

Bio-friendly and high-precision super-resolution imaging through self-supervised reconstruction structured illumination microscopy

Received: 25 April 2025

Accepted: 29 October 2025

Published online: 12 December 2025



Jiahao Liu^{1,6}, Xue Dong^{2,6}, Huaide Lu^{3,6}, Tao Liu^{1,6}, Wei Liu^{4,6}, Xinyao Hu³, Quan Meng², Amin Jiang³, Tao Jiang³, Xiaohan Geng³, Haosen Liu⁵, Jun Cheng¹, Edmund Y. Lam⁵, Yan-Jun Liu⁴✉, Shan Tan¹✉ & Dong Li²✉

Deep-learning-based structured illumination microscopy (SIM) has demonstrated substantial potential in long-term super-resolution imaging of biostructures, enabling the study of subcellular dynamics and interactions in live cells. However, the acquisition of ground-truth (GT) data for training poses inherent challenges, limiting its universal applicability. Current approaches without using GT training data compromise reconstruction fidelity and resolution, and the lack of physical priors in end-to-end networks further limits these qualities. Here we developed self-supervised reconstruction (SSR)-SIM by combining statistical analysis of reconstruction artifacts with structured light modulation priors to eliminate the need for GT and improve reconstruction precision. We validated SSR-SIM on common biological datasets and demonstrated that SSR-SIM enabled long-term recording of dynamic events, including cytoskeletal remodeling in cell adhesion, mitochondrial cristae remodeling, interactions between viral glycoprotein and endoplasmic reticulum, endocytic recycling of transferrin receptors, vaccinia-virus-induced actin comet remodeling, and mitochondrial intercellular transfer through tunneling nanotubes.

Super-resolution (SR) fluorescence microscopy is a powerful tool for visualizing subcellular structures and their dynamic behaviors^{1–4}. However, the substantially higher photon budget required for SR imposes critical limitations on live-cell applications, stemming primarily from cumulative phototoxicity, accelerated photobleaching and compromised temporal resolution. Among existing SR techniques, structured illumination microscopy (SIM)² stands out through superior photon efficiency combined with compatibility with diverse fluorescent probes, making it well suited to low-perturbation, long-term and multicolor SR imaging⁵. Nevertheless, SIM SR images often encounter reconstruction artifacts that ruin spatial resolution, particularly under

low signal-to-noise ratio (SNR) conditions. These artifacts manifest as structured noise patterns that often mislead interpretations of biological structures, thereby raising fundamental concerns about the reliability of the SIM method.

To complement SIM techniques, regularized algorithms were developed initially by leveraging domain-specific prior knowledge to distinguish between biological structures and stochastic noise^{6,7}, yet these exhibit a limited capacity for resolving fine-grained subcellular features. Recent advancements in deep-learning-based (DL)-SIM methods^{8–11} have demonstrated high reconstruction accuracy in a data-driven way, yet their practical implementation faces fundamental constraints: acquiring

A full list of affiliations appears at the end of the paper. ✉e-mail: Yanjun_Liu@fudan.edu.cn; shantan@hust.edu.cn; li-dong@tsinghua.edu.cn

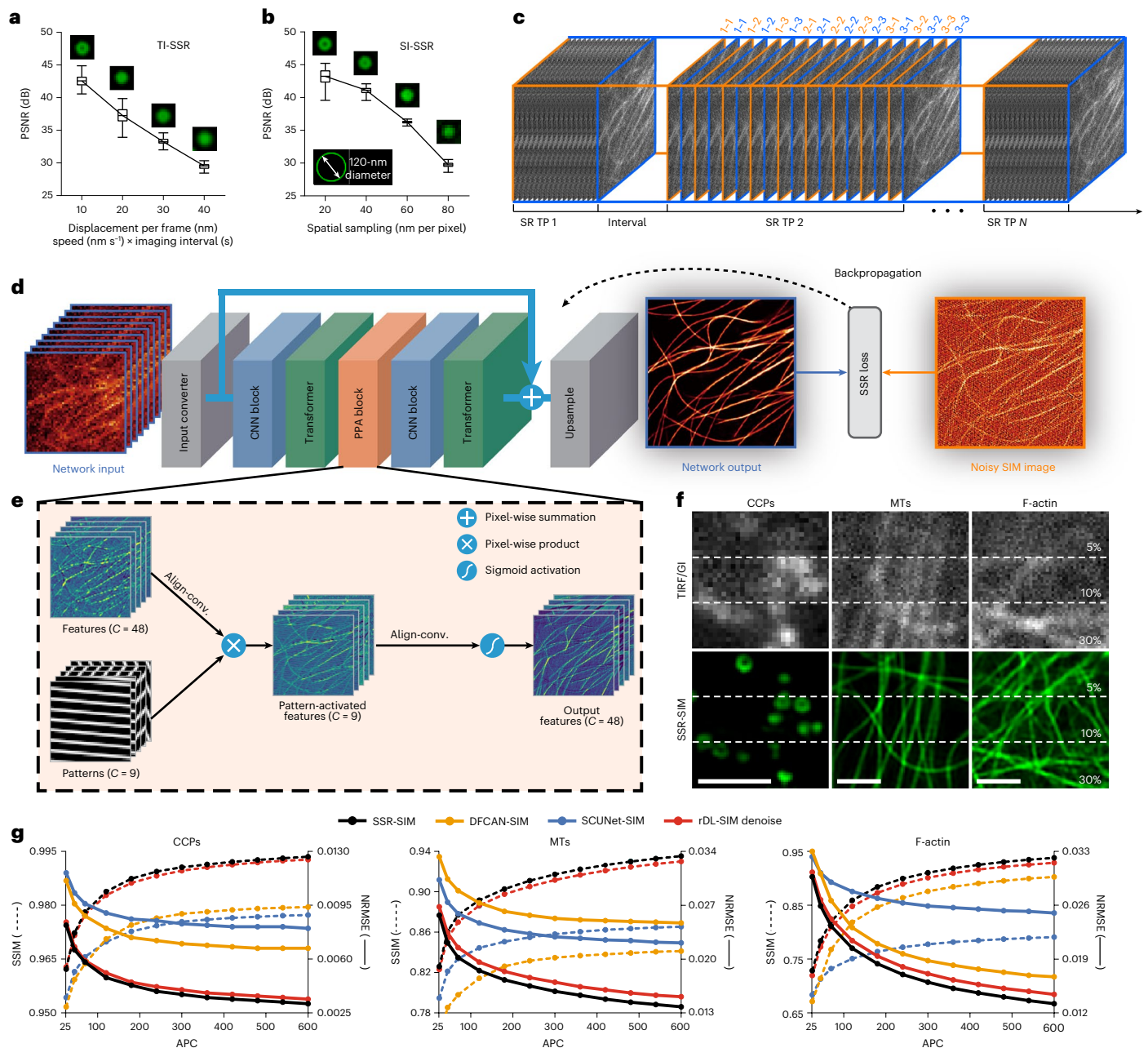


Fig. 1 | SSR-SIM method. a, b, Assessments of temporal interval (TI)-SSR (**a**) and spatial interval (SI)-SSR (**b**) strategies using synthetic hollow vesicles (diameter: 120 nm). Central line: medians; limits: 75% and 25%; whiskers: maximum and minimum. **c**, The SSR-SIM signal-consistent imaging sequence, halving each SR timepoint (TP) to yield training pairs with consistent and aligned biological structures. The orders of orientations and phases of the SIM patterns are indicated above. **d**, SSR-SIM training workflow and neural network architecture. **e**, The PPA block delivers the structured light pattern priors into activating

informative image features. Convolutional layers (Align-conv.) are used to adjust feature map dimensions. **f**, SSR-SIM reconstruction of CCPs, MTs and F-actin from the BioSR dataset. Experiments were repeated independently for five cells with similar results. Scale bars, 1 μm. **g**, SSIM (left axis, dashed line) and NRMSE (right axis, solid line) comparisons of learning-based SIM reconstruction ($n = 10$ per APC). The higher the SSIM or the lower the NRMSE, the higher the reconstruction precision. SCUNet-SIM, U-Nets through skip-layer connections for SIM⁶; rDL-SIM, rationalized DL-SIM.

reliable ground-truth (GT) data remains prohibitively challenging for light-sensitive and fast-evolving bioprocesses under diverse imaging scenarios (Supplementary Note 1). Although machine learning-SIM¹² attempts to circumvent this limitation by generating training sets from natural image datasets, its effectiveness is constrained by the inherent domain gap between natural scene statistics and biological microstructure characteristics. An alternative approach involves the incorporation of temporally or spatially interleaved self-supervised denoising frameworks^{13,14}, which depend critically on the theoretical assumption that noise components are temporally or spatially uncorrelated

whereas the biostructures maintain adequate consistency. However, our systematic analysis reveals that these prerequisites are violated in live-cell SR imaging due to the low temporal consistency of active microstructures (Extended Data Fig. 1a–e) and long-range spatial correlation of SIM artifacts (Supplementary Fig. 1). When self-supervised methods are deployed without addressing these intrinsic challenges, their integration into SIM reconstruction can produce biased outputs with degraded accuracy and resolution. Current end-to-end neural network architectures further exacerbate the degradation by neglecting the underlying physical principles, leading to incomplete high-frequency

component recovery. Collectively, these limitations highlight the need for a physics-informed SIM reconstruction framework capable of high-precision SR imaging without GT dependency.

In this study we present SSR-SIM to address these challenges. We first investigated factors influencing the performance of existing self-supervised image restoration algorithms upon their introduction into live-cell SR imaging. By applying theoretical constructs, we developed a signal-consistent imaging sequence coupled with an end-to-end SSR framework, enabling direct learning of high-quality reconstruction models from noisy raw SIM data without external GT datasets. Moreover, we designed a ‘physics-informed hybrid architecture integrating convolutional neural networks (CNNs) and transformers’ (PHCT) to perform high-precision SIM reconstruction. PHCT simultaneously captures local and long-range subcellular features through convolutional layers and self-attention mechanisms, while enhancing high-frequency components using a physics-based pattern activation (PPA) block rooted in harmonic SR theory. Validation on the BioSR⁹ dataset demonstrates that SSR-SIM outperforms nonlearning SIM denoising and deconvolution algorithms as well as supervised and unsupervised DL-SIM models. Extensive experiments further confirm SSR-SIM’s capability for artifact-free visualization with both high-spatiotemporal resolution and long-term imaging capability across diverse biological systems, including real-time monitoring of cytoskeletal and organelle remodeling, intracellular protein transport, intercellular communication and viral propagation.

Results

Principle of SSR

Unlike supervised methods that rely on GT data, self-supervised image restoration algorithms enable learning high-quality biological image representation from the degraded data themselves and have demonstrated efficacy in diffraction-limited microscopy denoising. These approaches train neural networks using spatially^{10,14,15} or temporally^{10,16} sampled imaging sequences, to preserve components consistent across sampling intervals (presumed to reflect latent biostructures) while eliminating stochastic components such as noise. To ensure effectiveness, three fundamental prerequisites must be met: (1) biological signal identity between source data and their sampled components, (2) biological signal identity between sampled pairs and (3) statistical noise and artifact independence between sampled pairs. However, SIM artifacts manifest as complicated visual patterns distinct from common noise in conventional diffraction-limited imaging (Supplementary Fig. 1), and matrix-vector analysis further revealed that the data reconstruction pipeline integrates noise from several acquisitions into artifacts with long-range spatial correlation (Supplementary Note 2). Consequently, spatially sampling reconstructed SIM images yields sample pairs where noise and artifacts lack statistical independence, which violates prerequisite (3). Second, the imaging intervals and camera sampling rate are in practice configured to maximize total imaging time and field of view (FOV). However, when tuning these parameters from ideal values to practical values (where interframe displacement occurs and sampling rate approaches the Nyquist rate; Supplementary Note 3), reconstruction precision and the resolving proportion of hollow vesicles deteriorated substantially (Fig. 1a,b and Extended Data Fig. 1f–k), due to increased signal difference violating prerequisites (1) and (2).

To solve the issues, we propose a signal-consistent imaging sequence (Fig. 1c) that yields two groups of raw images, $d_1(x)$ and $d_2(x)$, with identical excitation patterns and biological signals while ensuring independent occurrences of noise. Unlike image pre- or postdenoising (Supplementary Note 4), SSR-SIM directly executes SR reconstruction. To this end, we construct the loss function of SSR-SIM by introducing the linear reconstruction operator $\mathcal{R}$ (built on conventional SIM reconstruction implementation¹⁷) as the explicit prompt. The loss function is formulated as $\|\mathcal{R}[d_2(x)] - \Gamma_\Theta[d_1(x)]\|_2^2$ where Γ_Θ is a neural network with learnable parameters Θ (Supplementary Fig. 2), and we show the

learning scheme of SSR-SIM (Fig. 1d). By sampling using hardware adaptations rather than image postprocessing, SSR-SIM guarantees biological signal identity and noise and artifact independence between sampled pairs, enabling effective removal of spatially correlated artifacts without compromising reconstruction precision. We proved that optimized SSR-SIM models have the same theoretical reconstruction accuracy as GT-supervised models and validated this by training PHCT models in a supervised and self-supervised manner (Supplementary Notes 5 and 6). Throughout the optimizations, the trends of validation indices are almost identical (Supplementary Fig. 3) and the structural details and cut-off frequency of SR images inferred by the two models are identical (Extended Data Fig. 2). SSR-SIM model training and inference can be conducted on a single imaging dataset, yet its implementation may encounter an additional challenge: how to train robust SIM models using a small-scale training dataset. Regarding this, besides conventional data augmentation such as random cropping, we proposed two ad hoc data augmentation methods for SIM, shuffling training data that possess the same illumination patterns and using arbitrary-angle rotation before random cropping (Supplementary Fig. 4). The proposed data augmentation method reduces training data requirements from 50–100 pairs of raw images to as few as five pairs (Supplementary Fig. 5), facilitating zero-shot and few-shot reconstruction of a single time-lapse image sequence or small-scale image datasets.

Development and validation of PHCT architecture

Unlike wide-field microscopy, which uses uniform illumination, SIM illuminates the samples using dense periodic patterns to yield frequency-space modulation of SR components. In high numerical aperture (NA) SIM systems, the excitation patterns become too dense to be resolved in raw images (Extended Data Fig. 3a). For instance, our homebuilt SIM system operates with an excitation NA of up to 1.42, close to the objective NA of 1.49, resulting in line spacing of 0.172 μm (488-nm excitation), which is below the Rayleigh resolution limit of 0.215 μm . Furthermore, the line spacing, angles and initial phases of the sinusoidal illumination patterns vary over time due to mechanical and environmental instability, as well as temperature fluctuations, complicating data reconstruction. As the excitation NA increases, the reconstruction accuracy of DL-SIM models declines sharply, characterized by misleading structures and a reduction in peak SNR (PSNR) indexes (Extended Data Fig. 3b). This degradation is probably because commonly used CNN architectures struggle to detect the increasingly dense excitation patterns. Despite their effectiveness at handling processes in accord with translation invariance¹⁸, CNNs are limited in extracting SR information from modulated Moiré fringes with dense and translation-sensitive patterns. Furthermore, CNN-based image restoration architectures possess a limited receptive field, typically smaller than 64×64 pixels (about $4 \times 4 \mu\text{m}^2$ in our SIM system), which is insufficient for capturing nonlocal, long-range pixel dependencies of biostructures in large-FOV systems (for example, up to $1,536 \times 1,536$ pixels and $94 \times 94 \mu\text{m}^2$).

