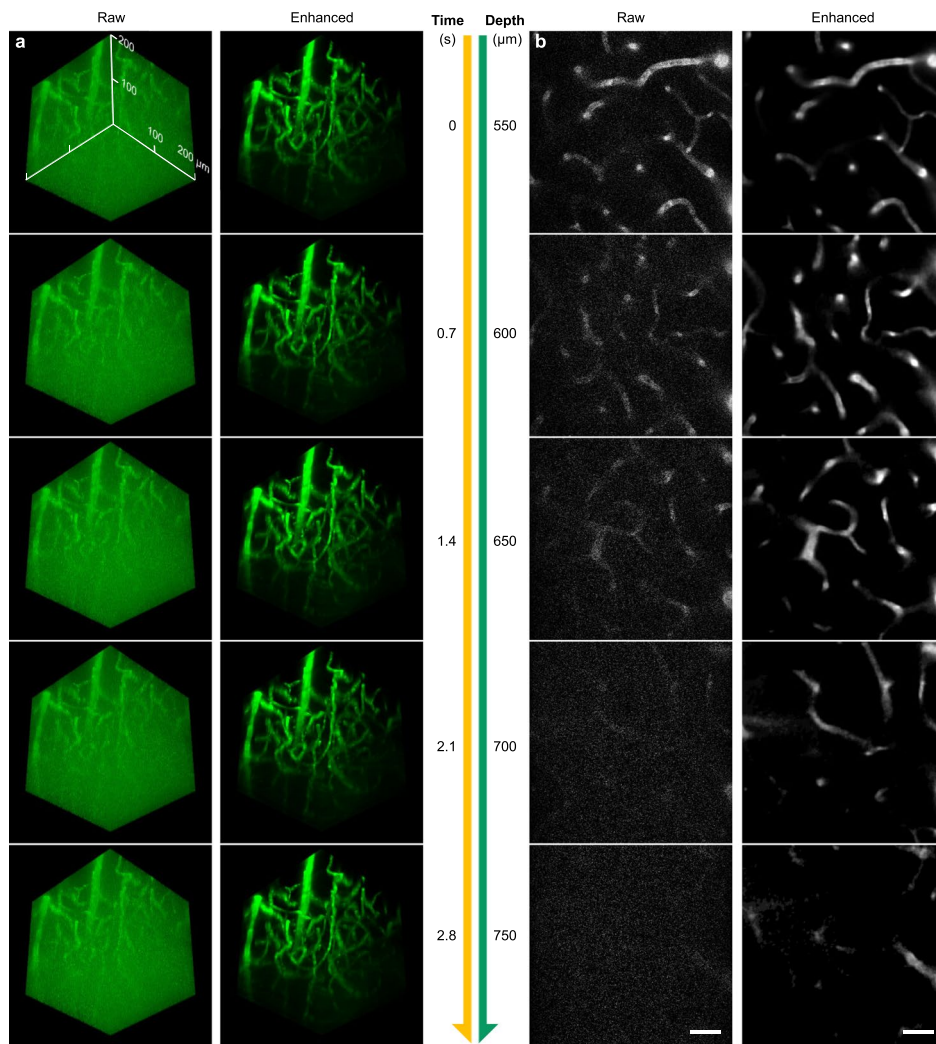


Fig. 4 Enhanced deep angiography with the MP-SSL network. **a**, Temporal evolution of 3D vascular network reconstructions within the depth range of 550–750 μm . **b**, Cross-sectional views of the vascular volumes at various depths, with yellow and green arrows indicating the time and depth axes, respectively. Scale bar: 30 μm

Reconstructing heartbeat and cardiovascular system

Understanding cardiac development and function is of utmost importance, given that the heart is the first organ to form and initiate blood circulation during vertebrate embryo development [8, 51, 52]. The visualization of the rapid and intricate motions of the beating heart presents significant imaging challenges, yet offers invaluable biomedical insights [8, 52, 53]. We extended the application of our model to address the challenge of denoising the photon-limited and motion-prone dynamics of the beating zebrafish heart. In this pursuit, we captured the cardiac dynamics across nearly 28 cycles within a 10-second span (Fig. 5a–d), utilizing transgenic lines that label the vasculature (Methods). The restored clarity achieved through MP-SSL is correspondingly showcased in Fig. 5e–h. The complete cardiac cycle can be observed in Fig. S13a–d and Visualization 8.

