THE SHARP PLANE IS TILTED: ANISOTROPY PRIORS MUST FOLLOW THE STORED AFFINE, NOT THE GRID

Anonymous authors
Paper under double-blind review

ABSTRACT

Restoration models for anisotropic 3-D data need one piece of physics: the direction of worst resolution. Recent 3-D microscopy restoration methods hard-code it as the grid axis z. We show that this is a statement about the resampling that produced the stored volume: under a linear resample A of the acquisition, the plane of best resolution $\mathcal{S}$ maps to $A\mathcal{S}$ and its normal $\mathbf{a}_0$ to $A^{-\top}\mathbf{a}_0$, so a grid-axis prior is exact only when the transported axis still lands on a grid axis. For stage-scanned light-sheet data A is the deskew affine the file records; when it leaves the camera plane tilted, the worst axis is $\mathbf{n} = A^{-\top}e_{\mathrm{scan}}$ or $\mathbf{s} = Ae_{\mathrm{scan}}$, both read from the file and ordered by the ratio of A's column norms. On 107 blocks from 18 open-top light-sheet biopsies (7 patients) the spectrum's sharp lobe lies along the camera axis A predicts in 103/107 blocks, and a sweep that uses no network finds $\mathbf{n}$ the cheapest direction to under-sample on the median over blocks. On held-out blocks decimated along $\mathbf{n}$, the restorer that assumes z leaves them worse than unrestored in explained variance (EV) at decimation factors $k = 4$ and 8 (0/4 patient-disjoint folds improve them). The same network pointed at $\mathbf{n}$ improves them in 4/4 folds and beats the convention by $+.025$, $+.060$ and $+.040$ in EV at $k = 4$, 8 and 16, and by .047 F1 on a nucleus-detection endpoint that does not use EV ($k = 8$). The gap survives the published CARE recipe, and eight restorers trained at fixed axes from $-45°$ to $90°$ trace a curve that peaks at the file's angle ($k \leq 8$). On a second instrument (oblique-plane microscopy) a degradation whose stored-frame direction the file sets (every second camera frame dropped before the deskew) is improved only by the restorer pointed at the file's scan direction once its kept-slice mask input is removed; on brain cone-beam CT with an injected oblique axis and a true reference the rotated prior beats the z prior in 9/9 fold$\times$seed pairs. The fix is one signed angle per stored volume, read from the file.

1 INTRODUCTION

A restoration or super-resolution model for anisotropic 3-D data needs one piece of physics: the direction along which resolution is worst. The forward model requires that direction, not that it be a grid axis; aligning it with z is how the field implements it. Every one of the eight recent 3-D microscopy restoration methods we read in full (§2; we claim nothing beyond these eight) hard-codes the low-resolution axis as z: EMGauss (He et al., 2025, Eq. 6), Volume Tells (Li et al., 2025, §3.1), SkelEM (Chen et al., 2026), and the official PyTorch implementations of IsoVEM (He et al., 2023) and niiv (Troidl et al., 2025), whose data loaders average-pool along one fixed array axis with the anisotropy factor in that slot.

For a large class of data, z is not the worst direction, and the reason is written in the file. Stage-scanned light-sheet volumes (open-top, lattice, oblique-plane; Glaser et al., 2019; Chen et al., 2014; Dunsby, 2008) are acquired on a sheared lattice: the camera images a plane tilted by the sheet angle, and the stage translates the sample between frames. The stored volume is produced by a deskew affine A, recorded in the metadata, that maps the sheared lattice to a Cartesian grid (Pietzsch et al., 2015; Hörl et al., 2019). The deskew stored in the open-top light-sheet (OTLS) and oblique-plane microscopy (OPM) files we use does not bring the camera plane onto the grid (the OTLS affine shears along the scan axis, the OPM affine rotates by the sheet angle), so the plane of best resolution

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

stays tilted; a deskew that shears along the camera axis, the lattice light-sheet convention, would keep it on the grid, and only the file says which was applied. The normal of that plane, $\mathbf{n} = A^{-\top} e_{\mathrm{scan}}$, and the physical scan direction Ae_{scan} are the two candidate axes for worst resolution, separated by an angle the affine fixes ($45°$ on the OTLS, $17.5°$ on the OPM), and a ratio of A's own column norms says which of the two dominates—all of it read from one matrix in the file (§3). A method that assumes z is the bad axis places its high-resolution planes at x–y, where only one of the two sharp directions lies. The statement is not about microscopes (§3); deskewed light-sheet data are the case where A is in the file and contains a shear (on the OPM also a rotation).

We then ask four questions of real data, each against a null model or a control arm; the first two against criteria fixed in the script headers before any real run.

Is the geometry there? On the primary instrument, 107 blocks from 18 OTLS prostate biopsies of 7 patients (Xie et al., 2022), the sharp lobe of the spectrum lies within $10°$ of the predicted camera axis in 103/107 blocks, and, measured without any network, the cheapest direction to under-sample lies at the predicted normal $\mathbf{n}$ on the median curve, and z costs more than $\mathbf{n}$ in 18/18 biopsies ($1.6\times$ at the median block; §5.1). A second instrument (dual-view OPM) reproduces the geometry from its own affine and point-spread functions (PSFs), and when every second camera frame of its raw stack is discarded before the deskew, only the arm pointed at the file's scan direction restores the block once the mask channel (the kept-slice indicator the network receives as a second input) is removed (§5.5).

Does the convention fail, or merely underperform? With one network and one pipeline we train restorers under seven assumed axes (Table 1) and apply every one of them to the *same* block degraded along the file's axis. The z restorer returns the block worse than it received it at $k = 4$ and 8 (0/4 patient-disjoint folds improve it); the file-angle restorer improves it in 4/4 folds and beats the convention by $+.025$, $+.060$ and $+.040$ in explained variance (EV), 12/12 fold×seed pairs at every k; no capacity or recipe change closes the gap (Table 2); the axis-conditioned network of §5.6 does so only at $k = 16$. Eight restorers trained at fixed axes from $-45°$ to $90°$ trace a curve on that block that peaks at the file angle at $k \leq 8$ (Fig. 2a), and when the block is decimated along z instead, the convention and the file-angle arm swap places (App. J). That curve measures the cost of misalignment; that the cost is lowest at the file angle is confirmed independently, on the median over blocks, by the network-free sweep of §5.1.

Is it the axis that helps, or the rotation that hands it to the network? On the primary instrument at $k \leq 8$, the axis: an axis-conditioned network that never rotates its input, trained across random axes and told the file angle, matches the rotate-input restorer within .005, and told z it loses $+.018/+.027$ ($k = 4/8$; $+.003$ at $k = 16$); on cone-beam CT it uses the axis only weakly (§5.6).

What is the fix? One signed angle per stored volume, which matches per-block measurement; on brain cone-beam CT with a true reference and an injected oblique axis the rotated prior beats the one assuming z in 9/9 pairs at $k = 8$ and 16 (§5.6).

Contributions. (i) A transport statement for anisotropy priors: under a recorded linear resample A the sharp plane maps to $A\mathcal{S}$, its normal $\mathbf{a}_0$ to $A^{-\top}\mathbf{a}_0$ and the scan smear to Ae_{scan}, so a grid-axis prior is exact only when the transported axis is still a grid axis; for deskewed volumes the two candidates $A^{-\top}e_{\mathrm{scan}}$ and Ae_{scan} are read from the file with no fitted parameter and ordered by the column-norm ratio ρ (§3). No learned restoration method we could find asks where the resolution axis lands in the stored volume; multi-view deconvolution does transport the PSF (§2, App. D). (ii) A measurement protocol—signed angular spectra, a spectral tensor, and a network-free direction sweep (DSDR) with a rotate-only control and an isotropized null, validated by known-answer guards (§4, App. A)—which places the axes where the file predicts (103/107 blocks). (iii) The cost of the misaligned prior on identical inputs: on blocks degraded along $\mathbf{n}$ the grid-axis restorer does worse than no restoration at $k \leq 8$ while the file-angle restorer improves them; its lead survives capacity and recipe changes, removal of the mask channel and a test block decimated along each block's own axis, holds on a nucleus-detection endpoint ($k = 8$), and eight fixed-axis restorers trace a curve that peaks at the file angle (§5.2, §5.4). (iv) Replication and a fix: a second instrument whose frame-drop degradation lands along a stored-frame direction the file sets, a true-reference test on brain cone-beam CT, and an axis-conditioned network that separates the information from the rotation (§5.5, §5.6); official IsoVEM and niiv code reproduce where the resolution is lost but, because they

re-pool their input, not the cost of the prior (App. I). The fix is one signed angle per stored volume, read from the file.

2 Related work

Anisotropic restoration in 3-D microscopy. Self-supervised isotropic reconstruction trains on the high-resolution lateral plane and applies the model axially (CARE-isotropic, Weigert et al., 2017; 2018); the same lateral-to-axial protocol underlies reference-free and cross-plane variants (Park et al., 2022; Ahituv et al., 2026) and, recently, EMGauss, Volume Tells, SkelEM, IsoVEM and niiv (He et al., 2025; Li et al., 2025; Chen et al., 2026; He et al., 2023; Troidl et al., 2025). Six of the eight date from 2023 or later, and all fix the axial direction to a grid axis by construction—as a data-generation choice (pool along one array axis), a coordinate declaration ("H and W are high resolution"), or frozen parameters ($z_i = z_0$, $\Delta R_i = 0$). Several of them target volume microscopy acquired by physical sectioning, where the worst axis really is z; the convention is right there, and travels unexamined into optically sheared data, where the deskew convention the file records decides whether it is.

What optics already knows. That the PSF of an oblique-plane system is anisotropic with respect to its tilted image plane is established (Dunsby, 2008), and classical light-sheet deconvolution avoids the question by ordering: cudaDecon and LLSpy deskew by shearing along the camera axis, which keeps the camera plane on the grid where a z-aligned optical transfer function (OTF) has the right axis, and rotate only afterwards (Lambert, 2023). Searches of PubMed and of recent ML and vision proceedings returned no work on where the resolution axis lands in the stored volume (App. A); multi-view deconvolution propagates the PSF through the registration affine (Preibisch et al., 2014) without asking which grid axis a learned prior should take (App. D).

Positioning. Recent work argues that a common simulation assumption is wrong—naive downsampling for anisotropy (Gao & Gao, 2026), or morphological versus optical anisotropy (He et al., 2025). We add the assumption none of them questions: that the degraded direction is a coordinate axis. Soft-equivariance work asks what a model loses when the assumed symmetry is only approximately right (Finzi et al., 2021); ours concerns an axis rather than a symmetry, and it is not small: the assumed and actual axes differ by the sheet angle on the primary instrument. We follow recent work that separates what a formulation requires from what convention imposed (Li & He, 2025; Geng et al., 2025; Lu et al., 2025): our convention is "the anisotropy axis is z", and on blocks degraded along the axis §5.1 finds worst, its restorer is worse than no restoration in EV at $k \leq 8$; by the same protocol the file-angle restorer is worse than none on z-degraded blocks at $k \leq 8$ (App. J), so the data decide which degradation is real (§5.1, without a network).

3 Setup: the sharp plane the grid does not know about

What the forward model requires vs. what convention imposes. Any anisotropic-resolution model needs the worst-resolution direction $\mathbf{a}$; nothing in the forward model requires it to be a grid axis.

> **Proposition (transport of an anisotropy prior).** Let the prior be given in the acquisition frame by a plane $\mathcal{S}$ of best resolution with normal $\mathbf{a}_0$ and, for stage scanning, a smear along the index axis e_{scan}. If the stored volume is a linear resample A of the acquisition, the prior must be transported: $\mathcal{S} \mapsto A\mathcal{S}$, $\mathbf{a}_0 \mapsto A^{-\top}\mathbf{a}_0$ (normals are covectors), $e_{\mathrm{scan}} \mapsto Ae_{\mathrm{scan}}$. A grid-axis prior is exact iff the transported axis is a grid axis; an axis-aligned rescaling preserves this, a shear or rotation generically breaks it.

The statement is not specific to microscopes, and we test it only there: an oblique MRI acquisition resampled to the axial grid or a tilted-gantry CT reformatted to an orthogonal grid stores such an A, and any volumetric model whose anisotropic pooling axis is fixed to the array inherits the question after any recorded linear resample.

The deskew in these files leaves the camera plane tilted. Let the raw stack have index axes $(e_x, e_{\mathrm{row}}, e_{\mathrm{scan}})$: two camera axes and the stage-scan axis. The PSF is axis-aligned in *physical camera coordinates*—sharp within the camera plane, worst along its normal—but the raw lattice is sheared relative to physical space because the stage moves obliquely to the camera plane. The deskew affine A maps lattice indices to physical Cartesian coordinates; throughout, A is its 3×3 linear part on index vectors ordered (x, y, z); translations do not affect directions. Because A is exactly the map that makes the lattice physical, the camera plane in the stored volume is spanned by the transported camera axes Ae_x and Ae_{row}, and two axes fixed by the affine compete for worst resolution:

$$\mathbf{a} = \mathbf{n} \;\propto\; A^{-\top} e_{\mathrm{scan}} \qquad \text{or} \qquad \mathbf{a} = \mathbf{s} \;\propto\; A e_{\mathrm{scan}}. \tag{1}$$

The normal $\mathbf{n}$ carries the axial PSF, since $A^{-\top} e_{\mathrm{scan}}$ is orthogonal to Ae_x and Ae_{row}; the scan direction $\mathbf{s}$ carries the smear of the stage step. For the OTLS data ($x' = x,\ y' = y/\sqrt{2},\ z' = z - y/\sqrt{2}$, identical in all 18 specimens) the sharp axes are x and $Ae_{\mathrm{row}} \propto (-1, 1)$ in (z, y), i.e. 135° from z toward y; the normal $\mathbf{n} \propto (1, 1)$, i.e. 45°: the sheet angle. For the dual-view OPM data ($A = R_x(35^\circ)\,\mathrm{Shear}(y{+}0.315z)\,\mathrm{Scale}_z(11.4)$ from its BigStitcher XML; R_x rotates about x) the camera row lands at 55° and the scan axis at 162.5° (both marked in Fig. 3b), and $\mathbf{n}$ at 145°: again the sheet angle from the z axis.

