nature biotechnology

Supplementary information

https://doi.org/10.1038/s41587-025-02553-8

A neural network for long-term super-resolution imaging of live cells with reliable confidence quantification

In the format provided by the authors and unedited

Contents

Section **Page**

Supplementary Notes **2**

1. General consideration of TISR models 2
2. Generation of simulated time-lapse images 4
3. The design of DPA-TISR 6
4. Generalization capability of DPA-TISR 11
5. Extension of DPA-TISR to SIM reconstruction and volumetric SR 14
6. Comparison between DPA-TISR with AiryScan 17
7. The design of Bayesian DPA-TISR 18
8. Rolling Fourier ring correlation analysis 25

Supplementary Tables **26**

Supplementary Figures **28**

Captions for Supplementary Videos **50**

Supplementary References **56**

Supplementary Notes

1. General consideration of TISR models

The resolution of an optical microscope is physically constrained by the diffraction limit, leading to a finite-size point spread function (PSF) in the spatial domain. This inherent limitation implies that a conventional optical microscope cannot distinguish structures finer than ~200 nm, even with the higher magnification or spatial sampling rate beyond the Nyquist criterion.

Super-resolution (SR) microscopy has revolutionized imaging capabilities by surpassing the long-standing limitations of diffraction-limited resolution. This breakthrough is achieved through precise modulation of fluorescence excitation or emission in spatially coordinated manners, such as stimulated emission depletion (STED) microscopy[1, 2], structured illumination microscopy (SIM)[3], or in a stochastic excitation manner, such as stochastic optical reconstruction microscopy (STORM)[4] and photoactivated localization microscopy (PALM)[5], followed by computational reconstruction to obtain the final SR images. However, it is noteworthy that the increase in spatial resolution obtained through any hardware SR approach necessitates longer acquisition time and/or higher illumination intensity, compromising other equally important performance, e.g., speed and duration, in live-cell imaging. While SIM has been gaining popularity for SR live-cell imaging with its low phototoxicity and high temporal resolution, its post-reconstruction process requires relatively high signal-to-noise ratios (SNR) for each raw image to produce high-quality SR-SIM images, impairing fast, low-light, and quantitative imaging. Although multiple improvements in all aspects of SR techniques have been exploring in recent years, the fundamental problems of conventional hardware-based super-resolution imaging methods, including the long acquisition time and severe phototoxicity, still hinder their wider application.

On the other hand, the rapid development of artificial intelligence has assisted conventional SR techniques in surpassing many traditional hardware limitations. For example, various deep learning networks have displayed excellent performance in the single-image super-resolution (SISR) task, which usually transforms a single low-resolution (LR) image to its high-resolution (HR) counterpart. Recently, state-of-the-art neural network models or their variants in computer vision have been adapted to enhance the resolution of microscopic images, including cross-modality transformation[6, 7], self-learning-based axial resolution enhancement[8-10], and SR reconstruction from raw images[11-13]. The integration and innovation of artificial intelligence algorithms with conventional microscopy technologies have emerged as a

fast-growing interdisciplinary research field, significantly affecting the development of SR microscopy.

Nevertheless, SISR-based computational approaches also pose substantial challenges and limits. First, the inherent quantum nature of photons introduces unavoidable stochasticity in optical measurements. In fluorescent imaging and neural network reconstruction, detection noise exacerbates the uncertainty of the results and hinders the visualization of underlying structures. As is shown in Supplementary Fig. 8, the quality of the network output deteriorated as the SNR of the input wide-field images decreased. Second, previous SISR models neglect the temporal consistency of biological imaging, resulting in large difference between two consecutive frames, thereby disrupting the observation of crucial bioprocesses. This lack of temporal stability further deteriorates long-term imaging quality.

Taken together, we believe that incorporating temporal information is crucial to improving spatial resolution, reconstruction fidelity, and temporal consistency in image super-resolution tasks. When processing time-lapse images, a more effective approach compared to existing SISR methodology is to input multiple consecutive frames simultaneously into the network, which allows information from different frames to propagate through each other, leading to optimal results.

2. Generation of simulated time-lapse images

I. Simulation of ground truth images of the tubular structure

We generate the ground truth (GT) images I_{GT} of tubular structures following our previously described procedures[11], which briefly consists of three steps:

(i) Utilize MATLAB's *randi()* function to randomly generate positional parameters for tubules and then employ the *insertShape()* function to sketch them onto a blank image of 384×384 pixels;

(ii) Assign a random elastic deformation field comprised of 4×4 elements and resize it to the same size as the generated image of tubular structure with bicubic interpolation;

(iii) Utilize the MATLAB function *interp2()* to implement the resized elastic deformation field onto the sample image, thereby inducing curvature in the tubular structures.

To mimic the active dynamics of microtubules in live cells and generate corresponding time-lapse image sequence, we simulate a displacement for each simulated tubules between adjacent two frames. Specifically, an initial sample image is generated following the aforementioned procedure. Subsequently, a displacement $D_{i,t}$ is assigned along a randomly specified orientation for each simulated tubule i at each time point t through the MATLAB function *imwarp()*. The displacement $D_{i,t}$ is defined as

$$D_{i,t} = \frac{1 + 2u_{i,t}}{3} v, \tag{1}$$

where $u_{i,t}$ is randomly sampled from the uniform distribution U[0,1], and v is the pre-defined velocity parameter representing the moving speed of current image sequence, which is set to $\sim 1 \mu m/s$, consistent with the growth velocity of microtubules in live COS-7 cells. The pixel size of the generated sample images is 30.3 nm, which is set to be half the pixel size of our Multi-SIM system.

II. Simulation of diffraction-limited wide-field images and GT-SIM images

To simulate the diffraction-limited wide-field (WF) image sequence of tubular structures, each of the ground truth images is convolved with a wide-field point spread function (PSF) based on our imaging configuration of $\mathrm{NA}_{detection} = 1.35$, $\lambda_{emission} = 515\mathrm{nm}$. The PSF is normalized so that the summation of all pixel values is equal to 1. The resulted images are then down-sampled by 2-fold via bicubic interpolation to generate diffraction-limited wide-field images without noise, denoted as $I_{clearWF}$. Finally, the wide-field images contaminated by Poisson noise is generated via:

$$I_{noisyWF} = N_{Poisson}(\alpha I_{clearWF}) + G(0, {\sigma_r}^2) + b_c \quad (2)$$

where, α is the coefficient used to tune the signal level; b_c is the camera background set as 100; $N_{Poisson}$ denotes Poisson corruption, $G(0, {\sigma_r}^2)$ denotes a white noise image with zero mean and variance of ${\sigma_r}^2$ measured from background frames of our sCMOS camera.

On the other hand, the GT-SIM image sequence, I_{GT-SIM}, is generated by simply convolving I_{GT} with SIM PSF, which is 2 fold narrower than the wide-field PSF:

$$I_{GT-SIM} = I_{GT} \otimes PSF_{GT-SIM} \quad (3)$$

where $\otimes$ is the convolution operation.

3. The design of DPA-TISR

Compared to SISR, which concentrates on exploiting the intrinsic properties of a single image to recover as much as the high-frequency information, TISR presents an additional challenge that it involves aggregating information from multiple highly-related but misaligned frames in time-lapse inputs. To this end, we firstly examined whether the state-of-the-art (SOTA) deep neural networks (DNNs) designed for natural video super-resolution (VSR), i.e., super-sampling, tasks could serve as viable solutions for TISR models in fluorescence imaging.

Most existing deep neural networks for VSR consist of four interrelated components: propagation, alignment, aggregation, and up-sampling, among which, the selection of propagation and alignment strategies can significantly impact the performance[14, 15]. Therefore, here we concentrated on evaluating these two essential components, propagation and alignment. Specifically, we explored two mainstream propagation schemes, i.e., sliding window-based propagation and recurrent network-based propagation, and three representative alignment methods based on optical flow (OF), non-local attention (NA), and deformable convolution (DC). After validating these different schemes and methods (Fig. 1, Extended Data Figs. 2, 3, and Supplementary Fig. 2), we found that recurrent- and deformable convolution-based deep neural networks outperformed other combinations in aggregating cross-frame features and demonstrated greater potential for achieving superior super-resolution reconstruction for time-lapse biological data.

Although the recurrent network- and DC-based DNN architecture demonstrates convincing TISR performance, we noticed that the rudimentary designs in DC-based alignment limit the efficacy of information aggregation in two aspects. First, in live-cell fluorescence imaging, the acquisition frame rate is typically controlled carefully to a relatively low value to mitigate photobleaching or phototoxicity. Consequently, the motion-induced differences between two adjacent biological images are often larger than those observed in natural images. Deformable convolution is designed to learn additional offsets to allow the network to gather information beyond its regular local neighborhood[16]. Despite its offset diversity compared to conventional optical flow-based methods, DC still relies on relatively local convolutional operations and has a limited receptive field, producing unsatisfactory results when dealing with the large motion of biological structures. Second, DC-based alignment module is found to be difficult to train. The training instability of DC often results in offset overflow, deteriorating the final performance. To conquer this issue, Chan *et* al.[14] directly incorporated optical flow into the alignment module as initial offsets. However, in

biological image TISR task, we adopted a similar method and observed that the optical flow learned by the network often fails to accurately present the actual motion of cell structures (Supplementary Fig. 1a). This discrepancy may be attributed to the heavier photon noise present in biological images compared to natural images, which substantially interferes with the accuracy of optical flow calculation and distorts the reconstruction results.

As delineated earlier, we contended that an effective alignment mechanism should possess several key attributes: (i) the capacity to model sub-pixel shifts and harness information from neighboring pixels and adjacent frames; (ii) the capability to accommodate substantial motion of biological structures, facilitating global alignment of neighboring frames; (iii) robustness against heavier photon noise, which can degrade optical flow estimation and compromise overall reconstruction quality.

To attain these objectives, especially for achieving global alignment and noise robustness, we directed our focus towards the Fourier domain. Firstly, every component within the Fourier frequency domain encapsulates information from the entire image, facilitating global analysis. Secondly, stochastic photon noise exhibits distinctive characteristics in the frequency domain, owing to its ultra-high frequency, enabling effective differentiation from the image signal. Our intuitive thinking is that alignment in frequency domain might be helpful for biological time-lapse image restoration.

In Fourier analysis, the translation of a signal can be expressed as the change of phase in the frequency domain while the amplitude remains unchanged, which is recognized as the frequency shifting property of Fourier transform. To illustrate this concept, let's consider a one-dimensional signal:

$$\mathrm{F}_t(\omega) = \mathrm{F}(\omega) * e^{-i\omega t} \tag{4}$$

where $\mathrm{F}(\omega)$ and $\mathrm{F}_t(\omega)$ represent the Fourier transform of the original signal and the signal after shifting by t in spatial domain. Extending this concept into a two-dimensional image, the Fourier transform of a two-dimensional image after sptial-shifting can be denoted as:

$$\mathrm{F}_{x,y}(u, v) = \mathrm{F}(u, v) * e^{-i2\pi\left(\frac{ux}{\mathrm{M}}+\frac{vy}{\mathrm{N}}\right)}, \tag{5}$$

$$\left|\mathrm{F}_{x,y}(u, v)\right| = \left|\mathrm{F}(u, v)\right| \tag{6}$$

where x, y represent two-dimensional translations and M, N represent height and width of the image. It is evident that the translation of a signal is equivalent to adding a linear increment to its phase term in frequency space. In essence, by altering the phase of the

Fourier transform of feature maps, we can reposition signals of different frequencies and achieve global movement of the feature maps. Notably, components of different frequencies can be relocated separately at subpixel accuracy, enabling a diverse range of translation schemes.

Taken together, we designed the deformable phase-space alignment mechanism, which involves the following steps: (i) Apply Fourier transform on feature maps, i.e., current feature maps and neighbor feature maps to be aligned; (ii) Adaptively enhance the phase components via the phase convolution; (iii) Apply inverse Fourier transform with the enhanced phase and unchanged amplitude; (iv) Spatially warp the phase-enhanced feature maps using optical flow; (v) Generate offsets for deformable convolution; (vi) Apply deformable convolution to the feature maps for final alignment. See Fig. 2a and Extended Data Fig. 4 for the schematic illustration of DPA mechanism.

To test whether DPA outperforms conventional spatial DC mechanism and its potential variants, we conducted a comparative experiment involving four distinct alignment mechanisms, outlined as follows:

(I) Deformable phase-space alignment (DPA) and its variants

Real-valued fast Fourier transform (denoted as $\mathrm{FFT}(\cdot)$, implemented with *torch.fft.rfft*) is firstly applied on the current feature maps (denoted as h_i) and neighborhood feature maps (denoted as h_{i-k}):

$$h_i^{phase} = Angle(\mathrm{FFT}(h_i))$$

$$h_{i-k}^{phase} = Angle(\mathrm{FFT}(h_{i-k}))$$

$$h_i^{amplitude} = Abs\big(\mathrm{FFT}(h_i)\big)$$

$$h_{i-k}^{amplitude} = Abs\big(\mathrm{FFT}(h_{i-k})\big), \tag{7}$$

where $Angle(\cdot)$ and $Abs(\cdot)$ represent the operation to obtain element-wise angle and absolute value of the features.

(i) Amplitude convolution (Extended Data Fig. 5a): the concatenation of amplitudes, i.e., $h_i^{amplitude}$ and $h_{i-k}^{amplitude}$ goes through a convolutional layer (denoted as $f(\cdot)$) and two consecutive residual blocks (denoted as $\mathrm{RB}(\cdot)$) to learn the amplitude residual:

$$\delta(h_{i-k}^{amplitude}, h_i^{amplitude}) = \mathrm{RB}(\mathrm{RB}(f(p))) + h_{i-k}^{amplitude}$$

$$p = h_i^{amplitude} \oplus h_{i-k}^{amplitude}. \tag{8}$$

An inverse real-valued fast Fourier transform is utilized to reconstruct the amplitude-refined feature maps:

$$h_{i-k}^{refined,amplitude} = \text{iFFT}\left(\delta(h_{i-k}^{amplitude}, h_i^{amplitude}) * e^{i*h_{i-k}^{phase}}\right). \quad (9)$$

(ii) Phase convolution (Extended Data Fig. 5b): the concatenation of phases, i.e., h_i^{phase} and h_{i-1}^{phase} goes through a convolutional layer and two consecutive residual blocks to learn the phase residual:

$$\delta(h_{i-k}^{phase}, h_i^{phase}) = \text{RB}(\text{RB}(f(p))) + h_{i-k}^{phase}$$

$$p = h_i^{phase} \oplus h_{i-k}^{phase}. \quad (10)$$

An inverse real-valued fast Fourier transform is utilized to reconstruct the phase-refined feature maps:

$$h_{i-k}^{refined,phase} = \text{iFFT}\left(h_{i-k}^{amplitude} * e^{i*\delta(h_{i-k}^{phase}, h_i^{phase})}\right). \quad (11)$$

(iii) Phase & amplitude convolution (Extended Data Fig. 5c): both the concatenation of phase and the concatenation of amplitude go through a convolutional layer and a single residual block to learn the phase residual and amplitude residual, respectively.

$$\delta(h_{i-k}^{phase}, h_i^{phase}) = \text{RB}(f(p_1)) + h_{i-k}^{phase}$$

$$p_1 = h_i^{phase} \oplus h_{i-k}^{phase}$$

$$\delta(h_{i-k}^{amplitude}, h_i^{amplitude}) = \text{RB}(f(p_2)) + h_{i-k}^{amplitude}$$

$$p_2 = h_i^{amplitude} \oplus h_{i-k}^{amplitude}, \quad (12)$$

An inverse real-valued fast Fourier transform is utilized to reconstruct the phase & amplitude-refined feature maps:

$$h_{i-k}^{refined,phase\,\&\,amplitude} = \text{iFFT}\left(\delta(h_{i-k}^{amplitude}, h_i^{amplitude}) * e^{i*\delta(h_{i-k}^{phase}, h_i^{phase})}\right). \quad (13)$$

(II) Deformable alignment (DA)

Spatial domain convolution (Extended Data Fig. 5d): the concatenation of current and neighborhood feature maps directly goes through a convolutional layer and two consecutive residual blocks to learn the spatial domain residual.

$$h_{i-k}^{refined,spatial-domain} = \text{RB}(\text{RB}(f(p))) + h_{i-k}$$

$$p = h_{i-k} \oplus h_i, \quad (14)$$

The refined feature maps are then fed into the deformable convolution module as is depicted in Extended Data Fig. 2e for subsequent spatial alignment.

We then examined four TISR models equipped with four alignment mechanisms detailed above on linear SIM data of MTs and simulated data of tubular structures with infallible GT references (Extended Data Fig. 5e-h). We found that feature alignment in the Fourier space typically outperforms alignment conducted solely in the spatial domain. Among the three frequential alignment mechanisms evaluated, phase convolution-based alignment consistently outperforms other configurations with comparable convolutional depth and computational complexity. This trend underscores the superiority of the proposed phase-space alignment mechanism.

4. Generalization capability of DPA-TISR

I. Generalization across different organelles

To test the generalization ability of the DPA-TISR model in processing new type of biological specimens that were not seen in its training process, we applied a DPA-TISR model trained with F-actin data to improve the resolution of time-lapse WF images of other three organelles acquired in the same cell type: lysosomes (Lyso), mitochondria (Mito), and microtubules (MTs) (Supplementary Fig. 9). Despite that Lyso, outer mitochondrial membrane, and microtubules are different from F-actin in morphology, the well-trained DPA-TISR model by F-actin data could accurately infer most of the fine structures of these three unseen specimens. Moreover, we applied the F-actin model as a general initialization, and implemented transfer learning to independently train a specific DPA-TISR model for each specimen. The transfer learning procedures typically took 500 epochs within 3 hours using training data augmented from 5 groups of WF-SR sequence pairs, i.e., ~1/5 time and data requirement compared to ordinary model training with random initialization. As is shown in Supplementary Fig. 9c, the outcome of DPA-TISR models after transfer learning presented finer structures and higher fidelity, e.g., more continuous microtubules and clearer outer membrane of Lyso and Mito, than the corresponding images generated by the F-actin model.

II. Generalization across different cell types

Next, we examined the generalization ability of DPA-TISR models across different cell types by applying them to data of the same biological structures but from different types of cells. Specifically, we independently trained four DPA-TISR models using the Lyso, Mito, MTs, and F-actin data from the established BioTISR dataset, which were acquired from COS-7 cells (Methods and Supplementary Table 1). Then we employed the same plasmids, i.e., Lamp1-Halo for Lyso, 2×mEmerald-Tomm20 for Mito, 3×mEmerald-Ensconsin for MTs, and Lifeact-mEmerald for F-actin, as those used in BioTISR to transfect HeLa cells, thereby obtaining Hela cells with same biological structure labelled. From the original wide-field observation of these two types of cells, we find that the COS-7 cells usually engage a larger area during adhesion while the HeLa cells present higher background fluorescence. Despite these differences in input images, DPA-TISR models trained with COS-7 cells are generally able to improve the resolution of various biological structures in HeLa cells (Supplementary Fig. 10). In most regions, the super-resolved structures by DPA-TISR are of high fidelity, however, there may be some blur or wrong predictions (indicated by red arrows in Supplementary Fig 10b) due to high background in corresponding regions of WF input images.