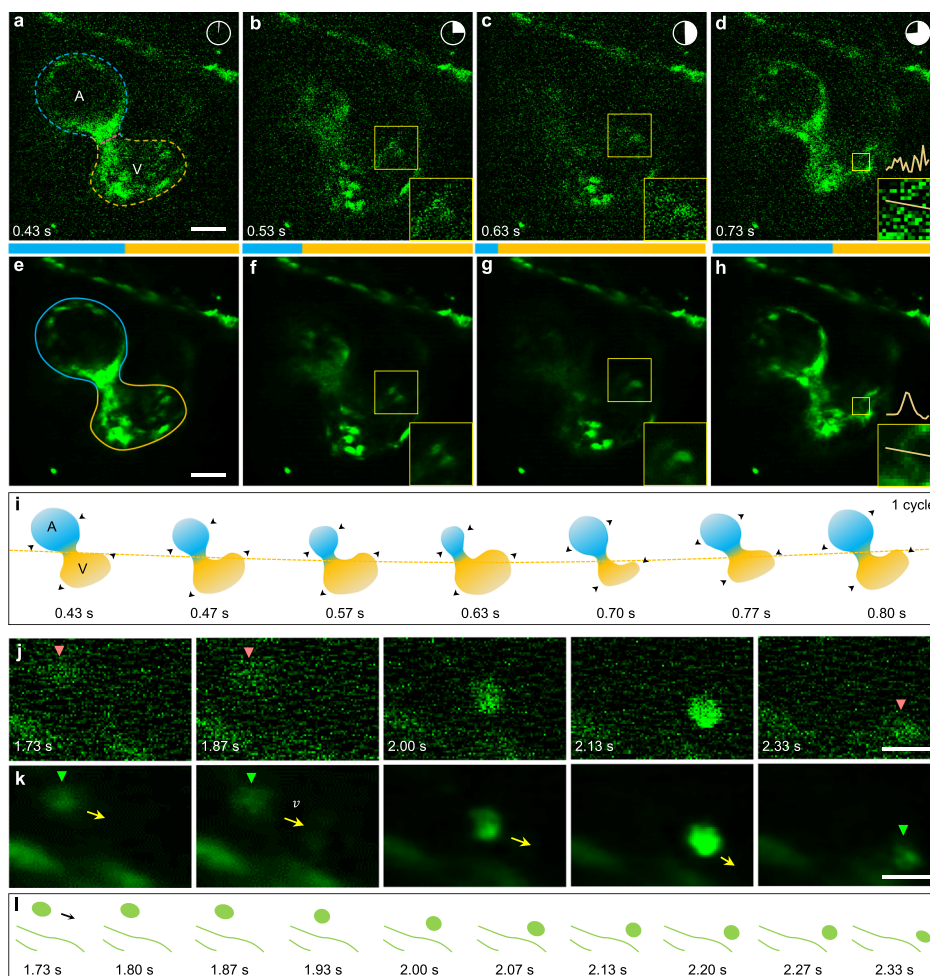


Fig. 5 Spatiotemporal enhancement of cardiac dynamics imaging using DeepBID. Additionally, see Fig. S13. **a–d**, Image of the heartbeat at various phases (with time indicated relatively by clock schematic) during the cardiac cycle. A, atrium; V, ventricle. **e–h**, The corresponding restoration of the heartbeat at various time points using MP-SSL. Blue-yellow dashed line indicates the vague cardiac silhouette, which was resolved (blue-yellow solid line) by the network. The color composition of blue and yellow in the middle bar signifies the atrium and ventricle size ratios during their systolic and diastolic processes. Yellow boxes correspond to the magnified views. Yellow line in the inset images refers to the line of the shown cross-section. **i**, Contours of the atrium and ventricle depicted using the denoised images, with black arrowheads indicating contraction and relaxation directions. **j**, Zoom-in views of an erythrocyte adjacent to the heart, gradually moving out of view. Magenta arrowheads indicate the vague erythrocyte. **k**, Denoised images courtesy of MP-SSL, revealing a clarified erythrocyte (green arrowheads). The motion direction of the erythrocyte is marked by yellow arrows, with v denoting velocity. **l**, Illustrations of erythrocyte locations at distinct instances based on its clear motion (**k**), a challenge with noisy stacks (**j**). Scale bars, 30 μm in **d, h** and 10 μm in **j, k**

Within the context of the beating heart, the cardiac microstructures and weak expression patterns tend to be elusive (Fig. 5b, c), mainly due to the insufficient frame rate of 30 Hz exceeding the exposure time required to minimize motion blur [8]. Nevertheless, the high-SNR network restoration effectively unveiled native cardiac structures across the entire cycle (Fig. 5f, g). Notably, the dynamic changes in the size of the atrium and ventricle can be observed throughout the cycle, indicative of the systolic and diastolic processes of the heartbeat. The self-supervised model recovered fine-details (insets in Fig. 5f–h) otherwise obscured by noise and motion-blurring (insets in Fig. 5b–d). The

noise that overwhelms the signal shown in the intensity profiles in the raw images (Fig. 5d) were highly suppressed by the self-supervised learning inference, producing real connecting filament (Fig. 5h). The background SFA were also suppressed (Fig. S13e, f). Leveraging the denoised images, the contours of the atrium and ventricle, as well as the beating pattern of the heart, were discerned more distinctly (Fig. 5i). This level of clarity has the potential to uncover insights into the cardiac systolic and diastolic global and regional functions.

We conducted a focused examination on an individual erythrocyte adjacent to the heart (Fig. 5i), revealing that at $t = 1.73$ s, it presented with rather indistinct details. Over the next ~ 0.6 seconds, it navigated along the fiber at a leisurely pace, characterized by unsharp and irregular structural features. Eventually, it descended beyond the field of view in the raw images. Post-denoising, the faint outline of the erythrocyte emerged into clarity (Fig. 5j), with precise spatial locations and trajectory meticulously extracted in Fig. 5k. These denoised and deblurred images of erythrocytes complement the comprehensive high-speed hematic data, thereby providing a more holistic understanding of zebrafish cardiac dynamics. In light of these achievements, our high-speed imaging technique, proficient in capturing both the intricate structural dynamics of the heart and its beat patterns, boasts considerable potential for advancing our grasp of physiological processes.

Conclusions

This study demonstrates a powerful self-supervised learning framework for denoising and deblurring of biological imaging data corrupted by noise and motion artifacts. The proposed approach focuses on exploiting spatiotemporal correlations within imaging data to suppress noise and recover clear structures. Quantitative evaluation on synthetic calcium imaging data showed significant improvement in temporal trace correlation, spatial motion correlation, SNR, and segmentation accuracy compared to raw data, even at very low SNRs and with motion artifacts. The x - y and y - t correlation benchmarking quantitatively validated the robustness to motion artifacts, showing both spatial and temporal structure are retained despite distortions. Moreover, DeepBID generalization was evidenced by consistent gains across varying noise, motion, and imaging object. Segmentation and matrix factorization further validated extraction of more accurate structures from the enhanced data.