To address these limitations, we developed a PHCT architecture that incorporates physical information of illumination patterns as priors into a hybrid CNN and transformer framework (Fig. 1d and Supplementary Fig. 2). PHCT begins with a residual local feature block (RLFB)¹⁹ to extract shallow features from diffraction-limited raw data. These features are then delivered into an efficient transformer block (ETB)²⁰ to capture long-range pixel dependencies within contextual features. Unlike conventional vision transformers²¹ requiring quadratic computational complexity, ETB has linear computational complexity by using an additional convolution-assisted channel integration and is advantageous for efficient large-FOV image reconstruction. The effectiveness of RLFB and ETB was validated quantitatively through ablation studies (Supplementary Fig. 6). To incorporate physical knowledge of excitation patterns and structured light modulation explicitly, we designed a PPA block (Fig. 1e) that enriches the intermediate features with structurally coded frequency information in a $2\times$ upsampled

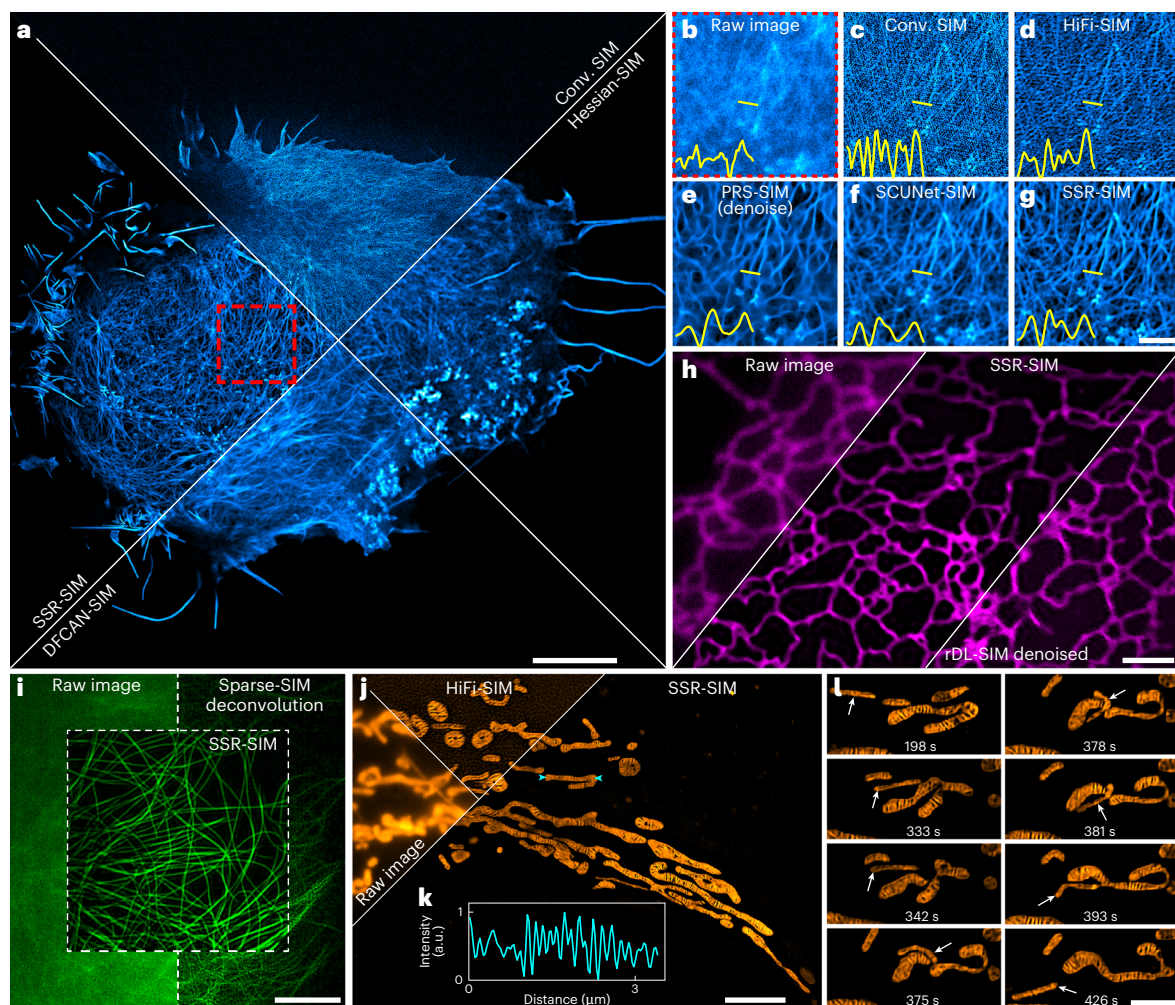


Fig. 2 | SSR-SIM imaging of cytoskeletons and organelles. **a**, A snapshot from time-lapse SSR-TIRF-SIM images of F-actin in COS-7 cells expressing LifeAct-mEmerald over 4.5 h, compared with conventional SIM (Conv. SIM), Hessian-SIM and DFCAN-SIM. **b–g**, Magnified views of the red-boxed regions in **a**, including the raw image (**b**), Conv. SIM (**c**), high-fidelity SIM (HiFi-SIM) (**d**), pixel-realignment-based self-supervised denoising for SIM (PRS-SIM)¹⁵ (**e**), SCUNet-SIM (**f**) and SSR-SIM (**g**). **h**, SSR-GI-SIM imaging of ER (Calnexin-Halo), compared with rDL-SIM. **i**, SSR-GI-SIM imaging of microtubules (Enscosin-mEmerald)

at low-SNR conditions, compared with Sparse-SIM deconvolution. **j**, SSR-GI-SIM imaging of mitochondrial cristae in PKMO-stained COS-7 cells, compared with HiFi-SIM. **k**, Intensity profile along the lines indicated by cyan arrowheads in **j**. **l**, Time-lapse images showing mitochondrial dynamics (white arrow). All experiments were repeated independently for more than three cells with similar results (**a–j**). Gamma, 0.5 (**a–g**) and 0.7 (**h–l**). Scale bars, 10 μm (**a**), 2 μm (**g,h,i**) and 5 μm (**j**).

feature space. Ablation studies on the PPA, along with experiments on simulated and biological images (Extended Data Figs. 3 and 4), demonstrated that PPA is effective and essential for high-NA SIM imaging. Finally, a RLFB-ETB cascade refines the output from the PPA block. The ultimate features are upsampled using a pixel-shuffle layer and convolved to generate the SR image in upsampled space.

We evaluated the overall performance of the PHCT-powered SSR-SIM on the BioSR⁹ dataset. In addition to the conventional SIM, competitive algorithms with different data processing pipelines (reconstruction, predenoising, postdenoising and postdeconvolution) and different learning manners (nonlearning, supervised learning and unsupervised learning) are involved in the comparison experiments, and we provide a brief introduction to these algorithms (Supplementary Note 7). Statistical metrics such as PSNR, structured similarity index (SSIM) and normalized root mean square error (NRMSE), demonstrated that the proposed SSR-SIM quantitatively outperforms other state-of-the-art (SOTA) SIM methods across varying fluorescence conditions (Fig. 1g and Supplementary Tables 1–3), with average photon counts (APCs) of raw data ranging from 25 to 600. Representative patches illustrate that SSR-SIM results effectively preserve recognizable,

high-resolution biological microstructures and closely resemble the GT (Extended Data Fig. 5). By integrating Monte Carlo dropout (MCD) layers²², SSR-SIM-MCD further provides self-supervised model uncertainty characterization, which is helpful for determining the applicability of pretrained models (Extended Data Fig. 6). PHCT has only 0.33 million (total internal reflection fluorescence (TIRF)/grazing incidence (GI)-SIM) and 0.80 million parameters (three-dimensional (3D)-SIM), making it lightweight and efficient. PHCT reconstruction of $512 \times 512 \times 9$ raw images (TIRF/GI-SIM) and $512 \times 512 \times 15 \times 8$ raw volumes (3D-SIM) takes about 17 ms and 194 ms, respectively, achieving SOTA reconstruction speeds (Supplementary Fig. 7). Altogether, these results affirm the feasibility and effectiveness of the SSR-SIM framework and highlight the architectural superiority of PHCT over other SOTA SIM methods.

SSR-SIM enables high-fidelity imaging of diverse cytoskeletons and organelles

The cytoskeleton and organelles constantly undergo dynamic remodeling while executing and coordinating various cellular functions. Filamentous actin (F-actin), endoplasmic reticulum (ER), microtubules (MTs) and mitochondrial cristae exhibit intricate subcellular structures

that require SR microscopy for precise visualization. However, high fluorescence doses used in conventional SR imaging often damage the biological structures and affect their dynamic behaviors. DL methods can reduce the photon budgets by improving low-SNR reconstruction, yet they necessitate clean SR images as GT. Here we assessed the capability of SSR-SIM on these samples without collecting extra training data. First, we transfected COS-7 cells with LifeAct-mEmerald and imaged F-actin structures over 4 h and 15 mins with the signal-consistent imaging sequence (Fig. 2a and Supplementary Video 1) and TIRF-SIM illumination. Actin dynamics exhibit active remodeling throughout imaging cycles, quantified by interframe alterations in filament similarities and segmentation masks (Supplementary Fig. 8). We found that SSR-TIRF-SIM resolves dense F-actin, whereas other SOTA SIM denoising and reconstruction methods resulted in artifact-contaminated and resolution-compromised F-actin structures (Fig. 2b–g and Supplementary Figs. 9a, 10 and 11). With high spatiotemporal resolution and extended imaging period, we further observed the formation, development and dissolution of a nanoscale ring of F-actin in cell anchoring and migration (Supplementary Fig. 9b), monitored the cytoskeleton remodeling in cell adhesion (Supplementary Fig. 12 and Supplementary Video 2), and captured structural changes in the cytoskeleton through rapid imaging of 10,000 frames (Supplementary Video 3).

When combined with GI illumination²³, SSR-GI-SIM can observe ER structures (Calnexin-Halo, COS-7 cells) while minimizing motion artifacts (Fig. 2h and Supplementary Fig. 13). Furthermore, SSR-GI-SIM enables clear identification of microtubule skeleton (Enscosin-mEmerald, COS-7 cells) under extremely low light illumination (APCs: 33.3 ± 4.9 (mean $\pm$ s.d.), $n = 9,000$), whereas other methods fail to resolve complete tubules (Fig. 2i, Supplementary Fig. 14 and Supplementary Video 4). Mitochondrial cristae are densely stacked within mitochondria, resulting in small cristae-to-cristae distances. For instance, recent stimulated emission depletion microscopy studies²⁴ reported an average distance of 90 ± 24 nm (mean $\pm$ s.d.) between adjacent mitochondrial cristae in COS-7 cells. Mitochondria are particularly susceptible to phototoxicity, and such damage can alter their membrane permeability, leading to structural deformation characterized by spherical swelling and matrix dilution, and ultimately causing cell apoptosis. In the context of conventional DL-SIM, obtaining GT data for mitochondrial components is challenging due to their sustained and rapid remodeling. Nonetheless, the SSR-GI-SIM effectively learns to reconstruct noisy imaging data autonomously, enabling clear visualization of mitochondrial cristae and their remodeling events, including penetration, fission, fusion and rotation (Fig. 2j–l and Supplementary Fig. 15). SSR-GI-SIM facilitates continuous recording of mitochondrial cristae over 2,000 timepoints, with high spatiotemporal resolution and resolving internal cristae junctions effectively (Supplementary Video 5). Furthermore, we evaluated the capacity of SSR-SIM to maintain cell viability in live-cell SR imaging by monitoring Ca^{2+} levels using FluoForte dye, because large changes in Ca^{2+} levels often correspond to cellular stress and the possible triggering of apoptosis or death^{25,26}. In negative controls, imaging was restricted to FluoForte (488 nm) with the ER channel (561 nm) disabled. For the others, the 561-nm laser power was adjusted to ensure clear visualization of ER mesh and filamentous structures. FluoForte fluorescence analyses (Extended Data Fig. 7a–c) show that Ca^{2+} levels under SSR-GI-SIM irradiation conditions exhibited minor fluctuations comparable to negative controls. In contrast, conventional GI-SIM irradiation frequently induced sudden Ca^{2+} level changes, indicating that cells are in stressed states. We further demonstrate that SSR-GI-SIM irradiation maintained stable ER fluorescence without observed photobleaching, indicating sustained long-term SR imaging (Extended Data Fig. 7d–j).

Multicolor SSR-SIM imaging of rapid subcellular interactions

In mammalian cells, ER forms a complex network spanning the entire cell, which is crucial for protein modification and translocation. Synthesized proteins are directed to various destinations for further

modification. The fluorescently labeled vesicular stomatitis virus glycoprotein (VSVG) serves as a valuable tool for tracking the bioprocesses of individual transporters²⁷. However, VSVG possesses a high moving speed, the ER undergoes constant vibration and chemical fixation can fragment ER structures, together making it challenging to collect GT data for conventional DL-SIM algorithms. One compromise⁹ is to use a Gaussian low-pass filter on SR images of fast-moving biostructures to smooth moving artifacts to improve supervised data, but at the cost of reduced spatial resolution. Conversely, our proposed SSR-SIM removes the need to obtain GT images and allows for high spatiotemporal resolution imaging of the colocalization and proximity and interactions between VSVG and ER (Calnexin-Halo) (Fig. 3a and Supplementary Video 6). The results obtained by SSR-GI-SIM showed that VSVGs were enveloped by the connection points of ER networks (Fig. 3b), the ER rings (Fig. 3c) and the fenestrations of ER sheets (Fig. 3d) suspended between the centers of two ER tubules (Fig. 3e), and attached to a single isolated ER tubule (Fig. 3f, over 80.75 s). We noted that the rapid movement of VSVG can lead to the remodeling of ER networks (Fig. 3g) and even the fission processes of ER sheets (Fig. 3h,i and Supplementary Video 7). Furthermore, the paradigm of ER hitchhiking, initially discovered in fungi²⁸ and observed recently in mammalian cells, involves the participation of ER in the vesicle movement of lysosomes²³ and peroxisomes¹⁰. Here we observed ER hitchhiking during interactions between ER and VSVG (Extended Data Fig. 8a,b), suggesting that ER hitchhiking is probably a widespread mechanism in protein translocation. We also observed that ER forms a ring-like structure to assist in the movement of VSVG, and the ER ring was depolymerized following the hitchhiking process (Extended Data Fig. 8c–e).

SR imaging of endocytic recycling, involving Rab11a and transferrin receptor (TfR) complexes²⁹, poses challenges due to their nanoscale structures, rapid movement and quick photobleaching. With SSR-GI-SIM, we were able to record continuously more than 1,500 SR timepoints of endocytic recycling by reducing the fluorescence dose (Supplementary Video 8). Specifically, we introduced a TfR1-mCherry fluorescent label in SUM159 cells stably expressing Rab11a-GFP and performed SR imaging within approximately 1 μm from the slide (near the bottom plasma membrane) by using a GI light sheet (1.37 excitation NA) at 0.2-s intervals. We show that SSR-GI-SIM enables artifact-free reconstruction of Rab11a and TfR1 raw data that are too noisy to be reconstructed by other SIM methods (Fig. 3j). Compared with SOTA deconvolution methods (for example, sparse deconvolution SIM (Sparse-SIM)), SSR-GI-SIM prevents over-smoothing and maintains reconstruction accuracy of biostructures in low-intensity areas riddled with artifacts (Supplementary Fig. 16). We observed two distinct morphologies of the Rab11a and TfR1 complex: tubular recycling cargo and disk-shaped structures (Fig. 3j,k). During rapid transport, Rab11a and TfR1 displayed a tubular morphology, with lengths ranging from 0.4 μm to 2.4 μm . Upon reaching the plasma membrane, the complexes exhibited solid disk or hollow ring-like structures. Further, we observed that the fluorescence signal of TfR1-mCherry disappeared before that of Rab11a-GFP in the final stage of internal recycling (Fig. 3l,m). We observed that the disappearance of TfR1-mCherry and Rab11a-GFP signals coincided with the occurrence of out-of-focus fluorescence (Fig. 3n), indicating the recycling of TfR1 and the cytoplasmic release of Rab11a. Taken together, these results demonstrated that SSR-SIM allows for high-quality, high-spatiotemporal resolution imaging of rapid subcellular interactions in light-sensitive samples.

