Supplementary Material

List of Figures

Contents

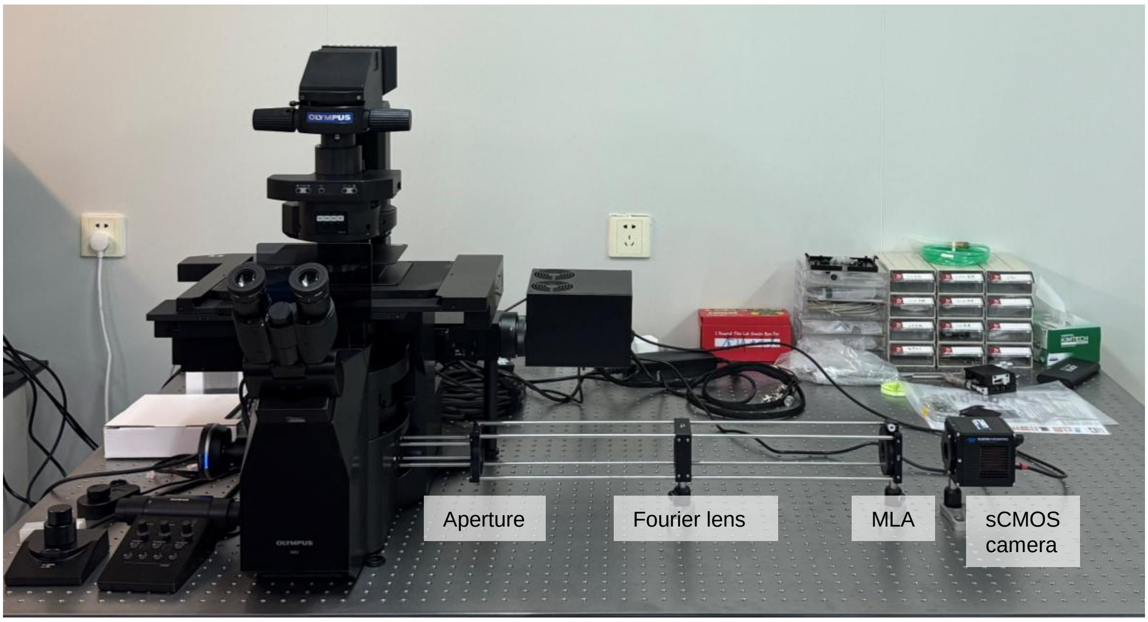

Fig. S1 | **Fourier light-field microscopy setup.** Overview of the FLFM system coupled to an inverted microscope, with the detection path consisting of an aperture, Fourier lens, microlens array (MLA), and sCMOS camera.

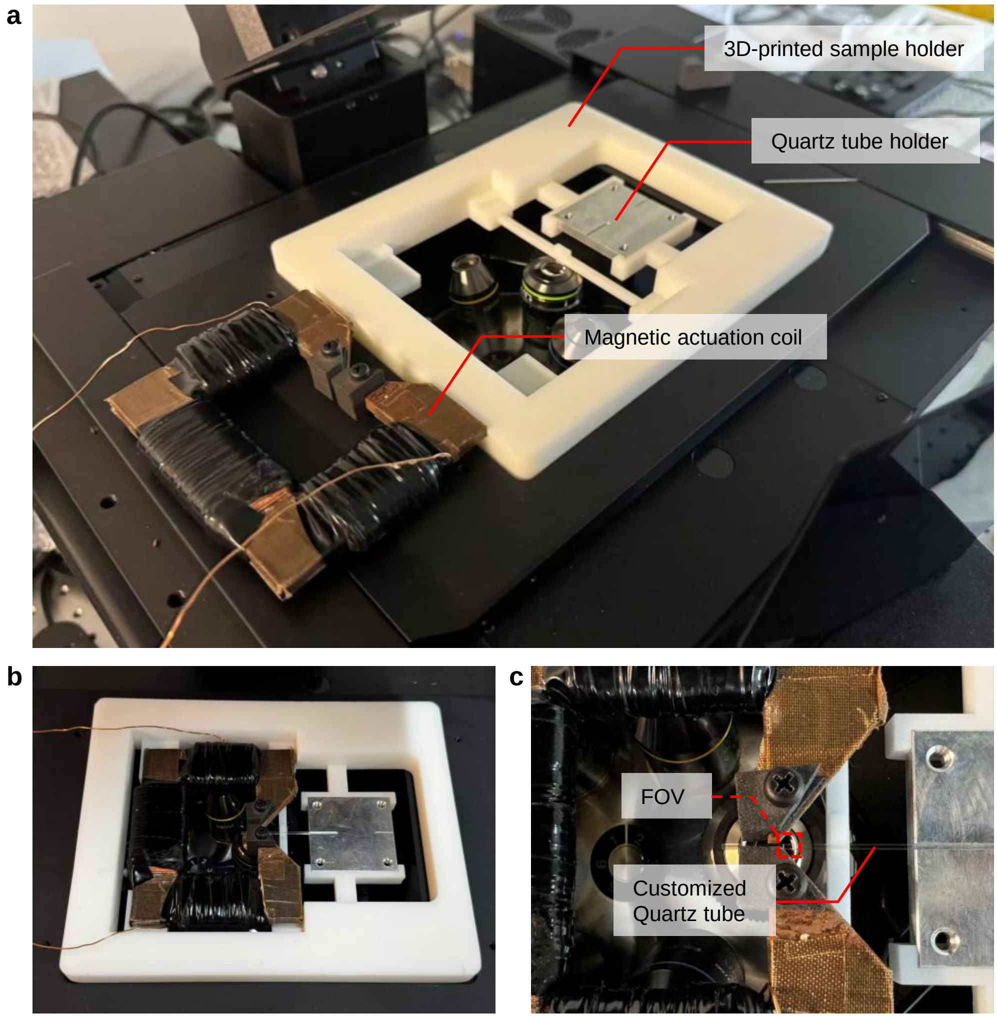

Fig. S2 | Custom microscope-stage insert for magnetic actuation during quartz-tube imaging. a, Overview of the assembled setup mounted on the microscope stage, showing the 3D-printed sample holder, quartz tube holder, and magnetic actuation coil. **b,** Top view of the integrated module. **c,** Close-up of the sample region. The customized quartz tube is positioned adjacent to the magnetic actuation coil and aligned with the microscope field of view, indicated by the dashed red box.

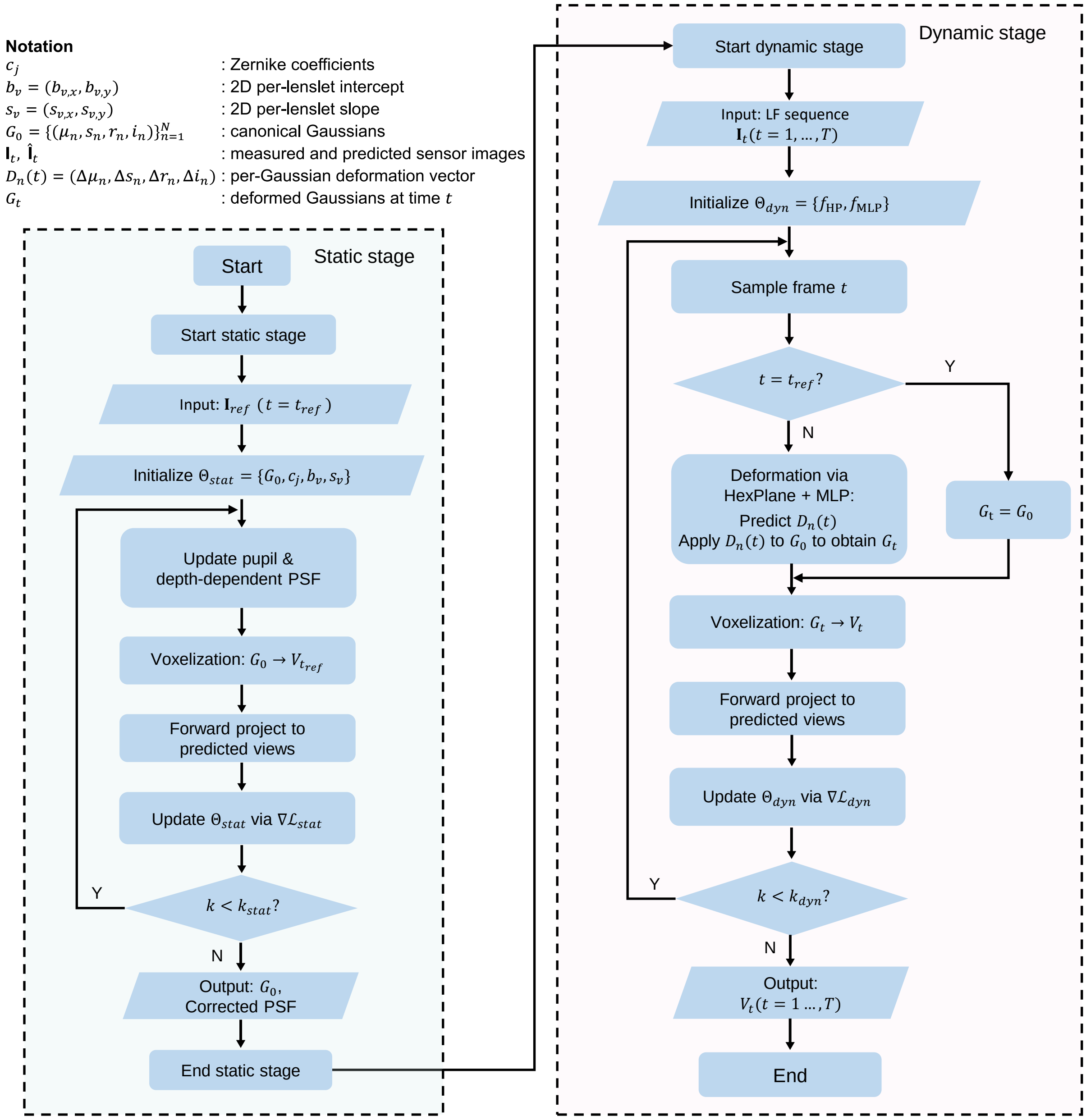

Fig. S3 | Two-stage 4DGAT reconstruction pipeline. In the static stage, the reference-frame FLFM measurement $\mathbf{I}_{\mathrm{ref}}$ at $t = t_{\mathrm{ref}}$ is used to jointly optimize the static-stage parameters $\Theta_{\mathrm{stat}} = \{G_0, c_j, b_v, s_v\}$, comprising the canonical Gaussians and JPGM parameters. At each iteration, the JPGM-derived depth-dependent PSF stack is updated, the canonical Gaussians G_0 are voxelized, forward projected to the predicted views, and refined through the static-stage loss $\mathscr{L}_{\mathrm{stat}}$. This stage outputs the canonical Gaussian representation G_0 together with the optimized JPGM-derived PSF stack, both of which are held fixed during dynamic reconstruction. The full sequence $\{\mathbf{I}_t\}_{t=1}^{T}$ is then used to learn deformation parameters $\Theta_{\mathrm{dyn}} = \{f_{\mathrm{HP}}, f_{\mathrm{MLP}}\}$. For each sampled frame t, the model keeps $G_t = G_0$ at the reference time, and otherwise predicts the per-Gaussian deformation vectors $\{D_n(t)\}_{n=1}^{N}$ and applies them attribute-wise to obtain G_t. The deformed Gaussians are then voxelized, forward projected to the predicted views, and optimized with $\mathscr{L}_{\mathrm{dyn}}$ to reconstruct the time-varying volumes.

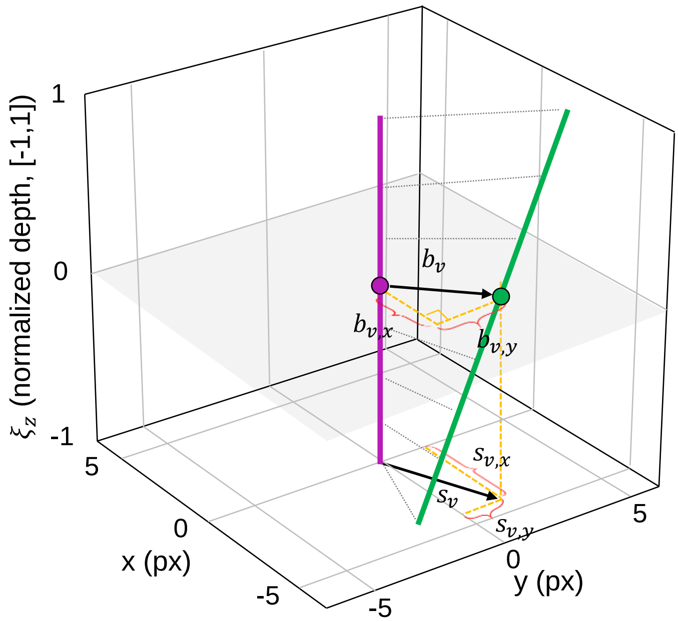

Fig. S4 | **Geometric interpretation of the per-lenslet PSF-shift model.** Motivated by hybrid-PSF FLFM calibration,[1] we model the shift of each lenslet's PSF center as a linear function of the normalized depth coordinate $\xi(z) \in [-1,1]$, $\Delta x_v(z) = b_{v,x} + s_{v,x}\,\xi(z)$, $\Delta y_v(z) = b_{v,y} + s_{v,y}\,\xi(z)$, where v indexes the lenslet view. The purple vertical line denotes the ideal PSF-center trajectory (no aberration, no shift), while the green tilted line is the actual trajectory produced by the linear model. The intercept vector $b_v = (b_{v,x}, b_{v,y})$ (drawn in the $\xi(z) = 0$ plane) represents the depth-independent in-plane misalignment of the lenslet, i.e., the lateral offset of the actual PSF center from the ideal position at the reference depth. The slope vector $s_v = (s_{v,x}, s_{v,y})$ represents the depth-dependent drift per unit change in $\xi(z)$, geometrically captured by the tilt of the green trajectory in the sensor plane.

