Self-supervised denoising for multimodal structured illumination microscopy enables long-term super-resolution live-cell imaging

Xingye Chen[1,2,3] , Chang Qiao[1,3]*, Tao Jiang[4,5], Jiahao Liu[4,6], Quan Meng[4,5], Yunmin Zeng[1], Haoyu Chen[4,5], Hui Qiao[1,3], Dong Li[4,5]* and Jiamin Wu[1,3]*

Xingye Chen, Chang Qiao, Tao Jiang and Jiahao Liu contributed equally to these works

*Correspondence:
qc17@tsinghua.org.
cn; lidong@ibp.ac.cn;
wujiamin@tsinghua.edu.cn

[1] Department of Automation, Tsinghua University, Beijing 100084, China
[2] Research Institute for Frontier Science, Beihang University, Beijing 100083, China
[3] Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing 100084, China
[4] National Laboratory of Biomacromolecules, New Cornerstone Science Laboratory, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[5] College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[6] Laboratory of Image Processing and Intelligent Control of Ministry of Education of China, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract

Detection noise significantly degrades the quality of structured illumination microscopy (SIM) images, especially under low-light conditions. Although supervised learning based denoising methods have shown prominent advances in eliminating the noise-induced artifacts, the requirement of a large amount of high-quality training data severely limits their applications. Here we developed a pixel-realignment-based self-supervised denoising framework for SIM (PRS-SIM) that trains an SIM image denoiser with only noisy data and substantially removes the reconstruction artifacts. We demonstrated that PRS-SIM generates artifact-free images with 20-fold less fluorescence than ordinary imaging conditions while achieving comparable super-resolution capability to the ground truth (GT). Moreover, we developed an easy-to-use plugin that enables both training and implementation of PRS-SIM for multimodal SIM platforms including 2D/3D and linear/nonlinear SIM. With PRS-SIM, we achieved long-term super-resolution live-cell imaging of various vulnerable bioprocesses, revealing the clustered distribution of Clathrin-coated pits and detailed interaction dynamics of multiple organelles and the cytoskeleton.

Keywords: Super-resolution microscopy, Structured illumination microscopy, Deep learning, Self-supervised learning, Live-cell imaging

Introduction

Studying biological dynamics and functions in live cells requires imaging with high spatiotemporal resolution and low optical invasiveness. Structured illumination microscopy (SIM) is commonly recognized as a well suitable tool for live imaging because of its ability to acquire a super-resolution (SR) image from only a small number of illumination pattern-modulated images [1, 2]. However, conventional SIM reconstruction algorithm is prone to generate photon noise-induced artifacts especially under low light conditions, which substantially degrades the image quality and overwhelms useful structural information, thereby inhibiting us from fully exploring the underlying biological processes [3, 4]. To alleviate the reconstruction noise, a long camera exposure time and high excitation power are usually applied in SIM imaging experiments, which reduce the

image acquisition speed and introduce considerable photobleaching and phototoxicity. This tradeoff severely limits the application of SIM in live-cell imaging.

Alongside the development of SIM instruments [5–7], many techniques and algorithms aiming to reconstruct high-quality SR-SIM images with low signal-to-noise ratio (SNR) inputs have been proposed. Some algorithms have been developed to analytically improve the estimation precision of the illumination pattern [8, 9] or iteratively denoise the reconstructed SR images under certain optical models and assumptions [10–12]. However, since the imaging process is complex and the image restoration/denoising problem is theoretically ill-posed, these analytical algorithms cannot fully address the statistical complexity and have limited noise suppression capability [13]. Recently, deep neural networks (DNNs) have shown outstanding performance in various optical imaging tasks [14–16], especially in microscopic images restoration [17–19]. Various deep-learning-based SIM algorithms have demonstrated great potential in reconstructing high-quality SR images, even under extreme imaging conditions. Nevertheless, existing methods still face several challenges. First, some existing techniques employ "end-to-end" schemes [17, 20–22], which directly transform wide-filed or raw SIM images into the SR-SIM image without fully exploiting the high-frequency information modulated by the illumination pattern, i.e., the Moore fringes. As a result, the entire framework degrades to an SR inference task (termed "image super-resolution" [23, 24]) instead of analytical SR reconstruction [25], which may suffer from the spectral bias issue [26, 27] and result in a compromised the resolution [25–27]. Second, a large number of well-matched low- and high-SNR image pairs are necessary to construct the training dataset [28, 29], which is laborious and even infeasible for biological specimens of low fluorescent efficiency or high dynamic. Third, the generalizability of the neural network is limited because in the supervised training scheme, a pre-trained denoising model cannot be reliably transferred to unseen domain with only noisy data, which inhibits the discovery of unprecedented biological structures and bioprocesses.

Here we proposed a pixel-realignment-based self-supervised method for structured illumination microscopy (PRS-SIM), which employs a deep neural network to achieve artifact-free reconstruction with ∼ 20 fold fewer collected photons than that used for conventional SIM algorithms [7]. The proposed PRS-SIM framework has several key advantages: first, because the analytical SIM reconstruction principle is embedded in the training and inference framework, the resolution enhancement is physically guaranteed by the SIM configuration rather than computationally achieved via data-driven supervised learning [19, 23, 30, 31]. Second, the PRS-SIM models are trained on low-SNR raw images only, without the requirement for either high-SNR ground-truth data or repeated acquisition of the same sample, resulting in a more feasible data acquisition process. Third, for time-lapse imaging, PRS-SIM can be implemented in an adaptive training mode, in which the collected low-SNR data are used to train a new customized model or fine-tune a pretrained model. Finally, PRS-SIM is compatible with multimodal SIM configurations, including total internal reflective fluorescence SIM (TIRF-SIM) [5], grazing incidence SIM (GI-SIM) [7], three dimensional SIM (3D-SIM) [2], lattice light-sheet SIM (LLS-SIM) [32], and non-linear SIM (NL-SIM) [33, 34]. Benefiting from these advances, PRS-SIM instantly enables long-term volumetric SR imaging of live cells with extremely low photo-damage to the biological samples.

Results and discussion

The principle and evaluation of PRS-SIM

The principle of PRS-SIM is schematized in Fig. 1. The PRS-SIM framework involves self-supervised neural network training (Fig. 1a) and the corresponding inference phase (Fig. 1b). Specifically, the training dataset is constructed via a novel pixel-realignment strategy, whose underlying mechanism is to utilize the spatial redundancy and statistical independence between adjacent pixels in noisy raw images [35, 36]. For each noisy raw SIM image stack, we firstly applied pixel-realignment strategy, which includes three operations of pixel extraction, up-sampling and sub-pixel registration (Materials and methods), to generate four raw image stacks of the same scene. Then by applying conventional SIM algorithm, four well-aligned raw SR images are reconstructed, which are subsequently arranged as the input and target reciprocally for network training. Notably, although each noisy SR images are generated from the same raw images, we theoretical proved the effectiveness of adopting these SR images into the loss calculation to train a SIM denoiser (Supplementary Note 1). By iteratively optimizing the L2-norm loss
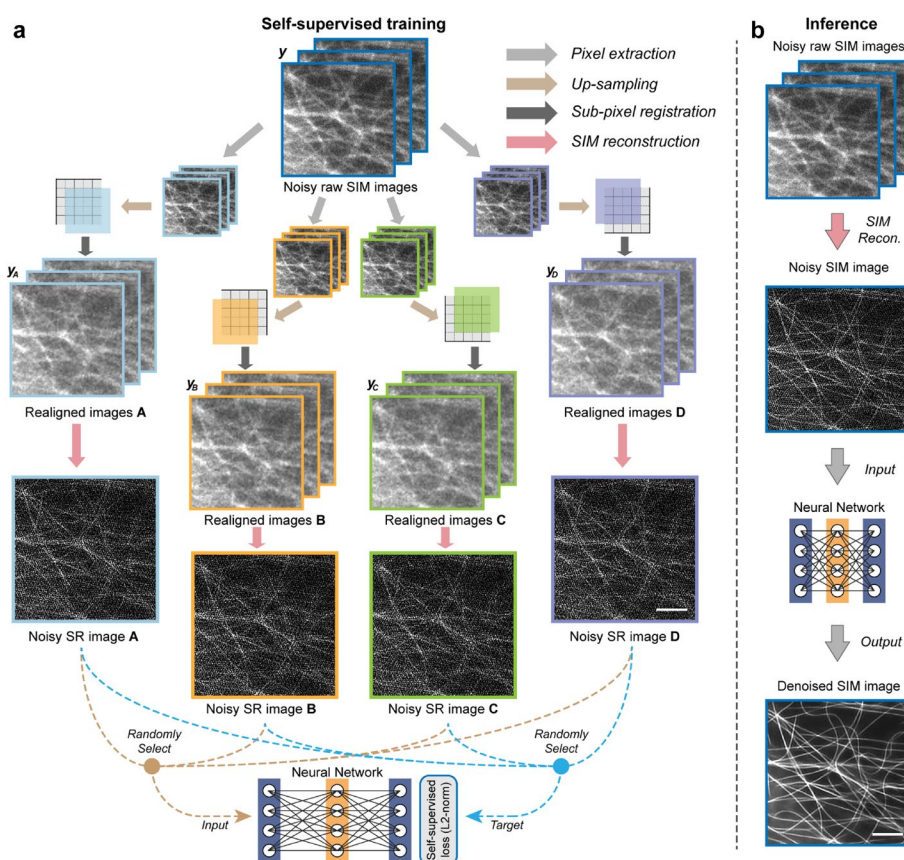

Fig. 1 Schematic of PRS-SIM. **a** Self-supervised training strategy of PRS-SIM. Four matched image groups y_A, y_B, y_C, and y_D are generated by applying pixel-realignment operation to a noisy low-resolution (LR) raw SIM image group y. Then with conventional SIM algorithm, four super-resolution (SR) images are reconstructed, which are further randomly arranged as the input and target for neural network training. **b** Inference pipeline of PRS-SIM. The noisy raw SIM image group are firstly reconstructed into a noisy SR image by conventional SIM algorithm. Then by inputting this noisy SR image into the pre-trained PRS-SIM model, the corresponding noise-free SR SIM image will be generated. Scale bar, 2 μm

function, the neural network will acquire the denoising ability that to transform noisy SIM images into their corresponding clean counterparts. In the inference phase, the raw images are firstly reconstructed into the noisy SR images via the conventional SIM algorithm, then the well-trained PRS-SIM model takes these noisy SIM images as inputs and outputs the final noise-free SR images.

