Single-shot volumetric fluorescence imaging with neural fields: supplemental document

List of Figures

List of Notes

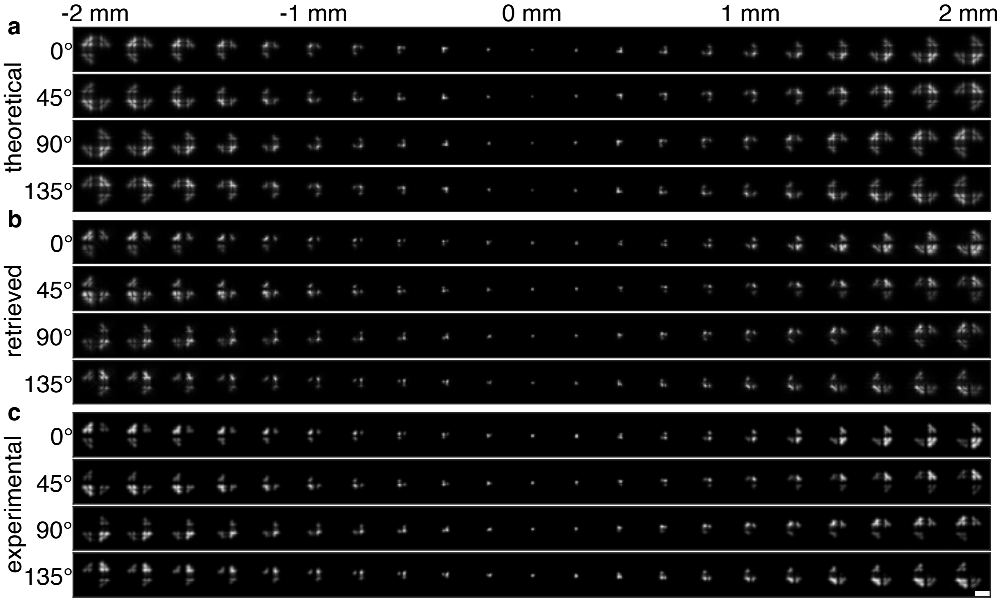

Fig S1 PSFs in each polarization channel at different height z. (a) Theoretical PSFs without aberration, (b) simulated PSFs using the retrieved phase, and (c) experimental PSFs. Scale bar: 0.1 mm.

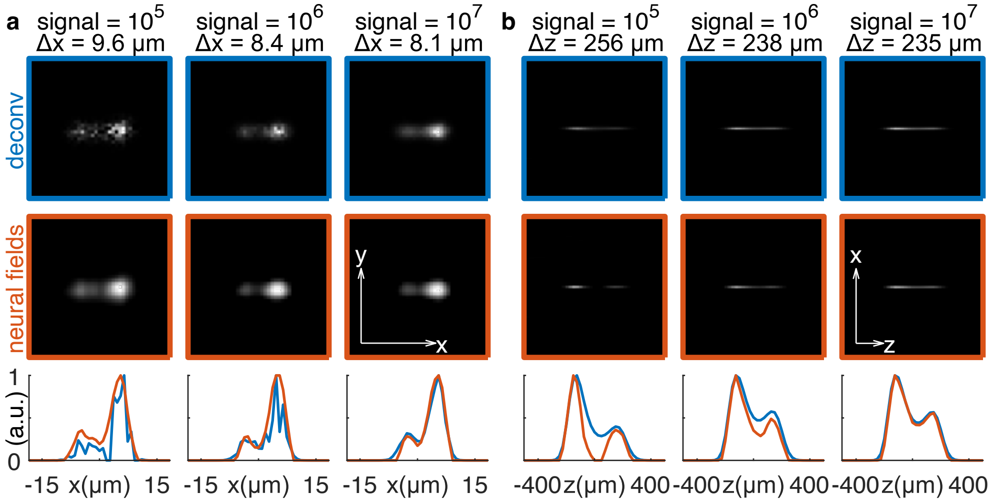

Fig S2 Cross sections and line profiles of representative reconstructions from simulated images of two point sources. Scale arrows: 20 μm in the xy cross sections and 200 μm in the xz cross sections.

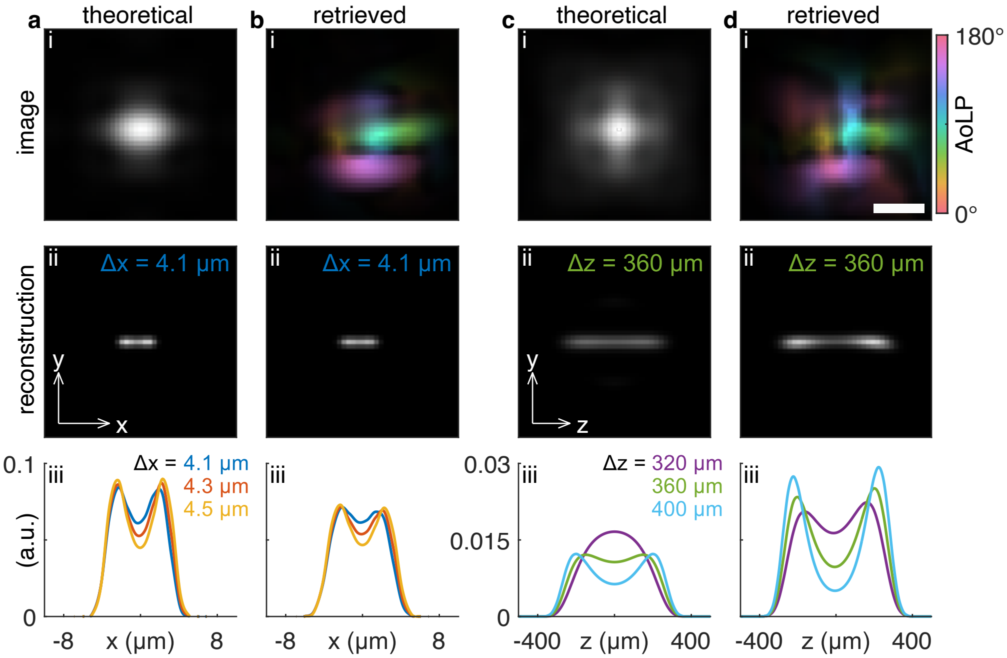

Fig S3 Evaluation of the (a,b) lateral and (c,d) axial resolutions using (a,c) theoretical and (b,d) retrieved QuadraPol PSFs. (i) Simulated raw images, where the retrieved PSFs exhibit more pronounced polarization features. (ii) Cross sections and (iv) line profiles of the reconstructions. Scale bar: 10 μm; scale arrows: 10 μm in x and y, and 200 μm in z. The reconstruction using retrieved PSF shows a slight degradation in lateral resolution but a significant improvement in axial resolution.