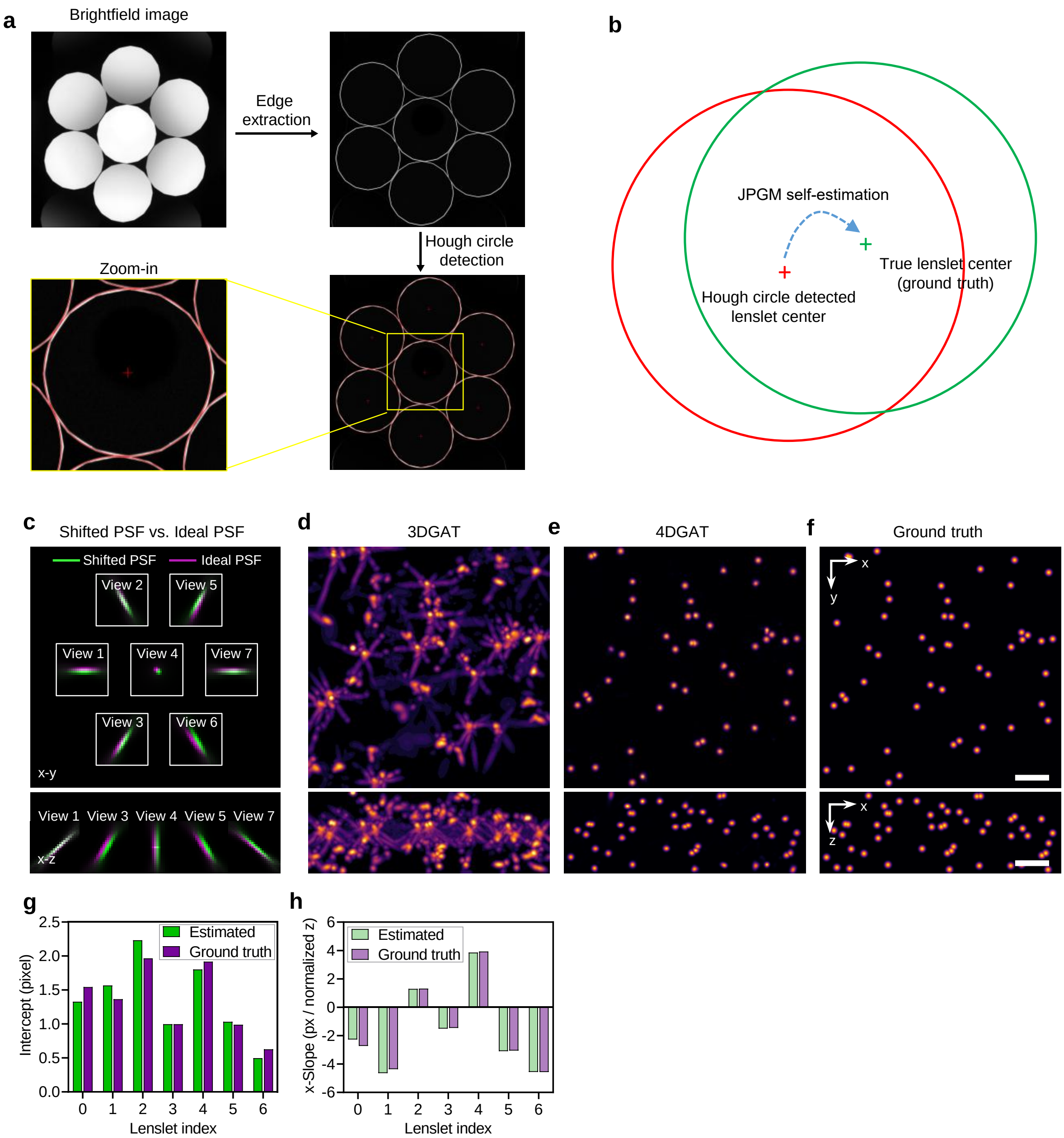

Fig. S5 | Hough-based lenslet-center initialization and shift-only validation of PSF-shift self-estimation in 4DGAT. a, Pre-processing pipeline for microlens-array (MLA) geometry calibration. A brightfield image of the MLA (top left) is used as the geometric reference; an image-gradient edge map (top right) is fed to a circular Hough transform to detect the seven lenslet apertures. The zoom-in (bottom left) shows the detection for the central lenslet. **b,** Schematic of the residual lenslet-center offset and its self-estimation by 4DGAT. The center obtained by Hough-circle detection (red circle and cross) provides the initial geometric estimate, whereas the green circle and cross denote the ground-truth lenslet aperture and center. Finite pixel sampling and circle-fitting uncertainty can produce a residual error between the detected and true geometric centers; optical misalignment and MLA fabrication tolerances can additionally displace the effective PSF center from the nominal geometric center. Rather than treating this offset as fixed, 4DGAT learns a per-lenslet correction through the intercept and slope parameters $\{b_v, s_v\}$. The blue dashed arrow schematically indicates the correction inferred from the Hough-based initialization toward the ground-truth center. **c–h,** Shift-only simulation isolating the estimation of lenslet-dependent PSF-center displacement. No Zernike aberrations are introduced. Among the learnable optical parameters, only these per-lenslet shift parameters are optimized, while the object representation is reconstructed simultaneously. **c,** Overlays of the shifted and ideal PSFs for different lenslet views, shown as x–y MIPs (upper block) and x–z MIPs (lower block). The view-dependent lateral displacement and axial drift illustrate the mismatch introduced solely by the intercept–slope shifts. **d–f,** x–y MIPs (upper) and x–z MIPs (lower) of the volumes reconstructed by 3DGAT (**d**) and 4DGAT (**e**), together with the ground truth (**f**). Because 3DGAT assumes an unshifted ideal PSF, the model mismatch produces severe lateral streaking and axial elongation. By self-estimating the per-lenslet intercepts and slopes, 4DGAT restores localized point objects in close agreement with the ground truth. **g,** Estimated and ground-truth per-lenslet intercepts. **h,** Estimated and ground-truth x-slope components. Scale bars, 50 μm in **d–f**.

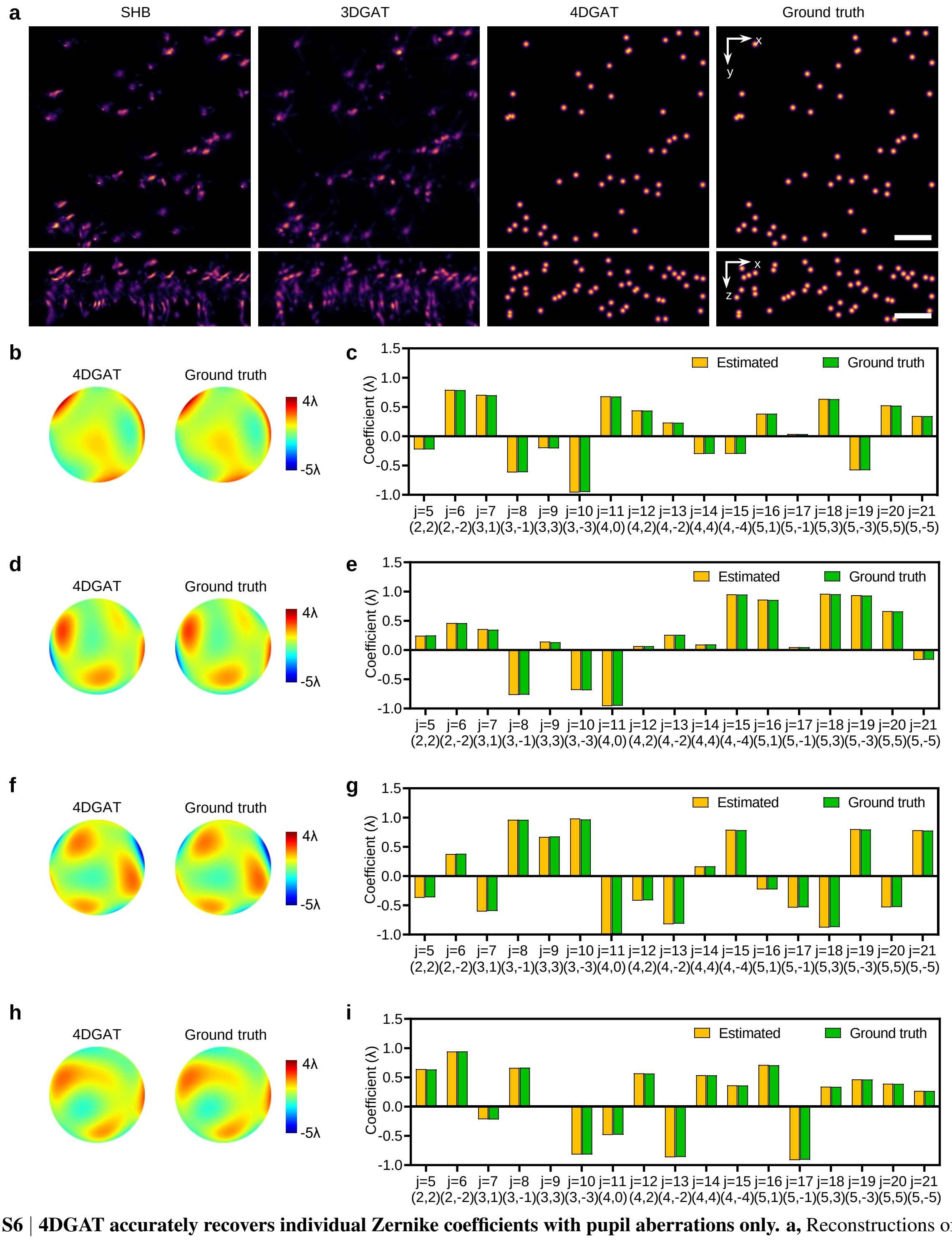

Fig. S6 | 4DGAT accurately recovers individual Zernike coefficients with pupil aberrations only. a, Reconstructions of the same synthetic bead object used in Fig. S7. The pupil aberrations were generated using the same Zernike-coefficient sampling protocol as in Fig. S7 (all 17 terms from $j = 5$ to 21 active, with coefficients sampled uniformly from $[-1, 1]$ waves). The forward model contains pupil aberrations only. SHB deconvolution and 3DGAT assume an ideal PSF and produce streaked, smeared reconstructions, whereas 4DGAT restores localized points that match the ground truth. **b,** 4DGAT-estimated pupil phase (left) and ground-truth pupil phase (right) on a common color scale spanning -5λ to 4λ (shared by all pupil-phase maps in **b, d, f, h**); the two maps are visually indistinguishable. **c,** Per-term Zernike coefficients estimated by 4DGAT (orange) and the corresponding ground-truth values (blue). The recovered values closely agree with the ground truth across all 17 coefficients. **d–i,** Three additional randomly sampled aberration realizations. For each row, the 4DGAT-estimated pupil phase is compared with the corresponding ground-truth phase (**d, f, h**), and the recovered per-term Zernike coefficients are plotted against the ground truth (**e, g, i**). Across all four aberration-only realizations, the phase maps and individual coefficients closely agree with the corresponding ground truth. Scale bars, 50 μm in **a**.

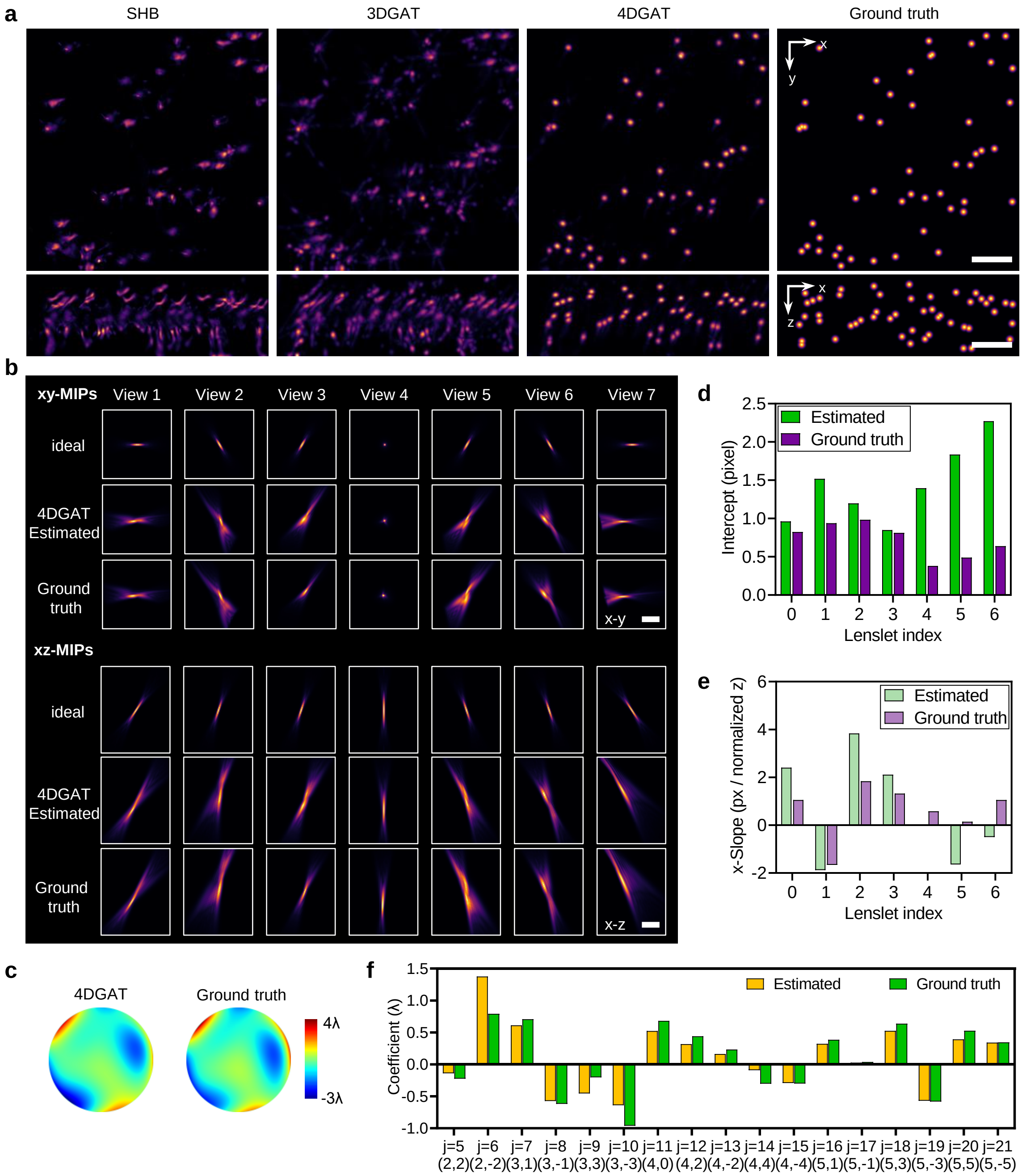

Fig. S7 | 4DGAT recovers the effective optical response and volume under joint estimation of strong pupil aberrations and per-lenslet PSF shifts. This simulation evaluates a fully flexible joint-estimation regime in which the ground-truth forward model contains both strong pupil aberrations (all 17 Zernike terms from $j = 5$ to 21 active, with coefficients sampled uniformly from $[-1, 1]$ waves) and random per-lenslet PSF shifts parameterized by lateral intercepts and depth-dependent slopes. 4DGAT jointly optimizes both parameter groups. Although cross-information between the Zernike and PSF-shift parameterizations can lead to parameter trade-offs, the effective per-view PSFs, pupil-phase distribution, and reconstructed volume remain well constrained and are recovered close to the ground truth. **a,** Reconstructions of synthetic beads under the joint optical perturbations. SHB deconvolution and 3DGAT assume an ideal PSF, producing streaked and smeared reconstructions, whereas 4DGAT recovers localized, correctly positioned points close to the ground truth. **b,** Per-view PSF comparisons across all seven lenslet views, shown as x–y MIPs (upper block) and x–z MIPs (lower block). Despite severe deviations from the ideal PSF, the 4DGAT estimates reproduce the shape, orientation, and axial elongation of the ground-truth PSFs across views. **c,** Pupil phase estimated by 4DGAT and the corresponding ground truth. **d,** Per-lenslet intercepts and **e,** per-lenslet x-slopes estimated by 4DGAT versus the ground truth. **f,** Zernike coefficients estimated by 4DGAT (orange) and the ground truth (green). The parameter-level discrepancies in **d–f** reflect trade-offs between the Zernike and PSF-shift parameterizations in the joint model, rather than a failure to recover the effective optical response. Scale bars, 50 μm in **a**; 20 μm in **b**.