We first systematically evaluated PRS-SIM on the publicly available biological image dataset BioSR [17, 37]. To quantify the performance of PRS-SIM, we calculated the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) referring to ground-truth (GT) SIM images as the criteria (Materials and methods). Three individual neural networks were trained separately for clathrin-coated pits (CCPs), endoplasmic reticulum (ER), and microtubules (MTs), as representative examples of hollow, reticular, and filament structures, respectively. The training dataset was augmented from the low SNR raw data in BioSR whose signal level ranging from 1 to 4 (MTs and CCPs) or 1 to 3 (ER). The average effective photon counts of these samples are 10 to 30-fold less than those used in artifact-free GT-SIM images. We compared PRS-SIM with conventional SIM (conv. SIM) and sparse-deconvolution SIM (Sparse-SIM) (Fig. 2a) and found that the detailed information can hardly be distinguished in conv. SIM and Sparse-SIM due to severe reconstruction artifacts. In contrast, PRS-SIM can clearly super-resolve ring-shaped CCPs and densely interlaced MTs, resulting in an image quality comparable to GT-SIM in both space and frequency domain (Fig. 2b). The statistical results in terms of the PSNR and SSIM of 40 individual cells for each sample showed that PRS-SIM substantially boosted the image quality for various types of specimens (Fig. 2c). The intensity profiles and Fourier ring correlation (FRC) analysis [38] (Fig. 2d, e) indicated that PRS-SIM achieves comparable spatial resolution as GT-SIM images, and successfully distinguishes several adjacent microtubules, which are indistinguishable by the other methods.

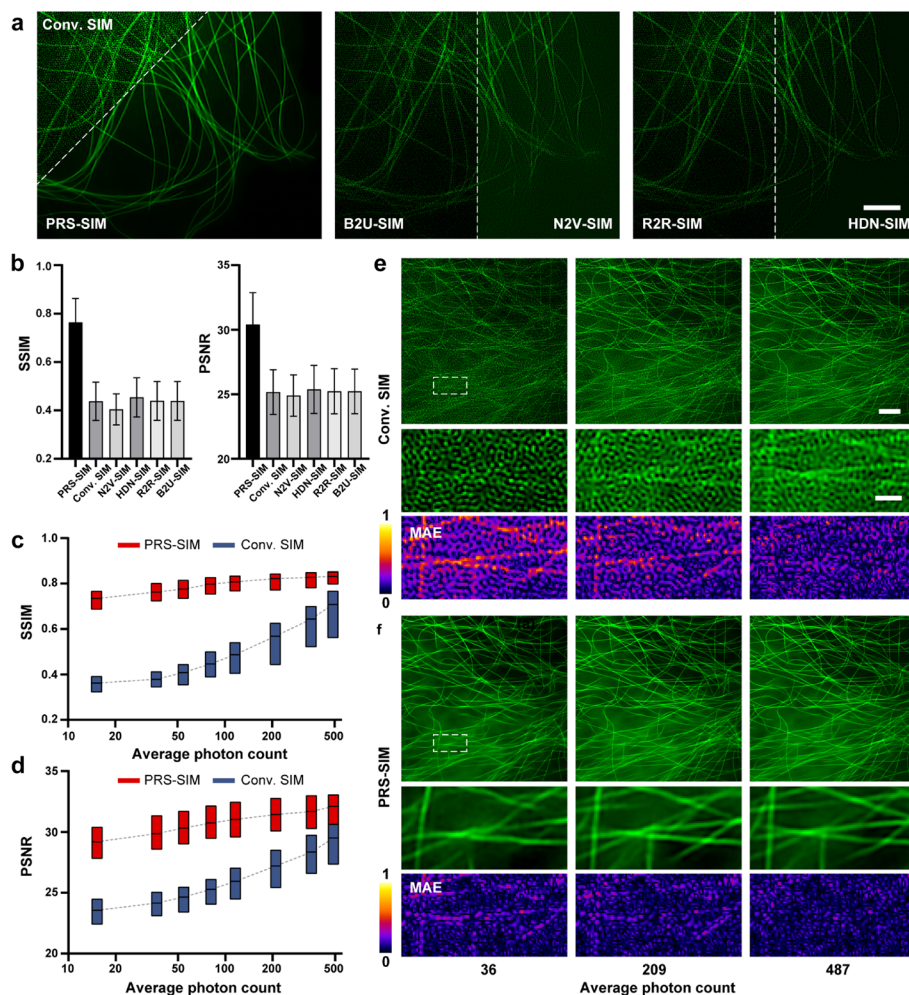**Comparison of PRS-SIM with other existing methods**

In addition to PRS-SIM, many other self-supervised denoising methods for fluorescence microscopy have been developed in recent years, such as noise2void (N2V) [39], hierarchical diverse denoising (HDN) [40], recorrupted-to-recorrupted (R2R) [41], and Blind-2unblind (B2U) [42]. Each of them took a specific characteristic or assumption of the noise to establish a self-supervised mechanism. Although these methods have shown great denoising performance for natural and microscopic images, they are not applicable to SIM images for two critical reasons. First, if the denoising algorithms are applied to raw SIM images (Supplementary Fig. 1a), i.e., images that are captured directly by the sensor, the algorithms have difficulty recognizing the illumination patterns and restoring the subtle Moiré fringes, thereby missing high-frequency information and generating reconstructed images with riddling artifacts (Supplementary Fig. 2). Although the recently proposed rDL-SIM [25] is capable to denoised the raw SIM images as well as maintaining the information of the structured illumination, it is implemented in a supervised manner and necessitates noise-free ground-truth data. Second, if these algorithms are employed in the post-reconstruction procedure (Supplementary Fig. 1b), taking N2V [39] or its 3D-form derivative DeepSeMi [43] for instance, the strongly spatially correlated noise patterns in the reconstructed SIM images are inconsistent with its blind-spot
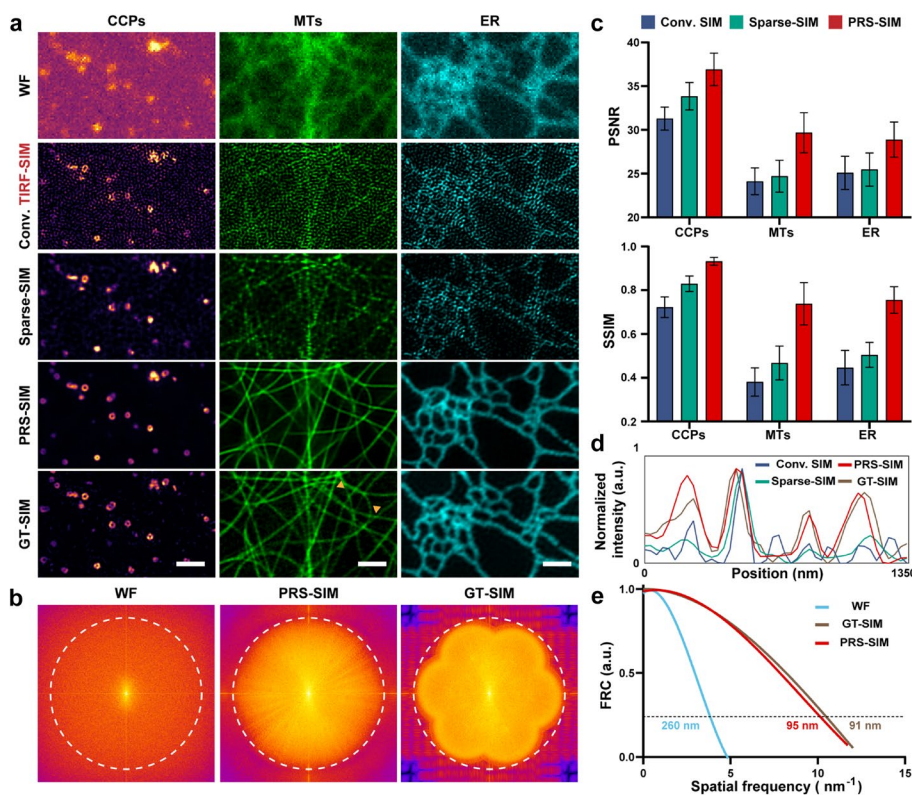
Fig. 2 Fidelity and resolution evaluation of PRS-SIM. **a** TIRF-SIM images of clathrin coated pits (CCPs), microtubules (MTs), and endoplasmic reticulum (ER) reconstructed and processed with Conv. SIM, Sparse-SIM, and PRS-SIM. Corresponding WF and GT-SIM images are provided for reference. Scale bar, 1 μm. **b** Fourier spectra of WF, PRS-SIM and, GT-SIM images of a MT sample. The dashed circle denotes the cutoff frequency (corresponding to the spatial frequency of 94 nm). **c** Quantitative comparison among PRS-SIM, Conv. SIM and Sparse-SIM. The PSNR and SSIM values are calculated referring to GT-SIM images ($N = 40$ for each data point). **d** Intensity profiles of Conv. SIM (blue), Sparse-SIM (green), PRS-SIM (red), and GT-SIM (brown) along the line indicated by the yellow arrowheads in **a**. **e** Fourier ring correlation curves of the WF, PRS-SIM and GT-SIM images in **b**. The resolution is calculated according to the cutoff frequency with an FRC threshold of 0.24