Notably, by primarily utilizing 64- μ s scan line priors, MP-SSL avoids spatiotemporal confusion that can introduce artifacts when using temporal correlation of 33 ms. This was evidenced by clearer traces and higher spatiotemporal similarity compared to TP-SSL, particularly for rapid biodynamics with large inter-frame drift. Ghosting artifacts and blurring were mitigated in MP-SSL restorations. The strength of spatial-focused learning was further shown in angiography. MP-SSL achieved sharp restoration of morphology and hemodynamics in microvessels, accurately resolving velocity differences between vessel sizes. Fine transient phenomena like trailing artifacts at high flow speeds could also be recovered. Deep-tissue volumetric imaging showed high-fidelity enhancement down to over 550- μ m depth. Multi-target segmentation and tracking was also enabled by suppressing signal fluctuations. Moreover, characteristic structural changes

throughout the rapid cardiac cycle were visualized with high resolution and SNR. Subtle intracellular endocardial-myocardial distance variations, folding motions, valve formations, and sarcomere details were resolved. The generalized applicability across diverse motion-affected imaging contexts highlights the power of data-driven self-supervised learning, without requiring task-specific optimizations.

Additionally, the enhancement efficacy increased for more challenging cases with lower SNR or fewer active neurons. This suggests the network is effectively learning complementary information to the scarce signals directly available in noisy raw data. The data-derived spatial priors act as powerful constraints to reconstruct high-fidelity structures. This principle of exploiting correlations in unaffected dimensions could be extended to temporal, spectral, or radial domains for denoising in other modalities.

For further confirmation, evaluating the method on diverse sample types and microscopy modalities could reveal generalizability limitations. We tested our calcium imaging trained model on different sample densities. The analysis in Note S3 and results in Fig. S14 show that the denoising quality declined slightly for very dense data, which may be improved by incorporating temporal smoothing to assist spatial filtering and applying super-resolution techniques to provide higher sampling density. We also applied our trained model on confocal microscopy [10]. Example images of BPAE cells, mouse brain, and zebrafish were corrupted with synthetic noise to generate low SNR inputs for denoising (Fig. S15). Despite differences in imaging modality, sample type, and noise characteristics from the original training data, our model effectively suppressed noise and enhanced visualization (Note S3). By learning to extract maximal useful information from the mirror-perspective signals, the network can be applied to strengthen visualization for data far beyond its original training distribution.

While showing significant improvements in denoising quality, the model still struggles with extremely noisy data with nonrandom background and non-fluorescent image modalities. Limitations remain to be addressed in future work. Firstly, the method currently relies on bidirectional scanning during image acquisition to provide pairs of dual-perspective information, which requires a high content match between adjacent scan lines. Advanced motion correction techniques could potentially enable learning from non-aligned lines acquired in a large FOV. A blend of real and synthetic training data may balance robustness and accuracy. Secondly, the use of spatial priors makes the model susceptible to missing small-scale signals that are corrupted and lack identifiable relationships in the raw data. While our analysis demonstrated the high spatial resolution increase post denoising (Fig. S16) compare to the input, it may come at the inherent cost of marginally reducing spatial resolution compared to pristine noise-free image when two sampling lines are not identical. Thirdly, the model was predominantly demonstrated on time-lapse imaging data, although the volumetric vessel imaging showed potential 4D application. Evaluating performance on full 4D volumetric data could better characterize enhancements to morphological quantification. The framework may also need adaptations to scale efficiently to 4D datasets. Additionally, manual labeling was used in a limited capacity in this work, primarily for segmenting neurons to quantify motion artifact suppression performance. While manual annotation does introduce some subjectivity and human error, we aimed to minimize the impact on results. Some alternatives include: using computational spot detection and segmentation algorithms

as an initial automated labeling pass, followed by human curation; employing simulated data with programmatically generated annotations for training and evaluation; active learning approaches that select optimized samples to minimize labeling needs.

In the future, combining complementary strengths of model-based and data-driven approaches could improve restoration fidelity, leveraging versatile neural networks along with optical models and signal priors. Joint denoising across multimodal datasets could harness correlated structural information. Extending the self-supervision concept to other imaging domains such as medical, satellite, or computational imaging could further demonstrate broad utility. Overall, this deep learning framework pushes the boundaries of live fluorescence microscopy. Enhanced SNR and correlation unlock richer quantitative insights into microscale biological dynamics. With optimized models and hardware, the approach promises real-time video enhancement during experiments. By overcoming photon limitations, this methodology helps realize full potential of fluorescence microscopy for biological discovery.

Methods

An overall technical roadmap (including software, hardware, experiment, and analysis) was presented in Fig. S17.

Optical setups

The *in vivo* multiphoton upright microscope was equipped with a galvo-resonant scanner for high-speed imaging at 30 Hz with 512×512 pixels. Excitation was provided by a femtosecond laser (Chameleon Discovery, Coherent) with pulsewidth around 100 fs and repetition rate of ~ 80 MHz. Group delay dispersion was pre-compensated to 8000 fs^2 , which ensured low power of $<80 \text{ mW}$ for mouse and $<20 \text{ mW}$ for zebrafish at 890 nm excitation to minimize photochemical and thermal stress as well as image distortion. In contrast, the reported photodamage power was about 120 mW (1,080–1,180 nm, 80 MHz, 100–250 fs, $3.3 \mu\text{s}/\text{px}$)[54–56]. The collimated laser beam was guided to the fast axis (resonant mirror) and slow axis (galvo mirror) of the galvo-resonant scanner. The scanner provided fast two-dimensional raster scanning under the control of two voltage signals. At a resonant frequency of 8 kHz and image pixels of 512×512 , the imaging time per frame was 33 ms for slow TP-SSL, and the scan speed per line was about $64 \mu\text{s}$ for ultrafast MP-SSL. The fast scan mode with brief pixel dwell times helped reduce cumulative energy deposition and associated phototoxicity. However, such rapid resonant scans suffered from poor SNR, which was overcome by self-supervised deep learning without needing prolonged integration times.