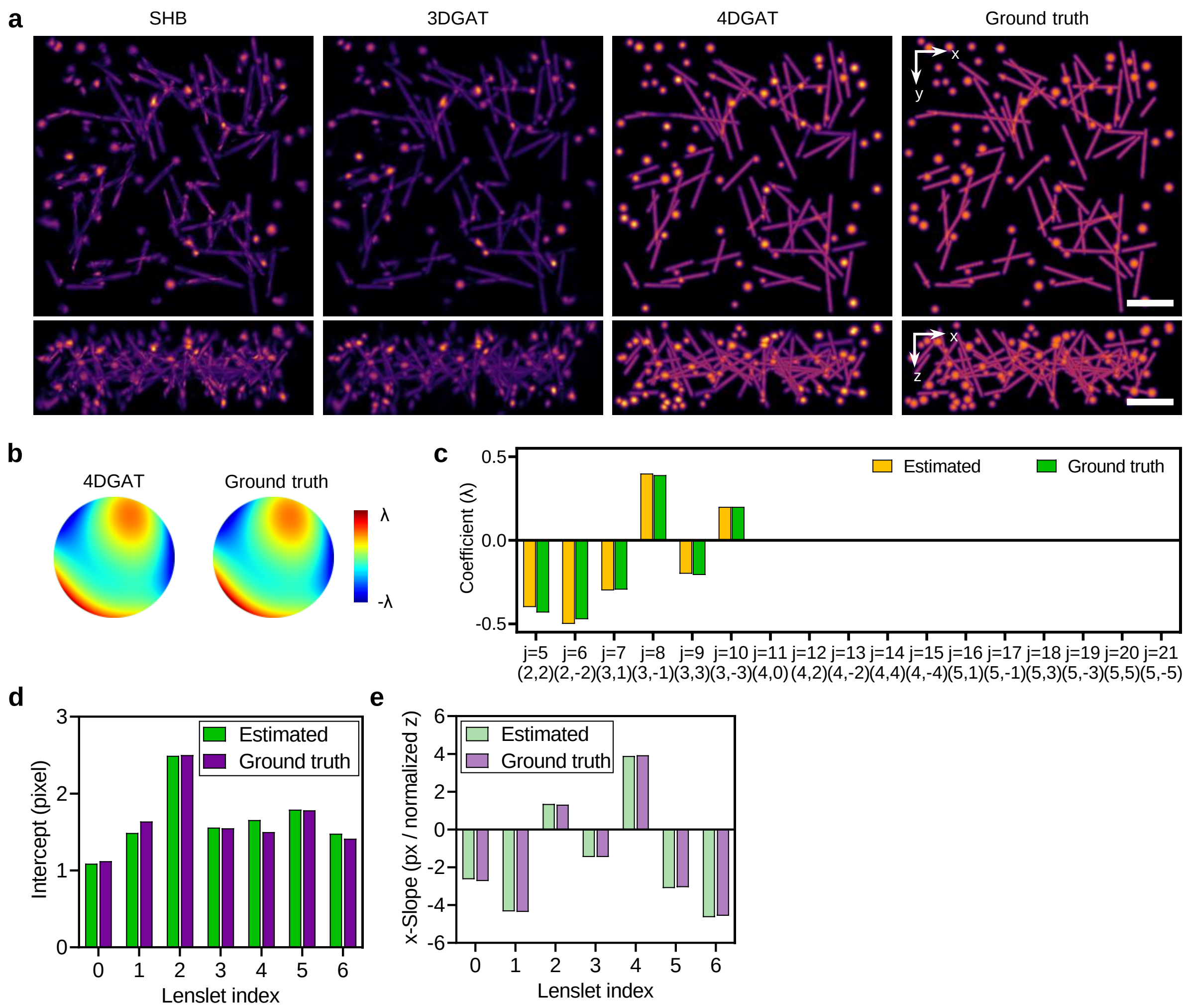

Fig. S8 | Joint recovery of pupil aberration and PSF shifts in a restricted aberration subspace. The ground-truth forward model contains both pupil aberration and per-lenslet PSF shifts, but only the first seven Zernike terms are active in the ground truth. During reconstruction, 4DGAT is allowed to optimize all 17 Zernike coefficients together with the per-lenslet shift parameters. In this simulated case, 4DGAT accurately reconstructs the object, recovers the active aberration terms, drives the inactive higher-order terms close to zero, and correctly estimates the per-lenslet PSF shifts. The result demonstrates successful joint parameter recovery in this lower-dimensional aberration regime; it does not establish general identifiability for the fully flexible joint model.

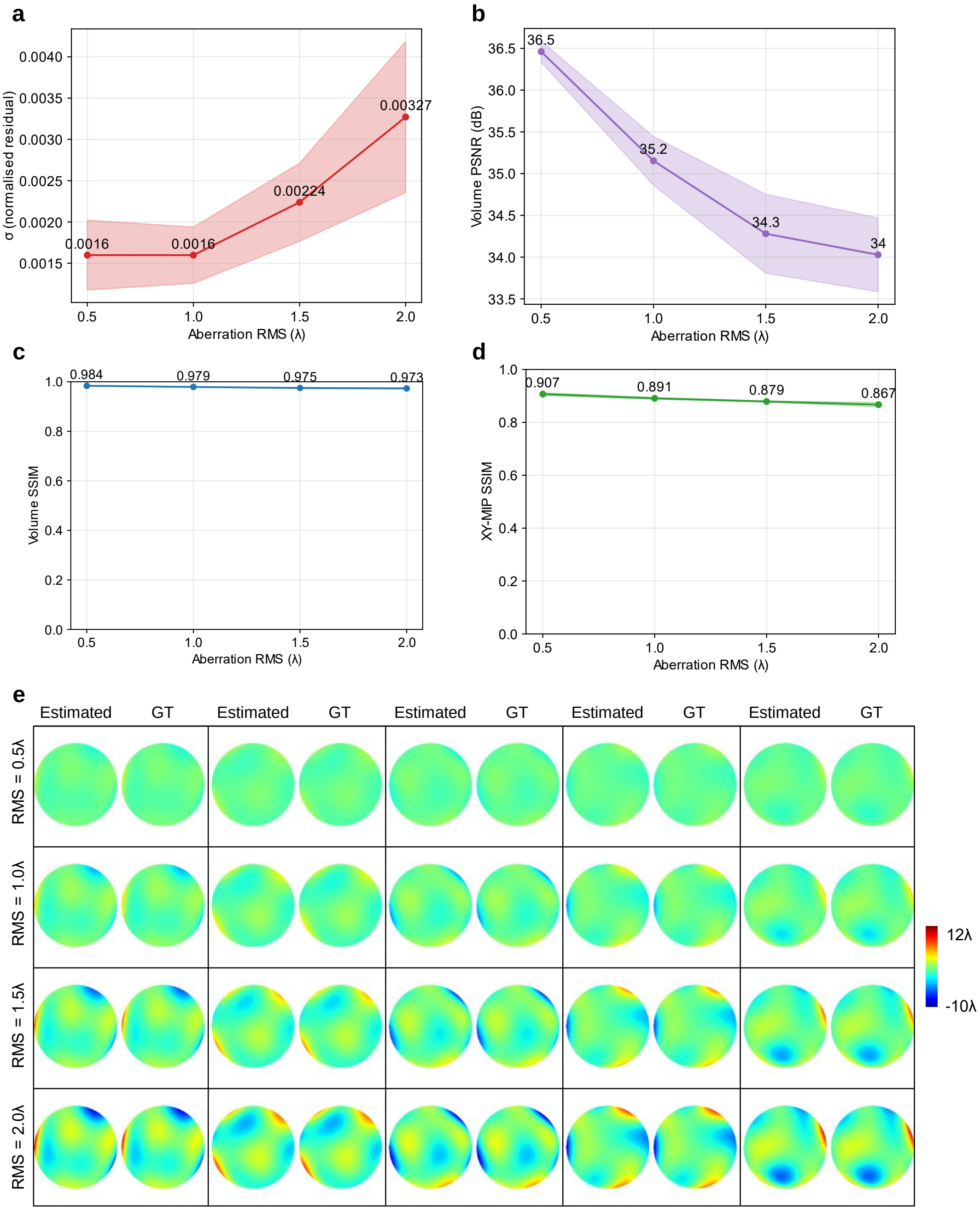

Fig. S9 | **RMS-sweep evaluation of aberration estimation and reconstruction fidelity.** 4DGAT was evaluated under increasing simulated pupil-aberration amplitudes from 0.5 to 2.0 λ RMS. For each RMS level we ran five independent trials using five different random seeds for the aberration sampling. **a,** Normalized residual aberration error, σ, remains low but increases with aberration strength. **b,** Volume PSNR decreases as RMS increases, indicating the expected degradation under stronger aberration. **c,d,** Volume SSIM and x–y MIP SSIM remain high across the sweep, showing that 4DGAT preserves reconstruction fidelity over this aberration range. Lines show the mean across five trials and shaded regions indicate s.d. **e,** Comparisons between estimated and ground-truth pupil wavefronts for each trial; each cell shows the 4DGAT-estimated wavefront paired with its corresponding GT wavefront. All pupil maps are displayed using a common color scale set by the RMS $= 2.0\,\lambda$ row.

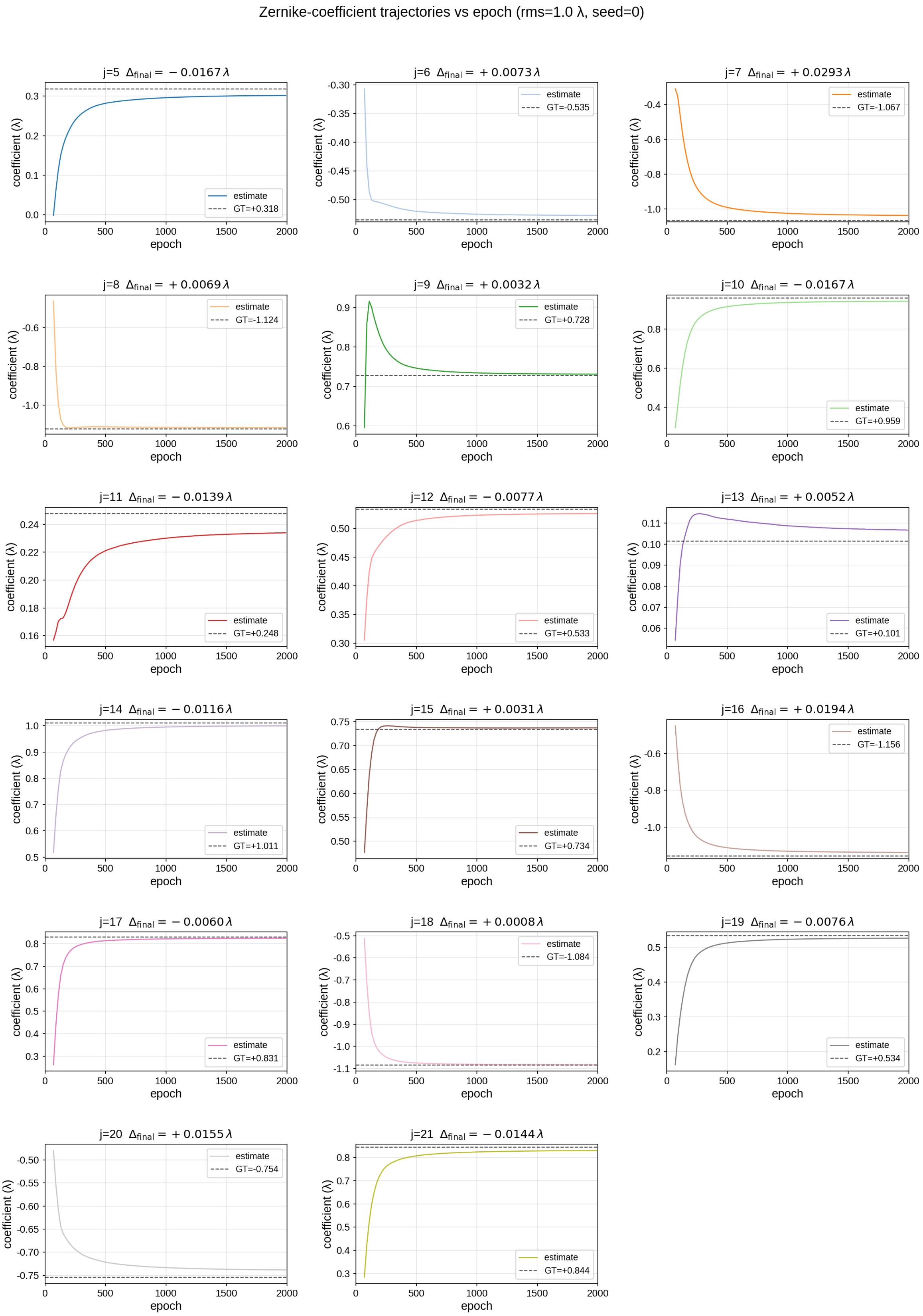

Fig. S10 | Zernike-coefficient convergence during 4DGAT static stage for the RMS $= 1.0\,\lambda$ aberration trial (one of the trials in Fig. S9). Each panel shows the trajectory of one recovered Zernike coefficient, from Noll indices $j = 5$ to $j = 21$, as a function of training epoch. Solid colored curves denote the estimated coefficients, and dashed black horizontal lines indicate the corresponding ground-truth values. The value reported in each panel title, Δ_{final}, is the final coefficient error, defined as $c_j(\text{final}) - c_j(\text{GT})$, in units of wavelength λ.

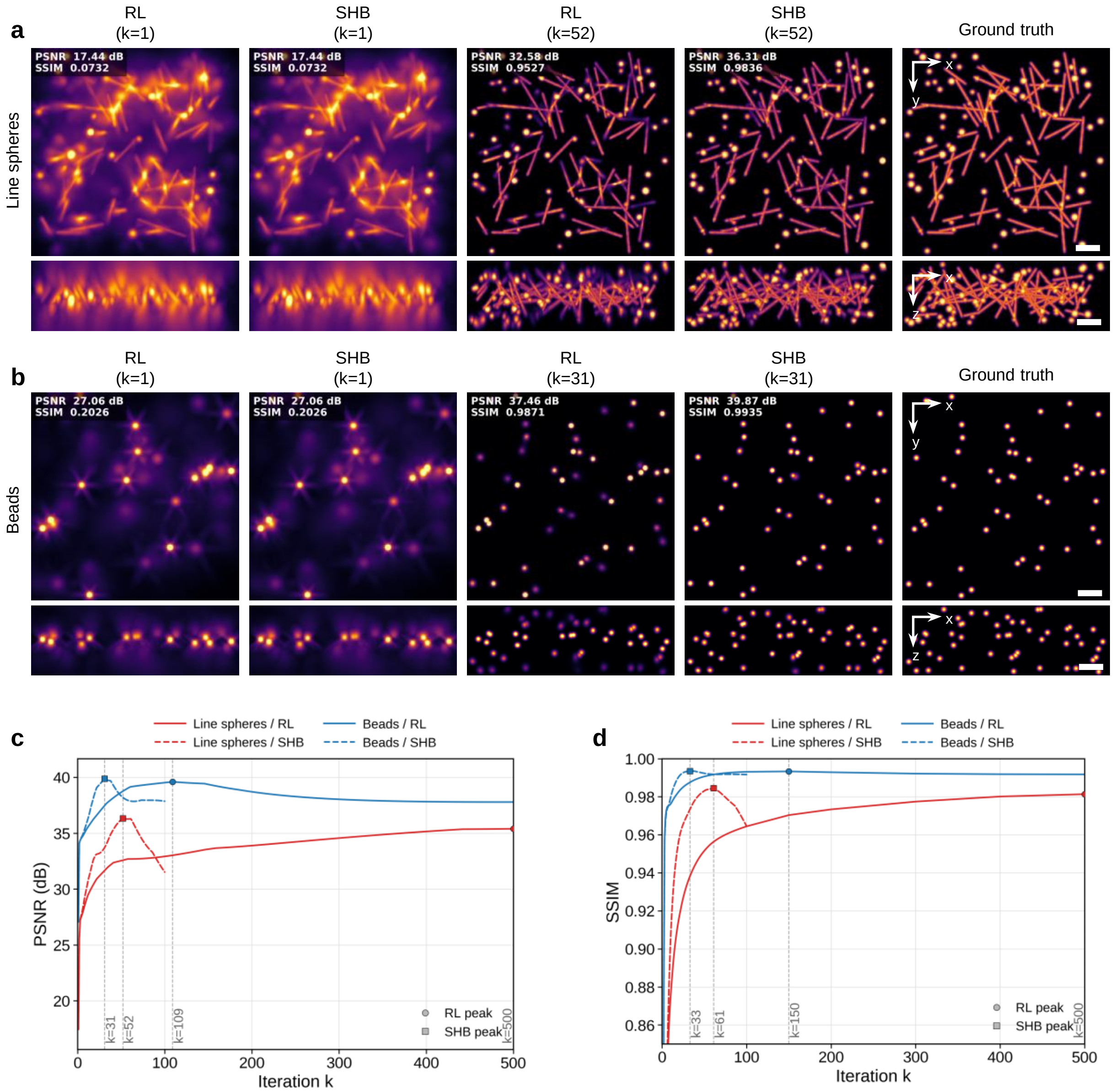

Fig. S11 | SHB accelerates FLFM deconvolution compared with RL. a,b, Representative FLFM reconstructions of simulated line-sphere (**a**) and bead (**b**) samples using Richardson–Lucy (RL) and SHB deconvolution at the indicated iterations, shown as *x*–*y* and *x*–*z* views and compared with the corresponding ground truth. **c,d,** PSNR (**c**) and SSIM (**d**) curves over iterations. SHB reaches high-fidelity reconstructions in substantially fewer iterations than RL, with earlier peak image quality and improved reconstruction sharpness.

