
Supplementary information

Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy

In the format provided by the authors and unedited

Supplementary Information for

Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy

Weisong Zhao, Shiqun Zhao, Liuju Li, Xiaoshuai Huang, Shijia Xing, Yulin Zhang, Guohua Qiu, Zhenqian Han, Yingxu Shang, De-en Sun, Chunyan Shan, Runlong Wu, Lusheng Gu, Shuwen Zhang, Riwan Chen, Jian Xiao, Yanquan Mo, Jianyong Wang, Wei Ji, Xing Chen, Baoquan Ding, Yanmei Liu, Heng Mao, Bao-Liang Song, Jiubin Tan, Jian Liu, Haoyu Li, Liangyi Chen

Correspondence to: lychen@pku.edu.cn (L.C.); lihaoyu@hit.edu.cn (H.L.)

Content

Supplementary Notes	3
Supplementary Note 1 Image formation and the resolution limit to the Wiener deconvolution.....	3
Supplementary Note 2 Computational super-resolution.	4
Supplementary Note 3 The development of the sparse deconvolution model.	8
Supplementary Note 4 The details of algorithm execution.....	14
Supplementary Note 4.1 Iterative approximation based on the weighted constraints.	14
Supplementary Note 4.2 Iterative deconvolution.	16
Supplementary Note 4.3 Background estimation.	17
Supplementary Note 4.4 Upsampling operation.	18
Supplementary Note 5 Using the sparse and continuity prior knowledge improves resolution and contrast.	20
Supplementary Note 6 Parameters in the sparse deconvolution software.	26
Supplementary Note 7 Optimal values of the sparsity and the fidelity under different conditions.	29
Supplementary Note 8 Parameters finetuning	33
Supplementary Note 8.1 Procedures for parameters finetuning.	33
Supplementary Note 8.2 Step-by-step examples.	33
Supplementary Note 8.3 More explanations on the background estimation.....	36
Supplementary Note 8.4 The effects of t-axial continuity.	36
Supplementary Note 8.5 Relationship between pixel sizes and the sparsity and fidelity values.	36
Supplementary Note 9 The calibration of nuclear pore diameters and the FWHM resolution.	47
Supplementary Note 9.1 Calibrating fitted diameters of ring-shaped pores.	47
Supplementary Note 9.2 The bead correction factor.	47
Supplementary Note 10 Other sparsity-based reconstructions that improve spatial resolution in theory.....	51
Supplementary Note 10.1 FISTA method.	51
Supplementary Note 10.2 NLHT method.	53
Additional Supplementary Figures	55
Supplementary Tables	69
Supplementary Table 1 SSIM of different reconstructions from images corrupted with varying levels of noise.	69
Supplementary Table 2 PSNR of different reconstructions from images corrupted with varying levels of noise.	69
Supplementary Table 3 Imaging conditions.....	70
Supplementary Table 4 Non-zero ratios and background selections of different images.	71
Supplementary Table 5 Computing time of sparse deconvolution.	73
Captions for Supplementary Videos.	74
References	92

Supplementary Notes

Supplementary Note 1 | Image formation and the resolution limit to the Wiener deconvolution.

The image formation of an aberration-free optical system is governed by simple propagation models of electromagnetic waves interacting with geometrically ideal surfaces of lenses and mirrors. Under such conditions, the process can be described as a two-dimensional Fourier transformation¹,

$$G(u, v) = H(u, v) \times F(u, v), \quad (1)$$

in which G is the two-dimensional Fourier transform of the recorded *image*, F is the transform of the intensity distribution of the *object* that gave rise to the image, and H is the spatial frequency transfer function of the *optical system*, the OTF. Because the total amount of light energy entering an optical system is limited by a physically limiting pupil or aperture, the OTF is controlled by the aperture function under aberration-free conditions. Therefore, the OTF $H(u, v)$ can be calculated from the aperture function $a(x, y)$ by

$$H(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(x, y) a(x+u, y+v) dx dy. \quad (2)$$

Therefore, the OTF is the autocorrelation of the aperture function, and x and y denote coordinates in the spatial domain, while u and v represent coordinates in the spatial frequency domain. The transmission of electromagnetic waves within the aperture is perfect, with $a(x, y) = 1$, while outside of the aperture, no wave can propagate and $a(x, y) = 0$. Therefore, the OTF changes to zero outside of a boundary defined by the autocorrelation of the aperture function. Because no spatial frequency information from the object outside of the OTF support can be transmitted to the recording device, the optical system itself is always bandwidth limited¹.

To recover the information of the original *object* from the recorded *image*, **Eq. (1)** can be rearranged into the following form:

$$F(u, v) = \frac{G(u, v)}{H(u, v)}. \quad (3)$$

In the preceding equation, it is obvious that the retrieved object spatial frequency exists only where $H(u, v) \neq 0$, and **division by null outside of the region of OTF support is impossible**. This is why diffraction limits the image resolution and causes the general belief that the information lost due to the cut-off of system OTF cannot be retrieved by mathematical inversion in the frequency domain.

Supplementary Note 2 | Computational super-resolution.

Richardson and Lucy^{2, 3} (RL) deconvolution, followed later by Shepp and Vardi⁴ regarded the problem of recovering the underlying image blurred by a known PSF as one of statistical estimation rather than an exercise in direct solution. They proposed an iterative deconvolution method based on the Bayes' theorem on conditional probabilities, and the following equation gives the basic algorithmic computation:

$$f^{n+1}(x, y) = f^n(x, y) \left[\left(\frac{g(x, y)}{h(x, y) \otimes f^n(x, y)} \right) \odot h(x, y) \right]. \quad (4)$$

For image data $g(x, y)$ and the PSF $h(x, y)$, the object at iteration $n+1$ is calculated from its estimation at n . The starting approximation $f^0(x, y)$ should be a smooth, nonnegative function having the same integrated intensity as the observed image, thus adding a nonnegative a priori knowledge to the approximation. From this equation, it is obvious that a large deviation of $\frac{g(x, y)}{h(x, y) \otimes f^n(x, y)}$ from unity on a large length scale compared to that of $h(x, y)$ would result in a large correction to yield the new $f^{n+1}(x, y)$, while small deviations lead to small corrections³. Thus the algorithm nonlinearly approximates the maximum-likelihood estimation of the observed sample.

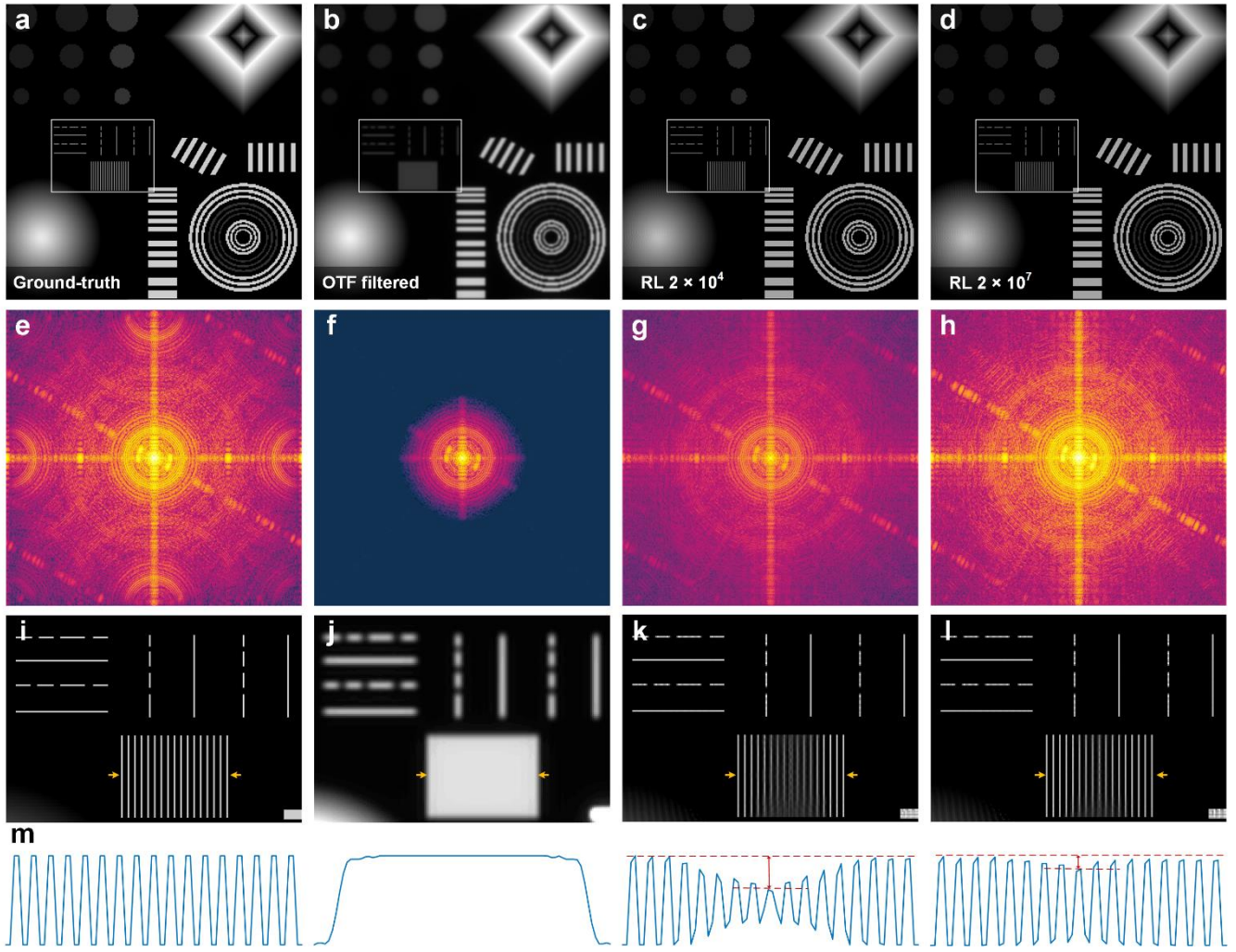
Unlike direct inversion in the frequency domain as Wiener deconvolution in **Eq. (3)**, this computational process in **Eq. (4)** is conducted in the spatial domain, avoiding the problem of division by null outside of the region of OTF support. **In other words, instead of the direct inversion of the low-pass filter $H(u, v)$ in the frequency domain ($G(u, v) = H(u, v) \times F(u, v)$), the RL deconvolution iteratively approximates with its inverse Fourier transform, i.e., the convolution kernel $h(x, y)$ in the spatial domain ($g(x, y) = h(x, y) \otimes f(x, y)$).** In principle, the extension beyond the diffraction limit is possible. It may occur for two reasons: the nonlinearity of the product of f^n with the other quantities in the square brackets and the prior information of nonnegative values of the object.

Because RL deconvolution will quickly converge to a solution dominated by the noise⁵, it is recommended to stop the reconstruction after a small number of iterations in practice. However, if the spatial frequency is sufficiently enormous, many iterations will be necessary to extract all the information⁶. To test this idea, we have synthesized a ground-truth, noise-free image containing various-shaped structures. After the image was convolved with a Gaussian PSF (FWHM of ~ 90 nm), we manually set spatial frequency beyond 90 nm to zero, which trimmed off much of the high-frequency information. Thus the blurred image resembled that obtained by optical microscopy of limited bandwidths. Unexpectedly, parallel lines with the spacing of 60 nm can be resolved after RL deconvolution for 2×10^4 iterations, whereas their intensities also became more uniform upon even more iterations up to 2×10^7 (**Supplementary Fig. 1**). Therefore, our simulation confirmed that many iterations of RL deconvolution extended bandwidth in synthetic images without noise, thus recovering the missing information⁷. This simulation echoed the mathematical bandwidth extrapolation idea first proposed by Wolter⁸, and later summarized in Goodman's book⁹. For an optical system, it follows that if an

object has a finite size, a unique analytic function exists that coincides inside $G(u, v)$. By extrapolating the observed spectrum using *analytic continuation*¹⁰ or the Gerchberg algorithm¹, it is possible in principle to reconstruct the object with arbitrary precision under noise-free conditions.

Despite the theoretical feasibility of mathematical SR, out-of-band extrapolation is usually unstable due to noise in data. The noise in actual experiments may not be an analytic function, thus allowing no solution to the bandwidth extrapolation based on analytic continuation¹⁰. From the perspective of information capacity, bandwidth extrapolation based on analytic continuation extends the spatial bandwidth at the price of the signal-to-noise ratio (SNR). For object fields containing large space-bandwidth products, even a small extension of spatial bandwidth leads to greatly amplified noise in the final reconstruction^{11, 12}.

To demonstrate possible effects of noise on RL deconvolution, we simulated parallel lines blurred with PSF (FWHM of ~90 nm), spacing from 10 nm to 100 nm, and increasing at the interval of 10 nm (**Supplementary Fig. 2**). Without noise, 2000 iterations of RL deconvolution correctly resolved lines 10 nm apart (**Supplementary Fig. 2c**); in contrast, adding the Poisson noise rendered excessive RL iterations improper, thus parallel lines with a distance smaller than 70 nm could not be resolved without many artifacts. The RL deconvolution became less effective in increasing resolution as more Gaussian noise was added to the image (**Supplementary Fig. 2e1-2e3**). Therefore, classic RL deconvolution may not enhance the resolution of images corrupted with noise such as those captured by real-world microscopes, which agrees with the majority of the imaging community's belief-the Rayleigh diffraction limit represents a practical frontier that cannot be overcome with a conventional imaging system⁹.



Supplementary Fig. 1 | Low-pass filtered image recovered by RL deconvolution under noise-free condition. (a) The ground-truth contains synthetic various-shaped structures with a pixel size of 20 nm. (b) The low-pass filtered (equal to 90 nm resolution) image of (a). (c) RL deconvolution result of (b) with 2×10^4 iterations. (d) RL deconvolution with 2×10^7 iterations. (e-h) Fourier transforms of (a-d). (i-l) Magnified views of white boxes from (a-d). (m) The corresponding profiles are taken from the regions between the yellow arrowheads in (i-l).



Supplementary Fig. 2 | The evolution of noise effects on parallel paired-line structures resolved by the RL-deconvolution. (a-b) We created a ground-truth containing twelve parallel paired lines departed with different distances (from 10 nm to 120 nm), and convoluted by a PSF with a 90 nm FWHM. (c) Without any synthetic noise (and sufficient spatial sampling, 10 nm/pixel), the RL-deconvolution could distinguish paired lines 10 nm apart. (d1-d4) We simulated images captured by real-world microscopes with large pixel sizes and different noise levels. After downsampling the image in (b) by two folds (20 nm/pixel), we added Poisson (d1), Poisson + 10% Gaussian (d2), Poisson + 50% Gaussian (d3), and Poisson + 100% Gaussian (d4) noise. (e1-e4) Upon introduction of even weak noise, RL-deconvolution failed to resolve paired lines 70 nm apart. In images corrupted with more than 50% Gaussian noise, the RL-deconvolution failed to improve the resolution. We used 2000, 15, 15, 5, and 5 times of iterative deconvolution for images of different noise levels (from 0% to 100%).

Supplementary Note 3 | The development of the sparse deconvolution model.

As explained in **Supplementary Note 2**, noise ultimately limited any possible extension in resolution by the RL deconvolution. Under any fluorescence microscopes, images captured by the photoelectric sensor¹³ can be regarded as the convolution of the object with the point spreading function (PSF) of the microscope. Thus a fluorescent image can be expressed as:

$$image = \mathcal{N} \left\{ \mathcal{S} \left[(I(\omega) \times x(\omega)) \otimes A(\omega, \theta) \right] \right\}, \quad (5)$$

where ω is the spatial coordinate, and I , x , and A represent the illumination, the objective, and the PSF of the microscope, respectively. The $\mathcal{S}$ and $\mathcal{N}$ denote the sampling and noise. The *image* is the final collected image of the microscope, and θ is the aberration in the microscope.

In real optical instruments, deviations may arise from the Poisson noise intrinsic to the observation, the inevitable read-out noise of the hardware¹⁴⁻¹⁸, the existence of aberration in the optical path¹⁹, and the effect of finite sampling^{20, 21}. To accurately deconvolve the object from images corrupted with noise always confers an ill-posed inverse problem. Therefore, by building up a constraint model using the prior knowledge widely shared by fluorescent microscopes, we wanted to decode the fluorescent signals as close to their ground-truth positions as possible. Taken an example of single-molecule localization microscopy (SMLM), it achieves super-resolution down to 10~20 nm by assuming each isolated fluorescence puncta stemmed from a single molecule^{22, 23}. However, such strong prior knowledge needs specifically designed instruments and fluorescent probes, which may not apply to conventional microscopes. Therefore, we introduce two pieces of prior knowledge, i.e., sparsity and continuity, which are common features of fluorescence generation that can be used in our reconstruction model to achieve super-resolution.

Continuity in xy - $t(z)$ axes:

For any microscope systems, to ensure sufficient Nyquist sampling criteria for maximal spatial resolution dictated by the optics, their PSF must occupy more than 3×3 pixels in space (**Supplementary Fig. 3**), which constitutes the basis for the continuity along x and y axes of any fluorescence microscope²⁴. For example, the Nup96 proteins within a nuclear pore complex could manifest either as fluorescence puncta (wide-field, confocal, Airy scan), regular or irregular ring-shaped structure (STED, expansion microscopy), or 8~14 nearly symmetric puncta overlaid on the ring (single-molecule localization microscopy)²⁵. Despite different shapes of Nup96 proteins under different microscopes, structures should always be regarded as continuous in which more than two spatial samplings are needed to achieve that resolution dictated by the Nyquist criteria. Moreover, if the distance that the biological sample moved between two consecutive time points was smaller

than the microscope's lateral resolution, structures were considered to be continuous along the t axis. Given that cells may demonstrate high-speed dynamics sometimes (such as Ca^{2+} diffusion within live cells, **Extended Data Fig. 10**), the parameter for the continuity along t axis can be turned off under those conditions. As we have demonstrated earlier, the application of continuity constraint substantially suppresses reconstruction artifacts caused by Gaussian-type random noise²⁴.

Sparsity:

Although the application of continuity a priori can increase the SNR, it also obscures the image and reduces the resolution. For any fluorescence microscopes, an increase in spatial resolution always leads to smaller PSF. Thus, as compared with a conventional microscope, the convolution of the object with the smaller PSF in super-resolution images always confers relative sparsity. Therefore, we propose introducing the sparsity as another prior knowledge to extract the fluorescent object's high-frequency information.

Absolute sparsity (ℓ_0 norm, the total number of non-zero elements) has been linked with compressive sensing (CS), which has been successfully applied to different fields of signal processing²⁶⁻²⁸. If the signal is sparse (mostly zeros) or sparse after some transformations, CS can precisely recover the signal from the highly noisy or corrupted measurements²⁶. However, instead of ℓ_0 norm (only absolute sparsity) used as the start point in CS theory, we directly used the ℓ_1 norm (sparsity score in **Supplementary Fig. 3a**), i.e., the sum of absolute values of all elements.

We simulated an array of sparsely distributed fluorescence molecules and the diffusive square fluorescence structures without fluorescence at the array complementary to the previous structure, and convolve these ground truth structures with PSFs of variable sizes to mimic images detected by microscopes of different spatial resolutions. No matter whether these structures were absolutely sparse or not, we showed that the sparsity scores of images of the high resolution were always smaller than those of low resolution (**Supplementary Fig. 3a, 3b**). Therefore, our sparsity term can handle both absolutely and relatively sparse structures and constrain the extraction of high spatial frequency information, as shown in numerous examples listed in **Supplementary Fig.4**.

In summary, the continuity means positively correlated fluorescence among neighboring 2×2 or 3×3 pixels, which equals the size of PSF of the microscope under the efficient sampling condition required by the Nyquist sampling theorem. On the other hand, any captured microscope images are always denser than the real object because of the convolution of the object with the system PSF, and high resolution always means an increase in relative sparsity (**Supplementary Fig. 3c**). Therefore, we believe that continuity and sparsity

are general features of the fluorescence microscope, which could be used as the prior knowledge to suppress noise and facilitate high-frequency information extraction collaboratively (**Supplementary Fig. 3c-3d**).

The sparse deconvolution model:

Therefore, we have proposed the following loss function containing these two priors, in which the Hessian matrix continuity to reduce artifacts and increase robustness at the price of reduced resolution, and the sparsity to balance the extraction of high-frequency information, which gives:

$$\arg \min_{\mathbf{x}, \mathbf{b}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + R_{\text{Hessian}}(\mathbf{x}) + \lambda_{L1} \|\mathbf{x}\|_1 \right\}. \quad (6)$$

The first term on the left side of the equation is the fidelity term, representing the distance between the recovered image $\mathbf{x}$ and the result obtained with the Wiener filtered SIM image $\mathbf{f}$. $\mathbf{A}$ is the PSF of the optical system. Here, $\mathbf{b}$ is the background, which is assumed to be smooth, varying in the whole field-of-view. We involved such background $\mathbf{b}$ in **Eq. (6)**, considering that the fluorescent images may contain strong background components from the out-of-focus emission light, fluorescent background, and cellular autofluorescence. We estimated and removed such background fluorescence, which is critical for the high-fidelity reconstruction of objects inside the cell. The second and third terms are the continuity and sparsity priors, respectively. $\|\cdot\|_1$ and $\|\cdot\|_2$ are the ℓ_1 and ℓ_2 norms, respectively, and λ and λ_{L1} denote weight factors to balance image fidelity with sparsity. The continuity prior regarding the structural continuity along the xy - t (z) axes is defined as follows:

$$\begin{aligned} R_{\text{Hessian}}(\mathbf{g}) &= \left\| \begin{bmatrix} \mathbf{g}_{xx} & \mathbf{g}_{xy} & \sqrt{\lambda_t} \mathbf{g}_{xt} \\ \mathbf{g}_{yx} & \mathbf{g}_{yy} & \sqrt{\lambda_t} \mathbf{g}_{yt} \\ \sqrt{\lambda_t} \mathbf{g}_{tx} & \sqrt{\lambda_t} \mathbf{g}_{ty} & \lambda_t \mathbf{g}_{tt} \end{bmatrix} \right\|_1. \\ &= \|\mathbf{g}_{xx}\|_1 + \|\mathbf{g}_{yy}\|_1 + \lambda_t \|\mathbf{g}_{tt}\|_1 + 2\|\mathbf{g}_{xy}\|_1 + 2\sqrt{\lambda_t} \|\mathbf{g}_{xt}\|_1 + 2\sqrt{\lambda_t} \|\mathbf{g}_{yt}\|_1 \end{aligned} \quad (7)$$

In **Eq. (7)**, λ_t denotes the regularization parameters that present the continuity along the t axis, which can be turned off upon imaging objects moving at high speed. Besides, we split the prior knowledge constrain and deconvolution subproblems into a two-step operation to gradually and effectively search the final solution $\mathbf{x}$. The full three-step is given as follow:

Step. 0 (optional): Because not all of the samples may have the background (listed in **Supplementary Table 4**), we estimate the background $\mathbf{b}$ in **Eq. (6)** by an optional isolate step before iterative optimization. We used a modified iterative wavelet transform method (lowest frequency wavelet bands of $\mathbf{f}$) to estimate $\mathbf{b}$ ²⁹.

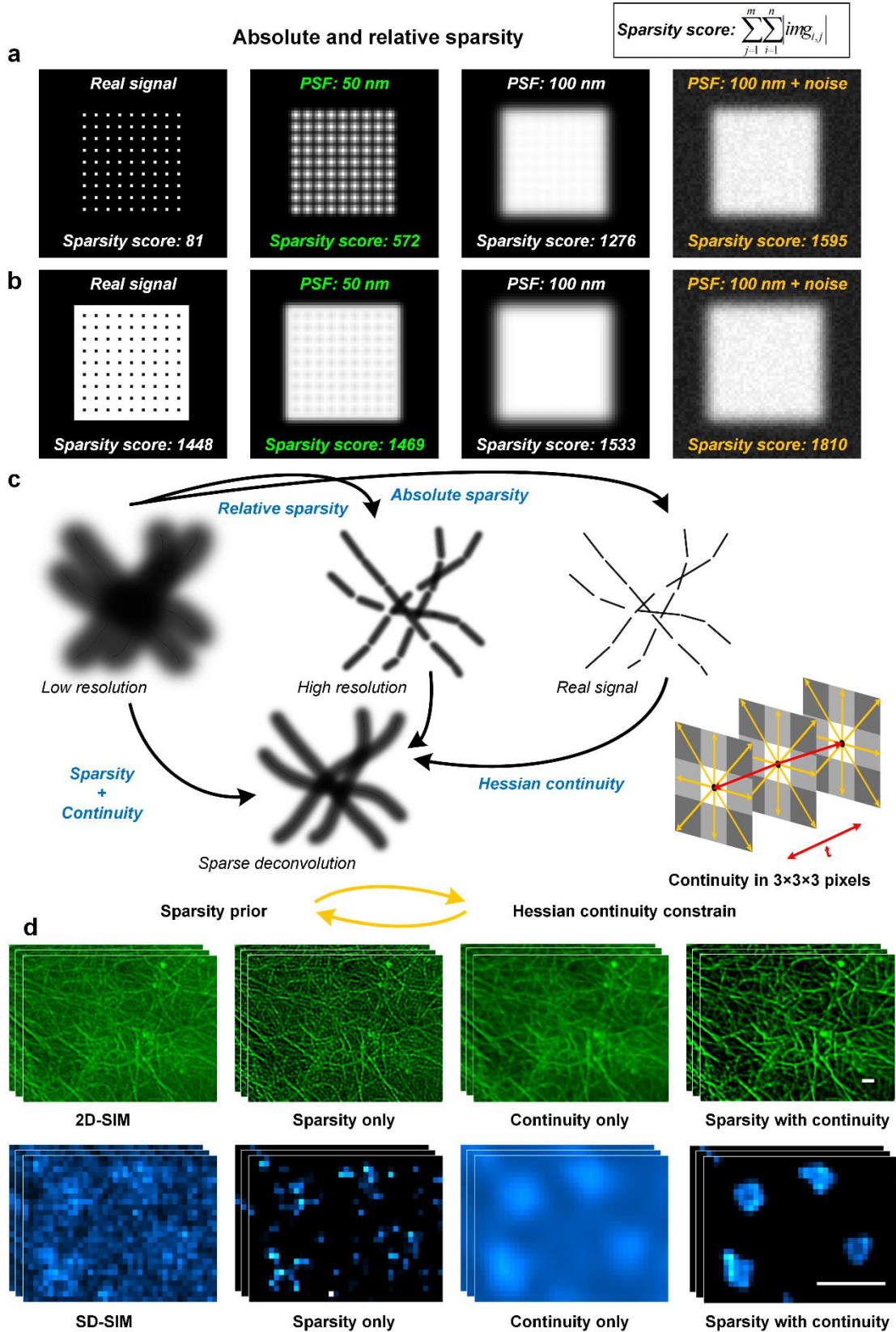
Step. 1: We solve the **Eq. (6)** gradually with the following **Eqs. (8)** and **(9)**. The detailed procedure of **Step. 1** can be found in **Supplementary Note 4.1**.

$$\arg \min_{\mathbf{g}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + R_{\text{Hessian}}(\mathbf{g}) + \lambda_{L1} \|\mathbf{g}\|_1 \right\}. \quad (8)$$

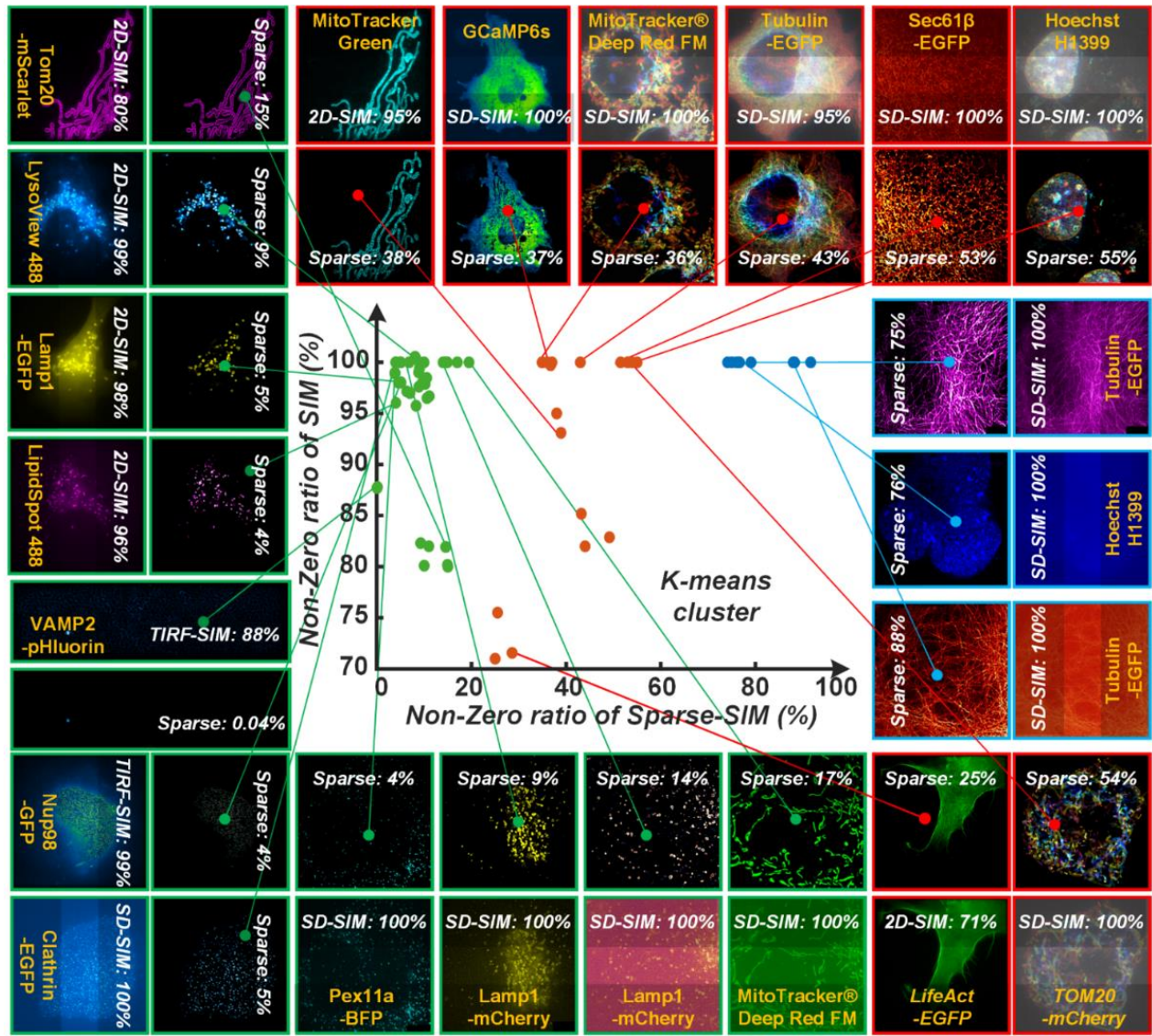
Step. 2: Iterative deconvolution to achieve final solution $\mathbf{x}$. The detailed procedure can be found in **Supplementary Note 4.2**.

$$\arg \min_{\mathbf{x}} \left\{ \|\mathbf{g} - \mathbf{Ax}\|_2^2 \right\}. \quad (9)$$

Because the pixel size reconstructed may not match with the improved resolution, we introduce a pre upsampling step (optional, marked as $\times 2$) to reconstruct the structure on smaller grids before the execution of **Step. 1**. The detailed notation is provided in **Supplementary Note 4.4**.



Supplementary Fig. 3 | The concept of sparse deconvolution. (a, b) Specific examples for absolute (a) and relative (b) sparsity. (c) Visualization of the working process of our sparse deconvolution. (d) Reconstructions with only sparsity, only continuity, or both the sparsity and continuity priors. Actin filaments labeled with LifeAct-EGFP in a live COS-7 cell in **Fig. 4a** under 2D-SIM (left, top panel) and CCPs labeled by Clathrin-EGFP in **Fig. 5a** under SD-SIM (left, bottom panel), followed by sparsity-constraint deconvolution (middle left), continuity-constraint deconvolution (middle right), or both sparsity and continuity-constraint deconvolution (right). Scale bars: 500 nm.



Supplementary Fig. 4 | Different decreases in non-zero ratios of images after the sparse deconvolution. We collected the number of non-zero values of images of this paper before and after our sparse deconvolution (details also in **Supplementary Table 4**). To demonstrate these numbers more clearly, they are normalized by the total pixel number of images. The resulting ratios could be clustered by the K-means method³⁰ into three categories in different colors (green, red, and blue). Each scatter denotes a non-zero ratio of one frame before (*x*-axis) and after (*y*-axis) the sparse deconvolution.

Supplementary Note 4 | The details of algorithm execution.

Supplementary Note 4.1 | Iterative approximation based on the weighted constraints.

Searching for the best solution of the loss function in **Eq. (8)** can be translated into a convex optimization problem (COP). To do that, we adopted the split Bregman algorithm widely used in total variation (TV) problems³¹ due to its fast convergence speed. Using Bregman splitting, we first replaced variable $\mathbf{g}$ with the intermediate variable $\mathbf{u}$. Then, the loss function could be transformed into a constrained minimization problem:

$$\min_{\mathbf{g}, \mathbf{u}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + \varphi(\mathbf{u}) \right\} = \min_{\mathbf{g}, \mathbf{u}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + \|\mathbf{u}_{xx}\|_1 + \|\mathbf{u}_{yy}\|_1 + \|\mathbf{u}_{tt}\|_1 + \|\mathbf{u}_{xy}\|_1 + \|\mathbf{u}_{xt}\|_1 + \|\mathbf{u}_{yt}\|_1 + \|\mathbf{u}\|_1 \right\}, \quad (10)$$

in which $\mathbf{u}_{xx} = \mathbf{g}_{xx}$, $\mathbf{u}_{yy} = \mathbf{g}_{yy}$, $\mathbf{u}_{tt} = \lambda_t \cdot \mathbf{g}_{zz}$, $\mathbf{u}_{xy} = 2 \cdot \mathbf{g}_{xy}$, $\mathbf{u}_{xt} = 2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt}$, $\mathbf{u}_{yt} = 2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt}$, $\mathbf{u} = \lambda_{L1} \cdot \mathbf{g}$.

By using the Lagrange multiplier method to enforce the constraints weakly, we obtained the following unconstrained problem:

$$\min_{\mathbf{g}, \mathbf{u}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + \varphi(\mathbf{u}) + \frac{\mu}{2} \cdot \left[\|\mathbf{u}_{xx} - \mathbf{g}_{xx}\|_2^2 + \|\mathbf{u}_{yy} - \mathbf{g}_{yy}\|_2^2 + \|\mathbf{u}_{tt} - \lambda_t \cdot \mathbf{g}_{tt}\|_2^2 + \|\mathbf{u}_{xy} - 2 \cdot \mathbf{g}_{xy}\|_2^2 + \|\mathbf{u}_{xt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt}\|_2^2 + \|\mathbf{u}_{yt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt}\|_2^2 + \|\mathbf{u} - \lambda_{L1} \cdot \mathbf{g}\|_2^2 \right] \right\}, \quad (11)$$

where μ is the Lagrange multiplier. Finally, we strictly enforced the constraints by applying the simplified Bregman iteration, which was proved by Goldstein and Osher³²:

$$\min_{\mathbf{g}, \mathbf{u}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + \varphi(\mathbf{u}) + \frac{\mu}{2} \cdot \left[\|\mathbf{u}_{xx} - \mathbf{g}_{xx} - \mathbf{v}_{xx}\|_2^2 + \|\mathbf{u}_{yy} - \mathbf{g}_{yy} - \mathbf{v}_{yy}\|_2^2 + \|\mathbf{u}_{tt} - \lambda_t \cdot \mathbf{g}_{tt} - \mathbf{v}_{tt}\|_2^2 + \|\mathbf{u}_{xy} - 2 \cdot \mathbf{g}_{xy} - \mathbf{v}_{xy}\|_2^2 + \|\mathbf{u}_{xt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt} - \mathbf{v}_{xt}\|_2^2 + \|\mathbf{u}_{yt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt} - \mathbf{v}_{yt}\|_2^2 + \|\mathbf{u} - \lambda_{L1} \cdot \mathbf{g} - \mathbf{v}\|_2^2 \right] \right\}, \quad (12)$$

where $\mathbf{v}$ is used to reduce the computational complexity in the iteration.