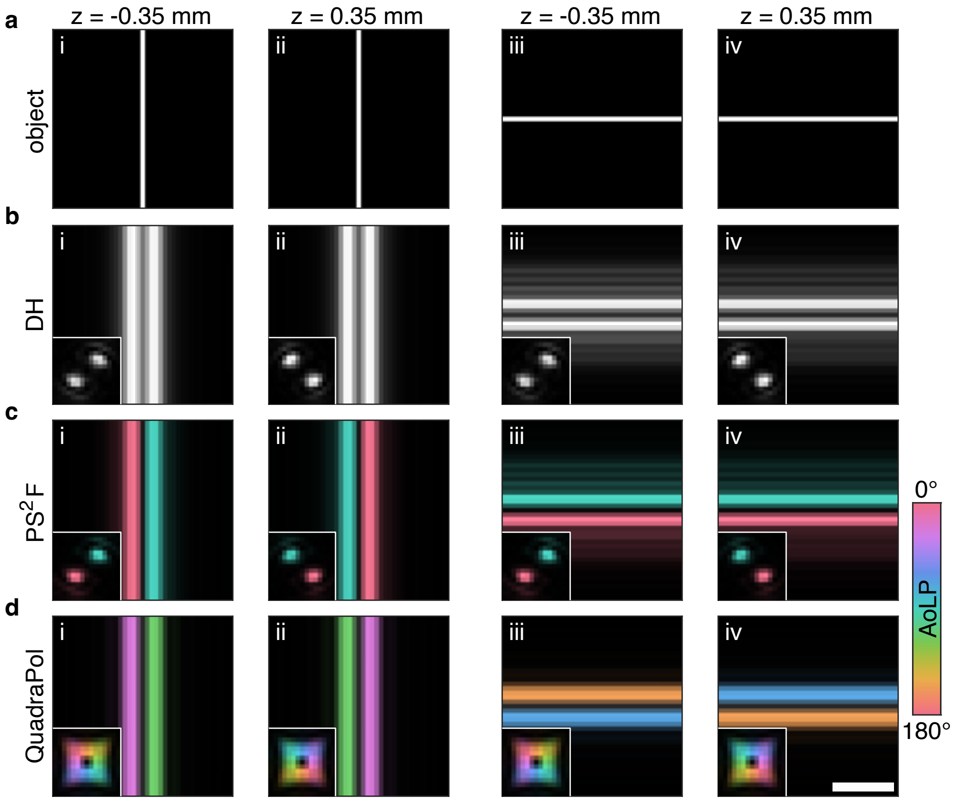

Fig S4 Estimation ambiguity evaluation using the double-helix, polarized spiral, and QuadraPol PSFs. Panels (a), (b), and (d) are identical to those presented in Figure 4 from the main text. (a) Simulated line objects; (i) vertical line at $z = -0.35$ mm; (ii) vertical line at $z = 0.35$ mm; (iii) horizontal line at $z = -0.35$ mm; (iv) horizontal line at $z = 0.35$ mm. (b-d) Images using (b) the double-helix (DH) PSF, (c) the polarized spiral PSF (PS^{2}F), and (d) the QuadraPol PSF. Scale bar: 50 µm; color bar: AoLP in (c,d). Insets show the PSFs.

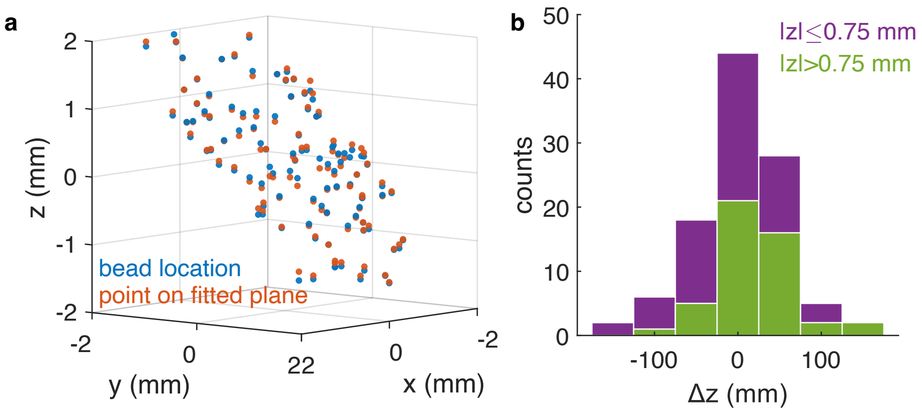

Fig S5 Depth accuracy of SVF using QuadraPol PSF. (a) Centroid locations of reconstructed fluorescent beads on a tilted coverslip, as shown in Figure 5. Blue dots represent the actual locations of the reconstructed beads; red dots indicate their projections on a fitted plane with a tilt angle of 46.1°, projected along the z-axis. (b) Distribution of distances between the actual locations of reconstructed beads and their projections. These distances fall within the axial resolution of the imaging system (~240 µm) for both small ($|z| \leq 0.75$ µm) and large ($|z| > 0.75$ µm) defocus distances, confirming the depth accuracy of the system.

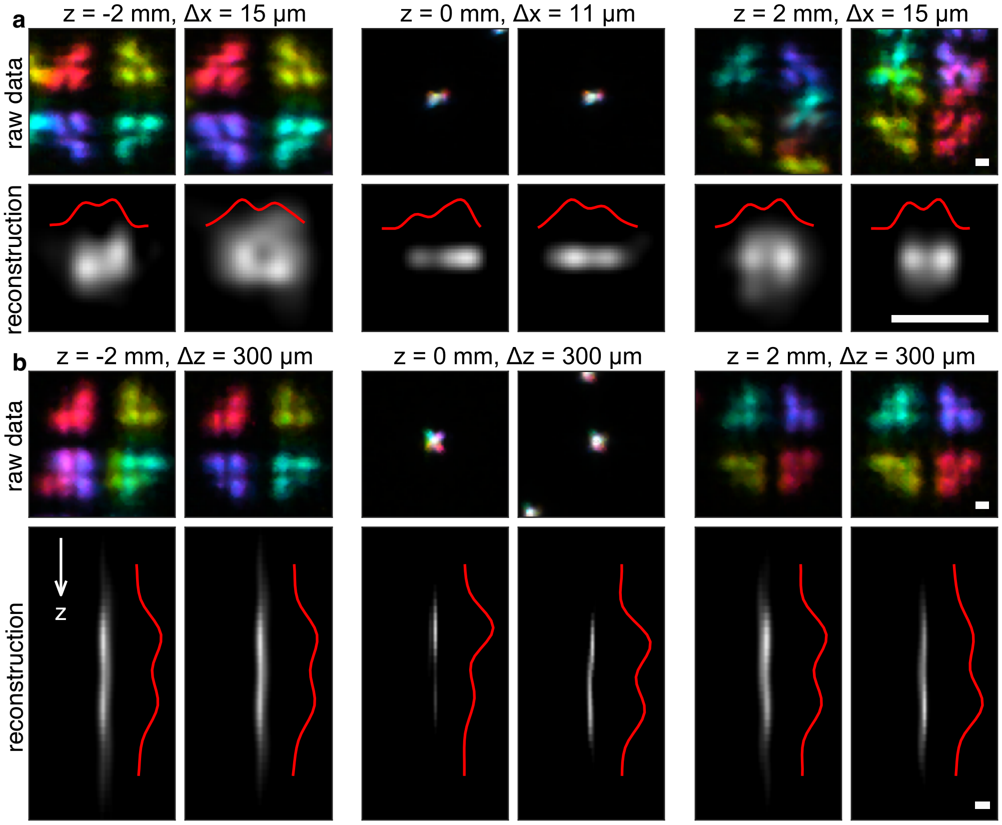

Fig S6 Determining the resolution using Rayleigh criterion for fluorescent beads data. Raw data (a) with multiple lateral shifts and (b) from various depths are combined to synthesize images of two beads with varying lateral and axial separations. Scale bar: 20 µm. The reconstructions indicate that lateral separations of 11 µm and 15 µm are distinguishable for two in-focus beads and two beads defocused by 2 mm, respectively. For an axial separation of 300 µm, both in-focus beads and those defocused by 2 mm are distinguishable. These measurements, after accounting for the bead diameter of 4 µm, are consistent with our simulation results in Figure 4.

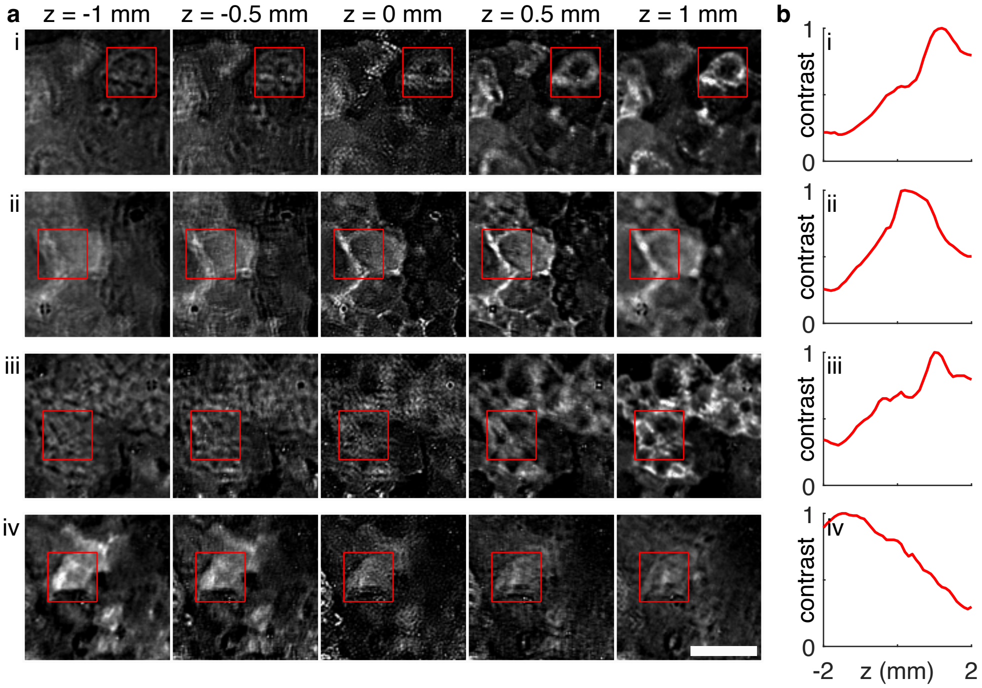

Fig S7 Demonstration of all-in-focus imaging for resolving *E. coli* tagged with mScarlet on sand surfaces. (a) Focal stack of four zoomed regions of interest in Figure 5. Scale bar: 0.5 mm. (b) Image contrast, defined as the difference between the 99th and 1st percentile values in the boxed areas, as a function of depth z.