Which of the two, from the same matrix. The two candidates are separated by an angle the affine fixes, and the choice between them is also read off A: the column norms of the affine give

$$\rho \;=\; \frac{\|Ae_{\mathrm{scan}}\|}{\|Ae_{\mathrm{row}}\|}, \tag{2}$$

how much coarser one stage step is than one camera pixel. When $\rho \approx 1$ the scan samples as finely as the camera does, the optical axial PSF is the widest thing in the volume, and the worst axis is the normal $\mathbf{n}$; when $\rho \gg 1$ the stage step smears more than the PSF does and the worst axis is $\mathbf{s}$. The two instruments sit at opposite ends: $\rho = 1.00$ for the OTLS affine and $\rho = 11.97$ for the OPM affine (from the XML's unrounded entries), and their measured worst axes are 41° (spectral tensor) against the predicted $\mathbf{n}$ at 45°, and 165° (direction sweep) against the predicted $\mathbf{s}$ at 162.5°, respectively (§5); on the OPM the measured bead PSFs, propagated through the affine as $\Sigma \mapsto A\Sigma A^\top$, land their long axis at 162.5° (App. D), a prediction that involves no network and no self-reference. We simulate the crossover: an isotropic field blurred by a camera-plane PSF, sampled on a 45° lattice with scan step ρ times the pixel and deskewed, makes $\mathbf{s}$ cheaper than $\mathbf{n}$ from $\rho \approx 3.9$ on, the cheapest direction moving from $\mathbf{n}$ through 30° toward $\mathbf{s}$ (Fig. 3c, App. F). On both instruments the answer is not z: on the first OPM dataset z is neither candidate; on OTLS z *is* $\mathbf{s}$ and the data reject it for $\mathbf{n}$ (§5). *Signs matter*: $+45^\circ$ and -45° are mirror images that no sheet angle distinguishes; only the affine does.

Boundary. On the primary instrument we have no raw frames, so the sharp lobe's origin (optical resolution vs. interpolation imprint of the lattice) cannot be decomposed: above spatial frequency $|\nu| > 0.3$ cycles/voxel the normal direction carries a lattice imprint visible as diagonal hatching, and all spectral readings below use the mid band 0.08–0.27. All operational claims concern the stored volume, where downstream methods operate.

4 METHOD: MEASURE THE AXIS, THEN POINT THE PRIOR AT IT

Terms. *OTLS*: open-top light-sheet. *OPM*: oblique-plane microscopy. *DSDR*: direction-swept decimation–reconstruction. *EV*: explained variance $1 - \mathrm{MSE}/\mathrm{Var}$, computed on the central cube of a block. k: decimation factor. ν: spatial frequency in cycles per voxel ($|\nu| \le 0.5$), with x component ν_x. ΔEV: *decimation cost*, $\Delta\mathrm{EV} = \mathrm{EV}(\text{rotate-only control}) - \mathrm{EV}(\text{rotate, decimate, restore})$. Angles are signed, from $+z$ toward $+y$; OPM angles are quoted in $[0^\circ, 180^\circ)$ ($162.5^\circ \equiv -17.5^\circ$). $\mathbf{n}$ (or $\mathbf{s}$) is the worst axis *predicted* from the file, θ^* the one *measured* per block by the spectral tensor; "the measured axis" means θ^*. *Mask channel*: the kept-slice indicator the restorer receives as a second input. *Cell*: one block under one seed (107 blocks $\times$ 3 seeds $= 321$ cells); *pairs*: fold$\times$seed pairs ($4 \times 3 = 12$ on the OTLS, $3 \times 3 = 9$ on the cone-beam CT test); "the shared block" is each held-out block under the shared degradation. *Wrong sign*: the tilt of the same size on the other side of z, realised as $-\mathbf{n}$ in the DSDR sweep (Fig. 1e), as $\theta^* - 90^\circ$ for the learned arm F (orthogonal

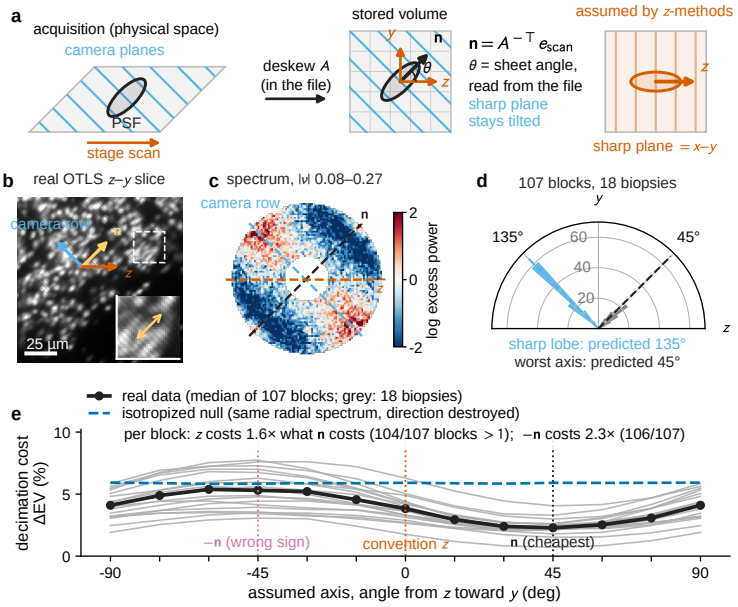


Figure 1: **The sharp plane is tilted.** (a) The stage scan acquires a sheared lattice of tilted camera planes; the deskew affine A (stored with the data) makes the grid Cartesian but leaves the camera plane tilted, so the worst axis is $\mathbf{n} = A^{-\top} e_{\mathrm{scan}}$, at the sheet angle from z (OTLS: $\rho = 1.00$, so the axial PSF dominates; §3). (b) A real OTLS z–y slice: nuclei are smeared along $\mathbf{n}$ and sharp along the deskewed camera row (135°). (c) Its mid-band spectrum (log power minus radial mean, $|\nu_x| < 0.03$): excess along the camera row, deficit in the sector between z and $\mathbf{n}$ (trough median 25° over the 107 blocks). (d) Signed angles over 107 blocks from 18 biopsies: the sharp lobe peaks at 135° (103/107 within 10° of the prediction), the spectral-tensor worst axis at 41°. (e) Direction-swept decimation cost (k=4, rotate-only control subtracted) versus the isotropized null: the cheapest direction is $\mathbf{n}$ (107-block median curve; grey: 18 biopsy means); per block, z costs 1.6× what $\mathbf{n}$ costs (IQR 1.4–1.9; 104/107 blocks > 1) and $-\mathbf{n}$ (wrong sign) 2.3× (IQR 1.8–2.8; 106/107). The null is flat to 0.10%; it sits above the data because isotropization spreads high-frequency power into every direction.

to θ^*, which is the mirror $-\theta^*$ only when $\theta^* = 45°$; App. H), as $-45°$ in the fixed-angle sweep (Fig. 2a), and as 17.5°, the mirror of $\mathbf{s}$, on the OPM.

(1) Signed angular spectrum in the z–y plane (Hann window, mid band): the peak locates the sharp lobe Ae_{row}; a second-order tensor cannot represent a lobe and a deficit that are not orthogonal, so we report the peak directly. **(2) Spectral inertia tensor** on the same band: the smallest-eigenvalue direction, orthogonal to the sharp lobe, estimates $\mathbf{n}$; robust in angle (0.52° drift under 2× noise), not in ratio. **(3) DSDR**: rotate so a test direction aligns with z, mean-pool k slices along z and interpolate back cubically (no network; learned methods use their own degradation model, App. H), rotate back and score EV on the central cube. Content already lost along the worst-resolution axis cannot be lost again, so the direction that is cheapest to under-sample marks that axis; a rotate-only control arm and an isotropized null (radial spectrum preserved, phases randomized with conjugate symmetry) are subtracted per block. **(4) Fix**: rotate the anisotropy prior of any z-assuming method to the predicted axis; a single signed angle per stored volume suffices (§5.4). Recipe: read A from the dataset's BigDataViewer/BigStitcher XML; form $\mathbf{n} \propto A^{-\top} e_{\mathrm{scan}}$ and $\mathbf{s} \propto Ae_{\mathrm{scan}}$; compute $\rho = \|Ae_{\mathrm{scan}}\|/\|Ae_{\mathrm{row}}\|$ and take $\mathbf{n}$ if $\rho \leq 2$, $\mathbf{s}$ if $\rho \geq 5$, and sweep a few blocks in between (the synthetic crossover is at $\rho \approx 3.9$, App. F); rotate the volume so that axis meets the method's hard-coded pooling axis (or, equivalently, rotate the pooling), train, rotate back. On all three instruments the tilt is planar, so this is one signed angle in the z–y plane. When A is orthogonal the two candidates collapse onto one direction, and for a natively grid-aligned acquisition that direction is z, so the recipe returns the convention where the convention is already right.

Why the proxy is trustworthy. No isotropic ground truth exists for these volumes, so $\Delta\mathrm{EV}$ is a proxy, and we validate it rather than assume it. The rotation enters both branches of $\Delta\mathrm{EV}$ and cancels; for the learned arms its cost is measured (.0003 EV, App. H). On isotropic synthetic fields arms A, B, C and $\mathrm{D}_{\mathrm{sign}}$ tie to 0.001, and on 48 *real* blocks (App. H) isotropized—same radial spectrum, same noise, direction destroyed—the arms tie to ± 0.002 in every fold, so nothing in the pipeline prefers an oblique axis a priori. Where ground truth exists (synthetic fields blurred along 41°) the proxy's verdict is reproduced against the *unblurred* original (App. A) and by Richardson–Lucy deconvolution (§5.3).

5 Experiments

§5.1 asks whether the geometry is there; §5.2–5.3 what assuming z costs a restorer and a deconvolution; §5.4–5.5 what does and does not close that gap, and §5.6 what a true reference says. Data: 18 OTLS prostate-biopsy volumes from 7 patients in the public PCa_Bx_3Dpathology collection (Xie et al., 2022), cut into 107 blocks of 160^3 ($0.896\,\mu$m isotropic voxels); two dual-view OPM datasets (an organoid with raw ND2, BigStitcher affine and official deskewed views; a symmetrically fused stage-scanned organoid culture, App. E); one lattice light-sheet cell (raw stage-scanned frames only). Blocks of one biopsy never straddle a fold; the four folds are patient-disjoint and intervals on arm differences are 95% t-intervals over the four fold means (App. H).

5.1 The geometry is there, and it is operational

The sharp lobe lies within 10° of 135° in 103/107 blocks (median 135°, IQR 133–137; per-biopsy medians 132–140 in 18/18), and the spectral tensor puts the worst axis at 41° (IQR 35–50), in the $+z{+}y$ quadrant in 104/107 blocks. In the network-free DSDR sweep, after null subtraction, the tensor axis tracks the operational optimum (block-level rank correlation $+0.55$ $[+0.36, +0.71]$ (block bootstrap) on the 15° sweep grid; biopsy-level $+0.64$ over the 18 biopsies); and the signed sweep (Fig. 1e) puts the cheapest direction at $\mathbf{n}$ in 64/107 blocks and 12/18 biopsies. On raw cost the ordering is robust: z is $1.6\times$ the cost along $\mathbf{n}$ (block median; 104/107 blocks above 1, biopsy medians above 1 in 18/18) and $-\mathbf{n}$ at $2.3\times$ (106/107); equivalently, 39% of the variance lost by decimating along z is avoidable by decimating along $\mathbf{n}$ instead ($1 - 1/1.64$; biopsy-level median 37%, IQR 34–43%).

5.2 What the assumed axis costs on the same degraded block

Every arm is the same 3-D U-Net restorer (App. H) that maps a block decimated by k along one axis and linearly refilled, with the mask channel, to the full block; arms differ only in the axis their training pairs are decimated along: z (A, convention), measured θ^* (B), $\theta^* + 45^\circ$ (C, deliberately wrong), a random assumed axis per block drawn from $[0^\circ, 90^\circ)$, the half the affine's sign selects ($\mathrm{D}_{\mathrm{sign}}$), the same drawn from the full $[-90^\circ, 90^\circ)$ ($\mathrm{D}_{\mathrm{none}}$), one global angle read from the file, $\mathbf{n} = 45^\circ$ (E), and the wrong sign $\theta^* - 90^\circ$ (F); three seeds at every k. Every arm is then applied to the *same* held-out blocks (107, in four patient-disjoint folds), decimated once along the file's axis $\mathbf{n}$—the direction §5.1 finds cheapest to under-sample—and linearly refilled (App. H). Three criteria were pre-registered on the proxy-task protocol of App. A, in which each arm also decimates along its own axis: (1) B$-$A>0 with an interval excluding 0 and exceeding .011, the gain the same correction gives with cubic interpolation and no network, (2) C$<$B, (3) A$<\mathrm{D}_{\mathrm{sign}}$ at some k. Under the shared-degradation protocol, which we report because the proxy-task protocol confounds the prior with where the decimation lands, (1) holds at every k (the B$-$A interval clears .011) and (2) and (3) hold in 4/4 folds. Under the proxy-task protocol (2) and (3) hold, and (1) holds at $k = 8$ and 16 but not at $k = 4$, where B$-$A is $+.002$ $[-.000, +.004]$, an interval that does not reach the margin; A also exceeds $\mathrm{D}_{\mathrm{sign}}$ by .0014 (4/4 folds) there, which (3), stated for some k, allows.