SSR-SIM investigation of intercellular dynamics and VACV-induced actin comet tail

Vaccinia virus (VACV)—a close relative of the monkeypox virus³⁰ and an invaluable model for studying virus–host interactions—uses actin-based fast motility (VACV-induced actin tail) to enhance viral dissemination³¹. However, SR imaging of VACV infection is challenging due to the potential inactivation of the virus by strong lasers. As a result,

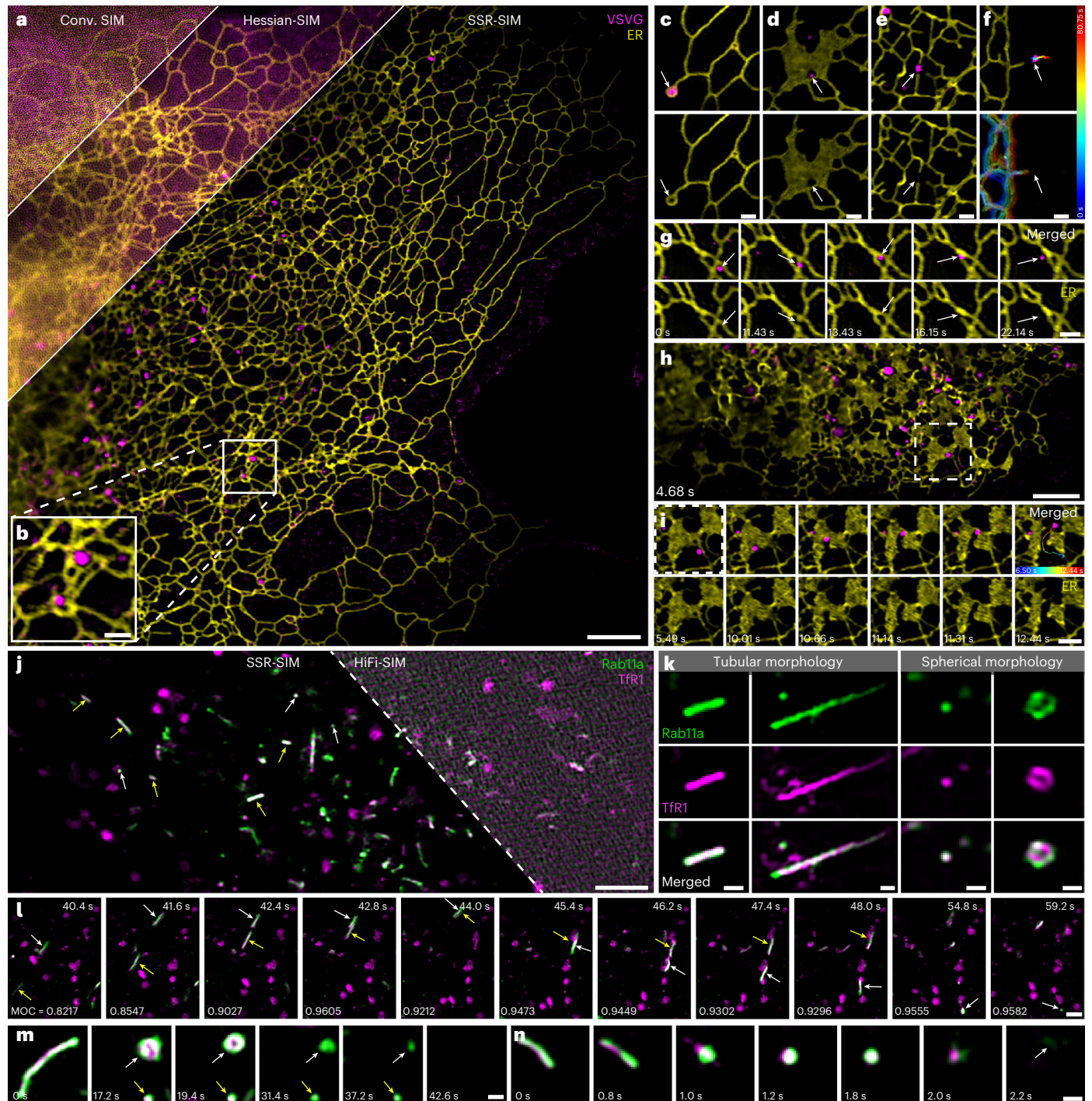


Fig. 3 | SSR-SIM imaging of ER–VSVG interactions and recycling endosome dynamics. **a**, A representative SSR-GI-SIM image of a COS-7 cell expressing Calnexin-Halo (yellow) and VSVG-mEmerald (magenta), compared with Conv. SIM and Hessian-SIM. Experiments were repeated independently for seven cells with similar results. **b**, An enlarged region from **a** highlighting the spatial relationship between VSVG and ER. **c–e**, Representative regions of interest (ROIs) showing VSVG in proximity to distinct ER morphologies: ER rings (**c**), ER sheets (**d**) and ER tubules (**e**). **f**, A representative ROI showing the interaction between VSVG and an isolated ER tubule, with color-coded motion trajectories. **g**, Time-lapse images showing ER fission and fusion events. **h**, SSR-GI-SIM imaging of ER and VSVG. **i**, Time-lapse images of the white-boxed regions in **h** showing ER fission when VSVG traverses ER sheets. The motion trajectory of

VSVG is plotted in the right image. Experiments shown in **b–f** and **h** were repeated independently for five ROIs with similar results. **j**, SSR-GI-SIM imaging of SUM159 cells with Rab11a-GFP knocked-in and expressing TfR1-mCherry, compared with HiFi-SIM. Yellow and white arrows point to example structures with different morphologies. Experiments were repeated independently for four cells with similar results. **k**, Representative images of Rab11a and TfR1 with different morphologies. **l**, Representative time-lapse images of tubular structures. MOC, Mander's overlap coefficient. **m**, Time-lapse images showing the TfR1-mCherry signal dissipating prior to Rab11a-GFP. **n**, Loss of TfR1 and Rab11a signals, accompanied by out-of-focus fluorescence. Gamma, 0.7 (**a–i**). Scale bars, 5 μm (**a,h**), 1 μm (**b–g,l**), 2 μm (**i,j**) and 300 nm (**k,m,n**).

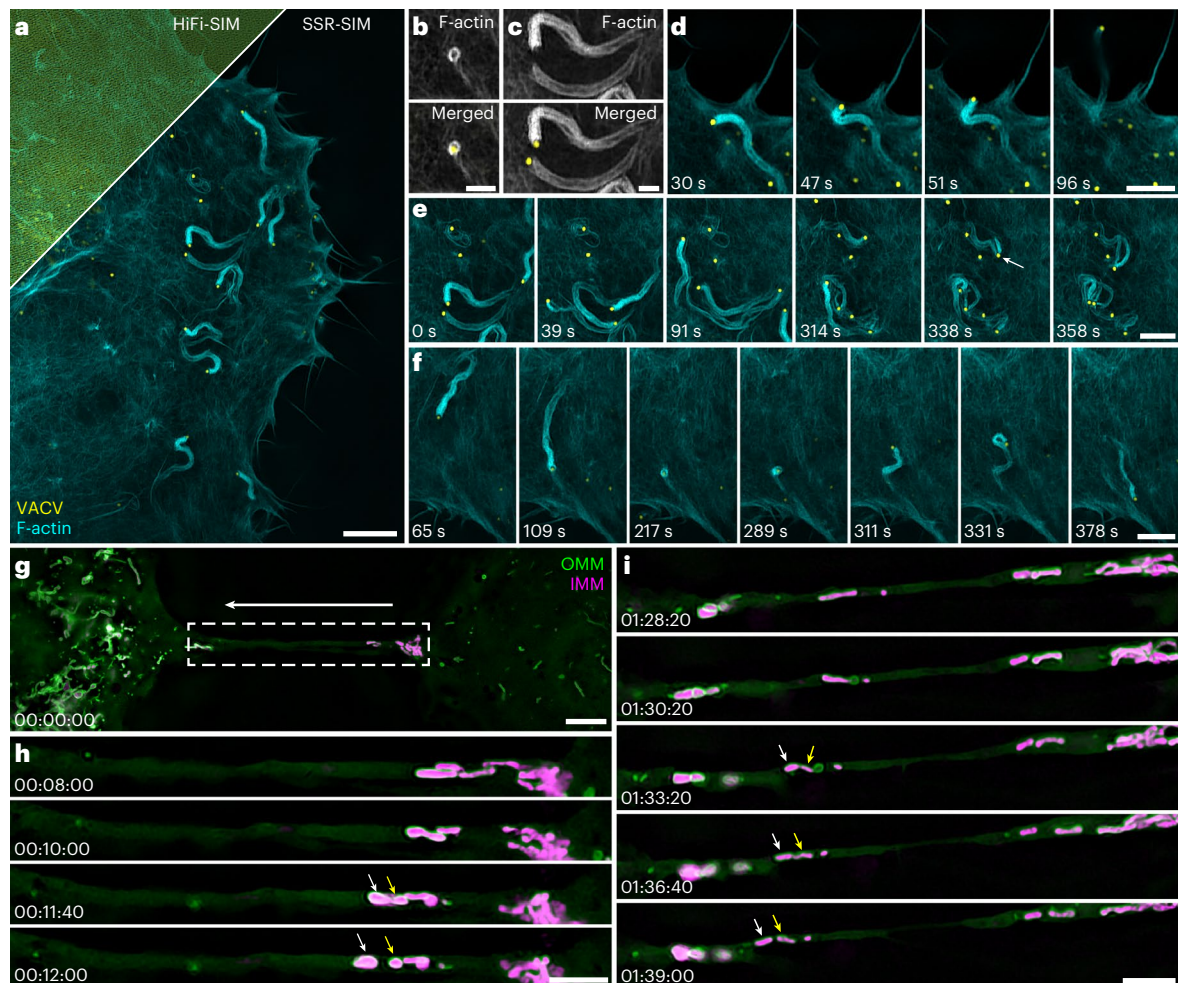


Fig. 4 | SSR-SIM imaging of actin structures in VACV induction and mitochondrial events in intercellular nanotubes. **a**, SSR-TIRF-SIM imaging of actin comet tails in COS-7 cells expressing LifeAct-mEmerald (cyan) 24 h after the induction of VACV expressing A4-mCherry (yellow), compared with HiFi-SIM. **b**, Actin cloud surrounding a static VACV particle. Experiments were repeated independently for eight cells with similar results. **c**, Hollow actin comet tails during viral particle movement. **d**, Viral particles movement out of the cell membrane and imaging plane. **e**, A single viral particle associated with several

actin comet tails (white arrow). **f**, Morphological changes of actin clouds and actin comet tails during viral particle movement. **g**, Time-lapse SSR-GI-SIM images of mitochondrial transfer through TNT across COS-7 cells expressing TOMM20-mEmerald (green) and stained by PKMO (magenta) over 3 h and 12 min. Experiments were repeated independently for three cells with similar results. **h,i**, Mitochondrial fission events at 12 min (**h**) and 1.5 h (**i**) after mitochondria enter the tunneling nanotube. Gamma, 0.7 (**a–f**) and 0.6 (**g–i**). Scale bars, 5 μ m (**a,g**), 1 μ m (**b,c**) and 3 μ m (**d–f,h,i**).

currently published results by SIM suffer from excessive artifacts and insufficient resolution enhancement^{31,32}, and manifested VACV-induced actin comet tails in solid and single-tail structures, just like those results obtained by diffraction-limited microscopy. Unexpectedly, through SSR-TIRF-SIM imaging, we observed that the actin comet tails in COS-7 cells stably expressing LifeAct-mEmerald and A4-mCherry (a viral core protein) appeared hollow, with several tails trailing behind the viral particles (Fig. 4a–c and Supplementary Video 9). This was further validated in Vero cells (Extended Data Fig. 9). The previous misunderstanding of the VACV-induced actin comet tail structures is probably owing to the degenerated SR imaging precision, as we observed with other SIM methods. After initiating the actin comet tails and commencing movement, the viral particles may pause occasionally, only to resume their movement after several seconds. We also observed virus-actin circling before and after viral anchoring (Extended Data Fig. 9b,c), which might be owing to the instability caused by actin fast polymerization and depolymerization. Furthermore, SSR-TIRF-SIM recorded that, in the extracellular environment, both the virus and actin comet tails became unstable and moved out of the imaging focal plane, possibly due to the inability to sustain actin recruitment (Fig. 4d). SSR-TIRF-SIM

also captured that a single viral particle could induce the generation of more than one cluster of actin comet tails (Fig. 4e). Furthermore, SSR-TIRF-SIM recorded the consecutive event that viral particles anchored near the cell membrane, recruiting actin to form actin clouds, inducing the formation of hollow actin comet tails and surfing along the cell membrane (Fig. 4f and Extended Data Fig. 9d,e).

Tunneling nanotubes (TNTs)—a recently uncovered mode of cell-to-cell communication—enable the selective exchange of organelles between cells, playing a pivotal role in maintaining cellular homeostasis during both physiological functions and pathological conditions^{33,34}. Achieving a comprehensive and detailed record of mitochondrial dynamics within TNTs demands continuous SR imaging lasting for hours with large imaging FOV. However, excessive laser excitation poses the risk of inducing vibrations and potentially leading to ruptures within the TNTs themselves. SSR-GI-SIM demonstrates the potential for offering prolonged SR monitoring of TNTs during invasive processes. In our study, we used SSR-GI-SIM imaging to visualize the mitochondria (TOMM20-mEmerald and PKmito ORANGE (PKMO)) in COS-7 cells, capturing images every 20 s for more than 3 h (Fig. 4g, Supplementary Fig. 17a,b and Supplementary Video 10).

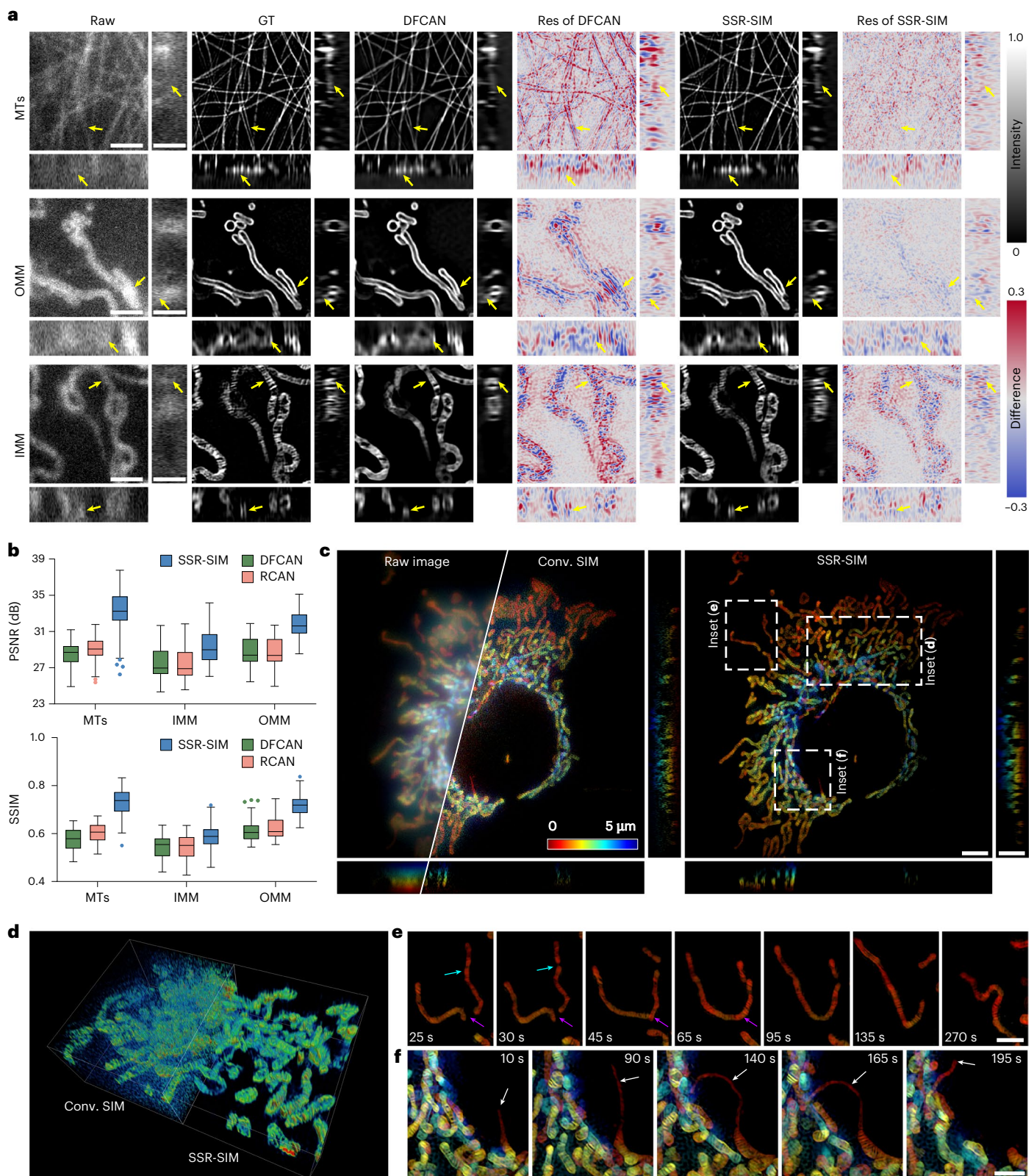


Fig. 5 | SSR-SIM imaging of 3D cellular structures. a, Orthogonal views of raw and SR reconstructed images showing MTs, OMM and IMM, compared with other 3D-SIM methods (representative structures indicated by yellow arrows). The residual (Res) between DFCAN/SSR-SIM and GT is provided for comparison. **b**, Statistical PSNR and SSIM metrics for different reconstruction methods ($n = 72$). Central line: medians; limits: 75% and 25%; whiskers: maximum and

minimum. **c**, Depth-color-coded orthogonal views of raw, Conv. 3D-SIM and SSR-3D-SIM volume of IMM in PKMO-stained COS-7 cells. **d**, Volume rendering comparing conventional SIM. **e**, Time-lapse SSR-3D-SIM images of mitochondrial fusion and fission events. **f**, Time-lapse SSR-3D-SIM images of mitochondrial dynamic tubulation. Gamma, 0.6 (**c**, **e**, **f**). Scale bars, 2 μm (**a**, **e**, **f**) and 4 μm (**c**).