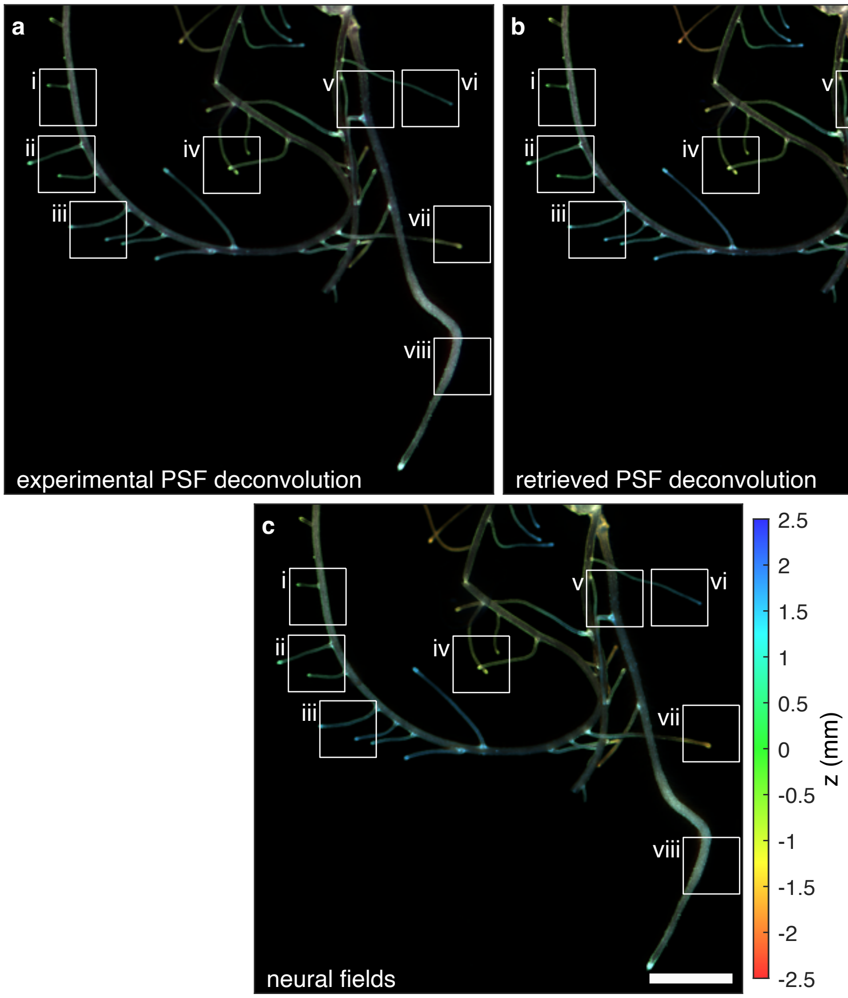

Fig S8 Volumetric reconstruction of wheat roots colorcoded by depth z. Reconstruction using (a) deconvolution with experimental PSF, (b) deconvolution with retrieved PSF, (c) and neural fields. Scale bar: 5 mm. The zoom-in regions in the white boxes are displayed in Figures S10 and S11.

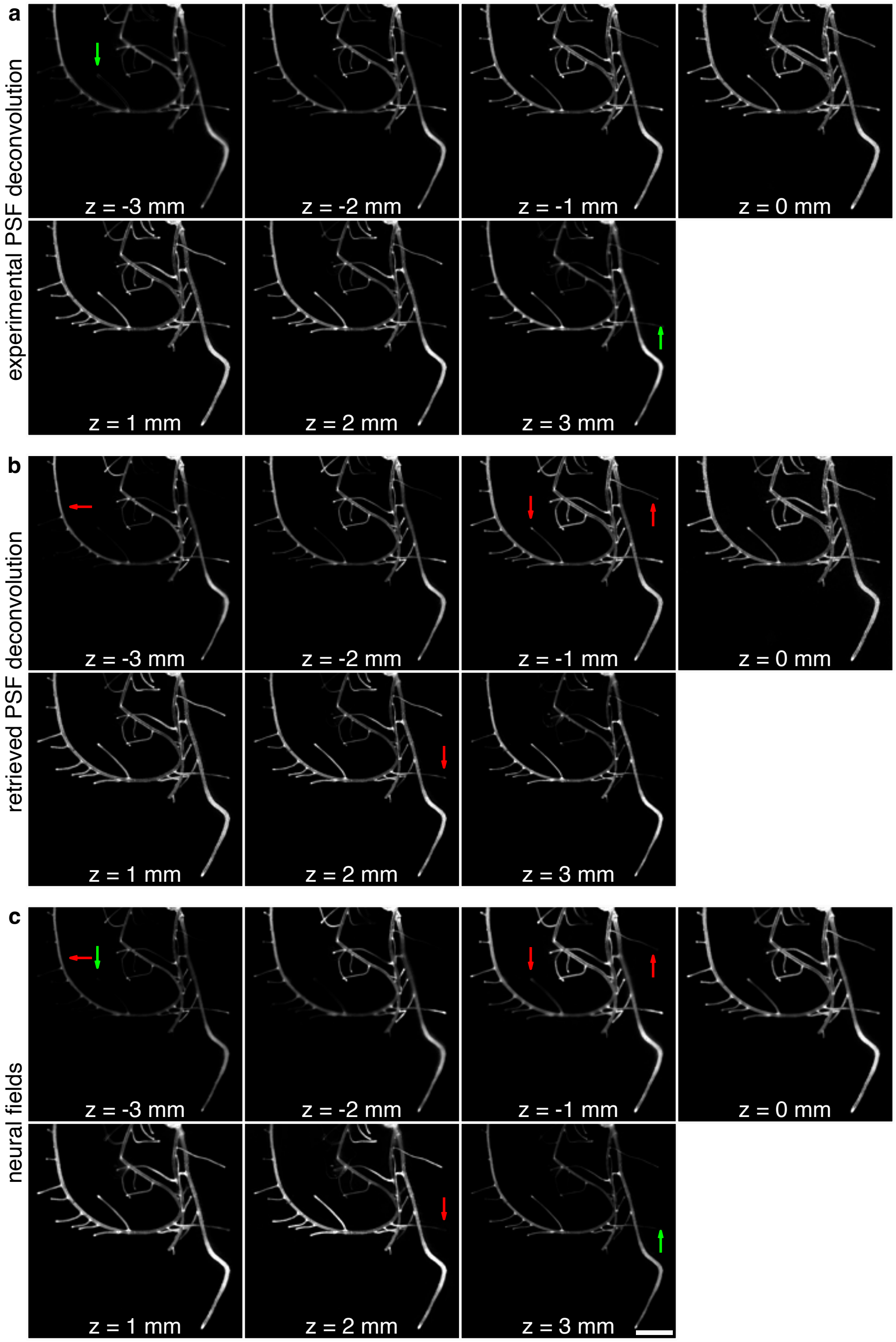

Fig S9 Focal stack from volumetric imaging of wheat roots. Reconstruction using (a) deconvolution with experimental PSF, (b) deconvolution with retrieved PSF, and (c) neural fields. Scale bar: 5 mm. Green and red arrows mark representative areas where neural fields provide sharper reconstruction in z compared to deconvolution with experimental and retrieved PSFs, respectively.