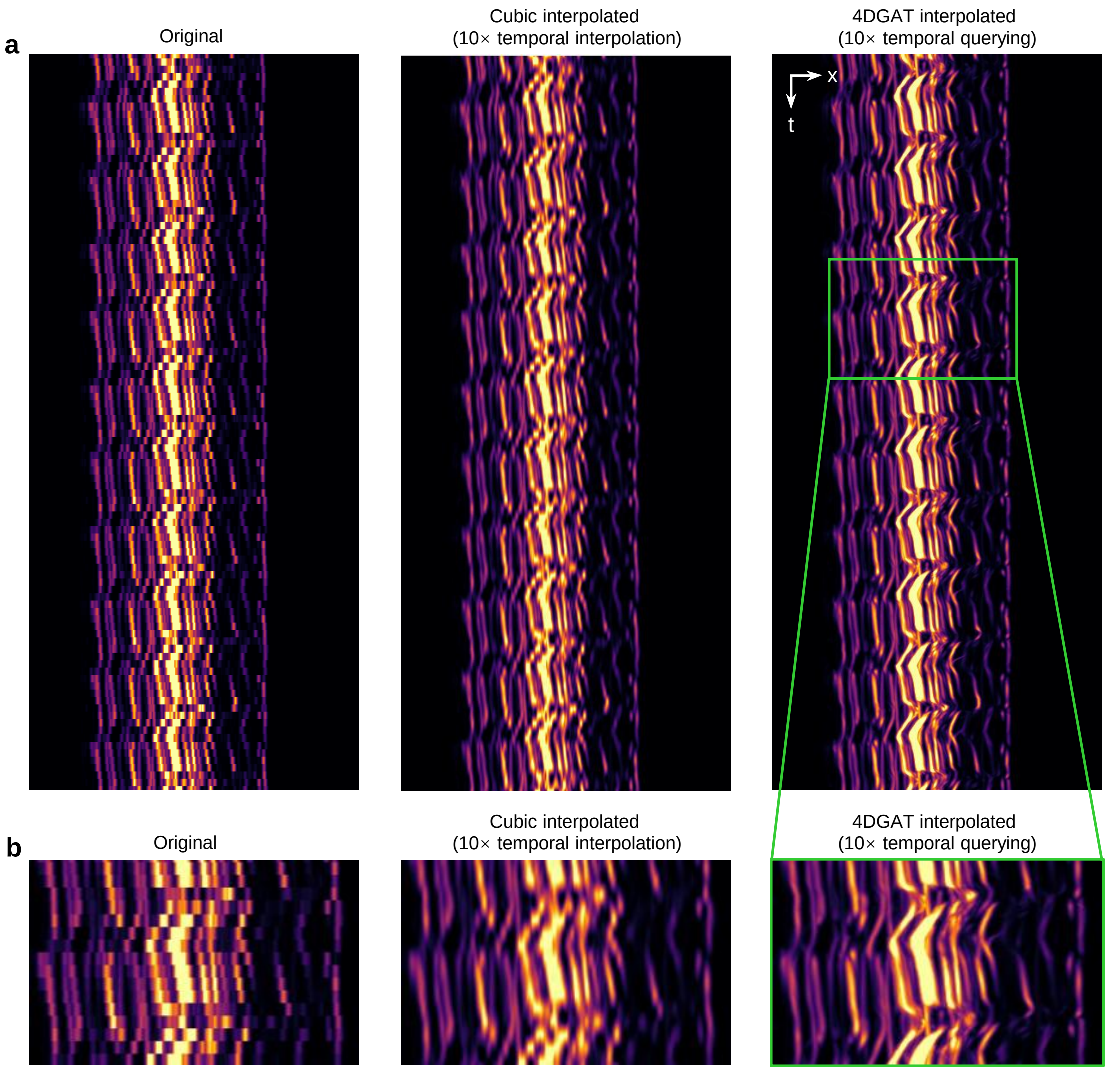

Fig. S12 | Continuous temporal interpolation of beating cardiac dynamics with 4DGAT. a, Kymographs extracted along a horizontal line through the beating zebrafish heart, comparing the original acquisition, conventional cubic temporal interpolation at 10× frame-rate upsampling, and 4DGAT interpolation queried at 10× denser timestamps. Cubic interpolation smooths the discrete frame sequence in image space, whereas 4DGAT samples the learned continuous temporal deformation field, yielding sharper and more coherent contraction–relaxation traces. **b,** Enlarged views of the boxed region in (**a**), highlighting that 4DGAT better preserves rapid motion trajectories and fine temporal structures compared with cubic interpolation.

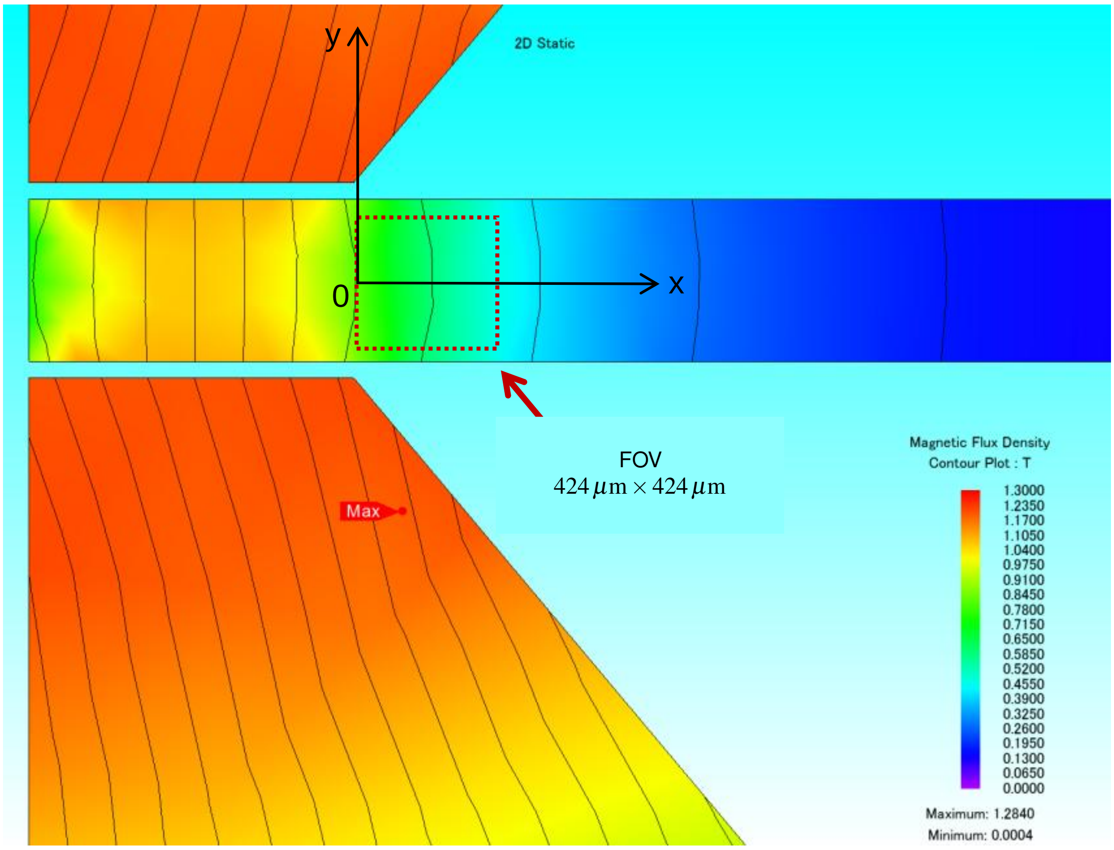

Fig. S13 | **Magnetic flux density distribution near the imaging region.** Two-dimensional magnetostatic simulation showing the spatial distribution of magnetic flux density $|B|$ in the x–y plane. The color map and contour lines represent $|B|$ in tesla. The red dashed square indicates the $424\,\mu\text{m} \times 424\,\mu\text{m}$ field of view used for imaging.

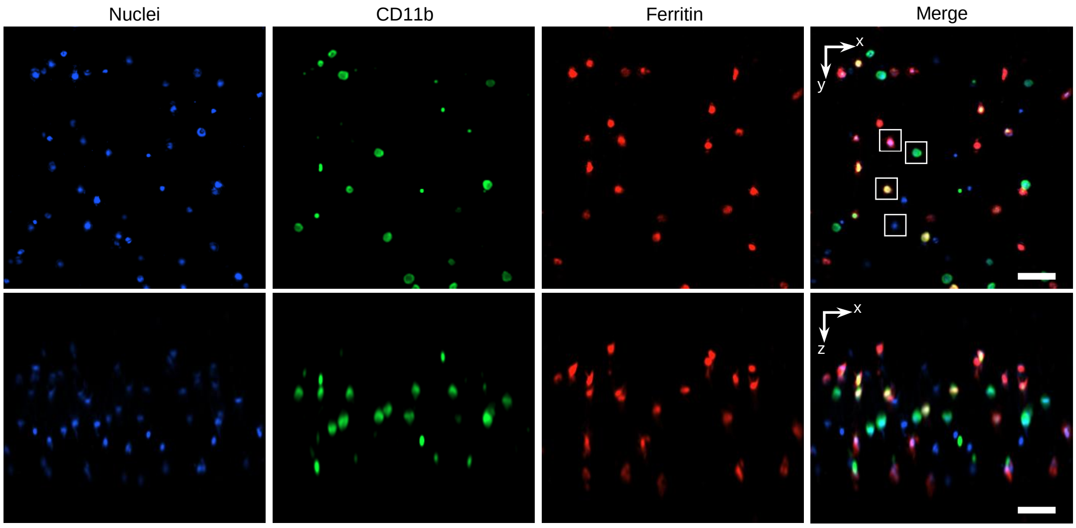

Fig. S14 | Multicolor 4DGAT reconstruction of an equal-ratio (1:1:1:1) mixture of four RAW264.7 macrophage populations representing the CD11b/Ferritin phenotypic classes at $t = 0\,$s. x–y and x–z MIPs of nuclei, CD11b, ferritin, and the merged results are shown. White boxes in the merged image indicate representative cells selected for the single-cell phenotyping examples shown in Fig. 5e. Scale bar, 50 μm.

Supplementary Note S1 Multi-objective-compatible optical design of the FLFM system

The Fourier light-field microscopy (FLFM) detection module supports objective switching across subcellular, cellular, and organ-scale imaging without changes to the downstream hardware. The fixed module comprised a 300-mm Fourier lens (ACT508-300-A, Thorlabs), a custom seven-lens hexagonal microlens array (MLA; APH-Q-P4000-R45.7, aµs; 4.0-mm pitch and 100-mm focal length), and a 2048 × 2048 sCMOS camera with 6.5-µm pixels (01-PRIME-BSI-EXP, Teledyne Photometrics). Compatibility across objectives follows from the pupil-image diameter at the MLA plane, which depends on the objective NA-to-magnification ratio (NA/M). The 20×/0.45 NA and 60×/1.25 NA objectives used in this study produce closely matched pupil diameters and therefore share the same Fourier lens, MLA, aperture, and camera. The same pupil-matching criterion also supports an intermediate 40×/0.90 NA configuration.

The 20×/0.45 NA long-working-distance objective (working distance, 6.6–7.8 mm) was selected specifically for the magnetic-actuation experiment in Fig. 5. The magnetic actuation assembly is relatively bulky, and the excitation coil is positioned close to the sample. An objective with a shorter working distance would mechanically interfere with, or contact, the coil and would prevent safe positioning of the sample. This objective therefore provided both the required mechanical clearance and the field coverage for cell-population- and organ-level imaging. The 60×/1.25 NA oil-immersion objective supported subcellular mitochondrial imaging, whereas the 40×/0.90 NA objective provided an intermediate cell-scale configuration.

For an infinity-corrected microscope, the pupil-image diameter at the MLA plane is[2]

$$D_{\text{pupil}} = 2f_{\text{FL}}\frac{\text{NA}}{M} \propto \frac{\text{NA}}{M}. \tag{S1}$$

Thus, a common Fourier module can be used when NA/M is conserved or remains closely matched. With $f_{\text{FL}} = 300$ mm, the 20×/0.45 NA and 40×/0.90 NA objectives have the same NA/$M = 0.0225$ and the same nominal 13.5-mm pupil image. The 60×/1.25 NA objective gives a 12.5-mm pupil image, 7.4% smaller, and remains compatible with the same seven-lens MLA and camera assembly. No downstream optical or mechanical component is changed between these configurations; only the objective and, where applicable, the immersion medium are changed.

The remaining design quantities in Supplementary Table S1 were calculated from the FLFM design relations reported by Guo et al.[2] Notably, we distinguish the nominal optical DOF from the effective reconstruction depth, because FLFM deconvolution can recover diffracted information beyond the axial range defined by the conventional PSF-FWHM criterion.[2,3] Objective-specific nominal PSFs and reconstruction grids were used, followed by the JPGM self-estimation described in Supplementary Note S2.

Table S1. Nominal multi-objective design values for the fixed FLFM module.

Objective	NA/M	D_{pupil} (mm)	M_{T}	p_{eff} (µm)	FOV (µm)	R_{xy} (µm)	R_z (µm)	DOF (µm)
20×/0.45	0.02250	13.50	6.667	0.9750	600	1.9125	7.1719	72.0
40×/0.90	0.02250	13.50	13.333	0.4875	300	0.9563	1.7930	18.0
60×/1.25	0.02083	12.50	20.000	0.3250	200	0.6375	0.7969	8.0

Note. Calculations use $\lambda = 510$ nm, $f_{\text{FL}} = 300$ mm, $f_{\text{ML}} = 100$ mm, an MLA pitch and outer-lenslet radial displacement of 4.0 mm, and a camera pixel pitch of 6.5 µm. Here, M_{T} is the total system magnification, p_{eff} is the effective sample-plane pixel pitch, FOV is the lateral field of view, R_{xy} and R_z are the nominal lateral and axial resolutions, and DOF is the nominal optical depth of field. The resolution and DOF values are design estimates calculated using the relations in Ref.[2]

Supplementary Note S2 Joint pupil-geometry model for Fourier light-field microscopy

S2.1 Definition and scope of JPGM

The joint pupil-geometry model (JPGM) is the differentiable FLFM image-formation operator that embeds the two learnable parameter groups defined by Eq. (2) of the main text:

$$\widehat{\mathbf{I}}_t = \mathrm{JPGM}(\mathbf{V}_t;\Theta_{\mathrm{P}},\Theta_{\mathrm{G}}), \tag{S2}$$

where $\mathbf{V}_t$ is the voxelized fluorescence volume at time t; the semicolon separates the volume supplied to the operator from the JPGM parameter groups. Here $v = 1,\ldots,N_{\mathrm{view}}$ indexes the lenslet views, consistent with Eq. (2) of the main text, and $N_{\mathrm{view}} = 7$ in the present system. The ordered parameter group $\Theta_{\mathrm{P}} = (c_j)_{j\in J}$ contains the Zernike coefficients in the complex pupil $P(x,y)$, and $\Theta_{\mathrm{G}} = (\mathbf{b},\mathbf{s})$ contains the per-lenslet intercept and slope vectors in $\Delta x_v(z)$ and $\Delta y_v(z)$, where $\mathbf{b} = (b_v)_{v=1}^{N_{\mathrm{view}}}$, $\mathbf{s} = (s_v)_{v=1}^{N_{\mathrm{view}}}$, $b_v = (b_{v,x}, b_{v,y})$, and $s_v = (s_{v,x}, s_{v,y})$. Thus, Θ_{P} and Θ_{G} correspond directly to the pupil-phase and PSF-center-shift parameters of Eq. (2), and together form the optical component $\{c_j, b_v, s_v\}$ of Θ_{stat}. Here, "pupil" refers to objective-pupil phase aberrations, whereas "geometry" refers to residual per-lenslet PSF-center offsets and their first-order variation with depth. JPGM is a compact image-formation model intended to reproduce the observable per-view, per-depth PSF response; the joint parameterization does not imply unique recovery of every underlying physical misalignment.

Equation (S2) defines the complete sensor-level prediction. Its elemental patch associated with lenslet view v is

$$\big[\widehat{\mathbf{I}}_t\big]_v(u_x,u_y) = \mathscr{F}^{-1}\left[\sum_{z=1}^{N_z}\mathscr{F}\{\mathbf{V}_t(z)\}\,\mathrm{OTF}_{v,z}(\Theta_{\mathrm{P}},\Theta_{\mathrm{G}})\right](u_x,u_y), \tag{S3}$$

where $[\cdot]_v$ denotes restriction to the lenslet-view patch indexed by v. The complete prediction is obtained by placing these elemental patches at their nominal sensor locations. The JPGM-derived PSF $H_{v,z}$ is generated by (i) constructing the learnable complex pupil, (ii) propagating the field by the angular-spectrum method, (iii) converting field to intensity, and (iv) sampling the resulting intensity PSF along a learnable lenslet-view geometry trajectory.