When comparing SSR-GI-SIM with other SIM methods, we observed that SSR-GI-SIM consistently identified the mitochondrial cristae, even when mitochondria were compressed by the nanotubes. In contrast, other SOTA SIM methods failed to detect these cristae structures (Supplementary Fig. 17c–g). Notably, the high-quality SSR-GI-SIM images reveal that mitochondrial fission events occur during transport through the TNT, approximately 12 min and 93 min after mitochondria exited the original cell and passed through the TNT (Fig. 4h,i), mirroring the fission dynamics observed within cells. Collectively, these results uncovered unprecedented structures and dynamics in virus–host interactions and in TNTs, indicating that SSR-SIM is likely to serve as a complement to techniques in the field of virology and intercellular communication.

SSR-3D-SIM visualization of mitochondrial dynamics

3D-SIM offers a twofold resolution improvement across all spatial dimensions and enables optical sectioning. Although it facilitates time-lapse volume SR imaging, 3D-SIM is applied predominantly to fixed samples due to limitations in imaging speed and risks of excessive photodamage. For instance, imaging a volume with a height of 5 μm requires approximately 38 axial scans and 570 raw images, resulting in an imaging cycle exceeding 6.8 s (comprising 10 ms for each exposure and 2 ms for each readout), longer than the roughly 0.1 s needed for a TIRF and GI-SIM snapshot. Furthermore, the photobleaching and phototoxicity from 570 instances of wide-field illumination are greater than those from nine instances of TIRF and GI illumination. Consequently, 3D-SIM imaging of dynamic or light-sensitive samples presents considerable challenges. By reducing each exposure we can shorten the imaging cycle and decrease photodamage, but it causes noisy raw data that are hard to reconstruct. Here we highlight the applicability of SSR-SIM to improve 3D-SIM imaging. We constructed 3D-PHCT by replacing the two-dimensional (2D) convolutional layers in PHCT with 3D convolutional layers to extract and integrate axial SR information. To further reduce the computational burden caused by the 3D convolution, we streamlined the network architecture and used only half the 3D-RLFBs and 3D-ETBs. We evaluated 3D-PHCT-powered SSR-SIM on a benchmark dataset containing three types of biological structure: MTs, the outer mitochondrial membrane (OMM) and the inner mitochondrial membrane (IMM), compared with 3D-RCAN¹¹ and 3D deep Fourier channel attention network (3D-DFCAN)⁹. Representative image patches (Fig. 5a) and PSNR and SSIM metrics (Fig. 5b) showcased the visual and statistical superiority of SSR-SIM. Next, we conducted SSR-3D-SIM imaging of mitochondrial cristae with an axial scanning depth of 5 μm , comparing it with conventional 3D-SIM through depth-color-coded orthogonal views (Fig. 5c) and volume rendering (Fig. 5d). SSR-3D-SIM yielded high-fidelity SR volumes with high spatiotemporal resolution, enabling clear identification of mitochondrial cristae and mitochondrial network remodeling events (Fig. 5e,f). Furthermore, with ultra-low excitation, SSR-3D-SIM facilitates the recording of more than 300 3D-SR volumes of mitochondrial dynamics (Supplementary Video 11).

Discussion

Advances in DL have enhanced the SR precision and image quality when GT data are available. In this study, we developed an SSR framework and a physics-informed hybrid reconstruction model for SIM to address challenges related to data availability and reconstruction fidelity. Our approach, termed SSR-SIM, eliminates the dependency on additional training datasets, achieves higher-precision reconstruction visually and quantitatively, and extends SIM's advantages of imaging speed, minimal photodamage and fluorescence compatibility to a wide range of samples. Nevertheless, the SSR-SIM method still has limitations. As an image reconstruction algorithm rather than an image generation algorithm, SSR-SIM imaging requires a minimum number of photons to ensure reconstruction feasibility. To overcome spatially correlated reconstruction artifacts, SSR-SIM uses repeated sampling in acquiring

training data, necessitating double exposure times. Despite this constraint, the SSR-SIM system supports exposure times as short as 0.1 ms, which remains sufficient for the vast majority of bioprocesses. Using SSR-SIM, we have discovered several unusual biological phenomena. For instance, VSVG induces tearing deformations in ER sheets during the movement of VSVG, VACV hijacks actin to facilitate its propagation and mitochondria in TNTs can still undergo fission. These findings underscore the broader applicability of SSR-SIM in biological research. Moreover, SSR-SIM possesses remarkable generalizability to other microscope techniques: (1) the concept of the SSR framework applies to uniformly illuminated microscopy deconvolution tasks (for example, wide-field, confocal and light-sheet microscopy). Besides the sinusoidal structured excitation used in standard SIM systems, the concept of SSR can also be extended to generalized systems, such as multifocal SIM³⁵, lattice light-sheet microscopy³⁶ and imaging scanning microscopy (ISM)³⁷, with the theoretical constructs unchanged; (2) enabling training and inference on the same imaging dataset eliminates the generalization errors since the training data are also the test data, making SSR-SIM user-friendly and accessible without requiring users to have DL expertise and (3) the PPA block in PHCT illustrates how prior knowledge of structured light modulation can be integrated into end-to-end reconstruction networks. Combining PPA with the latest advances in the field of computer vision, researchers can continually construct a series of SOTA end-to-end SIM reconstruction networks to improve SIM techniques. In conclusion, SSR-SIM holds potential for broadly uncovering biological phenomena, owing to its superior spatiotemporal resolution, enhanced SR precision, comprehensive applicability, prolonged imaging capability and modular expandability.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-025-02966-y>.

References

1. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
2. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, 82–87 (2000).
3. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
4. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006).
5. Li, D. et al. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).
6. Chu, K. et al. Image reconstruction for structured-illumination microscopy with low signal level. *Opt. Express* **22**, 8687–8702 (2014).
7. Huang, X. et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nat. Biotechnol.* **36**, 451–459 (2018).
8. Jin, L. et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nat. Commun.* **11**, 1934 (2020).
9. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **18**, 194–202 (2021).
10. Qiao, C. et al. Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nat. Biotechnol.* **41**, 367–377 (2023).

11. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat. Methods* **18**, 678–687 (2021).
12. Ward, E. N. et al. Machine learning assisted interferometric structured illumination microscopy for dynamic biological imaging. *Nat. Commun.* **13**, 7836 (2022).
13. Lehtinen, J. et al. Noise2Noise: learning image restoration without clean data. In *Proc. International Conference on Machine Learning* (eds Dy, J. et al.) 2965–2974 (PMLR, 2018).
14. Huang, T., Li, S., Jia, X., Lu, H. & Liu, J. Neighbor2neighbor: self-supervised denoising from single noisy images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (eds Masson-Forsythe, M. et al.) 14776–14785 (IEEE, 2021).
15. Chen, X. et al. Self-supervised denoising for multimodal structured illumination microscopy enables long-term super-resolution live-cell imaging. *Photonix* **5**, 4 (2024).
16. Li, X. et al. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. *Nat. Methods* **18**, 1395–1400 (2021).
17. Gustafsson, M. G. et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical J.* **94**, 4957–4970 (2008).
18. McCann, M. T., Jin, K. H. & Unser, M. Convolutional neural networks for inverse problems in imaging: a review. *IEEE Signal Process. Mag.* **34**, 85–95 (2017).
19. Kong, F. et al. Residual local feature network for efficient super-resolution. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops* (eds Mortensen, E. et al.) 766–776 (IEEE, 2022).
20. Zamir, S. W. et al. Restormer: efficient transformer for high-resolution image restoration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (eds Mortensen, E. et al.) 5728–5739 (IEEE, 2022).
21. Dosovitskiy, A. et al. An image is worth 16×16 words: transformers for image recognition at scale. In *Proc. International Conference on Learning Representations* (eds Mohamed, S. et al.) 611–631 (ICLR, 2021).
22. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proc. International Conference on Machine Learning* (eds Balcan, M. F. et al.) 1050–1059 (PMLR, 2016).
23. Guo, Y. et al. Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales. *Cell* **175**, 1430–1442 (2018).
24. Liu, T. et al. Multi-color live-cell STED nanoscopy of mitochondria with a gentle inner membrane stain. *Proc. Natl Acad. Sci. USA* **119**, e2215799119 (2022).
25. Kilian, N. et al. Assessing photodamage in live-cell STED microscopy. *Nat. Methods* **15**, 755–756 (2018).
26. Orrenius, S., Zhivotovsky, B. & Nicotera, P. Regulation of cell death: the calcium–apoptosis link. *Nat. Rev. Mol. Cell Biol.* **4**, 552–565 (2003).
27. Presley, J. F. et al. ER-to-Golgi transport visualized in living cells. *Nature* **389**, 81–85 (1997).
28. Guimaraes, S. C. et al. Peroxisomes, lipid droplets, and endoplasmic reticulum ‘hitchhike’ on motile early endosomes. *J. Cell Biol.* **211**, 945–954 (2015).
29. Grant, B. D. & Donaldson, J. G. Pathways and mechanisms of endocytic recycling. *Nat. Rev. Mol. Cell Biol.* **10**, 597–608 (2009).
30. Lum, F.-M. et al. Monkeypox: disease epidemiology, host immunity and clinical interventions. *Nat. Rev. Immunol.* **22**, 597–613 (2022).
31. Humphries, A. C. et al. Clathrin potentiates vaccinia-induced actin polymerization to facilitate viral spread. *Cell Host Microbe* **12**, 346–359 (2012).
32. Horsington, J., Turnbull, L., Whitchurch, C. B. & Newsome, T. P. Sub-viral imaging of vaccinia virus using super-resolution microscopy. *J. Virol. Methods* **186**, 132–136 (2012).
33. Rustom, A., Saffrich, R., Markovic, I., Walther, P. & Gerdes, H.-H. Nanotubular highways for intercellular organelle. *Transp. Sci.* **303**, 1007–1010 (2004).
34. Pinto, G., Brou, C. & Zurzolo, C. Tunneling nanotubes: the fuel of tumor progression? *Trends Cancer* **6**, 874–888 (2020).
35. York, A. G. et al. Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy. *Nat. Methods* **9**, 749–754 (2012).
36. Chen, B.-C. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).
37. Müller, C. B. & Enderlein, J. Image scanning microscopy. *Phys. Rev. Lett.* **104**, 198101 (2010).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

¹Key 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, China. ²State Key Laboratory of Membrane Biology, Tsinghua-Peking Center for Life Sciences, Beijing Frontier Research Center for Biological Structure, IDG/McGovern Institute for Brain Research, New Cornerstone Science Laboratory, School of Life Sciences, Tsinghua University, Beijing, China. ³National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. ⁴Shanghai Key Laboratory of Medical Epigenetics, Institutes of Biomedical Sciences, Shanghai Xuhui Central Hospital, Zhongshan-Xuhui Hospital, Fudan University, Shanghai, China. ⁵Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China. ⁶These authors contributed equally: Jiahao Liu, Xue Dong, Huaide Lu, Tao Liu, Wei Liu.

✉ e-mail: Yanjun_Liu@fudan.edu.cn; shantan@hust.edu.cn; li-dong@tsinghua.edu.cn

Methods

Multi-SIM setup

The hardware part of SSR-SIM was built on our homemade Multi-SIM²³ based on an inverted fluorescence microscope (Ti2E, Nikon) equipped with a TIRF objective (CFI SR HP Apochromat TIRF 100XC Oil, 1.49 NA, Nikon). Four lasers of 405 nm (catalog number LBX-405-300-CSB, Oxixus), 488 nm (catalog number Sapphire 488-500 LPX, Coherent), 560 nm (catalog number Sapphire 488-500 LPX, Coherent), 640 nm (catalog number LPX-640L-500-CSB, Oxixus) were combined collinearly and then passed through an acoustic-optical tunable filter (catalog number AOTFnc-400.650-CPCH-Tn, AA Quanta Tech) to switch the desired laser and adjust the laser power and illumination duration. The lasers were expanded by using a 4 f optical system, and the expanded light was modulated by a system consisting of half-wave plates, a polarizing beam splitter and a spatial light modulator (SLM) (catalog number QXGA-3DM, Forth Dimension Display). The grating patterns on the SLM were generated with excitation NA ranging from 1.26 to 1.42, resulting in 50–200 nm light penetration depth (the distance of light entering the cover glass) in TIRF-SIM³⁸, 200 nm–5 μ m in GI-SIM²³ and more than 5 μ m in 3D-SIM¹⁷. In TIRF-SIM and GI-SIM, the grating patterns were repeated at three orientations with three phases, and in 3D-SIM the grating patterns were repeated at three orientations with five phases per slice. This yields nine (TIRF/GI-SIM) or 15 (3D-SIM) raw frames per SR timepoint. In this study, we constructed signal-consistent imaging sequences by repeating SLM grating patterns, and raw exposure twice and halving each illumination duration, to generate samples with consistent signals and differentiated noise. The grating patterns and their constructed sequences were preburned into SLM, and the running order of these sequences of grating patterns can be tuned dynamically by the upper computer. The diffracted light passed through a mask that permits the passage of ± 1 order beams in TIRF/GI-SIM cases, and 0 order and ± 2 order beams in 3D-SIM cases, respectively. The lights were then delivered into a homemade half-wave plates and quarter-wave plates for changing the polarization directions and focused on the back focal plane of the objective lens. The beams were interfered ultimately at the objective's focus plane to generate harmonic illumination. In the 3D-SIM setup, the conjugation of the SLM plane to the objective focus plane was adjusted carefully by using 100-nm fluorescence beads. The 0 order, ± 1 orders and ± 2 orders optical transfer functions were recorded for the reconstruction. The emission fluorescence was collected by the same objective and separated by a multiband dichroic mirror (catalog number ZT 405/488/560/640 tpc-22.5deg-UF3, Chroma), and finally detected by a scientific complementary metal oxide semiconductor camera (Orca Flash 4.0 v.3, Hamamatsu, or Kinetix, Photometrics). Before each imaging experiment, we imaged multicolor 100-nm fluorescence beads and constructed affine matrices according to the relative positions of beads in camera targets. The affine matrices are then used in digitized calibration of different wavelengths. To measure the camera background, we acquired and averaged >100 noise images in a dark field with the laser turn-off to obtain the camera background, which is subtracted from the raw data to remove fixed pattern noise.

