nature methods

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41592-020-0816-x

In the format provided by the authors and unedited.

Three-dimensional nanoscopy of whole cells and tissues with in situ point spread function retrieval

Fan Xu[1,10], Donghan Ma[1,10], Kathryn P. MacPherson[2], Sheng Liu[1], Ye Bu[1], Yu Wang[3,4], Yu Tang[5,6], Cheng Bi[1], Tim Kwok[7], Alexander A. Chubykin[5,6], Peng Yin[3,4], Sarah Calve[1]✉, Gary E. Landreth[2,8]✉ and Fang Huang[1,6,9]✉

[1]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA. [2]Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN, USA. [3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [5]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. [6]Purdue Institute for Integrative Neuroscience, Purdue University, West Lafayette, IN, USA. [7]Birck Nanotechnology Center, Purdue University, West Lafayette, IN, USA. [8]Stark Neurosciences Research Institute, Indiana University School of Medicine, Indianapolis, IN, USA. [9]Purdue Institute of Inflammation, Immunology and Infectious Disease, Purdue University, West Lafayette, IN, USA. [10]These authors contributed equally: Fan Xu, Donghan Ma. ✉e-mail: scalve@purdue.edu; glandret@iu.edu; fanghuang@purdue.edu

Supplementary Materials for

Three dimensional nanoscopy of whole cells and tissues with *in situ* point spread function retrieval

Fan Xu[†], Donghan Ma[†], Kathryn P. MacPherson, Sheng Liu, Ye Bu, Yu Wang, Yu Tang, Cheng Bi, Tim Kwok, Alexander A. Chubykin, Peng Yin, Sarah Calve[*], Gary E. Landreth[*], and Fang Huang[*]

[†]These authors contributed equally.

*Correspondence to: fanghuang@purdue.edu (F.H.), scalve@purdue.edu (S.C.), glandret@iu.edu (G.E.L.)

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Supplementary Video 1: Performance demonstration of INSPR on simulated single-molecule emission patterns in biplane setup.

Supplementary Video 2: Performance demonstration of INSPR on simulated single-molecule emission patterns in astigmatism-based setup.

Supplementary Video 3: INSPR estimates wavefront distortions induced by the deformable mirror from single-molecule datasets of TOM20 immune-labeled with DNA-PAINT in fixed COS-7 cells.

Supplementary Video 4: INSPR reconstruction of immunofluorescence-labeled TOM20 in fixed COS-7 cells.

Supplementary Video 5: INSPR reconstruction of immunofluorescence-labeled Nup98 on an entire nucleus in fixed COS-7 cells.

Supplementary Video 6: INSPR reconstruction of immunofluorescence-labeled amyloid β plaque with low-density fibrils in a 30-μm-thick brain section from a 5XFAD mouse.

Supplementary Video 7: INSPR reconstruction of immunofluorescence-labeled amyloid β plaque with high-density fibrils in a 30-μm-thick brain section from a 5XFAD mouse.

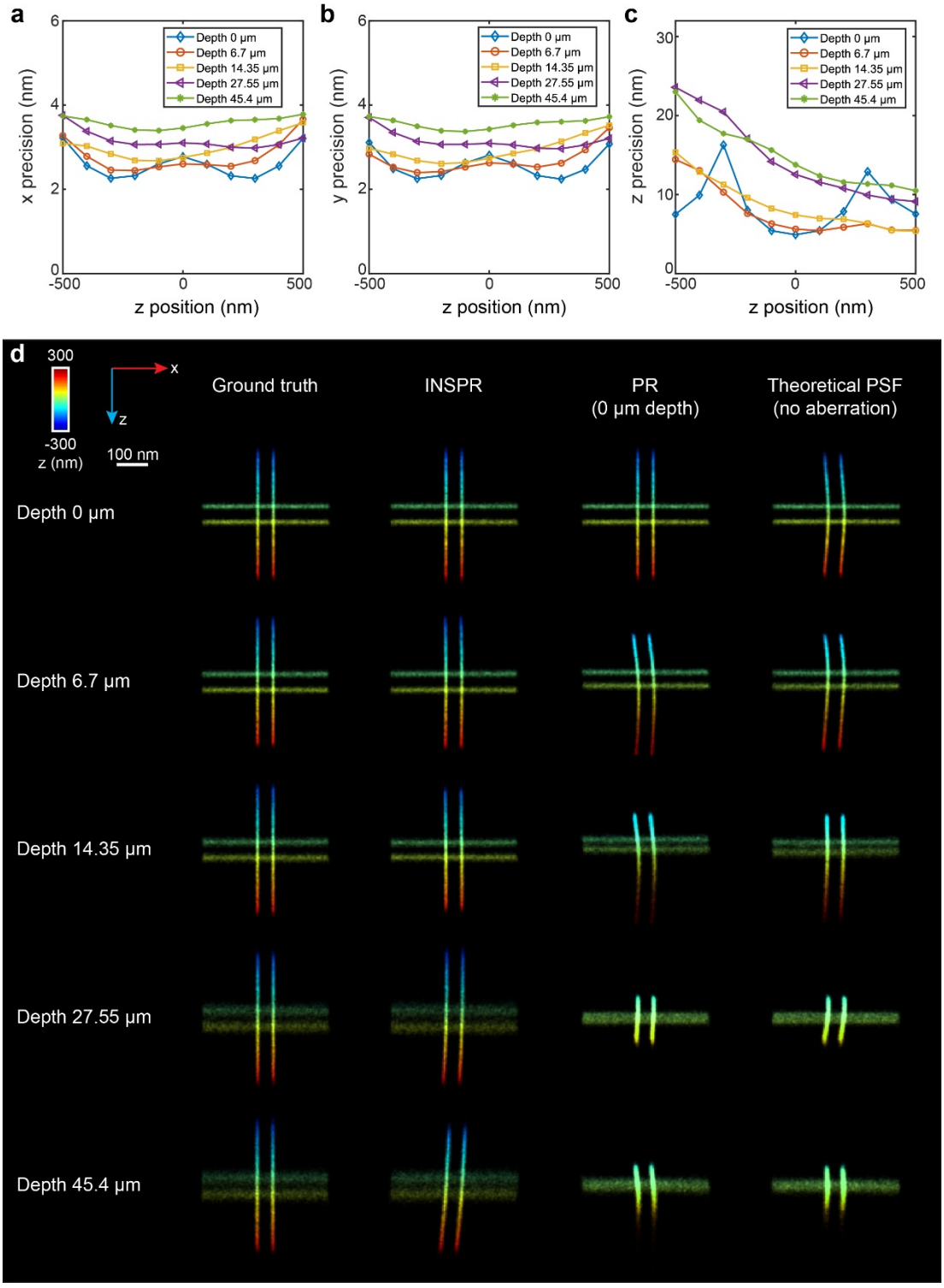
Supplementary Video 8: INSPR reconstruction of immunofluorescence-labeled ChR2-EYFP on dendrites in a 50-μm-thick brain section from mouse primary visual cortex.

Supplementary Video 9: INSPR reconstruction of immunofluorescence-labeled elastic fibers in developing cartilage from embryonic mouse forelimbs.

Supplementary Software 1: INSPR software package and user manual for biplane setup.

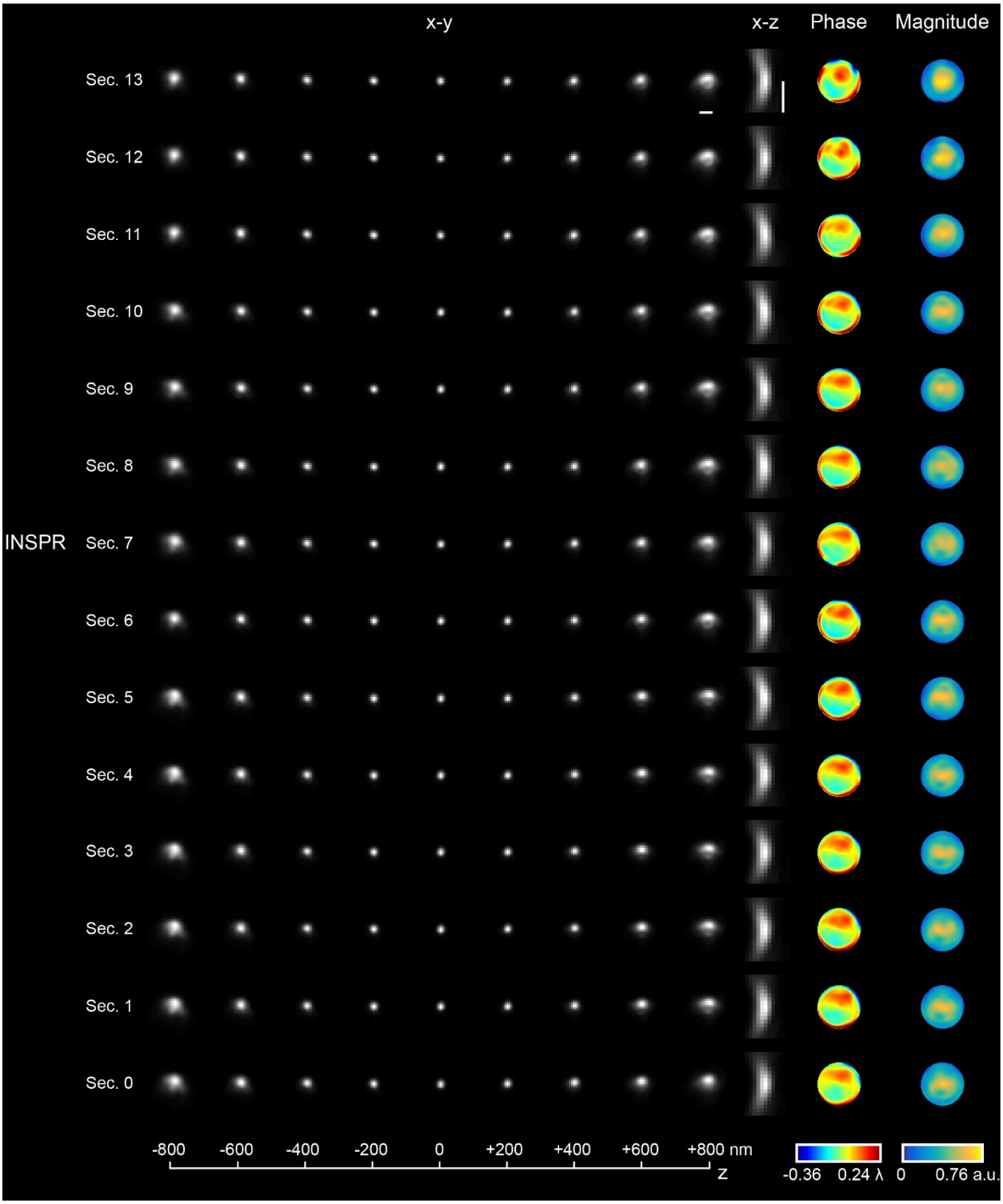
Supplementary Software 2: INSPR software package and user manual for astigmatism-based setup.

Table of Contents



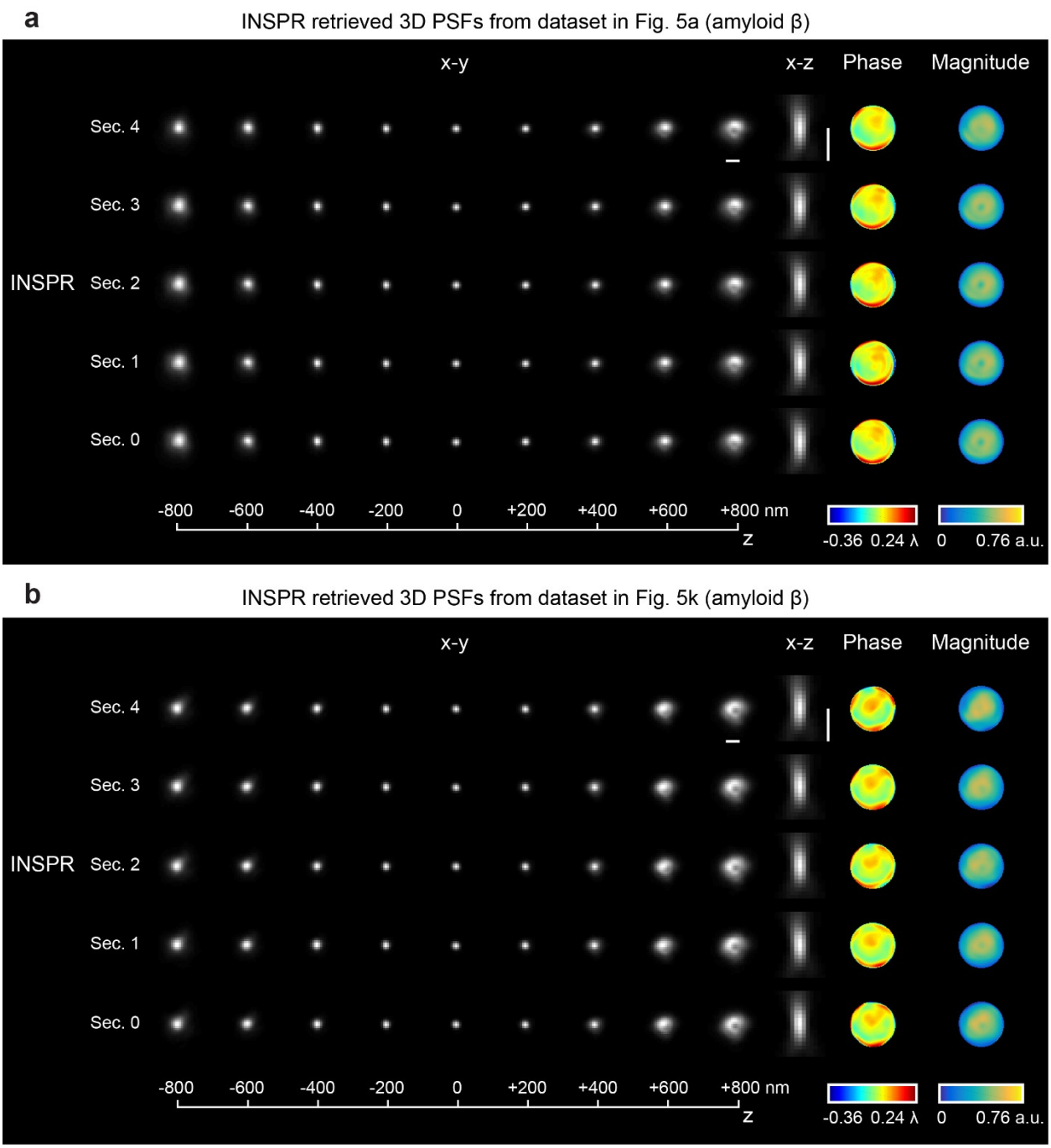
Supplementary Figure 1. Localization precisions and reconstructions of simulated, axially-distributed, double-line structures at different imaging depths using INSPR and *in vitro* methods.

(a–c) Localization precisions in the x, y, and z dimensions for different imaging depths (0, 6.7, 14.35, 27.55, and 45.4 μm). **(d)** x-z views of the reconstructed double-line structures using 3D PSF models obtained from the ground truth, INSPR, phase retrieval from PSFs at the depth of 0 μm (PR), and theoretical PSF (no aberration, constant pupil).



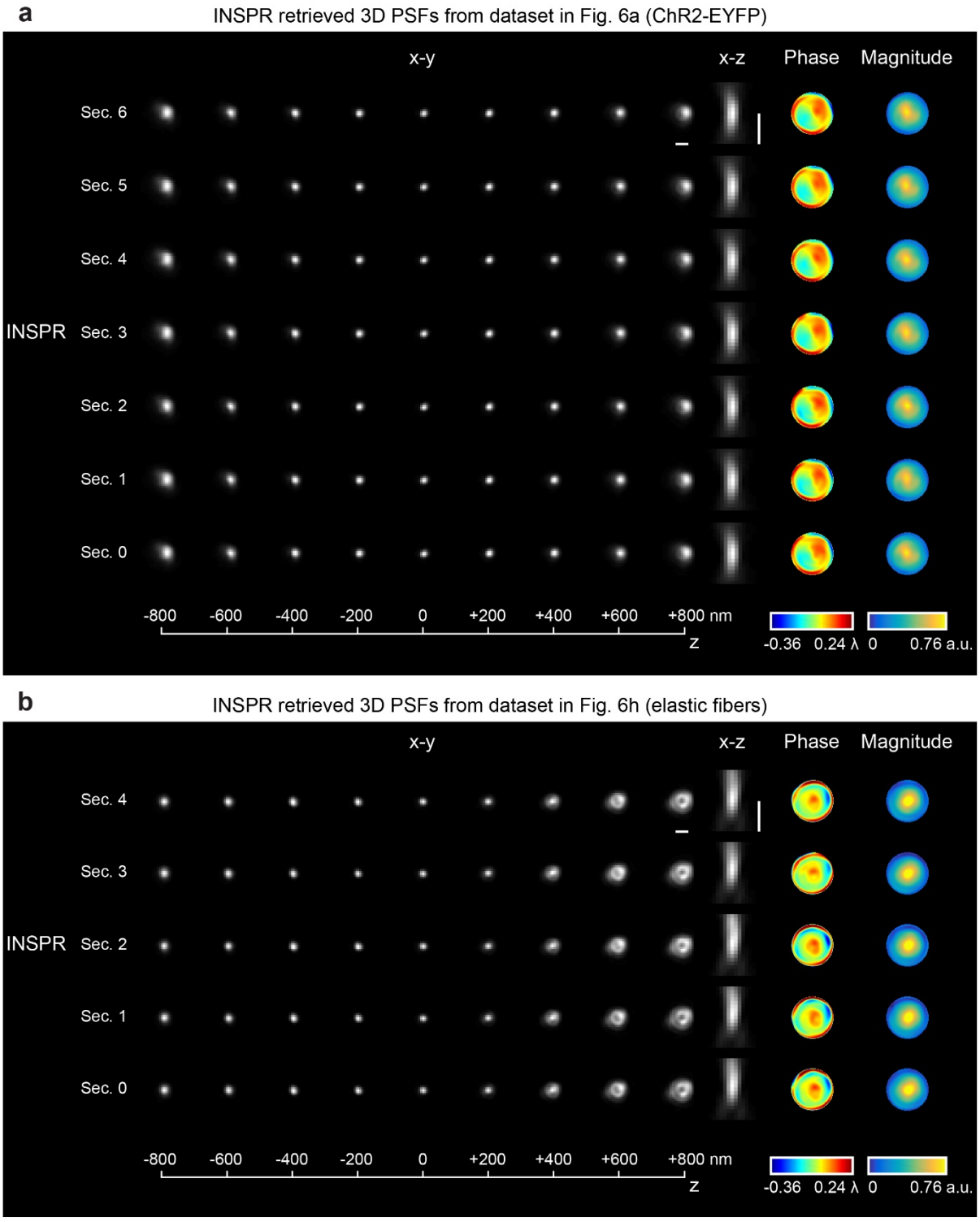
Supplementary Figure 2. 3D PSFs and their corresponding pupils retrieved by INSPR from the dataset of immunofluorescence-labeled Nup98 on the entire nuclear envelope in COS-7 cells in Fig. 4e.

Scale bar: 1 μm. Sec.: optical section.



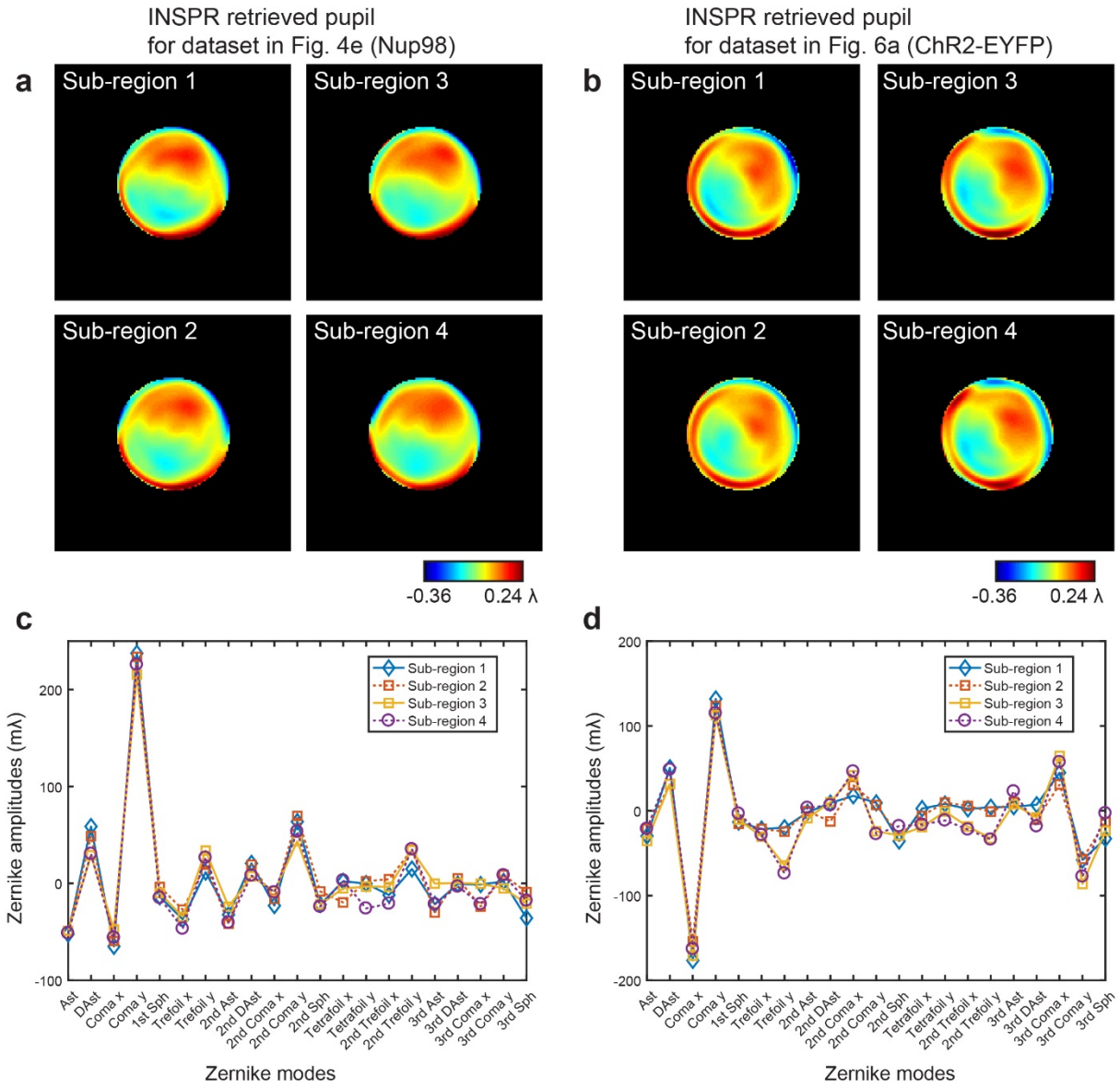
Supplementary Figure 3. 3D PSFs and their corresponding pupils retrieved by INSPR from the datasets of immunofluorescence-labeled amyloid β plaques in mouse brains in Fig. 5.

