

1 A general algorithm for consensus 3D cell segmentation from 2D segmented stacks

2

3 Authors

4 Felix Y. Zhou^{1,2}
5 Clarence Yapp^{3,4}
6 Zhiguo Shang¹
7 Stephan Daetwyler^{1,2}
8 Zach Marin^{1,2}
9 Md Torikul Islam⁵,
10 Benjamin Nanes^{1,2}
11 Edward Jenkins⁶
12 Gabriel M. Gihana^{1,2}
13 Bo-Jui Chang^{1,2}
14 Andrew Weems^{1,2}
15 Michael Dustin⁶
16 Sean Morrison⁵
17 Reto Fiolka^{1,2},
18 Kevin Dean^{1,2}
19 Andrew Jamieson¹
20 Peter K. Sorger^{3,4,7}
21 and Gaudenz Danuser^{1,2}

22

23

24 Affiliation

25 ¹Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX, USA.

26 ²Cecil H. & Ida Green Center for System Biology, University of Texas Southwestern Medical Center, Dallas,

27 TX, USA

28 ³Laboratory of Systems Pharmacology, Department of Systems Biology, Harvard Medical School, Boston,

29 MA, 02115, USA.

30 ⁴Ludwig Center at Harvard, Harvard Medical School, Boston, MA, 02115, USA.

31 ⁵Children's Research Institute and Department of Pediatrics, Howard Hughes Medical Institute, University of

32 Texas Southwestern Medical Center, Dallas, TX, USA.

33 ⁶Kennedy Institute of Rheumatology, University of Oxford, OX3 7FY UK

34 ⁷Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA.

35

36

37

38 Keywords

39 cell segmentation, instance segmentation, tissue segmentation, generalist 3D segmentation, cellular imaging

40

41

42 Correspondence

43

44 Correspondence to Felix Zhou (felix.zhou@utsouthwestern.edu) or Gaudenz Danuser

45 (gaudenz.danuser@utsouthwestern.edu)

46

47 **Abstract**

48
49 Cell segmentation is the fundamental task. Only by segmenting, can we define the quantitative spatial unit
50 for collecting measurements to draw biological conclusions. Deep learning has revolutionized 2D cell
51 segmentation, enabling generalized solutions across cell types and imaging modalities. This has been driven
52 by the ease of scaling up image acquisition, annotation and computation. However 3D cell segmentation,
53 which requires dense annotation of 2D slices still poses significant challenges. Labelling every cell in every
54 2D slice is prohibitive. Moreover it is ambiguous, necessitating cross-referencing with other orthoviews. Lastly,
55 there is limited ability to unambiguously record and visualize 1000's of annotated cells. Here we develop a
56 theory and toolbox, u-Segment3D for 2D-to-3D segmentation, compatible with any 2D segmentation method.
57 Given optimal 2D segmentations, u-Segment3D generates the optimal 3D segmentation without data training,
58 as demonstrated on 11 real life datasets, >70,000 cells, spanning single cells, cell aggregates and tissue.
59

60 **Main**

61 Instance segmentation is the problem of unambiguously assigning each pixel in a 2D or voxel in a 3D image
62 to unique objects of interest. Near universally, it is the first step in quantitative image analysis for many
63 scientific fields including medical imaging¹ and cell biology². It is only through segmentation that the objects
64 of interest to quantify, such as nuclei^{3,4}, organelles⁵, cells⁶, bacteria⁷, plants⁸, organs^{1,9} or vasculature¹⁰, are
65 explicitly identified and delineated within an image. The segmentation subsequently defines the quantitative
66 unit of analysis to extract desired quantitative object features such as morphology¹¹ (e.g. length, area, and
67 volume) and molecular expression (e.g. mean marker expression¹², subcellular patterns¹³) to perform
68 comparative analyses or in downstream processing such as surface unwrapping^{14,15}.

69 Segmentation is easy when cells are isolated, well-contrasted and uniformly illuminated, and amenable to
70 binary intensity thresholding and connected component analysis¹⁶. However, this is rare. In practice, in in-
71 vitro culture, in-situ tissues or in-vivo, cells of diverse morphologies may interact and aggregate together in
72 clusters that cannot be easily or accurately separated by traditional watershed techniques^{2,17}. This is further
73 compounded by inevitable variations in the imaging acquisition and staining used to visualize cellular
74 structures resulting in weak, partial, sparse or unspecific staining of desired foreground structures^{5,18}.

75 Thanks to advancements in GPU architecture, and increased availability of publically available labelled
76 datasets, generalist or ‘foundational’ 2D cell segmentation models have emerged both for interactive
77 segmentation using prompts such as μ-SAM¹⁹, CellsSAM²⁰ and dense segmentation of every cell such as
78 Cellpose⁶ and various transformer models². These methods leverage ‘big data’ and harness diversity in the
79 training data to demonstrate impressive ability to segment 2D cells acquired across modalities and cell types²
80 out-of-the box or with fine-tuning.