Cell maintenance and sample preparation

COS-7 (ATCC, CRL-1651) and Vero (NCACC, SCSP-520) cells were grown in DMEM (Gibco, catalog number 11965092) supplemented with 10% FBS (Gibco, catalog number 10099141C) and 1% penicillin and streptomycin (Thermo Fisher, catalog number 15140122); SUM159 cells (from K. He) were grown in DMEM/Ham's F-12, (GlutaMAX) supplemented with 10% FBS, 1% penicillin-streptomycin, 5 μ g ml⁻¹ bovine insulin, 10 mM of HEPES (Corning) and 1 μ g ml⁻¹ hydrocortisone. All cell lines were cultured at 37 °C and 5% CO₂. Cells were seeded onto collagen-coated coverslips (50 mg ml⁻¹) and cultured until ~70% confluency. Transfections were conducted using Lipofectamine 3000 (Invitrogen, catalog number L3000150) following standard protocol. The plasmid constructs included LifeAct-mEmerald, LifeAct-mRuby,

Ensconsin-mEmerald, Clathrin-GFP, Calnexin-Halo, KDEL-mCherry, VSVG-mEmerald, Rab11a-GFP, TfR1-mCherry, TOMM20-mEmerald and TOMM20-mScarlet. In ER experiments, cells transfected with HaloTag plasmids were treated with warm DMEM containing 30 nM JF646 or JF549 (Promega) for 30 min (followed by medium replacement and incubation) before imaging. The cells used for recording IMM dynamics were treated with warm DMEM containing 250 nM PKMO or PKMDR²⁴ (PKMITO ORANGE, Standard Imaging) at 37 °C for approximately 10 min before imaging. In experiments inducing and recording actin comet tails, COS-7 cells were first infected with a retrovirus system to express LifeAct-mEmerald, and then infected with A4-mCherry-labeled VACV approximately 24 h before imaging; Vero cells were infected with a retrovirus system to express LifeAct-mRuby, infected with VACV and stained with Hoechst stain before imaging. Lentiviral particles were generated by transient transfection of HEK293T cells using Lipofectamine 3000. HEK293T cells were seeded in six-well plates and transfected with 1.5 μ g of psPAX2, 0.5 μ g of pMD2.G and 2 μ g of a lentiviral vector plasmid when they reached 50–70% confluency. The medium was replaced with warm DMEM containing 10% FBS 10 h post-transfection, and viral supernatant was harvested twice at 48- and 72-h post-transfection. For lentivector transductions, cells at a density of 1×10^5 cells ml⁻¹ in six-well plates were infected with 1 ml viral supernatant and selected with 2 μ g ml⁻¹ puromycin at 72 h postinfection for several days. In experiments measuring cellular activity (with ER imaging), COS-7 cells were stained with HaloTag-BriDye for 30 min, followed by medium replacement and a 5-min incubation and then stained with Fluoroforte (ENZ-51017) for 1 h by following its protocol before imaging. During all imaging experiments, cells were maintained in a live-cell imaging system (Okolab) at 37 °C and 5% CO₂.

PHCT architecture and training details

We designed a PHCT to improve end-to-end SIM reconstruction. The detailed architecture is illustrated in Supplementary Fig. 2. Specifically, the RLFB¹⁹ is a lightweight CNN block consisting of three cascaded residual blocks followed by enhanced spatial attention (ESA)³⁹, used respectively to extract low-level features and aggregate the image features according to the spatial contents:

$$\text{RLFB}(x) = \text{ESA}\left(x + \text{RB}^{(M)}(x)\right), \quad (1)$$

where $x \in \mathbb{R}^{F \times H \times W}$ is the input features with channel number F ($F = 48$ in our experimental setting), and $\text{RB}^{(M)}$ denotes M cascaded residual blocks (RBs), each of which is given by:

$$\text{RB}(x) = x + \text{LReLU}\left(\text{Conv}^{3 \times 3}(x)\right), \quad (2)$$

where $\text{Conv}_{A \rightarrow B}^{k \times k}$ denotes the $k \times k$ convolutional layer mapping features with channel number A into features with channel number B , and LReLU denotes the LeakyReLU activation function with slope 0.05. The 1×1 convolution layers are first used at the beginning of ESA to reduce the channel number for reducing the computational complexity of the spatial attention map, yielding $y = \text{Conv}_{F \rightarrow 16}^{1 \times 1}(x)$, and the spatial attention map is then generated by:

$$\text{Att}_{\text{ESA}}(x) = \text{Conv}_{16 \rightarrow F}^{1 \times 1}\left(\text{Conv}^{1 \times 1}(y) + \text{Up}\left(\text{Conv}^{3 \times 3}\left(\text{MP}\left(\text{SConv}^{3 \times 3}(y)\right)\right)\right)\right), \quad (3)$$

where $\text{SConv}^{3 \times 3}$ denotes a 3×3 convolution layer with a stride of 2, and the MP represents the maximum pooling layer with 7×7 kernel and a stride of 3. The use of the large pooling window and the stride convolution layer enlarges the receptive field of spatial attention. An up-sample layer (denoted by Up) is then used to reshape the attention map to the same spatial dimensions as the input feature y , and a 1×1 convolutional layer is then used to align the channel number. Finally, the output feature is given by:

$$\text{ESA}(x) = x \times \text{Sigmoid}(\text{Att}_{\text{ESA}}(x)), \quad (4)$$

where Sigmoid denotes the sigmoid function.

Afterward, the ETB²⁰ is used, which includes a multihead transposed attention^{20,40} module (MDTA) for modeling fine-grained long-range pixel dependence among image contents, and a feed-forward network (FFN) used to enhance feature flow. MDTA applies a layer normalization on input features, expands the channel number of the features with 1×1 convolutional layer (from 48 to 144 in our experimental setting), and generates query ($q = W_q x$), key ($k = W_k x$) and value ($v = W_v x$) from channel expanded features by a 3×3 depth-wise convolutional (denoted by DConv) layer (that is, convolutional layer with group number equal to channel number), given by:

$$q, k, v = \phi_3 \left(\text{DConv}^{3 \times 3} \left(\text{Conv}_{F \rightarrow 3F}^{1 \times 1} (\text{LN}(x)) \right) \right), \quad (5)$$

where LN denotes the layer normalization and ϕ_M denotes the operation that splits a tensor into the specified number of M chunks. The attention map is then calculated as $\text{Softmax}(qk^T/\alpha)v$, where α is a learnable scaling parameter. Combined with a skip connection, the MDTA module is given by:

$$\text{MDTA}(x) = x + \text{Softmax}(qk^T/\alpha)v. \quad (6)$$

MDTA is followed by a gated FFN to improve the informative feature flow, which is given by:

$$\begin{cases} x_1, x_2 = \phi_2 \left(\text{DConv}^{3 \times 3} \left(\text{Conv}_{F \rightarrow 6F}^{1 \times 1} (\text{LN}(x)) \right) \right) \\ \text{FFN}(x) = x + \text{Conv}_{3F \rightarrow F}^{1 \times 1} (x_2 \times \text{GELU}(x_1)) \end{cases} \quad (7)$$

where GELU denotes the Gaussian error linear unit activation function:

$$\text{GELU}(x) = 0.5x \left(1 + \frac{2}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} e^{-t^2} dt \right). \quad (8)$$

Next, we used PPA to extract and enhance detailed structures with SR information spatial position encoding. The overall PPA is formulated as

$$\text{PPA}(x) = \text{Sigmoid} \left(\text{Conv}_{4P \rightarrow 3}^{3 \times 3} \left(\text{PS}_\downarrow \left(\text{Conv}_{F/4 \rightarrow P}^{3 \times 3} \left(\text{PS}_\uparrow(x) \times \text{pattern} \right) \right) \right) \right), \quad (9)$$

where $\text{PS}_\downarrow$ denotes the pixel-shuffle layer⁴¹ and $\text{PS}_\uparrow$ denotes the pixel unshuffle layer. PPA starts with determining the pattern parameters and constructing illumination patterns in $2 \times$ up-sample space. Parameter determination for each imaging is performed in the data preprocessing (Supplementary Note 8) and the generation of the pattern for each training patch is performed throughout the online data augmentation. A pixel-shuffle layer is then used to rescale the features, yielding features with channel number $F/4$, $2 \times$ spatial size, and a 3×3 convolutional layer is used to endow structurally coded position information of each pixel, yielding input features with channel number P , where P is 9 in TIRF/GI-SIM and 15 in 3D-SIM, respectively. Thereby, the features have the same dimension as the patterns. The features are multiplied by the patterns afterward and are delivered into a pixel unshuffle layer (the up-/down-sample ratio is 2), a 3×3 convolutional layer and a Sigmoid activation function to restore the original spatial and channel dimensions. The pixel-wise product of the encoded features and the illumination patterns integrates physical knowledge about harmonic light modulation into the DL models, thus allowing the network to grasp and boost the position-encoded SR information hiding in feature maps. Applying the convolution theorem amounts to performing band shifting at the prescribed frequency, very similar to what we do in conventional SIM reconstruction.

The output of the PPA block is integrated by a RLFB, an ETB and a global skip connection. The output of the last ETB is fed into an up-sample module consisting of a convolutional layer increasing the channel number of the feature by $4 \times$, a pixel-shuffle layer executing $2 \times$ upsample and a convolutional layer compressing the number of channels into 1 to output reconstructed SR images. Taking together, the overall PHCT is formulated jointly as:

$$\begin{cases} y = \text{Conv}_{P \rightarrow F}^{3 \times 3}(x) \\ z = \text{ETB}(\text{RLFB}(\text{PPA}(\text{ETB}(\text{RLFB}(y)))) \\ \text{PHCT}(x) = \text{Conv}_{F \rightarrow 1}^{3 \times 3}(\text{PS}_\uparrow(\text{Conv}_{F \rightarrow 4F}^{3 \times 3}(y + z))) \end{cases} \quad (10)$$

In 3D-PHCT (Supplementary Fig. 18), all the 2D convolutional layers are replaced with 3D convolutional layers, and the pattern activation of the 3D-PPA block is performed in the in-focus plane. Furthermore, considering that calculating 3D convolutional layers requires more CUDA memory capacity and the raw data of 3D-SIM can be much larger than TIRF/GI-SIM data, we need to simplify 3D-PHCT. Specifically, we removed the 3D-RLFB and 3D-ETB before the 3D-PPA block. The pre-processing of input data includes camera background removal and normalization. To avoid numeric instability caused by the impact of sporadic camera dead points and salt-and-pepper noise, data normalization is given by $d_i \leftarrow d_i / \text{percentile}(d_i, 99.9\%)$, where subscript i indexes the i th data point (that is, training sample) in a training minibatch and the percentile $(x, p\%)$ outputs the intensity value that ranks $p\%$. To augment the training set, we shuffled training raw frames with the same illumination patterns (offline; Supplementary Fig. 4) and performed arbitrary-angle random rotation and random cropping (online). To train 2D PHCT, each minibatch consists of 16 noisy raw image patches at $128 \times 128 \times 9$ and noisy reconstructed image patches at 256×256 , respectively. To train 3D-PHCT, each minibatch consists of four noisy raw image patches at $8 \times 96 \times 96 \times 15$ (z - x - y -15) and noisy reconstruction at $8 \times 192 \times 192$, respectively. The learning rate is typically initialized as 3×10^{-4} and decayed to 1×10^{-6} in 200,000 minibatch iterations with the cosine annealing strategy. The model weights are initialized with Kaiming normal distribution and trained with ADAM optimizer⁴² with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 1 \times 10^{-8}$. The training and inference were performed on a computer workstation equipped with an Intel(R) Core(TM) i9-10900x CPU at 3.7 GHz and an NVIDIA GeForce RTX 3090 graphic processing cards (GPU) with Python v.3.8.10 and PyTorch v.1.13.1.

Self-supervised uncertainty quantification

SSR-SIM allows for self-supervised assessment of neural network reconstruction uncertainty. Based on current uncertainty quantification methods reported in computer vision²², we implemented a GT-free approach using Monte Carlo dropout (MCD) for model uncertainty quantification. To assess model uncertainty of SSR-SIM models, we applied MCD with $P = 0.05$ before all learnable parameters (Extended Data Fig. 6a,b), specifically adding MCD layers preceding all convolutional and attention layers. MCD layers were enabled during both training and inference phases. Notably, using MCD can cause marginal performance degradation. For instance, on the F-actin dataset (BioSR), we observed a 0.07–0.18 dB decrease in validation PSNR after MCD integration. During inference, a trained SSR-SIM model with MCD generates several stochastic samplings (typically $n = 10$), yielding varied outputs. Calculating the mean and s.d. across these outputs produces the final reconstruction and its corresponding uncertainty map (Extended Data Fig. 6c). To demonstrate uncertainty quantification effectiveness, we trained MCD-enabled SSR-SIM models on microtubule data (Domain A) and Clathrin-coated pits data (Domain B), then tested both on microtubule data (Domain A). Results indicate that the model trained on out-of-distribution data (Domain B) exhibits suboptimal reconstructions and elevated uncertainty when tested on in-domain data (Domain A) (Extended Data Fig. 6d,e). Furthermore, we

computed model uncertainty across 10,000 frames of fast SSR-SIM F-actin imaging. Despite the absence of GT, the MCD-enabled SSR-SIM model delivers robust reconstruction characterization, maintaining consistently low uncertainty throughout the entire SR imaging sequence (Supplementary Fig. 11a,c and Supplementary Video 3).

Evaluation experiments on BioSR dataset

We evaluate the SSR-SIM and other considered methods quantitatively on the published BioSR⁹ dataset. The BioSR dataset includes four types of cytoskeleton and organelle: F-actin (LifeAct-mEmerald, COS-7 cells) imaged with TIRF-SIM, MTs (Ensconsin-mEmerald, COS-7 cells) imaged with GI-SIM, CCPs (Clathrin-GFP, SUM159 cells) imaged with TIRF-SIM and ER (KDEL-mEmerald, COS-7 cells) imaged with GI-SIM. However, considering that the ER images were contaminated by motion artifacts, we adopt only the first three biostructures. In our evaluation experiments, the GT data are obtained by the conventional Wiener-filter-based SIM reconstruction of the highest intensity (almost noise-free) raw images. To avoid edge effects, the SR images were cropped centrally from $1,004 \times 1,004$ to 960×960 , and correspondingly the raw images were centrally cropped from $502 \times 502 \times 9$ to $480 \times 480 \times 9$. The dataset of each structure contains more than 50 samples. The first ten samples are used for evaluation and the others are used for training. To evaluate a wide range of SNRs and to circumvent the mismatch between noisy raw data and GT that arises from sample movement in live-cell imaging, noisy raw images are generated by superimposing Poisson and Gaussian noise:

$$d(x) = \mathcal{P} \left[\alpha \times \frac{d^{(GT)}(x)}{\text{APC}(d^{(GT)}(x))} \right] + n_G(x) + b(x), \quad (11)$$

where $d^{(GT)}(x)$ is the noise-free raw images (we used the term GT because $d^{(GT)}(x)$ is the GT data for wide-field image denoising and rationalized DL-SIM denoising¹⁰), $\mathcal{P}$ denotes the Poisson process, α is the APC of aimed noisy raw images, $n_G(x) \sim \mathcal{N}(0, \sigma^2)$ where σ is the s.d. of the Gaussian noise ($\sigma = 10$ in our experimental setting) and $b(x)$ denotes the camera background. We used a series of α (25, 45, 70, 120, 180, 240, 300, 360, 420, 480, 540, 600) to produce noisy raw images at varying SNR levels.