(a,b) x-y and x-z views of the 3D PSFs retrieved by INSPR in different optical sections, as well as the phase and magnitude of the corresponding pupils, from the datasets of immunofluorescence-labeled amyloid β plaques with low-density (a) and high-density (b) fibrils in mouse brains. Scale bar: 1 μm. Sec.: optical section.



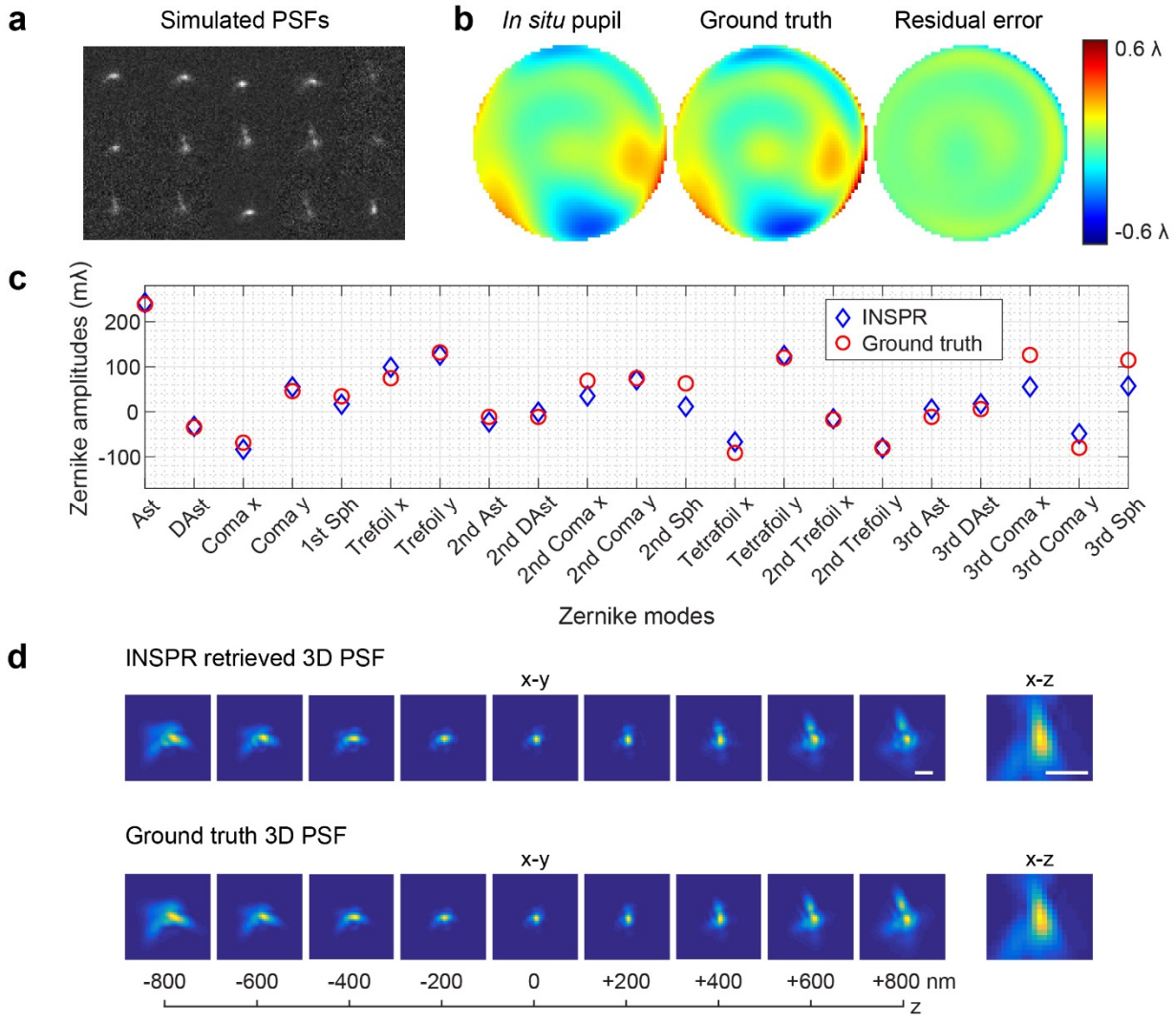
Supplementary Figure 4. 3D PSFs and their corresponding pupils retrieved by INSPR from the datasets of immunofluorescence-labeled ChR2-EYFP on dendrites and elastic fibers in developing cartilage in Fig. 6.

(a,b) x-y and x-z views of the 3D PSFs retrieved by INSPR in different optical sections, as well as the phase and magnitude of the corresponding pupils, from the datasets of immunofluorescence-labeled ChR2-EYFP on dendrites (a) and elastic fibers in developing cartilage (b). Scale bar: 1 μm. Sec.: optical section.



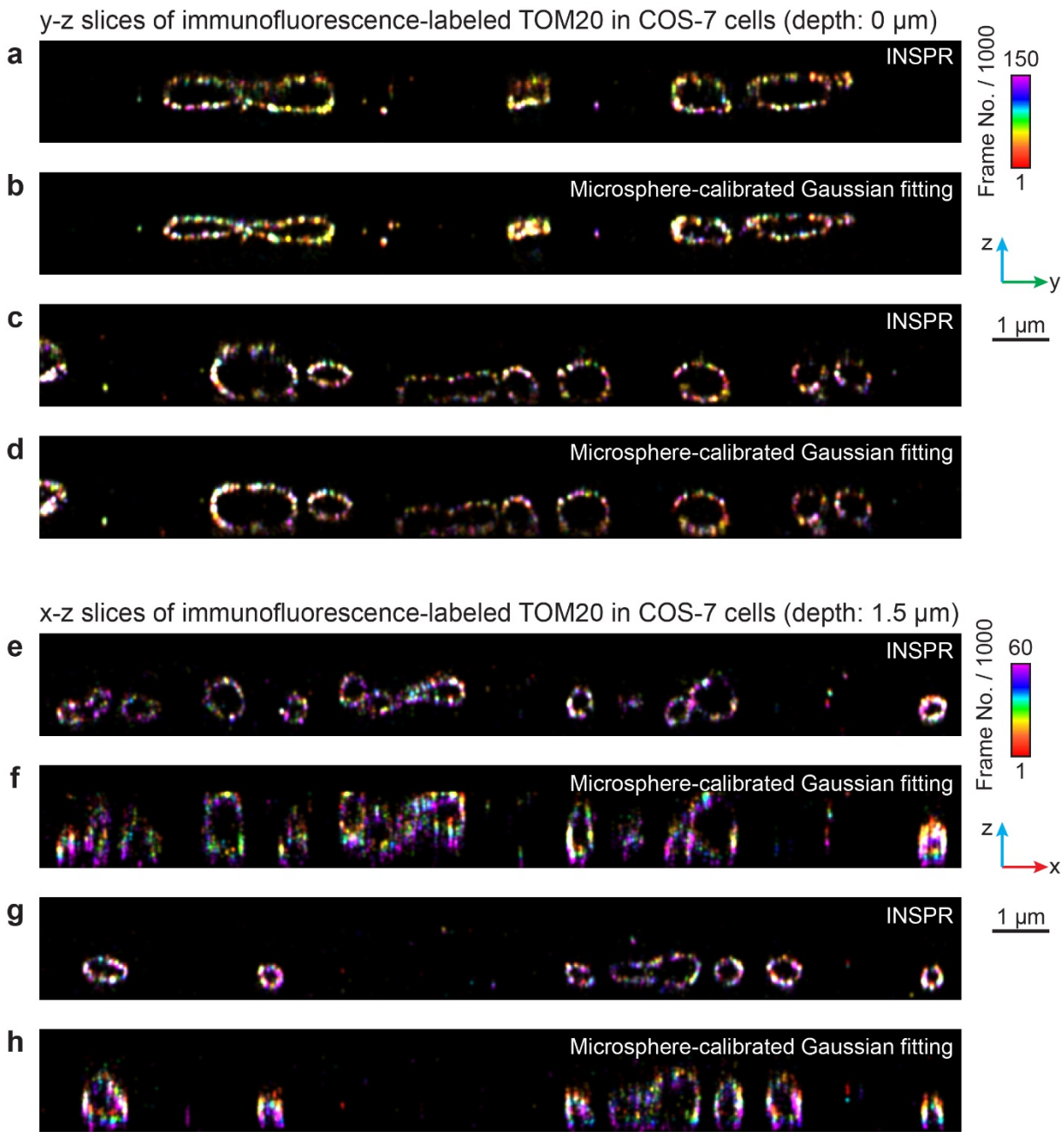
Supplementary Figure 5. Field-of-view dependent pupils retrieved by INSPR from the datasets of immunofluorescence-labeled Nup98 in COS-7 cells in Fig. 4e and ChR2-EYFP on dendrites in Fig. 6a.

(a,b) Phase of pupils retrieved by INSPR in four sub-regions of 30 μm × 30 μm imaging area from the datasets of immunofluorescence-labeled Nup98 in COS-7 cells (a) and ChR2-EYFP on dendrites (b). 6000 frames of blinking data from the first optical section are used for model generation in each dataset. **(c,d)** Amplitudes of 21 Zernike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration) decomposed from the INSPR retrieved pupils in (a,b). Results shown are representative of the seven datasets shown in the main text.



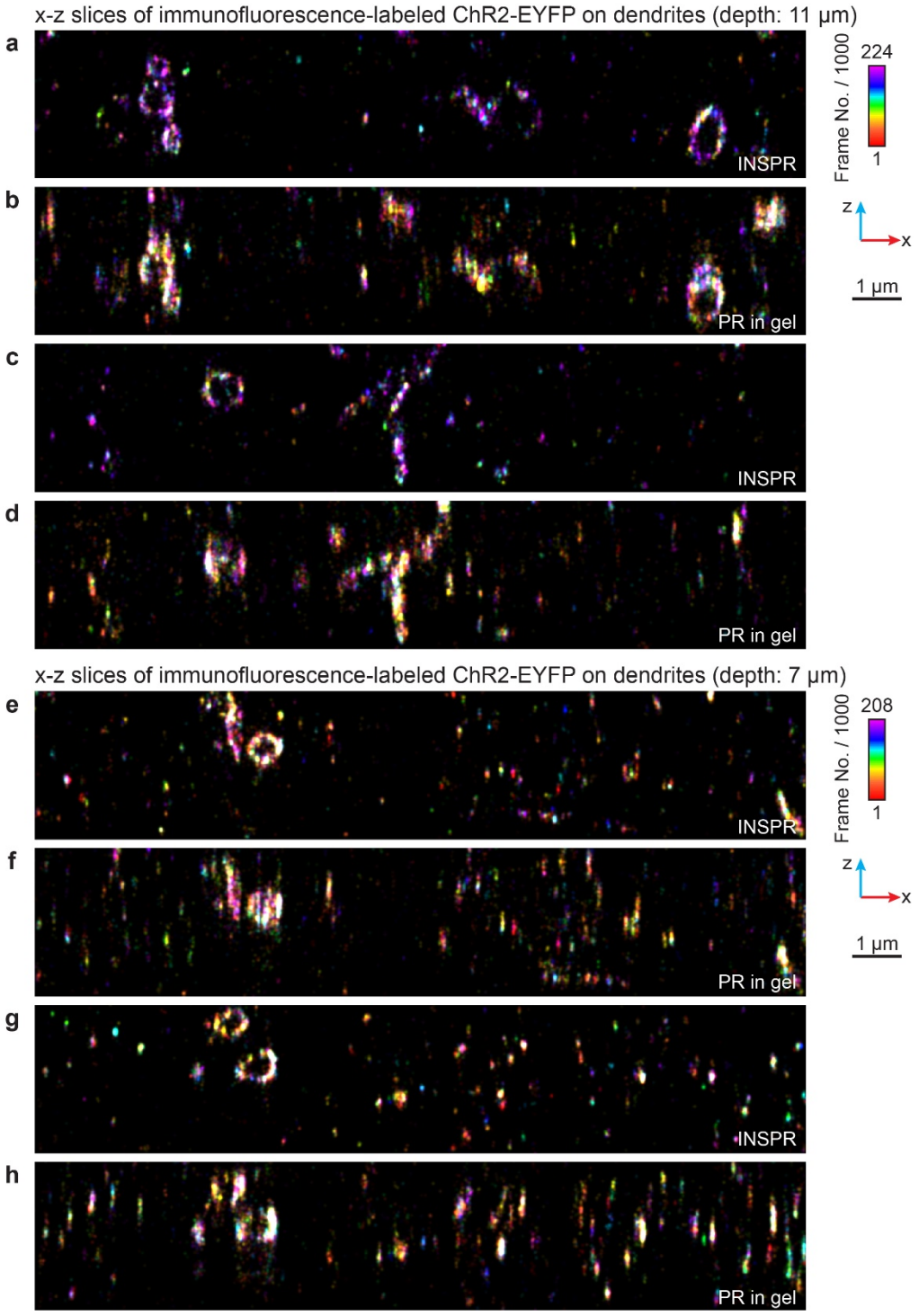
Supplementary Figure 6. Performance quantification of INSPR in astigmatism-based setup.

(a) Simulated astigmatism-based single-molecule emission patterns located randomly over an axial range from –800 to +800 nm with a known wavefront distortion. The amplitude of vertical astigmatism aberration is set to +1.5 (unit, λ/2π) as prior knowledge. **(b)** Phase of the *in situ* pupil retrieved by INSPR (left), the ground truth pupil (middle), and the residual error (right). The root-mean-square error (RMSE) is 40.2 mλ. **(c)** Amplitudes of 21 Zernike modes decomposed from the INSPR retrieved pupil (blue diamonds) compared with the ground truth (red circles). The RMSE is 27.4 mλ for the total 21 modes. **(d)** x-y and x-z views of the INSPR retrieved 3D PSF (top row) and the ground truth PSF (bottom row). Scale bar: 1 μm. Results shown are representative of 30 trials, whose animated demonstration is shown in **Supplementary Video 2**.

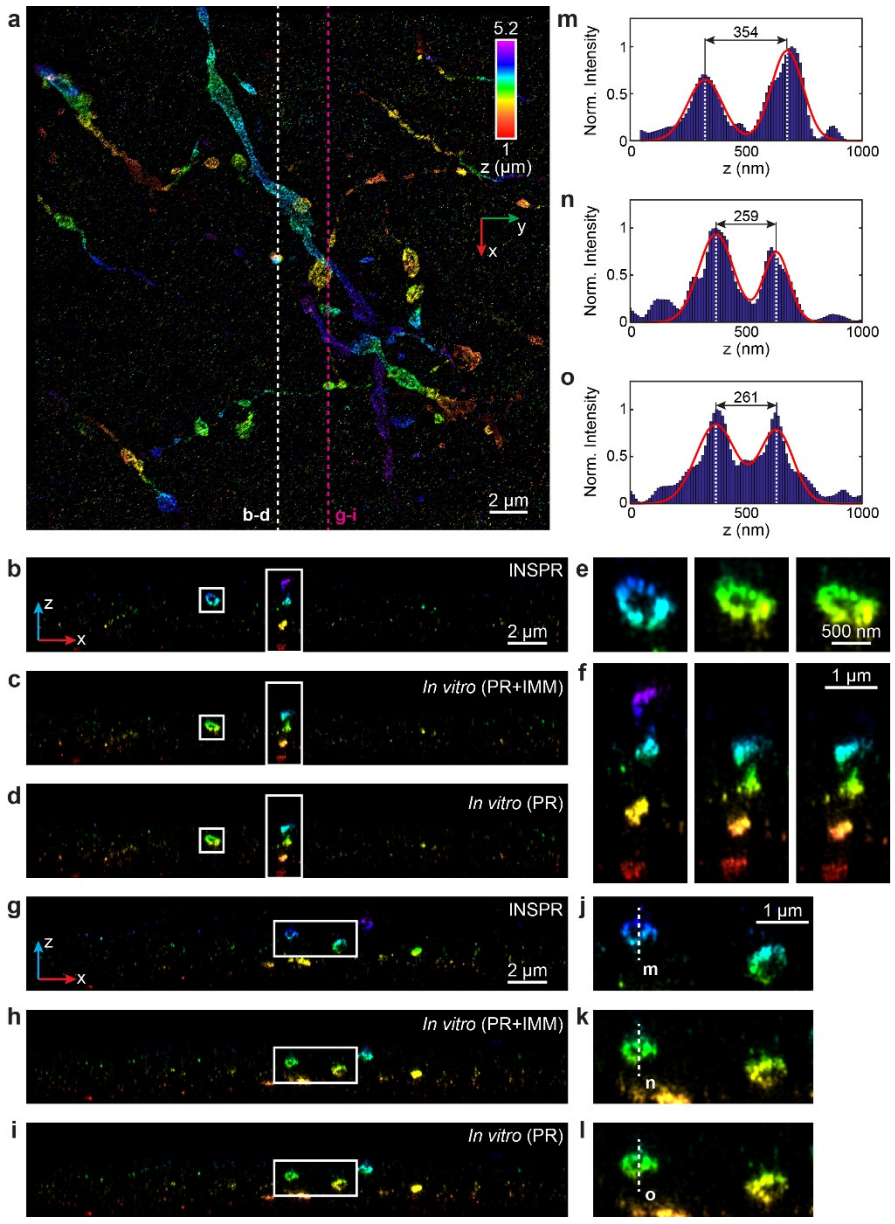


Supplementary Figure 7. Frame number color-coded cross sections of immunofluorescence-labeled TOM20 in COS-7 cells reconstructed using INSPR and microsphere-calibrated Gaussian fitting in astigmatism-based setup.

(a–d) y-z slices of the reconstructed mitochondrial network on the bottom coverslip using INSPR (a,c) and microsphere-calibrated Gaussian fitting (b,d). **(e–h)** x-z slices of the reconstructed mitochondrial network with a depth of 1.5 μm using INSPR (e,g) and microsphere-calibrated Gaussian fitting (f,h). The integration width of the slices in (a–h) in the third dimension is 200 nm. The localized molecules are color coded by their frame numbers in which the molecules are detected.

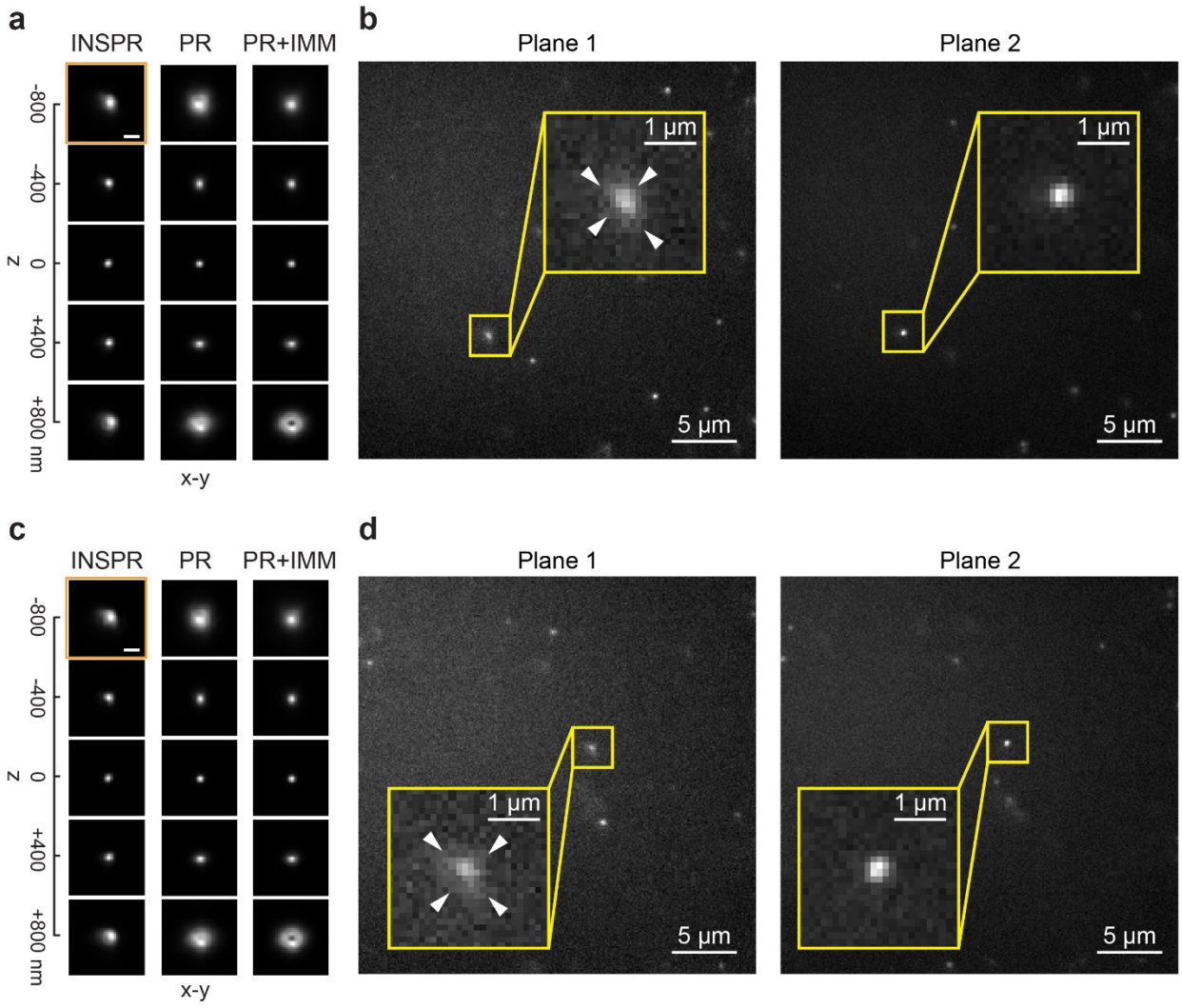


Supplementary Figure 8. Frame number color-coded cross sections of immunofluorescence-labeled ChR2-EYFP on dendrites reconstructed using INSPR and phase retrieval method based on beads embedded in agarose gel in biplane setup. (a–d) x-z slices of the reconstructed 11-µm-deep dendrites using INSPR (a,c) and phase retrieval method based on beads embedded in agarose gel (hereafter referred as 'PR in gel', b,d). **(e–h)** x-z slices of the reconstructed 7-µm-deep dendrites using INSPR (e,g) and PR in gel (f,h). The integration width of the x-z slices in (a–h) in the y direction is 200 nm. The localized molecules are color coded by their frame numbers in which the molecules are detected.