Left unrestored, the shared block scores .974, .845 and .559 at $k = 4$, 8 and 16. The restorer trained assuming z returns .963, .819 and .582: it makes the block *worse* than it found it at $k = 4$ and 8 ($-.011$ and $-.027$ against none; 0/4 folds improve it) and helps only at $k = 16$ ($+.023$), where the file-angle restorer gains $+.062$. The restorer pointed at the file's angle returns .988, .879 and .622—an improvement in 4/4 folds at every k and $+.025$, $+.060$ and $+.040$ above the convention (12/12 pairs; the per-block measured axis gives $+.022$, $+.051$, $+.035$). $\mathrm{D}_{\mathrm{none}}$, the augmentation-style

k	none	A (z)	B (θ^*)	E (file, 45°)	C	D_{sign}	D_{none}	F (wrong sign)	E−A per fold
4	.9736	.9627	.9844	.9878	.9701	.9779	.9671	.9562	+.0251 (.025/.028/.027/.020)
8	.8452	.8186	.8695	.8791	.8335	.8528	.8337	.8089	+.0605 (.061/.064/.060/.056)
16	.5591	.5819	.6172	.6215	.5819	.6072	.5848	.5572	+.0396 (.046/.037/.035/.042)

Table 1: Held-out EV of each arm's restorer on the *same* block, decimated along the file's axis **n**, the direction §5.1 finds without any network (App. H); "none" is that block left unrestored. E−A>0 and B−A>0 in 12/12 fold×seed pairs at every k; 95% t-intervals over the four fold means (3 df): E−A $+.025$ $[+.020, +.031]$, $+.060$ $[+.056, +.065]$, $+.040$ $[+.032, +.048]$ at $k = 4/8/16$; B−A $+.022$, $+.051$, $+.035$. Differences are computed before rounding; with four folds the per-fold differences are the primary evidence.

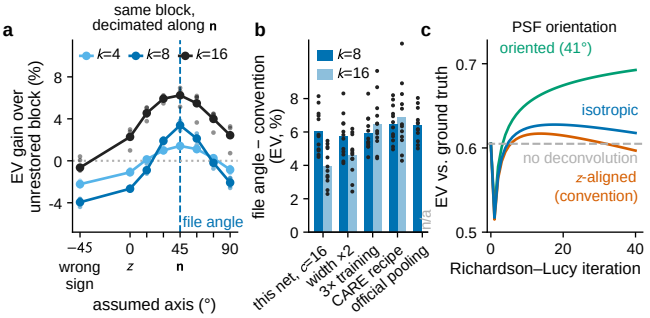


Figure 2: **What the assumed axis costs.** (a) The cost of misalignment: eight restorers, each trained assuming one fixed axis, applied to the same block decimated along **n**; EV gain over the unrestored block, fold means (points) and their average (lines). The curve peaks at the file angle, is symmetric about it and is lowest at $-45°$; the convention (0°) sits below zero at $k \leq 8$. (b) The file angle's gain over the convention survives doubling width, 3× training, the CARE recipe and its official pooling ($k = 8$ only). (c) Richardson–Lucy with three PSF orientations on volumes with a known 41° PSF: the z-aligned PSF falls below the un-deconvolved input.

baseline (a different orientation on every training block; App. H), beats the convention at $k \leq 8$ (4/4 folds) but at $k = 16$ only on average (1/4), and never the file's: E leads it by $+.021/+.045/+.037$ at $k = 4/8/16$ (4/4 folds each). Only the wrong-sign arm is worse than the convention (0/4 folds). Two further controls hold (App. H, J): E−A stays $+.029/+.083/+.064$ without the mask channel and $+.018/+.040/+.024$ with the test block decimated along each block's own θ^* (12/12 pairs at every k). On a downstream measurement that does not use EV, 3-D nucleus detection scored against the stored block's own detections, the file-angle restoration scores F1 .692 against .645 for the convention and .633 unrestored (E>A in 103/107 blocks and 4/4 folds; 100–105/107 under all four detector settings tried; App. J). The protocol is symmetric: on the same block decimated along z instead, the convention gains $+.028/+.057/+.112$ over the unrestored block (12/12 pairs at every k) and E−A is $-.032/-.068/-.072$ (0/12 pairs at every k; App. J). Eight restorers trained at fixed axes (0° to 90° in 15° steps, and $-45°$, the mirrored file angle) trace the curve directly (Fig. 2a, App. J): single-peaked at the file angle at $k \leq 8$ (best of the eight in 320/321 and 304/321 cells; at $k = 16$ the peak flattens, 166/321); symmetric, with restorers at equal misalignment on either side scoring within .007 EV of each other at every k; and lowest at the mirrored file angle, which sits below the file angle in 12/12, 12/12 and 12/12 fold×seed pairs; as a tolerance, a 15° error costs .013–.015 EV at $k = 8$ and a 45° error .055–.060.

5.3 Deconvolution falls below its input under the convention

Richardson–Lucy (RL) deconvolution (Richardson, 1972; Lucy, 1974) on synthetic volumes with a 41° PSF (ground truth known): after a first iteration that lowers all three below the input, the oriented PSF rises monotonically to EV .693; the isotropic PSF peaks at .628 then decays; the z-aligned PSF peaks at iteration 14 and declines to .597, *below the un-deconvolved input* (.605; Fig. 2c). Real-data RL is reported in App. C as ordering evidence only (oriented < z-aligned in 18/18 biopsies).

	$k = 8$				$k = 16$			
Recipe / capacity	A (z)	B (θ^*)	E (file)	E−A (pairs)	A (z)	B (θ^*)	E (file)	E−A (pairs)
This network (c=16)	.819	.870	.879	+.060 (12/12)	.582	.617	.622	+.040 (12/12)
Width ×2 (c=32)	.829	.875	.887	+.057 (12/12)	.578	.620	.624	+.046 (12/12)
Training ×3	.837	.881	.896	+.059 (12/12)	.587	.638	.652	+.065 (12/12)
CARE recipe	.829	.878	.893	+.065 (12/12)	.547	.593	.616	+.069 (12/12)
CARE, official pooling	.832	.880	.896	+.064 (12/12)	–	–	–	–

Table 2: The gap survives capacity and recipe changes on the shared block. Held-out EV of A, B and E on the same block decimated along $\mathbf{n}$ (unrestored: .845 at $k = 8$, .559 at $k = 16$); c=16, 1,500 steps unless stated (×3 = 4,500), CARE = csbdeep defaults, three seeds; parentheses count fold×seed pairs of 12 with E>A.

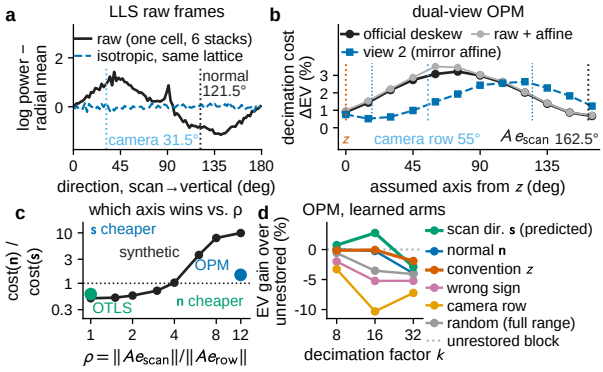


Figure 3: **Two more instruments.** (a) Lattice light-sheet raw frames: lattice-aware angular spectrum versus an isotropic field on the same lattice; peak and trough 100° apart, each 9–20° from its metadata angle. (b) Dual-view OPM: decimation cost by direction on the official product, on our own deskew of the raw frames (view 1), and on view 2, whose mirror-image affine predicts the mirrored curve; costliest within 5–20° of the predicted camera row. (c) Which candidate wins as a function of ρ on synthetic volumes (App. F; band = range over three fields): the cost ratio $\mathbf{n}$ over $\mathbf{s}$ crosses 1 at $\rho \approx 3.9$ and the two instruments (points) fall on the predicted sides. (d) Learned arms on the OPM product, every arm applied to the same block degraded along $\mathbf{s}$ (App. G): only the predicted scan direction improves the block at $k = 8$ and 16; the convention and the normal sit at zero, the others fall below, and at $k = 32$ no restorer recovers anything.

5.4 What is not sufficient, what is not necessary

Neither doubling width, training three times longer, nor the CARE recipe closes the gap (Table 2): E−A stays within +.057–.065 at $k = 8$ and +.040–.069 at $k = 16$, 4/4 folds and 12/12 pairs in every variant, and the convention's restorer stays below the unrestored block at $k = 8$ in every variant (0/4 folds). Per-block measurement is not necessary: the file's own angle, $\mathbf{n} = 45°$, and per-block θ^* stay within .01 of each other, the order following the test axis: on the shared block E leads at $k \leq 8$ (E−B +.0034 [+.002, +.005], +.0095 [+.005, +.014]; 4/4 folds) and the two tie at $k = 16$ (+.0044 [−.001, +.010]; 3/4 folds, 10/12 pairs), under the proxy-task protocol the two match to .002 (App. A), and along each block's own θ^* B leads by a similar margin (App. J).

5.5 A published recipe, official implementations, and two more instruments

CARE-isotropic (Weigert et al., 2018; App. H): on the shared block the convention loses +.065 to the file angle at $k = 8$ and +.069 at $k = 16$, 4/4 folds; with the official max-pool/nearest-upsampling variant, +.064 (Table 2). **IsoVEM** (He et al., 2023; official code, unmodified, App. H; $k = 8$; proxy-task protocol): the convention loses +.038 to the file angle and +.036 to the measured axis (4/4 folds each), all of it where its pooling lands: over their own pooled inputs both gain alike (App. I). **niiv** (Troidl et al., 2025; official model and training regime, App. I; $k = 8$): at three times

the official budget, with both loss curves flat, the convention loses $+.040$ to the file angle on the shared block and $+.058$ under the proxy-task protocol (4/4 folds each), but niiv re-pools its input along its own assumed axis, so neither arm reaches the unrestored .845; over that ingested input the file angle gains $+.008$ to $+.019$ more than the convention (2/4 folds with linear, 4/4 with nearest upsampling; App. I). At the official budget the convention's restorer underfits and the shared-block gap widens to $+.335$. **Dual-view OPM,** $35°$ **sheet** (BioImage Archive, 2025; Fig. 3b): on the official deskewed volume the most expensive direction to under-sample is $75°$ (predicted camera row $55°$), the cheapest $165°$ (the scan direction, App. D), and z costs $1.33\times$ the optimum; our own deskew of the raw frames gives the same curve (App. D). Learned arms on the same product, every one applied to the same block degraded along $\mathbf{s}$ (Fig. 3d, App. G): only the restorer pointed at $\mathbf{s}$ improves the block ($+.007$ and $+.028$ over no restoration at $k = 8$ and 16, 4/4 folds, 12/12 pairs; $+.009$ and $+.028$ over the convention); here z is only $17.5°$ from $\mathbf{s}$, and the convention's restorer sits at the unrestored block. Discarding every second camera frame of the raw stack before the deskew gives a degradation whose stored-frame direction is set by the file, not by us (App. G): the restorer pointed at $\mathbf{s}$ improves that block over its unrestored state ($+.003$ with the mask channel, $+.016$ without; 12/12 pairs) where the convention's does not; a random axis gains as much only with the mask channel; at every third frame the mask-free margin grows to $+.026$ (12/12) while the masked one thins to 7/12. A second OPM dataset, two mirror-image views fused, is the case where the convention is *right*: its ρ falls inside the band where §4 asks for a sweep, and the sweep selects z, where the two views' normals bisect (App. E). **Lattice light-sheet,** $31.5°$ (BioStudies, 2020; raw frames only; App. A; Fig. 3a): sharp lobe and deficit are $100°$ apart ($41°$, $141°$) but 9–$20°$ from the metadata values ($31.5°/121.5°$); one cell, no stored volume: a partial match.

5.6 The correction, an axis-conditioned alternative, and a true-reference test

The alternative to the rotation recipe hands the angle to the network: the U-Net of Table 1 works in the stored frame, never rotates its input, receives the assumed axis a as two constant channels $(\cos 2a, \sin 2a)$ and a FiLM modulation (Perez et al., 2018), and is trained once on blocks degraded along six random axes each (App. J); the *same weights* are then applied to the shared block and told the file angle, z or the wrong sign. Told the file angle the network scores .987/.875/.628 ($k = 4/8/16$), within $-.001/-.005/+.007$ of the rotate-input arm E; told z it loses $+.018/+.027/+.003$ (12/12, 12/12 and 11/12 pairs; 4/4 folds at every k), and told the wrong sign $+.040/+.058/+.007$ (12/12 pairs at every k); at $k = 16$ the gain comes from the axis-agnostic training, not from the told axis. We test the rotation where a true reference exists and no resample is involved: 27 central 160^3 crops of the SynthRAD2023 brain cone-beam CT validation set (Thummerer et al., 2023), each blurred along a random oblique axis θ_i (App. J); five rotate-input arms are applied to the same block decimated along θ_i and scored against the unblurred crop (three case-disjoint folds, three seeds). The restorer pointed at θ_i beats the convention by $+.025$ and $+.024$ at $k = 8$ and 16 (9/9 pairs each), and the convention gains only $+.002$ over the unrestored block at $k = 8$ (4/9) while the one pointed at θ_i gains $+.027$ (9/9). The axis-conditioned network on the same test uses the axis only weakly ($+.004$ over told z at $k = 8$, $-.001$ at $k = 16$) and trails the rotation recipe by .016 (App. J), so its validation is OTLS-only.