Statistics and reproducibility

To visualize image data with high dynamic ranges, colorization, multi-channel merging and gamma correction were performed by using Fiji v.1.5.4f⁴³. The gamma values used were provided in figure captions. The rolling ball⁴⁴ and wavelet-based background removal algorithms⁴⁵ were performed by using the Fiji plugin and PyWavelets v.1.7.0 Python scripts, respectively, to subtract excessive out-of-focus fluorescence background in SR images. In Fig. 3f,i and Extended Data Fig. 9e, the moving trajectory of VSVG proteins and VACV particles, along with temporal color-coded visualizations of their trajectories, were analyzed and plotted using TrackMate v.7.6.1 (ref. 46) within Fiji. To calculate the reconstruction errors (Fig. 5a and Extended Data Figs. 4 and 5), we first normalized GT images using $x \leftarrow x/\text{percentile}(x, 99.9\%)$, where $\text{percentile}(x, p\%)$ is the intensity value at the p th percentile. We then aligned the intensity range of the evaluated images to the normalized GT using least squares. Residuals were calculated by subtracting the aligned evaluated images from the normalized GT images. These residuals were visualized using the 'seismic' colormap (Extended Data Figs. 4 and 5) and the 'SQUIRREL-FRC' colormap (Fig. 5a, color bars are attached in figures). Both colormaps represent zero values as white and distinctly color positive/negative values, enabling simultaneous visualization of error direction without requiring absolute or squared residuals. Orthogonal views of 3D-SIM raw and reconstruction images (Fig. 5a,c) were generated by using Fiji. Depth-color-coded images from 3D-SIM reconstruction (Fig. 5c) were generated by rendering each slice with colors extracted from the Jet colormap (Fiji), followed by maximum intensity projection of the x - y - z - t stack in the z -dimension.

The calculation formula and implementation of the APC, PSNR/SSIM/NRMSE metrics and decorrelation analysis⁴⁷ are provided in Supplementary Note 9. The metric data in Figs. 1a,b and 5b and Extended Data Figs. 1f–i and 3b are presented in Tukey box-and-whisker format. The box's lower and upper bounds represent the 25th and 75th percentiles, with the interquartile range (IQR) defined as the box length. The lower whisker extended to the greater of the minimum value or the 25th percentile minus $1.5 \times \text{IQR}$, whereas the upper whisker extended to the lesser of the maximum value or the 75th percentile plus $1.5 \times \text{IQR}$. Intensity profiles of line segments of interest in the image (Fig. 2b–g,k and Supplementary Fig. 9a) were extracted using Fiji and normalized and plotted using Matplotlib. In Extended Data Fig. 1j,k, the resolving proportion of the synthetic hollow vesicles was visualized using color-coded matrices (colormap: coolwarm_r, Matplotlib). In Extended Data Fig. 2, to characterize frequency resolution, Fourier spectrum analysis and decorrelation analysis were performed and visualized using Fiji and MATLAB R2021a. The overall figure layout was finalized using Adobe Illustrator.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The published BioSR dataset⁹ was used to validate SSR-SIM and other SIM algorithms (Fig. 1f and Extended Data Figs. 2, 4 and 5). Due to the substantial size of long-term live-cell imaging data, depositing the complete dataset in a public repository is not currently feasible. However, representative F-actin imaging datasets for training and testing SSR-SIM models are publicly available via figshare at <https://doi.org/10.6084/m9.figshare.30406366.v1> (ref. 48). Any additional live-cell imaging datasets are available from the corresponding authors upon request. Source data are provided with this paper.

Code availability

The SSR-SIM Python code is available on GitHub at <https://github.com/HUST-Tan/SSR-SIM> (MIT license). Example training and testing datasets and pretrained models are available via figshare at <https://doi.org/10.6084/m9.figshare.30406366.v1> (ref. 48).

References

- Fiolka, R., Beck, M. & Stemmer, A. Structured illumination in total internal reflection fluorescence microscopy using a spatial light modulator. *Opt. Lett.* **33**, 1629–1631 (2008).
- Chen, L. et al. Sca-cnn: spatial and channel-wise attention in convolutional networks for image captioning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (eds Mortensen, E. et al.) 5659–5667 (IEEE, 2017).
- Vaswani, A. et al. Attention is all you need. In *Proc. International Conference on Neural Information Processing Systems* (eds Guyon, I. et al.) 5998–6008 (PMLR, 2017).
- Shi, W. et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (eds Mortensen, E. et al.) 1874–1883 (IEEE, 2016).
- Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *Proc. International Conference on Learning Representations* (eds Bengio, Y. et al.) 1–15 (ICLR, 2015).
- Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
- Sternberg, S. R. Biomedical image processing. *Computer* **16**, 22–34 (1983).
- Galloway, C. M., Le Ru, E. C. & Etchegoin, P. G. An iterative algorithm for background removal in spectroscopy by wavelet transforms. *Appl. Spectrosc.* **63**, 1370–1376 (2009).

46. Tinevez, J.-Y. et al. TrackMate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
47. Descoux, A., Grubmayer, K. S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nat. Methods* **16**, 918–924 (2019).
48. Liu, J. & Li, D. SSR-SIM dataset. *figshare* <https://doi.org/10.6084/m9.figshare.30406366.v1> (2025).

Acknowledgements

We thank K. He (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences) for providing the VSVG-mEmerald plasmid DNA and the SUM159 cell line expressing Rab11a-GFP (KI). This work was supported by grants from the National Natural Science Foundation of China (NSFC; 92254306 (D.L.), 32125024 (D.L.), 62071197 (S.T.), 62471192 (S.T.)); STI2030-Major Projects (2022ZD0206700 (D.L.)), the National Key R&D Program of China (2024YFA1307400 (D.L.)), Beijing Natural Science Foundation (Z240012 (D.L.)) and the New Cornerstone Science Foundation (D.L.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

D.L., S.T. and Y.-J.L. supervised the research; J.L. and H. Liu developed the algorithm under the supervision of D.L., S.T. and E.Y.L.; X.D., X.H. and H. Lu built the hardware systems under the supervision of D.L.; J.L., Q.M. and H. Lu developed the imaging and reconstruction software; W.L., T.J., A.J. and X.G. prepared the samples; J.L., H. Lu, W.L., T.J. and X.D. performed the imaging experiments under the supervision of

D.L. and Y.-J.L.; J.L., W.L., T.L. and J.C. analyzed the data and prepared the figures and videos with conceptual advice from D.L. and S.T.; D.L., S.T. and J.L. wrote the paper with input from all authors. All authors participated in discussions and commented on the paper.

Competing interests

S.T., J.L. and D.L. have a pending patent on the presented framework. The other authors declare no competing interests.

Additional information

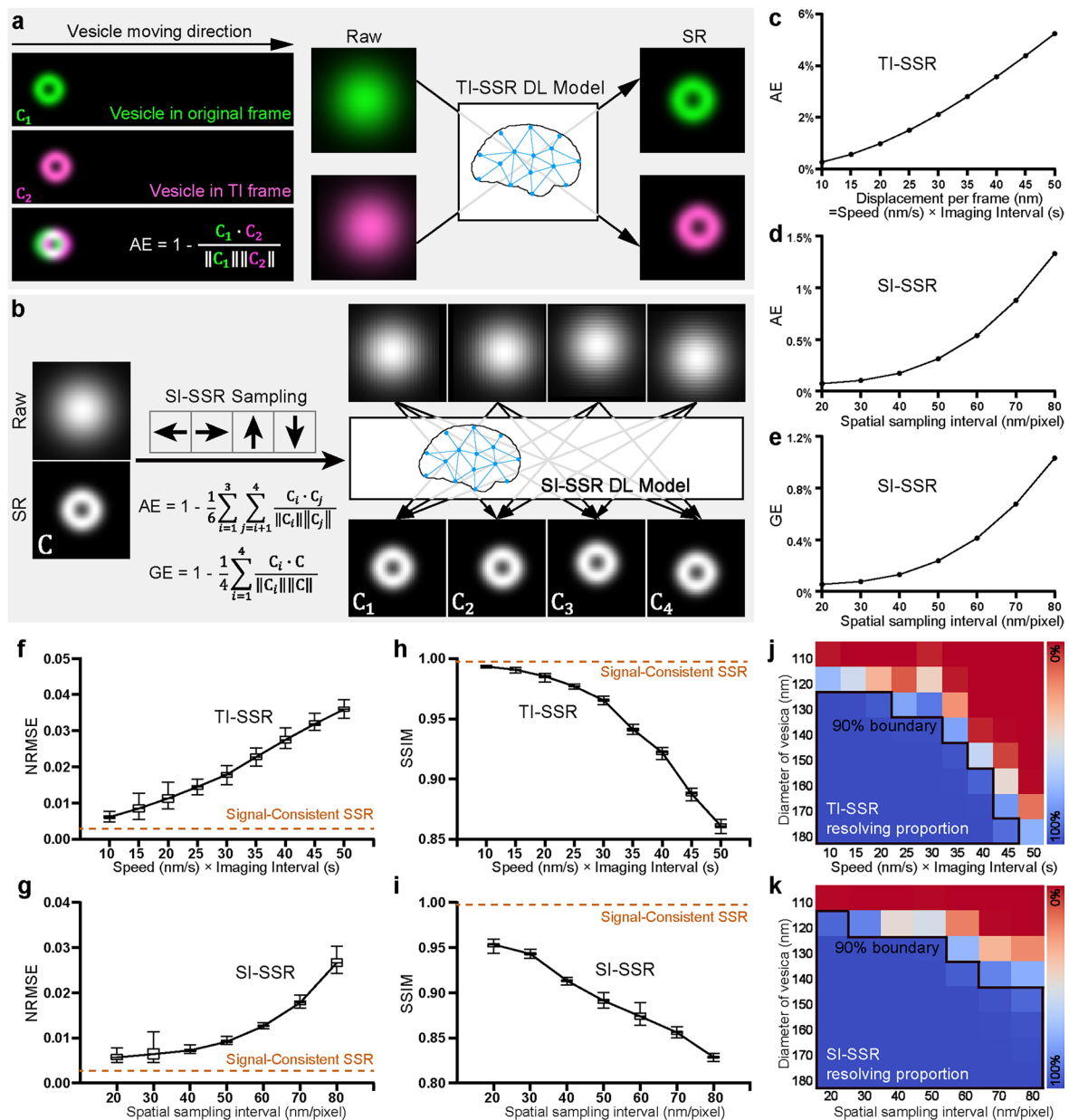
Extended data is available for this paper at <https://doi.org/10.1038/s41592-025-02966-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-025-02966-y>.

Correspondence and requests for materials should be addressed to Yan-Jun Liu, Shan Tan or Dong Li.

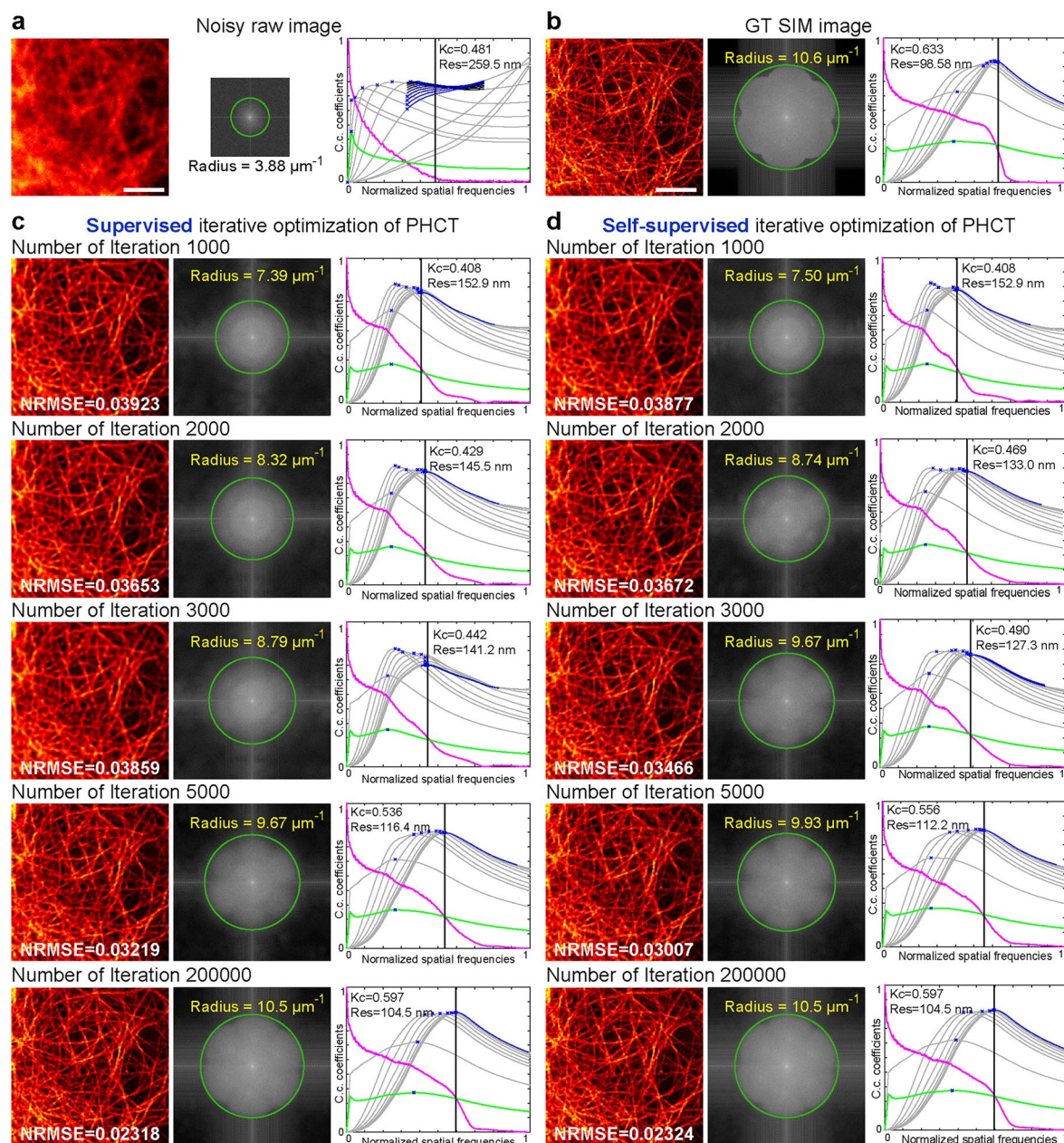
Peer review information *Nature Methods* thanks Ming Lei and Shian Zhang for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editors: Rita Strack and Nina Vogt, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.