Similarly, transfer learning could be further implemented to adapt the well-trained COS-7 models to HeLa cells, yielding better robustness and higher inference accuracy in the high background and low SNR regions referred to GT-SIM images (indicated by yellow arrows in Supplementary Fig. 10c).

III. Generalization across different imaging modalities

Finally, we made efforts to explore whether a well-trained DPA-TISR model could be applied to enhance the resolution of WF images acquired with another imaging modality, i.e., the light-sheet microscopy. We employed our home-built lattice light-sheet microscopy (LLSM) system[30] to acquire several volumetric image stacks of Lyso, Mito, MTs, and F-actin, the same specimens with that used for acquiring the BioTISR dataset. There are four major differences between WF images acquire by the Multi-SIM system (detailed in Methods section of the main text) and LLSM system: (i) the pixel size of LLSM images is ~1.5 times larger than that of Multi-SIM images, i.e., 92.6 nm vs. 60.4 nm; (ii) the diffraction-limited resolution of Multi-SIM system is higher than the LLSM system induced by different NA of the detection objectives, i.e., 1.49 for Multi-SIM and 1.1 for LLSM; (iii) the depth of field (DOF) of Multi-SIM images varied from ~100 nm for TIRF illumination to ~1 um for grazing incidence, while the DOF of LLSM is related to the thickness of the light-sheet, ranging in 600-800 nm for different excitation wavelengths, causing different background and out-of-focus fluorescence levels; (iv) the image plane of Multi-SIM system is parallel to the cover glass, whereas that of LLSM system is at an angle of 30 degree to the slide, resulting in discriminatory imaging region within each single capture. During SR processing of LLSM data, we reconfigured the DPA-TISR model to adopt multiple 2D slices of a single volumetric timepoint as the input, therefore allowing the models trained with 2D time-lapse data to handle volumetric data.

Representative raw LLSM images and their SR inference by DPA-TISR models are displayed in the Supplementary Fig. 11. Corresponding deconvolved results using the RL deconvolution algorithm (10 iterations) and lattice light-sheet structured illumination microscopy (LLS-SIM) images of the same regions are also shown for comparison. Leveraging the spatial continuity in axial dimension, the DPA-TISR models trained with Multi-SIM data successfully resolve the structural details of LLSM data with few reconstruction artifacts, such as the ring-like structure of Lyso, outer mitochondrial membrane, and the network of crisscrossing MTs or F-actin filaments. Moreover, benefiting from the laterally isotropic resolution of TIRF-SIM and GI-SIM, the DPA-TISR model could isotropically improve lateral resolution by approximately 2-fold as quantified by decorrelation analysis. In contrast, RL deconvolution and

conventional LLS-SIM only improve resolution by less than 1.5-fold isotropically (Supplementary Fig. 11b) or ~1.5-fold along a single lateral dimension (Supplementary Fig. 11d), respectively.

These results together demonstrate the generalization ability and transfer learning capability of DPA-TISR models, which allow us to quickly deploy the DPA-TISR model to enhance the resolution of data from unseen biological structures, cell types, and imaging modalities.

5. Extension of DPA-TISR to SIM reconstruction and volumetric SR

I. DPA-TISR-SIM for SIM reconstruction

Besides performing SR inference from wide-field images, SR-SIM reconstruction from time-lapse raw SIM images is another important potential application of deep learning-based TISR models. To study whether the proposed DPA-TISR scheme is able to benefit SIM reconstruction from noisy raw SIM images and outperform existing SISR-based SIM reconstruction neural network models such as DFCAN-SIM and CARE-SIM, we first slightly modified the original DPA-TISR model to make it adapt to time-lapse raw SIM image processing. The primary modifications are summarized as below: (i) the channel numbers of the input layer and first convolution layer of the feature extraction module were adjusted to 9 from 1, corresponding to 3 orientations × 3 phases; (ii) in the DPA module, the raw SIM images were averaged into time-lapse wide-field images before optical flow estimation to improve the input SNR and calculation accuracy, then the optical flow maps were used to warp features of each corresponding timepoint; (iii) meanwhile, the time-lapse wide-field images generated above were passed on to the end of the model via a long skip connection, resulting in a residual learning scheme. The schematic of DPA-TISR-SIM architecture is illustrated in Supplementary Fig. 12a.

Next, we prepared two datasets of microtubules and inner mitochondrial membrane, and for each type of structure, we independently trained a DPA-TISR-SIM model, a CARE-SIM model, and a DFCAN-SIM model. In the training procedure, instead of inputting (time-lapse) wide-field images, i.e., average of nine raw SIM images, the models were directly fed with (time-lapse) raw SIM images and supervised by GT-SIM images. Additionally, we trained a DPA-TISR model on each dataset by using time-lapse wide-field images as inputs for reference.

The perceptual and statistical comparison of the results generated by these models are shown in Extended Data Fig. 6, from which we found that (i) compared to CARE-SIM and DFCAN-SIM, the DPA-TISR-SIM substantially improved the SR reconstruction fidelity, achieving the highest PSNR and SSIM. The DPA-TISR-SIM model faithfully resolved the fine structure of microtubules and mitochondrial cristae, which were hardly distinguished in the noisy wide-field images or wrongly inferred by CARE-SIM and DFCAN-SIM models as indicated with the red arrows in Extended Data Fig. 6a. These results indicate that incorporating temporal continuity information significantly benefits the DL-based SIM reconstruction models. (ii) Compared to DPA-TISR, profiting from the abundant high-frequency information modulated in the raw SIM images, the DPA-TISR-SIM model could reconstruct detailed biological structures

with higher accuracy as pointed out by the white arrows in Extended Data Fig. 6a and 6b. This observation is also consistent with conclusions we have discussed for SISR models we have derived before[11]. (iii) Surprisingly, we noticed that the DPA-TISR model with seven time-lapse wide-field images (each frame was generated by averaging nine raw SIM images) as inputs outperformed CARE-SIM and DFCAN-SIM input with nine raw SIM images of the central frame for both biological specimens, implying that the information provided by adjacent frames is more prominent than that encoded in the raw SIM data within the DL-based SR reconstruction task.

II. 3D DPA-TISR for volumetric SR inference

3D-SIM is a common and powerful tool in revealing the volumetric dynamics of biological structures. SR reconstruction of 3D-SIM typically requires 15 raw SIM image stacks, i.e., 3 orientations × 5 phases, thereby a 3D version of DPA-TISR that converts 3D wide-field data into its 3D-SIM counterparts will reduce 14/15, i.e., over 90%, raw data requirement compared to conventional 3D-SIM.

Encouraged by this, we tried to modify the original 2D architecture of DPA-TISR into a 3D one, denoted as 3D DPA-TISR, while preserving the advancements of deformable phase-space alignment (DPA) and recurrent network-based propagation mechanisms. Specifically, we made three primary modifications as follows: (i) we replaced all 2D convolutional layers by 3D convolutional layers with a kernel size of 3×3×3, conferring better volumetric feature extraction and aggregation capability onto the backbone; (ii) in the phase-space alignment module, the volumetric features were processed as a whole via 3D fast Fourier transform (FFT), 3D phase convolution, and 3D feature recombination to capture and align tiny movements of the sample along both lateral and axial dimensions; (iii) in the subsequent optical flow-guided deformable convolution module, we reconfigured the SPyNet and deformable convolution into 3D versions so that they could work in the 3D feature space, achieving accurate 3D optical flow calculation and 3D deformable convolution.

To validate the effectiveness of 3D DPA-TISR and its superiority over its SISR counterparts, we employed our home-built Multi-SIM system to generate three new 3D time-lapse datasets of microtubule, F-actin, and inner membrane of mitochondria, respectively. For each type of specimens, we acquired more than 50 groups of WF sequences (512×512×11 voxels×10 timepoints) and their corresponding high-quality GT-SIM sequences (1024×1024×11 voxels×10 timepoints). Then for each biological structure, we independently trained one 3D DPA-TISR model and one 3D SISR model with the same training dataset for comparison. Similar to the 2D scenario, the network

architecture of 3D SISR model compared here was modified from 3D DPA-TISR, with the only difference been that it excludes all elements related to alignment and propagation (detailed in Methods section of the main text). Additionally, a SISR-based 3D RCAN model for each structure was trained for reference. The qualitative and quantitative comparisons among these models in the 3D TISR task are shown in Extended Data Fig. 7, which indicates that given time-lapse diffraction-limited volumetric data as input, the 3D DPA-TISR model significantly outperforms its SISR version and 3D RCAN both perceptually and statistically, revealing finer structures that are indistinguishable in wide-field and SISR images (indicated with yellow arrows in Extended Data Fig. 7). These results illustrate that the temporal continuity between adjacent frames also benefits volumetric data SR tasks and the proposed DPA-TISR framework is able to well adapt to 3D paradigm, demonstrating its broad applicability in both 2D and 3D implementations.

6. Comparison between DPA-TISR with AiryScan

There exist a number of optical solutions such as instant SIM[17] and image scanning microscopy (ISM)[18] that provide moderate resolution improvement without observably compromising other metrics of imaging performance. As one of the advanced implementations of ISM, AiryScan (Zeiss) has achieved a stunning commercial success and been widely adopted by life science laboratories due to its convenient deployment.

Similarly, a well-trained DPA-TISR model proposed here is expected to instantly improve the resolution of diffraction-limited images. To compare the optical hardware-based instant super-resolution techniques (taking the AiryScan system for example) with our DPA-TISR model, we imaged the same region-of-interests (ROIs) of three types of fixed biological specimens, CCPs, MTs, F-actin, with the AiryScan system and wide-field mode of the Multi-SIM system, respectively. Then we employed DPA-TISR models trained on BioTISR datasets to denoise and enhance the resolution for wide-field images. The Multi-SIM captures before and after DPA-TISR processing, and the AiryScan images after reconstruction concomitantly provided by the system are displayed in the Supplementary Fig. 13a-c. Moreover, we acquired high-quality GT-SIM images of the same ROIs using the SIM mode of the Multi-SIM system for reference (Supplementary Fig. 13d). Resorting to the area detector and corresponding reconstruction algorithm, the AiryScan system could achieve superior spatial resolution over wide-field captures, providing substantial expansion in Fourier power spectrums (Supplementary Fig. 13a vs. 13b). However, the AiryScan scheme or other ISM implementations improve resolution by typically less than 1.7-fold[19], thereby failing to resolve the hollow structure of CCPs or densely interlaced MTs and F-actin. In contrast, the DPA-TISR models could effectively improve spatial resolution by over 2-fold, resolving substantially finer details than AiryScan (Supplementary Fig. 13b vs. 13c), which are comparable to GT-SIM (Supplementary Fig. 13c vs. 13d).

It is noteworthy that the SR capability of DPA-TISR models demonstrated above is fundamentally endowed by the SIM modality of Multi-SIM hardware. If we trained DPA-TISR models using paired WF-AiryScan data, such models cannot gain better resolution than AiryScan. Therefore, in our opinion, the development of optical front-end and algorithm back-end are complementary and will synergize each other. On the one hand, the advancement in hardware provides deep learning models with better targets of higher resolution and quality. On the other hand, advanced algorithms such as DPA-TISR extend the capability of the hardware, i.e., substantially lowering the number or quality requirements of raw data for high-quality SR reconstruction, and enabling ultra-fast long-term live-cell SR imaging.

7. The design of Bayesian DPA-TISR

I. Reliability in deep neural networks

Within the domain of scientific investigation, the observation authenticity is of paramount importance. When the reliability of data cannot be confidently ascertained, it becomes evident that any subsequent conclusions drawn from such data are inherently susceptible to errors. This is particularly pertinent in the field of microscopy, where uncertainties in observations can lead to erroneous interpretations of biological phenomena.

This issue further exacerbates in the situation when computational techniques such as image restoration neural networks or the iterative deconvolution are adopted in the image processing pipeline. These advanced methods demonstrate exceptional efficacy in characterizing the statistical distributions of noises, resulting in notable enhancement in perceptual quality of the restored images. However, they also introduce a potential drawback: the risk of generating errors that closely resemble the ground truth, rendering them difficult to discern for scientists.

Therefore, in the development of deep learning-assisted microscopy technologies, as well as other fields, the evaluation and calibration of the confidence level of neural network outputs will play a significant role and have profound implications for future research and development. Particularly in the field of biological imaging, neural network models with quantifiable-confidence can assist biological researchers in better interpreting imaging results, thereby enhancing the rationality and accuracy of scientific conclusions drawn from the network outputs.

Consequently, it is of vital for a computational image processing method to measure the uncertainty or confidence of restored images. Some efforts have been made in evaluating the dependability and uncertainty of neural networks, and find that the reliability of neural network predictions can be quantified through two predictive uncertainties: model uncertainty and data uncertainty akin to epistemic and aleatoric uncertainty, respectively, in Bayesian analysis[20-22]. Model uncertainty (epistemic uncertainty) encompasses the uncertainty arising from deficiencies in the model, including errors in the training process, inadequacies in the model's architecture, or the absence of knowledge due to unknown samples. Data uncertainty (aleatoric uncertainty), on the other hand, is associated with uncertainty that directly originates from the data itself. This type of uncertainty arises due to the loss of information when representing the real world within a data sample, including observation noise and imperfect ground truth. To mitigate model uncertainty, researchers typically collect

more diversified training data or employ data augmentation techniques to expand the training dataset. Additionally, designing more robust network architectures and optimization algorithms can assist the networks in better handling aleatoric uncertainty.

II. Aleatoric uncertainty modelling

In the task of deep learning-based image SR reconstruction, the acquisition of ground truth, i.e., the GT-SIM data in this work, is a crucial step. The GT-SIM images are typically obtained by reconstructing raw SIM images acquired using high laser power and long camera exposure. However, even when ensuring that biological samples do not move during acquisition, the inherent quantum nature of fluorescence and imperfect photoelectric conversion of the camera lead to stochastic noise in each captured image and the corresponding ground truth image, contributing significantly to the aleatoric uncertainty in SR reconstructions of deep neural networks. Motion artifacts may also occur in GT-SIM images due to the movement of samples during exposure.

Inspired by previous work[23, 24], we assigned each pixel a Laplace distribution in the output SR image rather than a single intensity value:

$$p_{Laplace}(y;\hat{y},\hat{\sigma}) = \frac{1}{2\hat{\sigma}}\exp\left(-\frac{|y-\hat{y}|}{\hat{\sigma}}\right), \tag{15}$$

where $\hat{y}$ and $\hat{\sigma}$ denote the predicted value and scale of a certain pixel. In this way, the scale $\hat{\sigma}$ can be regarded as a measurement of the data uncertainty. Then the output SR image $\hat{y}$ and the scale $\hat{\sigma}$ can be simultaneously addressed by minimizing the negative log-likelihood (NLL) function:

$$\mathcal{L}_{NLL}(\hat{y},y) = \frac{1}{\mathrm{T}}\frac{1}{\mathrm{N}}\sum_{t=1}^{\mathrm{T}}\sum_{n=1}^{\mathrm{N}}\frac{|y_n^t-\hat{y}_n^t|}{\hat{\sigma}_n^t} + \log\hat{\sigma}_n^t\ , \tag{16}$$

where T denotes the number of timepoints and N denotes the number of pixels in a single image. Notably, with the scale parameters, the model has the flexibility to strike a balance between two extremes in minimizing the loss function above: achieving an accurate output prediction with low uncertainty or accommodating high uncertainty if an accurate output prediction is not feasible.

III. Epistemic uncertainty modelling

Compared to natural images, training data for microscopic images are more difficult to obtain. The diversity of biological specimens makes it challenging to collect a training dataset that encompasses all possible structures and fluorescent signal conditions. Furthermore, one of the primary objectives of biological imaging experiments is to

discover new biological structures and bioprocesses, rendering the approach of mitigating epistemic uncertainty through extensive data collection impractical. In such cases, SR neural networks may not adequately learn all possible biological structures, leading to inevitable epistemic uncertainty.

As previously discussed, we address aleatoric uncertainty by predicting a probability distribution, as opposed to a single scalar value. However, this method cannot capture uncertainty associated with the model parameters θ, which is induced by deficiencies in the model training procedure. For this purpose, we modified the DPA-TISR to a Bayesian neural network that employs a distribution over model parameters. Let {X, Y} represent the training dataset. The posterior distribution over model parameters, denoted as $\mathrm{p}(\theta|\mathrm{X},\mathrm{Y})$, can be expressed as follows:

$$\mathrm{p}(\theta|\mathrm{X},\mathrm{Y}) = \frac{\mathrm{p}(\mathrm{Y}|\theta,\mathrm{X})\mathrm{p}(\theta)}{\mathrm{p}(\mathrm{Y}|\mathrm{X})} \tag{17}$$

Given a test sample x^*, the restored image y^* can be predicted:

$$\mathrm{p}(\mathrm{y}^*|\mathrm{x}^*,\mathrm{X},\mathrm{Y}) = \int \mathrm{p}(\mathrm{y}^*|\mathrm{x}^*,\theta)\mathrm{p}(\theta|\mathrm{X},\mathrm{Y})\mathrm{d}\theta \tag{18}$$

While this equation is mathematically complete, the computation of $\mathrm{p}(\theta|\mathrm{X},\mathrm{Y})$ cannot be carried out analytically but can be approximated by variational parameters $\mathrm{q}(\theta)$. The objective is to approximate a distribution that closely resembles the posterior distribution obtained by the model. The degree of similarity between two distributions, measured by their KL divergence, can be expressed as follows:

$$\mathrm{KL}(\mathrm{q}(\theta)\|\mathrm{p}(\theta|X,Y)) = \int \log\frac{\mathrm{q}(\theta)}{\mathrm{p}(\theta|X,Y)}d\theta \tag{19}$$

KL divergence minimization can also be rearranged into the *evidence lower bound* (ELBO) maximization[25]:

$$L_{VI} \coloneqq \int \mathrm{q}(\theta)\log\mathrm{p}(\mathrm{Y}|\mathrm{X},\theta)\,d\theta - \mathrm{KL}\big(\mathrm{q}(\theta)\big\|\mathrm{p}(\theta|X,Y)\big), \tag{20}$$

where the first term is dedicated to fitting the data, while the second term for regularization is geared towards approaching the prior distribution. This methodology is commonly referred to as variational inference (VI). Among various approaches, Dropout VI[23, 26] stands out as one of the most prevalent methods, which uses dropout as a regularization term.