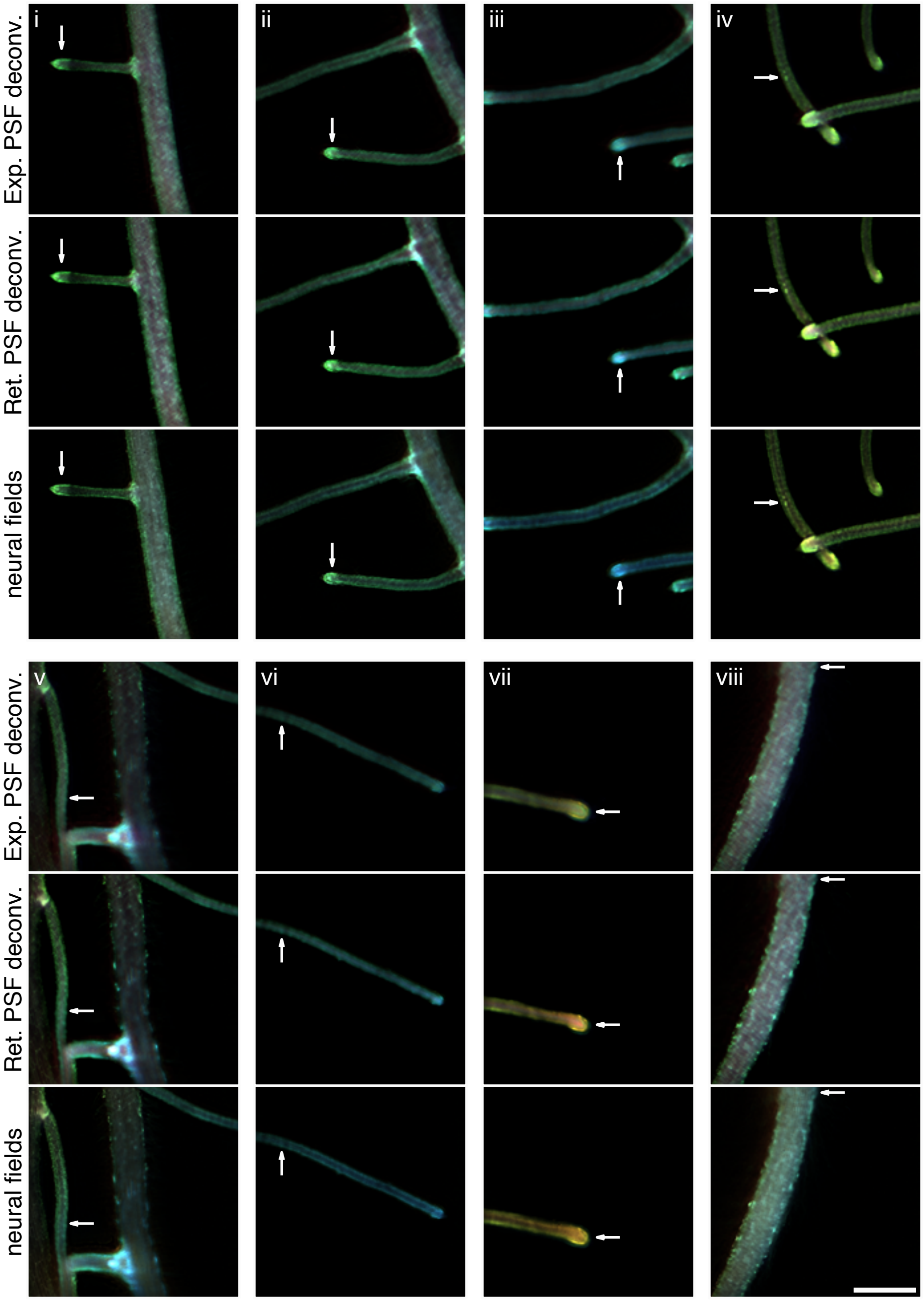

Fig S10 Zoomed regions of interest in Figure S8. Arrows indicate areas where neural fields produce sharper images compared to deconvolution. Scale bar: 1 mm.

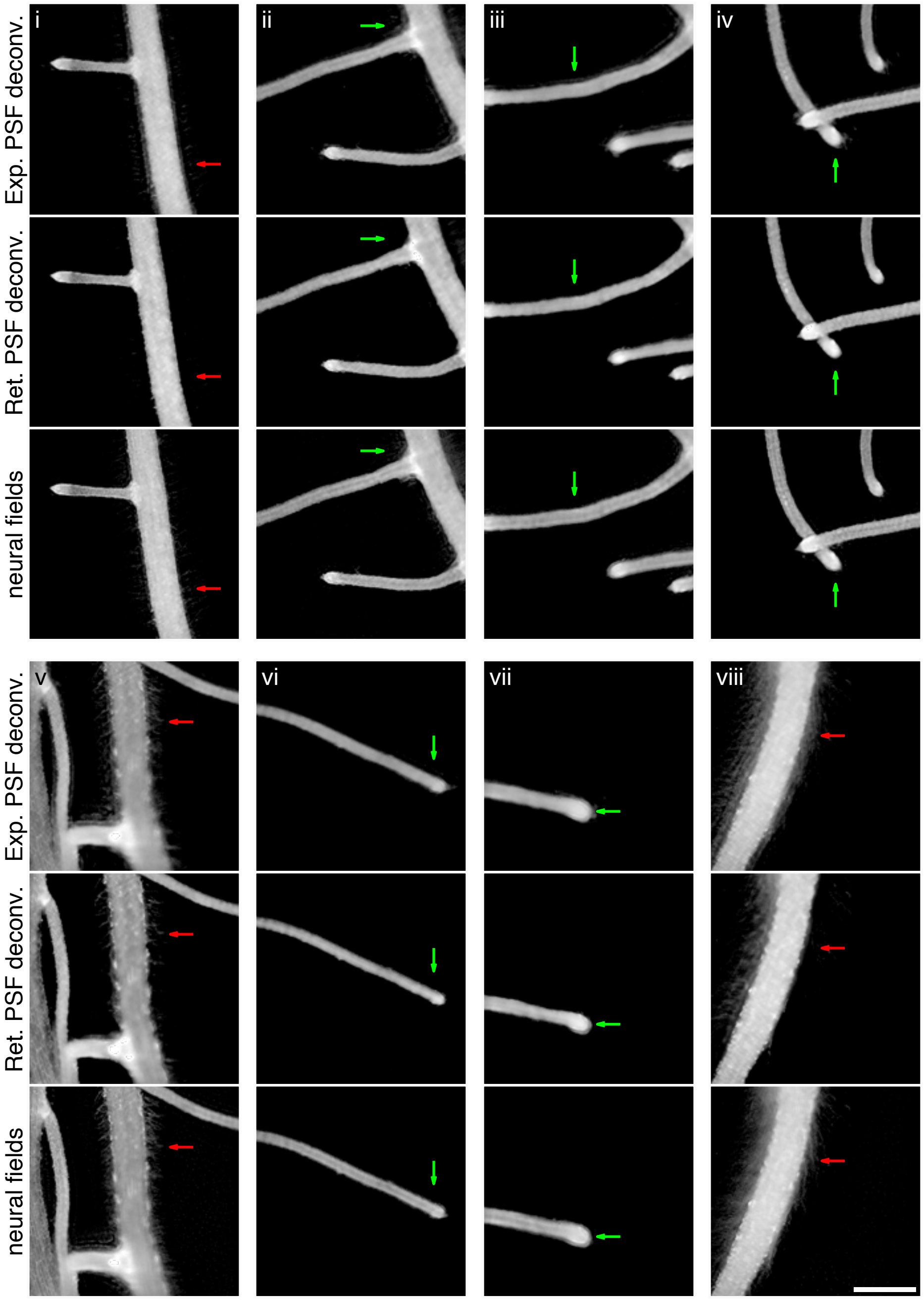

Fig S11 Zoomed regions of interest in Figure S8, plotted on a log scale. Green arrows highlight areas where neural fields mitigate artifacts from deconvolution with the experimental PSF; red arrows indicate areas where neural fields resolve fine root hair structures, which are not resolvable with deconvolution using the retrieved PSF. Scale bar: 1 mm.