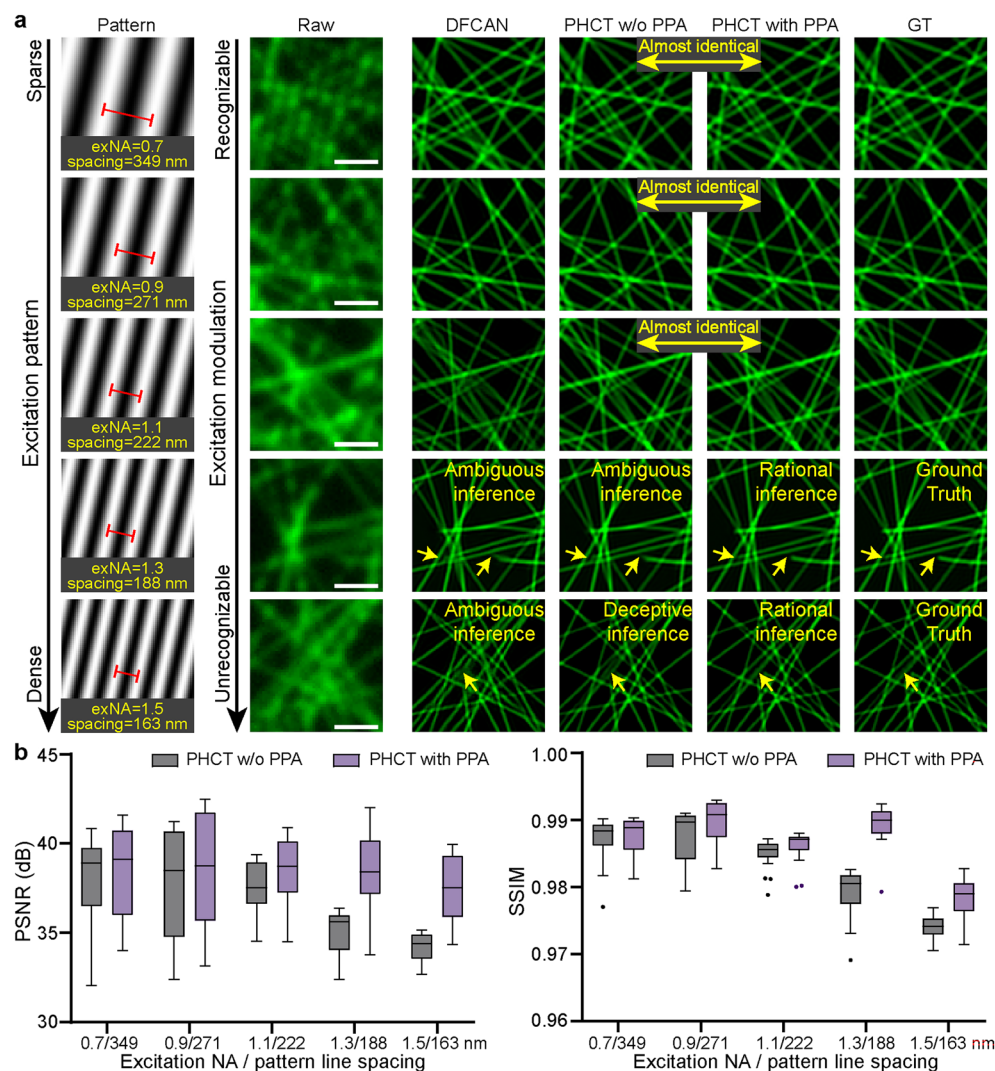
Extended Data Fig. 1 | Assessments of different SSR sampling strategies for high spatiotemporal resolution imaging. **a,b**, Simulation of moving hollow vesicles (diameter: 120 nm) under temporal (**a**, TI-SSR) and spatial (**b**, SI-SSR) sampling strategies. The alignment error (AE, defined as 100% minus the cosine similarity between paired sampled noisy images) and generalization error (GE, defined as 100% minus the cosine similarity between a sampled noisy image and its corresponding source image) are shown. **c**, AE for TI-SSR across

varying average displacements per frame (10–50 nm) at a fixed camera sampling rate of 60 nm/pixel. **d,e**, AE (**d**) and GE (**e**) for SI-SSR across varying camera spatial sampling rates (20–80 nm/pixel); no sample displacement was applied. **f–i**, NRMSE and SSIM metrics of TI-SSR and SI-SSR outputs under deviated self-supervised prerequisites ($n = 100$). Central line, medians; limits, 75% and 25%; whiskers, maximum and minimum. **j,k**, Proportion of resolved hollow vesicles under different sampling settings.



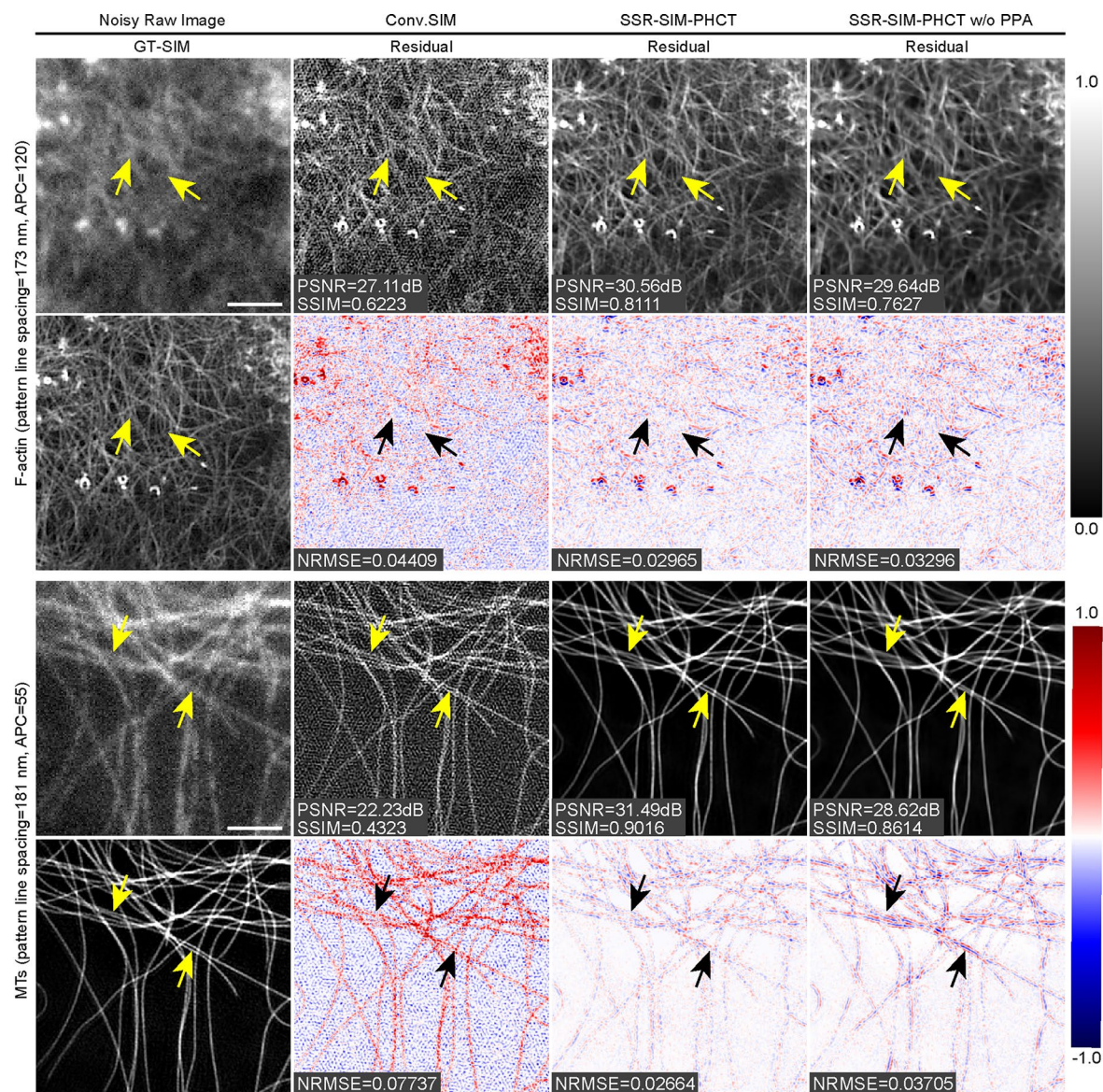
Extended Data Fig. 2 | Comparison of supervised and self-supervised iterative optimization of PHCT. a, The noisy raw image of F-actin with spectrum (middle) and decorrelation analysis (right). Kc denotes the cut-off frequency measured in the ratio of the Nyquist frequency. **b**, The corresponding ground-truth (GT)

image. **c,d**, Validation image patches and metrics for supervised optimization (**c**) and self-supervised optimization (**d**) at iteration numbers 1,000, 2,000, 3,000, 5,000, and 200,000. Scale bar, $2 \mu\text{m}$ (**a,b**).



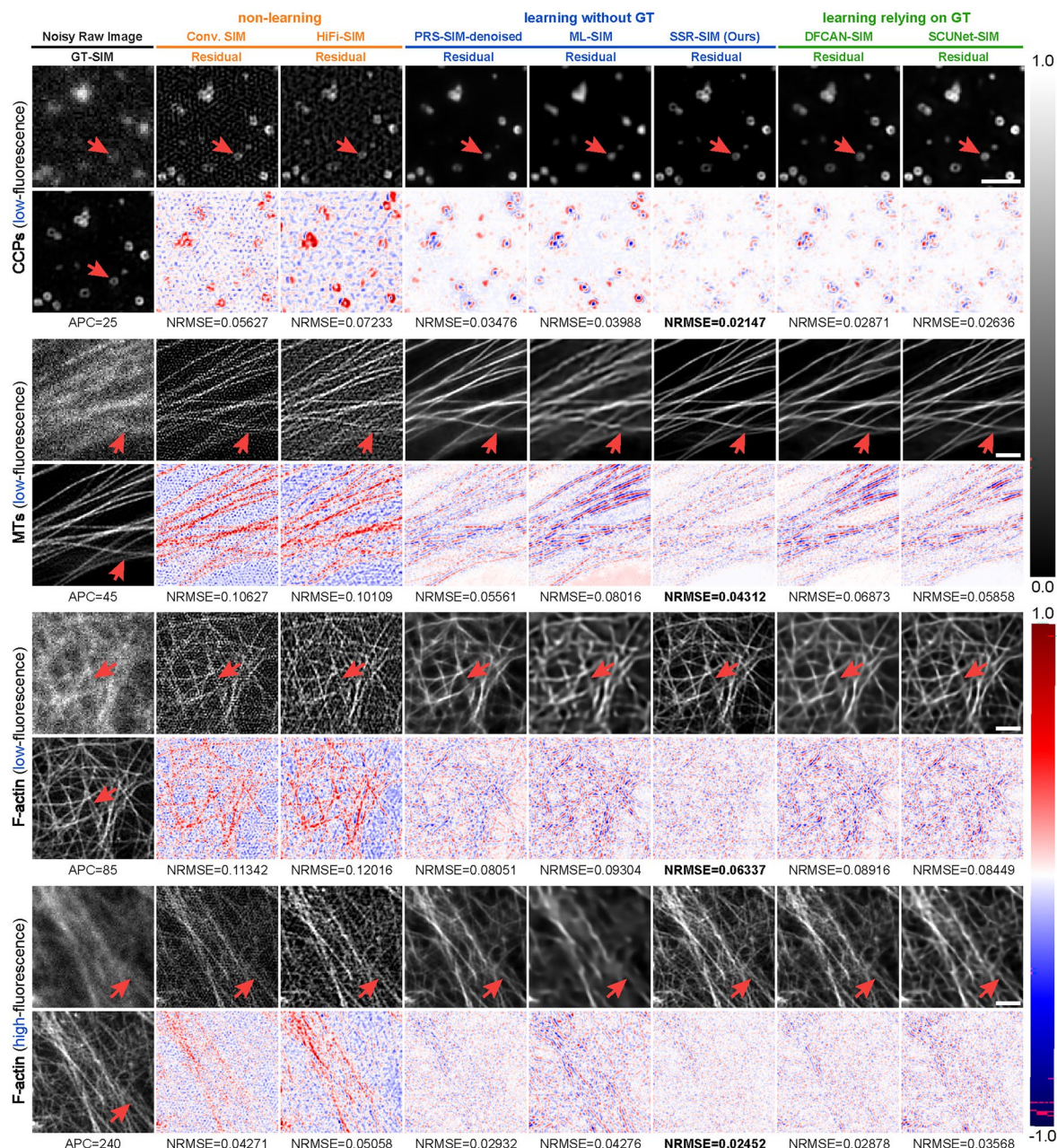
Extended Data Fig. 3 | Assessments of the impact of excitation pattern density on deep learning-based SIM. a, The pattern line spacing varies with the excitation NA. Yellow arrows point to representative regions in the high-excitation-NA reconstruction. Experiments were repeated independently

for 5 ROIs with similar results. **b,** PSNR and SSIM comparisons of PHCT with/without the PPA block at varying excitation NA and pattern line spacing, $n = 20$. Central line, medians; limits, 75% and 25%; whiskers, maximum and minimum. Scale bar, $1 \mu\text{m}$ (a).

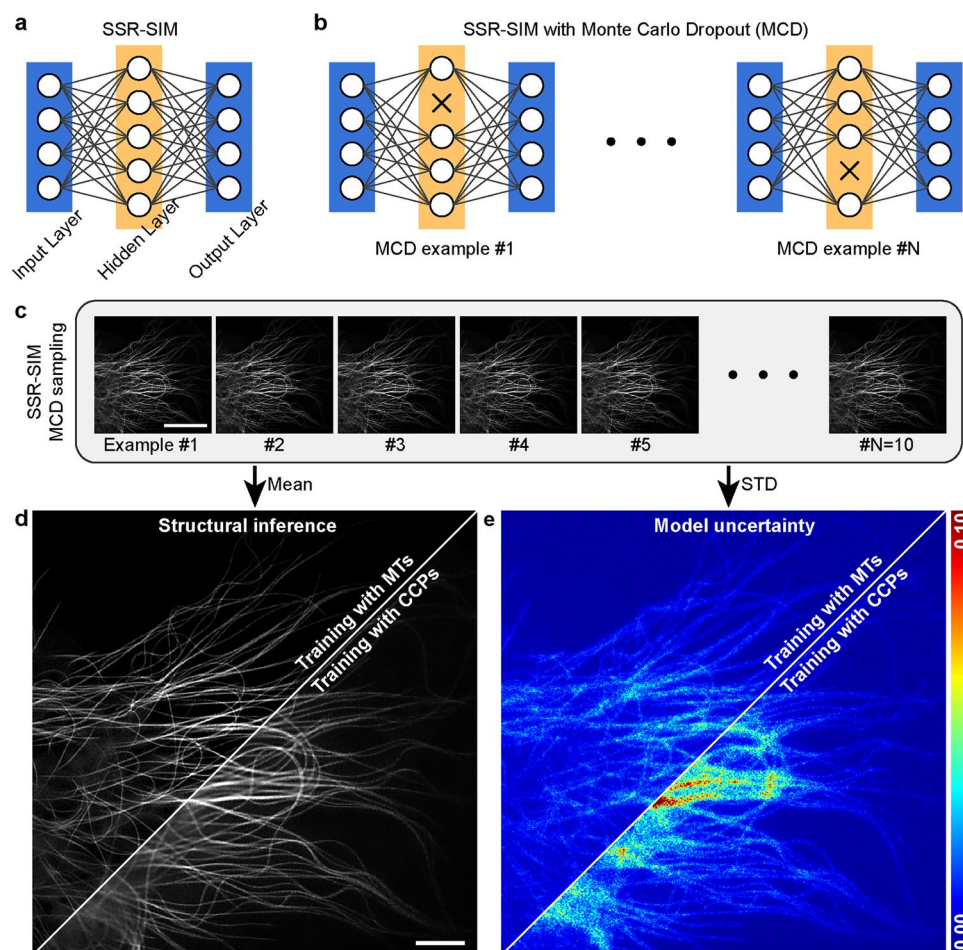


Extended Data Fig. 4 | Ablation study of the PPA block. Representative SSR-SIM images of microtubules and F-actin using PHCT, with and without the PPA block. Arrows highlight fine microfilament connections and adjacent microtubules.

PSNR, SSIM, and NRMSE values are attached for reference. Experiments were repeated independently for 5 ROIs (in different cells) with similar results. APC, average photon count of raw data. Scale bar, 2 μ m.