Leveraging the variational inference method, we can utilize a straightforward variational distribution $\mathrm{q}(\theta)$ by incorporating concrete dropout[23] before each weight

layer. This allows us to approximate the posterior distribution $\mathrm{p}(\theta|\mathrm{X},\mathrm{Y})$. By employing Monte Carlo (MC) integration over M samples that adhere to $\theta^{(\mathrm{m})}\sim\mathrm{q}(\theta)$, the equation above can be expressed as:

$$\mathrm{p}(y^*|x^*,\mathrm{X},\mathrm{Y}) \approx \int \mathrm{p}(y^*|x^*,\theta)\mathrm{q}(\theta)\mathrm{d}\theta \approx \frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p}\left(y^*\middle|x^*,\theta^{(\mathrm{m})}\right) \tag{21}$$

During the inference stage, the dropout layers in the Bayesian DPA-TISR randomly set input neurons to zero. By aggregating the outcomes of stochastic forward propagation through the trained model, the pixel-wise predictive mean can be computed as the super resolution result:

$$\hat{y}_{mean} = \frac{1}{\mathrm{M}}\sum_{m=1}^{\mathrm{M}} \hat{y}^{(m)}, \tag{22}$$

where $\hat{y}^{(m)}$ is the predicted mean from the m^{th} network $\theta^{(\mathrm{m})}$. The model uncertainty is then quantified by calculated the standard deviation of the predicted results:

$$\hat{\sigma}_{model} = \sqrt{\frac{1}{\mathrm{M}}\sum_{m=1}^{\mathrm{M}} (\hat{y}^{(m)} - \hat{y}_{mean})^2}\ . \tag{23}$$

Subsequently, the data uncertainty is assessed by computing the quadratic mean of the estimated variance:

$$\hat{\sigma}_{data} = \sqrt{\frac{1}{\mathrm{M}}\sum_{m=1}^{\mathrm{M}} (\hat{\sigma}^{(m)})^2}, \tag{24}$$

where $\hat{\sigma}^{(m)}$ is the predicted scale map from the m^{th} network $\theta^{(\mathrm{m})}$.

IV. Confidence calibration

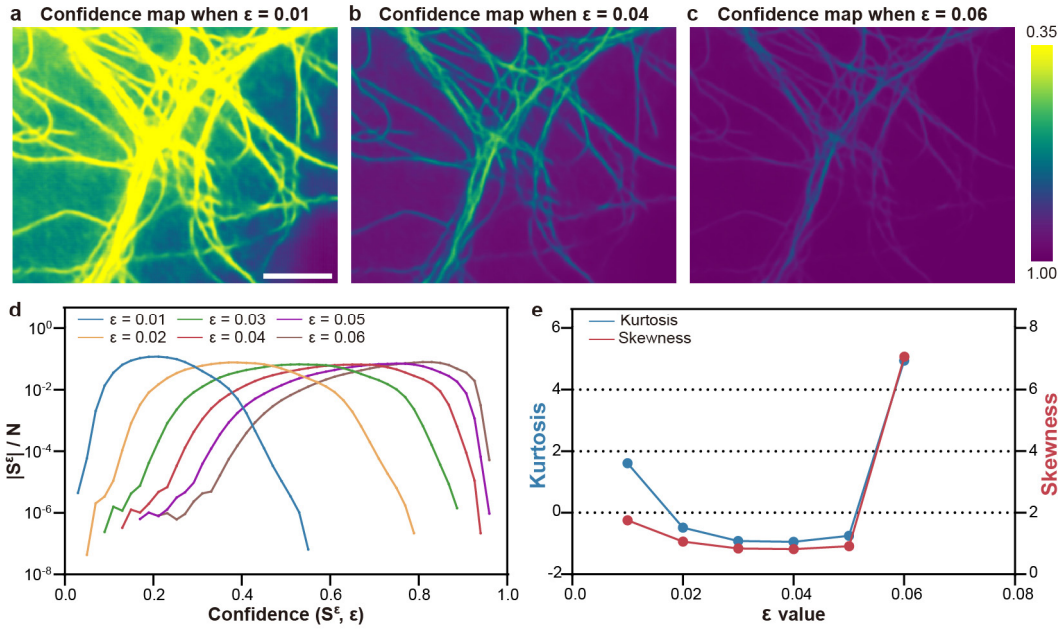
Confidence is deemed as well-calibrated when its value accurately reflects the actual probability of correctness. Consequently, to leverage uncertainty quantification methods effectively, it is imperative to ensure that the network is well calibrated. For regression tasks, the calibration can be defined such that predicted confidence intervals should match the accuracy intervals computed from the dataset, which can be effectively visualized through a reliability diagram. Notably, in calculation of confidence, accuracy, and reliability diagram following Eqs. (26)-(31) of Methods section in the main text, the credible interval length ε is an important scalar, which greatly influences the absolute value of confidence and accuracy, and fraction of pixels

that fall into each confidence range in reliability diagram. In our experiments, we chose the ε value based on two criteria. First, the ε should be a moderate value that is not too small or too large. A small value of ε results in a narrow credible interval, yielding generally low values cross the confidence map (Supplementary Fig. SN1a), while an overlarge ε will lead to a confidence map with unreasonably high values, e.g., very close to 1, with low discriminability across different pixels (Supplementary Fig. SN1c). We empirically found an ε value within [0.02, 0.05] could produce a relatively rational confidence map (Supplementary Fig. SN1b). Second, a good choice of ε should yield a uniform distribution of confidence ranging in [0, 1], so that there are enough pixels within each certain confidence range, which benefits both confidence discrimination and ECE calculation. To evaluate the uniformity of the confidence distribution under different ε values, we plotted the fraction of pixels that fall into each of the confidence ranges as shown in Supplementary Fig. SN1d. We further calculated the kurtosis and skewness to assess the uniformity of these curves using following formula:

$$Skewness = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{X_i - \mu}{\sigma}\right)^3\right], \tag{25}$$

$$Kurtosis = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{X_i - \mu}{\sigma}\right)^4\right], \tag{26}$$

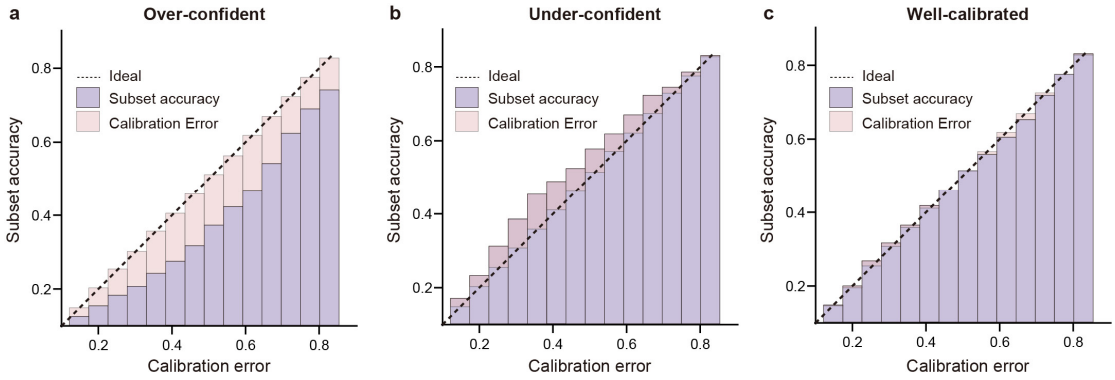
where μ and σ denote the mean value and standard deviation, respectively. Both analyses suggest an optical ε value of 0.03 or 0.04 that gives rise to the most uniform confidence distribution. Therefore, we selected 0.04 in all our experiments.



Supplementary Fig. SN1 | The selection of ε in confidence calculation. a-c, Representative confidence

maps calculated by Bayesian DPA-TISR using different credible interval length ε of 0.01 (a), 0.04 (b), and 0.06 (c), respectively. Scale bar, 3 μm. **d**, The fraction of pixels that fall into each of the confidence ranges with different ε selection of 0.01, 0.02, 0.03, 0.04, 0.05, and 0.06. **e**, Kurtosis and skewness values of the curves plotted in d, which suggests an optical ε value of 0.03 or 0.04 that gives rise to the most uniform confidence distribution.

Under a certain ε value, a DNN is considered under-confident when the predicted confidence is lower than the empirical accuracy (Supplementary Fig. SN2b), whereas it is regarded as over-confident when the predicted confidence exceeds the empirical accuracy (Supplementary Fig. SN2a). Well-calibrated metrics should yield credibility values similar to accuracy, resulting in a reliability diagram that is diagonal (Supplementary Fig. SN2c).



Supplementary Figure SN2 | Typical reliability diagrams of DNN. a, A representative reliability diagram of an over-confident model, where the subset accuracy is smaller than the corresponding confidence. **b**, A representative reliability diagram of an under-confident model, where the subset accuracy is larger than the corresponding confidence. **c**, A representative reliability diagram of a well-calibrated model, of which the subset accuracy is similar to the corresponding confidence.

Previous studies have highlighted that deeper networks often exhibit a tendency towards greater overconfidence compared to shallower ones[27]. After applying Bayesian neural networks to experimental data spanning multiple biological structures, we have similarly observed a propensity for estimated confidence levels to lean towards overconfidence (Extended Data Figs. 8, Supplementary Figs. 15, 16). In addressing this confidence mismatch with the actual accuracy, various approaches have been employed, including regularization methods and post-processing techniques[27]. Post-processing methods typically necessitate a separate calibration dataset for calibration, and their effectiveness can be contingent upon the size of the validation dataset. Conversely, regularization methods entail modifications to the objective, optimization, and/or regularization procedures to develop inherently calibrated DNNs. In the context of

biological image restoration tasks, our confidence correction method is structured upon regularization methods as is described in the Methods section of the main manuscript.

8. Rolling Fourier ring correlation analysis

In conjunction with the confidence generated by Bayesian DPA-TISR deep neural network, the Fourier ring correlation (FRC)[28] was developed to evaluate the overall effective resolution of a fluorescence image by characterizing the highest reliable cut-off frequency in the Fourier domain. Based on FRC analysis, rolling Fourier ring correlation (rFRC)[29] was recently developed to provide the local distance measurements at the pixel level by rolling a sliding window across the image, which offers insights into the resolvability and uncertainty.

In our manuscript, we conducted the rFRC analysis as a comparative method in generating confidence maps for TISR images. It was applied following the instructions of the original paper. First, an TISR image was down-sampled into two independent frames of identical contents with the same sampling rate by pixel shuffle, which is a commonly used operation to augment the raw data in FRC analysis. Next, the input images were padded symmetrically around a half size of the block to ensure accurate FRC calculation at the image boundaries. Then a background threshold was set for the center pixels to avoid FRC calculation in background areas. If the mean of the center pixels exceeded the threshold, the FRC was computed within a 32-pixel size block (which is the default size of the software), and the resulting FRC resolution was assigned to the center pixel of each block. Conversely, if the mean was below the threshold, a zero value was assigned to the central pixel. This process was repeated block by block across the entire image to generate a final rFRC confidence map.

Supplementary Tables

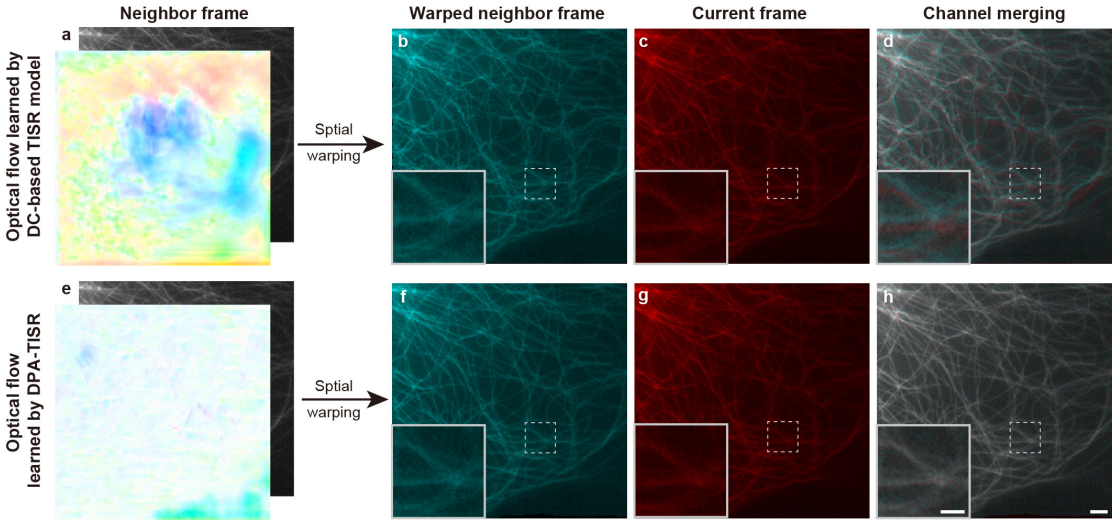
Supplementary Table 1 | Imaging conditions of BioTISR dataset

Data type	Structure	Imaging method (Acquisition mode)	Sample	Label	Excitation NA	Excitation λ (nm)	Exposure time per raw image (ms)	Total acquisition time per cycle (sec)	Volume size of raw data (X×Y×Z)	Cycle time (Acquisition + resting time) (sec)	Timepoints per ROI
2D	Lyso	GI-SIM	COS-7	Lamp1-Halo	1.35	561	2	~0.05	512 × 512	0.25	20
	Mito	GI-SIM	COS-7	2×mEmerald-Tomm20	1.35	488	2	~0.05	512 × 512	0.25	20
	MTs	GI-SIM	COS-7	3×mEmerald-Ensconsin	1.35	488	5	~0.14	512 × 512	0.5	20
	CCPs	TIRF-SIM	SUM-159	Clathrin-mEmerald	1.41	488	5	~0.14	512 × 512	1	20
	F-actin	TIRF-SIM	COS-7	Lifeact-mEmerald	1.41	488	5	~0.14	512 × 512	1	20
	F-actin	NL TIRF-SIM	COS-7	Lifeact-SkylanNS	1.41	488	5	~0.23	512 × 512	5	10
3D	MTs	3D-SIM	COS-7	3×mEmerald-Ensconsin	1.41	488	1	1.4 ~ 3.0	512 × 512 × (8 ~ 22)	1.4 ~ 3.0	10
	Mito	3D-SIM	COS-7	2×mEmerald-Tomm20	1.41	488	1	1.4 ~ 3.0	512 × 512 × (8 ~ 22)	1.4 ~ 3.0	10
	F-actin	3D-SIM	COS-7	Lifeact-mEmerald	1.41	488	1	1.4 ~ 3.0	512 × 512 × (8 ~ 22)	1.4 ~ 3.0	10

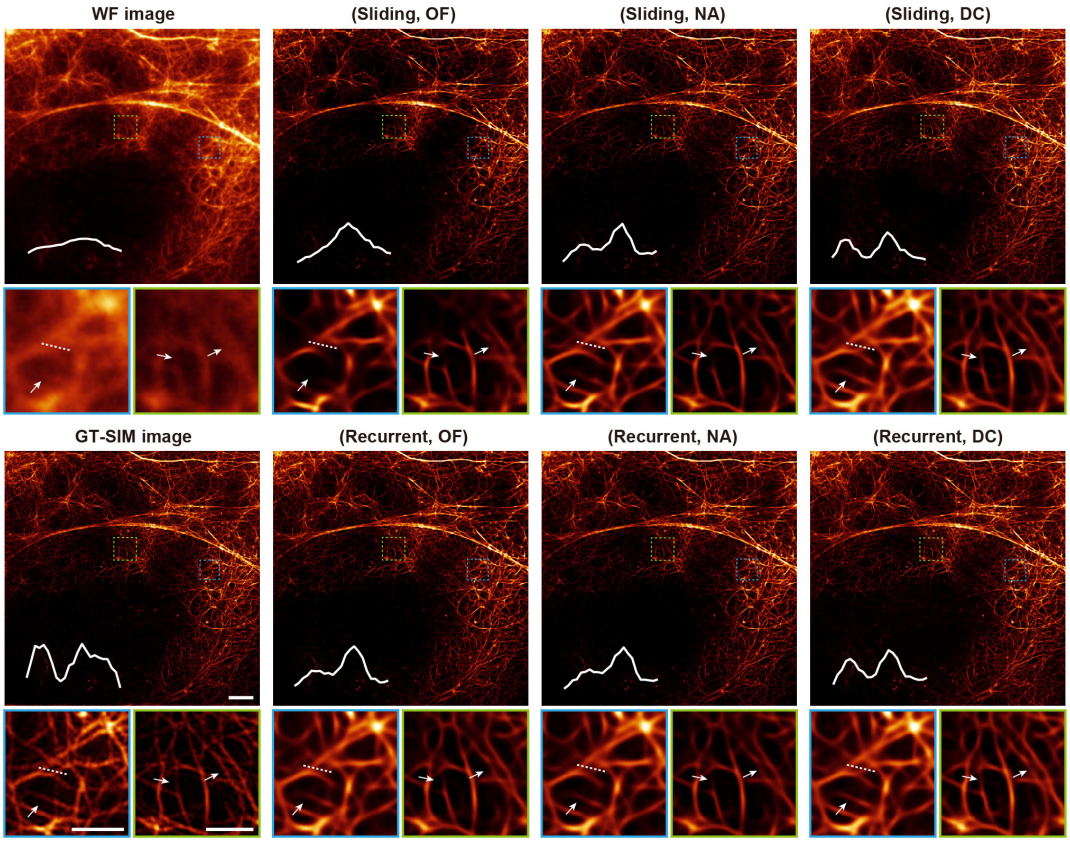
Supplementary Table 2 | Imaging conditions of live-cell experiments

Data	Imaging method	Sample	Label	Excitation NA	Excitation λ (nm)	Exposure time per raw image (ms)	Total acquisition time per cycle (sec)	Volume size of raw data (X×Y×C)	Cycle time (Acquisition + resting time) (sec)	Timepoints
Fig. 3a-e Supplementary Video 1	TIRF microscopy	COS-7	Lifeact-mEmerald Clathrin-mCherry	1.41	488 561	488: 5 561: 20	~0.03	768 × 768 × 2	1	5000
Fig. 3f-i Supplementary Video 2	GI microscopy	COS-7	TFAM-mEmerald PKMO-Halo	1.35	488 561	5	~0.01	1536 × 1536 × 2	0.33	5000
Fig. 5 Supplementary Video 4 Supplementary Video 5	GI-SIM	COS-7	TOMM20-mEmerald PMP-Halo	1.35	488 561	488: 5 561: 2	~0.06	768 × 768 × 2	1	1505
Extended Data Fig. 9 Supplementary Fig. 17 Supplementary Video 3	GI microscopy	COS-7	TOMM20-mEmerald Lamp1-Halo	1.35	488 561	20	~0.04	768 × 768 × 2	0.5	10000
Extended Data Fig. 10 Supplementary Video 6	GI microscopy	COS-7	TOMM20-mEmerald PMP-Halo	1.35	488 561	5	~0.01	1024 × 1024 × 2	0.25	12000