The iterative minimization procedure is given as follows:

$$\mathbf{g}^{k+1} = \min_{\mathbf{g}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + \frac{\mu}{2} \cdot \left[\|\mathbf{u}_{xx} - \mathbf{g}_{xx} - \mathbf{v}_{xx}\|_2^2 + \|\mathbf{u}_{yy} - \mathbf{g}_{yy} - \mathbf{v}_{yy}\|_2^2 + \|\mathbf{u}_{tt} - \lambda_t \cdot \mathbf{g}_{tt} - \mathbf{v}_{tt}\|_2^2 + \|\mathbf{u}_{xy} - 2 \cdot \mathbf{g}_{xy} - \mathbf{v}_{xy}\|_2^2 + \|\mathbf{u}_{xt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt} - \mathbf{v}_{xt}\|_2^2 + \|\mathbf{u}_{yt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt} - \mathbf{v}_{yt}\|_2^2 + \|\mathbf{u} - \lambda_{L1} \cdot \mathbf{g} - \mathbf{v}\|_2^2 \right] \right\}. \quad (13)$$

$$\begin{aligned}
\mathbf{u}_{xx}^{k+1} &= \min_{\mathbf{u}_{xx}} \left\{ \|\mathbf{u}_{xx}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u}_{xx} - \mathbf{g}_{xx}^{k+1} - \mathbf{v}_{xx}^k\|_2^2 \right\} \\
\mathbf{u}_{yy}^{k+1} &= \min_{\mathbf{u}_{yy}} \left\{ \|\mathbf{u}_{yy}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u}_{yy} - \mathbf{g}_{yy}^{k+1} - \mathbf{v}_{yy}^k\|_2^2 \right\} \\
\mathbf{u}_{tt}^{k+1} &= \min_{\mathbf{u}_{tt}} \left\{ \|\mathbf{u}_{tt}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u}_{tt} - \lambda_t \cdot \mathbf{g}_{tt}^{k+1} - \mathbf{v}_{tt}^k\|_2^2 \right\} \\
\mathbf{u}_{xy}^{k+1} &= \min_{\mathbf{u}_{xy}} \left\{ \|\mathbf{u}_{xy}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u}_{xy} - \lambda_z \cdot \mathbf{g}_{xy}^{k+1} - \mathbf{v}_{xy}^k\|_2^2 \right\} \\
\mathbf{u}_{xt}^{k+1} &= \min_{\mathbf{u}_{xt}} \left\{ \|\mathbf{u}_{xt}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u}_{xt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt}^{k+1} - \mathbf{v}_{xt}^k\|_2^2 \right\} \\
\mathbf{u}_{yt}^{k+1} &= \min_{\mathbf{u}_{yt}} \left\{ \|\mathbf{u}_{yt}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u}_{yt} - 2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt}^{k+1} - \mathbf{v}_{yt}^k\|_2^2 \right\} \\
\mathbf{u}^{k+1} &= \min_{\mathbf{u}} \left\{ \|\mathbf{u}\|_1 + \frac{\mu}{2} \cdot \|\mathbf{u} - \lambda_{L1} \cdot \mathbf{g}^{k+1} - \mathbf{v}^k\|_2^2 \right\}
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mathbf{v}_{xx}^{k+1} &= \mathbf{v}_{xx}^k + \mathbf{g}_{xx}^k - \mathbf{u}_{xx}^k \\
\mathbf{v}_{yy}^{k+1} &= \mathbf{v}_{yy}^k + \mathbf{g}_{yy}^k - \mathbf{u}_{yy}^k \\
\mathbf{v}_{tt}^{k+1} &= \mathbf{v}_{tt}^k + \lambda_t \cdot \mathbf{g}_{tt}^k - \mathbf{u}_{tt}^k \\
\mathbf{v}_{xy}^{k+1} &= \mathbf{v}_{xy}^k + 2 \cdot \mathbf{g}_{xy}^k - \mathbf{u}_{xy}^k \\
\mathbf{v}_{xt}^{k+1} &= \mathbf{v}_{xt}^k + 2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt}^k - \mathbf{u}_{xt}^k \\
\mathbf{v}_{yt}^{k+1} &= \mathbf{v}_{yt}^k + 2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt}^k - \mathbf{u}_{yt}^k \\
\mathbf{v}^{k+1} &= \mathbf{v}^k + \lambda_{L1} \cdot \mathbf{g}^k - \mathbf{u}^k
\end{aligned} \tag{15}$$

Using **Eqs. (13–15)**, the COP is split into three steps (smooth convex). Then, the final iterative steps can be written as follows:

$$\begin{aligned}
\mathbf{g}^{k+1} &= \text{iFFT} \left(\frac{\text{FFT}(\mathbf{L}_{sHe}^k) + \frac{\lambda}{\mu} \cdot \text{FFT}(\mathbf{f} - \mathbf{b})}{\text{FFT}(\nabla_{xx}^2 + \nabla_{yy}^2 + \lambda_t^2 \cdot \nabla_{tt}^2 + 4 \cdot \nabla_{xy}^2 + 4\lambda_t \cdot \nabla_{xt}^2 + 4\lambda_t \cdot \nabla_{yt}^2) + \lambda_{L1}^2 + \frac{\lambda}{\mu}} \right), \\
\mathbf{L}_{sHe}^k &= \left\{ \begin{aligned} &\nabla_{xx}^T (\mathbf{u}_{xx}^k - \mathbf{v}_{xx}^k) + \nabla_{yy}^T (\mathbf{u}_{yy}^k - \mathbf{v}_{yy}^k) + \lambda_z \cdot \nabla_{tt}^T (\mathbf{u}_{tt}^k - \mathbf{v}_{tt}^k) + 2 \cdot \nabla_{xy}^T (\mathbf{u}_{xy}^k - \mathbf{v}_{xy}^k) \\ &+ 2\sqrt{\lambda_t} \cdot \nabla_{xt}^T (\mathbf{u}_{xt}^k - \mathbf{v}_{xt}^k) + 2\sqrt{\lambda_t} \cdot \nabla_{yt}^T (\mathbf{u}_{yt}^k - \mathbf{v}_{yt}^k) + \lambda_{L1} (\mathbf{u}^k - \mathbf{v}^k) \end{aligned} \right\}
\end{aligned} \tag{16}$$

where the FFT (iFFT) denotes the (inverse) fast Fourier transforms, ∇_{xx} is the two-order derivation operator in the x -direction: $\nabla_{xx} = [1, -2, 1]$, and ∇_{yy} , ∇_{tt} , ∇_{xy} , ∇_{xt} , ∇_{yt} are similarly defined:

$$\begin{aligned}
\mathbf{u}_{xx}^{k+1} &= \begin{cases} \mathbf{g}_{xx}^{k+1} + \mathbf{v}_{xx}^k - \frac{1}{\mu}, & \mathbf{g}_{xx}^{k+1} + \mathbf{v}_{xx}^k \in \left(\frac{1}{\mu}, \infty\right) \\ 0, & \mathbf{g}_{xx}^{k+1} + \mathbf{v}_{xx}^k \in \left(-\frac{1}{\mu}, \frac{1}{\mu}\right) \\ \mathbf{g}_{xx}^{k+1} + \mathbf{v}_{xx}^k + \frac{1}{\mu}, & \mathbf{g}_{xx}^{k+1} + \mathbf{v}_{xx}^k \in \left(-\infty, -\frac{1}{\mu}\right) \end{cases} \\
&= \text{shrink}\left(\mathbf{g}_{xx}^{k+1} + \mathbf{v}_{xx}^k, \frac{1}{\mu}\right) \\
\mathbf{u}_{yy}^{k+1} &= \text{shrink}\left(\mathbf{g}_{yy}^{k+1} + \mathbf{v}_{yy}^k, \frac{1}{\mu}\right) \\
\mathbf{u}_{tt}^{k+1} &= \text{shrink}\left(\lambda_t \cdot \mathbf{g}_{tt}^{k+1} + \mathbf{v}_{tt}^k, \frac{1}{\mu}\right) \\
\mathbf{u}_{xy}^{k+1} &= \text{shrink}\left(2 \cdot \mathbf{g}_{xy}^{k+1} + \mathbf{v}_{xy}^k, \frac{1}{\mu}\right) \\
\mathbf{u}_{xt}^{k+1} &= \text{shrink}\left(2\sqrt{\lambda_t} \cdot \mathbf{g}_{xt}^{k+1} + \mathbf{v}_{xt}^k, \frac{1}{\mu}\right) \\
\mathbf{u}_{yt}^{k+1} &= \text{shrink}\left(2\sqrt{\lambda_t} \cdot \mathbf{g}_{yt}^{k+1} + \mathbf{v}_{yt}^k, \frac{1}{\mu}\right) \\
\mathbf{u}^{k+1} &= \text{shrink}\left(\lambda_{L1} \cdot \mathbf{g}^{k+1} + \mathbf{v}^k, \frac{1}{\mu}\right)
\end{aligned} \tag{17}$$

Using **Eqs. (15–17)**, the COP can easily be resolved.

Supplementary Note 4.2 | Iterative deconvolution.

To achieve the final solution $\mathbf{x}$, we iterative minimize the **Eq. (10)**. For the data with a high SNR, we applied the Landweber (LW) deconvolution³³, in which the gradient descent method with Nesterov momentum acceleration was used³⁴:

$$\begin{aligned}
\mathbf{y}^{(j+1)} &= \mathbf{x}^{(j+1)} + \beta_k (\mathbf{x}^{(j)} - \mathbf{x}^{(j-1)}) \\
\mathbf{x}^{(j+1)} &= \mathbf{y}^{(j+1)} + \mathbf{A}^T (\mathbf{g} - \mathbf{A} \mathbf{y}^{(j+1)}) \\
\gamma^j &= \frac{1}{2} \left(4\gamma_{j-1}^2 + \gamma_{j-1}^4 \right)^{\frac{1}{2}} - \gamma_{j-1}^2, \\
\beta_j &= 1 - \gamma_{j-1}^{-1}
\end{aligned} \tag{18}$$

where $\mathbf{g}$ is the image after sparsity reconstruction and $\mathbf{x}^{j+1}$ is the image after $j+1$ iterations. Iteration is terminated at an early stage to constrain the resulting data to avoid artifacts³⁵. The numbers of deconvolution iterations are listed in the parameter table in the user manual of **Supplementary Software**.

For the low SNR data corrupted with excessive noise, we used a Bayesian-based Richardson Lucy (RL)^{2,3} deconvolution algorithm to process these images. The accelerated RL-based deconvolution³⁶ algorithm

follows:

$$\begin{aligned}
\mathbf{y}^{j+1} &= \mathbf{x}^j \cdot \left(\mathbf{h}^T \cdot \frac{\mathbf{g}}{\mathbf{h} \cdot \mathbf{x}^j} \right) \\
\mathbf{v}^j &= \mathbf{x}^{j+1} - \mathbf{y}^j \\
\alpha^{j+1} &= \frac{\sum \mathbf{v}^j \cdot \mathbf{v}^{j-1}}{\sum \mathbf{v}^{j-1} \cdot \mathbf{v}^{j-1}} \quad , \\
\mathbf{x}^{j+1} &= \mathbf{y}^{j+1} + \alpha^{j+1} \cdot (\mathbf{y}^{j+1} - \mathbf{y}^j)
\end{aligned} \tag{19}$$

where $\mathbf{g}$ is the image after sparsity reconstruction and $\mathbf{x}^{j+1}$ is the image after $j+1$ iterations. Specifically, the adaptive acceleration factor (α) was introduced by Andrews *et al.*³⁶ and represented the computational iteration step length, which can be estimated directly from experimental results. The accelerated RL algorithm convergences rapidly, and the required number of iterations 5~10 times less than that of the classical RL method.

Supplementary Note 4.3 | Background estimation.

To estimate the background $\mathbf{b}$ appearing in **Eq. (8)**, we modified the iterative wavelet transform method²⁹. For the data that contains weak background or even no background (e.g., TIRF imaging mode), we remove such background estimate operation, i.e., set $\mathbf{b}$ as zero (corresponded to the *No Background* in the **Supplementary Table 4**) to avoid removal of information. Images from *low-dose illumination* exhibited a low and stable noise-like distribution of background fluorescence only. Thus, values over the mean value of the image were set to zero directly, and the resulting *residual image* was used for the following background estimation. Under other conditions, we estimated and removed the background as follows:

i. Original images from *high-dose illumination* containing a strong background fluorescence signal originating from out-of-focus emissions and cellular autofluorescence were regarded as the *input images* in the flowchart. The background can be iteratively estimated from the lowest frequency wavelet bands associated with the *input images* (Wiener filter). To extract the lowest band of the frequency domain, we used 2D Daubechies-6 wavelet filters to decompose the signal up to the 7th level.

ii. To prevent accidental removal of minor useful signals, we performed an inverse wavelet transform on the lowest band of the frequency information to the spatial domain and compared the result with half of the square root of the *input image*. We merged these two images by keeping the minimum values at each pixel. This operation removed high-intensity pixels resulting from inaccurate background estimation.

iii. The resulting *low band off-peak background* estimation data in **ii** were used again as the *input image* in **i**, and the abovementioned loop was repeated several times. In general, we set the number of iterations

rounds to three to estimate the background with minimal contribution from the real fluorescent signals.

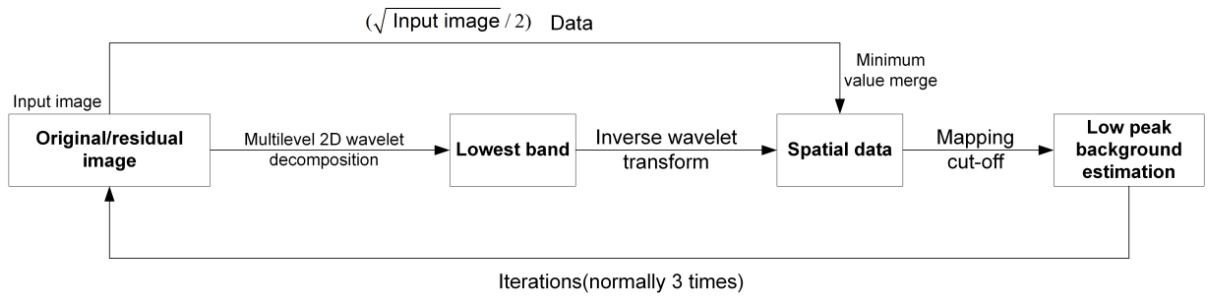
Supplementary Note 4.4 | Upsampling operation.

The pixel size could be the factor that limits SR image reconstruction. For example, although the Sparse-SIM's theoretical resolution could reach ~60 nm, the 32.5 nm pixel size may limit its resolution due to the insufficient spatial sampling. Likewise, when we used the EMCCD camera for the SD-SIM, 94 nm pixel size severely limited the resolution that could be achieved by the sparse deconvolution (down to ~90 nm). Therefore, we introduced an upsampling operation in the reconstruction pipeline.

As a result, the loss function in **Eq. (8)** for the sparse reconstruction can be rewritten as follows:

$$\arg \min_{\mathbf{g}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{D}\mathbf{g}\|_2^2 + R_{\text{Hessian}}(\mathbf{g}) + \lambda_{L1} \|\mathbf{g}\|_1 \right\}, \quad (20)$$

where $\mathbf{D}$ is the downsampling matrix. In practical, we used the Fourier interpolation to upsample the 2D-SIM data (pad the zeros out of SIM OTF)³⁷, and the spatial upsampling method (pad zeros around each pixel) to upsample the SD-SIM data, which was less sensitive to the SNR and less prone to snowflake artifacts.



Supplementary Fig. 5 | Flowchart of background estimation based on multilevel wavelets.

Key messages:

Algorithm | Sparse deconvolution.

Input: time-lapse/volumetric image stack $\mathbf{f}$

STEP 0.

if background estimation:

Background estimation using the procedure in **Supplementary Note 4.3**.

else:

$\mathbf{b} = 0$.

Output: $\mathbf{b}$

Input: $\mathbf{f}$, $\mathbf{b}$, $\mathbf{A}$

STEP 1.

if upsampling:

Minimize **Eq. (20)**:

$$\arg \min_{\mathbf{g}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{D}\mathbf{g}\|_2^2 + R_{\text{Hessian}}(\mathbf{g}) + \lambda_{\text{L1}} \|\mathbf{g}\|_1 \right\}.$$

else:

Minimize **Eq. (8)**:

$$\arg \min_{\mathbf{g}} \left\{ \frac{\lambda}{2} \|\mathbf{f} - \mathbf{b} - \mathbf{g}\|_2^2 + R_{\text{Hessian}}(\mathbf{g}) + \lambda_{\text{L1}} \|\mathbf{g}\|_1 \right\}.$$

STEP 2.

Minimize **Eq. (9)**:

$$\arg \min_{\mathbf{x}} \left\{ \|\mathbf{g} - \mathbf{A}\mathbf{x}\|_2^2 \right\}.$$

Output: $\mathbf{x}$

Supplementary Note 5 | Using the sparse and continuity prior knowledge improves resolution and contrast.

By adding the sparsity and the continuity prior knowledge to constrain the iterative deconvolution, we were able to enhance the resolution of synthetic images of different structures (**Supplementary Fig. 28-30, 35**), images captured by the wide-field microscope (**Supplementary Fig. 31**), and by SIM in resolving the DNA origami structures (**Fig. 1**), paired lines in the Argo slides (**Fig. 2a-2b, Extended Data Fig. 2**), fluorescence-labeled nuclear pores (**Fig. 2c-2g, Supplementary Fig. 36, 37**), actin mesh pores (**Fig. 4a-4i**), caveolae (**Fig. 4j-4l**), and intricate intermediate fusion pores (**Fig. 4o-4p**). However, most of the structures resolved were of regular shapes. To test whether the sparse deconvolution could be used to reconstruct erratic structures, we have synthesized four pairs of bisected-whole rings 120/80/40/0 nm apart in the same FOV (**Supplementary Fig. 6a**), convolved with a PSF of 110 nm, and corrupted with Poisson noise and 5% Gaussian noise (**Supplementary Fig. 6b**). The conventional LW deconvolution method reconstructed oversharpened puncta corrupted with excessive artifacts (**Supplementary Fig. 6d**). In contrast, although results of sparse deconvolution with two sets of fidelity and sparsity parameters (**Supplementary Fig. 6e, 6f**) did not generate identical images, it suppressed artifacts in both cases and resolved irregular bisected rings persisted even when the inter-ring distances were smaller than the SIM resolution. Admittedly, neither image was exactly the same as the convolution of the ground-truth image with the 45 nm PSF. Nevertheless, irregular bisected rings emerged even when the inter-ring distance was smaller than the SIM resolution in both cases. Similarly, irregular, periodic lines were resolved after the sparse deconvolution (**Supplementary Fig. 6g-6j**).

As the RL deconvolution failed to improve resolution in images corrupted with noise (**Supplementary Fig. 2**), we wondered how SNR conditions affect our sparse deconvolution performance. Thus, we synthesized parallel paired lines from 10 nm to 120 nm apart, convolved them with the PSF with 90 nm FWHM, and systematically added Poisson and Gaussian noise of different amplitudes (**Supplementary Fig. 7**). While the RL deconvolution failed to improve the resolution, the sparse deconvolution consistently outperformed the former in separating paired lines 50~60 nm apart. However, we also noticed that parallel lines under excessive noise sometimes manifested as twisted and fluctuated fluorescence profiles (**Supplementary Fig. 7**), indicative of artifacts. To quantitatively evaluate the quality of reconstruction, we have proposed a line restoration quality (LRQ) criterion:

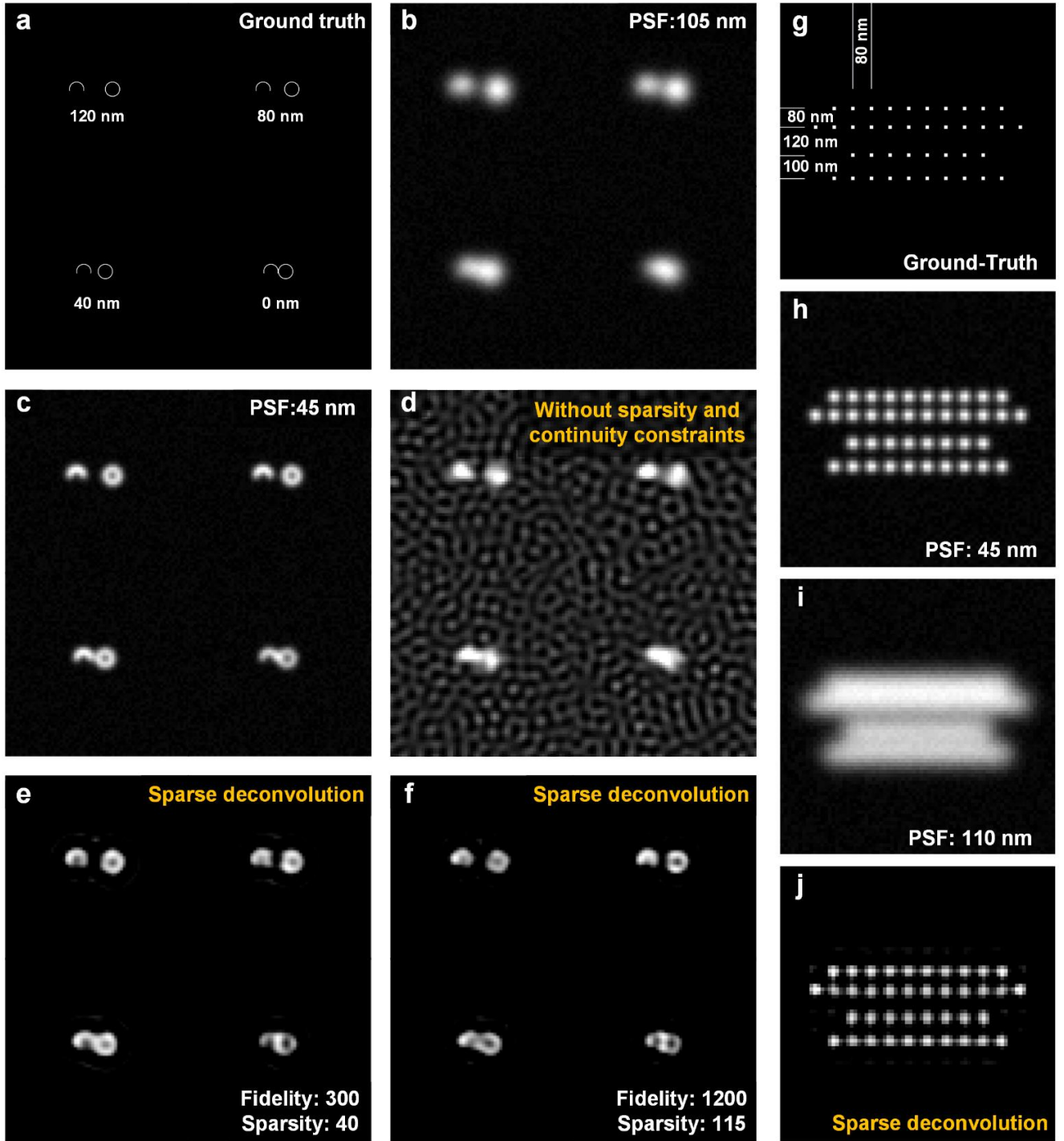
$$LRQ = \frac{2 \times \text{Avg}(\text{Region}_0)}{\text{Avg}(\text{Region}_1) + \text{Avg}(\text{Region}_2)} . \quad (21)$$

The *Region₁*, *Region₂*, and *Region₀* represent the parallel lines and the region in between, and *Avg* calculates the mean intensity of pixels within the corresponding area. Because the LRQ value always decayed rapidly from a stable high plateau to 0.4 as the distance between paired lines decreased, we selected it as the threshold of successful reconstruction (**Supplementary Fig. 8**). According to this criterion, RL deconvolution stopped at resolving lines 100 nm apart at 50% Gaussian noise, while the sparse deconvolution could separate lines 50 nm apart. Increasing Gaussian noise limited the ability of the sparse deconvolution in improving resolution. With 100% Gaussian noise, the sparse deconvolution could reach 60~70 nm resolution. With 150% Gaussian noise, the sparse deconvolution could not faithfully extract information beyond the OTF (**Supplementary Fig. 8**). However, compared to the RL deconvolution, it significantly increased the contrast of high-frequency information within the OTF and enabled lines 100 nm apart to be separated (**Supplementary Fig. 7**).

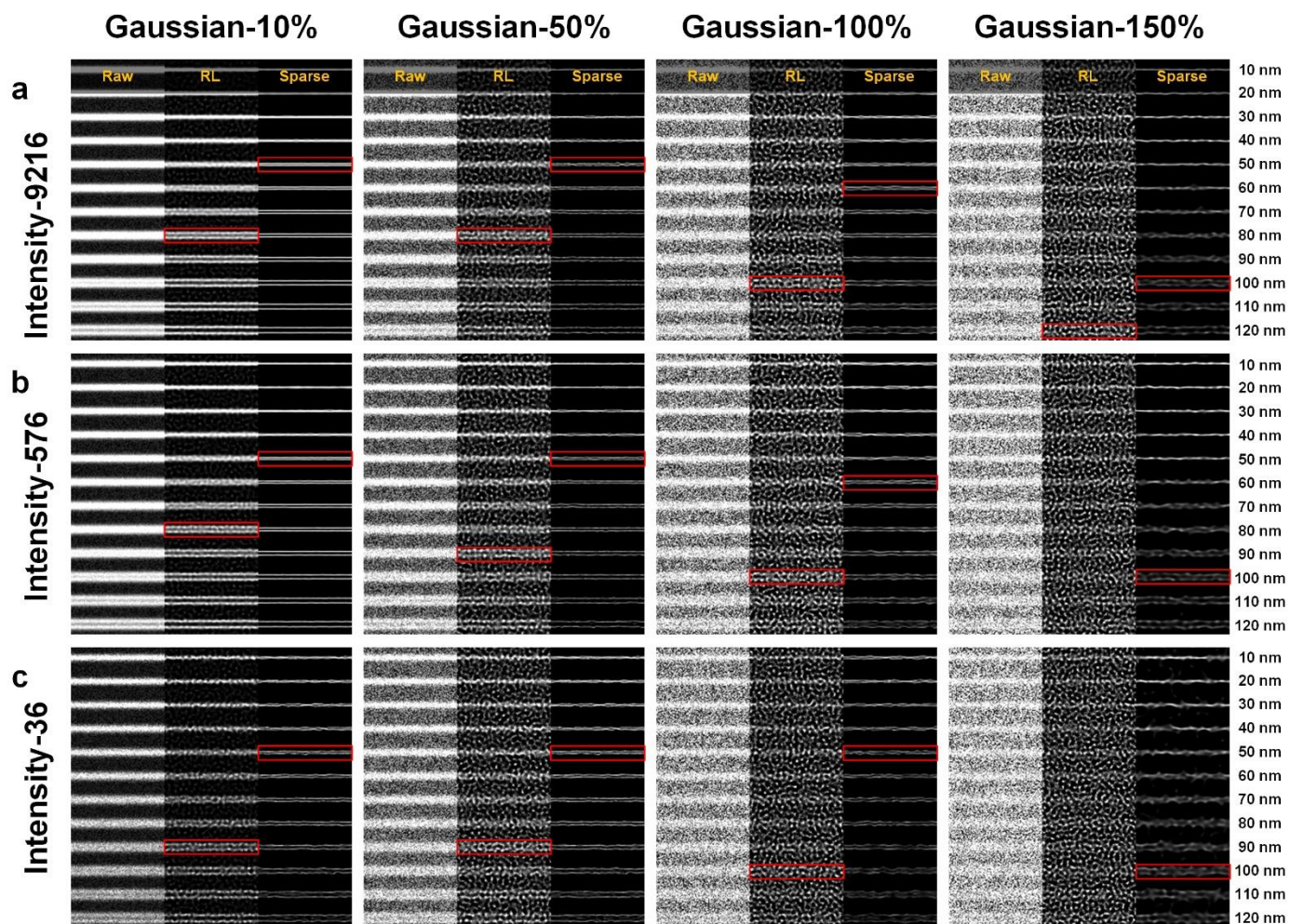
These data were also recapitulated in images of real samples taken by microscopes. Under the regular 200 mW illumination and 10 ms exposure (1 SNR), 2D-SIM, RL-SIM, and Sparse-SIM resolved fluorescent lines carved in the commercial Argo-SIM slide 150, 120, and 60 nm apart, respectively (**Extended Data Fig. 2**). Sparse-SIM resolution deteriorated to 90 nm under 25 mW illumination and 10 ms exposure (1/8 SNR) and returned to the optical limit of ~120 nm under 25 mW illumination and 5 ms exposure (1/16 SNR). Despite the failure to extend resolution at this extremely low SNR, sparse deconvolution was superior to the 2D-SIM and RL-SIM in generating reconstructions of good contrast and minimal artifacts. Finally, in the live COS-7 cell weakly expressed with tubulin-EGFP, images captured by SD-SIM were of low SNR (**Supplementary Fig. 9**). Under such a circumstance, the FRC resolutions of images under the SD-SIM and RL SD-SIM were similar (~250 nm). In contrast, although Sparse SD-SIM did not increase resolution beyond the system OTF, it increased the image contrast to ensure the maximal resolution allowed by the optics (~130 nm).

Key messages:

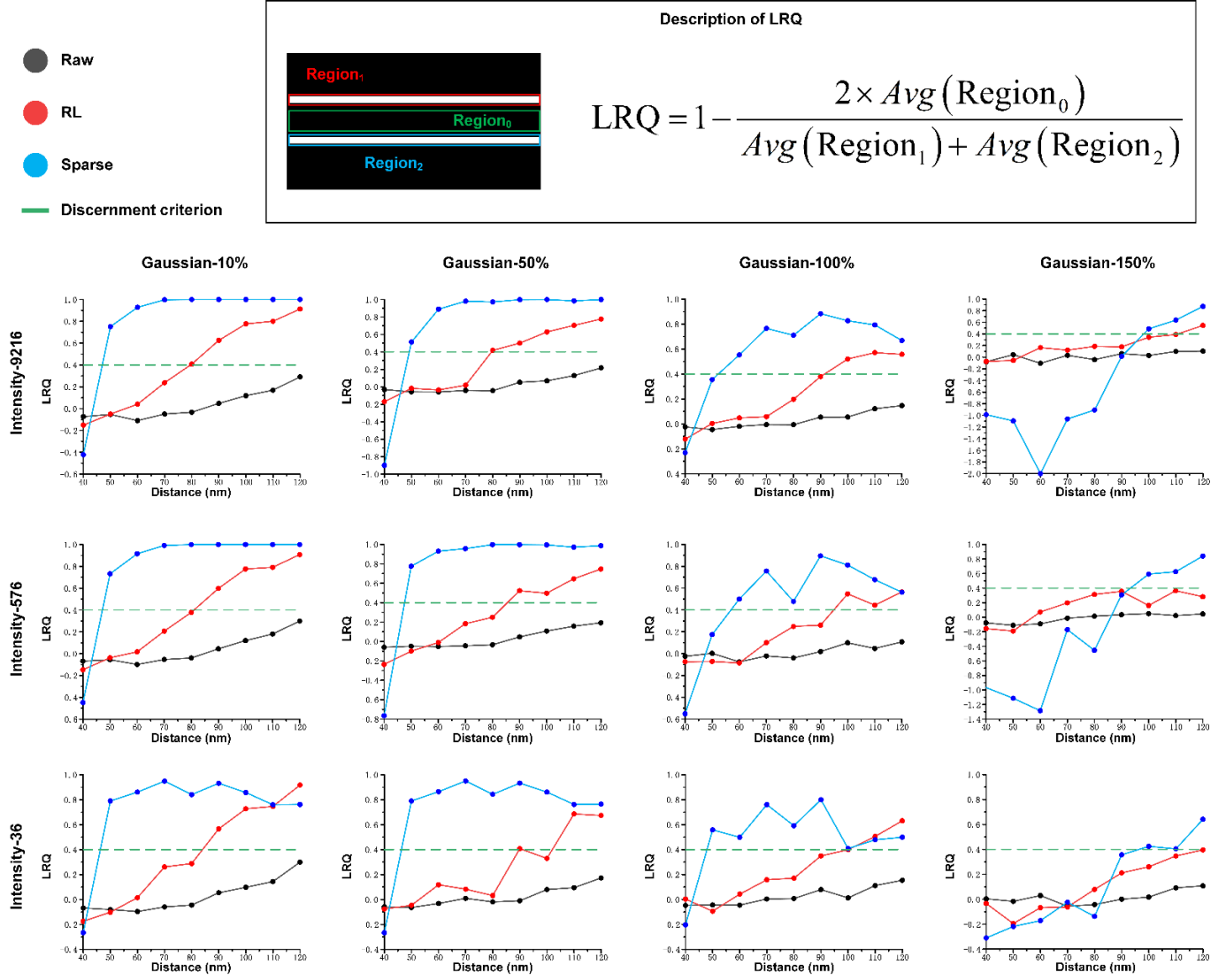
- 1) Reconstructing with the sparse and the continuity prior constrains the following RL deconvolution to suppress artifacts and increase resolution.
- 2) In images corrupted with excessive noise, the sparse deconvolution pipeline improves the contrast but not the resolution.



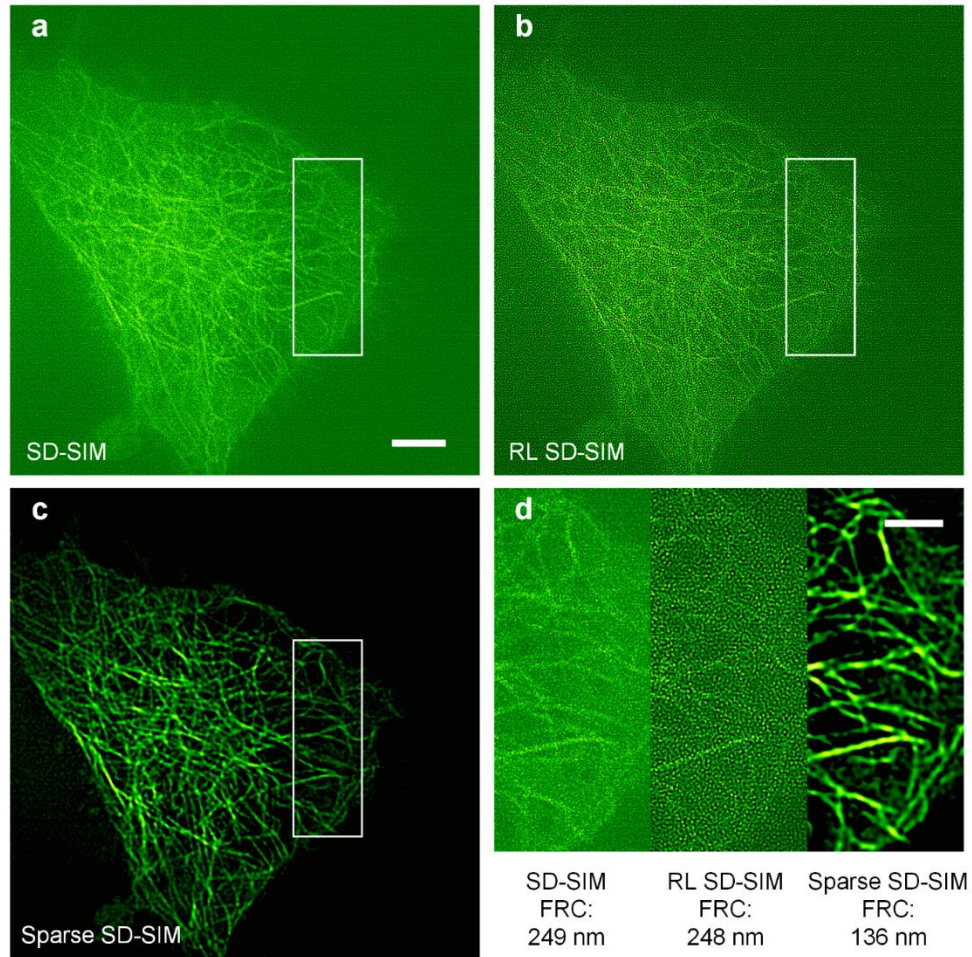
Supplementary Fig. 6 | Partial ring/line simulation. (a) We created synthetic partial and complete ring structures with an 80 nm diameter of 120 nm, 80 nm, 40 nm, and 0 nm apart as ground-truth. (b, c) The ground-truth image in (a) was convolved with a PSF with FWHM of either 110 nm (b) or 45 nm (c), and such images were subsequently subsampled 16 times to obtain the pixel size of 16 nm, and corrupted with Poisson noise and 5% Gaussian noise. (d) The images in (b) were reconstructed by LW deconvolution without sparsity and continuity constraints. (e, f) The images in (b) were reconstructed by Sparse deconvolution under different parameters. The fidelity and sparsity values were set as 300 and 40 for results in (e), and the results for 1200 and 115 parameter conditions were shown in (f). (g) We created four synthetic partial lines as ground-truth. Each partial line consists of multiple narrowed placed dots with a distance of 80 nm. The ground-truth image was convolved with PSF with the FWHM of either 45 nm (h) or 110 nm (i), and such images were subsequently subsampled 16 times, and corrupted with Poisson noise and 5% Gaussian noise. (j) Image in (i) was processed by sparse deconvolution.