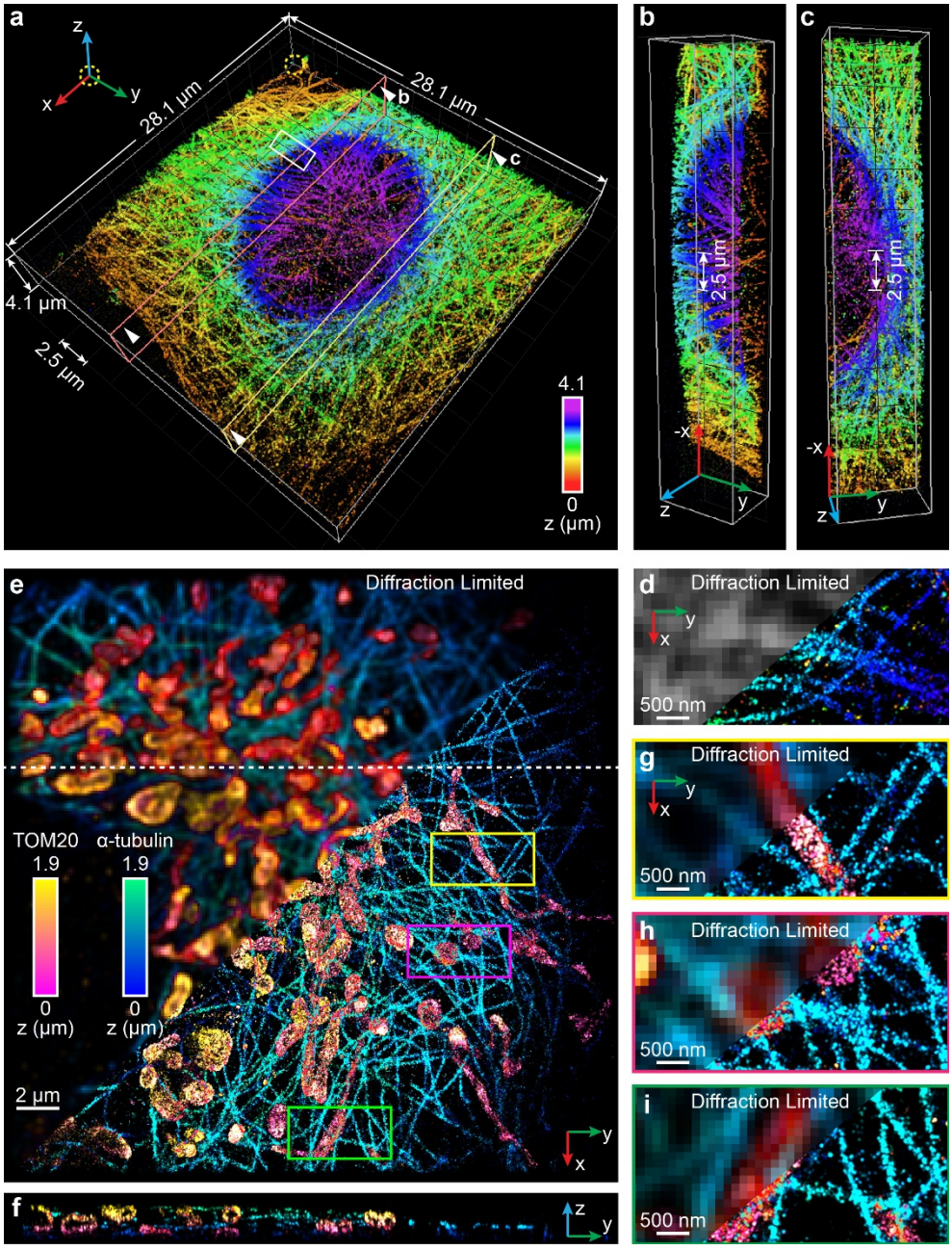
Supplementary Figure 9. 3D super-resolution reconstructions of immunofluorescence-labeled ChR2-EYFP on dendrites using INSPR and *in vitro* methods in biplane setup (depth: 1 – 5.2 μm).

(a) x-y overview of the super-resolution volume of immunofluorescence-labeled ChR2-EYFP on dendrites resolved by INSPR, with a depth of 1 μm from the coverslip. **(b–d)** x-z slices along the white dashed line in (a), reconstructed using INSPR (b), phase retrieval method based on beads on the coverslip with theoretical index mismatch model (hereafter referred as 'PR+IMM', c), and phase retrieval method based on beads on the coverslip (hereafter referred as 'PR', d). **(e)** Zoomed in x-z views of the areas (from left to right) as indicated by the left white boxed regions in (b–d). **(f)** Zoomed in x-z views of the areas (from left to right) as indicated by the right white boxed regions in (b–d). **(g–i)** x-z slices along the magenta dashed line in (a), reconstructed using INSPR (g), PR+IMM (h), and PR (i). **(j–l)** Zoomed in x-z views of the areas as indicated by the white boxed regions in (g–i). **(m–o)** Intensity profiles along the white dashed lines in (j–l), showing the difference in the axial width of the selected contour is ~26% for both PR+IMM and PR, as compared to INSPR. The integration width of the x-z slices in (b–l) in the y direction is 200 nm. The dataset shown is representative of five datasets of dendrites with a depth of ~2 μm. Norm.: normalized.



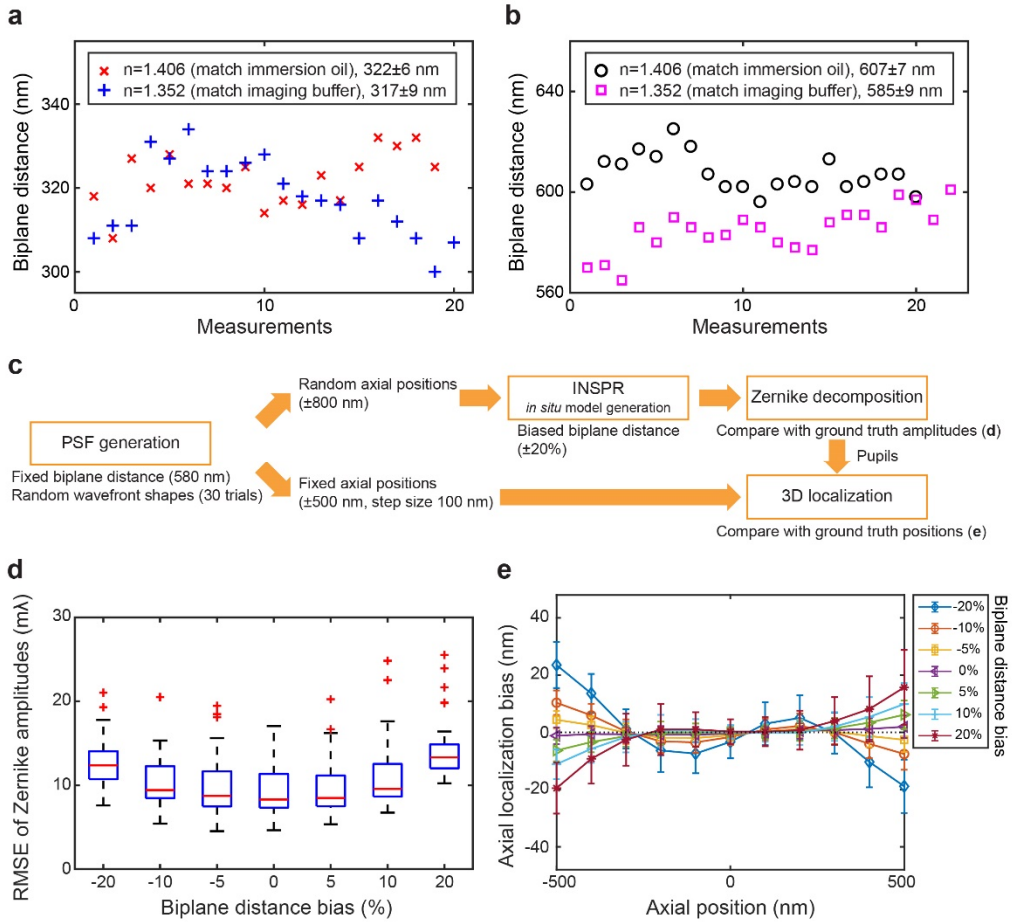
Supplementary Figure 10. Comparison of PSFs retrieved using INSPR and *in vitro* methods from the datasets of immunofluorescence-labeled ChR2-EYFP on dendrites in Fig. 6a and Supplementary Fig. 9.

(a) x-y view of the 3D PSF retrieved using INSPR (left), PR (middle), and PR+IMM (right) from the dataset of immunofluorescence-labeled ChR2-EYFP on dendrites (depth: 2 – 6.2 µm) in **Fig. 6a**. **(b)** Typical frame from biplane blinking data in **Fig. 6a**. The yellow boxed areas show a pair of detected emission patterns. **(c)** x-y view of 3D PSF retrieved using INSPR (left), PR (middle), and PR+IMM (right) from the dataset of immunofluorescence-labeled ChR2-EYFP on dendrites (depth: 1 – 5.2 µm) in **Supplementary Fig. 9**. **(d)** Typical frame from biplane blinking data in **Supplementary Fig. 9**. The yellow boxed areas show a pair of detected emission patterns. Scale bar in (a,c): 1 µm.



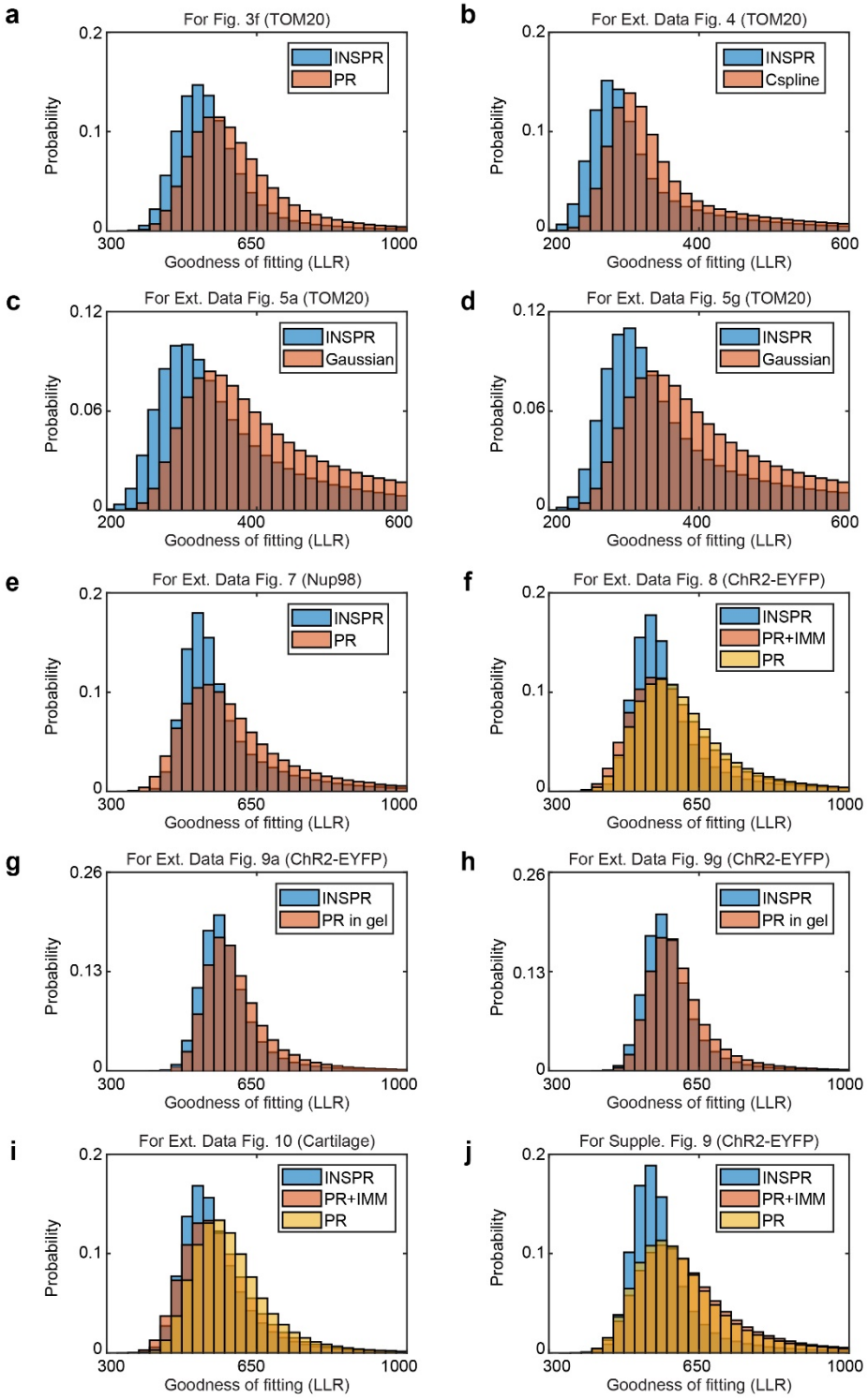
Supplementary Figure 11. INSPR reconstructions of microtubules in collagen embedded 3D-cultured BS-C-1 cells and TOM20 co-stained with α-tubulin using Exchange-PAINT in 2D cultured COS-7 cells in biplane setup.

(a) 3D overview of super-resolved microtubules in collagen embedded 3D cultured BS-C-1 cells with a total thickness of 4.1 µm. **(b,c)** Cross sections of the super-resolution reconstruction along the orange (b) and yellow (c) planes in (a). **(d)** Enlarged view of the area as indicated by the white boxed region in (a), showing fine microtubules (right) and its counterpart under the diffraction limit (left). **(e,f)** Reconstructions of mitochondria (TOM20) and microtubules (α-tubulin) in COS-7 cells immune-labeled with DNA-PAINT. An x-y overview (e) and y-z slice (f) along the white dashed line in (e) are shown. The integration width of the y-z slice in the x direction is 200 nm. **(g–i)** Enlarged views of the areas as indicated by the yellow (g), magenta (h), and green (i) boxed regions in (e), showing interactions between microtubules and mitochondria (right) and their counterparts under the diffraction limit (left). The datasets shown are representative of two datasets of microtubules in collagen and two datasets of Exchange-PAINT.



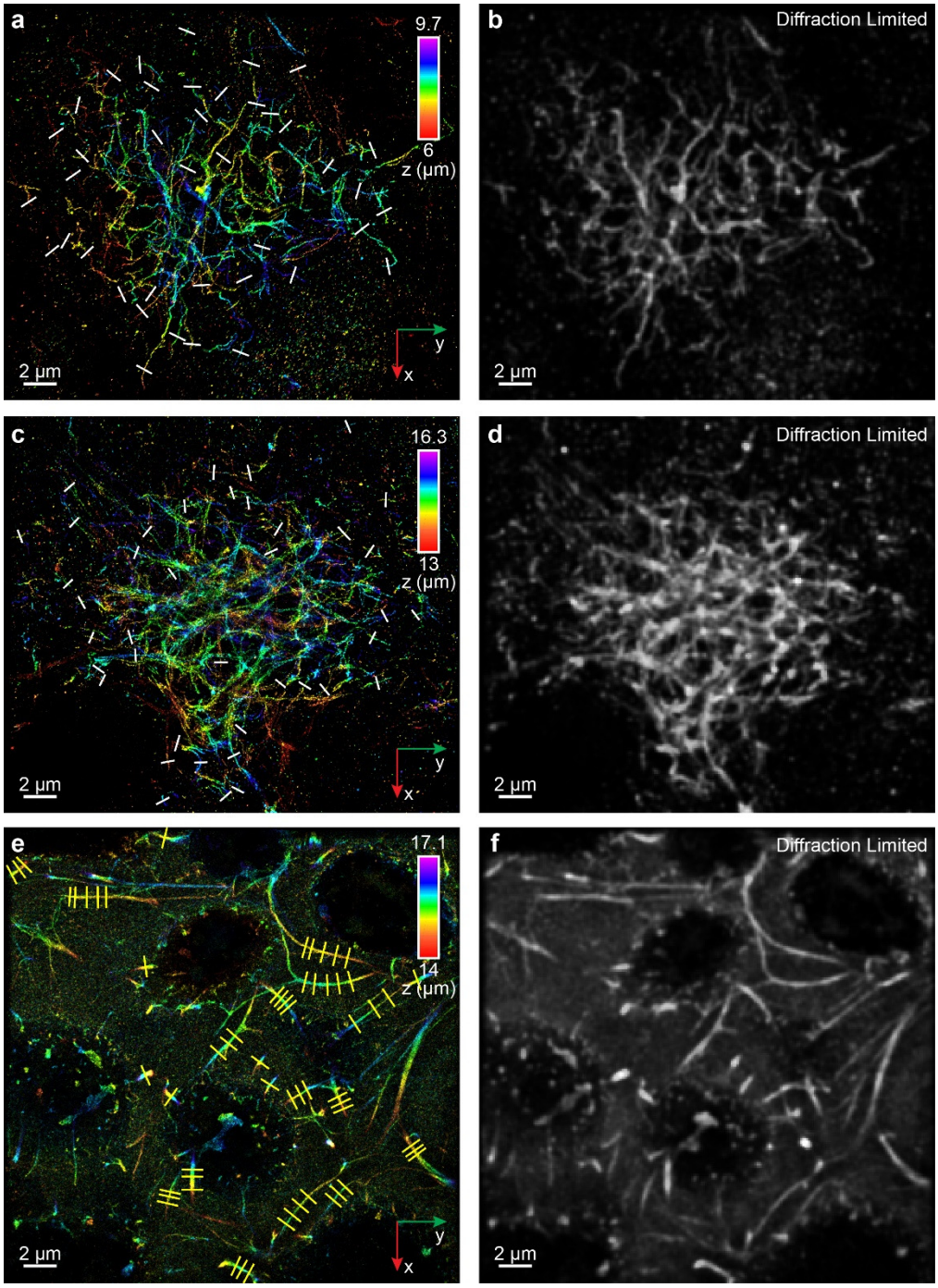
Supplementary Figure 12. Biplane distance measurement and its impact on INSPR.

(a) Distribution of biplane distance measured from beads in refractive index matched medium (mean±s.d.: 322±6 nm, 19 measurements, red crosses) and beads in refractive index mismatched medium (317±9 nm, 20 measurements, blue plus signs). **(b)** Distribution of biplane distance measured from beads in refractive index matched medium (607±7 nm, 20 measurements, black circles) and beads in refractive index mismatched medium (585±9 nm, 22 measurements, magenta squares). **(c)** Detailed process to estimate the impact of possible biplane distance bias. The ground truth biplane distance is set to 580 nm, and 30 sets of wavefront shapes are generated, consisting of 21 Zernike modes with their amplitudes randomly sampled from –1 to +1 (unit, $\lambda/2\pi$). These random wavefront shapes are used to generate PSFs with random axial positions from –800 to +800 nm. Subsequently, INSPR is used to retrieve pupils from these PSFs with a biased input of biplane distance (from –20% to 20%) mimicking a biplane distance measurement error. The decomposed Zernike amplitudes from the retrieved pupil are compared with the ground truth amplitudes, as shown in (d). Simulated single-molecule emission patterns are then generated using ground truth wavefront distortions and axial positions ranging from –500 to +500 nm with a step size of 100 nm. The retrieved pupils are used to localize these emission patterns and their localization biases caused by the inaccurate biplane distance inputs are shown in (e). **(d)** RMSE between the decomposed Zernike amplitudes of the INSPR retrieved pupils using an inaccurate biplane distance input and the ground truth amplitudes (30 trials for each bias). On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the plus signs. Points are considered outliers if they are greater than $q_3 + 1.5 \times (q_3 - q_1)$ or less than $q_3 - 1.5 \times (q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles of the data, respectively. **(e)** Axial localization biases at different axial positions introduced by assuming biased inputs of biplane distance when using INSPR (30 trials for each bias).