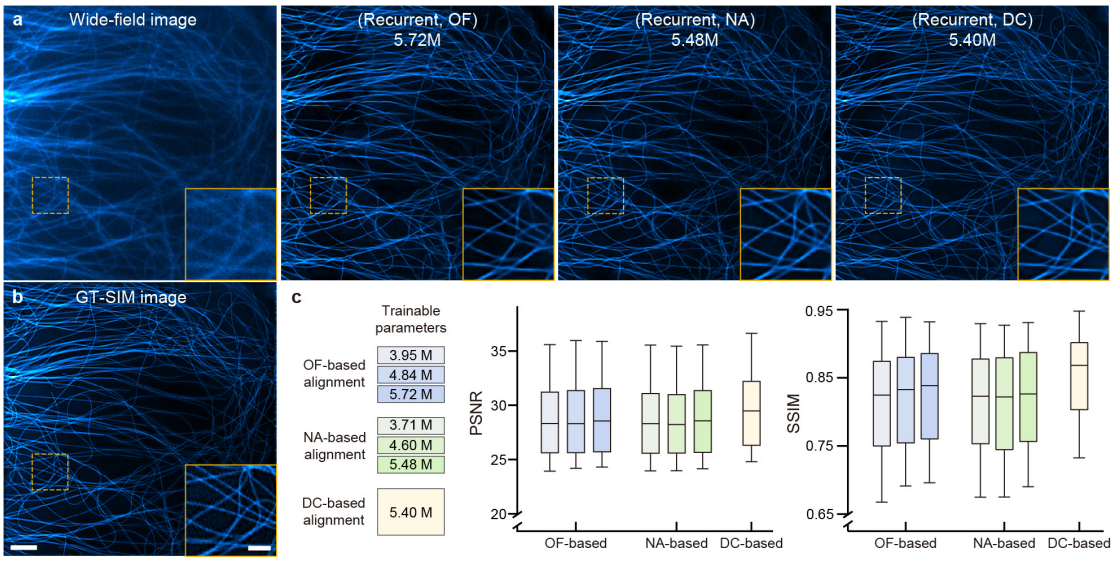
Supplementary Figures



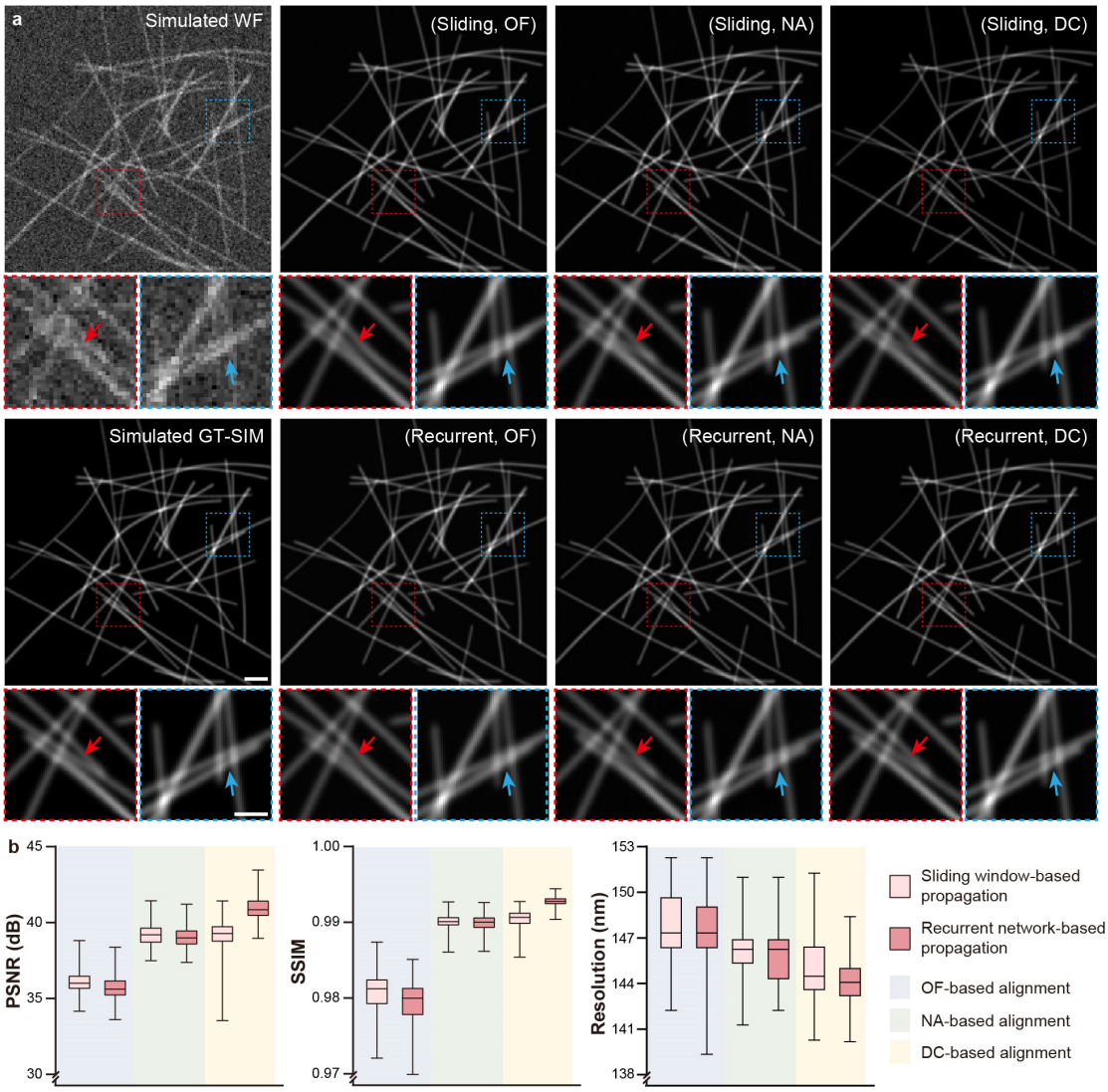
Supplementary Fig. 1 | Visualization of optical flow-based spatial warping. a, Wide-field neighbor frame of microtubules and corresponding optical flow relative to the current frame generated via a deformable convolution-based TISR model. **b**, The warped neighbor frame using the optical flow shown in a. **c**,**d**, Channel merging visualization (d) of the warped neighbor frame (b) and current frame (c). **e**, Wide-field neighbor frame of microtubules and corresponding optical flow relative to the current frame generated via the DPA-TISR model. **f**, The warped neighbor frame using the optical flow shown in e. **g**,**h**, Channel merging visualization (h) of the warped neighbor frame (f) and current frame (g). These results illustrate that the phase-space alignment mechanism in the DPA-TISR model can synergize with optical flow calculation, resulting in more accurate and finer feature alignment. Scale bar: 3 μm, 1 μm (zoom-in regions). Experiments in (a-d) and (e-h) were repeated independently each with 20 test images, all achieving similar results.



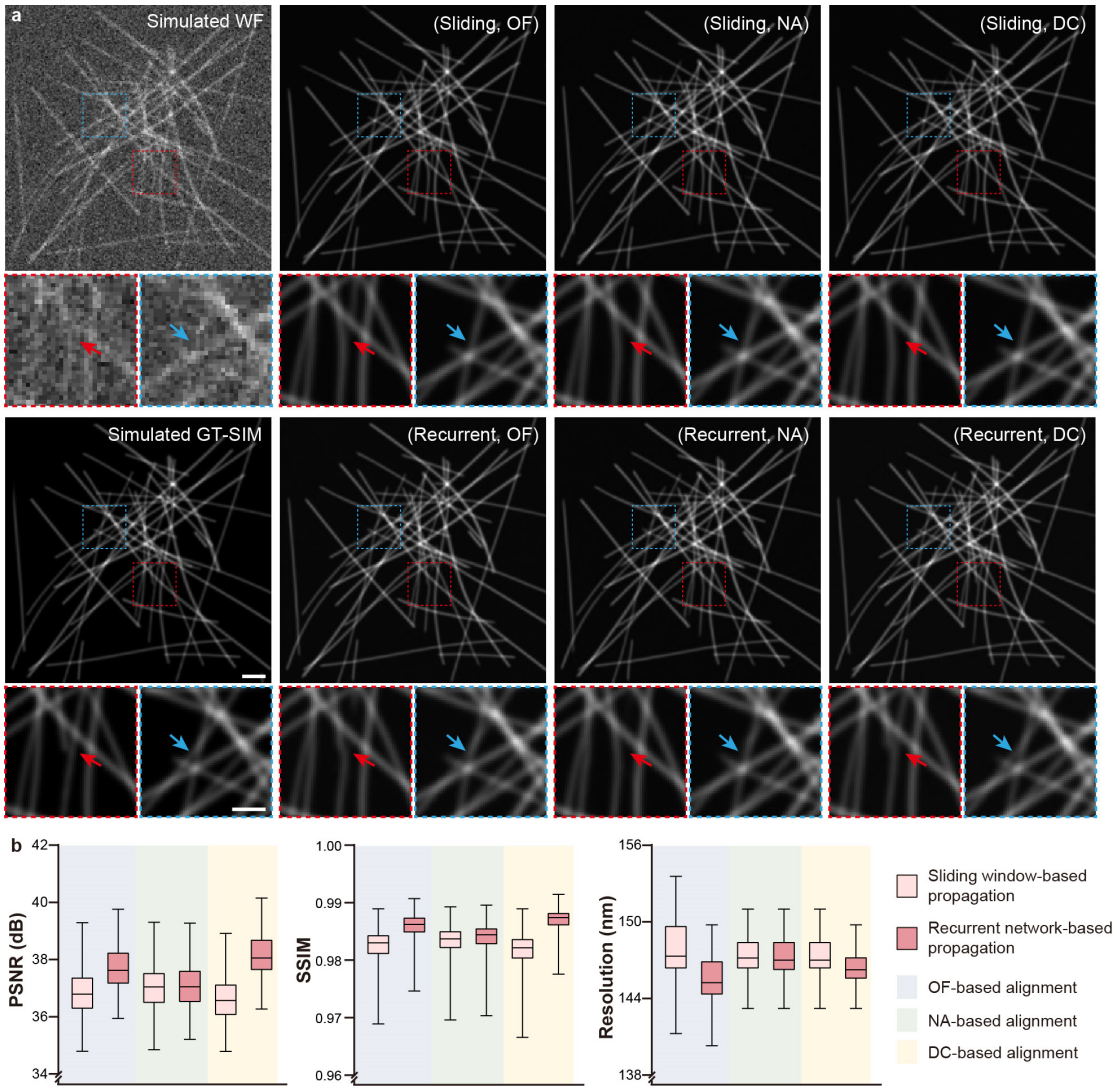
Supplementary Fig. 2 | Comparison of representative propagation and alignment mechanisms in TISR models with nonlinear SIM dataset of F-actin. Representative TISR images inferred by six models combined by two propagation methods, sliding window-based propagation and recurrent network-based propagation, and three alignment mechanisms based on optical flow (OF), non-local attention (NA) and deformable convolution (DC). All models are configured to up-scaling the input image by 3-fold and trained with the nonlinear SIM F-actin dataset in BioTISR. WF and nonlinear GT-SIM images are shown in the first column for reference. The intensity profiles along the white dotted lines in magnified images are shown on the bottom left corner of each image. Scale bar, 3 μm, 1 μm (zoom-in regions).



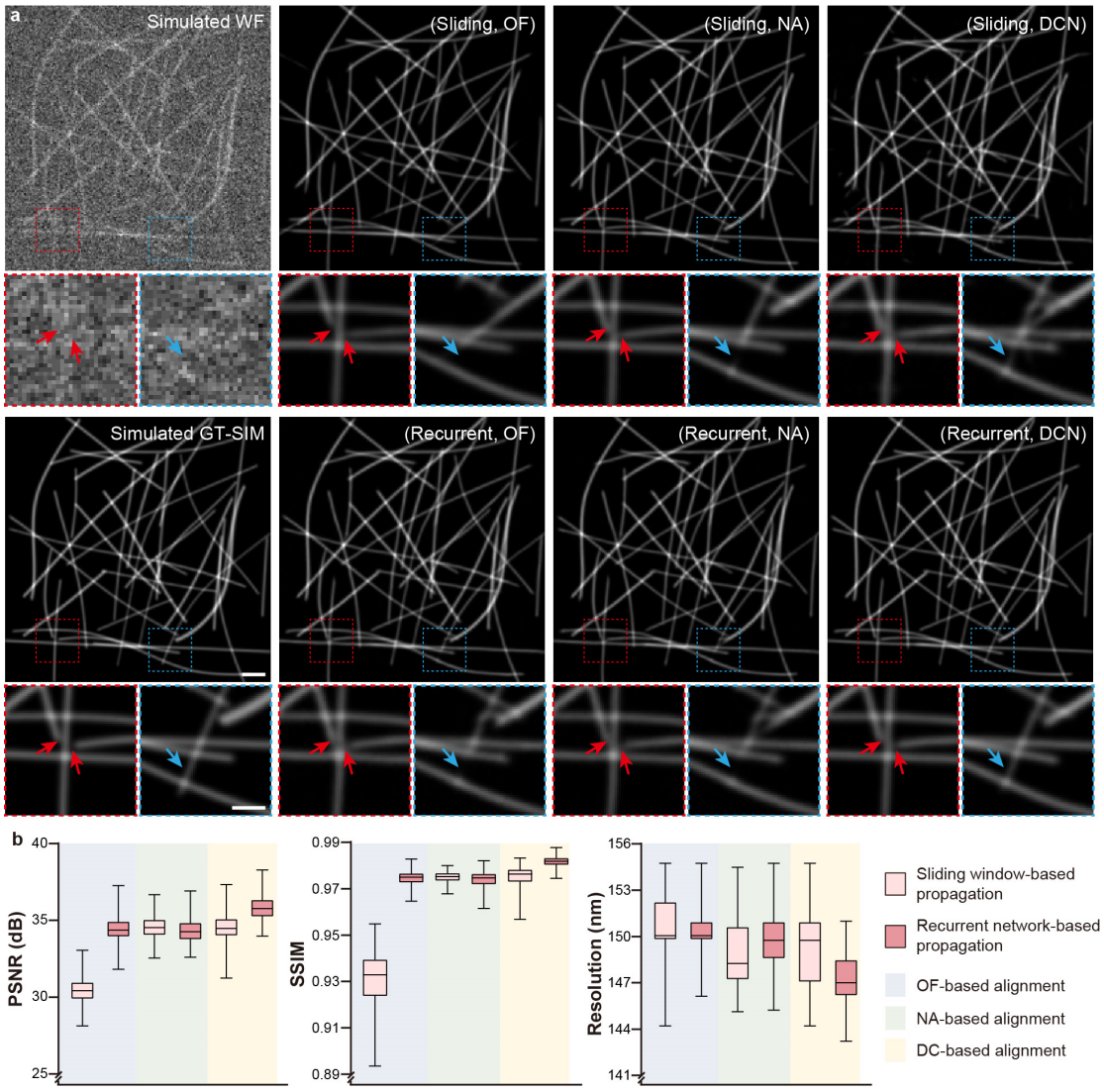
Supplementary Fig. 3 | Performance progression of TISR models constructed with optical flow-based and non-local attention-based alignment as their trainable parameters increase. a, Representative wide-field image of microtubules and its SR prediction via TISR models constructed by recurrent network-based propagation and alignment based on optical flow (OF, the second column), non-local attention (NA, the third column), and deformable convolution (DC, the fourth column) with similar trainable parameters of 5.72M, 5.48M, and 5.40M, respectively. **b**, GT-SIM image of the same region for reference. Scale bar, 3 μm, 1 μm (zoom-in regions). **c**, PSNR (left) and SSIM (right) evaluation of TISR models constructed with optical flow-based and non-local attention-based alignment as their trainable parameters increase (n=50). The box plots of PSNR and SSIM for DC-based TISR model are shown for reference. Center line, medians; limits, 75% and 25%; whiskers, the larger value between the largest data point and the 75th percentiles plus 1.5× the interquartile range (IQR), and the smaller value between the smallest data point and the 25th percentiles minus 1.5× the IQR. These results indicate that the parameter increase in the alignment module do not substantially influence the overall SR performance of TISR models, of which the reason we speculated is that without an effective image alignment mechanism, TISR models cannot fully exploit and benefit from the temporal dependency between adjacent frames despite of the model scale, which further underline the importance of developing an efficient alignment mechanism in the TISR neural network model.



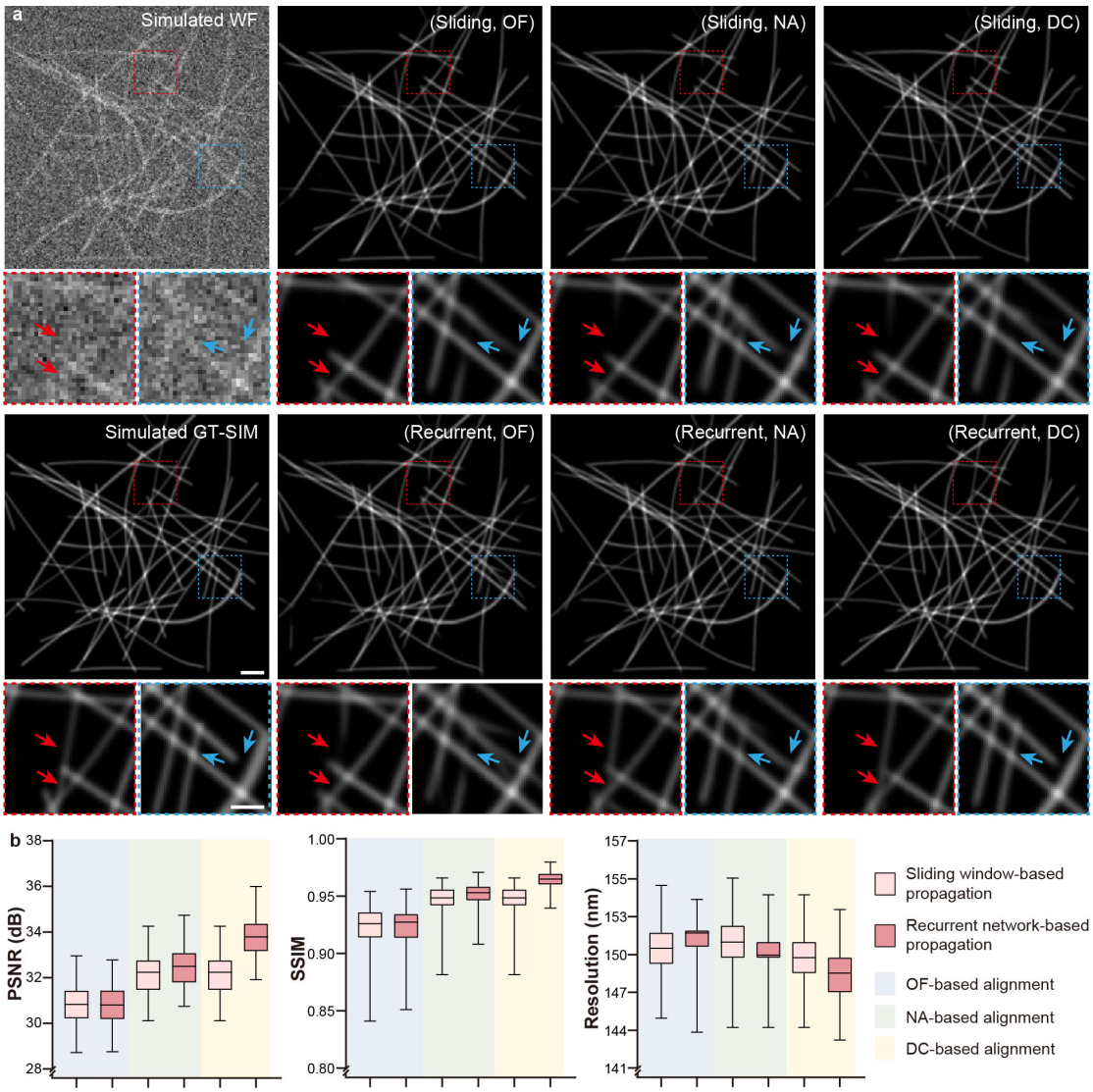
Supplementary Fig. 4 | Comparison of representative propagation and alignment mechanisms in TISR models with simulated data of high SNR ($\alpha = 20$) and low framerate (3 Hz). a, Representative TISR images of simulated tubular structures inferred by six models combined by two propagation methods, sliding window-based propagation and recurrent network-based propagation, and three alignment mechanisms based on optical flow (OF), non-local attention (NA) and deformable convolution (DC). WF and GT-SIM images are shown in the first column for reference. **b**, Statistical comparison of the six models in terms of PSNR, SSIM, and resolution quantified by decorrelation analysis[31] (n=200). Center line, medians; limits, 75% and 25%; whiskers, maximum and minimum. Scale bar, 1 μm (a), 0.5 μm (zoom-in regions in a).