Supplementary Fig. 7 | Effects of the combined Poisson-Gaussian noise on the paired lines resolved by the RL and Sparse deconvolution. We synthesized a ground-truth image containing ten parallel paired lines departed by different distances (from 10 nm to 100 nm), and convoluted it with a 90 nm FWHM PSF. We set the maximum fluorescence intensities of lines to 9216 (a), 576 (b), and 36 (c) a.u., respectively, and added a Poisson-distributed noise plus 10% (first column), 50% (second column), 100% (third column), and 150% (fourth column) Gaussian distributed noise (compared to the maximum intensities of lines). After that, we deconvolved these raw images (left, labeled as 'Raw') with RL (middle, labeled as 'RL') or Sparse deconvolution (right, labeled as 'Sparse').



Supplementary Fig. 8 | The line restoration quality (LRQ) values of images in Supplementary Fig. 7. *Avg* represents the averaged intensity of the pixels in the corresponding region. We use the value 0.4 (green lines) as the threshold (discernment criterion) to determine whether the lines are separated successfully.



Supplementary Fig. 9 | Sparse deconvolution increases the contrast of microtubules of low SNR captured by the SD-SIM. (a-c) A live COS-7 cell labeled with Tubulin-EGFP captured by SD-SIM. The raw image was shown in (a), which was deconvolved by the RL (b) or sparse deconvolution (c). (d) Magnified views of white boxes in (a-c). The FRC resolution of raw SD-SIM, RL SD-SIM, and Sparse SD-SIM were estimated to be 249 nm, 248 nm, and 136 nm, respectively. Scale bars: (a) 5 μm ; (d) 3 μm .

Supplementary Note 6 | Parameters in the sparse deconvolution software.

In the **Supplementary Software** user interface (UI), we included thirteen parameters to adapt to different hardware environments, experimental conditions, and fluorescence microscopes (**Supplementary Fig. 10**). To simplify the usage of this software, we have classified them into three categories: *fixed parameters*, *image property parameters*, and *content-aware parameters*.

The system hardware determines *fixed parameters*, and the *image property parameters* are associated with image quality, such as high-or-low SNR and strong-or-weak background. Ten parameters in these two categories are selected based on the system and the image property and need little tuning. The three parameters left belong to *Content-aware parameters*. They need to be adjusted carefully to achieve the optimal reconstruction results. We introduced a four-step workflow to finetune these *content-aware parameters* (shown in **Supplementary Fig. 10** and **14**), which serves as a general guide. We provided the detailed step-by-step illustration in **Supplementary Note 8**.

The parameters description:

Fixed parameters:

Pixel size: The physically equivalent pixel size of the final images.

Wavelength: The emission wavelength of the fluorescence probes.

Effective numerical aperture: Effective NA of the optical system based on its spatial resolution.

3D imaging: This is the option to choose whether the input images are volumetric or not.

GPU acceleration: This is the option to choose whether the calculation uses CUDA-GPU or not.

Image property parameters:

Background (Step 0): Based on the actual background fluorescence of the images processed, users can choose **no background** or the background under high-dose (*HI*) or low-dose of illumination (*LI*). Both the types *HI* and *LI* have *weak* and *strong* magnitude options. Thus five options of background are: *No*, *Weak-HI*, *Strong-HI*, *Weak-LI*, and *Strong-LI*. We want to emphasize that the **background estimation step** is an **optional operation** that may not be necessary. For example, when we processed Ca^{2+} imaging data in **Extended Data Fig. 17**, we selected **no** options to avoid removing the baseline signal. We have also provided parameters used for processing images in this manuscript for reference in **Supplementary Table 4**.

Upsampling (Step 0): In conditions of inadequate Nyquist sampling, we manually up-sample images to achieve the theoretical resolution increase posed by the sparse deconvolution. We usually choose the spatial upsampling method for **low-SNR** images and the Fourier upsampling method for **high-SNR** ones.

t (z)-axial continuity (Step 1): This parameter is for adjusting the continuity along the input dataset's *t* or *z*-

axis. We set the $t(z)$ -axial continuity less than or equal to 1 to avoid temporal blurring. For the fast time-lapse imaging, while the fidelity is less than 100, the $t(z)$ -axial continuity is usually assigned as one-hundredth of the fidelity. If the object being imaged has undergone fast movements, we need to set this parameter to a small number (0.1) or even zero to avoid causing motion artifacts.

Sparse iteration times (Step 1): Usually, we set it to 100 iterations. If spatial over-sampling is used, we need to increase the number to 200 or 300 to ensure the sparsity reconstruction convergence.

Iterative deconvolution (Step 2): We usually choose the Richardson-Lucy algorithm (RL) to deconvolve **low-SNR** images and the LandWeber deconvolution (LW) to deconvolve **high-SNR** ones.

Content-aware parameters:

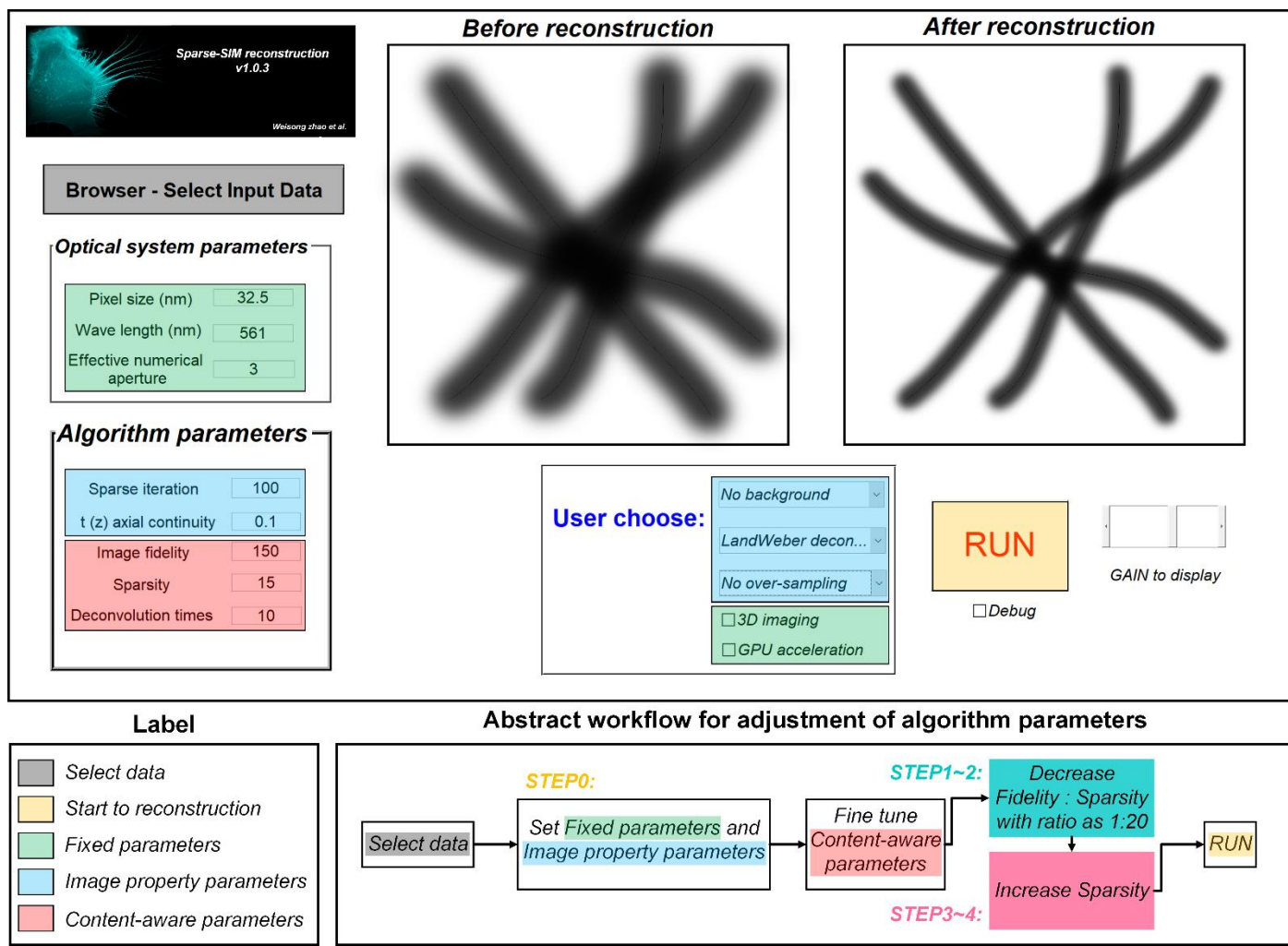
Image fidelity (Step 1): This parameter denotes the distance between the image before and after the sparse reconstruction and is the inverse of the **xy continuity**. Usually, we use a large value (1000~300) for high SNR images.

Sparsity (Step 1): This parameter represents the relative sparsity constraint enforced on the reconstruction (Step 1). Usually, we pre-set this value to **one-tenth** of the image fidelity term (**Supplementary Fig. 5**). However, an unnecessary high sparsity value may remove weak signals. Thus we need to finetune this parameter back-and-forth based on the final deconvolution result.

Iterative deconvolution times (Step 2): This parameter sets the times for the post iteration deconvolution. Because the LW method is slower than the vector extrapolation version of the RL method in reaching convergence, we choose 5~15 iteration times for the RL algorithm and 30~50 for the LW algorithm.

Key messages:

The ten parameters in the first two categories are primarily determined by the optical system and image property and need little tuning. Only the three content-aware parameters need to be adjusted back-and-forth carefully by visual examination of the reconstruction results.



Supplementary Fig. 10 | The sparse deconvolution software. The top panel shows the software UI, and the bottom panel shows the simplified workflow of parameters adjustment. The detailed workflow is shown in Supplementary Fig. 14.

Supplementary Note 7 | Optimal values of the sparsity and the fidelity under different conditions.

Among content-aware parameters, we usually set the iterative deconvolution time as a constant of 10, 50 for the RL, LW deconvolution method, respectively. As we used the fidelity term ($1/\text{continuity}$) to adjust the weight of continuity in the software, we only need to finetune the sparsity and the fidelity parameters in practice.

By collecting all fidelity and sparsity values carefully chosen for images used in this work, we find that the sparsity value roughly follows a linear relationship with the fidelity one (1:5~1:20, **Supplementary Fig. 11-13**). Although we do not understand the underlying mechanism, it represents the suitable range for processing different image datasets.

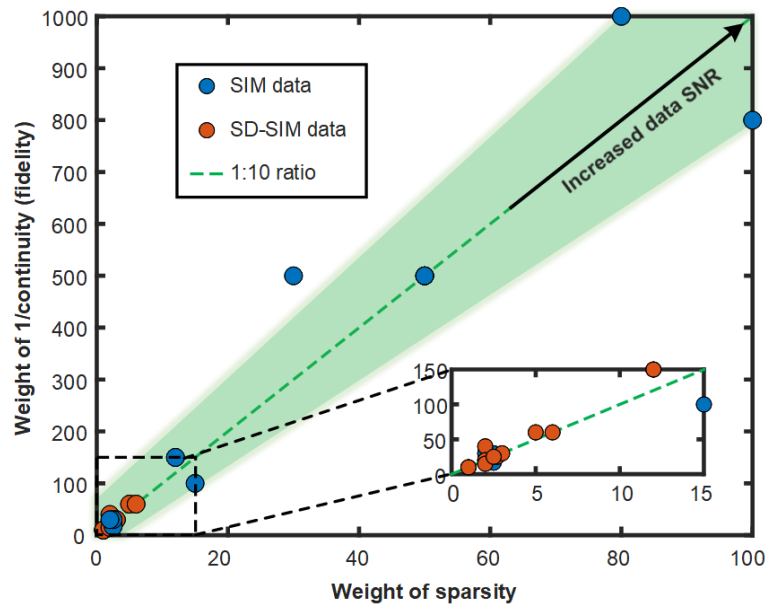
Because we want to improve the resolution in deconvolving high SNR images and the contrast in deconvolving low SNR images, these two parameters needed to be adjusted carefully. To explore the effects of different sparsity and fidelity values on reconstruction results, we synthesized ring-shaped and punctated structures of 80 nm in diameters, convoluted them with a PSF of 110 nm in full width at half maximum (FWHM) (**Supplementary Fig. 12a**), and corrupted them with the noise of various amplitudes (**Supplementary Fig. 12b**). With 2% noise, sparse deconvolution resolved both the ring-shaped and the punctuated structures with high fidelity and sparsity values, which approximately followed a 5:1 relationship (1000:200, 950:200, 400:80). However, at low fidelity numbers (smaller or equal to 100), the fluorescent ring was not resolved, indicating the failure to improve resolution (**Supplementary Fig. 12d**). With 10% noise, the ring became irregular using the fidelity: sparsity ratio of 1000:200 or 950:200, while a balance of 400:80 was optimal for the resolution enhancement. As the SNR of the image deteriorated, it was evident that large fidelity and sparsity values led to artifacts, while small numbers were better for the extraction of real structures from noise. Under 80% noise, a fidelity: sparsity ratio of 20:5 allowed the ring to be resolved with sharp edges. With 100% noise, we could only use small fidelity and sparsity values such as 10 and 2 to avoid artifacts. Despite the failure to improve resolution, we observed reduced background and much-improved contrast, which may help subsequent visualization, segmentation, tracking, and analysis.

Because the ground-truth for this synthetic image was known, we systematically examined different peak signal-to-noise ratios (PSNRs) obtained under different sets of fidelity and sparsity values (**Supplementary Fig. 13a**). If we selectively highlighted regions of PSNR larger than 25 (or 22.6 at 100% noise), the weight of fidelity linearly correlated with that of the sparsity within a narrow range at most noise levels (**Supplementary**

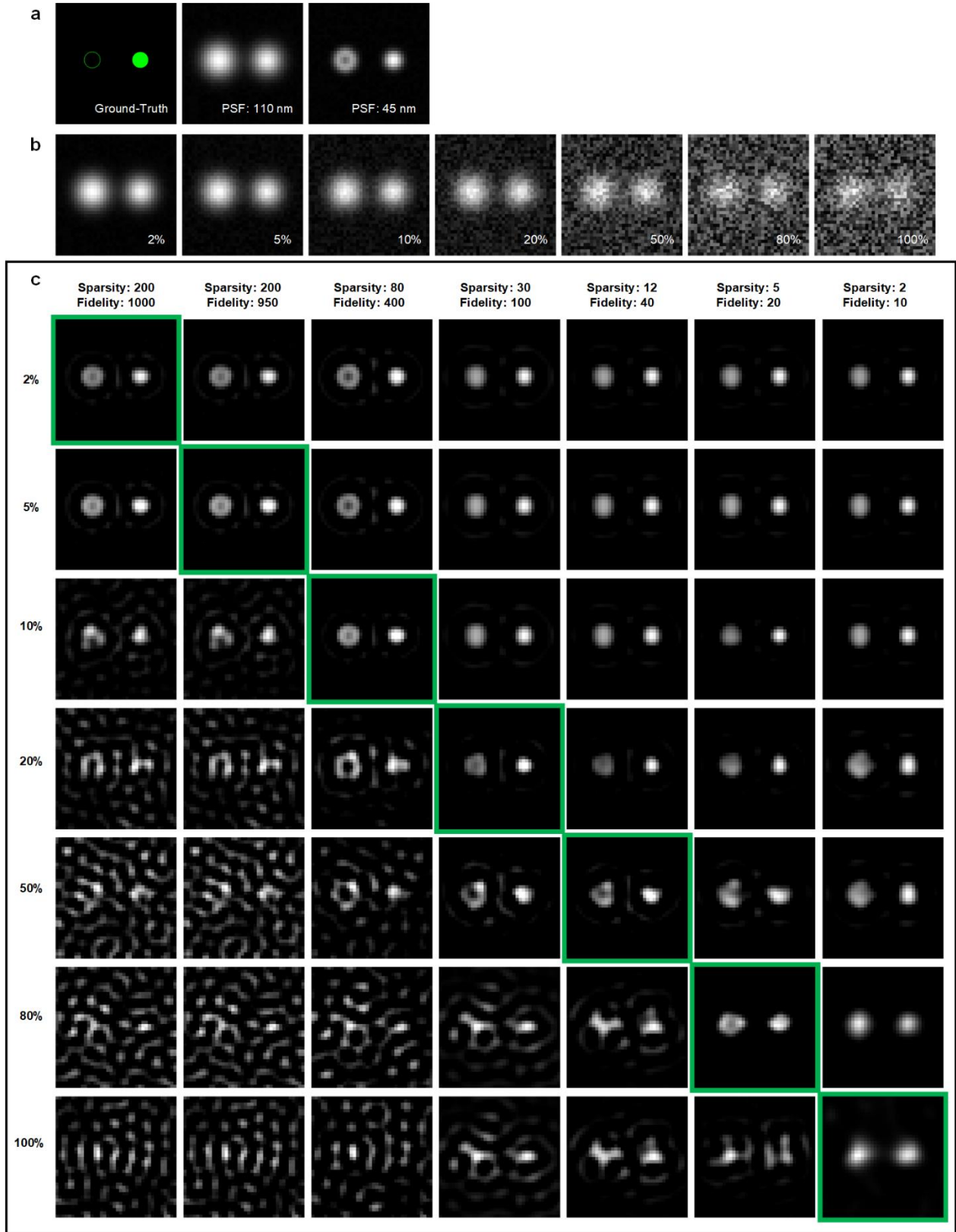
Fig. 13b). These data reinforce the roughly linear relationship between optimal choices of the fidelity and sparsity values.

Key messages:

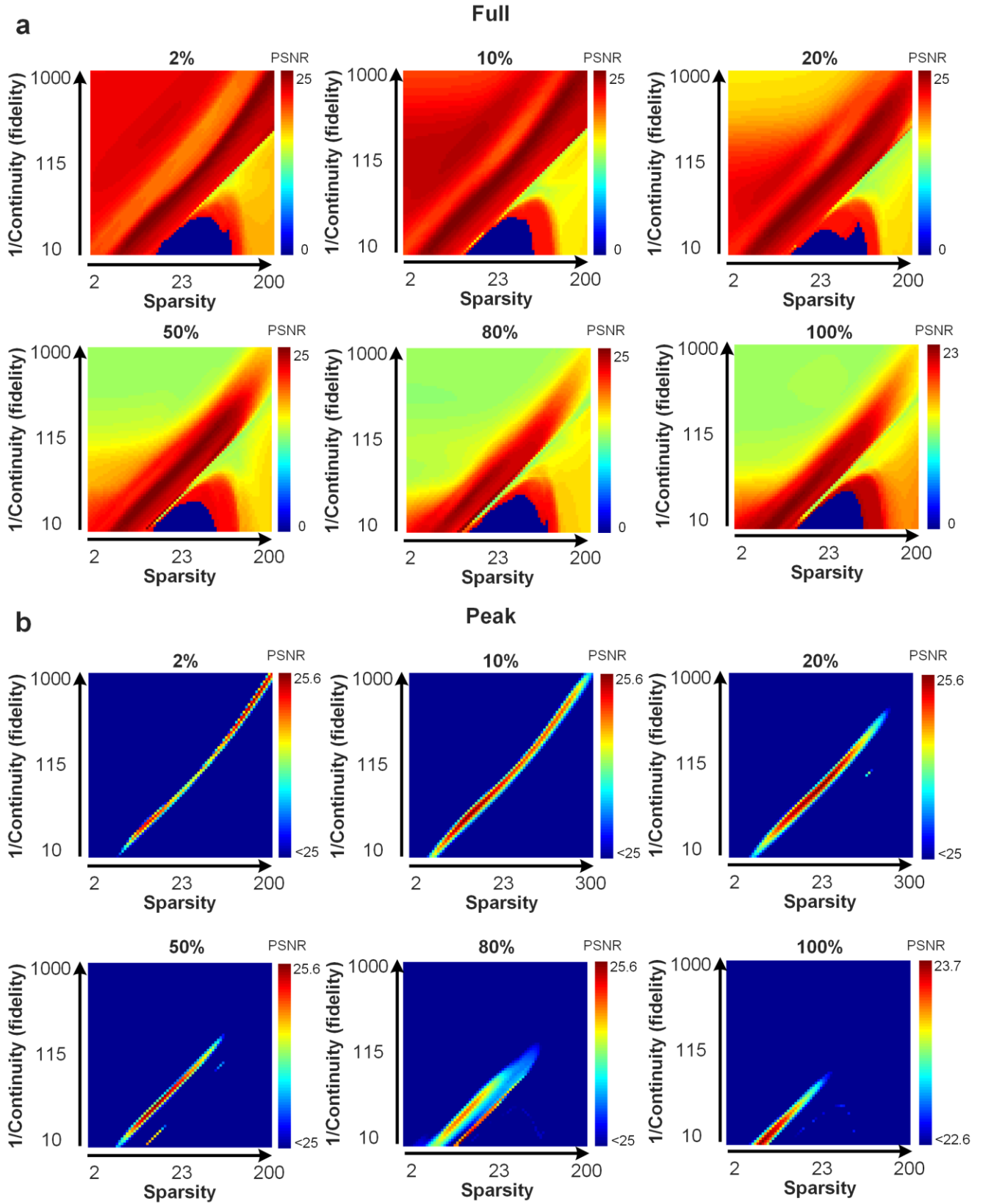
- 1) We only need to adjust the values of fidelity and sparsity carefully.
- 2) The optimal values of fidelity and sparsity follow an approximately linear relationship.
- 3) Images of high SNR afford large fidelity values, while low SNR images require small fidelity values.



Supplementary Fig. 11 | Linear relationship between optimal values of sparsity and fidelity. We have identified an approximately linear relationship of sparsity and fidelity ($\sim 1:5 \sim 1:20$) from the reconstruction results of SIM (blue) and SD-SIM (orange) in this work.



Supplementary Fig. 12 | Exploring optimal fidelity and sparsity choices for reconstructions of various SNR images. (a) We synthesized ring-shaped and punctuated structures of 80 nm in diameter as the ground-truth (left), convoluted them with a PSF of either 110 nm (middle) or 45 nm (right) in FWHM. (b) The image in the middle of (a) was subsequently subsampled 16 times (pixel sizes of 16 nm), and corrupted with Poisson noise plus 2%, 5%, 10%, 20%, 50%, 80%, and 100% Gaussian noise. (c) The 7×7 table of the reconstruction results. The column and row represent the noise amplitude and values of priors for the reconstruction, respectively.



Supplementary Fig. 13 | PSNR heat maps at different values of sparsity and fidelity. (a) The maps of the PSNR values (*c.f.*, images under different noise amplitudes in **Supplementary Fig. 12**) at different values of sparsity (row) and fidelity (column). (b) PSNR maps in (a) with values greater than 25 (2%, 5%, 10%, 20%, 50%, and 80% noise) or 22.6 (100% noise) at different conditions.

Supplementary Note 8 | Parameters finetuning

In this note, we intend to provide a comprehensive guide for the readers/users to process their data, including the detailed procedures to adjust the *Fidelity* and *Sparsity* values (**Supplementary Note 8.1**) and some step-by-step examples (**Supplementary Note 8.2**).

Supplementary Note 8.1 | Procedures for parameters finetuning.

STEP 0. Set the *Fixed parameters* and *Image property parameters* according to the microscope hardware, computing system, and properties of images collected as described in **Supplementary Note 6**.

STEP 1. Choose the initial *Fidelity* value according to the image SNR. We recommend starting from 1000 for high SNR images or 500 for low SNR ones. For starters, we also recommend choosing the *Sparsity* as one-twentieth of the *Fidelity*, which is to avoid over-filtering artifacts due to the inappropriately high *Sparsity* value.

STEP2. Adjust the *Fidelity* and *Sparsity* downwards simultaneously with the ratio maintained. Because the *Fidelity* is the opposite of the *xy continuity* term, its reduction constrains the reconstruction with more weight on the continuity. However, unnecessary small *Fidelity* values may over-smooth the structures and reduce the resolution. Therefore, we need to examine the reconstruction results and terminate the downwards adjustment once image over-smoothing is apparent.

STEP 3. Fine-tune *Sparsity* upwards. Large *Sparsity* values help us recover the high-frequency information on the one hand but also may over-filter the weak signals on the other hand. Therefore, we increase the weight of *Sparsity* until over-filtering artifacts emerge. Please note that adjusting the *Sparsity* value changes the ratio of *Fidelity* and *Sparsity*.

STEP 4. Stop at the value of *Sparsity* with the right balance of resolution and minimally removing weak fluorescence signals.

Supplementary Note 8.2 | Step-by-step examples.

Here, we provided two simulation examples and four experimental examples from imaging systems of 2D-SIM (*c.f.*, **Fig. 4a**), TIRF-SIM (*c.f.*, **Fig. 4j**), and SD-SIM (*c.f.*, **Extended Data Fig. 6**) microscopes, respectively, to illustrate the four-step parameters adjustments procedure in practice.

Examples for the finetuning of the *Fidelity* and *Sparsity* values

Example 1. *Sparse deconvolution of synthetic ring-shaped and punctuated structures corrupted with noise of different amplitudes* (**Supplementary Fig. 15**)

5% noise condition.

- 1) For synthetic ring-shaped and punctuated structures corrupted with 5% noise, we selected the **Fidelity** and **Sparsity** as 1000 and 50 as recommended.
- 2) Reducing the **Fidelity** from 1000 to 800 and maintaining the **Fidelity:Sparsity** ratio did not significantly change image quality. Therefore, we stopped the **Fidelity** at 1000.
- 3) Next, we increased the **Sparsity** while keeping the **Fidelity** stable. Once we raised the **Sparsity** to 150, the ring-shaped structure emerged. However, when the **Sparsity** was adjusted to 250, the FWHM of the ring membrane shrank to ~45 nm while the ring became interrupted. Therefore, we selected 1000 and 200 as the optimal values for the **Fidelity** and **Sparsity** for this image with low noise.

80% noise condition.

- 1) For synthetic ring-shaped and punctuated structures corrupted with 80% noise, we selected the **Fidelity** and **Sparsity** as 500 and 50 as recommended.
- 2) At this ratio, irregularly sharpened structures were apparent. Therefore, we reduced the **Fidelity** while maintaining the **Fidelity:Sparsity** ratio to suppress artifacts. The contrast of two fluorescence puncta increased at the 20:1 ratio, while reducing the **Fidelity** to an even smaller value did not improve further.
- 3) Thus we fixed the **Fidelity** to 20, and increased the **Sparsity** term. Once we raised the **Sparsity** to 6, the ring-shaped structure emerged, and even larger **Sparsity** led to over-sharpening artifacts. Therefore, we stopped at 20 and 6 as the optimal values for the **Fidelity** and **Sparsity** for this image with high noise.

Example 2. *Sparse deconvolution of dense actin filaments obtained by the 2D-SIM* (Supplementary Fig. 16)

- 1) For dense actin structures under the 2D-SIM, we selected the **Fidelity** and the **Sparsity** as 1000 and 50 as recommended. With these parameters, actin filaments were intermittent, and the snowflake-like artifacts emerged. Therefore, we needed to reduce the **Fidelity** (increase the continuity) to reduce artifacts caused by noise amplification.
- 2) We started with a large searching step for the general evaluation to enable rapid convergence and then gradually reduced the step size for fine adjustment. By decreasing the **Fidelity** (from 1000 to 150) and maintaining the **Fidelity:Sparsity** ratio, the snowflake-like artifacts were suppressed at the price of reduced resolution. When we selected 150 and 7 for the **Fidelity** and **Sparsity**, respectively, actin filaments were relatively continuous without much blurring.
- 3) Next, we kept the **Fidelity** to 150 and increased the **Sparsity** to further improve the resolution and contrast.

Once we selected a **Sparsity** value larger than 10, some weak actin filaments were trimmed off. Therefore, we set the **Fidelity** value of 150 and the **Sparsity** value of 10.

Example 3. Sparse deconvolution of the caveolae obtained by the TIRF-SIM (Supplementary Fig. 17)

- 1) For caveolae under the TIRF-SIM, we selected the **Fidelity** and **Sparsity** as 1000 and 50 as recommended. Interestingly, when we concomitantly reduced the **Fidelity** and the **Sparsity** values, the ring-shaped caveolae were blurred. Thus we stopped the **Fidelity** value at 1000.
- 2) Next, we kept the **Fidelity** to 1000 and increased the **Sparsity** to further improve the resolution and contrast. As we increased the **Sparsity** value, ring-shaped structures became more apparent. On the other hand, once we selected a **Sparsity** value of 130, the caveolae ring became interrupted. Therefore, we stopped the **Sparsity** value at 90 for optimal extraction of high-frequency caveolae structures.

Example 4. Sparse deconvolution of the actin filaments obtained by the SD-SIM (Supplementary Fig. 18)

- 1) Live-cell images obtained by our SD-SIM configurations (**Supplementary Table 3**) were usually of low SNR. Therefore, we set the values of the **Fidelity** and **Sparsity** as 500 and 25, respectively.
- 2) At the initial recommended values of the **Fidelity** and **Sparsity**, actin filaments were intermittent and corrupted with snowflake-like artifacts. By decreasing the **Fidelity** and maintaining the **Fidelity: Sparsity** ratio, the snowflake-like artifacts were suppressed. When we selected 40 and 2 for the **Fidelity** and **Sparsity**, respectively, actin filaments were relatively continuous without much blurring.
- 3) Next, we kept the **Fidelity** to 40 and increased the **Sparsity** to further improve the resolution and contrast. However, once we increased the **Sparsity** value, some weak actin filaments were trimmed off. Therefore, we set the optimal **Fidelity** and **Sparsity** values to 40 and 2.

Example 5. Sparse deconvolution of the CCPs obtained by the SD-SIM

- 1) For CCPs labeled by clathrin-DsRed and observed under the SD-SIM (*c.f.*, **Extended Data Fig. 6**), we needed to select the **Strong background (HI)** to remove the cytosol signal before sparse deconvolution (**Supplementary Fig. 19**). It was apparent that our background estimation method was able to remove different amplitudes of background fluorescence from different regions.
- 2) Upon selecting the **Fidelity** and **Sparsity** as 500 and 25, ring-shaped CCPs were oversharpened and brightening single pixels (**Supplementary Fig. 20**). As we reduced the **Fidelity** and **Sparsity** concomitantly, artifacts disappeared along with the blurring of rings. We selected 60 and 3 for the **Fidelity** and **Sparsity** without unnecessary blurring.
- 3) Next, we kept the **Fidelity** to 60 and increased the **Sparsity** to improve the resolution and contrast further.

We stop at the optimal **Fidelity** and **Sparsity** values of 60 and 5, which reached a good balance between resolving intricate signals and maintaining weak signals.

Supplementary Note 8.3 | More explanations on the background estimation.

To demonstrate the effects of different background estimation choices, we chose cytosolic lysosomes from **Fig. 4m** as the example. Upon selecting either **No background**, **Weak-HI**, or **Strong-HI**, it was apparent that the vesicular structure remained largely unchanged while the signal-to-background ratio increased substantially (**Supplementary Fig. 21a**).

Weakly fluorescent lysosomes at the upper left corner of the FOV, which was most likely due to the microscope's uneven illumination, might be removed by selecting the **Strong-HI** parameter (**Supplementary Fig. 21b**). Therefore, we could use the BaSiC correction³⁸ method to correct the skewed intensity distribution in the FOV so that obscure lysosomes with low intensities emerged in the image before the deconvolution (**Supplementary Fig. 21c**). Subsequent sparse deconvolution revealed multiple lysosomes with the resolution and contrast similar to lysosomes in the middle of the FOV (**Supplementary Fig. 21d**).

Finally, we also showed the effects of different background estimation choices on dense actin filaments revealed by the sparse deconvolution (**Supplementary Fig. 22**). Choosing either **Weak-HI** or **Strong-HI** removed the background without much affecting fluorescence profiles of filaments.

Supplementary Note 8.4 | The effects of t-axial continuity.

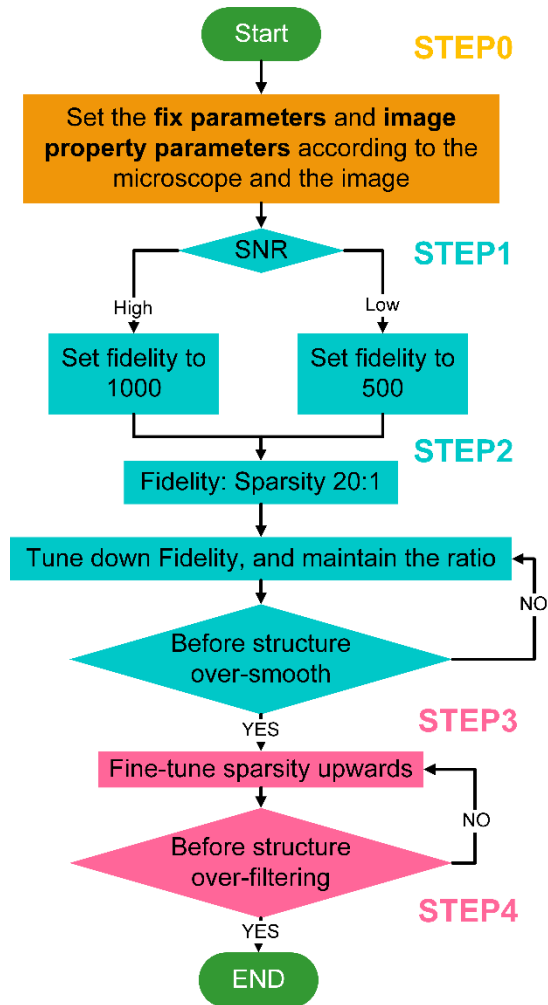
We conducted the simulation to show the effects of t-axial continuity on samples moving at different distances per interval (**Supplementary Fig. 23**). For a ring of ~80 nm in diameter, we simulated a frame-per-frame speed of 10 nm at 100% frame under the ultrafast sampling condition. After convolution with a PSF of 105 nm before corrupted with 30% Gaussian noise, sparse deconvolution with t-axial continuity at 0 yielded a partially broken ring due to the noise amplification. Raising the t-axial continuity to 0.5 suppressed noise and led to reconstruction close to the ground truth. When this dataset was resampled at 5×, 10×, and 20× slower, the frame-to-frame displacements of the ring were comparable to or larger than its diameter. Even under such circumstances, reconstruction with t-axial continuity at 0, 0.5, or 1 remained largely unaltered. In contrast, elongated and punctuated artifacts manifested with t-axial continuity at 3. Thus we recommend selecting a value of t-axial continuity less than 1 (as suggested in **Supplementary Note 6**), which usually did not cause artifacts.

Supplementary Note 8.5 | Relationship between pixel sizes and the sparsity and fidelity values.

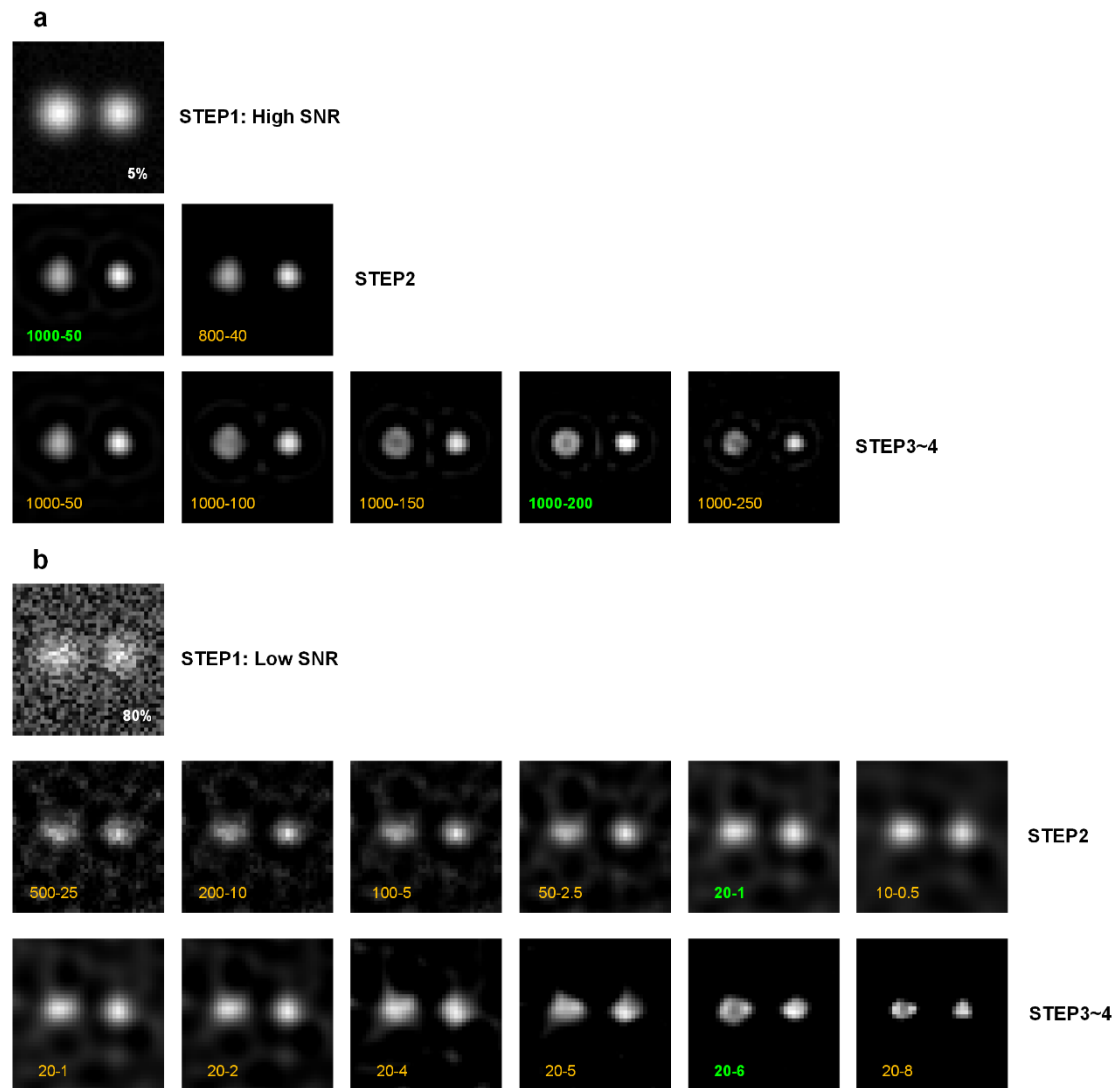
We simulated paired lines at different distances (**Supplementary Fig. 24a**), in which we convolved them with

100 nm PSF, corrupted them with 5% Gaussian noise, and downsampled 2 $\times$, 4 $\times$, and 8 $\times$ to the pixel size of 20 nm, 40 nm, and 80 nm, respectively (**Supplementary Fig. 24b**). To achieve maximal resolution, we had manually upsampled the images of 40 and 80 nm pixel size by 2 and 4 folds before underwent the sparse deconvolution pipeline (**Supplementary Fig. 24f**). As the reference, we also simulated the image of the ground truth objects as observed under a microscope with 50 nm resolution and 20 nm pixel size (**Supplementary Fig. 24e**).