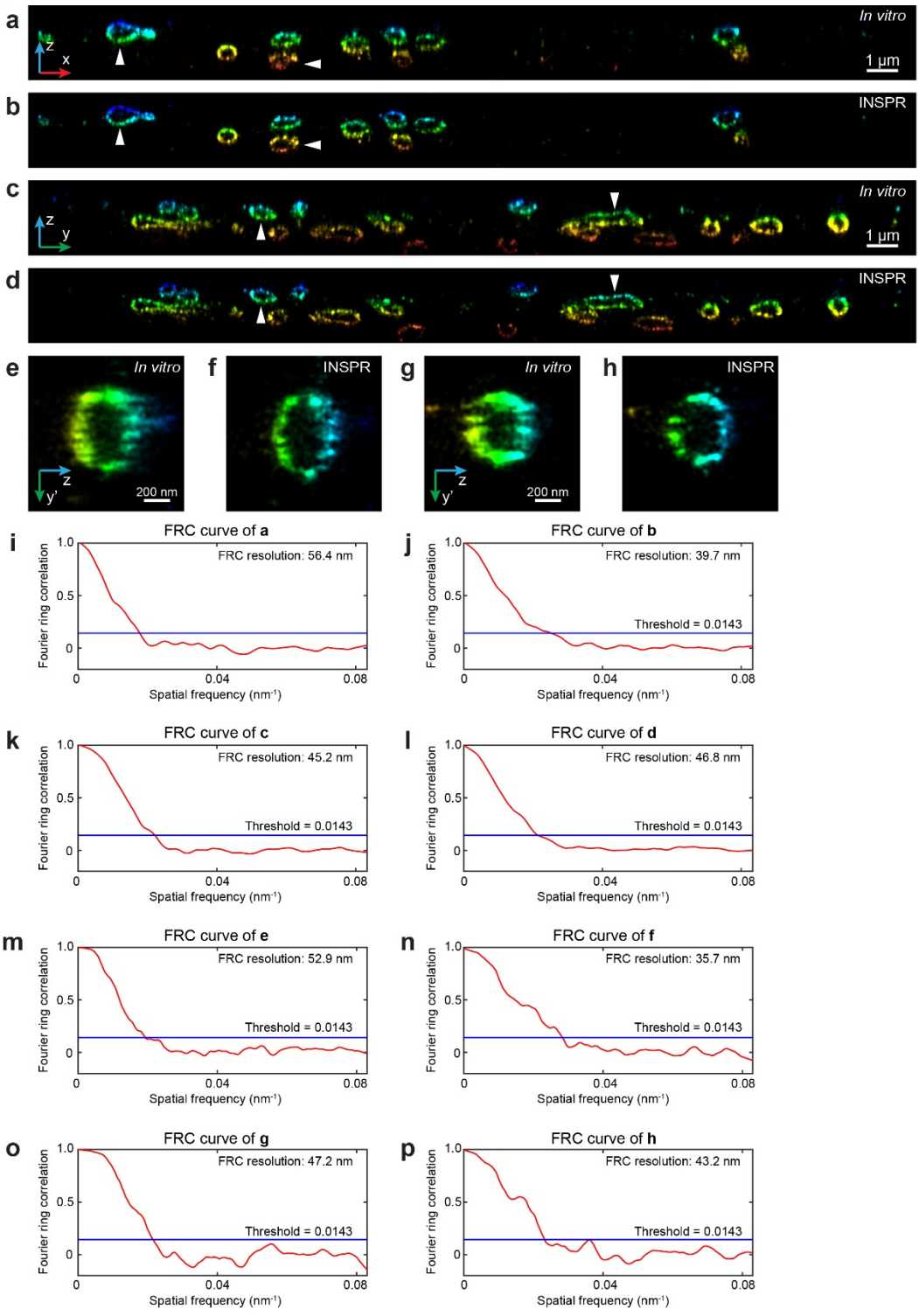
Supplementary Figure 13. Goodness-of-fitting of INSPR and *in vitro* methods.

(a–j) Comparisons of histograms of goodness-of-fitting metric of INSPR with *in vitro* methods cross datasets included in the manuscript. LLR: log-likelihood ratio described in **'Rejection methods'** section. Cspline: the PSF model is built by using cubic spline from fluorescent beads attached on the coverslip. Gaussian: the PSF model is built by using microsphere-calibrated Gaussian fitting. Ext.: extended. Supple.: supplementary.



Supplementary Figure 14. 3D super-resolution reconstructions of immunofluorescence-labeled amyloid β plaques in mouse brains and elastic fibers in developing cartilage using INSPR in biplane setup.

(a) x-y overview of the super-resolution volume of an amyloid β plaque with low-density fibrils in a mouse brain section, showing the positions of 40 typical individual fibrils (white lines). **(b)** Diffraction-limited image of (a). **(c)** x-y overview of the super-resolution volume of an amyloid β plaque with high-density fibrils in a mouse brain section, showing the positions of 40 typical individual fibrils (white lines). **(d)** Diffraction-limited image of (c). **(e)** x-y overview of the 3.1-µm-thick developing cartilage, showing the positions of 15 long elastic fibers with 3–5 measurements per fiber and 7 short fibers with single measurement (yellow lines). **(f)** Diffraction-limited image of (e).



Supplementary Figure 15. Resolution quantifications using Fourier ring correlation (FRC).

(a–d) x-z and y-z slices of 3D reconstructions of the mitochondrial network in **Fig. 3f** using the *in vitro* method (a,c) and INSPR (b,d). The white arrows highlight the difference between two methods. (**e–h**) Enlarged y'-z views of the outer membrane structures reconstructed using the *in vitro* method (e,g) and INSPR (f,h). (**i–p**) Typical FRC curves from 20 random measurements for the datasets in (a–h).

Supplementary Table 1. Imaging parameters for experimental data.

Datasets	Frames per step	Number of steps	Number of cycles	Total frames	Thickness (μm)	Center depth (μm)	Mean $\sqrt{\mathrm{CRLB}_{x,y}}$ (nm)	Mean $\sqrt{\mathrm{CRLB}_{z}}$ (nm)	Mean photon count	Mean background count	Number of localizations
Fig. 3f (TOM20)	2,000	6	20	240,000	~2.5	~10.3	8.2 (x) 7.2 (y)	21.3	5,951	144.2	1,052,610
Fig. 4a (Nup98)	2,000	5	10	100,000	~3.3	~1.7	9.2 (x) 10.4 (y)	33.8	2,531	34.2	622,142
Fig. 4e (Nup98)	1,000	14	10	140,000	~6.4	~3.2	9.3 (x) 10.4 (y)	35.0	2,498	34.9	381,054
Fig. 5a (amyloid β)	2,000	5	9	90,000	~3.7	~7.9	7.2 (x) 7.3 (y)	28.9	5,419	99.2	586,109
Fig. 5k (amyloid β)	2,000	5	14	140,000	~3.3	~14.7	8.7 (x) 8.6 (y)	32.8	4,138	104.2	1,090,760
Fig. 6a (ChR2-EYFP)	2,000	7	9	126,000	~4.2	~4.1	7.8 (x) 6.8 (y)	28.1	4,942	84.1	886,838
Fig. 6h (elastic fibers)	2,000	5	13	130,000	~3.1	~15.6	11.4 (x) 11.1 (y)	44.6	3,043	138.5	1,988,098
Extended Data Fig. 4 (TOM20)	2,000	6	25	300,000	~2.0	~14.0	8.0 (x) 7.9 (y)	26.1	3,434	84.8	2,237,989
Extended Data Fig. 5a (TOM20)	2,000	1	75	150,000	~1.3	~0.7	5.9 (x) 7.4 (y)	17.3	3,243	44.7	1,451,024
Extended Data Fig. 5g (TOM20)	2,000	1	30	60,000	~1.4	~2.2	6.7 (x) 5.7 (y)	20.4	6,351	172.9	805,269
Extended Data Fig. 9a (ChR2-EYFP)	2,000	8	14	224,000	~4.3	~13.2	10.7 (x) 10.1 (y)	35.9	3,875	141.1	735,193
Extended Data Fig. 9g (ChR2-EYFP)	2,000	8	13	208,000	~4.0	~9.0	11.1 (x) 9.8 (y)	36.1	3,816	122.9	990,732
Supple. Fig. 9 (ChR2-EYFP)	2,000	7	8	112,000	~4.2	~3.1	7.6 (x) 7.4 (y)	28.7	5,113	124.2	642,870
Supple. Fig. 11a (α-tubulin)	2,000	7	10	140,000	~4.1	~2.1	7.2 (x) 7.4 (y)	28.0	2,849	39.1	1,459,461
Supple. Fig. 11e (α-tubulin)	2,000	1	46	92,000	~1.7	~0.9	8.0 (x) 8.6 (y)	35.0	2,497	38.8	1,876,043
Supple. Fig. 11e (TOM20)	2,000	1	50	100,000	~1.9	~1.0	7.8 (x) 7.9 (y)	36.5	2,956	54.9	1,410,818

Supplementary Table 2. Parameter settings of INSPR for performance test based on wavefront distortions induced by a deformable mirror.

Zernike-based aberrations	Amplitude ($\lambda/2\pi$)	I_{seg} (photons)	Number of sub-regions	$\mathrm{Sim}_{\mathrm{min}}$	N_g	Z_shift_mode
Ast	+1	80	5,025	0.5	25	Shift
	–1	80	4,978	0.5	25	Shift
DAst	+1	80	4,948	0.5	25	Shift
	–1	85	4,933	0.5	25	Shift
Coma x	+1	70	5,053	0.5	25	Shift
	–1	85	4,990	0.5	25	Shift
Coma y	+1	75	4,959	0.5	25	Shift
	–1	70	4,888	0.5	25	Shift
1st Sph	+1	45	4,831	0.6	25	No shift
	–1	35	4,829	0.6	25	No shift
Trefoil x	+1	50	5,114	0.5	25	Shift
	–1	50	5,069	0.5	25	Shift
Trefoil y	+1	45	5,088	0.5	25	Shift
	–1	40	5,160	0.5	25	Shift
2nd Ast	+1	35	5,058	0.6	15	Shift
	–1	45	4,959	0.6	15	Shift
2nd DAst	+1	45	5,145	0.6	25	Shift
	–1	50	5,110	0.6	25	Shift
2nd Coma x	+1	60	5,055	0.5	15	No shift
	–1	55	4,942	0.5	15	No shift
2nd Coma y	+1	45	5,157	0.5	25	No shift
	–1	60	5,155	0.5	25	No shift
2nd Sph	+1	60	4,879	0.6	25	No shift
	–1	90	5,067	0.5	15	No shift
Tetrafoil x	+1	65	5,020	0.6	25	No shift
	–1	60	5,001	0.6	25	No shift
Tetrafoil y	+1	60	5,032	0.5	25	No shift
	–1	55	5,036	0.5	25	No shift
2nd Trefoil x	+1	35	5,053	0.5	25	No shift
	–1	40	4,815	0.5	25	No shift
2nd Trefoil y	+1	35	4,770	0.5	25	No shift
	–1	35	5,073	0.5	25	No shift
3rd Ast	+1	40	5,045	0.5	25	No shift
	–1	40	4,935	0.5	25	No shift
3rd DAst	+1	45	5,126	0.5	25	No shift
	–1	45	4,839	0.5	25	No shift
3rd Coma x	+1	60	4,880	0.5	25	No shift
	–1	50	4,949	0.5	25	No shift
3rd Coma y	+1	85	5,021	0.5	25	No shift
	–1	80	4,941	0.5	25	No shift
3rd Sph	+1	75	5,083	0.6	25	No shift
	–1	65	5,115	0.6	25	No shift

Common parameters used for INSPR include the sub-region size of 40 × 40 pixels, d_{thresh} of 28, I_{init} of 30, d_t of 145 nm, and XY_shift_mode of 'together shift'.

Ast: vertical astigmatism; DAst: diagonal astigmatism; Coma x: horizontal coma; Coma y: vertical coma; 1st Sph: primary spherical; Trefoil x: horizontal trefoil; Trefoil y: vertical trefoil; 2nd Ast: secondary vertical astigmatism; 2nd DAst: secondary diagonal astigmatism; 2nd Coma x: secondary horizontal coma; 2nd Coma y: secondary vertical coma; 2nd Sph: secondary spherical; Tetrafoil x: horizontal tetrafoil; Tetrafoil y: vertical tetrafoil; 2nd Trefoil x: secondary horizontal trefoil; 2nd Trefoil y: secondary vertical trefoil; 3rd Ast: tertiary vertical astigmatism; 3rd DAst: tertiary diagonal astigmatism; 3rd Coma x: tertiary horizontal coma; 3rd Coma y: tertiary vertical coma; 3rd Sph: tertiary spherical.

Supplementary Table 3. Resolution results using Fourier ring correlation (FRC).

Datasets	Reconstruction methods	Resolution results (mean±s.d., 20 trials, unit: nm)
Supple. Fig. 15a	*In vitro*	56±2
Supple. Fig. 15b	INSPR	40±2
Supple. Fig. 15c	*In vitro*	45±1
Supple. Fig. 15d	INSPR	47±2
Supple. Fig. 15e	*In vitro*	53±4
Supple. Fig. 15f	INSPR	36±2
Supple. Fig. 15g	*In vitro*	47±4
Supple. Fig. 15h	INSPR	43±3

Supplementary Notes

1. Supplementary demonstrations of INSPR

1.1. Comparison of simulated nanoscale structures at different imaging depths

We calculated the localization precision in x, y, and z dimensions according to our previously estimated optical aberrations[1] (**Supplementary Fig. 1a–c**). As the depth increased, the localization precision in the z dimension deteriorated significantly (**Supplementary Fig. 1c**). Then we evaluated the performance limits of INSPR by reconstructing simulated, axially-distributed, double-line structures at different imaging depths to show the impact of inaccurate fitting models on the reconstructions, especially along the axial dimension (**Supplementary Fig. 1d**). The x-z views of the structure were reconstructed using the ground truth pupil, 3D PSF models retrieved by INSPR, phase retrieved PSF at the depth of 0 μm[2,3] (PR), and theoretical PSF with no aberration. At 0 μm depth, the structures reconstructed using INSPR and PR were nearly identical with the ground truth, while the reconstruction using theoretical PSF was tilted. At depths of 6.7, 14.35, and 27.55 μm, the structures reconstructed using INSPR were still close to reconstructions using the ground truth PSF, while reconstructions using *in vitro* PR and theoretical PSF showed obvious distortions. At 45.4 μm depth, the structure reconstructed using INSPR started to appear tilted as compared to the ground truth. Moreover, as the depth increased, even if we used the ground truth PSF to carry out localization, the structure in the z direction still became much more blurred than that in the x direction, which also reflected the deteriorated axial localization precision (**Supplementary Fig. 1c**). It is because when imaging too deep, the information content of single-molecule emission pattern is reduced significantly in the axial direction and cannot be recovered by analytical/numerical techniques.

1.2. Comparison of PSFs and corresponding pupils among different specimens

To illustrate the difference of INSPR retrieved 3D PSFs among different samples, we listed the PSFs retrieved from the datasets of immunofluorescence-labeled TOM20 in COS-7 cells (**Extended Data Fig. 6c**), Nup98 in COS-7 cells (**Supplementary Fig. 2**), amyloid β plaques in mouse brains (**Supplementary Fig. 3**), ChR2-EYFP on dendrites (**Supplementary Fig. 4a**), and elastic fibers in developing cartilage (**Supplementary Fig. 4b**). We found these retrieved 3D PSFs varied significantly among different samples, which demonstrated the existence and variation of sample-induced aberrations.

Also, we compared the pupils in four sub-regions of the 30 μm × 30 μm imaging area (**Supplementary Fig. 5**), retrieved by INSPR from the datasets of immunofluorescence-labeled Nup98 in COS-7 cells in **Fig. 4e** and ChR2-EYFP on dendrites in **Fig. 6a**. 6000 frames from the first optical section of the blinking data

were used for model generation in each dataset. Comparing the phase of the four pupils (**Supplementary Fig. 5a,b**) and their decomposed Zernike amplitudes (**Supplementary Fig. 5c,d**), we observed that the pupil difference among these four sub-regions was small. These data suggest that due to the limited field of view (FOV) and emission path design, using the same pupil in the entire acquisition area for each optical section is valid. On the other hand, in cases where the designed FOV is large[4,5], INSPR can be useful to treat FOV-dependent aberrations by generating multiple *in situ* PSF models across the FOV.

1.3. Comparing INSPR with ZOLA-3D and cubic spline in astigmatism-based setup

We tested the performance of INSPR in astigmatism-based setup. We used INSPR to retrieve a known wavefront distortion from single-molecule emission patterns simulated randomly within an axial range of ±800 nm (**Supplementary Video 2** for 30 random trials and **Supplementary Fig. 6** for an example). The known wavefront distortion consisted of 21 Zenike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration), where the amplitude of vertical astigmatism was set to +1.5 (unit, $\lambda/2\pi$) as prior knowledge, and the amplitudes of the rest aberrations were randomly sampled from –1 to +1 (unit, $\lambda/2\pi$). INSPR successfully retrieved the *in situ* pupil with a phase error of 38±19 mλ (measured by root-mean-square error (RMSE), mean±s.d., **Supplementary Fig. 6b, Supplementary Video 2**), and a Zernike amplitude error of 26±12 mλ for the total 21 modes (measured by RMSE, **Supplementary Fig. 6c, Supplementary Video 2**). The INSPR retrieved 3D PSF showed high similarity with the ground truth PSF (**Supplementary Fig. 6d**).

We imaged immunofluorescence-labeled TOM20 in COS-7 cells in the astigmatism-based setup, where the deformable mirror was used to induce vertical astigmatism with an amplitude of +1.5 (unit, $\lambda/2\pi$). We reconstructed the interconnected mitochondrial network using INSPR, ZOLA-3D[6], and cubic spline from beads on the coverslip[7], where the imaged structures were at a depth of 13 μm from the coverslip (**Extended Data Fig. 4**). We observed improved performance using ZOLA-3D compared with cubic spline, as ZOLA-3D allowed incorporation of theoretically calculated index mismatch aberration in the PSF models. When compared with INSPR, reconstructions using ZOLA-3D and cubic spline methods resulted in broader distribution of membrane profiles and slightly compressed cross sections (~15%) which were caused by the sample-induced aberrations, – not accountable in these *in vitro* methods.

1.4. Comparing INSPR with microsphere-based fitting in astigmatism-based setup

We compared INSPR with a localization method which pinpointed the axial positions of single molecules by Gaussian fitting based on microspheres coated with fluorescent molecules[8]. The original development demonstrated its performance for astigmatism-based super-resolution imaging within 1.2 μm range above the coverslip. We followed this method to attach Alexa Fluor 647 labeled microspheres on the coverslip (**Online Methods**). To decrease statistical uncertainties, we recorded 2–3 microspheres in one FOV simultaneously. For each microsphere, we measured the center (x_0, y_0) and the radius R and then acquired single-molecule blinking data of this microsphere. We extracted the position (x, y) of each single molecule as well as its widths σ_x and σ_y by Gaussian fitting. By using x, y, x_0, y_0, and R, we obtained the depth z of each molecule from $\rho^2 + (R - z)^2 = R^2$, where $\rho = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ is the lateral radial position to the center of the microsphere. Then we generated a series of 60-nm sliding windows in the axial direction, and arranged all the detected emission patterns into these windows according to their depths. For the emission patterns in each sliding window, we calculated the mean (μ_w) and standard deviation (σ_w) of their σ_x and σ_y, and rejected those with σ_x or σ_y greater than $\mu_w + 1.5\sigma_w$ or less than $\mu_w - 1.5\sigma_w$. After that, the calibration curve was obtained by fitting the mean σ_x and σ_y in each sliding window as a function of z position[9,10].

We first compared the performance of INSPR and microsphere-calibrated Gaussian fitting within its expected working range by reconstructing immunofluorescence-labeled TOM20 in COS-7 cells on the bottom coverslip in the astigmatism-based setup (**Extended Data Fig. 5a–f**). The outer membrane contours resolved by using INSPR and microsphere-calibrated Gaussian fitting showed that both two methods worked well in this range (**Extended Data Fig. 5b–e**).

We then compared the performance of INSPR and microsphere-calibrated Gaussian fitting outside its working range (**Extended Data Fig. 5g–l**). We reconstructed immunofluorescence-labeled TOM20 in COS-7 cells in the astigmatism-based setup using these two methods, where the imaged structures were at a depth of 1.5 μm from the coverslip. We found the structures reconstructed using microsphere-calibrated Gaussian fitting were elongated along the axial direction, while INSPR still worked well (**Extended Data Fig. 5h–k**). Besides, we showed the cross sections with localized molecules color coded by their frame numbers in which the molecules were detected (**Supplementary Fig. 7**), which demonstrated that the observed elongation was not caused by the axial drift of the sample during data acquisition or alignment inaccuracies, but by localization imprecisions and biases due to the inaccurate fitting model.

1.5. Comparing INSPR with phase retrieval methods using beads in gel in biplane setup

We compared INSPR with another localization algorithm using 3D PSF obtained from beads embedded in agarose gel[11] (hereafter referred as 'PR in gel'), whose refractive index was similar to that of the imaging medium. We incubated fluorescent beads embedded in agarose gel (**Online Methods**), imaged isolated beads at different imaging depths, used phase retrieval to get the 3D PSF model, and then used this *in vitro* model for localization.