6 Discussion and limitations

Whenever data are linearly resampled by some A before learning, pinning an anisotropy prior to a grid axis silently assumes the transported axis (§3) is still a grid axis; in practice, read A from the file, form $\mathbf{n}$ and $\mathbf{s}$, choose by ρ (§4) and rotate the volume, or the pooling axis, by that one angle. At coarse decimation ($k = 16$) the axis-conditioned network of §5.6 matches the file angle even when told z, so the rotation is not the only remedy, and for ρ between 2 and 5 a sweep must decide. The two candidates are specific to stage-scanned acquisition; on the primary instrument the sheet angle is $45°$, so the mirrored file angle is also the axis orthogonal to $\mathbf{n}$. The evidence is one primary instrument (18 biopsies, 7 patients), two OPM datasets (learned arms on one organoid only) and one lattice light-sheet cell; the cone-beam CT test checks the rotation correction, not the transport, and official code separates the prior from where resolution is lost only weakly (App. I, J). With no isotropic reference (App. E), every OTLS score is taken against the stored block; the nucleus endpoint (App. J) changes the score, not the reference.

REPRODUCIBILITY STATEMENT

All measurement and training scripts, the guard logs (passing runs on the real pipeline; the broken-input checks each guard had to fail first are described in App. A), the pre-registered criteria written into the script headers, the seeds, and every result JSON behind every number in the paper are provided as anonymized supplementary material; the deskew affines are read from the datasets' own BigDataViewer/BigStitcher XML files (§3). The synthetic guards and null models are specified in App. A; the spectral band and estimator settings in §4 and App. B; the block extraction, folds and arms in §5 and App. H (OPM and IsoVEM arms: App. G, I). The dual-view OPM data (BioImage Archive, 2025) are public under CC0 and the lattice light-sheet frames (BioStudies, 2020) under BioStudies' terms; the OTLS volumes are the public PCa_Bx_3Dpathology collection (Xie et al., 2022; 2023; Clark et al., 2013) (CC BY 4.0); the brain cone-beam CT volumes are the SynthRAD2023 validation set (Thummerer et al., 2023), public under CC BY-NC 4.0 and used for research only; the block and crop coordinates used are listed in the released code.

ETHICS STATEMENT

All data are public: the OTLS volumes are de-identified prostate biopsies from The Cancer Imaging Archive (Clark et al., 2013; Xie et al., 2023) (CC BY 4.0); the OPM and lattice light-sheet data are from the BioImage Archive and BioStudies; the brain cone-beam CT volumes are the de-identified SynthRAD2023 challenge data (Thummerer et al., 2023) (CC BY-NC 4.0). No new human or animal data were collected. The work identifies a misspecified assumption in existing software and proposes a correction; we see no potential for harmful use.

AI USE STATEMENT

We used a generative AI coding assistant (an LLM) throughout this work, under author verification at every step.

Subtasks with required disclosure, used: proposing and refining hypotheses; helping develop the conceptual framework and formulating the closed-form statements of §3; designing and giving feedback on methodology and experiments, including the guard and null-model protocol and the arm design; implementing methods (the assistant drafted the analysis, training and figure code); generating the synthetic fields used by the guards; reading and reformatting the raw microscopy files into the analysis cubes; and interpreting results.

Subtasks with required disclosure, not used: writing proofs or supplying ingredients for proofs (the paper contains no theorems or proofs beyond the one-line Proposition of §3, whose orthogonality argument was derived and checked by hand); translation. *Not applicable*: qualitative and thematic data analysis.

Subtasks with recommended disclosure, used: drafting and editing parts of the manuscript; producing the figures; searching and summarizing literature; formatting references.

Verification. Every number in the paper was produced by a script that the authors ran and inspected against the result files released in the supplementary material. Every guard was first executed on a deliberately broken input and seen to fail for the right reason before it was allowed to pass; one such run exposed a sign error in the rotation convention that had survived an earlier guard (App. A). The closed-form statements of §3 were re-derived by hand and checked against the measured axes on two instruments. Every citation was verified against the publisher's or arXiv's own page. We take responsibility for the final content of this work, including all text, claims and artifacts produced with the aid of generative AI.

REFERENCES

Adi Ahituv, Anat Ilivitzki, and Moti Freiman. CRIS: Cross-plane self-supervised isotropic restoration for anisotropic volumetric imaging across modalities. *arXiv preprint arXiv:2606.15967*, 2026.

BioImage Archive. High content 3D imaging by dual-view oblique plane microscopy (BioImage Archive S-BIAD2314). doi: 10.6019/S-BIAD2314, 2025.

BioStudies. Lattice light sheet sample data – EB3/H2B (mitosis). BioStudies accession S-BSST435, ebi.ac.uk/biostudies, 2020.

Bi-Chang Chen, Wesley R Legant, Kai Wang, Lin Shao, Daniel E Milkie, Michael W Davidson, Chris Janetopoulos, Xufeng S Wu, John A Hammer, Zhe Liu, et al. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, 346(6208):1257998, 2014. doi: 10.1126/science.1257998.

Bohao Chen, Yanchao Zhang, Yanan Lv, Chenxun Deng, Hua Han, and Xi Chen. SkelEM: Training-signal decoupling of skeleton and diffusion for self-supervised axial super-resolution in volume microscopy. *arXiv preprint arXiv:2606.30012*, 2026.

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013. doi: 10.1007/s10278-013-9622-7.

Chris Dunsby. Optically sectioned imaging by oblique plane microscopy. *Optics Express*, 16(25): 20306–20316, 2008. doi: 10.1364/OE.16.020306.

Marc Finzi, Gregory Benton, and Andrew Gordon Wilson. Residual pathway priors for soft equivariance constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. arXiv:2112.01388.

Longmi Gao and Pan Gao. VEMamba: Efficient isotropic reconstruction of volume electron microscopy with axial-lateral consistent Mamba. *arXiv preprint arXiv:2603.00887*, 2026.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J. Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models. *arXiv preprint arXiv:2512.02012*, 2025.

Adam K Glaser, Nicholas P Reder, Ye Chen, Chengbo Yin, Linpeng Wei, Soyoung Kang, Lindsey A Barner, Weisi Xie, Erin F McCarty, Chenyi Mao, et al. Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues. *Nature Communications*, 10:2781, 2019. doi: 10.1038/s41467-019-10534-0.

Jia He, Yan Zhang, Wenhao Sun, Ge Yang, and Fei Sun. IsoVEM: Isotropic reconstruction for volume electron microscopy based on transformer. *bioRxiv*, 2023. doi: 10.1101/2023.11.22.567807.

Yumeng He, Zanwei Zhou, Yekun Zheng, Chen Liang, Yunbo Wang, and Xiaokang Yang. EMGauss: Continuous slice-to-3D reconstruction via dynamic Gaussian modeling in volume electron microscopy. *arXiv preprint arXiv:2512.06684*, 2025.

David Hörl, Fabio Rojas Rusak, Friedrich Preusser, Paul Tillberg, Nadine Randel, Raghav K Chhetri, Albert Cardona, Philipp J Keller, Hartmann Harz, Heinrich Leonhardt, Mathias Treier, and Stephan Preibisch. BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples. *Nature Methods*, 16(9):870–874, 2019. doi: 10.1038/s41592-019-0501-0.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. arXiv:1412.6980.

Talley Lambert. cudaDecon/pycudadecon: CUDA-accelerated 3D deconvolution. GitHub repository tlambert03/pycudadecon, 2023.

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. *arXiv preprint arXiv:2511.13720*, 2025.

Zelin Li, Chenwei Wang, Zhaoke Huang, Yiming Ma, Cunmin Zhao, Zhongying Zhao, and Hong Yan. Volume Tells: Dual cycle-consistent diffusion for 3D fluorescence microscopy de-noising and super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. arXiv:2503.02261.

Yiyang Lu, Qiao Sun, Xianbang Wang, Zhicheng Jiang, Hanhong Zhao, and Kaiming He. Bidirectional normalizing flow: From data to noise and back. *arXiv preprint arXiv:2512.10953*, 2025.

Leon B Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745–754, 1974. doi: 10.1086/111605.

Hyoungjun Park, Myeongsu Na, Bumju Kim, Soohyun Park, Ki Hean Kim, Sunghoe Chang, and Jong Chul Ye. Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy. *Nature Communications*, 13:3297, 2022. doi: 10.1038/s41467-022-30949-6.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11671.

Tobias Pietzsch, Stephan Saalfeld, Stephan Preibisch, and Pavel Tomancak. BigDataViewer: visualization and processing for large image data sets. *Nature Methods*, 12(6):481–483, 2015. doi: 10.1038/nmeth.3392.

Stephan Preibisch, Fernando Amat, Evangelia Stamataki, Mihail Sarov, Robert H. Singer, Eugene Myers, and Pavel Tomancak. Efficient Bayesian-based multiview deconvolution. *Nature Methods*, 11(6):645–648, 2014. doi: 10.1038/nmeth.2929.

William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972. doi: 10.1364/JOSA.62.000055.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28.

Adrian Thummerer, Erik van der Bijl, Arthur Galapon, Jr., Joost J. C. Verhoeff, Johannes A. Langendijk, Stefan Both, Cornelis A. T. van den Berg, and Matteo Maspero. SynthRAD2023 grand challenge dataset: Generating synthetic CT for radiotherapy. *Medical Physics*, 50(7):4664–4674, 2023. doi: 10.1002/mp.16529.

Jakob Troidl, Yiqing Liang, Johanna Beyer, Mojtaba Tavakoli, Johann Danzl, Markus Hadwiger, Hanspeter Pfister, and James Tompkin. niiv: Interactive self-supervised neural implicit isotropic volume reconstruction. In *MICCAI Workshops (EMA)*, 2025. doi: 10.1007/978-3-032-13961-0_26.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Martin Weigert, Loic Royer, Florian Jug, and Gene Myers. Isotropic reconstruction of 3D fluorescence microscopy images using convolutional neural networks. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2017. arXiv:1704.01510.

Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12):1090–1097, 2018. doi: 10.1038/s41592-018-0216-7.

Weisi Xie, Nicholas P Reder, Can Koyuncu, Patrick Leo, Sarah Hawley, Hongyi Huang, Chenyi Mao, Nadia Postupna, Soyoung Kang, Robert Serafin, Gan Gao, Qinghua Han, Kevin W Bishop, Lindsey A Barner, Pingfu Fu, Jonathan L Wright, C Dirk Keene, Joshua C Vaughan, Andrew Janowczyk, Adam K Glaser, Anant Madabhushi, Lawrence D True, and Jonathan T C Liu. Prostate cancer risk stratification via nondestructive 3D pathology with deep learning–assisted gland analysis. *Cancer Research*, 82(2):334–345, 2022. doi: 10.1158/0008-5472.CAN-21-2843.

Weisi Xie, Nicholas P. Reder, Can Fahrettin Koyuncu, Patrick Leo, Sarah Hawley, Hongyi Huang, Chenyi Mao, Nadia Postupna, Soyoung Kang, Robert Serafin, Gan Gao, Qinghua Han, Kevin Bishop, Lindsey Barner, Pingfu Fu, Jonathan Wright, Christopher Keene, Joshua Vaughan, Andrew Janowczyk, Adam Glaser, Anant Madabhushi, Lawrence True, and Jonathan Liu. 3D pathology of prostate biopsies with biochemical recurrence outcomes: raw H&E-analog datasets and image translation-assisted segmentation in 3D (ITAS3D) datasets (PCa_Bx_3Dpathology). The Cancer Imaging Archive, 2023.

A GUARDS AND NULL MODELS

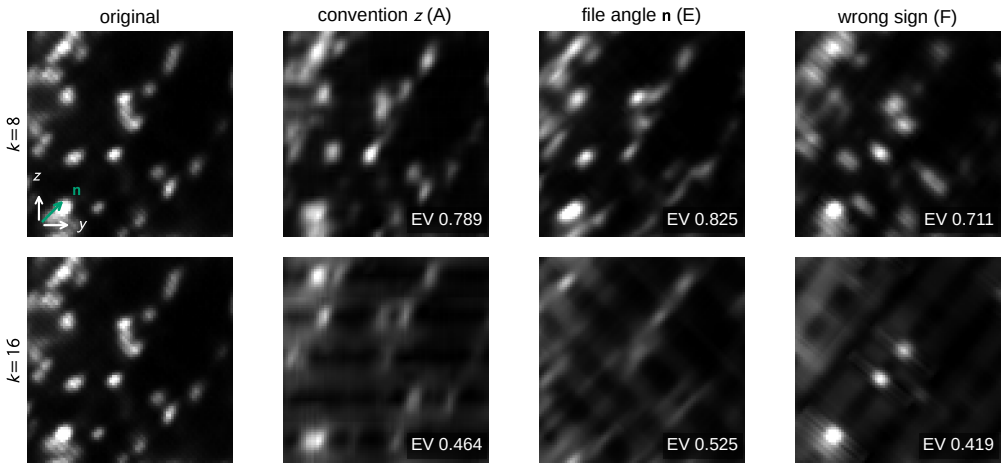


Figure 4: **What the numbers look like.** Central z–y slice of the scored 80^3 region of one held-out block (fold 0, seed 0; the block whose E−A gap is the median of its fold at $k = 8$, not the best case), restored by the same network under three assumed axes, each decimating along its own assumed axis (the proxy-task protocol of App. A, not Table 1's shared-degradation protocol). At $k = 8$ the file angle keeps the nuclei closest to the original; at $k = 16$ the convention leaves z-periodic banding and the wrong sign the heaviest residue, while the file angle, the lowest-MSE arm, is also the smoothest; its EV advantage does not show as brighter nuclei. EV is on the shown region.