principle, leading to poor denoising performance (Fig. 3a-b). Benefitting from the intrinsic linearity of the SIM algorithm, our proposed pixel realignment strategy offers a solution to generate SR image pairs, which contains SR information meanwhile meets the noise distribution requirement for self-supervised training (Supplementary Note 1). We experimentally compared PRS-SIM with the aforementioned methods (Fig. 3a-b, Supplementary Fig. 2). Both the perceptual comparisons and the quantitative analysis showed that PRS-SIM can generate SR images with considerably fewer artifacts, outperforming other self-supervised denoising methods by a large margin. Moreover, the framework of PRS-SIM is different from previously proposed self-supervised denosing method Neighbor2Neighor, and the latter one cannot achieve satisfactory performance for SIM images (Supplementary Fig. 3). PRS-SIM also demonstrated massively improved performance compared to existing advanced reconstruction algorithms, such as sparse-SIM [44], HiFi-SIM [11], JSFR-SIM [45] and direct-SIM [46] (Supplementary Fig. 4).

Next, we validated the robustness of PRS-SIM on both synthetic (Supplementary Note 2) and experimental data with different signal levels. We trained a PRS-SIM

Fig. 3 Comparison of PRS-SIM with state-of-the-art self-supervised denoising methods on input images of different signal levels. **a** SR-SIM images denoised by PRS-SIM, B2U-SIM, N2V-SIM, R2R-SIM and HDN-SIM from the same noisy input. Scale bar, 2 μm. **b** Quantitative comparison of the performance among the aforementioned methods. **c, d** Quantitative evaluation of PRS-SIM over different signal levels (indicated by the average photon counts of raw images). **e, f** Representative Conv. SIM images and PRS-SIM denoised images of different signal levels. The mean-absolute-error (MAE) maps of the zoom-in region are provided for an intuitive visualization. Scale bar, 2 μm (regular), 0.5 μm (zoom-in). $N = 40$ for each data point in (**b-d**)

model with a mixed training dataset containing images of various SNR, and then applied the trained model to process noisy SIM images of different signal levels. We demonstrated that a well-trained PRS-SIM model is applicable with a wide range of input SNRs, and significantly outperforms the conventional SIM reconstruction algorithm in all signal level conditions (Fig. 3c-f, Supplementary Fig. 5). Furthermore, we compared PRS-SIM with the classical noise2noise (N2N) method [47], which requires two independently captured images of the same scene to train a denoiser (Materials and methods). This requirement is impractical when the biological samples are highly dynamic or the total number of frames is limited due to photobleaching and phototoxicity. Resorting to the self-supervised training scheme, a single SIM capture for each scene is enough to train a PRS-SIM model. We compared PRS-SIM and

N2N-SIM using synthetic structure with different moving speeds (Supplementary Fig. 6) and noted that as the moving speed increased, N2N-SIM generated considerably deteriorated SIM images and was prone to oversmoothing the details of subcellular structures. Compared with N2N-SIM, the proposed PRS-SIM maintained a steady denoising performance regardless of the sample moving speed, indicating the superb live-cell imaging capability especially for samples of high dynamics.

PRS-SIM for multimodal SIM systems

Due to the internal similarity of the post-processing pipeline across various SIM modalities, besides TIRF/GI-SIM, PRS-SIM is compatible with other SIM configurations such as 3D-SIM, LLS-SIM, and NL-SIM, enabling higher resolution or volumetric SR imaging under low-light conditions. For 3D-SIM, we evaluated the performance of PRS-SIM by processing the images of lysosomes (Lyso) labelled with Lamp1-mEmerald in fixed COS7 cells (Fig. 4a-b and Supplementary Fig. 7). For each sample, 17 individual cells were imaged under low and high illumination conditions to acquire noisy data and the corresponding high SNR reference, respectively. The raw SIM data were first reconstructed
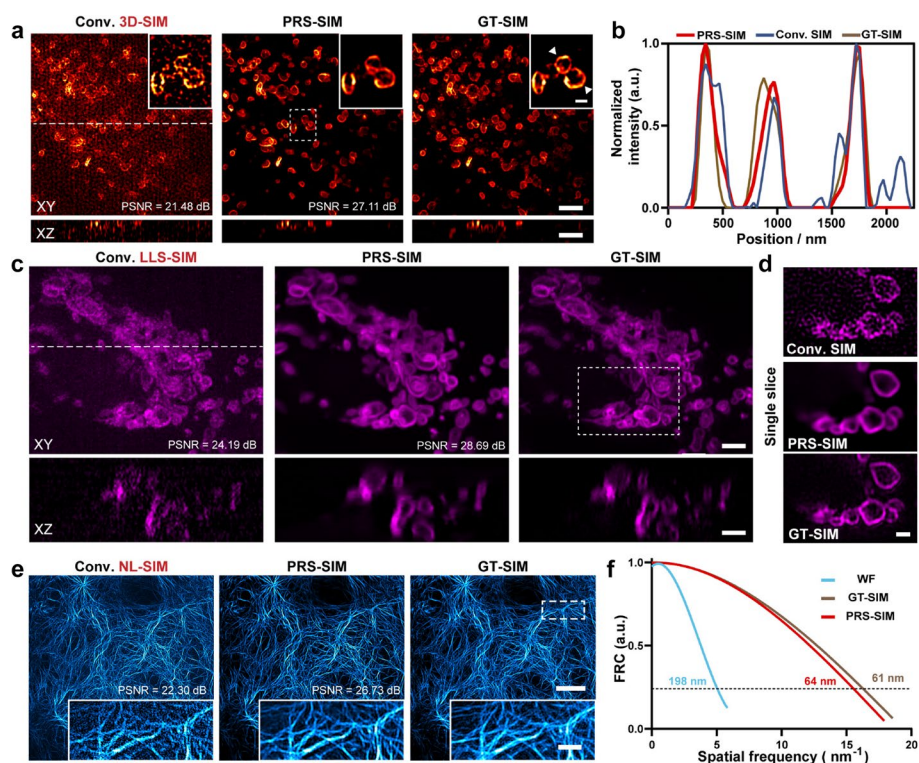
Fig. 4 PRS-SIM for multimodal SIM systems. **a** 3D-SIM images of lysosomes in fixed COS7 cells reconstructed with Conv. SIM and PRS-SIM accompanied with the corresponding GT-SIM image. Single slice view of the square region are provided for visualizing the details. Scale bar: 2 μm (regular), 0.5 μm (zoom-in regions). **b** Intensity profiles of Conv. SIM (blue), Sparse-SIM (green), PRS-SIM (red), and GT-SIM (brown) along the line indicated by the arrowheads in **a**. **c** LLS-SIM images of mitochondria in fixed COS7 cells reconstructed with Conv. SIM and PRS-SIM. Scale bar, 2 μm. **d** Representative single-slice image of the squared region in **c**. Scale bar, 1 μm. **e** NL-SIM images of F-actin in COS7 cells. WF, PRS-SIM and GT-SIM images are shown. **f** The FRC curves of the samples in **e**. The resolution is calculated based on the cutoff frequency with a threshold of 0.24. Scale bar, 5 μm (regular), 1 μm (zoom-in). In **a** and **c**, the XY plane is displayed in maximum intensity projection (MIP) view and the XZ plane is displayed in sectioned view (indicated by white dashed lines)

into 3D SR volumes via the conventional 3D-SIM algorithm and then denoised with 3D PRS-SIM models, which were modified into 3D U-net [48] architectures from the original 2D version (Materials and methods). The 3D PRS-SIM models were trained with the noisy data only. The orthogonal view of the representative PRS-SIM images (Fig. 4a) and line profiles (Fig. 4b) indicated that most of the noise-induced artifacts in the conventional 3D-SIM results were removed by PRS-SIM, and the reconstruction quality of PRS-SIM is comparable to that of GT-SIM in both the XY plane and Z-axis.

For the LLS-SIM configuration, we employed our home-built LLS-SIM system to acquire raw images of mitochondria (Mito) labelled with TOMM20-2xmEmerald, then trained a PRS-SIM model following a similar procedure. As is shown that PRS-SIM achieved a substantial improvement in both perceptual quality and statistical metrics (Fig. 4c-d and Supplementary Fig. 8) across a whole cell field-of-view (FOV) of $70\mu m \times 47\mu m \times 27\mu m$ (after de-skewing).

Compared with linear SIM, NL-SIM provide higher spatial resolution up to ~ 60 nm, however, at the expense of heavier photon budget, and is more subject to reconstruction artifacts [6]. To evaluate the effectiveness of PRS-SIM on NL-SIM denoising, we leveraged the NL-SIM images of F-actin labelled with Lifeact-SkylanNS in BioSR dataset [17] with signal levels ranging from 1 to 5 to train a PRS-NL-SIM model, then applied it to noisy NL-SIM images. The presented perceptual quality, quantitative evaluation, and FRC analysis (Fig. 4e-f, Supplementary Fig. 9) jointly indicate a superb capability of PRS-SIM to restore high frequency details of NL-SIM images without supervision. These results suggest that PRS-SIM shows a great potential for extending the application scope of multimodal SIM to low-light conditions without the need to acquire abundant training data.