We applied both methods to reconstruct immunofluorescence-labeled ChR2-EYFP on dendrites with depths of 11 μm and 7 μm inside a 50-μm-thick mouse brain section in the biplane setup (**Extended Data Fig. 9**). By using INSPR, the membrane contours labeled by ChR2-EYFP could be reconstructed in both lateral and axial dimensions, while the reconstruction using PR in gel showed distorted axial cross sections and broader membrane profiles (**Extended Data Fig. 9b–e, h–k**). Besides, we showed the cross sections with localized molecules color coded by their frame numbers in which the molecules were detected (**Supplementary Fig. 8**), which demonstrated that the observed distortion was not caused by the axial drift of the sample during data acquisition or alignment inaccuracies, but by localization imprecisions and biases due to the inaccurate PSF model.

This result shows that although using PR in gel considers the index mismatch aberration, it remains an *in vitro* calibration and therefore cannot account for sample-induced aberrations which vary from sample to sample, and are prominent in tissue specimens.

1.6. Comparing INSPR with phase retrieval methods with/without the theoretical index mismatch model in biplane setup

We compared INSPR with phase-retrieval based localization algorithm with/without theoretical index mismatch model[2,3] (hereafter referred as 'PR+IMM/PR') in a biplane setup by reconstructing tissues with a depth up to 17 μm (**Extended Data Figs. 8, 10, Supplementary Fig. 9**). INSPR clearly resolved the membrane contours immunofluorescence-labeled by ChR2-EYFP on dendrites as well as the immunofluorescence-labeled elastic fibers in developing cartilage, highlighted by the axial cross sections, while *in vitro* methods result in reconstruction artifacts and poor resolvability (**Extended Data Fig. 8, Supplementary Fig. 9**). We added comparison among the blinking datasets of immunofluorescence-labeled ChR2-EYFP on dendrites in mouse brains and the 3D PSFs retrieved by INSPR and *in vitro* methods (**Supplementary Fig. 10**), which provides additional evidence in raw detected images showing INSPR's ability of extracting observed emission shapes from single-molecule emission patterns.

We found that due to tissue-induced aberrations, incorporating theoretical IMM in PR-PSF plays a small role when the imaging depth is within 5 μm in tissues where dominate aberration modes are coma and diagonal astigmatism (**Extended Data Fig. 8, Supplementary Figs. 9, 10**). However, when imaging at larger depths, incorporating IMM will increase the reconstruction fidelity (**Extended Data Fig. 10**) by making its PSF closer to that retrieved from INSPR than that without considering index mismatch aberration.

1.7. Application of INSPR in 3D cultured cells and Exchange-PAINT imaging in biplane setup

As a supplementary application of INSPR in the biplane setup, we performed reconstructions of immunofluorescence-labeled α-tubulin in collagen embedded 3D-cultured BS-C-1 cells, as well as Exchange-PAINT imaging[12] in COS-7 cells (**Supplementary Fig. 11**). Compared to 2D monolayer cells cultured on flat coverslips, 3D cultured cells can mimic the *in vivo* cell behaviors realistically and provide useful physiological information[13]. A 28.1 μm × 28.1 μm × 4.1 μm volume of microtubules grown in collagen embedded 3D-cultured BS-C-1 cells was reconstructed, where fine and dense microtubules covering the nuclear envelope were resolved (**Supplementary Fig. 11a–d**), showing the capability of INSPR to resolve nanoscale structures in different culture systems. INSPR was further utilized in Exchange-PAINT imaging, where mitochondria and microtubules were sequentially resolved with high resolution (**Supplementary Fig. 11e–i**), showing the potential capability of INSPR when using multiplexed super-resolution labeling approaches.

1.8. Localization precision estimated by Cramér-Rao lower bound

We estimated the mean localization precision by Cramér-Rao lower bound[10] (CRLB, as described in '**Calculation of Cramér-Rao lower bound**' section) for each single-molecule dataset (**Supplementary Table 1**), which is ~8 nm in lateral and ~21 nm in axial dimensions for the dataset in **Fig. 3f** (TOM20), ~8 nm in lateral and ~26 nm in axial dimensions for the datasets in **Extended Data Figs. 4, 5 (**TOM20), ~10 nm in lateral and ~35 nm in axial dimensions for the datasets in **Fig. 4** (Nup98), ~9 nm in lateral and ~33 nm in axial dimensions for the datasets in **Fig. 5** (amyloid β plaques), ~11 nm in lateral and ~36 nm in axial dimensions for the datasets in **Fig. 6a**, **Extended Data Fig. 9**, and **Supplementary Fig. 9** (ChR2-EYFP), ~11 nm in lateral and ~45 nm in axial dimensions for the dataset in **Fig. 6h** (elastic fibers), ~7 nm in lateral and ~28 nm in axial dimensions for the dataset in **Supplementary Fig. 11a** (α-tubulin), ~8 nm in lateral and ~36 nm in axial dimensions for the datasets in **Supplementary Fig. 11e** (α-tubulin and TOM20).

2. Method details

2.1. Definition and characterization of biplane distance

We estimated the distance between two detection planes in the biplane configuration[14] (named as biplane distance δ, **Extended Data Fig. 1g**) by imaging an isolated fluorescent bead on the coverslip at different axial positions. Intensity profiles (*i.e.* PSFs) in each z position were fitted with a 2D Gaussian function

$$f(x,y) = I_0 e^{-\left[\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right]} + b, \tag{1}$$

where (x_0, y_0) is the center position of each PSF, I_0 is the peak intensity, b is the background, and σ is the standard deviation representing the width of the PSF. The modulation of the estimated standard deviation σ_i in each axial position z was modeled as

$$\sigma_i(z) = \sigma_{i0}\sqrt{1 + \left(\frac{z-c_i}{d_i}\right)^2 + A_i\left(\frac{z-c_i}{d_i}\right)^3 + B_i\left(\frac{z-c_i}{d_i}\right)^4}, \tag{2}$$

where i is the index of plane ($i = 1,2$), σ_{i0} is the standard deviation of the PSF when the emitter is at the focal plane, c_i is the position of the focal plane, d_i is the depth of focus, and A_i and B_i are empirical coefficients[9,10]. From the position of each focal plane, we obtained the biplane distance $\delta = |c_1 - c_2|$.

When imaging above the coverslip surface, due to the mismatched refractive indices between objective immersion and imaging media, the actual position c_i' of each focal plane is shifted to $c_i n_\mathrm{w}/n_\mathrm{o}$, where n_w and n_o are the refractive indices of the water-based imaging medium and the objective immersion medium, respectively[15]. Therefore, the distance between two detection planes in the imaging medium is changed to $\delta n_\mathrm{w}/n_\mathrm{o}$. This rescaling is only valid when the numerical aperture of the objective lens NA $\ll 1$. Since the NA of our objective lens is 1.35, a biplane distance bias will exist. To estimate this potential bias, we performed biplane distance measurement in both index matched and mismatched cases and compared the obtained mean distance with the theoretical calculation (**Supplementary Fig. 12a,b**). To mimic the conditions when imaging on or above the coverslip, we created a 5-μm-thick sample cavity filled with imaging medium (for the refractive index matched medium, $n_\mathrm{buffer} = n_\mathrm{o} = 1.406$, and for the refractive index mismatched medium, $n_\mathrm{buffer} = n_\mathrm{w} = 1.352$) between two coverslips, with fluorescent beads attached on the upper one. When the biplane distance in index matched medium was 322 nm (mean of 19 measurements), the rescaled biplane distance in index mismatched medium could be estimated as $322 \times \frac{1.352}{1.406} = 309.6$ nm, while the measured distance was 317 nm (mean of 20 measurements, **Supplementary Fig. 12a**). When the biplane distance in index matched medium was 607 nm (mean of 20 measurements),

the rescaled biplane distance in index mismatched medium could be estimated as $607 \times \frac{1.352}{1.406} = 583.7$ nm, while the measured distance was 585 nm (mean of 22 measurements, **Supplementary Fig. 12b**). Based on these data, we found that the biplane distance bias caused by rescaling was within 5%. By this calculation, we demonstrate that theoretical approximation can provide reasonable estimation of biplane distance in index mismatched cases for a silicone-oil-immersion objective lens while providing an alternative approach to measure the distance between two detection planes.

Further, we evaluated the resulting RMSE of Zernike amplitudes and localization bias induced by a possible biased input of biplane distance within INSPR (**Supplementary Fig. 12c–e).** The ground truth biplane distance was set to 580 nm, and 30 sets of wavefront distortions were generated, consisting of 21 Zernike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration) with their amplitudes randomly sampled from –1 to +1 (unit, $\lambda/2\pi$). We first used these random wavefront distortions to generate PSFs located randomly in an axial range of ±800 nm. Subsequently, we used INSPR to retrieve corresponding pupils from these PSFs with biased inputs of biplane distance (from –20% to 20%), and compared their decomposed Zernike amplitudes with the ground truth amplitudes. We reported the RMSEs of INSPR retrieved amplitudes at different bias levels of biplane distance inputs and found that they were generally within 20 mλ (**Supplementary Fig. 12d**). To quantify the potential localization biases introduced by inaccurate biplane distance inputs, we generated simulated PSFs using ground truth wavefront distortions and axial positions ranging from –500 to +500 nm with a step size of 100 nm. We then used the retrieved pupils (as described above) to localize these PSFs and compared with their ground truth positions to calculate the localization biases (**Supplementary Fig. 12e**). We found that inaccurate biplane distance introduced biases in axial localization, and these introduced biases were more prominent at large axial ranges and small near focus. Increasing biplane distance bias increases axial localization bias. When the input biplane distance bias was within 5% from the ground truth, the observed localization bias was generally within 5 nm in the axial direction (**Supplementary Fig. 12e**).

2.2. Calibration of deformable mirror

The calibration of the deformable mirror (DM) was achieved according to previously described methods[1,16]. The response of the membrane surface of the DM can be decomposed into an orthogonal set of mirror deformation modes (DM modes), resembling Zernike-based aberration modes for describing the aberrations in an optical system[17]. We selected 23 DM modes related to Zernike modes from vertical astigmatism to tertiary spherical aberration (Wyant order) for calibration. For each DM mode, we retrieved pupil functions using a phase retrieval algorithm[2,3] for 9 different input amplitudes (–4, –3, –2, –1, 0, 1, 2, 3, 4; unit, $\lambda/2\pi$)

by imaging isolated fluorescent beads on the coverslip at 11 different axial positions (in an axial range from –2 to +2 μm with a step size of 400 nm). The amplitudes of 32 Zernike modes (Wyant order, from vertical astigmatism to quaternary spherical aberration) were extracted for different input amplitudes in each DM mode. The amplitude relationship between DM mode and Zernike mode can be described as

$$\Phi_{\mathrm{DM}}(i,j) = \sum_{n=5}^{36} c_n(i,j) Z_n \,, \tag{3}$$

where $\Phi_{\mathrm{DM}}(i,j)$ describes the ith DM mode with input amplitude j (j = –4, –3, –2, –1, 0, 1, 2, 3, 4; unit, λ/2π), Z_n and $c_n(i,j)$ describe the nth Zernike mode and its corresponding amplitude (the first four Zernike modes, piston, x-tilt, y-tilt, and defocus, are removed here). By linearly fitting through the amplitudes of Zernike modes as a function of DM mode amplitudes, the relationship between DM mode and Zernike mode can be expressed as

$$\Phi'_{\mathrm{DM}}(i) = \sum_{n=5}^{36} c'_n(i) Z_n \,, \tag{4}$$

where $\Phi'_{\mathrm{DM}}(i)$ describes the ith DM mode, Z_n describes the nth Zernike mode, and $c'_n(i)$ is its fitting amplitude. By solving Eq. (4) with least-squares, each Zernike mode was expressed as a linear combination of DM modes.

We then tested the calibration accuracy for 21 Zernike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration). For each Zernike mode, we retrieved pupil functions for amplitudes at ±1 (unit, λ/2π) by imaging a fluorescent bead on the coverslip at different imaging depths, decomposed the pupil functions into 21 Zernike modes, and obtained their retrieved amplitudes. To eliminate the influence of the aberrations induced by the imaging system itself, we calculated the difference between the retrieved amplitudes of the test Zernike mode at amplitudes of +1 and –1 (unit, λ/2π), and divided this difference by 2. After processing 21 Zernike modes, we built a heat map representing the relationship between the input and phase retrieved amplitudes of Zernike modes (**Extended Data Fig. 2d**). We observed that this heat map always got the largest estimation along the diagonal elements. The average error between the input and phase retrieved amplitude along the diagonal elements was 6% in the first 14 Zernike modes and 27% in the last 7 Zernike modes. Besides, we calculated the RMSE between the 21 phase retrieved amplitudes and the 21 input amplitudes (only the test mode has an amplitude of 1, and the amplitudes of other modes are all 0, unit, λ/2π) for each Zernike mode, and got an average RMSE of 13 mλ for the total 21 test modes.

2.3. Characterization of sCMOS camera

The characterization of sCMOS camera (including the offset, variance, and gain for each pixel on the camera) was estimated according to previously described methods[18]. Briefly, the offset o_q for pixel q on the camera was obtained by performing a temporal average on 20,000 frames acquired with the camera's entrance port being covered, as

$$o_q = \frac{1}{M}\sum_{m=1}^{M} s_q^m, \tag{5}$$

where s_q^m is the ADU count at frame m for pixel q, and M is the total frame number.

The variance var_q for pixel q was obtained by calculating

$$var_q = \frac{1}{M}\sum_{m=1}^{M} \left(s_q^m\right)^2 - {o_q}^2. \tag{6}$$

To estimate the gain for each pixel, we imaged a fluorescent plastic slide (92001, Chroma) using the 642-nm laser at different illumination intensities (20,000 frames were recorded in each intensity level). By using the Moore-Penrose pseudo-inverse algorithm, the gain g_q for pixel q was obtained as

$$g_q = \left(B_q B_q^T\right)^{-1} B_q A_q^T, \tag{7}$$

$$A_q = \left\{\left(v_q^1 - var_q\right), \cdots, \left(v_q^n - var_q\right), \cdots, \left(v_q^N - var_q\right)\right\}, \tag{8}$$

$$B_q = \left\{\left(D_q^1 - o_q\right), \cdots, \left(D_q^n - o_q\right), \cdots, \left(D_q^N - o_q\right)\right\}, \tag{9}$$

where v_q^n and D_q^n stand for the temporal variance and average of the ADU counts (20,000 frames) for the nth illumination intensity in pixel q, respectively, and N is the total number of illumination intensity levels.

2.4. Data acquisition

The SMLM setup is extremely susceptible to sample drift in the axial direction for its long data acquisition time, typically from tens of minutes to hours. To compensate this drift, we implemented a focus stabilization module[19]. Before fluorescence imaging, we recorded a series of bright-field images of the sample along the axial direction (from –1 to +1 μm, with a step size of 100 nm) as reference images. During fluorescence imaging, we recorded a real-time bright-field image of the sample after each acquisition cycle (1000 or 2000 frames, depending on the sample stability), and compared the similarities between this real-time image

and reference images by calculating their 2D correlation. The correlation values of the most similar reference image and its nine adjacent images, together with their z positions, were fitted with third degree polynomials. The z position corresponding to the maximum correlation value in the fitting curve was treated as the sample drift. Then we moved the objective lens in the inverse direction to compensate this drift. In this way, focus stabilization can be achieved during data acquisition.

The single-plane dataset for DM calibration (as described in **'Calibration of deformable mirror'** section) was collected by imaging a fluorescent bead on the coverslip over an axial range from –2 to +2 μm with a step size of 400 nm, and taking 10 frames per step with a frame rate of 5 Hz. The amplitude of each DM mode was set to be –4, –3, –2, –1, 0, 1, 2, 3, and 4 (unit, $\lambda/2\pi$).

The biplane dataset for camera calibration (as described in **'Characterization of sCMOS camera'** section) was collected by imaging a fluorescent slide. With the cap covered on the camera's entrance port, 20,000 frames were collected with a frame rate of 50 Hz to calculate the offset and variance of the readout noise for each pixel. With illumination from the 642-nm laser at nine different intensities (from 20 to 300 photons), 20,000 frames were collected with a frame rate of 50 Hz for each illumination level to estimate the gain for each pixel. The datasets in astigmatism-based setup shared the same region with those of plane 1 in biplane setup.

The biplane datasets for testing the accuracy of DM calibration (**Extended Data Fig. 2d**), measuring the biplane distance (**Extended Data Fig. 1g**), and building the *in vitro* model (**Fig. 3f–s, Extended Data Figs. 6–10, Supplementary Fig. 9**) were separately collected by imaging fluorescent beads on the coverslip or in the agarose gel over an axial range from –1.5 to +1.5 μm with a step size of 100 nm, and taking 50 frames per step with a frame rate of 10 Hz. The biplane distance (as described in **'Definition and characterization of biplane distance'** section) was estimated to be 580 nm for distorted wavefront control (**Fig. 2e,f**), 286 nm for imaging Alexa Fluor 647 labeled mitochondria (**Fig. 3f–s**), 568 nm for imaging dendrites with depths of 7 μm and 11 μm (**Extended Data Fig. 9**), and 558 nm for all the other imaging sessions (**Figs. 4–6, Supplementary Figs. 9, 11**).

The biplane datasets for measuring the biplane distance in different conditions (**Supplementary Fig. 12a,b**) were collected by imaging fluorescent beads in different imaging media over an axial range from –1.5 to +1.5 μm with a step size of 50 nm, and taking 3 frames per step with a frame rate of 5 Hz. The biplane distance was set to ~600 nm (**Supplementary Fig. 12a**) and ~300 nm (**Supplementary Fig. 12b**).

The astigmatism-based dataset for building the *in vitro* cubic spline model (**Extended Data Fig. 4**) was collected by imaging fluorescent beads on the coverslip over an axial range from –1 to +1 μm with a step size of 50 nm, and taking 50 frames per step with a frame rate of 10 Hz (~5 beads in each dataset, 3 datasets

in total). Here we used DM to induce vertical astigmatism with an amplitude of +1.5 (unit, λ/2π). Due to instrument imperfections, the setup itself has vertical astigmatism with an amplitude of –0.3 (unit, λ/2π), so the resulting vertical astigmatism has an amplitude of +1.2 (unit, λ/2π) as prior knowledge.

The astigmatism-based dataset for obtaining the calibration curve from microspheres (**Extended Data Fig. 5**) was collected by imaging Alexa Fluor 647 labeled microspheres on the coverslip. The microsphere sample was first illuminated with the transmitted light to record a bright-field image at the equatorial plane of the microspheres, which was used to measure both the radius R and the center (x_0, y_0) of each microsphere. Then the objective lens was moved axially to the selected imaging depth. Before fluorescence imaging, bright-field images of this region were recorded over an axial range from –1 to +1 μm with a step size of 100 nm as reference images for focus stabilization. Then the blinking data were collected at the illumination of the 642-nm laser. The laser power was 17 kW/cm^2 to get low density of molecules. 1000 frames were collected per cycle with a frame rate of 50 Hz and ~15 cycles were collected.

In biological imaging (**Figs. 2f, 3f–s, 4–6, Extended Data Figs. 4, 5, 9, Supplementary Figs. 9, 11**), the sample was first excited with the 642-nm laser at a low intensity of ~50 W/cm^2 to find a region of interest. The depth from this region to the bottom coverslip was measured by recording a first position of the objective lens when the dusts on the bottom coverslip were in focus, then recording a second position of the objective lens when the region of interest was in focus. The difference between these two recorded positions was treated as the depth of this region. Before fluorescence imaging, bright-field images of this region were recorded over an axial range from –1 to +1 μm with a step size of 100 nm as reference images for focus stabilization. Then the blinking data were collected at a laser intensity of 2–6 kW/cm^2 and a frame rate of 50 Hz. For distorted wavefront control (**Fig. 2f**), 2000 frames were collected for each Zernike-based aberration mode with its amplitude set at ±1 (unit, λ/2π). For single-section imaging (**Extended Data Fig. 5, Supplementary Fig. 11e–i**), 2000 frames were collected per cycle and ~50 cycles were collected. For multi-section imaging (**Figs. 3f–s, 4–6, Extended Data Figs. 4, 9, Supplementary Figs. 9, 11a–d**), the sample was scanned axially by translating the objective lens with a step size of 400 nm in biplane setup and 250 nm in astigmatism-based setup from the bottom to the top of the sample. 1000 or 2000 frames were collected for each cycle in one optical section, 5–14 optical sections were collected according to the thickness of the sample, and 8–25 cycles were collected in total (**Supplementary Table 1**).

2.5. PSF generation

According to scalar diffraction theory[20], the PSF of an imaging system can be calculated from the Fourier transform of the pupil function as

$$\mu_0(x,y,z) = \left|\mathcal{F}\left[P(k_x,k_y)e^{i2\pi k_z z}\right]\right|^2, \tag{10}$$

where $\mu_0(x,y,z)$ describes the PSF at position (x,y,z) in the sample space, $\mathcal{F}$ denotes the Fourier transform operator, and $P(k_x,k_y)$ is the pupil function at the back focal plane of the objective lens. The size of the pupil function is limited by ${k_x}^2 + {k_y}^2 \le \left(\frac{\mathrm{NA}}{\lambda}\right)^2$, where NA is the numerical aperture of the objective lens and λ is the emission wavelength in air. The defocus phase is described by the factor $e^{i2\pi k_z z}$, where $k_z = \sqrt{\left(\frac{n}{\lambda}\right)^2 - {k_x}^2 - {k_y}^2}$, and n is the refractive index of the objective immersion medium.

The pupil function can be expressed as

$$P\left(k_x,k_y\right) = A\left(k_x,k_y\right)\cdot e^{i\Phi\left(k_x,k_y\right)}, \tag{11}$$

where $A\left(k_x,k_y\right)$ and $\Phi\left(k_x,k_y\right)$ are the magnitude and phase of the electric field at the pupil plane, respectively. $\Phi\left(k_x,k_y\right)$ describes the optical aberrations introduced by instrument imperfections and the local biological context, which can be decomposed into a series of Zernike modes[21] as

$$\Phi\left(k_x,k_y\right) = \sum_{n=1}^{N} c_n Z_n(k_x,k_y), \tag{12}$$

where $Z_n(k_x,k_y)$ describes the nth Zernike mode, c_n is its corresponding amplitude, and N is the number of Zernike modes. In our simulations, N was set to 25. Among these 25 Zernike modes, piston, x-tilt, and y-tilt do not influence the shape of the PSF, and defocus depends on the axial position, so we only considered the rest 21 Zernike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration).