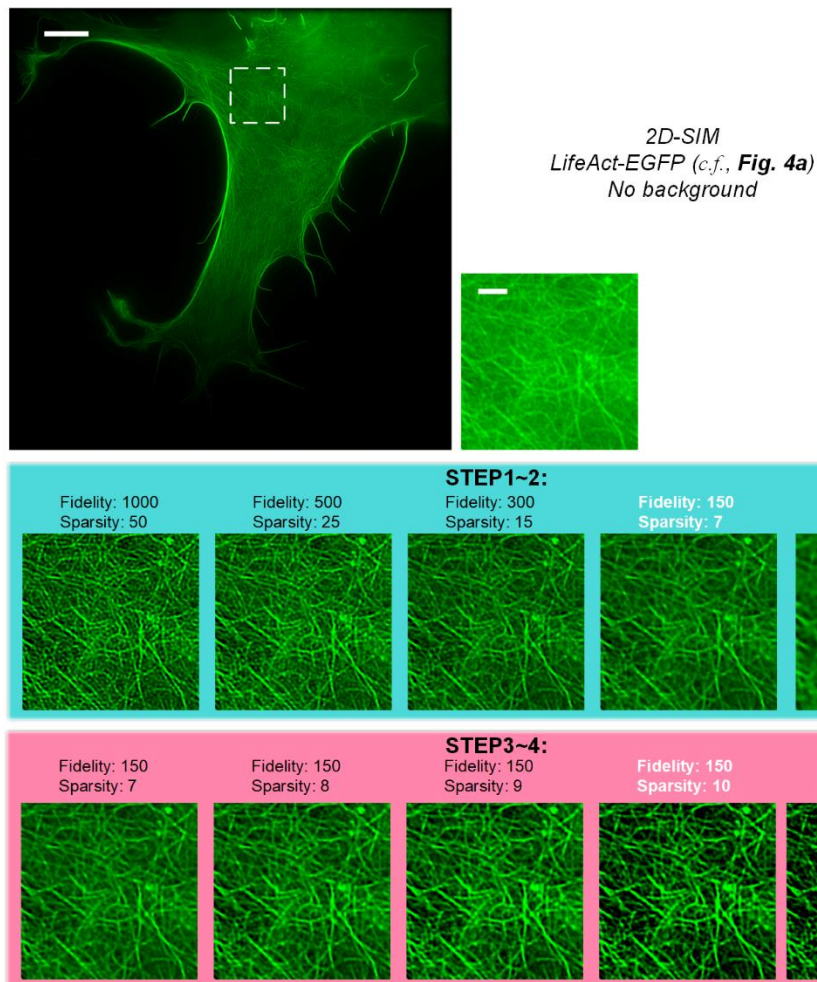
Interestingly, we found that the suitable value of the sparsity remained the same among raw images of different pixel sizes. **This suggests that the sparsity confers one intrinsic property of the object.** Instead, we needed to reduce the fidelity value to yield good deconvolution results. Given that the fidelity is the reciprocal of the continuity, which means **more continuity prior knowledge must be used to antagonize the loss of information due to inadequate Nyquist sampling.** On the other hand, although Sparse 2 $\times$ and 4 $\times$ could resolve separated lines 60 nm apart as Sparse 1 $\times$, we detected a gradual reduction of PSNR against synthetic ground truth of 50 nm resolution. The reduced PSNR manifested as increased curvatures in the straight lines in the images and corresponding striped patterns in their frequency components (white arrows in **Supplementary Fig. 24h**). It is worth noting that, even though the artificial upsampling operation can help sparse deconvolution to boost the achievable resolution, such algorithmic processing can not be equivalent to the optical sampling setup with actual small pixel sizes. In other words, a sparse spatial sampling rate will reduce the accuracy and quality of the reconstructed images (**Supplementary Fig. 24f**). Nevertheless, in this simple synthetic image containing structures under a single orientation, we could see an extension of the vertical high-frequency information in the frequency domain after the sparse deconvolution (green arrows in **Supplementary Fig. 24h**), no matter whether initial images were adequately sampled or undersampled. Thus, this experiment reinforces that the sparse deconvolution postulates the high-frequency information of the object, but not making random guesses.



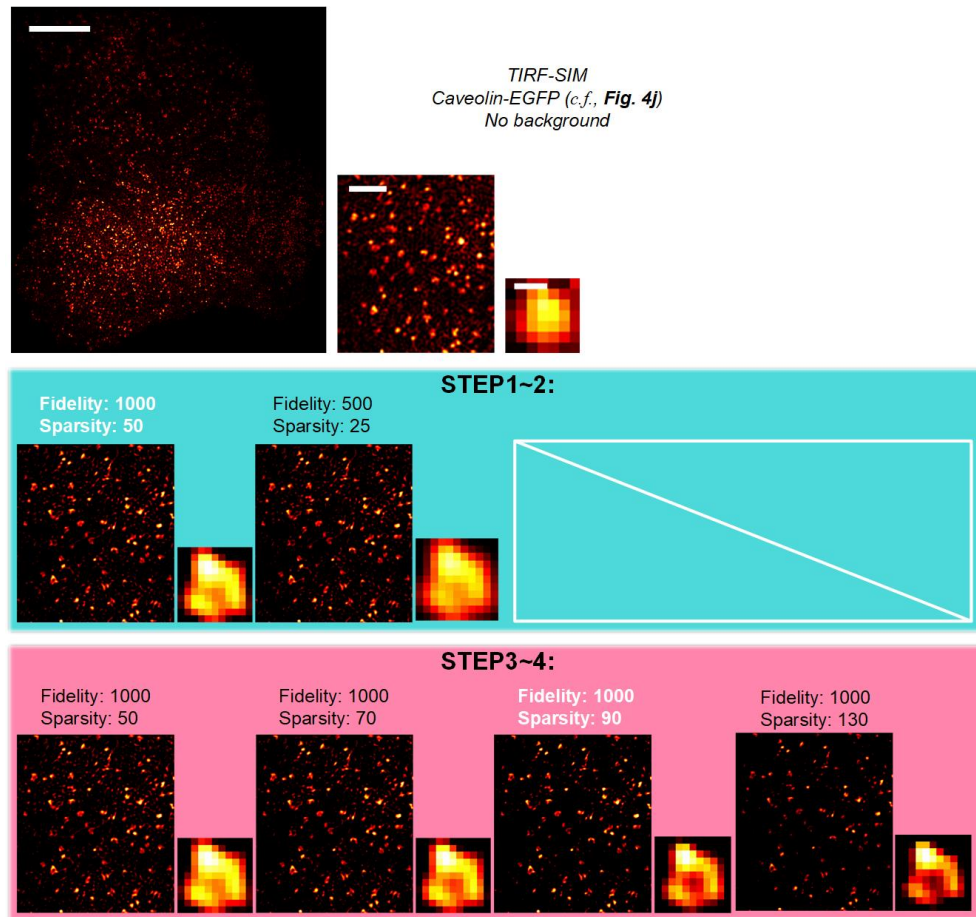
Supplementary Fig. 14 | The complete workflow of adjusting parameters in the sparse deconvolution software in handling various fluorescence images.



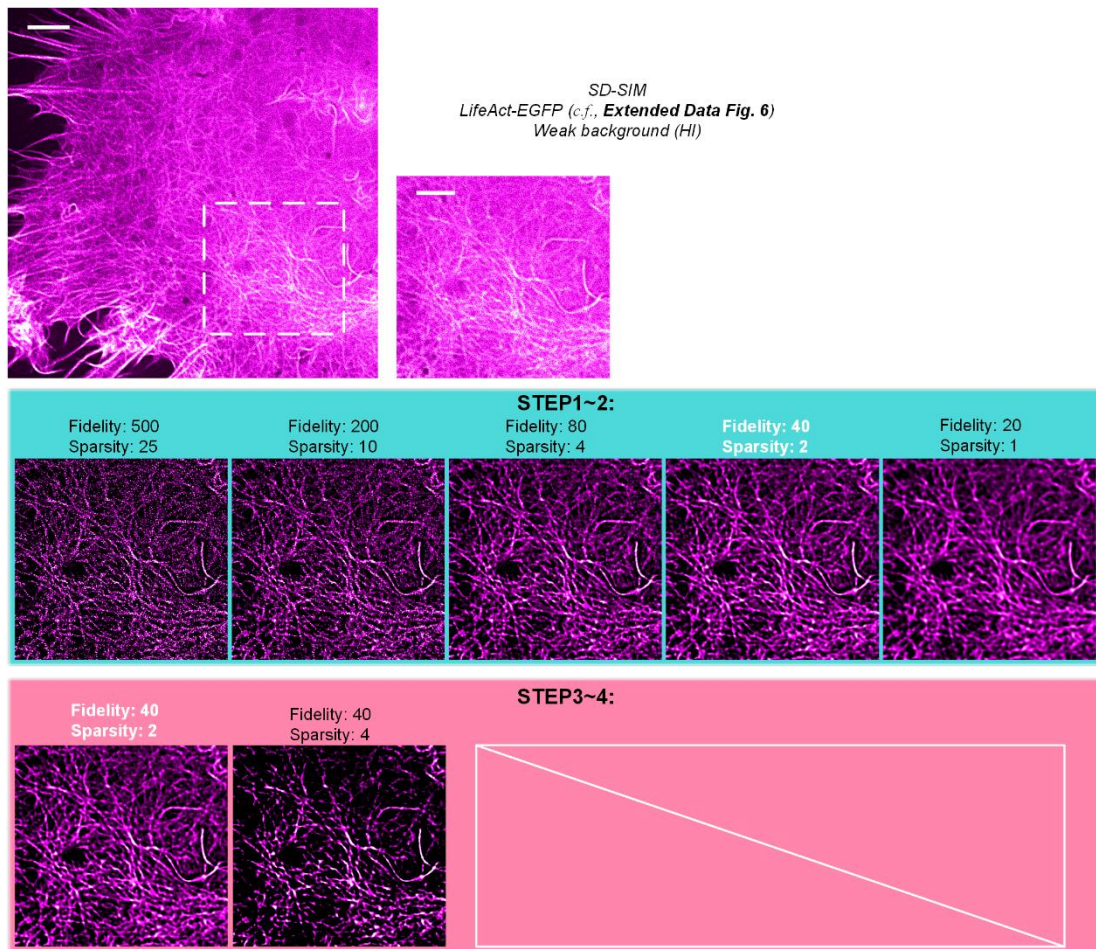
Supplementary Fig. 15 | The process of selecting optimal parameters for the sparse deconvolution of synthetic ring-shaped and punctuated structures corrupted with small (a) or large (b) noise (*c.f.*, Supplementary Fig. 12).



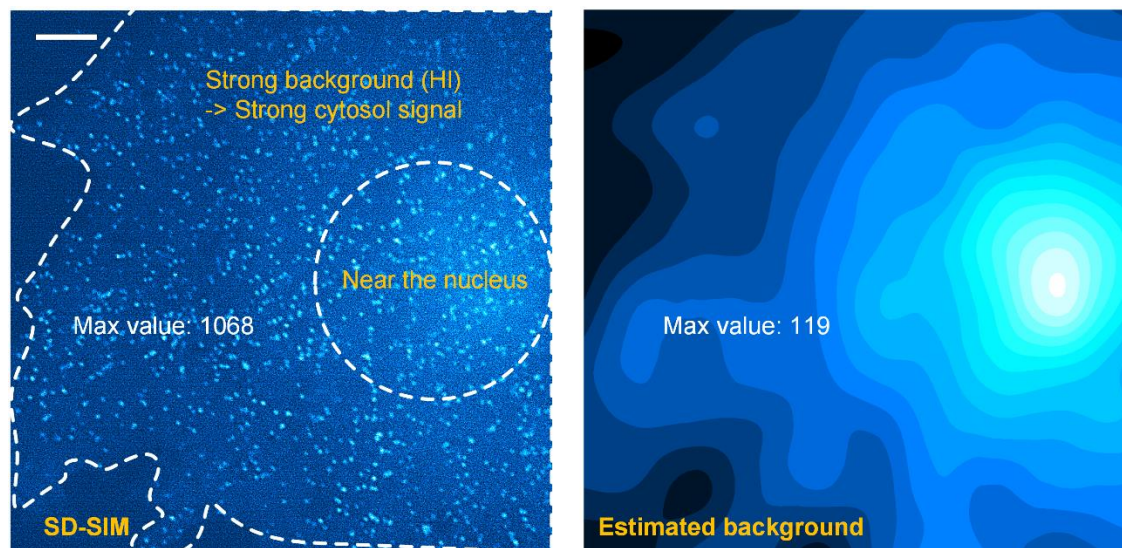
Supplementary Fig. 16 | The process of selecting optimal parameters for the sparse deconvolution of actin filaments obtained by the 2D-SIM. Scale bars: 5 μ m; 1 μ m.



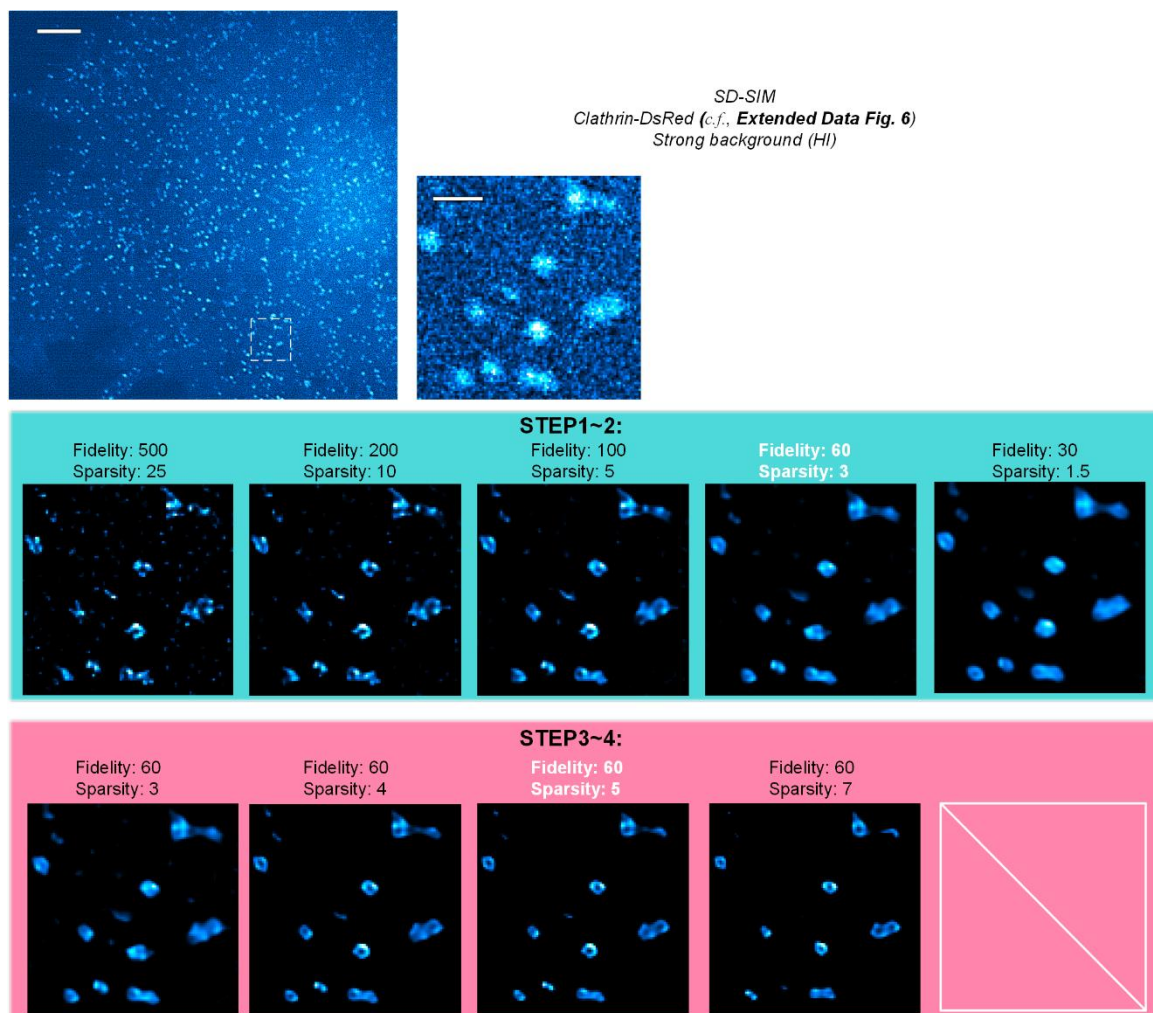
Supplementary Fig. 17 | The process of selecting optimal parameters for the sparse deconvolution of caveolae obtained by the TIRF-SIM. Scale bars: 5 μ m; 1 μ m; 100 nm.



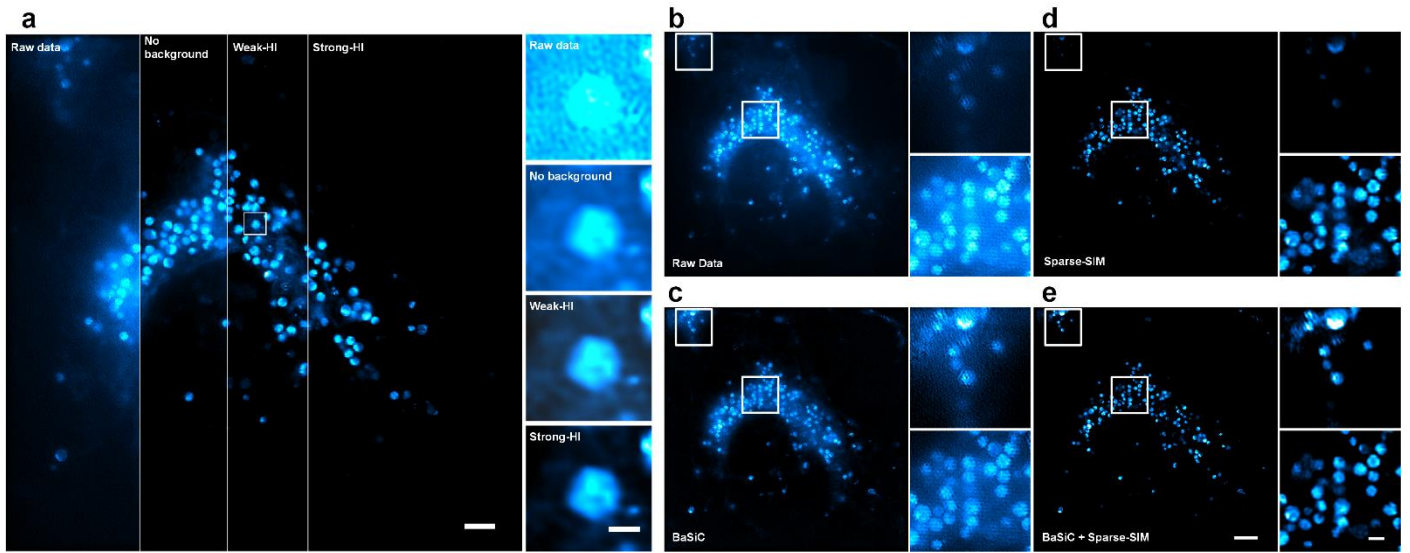
Supplementary Fig. 18 | The process of selecting optimal parameters for the sparse deconvolution of actin filaments obtained by the SD-SIM. Scale bars: 5 μ m; 3 μ m.



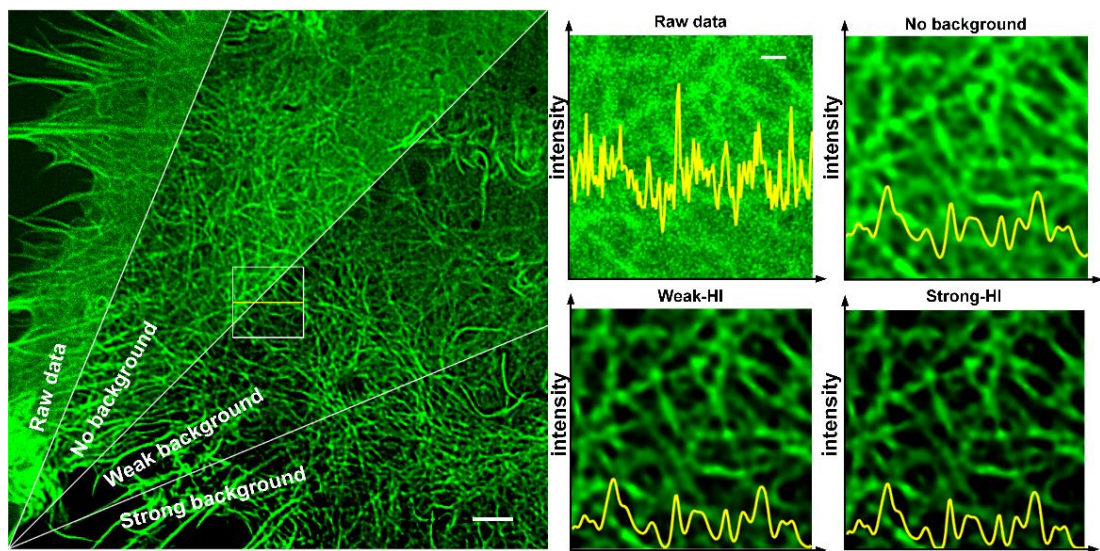
Supplementary Fig. 19 | Estimating background of different amplitudes at different regions. The raw SD-SIM image (left) and the estimated background by using the **Strong-HI** parameter (right). Scale bars: 5 μm .



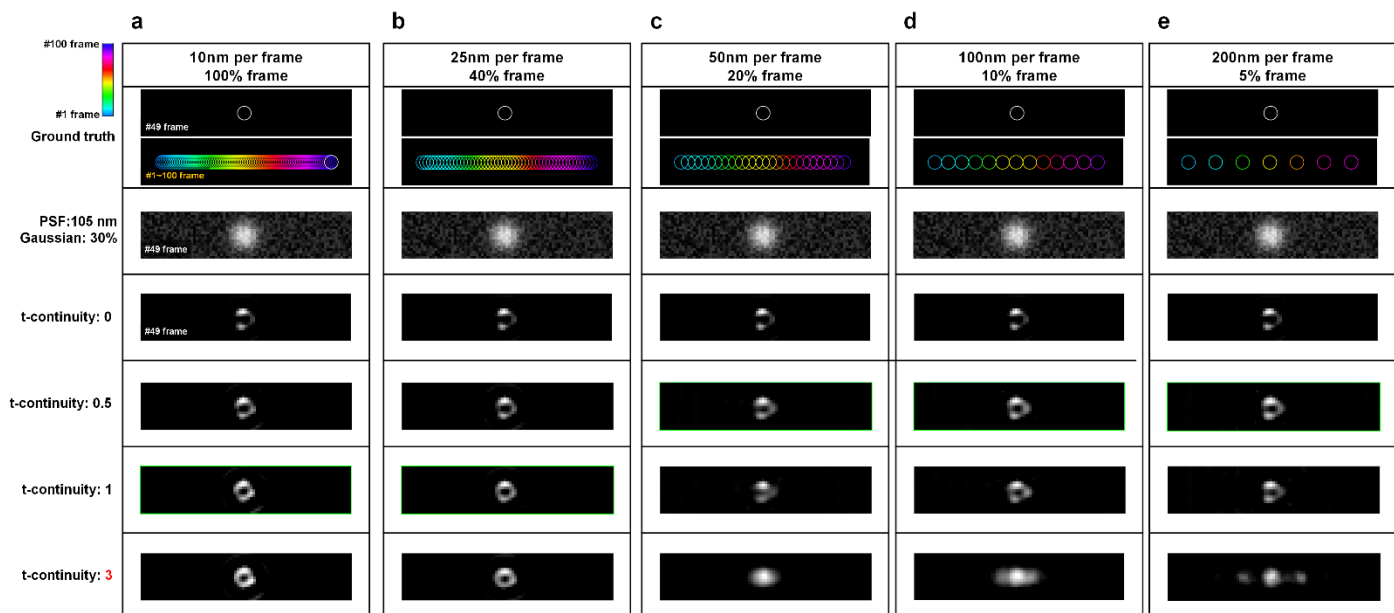
Supplementary Fig. 20 | The subsequent process of selecting optimal parameters for the sparse deconvolution of CCPs obtained by the SD-SIM. Scale bars: 5 μm ; 1 μm .



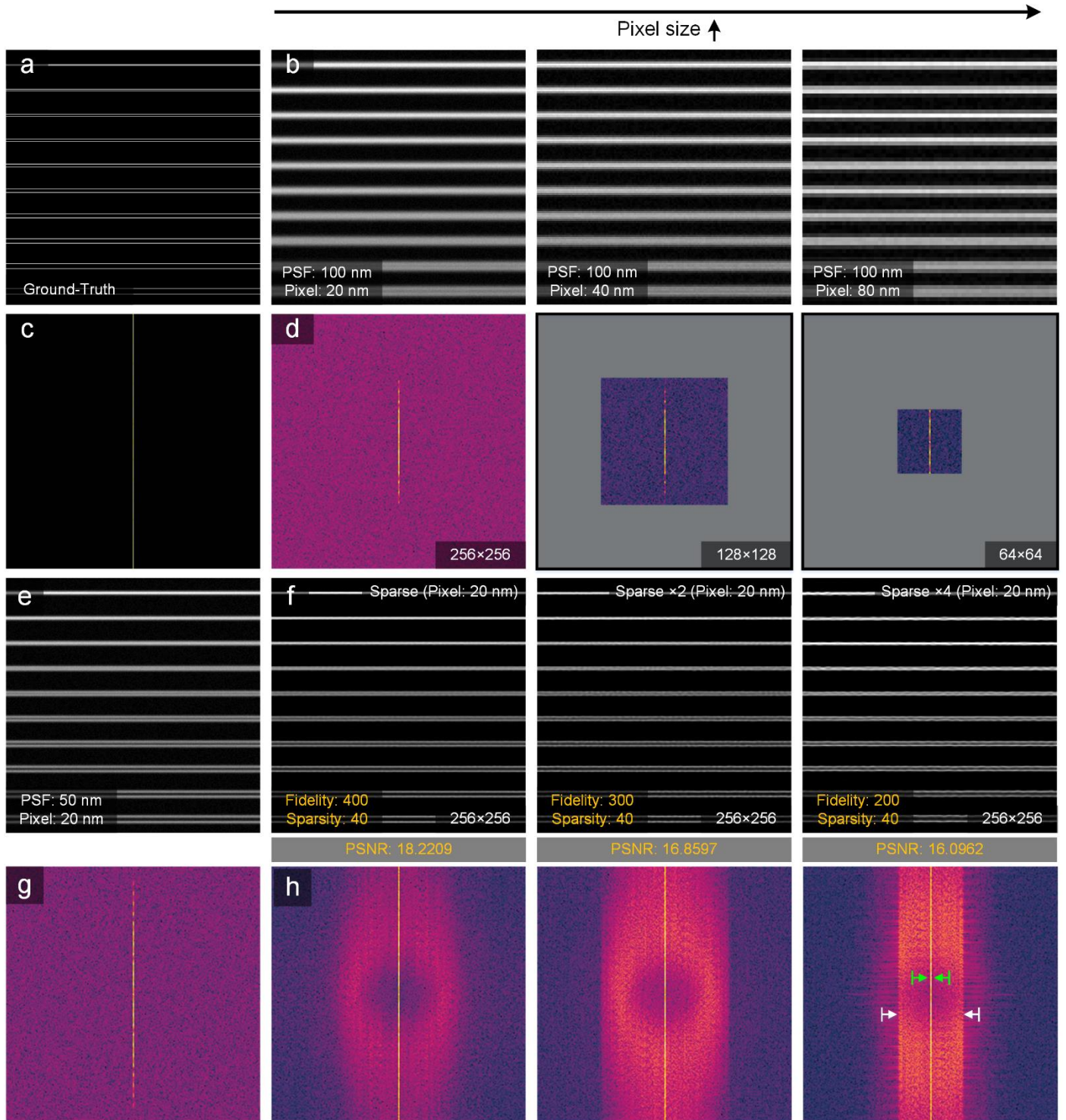
Supplementary Fig. 21 | Images reconstructed with different background estimations and the effect of the BaSiC correction. (a) The 2D-SIM image (*c.f.*, Fig. 4m) after reconstructed with *No background*, *Weak-HI*, or *Strong-HI*. Scale bars: left, 3 μm ; right, 500 nm. (b, c) The 2D-SIM image (*c.f.*, Fig. 4m) before (b) and after (c) BaSiC illumination correction. (d, e) The sparse deconvolution results of (b) and (c). Scale bars: (a) left, 3 μm ; right, 500 nm; (e) left, 5 μm ; right, 1 μm .



Supplementary Fig. 22 | Actin filaments reconstructed with different background estimations (*c.f.*, Extended Data Fig. 15). Scale bars: left, 3 μm ; right, 500 nm.



Supplementary Fig. 23 | The effects of t-axial continuity on the samples moving at different speeds. From top to bottom: The ground-truth structures at 4th frame (1st row), and the corresponding color-coded temporal projection view within 100 frames (2nd row) in total. We synthesized an image stack of ring structures with a diameter of 80 nm moving at 10 nm/frame (a) and resampled such image stack at 2.5 $\times$ (b), 5 $\times$ (c), 10 $\times$ (d), and 20 $\times$ (e) slower. The stacks were convolved by PSF with 105 nm FWHM, subsampled by 16 times, and corrupted by 30% Gaussian noise subsequently (3rd row). Then the images were reconstructed by Sparse deconvolution with t-axial continuity weight as 0, 0.5, 1, and 3 as shown in 4-7 rows, respectively.



Supplementary Fig. 24 | The reconstruction quality of sparse deconvolution under different pixel sizes.

(a) We created a ground-truth containing twelve parallel paired lines departed with different distances (from 10 nm to 100 nm). (b) The ground-truth in (a) was convoluted by a PSF with a 100 nm FWHM, and subsampled by 2 (left), 4 (middle), and 8 (right) times. The resulting images were then corrupted by 5% Gaussian noise. (c, d) The corresponding Fourier transforms of (a) and (b). (e) The ground-truth in (a) was convoluted by a PSF with a 50 nm FWHM, and subsampled by 2 times. (f) The images in (b) were reconstructed by sparse deconvolution with no upsampling (left), 2 times upsampling (middle), and 4 times upsampling (right). (g, h) The corresponding Fourier transforms of (e) and (f).

Supplementary Note 9 | The calibration of nuclear pore diameters and the FWHM resolution.

Unexpectedly, we found that when the sizes of system PSF and the size of camera pixel were comparable to the size of the structure been imaged, fitted diameters of punctated and ring-shaped structures differently deviated from their real values. For large punctated fluorescence beads, the fitted FWHMs were larger than the PSF of the microscope. For ring-shaped pores under the Sparse-SIM, their fitted diameters were smaller than their real values. Therefore, we used simulations to derive correction factors under different situations.

Supplementary Note 9.1 | Calibrating fitted diameters of ring-shaped pores.

To examine the relationship between fitting diameter and the real diameter, we created synthetic ring structures (1 nm sampling) with different diameters (60, 70, 80, 90, 100, 110, 120 nm). These ground truth structures are then convolved with different sized PSF (10, 20, 30, 40, 50 nm $FWHM_{PSF}$). The results are subsampled by 16 times to match the experimental condition (**Supplementary Fig. 25a**). Resultant images are fitted with a double-peak Gaussian function to extract the fitted diameters (**Supplementary Fig. 25a**). As can be seen in **Supplementary Fig. 25b**, with 1 nm pixel size and no subsampling, the fitted diameter (FD) was smaller than the real diameter (RD) due to the convolution of system PSF ($FWHM_{PSF}$) with the ring, which could be approximated by the equation:

$$RD = \sqrt{FD^2 + FWHM_{PSF}^2} . \quad (22)$$

Interestingly, if the pixel was subsampled to be 16 nm per pixel, the apparent FD was further reduced by the aliasing effect of a large pixel. Under such circumstance, the relationship between FD and RD could be approximated by a linear function:

$$RD = a \times FD + b . \quad (23)$$

Therefore, we used a linear regression with parameters ($a = 0.9335$, $b = 22.8085$) to correct for the nuclear pore measurement in Sparse-SIM $2\times$ configuration (50 nm resolution, 16 nm pixel size). To correct the pore size measured by TIRF-SIM (90 nm resolution, 32 nm pixel size), we used another set of parameters ($a = 0.7740$, $b = 62.9420$).

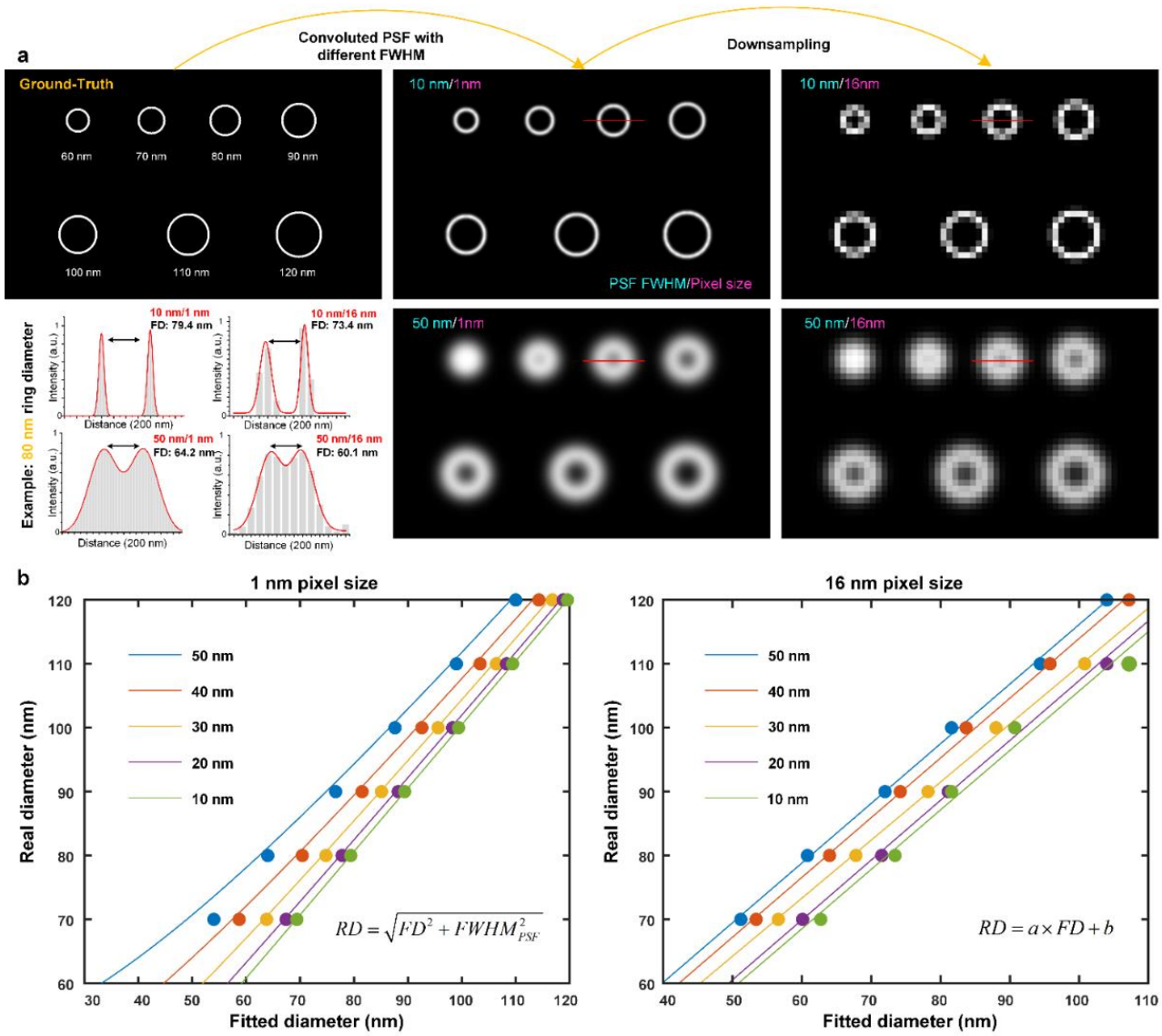
Supplementary Note 9.2 | The bead correction factor.