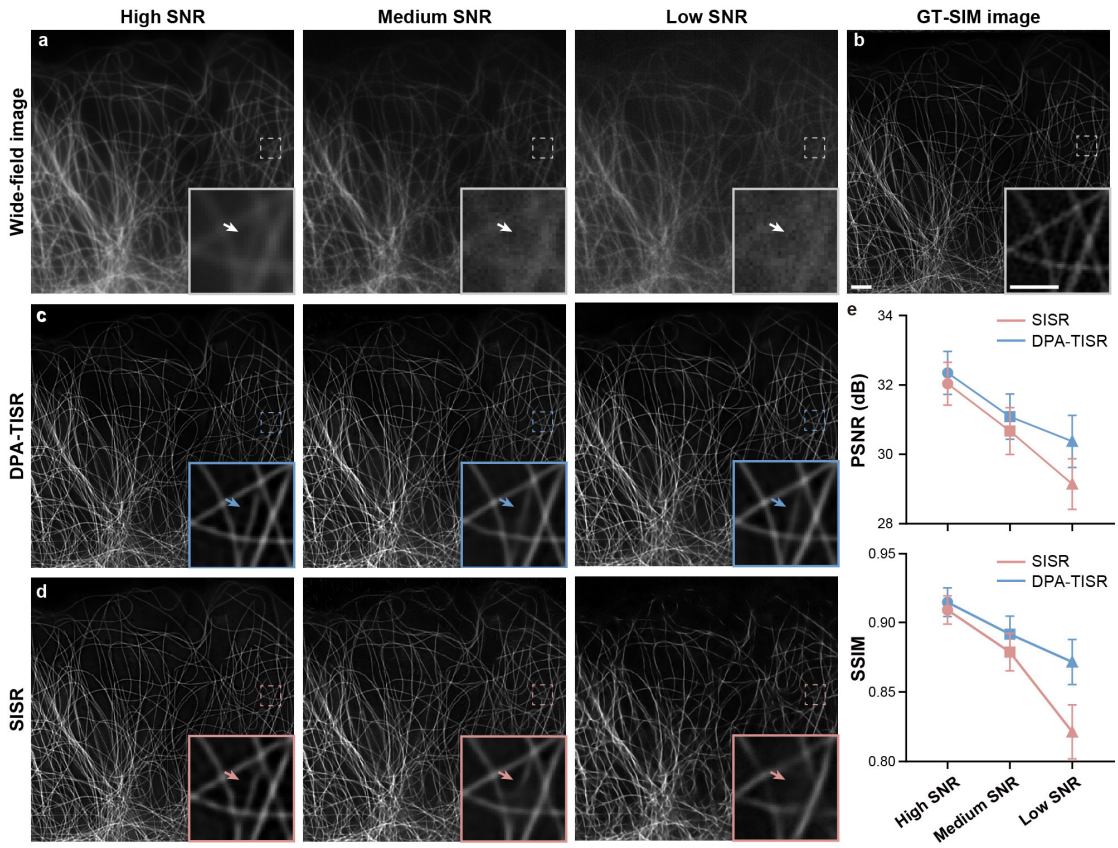
Supplementary Fig. 5 | Comparison of representative propagation and alignment mechanisms in TISR models with simulated data of high SNR ($\alpha = 20$) and high framerate (30 Hz). a, Representative TISR images of simulated tubular structures inferred by six models combined by two propagation methods, sliding window-based propagation and recurrent network-based propagation, and three alignment mechanisms based on optical flow (OF), non-local attention (NA) and deformable convolution (DC). WF and GT-SIM images are shown in the first column for reference. **b**, Statistical comparison of the six models in terms of PSNR, SSIM, and resolution quantified by decorrelation analysis[31] (n=200). Center line, medians; limits, 75% and 25%; whiskers, maximum and minimum. Scale bar, 1 μm (a), 0.5 μm (zoom-in regions in a).



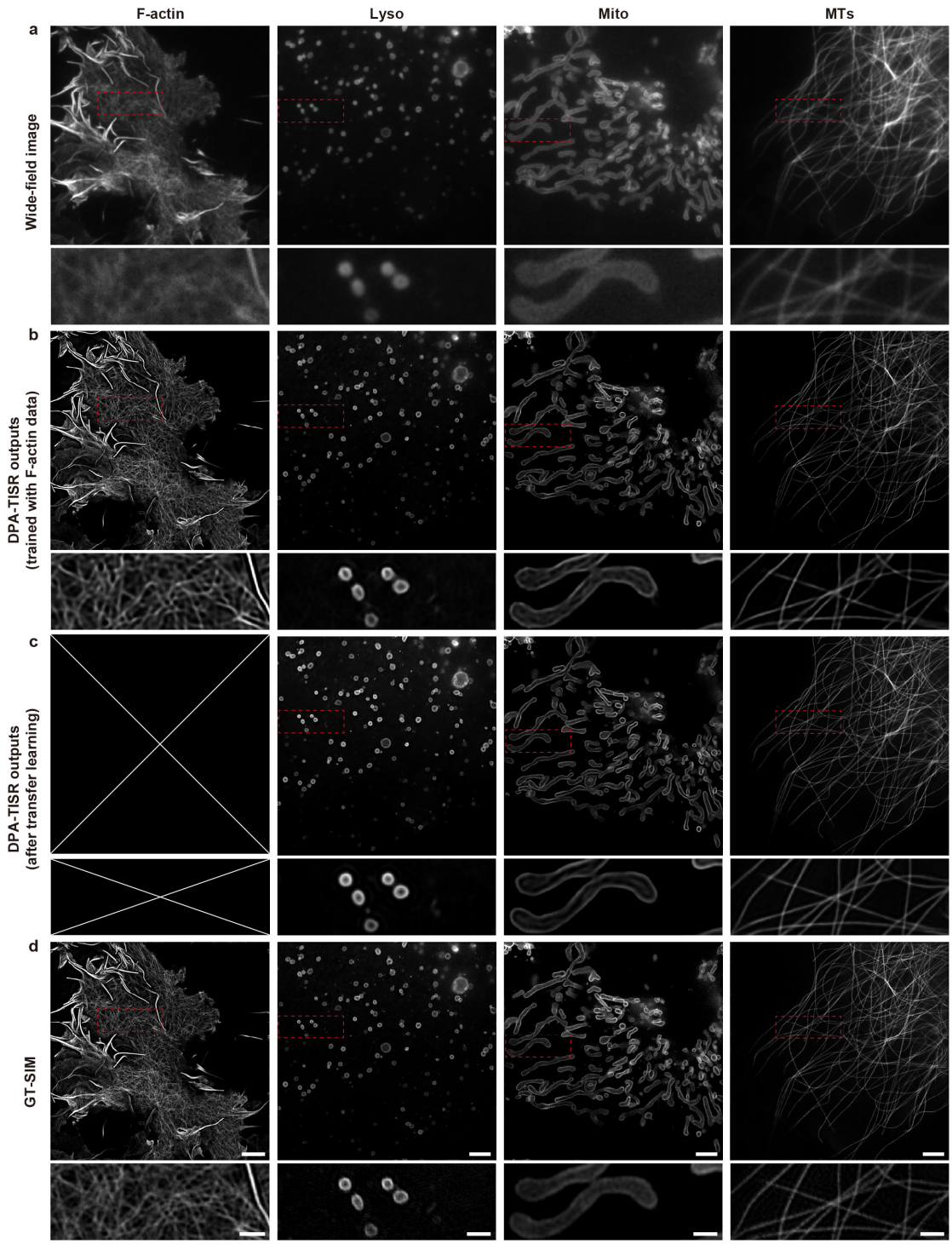
Supplementary Fig. 6 | Comparison of representative propagation and alignment mechanisms in TISR models with simulated data of low SNR ($\alpha = 5$) and low framerate (3 Hz). a, Representative TISR images of simulated tubular structures inferred by six models combined by two propagation methods, sliding window-based propagation and recurrent network-based propagation, and three alignment mechanisms based on optical flow (OF), non-local attention (NA) and deformable convolution (DC). WF and GT-SIM images are shown in the first column for reference. **b**, Statistical comparison of the six models in terms of PSNR, SSIM, and resolution quantified by decorrelation analysis[31] (n=200). Center line, medians; limits, 75% and 25%; whiskers, maximum and minimum. Scale bar, 1 μm (a), 0.5 μm (zoom-in regions in a).



Supplementary Fig. 7 | Comparison of representative propagation and alignment mechanisms in TISR models with simulated data of low SNR ($\alpha = 5$) and high framerate (30 Hz). a, Representative TISR images of simulated tubular structures inferred by six models combined by two propagation methods, sliding window-based propagation and recurrent network-based propagation, and three alignment mechanisms based on optical flow (OF), non-local attention (NA) and deformable convolution (DC). WF and GT-SIM images are shown in the first column for reference. **b**, Statistical comparison of the six models in terms of PSNR, SSIM, and resolution quantified by decorrelation analysis[31] (n=200). Center line, medians; limits, 75% and 25%; whiskers, maximum and minimum. Scale bar, 1 μm (a), 0.5 μm (zoom-in regions in a).

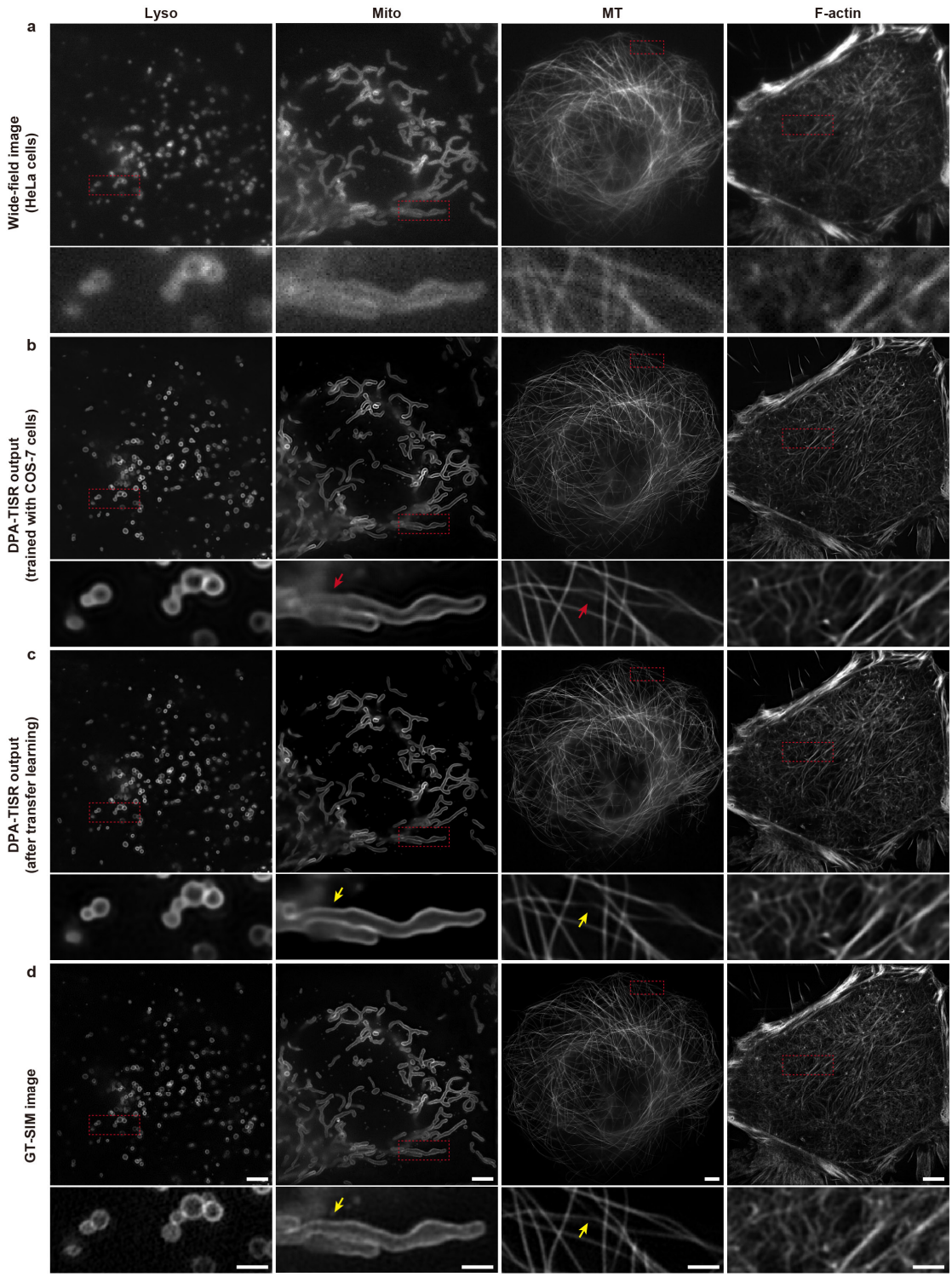


Supplementary Fig. 8 | Comparison of DPA-TISR and SISR models with microtubule images under different SNR conditions. a, Wide-field images of high (first column), medium (second column) and low SNR (third column) from the BioTISR dataset. **b**, GT-SIM image of the same region shown in a. **c**,**d**, SR results inferred from WF images via DPA-TISR (c) and its SISR (d) variant (see Methods section in the main manuscript for details of the modification). **e,** Statistical comparison of DPA-TISR and SISR models in terms of PSNR (upper panel) and SSIM (lower panel) calculated at 3 levels of SNR for microtubule data (n=50). Central point, medians; whiskers, mean±standard deviation. Scale bar: 3 μm, 1 μm (zoom-in regions).



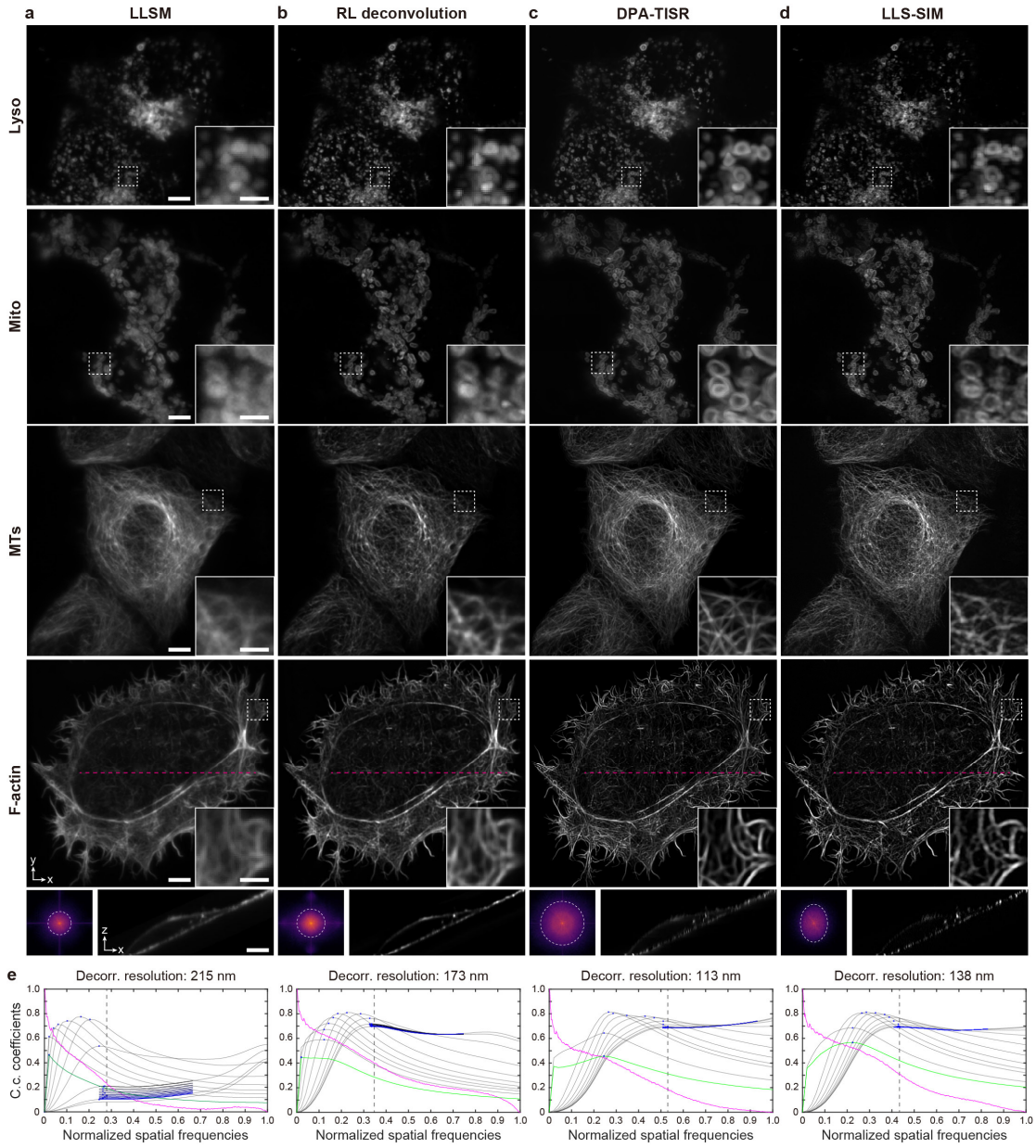
Supplementary Fig. 9 | Generalization and transfer learning capability of DPA-TISR across different organelles. **a**, Representative wide-field images of F-actin, lysosomes (Lyso), outer mitochondrial membrane (Mito), and microtubules (MTs). **b**, The SR images inferred by the same DPA-TISR network, which is trained with only F-actin images in BioTISR dataset. **c**, SR images reconstructed by the DPA-TISR networks after transfer learning for each type of specimens, which are initialized by the DPA-TISR model trained with only F-actin data. **d**, The ground truth SIM images shown for reference. Scale bar, 3 μm (a-d), 1 μm (magnified regions in a-d). Refer to Supplementary Note 4 for more

discussion on the generalization ability of DPA-TISR. Experiments were repeated with more than 50 test images for each type of specimens, achieving similar results.