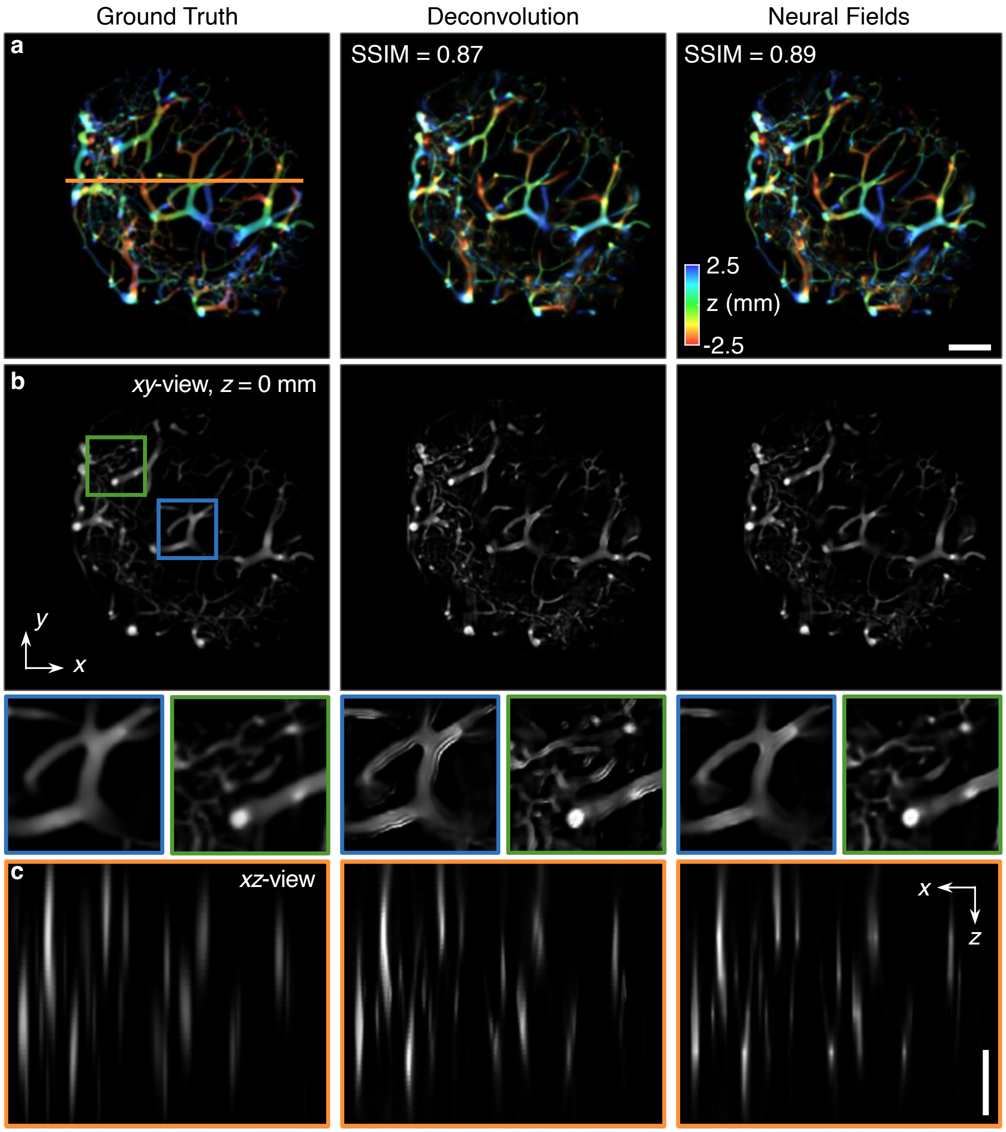

Fig S12 SVF imaging demonstrated with synthetic lymph node vasculature data [67]. (a) The 3D reconstruction using neural fields achieves a structural similarity index measure (SSIM) of 0.89, which is higher than that achieved using RL deconvolution (0.87). Both the xy- and xz-cross sections show that the reconstruction using neural fields exhibits fewer artifacts and is closer to the ground truth compared to RL deconvolution. Scale bar: 100 μm in (a) and 1 mm in (c).

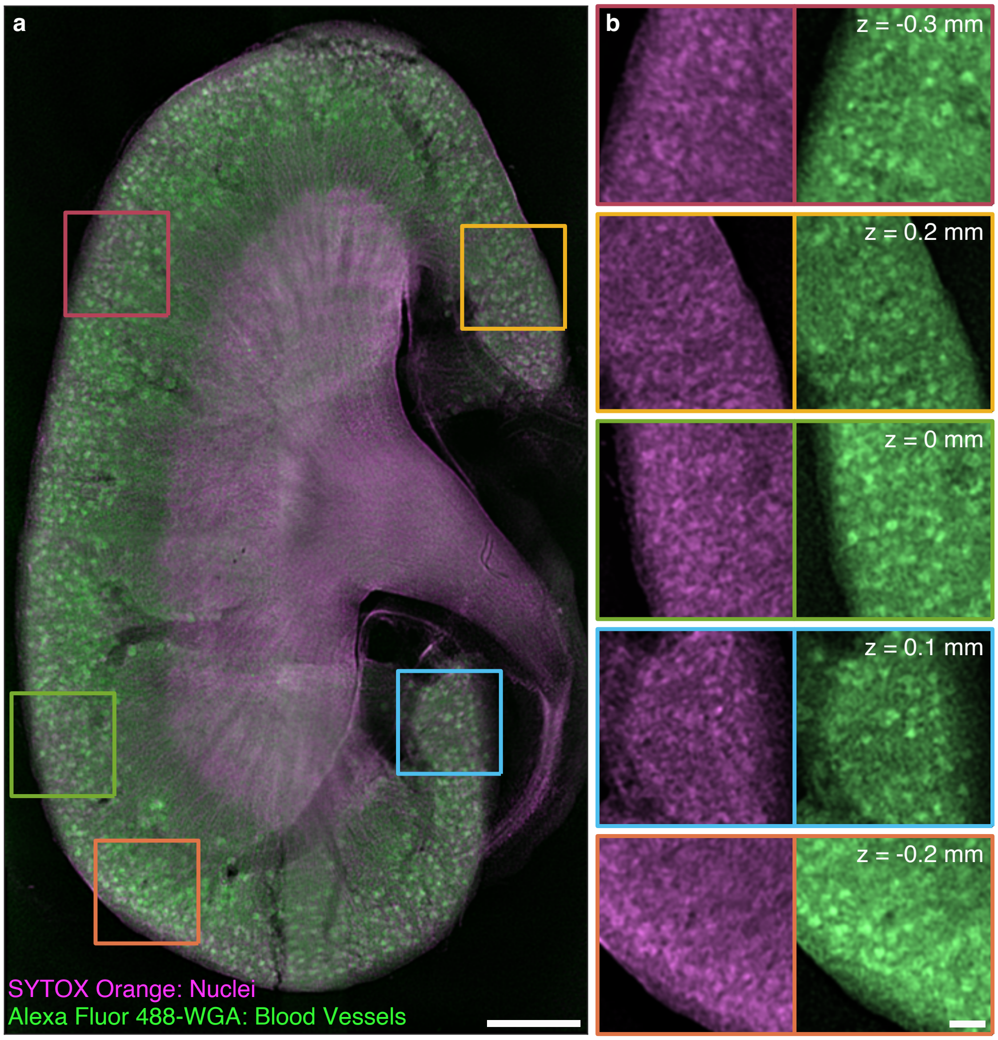

Fig S13 Two-channel SVF imaging for a mouse kidney section. (a) SVF imaging of a 550-μm-thick mouse kidney section (SunJin Lab Co., PS001) labeled with SYTOX Orange and Alexa Fluor 488-WGA. The dual-color imaging results show nuclei in magenta and blood vessels in green. Within the renal capsule, the cortical region contains richly vascularized structures, notably the spherical glomeruli distributed throughout. (b) Zoomed-in areas demonstrate the ability to resolve glomeruli located at various depths. Scale bar: 1 mm in (a) and 200 μm in (b).