Figure 4 shows what one restored block looks like under three assumed axes. Every guard below was first run on a deliberately broken input and seen to fail for the right reason before it was allowed to pass on the real pipeline.

Direction-swept decimation (DSDR). Three known-answer synthetic fields (160^3, power-law spectrum β=3): an isotropic field must give a flat cost curve (peak-to-trough 0.016 EV, 0.9 standard errors over 8 angles); a field blurred along $0°$ must return $\arg\min = 0°$ (rank correlation of cost with angular distance $+1.000$); a field blurred along $41°$ must return $\arg\min = 41°$ ($+1.000$). The blur is a 1-D Gaussian along the injected direction, $\sigma = 2.5$ voxels for the model-free guard and $\sigma = 1.1$ for the learned-arm guard (the latter chosen to give a spectral anisotropy ratio of 1.22, below the real blocks'). The first version of the $41°$ guard failed and exposed a sign error (`ndimage.rotate` moves content from ϕ to $\phi{+}a$, so aligning direction a with z requires rotating by $-a$); the $0°$ guard is a fixed point of that error and had passed. Every $\arg\min$ guard therefore includes a non-fixed-point injected angle.

Learned arms. The network of Table 1 was run on the same synthetic fields (arms A, B, C and $\mathrm{D_{sign}}$): isotropic input gives a peak-to-trough of 0.0010 across arms; an injected $41°$ axis with spectral anisotropy ratio 1.22 (below the ≈ 1.4 of the real blocks, so the guard is the harder case) gives B−A $+.0138$ and C−B $-.0172$. The CARE-recipe port gives identical values on these fields (arm A: .7831 isotropic, .9584 injected), because a stationary Gaussian field has a unique linear MMSE restorer that any sufficiently trained network converges to; the recipes separate only on real, non-Gaussian data. On the 48 real blocks of the earlier pass isotropized as in the null below (same radial spectrum and noise, direction destroyed), the four arms at $k = 8$ give A .7670, B .7669, C .7677, F .7670 (per-fold B−A $-.002, +.001, -.002, +.002$): the rotation–interpolation step confers no advantage at the real spectral content. The IsoVEM official code passes the same guard (isotropic peak-to-trough .0014; injected B−A $+.0169$, C−B $-.0210$) once its own average-pooling degradation is used: feeding decimated slices instead shifts its output by 3.5 planes (the pooling-centre convention) and drives EV to ≈ 0, which we found with a cross-correlation diagnostic before running any real arm.

Each arm's own proxy task carries its ordering before any network runs. In the protocol on which the criteria were pre-registered, every arm decimates along its *own* assumed axis, so the arms do not receive the same degraded block. Held-out EV under that protocol (same folds, seeds and network as Table 1):

k	A (z)	B (θ^*)	E (45°)	C	$\mathrm{D}_{\mathrm{sign}}$	$\mathrm{D}_{\mathrm{none}}$	F
4	.9867	.9884	.9887	.9841	.9853	.9769	.9671
8	.8447	.8785	.8792	.8402	.8517	.8130	.7645
16	.5848	.6229	.6209	.5745	.5951	.5624	.5169

Scoring the linearly filled input itself, with no network, against the stored block gives the same ordering:

	$k=4$		$k=8$		$k=16$	
$X-\mathrm{A}$	no network	restored	no network	restored	no network	restored
B	$+.0152$	$+.0018$	$+.0591$	$+.0338$	$+.0874$	$+.0381$
E	$+.0151$	$+.0020$	$+.0577$	$+.0346$	$+.0848$	$+.0362$
$\mathrm{D}_{\mathrm{sign}}$	$+.0094$	$-.0014$	$+.0331$	$+.0070$	$+.0414$	$+.0104$
$\mathrm{D}_{\mathrm{none}}$	$-.0009$	$-.0097$	$-.0022$	$-.0316$	$+.0010$	$-.0223$
C	$+.0005$	$-.0026$	$-.0011$	$-.0045$	$-.0086$	$-.0103$
F	$-.0176$	$-.0196$	$-.0581$	$-.0801$	$-.0625$	$-.0679$

The gap between the measured axis and the convention is $+.059$ at $k = 8$ before restoration and $+.034$ after it: the restorer *narrows* the gap, because the z-decimated input is the most damaged and has the most room to recover. Under that protocol B$-$A measures where the decimation lands, not what the prior contributes; the same holds for $\mathrm{D}_{\mathrm{sign}}$ and for the OPM arms' own proxy tasks (App. G). Only $\mathrm{D}_{\mathrm{none}}$ and F show a cost beyond the decimation ($-.029$ and $-.022$ at $k = 8$: training on a mixture of directions, and the wrong-sign prior). What the prior contributes is measured by giving every arm the same degraded block (Table 1, App. H); the synthetic version of that separation follows below.

Two confounds that cannot produce the ordering (proxy-task protocol). Every arm except A rotates the volume with cubic interpolation, which smooths it and could make restoration easier; but C ($\theta^* + 45^\circ$) and F ($\theta^* - 90^\circ$) rotate by as much as B and land at A and about 0.11 below B respectively ($k = 8, 16$), on isotropic synthetic input arms A, B, C and $\mathrm{D}_{\mathrm{sign}}$ tie to 0.001, and on the 48 real blocks *isotropized* (radial spectrum and noise preserved, direction destroyed) A/B/C/F tie to ± 0.002 in every fold (above), so rotation is not what B gains from. Second, the stored volume carries a high-frequency lattice imprint along $\mathbf{n}$ (App. B); self-referenced scoring counts its loss as a cost, which penalizes decimation along $\mathbf{n}$—the direction the data nevertheless find cheapest. Both confounds push against the result, not for it.

A true isotropic reference on synthetic data. The proxy scores a restoration against the stored volume, which is already blurred along $\mathbf{n}$; decimating along $\mathbf{n}$ therefore destroys less high-frequency content than decimating along a sharp direction, which could favour arm B for a reason unrelated to the geometry. We test that where an isotropic reference exists: eight isotropic power-law fields (160^3, $\beta = 3$) are blurred along 41° ($\sigma = 1.6$ voxels) to stand in for the acquired volume, the arms are trained and evaluated on that volume exactly as on real data, and the restored held-out blocks are scored against the *unblurred* field. The ordering is the one the geometry predicts, under both scorings, EV and SSIM (Wang et al., 2004) (three seeds, two held-out fields each). Against the true reference at $k = 8$: EV B .659, A .633, C .627, F .586; SSIM B .502, A .457, C .447, F .378. At $k = 16$: EV B .500, A .477, C .474, F .450; SSIM B .297, A .273, C .266, F .243. B$-$A is $+.027$ (EV, $k = 8$) and $+.023$ ($k = 16$), and $+.045$ and $+.024$ in SSIM, positive in 3/3 seeds in every case. Self-referenced scoring of the same runs gives B$-$A $+.038$ and $+.034$, so it inflates the gap by 40–50% but does not manufacture it.

Holding the degradation fixed. In every arm above the assumed axis is also the axis the block is decimated along, so the comparison mixes a misspecified prior with a difference in how much content the decimation destroys. On the synthetic fields we separate the two: each arm is still

trained on pairs it synthesizes along its own assumed axis (what the methods themselves do), but all four are then evaluated on the *same* volume, decimated once along the true 41° axis, and scored against the unblurred field (eight fields, two held out, three seeds, $1{,}500$ steps). Running the same trained networks under the proxy-task protocol reproduces the numbers above to 10^{-5}, so the two protocols differ only in the evaluation. EV against the isotropic reference is A .6383, B .6579, C .6373, F .6357 at $k = 8$ and A .4743, B .4992, C .4637, F .4593 at $k = 16$. Arm A undergoes two fewer rotations than the others; a rotate-only control (the same rotations, no decimation, no network) puts that handicap at .0025 EV, and subtracting it gives B$-$A $= +.0218$ and $+.0270$, positive in 3/3 seeds at both k, against $+.0268$ and $+.0233$ under the proxy-task protocol—the measured-axis advantage is not what the coupling was producing. What the coupling *was* producing is the wrong sign's severity: corrected F$-$A is $-.0005$ ($k = 8$) and $-.0129$ ($k = 16$), against $-.047$ and $-.027$ when F also decimates along its own axis, so most of F's margin in the proxy-task table above is the price of decimating a sharp direction rather than of the prior; C behaves the same way ($+.0021$, $-.0074$). With the degradation shared, on these synthetic fields being wrong by 41°, 45° or 90° costs about the same (on the real blocks the intermediate arms do sit between, Table 1): what the file buys is being right. In SSIM the B$-$A gap is $+.0149$ and $+.0134$ (3/3 seeds, uncorrected). One caveat on the mechanism: our network reads the sampling pattern as a second input channel, and replacing that channel with a constant at test time—out of distribution for every arm, so indicative only—shrinks B$-$A to $+.0064$ ($k = 8$, 3/3 seeds) and $+.0030$ ($k = 16$, 2/3), both rotate-only corrected; a test-time substitution cannot separate the prior from the pattern, and a network trained without the channel shows that the gap is not the channel (E$-$A is larger without it; App. H). The unmodified IsoVEM code, which takes no such channel, shows the same proxy-task gap, although App. I traces it to where its pooling lands; the one-channel network keeps E$>$B$>$D$_{\text{sign}}>$A on the real blocks, with the wrong-sign arm below the convention at $k = 4$ and 16 and level with it at $k = 8$ (App. H).

Isotropized null. The null keeps the radial power spectrum of each block ($\sqrt{\text{mean}|F|^2}$ per shell—averaging $|F|$ instead biases power down by 22% by Jensen's inequality, which the guard caught) and randomizes phases with the correct conjugate symmetry. A phase-randomized-only control that keeps $|F|$ preserves direction and is used as the guard's "must fail" input. On those 48 blocks the null's per-block angular depth (peak-to-trough over the $\pm 90^\circ$ sweep) is 0.002 EV (median) against 0.029 for the data; on the 107 blocks it is 0.002 against 0.030, and the null's median curve is flat to 0.001 (Fig. 1e).

Where D$_{\text{none}}$'s deficit comes from (proxy-task protocol). The four fixed-angle arms sit at offsets $-\theta^*$ (A), 0 (B), $+45^\circ$ (C) and $+90^\circ$ (F) from each block's measured axis, so their mean estimates what a model trained on one fixed direction loses on average—separately from the cost of training on a mixture of directions. On the model-free sweep, where the full 12-point curve is available, this four-point estimate reproduces the directional mean to 0.06 percentage points under the measured θ^* distribution, so we apply it to the learned arms. Of D$_{\text{none}}$'s gap to A, the per-direction term accounts for $53/40/45\%$ at $k = 4/8/16$ and training heterogeneity for the remaining $47/60/55\%$. The per-direction term is the wrong sign's: dropping F, the mean of A, B and C alone is within .0003 of A at $k = 4$ and lies *above* it at $k = 8$ and $k = 16$ ($.9864/.8544/.5941$ vs. $.9867/.8447/.5848$).

Literature search. Title/abstract searches of PubMed (three queries: deskew* + point-spread function; light-sheet + deconvolution + tilt/oblique/skew + PSF; lattice light-sheet + PSF + deskew/shear; 0/0/0 hits) and title scans of the CVPR 2026 (4,984 papers) and ICLR 2026 proceedings and of recent ICML, NeurIPS and ICCV proceedings (oblique, deskew, shear, through-plane, slice thickness, voxel spacing) returned no work on where the resolution axis lands in the stored volume.

Lattice-aware estimator for raw frames. For the lattice light-sheet stacks the raw index axes (scan, camera y, camera x) are not orthogonal in physical space; we map index frequencies through the reciprocal lattice, $\nu = L^{-\top} f$ (L the lattice matrix in physical units, f the index-space frequency), and validate on an isotropic field sampled on the same lattice, which returns an angular depth of 0.26–0.48 against 1.5–3.2 for the data.

B SPECTRAL READINGS: WHERE THE IMPRINT IS

Along rays in the z–y plane (all 107 blocks; median [IQR] of log power minus the all-direction radial mean): the camera axes are broadband sharp (x: $+2.4$ $[+2.0, +2.9]$ at $|\nu| = 0.2$–0.3; the deskewed camera row at $135°$: $+0.9$ $[+0.7, +1.2]$ at $|\nu| = 0.1$–0.3, peaking at $|\nu| \approx 0.27$). The normal $\mathbf{n}$ ($45°$) is a deficit of -1.3 $[-1.6, -1.0]$ in the sub-band $|\nu| = 0.1$–0.25 of the mid band (z itself: -0.9 $[-1.2, -0.7]$) but carries a $+3.3$ $[+2.4, +4.5]$ excess above $|\nu| = 0.3$: a periodic interpolation imprint of the raw lattice, visible as diagonal hatching in z–y slices. Averaging over the full band 0.05–0.35 therefore misplaces the deficit; all directional readings in the paper use 0.08–0.27. The tensor axis and the operational optimum are unaffected because both are governed by orthogonality to the sharp lobes, not by the deficit itself. The tensor axis ($41°$) nonetheless sits systematically a few degrees closer to z than the operational optimum ($45°$, the normal); grid refinement is ruled out and regression dilution explains only part of the offset.