The simulated PSFs in the biplane configuration were generated as follows. (1) Amplitudes $(c_1, c_2, \cdots, c_N)$ was set to obtain a pupil function $P\left(k_x,k_y\right)$. (2) Pairs of normalized PSFs $\mu_0(x,y,z)$ and $\mu_0(x,y,z+\delta)$ were generated for two planes on basis of $P\left(k_x,k_y\right)$, where δ is the biplane distance. Here we assume the two planes share the same pupil function. (3) μ_0 was multiplied with total photon count I and added with a background count bg to obtain ideal PSFs $\mu = I\mu_0 + bg$. (4) μ was corrupted with Poisson noise.

For simulations in biplane setup, the biplane distance δ was set to 580 nm. The simulation parameters including the emission wavelength, the numerical aperture of the objective lens, the effective pixel size on the detection plane, and the refractive indices of the objective immersion and imaging media were consistent with the experimental conditions.

For simulations in **Fig. 2a–d** and **Supplementary Video 1** (randomly aberrated PSFs), the amplitudes of 21 Zernike modes were randomly sampled from –1 to +1 (unit, $\lambda/2\pi$) in each trial (30 trials in total). 2000 PSFs were generated with a sub-region size of 40 × 40 pixels, whose axial positions were randomly sampled from –800 to +800 nm and lateral positions from –3 to +3 pixels to the center of the sub-region. The total photon count per emission event was sampled from a Gaussian distribution with a mean of 2000 and a standard deviation of 500, and the background count per pixel was uniformly sampled from 10 to 20.

For simulations in **Extended Data Fig. 2a,b** (PSFs at different imaging depths), the amplitudes of Zernike modes were set according to our previously estimated optical aberrations (consisting of vertical astigmatism, diagonal astigmatism, horizontal coma, vertical coma, and primary and secondary spherical aberrations) at various imaging depths[1] (0, 6.7, 14.35, 27.55, and 45.4 μm). The axial positions of PSFs were randomly sampled from –800 to +800 nm, but with a defocus offset at a certain depth (**Extended Data Fig. 2b**). The number of PSFs, sub-region size, lateral range to the center, total photon count, and background count were the same with those in **Fig. 2a–d**.

For simulations in **Extended Data Fig. 2c** (different signal to background ratio (SBR) conditions), the amplitudes of 21 Zernike modes were randomly sampled from –1 to +1 (unit, $\lambda/2\pi$) in each trial (11 trials in total). The pairs of the total photon count per emission event and background count per pixel were set to (5000, 5), (1000, 5), (5000, 20), (1000, 20), (5000, 100), and (1000, 100), and the number of PSFs ranged from 100 to 2900 with an increment of 200. The sub-region size, axial range, and lateral range to the center were the same with those in **Fig. 2a–d**.

For simulations in **Extended Data Fig. 2e,f** (channel-specific PSFs), PSFs with vertical astigmatism aberration at an amplitude of 2 (unit, $\lambda/2\pi$) were generated at axial positions from –500 to +500 nm, with a step size of 100 nm, and 1000 PSFs were generated for each z position. The total photon count per emission event was set to 2000, and the background count per pixel was set to 30. Plane 1 and plane 2 were related with an affine transformation including a rotation of 30 degrees. Then, besides Poisson noise, pixel-dependent Gaussian noise was added to each pixel of the simulated datasets (the variance distribution was shown in the inset of **Extended Data Fig. 2e**).

For simulations in **Supplementary Fig. 1** (double-line structures at different depths), the amplitudes of Zernike modes were set according to our previously estimated optical aberrations at various imaging depths[1]. The double lines were separated by 50 nm. 20,000 PSFs from molecules within the simulated structure were generated randomly from –200 to +200 nm in both x and z dimensions. The total photon count per emission event was set to 4000 and the background count per pixel was set to 10.

For simulations in **Supplementary Fig. 12c–e** (influence of biplane distance bias), the amplitudes of 21 Zernike modes were randomly sampled from –1 to +1 (unit, λ/2π) in each trial (30 trials in total). For pupil estimation, the simulated single molecules were randomly distributed from –800 to +800 nm in the axial direction. The total photon count per emission event was sampled from a Gaussian distribution with a mean of 2000 and a standard deviation of 500, and the background count per pixel was uniformly sampled from 10 to 20. For localization, the simulated single molecules were axially distributed from –500 to +500 nm with a step size of 100 nm (1000 single molecules per step). The total photon count per emission event was set to 4000, and the background count per pixel was set to 10.

For simulations in astigmatism-based setup (**Supplementary Fig. 6, Supplementary Video 2**), the amplitude of vertical astigmatism was set to +1.5 (unit, λ/2π) as prior knowledge, and the amplitudes of the rest 20 Zernike modes were randomly sampled from –1 to +1 (unit, λ/2π) in each trial (30 trials in total). 2000 PSFs with a sub-region size of 40 × 40 pixels were randomly sampled with axial positions from –800 to +800 nm and lateral positions from –3 to +3 pixels to the center. The total photon count per emission event was sampled from a Gaussian distribution with a mean of 2000 and a standard deviation of 500, and the background count per pixel was uniformly sampled from 10 to 20.

2.6. INSPR framework

INSPR constructs an *in situ* 3D PSF response directly from single molecules located in the cellular volume. The key of constructing the 3D PSF is to find the relationship between detected single molecules and their axial positions. Drawing inspiration from *k*-means algorithm[22], we assigned n detected single molecules (*i.e.* PSF library $X = \{\mathrm{PSF}_1, \mathrm{PSF}_2, \cdots, \mathrm{PSF}_n\}$) into k reference z-stack PSFs (*i.e.* templates $T = \{\mathrm{PSF}'_1, \mathrm{PSF}'_2, \cdots, \mathrm{PSF}'_k\}, k \ll n$) by calculating

$$\arg\max_{S} \sum_{i=1}^{k} \sum_{j \in S_i} \mathrm{Sim}\left(\mathrm{PSF}_j, \mathrm{PSF}'_i\right), \tag{13}$$

where Sim is a function to measure the similarity between detected single molecule PSF_j and template PSF'_i, and S_i is a set of PSFs that are assigned into the ith template group. Our goal is to construct k group sets $S = \{S_1, S_2, \cdots, S_k\}$ so that the PSFs within each group are similar to each other, while the averaged PSF of each group is dissimilar to that of other groups.

To ensure this PSF assignment is unique, we used the biplane configuration to detect n pairs of PSFs from two axially separated detection planes $X_{bp} = \left\{\left(\mathrm{PSF}_{1,1}, \mathrm{PSF}_{1,2}\right), \left(\mathrm{PSF}_{2,1}, \mathrm{PSF}_{2,2}\right), \cdots, \left(\mathrm{PSF}_{n,1}, \mathrm{PSF}_{n,2}\right)\right\}$.

INSPR started with a constant pupil and generated k pairs of templates $T_{bp} = \{(\mathrm{PSF}'_{1,1}, \mathrm{PSF}'_{1,2}), (\mathrm{PSF}'_{2,1}, \mathrm{PSF}'_{2,2}), \cdots, (\mathrm{PSF}'_{k,1}, \mathrm{PSF}'_{k,2})\}$ according to '**PSF generation**' section. We also demonstrated it in astigmatism-based setup with prior knowledge. All detected PSFs in the library were compared with these templates and then assigned into the most similar template group (assignment step in **Fig. 1**). The process then updated the pupil by using aligned and averaged PSFs from the axially assigned group (update step in **Fig. 1**), and the updated pupil was used to generate new templates. This process usually converged in 6–10 iterations (convergence criteria: the phase difference (measured by RMSE) between two adjacent iterations is smaller than 20 mλ).

INSPR includes three key components: PSF library construction, PSF library assignment, and 3D model estimation (**Extended Data Fig. 1a**), which are explained as follows.

PSF library construction. For biplane setup, the PSF library was constructed from the single-molecule dataset, including pairs of emission patterns at random depths in a certain axial range. The raw datasets from two planes were first aligned to the same regions of interest (biplane registration), and then cropped into individual sub-regions containing single emitters (segmentation). For astigmatism-based setup, the PSF library was directly segmented from the raw dataset. This process includes the following two steps.

1. Biplane registration. For biplane setup, images in plane 1 were treated as reference, and images in plane 2 were aligned to plane 1 using affine transformation (including translation, scale, shear, and rotation)[23]. Transformation between two planes can be obtained either by imaging 10–15 fluorescent beads on the coverslip with an axial range from –1 to +1 μm and a step size of 100 nm (50 frames in each step), or by collecting a single-molecule blinking dataset (1000 or 2000 frames). The image sequences of beads or blinking datasets in two planes were individually projected into the lateral plane by maximum-intensity projection. Then we calculated the affine matrix based on these projection images in two planes (using 'imregtform' function in MATLAB), and registered the images from plane 2 to plane 1 according to the affine matrix (using 'imwarp' function in MATLAB). For astigmatism-based setup, this step was ignored.
2. Segmentation. For biplane setup, after summing the images from registered planes, sub-regions with single molecules were cropped using a segmentation algorithm[18], where two uniform filters with different kernel sizes were used to reduce noise, and a maximum filter was used to find local maximum intensities. The positions of these maximum intensities were localized at the centers of candidate sub-regions. We utilized two intensity thresholds (initial intensity threshold I_{init} and segmentation threshold I_{seg}) and a distance threshold d_{thresh} to make sure that each selected sub-region only contained one molecule with enough brightness. First, the candidate sub-regions were selected if their

maximum intensities were higher than I_{init}. Second, the overlapped molecules were rejected if the centers of two sub-regions were closer than d_{thresh}. Third, the maximum intensities of the rest sub-regions were filtered out if they were below I_{seg}. For *in situ* model estimation in **Fig. 2e,f**, **Supplementary Table 2**, and **Supplementary Video 3** (distorted wavefront control), we set the sub-region size to 40 × 40 pixels, d_{thresh} to 28, I_{init} to 30, and I_{seg} ranging from 35 to 90 in order to get ~5000 single molecules. For *in situ* model estimation in **Figs. 3f**, **4–6**, **Extended Data Fig. 9**, and **Supplementary Figs. 9, 11**, we set the sub-region size to 32 × 32 pixels, d_{thresh} to 26, I_{init} to 25, and I_{seg} to 40. For 3D single-molecule localization (**Figs. 3f, 4–6, Extended Data Fig. 9, Supplementary Figs. 9, 11**), we set the sub-region size to 16 × 16 pixels, d_{thresh} to 10, I_{init} to 25, and I_{seg} to 40. For datasets acquired in the astigmatism-based setup (**Extended Data Figs. 4, 5**), single molecules were directly cropped from the raw dataset. The sub-regions of single molecules in adjacent optical sections were merged to increase the axial range. We set the sub-region size to 32 × 32 pixels, d_{thresh} to 26, I_{init} to 25, and I_{seg} to 40 for *in situ* model estimation, and the sub-region size to 16 × 16 pixels, d_{thresh} to 10, I_{init} to 25, and I_{seg} to 40 for 3D single-molecule localization.

In single-molecule localization techniques, improving the quality of the detected PSFs will improve the accuracy of *in situ* model estimation and 3D localization. This can be realized with combination of adaptive optics[1,24–26], tissue clearing[27] and expansion methods[28,29], and light-sheet illumination approaches[30–32].

PSF library assignment. The detected PSFs in the library were assigned to temporary axial positions following three steps. First, the reference z-stack PSFs in one or two planes were generated from the initial/estimated pupil function. Second, the detected PSFs were classified into different groups based on their similarities with the reference PSFs. Third, 2D alignment was carried out by calculating the shift distance between detected PSFs and the most similar reference PSFs. The details are as follows.

1. Reference z-stack PSF (template) generation. Templates were generated from the initial/estimated pupil function, which was a constant pupil in the first iteration, and iteratively optimized in assignment and update steps. For biplane setup, INSPR generated k templates in each plane with an axial range from –1.4 to +1.4 μm ($T_{bp} = \{(\text{PSF}'_{1,1}, \text{PSF}'_{1,2}), (\text{PSF}'_{2,1}, \text{PSF}'_{2,2}), \cdots, (\text{PSF}'_{k,1}, \text{PSF}'_{k,2})\}$), which is sufficient to cover all the detected PSFs. The axial step size of the templates d_t was set to $\delta/2$ or $\delta/4$, where δ is the biplane distance. The setting of the axial step size made the templates generated from the pupil function of each plane overlap around the center of the axial range. For simulations in biplane setup (**Fig. 2a–d, Extended Data Fig. 2a–c, Supplementary Video 1**), we set δ to 580 nm and d_t to $\delta/4$ (145 nm). For the datasets in **Fig. 2e,f**, **Supplementary Table 2**, and **Supplementary Video 3** (distorted wavefront control), δ was obtained from measurement as 580 nm, and we set d_t to $\delta/4$ (145

nm). For the dataset in **Fig. 3f** (TOM20), δ was obtained from measurement and rescaling as 286 nm, and we set d_t to $\delta/2$ (143 nm). For the dataset in **Extended Data Fig. 9** (ChR2-EYFP), δ was obtained from measurement and rescaling as 568 nm, and we set d_t to $\delta/4$ (142 nm). For other datasets (**Figs. 4–6, Supplementary Figs. 9, 11**), δ was obtained from measurement and rescaling as 558 nm and we set d_t to $\delta/4$ (139 nm). For simulations and datasets in the astigmatism-based setup (**Extended Data Figs. 4, 5, Supplementary Fig. 6, Supplementary Video 2**), the templates were generated at axial positions from –1 to +1 μm, with a step size of 100 nm. We set the amplitude of vertical astigmatism to +1.5 (unit, λ/2π) for simulation and +1.2 (unit, λ/2π) for experimental data as initial guess.

2. Classification. For biplane setup, each pair of detected single molecules in the PSF library ($X_{bp} = \{(\mathrm{PSF}_{1,1}, \mathrm{PSF}_{1,2}),\ (\mathrm{PSF}_{2,1}, \mathrm{PSF}_{2,2}),\ \cdots,\ (\mathrm{PSF}_{n,1}, \mathrm{PSF}_{n,2})\}$) was assigned to a certain template group with the highest similarity. The similarity is defined as

$$\mathrm{Sim}(\mathrm{PSF}_j, \mathrm{PSF}'_i) = \frac{1}{2}\left[\mathrm{NCC}(\mathrm{PSF}_{j,1}, \mathrm{PSF}'_{i,1}) + \mathrm{NCC}(\mathrm{PSF}_{j,2}, \mathrm{PSF}'_{i,2})\right], \quad (14)$$

where NCC is the normalized cross correlation between detected PSFs and templates in each plane. The value of NCC varies from –1 to +1, where high NCC represents high similarity between detected PSFs and templates. Therefore, each detected PSF in the library was classified into a certain template group, which formed k group sets $S = \{S_1, S_2, \cdots, S_k\}$. We used similarity threshold $\mathrm{Sim}_{\min}$ and number threshold N_g to select high-similarity PSFs in each group, where PSFs with a similarity lower than $\mathrm{Sim}_{\min}$ or groups with fewer than N_g PSFs were rejected. For simulations (**Fig. 2a–d, Extended Data Fig. 2a–c, Supplementary Fig. 6, Supplementary Videos 1, 2**), we set $\mathrm{Sim}_{\min}$ to 0.5, and N_g ranging from 5 to 30. For the datasets in **Fig. 2e,f**, **Supplementary Table 2**, and **Supplementary Video 3** (distorted wavefront control), we set $\mathrm{Sim}_{\min}$ to 0.5 or 0.6, and N_g to 15 or 25. For other datasets acquired in the biplane setup (**Figs. 3f, 4–6, Extended Data Fig. 9, Supplementary Figs. 9, 11**), we set $\mathrm{Sim}_{\min}$ to 0.6, and N_g to 30. For the datasets acquired in the astigmatism-based setup (**Extended Data Figs. 4, 5**), each detected single molecule was assigned to a certain template group. We set $\mathrm{Sim}_{\min}$ to 0.6, and N_g to 50.

3. 2D alignment. For each detected PSF, 2D cross correlation was used to calculate the shift distance with its corresponding template. To find the correlation peak from the 2D cross correlation image, Fourier interpolation with 10 times up sampling was used to identify the peak with a sub-pixel size of 12 nm. Then the detected PSF was aligned to its template according to the shift distance. Here two shift modes were used in biplane setup. (1) XY_shift_mode = 'separate shift', meaning that the shift distance was calculated individually for each plane, and the PSFs were aligned to their corresponding templates separately. This mode was used in simulations (**Fig. 2a–d, Extended Data Fig. 2a–c, Supplementary**

Video 1) and distorted wavefront control (**Fig. 2e,f, Supplementary Table 2, Supplementary Video 3**). (2) XY_shift_mode = 'together shift', meaning that the shift distances were calculated together for two planes, and the PSFs of two planes were aligned to the corresponding pair of templates. This mode is more robust especially for data with low SBR, so it was used for experimental data.

In this work, we used biplane configuration and astigmatism configuration with prior knowledge to avoid degeneracies (**Extended Data Fig. 1d,e**). Besides, the framework of INSPR can also be generalized to other configurations, such as using a phase mask to generate non-degenerate PSF shapes[33,34].

3D model estimation. The classified PSFs in each group were averaged to improve SBR, and then re-arranged by their axial positions. The phase retrieval method[2,3] was carried out to estimate a new pupil function, which described the *in situ* 3D model and was used to generate reference z-stack PSFs in '**PSF library assignment**' section. The details are as follows.

1. Group average. Here k group sets $S = \{S_1, S_2, \cdots, S_k\}$ were formed by PSF library assignment. In each group, the assigned PSFs were similar to each other and aligned to the center of the template. In order to obtain high-contrast images, these assigned PSFs were first normalized by z-score normalization and then averaged together as

$$\mathrm{PSF}_{i,m}^{\mathrm{Ave}} = \mathrm{Ave}\left(\left\{\mathrm{PSF}_{j,m} : j \in S_i\right\}\right), (m = 1,2), \tag{15}$$

   where Ave is the image averaging operation, $\mathrm{PSF}_{i,m}^{\mathrm{Ave}}$ is the averaged PSF of plane m in group i, and $\mathrm{PSF}_{j,m}$ is the normalized PSF of plane m in the library. Thus we obtained $2k$ average images $A_{bp} = \left\{\left(\mathrm{PSF}_{1,1}^{\mathrm{Ave}}, \mathrm{PSF}_{1,2}^{\mathrm{Ave}}\right), \left(\mathrm{PSF}_{2,1}^{\mathrm{Ave}}, \mathrm{PSF}_{2,2}^{\mathrm{Ave}}\right), \cdots, \left(\mathrm{PSF}_{k,1}^{\mathrm{Ave}}, \mathrm{PSF}_{k,2}^{\mathrm{Ave}}\right)\right\}$, and their axial positions $Z_{bp} = \left\{\left(Z_{1,1}, Z_{1,2}\right), \left(Z_{2,1}, Z_{2,2}\right), \cdots, \left(Z_{k,1}, Z_{k,2}\right)\right\}$ for biplane setup. For astigmatism-based setup, we obtained k average images and their axial positions.
2. PSF stack re-arrangement. For biplane setup, the $2k$ averaged PSFs were re-arranged into an axial range from –1.4 to +1.4 μm. The PSFs with the same axial positions in two planes were merged together. Thus a 3D PSF stack was formed, which was usually from –1 to +1 μm with a step size of 100–200 nm. For astigmatism-based setup, this step was ignored.
3. Pupil generation. The 3D PSF stack was used to generate the *in situ* 3D PSF model by phase retrieval method[2,3], which is based on Gerchberg-Saxon algorithm and outputs a pupil function to generate the retrieved PSFs within an axial range of ~2 μm. The phase retrieval process was carried out with a stack of averaged PSFs, their corresponding axial positions, and system parameters including the numerical aperture of the objective lens, the emission wavelength, the refractive index of the objective immersion medium, and the pixel size on the detection plane. Here we used two update modes. (1) Z_shift_mode

= 'shift'. In this mode, phase retrieval was carried out three times iteratively. For each time, phase retrieval found x tilt, y tilt, and defocus aberrations from decomposed Zernike modes and compensated these aberrations by shifting the lateral and axial positions of averaged PSFs. We used this mode in simulations, imaging experiments, and distorted wavefront control for low-order aberrations (from vertical astigmatism to secondary diagonal astigmatism except for the spherical aberration). (2) Z_shift_mode = 'no shift'. In this mode the positions were not updated. We used this mode in cases where wavefront distortions were significantly large, since the induced phase wrapping could make Zernike decomposition unreliable.

In fact, phase retrieval is not the only way to estimate the 3D model in INSPR. Any model-estimation tools developed for single-molecule localization, such as feature-based mapping[14,35], interpolation[7,36], and deep learning[37,38], can be utilized to build the 3D model in generalized INSPR.

2.7. 3D single-molecule localization using INSPR model transformation

INSPR models the 3D PSFs through the pupil function (as described in '**PSF generation**' section). For the biplane configuration, the PSFs in each plane can be described as

$$\begin{cases} \mu_1(x,y,z) = I_1 \cdot \mu_0(x,y,z) + bg_1 \\ \mu_2(x,y,z) = I_2 \cdot \mu_0(x,y,z+\delta) + bg_2 \end{cases}, \tag{16}$$

where μ_1 and μ_2 represent the PSF models in two planes, $\mu_0(x,y,z)$ and $\mu_0(x,y,z+\delta)$ are normalized PSFs generated by the pupil function at positions (x,y,z) and $(x,y,z+\delta)$, δ is the biplane distance, I_1 and I_2 are the total photon counts, and bg_1 and bg_2 are the background counts.