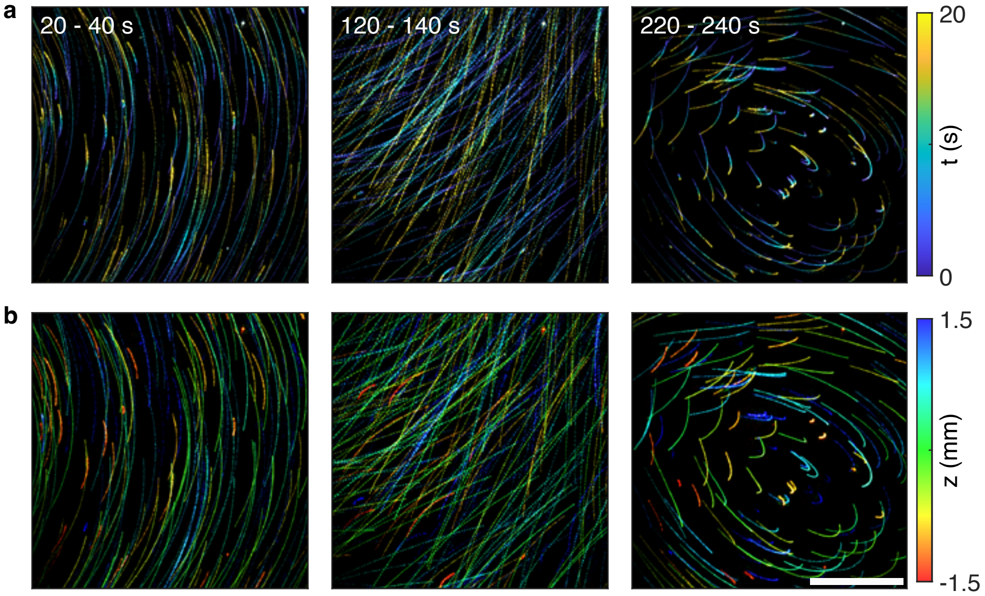

Fig S14 Three-dimensional particle image velocimetry using the QuadraPol PSF. This demonstrates the video-rate imaging capability of our SVF system with fluorescent beads suspended in water within a $50 \times 50 \times 5$ mm^3 container. Images are captured at a rate of 5 frames per second. (a,b) Trajectories of fluorescent beads over 20-second intervals, color-coded by (a) time and (b) depth. Scale bar: 1 mm. Over time, we observed various vortices and noted that beads with greater defocus tend to exhibit slower movements, likely due to increased hydrodynamic drag as they move closer to the container surface.

S1 Forward imaging model

To model the images captured by the polarization camera, we use a vectorial forward model to describe the light emitted from fluorescent molecules. A fluorescent molecule is modeled as an oscillating dipole with orientation $[\mu_x, \mu_y, \mu_z]^\top$. The polarized optical field for an in-focus dipole, $\boldsymbol{E}_{\mathrm{bfp}}$, at the back focal plane (BFP) is given by [46,47]

$$\begin{aligned}\boldsymbol{E}_{\mathrm{bfp}}(\phi,\rho) &= \begin{bmatrix} E_{x,\mathrm{bfp}}(\phi,\rho) \\ E_{y,\mathrm{bfp}}(\phi,\rho) \end{bmatrix} \\ &= \frac{E_0}{(1-\rho^2)^{1/4}} \begin{bmatrix} \sin^2(\phi)+\cos^2(\phi)\sqrt{1-\rho^2} & \sin(2\phi)(\sqrt{1-\rho^2}-1)/2 & -\rho\cos(\phi) \\ \sin(2\phi)(\sqrt{1-\rho^2}-1)/2 & \cos^2(\phi)+\sin^2(\phi)\sqrt{1-\rho^2} & -\rho\sin(\phi) \end{bmatrix} \begin{bmatrix} \mu_x \\ \mu_y \\ \mu_z \end{bmatrix},\end{aligned} \tag{1}$$

where (ϕ, ρ) represent the polar coordinates at the BFP, and E_0 is a complex scaling factor. Given that ρ is constrained by the low numerical aperture (NA) of our imaging system, i.e., $\rho \leq \mathrm{NA} = 0.056$, the field can be approximated as

$$\boldsymbol{E}_{\mathrm{bfp}}(\phi,\rho) = \begin{bmatrix} E_{x,\mathrm{bfp}}(\phi,\rho) \\ E_{y,\mathrm{bfp}}(\phi,\rho) \end{bmatrix} = E_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}. \tag{2}$$

For an isotropic emitter, the emission pattern is the incoherent mixture of images produced by dipoles oriented along the x-, y-, and z-axes. Given that the contribution from the z-oriented dipole is negligible, we focus on the optical field at the BFP for the x- and y-oriented dipoles before modulation by the polarizers,

$$\begin{aligned}\boldsymbol{E}_{\mathrm{bfp}}(\phi,\rho)|_{[\mu_x,\mu_y]=[1,0]} &= E_0[1,0]^\top,\\ \boldsymbol{E}_{\mathrm{bfp}}(\phi,\rho)|_{[\mu_x,\mu_y]=[0,1]} &= E_0[0,1]^\top.\end{aligned} \tag{3}$$

After the fluorescence passes through linear polarizers, the optical fields are given by

$$\boldsymbol{E}'_{\mathrm{bfp}}(\phi,\rho)|_{[\mu_x,\mu_y]=[1,0]} = E_0 \times \begin{cases} [1,0]^\top & 0^\circ \text{ polarizer} \\ [1/2,1/2]^\top & 45^\circ \text{ polarizer} \\ [0,0]^\top & 90^\circ \text{ polarizer} \\ [1/2,-1/2]^\top & 135^\circ \text{ polarizer} \end{cases} \tag{4}$$

for x-oriented dipoles, and

$$\boldsymbol{E}'_{\mathrm{bfp}}(\phi,\rho)|_{[\mu_x,\mu_y]=[0,1]} = E_0 \times \begin{cases} [0,0]^\top & 0^\circ \text{ polarizer} \\ [1/2,1/2]^\top & 45^\circ \text{ polarizer} \\ [0,1]^\top & 90^\circ \text{ polarizer} \\ [-1/2,1/2]^\top & 135^\circ \text{ polarizer} \end{cases} \tag{5}$$

for y-oriented dipoles.

The image captured at the detector corresponding to the x-polarization channel is given by

$$E'_{x,\mathrm{img}} = \mathcal{F}\{E'_{x,\mathrm{bfp}}|_{[\mu_x,\mu_y]=[1,0]}\}^2 + \mathcal{F}\{E'_{x,\mathrm{bfp}}|_{[\mu_x,\mu_y]=[0,1]}\}^2, \tag{6}$$

where the first term corresponds to the p-polarization in Eq. (1), and the second term corresponds to the s-polarization. Similar expressions can be derived for the other three detection channels.

S2 Hyperparameters in neural fields

The raw images have sizes of $N \times N = 2048 \times 2048$ pixels in our demonstrated imaging system. The feature space tensor M with a size of $N/2 \times N/2 \times Q$ is designed based on the raw images' sizes ($N = 2048$). The number of feature channels Q is 32, determined by the feature capacity for the neural rendering. We gradually increase the value of Q and find $Q = 32$ appropriate. The feature tensor u has a size of $Z \times Q$, where $Z = 8$ is a predefined number of z-features correlated to z-coordinates. In our demonstrated experiment, we sampled feature tensor u at 81 planes uniformly. The z planes between the predefined z-features are obtained through linear interpolations. The neural network consists of two nonlinear layers with ReLU activation function and one linear layer. Each layer has 32 neurons, which match the number of feature channels, and one bias (offset) neuron. The output layer only contains one neuron for the image volume output.

The two-step optimization process runs 200 epochs in total, with 100 epochs for initialization and 100 epochs for further optimization. Note that the Richard-Lucy deconvolution before the initialization stage was iterated 100 times in all experiments. An initial learning rate of 0.005 is implemented for good convergence. From our hyperparameter tuning experience, a learning rate between 0.003 and 0.008 can support fast convergence. We implemented a cosine annealing learning scheduler starting at the initial learning rate and ending at $1/6$ of the initial learning rate. The AdamW [49] optimizer is adopted for adaptive gradient descent together with a SmoothL1Loss [50].

S3 Relaxed sparsity constraints and SNR analysis using the QuadraPol PSF

Here, we compare the performance using QuadraPol PSF to that of the Miniscope3D. We note that the Miniscope3D divides the pupil into 36 parts, whereas the QuadraPol PSF divides it into 4 parts. As a result, the DOF using the Miniscope3D is $\sim$3 to 9 times that of the QuadraPol PSF, which comes with the trade-off of lower lateral resolution. Therefore, we compared the PSFs of the Miniscope3D within a DOF of 0.3 mm and the QuadraPol PSF within a DOF of 0.1 mm.