In our experiments, the diameters of fluorescent beads (100 nm) were comparable in scale to the spatial resolution of the Sparse SD-SIM (~ 90 nm). Under such circumstances, the measured FWHMs were also affected by the size of the beads, which did not accurately report the actual microscope resolution. Therefore,

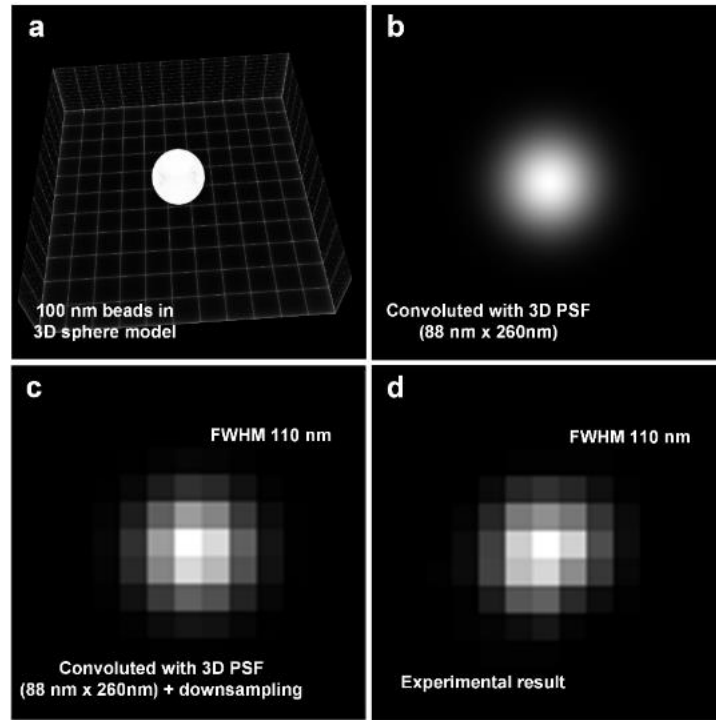
we intend to use the 'beads correction factor' equation derived in the PSFj³⁹ to correct for such effects (details in PSFj), and retrieving the real resolution of the optic system (FWHM_{PSF}). In addition, per personal communication with the authors of the PSFj, we also tried to simulate a fluorescent bead to be a 3D sphere with constant fluorophore concentration over the bead volume, in which gives:

$$f(x, y, z) = \begin{cases} C_0, & 2\sqrt{x^2 + y^2 + z^2} \leq d_{\text{bead}} \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

Since there is no analytical solution for the convolution of 3D sphere bead and PSF, we solved this convolution numerically with the specific experiment parameters as shown in **Supplementary Fig. 26**. In our simulations, the 100 nm 3D sphere was convolved with a 3D PSF (FWHM_{PSF} of $88 \text{ nm} \times 88 \text{ nm} \times 260 \text{ nm}$), and subsequently subsampled to match the experimental condition (the pixel size of 38 nm). The resulting image had a measured bead FWHM as 110 nm, which was the same as our experimental fitted result 110 nm. For a system with higher or lower spatial resolution (FWHM_{PSF} : 80 nm or 100 nm) would lead to measured bead FWHM varied between 102 nm and 120 nm. Therefore, we concluded that the actual FWHM_{PSF} of our Sparse SD-SIM should be approximated 90 nm.



Supplementary Fig. 25 | Relationships between fitted and real diameters of ring structures of various sizes as observed under microscopes of different resolutions and spatial sampling frequencies. (a) The simulation workflow. A representative example of ring structures with different diameters was convolved with the PSF with FWHM of either 10 nm (upper panels) or 50 nm (bottom panels). These images were shown without (pixel size of 1 nm, left) and with spatial subsampling (pixel size of 16 nm, right). The histogram in the left-bottom corner showed examples of fitted diameters (FD) of the 80 nm ring obtained under different conditions. **(b)** Linear relationships between fitted and real diameters (RD) under 1 nm (left) or 16 nm (right) spatial sampling. Different colors indicate relationships obtained with microscopes of different spatial resolution (FWHM_{PSF}). With sufficient spatial sampling, RD could be derived from the fitted result using the equation: $RD = \sqrt{FD^2 + FWHM_{PSF}^2}$. When the pixel size of the final image was comparable to the FWHM_{PSF} and the diameter of the ring, the relationships were approximated by a linear function offset by a constant, $RD = a \times FD + b$ ($a = 0.9335$, $b = 22.8085$).



Supplementary Fig. 26 | A similar size of the synthetic bead after convolution with the PSF of the system as compared with that detected in the real experiment. (a) The 3D rendering of a synthetic 100 nm bead in three dimensions. **(b)** The 2D image of the bead convolved with a 3D PSF with the FWHM_{xyz} of $88 \text{ nm} \times 88 \text{ nm} \times 260 \text{ nm}$ (1 nm pixel size). **(c)** The image in (b) was downsampled by 38 times. **(d)** The experimental image of 100 nm fluorescent bead (*c.f.*, **Extended Data Fig. 5a**), which was nearly identical to the simulated image in (c).

Supplementary Note 10 | Other sparsity-based reconstructions that improve spatial resolution in theory.

In this part, we implemented two previous reconstruction procedures based on the sparsity concept, FISTA (Fast Iterative Shrinkage-Thresholding Algorithm)⁴⁰, and NLHT (Non-Local Hard Thresholding)^{27, 41} algorithms, which have been proposed to extend spatial resolution under coherent imaging of very sparse samples.

Supplementary Note 10.1 | FISTA method.

Similar to bandwidth interpolation methods that were based on compressive sensing to reduce the acquisition time^{27, 41}, FISTA reconstructs with l_1 constraint that is mathematically expressed as:

$$\arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \lambda \|\mathbf{x}\|_1 \right\} = \arg \min_{\mathbf{x}} \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}, \quad (25)$$

where the first term on the left side is the fidelity term, representing the distance between recovered image $\mathbf{x}$ and the image obtained after the Wiener filtered result $\mathbf{b}$. $\mathbf{A}$ is the point spread function (PSF) of the imaging system. The second term represents the sparsity prior, $\|\cdot\|_1$ and $\|\cdot\|_2$ are the l_1 and l_2 norms, respectively. λ denotes the weight factor, balancing the images fidelity and the sparsity prior.

We reconstructed the SR images with the sparsity prior based on **Eq. (25)**, a convex optimization problem resolved with the FISTA (with backtracking) algorithm⁴⁰.

For the simplicity of derivation, we termed $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2$. First, consider the unconstrained minimization problem of a continuously differentiable function $f(\mathbf{x})$:

$$\min \{ f(\mathbf{x}) \}. \quad (26)$$

The simplest methods to solve **Eq. (26)** is the gradient descent method:

$$\mathbf{x}_k = \mathbf{x}_{k-1} - t_k \nabla f(\mathbf{x}_{k-1}), \quad (27)$$

where $t_k > 0$ is a suitable step distance. The gradient iteration **Eq. (27)** can be regarded as an extremal solution of a quadratic equation (proximal regularization of the linearized function $f(\mathbf{x})$ at $\mathbf{x}_{k-1}$), which could be rewritten in the equivalent form as

$$\mathbf{x}_k = \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}_{k-1}) + \langle \mathbf{x} - \mathbf{x}_{k-1}, \nabla f(\mathbf{x}_{k-1}) \rangle + \frac{1}{2t_k} \|\mathbf{x} - \mathbf{x}_{k-1}\|^2 \right\}, \quad (28)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors.

Adopting this same basic gradient solution to the non-smooth l_1 regularized problem equal to **Eq. (25)**

$$\mathbf{x}_k = \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}_{k-1}) + \langle \mathbf{x} - \mathbf{x}_{k-1}, \nabla f(\mathbf{x}_{k-1}) \rangle + \frac{1}{2t_k} \|\mathbf{x} - \mathbf{x}_{k-1}\|^2 + \lambda \|\mathbf{x}\|_1 \right\}. \quad (29)$$

After removing constant terms, the equation can be rewritten as

$$\mathbf{x}_k = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2t_k} \left\| \mathbf{x} - (\mathbf{x}_{k-1} - t_k \nabla f(\mathbf{x}_{k-1})) \right\|^2 + \lambda \|\mathbf{x}\|_1 \right\}. \quad (30)$$

For any $L > 0$, considering the quadratic approximation of $F(\mathbf{x}) \equiv f(\mathbf{x}) + \lambda \|\mathbf{x}\|_1$ at a given point $\mathbf{x}_{k-1}$, we define

$Q_L(\mathbf{x}, \mathbf{x}_{k-1})$ as:

$$Q_L(\mathbf{x}, \mathbf{x}_{k-1}) \equiv f(\mathbf{x}_{k-1}) + \langle \mathbf{x} - \mathbf{x}_{k-1}, \nabla f(\mathbf{x}_{k-1}) \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{x}_{k-1}\|^2 + \lambda \|\mathbf{x}\|_1, \quad (31)$$

where L is Lipschitz constant of smooth convex function f . Through formula derivation, we rewrite the formula above as follows:

$$\mathbf{x}_k = \arg \min_{\mathbf{x}} \left\{ \frac{L}{2} \left\| \mathbf{x} - \left(\mathbf{x}_{k-1} - \frac{1}{L} \nabla f(\mathbf{x}_{k-1}) \right) \right\|^2 + \lambda \|\mathbf{x}\|_1 \right\}. \quad (32)$$

where **Eq. (32)** can be simply interpreted that the constant terms t_k of **Eq. (30)** are replaced by Lipschitz constant terms L to ensure the convergence. Based on **Eq. (32)**, we defined a unique minimization operator $p_L(y)$ as:

$$p_L(\mathbf{x}_{k-1}) \equiv \arg \min_{\mathbf{x}} Q_L(\mathbf{x}, \mathbf{x}_{k-1}) = \arg \min_{\mathbf{x}} \left\{ \lambda g(\mathbf{x}) + \frac{L}{2} \left\| \mathbf{x} - \left(\mathbf{x}_{k-1} - \frac{1}{L} \nabla f(\mathbf{x}_{k-1}) \right) \right\|^2 \right\}. \quad (33)$$

Considering $g(\mathbf{x}) = \|\mathbf{x}\|_1$, we can use the soft-threshold algorithm (1) to solve $p_L(y)$, and let $u = y - \frac{1}{L} \nabla f(y)$,

$a = \frac{\lambda}{L}$, the expression is given as,

$$\text{soft}(u, a) \equiv \text{sign}(u) \max\{|u| - a, 0\}. \quad (34)$$

As a result, the complete steps of the FISTA algorithm can be concluded as follows:

Step 0. Let $L_0 > 0, \delta > 1$, and $\mathbf{x}_0 \in \mathbb{R}^n$. Take $y_1 = x_0, m_1 = 1$.

Step k. ($k \geq 1$) search the smallest nonnegative integers i_k subject to $L' = \delta^{i_k} L_{k-1}$

$$F(p_L(\mathbf{y}_k)) \leq Q_L(p_L(\mathbf{y}_k), \mathbf{y}_k). \quad (35)$$

Let $L' = \delta^{i_k} L_{k-1}$ and proceed

$$x_k = \text{soft}(y_k - \frac{1}{L} \nabla f(y_k), \frac{\lambda}{L}). \quad (36)$$

Combine the Nesterov³⁴ method to accelerate gradient update:

$$m_{k+1} = \frac{1 + \sqrt{1 + 4m_k^2}}{2}. \quad (37)$$

$$y_{k+1} = x_k + \left(\frac{m_k - 1}{m_{k+1}} \right) (x_k - x_{k-1}). \quad (38)$$

Supplementary Note 10.2 | NLHT method.

Adding some constraints to the FISTA algorithm yields the Non-Local hard Thresholding algorithm (NLHT)⁴¹. Initially, the results of l_1 constraint are solved by the Basis Pursuit De-Noising algorithm (BPDN), in which the pixels with values close to zero are located by performing a nonlocal thresholding step. Each element of $\mathbf{x}$ will be zeroed out along with its neighbors when its value is below a fixed threshold. The BPDN step is repeated with the additional constraint that the locations corresponding to the off-support are set to zero⁴¹. Because of the slow convergence speed, we replaced the BPDN with the FISTA to solve the convex optimization problem and conducted non-local hard thresholding. A more detailed version of the algorithm is provided as follows:

Step 1: Use FISTA to solve

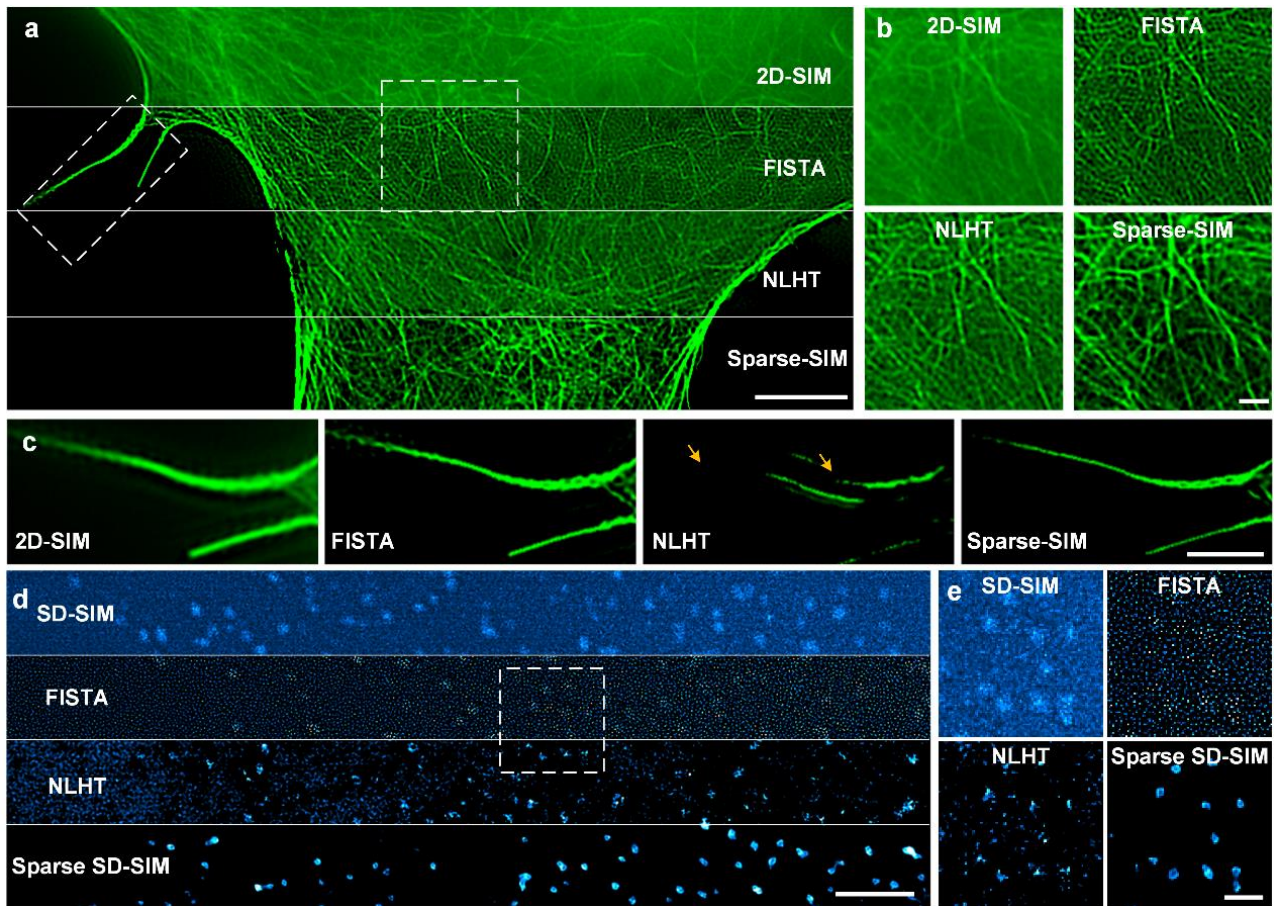
$$\arg \min_{\mathbf{x}} \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}, \quad \mathbf{x}[j] = 0, \forall j \in S. \quad (39)$$

where S is the index set of the zero value.

Step 2: Search all j' such that $x[j] \leq \varsigma \max(x)$ which are distanced from j' to the right or left by 1 pixel.

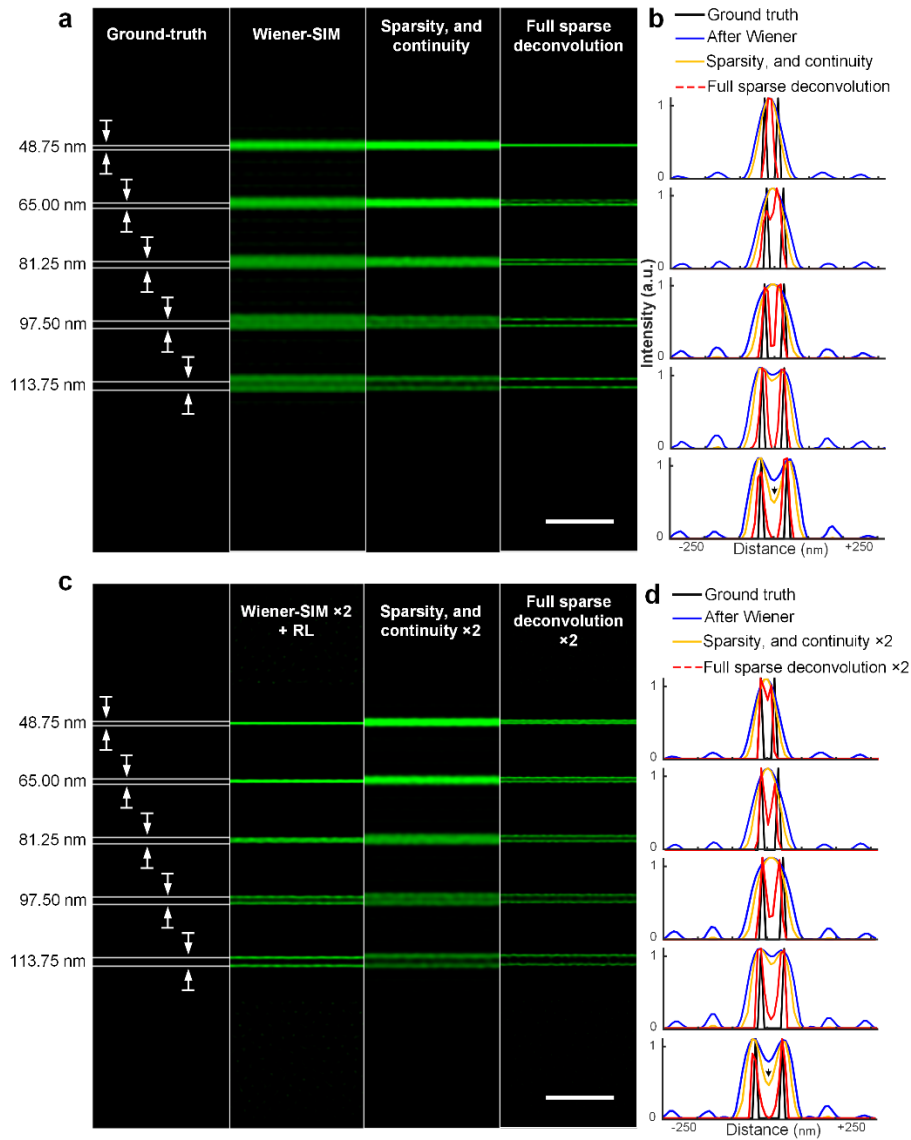
Then add j' to S' and set $S = S \cup S'$. If the index set S was not updated, increase ς by $\Delta\varsigma$. It is worth denoting that, ς and $\Delta\varsigma$ are threshold and increment in the threshold respectively.

Comparison of the reconstruction results using Sparse-SIM based algorithm with FISTA and NLHT are shown in **Supplementary Fig. 27**. It is worth noticing that, to avoid over-filtering structural information by the NLHT (**Supplementary Fig. 27c**, pointed by yellow arrows), the iteration of these two steps needed to be terminated after several rounds.

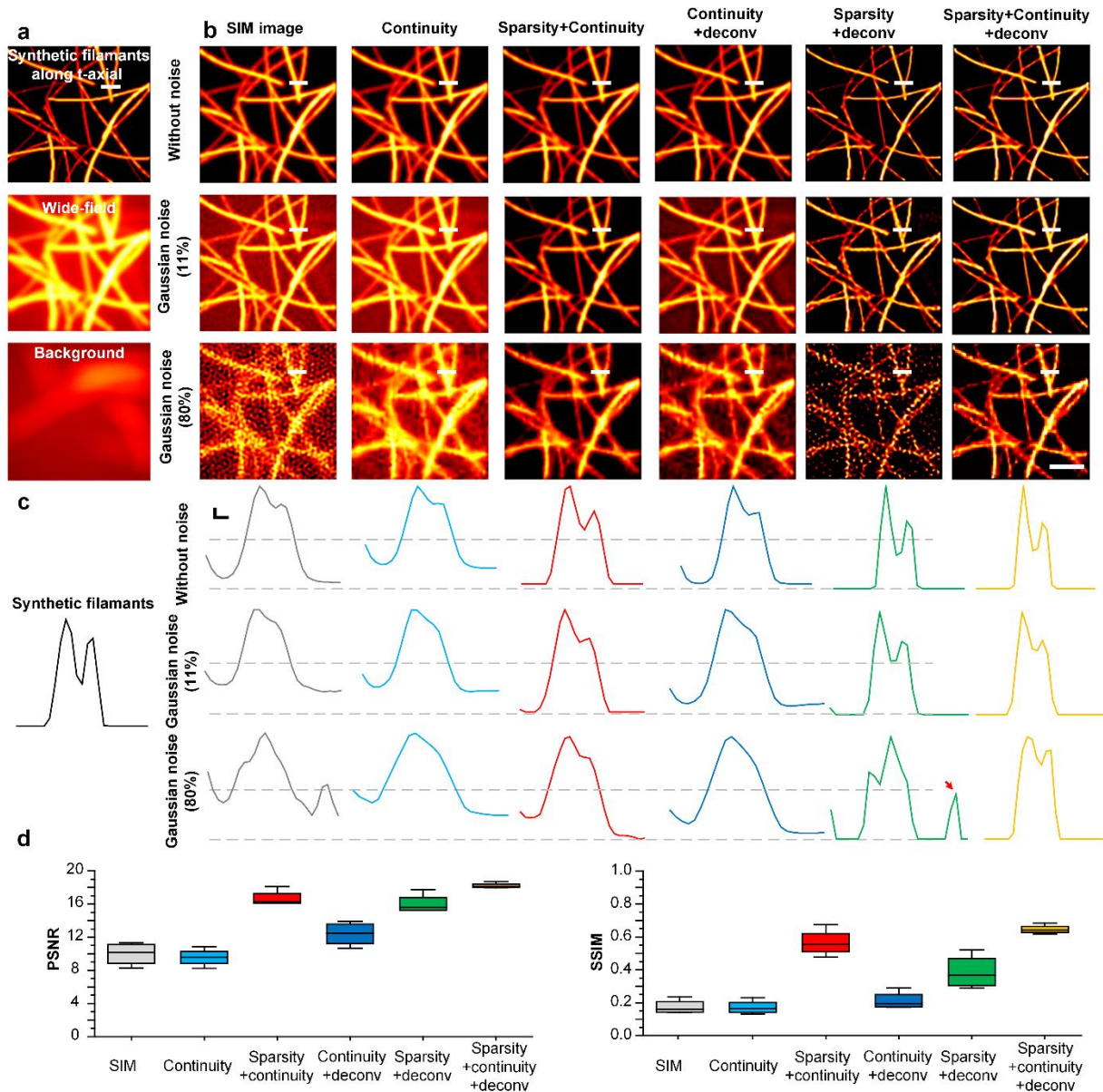


Supplementary Fig. 27 | Comparisons of our sparse deconvolution pipeline along with pre-existing reconstruction methods based on the sparsity a priori. (a-c) Actin filaments labeled with LifeAct-EGFP in a live COS-7 cell in **Fig. 4a** under the 2D-SIM, followed by the FISTA, or NLHT reconstruction, or reconstruction with our sparse algorithm. (d and e) CCPs labeled by Clathrin-EGFP under the SD-SIM in **Extended Data Fig. 6**, followed by the FISTA, or NLHT reconstruction, or reconstruction with our sparse algorithm. Under this condition of compromised SNR, neither FISTA nor NLHT method could resolve any ring-shaped structures. Scale bars: (a) 1 μm ; (b and c) 500 nm; (d) 3 μm ; (e) 1 μm .

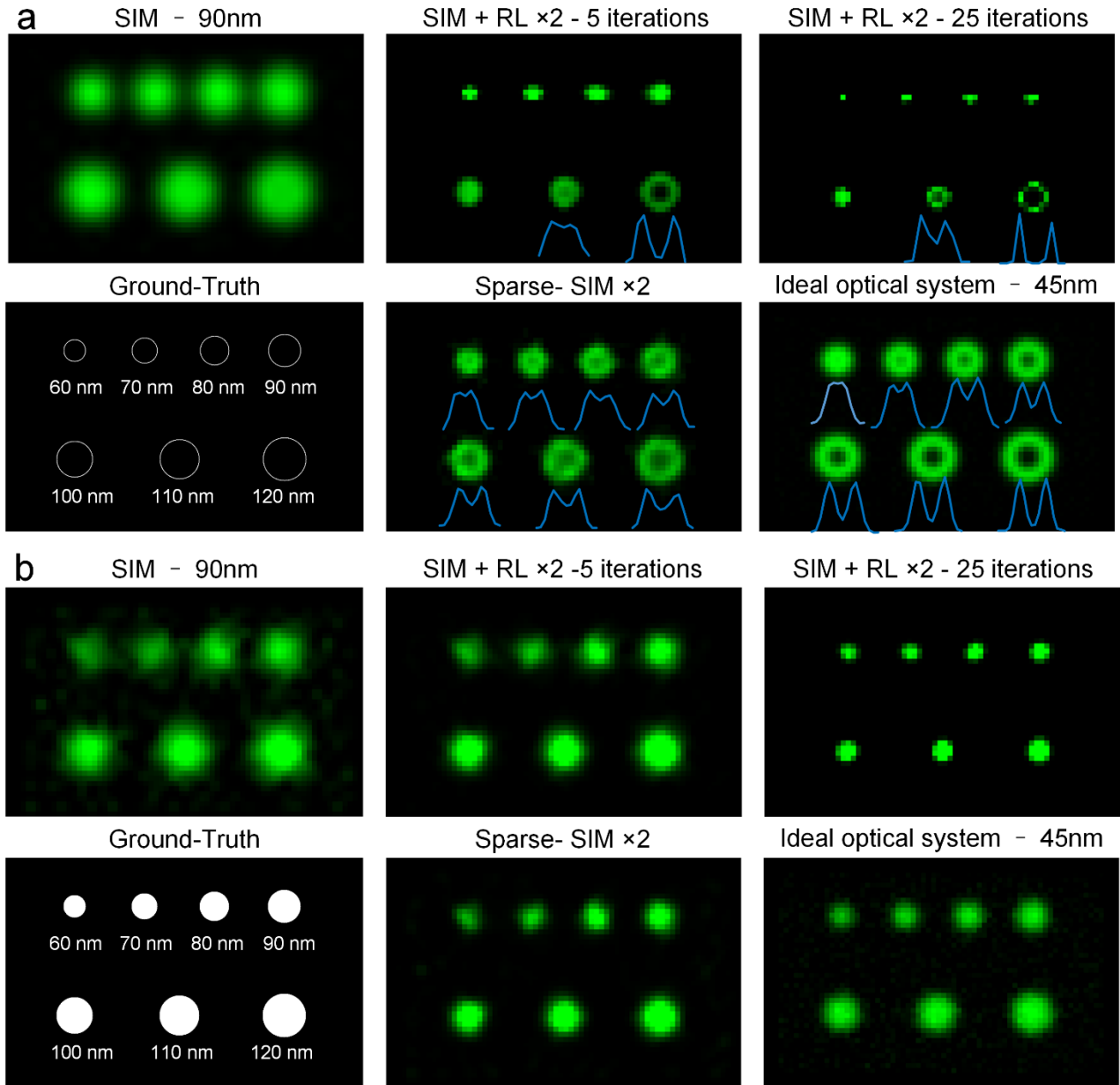
Additional Supplementary Figures



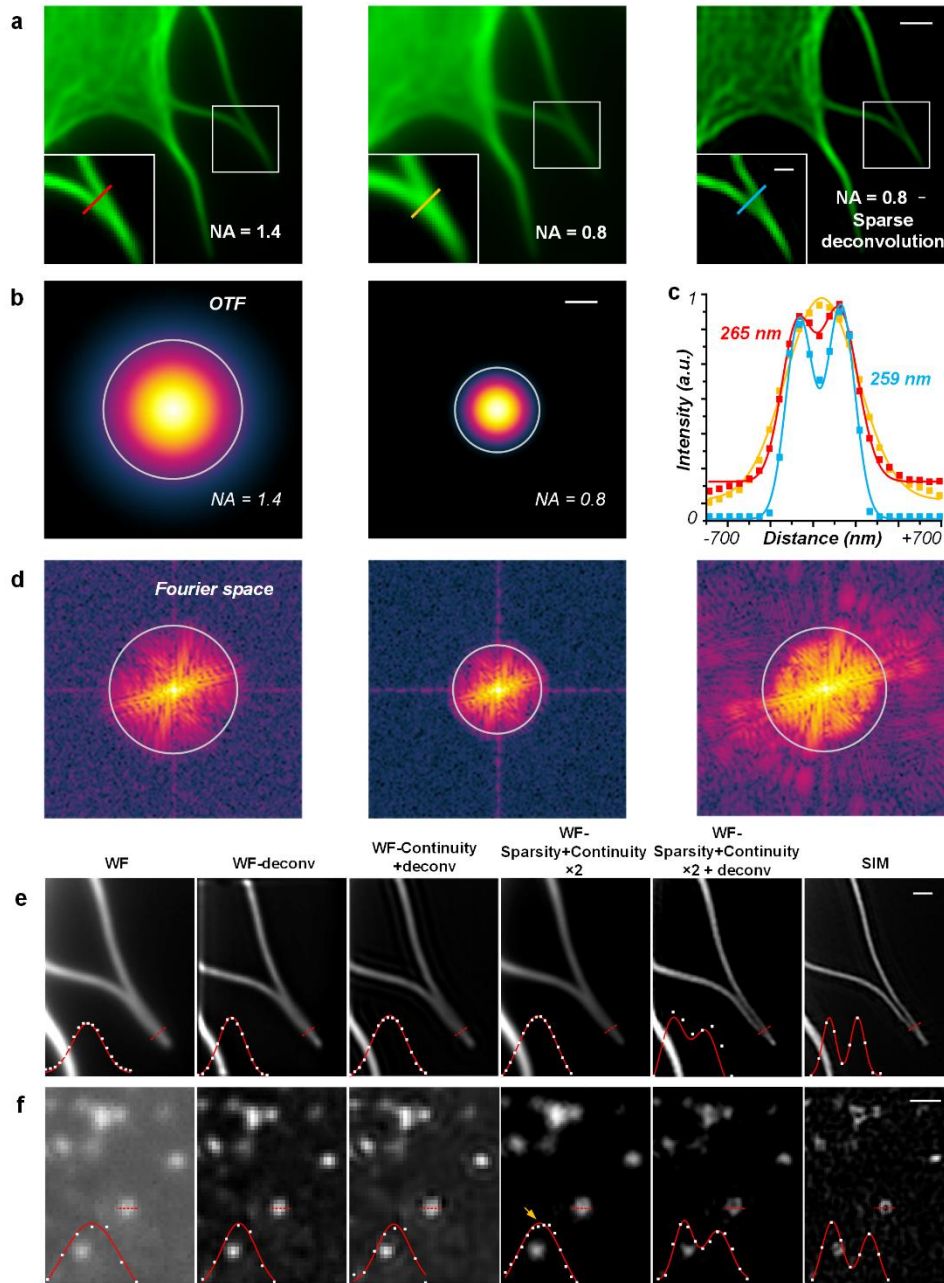
Supplementary Fig. 28 | Benchmark of spatial resolution at different steps of sparse deconvolution according to the synthetic images. (a) The resolution plate with a pixel size of 16.25 nm, which contained five pairs of lines at distances of 48.75 nm, 65.00 nm, 81.25 nm, 97.50 nm, and 113.75 nm. The synthetic image (512×512 pixels) is first illuminated by pattern excitation and then convolved with a microscope PSF of (1.4 NA, 488 nm excitation). The signal is recorded with an sCMOS camera with a pixel size of 65 nm, which meant $4\times$ downsampling of the original image (128×128 pixels). We also included *Gaussian* noise with a variance of 5% of the peak intensity of the line to the raw image. Next, (from left to right) we used inverse Wiener filtering to obtain a conventional SIM image (256×256 pixels), followed by the reconstruction constrained by continuity and sparsity a priori information and final deconvolution. The theoretical limit resolution of Wiener SIM was calculated to be 97.6 nm by following the equation $\lambda/2(NA_i + NA_d)$, in which i and d respectively represent the illumination and detection NA. (b) The corresponding intensities across different pairs of lines in (a) are shown here. Two lines separated by 65 nm could be resolved only when the raw image underwent the full sparse deconvolution pipeline. (c) We also upsampled the SIM image obtained after Wiener inverse filtering to 512×512 pixels and processed it with RL deconvolution (20 iterations), and other steps in the sparse deconvolution pipeline thereafter. (d) The corresponding intensities across different pairs of lines in (c) are shown here. Note that two lines 48.75 nm apart could be separated by sparse deconvolution in the upsampled image. Scale bars: 1 μm .



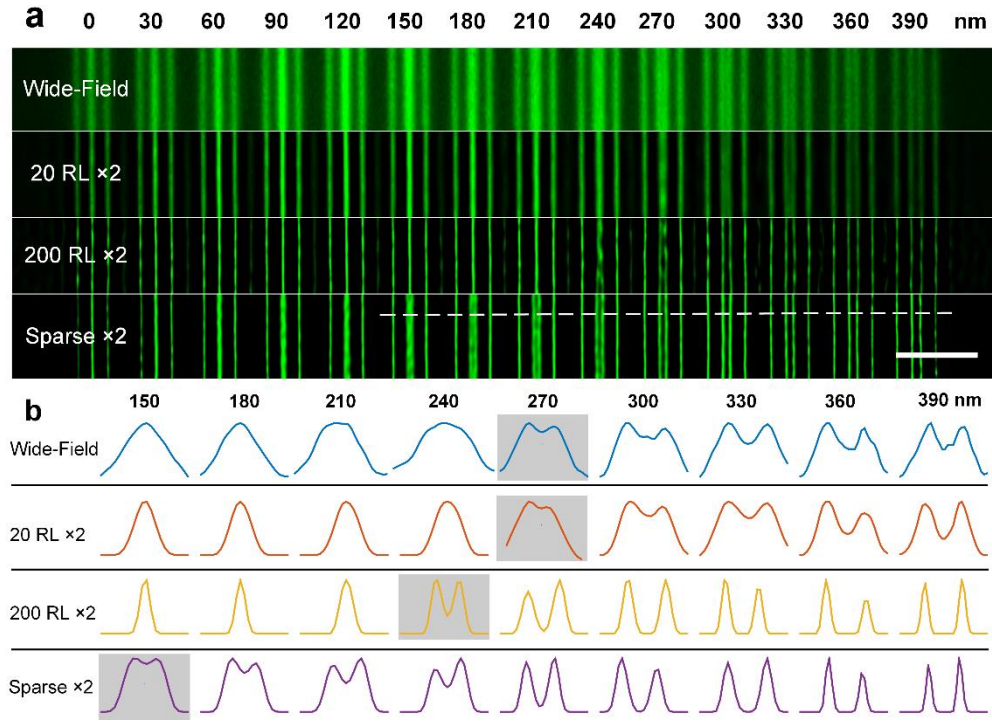
Supplementary Fig. 29 | Contributions of different steps in sparse deconvolution of synthetic images corrupted with different noise extents. (a) From top to bottom: the ground-truth filament structures, corresponding wide-field image, and out-of-focus background (Methods). (b) Performance comparisons among five different types of SR reconstructions at 0% 11% and 80% noise levels (from top to bottom). Specifically, the raw images were reconstructed with inverse Wiener filtering (column 1), followed by continuity-constraint reconstruction only (column 2), sparsity-plus-continuity-constraint reconstruction (column 3), continuity-constraint reconstruction and deconvolution (column 4), sparsity-constraint reconstruction only and deconvolution (column 5), or sparsity-plus-continuity-constraint reconstruction and deconvolution (column 6). (c) The fluorescence intensity profiles along the two opposing synthetic filaments in (b) under different conditions. While the continuity reconstruction followed by the deconvolution did not separate these filaments, the sparsity reconstruction followed by the deconvolution caused artifacts (an example indicated by the asterisk) in raw images with 80% noise. (d) Compared to the synthetic ground truth, the average structural similarity (SSIM) values⁴² and peak signal-to-noise ratio (PSNR) of SR images reconstructed with different methods from raw images corrupted with different levels of noise (0%, 11%, 25%, 50%, and 80%). More details are given in **Supplementary Tables 1 and 2**. Centerline, medians; limits, 75% and 25%; whiskers, maximum and minimum; error bars, s.e.m. Experiments were repeated five times independently with similar results. Scale bars: (b) 1 μm ; (c) axial: 0.1 arbitrary units (a.u.); lateral: 100 nm.