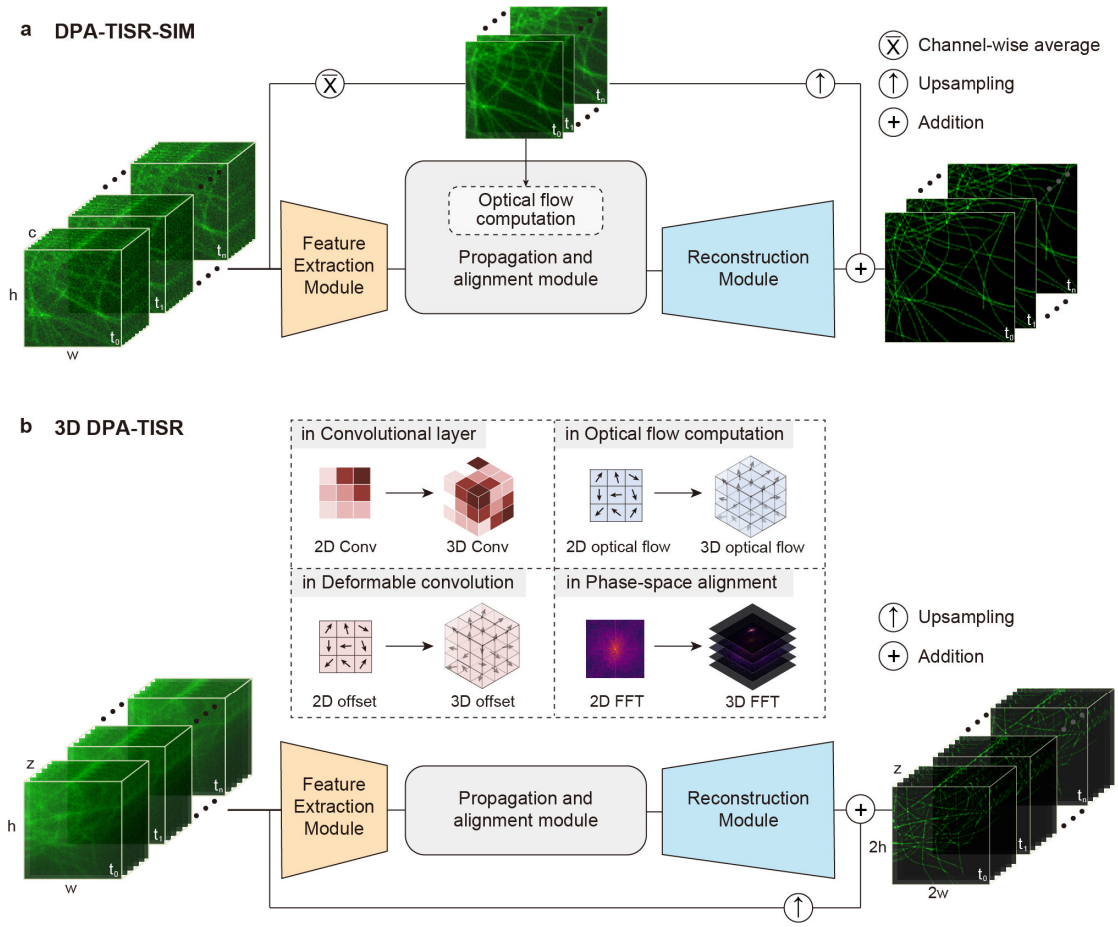


Supplementary Fig. 10 | Generalization and transfer learning capability of DPA-TISR across different cell types. a, Representative wide-field images of lysosomes (Lyso), outer mitochondrial membrane (Mito), microtubules (MTs), and F-actin acquired in HeLa cells. **b**, The SR images inferred by DPA-TISR models, which are trained with images of corresponding biological structures in COS-7 cells. **c**, The SR images reconstructed by the DPA-TISR networks after transfer learning using data acquired from HeLa cells, which are initialized by the DPA-TISR models used in b. **d**, The ground truth

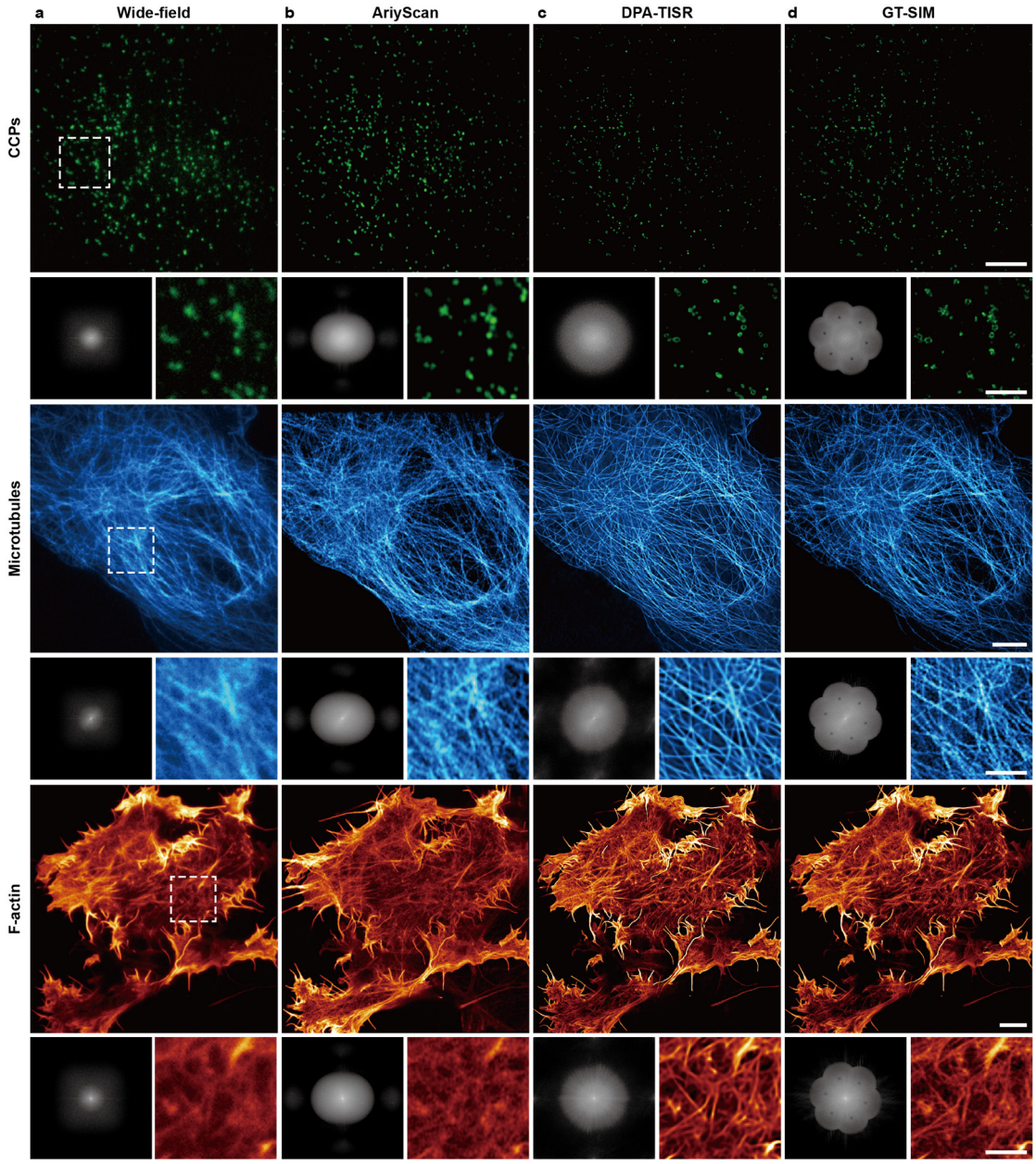
SIM images shown for reference. Scale bar, 3 μm (a-d), 1 μm (magnified regions in a-d). Refer to Supplementary Note 4 for more discussion on the generalization ability of DPA-TISR. Experiments were repeated with more than 50 test images for each type of specimens, achieving similar results.



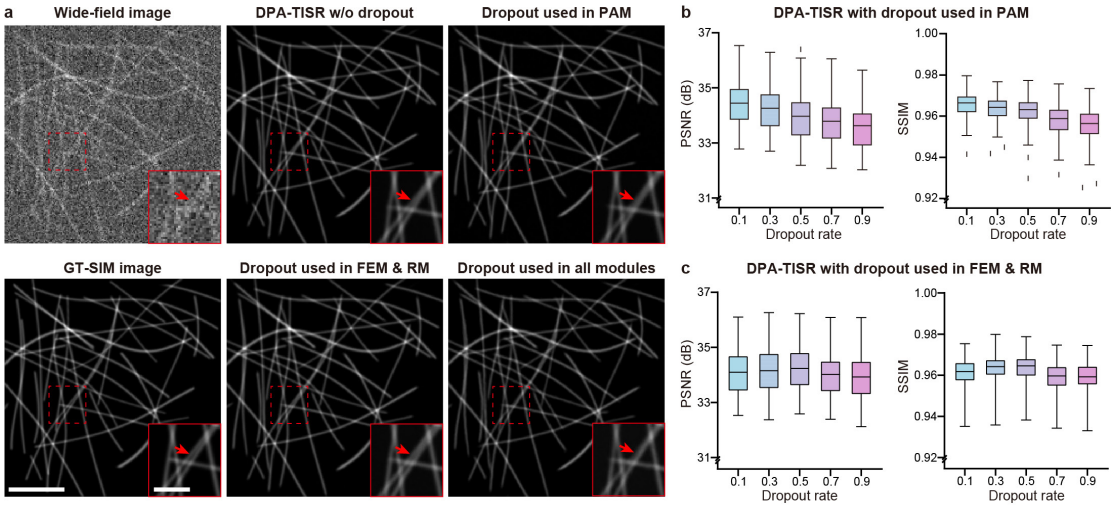
Supplementary Fig. 11 | Generalization capability of DPA-TISR across different imaging modalities. a, Representative images (maximum intensity projection, MIP) of lysosomes (Lyso), outer mitochondrial membrane (Mito), microtubules (MTs), and F-actin acquired with a lattice light-sheet microscopy system. **b**, The deconvolved results (MIP) of images shown in a using the RL deconvolution algorithm (10 iterations). **c**, The SR images inferred by the DPA-TISR networks trained on BioTISR dataset, which is acquired with TIRF-SIM and GI-SIM modes of our Multi-SIM system. **d**, The lattice light-sheet structured illumination microscopy (LLS-SIM) images (MIP) of the same regions shown for reference. Scale bar, 3 μm (a-d), 1 μm (magnified regions in a-d). **e**, Decorrelation analysis of the F-actin images in a-d, demonstrating the resolution improvement over the diffraction-limit by RL deconvolution, DPA-TISR, and LLS-SIM methods. Refer to Supplementary Note 4 for more discussion on the generalization ability of DPA-TISR. Experiments were repeated with more than 50 test images for each type of specimens, achieving similar results.



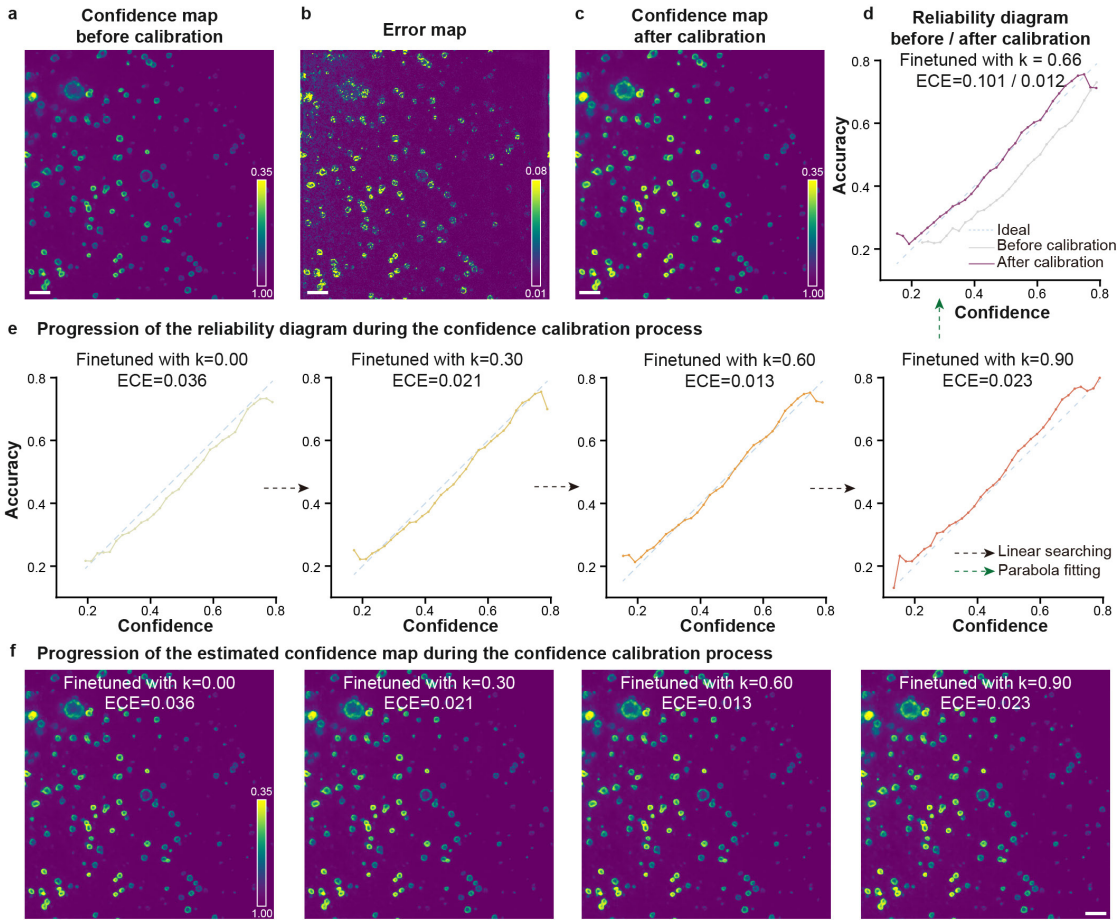
Supplementary Fig. 12 | Schematic of the network architecture of DPA-TISR-SIM and 3D DPA-TISR. a, Schematic of the network architecture of DPA-TISR-SIM. b, GT-SIM image of the same region shown in a. **b**, Schematic of the modifications from DPA-TISR to 3D DPA-TISR.



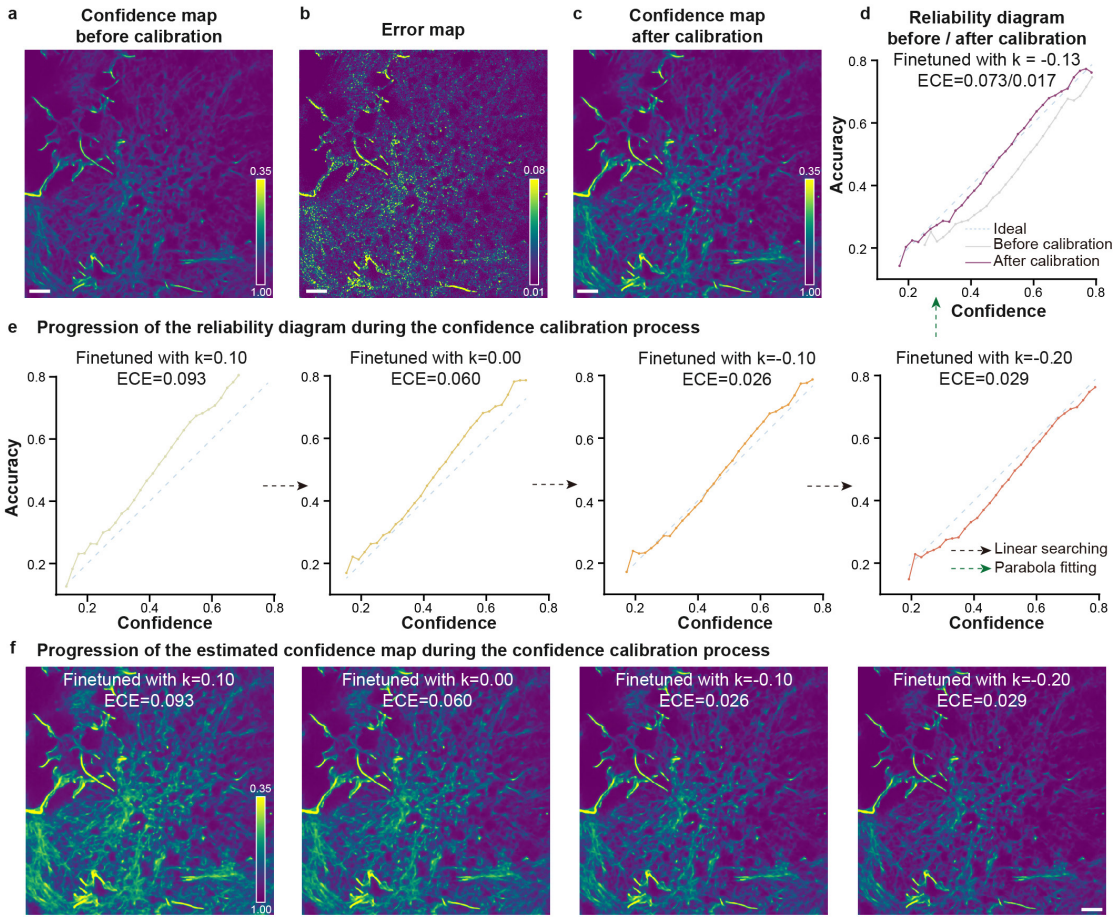
Supplementary Fig. 13 | Comparison of SR images generated by DPA-TISR and AiryScan. a, Representative images of clathrin-coated pits (CCPs), microtubules (MTs), and F-actin acquired with the wide-field mode of the Multi-SIM system. **b**, SR images of the same regions shown in a acquired and reconstructed using the AiryScan system and concomitant reconstruction algorithm (Zeiss). **c**, The SR images generated by the DPA-TISR networks trained on data of corresponding biological structures. **d**, The SR-SIM images obtained with the SIM mode of the Multi-SIM system and reconstructed with the conventional SIM reconstruction algorithm. Scale bar, 5 μm (a-d), 2 μm (magnified regions in a-d). Gamma value, 0.9 for SR images of microtubules, 0.7 for SR images of F-actin. Experiments in (a-d) were repeated with 10 test images for each type of specimens, all achieving similar results.