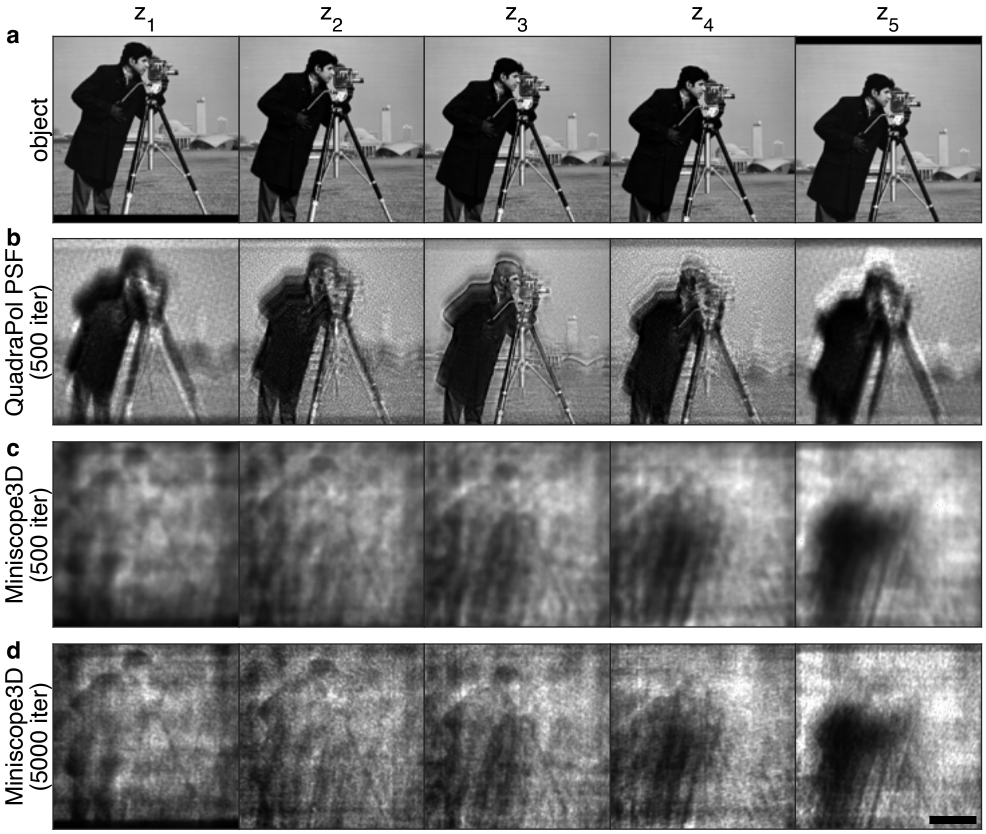

Fig S15 Evaluating the reconstruction quality using the Richardson-Lucy deconvolution algorithm without sparsity constraints. (a) Simulated five-layered object. (b) Reconstruction using the QuadraPol PSF after 500 iterations. (c,d) Reconstruction using the Miniscope3D after (c) 500 and (d) 5000 iterations. Scale bar: 0.1 mm.

To further illustrate the relaxation of the sparsity constraint beyond the examples provided in the main text, we simulate a five-layered object featuring a slightly shifted cameraman pattern in each layer (Figure S15a). For the QuadraPol PSF, the layers z_1 through z_5 are uniformly sampled from -0.05 to 0.05 mm, while for the Miniscope3D, they range from 0 to 0.3 mm. The reconstruction using the QuadraPol PSF clearly resolves all five layers (Figure S15b), exhibiting only PSF-induced blurring and much fewer artifacts compared to the Miniscope3D (Figure S15c). Even with 10 times more iterations, the same algorithm using PSFs from the Miniscope3D still fails to accurately recover the object (Figure S15d).

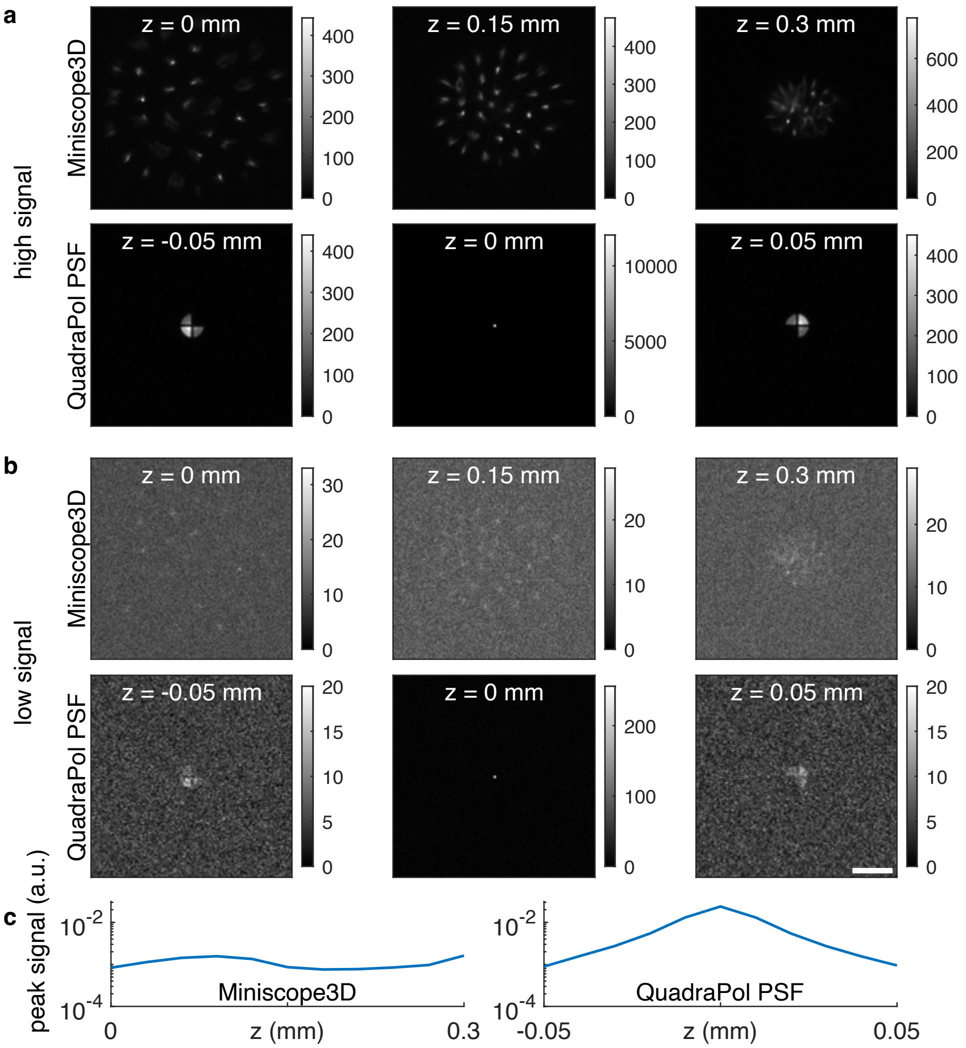

Fig S16 SNR comparison between QuadraPol PSF and Miniscope3D [13]. (a) Representative PSFs with Poisson shot noise under high SNR conditions, with 500,000 signal photons and 10 background photons per camera pixel. (b) Representative PSFs with Poisson shot noise under low SNR conditions, with 15,000 signal photons and 10 background photons per camera pixel. (c) Peak signal of the PSF using the Miniscope3D and the QuadraPol PSF. Scale bar: 0.1 mm.

For accurate comparison of the SNR, the signal for QuadraPol PSF is reduced by a factor of

16 to account for fluorescence losses due to the custom polarizer (which rejects half of the fluorescence) and the polarization camera (which rejects half of the remaining fluorescence and separates it across four channels). The background is halved since it is also filtered by the polarization camera. Simulated images at a high signal level (Figure S16a) show the shape of the PSFs, while images at a low signal level (Figure S16b) demonstrate their SNR performance. The peak signal level indicates that the QuadraPol PSF (Figure S16c) outperforms the Miniscope3D across their respective DOFs.