C RICHARDSON–LUCY ON REAL BLOCKS

Real-data RL has no reference and its anisotropy readout is confounded by noise amplification, so we report ordering only (107 blocks, 18 biopsies; biopsy-level paired comparisons of the residual spectral anisotropy ratio): oriented $<$ z-aligned in 18/18 biopsies (median difference -0.074; 97/107 blocks), oriented $<$ isotropic in 14/18 (-0.034; 71/107), isotropic $<$ z-aligned in 15/18 (-0.051; 79/107).

D MEASURED PSF THROUGH THE AFFINE (OPM)

Applying the file's affine to the raw frames ourselves gives the same decimation-cost curve as the official deskewed product (cheapest $165°$, costliest $60°$, z at $1.39\times$). The BigStitcher bundle ships six measured bead PSFs in the raw index frame (standard deviations 3.2–3.3 px, 3.2–3.3 px, 2.1–2.3 planes; no y–z covariance). Propagating their second moments through the file's affine, $\Sigma_{\text{stored}} = A\Sigma A^{\top}$ (as multi-view deconvolution does, Preibisch et al., 2014), gives a stored-frame ellipsoid with long axis $(x, y, z) = (0, -0.30, 0.95)$, i.e. $162.5° \pm 0.2°$ in the z–y plane for all six PSFs, and an $8\times$ aspect ratio: exactly Ae_{scan}, and the direction the operational sweep finds cheapest ($165°$). With a $1\ \mu$m step against $0.0876\ \mu$m camera pixels, the 2.1–2.3-plane width along the scan becomes 2.1–2.3 planes $\times\ \|Ae_{\text{scan}}\| = 11.97$ voxels per plane, a 25–28-voxel smear in the stored frame against 3.2–3.3 px in the camera plane, so the scan direction, not the camera-plane normal, is the worst axis. This is the optics-only prediction, without any lattice imprint, and it agrees with the stored-volume measurement.

E WHY A SECOND VIEW IS NOT AN ISOTROPIC REFERENCE

The two OPM views are stored on the same grid (cross-correlation peak within one voxel of zero offset; $r = 0.921$ between views, $r = 0.997$ between view 1 and the shipped fusion, $r = 0.948$ between view 2 and the fusion). Scoring the direction-swept decimation of view 1 against view 2 instead of against itself gives a cost curve that is *negative* for directions 0–90° (decimating view 1 there makes it agree better with view 2) and positive for 105–165° (peak at 135°, $+0.71\%$). The sign flip follows view 2's own geometry: its affine is the mirror image of view 1's, so it is blind along 17.5° (its scan) and 35° (its normal) and sharpest along its camera row at 125°; removing view-1 content that view 2 cannot see—including noise and imprint—raises agreement, and only structure view 2 confirms (105–165°) costs. The cross-view curve is rank-correlated -0.36 with the self-referenced one and measures the reference's anisotropy rather than the target's. The same happens to learned restoration: a CARE-isotropic 2-D restorer trained on view 1 under any assumed axis lowers agreement with view 2 by ≈ 0.10 EV (noise amplification into the reference's blind directions), with axis differences an order of magnitude smaller. The joint two-view deconvolution shipped alongside that fusion does not supply one either. It is a different volume: it correlates 0.974 with view 1, and its z–y angular spectrum is view 1's (peak 69°, trough 175°, depth 5.67 vs. 5.68), so it inherits the first view's anisotropy rather than removing it. A reference-based test of anisotropy priors needs an isotropic reference, which neither a single tilted view nor either shipped product can supply.

A symmetric two-view fusion where the convention is right. The archive's stage-scanned colorectal-organoid dataset (directory `S_Figure_10` of S-BIAD2314; two views at $\pm 35°$, fused, $2\times$ binned to $0.706\,\mu$m, $162 \times 1641 \times 1171$) has, in its BigStitcher XML, the mirror-image affines $y' = 0.819y - 1.743z$, $z' = \mp 0.574y$ after a z-calibration of 1.639: scan direction Ae_{scan} along y ($90°$; the calibration scales the scan column, so $\rho = 1.743 \times 1.639 = 2.86$ for each view), camera rows at $55°$ and $125°$ with normals at $145°$ and $35°$. For a symmetric fusion the two sharp planes are bisected by y and the two normals by z, so the file predicts that z is the worst axis of the fused product; $\rho = 2.86$ is below the synthetic crossover, so each view's worst axis is its own normal, and the two normals bisect at z. It does fall in the band where §4 asks for a sweep, and here the two candidates make opposite predictions for the fused volume: the two normals bisect at z, whereas both views' scan directions coincide at y (the two camera rows bisect at y as well). The measurements below select the normal, consistent with $\rho = 2.86$ below the crossover. Measured on five 160^3 blocks of the brightest channel (mid band 0.08–0.27): sharp lobe at $89/87/83/87/101°$, trough at $9/173/9/173/9°$; DSDR cheapest direction $0°$ in 5/5, costliest 90–$105°$. The convention is correct here because the file says it is, not by default.

F Synthetic crossover of the axis-selection criterion

Three isotropic power-law fields ($\beta = 3$, 160^3) are blurred by a camera-plane PSF (lateral $\sigma = 0.8$ voxel, $3\times$ wider along the plane normal), sampled on a sheared lattice whose camera row is tilted $45°$ and whose scan step is ρ pixels ($\rho \in \{1, 1.5, 2, 3, 4, 6, 8, 12\}$), and resampled to an isotropic Cartesian grid by the same linear deskew used for real data. The direction-swept decimation cost ($k = 4$) then gives the cheapest direction for each ρ. The cost ratio cost($\mathbf{n}$)/cost($\mathbf{s}$) rises monotonically with ρ—$0.50, 0.52, 0.57, 0.71, 1.03, 3.6, 7.9, 9.9$—crossing 1 between $\rho = 3$ and 4 ($\rho \approx 3.9$ by log-linear interpolation of the median; 3.9–4.0 in the three fields); the cheapest direction moves from $45°$ ($\mathbf{n}$; $\rho \leq 2$) through $30°$ ($\rho = 3, 4$) to between $-15°$ and $0°$ (toward $\mathbf{s}$; $\rho \geq 6$). The OTLS affine ($\rho = 1.00$, measured ratio 0.61) and the OPM affine ($\rho = 11.97$, measured ratio 1.5 on the official product, 1.4 from the raw frames; both taken at the predicted $\mathbf{n}$ and $\mathbf{s}$ by linear interpolation of the $15°$ sweep) lie on the two sides. The crossover value depends on the PSF anisotropy and the interpolation kernel and is not claimed to be universal; the ordering is.

G Learned arms on the dual-view OPM product

Model-free sweep on view 2 first: view 2, whose mirror-image affine predicts scan $17.5°$ and camera row $125°$, answers $15°$ and $120°$ on its official product, with z at $1.46\times$ the optimum.

Official deskewed view 1 ($0.086\,\mu$m voxels as stored; the raw camera pixel is $0.0876\,\mu$m), $2\times$ block-averaged ($0.172\,\mu$m; one scan step ≈ 6 voxels, $\|Ae_{\text{scan}}\|/2$ in camera-pixel units), eight non-overlapping 160^3 blocks in the organoid, four spatial folds, three seeds, the network and pipeline of Table 1, every arm applied to the same block decimated along the predicted scan direction $\mathbf{s} = 162.5°$ (App. H). Assumed axes: z (A), $\mathbf{s} = 162.5°$ (predicted), $\mathbf{n} = 145°$, wrong sign $17.5°$, camera row $55°$, random.

k	none	A (z)	$\mathbf{s}$ (pred.)	$\mathbf{n}$	wrong sign	camera row	random
8	.9861	.9846	.9934 (4/4)	.9863	.9661	.9533	.9803
16	.9114	.9106	.9390 (4/4)	.9090	.8591	.8083	.8761
32	.6264	.6073	.5976	.5867	.5742	.5541	.5856

"none" is the block left unrestored; parentheses count folds in which $\mathbf{s}$ beats A. At $k = 8$ and 16 the restorer pointed at the predicted scan direction is the only one that improves the block ($+.007$ and $+.028$ over none, 4/4 folds, 12/12 fold×seed pairs; $+.009$ and $+.028$ over the convention), the convention's restorer sits at none ($-.001$, $-.001$; above it in 1/4 and 1/4 folds), and the normal $\mathbf{n}$ ties with the convention ($+.002$, $-.002$). At $k = 32$ five planes remain and no restorer, $\mathbf{s}$ included, recovers anything (every arm $-.019$ to $-.072$ below none); an earlier single-seed run had read the normal as joining $\mathbf{s}$ at $k = 32$, which three seeds do not support. Under each arm's own proxy task the three-seed means reproduce the earlier single-seed values to .008 at $k \leq 16$ and to .028 at $k = 32$. With one organoid this is replication evidence, not a second statistical unit.

A degradation applied in the raw frame. The test degradation is not chosen by us if it is applied to the raw frames before the affine: from view 1 of the raw stack (actin channel) we keep camera frames $0, 2, 4, \ldots$ and linearly refill the discarded ones along the scan index, deskew the full and the frame-dropped stack with the file's affine (linear interpolation) onto one isotropic grid of two raw pixels ($0.0876\,\mu$m each) per voxel, cut eight non-overlapping 160^3 blocks around the organoid on this grid (not the official product's), and deskew the kept-frame indicator the same way (left fractional) to stand in for the mask channel; the random arm's per-block angles are those of the released code. One frame step is 6 voxels along Ae_{scan}, so the frame-dropped block has 12-voxel gaps and the arms are trained at $k = 12$ on the full-frame blocks by decimating along their own axes (six axes, four folds, three seeds; one-channel arms without the mask as well); every arm receives the same frame-dropped block rotated into its own frame and its output is scored against the full-frame block on the central 80^3 cube (no rotation is involved in producing the degraded block). Against the full-frame deskew (unrestored .954), the restorer pointed at $\mathbf{s}$ improves it by $+.003$ with the mask channel and $+.016$ without (12/12 and 12/12 pairs); the convention's sits at $-.003$ and $-.011$ (4/12 and 0/12); the normal $\mathbf{n}$ at $-.004$ and $-.018$, the wrong sign at $-.007$ and $-.071$, the camera row at $-.013$ and $-.108$, a random axis at $+.003$ and $-.024$ (one fixed draw per block, four of the eight within 15° of $\mathbf{s}$); $\mathbf{s}$ leads the convention by $+.006$ and $+.028$ (11/12 and 12/12). Dropping two frames in three is harsher: at $f = 3$ (18-voxel gaps, arms at $k = 18$; unrestored .873) $\mathbf{s}$ gains $+.009$ with the mask channel (7/12 pairs) and $+.026$ without (12/12), the convention $-.012$ and $-.012$ (6/12 and 3/12), the normal $-.046$ and $-.025$, the wrong sign $-.044$ and $-.092$, the camera row $-.034$ and $-.139$, the random axis $+.001$ and $-.047$ (8/12 and 2/12), and $\mathbf{s}$ leads the convention by $+.021$ and $+.038$ (10/12 and 12/12). The direction along which frames were dropped is whatever the affine makes of the scan axis. Without the mask channel the file's prediction—not z and not the normal—is the only arm that restores the block at both factors; with the mask channel its margin is small and a fixed random draw that lies near $\mathbf{s}$ half the time matches it. A plausible reason, untested, is that the deskewed, fractional indicator is unlike the binary masks every arm trained on, and that the mask alone tells any arm where the gaps are. The reference is the full-frame deskew, itself interpolated between frames 6 voxels apart, so this tests the axis, not absolute quality.

H Training and evaluation protocol

Data and folds. Eighteen OTLS biopsy volumes (PCa_Bx_3Dpathology (Xie et al., 2022): cases 1A–1D, 2B, 2D–2F, 3A, 3B, 3F, 4A, 4B, 4D, 4E, 5B, 7A and 49E, i.e. seven patients); six 160^3 blocks per biopsy ($0.896\,\mu$m isotropic voxels), chosen by one rule for every volume (the six highest-signal z positions at least 480 slices apart, centred in x–y, tissue mean above threshold), 107 in all (case 2B contributes five: one candidate sits on a damaged HDF5 chunk and was skipped). Folds are patient-disjoint: four folds holding out {1A–1D}, {4A, 4B, 4D, 4E}, {2B, 2D–2F, 5B} and {3A, 3B, 3F, 7A, 49E} (24/24/29/30 blocks) and training on the remaining biopsies. An earlier pass of this study used 8 of these biopsies (5 patients, 48 blocks) with only two of four folds patient-disjoint; the gap was the same there (B$-$A $+.036/+.041$ at $k = 8/16$ on the patient-disjoint folds, $+.030/+.044$ on the others), and the guards of App. A that quote 48 blocks were run on that set. Three seeds $(0, 1, 2)$ per arm. Arm E uses the file's normal (45°); a global 40° (the rounded median of the per-block θ^*) was run as well and differs from it by $+0.0006$, $+0.0010$, -0.0008 at $k = 4, 8, 16$ ($\mathrm{E}_{45}-\mathrm{E}_{40}$). EV is scored on the central 80^3 cube of each held-out block after rotating back to the stored frame, so that rotation padding never enters the score. Differences from A on the shared block (Table 1): $\mathrm{D}_{\mathrm{sign}}-$A $+.015/+.034/+.025$ (4/4 folds at every k); $\mathrm{D}_{\mathrm{none}}-$A $+.004/+.015/+.003$ (4/4, 4/4, 1/4); C$-$A $+.007/+.015/-.000$ (4/4, 4/4, 2/4); F$-$A $-.007/-.010/-.025$ (0/4 folds at every k); C$<$B in 4/4 folds at every k. Intervals on arm differences are 95% t-intervals (3 df) over the four patient-disjoint fold means of the paired difference, each fold mean averaging its three seeds; with only four clusters a percentile cluster bootstrap has 35 distinct resamples and is not reported.