Visualization of bioprocesses sensitive to phototoxicity

One major limitation of SIM is the requirement of high-intensity illumination, resulting in substantial phototoxic side effects. This phototoxicity largely limits the SR imaging duration for live specimens, particularly when imaging molecules with low expression levels or processes that are vulnerable to high-dose illumination. To demonstrate the potential of our method in reducing the required light dose, we first applied PRS-SIM to visualize clathrin-mediated endocytosis in gene-edited SUM159 cells expressing clathrin-EGFP at endogenous levels. The limited fluorescence of these cells prevents conventional TIRF-SIM (conv. TIRF-SIM) imaging from more than 150 frames, corresponding to an imaging time of ~ 3 min [6], because under low SNR conditions, conv. TIRF-SIM image contained substantial reconstruction artifacts (Fig. 5a). Although the fluorescence intensity of each raw image was 20-fold less than that of the high-SNR GT-SIM image (average photon count in raw images: 21.6±2.6 vs. 454.0±18.2, GT images collected only for the first 20 frames), PRS-SIM was still able to reconstruct high-fidelity SR information of the hollow, ring-like structure of CCPs (Fig. 5a). Therefore, PRS-SIM allowed us to characterize clathrin-mediated endocytosis at high spatiotemporal resolution for an unprecedented imaging duration of more than 5,000 frames, corresponding to an imaging time of more than 40 min (Supplementary Video 1). Previous studies have reported that clathrin-mediated endocytosis is initiated randomly based on analyses of the distribution of all CCP nucleation events over the limited observation window of ~ 7 min [49,
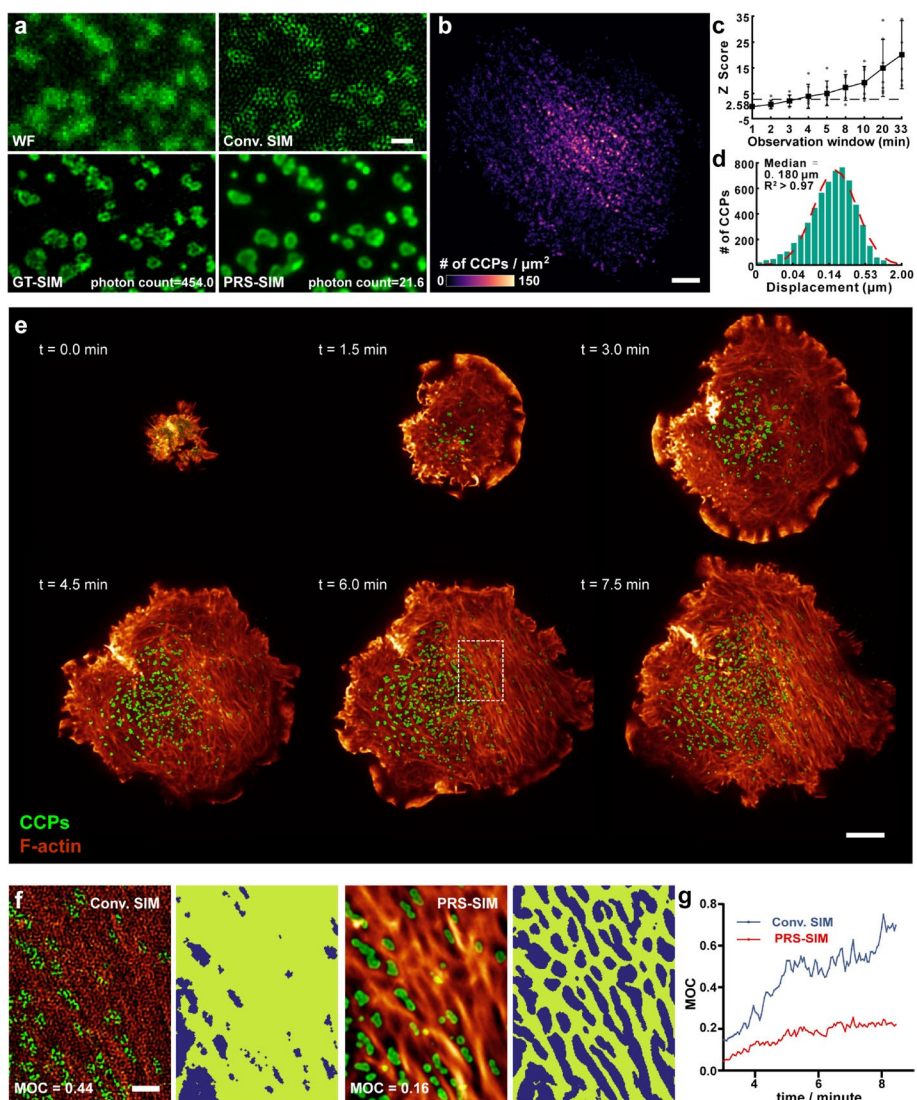
Fig. 5 Long-term observation of the bioprocesses sensitive to phototoxicity via PRS-SIM under low excitation power. **a** TIRF-SIM imaging of clathrin coated pits (CCPs) over 5000 frames (Supplementary Video 1). Although the collected fluorescence was ~20-fold lower than those used for acquiring artifact-free GT-SIM image, PRS-SIM image still conveys high-fidelity ring-like structure and prevents most artifacts fulfilled in conventional SIM image. **b** Spatial distribution of CCP nucleation events across the plasma membrane of a SUM-159 cell over the whole imaging duration. **c** z-score of CCP nucleation calculated from 7 cells rapidly increases as extending the observation window. z-score gets larger than 4.95 when observation window is longer than 4 min, indicating that there is a less than 1% likelihood that the clustered pattern of CCPs' nucleation could be the result of random occurrence. **d** Histogram of mean square displacement (MSD) of 3572 CCP tracks from 3 cells. **e** Dual-color time-lapse imaging of CCPs (green) and F-actin (red) in a live SUM159 cell during the adhesion process (Supplementary Video 2). The whole imaging duration is ~8 min and representative PRS-SIM denoised frames are displayed. **f** Zoom-in visualization of the interaction between CCPs and F-actin. The SR images (left) and the segmentation result of F-actin (right) are displayed. **g** Mander's overlapped coefficient (MOC) of the CCPs referring to F-actin during the cell adhesion. Lower MOC values indicated most CCPs are located in the gap of F-actin filament. Two curves are calculated based on the segmentation results from conv. SIM (blue) and PRS-SIM (red) images, respectively. Scale bar, 0.5 μm (**a**), 5 μm (**b**, **e**), 1 μm (**f**)

50]. By imaging the same process over 40 min, we found that most CCP nucleation sites tended to be spatially clustered (Fig. 5b, c, z-score > 20, $n = 7$ cells; Materials and methods), with many events occurring in confined regions, possibly at stable clathrin coated plaques [51]. Moreover, after tracking the CCP trajectories from their initiation to their detachment from the plasma membrane, we noted that the displacement of most CCPs was relatively small (Fig. 5d, Median = 0.180 μm). This finding is consistent with clathrin uncoating occurring near the site of invagination of the coated pit.

We also utilized PRS-SIM to investigate dynamic interactions between subcellular organelles and the cytoskeleton in SUM159 cells. Since the growing cells are light-sensitive and fragile, we decreased the illumination power to 10% of that used for usual experiments to image the entire adhesion process after dropping a SUM159 cell onto a coverslip. Under the low excitation intensity conditions, we successfully recorded the detailed interactions between CCPs and F-actin during the cell adhesion and migration for ~8 min with more than 170 SR-SIM frames (Fig. 5e, Supplementary Video 2). As shown in Fig. 5f, the hollow structure of CCPs (green) and the densely interlaced F-actin (orange) cannot be resolved in conventional SIM (conv. SIM) images due to the noise-induced artifacts. In contrast, the fine structures of CCPs and F-actin were both clearly distinguished by PRS-SIM, enabling further study of their detailed interactions. We next applied the Weka segmentation algorithm [52] to extract the filament skeleton and calculated the Mander's overlap coefficient (MOC) between the two structures in each frame (Materials and methods; Fig. 5f). We found that the MOC remained in a relatively small value during the whole adhesion process, indicating that most CCPs stayed at the interspace of actin filaments and were intensively regulated by the cytoskeleton throughout the adhesion process (Fig. 5g).