The data collected by using the sCMOS camera come with statistical properties including Poisson noise and pixel-dependent readout noise[18]. If we directly transform and interpolate data between two detection planes in 3D localization (*i.e.*, treating plane 1 as reference and transforming the data from plane 2 to plane 1, as described in '**INSPR framework**' section), the noise distribution will no longer maintain these statistical properties, resulting in imaging artifacts and localization imprecisions (**Extended Data Fig. 2e,f**). To solve this problem, we generated a channel-specific PSF model (*i.e.*, transforming the model instead of transforming the data) for single-molecule localization.

First, we carried out segmentation for the raw data in plane 2 as follows. (1) The center position (X_1, Y_1) of a cropped sub-region in plane 1 was recorded as described in '**INSPR framework**' section. (2) (X_1, Y_1) was transformed by affine transformation to find its corresponding position (X_2, Y_2) in plane 2. (3) A sub-region of the raw data in plane 2 was cropped with an integer center $(X_{2\text{int}}, Y_{2\text{int}}) = floor(X_2, Y_2)$, and the

non-integer offset was calculated as $(\Delta x, \Delta y) = (X_2, Y_2) - (X_{2\text{int}}, Y_{2\text{int}})$. The noise calibration map (including offset, variance, and gain for each pixel) of the sCMOS camera for each sub-region was cropped in a similar way.

For one single molecule, the relationship between its positions in plane 1 and plane 2 (**Extended Data Fig. 1b**) can be expressed as

$$(X_{2\text{int}} + x', Y_{2\text{int}} + y', 1) = (X_1 + x, Y_1 + y, 1) \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e & f & 1 \end{bmatrix}, \tag{17}$$

where (x, y) and (x', y') are the positions of the single molecule in the cropped sub-regions of two planes. $\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e & f & 1 \end{bmatrix}$ is the matrix of affine transformation[23] including six parameters, where $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ represents scale, shear, and rotation operations, and (e, f) represents the translation operation. Affine transformation is a linear transformation and can be written as

$$\begin{cases} X_{2\text{int}} + x' = a(X_1 + x) + c(Y_1 + y) + e = aX_1 + cY_1 + e + ax + cy \\ Y_{2\text{int}} + y' = b(X_1 + x) + d(Y_1 + y) + f = bX_1 + dY_1 + f + bx + dy \end{cases}. \tag{18}$$

Since the center position of the cropped sub-region was transformed from plane 1 to plane 2 by the same affine matrix

$$\begin{cases} X_2 = aX_1 + cY_1 + e = X_{2\text{int}} + \Delta x \\ Y_2 = bX_1 + dY_1 + f = X_{2\text{int}} + \Delta y \end{cases}, \tag{19}$$

by combining Eqs. (18) and (19), the relationship between (x, y) and (x', y'), the positions in cropped sub-regions, can be described as

$$\begin{cases} x' = ax + cy + \Delta x \\ y' = bx + dy + \Delta y \end{cases}, \tag{20}$$

showing that the raw single-molecule data and the cropped sub-region share the same scale, shear, and rotation parameters $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ between two planes, except for the translation parameters (e, f).

Second, to generate the channel-specific PSF model in plane 2, the model should satisfy two conditions. First, it shares the same shape information (scale, shear, and rotation) with the cropped sub-regions. Second, it has the same center position before and after transformation. Therefore, the affine transformation applied to the model $\text{Affine}_c = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e' & f' & 1 \end{bmatrix}$ should satisfy

$$(X_c, Y_c, 1) = (X_c, Y_c, 1)\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ e' & f' & 1 \end{bmatrix}, \tag{21}$$

where (X_c, Y_c) is the center position of the cropped sub-region (when performing affine transformation, the upper left corner of the image is defined as the origin, and X_c is equal to half of the sub-region size). The translation parameters (e', f') in Affine_c can be calculated as $\begin{cases} e' = (1-a)X_c - cY_c \\ f' = (1-d)Y_c - bX_c \end{cases}$. Thus, the channel-specific PSF model in plane 2 can be described as

$$\mu_2'(x', y', z) = \text{Trans}\{\text{Affine}_c[\mu_2(0,0,z)], (x', y')\}, \tag{22}$$

where Trans is the translation operation and $\mu_2(0,0,z)$ represents the PSF model of plane 2 at position $(0,0,z)$. The model $\mu_2(0,0,z)$ was first transformed to the channel-specific model $\mu_2{}'(0,0,z)$ by affine transformation Affine_c, and then translated to the position (x', y') given by Eq. (20).

Third, we directly incorporated the channel-specific PSF model inside the maximum likelihood estimator (MLE)[10] to estimate seven parameters $(x, y, z, I_1, I_2, bg_1, bg_2)$ by considering the Poisson noise and pixel-dependent sCMOS noise as

$$L(\theta|D) = \prod_q \frac{(\mu_{1q} + \gamma_{1q})^{(D_{1q}+\gamma_{1q})} e^{-(\mu_{1q}+\gamma_{1q})}}{\Gamma(D_{1q} + \gamma_{1q} + 1)} \prod_q \frac{(\mu_{2q}' + \gamma_{2q})^{(D_{2q}+\gamma_{2q})} e^{-(\mu_{2q}'+\gamma_{2q})}}{\Gamma(D_{2q} + \gamma_{2q} + 1)}, \tag{23}$$

$$\theta \in (x, y, z, I_1, I_2, bg_1, bg_2),$$

where D is the cropped sub-region of two planes, q is the pixel index, μ_1 and $\mu_2{}'$ represent the PSF models in planes 1 and 2, respectively. $\gamma_{mq} = \frac{var_{mq}}{g_{mq}^2}$, where var_{mq} and g_{mq} are the variance and gain for pixel q in plane m ($m = 1,2$). θ denotes the fitting parameters including the same position (x, y, z), and different total photon counts (I_1, I_2) and background counts (bg_1, bg_2) for two planes. A modified Levenberg-Marquadt method[39] was used to optimize θ by minimizing the negative log-likelihood function

$$-\ln(L) = \sum_q \mu_{1q} - (D_{1q} + \gamma_{1q})\ln(\mu_{1q} + \gamma_{1q}) + \sum_q \mu_{2q}' - (D_{2q} + \gamma_{2q})\ln(\mu_{2q}' + \gamma_{2q}). \tag{24}$$

The first and second derivatives are

$$f = -\frac{\partial \ln(L)}{\partial \theta} = \sum_q \left(1 - \frac{D_{1q} + \gamma_{1q}}{\mu_{1q} + \gamma_{1q}}\right)\frac{\partial \mu_{1q}}{\partial \theta} + \sum_q \left(1 - \frac{D_{2q} + \gamma_{2q}}{\mu_{2q}' + \gamma_{2q}}\right)\frac{\partial \mu_{2q}'}{\partial \theta}, \tag{25}$$

$$f' = \frac{\partial f}{\partial \theta} = \sum_q \frac{D_{1q} + \gamma_{1q}}{\left(\mu_{1q} + \gamma_{1q}\right)^2}\left(\frac{\partial \mu_{1q}}{\partial \theta}\right)^2 + \left(1 - \frac{D_{1q} + \gamma_{1q}}{\mu_{1q} + \gamma_{1q}}\right)\frac{\partial^2 \mu_{1q}}{\partial \theta^2} + \sum_q \frac{D_{2q} + \gamma_{2q}}{\left(\mu'_{2q} + \gamma_{2q}\right)^2}\left(\frac{\partial \mu'_{2q}}{\partial \theta}\right)^2 + \left(1 - \frac{D_{2q} + \gamma_{2q}}{\mu'_{2q} + \gamma_{2q}}\right)\frac{\partial^2 \mu'_{2q}}{\partial \theta^2}, \quad (26)$$

where the second derivatives $\frac{\partial^2 \mu_{1q}}{\partial \theta^2}$ and $\frac{\partial^2 \mu'_{2q}}{\partial \theta^2}$ were set to 0, and the fitting parameters were updated from

$$\theta_{n+1} = \theta_n - \frac{f}{f'(1+\beta)}, \quad (27)$$

where β is a damping factor to adjust the convergence speed, and was set to 0 here.

For the astigmatism-based setup, we directly incorporated the PSF model in one plane inside the MLE to estimate five parameters (x, y, z, I_1, bg_1).

The localization speed in INSPR mainly depends on the speed of fitting the parameters in the 3D PSF model generated by the Fourier transform of the retrieved pupil function. To speed up this process, we used the cubic interpolation methods[40] to pre-generate 3D PSF models along the axial direction for each voxel of 0.25 pixel × 0.25 pixel × 4 nm in the whole range of 25 pixels × 25 pixels × 2.6 μm. The GPU implementation of cubic interpolation achieves a localization speed of 240 PSFs per second, ~400 times faster than the CPU implementation using MATLAB. The code was tested on a computer with an Intel Core i7-8700K processor at 3.70 GHz with 32 GB memory and an NVIDIA GeForce GTX 1070 graphics card with 8.0 GB memory.

2.8. Calculation of Cramér-Rao lower bound

To quantify the Fisher information content of detected PSFs in INSPR, the CRLB[10] for estimating localization precision in an unbiased estimator was calculated as

$$var(\theta_i) \geq [F(\theta)^{-1}]_{ii}, \quad (28)$$

where $var(\theta)$ is the estimation variance of an estimator, $F(\theta)$ is the Fisher information matrix, θ is the vector of estimation parameters, and i denotes the index of each parameter.

For biplane setup, by incorporating the noise characteristic (Poisson noise and pixel-independent readout noise) of the sCMOS camera and the channel-specific PSF model, the relevant Fisher information in each element can be calculated as

$$F_{ij}(\theta) = \sum_q \frac{1}{\mu_{1q} + \gamma_{1q}} \frac{\partial \mu_{1q}}{\partial \theta_i} \frac{\partial \mu_{1q}}{\partial \theta_j} + \sum_q \frac{1}{\mu'_{2q} + \gamma_{2q}} \frac{\partial \mu'_{2q}}{\partial \theta_i} \frac{\partial \mu'_{2q}}{\partial \theta_j}, \tag{29}$$

$$\theta \in (x, y, z, I_1, I_2, bg_1, bg_2),$$

where μ_1 and μ'_2 represent the PSF models in planes 1 and 2, respectively; γ is the noise characteristic of the sCMOS camera, and q is the pixel index. For astigmatism-based setup, the Fisher information in each element was calculated with parameters $\theta \in (x, y, z, I_1, bg_1)$.

Furthermore, the Fisher information in the x and y dimensions in biplane setup was changed by considering the position relationship in the channel-specific PSF model $\begin{cases} x' = ax + cy + \Delta x \\ y' = bx + dy + \Delta y \end{cases}$ (as described by Eq. (20) in '**3D single-molecule localization**' section), where (x, y) and (x', y') are the positions of the PSF model of two planes, $(\Delta x, \Delta y)$ is the non-integer offset in plane 2, and $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ represents scale, shear, and rotation operations in affine transformation.

By calculating the derivative of the x dimension in plane 2

$$\frac{\partial \mu'_{2q}}{\partial x} = \frac{\partial \mu'_{2q}}{\partial x'} \frac{\partial x'}{\partial x} + \frac{\partial \mu'_{2q}}{\partial y'} \frac{\partial y'}{\partial x} = a \frac{\partial \mu'_{2q}}{\partial x'} + b \frac{\partial \mu'_{2q}}{\partial y'}, \tag{30}$$

the Fisher information in the x dimension can be written as

$$\begin{aligned} F_{xx} &= \sum_q \frac{1}{\mu_{1q} + \gamma_{1q}} \left(\frac{\partial \mu_{1q}}{\partial x} \right)^2 + \sum_q \frac{1}{\mu'_{2q} + \gamma_{2q}} \left(\frac{\partial \mu'_{2q}}{\partial x} \right)^2 \\ &= \sum_q \frac{1}{\mu_{1q} + \gamma_{1q}} \left(\frac{\partial \mu_{1q}}{\partial x} \right)^2 + \sum_q \frac{1}{\mu'_{2q} + \gamma_{2q}} \left(\frac{\partial \mu'_{2q}}{\partial x'} \frac{\partial x'}{\partial x} + \frac{\partial \mu'_{2q}}{\partial y'} \frac{\partial y'}{\partial x} \right)^2 \\ &= \sum_q \frac{1}{\mu_{1q} + \gamma_{1q}} \left(\frac{\partial \mu_{1q}}{\partial x} \right)^2 + \sum_q \frac{1}{\mu'_{2q} + \gamma_{2q}} \left(a \frac{\partial \mu'_{2q}}{\partial x'} + b \frac{\partial \mu'_{2q}}{\partial y'} \right)^2. \end{aligned} \tag{31}$$

Similarly, the Fisher information in the y dimension can be written as

$$F_{yy} = \sum_q \frac{1}{\mu_{1q} + \gamma_{1q}} \left(\frac{\partial \mu_{1q}}{\partial y} \right)^2 + \sum_q \frac{1}{\mu'_{2q} + \gamma_{2q}} \left(c \frac{\partial \mu'_{2q}}{\partial x'} + d \frac{\partial \mu'_{2q}}{\partial y'} \right)^2. \tag{32}$$

2.9. Rejection methods

CRLB is one criterion to measure the localization uncertainty with retrieved 3D PSF model given by the position (x, y, z), photon count I, and background count bg. Here we focused on the localization uncertainty in the z dimension ($\sqrt{\mathrm{CRLB_z}}$). In order to improve the quality of reconstructed images, localizations with $\sqrt{\mathrm{CRLB_z}}$ larger than a certain threshold were rejected. We set this threshold to 30 nm for the dataset in **Fig. 3f** (TOM20), 35 nm for the datasets in **Extended Data Figs. 4, 5** (TOM20), 40 nm for the datasets in **Fig. 6a** and **Supplementary Fig. 9** (ChR2-EYFP) and the dataset in **Supplementary Fig. 11a** (α-tubulin), 45 nm for the datasets in **Fig. 5a** (low-density amyloid β plaque), 50 nm for the datasets in **Fig. 4** (Nup98), **Fig. 5k** (high-density amyloid β plaque), **Extended Data Fig. 9** (ChR2-EYFP), and **Supplementary Fig. 11e** (TOM20 and α-tubulin), and 70 nm for the dataset in **Fig. 6h** (elastic fibers).

Log-likelihood ratio (LLR) is a goodness of fitting which reflects the similarity between each detected PSF and its corresponding PSF model after fitting, and can be expressed as

$$\begin{aligned}\mathrm{LLR} &= -2\ln\left(\frac{L(\mu|D)}{L(D|D)}\right)\\ &= \sum_q 2\left[\mu_{1q} - D_{1q} - (D_{1q}+\gamma_{1q})\ln(\mu_{1q}+\gamma_{1q}) + (D_{1q}+\gamma_{1q})\ln(D_{1q}+\gamma_{1q})\right]\\ &+ \sum_q 2\left[\mu'_{2q} - D_{2q} - (D_{2q}+\gamma_{2q})\ln(\mu'_{2q}+\gamma_{2q}) + (D_{2q}+\gamma_{2q})\ln(D_{2q}+\gamma_{2q})\right],\end{aligned} \tag{33}$$

where D is the cropped sub-region of single molecule, μ is the PSF model, γ is the noise characteristic of the sCMOS camera, and q is the pixel index. Lower LLR means higher similarity between each detected PSF and its corresponding PSF model. Here we set LLR to 1000 for each 16 × 16 pixels in the single-molecule dataset in biplane setup, and LLR to 600 in astigmatism-based setup. LLR shows high similarity with the detected emission patterns in camera frames in contrast to the *in vitro* model (**Supplementary Fig. 13**).

Besides, we rejected single molecules more than 800 nm out of focus in biplane setup, and 500 nm out of focus in astigmatism-based setup. Single molecules with total photon count lower than 1000 were rejected for both biplane and astigmatism-based setups in our reconstructions.

2.10. Drift correction, optical section alignment, and Exchange-PAINT alignment

To image thick samples, optical sections were recorded as described in **'Data acquisition'** section. The drift correction and optical section alignment were carried out according to a previously described method[1]. In each optical section, drift was calibrated by calculating the correlation between each 3D volume consisting of localized single molecules from 1000 frames using a redundancy-based correction method[16,41,42]. These calibrated 3D volumes formed an ~1.6-μm-thick optical section. Whole cells or tissue specimens were scanned axially by translating the objective lens with a step size of 250 nm for astigmatism-based setup and 400 nm for biplane setup, which ensured enough overlapped regions between adjacent optical sections. We performed the 3D-correlation-based method[16] to align two adjacent optical sections, and finally reconstructed a super-resolution 3D volume.

Exchange-PAINT imaging was performed sequentially, first mitochondria and then microtubules. Before fluorescence imaging, we recorded a series of bright-field images of the sample over an axial range from –1 to +1 μm with a step size of 100 nm as reference images for both focus stabilization (as described in **'Data acquisition'** section) and Exchange-PAINT alignment. After calibrating the drift in each imaging process, we used these reference images to align the calibrated results from mitochondria to microtubules by performing the 3D-correlation-based method[16].

2.11. Supercritical angle fluorescence and polarization effects on INSPR

When using a high-NA-oil-immersion objective (*e.g.* when $NA \geq 1.4$) to image into a watery environment, supercritical angle fluorescence (SAF) effect[43,44] is enabled close to the coverslip-water interface, resulting that the evanescent wave that originates from fluorophores can penetrate the interface and become a propagating wave at an angle larger than the critical angle. This SAF effect could also be retrieved by using phase retrieval[3]. However, to retrieve this SAF pupil, fluorescent beads need to be attached on the coverslips and moved together with the coverslip at different axial positions for phase retrieval. When imaging fluorescent-labeled cellular structures, single molecules with different axial positions locate at different depths from the coverslip, and therefore, the strength of SAF or the SAF pupil will differ drastically (due to the sensitivity of the SAF effect with the distance) for different molecules. In this case, using a single pupil within the near interface region may not be suitable. To resolve this, the algorithm needs to retrieve the varying SAF pupils at different depths from the coverslip to reconstruct the 3D SAF-PSF.

In this work, we assume that the fluorophore is freely rotating with a frequency much higher than the camera frame rate, and therefore, the optical transfer function (OTF) of the detected fluorophore is radial-

symmetrically modified. In 3D PSF model estimation using INSPR, we account for this effect through a process named OTF rescaling[2,3].

In cases where single molecules are considered dipoles with fixed orientations, it has been shown that using a more sophisticated PSF model allows to take these orientations into account[45,46], or adding an azimuthal polarization filter located at the back focal plane can remove localization bias[47]. The possibility of co-estimating both location and orientation of the single molecules is of great interest and can be quite complicated in presence of aberration due to the high dimensionality of the PSF model.

3. Quantification analysis

3.1. Simulation analysis

We tested the performance of INSPR in different simulation conditions by generating PSFs with random aberrations in biplane setup (**Fig. 2a–d, Supplementary Video 1**), PSFs at different imaging depths above the coverslip surface (**Extended Data Fig. 2a,b**), PSFs with different SBR conditions (**Extended Data Fig. 2c**), channel-specific PSFs (**Extended Data Fig. 2e,f**), impact of biplane distance bias (**Supplementary Fig. 12c–e**), PSFs with random aberrations in astigmatism-based setup (**Supplementary Fig. 6, Supplementary Video 2**), and 3D training datasets of microtubules (MT0.N1.LD) from the SMLM challenge[48] (**Fig. 3a–e, Extended Data Fig. 3**).

For simulations in **Fig. 2a–d** and **Supplementary Video 1** (random aberrated PSFs in biplane setup), we decomposed the INSPR retrieved pupil into 21 Zernike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration) and recorded their amplitudes. For each trial, the phase and Zernike amplitude errors were defined as the RMSE between the INSPR retrieved pupil and the ground truth pupil. We then calculated their mean and standard deviation in the total 30 trials.

For simulations in **Extended Data Fig. 2a,b** (PSFs at different imaging depths above the coverslip surface), the defocus offset (*i.e.*, the axial shift from the actual focal plane) was estimated as follows. First, we retrieved the 3D PSF in a range from –1.5 to +1.5 μm with a step size of 20 nm. Then, we plotted the PSF intensity curve at the center of the lateral position along the axial direction and recorded the axial position of the maximum intensity. After that, we calculated the axial offset between the plane with the maximum intensity and the actual focal plane. In the actual imaging process, we only collected single molecules at the axial plane with the maximum intensity and treated this plane as the focal plane. To simulate this process, we generated 2000 PSFs located randomly with an axial range from –800 to +800 nm but with a certain defocus offset at each imaging depth, and retrieved *in situ* pupils using INSPR. For each imaging depth, we

first generated the 3D PSF by our previously estimated optical aberrations[1] with an axial range from –800 to +800 nm and a step size of 100 nm as ground truth. Then we calculated the 3D normalized cross correlations (NCCs) between the ground truth PSFs and the PSFs retrieved using the Gaussian model, theoretical index mismatch model[2,3], and INSPR for different axial offsets. Finally, we recorded the maximum NCC values and their corresponding 3D PSF at different imaging depths.

For simulation in **Extended Data Fig. 2c** (different SBR conditions), the amplitude error of Zernike modes was calculated at different photon and background conditions (11 trials in each condition). In each condition, we calculated the mean and standard deviation of the amplitude error in the total 11 trials for different numbers of single molecules. After calculating all conditions, we obtained the convergence curves.

For simulations in **Extended Data Fig. 2e,f** (channel-specific PSFs), the $\sqrt{\mathrm{CRLB}}$ in the x, y, and z dimensions was calculated as described in '**Calculation of Cramér-Rao lower bound**' section. In the measurement system, the achievable estimation precision is limited by CRLB for an unbiased estimator. The precision in each dimension at a certain position (ranging from –500 to +500 nm with a step size of 100 nm) was calculated from the standard deviation of the estimated positions of 1000 PSFs. The bias in each dimension was calculated as the difference between the averaged position of 1000 PSFs and the ground truth position.