81 Physiologically, however, cells interact within complex 3D environments. The importance of studying cell
82 biology processes in the relevant physiological 3D environments is well-documented^{11,21-24}. Moreover the
83 emergence of 3D *in-situ* tissue imaging has further provided unprecedented insights into the complex nature
84 of the tissue microenvironment and its role in development and disease; including novel cell-cell interaction,
85 tissue organization, and diverse cell morphologies^{12,25}. Unlocking the potential of 3D imaging necessarily
86 requires reliable, general and scalable 3D cell segmentation solutions. Simply replicating the training strategy
87 of 2D foundation models is likely prohibitive, requiring significant amounts of well-labelled, diverse 3D cell
88 datasets and dedicated, specialized GPU computing.

89 Despite the relative ease of acquisition, abundance of industrial annotation tools in 2D^{26,27} and ease of crowd-
90 sourcing and proofreading in a single field-of-view²⁸, the Cellpose training dataset comprises just 540 training
91 images (total ~70,000 cells, 5 modalities) and the most recent and largest multimodal challenge² only 1000
92 training images (total 168,491 cells, 4 modalities). Replicating a densely labelled 3D dataset with comparable
93 level of cell diversity and numbers, given more complex microenvironments, more variable image quality, and
94 more diverse morphologies and cell packing is formidable^{5,12,25,29}. Despite ongoing efforts to develop scalable
95 3D annotation tools³⁰⁻³³ with AI assistance^{34,35} and proofreading³⁶, it still requires significant manual expertise
96 and intervention³⁷⁻⁴¹. Moreover, labelling suffers from inter- and intra- annotator variation⁴²⁻⁴⁴ and be biased
97 towards easy cases. Consequently, both classical⁴⁵⁻⁴⁸ and deep-learning based⁴⁹⁻⁵¹ 3D segmentation method
98 development focus primarily on nuclei, which have well-defined round shapes, are separated from neighbors,

99 and visualizable with high clarity by nuclear dyes⁵². The scarce densely annotated, proofread datasets of 3D
100 cells⁵³ have primarily been of plant tissue^{8,54}, or few cell aggregates^{17,55} or of embryos⁵⁶. Even these have
101 few unique images and have been assembled from crops or are of different timepoints of limited true
102 biological replicates. Synthetic^{57,58}, partial⁵⁹ or generative model^{60,61} synthesized datasets have been
103 proposed to alleviate the need for fully labelled data, but have so far only been demonstrated to star-convex
104 morphologies. It is unclear how they generalize to more complex morphologies, image background and future,
105 novel 3D imaging modalities.

106 High quality, annotated datasets with solid ground truth and minimal noise⁶² are not the only limitation. The
107 time to train or fine-tune foundation models is already a major consideration in 2D, requiring significant time
108 investment, memory and specialized GPUs^{1,2,19,20,63} or careful dataset curation⁶⁴. Training comparable 3D
109 models will not only require more time and dedicated resources, but suffer additional challenges such as
110 model overparameterization, necessitating more efficient, revised architecture designs^{10,17,65}. Lastly, even if
111 trained on a vast dataset, foundation models still cannot guarantee generalization nor robustness^{66,67}. SAM
112 models are nevertheless fine-tuned for medical¹ and microscopy¹⁹ and microscopy¹⁹ datasets. Cellpose
113 models also require retraining for best performance⁶⁸. Moreover there is no way to choose what will work *a priori*⁶². Consequently, at the expense of reduced segmentation coverage or accuracy, it is more efficient for
114 academic labs to adopt human-in-the-loop, interactive segmentation tools like ilastik⁶⁹ or to use segmented
115 nuclei as seeds for 3D watershed²⁹.

117 To address the shortcomings of directly training 3D segmentation models, we revisit the idea of leveraging
118 2D segmentations to generate consensus 3D segmentations without data retraining. Using 2D predictions
119 to assist 3D inference is common, largely to minimize computation and training. Primarily this involves
120 adapting pretrained 2D models to 3D for example by inflating 2D convolutional kernels followed by fine-tuning
121 in 3D⁶⁵, or are applied to process the 3D data slice-by-slice and the outputs are combined and processed by
122 a separately trained 3D model⁷⁰. Few works examine no-training approaches. For segmentation, near-
123 universally 2D segmentation tools generate a 3D segmentation by matching and stitching 2D segmentations
124 across xy slices^{6,28} whereby stitching is controlled by an overlap score. Relying on a single view these 3D
125 segmentations are notoriously rasterized and often erroneously join multiple touching cells as
126 tubes^{28,35,52,71-73}. CellStitch⁷⁴ and 3DCellComposer⁵² propose matching across orthogonal xy, xz, yz views
127 to find a consensus 3D segmentation. However these discrete matching approaches are inherently difficult
128 to computationally scale-up with cell numbers and cannot easily handle missing, undersegmented or
129 oversegmented cells across slices. Interestingly, Cellpose⁶ proposed to average predicted 2D flow vectors
130 along the xy, xz and yz directions to construct a 3D gradient map. By tracing the gradient map to the simulated
131 heat origin, the 3D cell instances are found by grouping all voxels ending up in the same sink. Whilst
132 conceptually elegant, its execution has been restricted to Cellpose predicted gradients and demonstrates
133 limited performance on anisotropic⁷⁴, noisy or morphologically non-ellipsoidal datasets¹⁷ despite training. We
134 also empirically observe puzzling fragmentation artefacts around 3D cell centroids in the predicted
135 segmentation, incompatible with its theory and inconsistent with simply stitching the equivalent cellpose 2D
136 cell masks⁷⁴.