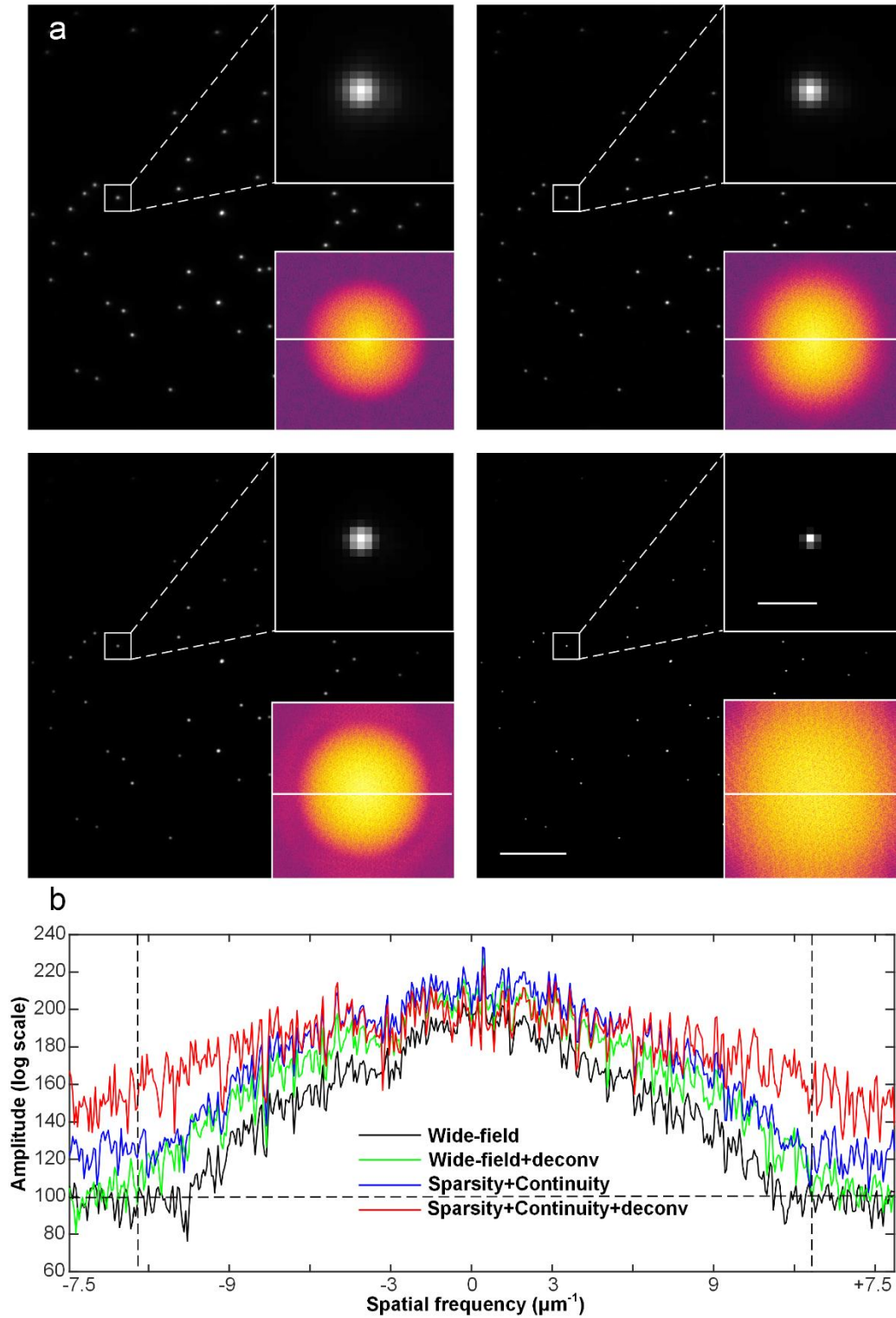
Supplementary Fig. 30 | Comparisons of reconstruction results of synthetic ring-shaped and punctated structures by the sparse and RL deconvolutions. Ring-shaped (**a**) and punctated structures (**b**) with different diameters were convolved with PSF with FWHM of either 90 nm (top left corner) or 45 nm (bottom right corner). These images were subsequently subsampled 32 times (pixel sizes of 32 nm) and corrupted with Poisson noise and 2% Gaussian noise to simulate ground truth images observed under microscopes with different spatial resolutions. For image reconstruction, we first Fourier interpolated images, and then deconvolved with conventional RL (5 and 25 iterations) or the sparse deconvolution algorithm, respectively (same parameters for punctated and ring-shaped structures).



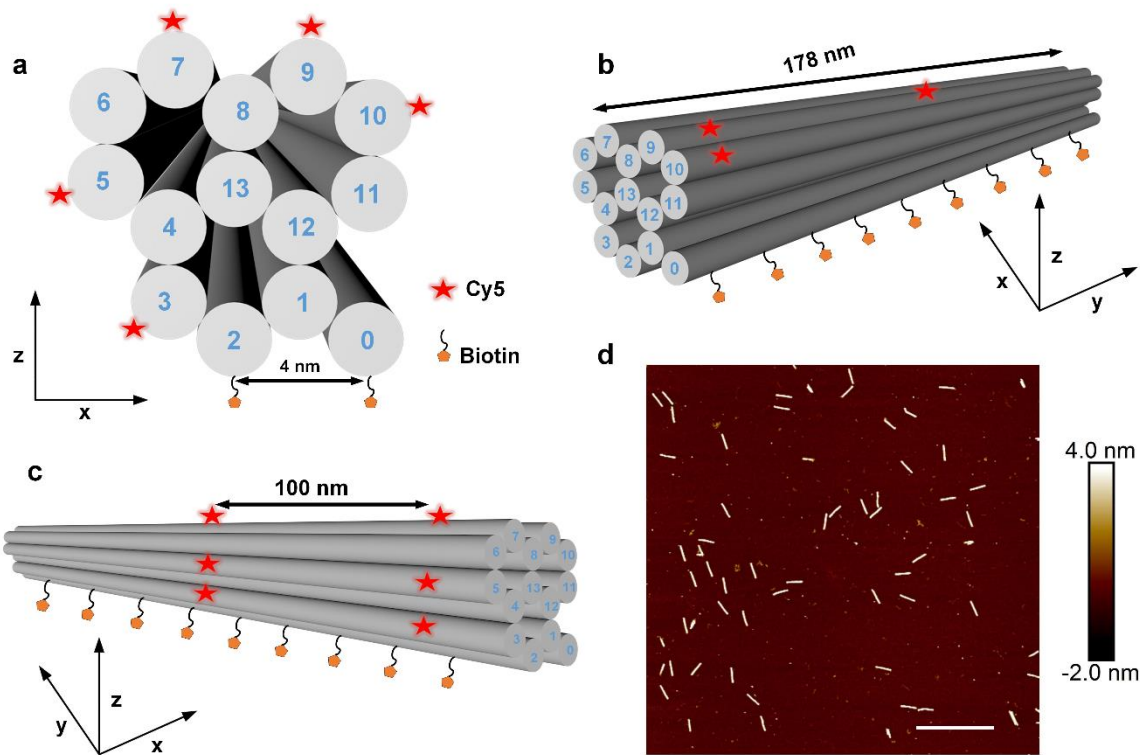
Supplementary Fig. 31 | Bona fide extension of spatial resolution by the sparse deconvolution when processing real biological structures. (a) From left to right are the high NA (1.4) image, the simulated low NA (0.8) image, and the low NA (0.8) image after sparse deconvolution (same data as in **Fig. 4a**). Here, the low NA image was generated using the formula $\text{iFFT}[\text{FFT}(\text{IMG}_{1.4}) \times (\text{OTF}_{0.8}/\text{OTF}_{1.4})]$. In the formula, the (i) FFT denotes the (inverse) fast Fourier transforms, and $\text{IMG}_{1.4}$ represents the high NA image. (b) The optical transform functions of $\text{NA} = 1.4$ ($\text{OTF}_{1.4}$) and $\text{NA} = 0.8$ ($\text{OTF}_{0.8}$). (c) The corresponding Gaussian-fitted cross-sections of the insets in (a). The two peaks with a distance of 259 nm are resolved under the low NA after sparse deconvolution, which is similar to 265 nm in the original high NA image. This indicates that sparse deconvolution achieves high NA resolution with better contrast. (d) The corresponding Fourier space images are presented from left to right. (e, f) We used live COS-7 cells overexpressing LifeAct-EGFP (e, *c.f.*, **Fig. 4a**, see also **Supplementary Video 15**) or Clathrin-EGFP (f) and imaged them with 2D-SIM, TIRF-SIM respectively. From the raw dataset, we used regular deconvolution, continuity reconstruction followed by deconvolution, upsampling plus sparsity, and continuity reconstruction, or upsampling plus sparsity and continuity reconstruction followed by deconvolution. Centerline, medians; limits, 75% and 25%; whiskers, maximum and minimum; error bars, s.e.m. Experiments were repeated five times independently with similar results. Scale bars: (a, b, d-f) 500 nm; (a, inset) 200 nm.



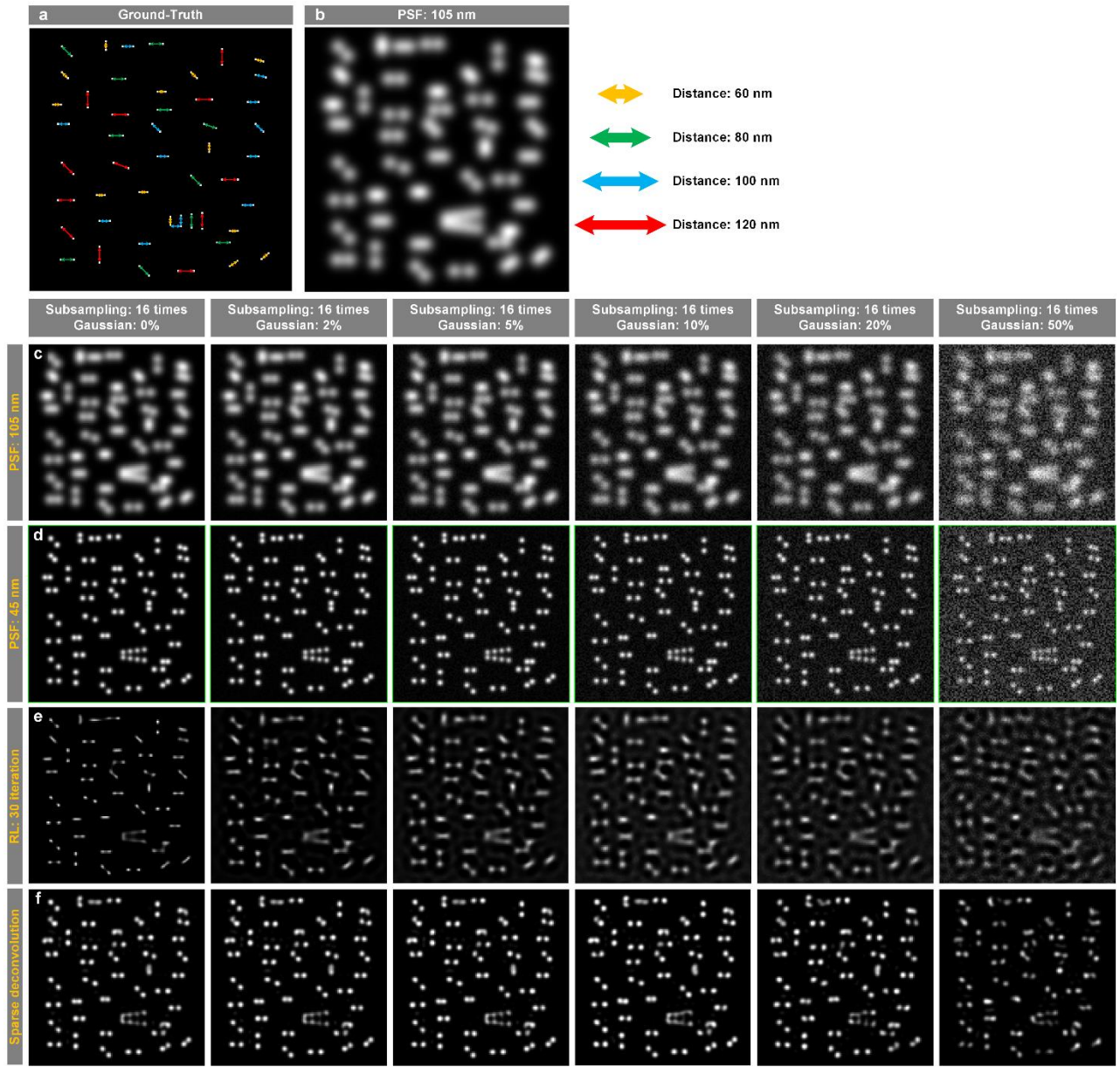
Supplementary Fig. 32 | Benchmarking of the resolution improvement of wide-field microscopy by the sparse deconvolution using pre-made pairs of horizontal lines of decreasing space. (a) The standard horizontal gradually spaced lines observed either by the wide-field microscope (NA 1.4), or Fourier interpolated, followed by the RL deconvolution with 20 (20 RL $\times 2$) or 200 iterations (200 RL $\times 2$), or Fourier interpolated followed by the sparse deconvolution (Sparse $\times 2$). (b) The wide-field, 20 RL $\times 2$, 200 RL $\times 2$, Sparse $\times 2$ configurations were able to distinguish pair lines up to 270 nm, 300 nm, 240 nm, and 150 nm, respectively. Experiments were repeated five times independently with similar results. Scale bar: 3 μm .



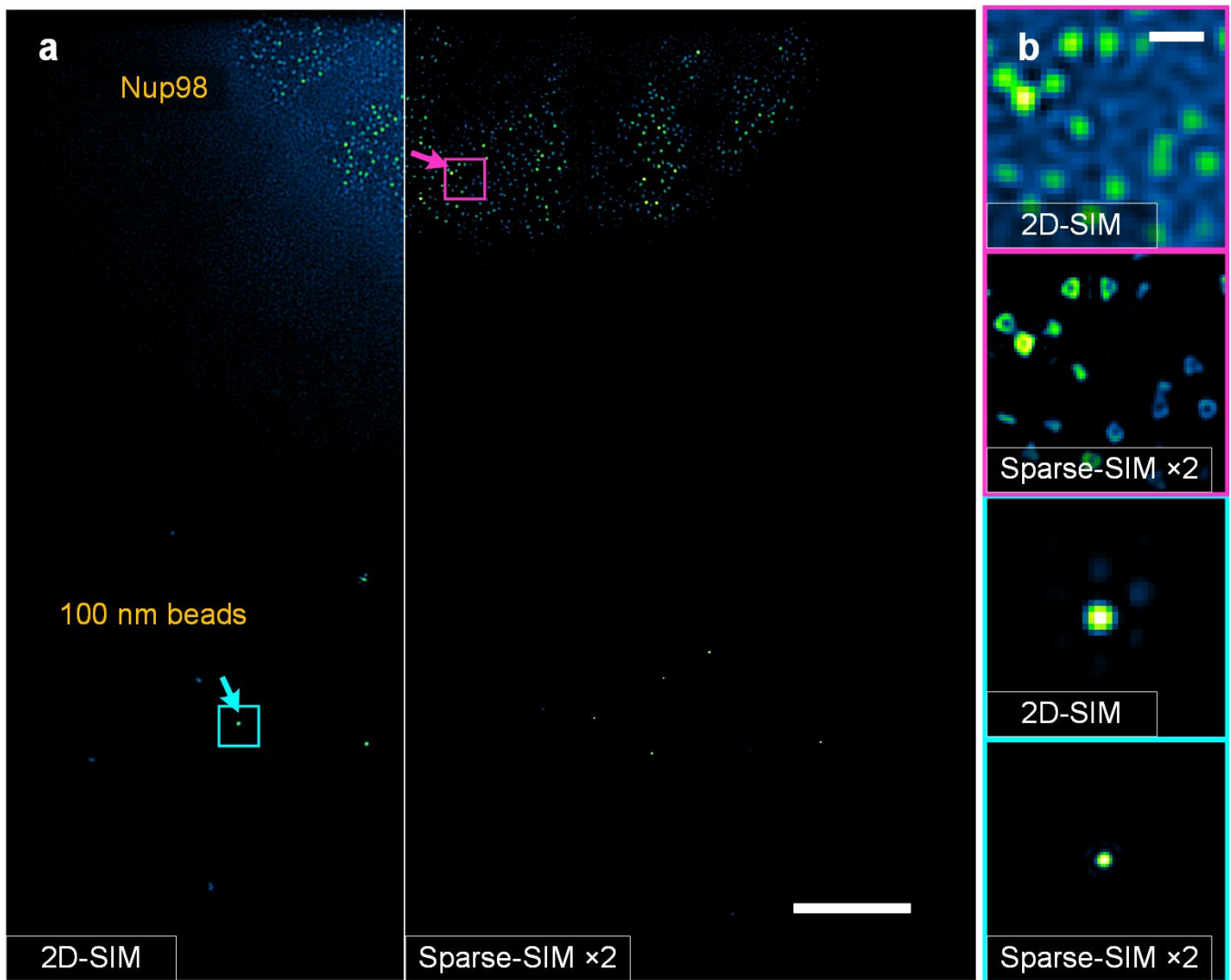
Supplementary Fig. 33 | OTFs obtained by the Fourier transform of fluorescent beads visualized under different conditions. (a) Images of fluorescent beads (48 nm in diameter) under wide-field imaging: (left upper), wide-field followed by deconvolution (right upper), wide-field plus sparsity and continuity reconstruction (left bottom), and wide-field plus sparsity and continuity reconstruction followed by deconvolution (right bottom). Upper insets: A single bead observed under different conditions; Lower insets: the corresponding Fourier transform of the image on the left. **(b)** The amplitudes along the white lines in the frequency domain from the low insets in **a**, while the two vertical dashed lines indicate the wide-field microscope's spatial frequency limits. Experiments were repeated five times independently with similar results. Scale bar: 5 μm ; inset: 500 nm.



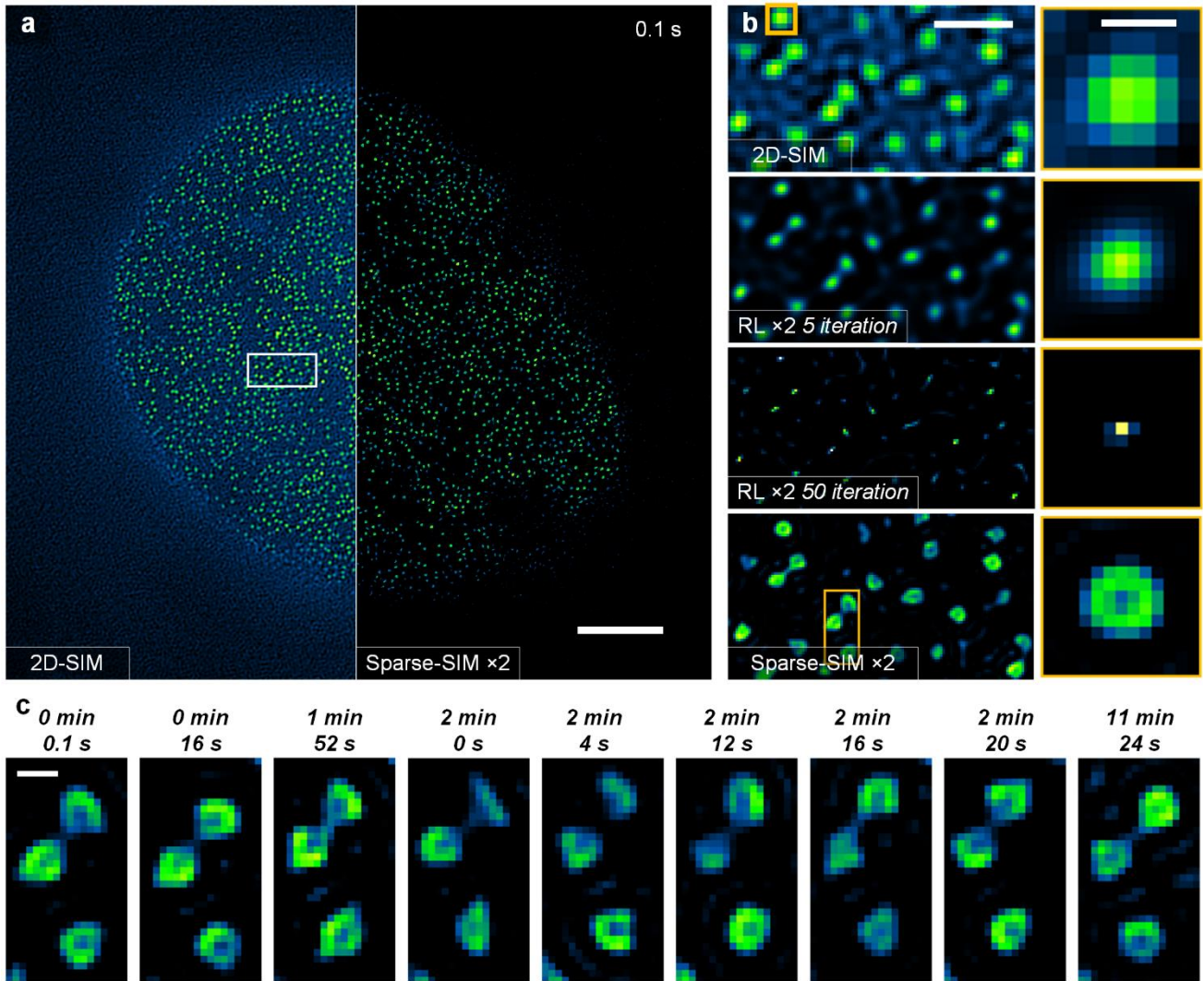
Supplementary Fig. 34 | Design and representative images of the 100 nm tubular DNA origami nanostructure. (a) Front-view schematic diagram of 14-HB DNA origami, including incorporating the Cy5/Alexa Fluor 647 dyes and biotin at 3' ends of the selected staple oligonucleotides. We attached biotin to two helix bundles (HB0 and HB2), and Cy5 to five (HB3, HB5, HB7, HB9, HB10), or four helix bundles (HB3, HB5, HB7, HB9). (b) Right-side view of the schematic diagram. (c) Left-side view of the schematic diagram. (d) DNA origami over a large field of view under the atomic force microscope. Experiments were repeated five times independently with similar results. Scale bar: 1 μm .



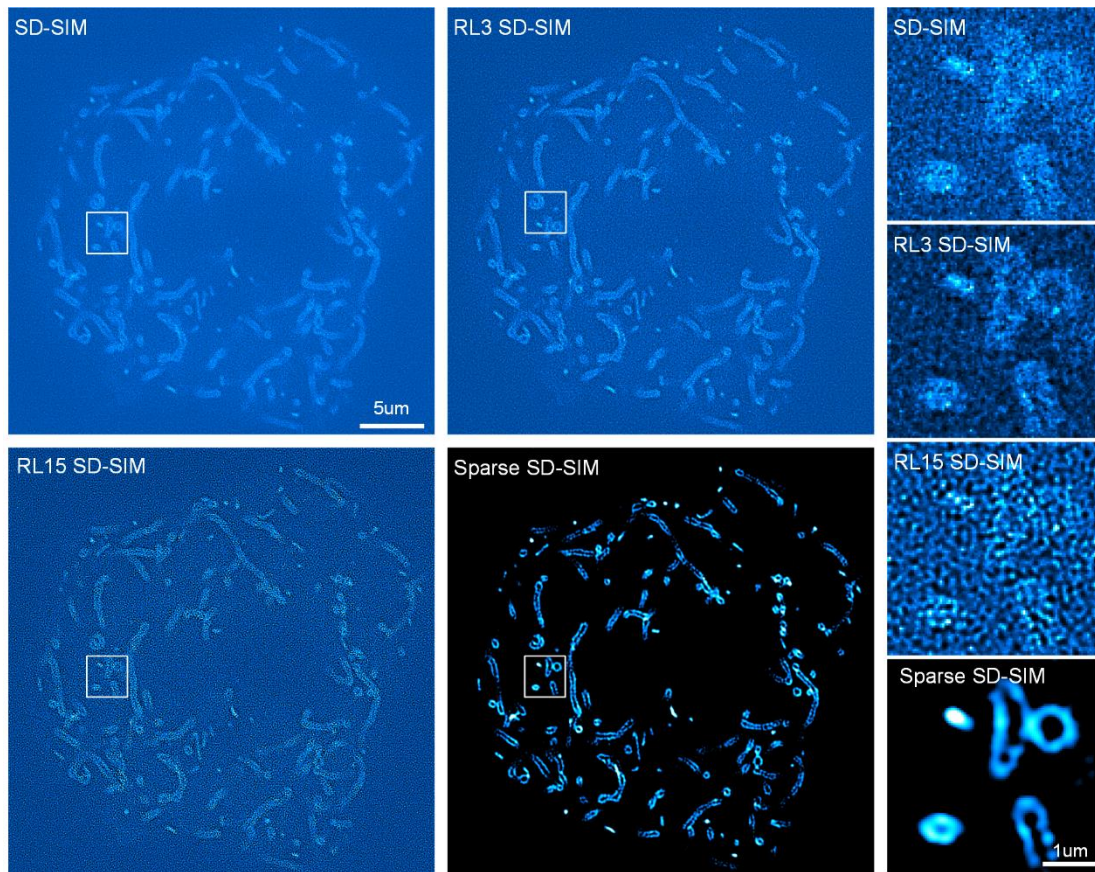
Supplementary Fig. 35 | The evolution of noise effects on simulated fluorescence pairs resolved by the RL deconvolution and Sparse deconvolution. (a) We simulated fluorescence pairs of different distances (60/80/100/120 nm) mixed in the same field-of-view as ground-truth. The ground truth in (a) was convolved with PSF with FWHM of 105 nm (b). (c, d) The ground-truth in (a) was convolved with PSF with FWHM of either 105 nm (c) or 45 nm (d), and such images were subsampled 16 times and corrupted with 0%, 2%, 5%, 10%, 20%, and 50% Gaussian noise subsequently. (e, f) The images in (c) were reconstructed by RL deconvolution with 30 iterations (e) and Sparse deconvolution (f).



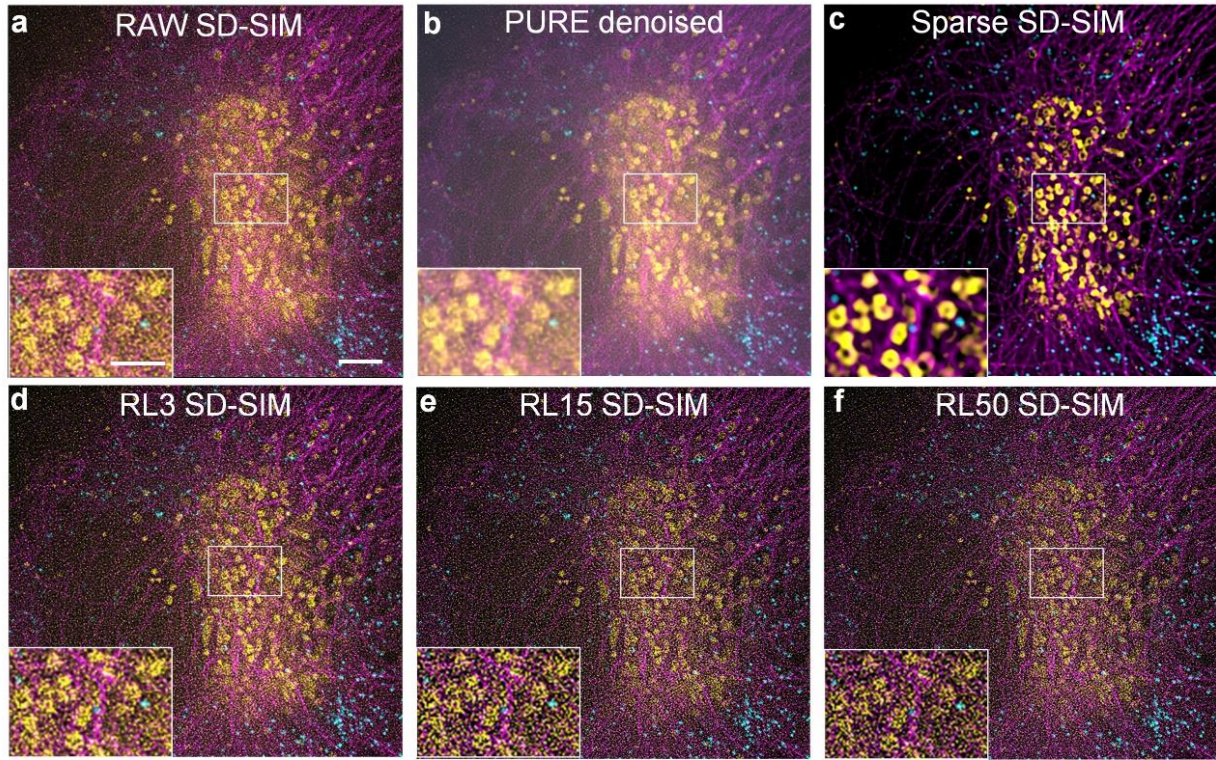
Supplementary Fig. 36 | Nup98 pores and 100 nm beads in the same FOV under the 2D-SIM and the Sparse-SIM $\times 2$. (a) Snapshots under the 2D-SIM and the Sparse-SIM $\times 2$ configurations were shown on the left and right side. (b) Magnified view of the regions in (a) under 2D-SIM and Sparse-SIM $\times 2$. The magenta and cyan boxes represent the view of Nup98 and 100 nm bead, respectively. Experiments were repeated five times independently with similar results. Scale bars: (a) 3 μm ; (b) 300 nm.



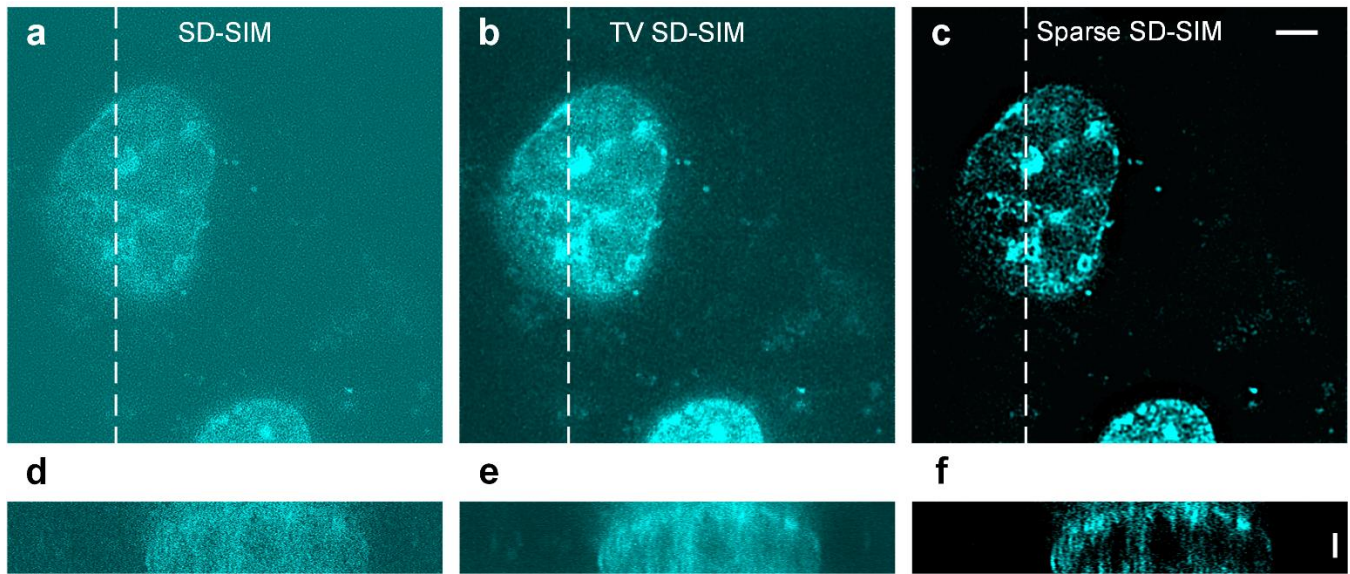
Supplementary Fig. 37 | Sparse-SIM detected long-term dynamics of ring-shaped nuclear pores labeled by Nup98-GFP in a live COS-7 cell (the holistic view of Fig. 2c). (a) Snapshots were acquired in 0.1 sec under the 2D-SIM, and the Sparse-SIM $\times 2$ configurations were shown on the left and right sides. (b) Magnified view of the region in (a) under different configurations were shown in the left, from which the nuclear pore enclosed in the yellow box was further zoomed up and shown in the right panels. (c) The yellow box in (b) is enlarged and shown at nine time points. Experiments were repeated five times independently with similar results. Scale bars: (a) 3 μm ; (b) 500 nm; 100 nm; (c) 100 nm.



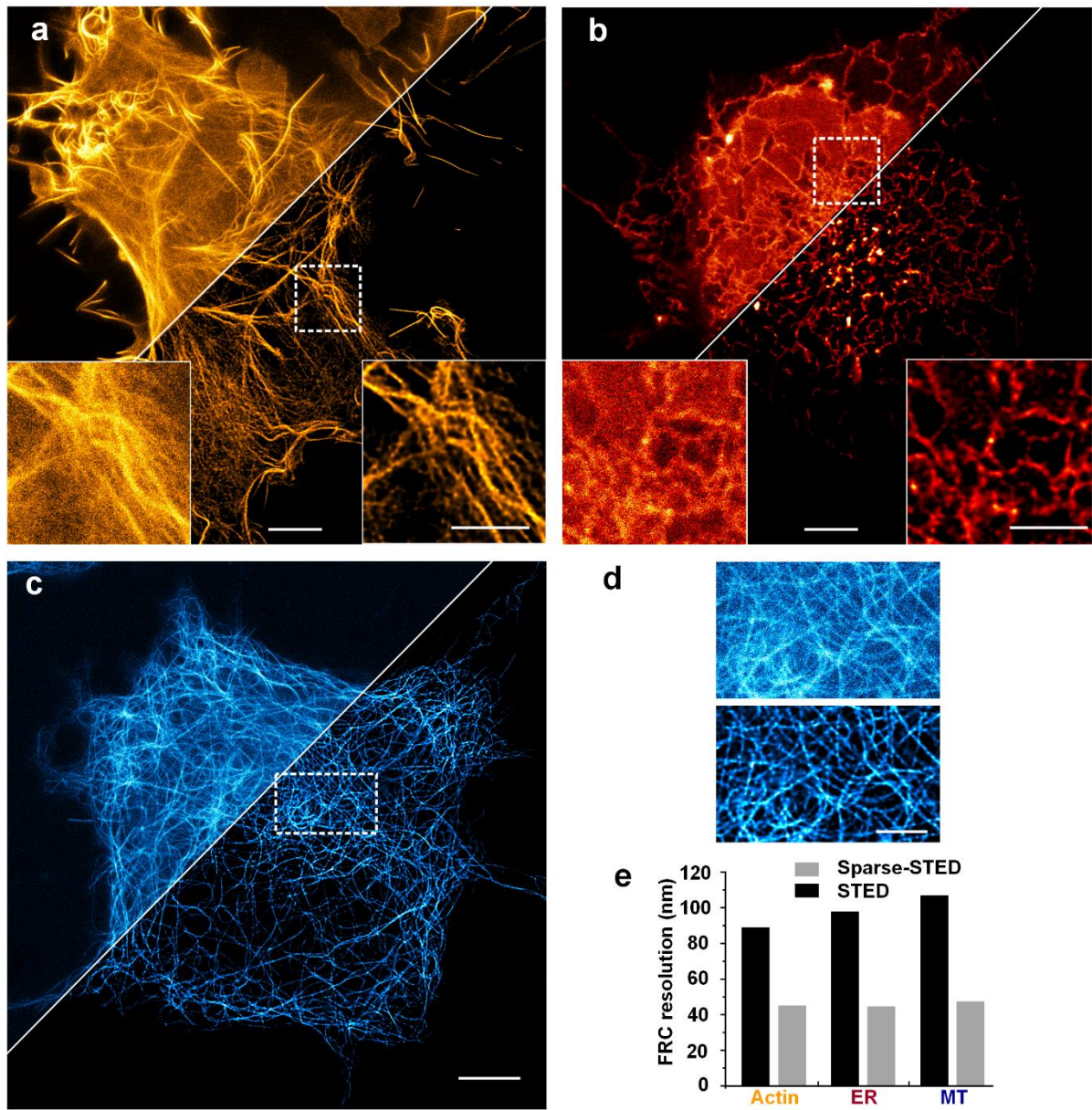
Supplementary Fig. 38 | Comparing results of SD-SIM, RL deconvolved SD-SIM, and Sparse SD-SIM (*c.f.*, Fig. 5k). (a) OMM in a live COS-7 cell (labeled with TOM20-mCherry) under SD-SIM, SD-SIM deconvolved by RL with 3 and 15 iterations, and Sparse SD-SIM, respectively. (e) Magnified views from white boxes in (a-d) respectively. Scale bars: (a-d) 5 μm ; (e) 1 μm .