Long-term volumetric SR imaging of subcellular dynamics with adaptive trained PRS-SIM

Volumetric SIM imaging, such as 3D-SIM and LLS-SIM, causes severer photo-damage to live specimens than 2D-SIM (TIRF-SIM) [12]. To realize long-term volumetric SR live-cell imaging, we equipped our multi-SIM system with PRS-SIM and imaged a live COS7 cell expressing 3xmEmerald-Ensconsin (green) and Lamp1-Halo (red) in 3D-SIM mode under ~20-fold lower excitation power than typical imaging conditions (Fig. 6a-c). The data were acquired over 1 h (400 two-color 3D-SIM volumes at an interval of 10 s, Supplementary Video 3). During the data acquisition process, no decrease in cell activity was observed, indicating negligible phototoxicity effects. Although conventional SIM reconstruction reduces the out-of-focus fluorescence and improves the axial resolution, the detection noise severely degrades the image quality, preventing us from investigating the underlying bioprocesses. In contrast, the PRS-SIM model, which was trained by ~20 selected frames/volumes from the noisy time-lapse data, substantially removed the reconstruction artifacts and restored the fine structures of both organelles including continuous microtubule filaments and the hollow lysosomes. These advantages of PRS-SIM enable a clear volumetric observation of the dynamic interaction between microtubules and lysosomes, e.g., the directional movement of a lysosome along the MT filaments (Fig. 6b) and the hitchhiking remodeling mechanism of MT filaments under the traction of lysosomes (Fig. 6c).
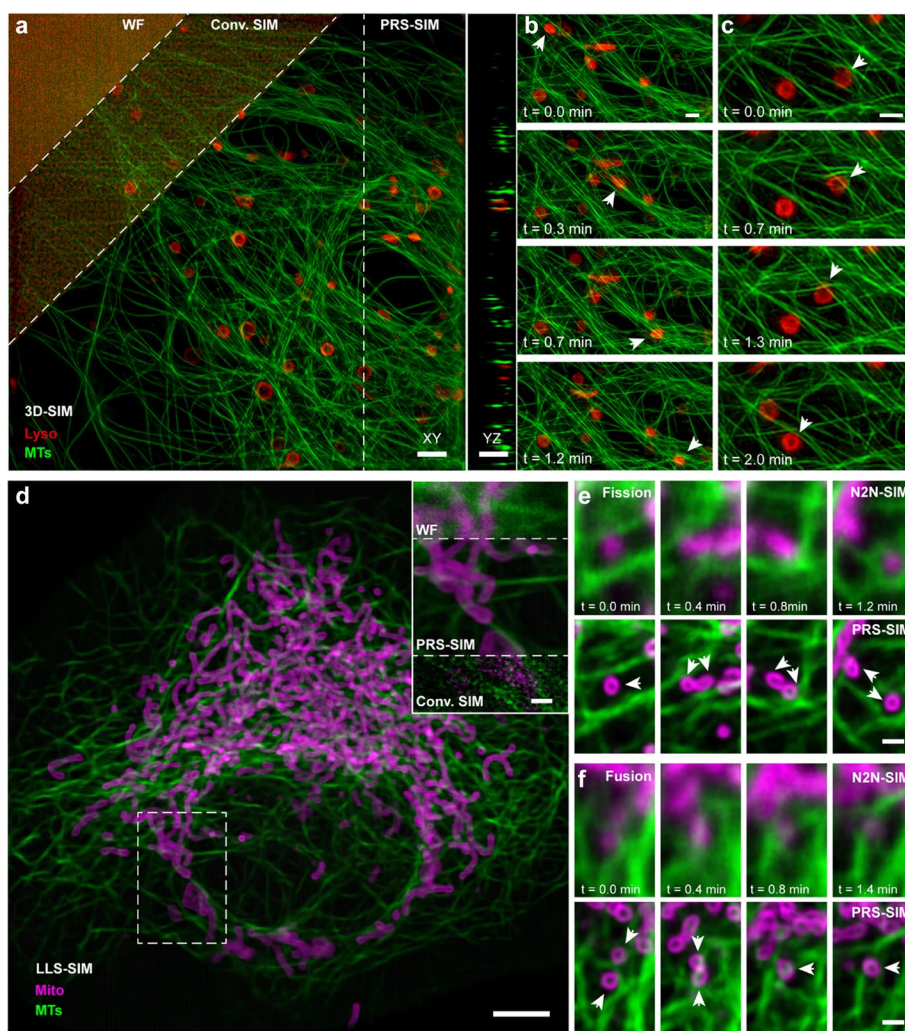
Fig. 6 Long-term volumetric super-resolution imaging of live cells with adaptively trained PRS-SIM. **a** 3D-SIM imaging of a live COS7 cell expressing 3xmEmerald-Ensconsin (green) and Lamp1-Halo (red) (Supplementary Video 3). The WF, Conv. 3D-SIM, and PRS-SIM results are compared. Scale bar, 2 μm. **b, c** Time-lapse PRS-SIM images of the dynamic interaction between lysosomes (Lyso) and microtubules (MTs) as the Lyso is moving along adjacent MTs (**b**) or deforming under the traction of MTs (**c**). Scale bar, 1 μm. **d** PRS-SIM enhanced LLS-SIM images of a live COS7 cell expressing TOMM20-2xmEmerald (magenta) and 3xmCherry-Ensconsin (green) (Supplementary Video 4). The zoom-in comparison of WF, Conv. SIM and PRS-SIM are displayed in the corner. Scale bar, 5 μm (regular), 1 μm (zoom-in). **e, f** Time-lapse recording of the rapid fission (**e**) and fusion (**f**) processes of mitochondria (Mito) interacted with MTs. The denoised images of N2N-SIM and PRS-SIM are compared. Scale bar, 1 μm

We next applied the PRS-SIM enhanced LLS-SIM system to record the volumetric subcellular dynamics of COS7 cells expressing TOMM20-2xmEmerald and 3xmCherry-Ensconsin (Fig. 6d-f). Two PRS-SIM models for Mito and MTs were independently trained with the noisy time-lapse data themselves, which consisted of ~310 two-color SIM volumes acquired at an interval of 12 s (Supplementary Video 4). We demonstrated that the adaptively trained PRS-SIM models removed most noise-induced artifacts and resolved the delicate structures of Mito and MTs (Fig. 6d). However, due to the rapid movement and deformation of the two observed structures, the classical denoising algorithm N2N [47] and its derivative DeepCAD [53, 54], which are based on the temporal

continuity between adjacent frames (Materials and methods), generated oversmoothed images with severe motion blur (Fig. 6e-f, Supplementary Fig. 10). With the prolonged observation window provided by PRS-SIM, we clearly identified the fission and fusion processes of Mito (Fig. 6e, f), which are some of the most common yet very important bioprocesses in live cells. Moreover, we emphasized that since the adaptive training mode of PRS-SIM utilizes only the noisy collected data for network training and then denoises themselves, there is no domain shift problem. Thus, the adaptively trained PRS-SIM models provide a high denoising fidelity and show great potential in the discovery of previously unseen biological structures and phenomena.

Conclusions

In summary, PRS-SIM is a novel self-supervised learning-based method for SIM image restoration, which trains the denoiser with only noisy data and reconstructs artifact-free SR-SIM images with ~20-fold less fluorescence than routine SIM imaging conditions. The proposed self-supervised strategy eliminates the need for high-SNR GT data or repeated acquisition to construct the training dataset. Consequently, this easy-to-implement data acquisition scheme is applicable to biological specimens of high dynamics or with low fluorescence efficiency. Moreover, although both PRS-SIM and previously proposed Neighbor2Neighbor [35] utilize the similarity between adjacent pixels to for self-supervised denoising, we emphasize that their implementation is quite different and PRS-SIM exhibits superior performance in SIM application (Supplementary Fig. 3). For long-term live-cell imaging, PRS-SIM can be applied in the adaptive training mode, where acquired noisy data are directly used to train the denoising model. Therefore, no pre-trained models for the same samples are needed, and with this advance, PRS-SIM is potential to discover previously unknown biological structures and phenomena. Finally, our method is applicable to multiple SIM modalities, including TIRF/GI-SIM, 3D-SIM, LLS-SIM, and even NL-SIM. With PRS-SIM, we achieved long-term live observations of subcellular dynamics and diverse bioprocesses with extremely low invasiveness, demonstrating the broad applicability of our method. Furthermore, to enhance the accessibility of PRS-SIM for biological research, we developed an easy-to-use Fiji plugin [55] (Supplementary Note 3, Supplementary Figs. 11–12), with which the training and inference of PRS-SIM models can be easily carried out by several clicks for users of biological background.

PRS-SIM can be improved in several ways. First, successful PRS-SIM reconstruction relies on accurate estimations of the SIM patterns, which is challenging under extremely low-light conditions for conventional SIM parameter estimation algorithm. Therefore, employ advanced algorithm such as PCA-SIM [56], or an additional neural network for more precise parameter estimation may improve the robustness of PRS-SIM. Second, incorporating other advanced analytical SIM reconstruction algorithms into PRS-SIM framework, e.g., HiFi-SIM [11], True-Wiener-SIM [10], JSFR-(AR)-SIM [45, 57] etc., is potential to further improve the fidelity and quality. Third, while PRS-SIM effectively mitigates noise-induced artifacts, it cannot solve the artifacts caused by other factor, such as imperfect optical system, uneven illumination patterns or sample scattering. For instance, when imaging thick samples, both the illumination pattern and the detected fluorescence will be strongly deviated, resulting in deteriorated image quality,

which cannot be restored by PRS-SIM. Integrating PRS-SIM into an optical system embedded with adaptive optics [58–60] or multi-focus detection [61] modules could further address this issue. We believe with continuous evolution, PRS-SIM is potential to become a universal tool for SR-SIM users, and it can play important roles in revealing complicated biological processes such as rapid dynamics and interactions of organelles during light-sensitive bioprocesses.

Materials and methods

Optical setup

All the experiments in this work were performed on our home-built multi-modality SIM system (Multi-SIM) or lattice light-sheet SIM (LLS-SIM) system. The Multi-SIM system is extended our previous study [7], where the TIRF-SIM, GI-SIM and 3D-SIM mode are integrated. Briefly, three laser beams (488 nm, 560 nm, and 640 nm) were collimated for multi-channel excitation and controlled by an AOTF for rapid switching. The structured illumination patterns were generated by a ferroelectric spatial light modulator (SLM, QXGA-3DM, Forth Dimension Display) placed conjugated to the sample plane. In our experiments, the illumination patterns of 3-phase×3-orientation for TIRF-SIM mode and 3-phase×5-orientation for 3D-SIM mode were generated. The effective NA of the excitation pattern is 1.43 for TIRF-SIM and 1.2 for 3D-SIM. An excitation path and detection path shared the same objective (oil immersion, 1.49 NA, Nikon). The final images were detected by a sCMOS camera (Hamamatsu, Orca Flash 4.0 v3).