The excitation laser beam was then relayed, scaled and corrected by scan lenses (SL50-2P2, Thorlabs) and tube lenses (TTL200MP, Thorlabs) to match the back pupil of the objective and produce a planar image plane. A high numerical aperture (NA) water dipping objective (N20X-PFH, $20\times$, 1.0 NA, Olympus) with 2 mm working distance was used for *in vivo* imaging. The fluorescence photons emitted from the sample were collected by the objective and separated from the excitation by a long-pass dichroic (DMSP680B, Thorlabs). Another short-pass dichroic (DMSP567R, Thorlabs)

was installed in the detection path to split green and red fluorescence. High sensitivity GaAsP photomultiplier tubes (PMTs; PMT2101/M, Thorlabs) with transimpedance amplifier collected the fluorescence signals, providing voltage outputs. Scanner and detector I/O were synchronized via a high-speed DAQ (ATS9440, 125M/s) and a high-performance computing workstation (48 Gb memory, solid state drives).

Animal preparation

Male mice (Balb/c) aged 8-12 weeks were purchased from the Guangdong Medical Laboratory Animal Center. These mice were housed in the animal facilities at the Institute of Optoelectronic Engineering, Shenzhen University. All animal procedures were approved by the Ethics Committee of Experimental Animals, Medical Department, Shenzhen University.

Detailed protocols regarding cranial window procedures have been previously published [57–59]. Briefly, the mice were anesthetized using a gas anesthesia system (R500IP, RWD) with 1.5-2% isoflurane, and a heating blanket was used to maintain a body temperature of 37°C during surgery. After removing the fur and scalp, a small section of skull bone with a diameter of approximately 3 mm was excised using a dental drill. Subsequently, a glass cover glass and a homemade titanium alloy ring were affixed to the cranial window using dental cement. For vascular imaging, we synthesized the reported photosensitizer with aggregation-induced emission (AIE) characteristics, TPETPABT [60], which were prepared according to previous work [61]. The Chemical structure of TPETPABT was presented in Fig. S18a, as well as its ¹H NMR and ¹³C NMR spectra were probed and shown in Fig. S18b. After inflammation subsided and the cranial window cleared, this more established AIE fluorophore (5 mg/kg) were administered to the mice via orbital injection. The vascular two-photon imaging in the mouse brain was performed immediately after the injection.

For astrocyte imaging, we first performed craniotomy surgery on the mice following the above procedures. The acute brain injury caused by the craniotomy induced an immune response in the brain. After surgery, 100 µL of Sulforhodamine 101 (SR101) at a concentration of 3.3 mg/ml [57, 62] was injected to label astrocytes. The labeling efficiency and brightness of astrocytes peaked about 180 minutes after injection. Thus, the astrocytes imaging was performed three hours after the injection.

Zebrafish embryos (CZ62:s843Tg/(AB)) expressing vascular epithelial cells indicator (Tg(kdrl:EGFP)) were purchased from the China Zebrafish Resource Center. The s843Tg allele was generated by random integration of pT2(kdrl:EGFP) construct. The zebrafish embryos were raised in E3 solution containing 0.003% N-phenylthiourea (P7629, Sigma) to inhibit pigmentation after 20 hours post-fertilization. Prior to *in vivo* imaging, zebrafish were anesthetized with 600 µM Tricaine (E10521, Sigma) and mounted in 1% low-melting-point agarose (NuSieve™ GTG™ Agarose, 50080, Cambrex BioScience) for vascular two-photon imaging. As identified by the microscope, the GFP fluorescence signal was detected in cardiovascular system.

Data processing

We used a flexible framework to integrate modules including data reading/writing, model building (training, validation and testing), network architecture (Fig. S1b), loss

functions, etc. This framework allows convenient integration of a customized modules. For example, we wrote a standalone dataset reader to process the collected noise datasets (3D stacks) without requiring target images for network training. Training can choose between MP-SSL and TP-SSL modes. The self-supervised training data preprocessing allows automatic data partitioning. For collinearly bidirectional scanning, where the forward scan path $x = bt$ and the backward scan path $x = 2a - bt$ are conjugated, with a being the sum of the start point and end point of the scan path, and b being the slope. The pixel number collected in the y direction is twice that in the x direction, and the stack dimensions are $N \times 2N \times T$, where N is the number of x -direction pixels, T is the number of frames. The data preprocessing module partitions them into two sets of $N \times N \times T$ for network input and $N \times N \times T$ for target (Fig. S1c). This self-supervised learning utilizes the high semantic information correlation between the conjugated lines, the randomness of noise, and the frequency mismatch of fringe artifacts across lines to achieve denoising and background removal.

For normal unidirectional or bidirectional microscopic scanning modes, we also provide another way to construct the dataset. The stack of dimensions $N \times N \times T$ is partitioned into two sets of $N \times N \times T/2$ for network input) and $N \times N \times T/2$ for target, where the odd rows of each frame and the odd rows of the next frame constitute a new frame of the input set, and the even rows of each frame and the even rows of the next frame constitute a new frame of the target set (Fig. S1d). This ensures pixel spatial uniformity within images and avoids dimension mismatch with internal network operations.

The input and GT images were produced in 8-bit TIFF files with customized macro processing algorithm in Fiji [63] to reduce storage requirements, speed up data read, write and transfer, and accelerate network train and test. To cope with intensity variations across different samples and imaging platforms, the mean of the entire stack is subtracted from each input stack after reading. To alleviate data dependence of the method and further eliminate overfitting, we adopted 12 times data augmentation to generate sufficient training pairs from small amounts of data. The spatial overlap ratio is set to 0.25 for 512×512 pixels. The dimension of each substack is $150 \times 150 \times 100$. For each training pair, one random transform is chosen, including horizontal flip, vertical flip, 90° left rotation, 180° rotation, 90° right rotation, and no transformation. Additionally, the input and target are randomly swapped with a probability of 0.5.