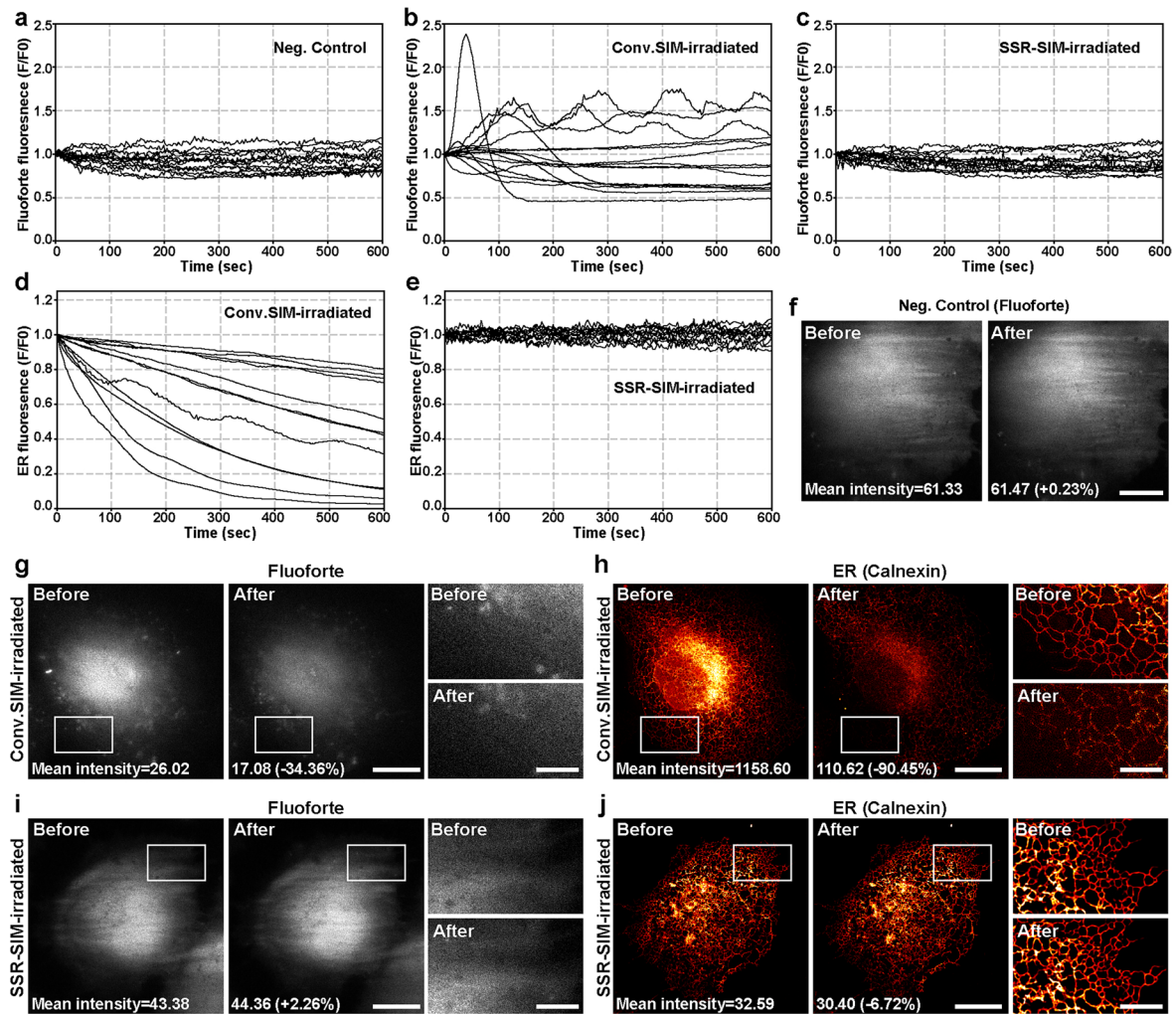


Extended Data Fig. 5 | Comparison of SSR-SIM with other SIM algorithms. Red arrows indicate representative hollow ring structures of CCPs, adjacent microtubules, and dense F-actin. The NRMSE metric for each reconstruction is provided. Experiments were repeated independently for 5 ROIs with similar results. APC, average photon count. Scale bar, 1 μ m.



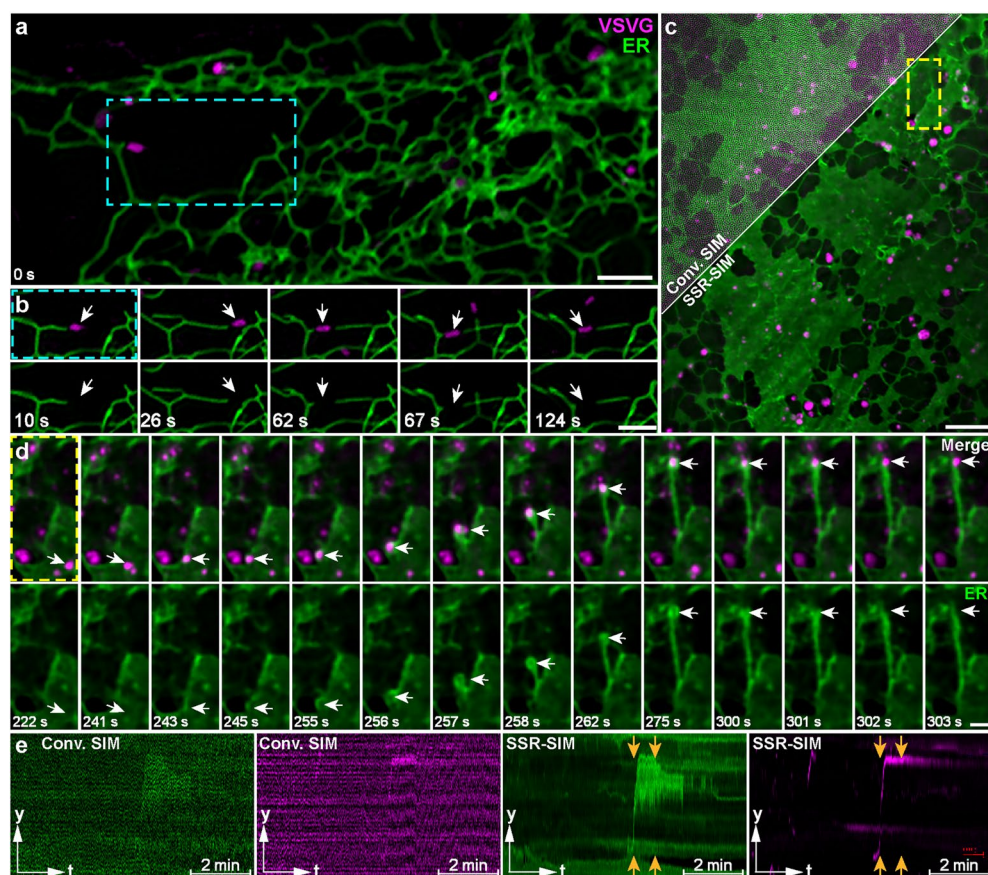
Extended Data Fig. 6 | Reconstruction and uncertainty quantification of SSR-SIM models via Monte Carlo Dropout. **a**, Schematic of the SSR-SIM architecture. Input/output layers are shown in blue and hidden layers in orange. **b**, Diagram of the Monte Carlo Dropout (MCD) strategy used for uncertainty quantification. MCD (dropout rate $p = 0.05$) was applied before all learnable layers during both training and inference. **c**, Multiple stochastic reconstructions generated from a single input using the trained SSR-SIM model with MCD.

d,e, The mean reconstruction (**d**) and standard deviation (uncertainty, **e**) computed from the multiple outputs shown in **c**. Models were trained on microtubule data (MTs, Domain A) and Clathrin-coated pits data (CCPs, Domain B), and tested on microtubule data (Domain A). For each training condition, the final mean reconstruction and the corresponding uncertainty maps are displayed. Experiments were repeated independently for 5 cells with similar results. Scale bars, 10 μm (**c**) and 3 μm (**d**).



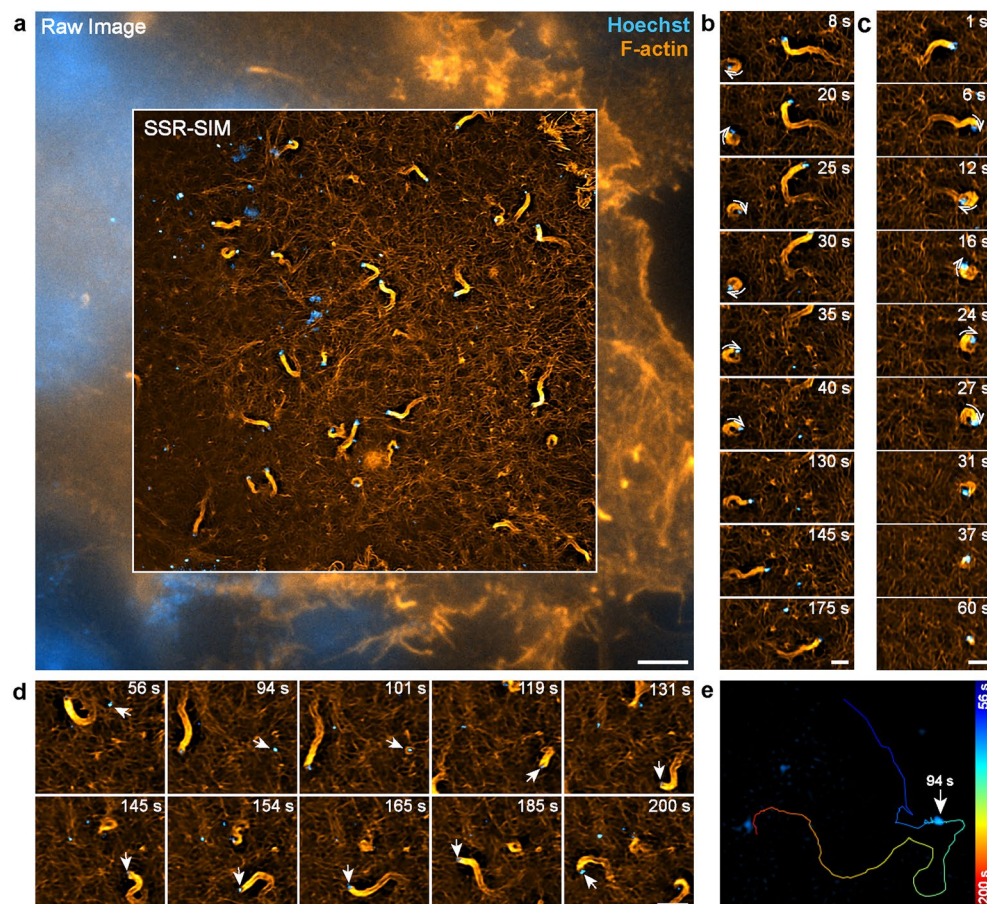
Extended Data Fig. 7 | Assessment of cellular activity and photobleaching under conventional SIM and SSR-SIM imaging. **a**, Ca^{2+} -level response in Fluorescein-dyed and Calnexin-Halo-expressing COS-7 cells under negative control condition (Fluorescein excited in 488-nm laser, without 561-nm-channel excitation or SIM illumination) ($n = 14$). Imaging for each ROI was acquired over 10 minutes (600 time points); same for all subsequent panels. **b,c**, Ca^{2+} -level response over time during conventional SIM (**b**) and SSR-SIM (**c**) live-cell imaging ($n = 15$

each). **d,e**, Endoplasmic reticulum (ER) fluorescence signals over time during conventional SIM (**d**) and SSR-SIM (**e**) imaging. ($n = 15$ each). **f**, Representative Ca^{2+} signal images under negative control condition. **g-j**, Representative Ca^{2+} signal (**g,i**) and ER images (**h,j**) before and after conventional SIM and SSR-SIM imaging. Experiments (**f-j**) were repeated independently for 15 cells with similar results. Scale bars: 15 μm (full-field views in **f-j**) and 5 μm (zoomed-in regions in **g-j**).



Extended Data Fig. 8 | SSR-SIM imaging of VSVG and sheet-like ER interaction. **a**, SSR-SIM imaging of co-movement of VSVG and ER structures. **b**, Time-lapse images of the dynamic events in cyan-boxed region (**a**). **c**, A separate image showing additional VSVG and ER interactions. **d**, Time-lapse images of the dynamic events in yellow-boxed regions (**c**). **e**, The y - t plots corresponding to the

trajectories indicated in **d**. Orange arrows indicate the 256 s and 300 s time points in the sequence. All experiments were repeated independently for more than 3 ROIs with similar results. Gamma, 0.7 (**a-e**). Scale bars, 2 μm (**a,b**), 5 μm (**c**), and 1 μm (**d**).



Extended Data Fig. 9 | SSR-SIM imaging of actin comet tails in Vero cells infected with Vaccinia Virus (VACV). **a**, Vero cells stably expressing LifeAct-mRuby (Orange Hot) and stained with Hoechst dyes (Cyan Hot). Experiments were repeated independently for 3 cells with similar results.

b-d, Time-lapse images of VACV particles and actin comet tails. **e**, Movement trajectory of the VACV particle indicated in **d**. The white arrow denotes the position at the 94-second time point. Gamma, 0.7 (**a-d**). Scale bars, 5 μ m (**a**), 2 μ m (**b-d**).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No collected data. All data were acquired using our home-built systems where the details of optical and software configurations can be found in the Methods section.
Data analysis	The deep learning models were developed using Python v.3.8.10 and Pytorch v.1.13.1, and the source code has been uploaded onto Github (https://github.com/HUST-Tan/SSR-SIM). The decorrelation analysis was performed using Matlab R2021a. Intensity profiles of line segments of interest in the image were extracted using Fiji v.1.5.4f. and normalized and plotted using Python v.3.8.10 and Matplotlib v.3.9.2. In Extended Data Figure 1j,k, resolving proportion of the synthetic hollow vesicles were visualized using color-coded matrices using Matplotlib v.3.7.5. The rolling ball and wavelet-based background removal algorithms were performed by using the built-in plugin in Fiji v.1.5.4f and PyWavelets v.1.7.0, respectively. The colorization, gamma correction, and multi-channel merging were performed by using Fiji v.1.5.4f. The moving trajectories of VSVG proteins and VACV particles were analyzed and plotted using TrackMate v7.6.147 within Fiji v.1.5.4f. Depth color-coded images from 3D-SIM reconstruction (Fig. 5c) were generated by rendering each slice with colors extracted from the Jet colormap (Fiji v.1.5.4f). The overall figure layout was finalized using Adobe Illustrator (CS6).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The published BioSR dataset was used to validate SSR-SIM and other SIM algorithms. Due to the substantial size of long-term live-cell imaging data, depositing the complete dataset in a public repository is not currently feasible. However, representative F-actin imaging datasets for training and testing SSR-SIM models are publicly available on Figshare repository at https://figshare.com/articles/dataset/SSR-SIM_Dataset/30406366. Any additional live-cell imaging datasets are available from the corresponding author upon request. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender Sex and gender were not considered in study design and no finds were sex- or gender-dependent.

Reporting on race, ethnicity, or other socially relevant groupings No socially constructed or socially relevant categorization variable(s) were used.

Population characteristics No human research participants.

Recruitment No human research participants.

Ethics oversight No human research participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size The sample size (n) of each experiment is provided in the figure legends/captions. To confirm statistical significance, for biological findings the sample sizes (n; imaging regions of interest) were all set to $n \geq 3$, and for statistical comparison of different computational methods, the sample sizes (n) were all set to $n \geq 10$. n varied across experiments, depending on the number of n that was actually detected or collected.

Data exclusions No data was excluded from the analysis.

Replication The number of repetitions for each experiment is provided in the figure legends/captions and all repetitions showed similar results.

Randomization Samples were randomly assigned.

Blinding Data acquisition and analysis were being blinded to the experimental group.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

COS-7 cell line (Cat. No.:CRL-1651) was originally from ATCC. COS-7 cell line is the gift from Dr. Junjie Hu's lab in Institute of Biophysics, Chinese Academy of Sciences. Vero cell line (Cat. No.:SCSP-520) was purchased from National Collection of Authenticated Cell Cultures (NCACC, Shanghai, China). SUM159 cell line is the gift from Dr. Kangmin He's lab in Institute of Genetics and Developmental Biology, Chinese Academy of Sciences.

Authentication

COS-7 cells were authenticated by ATCC. Vero cells were authenticated by NCACC. SUM159 cells were authenticated by STR profiling.

Mycoplasma contamination

The cell lines used in this study are not contaminated by mycoplasma.

Commonly misidentified lines
(See [ICLAC](#) register)

There is no misidentified cell line used in this study.

Plants

Seed stocks

No plants used in this study.

Novel plant genotypes

No plants used in this study.

Authentication

No plants used in this study.