The LLS-SIM system is developed based the the original setup [32]. For our configuration, three laser beams (488 nm, 560 nm, and 640 nm) were used for multi-color excitation. The illumination pattern is displayed on the SLM (the same as Multi-SIM) and then filtered by an annular mask (outer NA: 0.5; inner NA: 0.375) to obtain a balanced axial and lateral resolution. The z-axis scanning is implemented by a high-speed galvo mirror (Cambridge Technology). The emission fluorescence is collected by a water-immersion objective (1.1 NA, Nikon) and then imaged by a sCMOS camera (Hamamatsu, Orca Fusion). During the imaging, each z-slice is illuminated by the patterns of 3-phase×1-orientation. The oblique angle between the illumination path and the detection path is 29.7°.

Data acquisition

The experiments in this work can be categorized as static sample imaging and time-lapse live-cell imaging. For static sample imaging, we utilized the data from the open-source dataset BioSR [17] or acquired via our home-built SIM systems. For TIRF-SIM experiments, the CCPs, ER, and MTs images whose signal levels range from 1 to 4 (fixed cells of MTs and CCPs) or 1 to 3 (live cells of ER) in BioSR were used to create the training dataset. The GT-SIM images are accompanied in the dataset. For NL-SIM experiments, the F-actin images whose signal levels range from 1 to 5 in BioSR were used to create the training dataset. For 3D-SIM and LLS-SIM experiments, the dataset used for both training and inference was acquired with our home-built Multi-SIM and LLS-SIM systems. Specifically, for each type of specimen, we acquired ∼ 20 sets of raw SIM images at three or four escalating levels of excitation light intensity to create the training dataset, and then tuned the laser power to the maximum to capture the

high-SNR images as the corresponding GT data. Notably, the training dataset is generated purely with the low-SNR data, and the high-SNR GT data are only used as the reference for quantitative analysis.

For time-lapse imaging, the 2D and 3D experiments were carried out with the TIRF-SIM and 3D-SIM mode of the Multi-SIM system, respectively. The excitation light power used in all live experiments was set to 5-10% as that used in common imaging conditions with short exposure durations, corresponding to an average photon count of $20 \sim 50$ for each raw SIM image, to minimize the phototoxicity and photobleaching effects. The specific imaging conditions for each time-lapse experiment were listed in Supplementary Table 1.

Pixel-realignment strategy

The self-supervised training dataset was generated with the pixel-realignment strategy. The raw dataset consists of a series of low-SNR raw SIM image groups. Each individual image in a group is a WF image under a specific illumination pattern (e.g., 3-orientation $\times$ 3-phase for 2D/TIRF-SIM and 3-orientation $\times$ 5-phase $\times$ Z-slice for 3D-SIM). For each raw SIM image group, the generation of the training dataset of PRS-SIM models mainly takes the following steps:

(i) Each raw image is divided into 4 sub-images by applying a 2×2 down-sampler, forming four sub-image groups.
(ii) The augmented four sub-image groups are up-sampled into the original size with the nearest interpolation.
(iii) Based on the position of the valid pixel in each 2×2 cell, a sub-pixel translation is applied to each raw image, which guarantees that they are well spatially calibrated with each other.
(iv) The generated sub-images groups are reconstructed into four noisy SIM images using the conventional SIM algorithm.
(v) Then several image patched pairs are generated by randomly selecting two out of four noisy SIM images as the input and target, respectively.

For 3D-SIM stacks, both the down-sampling, up-sampling and translation operations in step (i)-(iii) are implemented in a slice-by-slice manner and the input and the target of the network are accordingly changed to a 3D-image stack. By applying step (i)-(v) to all noisy SIM image groups, the complete training dataset is generated. Typically, ~ 5 or even fewer individual image groups are adequate for training a robust PRS-SIM model (Supplementary Figs. 13–14).

Please note that in PRS-SIM, both the input and target data for network training is the aligned SR image after SIM reconstruction. Although it is also feasible to train a denoiser by replacing the input SR image with the aligned raw images (without SIM reconstruction), in this situation, the resolution of the denoised image will be reduced due to the spectral bias issue [26, 27] (Supplementary Fig. 15). Moreover, the up-sampling operation embedded in PRS-SIM can be implemented by any common used interpolation method, including nearest interpolation, bilinear interpolation, and

bicubic interpolation, which demonstrated similar denoising performance based on our experiment (Supplementary Fig. 16).

Conventional SIM reconstruction

As demonstrated in Fig. 1, PRS-SIM employs conventional SIM reconstruction [1, 2] in its framework to generated the noisy SR image. Briefly, it takes the following steps (detailed discussions in Supplementary Note 1:

  (i)   Determine the pattern modulation parameters by loading the system configuration file or estimating from the detected images.
 (ii)   Apply Fourier transformation to each raw image.
(iii)   Separate the low-frequency components and high-frequency components of each raw image.
 (iv)   Shift the high-frequency component of each raw image to the corresponding position in Fourier domain based on the pattern modulation parameters.
  (v)   Combine all the components together with a generalized Wiener filter and an apodization function.
 (vi)   Apply inversed Fourier transformation on the expanded Fourier spectrum and generate the final super-resolution image.

The open-source code of conventional SIM reconstruction applicable in PRS-SIM includes 3-beam-SIM, [2] fairSIM [62], Open-3DSIM [63] and PCA-SIM [56].

Network architecture

PRS-SIM employs U-net [48] as the backbone architecture, which has already shown superior performance in denoising task elsewhere [47] (Supplementary Fig. 17). The network is composed of an encoder module and a decoder module. For the encoder module, the input data is firstly fed into a convolutional layer with 48 kernels and then encoded by five consecutive encoding blocks. Each encoding block consists of a convolutional layer followed by a non-linear activation layer and a max-pooling layer for spatial down sampling. For the decoder module, five decoding blocks are involved, each of which consists of two consecutive convolutional layer and a nearest interpolation layer for spatial up sampling. Skip-connections were embedded between the encoding and decoding blocks to prevent over-fitting. Two additional convolutional layers were placed at the end of the network to transfer the final denoised image into the same shape as the input image. Concretely, the kernel size of all the convolutional layers is 3×3 and the activation function used is Leaky-ReLU, which is defined as:

$$LeakyReLU(x) = \max(0, x) + \gamma \cdot \min(0, x), \tag{1}$$

where γ denotes the negative slope coefficient (set as 0.1 in our experiments). For 3D-SIM applications, all the convolutional layers and pooling layers were replaced with the corresponding 3D versions and the other parts remained unchanged. It is noted that although all the experiments in this manuscript is implemented with U-net, PRS-SIM is also compatible with other network backbone, including RCAN [64], RDN [65] and uFormer [42] (Supplementary Fig. 18).

Data processing and network training

The training dataset of PRS-SIM consist of a series of image pairs generated only from the low-SNR raw images as described in previous sections. For pre-trained PRS-SIM models, 20–40 distinct ROIs of each type of specimens were imaged to create the training dataset. For adaptive training mode of PRS-SIM, ~20 frames/volumes were randomly selected from the entire time series. The aligned images were generated by the proposed pixel-realignment strategy from the noisy raw images. For data augment, the input and target image pair were first to be randomly cropped into patches that match the size of network input ($128\times$ 128 pixels for 2D/TIRF/LLS-SIM and $64\times64\times8$ voxels for 3D-SIM). Random rotation and flipping were optional to be employed to enrich the dataset and avoid overfitting. The total number of mini-patch used for training is ~100,000 (Supplementary Fig. 19).

During the network training, Adam optimizer with an initial learning rate of 10^{-4} was adopted to accelerate the convergence. Although based on the theoretical derivation of PRS-SIM (Supplementary Note 1), L2-norm was set as the default loss function, it can also be alternated with L1-norm without obvious compromised performance (Supplementary Fig. 20). A multi-step scheduler was employed to decrease the learning rate by a factor of 0.5 at the designated epochs. The training processes were performed on a workstation equipped with a graphics processing unit (Nvidia GeForce RTX 3090Ti, 24GB memory). The source codes were written based on PyTorch v1.5 framework in Python v3.6. The typical training time for a dataset of ~100,000 mini-patch pairs is about 2 h for 2D batches and 4 h for 3D batches. More training details of the experiments in this work were listed in Supplementary Table 2.

For the inference phase, the noisy raw SIM images were reconstructed into SR images via conventional SIM algorithm, divided into several tiled patches of 256×256 pixels with 10% overlap, fed into the pre-trained network, and finally stitched together to form the denoised SR images. For adaptive training mode of PRS-SIM, the time-lapse data was denoised with the model trained by itself, while in other experiments the data was denoised with the pre-trained network of the same type of specimens.

For N2N-SIM training in Fig. 6e-f, Supplementary Fig. 6, and Supplementary Fig. 10, we randomly selected two consecutive frames/volumes from the time-lapse data used as the input and target, respectively. The whole training dataset are generated from ~20 independent frame/volume pairs. Other operations and configurations during training and inference are the same as PRS-SIM.