S2.2 Pupil modeling

We represent the system's complex pupil as the product of an ideal amplitude mask and a phase that captures wavefront aberrations.

Pupil coordinates. Let $M(x,y) \in \{0,1\}^{H\times W}$ denote the binary aperture mask and let (c_x, c_y) and R be the aperture center and radius inferred from M. Define the normalized radial and angular coordinates

$$\rho(x,y) \;=\; \frac{\sqrt{(x-c_x)^2+(y-c_y)^2}}{R}, \qquad \theta(x,y) \;=\; \mathrm{arctan2}(y-c_y,\, x-c_x). \tag{S4}$$

Zernike phase. We parameterize the low-order wavefront on the aperture by a linear combination of Zernike polynomials under the Noll linear index j:[4]

$$\varphi_Z(x,y) \;=\; 2\pi\sum_{j\in J} c_j Z_j\big(\rho(x,y),\,\theta(x,y)\big), \qquad J=\{5,\ldots,21\}, \quad K=17, \tag{S5}$$

where Z_j employs the classical Zernike radial polynomial. The sum starts at $j = 5$ because $j = 1,\ldots,4$ correspond to piston, tip, tilt, and defocus, which are either trivial or absorbed into the nominal axial calibration. The coefficients c_j are expressed in wavelengths and are treated as learnable parameters. In our default configuration the active Noll index set is $J = \{5,\ldots,21\}$, giving $K = 17$ coefficients.

Complex pupil. The complex pupil is then

$$P(x,y) \;=\; M(x,y)\,\exp\big(\mathrm{i}\,\varphi_Z(x,y)\big), \tag{S6}$$

where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit.

S2.3 Depth-dependent field propagation

For each depth index z, $E_{\mathrm{MLA},z}(x,y)$ represents the complex field before the pupil plane. After modulation by the pupil

$$E_{0,z}(x,y) \;=\; E_{\mathrm{MLA},z}(x,y)\cdot P(x,y), \tag{S7}$$

the field is propagated to the camera plane via the angular-spectrum method (ASM).[5] We denote the propagation factor by $H_{\mathrm{ASM}}(f_x,f_y)$, with spatial frequencies (f_x,f_y) defined on the discrete pupil grid. The camera-plane field is

$$E_{\mathrm{cam},z}(x,y) \;=\; \mathscr{F}^{-1}\Big\{\mathscr{F}[E_{0,z}](f_x,f_y)\cdot H_{\mathrm{ASM}}(f_x,f_y)\Big\}(x,y). \tag{S8}$$

From field to intensity. Under incoherent fluorescence the PSF is the squared-magnitude field

$$I_z(x,y) \;=\; \left|E_{\mathrm{cam},z}(x,y)\right|^2, \tag{S9}$$

and the final PSF $H_z(x,y)$ is the result of normalizing $I_z(x,y)$ layer-by-layer along the z-direction.

S2.4 Per-view PSF extraction with shift calibration

Each of the N_{view} lenslets contributes one elemental image centered at a known lenslet position $(c_{v,x}, c_{v,y})$. To obtain the per-view PSF we crop a patch of half-width r (default $r = 80$ pixels) around that position. In practice the effective center of each view drifts slightly with depth because of (i) residual geometric misalignment between the nominal and actual lenslet positions and (ii) first-order axial-dependent aberrations. We model this drift as a linear function of a normalized depth coordinate

$$\xi(z) \;=\; \frac{2(z-1)}{N_z-1} - 1 \;\in\; [-1,1], \tag{S10}$$

assigning to each view a learnable intercept $b_v = (b_{v,x}, b_{v,y}) \in \mathbb{R}^2$ and a learnable slope $s_v = (s_{v,x}, s_{v,y}) \in \mathbb{R}^2$:

$$\Delta x_v(z) = b_{v,x} + s_{v,x}\,\xi(z), \qquad \Delta y_v(z) = b_{v,y} + s_{v,y}\,\xi(z). \tag{S11}$$

The per-view per-depth PSF is then obtained as a sub-pixel bilinear crop

$$H_{v,z}(u_x,u_y) \;=\; H_z\big(c_{v,x} + \Delta x_v(z) + u_x,\; c_{v,y} + \Delta y_v(z) + u_y\big), \tag{S12}$$

implemented via differentiable grid-sampling so that $\{b_v, s_v\}$ can be optimized jointly with the Zernike coefficients as part of Θ_{stat}. Keeping the model linear in $\xi(z)$ keeps the number of calibration parameters small and is sufficient in the paraxial regime. A geometric illustration of the intercept-slope shift model is shown in Fig. S4.

Lenslet-center initialization. The nominal centers $(c_{v,x}, c_{v,y})$ are initialized from a brightfield image of the MLA. An image-gradient edge map is supplied to a circular Hough transform to locate the seven lenslet apertures, and the detected circle centers define the nominal crop locations. Residual center offsets caused by finite pixel sampling, circle fitting, optical alignment and MLA fabrication tolerances are subsequently represented by b_v and s_v and optimized jointly with the canonical Gaussian scene (Fig. S5).

Supplementary Note S3 4DGAT Algorithm

Building on the forward model of Section S2, 4DGAT reconstructs a time-varying fluorescent volume from an FLFM image sequence $\{\mathbf{I}_t\}_{t=1}^{T}$ using an anisotropic, time-deformed Gaussian mixture. The pipeline is trained in two stages: a static stage that jointly recovers the canonical scene at the reference time t_{ref} (the middle frame) and estimates the JPGM parameters, and a dynamic stage that holds the canonical scene and JPGM fixed while learning a 4D deformation field that carries the canonical Gaussians to every other time instant. The detailed workflow is also displayed in Fig. S3.

S3.1 Canonical Gaussian representation

We represent the reference-frame scene by a set of N anisotropic 3D Gaussians[6]

$$G_0 \;=\; \Big\{ \big(\mu_n \in \mathbb{R}^3,\; s_n \in \mathbb{R}^3_{>0},\; r_n \in \mathbb{S}^3,\; i_n \in \mathbb{R}_{\geq 0}\big) \Big\}_{n=1}^{N}, \tag{S13}$$

where μ_n is the center (world coordinates), s_n the axis-aligned scale, r_n a unit quaternion encoding orientation, and i_n represents fluorescence intensity.

Covariance. Let $S_n = \mathrm{diag}(s_n)$ and let $R_n \in SO(3)$ be the rotation matrix built from the unit quaternion r_n. The covariance factors as

$$\mathbf{\Sigma}_n \;=\; R_n S_n S_n^{\top} R_n^{\top}, \tag{S14}$$

which follows the canonical anisotropic covariance parameterization of 3D Gaussian splatting.[6] Equation (S14) is automatically positive semi-definite and avoids storing nine covariance entries directly.

Activations. To keep every parameter on the real line during optimization, the raw tensors $\tilde{s}_n, \tilde{r}_n, \tilde{i}_n$ stored by the model are mapped to their physically valid ranges by

$$\begin{aligned} s_n &= \sigma(\tilde{s}_n)\cdot(s_{\max} - s_{\min}) + s_{\min}, \\ r_n &= \tilde{r}_n / \|\tilde{r}_n\|_2, \\ i_n &= \mathrm{softplus}(\tilde{i}_n), \end{aligned} \tag{S15}$$

where σ is the sigmoid. By default $s_{\min} = 0.8$ and $s_{\max} = 20$ in voxel units.

S3.2 Gaussian voxelization

To feed the Gaussian set into JPGM (Supplementary Note S2) we rasterize G_0 into a dense voxel grid $\mathbf{V}$ of shape (N_z, N_y, N_x) centered at a world location $c_V \in \mathbb{R}^3$ with anisotropic voxel size $s_V = (s_{V,x}, s_{V,y}, s_{V,z})$. Writing $\mathbf{x} \in \mathbb{R}^3$ for a voxel center, the rendered volume in continuous-domain is

$$\mathbf{V}(\mathbf{x}) \;=\; \sum_{n=1}^{N} i_n \exp\Big(-\tfrac{1}{2}(\mathbf{x}-\mu_n)^{\top}\mathbf{\Sigma}_n^{-1}(\mathbf{x}-\mu_n)\Big). \tag{S16}$$

A custom CUDA kernel implements Eq. (S16) and its gradients with respect to $\{\mu_n, s_n, r_n, i_n\}$. The Gaussian-to-volume voxelization routine is adapted from the differentiable voxelizer of R2-Gaussian,[7] which was developed for Gaussian-based tomographic reconstruction. In 4DGAT, we use this voxel-level accumulation strategy to render a fluorescence-density volume, while replacing the original X-ray/radiative projection model with JPGM described in Supplementary Note S2. Gaussians whose 3σ ellipsoid does not overlap the grid are skipped for efficiency. At time t, we apply exactly the same rasterizer to the deformed set G_t to obtain $\mathbf{V}_t$.

S3.3 HexPlane 4D feature field

To efficiently encode how each canonical Gaussian deforms in time we follow the factorized HexPlane representation.[8] The 4D ambient space (x,y,z,t) is factorized into the six 2D planes indexed by the unordered pairs

$$\mathscr{I}^2 \;=\; \big\{(x,y),\,(x,z),\,(y,z),\,(x,t),\,(y,t),\,(z,t)\big\}, \tag{S17}$$

with the three spatial planes $(x,y),(x,z),(y,z)$ covering purely spatial correlations and the three spatiotemporal planes $(x,t),(y,t),(z,t)$ capturing spatiotemporal variation.

Feature query. Given a canonical Gaussian center $\mu_n = (x,y,z)$ and a timestamp $t \in [0,1]$ mapped to normalized coordinates $(\bar{x},\bar{y},\bar{z},\bar{t}) \in [-1,1]^4$, the HexPlane feature is the Hadamard product of bilinear samples across the six planes of each level, concatenated across levels:

$$f_{\mathrm{HP}}(\mu_n,t) \;=\; \bigoplus_{l=1}^{3} \;\bigodot_{(a,b)\in\mathscr{I}^2} \mathrm{GRID\_SAMPLE}\big(P^l_{(a,b)},\, p_{n,a},\, p_{n,b}\big) \;\in\; \mathbb{R}^{3C}. \tag{S18}$$

Here $\bigodot$ denotes elementwise product across planes, $\bigoplus$ denotes channel-wise concatenation across levels, and GRID_SAMPLE is the standard differentiable bilinear sampler.

S3.4 Deformation MLP

The HexPlane feature f_{HP} is fed through a small deformation MLP that outputs a per-Gaussian deformation vector

$$D_n(t) \;=\; \big(\Delta\mu_n,\, \Delta s_n,\, \Delta r_n,\, \Delta i_n\big) \;\in\; \mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^4\times\mathbb{R}. \tag{S19}$$

The MLP consists of a shared trunk ϕ_θ of depth $D=2$ and width $W=128$ with ReLU activations, followed by four parallel heads h_μ, h_s, h_r, h_i:

$$D_n(t) \;=\; \big(h_\mu, h_s, h_r, h_i\big)\circ\phi_\theta\big(f_{\mathrm{HP}}(\mu_n,t)\big). \tag{S20}$$

Together, f_{HP} and f_{MLP} constitute the dynamic-stage parameters

$$\Theta_{\mathrm{dyn}} \;=\; \{f_{\mathrm{HP}},\, f_{\mathrm{MLP}}\}. \tag{S21}$$

Deformation-application operator. The deformation field carries G_0 to its time-t counterpart G_t by applying $D_n(t)$ to each Gaussian attribute-wise:

$$\begin{aligned} \mu_n(t) &= \mu_n + \Delta\mu_n, \quad & s_n(t) &= s_n + \Delta s_n,\\ r_n(t) &= \Delta r_n \otimes r_n, \quad & i_n(t) &= i_n + \Delta i_n, \end{aligned} \tag{S22}$$

where $\otimes$ denotes quaternion multiplication. We further enforce $\Delta\mu_n(t_{\mathrm{ref}})=0$, $\Delta s_n(t_{\mathrm{ref}})=0$, $\Delta i_n(t_{\mathrm{ref}})=0$, and $\Delta r_n(t_{\mathrm{ref}})=(1,0,0,0)$, so that $G_{t_{\mathrm{ref}}}\equiv G_0$, anchoring the canonical representation to the reference-frame state.

S3.5 Loss design

We denote the training iteration by k, with k_{stat} and k_{dyn} the maximum iteration counts for the static and dynamic stages respectively.

S3.5.1 Image-domain fidelity terms

For any pair of images A,B with size $H\times W$, we use three pixel-level similarity measures

$$\mathscr{L}^{(p)}_{\mathrm{pix}}(A,B) \;=\; \tfrac{1}{HW}\sum_{x,y}\big|A(x,y)-B(x,y)\big|^p, \quad p\in\{1,2\}, \tag{S23}$$

$$\mathscr{L}_{\mathrm{fft}}(A,B) \;=\; \tfrac{1}{HW}\sum_{f_x,f_y}\big|\mathscr{F}(A)(f_x,f_y)-\mathscr{F}(B)(f_x,f_y)\big|, \tag{S24}$$

$$\mathscr{L}_{\mathrm{ssim}}(A,B) \;=\; 1-\mathrm{SSIM}(A,B). \tag{S25}$$

Here $\mathscr{L}^{(p)}_{\mathrm{pix}}$ is the image-space data term (MAE for $p=1$ and MSE for $p=2$), $\mathscr{L}_{\mathrm{fft}}$ penalizes mismatch in the Fourier domain and emphasizes high-frequency structure, and $\mathscr{L}_{\mathrm{ssim}}$ is the classical structural-similarity loss.

S3.5.2 Static stage

With only the reference frame in use, the static-stage parameters

$$\Theta_{\mathrm{stat}} \;=\; \big\{G_0,\, c_j,\, b_v,\, s_v\big\} \tag{S26}$$

are jointly optimized by minimizing

$$\boxed{\mathscr{L}_{\mathrm{stat}} \;=\; \mathscr{L}^{(p)}_{\mathrm{pix}}\big(\widehat{\mathbf{I}}_{\mathrm{ref}},\mathbf{I}_{\mathrm{ref}}\big) + \lambda_F(k)\,\mathscr{L}_{\mathrm{fft}}\big(\widehat{\mathbf{I}}_{\mathrm{ref}},\mathbf{I}_{\mathrm{ref}}\big) + \lambda_S\,\mathscr{L}_{\mathrm{ssim}}\big(\widehat{\mathbf{I}}_{\mathrm{ref}},\mathbf{I}_{\mathrm{ref}}\big)}, \tag{S27}$$

with a time-annealed Fourier weight

$$\lambda_F(k) \;=\; \lambda_F^0\cdot\Big(0.8\,\big(1-\tfrac{k}{k_{\mathrm{stat}}}\big)+0.2\Big), \tag{S28}$$

which is strong early on (when coarse structure matters most) and decays to 20% of its peak by the end of the stage. The static stage is additionally responsible for Gaussian densification and pruning: every Δk iterations, Gaussians with large positional gradients are cloned or split and those with vanishing intensity i_n are pruned, following the densification and pruning strategy of 3D Gaussian splatting.[6]

S3.5.3 Dynamic stage

In the dynamic stage G_0 and the JPGM parameters $\{c_j, b_v, s_v\}$ are held fixed, and only Θ_{dyn} is optimized against the full video $\{\mathbf{I}_t\}_{t=1}^{T}$. The total loss is

$$\begin{aligned}
\mathscr{L}_{\text{img}}^{(t)} &= \tfrac{1}{1+\lambda_S}\Big[\mathscr{L}_{\text{pix}}^{(p)}(\widehat{\mathbf{I}}_t, \mathbf{I}_t) + \lambda_F\, \mathscr{L}_{\text{fft}}(\widehat{\mathbf{I}}_t, \mathbf{I}_t) + \lambda_S\, \mathscr{L}_{\text{ssim}}(\widehat{\mathbf{I}}_t, \mathbf{I}_t)\Big], \\
\mathscr{L}_{\text{reg}} &= \lambda_{\text{TV}}\, \mathscr{L}_{\text{TV}} + \lambda_{\text{sm}}\, \mathscr{L}_{\text{sm}} + \lambda_{11}\, \mathscr{L}_{11}, \\
&\boxed{\mathscr{L}_{\text{dyn}} = \mathscr{L}_{\text{img}}^{(t)} + \mathscr{L}_{\text{reg}}}.
\end{aligned} \tag{S29}$$

The individual terms are defined as follows.