In the absence of experimental noise-free and artifact-free images, temporal averaging was utilized to obtain approximations of the ground truth morphology. For astrocyte data, 300 raw input frames were averaged to estimate the static cellular structure (Fig. 2j). This provided a reference for computing error maps and correlations to quantify denoising performance. For vasculature, high laser power was utilized to acquire high SNR data visualizing clear morphology (Fig. 3a). Mixed Poisson-Gaussian noise was then digitally added to simulate low SNR input data (Fig. 3b, g). The original high SNR images served as pseudo-ground truth references for evaluating SNR and SSIM improvement.

Network architecture, training and inference

The network adopts a 3D U-Net topology [32, 64] with an encoder-decoder architecture. All operations inside the network are performed in 3D, including convolution, max pooling, and interpolation. The network consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. The contracting path contains 4 encoder blocks, each with two $3 \times 3 \times 3$ convolutions followed by a $2 \times 2 \times 2$ max pooling operation with stride 2 for downsampling. At each downsampling step, the number of feature channels is doubled starting from an initial 16 channels. Pooling reduces volume size while expanding feature channels to capture context and abstract representations of the input.

The expanding path consists of 4 decoder blocks, each with an upsampling of the feature map followed by a $2 \times 2 \times 2$ convolution (up-convolution) to reduce number of channels by half. This is followed by concatenation with the correspondingly cropped feature map from the contracting path. Two $3 \times 3 \times 3$ convolutions are then applied to integrate localization information from the contracting path. Cropping and concatenation enables precise localization by integrating high-resolution features from earlier layers. The last decoder output goes through a $1 \times 1 \times 1$ convolution to reduce channels to number of desired output classes. This is passed through a final 3D convolution to generate the predicted output with the same spatial dimensions as the input.

Skip connections between the contracting and expanding paths provide global context as well as localized information to enable precise voxel-level prediction. The overlapping field of views at different depths give the 3D U-Net a large receptive field for incorporating extensive context.

The hyperparameters model architecture and training include: self-supervised mode: MP-SSL or TP-SSL; spatial patch size for reducing memory: 150 pixels; patch overlap factor: 0.2; temporal patch size for reducing memory: 100 frames; batch size per GPU: 1; number worker per GPU: 2; optimizer: Adam; learning rate: 0.00005; loss function: L1 + MSE; training iterations: 100,000; no warm up; validation frequency: 10,000; save image during validation: true; data augmentation: flips, rotations, crops.

The abovementioned data augmentation strategies are applied to each training pair. The weight of L1 loss is $\lambda_1 = 0.5$ and the weight of MSE loss is $\lambda_2 = 0.5$ for the loss function in training. We used adaptive moment estimation (Adam) [65] as the optimizer of the generator, $\beta_1 = 0.5$, $\beta_2 = 0.999$. More details could be found in the public codes. This flexible framework allows convenient switching of networks. In the public codes, we also provided alternative networks such as 3D RCAN.

Synthesis of neuronal imaging data

We quantitatively evaluated the MP-SSL method on synthetic neuronal imaging data and for comparisons with TP-SSL. The simulated processes involve generation of neural volume and activity, modeling of light propagation through scattering volume, and microscopic scanning and image formation [33]. The vasculature was first generated throughout the volume, followed by somata, dendritic and axon. The spiking calcium dynamics of each neuron were simulated and converted to fluorescence. Then,

the optical wavefront corresponding to the TPLSM optics propagated through the scattering volume, producing a spatially variant point spread function to create relative intensity masks. Realistic images are rendered by scanning the composite volume using the optical model output, incorporating noise and motion arising from light collection, amplification and digitization processes. The simulation parameters are listed in Table S2, with defaults used for those unspecified. The simulated data exhibits spatiotemporal realism highly similar to experimentally obtained data and were used for network performance and generalization verification.

Neuronal and vascular segmentation

We segmented neurons manually with a limited view that ensuring one neuron inside the cropped view. The calcium trace of the segmented neurons was extracted using peak matching to avoid motion drifts. We also employed CNMF [39] algorithm with motion correction [40]. The processing pipeline included motion correction, source extraction and deconvolution. To ensure that the algorithm extracted the same neurons from the raw input, denoised output, and pristine reference images for comparison, we concatenated three types of images, i.e. reference-noise-restoration in the temporal dimension: $N \times N \times 3T$. The parameters for calcium imaging data analysis are listed in Table S3, with defaults used for those unspecified. After processing, the extracted calcium traces were divided into three corresponding segments. Nevertheless, this method would extract more ambiguous neurons due to fluorescent instability caused by dramatic motional drifts.

To segment the time-lapse vascular stacks, we used SAM-Track, which combines segment-anything model (SAM) [48] for automatic key-frame segmentation, and decoupling features in associating objects with transformers (DeAOT) [49] for efficient multi-object tracking and propagation. The pretrained model for global segmentation was “r50_deatl”, with a SAM gap of 4, 16 points per slide and a max objective number of 255.

4D volumetric visualization

For 4D visualization to reveal spatiotemporal dynamics of the astrocytic and vascular volumes, we implemented custom Matlab scripts and built-in functions to generate time-lapse volumetric images. The brightness of the images before and after denoising is adjusted to have similar visual effects [32]. These volumetric images were shown in the movies. 2D visualization of the volumetric data were obtained with “3D Project” (Brightest Point) in Fiji. Orthogonal views were also obtained using Fiji. Images with a relatively low brightness were regulated by adjusting the dynamic ranges (brightness/contrast) in Fiji to better display the indiscernible morphological features [15].