Image assessment metrics

To quantitatively evaluate the denoised images output by PRS-SIM, we employed the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) between the denoised image I referring to the GT image I_{gt} as the metric. Since the signal intensity of the denoised and GT images is of different dynamic range, we first applied percentile normalization to I and I_{gt} as:

$$\widetilde{I} = \frac{I - \mathrm{prctile}(I, p_{min})}{prctile(I, p_{max}) - prctile(I, p_{min})}, \tag{2}$$

$$\widetilde{I}_{gt} = \frac{I_{gt} - \mathrm{prctile}(I_{gt}, p_{min})}{\mathrm{prctile}(I_{gt}, p_{max}) - \mathrm{prctile}(I_{gt}, p_{min})}, \tag{3}$$

where $prctile(I, p)$ denotes the intensity of the pixel ranking at $p\%$ of image I, and $\widetilde{I}$ denotes the corresponding normalized image. The p_{min} and p_{max} are set as 0.1 and 100 in our analysis. To further alleviate the disturbance in metric calculation, we implemented a linear transformation to the normalized image $\widetilde{I}$ by:

$$\widetilde{I}_{trans} = \alpha \, \widetilde{I} + \beta \tag{4}$$

where α and β denote the transformation coefficients to minimize the square root error between the transformed image and the normalized GT image, which can be formulated as a linear regression problem:

$$\min \|\alpha \widetilde{I} + \beta - \widetilde{I}_{gt}\|_2^2, \tag{5}$$

where $\| \cdot \|_2$ is the L2-norm. The closed solution of this problem is:

$$\widehat{\alpha} = \frac{\sum \widetilde{I}_{gt} \cdot \left(\widetilde{I} - \mathrm{mean}\left(\widetilde{I}\right)\right)}{\sum \widetilde{I}^2 - N \cdot \mathrm{mean}\left(\widetilde{I}\right)^2}, \tag{6}$$

$$\widehat{\beta} = N \cdot \sum (\widetilde{I}_{gt} - \widehat{\alpha} \bullet \widetilde{I}), \tag{7}$$

where N is the pixel number of the image, $\sum \cdot$ denotes the pixel-wise sum, $\widehat{\alpha}$ and $\widehat{\beta}$ denote the optimal values of the transformation coefficients α and β, respectively. Then the final PSNR and SSIM are calculated as:

$$PSNR\left(\widetilde{I}_{trans}, \widetilde{I}_{gt}\right) = 10 \bullet log10\left(\frac{1}{\frac{1}{N}\sum\left(\widetilde{I}_{trans} - \widetilde{I}_{gt}\right)^2}\right) \tag{8}$$

$$SSIM\left(\widetilde{I}_{trans}, \widetilde{I}_{gt}\right) = \frac{\left(2\mu_{\widetilde{I}_{trans}}\mu_{\widetilde{I}_{gt}} + C_1\right)\left(2\sigma_{\widetilde{I}_{trans}\widetilde{I}_{gt}} + C_2\right)}{\left(\mu_{\widetilde{I}_{trans}}^2 + \mu_{\widetilde{I}_{gt}}^2 + C_1\right)\left(\sigma_{\widetilde{I}_{trans}}^2 + \sigma_{\widetilde{I}_{gt}}^2 + C_2\right)} \tag{9}$$

where $\mu_{\widetilde{I}_{trans}}$, $\mu_{\widetilde{I}_{gt}}$ and $\sigma_{\widetilde{I}_{trans}}$, $\sigma_{\widetilde{I}_{gt}}$ denote the mean values and standard deviations of image $\widetilde{I}_{trans}$ and $\widetilde{I}_{gt}$, respectively; $\sigma_{\widetilde{I}_{trans}\widetilde{I}_{gt}}$ denotes the cross-covariance between $\widetilde{I}_{trans}$ and $\widetilde{I}_{gt}$. The constant C_1 and C_2 used in this paper is 0.01^2 and 0.03^2, respectively.

To characterize the resolution of the images output by PRS-SIM, we employed single-image based Fourier ring correlation (FRC) method [38]. The raw image I is split into

two sub-images I_1 and I_2 by interleaved pixel extraction. Then the FRC value of the central ring region with radius R is calculated as:

$$FRC(R) = \frac{\sum_{r<R} \mathcal{F}(I_1)\mathcal{F}(I_2)^*}{\sqrt{\sum_{r<R}|\mathcal{F}(I_1)^2|}\cdot\sqrt{\sum_{r<R}|\mathcal{F}(I_1)^2|}}, \tag{10}$$

where the symbol $\mathcal{F}$ denotes Fourier transformation. By calculating the FRC value from 0 to R_{max} (the reciprocal of the pixel size), a generally declining curve is formulated. The resolution can be measured as the reciprocal of the Fourier cutoff frequency R_{cutoff}, where $FRC(R_{cutoff}) < tsh$, where tsh represents the spectral intensity threshold. In our analysis, the tsh is set as a typical value of 0.25.

Data analysis

We utilized the spatial autocorrelation (i.e., Global Moran's Index [66]) to evaluate whether the distribution of clathrin coated pit (CCP) nucleation sites is clustered, dispersed, or random. For each time-lapse dataset, we first localized the centroid positions of all CCPs at each time point, and then linked them temporally in the whole time series using the ImageJ plugin TrackMate [67], thus yielded trajectories of all detected CCPs. To rule out the false-positive events, the trajectories of less than 40 time points corresponding to a duration of 20 s were excluded from following computation. Subsequently, for each time-lapse data, the initial locations of the CCP trajectories detected in the designated observation window were projected onto the same image as the CCP nucleation sites' map (Fig. 3b). Then, the Moran's Index can be calculated as:

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}d_{i,j}z_iz_j}{\sum_{i=1}^{n}z_i^2}, \tag{11}$$

where $z_i = (x_i - \bar{X})$ is the deviation of the event count of the ith pixel from the average count; $d_{i,j}$ refers to the inverse Euclidean distance between pixel i and j; n is the total pixel number of the map and $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n}d_{i,j}$ is the summation of $d_{i,j}$. Finally, the z-score was calculated for each nucleation sites map to evaluate the significance of the Moran's Index (Fig. 3c):

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}, \tag{12}$$

where $E[\bullet]$ and $V[\bullet]$ are the expectation and the variance of I, respectively. In general, the larger z-score indicates the stronger tendency of clustering.

To quantitatively investigated the interaction of organelles during the cell adhesion (Fig. 3e-g), we calculated the Mander's overlapped coefficient (MOC) [68] of CCPs referring to F-actin. For each frame, a binary mask (denoted as M) is firstly generated by applying a threshold tsh_M to the F-actin channel, which represents the F-actin skeleton:

$$M = I_{F-actin} > tsh_M \tag{13}$$

Then the *MOC* value is calculated as:

$$MOC = \frac{\sum_M I_{CCPs}}{\sum_{total} I_{CCPs}},\tag{14}$$

where $\sum_M I_{CCPs}$ and $\sum_{total} I_{CCPs}$ denote the intensity summation of the CCP channel within the masked region and the entire image, respectively.

Cell culture, transfection, and stain

Cos7 cells were cultured in DMEM (Gibco), supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin in 37℃ with 5% CO_2. For live cell imaging, the coverslips were pre-coated with 50 μg ml^{-1} of collagen and cells were seeded onto coverslips with about 70% density before transfection. After 12 h, cells were transfected with plasmids using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol. Cells were imaged 12–24 h after transfection in a stage top incubator (Okolab) to maintain condition at 37℃ with 5% CO_2. The plasmid constructs used in this work were 3xmEmerald-Ensconsin, Lamp1-mEmerald, TOMM20-2xmEmerald, calnexin-mEmrald, and Lamp1-Halo.

SUM159 cells were genome edited to incorporate EGFP to the C-terminus of clathrin light chain A (clathrin-EGFP) using the TALEN-based approach [69]. The clathrin-EGFP expressing cells were enriched by two sequential bulk sorting. The cells were cultured in DMEM/F-12 (Gibco) medium supplemented with 5% fetal bovine serum (Gibco), 5 μg/ml Bovine insulin (Cell Applications), 10 mM HEPES (Gibco), 1 μg/mL Hydrocortisone (Sigma) and 1% Penicillin-Streptomycin (Gibco) in 37℃ with 5% CO2. For dual-color experiments, these SUM-ki-CLAT-GFP cells were further transfected with the lifeact-Halo. Before imaging, we digested the cells using 0.25% Trypsin, and then dropped cell suspension onto the coverslip pre-treated with 50 μg/mL collagen.

Abbreviations
SIM	structured illumination microscopy
PRS-SIM	pixel-realignment based self-supervised reconstruction for SIM
N2N	noise-to-noise
TIRF	total internal reflection fluorescent
WF	wide field
LLS	lattice light sheet
SR	super-resolution
HR	high-resolution
LR	low-resolution
SNR	signal-to-noise ratio
SSIM	structural similarity
PSNR	peak signal-to-noise ratio
PSF	point spread function
OTF	optical transfer function
GT	ground truth
MT	microtubule
CCP	clathrin coated pit
ER	endoplasmic reticulum
Mito	mitochondrion
Lyso	lysosome

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43074-024-00121-y.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Availability of data and materials
The dataset supporting this work is partial from public dataset bioSR [37] at https://figshare.com/articles/dataset/BioSR/13264793. All other data are available from the corresponding authors upon reasonable request. The relevant code is available at https://github.com/IntelIM-THUIBCS/PRS-SIM.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
X.C., C.Q. and D.L. have a pending patent application on the presented frameworks.