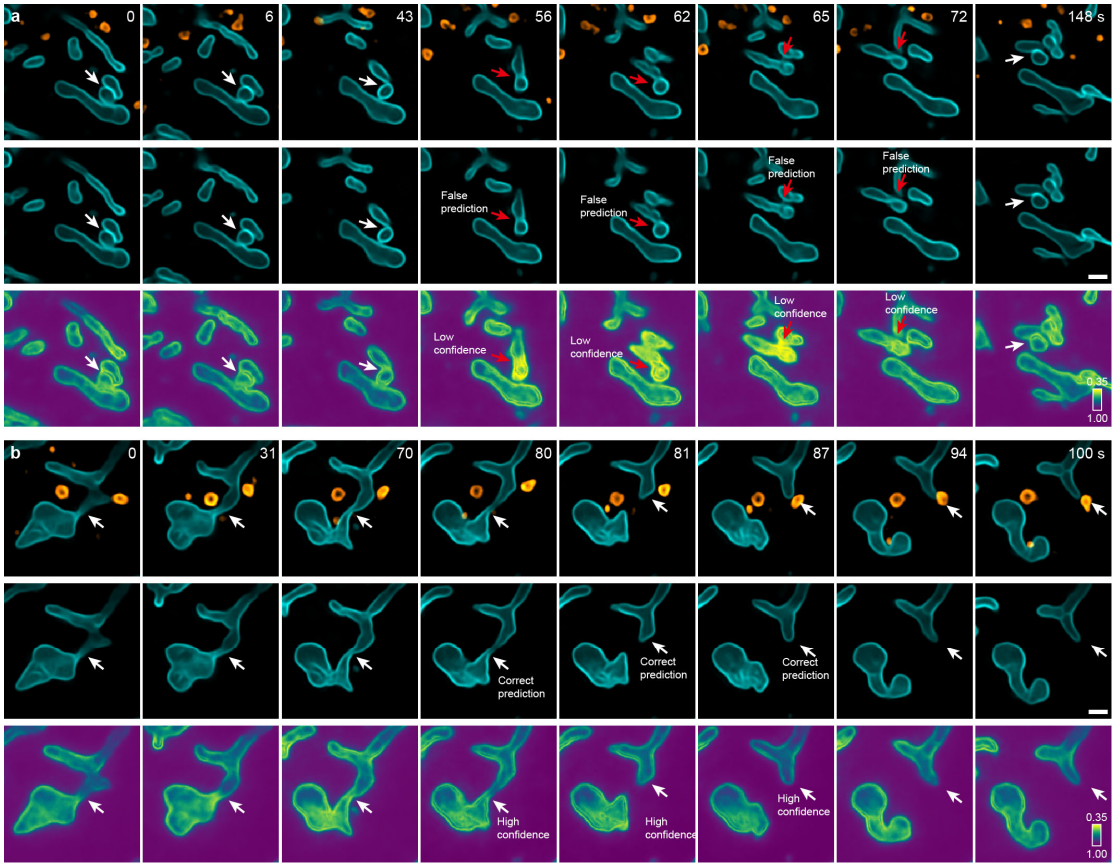
Supplementary Fig. 14 | Comparison of DPA-TISR models with dropout used in different modules. **a**, Representative wide-field image of simulated tubular structure and SR images generated by DPA-TISR without dropout, DPA-TISR with dropout used in the propagation and alignment module (PAM), DPA-TISR with dropout used in the feature extraction module (FEM) and reconstruction module (RM), and DPA-TISR with dropout used in all three modules. The GT-SIM image is provided for reference. Scale bar, 3 μm, 1 μm (zoom-in regions). **b**, PSNR and SSIM evaluation for DPA-TISR models with dropout used in FEM and RM as the dropout rate increases (n=200), indicating that moderate usage of dropout in FEM and RM benefits the overall SR performance of DPA-TISR. **c**, PSNR and SSIM evaluation for DPA-TISR models with dropout used in PAM as the dropout rate increases (n=200), revealing that the introduction of dropout in PAM hinders SR capability of the DPA-TISR model. Center line, medians; limits, 75% and 25%; whiskers, the larger value between the largest data point and the 75th percentiles plus 1.5× the interquartile range (IQR), and the smaller value between the smallest data point and the 25th percentiles minus 1.5× the IQR; outliers, data points larger than the upper whisker or smaller than the lower whisker.



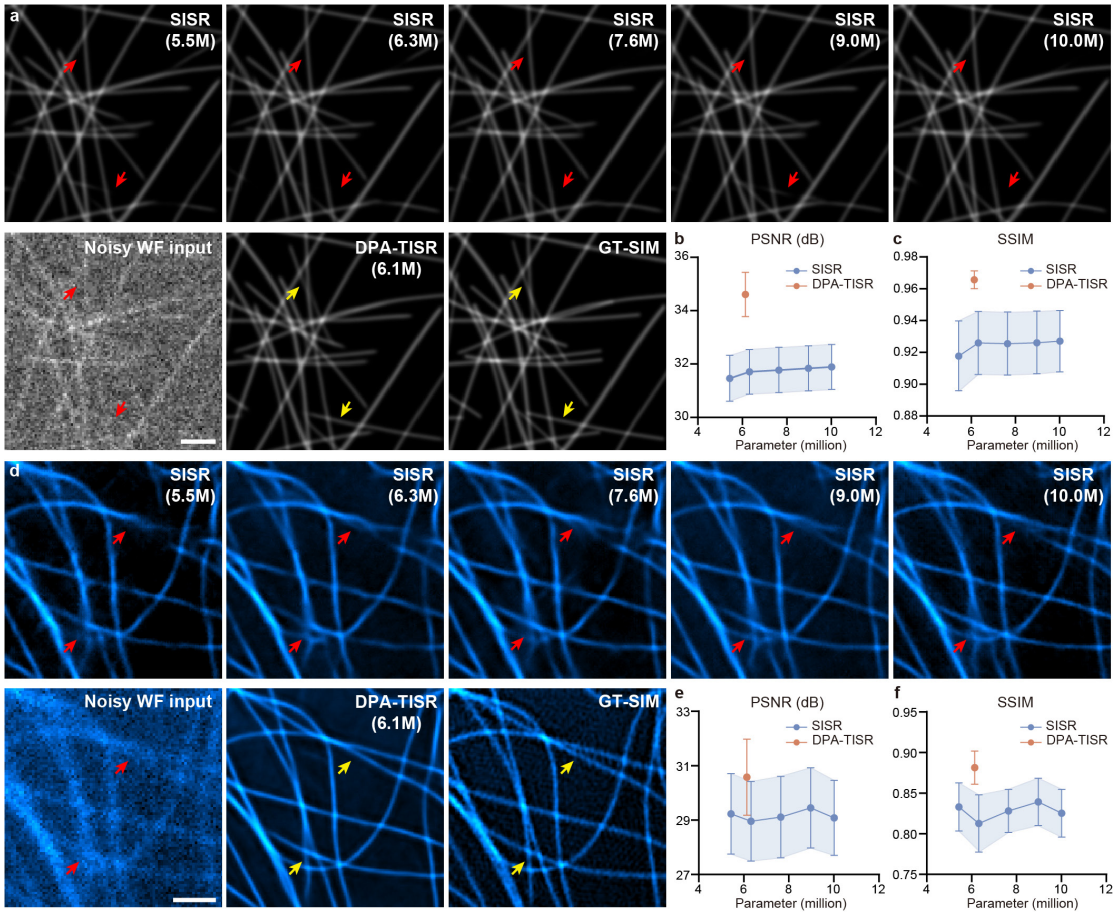
Supplementary Fig. 15 | Confidence calibration for the Bayesian DPA-TISR model trained with lysosome images. a, Confidence map generated by the Bayesian DPA-TISR model before confidence calibration. **b**, The absolute error map calculated with the inferred TISR image and GT-SIM image shown for reference. **c**, Confidence map generated by the Bayesian DPA-TISR model after confidence calibration. **d**, Reliability diagrams presented by accuracy versus average confidence before and after confidence correction with corresponding expected calibration error (ECE) and k value labeled. **e**,**f**, Representative detailed searching steps of the iterative finetuning process. The reliability diagrams (e) and corresponding confidence maps (f) are shown. Scale bar, 1 μm (a-c and f).



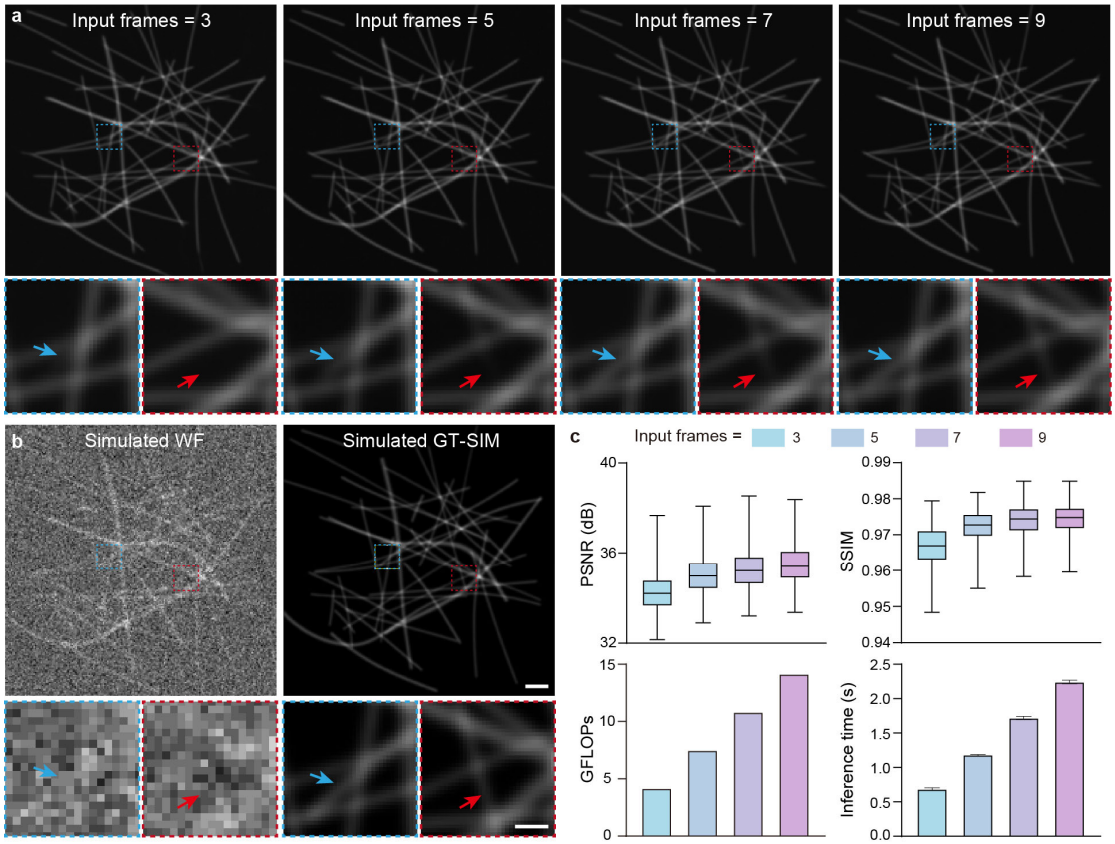
Supplementary Fig. 16 | Confidence calibration for the Bayesian DPA-TISR model trained with F-actin images. a, Confidence map generated by the Bayesian DPA-TISR model before confidence calibration. **b**, The absolute error map calculated with the inferred TISR image and GT-SIM image shown for reference. **c**, Confidence map generated by the Bayesian DPA-TISR model after confidence calibration. **d**, Reliability diagrams presented by accuracy versus average confidence before and after confidence correction with corresponding expected calibration error (ECE) and k value labeled. **e**,**f**, Representative detailed searching steps of the iterative finetuning process. The reliability diagrams (e) and corresponding confidence maps (f) are shown. Scale bar, 1 μm (a-c and f).



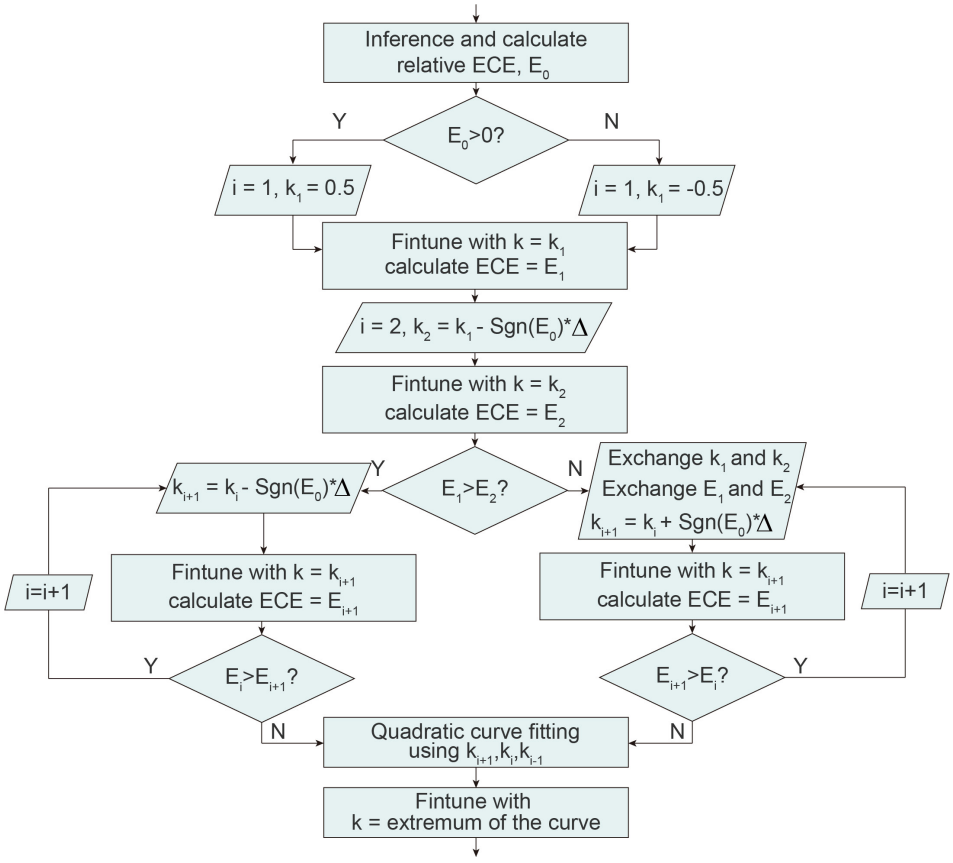
Supplementary Fig. 17 | Representative correct and false predictions indicated by the confidence map generated via Bayesian DPA-TISR. a, Time-lapse SR images of mitochondrial and lysosomes, as well as corresponding confidence map of mitochondrial images generated by Bayesian DPA-TISR, showcasing an equivocal mitochondrial fusion event. The red arrows indicate the potential false predictions of DPA-TISR identified by the low confidence value, which alerts us it may not be a real fusion event. **b**, Time-lapse SR images of mitochondrial and lysosomes, as well as corresponding confidence map of mitochondrial images generated by Bayesian DPA-TISR, showcasing a typical mitochondrial fission event. The high confidence of the prediction indicates a high reliability of this observation. Scale bar, 1 μm (a, b).



Supplementary Fig. 18 | Evaluation of SISR models with escalating numbers of trainable parameters on simulated and experimental datasets. a, Representative SR images of simulated tubular structure inferred by the DPA-TISR model and SISR models with escalating numbers of trainable parameters from 5.5 to 10 million. A single timepoint of noisy wide-field (WF) input and corresponding GT-SIM image are shown for reference. **b**,**c**, Statistical comparison of DPA-TISR and SISR models trained and evaluated with simulated dataset of tubular structures (n=200) in terms of PSNR (b) and SSIM (c). **d**, Representative SR images of microtubules inferred by the DPA-TISR model and SISR models with escalating numbers of trainable parameters from 5.5 to 10 million. A single timepoint of noisy wide-field (WF) input and corresponding GT-SIM image are shown for reference. **e**,**f**, Statistical comparison of DPA-TISR and SISR models trained and evaluated with experimental dataset (n=20) of microtubules in terms of PSNR (e) and SSIM (f). Center point, mean; whiskers, mean ± standard deviation. Scale bar, 1 μm (a and d).

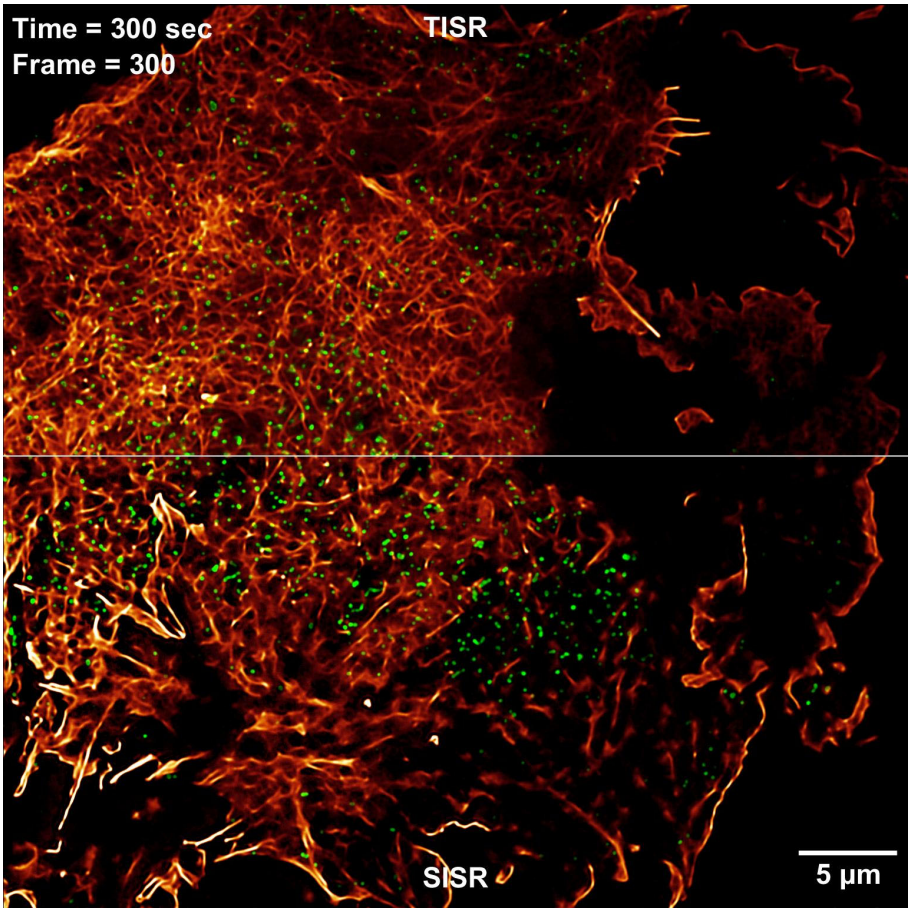


Supplementary Fig. 19 | Evaluation of DPA-TISR trained with different input configuration. a, TISR images inferred by DPA-TISR models trained and tested with different input sequence length ranging from 3 to 9. **b**, Corresponding simulated wide-field image as inputs and GT-SIM image shown for reference. **c**, Statistical comparison of DPA-TISR models trained and tested with different input sequence length in terms of PSNR, SSIM, GFLOPs and inference time on simulated data (n=200). For the box plots in the upper row: center line, medians; limits, 75% and 25%; whiskers, maximum and minimum. These evaluations demonstrate that the performance of DPA-TISR is generally enhanced with increasement of the input sequence length, however, at the expense of higher computation complexity. According to the evaluation, we selected an input sequence length of 7 in our experiments of this paper for compromise between efficiency and performance. Scale bar, 1 μm, 0.25 μm (zoom-in regions).