Plane total variation. Applied to every HexPlane across all resolution levels to encourage spatial and spatiotemporal smoothness of the feature field

$$\mathscr{L}_{\text{TV}} = \sum_{l=1}^{3} \sum_{(a,b)\in\mathscr{I}^2} \text{TV}\big(P_{(a,b)}^{l}\big), \qquad \text{TV}(P) = \tfrac{1}{|P|}\sum_{i,j,c}\Big[(P_{c,i+1,j} - P_{c,i,j})^2 + (P_{c,i,j+1} - P_{c,i,j})^2\Big]. \tag{S30}$$

Temporal second-order smoothness. Applied only on the spatiotemporal planes $\mathscr{P}_{\text{st}} = \{(x,t), (y,t), (z,t)\}$ to suppress jitter:

$$\mathscr{L}_{\text{sm}} = \sum_{l,\,(a,b)\in\mathscr{P}_{\text{st}}} \tfrac{1}{|P_{(a,b)}^{l}|} \sum_{c,i,j}\Big(P_{(a,b)}^{l}[c,i,j+1] - 2P_{(a,b)}^{l}[c,i,j] + P_{(a,b)}^{l}[c,i,j-1]\Big)^2, \tag{S31}$$

where j indexes the time axis of each spatiotemporal plane.

Temporal L_1 sparsity. To keep the spatiotemporal planes close to their unit initialization (identity deformation) and promote sparse temporal variation:

$$\mathscr{L}_{11} = \sum_{l,\,(a,b)\in\mathscr{P}_{\text{st}}} \tfrac{1}{|P_{(a,b)}^{l}|} \sum_{c,i,j}\big|\,1 - P_{(a,b)}^{l}[c,i,j]\,\big|. \tag{S32}$$

Default weights. Unless explicitly stated, we use λ_F^0 and λ_S from configuration ($\lambda_S = 0$ by default, λ_F^0 task-dependent), $\lambda_{\text{TV}} = 10^{-4}$, $\lambda_{\text{sm}} = 10^{-3}$, $\lambda_{11} = 10^{-4}$.

Supplementary Note S4 Local cross-information between Zernike aberrations and lenslet intercept-slope calibration

This section describes the local interaction between the pupil and geometry parameter groups of JPGM. The goal is to identify which image-response directions are shared between them and how a small geometry-calibration error can leak into the recovered Zernike coefficients.

S4.1 Purpose, notation, and the forward path

Supplementary Note S2 defines the depth-dependent lenslet shift as

$$\Delta x_v(z) = b_{v,x} + s_{v,x}\,\xi(z), \qquad \Delta y_v(z) = b_{v,y} + s_{v,y}\,\xi(z). \tag{S33}$$

The Zernike phase is

$$\varphi_Z(x,y) \;=\; 2\pi \sum_{j\in J} c_j\, Z_j\big(\rho(x,y),\theta(x,y)\big), \qquad J=\{5,\dots,21\}, \quad K=17. \tag{S34}$$

For local differentiation, the two parameter groups defined in Supplementary Note S2 are written in an ordered-vector form as

$$\Theta_{\mathrm{P}} \;=\; (c_j)_{j\in J} \in \mathbb{R}^K, \qquad \Theta_{\mathrm{G}} \;=\; (\mathbf{b},\mathbf{s}) \in \mathbb{R}^{4N_{\mathrm{view}}}. \tag{S35}$$

Here $\mathbf{b}$ contains all intercepts $b_v = (b_{v,x}, b_{v,y})$ and $\mathbf{s}$ contains all slopes $s_v = (s_{v,x}, s_{v,y})$. These are the same Θ_{P} and Θ_{G} used in Eq. (S2), with their dimensions made explicit. The final sensor prediction is written as the vector

$$\mathbf{y}(\Theta_{\mathrm{P}},\Theta_{\mathrm{G}}) \;=\; \mathrm{vec}(\widehat{\mathbf{I}}_t). \tag{S36}$$

For the local analysis, the current object estimate $\mathbf{V}_t$ is held fixed, and the forward prediction is regarded as a differentiable map

$$\mathbf{y}(\Theta_{\mathrm{P}},\Theta_{\mathrm{G}}) \;=\; \mathrm{vec}\big[\mathrm{JPGM}(\mathbf{V}_t;\Theta_{\mathrm{P}},\Theta_{\mathrm{G}})\big]. \tag{S37}$$

Here JPGM denotes the same differentiable image-formation model introduced in Supplementary Note S2. The volume $\mathbf{V}_t$ is held fixed, and only the parameter groups $(\Theta_{\mathrm{P}},\Theta_{\mathrm{G}})$ are differentiated against in the local analysis below.

S4.2 Local image responses of the two parameter groups

Around the current optimization point, the predicted image can be locally linearized using the standard first-order approximation:[9]

$$\Delta\mathbf{y} \;\approx\; J_{\mathrm{P}}\Delta\Theta_{\mathrm{P}} + J_{\mathrm{G}}\Delta\Theta_{\mathrm{G}}, \qquad J_{\mathrm{P}} = \frac{\partial \mathbf{y}}{\partial \Theta_{\mathrm{P}}}, \quad J_{\mathrm{G}} = \frac{\partial \mathbf{y}}{\partial \Theta_{\mathrm{G}}}. \tag{S38}$$

Each column of J_{P} is the image change caused by one Zernike coefficient. Each column of J_{G} is the image change caused by one intercept or slope parameter. Since b_v and s_v move the sub-pixel crop center in Eq. (S33), their local derivatives are image-gradient directions

$$\begin{aligned}
\frac{\partial H_{v,z}}{\partial b_{v,x}} &\approx \partial_x H_{v,z}, & \frac{\partial H_{v,z}}{\partial b_{v,y}} &\approx \partial_y H_{v,z},\\
\frac{\partial H_{v,z}}{\partial s_{v,x}} &\approx \xi(z)\,\partial_x H_{v,z}, & \frac{\partial H_{v,z}}{\partial s_{v,y}} &\approx \xi(z)\,\partial_y H_{v,z}.
\end{aligned} \tag{S39}$$

The meaning of J_{P} is different. A perturbation of a Zernike coefficient changes the pupil phase

$$P(x,y) \;=\; M(x,y)\exp\big(\mathrm{i}\,\varphi_Z(x,y)\big), \qquad \frac{\partial P}{\partial c_j} \;=\; \mathrm{i}\,2\pi Z_j P. \tag{S40}$$

J_{P} describes aberration-induced shape changes, while J_{G} describes shift-like and shear-like changes.

S4.3 Block Hessian and cross-information criterion

For a local analysis, the image loss can be approximated by the Gauss–Newton quadratic form

$$\mathscr{L}(\Delta\Theta_{\mathrm{P}},\Delta\Theta_{\mathrm{G}}) \;=\; \tfrac{1}{2}\big\|W^{1/2}(J_{\mathrm{P}}\Delta\Theta_{\mathrm{P}} + J_{\mathrm{G}}\Delta\Theta_{\mathrm{G}} - \mathbf{e})\big\|_2^2 + \tfrac{1}{2}\Delta\Theta_{\mathrm{P}}^{\top}\Lambda_{\mathrm{P}}\Delta\Theta_{\mathrm{P}} + \tfrac{1}{2}\Delta\Theta_{\mathrm{G}}^{\top}\Lambda_{\mathrm{G}}\Delta\Theta_{\mathrm{G}}, \tag{S41}$$

where $\mathbf{e}$ is the current residual to be explained. The matrix $W \succeq 0$ is the local image-space metric induced by the selected loss. For a pure pixel L_2 loss, $W = I$. For pixel L_1, FFT loss, or SSIM terms, W can be interpreted as the local quadratic approximation of that loss. The matrices Λ_{P} and Λ_{G} represent explicit regularization on Zernike coefficients and on intercept-slope parameters.

Expanding Eq. (S41) gives the local Hessian in block form:

$$H = \begin{bmatrix} A & C \\ C^{\top} & D \end{bmatrix}, \qquad A = J_{\mathrm{P}}^{\top} W J_{\mathrm{P}} + \Lambda_{\mathrm{P}},\; C = J_{\mathrm{P}}^{\top} W J_{\mathrm{G}},\; D = J_{\mathrm{G}}^{\top} W J_{\mathrm{G}} + \Lambda_{\mathrm{G}}. \tag{S42}$$

The blocks A and D are self-information terms for the Zernike and shift parameters, respectively. The off-diagonal block

$$C = J_{\mathrm{P}}^{\top} W J_{\mathrm{G}} \tag{S43}$$

is the *cross-information block*. Its entries are weighted inner products between local image-response directions:

$$C_{\alpha\beta} = \left\langle \frac{\partial \mathbf{y}}{\partial (\Theta_{\mathrm{P}})_{\alpha}}, \frac{\partial \mathbf{y}}{\partial (\Theta_{\mathrm{G}})_{\beta}} \right\rangle_{W}. \tag{S44}$$

If $C = 0$, the two parameter groups are locally orthogonal in the selected image metric, and the first-order normal equations decouple. If C is nonzero, the update for one group depends on the other:

$$\begin{bmatrix} A & C \\ C^{\top} & D \end{bmatrix} \begin{bmatrix} \Delta\Theta_{\mathrm{P}} \\ \Delta\Theta_{\mathrm{G}} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_{\mathrm{P}} \\ \mathbf{g}_{\mathrm{G}} \end{bmatrix}, \qquad \mathbf{g}_{\mathrm{P}} = J_{\mathrm{P}}^{\top} W \mathbf{e},\; \mathbf{g}_{\mathrm{G}} = J_{\mathrm{G}}^{\top} W \mathbf{e}. \tag{S45}$$

This is the direct first-order criterion used here: the interaction is described by the matrix C itself rather than by a single scalar summary. Exact first-order indistinguishability would require a stronger condition, namely nonzero $\Delta\Theta_{\mathrm{P}}$ and $\Delta\Theta_{\mathrm{G}}$ such that

$$J_{\mathrm{P}} \Delta\Theta_{\mathrm{P}} + J_{\mathrm{G}} \Delta\Theta_{\mathrm{G}} = 0. \tag{S46}$$

That condition makes a Zernike perturbation and an intercept-slope perturbation produce the same first-order image change. In the present implementation, the most dangerous exact translation modes are already weakened because tip and tilt are excluded by starting the Zernike sum at $j = 5$.

S4.4 Intercept-slope decomposition and modal profile

Since $\Theta_{\mathrm{G}} = [\mathbf{b}; \mathbf{s}]$, the geometry Jacobian and cross-information block decompose as

$$J_{\mathrm{G}} = [J_b \;\; J_s], \qquad C = [C_b \;\; C_s], \qquad C_b = J_{\mathrm{P}}^{\top} W J_b, \qquad C_s = J_{\mathrm{P}}^{\top} W J_s. \tag{S47}$$

The row-wise modal profile is defined as

$$\eta_j^{b} = \frac{\left\| C_b[k_j, :] \right\|_2}{\| C_b \|_F}, \qquad \eta_j^{s} = \frac{\left\| C_s[k_j, :] \right\|_2}{\| C_s \|_F}, \qquad k_j = j - 5. \tag{S48}$$

This profile answers which Zernike modes are most shift-like in the local image metric. The slope block is expected to be weaker than the intercept block because slope derivatives contain the approximately symmetric depth factor $\xi(z)$, which leads to cancellation across depth.

S4.5 Schur complements and leakage map

The matrix-level consequence of joint fitting can be stated through the Schur complements without introducing a worst-direction scalar. Eliminating $\Delta\Theta_{\mathrm{G}}$ gives

$$S_{\mathrm{P}} = A - C D^{-1} C^{\top}, \qquad S_{\mathrm{P}} \Delta\Theta_{\mathrm{P}} = \mathbf{g}_{\mathrm{P}} - C D^{-1} \mathbf{g}_{\mathrm{G}}. \tag{S49}$$

Similarly, eliminating $\Delta\Theta_{\mathrm{P}}$ gives

$$S_{\mathrm{G}} = D - C^{\top} A^{-1} C, \qquad S_{\mathrm{G}} \Delta\Theta_{\mathrm{G}} = \mathbf{g}_{\mathrm{G}} - C^{\top} A^{-1} \mathbf{g}_{\mathrm{P}}. \tag{S50}$$

Thus, the effective Zernike curvature in a joint fit is not simply A but the Schur complement S_{P}; the effective intercept-slope curvature is not simply D but S_{G}. One useful diagnostic follows from considering a small geometry-calibration perturbation $\Delta\Theta_{\mathrm{G}}$ while holding the image residual fixed. The first-order Zernike leakage induced by that perturbation is

$$\boxed{\Delta\Theta_{\mathrm{P,leak}} = -A^{-1} C \Delta\Theta_{\mathrm{G}}.} \tag{S51}$$

Equation (S51) is the leakage map used in the numerical experiments below. It predicts the direction and scale of Zernike coefficient changes that can compensate a small error in the lenslet intercept or slope calibration.

Interpretation of reconstruction-level observations. The leakage analysis above also clarifies the parameter-level behavior observed in the main-text Fig. 2 and in Fig. S7. In the joint-calibration setting, small term-by-term discrepancies between the recovered and ground-truth Zernike coefficients and lenslet-shift parameters should not be interpreted as a failure of optical correction. Rather, they reflect the nonzero overlap between the image-response subspaces spanned by J_P and J_G: a shift-like image response can be partially redistributed between $\Delta\Theta_\mathrm{P}$ and $\Delta\Theta_\mathrm{G}$ while leaving the aggregate sensor response, effective PSF, and reconstructed volume nearly unchanged. This behavior is consistent with nonlinear least-squares geometry, where well-constrained model predictions can coexist with poorly identifiable parameter combinations.[10] This explains why Fig. S7 shows accurate recovery of the bead volume and per-view PSF shape, but only approximate agreement of the individual Zernike, intercept, and slope parameters with the ground truth. Conversely, in the aberration-only control of Fig. S6, the competing per-lenslet shift degrees of freedom are removed; the leakage channel $C\Delta\Theta_\mathrm{G}$ in Eq. (S51) is therefore absent, and the same Zernike parameterization recovers the individual coefficients accurately. Taken together, these two regimes indicate that the residual parameter-level offsets observed during joint calibration mainly arise from Zernike–shift cross-information, rather than from an intrinsic inability of 4DGAT to estimate Zernike aberrations.

Supplementary Note S5 Numerical analysis of JPGM pupil-geometry cross-information

We numerically validate the local cross-information framework using simulation experiments. The analysis focuses on three diagnostics: modal profiles of C_b and C_s, direct shift-injection tests using the leakage map, and empirical repeatability under Gaussian image noise. No single worst-direction scalar is reported; instead, the experiments describe the distribution and practical effect of the cross-information block. Directional agreement is not inferred from norm-only tables.

S5.1 Reconstruction-level controls and qualitative validation

Before analyzing the local cross-information blocks, we first examine three reconstruction-level simulations that separate different calibration regimes. These experiments are not intended to replace the local analysis below; rather, they provide an observable-level interpretation of when individual Zernike coefficients can be recovered accurately and when only the aggregate optical response is tightly constrained.

In the *aberration-only* control, the per-lenslet intercept and slope perturbations are absent, and the simulated PSFs contain only high-order Zernike aberrations. Across multiple randomly sampled phase realizations, 4DGAT recovers both the pupil phase maps and the individual Zernike coefficients in close agreement with the ground truth, while methods assuming an ideal PSF produce smeared or streaked bead reconstructions (Fig. S6). This control confirms that the Zernike parameterization and optimizer can recover the individual coefficients when the competing intercept-slope calibration degrees of freedom are removed.

We next consider a *joint but constrained* calibration setting, where the ground-truth forward model contains both pupil aberration and per-lenslet intercept/slope shifts, but the aberration lies in a restricted seven-term Zernike subspace. Although 4DGAT is allowed to optimize the full set of Zernike coefficients together with the shift parameters, the recovered solution preserves the sparse active aberration structure, keeps the inactive higher-order coefficients close to zero, and accurately estimates the per-lenslet intercepts and slopes (Fig. S8). In this simulated seven-term regime, the recovered parameter values agree closely with the ground truth. This case demonstrates successful joint recovery under a restricted aberration model, but does not by itself establish general identifiability.