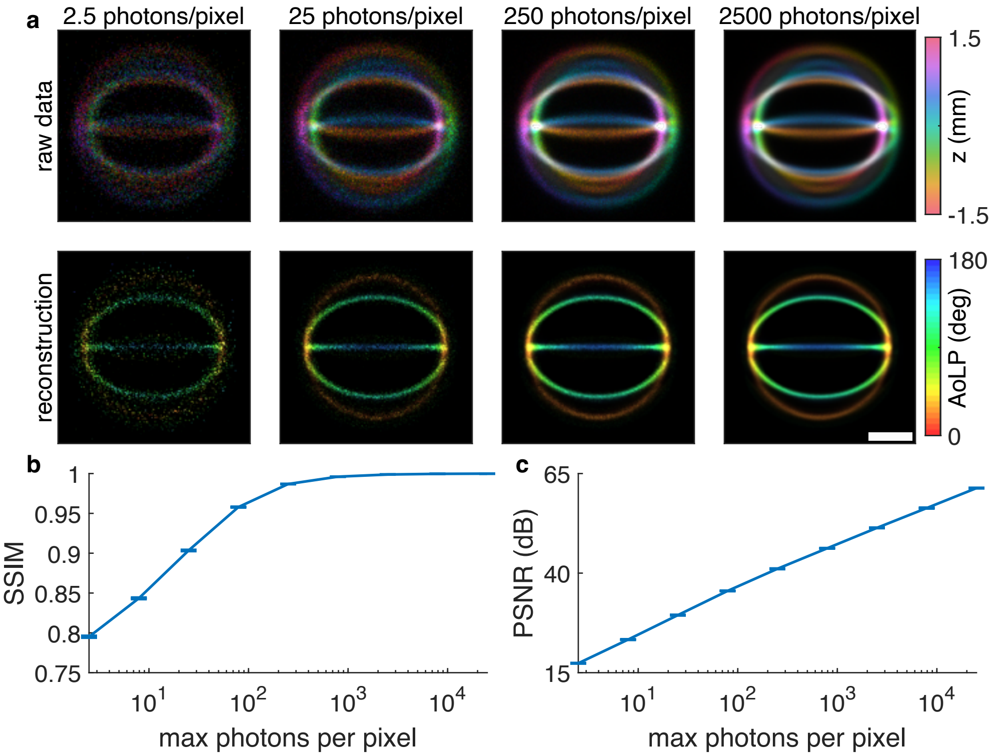

Fig S17 SNR analysis of the QuadraPol PSF. (a) Representative simulated images with Poisson shot noise and reconstructions of the object shown in Figure 4d under various signal levels, indicated by the maximum expected photons per pixel. Scale bar: 200 µm. (b) Structural similarity index measure (SSIM) and (c) peak signal-to-noise ratio (PSNR) as functions of maximum photons per pixel in the raw images. Error bars represent one standard deviation across 20 realizations.

We further evaluate the SNR performance of the QuadraPol PSF as implemented in our study. We simulated the line structure in Figure 4d and generated images with Poisson shot noise at various signal levels (Figure S17(a)). We assessed the reconstruction quality using the structural similarity index measure (SSIM, Figure S17(b)) and peak signal-to-noise ratio (PSNR, Figure S17(c)). At an extremely low signal level where the maximum photons per pixel in the simulated images is 2.5, the reconstruction has an SSIM of 0.80 and a PSNR of 17.4 dB, indicating barely acceptable quality. At a higher but still low signal level of 25 photons per pixel, the reconstruction quality improved significantly, achieving an SSIM of 0.90 and a PSNR of 29.5 dB. These results indicate that the QuadraPol PSF can achieve good reconstruction quality even under low light conditions.

S4 PSFs and merged PSFs for image volume reconstruction

In Section 2.2 of the paper, we report the theoretical PSFs, the retrieved PSFs, and the experimental PSFs. The application of these PSFs in volumetric reconstructions with neural fields falls into two scenarios: (1) When the experimental PSFs exhibit good SNR within the sample's depth range, they can accurately model the imaging system across every z-plane throughout the sample volume. In such cases, we exclusively use the experimental PSFs for image reconstructions. (2) When dealing with thicker samples (thickness exceeding 4 mm for our system), the PSFs at greater depth positions are significantly noisier. Under these circumstances, we combine experimental PSFs

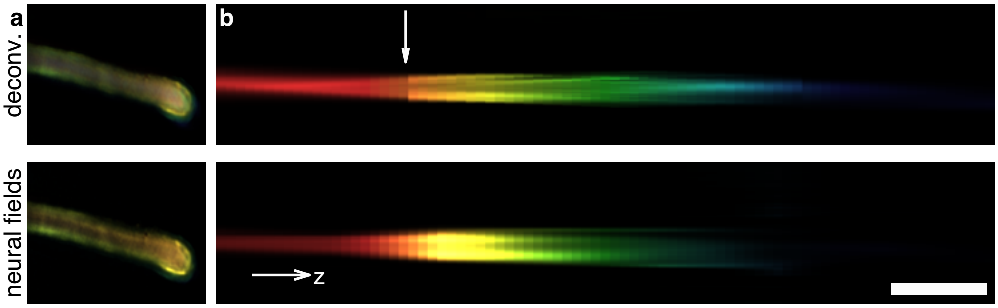

Fig S18 Reconstruction of wheat roots using deconvolution and neural fields with experimental PSF for $|z| \leq 2$ mm and retrieved PSF otherwise. (a) The xy and (b) the xz views of zoomed regions in Figure 7(e). Scale bar: 1 mm.

from near the focal plane with retrieved PSFs from positions far from the focal plane to initialize the algorithm. This approach is particularly effective for managing noise and ensuring accurate reconstruction. An application of this strategy is demonstrated in the plant root experiment, as detailed below.

In the plant root imaging experiment (Section 3.4 and Figure 7 in the main text), the experimental PSFs at large defocus distances are degraded with low SNR, which significantly reduces the quality of the deconvolved image volume. Alternatively, using retrieved PSFs allows for an extended depth range without such limitations. These retrieved PSFs (Figure 2(b)) are calculated using phase retrieval, which uses a limited number of experimental PSFs at various depths to estimate the pupil phase. Nonetheless, the retrieved pupil phase and the theoretical amplitude cannot exactly replicate the experimental PSF due to the non-uniform amplitude of the polarizers in the experiments. On the other hand, a possible approach is taking high SNR experimental PSFs from small defocus distances and retrieved PSFs at larger defocus distances, and merging them to construct a new set of PSFs for deconvolution. However, this method can introduce discontinuities along the axial direction, as shown in Figure S18.

To overcome the aforementioned challenges, we incorporate a merged PSF set in our neural fields algorithm to achieve improved results. Specifically, we first deconvolved the captured images using experimental PSFs with defocus distances $|z| \leq 2$ mm and retrieved PSFs with defocus distances $|z| \leq 4$ mm, respectively (demonstrated in Figure S19(a)). We then merged the two image volumes, one from experimental PSFs with $|z| \leq 2$ mm and the other from retrieved PSFs with 2 mm $< |z| \leq 4$ mm to create a merged image volume. During the initialization, the neural field is fitted to the merged image volume (Figure S19(b)). Next, the corresponding experimental and retrieved PSFs are implemented in the forward model of the imaging system during the opti-

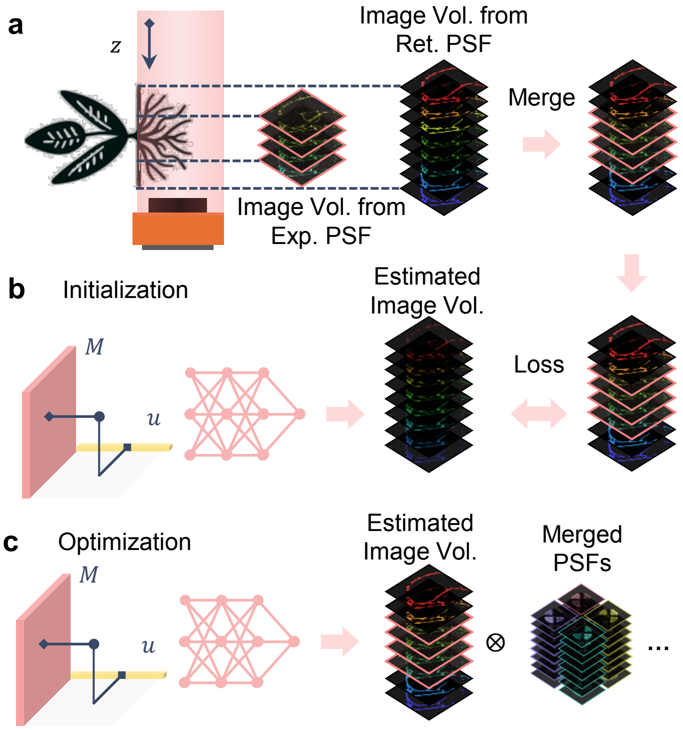

Fig S19 Neural fields algorithm with merged PSFs. (a) Merging of image volumes from experimental and retrieved PSFs. (b) Initialization with merged image volume. (c) Optimization with merged PSFs. Image Vol. stands for image volume; Exp. PSF denotes experimental PSF; Ret. PSF represents retrieved PSF.

mization, as depicted in Figure S19(c). This joint optimization using different PSFs for different regions effectively addresses the problems in deconvolution.