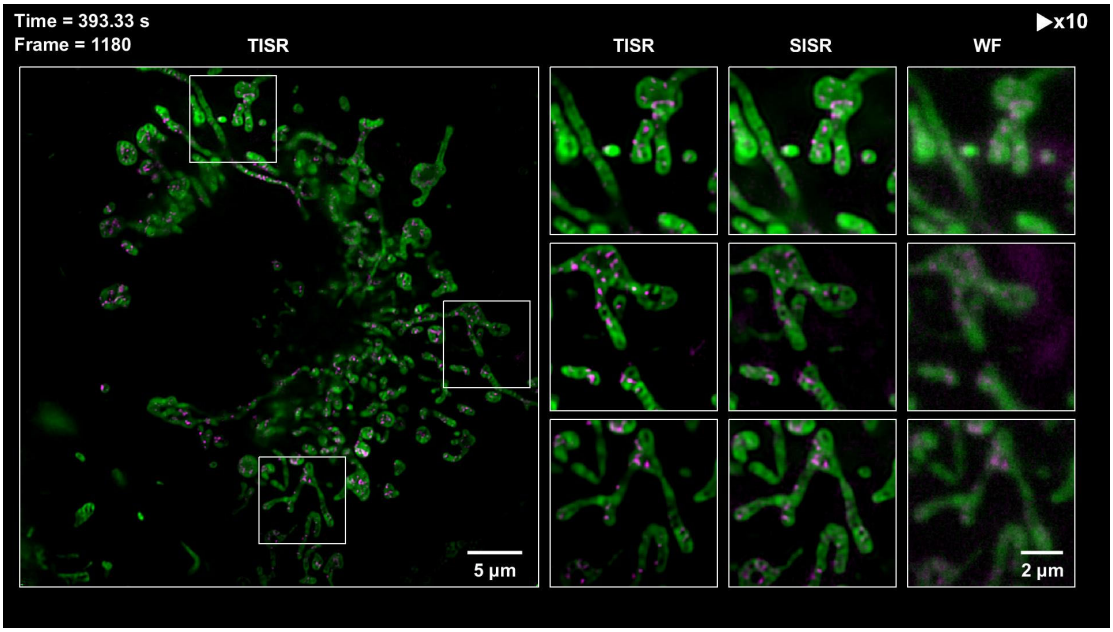


Supplementary Fig. 20 | Flow diagram of the confidence calibration procedure.

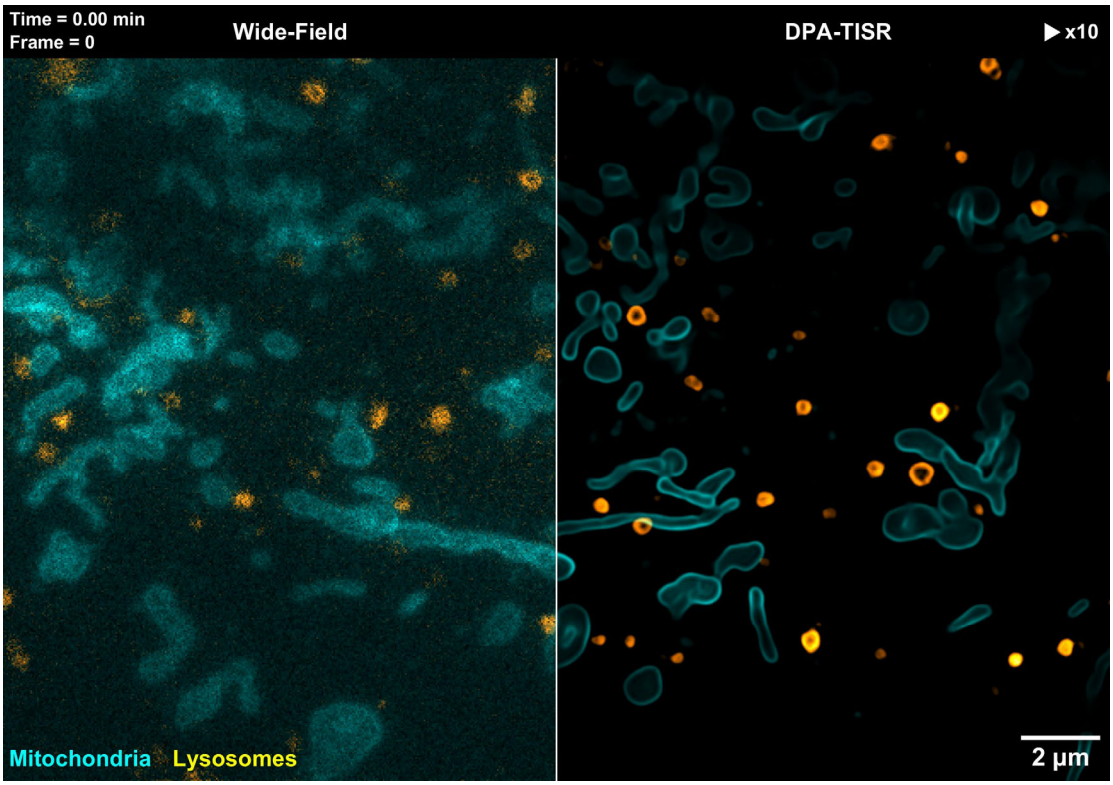
Captions for Supplementary Videos



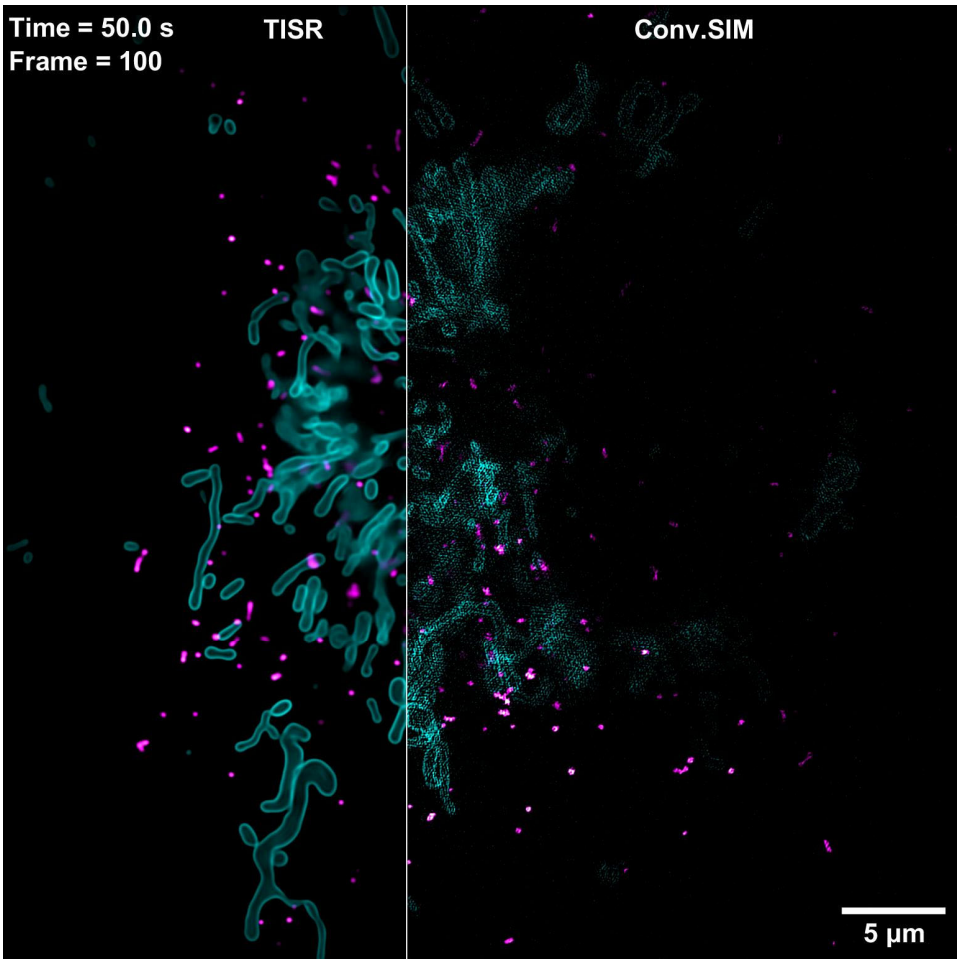
Supplementary Video 1 | Two-color SR live-imaging of CCPs and F-actin enabled by DPA-TISR. DPA-TISR enhanced TIRF imaging of CCPs and F-actin over 4,800 timepoints at 1-sec interval of a COS-7 cell co-expressing Lifeact-mEmerald (red) and Clathrin-mCherry (green), visualizing the rapid dynamics and interactions between the two subcellular structures.



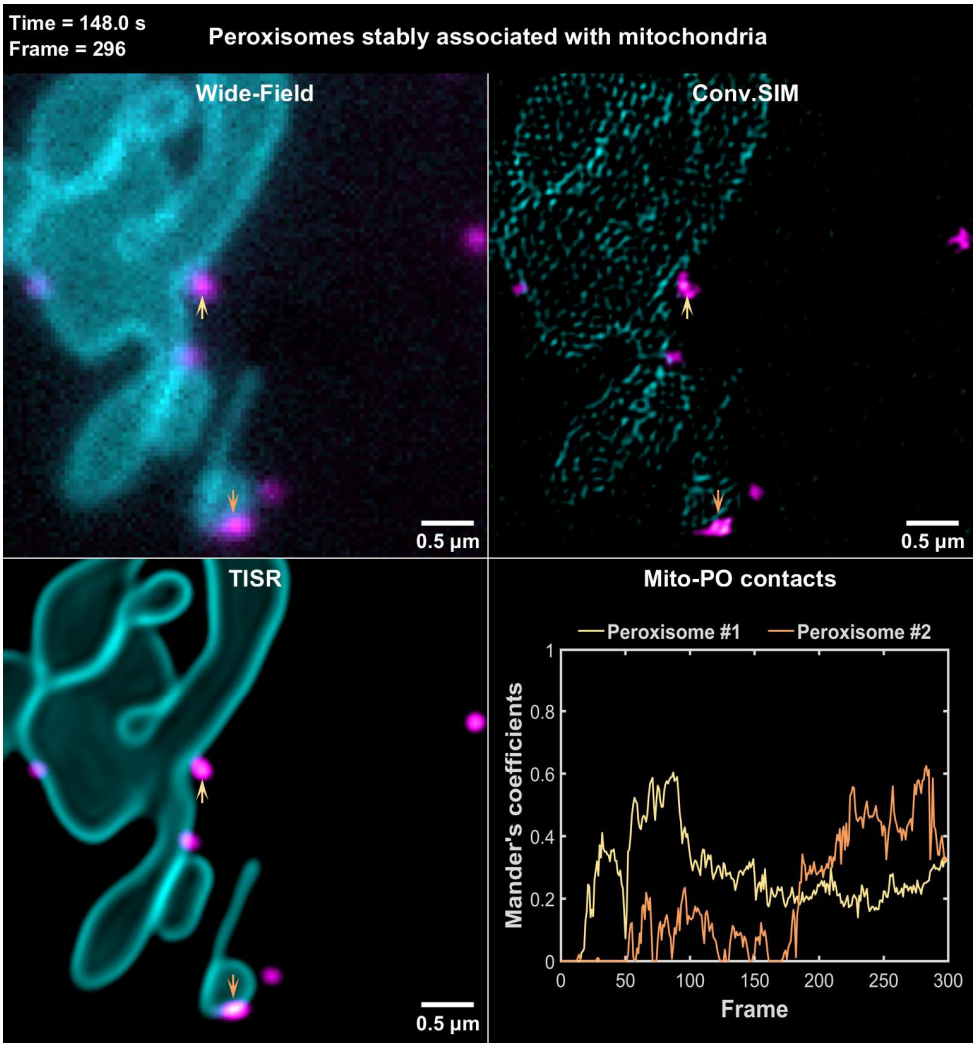
Supplementary Video 2 | Long-term SR recordings of mitochondrial cristae and nucleoid dynamics with DPA-TISR. Two-color SR live-imaging of a COS-7 cell labelled with PKMO-Halo (green) and TFAM-mEmerald (magenta) by DPA-TISR over 1,680 timepoints at a framerate of 3Hz, recoding mitochondrial fission and fusion and reorganizations of nucleoids concomitant with cristae deformation.



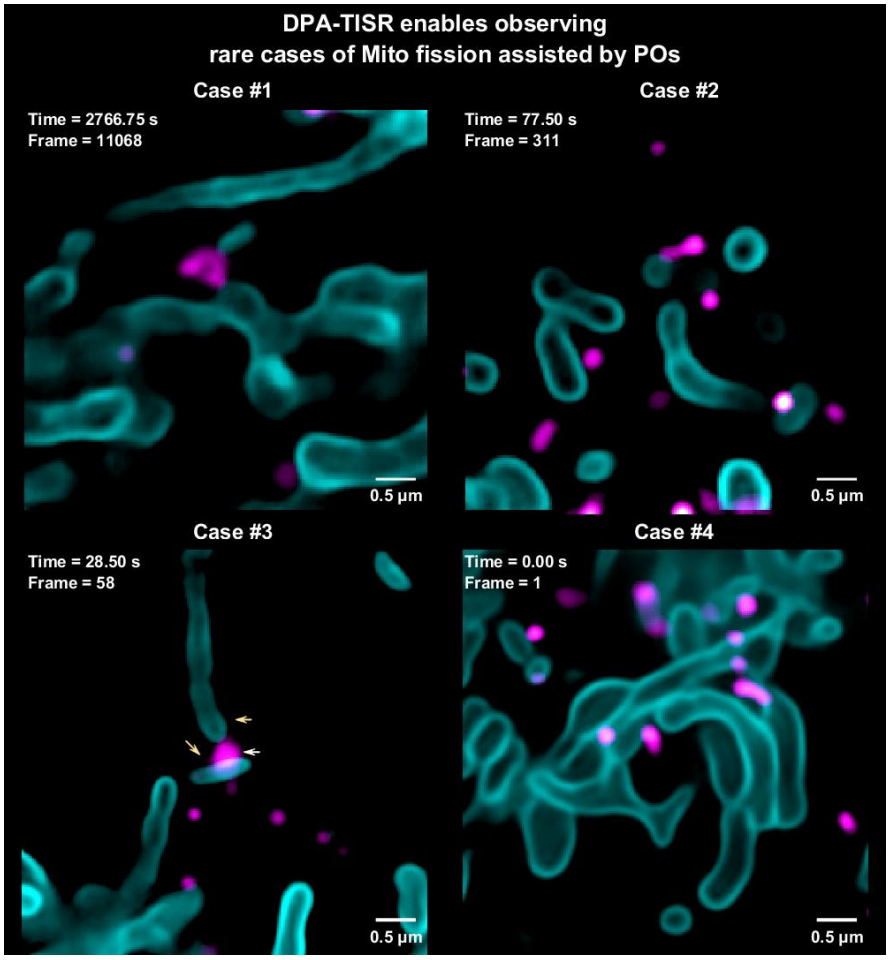
Supplementary Video 3 | Long-term SR live-imaging of outer mitochondrial membrane and lysosomes enabled by DPA-TISR. Two-color SR live-imaging of a COS-7 cell labelled with 2×mEmerald-Tomm20 (cyan) and Lamp1-Halo (yellow) over a ultralong time course of ~10,000 timepoints at 0.5-second intervals, showcasing multiple lysosome-mediated mitochondrial dynamics such as hitchhiking and fissions.



Supplementary Video 4 | Dynamic interactions between mitochondria (Mito) and peroxisomes (PO) visualized by DPA-TISR. SR live-imaging reconstructed via DPA-TISR of a COS-7 cell co-expressing 2×mEmerald-Tomm20 (cyan) and PMP-Halo (magenta) with high spatiotemporal resolution and long duration, revealing intricacy of complex Mito-PO contact behaviors.



Supplementary Video 5 | Examples of different types of Mito-PO contacts. Representative examples of four types of Mito-PO contacts: (i) POs stably associated with a single Mito; (ii) POs served as the bridge that simultaneously tethered two Mito; (iii) POs unexpectedly changing their contact sites from one Mito to another; and (iv) POs with no contact with Mito. Wide-field images (top left corner), conventional SIM images (top right corner), DPA-TISR images reconstructed from WF images (bottom left corner) are displayed for comparison. The Mander's overlap coefficient plots are shown in the bottom right panel, which serves as a quantitative indicator of the strength of Mito-PO contacts.



Supplementary Video 6 | DPA-TISR enables observation of rare biological events showing peroxisome-mediated mitochondrial fission. SR live-imaging enabled by DPA-TISR of COS-7 cells co-expressing 2×mEmerald-Tomm20 (cyan) and PMP-Halo (magenta) with ultra-long duration of more than 11,000 timepoints, showcasing four events where mitochondria split at the contact sites with peroxisomes.

Supplementary References

1. Hell, S.W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters* **19**, 780-782 (1994).
2. Klar, T.A., Jakobs, S., Dyba, M., Egner, A. & Hell, S.W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences* **97**, 8206-8210 (2000).
3. Gustafsson, M.G. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy* **198**, 82-87 (2000).
4. Rust, M.J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793-795 (2006).
5. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *science* **313**, 1642-1645 (2006).
6. Wang, Z. et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nature Methods* **18**, 551-556 (2021).
7. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* **16**, 103-110 (2019).
8. Ning, K. et al. Deep self-learning enables fast, high-fidelity isotropic resolution restoration for volumetric fluorescence microscopy. *Light: Science & Applications* **12**, 204 (2023).
9. Li, X. et al. Three-dimensional structured illumination microscopy with enhanced axial resolution. *Nature Biotechnology* (2023).
10. Park, H. et al. Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy. *Nature Communications* **13**, 3297 (2022).
11. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* **18**, 194-202 (2021).
12. Qiao, C. et al. 3D Structured Illumination Microscopy via Channel Attention Generative Adversarial Network. *IEEE Journal of Selected Topics in Quantum Electronics* **27**, 1-11 (2021).
13. Jin, L. et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nature Communications* **11**, 1934 (2020).
14. Chan, K.C., Zhou, S., Xu, X. & Loy, C.C. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 5972-5981 (2022).
15. Chan, K.C., Wang, X., Yu, K., Dong, C. & Loy, C.C. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 4947-4956 (2021).
16. Dai, J. et al. in Proceedings of the IEEE international conference on computer vision 764-773 (2017).
17. York, A.G. et al. Instant super-resolution imaging in live cells and embryos via analog image processing. *Nat Methods* **10**, 1122-1126 (2013).
18. Muller, C.B. & Enderlein, J. Image scanning microscopy. *Physical Review Letters* **104**, 198101 (2010).
19. The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. (2015).
20. Gawlikowski, J. et al. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 1-77 (2023).
21. Abdar, M. et al. A review of uncertainty quantification in deep learning: Techniques, applications

and challenges. *Information Fusion* **76**, 243-297 (2021).

22. Kendall, A. & Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Advances in Neural Information Processing Systems* **30**, 5574-5584 (2017).
23. Gal, Y. & Ghahramani, Z. in international conference on machine learning 1050-1059 (PMLR, 2016).
24. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090-1097 (2018).
25. Blei, D.M., Kucukelbir, A. & McAuliffe, J.D. Variational inference: A review for statisticians. *Journal of the American statistical Association* **112**, 859-877 (2017).
26. Gal, Y. & Ghahramani, Z. Bayesian convolutional neural networks with Bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158* (2015).
27. Guo, C., Pleiss, G., Sun, Y. & Weinberger, K.Q. in International conference on machine learning 1321-1330 (PMLR, 2017).
28. Banterle, N., Bui, K.H., Lemke, E.A. & Beck, M. Fourier ring correlation as a resolution criterion for super-resolution microscopy. *Journal of structural biology* **183**, 363-367 (2013).
29. Zhao, W. et al. Quantitatively mapping local quality of super-resolution microscopy by rolling Fourier ring correlation. *Light: Science & Applications* **12**, 298 (2023).
30. Chen, B.C. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998 (2014).
31. Descloux, A., Grussmayer, K.S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nat Methods* **16**, 918-924 (2019).