Supplementary Fig. 39 | Comparing results of SD-SIM, PURE denoised SD-SIM, Sparse SD-SIM, and RL deconvolved SD-SIM. (c.f., Fig. 6e). (a-c) HeLa cell labeled with tubulin-EGFP (magenta), Pex11a-BFP (cyan), and Lamp1-mCherry (yellow) captured by SD-SIM, PURE denoised⁴³ SD-SIM, Sparse SD-SIM. (d) Raw SD-SIM deconvolved by RL with 3 iterations. (e) Raw SD-SIM deconvolved by RL with 15 iterations. (f) Raw SD-SIM deconvolved by RL with 50 iterations. PURE: Poisson Unbiased Risk Estimate⁴³. Scale bars: (a-f) 5 μm ; (a-f, inset) 3 μm .



Supplementary Fig. 40 | Comparing results of SD-SIM, TV SD-SIM, and Sparse SD-SIM (c.f., Fig. 6j). (a-c) Hoechst in a live COS-7 cell under the SD-SIM (axial position at +3 μm) (a), TV³¹ SD-SIM (b), and Sparse SD-SIM (c), respectively. (d-f) Corresponding z-axial views from the white dashed lines in (a-c). TV: Total Variation³¹; Scale bars: (a-c) 5 μm ; (d-f) 2 μm .



Supplementary Fig. 41 | Extending the spatial resolution of STED microscopy with sparse deconvolution. (a, b) A live COS-7 cell expressing either Lifeact-GFP (a) or Sec61 β -GFP (b) visualized under the STED microscope (top), or after the sparse deconvolution (bottom). Insets show magnified views of the same region seen under the STED (lower left inset) and after the sparse deconvolution (lower right inset). (c-d) A snapshot of SiR-Tubulin-labeled microtubules seen under the STED (top) and Sparse-STED (bottom, c). The region enclosed by the white dashed box in (c) was magnified and shown in d (STED, top; Sparse-STED, bottom). (e) FRC minimum resolutions of actin, ER, and microtubules in (a-c) under STED and the Sparse-STED. ER: endoplasmic reticulum; MT: microtubule; Scale bars: (a, b, c) 5 μ m; (a inset, b inset and d) 2 μ m.

Supplementary Tables

Supplementary Table 1 | SSIM of different reconstructions from images corrupted with varying levels of noise.

	Without noise	G (11%)	G (25%)	G (50%)	G (80%)
Wiener-SIM	0.5612	0.2352	0.1773	0.1417	0.1445
Continuity	0.5500	0.2303	0.1724	0.1311	0.1545
Continuity+sparsity	0.8276	0.6741	0.4766	0.5652	0.5418
Continuity^{+deconv}	0.8311	0.2886	0.1724	0.1790	0.2107
Sparsity^{+deconv}	0.9430	0.5203	0.4178	0.2884	0.3191
Sparse-SIM	0.9570	0.6832	0.6431	0.6167	0.6376

G represented the Gaussian noise. From **Supplementary Fig. 29**.

Supplementary Table 2 | PSNR of different reconstructions from images corrupted with varying levels of noise.

	Without noise	G (11%)	G (25%)	G (50%)	G (80%)
Wiener-SIM	18.08	10.90	9.39	8.27	11.32
Continuity	17.94	10.85	9.40	8.23	9.73
Continuity+sparsity	21.63	18.10	16.09	16.13	16.40
Continuity^{+deconv}	23.01	13.93	11.77	10.68	13.22
Sparsity^{+deconv}	27.04	17.74	15.88	15.29	15.24
Sparse-SIM	27.92	18.72	18.08	17.97	18.11

G represented the Gaussian noise. From **Supplementary Fig. 29**.

Supplementary Table 3 | Imaging conditions.

Microscope	Figure	Label	Wavelength (nm)	Exposure (ms)	Illumination intensity (W/cm ²)
SIM	Fig. 2c	Nup-98GFP	488	7×9	21
	Fig. 4a	LifeAct-EGFP	488	20×9	2
	Fig. 4j	Caveolin-EGFP	488	7×9	14
	Fig. 4m	LAMP1-EGFP	488	20×9	2
		LysoView 488	488	20×9	2
		LipidSpot 488	488	20×9	2
	Fig. 4o	VAMP2-pHluorin	488	0.2×9	187
	ExFig. 5f	Clathrin-EGFP	488	7×9	14
	ExFig. 14a	MitoTracker Green	488	7×9	14
		Tom20-mScarlet	561	7×9	50
	ExFig. 14f	MitoTracker Green	488	7×9	14
		Sec61β-mCherry	561	7×9	100
Expansion-SIM	Fig. 3	α-tubulin Sec61β-GFP	488	20×9	20
SD-SIM	Fig. 5a	Clathrin-EGFP	488	200	0.91
	Fig. 5g, ExFig. 6	Lifeact –EGFP	488	200	0.89
		Clathrin-DsRed	561	300	12.11
		Hoechst H1399	405	200	0.87
	Fig. 5h	Tubulin-EGFP	488	200	0.91
		Lamp1-mCherry	561	200	8.84
		MitoTracker® Deep Red FM	640	50	1.07
	Fig. 5k	Tom20-mCherry	561	200	8.84
	Fig. 6a	Sec61β-EGFP	488	200	0.91
	Fig. 6e	Pex11a-BFP	405	100	0.87
		Tubulin-EGFP	488	300	0.91
		Lamp1-mCherry	561	200	4.96
	Fig. 6h	Hoechst H1399	405	50	0.87
		Tubulin-EGFP	488	200	0.89
		MitoTracker® Deep Red FM	640	50	0.27
	ExFig. 7	Sec61β-EGFP	488	200	0.77
		Lysotracker Deep Red	640	200	0.92
	ExFig. 10	GCaMP6s	488	90	0.27
STED	ExFig. 8a	Mab414	590		Et.: 112 Dt.: 229×10 ⁶
	ExFig. 8f	Tom20	594		Et.: 74 Dt.: 48×10 ⁶
		Tubulin	635		Et.: 97 Dt.: 14×10 ⁶
	SI Fig. 41a	Lifeact-GFP	488		Et.: 109 Dt.: 16×10 ⁶
	SI Fig. 41b	Sec61β-GFP	488		Et.: 156 Dt.: 13×10 ⁶
	SI Fig. 41c	SiR-Tubulin	651		Et.: 307 Dt.: 158×10 ⁶
MTPM	ExFig. 9	Thy1-GFP	920		3.98×10 ⁶

ExFig.: Extended Data Figure; SI Fig.: Supplementary Figure; Et.: Excitation laser; Dt.: Depletion laser.

Supplementary Table 4 | Non-zero ratios and background selections of different images.

Figure	Label	Modality	Non-zero number (Raw / Sparse)	Non-zero ratio (Raw / Sparse)	Pixel number	Background
Fig. 2c	Nup98-GFP	TIRF-SIM	1037042 / 159384	99 % / 4 %	1024×1024 ($\times 2$)	Weak-HI
Fig. 4a	LifeAct-EGFP	2D-SIM	1675100 / 589824	71 % / 25 %	1536×1536	No
Fig. 4j	Caveolin-EGFP	TIRF-SIM	566376 / 1215637	82 % / 44 %	793×871 ($\times 2$)	No
Fig. 4m	LAMP1-EGFP	2D-SIM	2312110 / 117965	98 % / 5 %	1536×1536	Strong-HI
Fig. 4m	LysoView 488	2D-SIM	2345140 / 212336	99 % / 9 %	1536×1536	Strong-HI
Fig. 4m	LipidSpot 488	2D-SIM	2264924 / 94371	96 % / 4 %	1536×1536	Strong-HI
Fig. 4o	VAMP2-pHluorin	TIRF-SIM	64659 / 1180	88 % / 0.04 %	512×144 ($\times 2$)	No
Fig. 5a	Clathrin-EGFP	SD-SIM	751564 / 37578	100 % / 5 %	899×836	Strong-HI
Fig. 5h	Lamp1-mCherry	SD-SIM	1327104 / 185795	100 % / 14 %	1152×1152	Strong-HI
Fig. 5h	MitoTracker® Deep Red FM	SD-SIM	1327104 / 225608	100 % / 17 %	1152×1152	Strong-HI
Fig. 5h	Hoechst H1399	SD-SIM	1327104 / 1008599	100 % / 76 %	1152×1152	Strong-LI
Fig. 5h	Tubulin-EGFP	SD-SIM	1327104 / 1167852	100 % / 88 %	1152×1152	Weak-LI
Fig. 5k	Tom20-mCherry	SD-SIM	1015950 / 547597	100 % / 54 %	1042×975 3D-MIP	Strong-HI

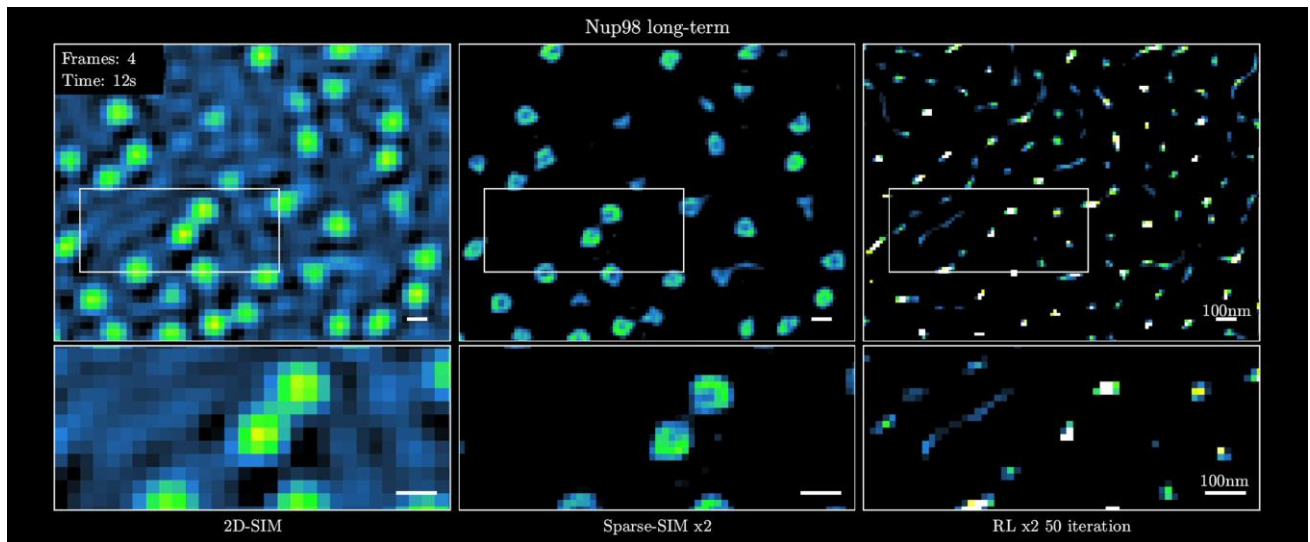
Continued Table of Supplementary Table 4...

Figure	Label	Modality	Non-zero number (Raw / Sparse)	Non-zero ratio (Raw / Sparse)	Pixel number	Background
Fig. 6a	Sec61 β -EGFP	SD-SIM	262144 / 555745	100 % / 53 %	512 $\times$ 512 ($\times$ 2)	Weak-LI
Fig. 6e	Tubulin-EGFP	SD-SIM	262144 / 786432	100 % / 75 %	512 $\times$ 512 ($\times$ 2)	No
Fig. 6e	Lamp1-mCherry	SD-SIM	262144 / 94371	100 % / 9 %	512 $\times$ 512 ($\times$ 2)	Strong-LI
Fig. 6e	Pex11a-BFP	SD-SIM	262144 / 41943	100 % / 4 %	512 $\times$ 512 ($\times$ 2)	No
Fig. 6h	Hoechst H1399	SD-SIM	262144 / 576717	100 % / 55 %	512 $\times$ 512 ($\times$ 2) 3D-MIP	Strong-HI
Fig. 6h	MitoTracker® Deep Red FM	SD-SIM	262144 / 377487	100 % / 36 %	512 $\times$ 512 ($\times$ 2) 3D-MIP	Strong-HI
Fig. 6h	Tubulin-EGFP	SD-SIM	262144 / 450888	100 % / 43 %	512 $\times$ 512 ($\times$ 2) 3D-MIP	Strong-LI
ExFig.4a	MitoTracker Green	2D-SIM	249037 / 99615	95 % / 38 %	512 $\times$ 512	No
ExFig.4a	Tom20-mScarlet	2D-SIM	209715 / 39322	80 % / 15 %	512 $\times$ 512	No
ExFig.4e	MitoTracker Green	2D-SIM	254280 / 18350	97 % / 7 %	512 $\times$ 512	No
ExFig.4e	Sec61 β -mCherry	2D-SIM	214958 / 28836	82 % / 11 %	512 $\times$ 512	No
ExFig.10	GCaMP6s	SD-SIM	2303136 / 852160	100 % / 37 %	1454 $\times$ 1584	No

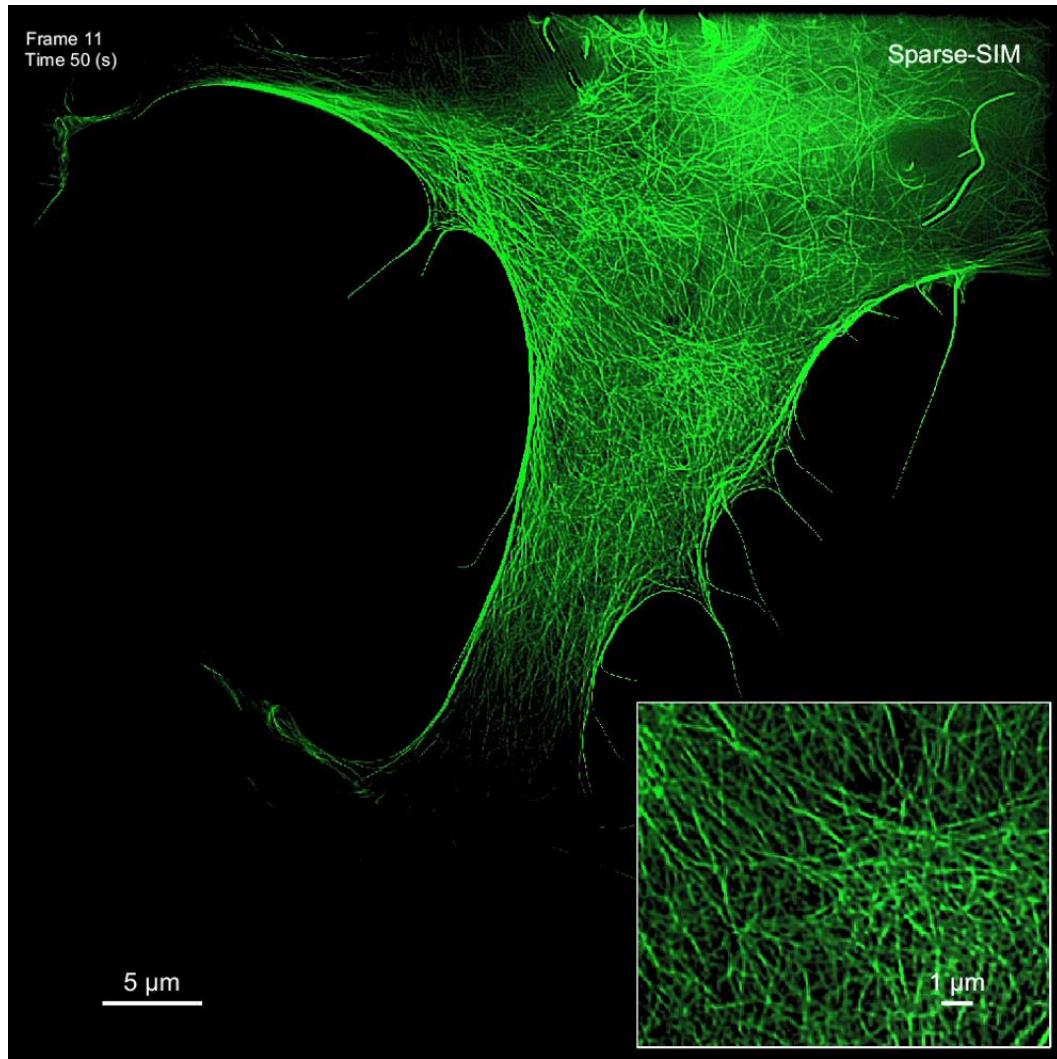
Supplementary Table 5 | Computing time of sparse deconvolution.

Figure	Label	Size ($x \times y \times t/z$)	GPU time	CPU time
Fig. 2c	Nup-98GFP	$2048 \times 2048 \times 200$ (t)	~1800 s	~1812 min
Fig. 3a	α -tubulin	$2048 \times 2048 \times 10$ (t)	~90 s	~92 min
Fig. 3d	Sec61 β -GFP	$2048 \times 2048 \times 10$ (t)	~90 s	~92 min
Fig. 4a	LifeAct-EGFP	$1536 \times 1536 \times 20$ (t)	~90 s	~46 min
Fig. 4j	Caveolin-EGFP	$2048 \times 2048 \times 200$ (t)	~1680 s	~1708 min
Fig. 4m	LAMP1-EGFP	$1536 \times 1536 \times 50$ (t)	~188 s	~120 min
Fig. 5a	Clathrin-EGFP	$873 \times 873 \times 100$ (t)	~140 s	~80 min
Fig. 5g	Clathrin-DsRed	$1152 \times 1152 \times 200$ (t)	~425 s	~329 min
Fig. 5k	Tom20-mCherry	$1152 \times 1152 \times 110$ (z)	~231 s	~181 min
Fig. 5h	Hoechst H1399	$1152 \times 1152 \times 100$ (t)	~210 s	~165 min

Captions for Supplementary Videos.



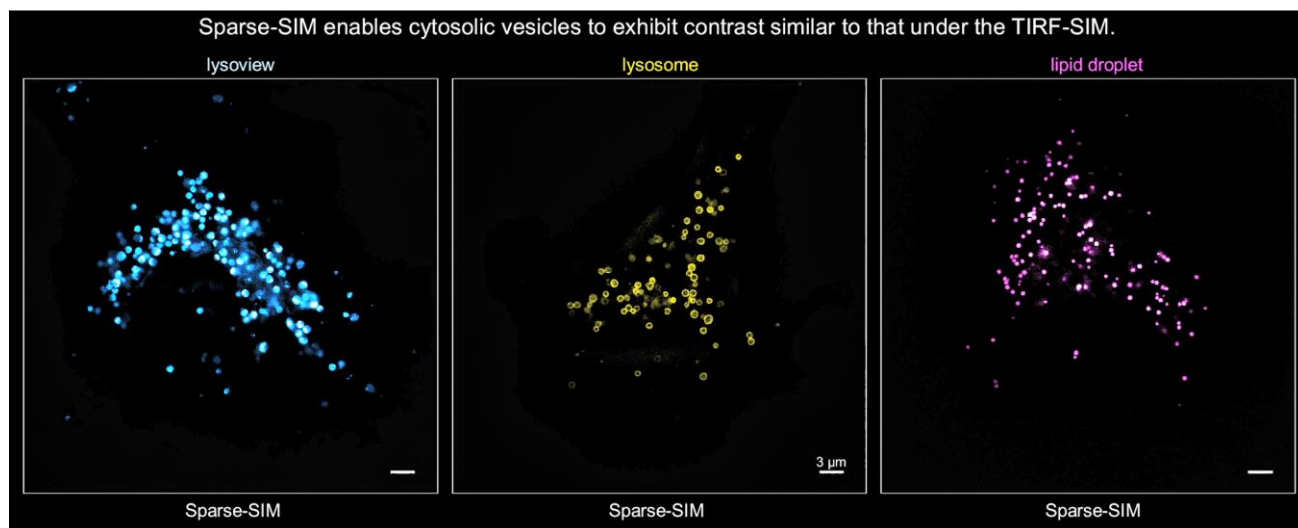
Supplementary Video 1 | Ring-shaped Nup98 pores resolved by Sparse-SIM 2 $\times$. Nuclear pores labeled with Nup98-GFP in a live COS-7 cell with 4 s interval imaged by 2D-SIM, 2D-SIM followed by RL deconvolution and Sparse-SIM 2 $\times$ (*c.f.*, **Fig. 2c**) demonstrate that only Sparse-SIM 2 $\times$ can resolve the ring-shaped Nup98.



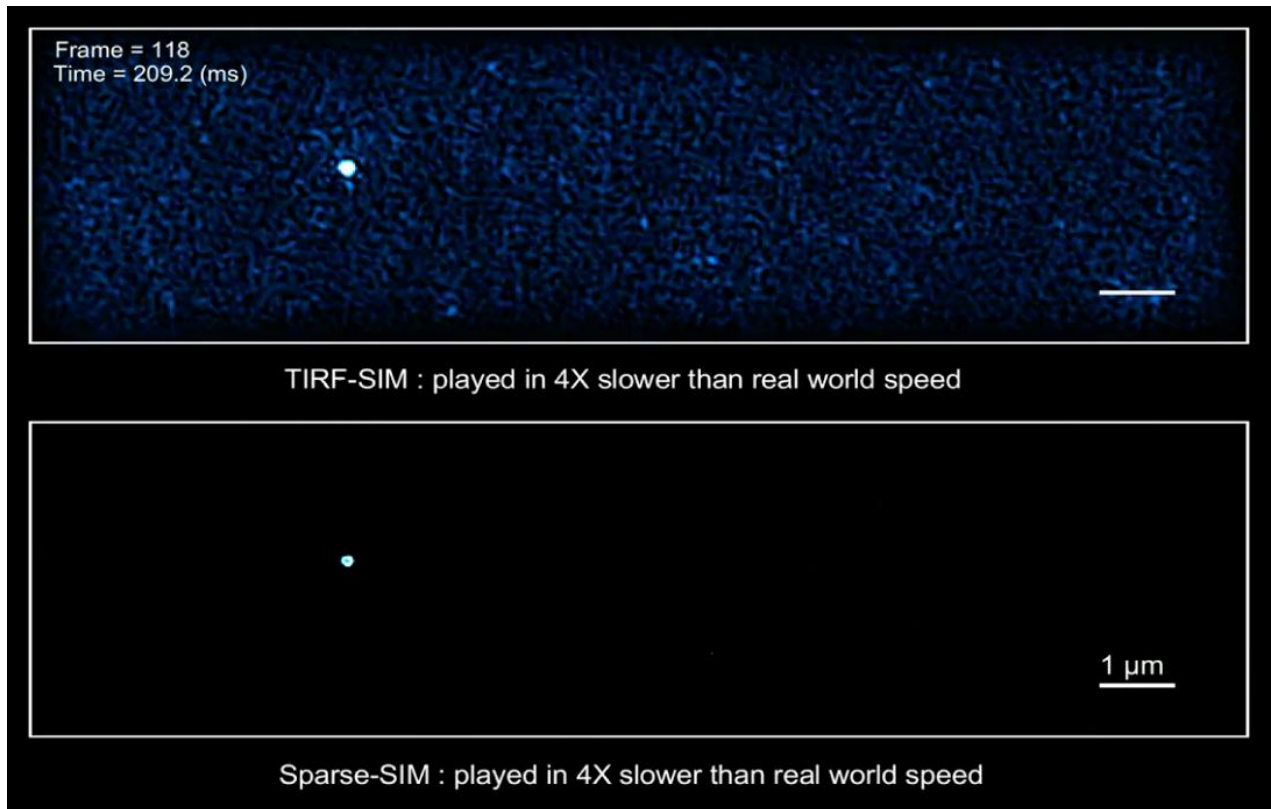
Supplementary Video 2 | Dense actin mesh network revealed due to improved spatial resolution and enhanced contrast. Part I compares the actin results by 2D-SIM, Hessian-SIM and Sparse-SIM, demonstrating that finer structures of the dense actin mesh network can be revealed only with Sparse-SIM (*c.f.*, **Fig. 4a**). Part II exhibits the corresponding profiles of the red and blue lines from 2D-SIM (left) and Sparse-SIM (right) over 20 frames at 5 s intervals, showing the enhanced contrast available with Sparse-SIM.



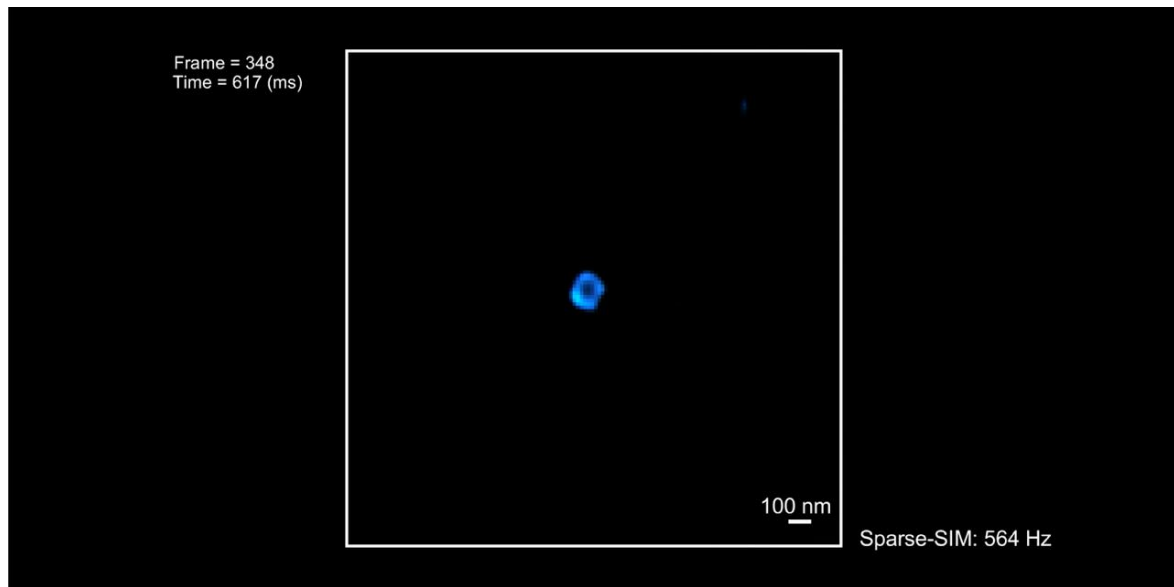
Supplementary Video 3 | Ring-shaped caveolae revealed by Sparse-SIM 2×. The caveolae in a COS-7 cell at 37°C transfected with Caveolin-EGFP by TIRF, TIRF-SIM, Sparse-SIM and Sparse-SIM 2× (*c.f.*, **Fig. 4j**).



Supplementary Video 4 | Improved contrast of fluorescent vesicles deep in the cytosol under Sparse-SIM. Part I the structures labeled with LAMP1-EGFP, LysoView, and LipidSpot in COS-7 cells are shown from left to right at 5 s intervals (*c.f.*, **Fig. 4m**). Part II compares the 2D-SIM, Hessian-SIM and Sparse-SIM performances on three individual organelles and the corresponding magnified regions of interest.

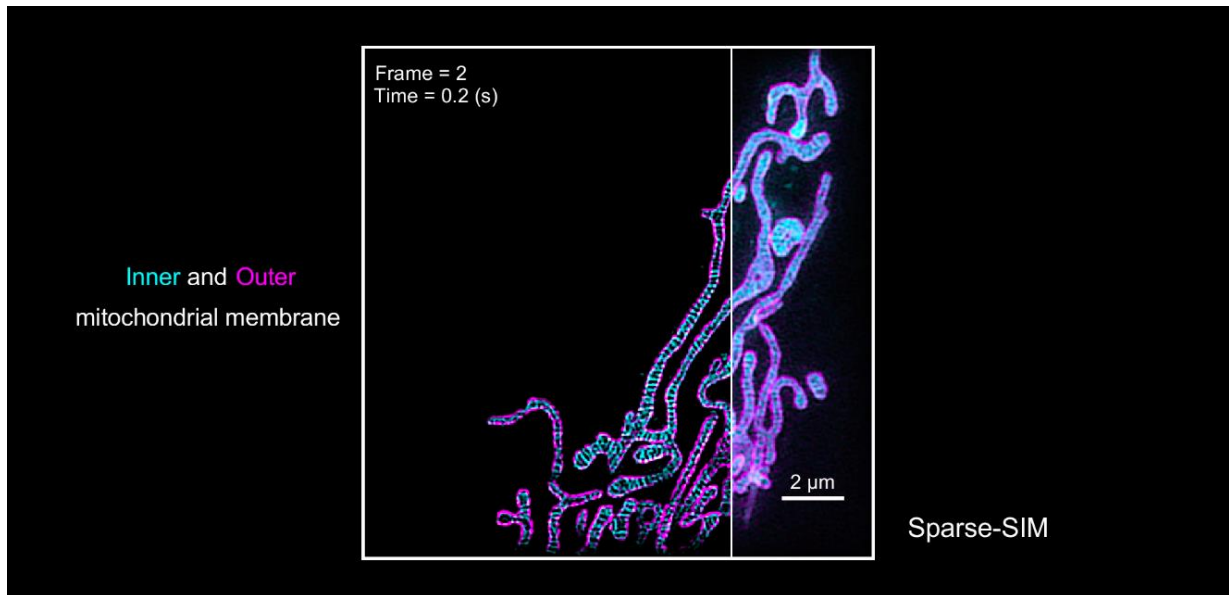


Supplementary Video 5 | Visualization of ultrafast vesicle fusion (played 4× slower than real-world speed). Fusion pores by TIRF-SIM (top) and Sparse-SIM (bottom) played at a speed 4x slower than real world rates. Vesicle in an INS-1 cell labeled with VAMP2-pHluorin by TIRF-SIM and Sparse-SIM, respectively (*c.f.*, **Fig. 4o**).

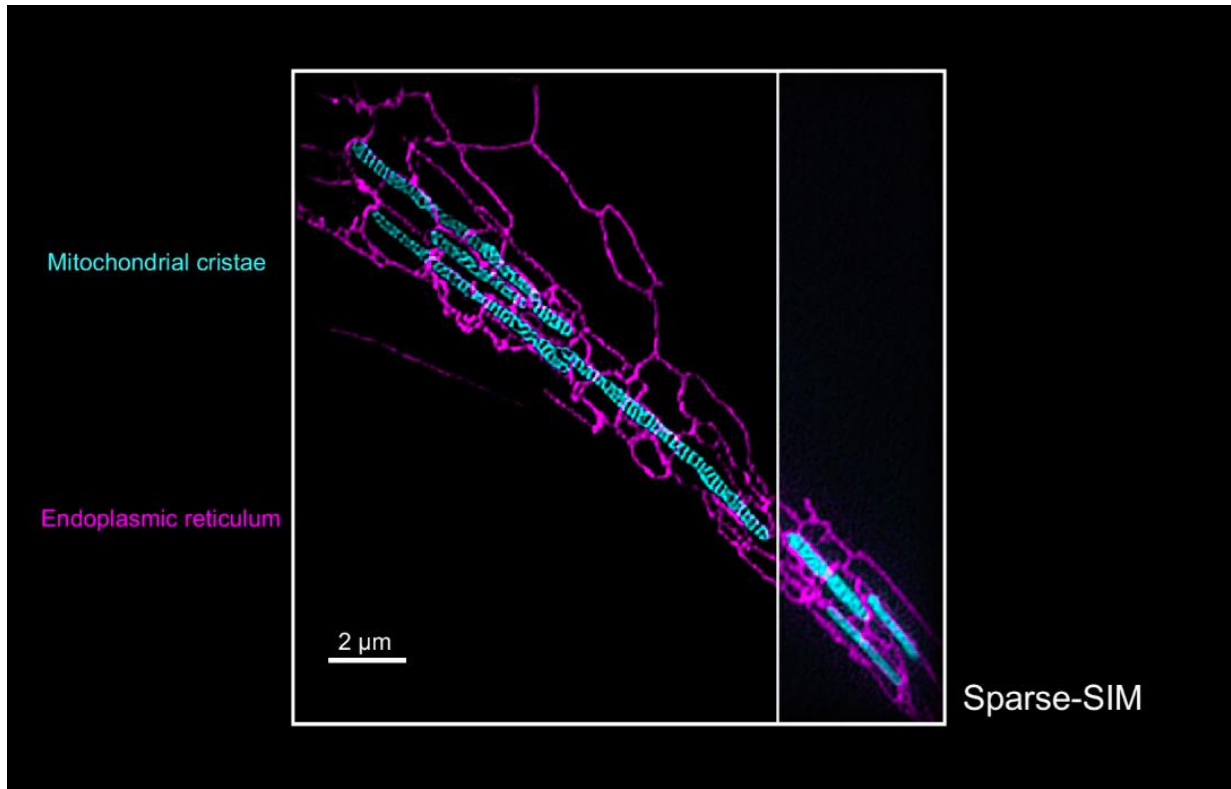


Supplementary Video 6 | Vesicle with a 60 nm pore diameter captured by the Sparse-SIM at 564 Hz.

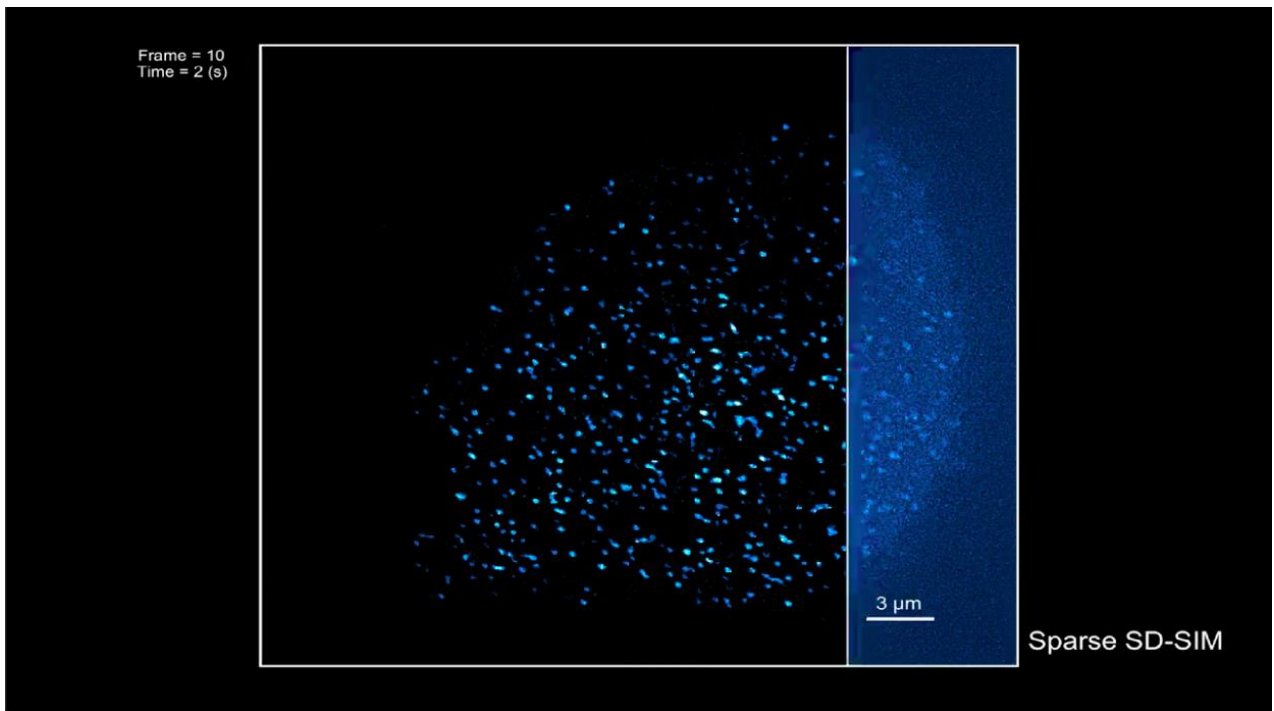
The initial opening of the fusion pore occurred much earlier, and its duration was much longer under Sparse-SIM than under TIRF-SIM (same data as in **Supplementary Video 5**).