Performance metrics

The quality metrics, including 2D and 3D correlation and SSIM, 3D SNR were calculated between the signal (input or output) intensity, I_{sig} and the reference intensity, I_{ref} . Pearson correlation coefficient ρ is formulated as

$$\rho = \frac{\sum_{\text{dims}=1,2,3} (I_{sig} - \bar{I}_{sig})(I_{ref} - \bar{I}_{ref})}{(N_{total} - 1)\sigma_{sig}\sigma_{ref}} \quad (1)$$

where \bar{I}_{sig} are σ_{sig} are the mean and SD of I_{sig} , respectively. \bar{I}_{ref} are σ_{ref} are the mean and SD of I_{ref} , respectively. Dimensional subscript $dims = 1,2,3$ correspond to trace ($\sum_t I$), frame ($\sum_{x,y} I$), and stack ($\sum_{x,y,t} I$). N_{total} is the total pixel number.

SNR is obtained by computing the ratio of summed squared magnitude of I_{sig} to that of the noise $I_{noise} = I_{sig} - I_{ref}$:

$$SNR = 20 \log_{10} \frac{RSS(I_{sig})}{RSS(I_{noise})} \quad (2)$$

where RSS is the root-sum-of-squares:

$$RSS = \sqrt{\sum_{x,y,t} |I|^2} \quad (3)$$

SSIM is based on the computation of luminance, contrast, and structure. The overall index is calculated as

$$SSIM = \frac{(2\bar{I}_{sig}\bar{I}_{ref} + C_1)(2\sigma_{sig,ref} + C_2)}{(\bar{I}_{sig}^2 + \bar{I}_{ref}^2 + C_1)(\sigma_{sig}^2 + \sigma_{ref}^2 + C_2)} \quad (4)$$

where $\sigma_{sig,ref}$ is the cross-covariance for I_{sig} and I_{ref} . By default, $C_1 = (0.01 \times L)^2$ and $C_2 = (0.03 \times L)^2$, where L is the specified dynamic range value. For example, the default dynamic range is 255 for images of data type uint8. The function uses these regularization constants to avoid instabilities in image regions where the mean or SD are close to zero.

The quantitative assessment scores, RSE and RSP, as well as the error maps were obtained by estimating the resolution scaling function, registering the restored image against the reference image, and rescaling the restored image intensity with the resolution scaling function estimation. RSE and RSP can be calculated through a root-mean-square error between I_{RS} (created by applying the RSF to the restored image) and the reference image I_{ref} :

$$RSE = \sqrt{\frac{\sum_{x,y} [I_{ref}(x,y) - I_{RS}(x,y)]^2}{N_{total}}} \quad (5)$$

$$RSP = \frac{\sum_{x,y} [I_{ref}(x,y) - \bar{I}_{ref}] [I_{RS}(x,y) - \bar{I}_{RS}]}{\sqrt{\sum_{x,y} [I_{ref}(x,y) - \bar{I}_{ref}]^2} \sqrt{\sum_{x,y} [I_{RS}(x,y) - \bar{I}_{RS}]^2}} \quad (6)$$

The error map is the pixel-wise absolute difference between the restored and reference image. We used NanoJ-Squirrel Plugin [41, 42] in Fiji to compute these metrics and visualize the discrepancy of the input and output images compared to the average images.

Statistics and reproducibility

Sample sizes and statistical analyses including the mean, SD, and significant difference were specified in figure legends and text for each experiment. Tukey box and whisker plots showed the statistical correlations and SNR, where box indicated the upper and lower quartiles, while the line inside the box represented the median. The lower whisker extended to the first data point greater than the lower quartile minus 1.5 times the interquartile range. Similarly, the upper whisker extended to the last data point less than the upper quartile plus 1.5 times the interquartile range. Outliers were marked with small dots. Three black lines in the violin plot indicate quartile positions, where solid line represents median. The statistical differences, p values were located above the data. Representative frames were shown in the figures, with similar conclusions for other frames.

Abbreviations

TPLSM	Two-photon laser-scanning microscopy
FOV	Field of view
SNR	Signal-to-noise ratio
DeepBID	Deep learning for biodynamics imaging denoising and deblurring
MP-SSL	Mirrored perspective self-supervised learning
TP-SSL	Time-lapse perspective self-supervised learning
PSF	Point spread function
SSIM	Structural similarity index measure
CNMF	Constrained non-negative matrix factorization
RSE	Resolution scaled error
RSP	Resolution scaled Pearson coefficient
RTLS	Residence time line scanning
SFA	Scanning fringe artifacts
NA	Numerical aperture
PMT	Photomultiplier tube
AIE	Aggregation-induced emission
Adam	Adaptive moment estimation
SAM	Segment-anything model
DeAOT	Decoupling features in associating objects with transformers

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43074-023-00117-0>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Additional file 8.

Additional file 9.

Acknowledgements

Not applicable.

Authors' contributions

J.Q. and L.L. supervised this research and designed detailed implementations. B.S. constructed and tested the deep network, developed the Python, Matlab, Fiji analysis codes, performed simulations, analyzed the imaging data. B.S. and C.L. built the two-photon microscope, performed imaging experiments, and processed the data. C.L., W.P., and Y.J. provided animal models, performed craniotomy on mice, and prepared samples. W.P., Y.J., and W.W. synthesized the dyes and measured the material properties. R.H., B.G., J.Q. and L.L. supervised the data analysis and edited the manuscript. All authors contributed to writing the manuscript.

Funding

We thank the National Natural Science Foundation of China (62225505/61935012/ 62175163/61835009/62127819/62205220), Shenzhen Key Projects (JCYJ20200109105404067), Shenzhen International Cooperation Project (GJHZ20190822095420249), and Shenzhen Medical Research Project (A2303018) for financial support.