Finally, we test the *fully flexible high-aberration* setting, where the PSF contains large Zernike aberrations with all 17 active terms and random per-lenslet intercept/slope drift. In this over-parameterized joint calibration regime, 4DGAT recovers the reconstructed volume, the per-view PSF shape, and the aggregate pupil phase in close agreement with the ground truth (Fig. S7). However, the individual Zernike coefficients and per-lenslet shift parameters need not match the ground truth term by term. This behavior is expected: the Zernike basis and the intercept-slope shift model are not strictly orthogonal in the image-space measurement metric, so the same observable PSF response can be decomposed in more than one way across the two parameter groups.

S5.2 Modal cross-block profiles

The high-aberration result in Fig. S7 motivates a local analysis of which Zernike modes are most shift-like under the image-space metric.

This experiment decomposes the cross-information blocks row by row in Zernike space using Eq. (S48). The global Frobenius ratio is

$$\|C_s\|_F \,/\, \|C_b\|_F \;=\; 0.1919.$$

Thus, the slope block carries only $(19.2\%)^2 \approx 3.7\%$ of the intercept-block cross-information energy. Equivalently, the slope block is about 5.2 times weaker than the intercept block. This agrees with the cancellation argument in Section S4.4: slope derivatives are weighted by the approximately symmetric depth coordinate $\xi(z)$.

The top Zernike contributors to the intercept block are listed in Table S2. These modes have local image responses with the strongest overlap with depth-independent lenslet shifts.

Table S2. Top-5 Zernike contributors to the aberration–intercept cross-information block C_b.

Rank	Noll j	(n,m)	Family	η_j^b
1	10	$(3,3)$	Trefoil	0.3203
2	8	$(3,1)$	Coma	0.3193
3	19	$(5,-3)$	Sec. trefoil	0.3110
4	6	$(2,2)$	Astigmatism	0.2725
5	5	$(2,-2)$	Astigmatism	0.2721

The top Zernike contributors to the slope block are listed in Table S3. The slope profile is not identical to the intercept profile: coma-like modes become more prominent, while the overall slope-block energy remains weaker.

Table S3. Top-5 Zernike contributors to the aberration–slope cross-information block C_s.

Rank	Noll j	(n,m)	Family	η_j^s
1	8	$(3,1)$	Coma	0.3441
2	7	$(3,-1)$	Coma	0.3002
3	6	$(2,2)$	Astigmatism	0.2881
4	10	$(3,3)$	Trefoil	0.2675
5	16	$(5,1)$	Sec. coma	0.2644

Primary spherical aberration ($j = 11$) is weak in the intercept profile, with $\eta_{11}^b = 0.0795$, below the median value 0.2472. This supports the qualitative expectation that symmetric broadening is less shift-like than the leading asymmetric modes.

S5.3 Direct shift-injection leakage test

To test Eq. (S51), we introduce controlled perturbations into Θ_G and refit the Zernike coefficients while holding the simulated object and forward model otherwise fixed. We compare the measured Zernike change with the first-order prediction $-A^{-1}C\,\Delta\Theta_G$.

For the $\Delta b_x = 0.3$ and $\Delta s_y = 0.3$ cases, the perturbation amplitudes are expressed in pixels and pixels per unit normalized depth, respectively. The same perturbation was applied to the corresponding coordinate of all active lenslets. Because ξ is dimensionless, $\|\Delta\Theta_G\|_2$ is reported in pixels and denotes the norm of the full perturbation vector.

Table S4. Direct leakage test. The linear leakage map predicts the measured Zernike response to small intercept/slope perturbations with the correct order of magnitude and direction-dependent scale.

Case	$\|\Delta\Theta_G\|_2$ (pixels)	Measured $\|\Delta\Theta_P\|_2$ (waves)	Predicted $\|\Delta\Theta_{P,\text{leak}}\|_2$ (waves)	Ratio
$\Delta b_x = 0.3$	0.7937	4.92×10^{-2}	5.58×10^{-2}	0.882
$\Delta s_y = 0.3$	0.7937	2.81×10^{-2}	2.52×10^{-2}	1.115
Random $\Delta\Theta_G$	1.5870	2.62×10^{-1}	2.30×10^{-1}	1.141

Across the three representative cases, the measured-to-predicted ratios lie between 0.88 and 1.14. This supports the first-order local analysis: small lenslet-shift calibration errors can be partially absorbed by Zernike coefficients, and the leading leakage direction is captured by the cross-information block C.

S5.4 Empirical repeatability under image noise

We further evaluate joint refitting stability under additive Gaussian image noise. The noise level is set to 1% of the peak clean image intensity. Over 30 noisy refits, the empirical perturbation norms (mean ± s.d.) are

$$\|\Delta\Theta_P\|_2 = (5.22\pm1.19)\times10^{-3}\ \text{waves}, \qquad \|\Delta\Theta_G\|_2 = (5.08\pm0.74)\times10^{-2}\ \text{pixels}.$$

The Zernike coefficients with the largest empirical standard deviations are listed below.

Table S5. Zernike coefficients with the largest empirical variability over 30 noisy joint refits.

Rank	Noll j	(n,m)	Family	Empirical std. dev. (waves)
1	6	$(2,2)$	Astigmatism	2.30×10^{-3}
2	9	$(3,-3)$	Trefoil	1.79×10^{-3}
3	5	$(2,-2)$	Astigmatism	1.68×10^{-3}
4	7	$(3,-1)$	Coma	1.47×10^{-3}
5	10	$(3,3)$	Trefoil	1.45×10^{-3}

The empirical standard deviation of the primary spherical mode is 5.32×10^{-4} waves for $j = 11$, which is substantially lower than the largest shift-like modes. This is consistent with the modal cross-block profile: modes with weaker shift-like image responses are more stable under joint refitting.

These experiments support three practical conclusions. First, the relevant local interaction is captured by the cross-information block $C = J_P^\top W J_G$. Second, the intercept component C_b dominates the slope component C_s, consistent with cancellation over the approximately symmetric depth coordinate. Third, the leakage map $-A^{-1}C\,\Delta\Theta_G$ provides a direct diagnostic for how small intercept/slope calibration errors can be absorbed by Zernike coefficients.

Supplementary Note S6 Scaled Heavy-Ball (SHB) Acceleration of the Richardson–Lucy (RL) Algorithm for FLFM Deconvolution

S6.1 Background and motivation

The Richardson–Lucy (RL) algorithm[11,12] is the standard iterative scheme for fluorescence-microscopy deconvolution under Poisson noise. It updates the object multiplicatively, automatically preserves non-negativity, and is robust in the low-photon-count regime. Its drawback is a first-order convergence rate of only $\mathscr{O}(1/k)$, which makes it slow on volumetric problems where each iteration involves multiple FFT pairs. For Fourier light-field microscopy (FLFM), the forward operator $\mathbf{A}$ evaluates one Fourier multiplication per (lenslet, depth) pair, so the iteration count translates almost linearly into wall-clock time.

The Scaled Heavy-Ball (SHB) method proposed by Wang and Miller[13] for 3D confocal-microscopy RL deconvolution accelerates RL by injecting a Polyak heavy-ball momentum term while retaining the multiplicative, non-negative update. It achieves the optimal first-order convergence rate of $\mathscr{O}(1/k^2)$ for the Poisson cost function under a positive background, includes the Biggs–Andrews (BA) extrapolation[14] (the default in MATLAB's `deconvlucy`) as a special case, and adds essentially no per-iteration overhead. SHB has been used in confocal and widefield deconvolution for over a decade, but to the best of our knowledge the present work is the first application of SHB to FLFM Richardson–Lucy deconvolution.

S6.2 Forward and adjoint operators

We briefly summarize the generic forward and adjoint operators used in the deconvolution iteration. For a fixed per-view, per-depth PSF stack, let $\mathbf{A}$ denote the linear FLFM forward operator and $\mathbf{A}^\top$ its adjoint. Given a 3D object $\mathbf{V} = \{V_z\}_{z=1}^{N_z}$ and the per-view, per-depth PSFs $\{H_{v,z}\}$, the per-view forward map is

$$\big[\widehat{\mathbf{I}}\big]_v \;=\; \sum_{z=1}^{N_z} \mathscr{F}^{-1}\big[\,\mathscr{F}(V_z)\odot \mathrm{OTF}_{v,z}\,\big], \tag{S52}$$

stacked over views by the PATCHER operator to give the sensor prediction $\widehat{\mathbf{I}} = \mathbf{AV}$. The adjoint $\mathbf{A}^\top$ applied to a residual $\mathbf{r} = \{r_v\}$ returns

$$\big[\mathbf{A}^\top \mathbf{r}\big]_z \;=\; \sum_{v=1}^{N_{\text{view}}} \mathscr{F}^{-1}\big[\,\mathscr{F}(r_v)\odot \mathrm{OTF}^*_{v,z}\,\big], \tag{S53}$$

since the convolution adjoint is multiplication by the conjugate OTF in the Fourier domain. We adopt this conjugate-OTF form rather than convolving with the rotated PSF: the two are mathematically identical, but the conjugate-OTF form avoids storing an additional flipped-PSF buffer of size $\mathscr{O}(N_{\text{view}}\cdot N_z\cdot H\cdot W)$.

S6.3 Richardson–Lucy as a scaled gradient-projection method

In our implementation, a constant camera/background offset can be subtracted from the raw measurement before deconvolution, after which the image is clipped to be non-negative and normalized. The RL/SHB iteration itself uses the background-subtracted model with no additive background term. Under this Poisson model, the negative log-likelihood is

$$C(\mathbf{V}) \;=\; \langle \mathbf{1}, \mathbf{AV}\rangle \;-\; \sum_i I_i \log\!\big([\mathbf{AV}]_i\big), \tag{S54}$$

with gradient $\nabla C(\mathbf{V}) = \mathbf{A}^\top\mathbf{1} - \mathbf{A}^\top\big(\mathbf{I}/(\mathbf{AV})\big)$. Using the standard PSF normalization $\mathbf{A}^\top\mathbf{1} = \mathbf{1}$, the projected-gradient iterate

$$\mathbf{V}^{k+1} \;=\; \mathscr{P}_+\Big[\mathbf{V}^k - \alpha^k \odot \nabla C(\mathbf{V}^k)\Big] \tag{S55}$$

reduces to the multiplicative RL form once the diagonal step size is chosen as $\alpha^k = \mathbf{V}^k$:

$$\mathbf{V}^{k+1} \;=\; \mathbf{V}^k \odot \mathbf{A}^\top\Big(\mathbf{I} \oslash \max\!\big(\mathbf{AV}^k,\, 10^{-10}\big)\Big), \tag{S56}$$

where the small numerical floor 10^{-10} prevents division by zero (no physical background is added). The key observation of Wang and Miller is that the diagonal step $\alpha^k = \mathbf{V}^k$ is in fact a diagonal approximation of the inverse Hessian of C, so RL is implicitly a *scaled* gradient method in which the metric tensor adapts to the current estimate. This justifies replacing the loose Lipschitz constant L^{-1} with the scaling matrix $\mathbf{D} = \mathrm{diag}(\mathbf{V}^k)$ in the convergence analysis, and is what gives "scaled" heavy-ball its name.

S6.4 Heavy-ball extrapolation

SHB inserts a Polyak heavy-ball momentum step before the gradient is evaluated. With momentum coefficient $\beta^k \in [0,1)$ the iteration becomes

$$\mathbf{p}^k = \mathbf{V}^k + \beta^k\big(\mathbf{V}^k - \mathbf{V}^{k-1}\big), \tag{S57}$$

$$\mathbf{V}^{k+1} = \mathscr{P}_+\Big[\mathbf{p}^k \odot \mathbf{A}^\top\Big(\mathbf{I} \oslash \max\big(\mathbf{A}\mathbf{p}^k,\, 10^{-10}\big)\Big)\Big]. \tag{S58}$$

Setting $\beta^0 = 0$ makes the very first step coincide with standard RL, which seeds the heavy-ball trajectory from a clean RL update. Wang and Miller analyze the scaled heavy-ball scheme under a positive-background Poisson model, which ensures a Lipschitz-gradient condition and, as shown in Theorem 2 (ref. 13), yields

$$C(\mathbf{V}^k) - C(\mathbf{V}^*) \le \frac{2}{(k+1)^2}\left\|\mathbf{V}^0 - \mathbf{V}^*\right\|_{\mathbf{D}}^2, \quad \forall k \ge 1, \tag{S59}$$

where $\|\cdot\|_{\mathbf{D}}$ is the weighted norm associated with the diagonal scaling $\mathbf{D}$, a quadratic improvement over the $\mathscr{O}(1/k)$ bound that applies to plain RL. Our implementation follows the same scaled heavy-ball update structure but uses background-subtracted measurements and a small numerical floor in the denominator for stability.

S6.5 Algorithm implementation details

The complete SHB-accelerated RL iteration we use is summarized in Algorithm S1.

Algorithm S1 SHB-accelerated Richardson–Lucy for FLFM

Require: measurement $\mathbf{I}$, PSF stack $\{H_{v,z}\}$, max iterations K
Ensure: reconstruction $\mathbf{V}^K$

precompute $\mathrm{OTF}_{v,z} \leftarrow \mathrm{FFT}(H_{v,z})$ and $\mathrm{OTF}^*_{v,z}$
$\mathbf{V}^0 \leftarrow \mathbf{1}$ ▷ uniform initialization
$\mathbf{V}^{-1} \leftarrow \mathbf{V}^0$
$\varepsilon \leftarrow \max\big(\mathrm{mean}(\mathbf{I})\cdot 10^{-6},\, 10^{-8}\big)$
for $k = 0, 1, \ldots, K-1$ **do**
  $\beta \leftarrow \max\big(0,\, (k-1)/(k+2)\big)$
  $\mathbf{p} \leftarrow \max\big(\mathbf{V}^k + \beta(\mathbf{V}^k - \mathbf{V}^{k-1}),\, \varepsilon\big)$ ▷ extrapolate
  $\widehat{\mathbf{I}} \leftarrow \mathbf{A}\mathbf{p}$ ▷ forward prediction
  $\mathbf{r} \leftarrow \mathbf{I} \big/ \max\big(\widehat{\mathbf{I}},\, 10^{-10}\big)$ ▷ measurement-to-prediction ratio
  $\mathbf{V}^{k+1} \leftarrow \max\big(\mathbf{p} \odot \mathbf{A}^\top \mathbf{r},\, 0\big)$ ▷ multiplicative update
  $\mathbf{V}^{k-1} \leftarrow \mathbf{V}^k$ ▷ roll buffer in place
end for

S6.6 Practical observations on FLFM data

On our FLFM datasets the SHB iteration reaches the same value of the Poisson cost function and the same visual sharpness as standard RL with roughly 4–6× fewer iterations. This acceleration is illustrated in Fig. S11, where SHB reaches comparable PSNR, SSIM and visual reconstruction quality in substantially fewer iterations than RL for both line-sphere and bead phantoms. The observed iteration reduction is also in agreement with the speed-up factor of ≈5× reported by Wang and Miller[13] for confocal and widefield 3D data. Restored volumes are visually indistinguishable between RL and SHB-RL at convergence, confirming that, as in the original 3D-confocal study, acceleration does not come at the cost of reconstruction quality.

We emphasize that our contribution in this section is not the SHB scheme itself, but its application to accelerate Richardson–Lucy deconvolution for FLFM.

Supplementary Note S7 Computational cost of 4DGAT

S7.1 Computational cost of the static stage

The computational cost of the static stage is dominated by periodic JPGM self-estimation rather than by the ordinary Gaussian reconstruction updates, as summarized in Table S7. In a standard static training step, the current Gaussian representation is first rasterized into a dense 3D volume, which takes approximately 0.004 s. The volume is then projected to the camera plane using the cached optical propagation operator, with one complete forward projection taking approximately 0.03 s. Including image-domain loss evaluation, backpropagation and parameter updates, a regular static training step takes approximately 0.07 s. Therefore, the pure reconstruction cost of 2,000 ordinary static epochs is only on the order of 2–3 min.

By contrast, each JPGM self-estimation event is substantially more expensive. This event does not simply reuse the cached propagation operator. Instead, it regenerates the complete optical response according to the current Zernike pupil phase, lenslet-dependent lateral shifts and depth-dependent PSF-center drift, converts this response into the propagation operator used by the FLFM forward model, and backpropagates the image-domain error to the JPGM parameters. A single JPGM-update iteration typically takes approximately 0.6 s, more than an order of magnitude slower than a regular static training step. When each estimation event contains 100 iterations and is triggered 36 times throughout the static stage (every 50 epochs starting from epoch 200), the static-stage wall-clock time is therefore dominated by repeated JPGM self-estimation.