Where the pooling axis is hard-coded. IsoVEM (`dataset.py`, line 62, applied to a transposed volume) and niiv (`dataio.py`, line 68):

```
        AvgPool3d(kernel_size=(1, self.scale_factor, 1))
    AvgPool3d(kernel_size=[1, 1, int(self.anisotropic_factor)])
```

Which of the eight z-methods we could run. EMGauss, Volume Tells and SkelEM: code unavailable or removed. CARE-isotropic and its reference-free and cross-plane variants share one protocol, for which the csbdeep port below stands in. IsoVEM: run unmodified. niiv: run with its official model and training regime (App. I). So two official codes and one surrogate cover the protocol of five of the eight, and the other three are untested.

Assumed-axis pipeline (all arms; training pairs and the proxy-task evaluation). A block is rotated in the z–y plane by cubic interpolation (reflect padding) so that the assumed axis meets z, degraded along z by factor k, restored, and rotated back by the inverse angle. Arms differ only in the angle. Arm $\mathrm{D}_{\mathrm{sign}}$ draws one angle per block uniformly from $[0°, 90°)$, fixed for the run by the seed and used identically at training and test time, so the network sees many orientations while each held-out block is still scored along a definite assumed axis. That interval is half of the $180°$ axis space—the half containing $\mathbf{n}$, selected by the sign of the affine—so $\mathrm{D}_{\mathrm{sign}}$ encodes the tilt's sign and nothing else. On this instrument the sheet angle is $45°$, so $\mathbf{n}$ is the exact midpoint of that half and z its edge; on a shallower sheet z would sit closer to $\mathbf{n}$ inside the half and the $\mathrm{D}_{\mathrm{sign}}$ margin would shrink. $\mathrm{D}_{\mathrm{none}}$ is the same arm drawn from the full $[-90°, 90°)$ and encodes nothing. Arm F is $\theta^* - 90°$, i.e. orthogonal to the measured axis. It is the arm on the mirror side of z: were θ^* exactly the file's $45°$ it would be the mirrored tilt $-\theta^*$ itself, the ambiguity the affine resolves; with the measured per-block θ^* the two separate, by a median $16°$ over the blocks (the separation is $|2\theta^* - 90°|$, which grows in both directions away from $45°$, so the per-block median exceeds what the median θ^* alone would give). The arms use the per-block θ^* stored with the block cache, the tensor axis folded into $[0°, 90°]$ (median $40°$); §5 and Fig. 1d report the same estimator recomputed on the same blocks with its sign kept (median $41°$). The two are identical on 59 of the 107 blocks and within $1°$ on 71 (median per-block difference $0°$, mean $0.7°$), and the $16°$, $6.5°$ and $10°$ figures of this paragraph move by under $0.5°$ between them. What makes the penalty large is where the arm lands rather than the mirror as such: F sits a median $6.5°$ from each block's own sharp-lobe peak, and the mirrored tilt a median $10°$ from it, so both fall on the sharp lobe and the wrong-sign penalty here is the cost of decimating along it; on a different sheet angle it would be smaller. Runs used one NVIDIA RTX 4090: about 50 s of training and 10–15 s of tiled inference per arm $\times$ fold $\times$ seed for our network, and 27–46 min (median 29) per arm $\times$ fold for the unmodified IsoVEM loop.

Shared-degradation evaluation (Table 1). Training is unchanged: each arm synthesizes its pairs along its own assumed axis. At test every arm receives the *same* input—the held-out block decimated once along the file's axis $\mathbf{n} = 45°$ and linearly refilled, with its observation mask, both rotated back to the stored frame—presented in the arm's own frame (rotated by $-a$), restored, rotated back, and scored on the central 80^3 cube against the stored block. A rotate-only control (the same rotations, no decimation, no network) is scored per arm; arm A undergoes two fewer rotations than the others and the control puts that difference at .0003 EV, so the table reports raw values. The coupled score of the same trained network is recorded alongside and reproduces the proxy-task values of App. A to .002 on the 12-cell means.

Degradation models. Model-free sweep (Figs. 1e, 3b,c): mean-pool k slices along z, then cubic interpolation back to full size; the rotate-only control uses the identical rotation pair without pooling. Our network and the CARE port: keep every k-th slice, fill the gaps by linear interpolation along z, and feed the network two channels—the filled volume and a binary mask of kept slices—with the target being the original block. IsoVEM: its own average pooling (k slices $\to$ one), which is what its official training code assumes; feeding decimated slices instead shifts its output by $(k-1)/2$ planes (3.5 at $k = 8$) and drives EV to ≈ 0 (App. A).

Our network. A three-level 3-D U-Net (Ronneberger et al., 2015) with base width $c = 16$ (channels 16–32–64), two 3^3 convolutions and GELU per level, average pooling down and trilinear up, a 1^3 output convolution, and a residual connection to the linearly filled input channel. Loss MSE; Adam (Kingma & Ba, 2015), learning rate 2×10^{-4}; 1,500 steps of batch 4 on random 64^3 crops (about 6k crops per run); inference tiled with 64^3 windows and 8-voxel margins discarded. The "width $\times 2$" arm sets $c = 32$; the "3$\times$ training" arm uses 4,500 steps. Longer training helps both arms about equally at $k = 8$: from 1,500 to 4,500 steps the file angle's gain over the unrestored block goes from $+.034$ to $+.051$ and the convention's from $-.027$ to $-.008$ ($k = 8$; Table 2), so the convention stays below the unrestored block.

Without the mask channel. The network above reads the kept-slice mask as a second input channel, which App. A flags as a possible source of the convention's deficit under the shared degradation (an oblique mask pattern is out of distribution for arm A). A one-channel network fed only the linearly refilled volume, everything else unchanged (same blocks, folds, seeds, steps and arms), passes the synthetic guards of App. A (isotropic input flat to .002, injected 41° axis B$-$A $+.015$, C$-$B $-.018$) and gives E$-$A $+.029/+.083/+.064$ at $k = 4/8/16$ (12/12 pairs and 4/4 folds at every k), larger than with the mask; the convention's restorer sits .014 and .053 below the unrestored block at $k = 4$ and 8 (0/4 folds improve it) and level with it at $k = 16$ ($+.0001$), the file-angle restorer sits $+.015/+.030/+.064$ above it (4/4 folds), and E$-$B is $+.002/+.008/+.008$. The one qualitative change is the wrong-sign arm: F$-$A is $-.012$ and $-.055$ at $k = 4$ and 16 (F below A in 12/12 pairs) but $-.0003$ at $k = 8$ (F below A in 8/12 pairs and 2/4 folds), consistent with App. A, where most of F's margin is the decimation rather than the prior. The mask channel is not what separates the arms.

CARE-isotropic port. csbdeep 0.8.2 defaults transcribed to PyTorch: residual U-Net of depth 2, $n_{\text{first}} = 32$ (32–64–128), two convolutions and ReLU per level, linear 1^3 output with residual, MAE loss, Adam 4×10^{-4}, batch 16, 1.4M parameters; the "official pooling" variant replaces average pooling and trilinear upsampling by max pooling and nearest-neighbour upsampling, as in the original.

IsoVEM. The official repository's `train.py` and `test.py`, unmodified: 15 epochs, batch 2, learning rate 10^{-3}, its 8-fold rotation test-time averaging and tiled inference ($16 \times 128 \times 128$ tiles for $k = 8$, overlap 8). We only rotate the input block so that the assumed axis meets the array axis its pooling acts on, and rotate the output back.

Direction sweep. Angles -90° to 90° in 15° steps, $k = 4$, scored on the central 80^3 cube; the isotropized null is generated per block (App. A) and swept identically.

I Official code on real blocks: IsoVEM and niiv

IsoVEM. Same folds as Table 1, arms A, B and E (and a global 40°), seed 0, $k = 8$ (15 epochs of the official training loop, average-pooling degradation, official tiled inference with eight-fold rotation averaging), under the proxy-task protocol: IsoVEM's pipeline degrades along its own assumed axis, so its arms do not receive the same block. A (convention): .9131, .8916, .8935, .9048; B (measured axis): .9460, .9402, .9266, .9354; B$-$A $= +.033, +.049, +.033, +.031$ (4/4 folds positive, mean $+.036$). E (the file's angle, 45°, nothing measured): .9446, .9392, .9279, .9422; E$-$A $= +.032, +.048, +.034, +.037$ (mean $+.038$, 4/4 folds), within $+.0014$ of B; a global 40° gives the same (mean $E_{45}-E_{40}$ $-.002$). Absolute EVs are higher than under our own degradation because IsoVEM's average-pooling is milder than slice decimation at the same k; the gap between axes is the size seen under the proxy-task protocol with our network ($+.034$ at $k = 8$, App. A), and it is where the pooling lands: when each arm's pooled input is scored with no network (the same rotation, uint8 scaling and pooling, upsampled back; `niiv_ingest_baseline.py`), the convention and the file-angle arm gain alike over it ($+.088$ and $+.085$ with linear upsampling, a difference of $-.003$, 0/4 folds; $-.002$, 0/4, against nearest). IsoVEM therefore reproduces the direction in which resolution is lost, not a cost of its prior. We report IsoVEM at $k = 8$ only: its shipped architecture ties the attention window and inference tiling to the scale, and running $k = 16$ would have required editing the official model, which would forfeit the point of the comparison.

niiv. The official model and configuration (`config_cvr_s`), the official degradation (average pooling by k along the assumed axis), the official training sample (a random 64×64 lateral-plane crop, matching the side length of the official 64^3 volumes, average-pooled by k along one in-plane axis as input and the crop itself as target), loss (MAE), optimiser (Adam, 5×10^{-5}), batch (70) and epoch count (1500; one epoch is one pass over the training blocks) are used unchanged; our driver reads the blocks from our cache instead of from files, and the LPIPS, DISTS and PSNR that the official loop logs without entering the gradient are not computed (`niiv_arms.py`). Same folds as IsoVEM; arms A and E only; seed 0. Two things had to be settled first. With whole 160^2 planes as training samples (batch 16 for memory), the model plateaued in 13 of 16 runs (4 folds $\times$ 2 arms, seed 0 and one re-seed; training MAE .10–.17, a near-constant output), so we returned to the official 64^2 regime, under which every run converged; the one whole-plane pair in which both arms converged

(fold 3, seed 0) scored .880 (assumed axis z) against .873 (file angle) under the proxy-task protocol, the reverse of the windowed result on the same fold (.843 against .908 at the triple budget below), and we flag it. The 64^2-trained model does not transfer to whole-plane inference: applied to x–z slices of a training block, the official-budget weights fit a 64-pixel window better than the whole 160-pixel plane in 8/8 models (MAE .01–.13 against .05–.21; `niiv_window_vs_plane_probe.json`). Inference therefore tiles every x–z slice with 64×8 input windows and averages the overlaps, identically for both arms (`niiv_eval_tiled.py`). For reference, our network's input, the block decimated and linearly refilled, scores .8452 along $\mathbf{n}$ and .7876 along z. At the official budget (1500 epochs; 3000 steps, two per epoch: a batch of 70 planes and the remainder) the z arm underfits (training MAE .055 against .034 for the file angle) and scores EV .439 (assumed axis z) against .871 (file angle) under the proxy-task protocol and .433 against .768 on the shared block, E>A in 4/4 folds under each protocol and in 107/107 blocks on the shared block; the convention loses $+.335$ to the file angle on the shared block and $+.432$ under the proxy-task protocol. With three times the budget (4500 epochs; 9000 steps, two per epoch: a batch of 70 planes and the remainder) both loss curves are flat (training MAE .033 against .032) and the gap is the size seen with our network and with IsoVEM: EV .827 (assumed axis z) against .885 (file angle) under the proxy-task protocol and .738 against .778 on the shared block, E>A in 4/4 folds under each protocol and in 99/107 blocks on the shared block; the convention loses $+.040$ to the file angle on the shared block and $+.058$ under the proxy-task protocol. Under the proxy-task protocol niiv pools the block it is given along its own axis, so most of that gap is where the pooling lands: the input pooled along the file angle scores $+.040$ above the one pooled along z (.853 against .813 linearly upsampled), and over its own ingested input the file angle gains $+.017$ more ($+.018$ against nearest upsampling; 74/107 and 77/107 blocks, 3/4 folds each). Unlike the arms of Table 1, which receive an identical input, niiv first average-pools whatever it is given by k along its assumed axis (its official test path ingests isotropic data the same way), so on the shared block the two arms' inputs differ after ingestion: linearly upsampled, the input pooled along z scores .724 and the one pooled along the file angle .756 (nearest .705 and .726), and that difference accounts for $+.031$ of the $+.040$ shared-block gap. Measured as each arm's improvement, over its own ingested input each arm gains $+.014$ (z, 101/107 blocks) and $+.022$ (file angle, 105/107), a difference of $+.008$ (62/107 blocks, 2/4 folds; $+.019$, 91/107 blocks, 4/4 folds against nearest upsampling). Because of that pooling, neither arm reaches the unrestored .845. Single seed in the 64^2 regime; the two budgets are the only settings run there.