Availability of data and materials

The main data supporting the findings of this study are available within the paper and its [Supplementary Information](#). The synthetic Ca data, including training dataset with bidirectional collinear scan ($N_y = 2N_x$) for MP-SSL, training dataset with normal scan ($N_y = N_x$) for TP-SSL, and testing data ($N_y = N_x$) are publicly available at <https://zenodo.org/record/8248112>. Experimental data, including astrocyte data, rapid hemodynamic data (larger vessels), vascular data (smaller vessels), zebrafish cardiac data are publicly available at <https://zenodo.org/record/8312530>. The original codes for TP-SSL and MP-SSL network frameworks are readily accessible at <https://github.com/shenblin/DeepBID> with the modifications and customized parameters described in Methods. The model includes python codes for training and testing for the network and for automatically splitting input stacks and restitching output tiles which have been integrated in the model. Our data and codes will be made publicly available before publication.

Declarations

Ethics approval and consent to participate

All animal procedures were approved by the Ethics Committee of Experimental Animals, Medical Department, Shenzhen University.

Consent for publication

All authors agreed to publish this paper.

Competing interests

The authors declare that they have no conflicts of interest.

Received: 15 September 2023 Revised: 21 November 2023 Accepted: 18 December 2023

Published online: 05 January 2024

References

1. Drew PJ, Shih AY, Kleinfeld D. Fluctuating and sensory-induced vasodynamics in rodent cortex extend arteriole capacity. *P Natl Acad Sci USA*. 2011;108:8473–8.
2. Wang Z, et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nat Methods*. 2021;18:551–6.
3. Andrade-Talavera, Y, Fisahn, A. & Rodríguez-Moreno, A. Timing to be precise? An overview of spike timing-dependent plasticity, brain rhythmicity, and glial cells interplay within neuronal circuits. *Mol Psychiatry*. 2023;28:2177–88.
4. Fan JL, et al. High-speed volumetric two-photon fluorescence imaging of neurovascular dynamics. *Nat Commun*. 2020;11:6020.
5. Shen Z, Lu Z, Chhatbar PY, O'Herron P, Kara P. An artery-specific fluorescent dye for studying neurovascular coupling. *Nat Methods*. 2012;9:273–6.
6. O'Herron P, et al. Neural correlates of single-vessel haemodynamic responses in vivo. *Nature*. 2016;534:378–82.
7. Fan J, et al. Video-rate imaging of biological dynamics at centimetre scale and micrometre resolution. *Nat Photonics*. 2019;13:809–16.
8. Mickoleit M, et al. High-resolution reconstruction of the beating zebrafish heart. *Nat Methods*. 2014;11:919–22.
9. Huang XS, et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nat Biotechnol*. 2018;36:451–9.
10. Zhang, Y.D. et al. A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images. 2019 *Ieee/Cvfr Conference on Computer Vision and Pattern Recognition (Cvpr 2019)*, 2019: 11702–11710.
11. Lee S, Negishi M, Urakubo H, Kasai H, Ishii S. Mu-net: Multi-scale U-net for two-photon microscopy image denoising and restoration. *Neural Netw*. 2020;125:92–103.
12. Audier, X., Heuke, S., Volz, P., Rimke, I. & Rigneault, H. Noise in stimulated Raman scattering measurement: From basics to practice. *Apl Photonics*. 2020;5:011101.
13. Weigert M, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods*. 2018;15:1090–7.
14. Lin HN, et al. Microsecond fingerprint stimulated Raman spectroscopic imaging by ultrafast tuning and spatial-spectral learning. *Nat Commun*. 2021;12:3052.
15. Shen B, et al. Deep learning autofluorescence-harmonic microscopy. *Light*. 2022;11:76.
16. Liu S-B, et al. Deep learning enables parallel camera with enhanced-resolution and computational zoom imaging. *Photonix*. 2023;4:17.
17. Wang K, et al. Deep learning wavefront sensing and aberration correction in atmospheric turbulence. *Photonix*. 2021;2:8.
18. Li, Y.P. et al. Fast denoising and lossless spectrum extraction in stimulated Raman scattering microscopy. *J Biophotonics*. 2021;14:e202100080.
19. Jiang M, et al. Fiber laser development enabled by machine learning: review and prospect. *Photonix*. 2022;3:16.
20. Zhao Z, et al. Two-photon synthetic aperture microscopy for minimally invasive fast 3D imaging of native subcellular behaviors in deep tissue. *Cell*. 2023;186:2475–2491.e2422.