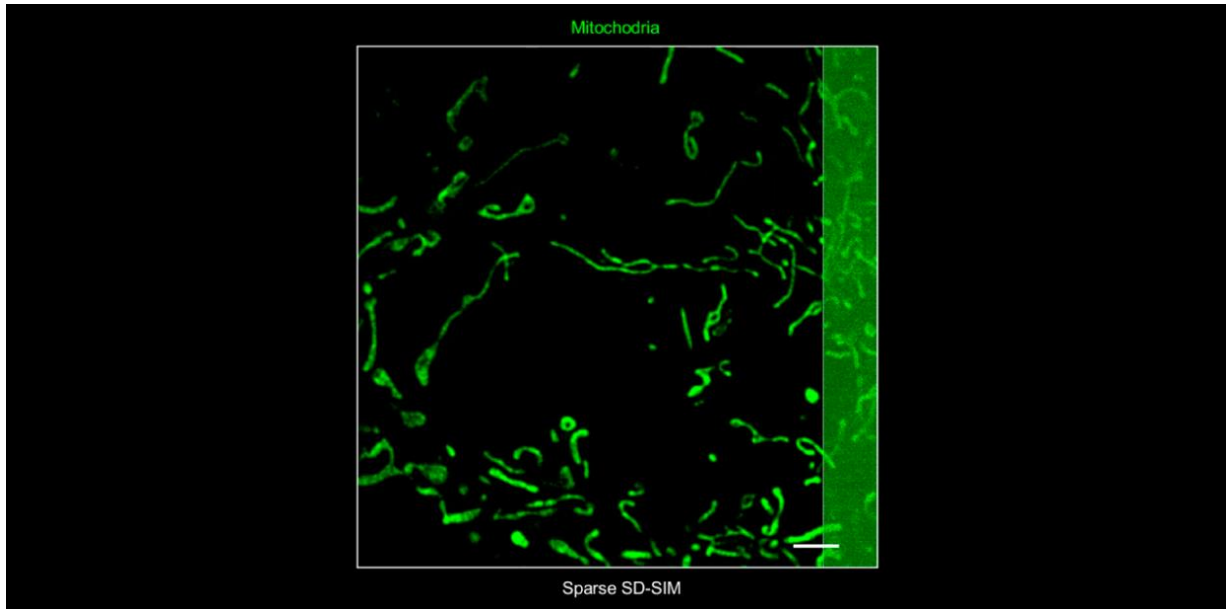
Supplementary Video 7 | Asynchronized movements between the outer and inner mitochondrial membranes revealed by dual-color Sparse-SIM. Outer and inner mitochondrial membranes (OMM and IMM, labeled by Tom20- mScarlet and Mito-Tracker Green, respectively) by 2D-SIM and Sparse-SIM for 30 frames with a 0.2 s acquisition time (*c.f.*, **Extended Data Fig. 4a**). The extension of the inner mitochondrial membrane was not enclosed by the Tom20-labeled structures captured by Sparse-SIM in a few frames, demonstrating the nonhomogenous distribution of Tom20 on the OMM.



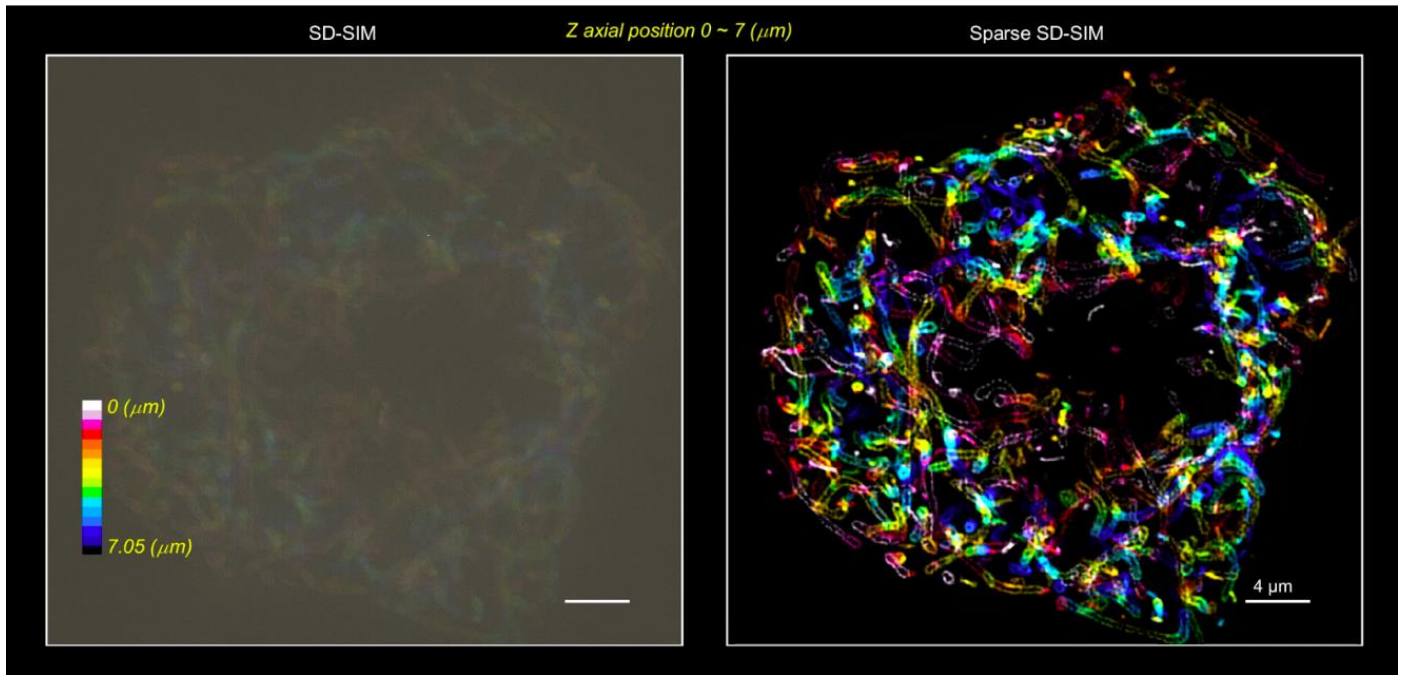
Supplementary Video 8 | Relative dynamics between ER tubules and inner mitochondrial membrane recorded by dual-color Sparse-SIM. The inner mitochondrial membrane (cyan) and ER (magenta) in a COS-7 cell labeled by MitoTracker Green and Sec61 β -mCherry (*c.f.*, **Extended Data Fig. 4f**). Part I compares the IMM, and ER by dual-color 2D-SIM and Sparse-SIM. In Part II, the ER contacts the IMM, rearranging the orientations of the inner cristae structures. Part III shows the equal probability of ER tubules contacting the mitochondria at cristae regions or in the matrix between cristae.



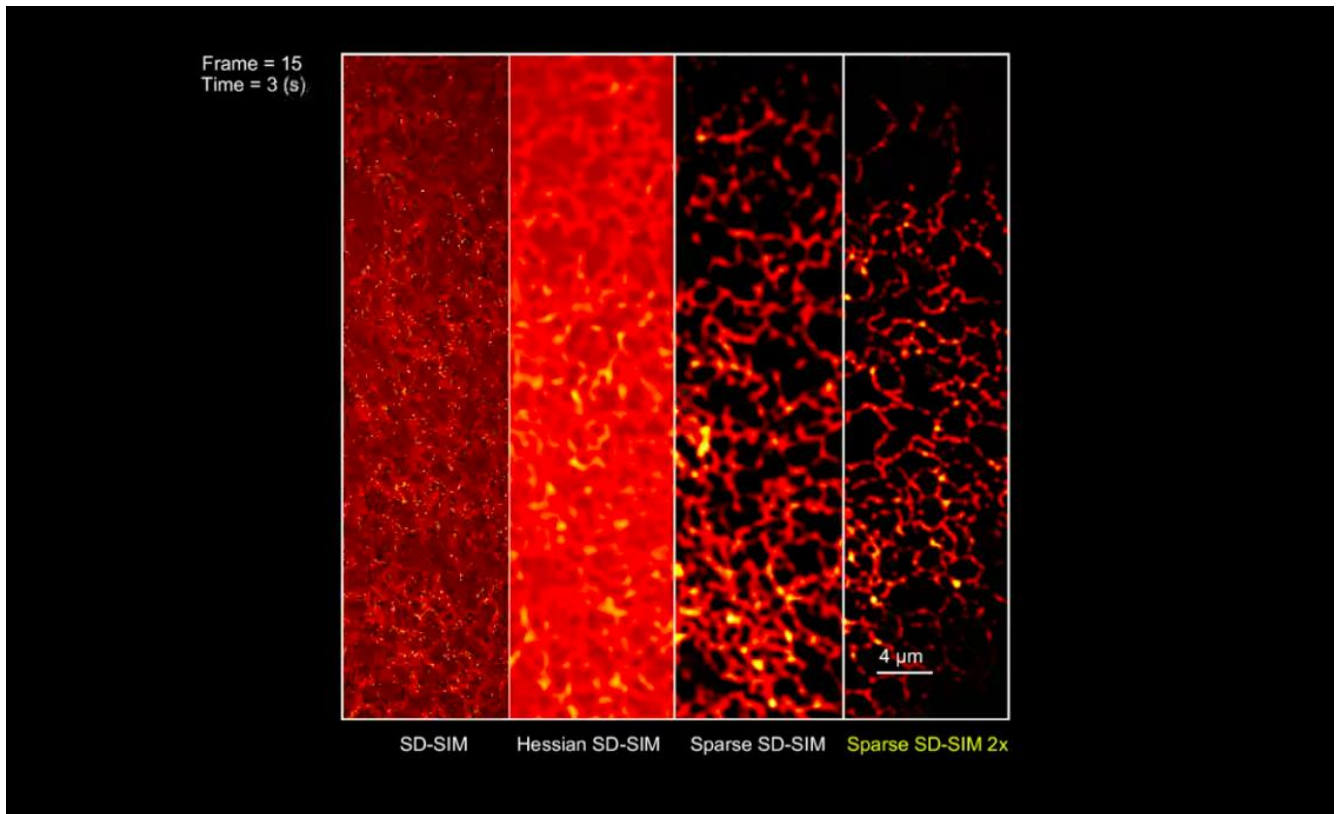
Supplementary Video 9 | Ring-shaped structures of clathrin-coated pits resolved by Sparse SD-SIM. Comparison of clathrin-coated pits (CCPs) in COS-7 cells expressing clathrin-EGFP light chains for 100 time points with a 0.2 s exposure time imaged by SD-SIM and Sparse SD-SIM, respectively (*c.f.*, **Fig. 5a**). Disintegration (part II of the video) and disappearance (part III) events of CCPs are resolved only by Sparse-SD-SIM.



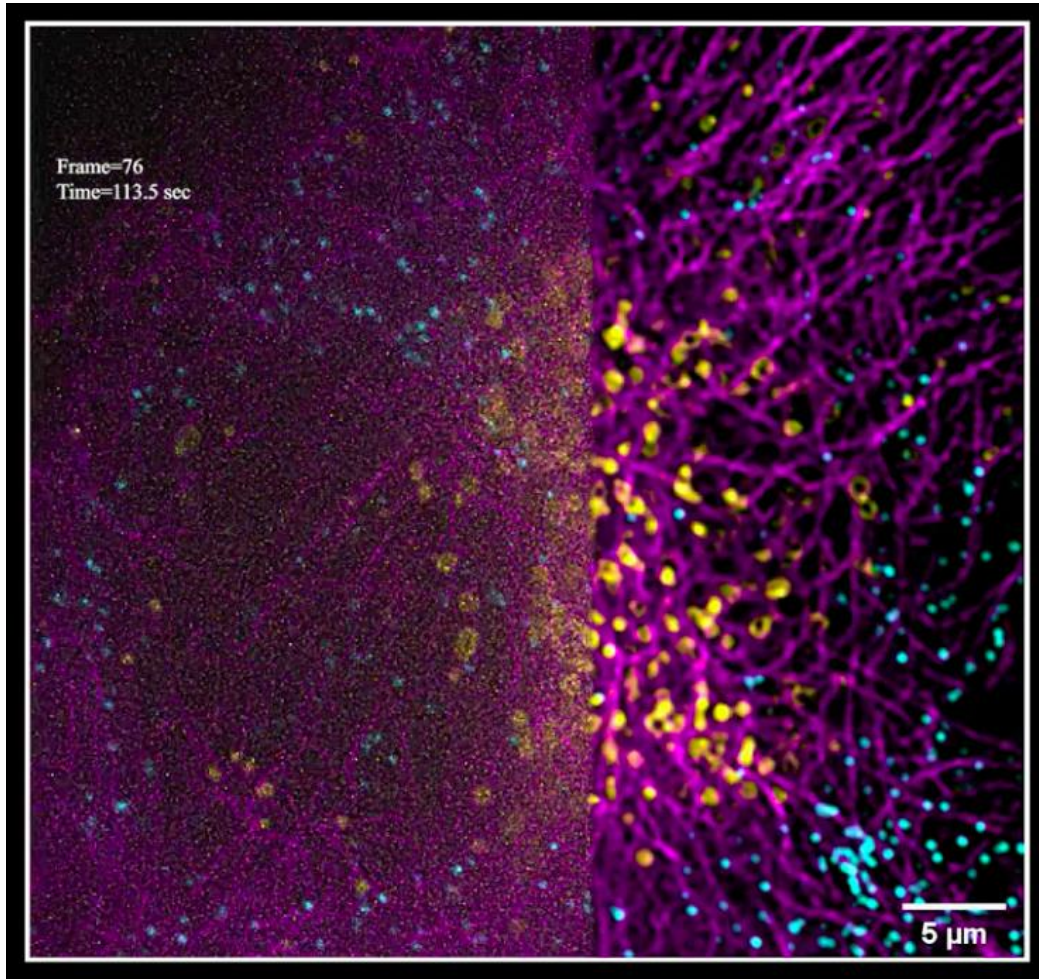
Supplementary Video 10 | Four-color sub-90-nm resolution live-cell imaging of lysosomes, mitochondria, nuclei, and microtubules. A COS-7 cell labeled with LAMP1-mCherry (yellow), Mito-Tracker (green), Hoechst (blue), and Tubulin-EGFP (magenta) imaged by SD-SIM and Sparse SD-SIM for 100 frames at a 2.3 s exposure time (*c.f.*, **Fig. 5h**). Part I shows the SD-SIM performance of the four organelles, which gradually transfer to Sparse SD-SIM in sequence. Then, the four organelles under SD-SIM and Sparse SD-SIM are compared simultaneously showing the striking contrast. Part II displays the dynamics of the magnified selection region in part I.



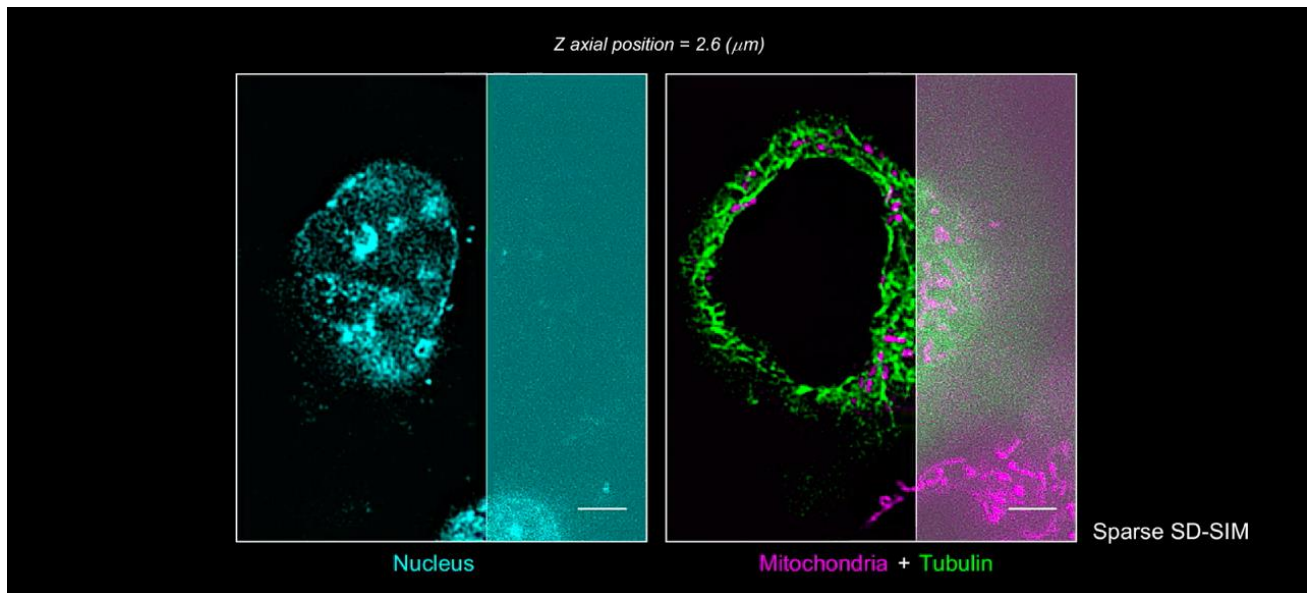
Supplementary Video 11 | Live-cell 3D image of outer mitochondrial membrane in a dividing cell with ~7 μm in depth. Mitochondrial expressions of the outer membrane marker TOM20-mCherry captured by SD-SIM and Sparse SD-SIM. (*c.f.*, **Fig. 5k**). Part I shows the color-coded projection of the mitochondrial membrane in the z direction as the depth increases gradually and the y-z orthoslices along the x-axis. Part II displays multi-angle views of the 3D rendering volume of the mitochondrial membrane showing the details recorded by SD-SIM (red) and Sparse SD-SIM (green).



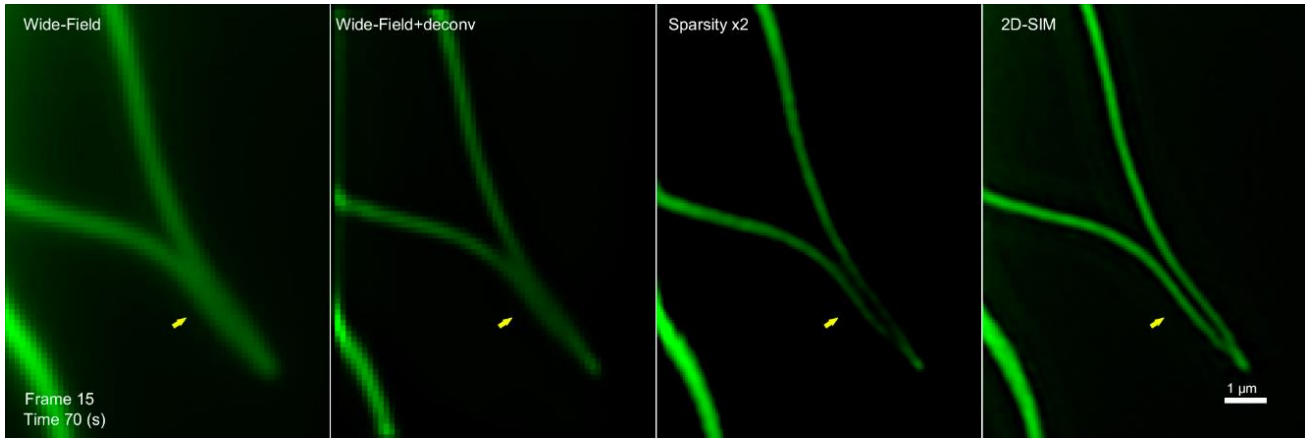
Supplementary Video 12 | ER tubular structures under Nyquist insufficient sampling paradigm resolved by Sparse SD-SIM 2 $\times$. Tubular ER in the periphery of a COS-7 cell labeled with Sec61 β -EGFP. Part I shows a comparison of the tubular ER by SD-SIM, Hessian SD-SIM, Sparse SD-SIM, and Sparse SD-SIM 2 $\times$ (*c.f.*, **Fig. 6a**). Part II compares the tubular ER by SD-SIM (left) with those by Sparse SD-SIM (right) 2 $\times$ for 500 time points at a 0.2 s acquisition time. A magnified view of the selected region in Part II is shown in Part III.



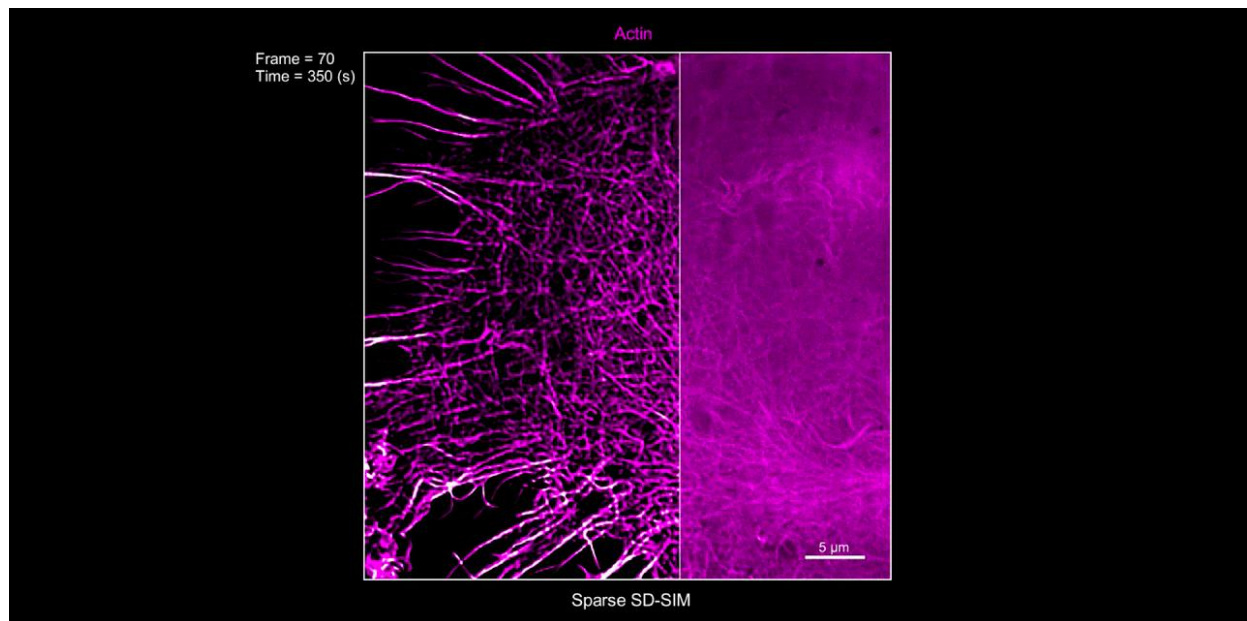
Supplementary Video 13 | Microtubules regulate lysosome-peroxisome membrane contacts in live cells captured by three-color Sparse SD-SIM 2× imaging. Part I: A time-lapse full-field view comparison of SD-SIM and Sparse SD-SIM 2×. Part II: The two zoomed-in examples show a comparison of the images by SD-SIM and Sparse SD-SIM 2× (*c.f.*, **Fig. 6e**). The first example shows a lysosome moving along the microtubule to make membrane contact the peroxisome for close to 2 mins; the deformation of the lysosome membrane during the contact period can be detected only by Sparse SD-SIM 2×. The second example shows a lysosome-peroxisome membrane in contact moving along microtubules together in a relatively dense microtubule region; the movement can be detected only by Sparse SD-SIM 2×.



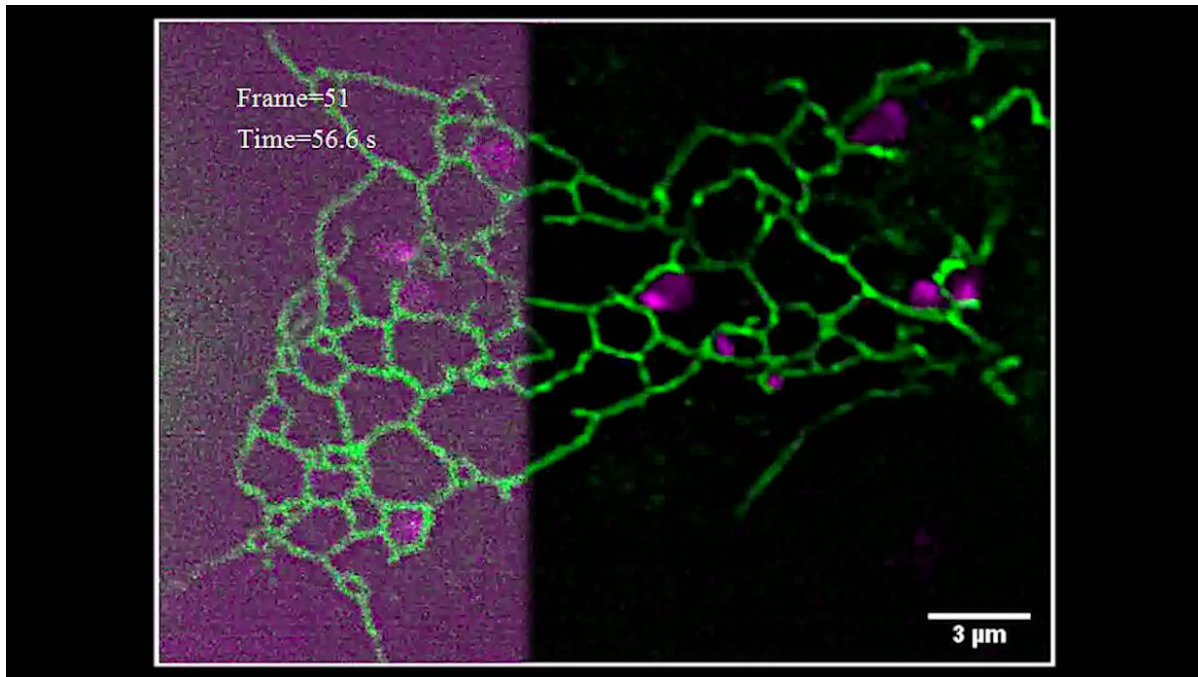
Supplementary Video 14 | Three-color 3D volume rendered views under Sparse SD-SIM 2 $\times$. Live-cell three-color 3D imaging of a COS-7 cell labeled with Tubulin-EGFP (green), Hoechst (cyan), and MitoTracker® Deep Red FM (magenta) under Sparse SD-SIM 2 $\times$ (*c.f.*, **Fig. 6h**). Part I shows x - y orthoslices gradually forming color-coded volumes of nuclei (left), mitochondria (middle), and microtubules (right). Part II displays three-color 3D volume rendered views of nuclei, mitochondria, and microtubules from various angles.



Supplementary Video 15 | Benchmarks of time-varying actin filaments under different approaches according to the ground truth 2D-SIM images. Two branched actin filaments were recorded by Wide-Field, Wide-Field + deconvolution, Sparse 2 $\times$, and 2D-SIM (*c.f.*, **Supplementary Fig. 31e**). Using 2D-SIM as the ground truth, Wide-Field with Sparse 2 $\times$ reveals fine structure such as two-branched actin filaments separated correctly.

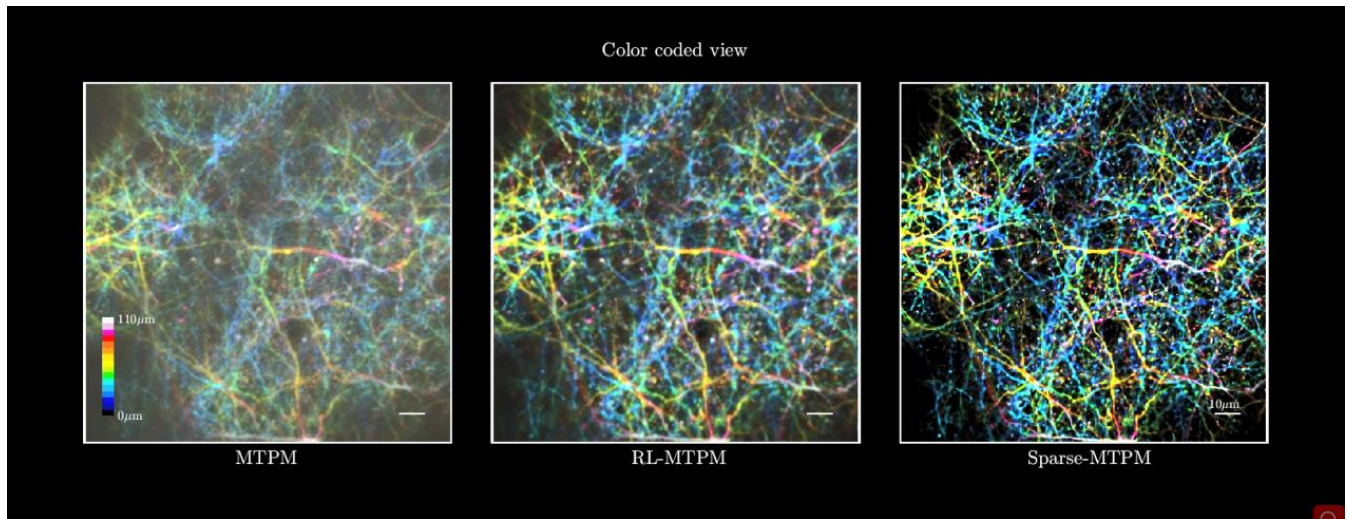


Supplementary Video 16 | Relative dynamics of both actin filaments and CCPs captured by dual-color Sparse SD-SIM. Actin filaments (labeled by LifeAct-EGFP) and CCPs (labeled by Clathrin-DsRed) in COS-7 cells for 200 time points at a 5 s interval (*c.f.*, **Extended Data Fig. 6**). A CCP stably docked at the intersection of two actin filaments that begins to disappear from the confocal plane as the neighboring filaments close up and meet (marked with a yellow circle).



Supplementary Video 17 | ER-lysosome contacts in live cells captured by dual-color Sparse SD-SIM.

Part I: The time-lapse full-field ER-lysosome (labeled with Sec61 β -EGFP and LysoTracker Deep Red) view comparison of SD-SIM and Sparse SD-SIM (*c.f.*, **Extended Data Fig. 7**). Part II: The zoom-in example shows typical lysosome-ER contact dynamics by SD-SIM and Sparse SD-SIM. The lysosome-ER contact dynamics can be clearly viewed under Sparse SD-SIM.



Supplementary Video 18 | Volumetric imaging of neuronal dendrites and spines by Sparse-MTPM.

Three-dimensional distributions of neuronal dendrites and spines within a volume of $190 \times 190 \times 110 \mu\text{m}^3$ from the brain of a Thy1-GFP transgenic mouse observed with MTPM, RL deconvoluted MTPM (RL-MTPM), and Sparse-MTPM (*c.f.*, **Extended Data Fig. 9**). Part I demonstrates x - y comparison of these three methods. Part II shows the x - y orthoslices gradually forming color-coded volumes of MTPM, RL deconvoluted MTPM (RL-MTPM), and Sparse-MTPM from left to right respectively. Part III displays 3D volume rendered views of neuronal dendrites and spines from various angles.

References

1. Hunt, B.R. Super-resolution of images: Algorithms, principles, performance. *International Journal of Imaging Systems and Technology* **6**, 297-304 (1995).
2. Richardson, W.H. Bayesian-based iterative method of image restoration. *Journal of The Optical Society of America A* **62**, 55-59 (1972).
3. Lucy, L.B. An iterative technique for the rectification of observed distributions. *The Astronomical Journal* **79**, 745 (1974).
4. Shepp, L.A. & Vardi, Y. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging* **1**, 113-122 (1982).
5. Dey, N. et al. Richardson–Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microscopy research and technique* **69**, 260-266 (2006).
6. Lucy, L.B. Resolution limits for deconvolved images. *The Astronomical Journal* **104**, 1260-1265 (1992).
7. Lucy, L.B. Statistical limits to super resolution. *Astronomy & Astrophysics* **261**, 706-710 (1992).
8. Wolter, H. On basic analogies and principal differences between optical and electronic information, Vol. 1. (Elsevier, 1961).
9. Goodman & W, J. Introduction to Fourier optics. (Roberts and Company Publishers, 2005).
10. Bertero, M. & De Mol, C. Super-Resolution by Data Inversion, Vol. 36. (Elsevier, 1996).
11. Lindberg, J. Mathematical concepts of optical superresolution. *Journal of Optics* **14**, 083001 (2012).
12. Cox, c.I. & Sheppard, C. Information capacity and resolution in an optical system. *Journal of The Optical Society of America A* **3**, 1152-1158 (1986).
13. Aspelmeier, T., Egner, A. & Munk, A. Modern Statistical Challenges in High-Resolution Fluorescence Microscopy. *Annual Review of Statistics and Its Application* **2**, 163–202 (2015).
14. Robbins, M. & Hadwen, B.J. The noise performance of electron multiplying charge-coupled devices. *IEEE Transactions on Electron Devices* **50**, 1227-1232 (2003).
15. Foi, A., Trimeche, M., Katkovnik, V. & Egiazarian, K. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data. *IEEE Transactions on Image Processing* **17**, 1737-1754 (2008).
16. Mortensen, K.I., Churchman, L.S., Spudich, J. & Flyvbjerg, H. Optimized localization-analysis for single-molecule tracking and super-resolution microscopy. *Nature Methods* **7**, 377 - 381 (2010).
17. Hirsch, M., Wareham, R.J., Martin-Fernandez, M., Hobson, M.P. & Rolfe, D. A Stochastic Model for Electron Multiplication Charge-Coupled Devices: From Theory to Practice. *PLoS ONE* **8**, e53671 (2013).
18. Liu, S. et al. sCMOS noise-correction algorithm for microscopy images. *Nature Methods* **14**, 760-761 (2017).
19. Born, M. & Wolf, E. Principles of optics, Edn. 7th. (Cambridge University Press, Cambridge; 1999).
20. Heintzmann, R. & Sheppard, C. The sampling limit in fluorescence microscopy. *Micron* **38**, 145-149 (2007).
21. Lohmann, A. et al. Space bandwidth product of optical signals and systems. *Journal of The Optical Society of America A* **13**, 470-473 (1996).
22. Betzig, E. et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642-1645 (2006).
23. Rust, M., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793-796 (2006).
24. Huang, X. et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. *Nature biotechnology* **36**, 451-459 (2018).
25. Thevathasan, J.V. et al. Nuclear pores as versatile reference standards for quantitative superresolution microscopy. *Nature Methods* **16**, 1045-1053 (2019).
26. Candes, E.J. & Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory* **52**, 5406-5425 (2006).
27. Szameit, A. et al. Sparsity-based single-shot subwavelength coherent diffractive imaging. *Nature materials* **11**, 455-

- 459 (2012).
28. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster STORM using compressed sensing. *Nature Methods* **9**, 721-723 (2012).
 29. Galloway, C., Le Ru, E. & Etchegoin, P. An iterative algorithm for background removal in spectroscopy by wavelet transforms. *Applied Spectroscopy* **63**, 1370-1376 (2009).
 30. Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129-136 (1982).
 31. Wang, Y., Yang, J., Yin, W. & Zhang, Y. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences* **1**, 248-272 (2008).
 32. Goldstein, T. & Osher, S. The split Bregman method for L1-regularized problems. *SIAM journal on imaging sciences* **2**, 323-343 (2009).
 33. Landweber, L. An iteration formula for Fredholm integral equations of the first kind. *American journal of mathematics* **73**, 615-624 (1951).
 34. Sutskever, I., Martens, J., Dahl, G. & Hinton, G. in International conference on machine learning 1139-1147 (2013).
 35. Veklerov, E. & Llacer, J. Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Transactions on Medical Imaging* **6**, 313-319 (1987).
 36. Biggs, D.S. & Andrews, M. Acceleration of iterative image restoration algorithms. *Applied optics* **36**, 1766-1775 (1997).
 37. Stein, S.C., Huss, A., Hähnel, D., Gregor, I. & Enderlein, J. Fourier interpolation stochastic optical fluctuation imaging. *Optics Express* **23**, 16154-16163 (2015).
 38. Peng, T. et al. A BaSiC tool for background and shading correction of optical microscopy images. *Nature Communications* **8**, 14836 (2017).
 39. Theer, P., Mongis, C. & Knop, M. PSFj: know your fluorescence microscope. *Nature Methods* **11**, 981-982 (2014).
 40. Beck, A. & Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sciences* **2**, 183-202 (2009).
 41. Gazit, S., Szameit, A., Eldar, Y.C. & Segev, M. Super-resolution and reconstruction of sparse sub-wavelength images. *Optics Express* **17**, 23920-23946 (2009).
 42. Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600-612 (2004).
 43. Luisier, F., Vonesch, C., Blu, T. & Unser, M. Fast interscale wavelet denoising of Poisson-corrupted images. *Signal Processing* **90**, 415-427 (2010).