References

1. Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc. 2000;198:82–7.
2. Gustafsson MG, Shao L, Carlton PM, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys J. 2008;94:4957–70.
3. Karras C, Smedh M, Förster R, et al. Successful optimization of reconstruction parameters in structured illumination microscopy–a practical guide. Opt Commun. 2019;436:69–75.
4. Demmerle J, Innocent C, North AJ, et al. Strategic and practical guidelines for successful structured illumination microscopy. Nat Protoc. 2017;12:988–1010.
5. Kner P, Chhun BB, Griffis ER, et al. Super-resolution video microscopy of live cells by structured illumination. Nat Methods. 2009;6:339–42.
6. Li D, Shao L, Chen B-C, et al. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. Science. 2015;349:aab3500.
7. Guo Y, Li D, Zhang S, et al. Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales. Cell. 2018;175:1430-1442. e1417.
8. Chu K, McMillan PJ, Smith ZJ, et al. Image reconstruction for structured-illumination microscopy with low signal level. Opt Express. 2014;22:8687–702.
9. Labouesse S, Negash A, Idier J, et al. Joint reconstruction strategy for structured illumination microscopy with unknown illuminations. IEEE Trans Image Process. 2017;26:2480–93.

10. Smith CS, Slotman JA, Schermelleh L, et al. Structured illumination microscopy with noise-controlled image reconstructions. Nat Methods. 2021;18:821–8.
11. Wen G, Li S, Wang L, et al. High-fidelity structured illumination microscopy by point-spread-function engineering. Light: Sci Appl. 2021;10:70.
12. Hagen N, Gao L, Tkaczyk TS. Quantitative sectioning and noise analysis for structured illumination microscopy. Opt Express. 2012;20:403–13.
13. Wu Y, Shroff H. Faster, sharper, and deeper: structured illumination microscopy for biological imaging. Nat Methods. 2018;15:1011–9.
14. Liu SB, Xie BK, Yuan RY, et al. Deep learning enables parallel camera with enhanced-resolution and computational zoom imaging. PhotoniX. 2023;4:1–20.
15. Wu Y, Han X, Su Y, et al. Multiview confocal super-resolution microscopy. Nature. 2021;600:279–84.
16. Kreiss L, Jiang S, Li X, et al. Digital staining in optical microscopy using deep learning - a review. PhotoniX. 2023;4:1–32.
17. Qiao C, Li D, Guo Y, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods. 2021;18:194–202.
18. Huang B, Li J, Yao B, et al. Enhancing image resolution of confocal fluorescence microscopy with deep learning. PhotoniX. 2023;4:1–22.
19. Wang H, Rivenson Y, Jin Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nat Methods. 2019;16:103–10.
20. Jin L, Liu B, Zhao F, et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. Nat Commun. 2020;11:1–7.
21. Qiao C, Chen X, Zhang S, et al. 3D structured illumination microscopy via channel attention generative adversarial network. IEEE J Sel Top Quantum Electron. 2021;27:1–11.
22. Shah ZH, Müller M, Wang TC, et al. Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. Photonics Res. 2021;9:B168-181.
23. Yang W, Zhang X, Tian Y, et al. Deep learning for single image super-resolution: a brief review. IEEE Trans Multimedia. 2019;21:3106–21.
24. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell. 2015;38:295–307.
25. Qiao C, Li D, Liu Y, et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. Nat Biotechnol. 2023;41:1–11.
26. Rahaman N, Baratin A, Arpit D, Draxler F, Lin M, Hamprecht F, Bengio Y, Courville A. On the spectral bias of neural networks. International Conference on Machine Learning. 2019. p. 5301–5310. PMLR.
27. Xu ZQ, Zhang Y, Luo T, Xiao Y, Ma Z. Frequency principle: Fourier analysis sheds light on deep neural networks. arXiv preprint arXiv:1901.06523. 2019.
28. Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process. 2017;26:3142–55.
29. Zhang K, Zuo W, Zhang L. FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans Image Process. 2018;27:4608–22.
30. Weigert M, Schmidt U, Boothe T, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. Nat Methods. 2018;15:1090–7.
31. Chen J, Sasaki H, Lai H, et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. Nat Methods. 2021;18:678–87.
32. Chen B-C, Legant WR, Wang K, et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. Science. 2014;346: 1257998.
33. Rego EH, Shao L, Macklin JJ, et al. Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution. Proceedings of the National Academy of Sciences. 2012;109:E135–43.
34. Gustafsson MG. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. Proceedings of the National Academy of Sciences. 2005;102:13081–6.
35. Huang T, Li S, Jia X, et al. Neighbor2neighbor: a self-supervised framework for deep image denoising. IEEE Trans Image Process. 2022;31:4023–38.
36. Lequyer J, Philip R, Sharma A, et al. A fast blind zero-shot denoiser. Nat Mach Intell. 2022;4:1–11.
37. Qiao C, Li D. BioSR: a biological image dataset for super-resolution microscopy. Figshare; 2022. https://figshare.com/articles/dataset/BioSR/13264793.
38. Nieuwenhuizen RP, Lidke KA, Bates M, et al. Measuring image resolution in optical nanoscopy. Nat Methods. 2013;10:557–62.
39. Krull A, Buchholz TO, Jug F. Noise2void-learning denoising from single noisy images. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 2129–2137.
40. Prakash M, Delbracio M, Milanfar P, Jug F. Interpretable unsupervised diversity denoising and artefact removal. arXiv preprint arXiv:2104.01374. 2021.
41. Pang T, Zheng H, Quan Y, Ji H. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021. p. 2043–2052.
42. Wang Z, Liu J, Li G, Han H. Blind2unblind: Self-supervised image denoising with visible blind spots. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 2027–2036.
43. Zhang G, Li X, Zhang Y, et al. Bio-friendly long-term subcellular dynamic recording by self-supervised image enhancement microscopy. Nat Methods. 2023;20:1957–70.
44. Zhao W, Zhao S, Li L, et al. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. Nat Biotechnol. 2022;40:606–17.
45. Wang Z, Zhao T, Hao H, et al. High-speed image reconstruction for optically sectioned, super-resolution structured illumination microscopy. Adv Photonics. 2022;4:026003–026003.

46. Wen G, Li S, Liang Y, et al. Spectrum-optimized direct image reconstruction of super-resolution structured illumination microscopy. PhotoniX. 2023;4:1–18.
47. Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, Aila T. Noise2Noise: Learning image restoration without clean data. arXiv preprint arXiv:1803.04189. 2018.
48. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. arXiv preprint arXiv:1505.04597. 2015.
49. Ehrlich M, Boll W, Van Oijen A, et al. Endocytosis by random initiation and stabilization of clathrin-coated pits. Cell. 2004;118:591–605.
50. Godlee C, Kaksonen M. From uncertain beginnings: initiation mechanisms of clathrin-mediated endocytosis. J Cell Biol. 2013;203:717–25.
51. Cureton DK, Massol RH, Saffarian S, et al. Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization. PLoS Pathog. 2009;5: e1000394.
52. Arganda-Carreras I, Kaynig V, Rueden C, et al. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics. 2017;33:2424–6.
53. Li X, Zhang G, Wu J, et al. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. Nat Methods. 2021;18:1395–400.
54. Li X, Li Y, Zhou Y, et al. Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. Nat Biotechnol. 2023;41:282–92.
55. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9:676–82.
56. Qian J, Cao Y, Bi Y, et al. Structured illumination microscopy based on principal component analysis. eLight. 2023;3:4.
57. Wang Z, Zhao T, Cai Y, et al. Rapid, artifact-reduced, image reconstruction for super-resolution structured illumination microscopy. The Innovation. 2023;4:100425.
58. Hampson KM, Turcotte R, Miller DT, et al. Adaptive optics for high-resolution imaging. Nat Reviews Methods Primers. 2021;1:1–26.
59. Zhou Z, Huang J, Li X, et al. Adaptive optical microscopy via virtual-imaging-assisted wavefront sensing for high-resolution tissue imaging. PhotoniX. 2022;3:1–20.
60. Turcotte R, Liang Y, Tanimoto M, et al. Dynamic super-resolution structured illumination imaging in the living brain. Proc Natl Acad Sci. 2019;116:9586–91.
61. Zhang Y, Wang Y, Wang M, et al. Multi-focus light-field microscopy for high-speed large-volume imaging. PhotoniX. 2022;3:1–20.
62. Müller M, Mönkemöller V, Hennig S, et al. Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ. Nat Commun. 2016;7:1–6.
63. Cao R, Li Y, Chen X, et al. Open-3DSIM: an open-source three-dimensional structured illumination microscopy reconstruction platform. Nat Methods. 2023;20:1183–6.
64. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image super-resolution using very deep residual channel attention networks. Proceedings of the European conference on computer vision (ECCV). 2018. p. 286–301.
65. Huang G, Liu Z, Pleiss G, et al. Convolutional networks with dense connectivity. IEEE Trans Pattern Anal Mach Intell. 2019;44:8704–16.
66. Ehlschlaeger CR, Shortridge AM, Goodchild MF. Visualizing spatial data uncertainty using animation. Computers Geosci. 1997;23:387–95.
67. Tinevez J-Y, Perry N, Schindelin J, et al. TrackMate: an open and extensible platform for single-particle tracking. Methods. 2017;115:80–90.
68. Manders E, Verbeek F, Aten J. Measurement of co-localization of objects in dual-colour confocal images. J Microsc. 1993;169:375–82.
69. Sanjana NE, Cong L, Zhou Y, et al. A transcription activator-like effector toolbox for genome engineering. Nat Protoc. 2012;7:171–92.

Publisher's Note