21. Sarder P, Nehorai A. Deconvolution methods for 3-D fluorescence microscopy images. *Ieee Signal Process Mag.* 2006;23:32–45.
22. Figueiredo MAT, Bioucas-Dias JM. Restoration of Poissonian images using alternating direction optimization. *Ieee Trans Image Process.* 2010;19:3133–45.
23. Chen JJ, et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat Methods.* 2021;18:678–87.
24. Qiao C, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat Methods.* 2021;18:194–202.
25. Chaudhary, S., Moon, S. & Lu, H. Fast, efficient, and accurate neuro-imaging denoising via supervised deep learning. *Nat Commun.* 2022;13:5165.
26. Zhang Y, et al. Multi-focus light-field microscopy for high-speed large-volume imaging. *Photonix.* 2022;3:30.
27. Junchao F, Xiaoshuai H, Liuju L, Shan T, Liangyi C. A protocol for structured illumination microscopy with minimal reconstruction artifacts. *Biophys Rep.* 2019;5:80–90.
28. Xue F, et al. Hessian single-molecule localization microscopy using sCMOS camera. *Biophys Rep.* 2018;4:215–21.
29. Lehtinen, J. et al. Noise2Noise: Learning Image Restoration without Clean Data. *Int Conference Machine Learn.* 2018;80:2965–74.
30. Lecoq J, et al. Removing independent noise in systems neuroscience data using DeepInterpolation. *Nat Methods.* 2021;18:1401–8.
31. Li XY, et al. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. *Nat Methods.* 2021;18:1395–400.
32. Li X, et al. Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nat Biotechnol.* 2023;41:282–92.
33. Song A, Gauthier JL, Pillow JW, Tank DW, Charles AS. Neural anatomy and optical microscopy (NAOMi) simulation for evaluating calcium imaging methods. *J Neurosci Methods.* 2021;358:109173.
34. Zhang Y, et al. Rapid detection of neurons in widefield calcium imaging datasets after training with synthetic data. *Nat Methods.* 2023;20:747–54.
35. Meiniel W, Olivo-Marin JC, Angelini ED. Denoising of Microscopy Images: A Review of the State-of-the-Art, and a New Sparsity-Based Method. *Ieee Trans Image Process.* 2018;27:3842–56.
36. Lorenz KS, Salama P, Dunn KW, Delp EJ. Digital correction of motion artefacts in microscopy image sequences collected from living animals using rigid and nonrigid registration. *J Microsc.* 2012;245:148–60.
37. Lee S, Vinegoni C, Sebas M, Weissleder R. Automated motion artifact removal for intravital microscopy, without a priori information. *Sci Rep-Uk.* 2014;4:4507.
38. Vinegoni C, Lee S, Aguirre AD, Weissleder R. New techniques for motion-artifact-free in vivo cardiac microscopy. *Front Physiol.* 2015;6:147.
39. Pnevmatikakis, Eftychios A, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron.* 2016;89:285–99.
40. Pnevmatikakis EA, Giovannucci A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *J Neurosci Methods.* 2017;291:83–94.
41. Culley S, et al. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat Methods.* 2018;15:263–6.
42. Wang H, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat Methods.* 2019;16:103–10.
43. Secomb TW. Hemodynamics. *Compr Physiol.* 2016;6:975–1003.
44. Peyrounette M, Davit Y, Quintard M, Lorthois S. Multiscale modelling of blood flow in cerebral microcirculation: Details at capillary scale control accuracy at the level of the cortex. *Plos One.* 2018;13:e0189474.
45. Munoz CJ, Lucas A, Williams AT, Cabrales P. A review on microvascular hemodynamics. *Crit Care Clin.* 2020;36:293–305.
46. Kamoun WS, et al. Simultaneous measurement of RBC velocity, flux, hematocrit and shear rate in vascular networks. *Nat Methods.* 2010;7:655–60.
47. Mannam V, et al. Real-time image denoising of mixed Poisson-Gaussian noise in fluorescence microscopy images using ImageJ. *Optica.* 2022;9:335–45.
48. Kirillov, A. et al. Segment anything. *arXiv:2304.02643.* 2023. <https://doi.org/10.48550/arXiv.2304.02643>.
49. Cheng, Y. et al. Segment and track anything. *arXiv:2305.06558.* 2023. <https://doi.org/10.48550/arXiv.2305.06558>.
50. Zhang X, et al. High-resolution mapping of brain vasculature and its impairment in the hippocampus of Alzheimer's disease mice. *Natl Sci Rev.* 2019;6:1223–38.
51. Hove JR, et al. Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis. *Nature.* 2003;421:172–7.
52. Hoage T, Ding Y, Xu X. Quantifying cardiac functions in embryonic and adult zebrafish. *Methods Mol Biol (Clifton, NJ).* 2012;843:11–20.
53. De Luca E, et al. ZebraBeat: a flexible platform for the analysis of the cardiac rate in zebrafish embryos. *Sci Rep-Uk.* 2014;4:4898.
54. Debarre D, Olivier N, Supatto W, Beaurepaire E. Mitigating Phototoxicity during Multiphoton Microscopy of Live *Drosophila* Embryos in the 1.0–1.2 μm Wavelength Range. *Plos One.* 2014;9:e104250.
55. You S, et al. Intravital imaging by simultaneous label-free autofluorescence-multiharmonic microscopy. *Nat Commun.* 2018;9:21–5.
56. Lefort C. A review of biomedical multiphoton microscopy and its laser sources. *J Phys D Appl Phys.* 2017;50:423001.
57. Liu H, et al. Visualizing astrocytes in the deep mouse brain in vivo. *J Biophotonics.* 2019;12:e201800420.
58. Samanta S, et al. AIE-active two-photon fluorescent nanoprobe with NIR-II light excitability for highly efficient deep brain vasculature imaging. *Theranostics.* 2021;11:2137–48.
59. Poon C, Teikari P, Rachmadi MF, Skibbe H, Hynynen K. A dataset of rodent cerebrovasculature from in vivo multiphoton fluorescence microscopy imaging. *Sci Data.* 2023;10:141.

60. Qin W, et al. Bright and Photostable Organic Fluorescent Dots with Aggregation-Induced Emission Characteristics for Noninvasive Long-Term Cell Imaging. *Adv Funct Mater.* 2014;24:635–43.
61. Wu W, et al. Polymerization-Enhanced Photosensitization. *Chem.* 2018;4:1937–51.
62. Florence CM, Baillie LD, Mulligan SJ. Dynamic volume changes in astrocytes are an intrinsic phenomenon mediated by bicarbonate ion flux. *Plos One.* 2012;7:e51124.
63. Schindelin J, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9:676–82.
64. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. (eds. S. Ourselin, L. Joskowicz, M.R. Sabuncu, G. Unal & W. Wells) 424-432 (Springer International Publishing, Cham; 2016).
65. Cortinas-Lorenzo B, Perez-Gonzalez F. Adam and the Ants: On the Influence of the Optimization Algorithm on the Detectability of DNN Watermarks. *Entropy.* 2020;22:1379.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