For simulations in **Supplementary Fig. 12c–e** (impact of a biased biplane distance input), we first used random wavefront shapes (30 trials in total) to generate PSFs located randomly in an axial range of ±800 nm with a biplane distance of 580 nm (2000 PSFs per trial). Subsequently, we used INSPR to retrieve the corresponding pupils from the generated PSFs with biased biplane distance inputs (from –20% to 20%), and compared their decomposed Zernike amplitudes with the ground truth amplitudes (**Supplementary Fig. 12d**). Then, for each trial (a single wavefront), we used the retrieved pupils to localize the PSFs generated from the ground truth wavefront at fixed axial positions from –500 to +500 nm with a step size of 100 nm (1000 single molecules per step). For each axial step, we calculated the mean axial position of 1000 single molecules to obtain the localization bias for a single trial. Subsequently, the mean and standard deviation of the resulting localization biases in the total 30 trials were reported (**Supplementary Fig. 12e**).

For simulations in **Supplementary Fig. 6** and **Supplementary Video 2** (randomly aberrated PSFs in astigmatism-based setup), the INSPR retrieved pupil was decomposed into 21 Zernike modes (Wyant order, from vertical astigmatism to tertiary spherical aberration) and the amplitudes in all modes were recorded. Then we calculated the mean and standard deviation of the phase and Zernike amplitude errors in the total 30 trials.

For 3D training datasets of microtubules from the SMLM challenge[48] (**Fig. 3a–e, Extended Data Fig. 3**), the system parameters including the numerical aperture of the objective lens, the emission wavelength, the refractive indices of the objective immersion and imaging media, and the effective pixel size, offset, and gain on the camera, were obtained from the SMLM challenge. We segmented the datasets into sub-regions to construct the PSF library, estimate the pupil, and localize their positions. We set initial intensity threshold I_{init} to 25, segmentation threshold I_{seg} to 40, similarity threshold $\mathrm{Sim}_{\mathrm{min}}$ to 0.5, and group threshold N_g to 30. For the biplane dataset, we set the sub-region size to 32 × 32 pixels, distance threshold d_{thresh} to 18, and the biplane distance to 500 nm, as well as shifting sub-regions in two planes together. The templates were generated at axial positions from –1 to +1 μm with a step size of 62.5 nm. For the astigmatism-based dataset, we set the sub-region size to 26 × 26 pixels and d_{thresh} to 14, in order to get more sub-regions of PSFs. The templates were generated at axial positions from –800 to +800 nm with a step size of 60 nm. We set the amplitude of vertical astigmatism to –1.5 (unit, $\lambda/2\pi$) as initial guess. For 3D single-molecule localization, we set the sub-region size to 16 × 16 pixels and d_{thresh} to 6. We observed that INSPR allowed blind reconstruction for both biplane and astigmatism datasets without calibration PSFs. We also observed the reconstruction from the biplane dataset appeared to be closer to the ground truth than that from the astigmatism-based dataset. This is likely caused by the larger axial range in the biplane dataset, which made phase retrieval more accurate. The 3D PSFs of the z-stack beads for both biplane and astigmatism-based setups obtained from the calibration files in the SMLM challenge were compared with INSPR retrieved ones in **Fig. 3e** and **Extended Data Fig. 3g,n**.

3.2. Distorted wavefront verification

We tested the accuracy of distorted wavefront estimation from single-molecule blinking datasets (**Fig. 2e,f, Supplementary Video 3**). The deformable mirror was calibrated to introduce individual Zernike-based aberration modes (Wyant order, from vertical astigmatism to tertiary spherical aberration, 21 modes in total) as described in '**Calibration of deformable mirror**' section. For each Zernike mode, the single-molecule emission patterns were distorted by the introduced Zernike-based aberrations (amplitudes at ±1, unit, $\lambda/2\pi$) and fed into INSPR after acquisition to retrieve the *in situ* PSF and its corresponding pupil function (parameters used for INSPR are shown in **Supplementary Table 2**). We decomposed the retrieved pupil functions into 21 Zernike modes, and obtained their amplitudes. To eliminate the influence of instrument imperfections which introduces an offset to the retrieved amplitude, we calculated the difference between the retrieved amplitudes when the amplitude of the input Zernike mode is equal to +1 and –1 (unit, $\lambda/2\pi$), and divided this difference by 2. After processing 21 Zernike modes, we built a heat map representing the relationship between the input and output amplitudes of Zernike modes. The estimation error between using

INSPR and the *in vitro* method for each Zernike mode was carried out by calculating the difference of the amplitudes at the diagonal elements of the heat map. Besides, we calculated the RMSE between the 21 *in situ* retrieved amplitudes and the 21 *in vitro* retrieved amplitudes for each Zernike mode, and got an average RMSE of 23 mλ for the total 21 Zernike modes.

3.3. 3D PSF quantification of 3D super-resolution reconstructions

The x-y and x-z views of 3D PSFs retrieved from the datasets in **Fig. 3r** and **Extended Data Fig. 6c** (TOM20), **Supplementary Fig. 2** (Nup98), **Supplementary Fig. 3** (amyloid β), **Supplementary Fig. 4a** (ChR2-EYFP), and **Supplementary Fig. 4b** (elastic fibers) were generated by the retrieved pupil functions (as described in '**PSF generation**' section). We first retrieved a 3D PSF by using the *in vitro* method (phase retrieval based on fluorescent beads on the coverslip) with an axial range from –800 to +800 nm and a step size of 100 nm as reference. Then, in each optical section, we generated a series of *in situ* PSFs in the same axial range, but with a defocus offset. We set the offset values from –500 to +500 nm away from the focal plane with a step size of 50 nm. By calculating the 3D cross correlation, we found the best matched 3D PSF and assigned this 3D PSF to the corresponding optical section. The magnitude of the retrieved pupil function $A(k_x, k_y)$ in each optical section is normalized so that $\iint \left|A(k_x, k_y)\right|^2 \mathrm{d}k_x \, \mathrm{d}k_y = 1$, and the size of the pupil function is limited by ${k_x}^2 + {k_y}^2 \leq \left(\frac{NA}{\lambda}\right)^2$, where NA is the numerical aperture of the objective lens and λ is the emission wavelength in air.

3.4. Profile quantification of 3D super-resolution reconstructions

Intensity profiles in reconstruction images were measured with the line profile tool in ImageJ[49], and then fitted with Gaussian functions. To obtain reliable profiles, for the datasets in **Fig. 4** and **Extended Data Fig. 7** (Nup98), we used a pixel size of 6 nm and a standard deviation of 1 pixel (Gaussian blur) in reconstruction images. For other datasets, we used a pixel size of 12 nm and a Gaussian blur of 1 pixel in reconstruction images.

For the datasets in **Fig. 3f–s** and **Extended Data Fig. 6** (TOM20), we generated reconstruction images of 25 outer membrane structures in the y'-z plane, and obtained their intensity profiles along both y' and z directions. Here the orientation of the cross section was rotated to allow projection of the 3D membrane bounded structures to the 2D image. For the datasets in **Fig. 4** and **Extended Data Fig. 7** (Nup98), we got intensity profiles of 40 Nup98 structures in the x-y plane and 20 Nup98 structures in the x-z plane. For the

datasets in **Fig. 5** and **Supplementary Fig. 14a–d** (amyloid β), we got intensity profiles of 40 fibrils in both x-y and x-z planes. For the dataset in **Fig. 6h** and **Supplementary Fig. 14e,f** (elastic fibers), we got intensity profiles of 15 long elastic fibers (3–5 measurements for each, 53 measurements in total) and 7 short fibers (single measurement for each) in the x-y plane and 40 fibers in the x-z plane.

Intensity profiles of immunofluorescence-labeled TOM20 in COS-7 cells in the y'-z plane (along the y' or z direction) and Nup98 in COS-7 cells in the x-y plane (along the y direction) were fitted with a linear combination of two Gaussian functions $f(x) = a_1 e^{-\frac{(x-\mu_1)^2}{2{\sigma_1}^2}} + a_2 e^{-\frac{(x-\mu_2)^2}{2{\sigma_2}^2}} + b$, where $(a_1, a_2, \mu_1, \mu_2, \sigma_1, \sigma_2, b)$ are fitting parameters, μ_1 and μ_2 are the positions of the centers of the peaks, $|\mu_2 - \mu_1|$ represents the width of the contour along the fitting direction, σ_1 and σ_2 are the standard deviations representing the widths of the structure boundary.

Intensity profiles of immunofluorescence-labeled Nup98 in COS-7 cells in the x-z plane (along the z direction), and amyloid β plaques and elastic fibers in both x-y and x-z planes (along the direction perpendicular to the fibers) were fitted with a Gaussian function $f(x) = ae^{-\frac{(x-\mu)^2}{2\sigma^2}} + b$, where (a, μ, σ, b) are fitting parameters, μ is the position of the center of the peak, σ is the standard deviation. The full width at half maximum (FWHM) of the Gaussian function is equal to $2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$.

3.5. Axial slice quantification of 3D super-resolution reconstructions

We compared the axial slices of images reconstructed by INSPR and *in vitro* approaches. For the dataset in **Extended Data Fig. 4** (TOM20), the images were reconstructed using INSPR, ZOLA-3D, and cubic spline from beads on the coverslip. For the datasets in **Extended Data Fig. 5** (TOM20), the images were reconstructed using INSPR and microsphere-calibrated Gaussian fitting. For the dataset in **Extended Data Fig. 7h,i** (Nup98), the thickness of the central cross section of the entire nuclear envelope was measured from reconstruction results using INSPR and PR. For the datasets in **Extended Data Fig. 8** and **Supplementary Fig. 9** (ChR2-EYFP), the images were reconstructed using INSPR, PR, and PR+IMM. For the datasets in **Extended Data Fig. 9** (ChR2-EYFP), the images were reconstructed using INSPR and PR in gel. For the dataset in **Extended Data Fig. 10** (elastic fibers), we measured the angles of individual elastic fibers in the x-z slices resolved by INSPR, PR, and PR+IMM.

Intensity profiles in these reconstruction images were also measured with the line profile tool in ImageJ. We used a pixel size of 12 nm and a Gaussian blur of 1 pixel in reconstruction images. We fitted intensity

profiles with a linear combination of two Gaussian functions, and then calculated the distance between the centers of the two peaks.

The diffraction limited images in axial slices (**Fig. 6c,f,g**) were generated by convolution between reconstructed images and *in situ* PSF models.

3.6. Axial thickness quantification of 3D super-resolution reconstructions

Inaccurate PSF models shrink or expand reconstructed cellular volume as shown in **Supplementary Fig. 1d**. Stacking multiple shrunk/expanded super-resolution volumes will result in either horizontal stripe artifacts in the axial cross section (when using a fixed distance between adjacent super-resolution volumes during stacking) or a shrunk axial range of the entire volume (when using 3D cross correlation for optical section alignment[16] – used in this work). Although it might be difficult to get the ground truth thickness of the imaged nuclear envelope, we can give a rough estimation as follows.

In biplane setup, the axial range of localization in one optical section is approximately ±0.8 μm (depending on the structure distribution), and the z-piezo stage movement between adjacent optical sections t_{sec} is 0.4 μm. When considering the refractive index mismatch, the rescaled thickness between adjacent optical sections t'_{sec} is equal to $t_{\mathrm{sec}} n_{\mathrm{w}}/n_{\mathrm{o}}$, where n_{w} and n_{o} are the refractive indices of the imaging medium and the objective immersion medium, respectively. In our experimental conditions, $n_{\mathrm{w}} = 1.352$ and $n_{\mathrm{o}} = 1.406$. Therefore, t'_{sec} can be estimated as $0.4 \times \frac{1.352}{1.406} = 0.385$ μm. If a dataset has N optical sections, the thickness is estimated as $0.8 + (N-1) \times t'_{\mathrm{sec}} + 0.8 = 1.6 + (N-1) \times 0.385$ μm.

The dataset in **Fig. 4e** (Nup98) has 14 optical sections, so the estimated axial thickness is $1.6 + (14-1) \times 0.385 = 6.6$ μm, which matches the thickness of the INSPR reconstructed structures 6.4 μm (**Supplementary Table 1**), rather than that of the *in vitro* reconstructed structures 5.3 μm. The dataset in **Fig. 6a** (ChR2-EYFP) has 7 optical sections, so the estimated axial thickness is $1.6 + (7-1) \times 0.385 = 3.91$ μm, which matches the thickness of the INSPR reconstructed structures 4.2 μm (**Supplementary Table 1**), rather than that of the *in vitro* reconstructed structures 3.4 μm (using PR/PR+IMM). The dataset in **Fig. 6h** (elastic fibers) has 5 optical sections, so the estimated axial thickness is $1.6 + (5-1) \times 0.385 = 3.1$ μm, which matches the thickness of the INSPR reconstructed structures 3.1 μm (**Supplementary Table 1**), rather than that of the *in vitro* reconstructed structures 2.3 μm (using PR) or 2.7 μm (using PR+IMM).

3.7. Resolution quantification of mitochondrial network in axial cross sections

We used Fourier ring correlation[50] (FRC) plugin tool in ImageJ to quantify the resolution of mitochondrial network in axial cross sections (**Fig. 3g–l, Supplementary Fig. 15a–h**). To increase the reliably of calculation, each dataset was calculated 20 times in random orders. Statistical results are shown in **Supplementary Table 3**, and typical FRC curves are shown in **Supplementary Fig. 15i–p**. Although we could visually verify the improvement using INSPR and the *in vitro* method, our FRC result from these cross-section reconstruction shows only slight improvement comparing INSPR to the *in vitro* method (**Supplementary Table 3**). This is likely caused by the requirement of FRC: isotropic resolution in both directions. The axial resolution in our cross-section images is typically 2–3 times poorer than lateral (**Supplementary Table 1**) and thus FRC might underestimate the improvement in the axial direction. As an example, the reconstructed y'-z section image shown in **Supplementary Fig. 15h** contains membrane contours that are sharper than **Supplementary Fig. 15g** in the axial direction, and their line profiles in the axial direction shown in **Fig. 3p** are measured to be ~30% thinner than those in **Fig. 3o**. However, when calculating FRC resolutions (**Supplementary Fig. 15o,p**), the difference is reported to be ~8%.

References

1. Mlodzianoski, M. J. et al. Active PSF shaping and adaptive optics enable volumetric localization microscopy through brain sections. *Nat. Methods* **15,** 583–586 (2018).
2. Hanser, B. M., Gustafsson, M. G. L., Agard, D. A. & Sedat, J. W. Phase-retrieved pupil functions in wide-field fluorescence microscopy. *J. Microsc.* **216,** 32–48 (2004).
3. Liu, S., Kromann, E. B., Krueger, W. D., Bewersdorf, J. & Lidke, K. A. Three dimensional single molecule localization using a phase retrieved pupil function. *Opt. Express* **21,** 29462–29487 (2013).
4. Douglass, K. M., Sieben, C., Archetti, A., Lambert, A. & Manley, S. Super-resolution imaging of multiple cells by optimized flat-field epi-illumination. *Nat. Photonics* **10,** 705–708 (2016).
5. Stehr, F., Stein, J., Schueder, F., Schwille, P. & Jungmann, R. Flat-top TIRF illumination boosts DNA-PAINT imaging and quantification. *Nat. Commun*. **10,** 1268 (2019).
6. Aristov, A., Lelandais, B., Rensen, E., & Zimmer, C. ZOLA-3D allows flexible 3D localization microscopy over an adjustable axial range. *Nat. Commun.* **9,** 2409 (2018).
7. Li, Y. et al. Real-time 3D single-molecule localization using experimental point spread functions. *Nat. Methods* **15,** 367–369 (2018).
8. Cabriel, C., Bourg, N., Dupuis, G. & Lévêque-Fort, S. Aberration-accounting calibration for 3D single-molecule localization microscopy. *Opt. Lett.* **43,** 174–177 (2018).
9. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319,** 810–813 (2008).
10. Smith, C. S., Joseph, N., Rieger, B. & Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods* **7,** 373–375 (2010).
11. Li, Y., Wu, Y.-L., Hoess, P., Mund, M. & Ries, J. Depth-dependent PSF calibration and aberration correction for 3D single-molecule localization. *Biomed. Opt. Express* **10,** 2708–2718 (2019).
12. Jungmann, R. et al. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11,** 313–318 (2014).
13. Edmondson, R., Broglie, J. J., Adcock, A. F. & Yang, L. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay Drug Dev. Technol*. **12,** 207–218 (2014).
14. Juette, M. F. et al. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nat. Methods* **5,** 527–529 (2008).
15. Hell, S., Reiner, G., Cremer, C. & Stelzer, E. H. K. Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *J. Microscopy* **169,** 391–405 (1993).
16. Huang, F. et al. Ultra-high resolution 3D imaging of whole cells. *Cell* **166,** 1028–1040 (2016).

17. Booth, M. J. Adaptive optical microscopy: the ongoing quest for a perfect image. *Light Sci. Appl*. **3,** e165 (2014).
18. Huang, F. et al. Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10,** 653–658 (2013).
19. McGorty, R., Kamiyama, D. & Huang, B. Active microscope stabilization in three dimensions using image correlation. *Opt. Nanoscopy* **2,** 3 (2013).
20. Goodman, J. W. *Introduction to Fourier optics* (Roberts & Company Publishers, Greenwood, Village, ed. 3, 2004).
21. Wyant, J. C. & Creath, K. "Basic wavefront aberration theory for optical metrology" in *Applied Optics and Optical Engineering* (Academic, New York, Vol. XI, 1–53, 1992).
22. Hartigan, J. A. & Wong, M. A. A *k*-means clustering algorithm. *J. R. Stat. Soc. Series C Appl. Stat.* **28,** 100–108 (1979).
23. da Silva, E. A. B. & Mendonca, G. V. "Digital image processing" in *Electrical Engineering Handbook* (Elsevier Academic Press, Boston, 891–910, 2005).
24. Booth, M. J., Andrade, D., Burke, D., Patton, B. & Zurauskas, M. Aberrations and adaptive optics in super-resolution microscopy. *Microscopy* **64,** 251–261 (2015).
25. Burke, D., Patton, B., Huang, F., Bewersdorf, J. & Booth, M. J. Adaptive optics correction of specimen-induced aberrations in single-molecule switching microscopy. *Optica* **2,** 177–185 (2015).
26. Ji, N. Adaptive optical fluorescence microscopy. *Nat. Methods* **14,** 374–380 (2017).
27. Gradinaru, V., Treweek, J., Overton, K. & Deisseroth, K. Hydrogel-tissue chemistry: principles and applications. *Annu. Rev. Biophys*. **47,** 355–376 (2018).
28. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347,** 543–548 (2015).
29. Chozinski, T. J. et al. Expansion microscopy with conventional antibodies and fluorescent proteins. *Nat. Methods* **13,** 485–488 (2016).
30. Chen, B.-C. et al. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* **346,** 1257998 (2014).
31. Power, R. M. & Huisken, J. A guide to light-sheet fluorescence microscopy for multiscale imaging. *Nat. Methods* **14,** 360–373 (2017).
32. Legant, W. R. et al. High-density three-dimensional localization microscopy across large volumes. *Nat. Methods* **13,** 359–365 (2016).
33. Pavani, S. R. P. et al. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Natl. Acad. Sci. USA.* **106,** 2995–2999 (2009).
34. Shechtman, Y., Sahl, S. J., Backer, A. S. & Moerner, W. E. Optimal point spread function design for 3D imaging. *Phys. Rev. Lett*. **113,** 133902 (2014).

35. Lew, M. D. et al. Three-dimensional superresolution colocalization of intracellular protein superstructures and the cell surface in live *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA*. **108,** 1102–1110 (2011).
36. Babcock, H. P. & Zhuang, X. Analyzing single molecule localization microscopy data using cubic splines. *Sci. Rep*. **7,** 552 (2017).
37. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5,** 458–464 (2018).
38. Zhang, P. et al. Analyzing complex single-molecule emission patterns with deep learning. *Nat. Methods* **15,** 913–916 (2018).
39. Laurence, T. A. & Chromy, B. A. Efficient maximum likelihood estimator fitting of histograms. *Nat. Methods* **7,** 338–339 (2010).
40. Liu, S. & Huang, F. Enhanced 4Pi single-molecule localization microscopy with coherent pupil based localization and light sheet illumination. Preprint at https://www.biorxiv.org/content/10.1101/586404v1 (2019).
41. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution singleparticle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).
42. Wang, Y. et al. Localization events-based sample drift correction for localization microscopy with redundant cross-correlation algorithm. *Opt. Express* **22,** 15982–15991 (2014).
43. Deschamps, J., Mund, M. & Ries, J. 3D superresolution microscopy by supercritical angle detection. *Opt. Express* **22,** 29081–29091 (2014).
44. Bourg, N. et al. Direct optical nanoscopy with axially localized detection. *Nat. Photonics* **9,** 587–593 (2015).
45. Mortensen, K. I., Churchman, L. S., Spudich, J. A. & Flyvbjerg, H. Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nat. Methods* **7,** 377–381 (2010).
46. Mortensen, K. I., Sung, J., Flyvbjerg, H. & Spudich, J. A. Optimized measurements of separations and angles between intra-molecular fluorescent markers. *Nat. Commun.* **6,** 8621 (2015).
47. Backlund, M. P. et al. Removing orientation-induced localization biases in single-molecule microscopy using a broadband metasurface mask. *Nat. Photonics* **10,** 459–462 (2016).
48. Sage, D. et al. Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nat. Methods* **16,** 387–395 (2019).
49. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
50. Nieuwenhuizen, R. P. J. et al. Measuring image resolution in optical nanoscopy. *Nat. Methods* **10,** 557–562 (2013).