J Further controls: the test axis, an axis-conditioned network, and a second modality with a true reference

Decimating the test block along each block's own measured axis. Table 1 decimates every held-out block along the file's $45°$, which is also arm E's training axis. Repeating the evaluation with the test block decimated along its own measured θ^* instead (training unchanged; unrestored block .974/.847/.562 at $k = 4/8/16$) makes B the matched arm, and E, whose single angle sits $5°$ from the median θ^* and $8°$ from the test axis at the median block (range 0.5–$42°$), still leads the convention: E−A $+.018/+.040/+.024$ (12/12, 12/12 and 12/12 pairs; 4/4 folds at every k), B−A $+.021/+.049/+.031$, and B leads E by only $+.003/+.008/+.007$. The convention again sits below the unrestored block at $k = 4$ and 8 ($-.007$ and $-.016$; 0/4 and 0/4 folds improve it) and above it at $k = 16$ ($+.031$); F−A is $-.013/-.022/-.040$ (F below A in 12/12, 12/12 and 10/12 pairs); C−A $-.001/-.005/-.024$ (6/12, 4/12 and 0/12 pairs above A) and $\mathrm{D_{none}}$−A $+.000/+.004/-.009$ (4/12, 6/12 and 3/12), so along θ^* the arms C and $\mathrm{D_{none}}$ no longer beat the convention, while $\mathrm{D_{sign}}$−A $+.010/+.022/+.012$ (12/12, 12/12 and 10/12) still does. The file angle's lead is therefore not an artefact of matching the test degradation.

An axis-conditioned network. The rotate-input recipe hands the prior to the network by rotating the volume. To separate the information from the rotation, we train the same U-Net without any rotation: it works in the stored frame, keeps the two image channels of Table 1's network (volume and kept-slice mask), receives the assumed axis a as two constant input channels $(\cos 2a, \sin 2a)$ and as a FiLM modulation (Perez et al., 2018) after every block, and sees each training block degraded along six random axes (App. H otherwise unchanged; guard on synthetic fields with an injected $41°$ or $-30°$ axis: told the true axis 0.9718/0.9718, told z 0.9607/0.9647, told the wrong sign 0.9589/0.9577; a FiLM-only variant failed this guard by margin and was replaced by the two-

channel design before any real-data run). Training pairs are synthesized with the same rotate–decimate–rotate-back operator that produces the shared block, so the two interpolations it carries are seen in training as well as at test. At test the *same weights* receive the shared block of Table 1 and are told the file angle, the block's θ^*, z, a random axis or the wrong sign. Told the file angle the network scores .987/.875/.628, within $-.001/-.005/+.007$ of the rotate-input arm E (told the block's θ^* instead, it stays within .003 of that score at every k); told z it loses $+.018/+.027/+.003$ (12/12, 12/12 and 11/12 pairs; 4/4 folds at every k), told a random axis $+.018/+.027/+.004$, and told the wrong sign $+.040/+.058/+.007$ (12/12 pairs at every k). With the weights fixed, the axis handed to the network is what moves the score at $k = 4$ (told z it sits only $+.006$ above the rotate-input convention); at $k = 8$ the told axis and the axis-agnostic training contribute about equally ($+.027$ and $+.029$); at $k = 16$ the training carries most of the gain (told z it still beats the rotate-input convention by $+.044$) and the told axis matters little ($+.003$).

The protocol is symmetric. Repeating Table 1 with the shared test block decimated along z instead of $\mathbf{n}$ (arms A, B and E; same 107 blocks, patient-disjoint folds, three seeds; unrestored .959/.788/.474 at $k = 4/8/16$) reverses the ordering: the convention gains $+.028/+.057/+.112$ over the unrestored block (12/12 pairs at every k), the file angle sits at $-.004/-.011/+.040$ (1/12, 1/12 and 12/12), and E$-$A is $-.032/-.068/-.072$ (0/12 pairs at every k). The arm whose training axis matches the test degradation wins, so the shared-degradation table measures the cost of a prior that points away from where the degradation lies; which direction that is comes from §5.1 without any network and from App. A on a known synthetic axis.

The cost of misalignment. For the five arms whose assumed axis is fixed per block (A: z; B: θ^*; C: $\theta^* + 45°$; E: $45°$; F: $\theta^* - 90°$), take the angle between the assumed axis and the test axis (undirected, in $[0°, 90°]$) and rank the five EVs against it within every block$\times$seed cell. Under the shared-degradation protocol (test axis $45°$) the rank correlation is Spearman within each block$\times$seed cell, median $-1.0/-0.9/-0.9$ over cells (negative in 321/321, 321/321 and 317/321 cells at $k = 4/8/16$); with the test block decimated along each block's own θ^* it is $-1.0/-0.9/-0.9$ (negative in 321/321, 321/321 and 319/321). A median cell has at most one adjacent pair out of order. The order is the angle's, not the table's fixed arm order: where the angle order of two arms depends on θ^*, the EVs follow the angle. Under the shared protocol C beats B in 30/30, 27/30 and 28/30 cells ($k = 4/8/16$) with $\theta^* < 22.5°$ (there C, at $\theta^* + 45°$, is the arm closer to the $45°$ test axis), and A beats C in 36/42, 33/42 and 42/42 cells with $\theta^* \geq 55°$, clear of the $45°$ crossover, and in 0/30 cells at every k with $\theta^* < 22.5°$. The one exception we found is under the θ^* protocol at $k = 16$: for $\theta^* > 45°$ the angle predicts C over A, and C wins 106/108 and 86/108 cells at $k = 4$ and 8 but wins only 40/108 cells at $k = 16$, where, under this protocol, the cost flattens beyond about $45°$ of misalignment. Restoration quality falls with how far the assumed axis is from the actual one for both test axes, and with the same sign on the block decimated along z (three arms, above); the file supplies the zero of that curve and §5.1 verifies it on the median curve without a network (`misalignment_cost.py`).

A downstream endpoint. EV is a proxy; a measurement that does not use it is 3-D nucleus detection, although its reference is still the stored block. On the shared blocks of every fold (seed 0, $k = 8$), Laplacian-of-Gaussian blob detection (`skimage.feature.blob_log`; $\sigma \in [1.5, 4]$ voxels, six scales, threshold .05) is run on the original block, on the unrestored shared block and on the restorations of arms A and E, saved by `downstream_save.py` (their EVs are within .004 EV of the seed-0 rows of Table 1). The original's detections are the reference (a median of 389 nuclei per 80^3 block); a detection counts as a hit if it lies within 3 voxels of an unmatched reference detection, and the endpoint is F1 against that reference (`downstream_detect.py`). The file-angle restoration detects nuclei more reliably than the convention's, with fewer spurious detections and more reference nuclei recovered: F1 .692 against .645 for the convention and .633 unrestored (precision/recall .732/.658 against .659/.635), E>A in 103/107 blocks and 4/4 folds. Under three further detector settings (threshold .10; match radius 2 voxels; $\sigma \in [2, 5]$) the ordering is unchanged: E>A in 100–105/107 blocks and 4/4 folds under all four detector settings, E$-$A .046–.073 in F1. The two endpoints disagree on one thing: on this one the convention's restoration sits .011–.023 above the unrestored block, whereas in EV it sits below it (Table 1); they agree on the ordering that matters, A<E.

A fixed-angle sweep. To draw the curve rather than sample it, six further restorers were trained under the shared-degradation protocol with assumed axes $15°$, $30°$, $60°$, $75°$, $90°$ and $-45°$ (the mirrored file angle, the only negative angle sampled) and placed beside A ($0°$) and E ($45°$): eight fixed axes, the same network, pipeline, folds and three seeds, every restorer applied to the same block decimated along $\mathbf{n}$ (`misalignment_sweep.py`). Their mean gains over the unrestored block, in the order $-45°$, $0°$, $15°$, $30°$, $45°$, $60°$, $75°$, $90°$, are $-.022$, $-.011$, $+.001$, $+.010$, $+.014$, $+.011$, $+.003$ and $-.008$ at $k = 4$; $-.039$, $-.027$, $-.009$, $+.019$, $+.034$, $+.021$, $-.002$ and $-.021$ at $k = 8$; and $-.006$, $+.023$, $+.045$, $+.059$, $+.062$, $+.055$, $+.040$ and $+.024$ at $k = 16$ (Fig. 2a). As a tolerance, a $15°$ error in the assumed axis costs .013–.015 EV at $k = 8$, a $30°$ error .036–.043 and a $45°$ error .055–.060. The fold-mean curve is single-peaked at every k; per cell the file angle is the best of the eight in 320/321, 304/321 and 166/321 cells (at $k = 16$ the peak is broad: $45°$ leads $30°$ by only .003 (9/12 pairs), and its neighbours at $30°$ and $60°$ take 91 and 48 of the rest). It is symmetric about the file angle: restorers at equal misalignment on either side ($30°$ vs $60°$, $15°$ vs $75°$, $0°$ vs $90°$) score within .007 EV of each other at every k. It is lowest at the mirrored file angle, which sits below the file angle in 12/12, 12/12 and 12/12 fold×seed pairs and below the convention in 12/12, 11/12 and 11/12: a restorer that assumes the right tilt with the wrong sign is worse than one that assumes no tilt at all, the point the sheet angle alone cannot settle and the affine does. On this instrument the mirrored file angle is also the axis orthogonal to $\mathbf{n}$, the far end of the curve; on a sheet angle other than $45°$ the two would separate (App. H). Within block×seed cells, EV falls with the folded angle between the assumed axis and $\mathbf{n}$: median Spearman $-0.98/-0.98/-0.93$ over the eight arms (negative in 321/321, 321/321 and 318/321 cells); the most negative value the tied offsets allow is -0.98, so at $k = 4$ and 8 the median cell is ordered perfectly.

A second modality with a true reference. On real anatomy the reference is the volume itself: 27 central 160^3 crops of the SynthRAD2023 (Thummerer et al., 2023) brain cone-beam CT validation volumes (1 mm isotropic voxels; CC BY-NC), each blurred along its own random oblique axis $\theta_i \sim U(-90°, 90°)$ (1-D Gaussian, $\sigma = 1.6$ voxels; the blur axis stands in for the one a header would record) and stored on the same axis-aligned grid. Arms A (z), E (the injected axis θ_i), C ($\theta_i + 45°$), F ($\theta_i - 90°$) and D (a random axis; seeds 0 and 2 only, since at seed 1 the script drew it with the injected axis's own random stream) train as in App. A ("Holding the degradation fixed"), every arm receives the same block decimated along θ_i, and all are scored against the unblurred crop (three folds by case, three seeds). At $k = 8$ the injected axis beats the convention by $+.025$ (9/9 pairs; $+.026$ after the rotate-only control), the wrong axes lose (C−A $-.013$, F−A $-.030$; below A in 9/9 and 9/9), the random axis sits at $+.000$ (3/6), and against the fold-matched unrestored block the convention gains $+.002$ (4/9) while the injected axis gains $+.027$ (9/9). At $k = 16$ the same holds: E−A $+.024$ (9/9; $+.025$ corrected), C−A $-.010$ (7/9 below A), F below A in 9/9 pairs (median $-.069$, the other eight cells between $-.041$ and $-.138$; one run diverges to EV $-.515$ and drags the mean to $-.185$), the random axis $-.016$ (2/6 above A), and both restorers now improve on the unrestored block ($+.024$ and $+.048$; 9/9 and 9/9), as on the OTLS data at $k = 16$. The crops carry anatomical anisotropy of their own (spectral ratio 1.2–2.6, axis varying from block to block), so an axis *measured* from these data would be confounded by content; an axis *read from a header* would not be, which is the deployable case. We did not use the lattice model of App. F for this test: when the crops are pushed through it (3:1 camera-plane PSF, $\rho = 1$) and directions are swept against the unblurred crop at $k = 8$, the crops' own anisotropy dominates: EV along $\mathbf{n}$ minus EV along z $-.007$, $\mathbf{n}$ cheapest in 0/27 crops (the wrong sign $-\mathbf{n}$ is cheapest on the mean curve), $+.008$ with a 6:1 PSF ($\mathbf{n}$ cheapest in 5/27), and $-.025$ at $\rho = 6$. A mild PSF is not resolvable against this anatomy at these scales, so the test above injects the axis directly.

The axis-conditioned network on the same test. Trained once per fold on the training crops degraded along six random axes each and applied with the same weights to the shared block, the axis-conditioned network above gains only $+.004$ (8/9 pairs) from being told the injected axis instead of z at $k = 8$ and $-.001$ (3/9) at $k = 16$; told the wrong sign it loses $+.012$ and $+.003$ (9/9 and 7/9); and it trails the rotate-input arm E by .016 at $k = 8$ (9/9 pairs) and .003 at $k = 16$. The network does use the axis—the wrong sign costs it, and the told-axis gain grows with $|\theta_i|$ (crop-level $r = 0.63$ at $k = 8$; $+.009$ over the 7 crops with $|\theta_i| > 60°$) and is near zero where θ_i is close to z—but weakly, and it does not reach the rotate-input arm, which matches each block's axis exactly; a plausible reason, untested, is the 18 training crops and a different test axis per crop. Its validation is therefore OTLS-only; the rotation recipe is the one verified against a true reference.