S7.2 Computational cost of the dynamic stage

Unlike the static stage, the dynamic stage keeps JPGM fixed and performs no further pupil or per-lenslet geometry updates. Its runtime is therefore mainly determined by the number of frame-wise optimization steps. In our implementation, one dynamic epoch iterates over all frames; thus, the total number of dynamic iterations is

$$N_{\mathrm{dyn}} = N_{\mathrm{epoch}}^{\mathrm{dyn}} \times N_{\mathrm{frame}}. \tag{S60}$$

At each frame-iteration, the canonical Gaussians are deformed by the HexPlane–MLP field, rasterized into a 3D volume, projected to the FLFM sensor, and optimized using image-domain, Fourier-domain and temporal regularization losses.

The average dynamic frame-iteration time was approximately 0.1 s, including training-loop overhead, loss evaluation, backpropagation and logging. For example, the dynamic stage of Fig. 3a,e,g contains 130 frames and took approximately 1.5 h (Table S6), corresponding to

$$N_{\mathrm{dyn}} \approx \frac{1.5 \times 3600}{0.1} \approx 5.4 \times 10^4 \tag{S61}$$

frame-iterations, or approximately 4.2×10^2 full passes over the 130-frame sequence.

Overall, the static stage is dominated by intermittent JPGM self-estimation, whereas the dynamic stage is dominated by the frame-wise repetition of HexPlane deformation, volume rendering and FLFM projection. With all other parameters held fixed (volume size, HexPlane resolution and $N_{\mathrm{epoch}}^{\mathrm{dyn}}$), the dominant dynamic-stage runtime is approximately linear in the number of frames N_{frame}.

Table S6. Dataset-specific imaging, reconstruction, and training parameters.

Dataset	Acquisition and reconstruction settings					Model resolution and training time			
Figure	**Frames used**	**Data size**	**Objective**	**Medium**	**Voxel size (μm)**	**HexPlane resolution**	**Static time**	**Dynamic time**	**Total time**
Fig. 1c	1	384 × 384 × 87	60×/1.25	Oil	0.325 × 0.325 × 0.325	–	∼30 min	–	∼30 min
Fig. 2a–b	40	384 × 384 × 134	20×/0.45	Air	0.975 × 0.975 × 1.5	64 × 64 × 64 × 20	∼30 min	∼1 h	∼1.5 h
Fig. 2h,k,l	40	384 × 384 × 134	20×/0.45	Air	0.975 × 0.975 × 1.5	64 × 64 × 64 × 20	∼30 min	∼1 h	∼1.5 h
Fig. 3a,e,g	130	384 × 384 × 62	60×/1.25	Oil	0.325 × 0.325 × 0.325	64 × 64 × 64 × 65	∼20 min	∼1.5 h	∼2 h
Fig. 3b,f,h	100	384 × 384 × 62	60×/1.25	Oil	0.325 × 0.325 × 0.325	64 × 64 × 64 × 50	∼20 min	∼1 h	∼1.5 h
Fig. 4	100	384 × 384 × 101	20×/0.45	Air	0.975 × 0.975 × 1.5	64 × 64 × 64 × 50	∼30 min	∼1 h	∼1.5 h
Fig. 5	166	384 × 384 × 151	20×/0.45	Air	0.975 × 0.975 × 2	64 × 64 × 64 × 83	∼40 min × 3 ch	∼2 h × 3 ch	∼2.5 h × 3 ch
Fig. S6 trials 1–4	1	300 × 300 × 101	20×/0.45	Air	0.975 × 0.975 × 1.5	–	∼30 min per trial	–	∼30 min per trial
Fig. S7	1	300 × 300 × 101	20×/0.45	Air	0.975 × 0.975 × 1.5	–	∼30 min	–	∼30 min
Fig. S8	20	300 × 300 × 101	20×/0.45	Air	0.975 × 0.975 × 1.5	64 × 64 × 64 × 10	∼30 min	∼30 min	∼1 h
Fig. S9	1	300 × 300 × 101	20×/0.45	Air	0.975 × 0.975 × 1.5	–	∼30 min per trial	–	∼30 min per trial

Table S7. Runtime breakdown of the static-stage optimization.

Operation	**Main computation**	**Time per operation**	**Typical count**	**Total time**
Gaussian voxelization	Rasterize current 3D Gaussians into a dense volume	∼0.004 s	2,000	∼8 s
Cached FLFM forward projection	Project the voxelized volume using cached optical propagation	∼0.03 s	2,000	∼1 min
Regular static training step	Voxelization, cached projection, loss evaluation, backpropagation and Gaussian update	∼0.07 s	2,000	∼2–3 min
JPGM self-estimation	Regenerate the PSF stack from the Zernike phase and per-lenslet geometry parameters, update the propagation operator and backpropagate to JPGM parameters	∼0.6 s	100 × 36	∼36 min
Full static stage	Regular Gaussian reconstruction plus periodic JPGM self-estimation	–	–	∼40 min

Note. The static stage runs 2,000 epochs in total. JPGM updates start after epoch 200 and are triggered every 50 epochs; they are therefore performed at epochs $250, 300, \ldots, 2000$ (36 updates in total). Each JPGM update internally runs 100 optimization steps, giving 36×100 JPGM iterations.

Supplementary Note S8 Sample preparation and imaging

S8.1 H9c2 cell culture, mitochondrial labeling, and live-cell imaging

H9c2 rat cardiomyocyte-derived cells were obtained from laboratory stock and maintained in basal growth medium consisting of Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Cells were cultured at 37 °C in a humidified incubator with 5% CO_2, and the medium was replaced every two days. To label mitochondria, H9c2 cells were transduced with a Mito-EGFP lentiviral vector (GeneChem, Shanghai, China) at a multiplicity of infection of 10. After 24 h, the medium was replaced with fresh basal growth medium. Cells were cultured for an additional 48 h and then selected with 1 µg/mL puromycin for 2 days to establish stable Mito-EGFP-expressing cell lines. For imaging, stable H9c2 cells were seeded in glass-bottom culture dishes (Nest, Cat. No. 801002) at a density of 5,000 cells per dish. After 24 h, cells were gently washed and maintained in basal growth medium for subsequent live-cell imaging.

S8.2 Mitochondrial tubule-width measurement

For the quantification in Fig. 3j, apparent mitochondrial tubule width was measured from intensity profiles across locally isolated tubule segments. The same tubule locations were used for the raw LF image, SHB, 3DGAT, SHB using the 4DGAT-estimated PSF, and 4DGAT whenever possible. For each selected segment, a line profile was placed approximately perpendicular to the tubule axis. The local background was estimated from the profile ends and subtracted, and the profile was normalized to its peak intensity. The apparent width was defined as the FWHM, calculated by linear interpolation between the two half-maximum crossings. Measurements from all selected tubule segments were pooled for the boxplot. Because this quantity is measured from image profiles after projection or reconstruction, it is reported as apparent tubule width rather than physical mitochondrial diameter.

S8.3 Zebrafish preparation, mechanical injury, and imaging

Zebrafish larvae expressing myl7:H2A-mCherry were imaged at 48 hours post-fertilization (hpf). To inhibit pigmentation, embryos were treated with 0.003% 1-phenyl-2-thiourea (PTU) from 24 hpf. Larvae were anesthetized in E3 medium containing 0.02% tricaine methanesulfonate (MS-222). For mechanical injury, anesthetized larvae were positioned laterally under a stereomicroscope, and localized compression was applied once to the anterior trunk/pericardial region using Dumont #5 forceps for approximately 1 s. After injury, larvae were allowed to stabilize for 5 min in E3 medium containing tricaine. For FLFM imaging, larvae were embedded laterally in 0.7% low-melting-point agarose prepared in E3 medium containing tricaine and PTU, with the heart facing the objective. The chamber-resolved mesh propagation and functional metrics are described in Section 4.6 of the main text.

S8.4 RAW264.7 and BMDM culture, induction of cellular magnetism, and immunofluorescence staining

Laboratory-stock RAW264.7 murine macrophages were maintained in DMEM containing 10% fetal bovine serum and 1% penicillin-streptomycin, with medium replacement every 2 days. BMDMs were differentiated from femur- and tibia-derived mouse bone marrow cells in the same medium containing recombinant murine M-CSF (40 ng/mL; PrimeGene). All cultures were maintained at 37 °C and 5% CO_2.

For BMDM magnetization, fresh mouse red blood cells (RBCs) collected in EDTA tubes were washed twice with Alsever's solution, opsonized with rabbit anti-mouse RBC antibody (0.25 mg/mL; Rockland Immunochemicals) for 30 min at 4 °C, and added to bone marrow cultures at a 10:1 RBC-to-bone marrow cell ratio. After 5 days of co-culture, ferric ammonium citrate (200 µM; Sigma-Aldrich) was added, and cells were cultured for a further 4 days with M-CSF. For RAW264.7 magnetization, cells were incubated with ferumoxytol (200 µg·Fe/mL) for 24 h, washed twice with PBS, fixed with 4% paraformaldehyde for 10 min at room temperature, permeabilized and washed with BD Perm/Wash Buffer (BD Biosciences), and blocked with 2% bovine serum albumin.

RAW264.7 cells were incubated overnight at 4 °C with rat anti-mouse CD11b (M1/70, eBioscience) and rabbit anti-ferritin (SC0620), whereas BMDMs were incubated with rat anti-mouse F4/80 (BM8, eBioscience) and rabbit anti-ferritin (SC0620); all primary antibodies were diluted 1:100. After washing with BD Perm/Wash Buffer, cells were incubated for 1 h at room temperature in the dark with Alexa Fluor 488-conjugated donkey anti-rat IgG and Alexa Fluor 555-conjugated donkey anti-rabbit IgG (H+L) secondary antibodies (Thermo Fisher Scientific), each diluted 1:500. Nuclei were counterstained with Hoechst 33342 (5 µg/mL, 10 min). Cells were washed, resuspended in PBS, and mixed with isotonic Percoll (Cytiva) to a final concentration of 25% to reduce sedimentation during imaging and magnetic measurements.

S8.5 Cell-suspension and mixture preparation

For CD11b/Ferritin volumetric classification, four RAW264.7 populations representing the CD11b$^-$Ferritin$^-$, CD11b$^-$Ferritin$^+$, CD11b$^+$Ferritin$^-$ and CD11b$^+$Ferritin$^+$ classes were combined at an equal 1:1:1:1 ratio. A matched aliquot of the same mixture was reserved for flow-cytometry validation. For magnetic-mobility analysis, magnetic and non-magnetic macrophage populations were combined at a predefined 1:5 ratio before loading into the quartz capillary.

S8.6 Magnetic-field actuation module and macrophage imaging

A magnetic-field actuation module adapted from the original Optical Tracking-based Magnetic Sensor (OTMS) design[15] was integrated with the inverted IX85 FLFM platform. The module consisted of a laminated silicon-steel core magnetized by a 1000-turn copper-wire coil. The shaped pole pieces formed a tapered air gap that concentrated the magnetic field and generated a lateral magnetic-field gradient. Cells were loaded into a rectangular quartz capillary tube with an inner cross-section of 0.6 mm $\times$ 0.3 mm, which was inserted through the air gap during measurement. In this configuration, the dominant magnetic-field gradient aligned approximately with the lateral imaging direction and remained perpendicular to the optical axis. The observation window coincided with the FLFM field of view (FOV). In routine measurements, the left edge of the FOV was placed adjacent to the pole-gap opening to maximize the magnetic-field gradient across the imaging region.

S8.7 Single-cell segmentation and marker classification of macrophages

For the volumetric cytometry analysis in Fig. 5e-g and Supplementary Fig. S14, 4DGAT-reconstructed multichannel volumes were used for 3D cell segmentation and marker assignment. The nuclear channel was used as the pan-cell reference. Nuclear objects were detected using a classical 3D blob-detection workflow. CD11b and ferritin marker objects were detected from the 488-nm and 555-nm channels using the same workflow with channel-specific thresholds. For each nuclear object, local CD11b and ferritin signals were assigned from spatially matched marker objects after local background correction.

S8.8 Flow cytometry validation of marker-defined population fractions

The matched aliquot of the four-population RAW264.7 mixture was analyzed using an Attune NxT Flow Cytometer (Thermo Fisher Scientific, USA). Cells were harvested, washed with PBS, and resuspended in flow cytometry staining buffer. Single-stained controls were used for fluorescence compensation, and unstained controls were included to define background signals. Data were acquired on the Attune NxT system under identical acquisition settings for all samples within each experiment. Flow cytometry data were analyzed using FlowJo software. Agreement between the 4DGAT-derived and flow-cytometry class distributions was assessed using the overall chi-squared test and the two-sided per-class Fisher's exact tests reported in Fig. 5g.

S8.9 Magnetic-mobility tracking and anomalous-diffusion analysis

For the magnetic-mobility analysis in Fig. 5h-j, cells in the predefined 1:5 magnetic/non-magnetic mixture were detected and tracked from the 555-nm ferritin channel. Representative speed traces were generated using the same sliding-window displacement calculation. Slow and fast populations were separated at the local density minimum between the two modes of the log-transformed mean-speed distribution, yielding a threshold of 8.27 μm/s. For each trajectory, the time-averaged mean squared displacement was computed as

$$\mathrm{MSD}(\tau) = \left\langle |\boldsymbol{r}(t+\tau) - \boldsymbol{r}(t)|^2 \right\rangle_t \tag{S62}$$

where $\boldsymbol{r}(t)$ denotes the centroid position of a tracked cell at time t and $\langle \cdot \rangle_t$ denotes averaging over all valid time windows within each trajectory. The apparent anomalous-diffusion exponent α was used to summarize the motion mode of each trajectory, following standard single-particle tracking analysis.[16] Values of α near 1 indicate Brownian-like diffusion, whereas larger values indicate more persistent or directed motion.

References

1. Liu, W., Kim, G.-A. R., Takayama, S. & Jia, S. Fourier light-field imaging of human organoids with a hybrid point-spread function. *Biosens. Bioelectron.* **208**, 114201 (2022).

2. Guo, C., Liu, W., Hua, X., Li, H. & Jia, S. Fourier light-field microscopy. *Opt. Express* **27**, 25573–25594 (2019).

3. Incardona, N., Tolosa, A., Saavedra, G., Martinez-Corral, M. & Sanchez-Ortiga, E. Fast and robust wave optics-based reconstruction protocol for Fourier lightfield microscopy. *Opt. Lasers Eng.* **161**, 107336 (2023).

4. Noll, R. J. Zernike polynomials and atmospheric turbulence. *J. Opt. Soc. Am.* **66**, 207–211 (1976).

5. Matsushima, K. & Shimobaba, T. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Opt. Express* **17**, 19662–19673 (2009).

6. Kerbl, B., Kopanas, G., Leimkühler, T. & Drettakis, G. 3D Gaussian Splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**, 139:1–139:14 (2023).

7. Zha, R. *et al.* R^2-Gaussian: rectifying radiative Gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems*, vol. 37, 44907–44934 (Curran Associates, Inc., 2024).

8. Cao, A. & Johnson, J. HexPlane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 130–141 (2023).

9. Nocedal, J. & Wright, S. J. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering (Springer, 2006), 2 edn.

10. Transtrum, M. K., Machta, B. B. & Sethna, J. P. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E* **83**, 036701 (2011).

11. Richardson, W. H. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* **62**, 55–59 (1972).

12. Lucy, L. B. An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745–754 (1974).

13. Wang, H. & Miller, P. C. Scaled Heavy-Ball acceleration of the Richardson–Lucy algorithm for 3D microscopy image restoration. *IEEE Trans. Image Process.* **23**, 848–854 (2014).

14. Biggs, D. S. C. & Andrews, M. Acceleration of iterative image restoration algorithms. *Appl. Opt.* **36**, 1766–1775 (1997).

15. Wu, L., Zhang, Y., Zhao, F., Sheng, J. & Gu, N. Magnetic characterization of human intrinsically magnetic monocytes through a novel optical tracking-based magnetic sensor. *Small* **20**, 2307306 (2024).

16. Saxton, M. J. & Jacobson, K. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 373–399 (1997).