137 To derive a formal framework for 2D-to-3D segmentation unifying stitching and Cellpose proposed gradient
138 aggregation, we revisited the instance cell segmentation problem from first principles. We find that the general
139 2D-to-3D aggregation can be formulated as an optimization problem, whereby we reconstruct the 3D gradient
140 vectors of the distance transform equivalent of each cell's 3D medial-axis skeleton. This problem is then
141 solved to obtain the corresponding 3D segmentation using gradient descent and spatial connected
142 component analysis. To generate consensus 3D segmentations for any 2D segmentation method and using
143 any of one, two or all three xy, xz, yz views, we developed a toolbox, u-Segment3D. u-Segment3D
144 implements robust methods and exposes hyperparameters to flexibly handle imperfect 2D segmentations.
145 Moreover it includes preprocessing and postprocessing methods to assist the application of pretrained
146 models on unseen datasets, to correct and to recover missing 3D segmentation features. We first describe
147 our formalism of 2D-to-3D segmentation. We then validate u-Segment3D by optimal, near-perfectly
148 reconstructing the reference 3D segmentations in 11 real-life datasets, >70,000 cells from few cell aggregates,
149 embryos, tissue, and entire vasculature networks from its 2D slice-by-slice segmentations. We then use
150 pretrained Cellpose 2D models to demonstrate how to use u-Segment3D for any 2D method. We further
151 validate u-Segment3D faithfully translates 2D segmentation performance and further exploits complementary
152 information from multiple views to deliver consistent and improved 3D segmentation. In short, the better the
153 2D segmentation, the better the resultant 3D segmentation. Finally using pretrained Cellpose 2D models, we
154 demonstrate the flexibility and capacity of u-Segment3D to segment unseen 3D volume data on sets of

155 anisotropic cell cultures, and unwrapped embryo surfaces¹⁵; high-resolution single cells and cell aggregates
156 with intricate surface protrusions⁷⁵; thin, sprouting vasculature in zebrafish, and tissue architectures imaged
157 with spatial multiplexing²⁵ and electron microscopy⁷⁶.

158 The u-Segment3D package is implemented in Python 3 using open-source packages. Scripts and
159 configuration files are available to facilitate parallel computing and deployment on CPU-based high-
160 performance computing (HPC) clusters for large 3D volumes. u-Segment3D is freely available and can be
161 installed locally from <https://github.com/DanuserLab/u-segment3D>.

162 Results

163 **A formal framework for 2D-to-3D segmentation**

164 Dense instance segmentation identifies every object instance in the image and assign a unique id to all voxels
165 comprising an instance. This is equivalent to: (i) binary labelling every image voxel as foreground (value 1)
166 or background (value 0), and (ii) further assigning to a foreground voxel, a unique positive integer ID, (Fig.
167 1a). Starting with an instance segmentation of touching 2D cells, if we erode each object by 1 pixel from its
168 border, then every cell would become spatially separated. Consequently, unique object IDs becomes
169 redundant. The cells are equally-well represented by a binary foreground/background image, whereby object
170 IDs are parsed by performing connected component analysis to identify spatially contiguous regions, (Fig 1b).
171 This binary image can be factorized and equally-well computed from its 1D slices in either x- and y- directions,
172 (Fig 1c i). Within each 1D slice we can independently identify spatially contiguous 1D regions as unique 1D
173 ‘cells’, (Fig 1c ii). However, when 1D slices are restacked into a 2D image, regardless of the 1D labelling,
174 only unique 2D objects retain spatial adjacency across 1D slices (Fig. 1c iii). Treating all 1D cells as
175 foreground and applying connected component analysis re-identifies 2D contiguous regions and fully
176 reconstructs the original 2D instance segmentation (Fig. 1d iv). This immediately implies that given the perfect
177 1D instance segmenter that accurately delineates the cell boundary even when touching in 1D, the ideal 2D
178 segmentation can be reconstructed: run the 1D segmentation slice by slice, eroding each unique 1D ‘cell’ to
179 ensure spatial separation, combine the 1D slices identified from x- and y- scans into a 2D binary image, apply
180 spatial connected components to identify spatially contiguous 2D objects and reverse the amount eroded in
181 1D. Combining orthogonal views is now necessary to resolve touching cell boundaries. These first principle
182 arguments hold equally in 3D and higher dimensions. In 3D however, instead of 1D ‘slices’ we have x-y, x-z,
183 y-z 2D ‘slices’. Generally, $n-1$ D segmentation implies n D segmentation.
184

185 However, how much do we need to erode to guarantee applicability to heterogeneous cell size and
186 morphology? Firstly, the ideal erosion process should be uniform with respect to individual cell boundary, as
187 this ensures maximal separation from neighbors at all points on the boundary. Secondly, we observe that
188 iterative application of uniform erosion evolves any cell shape to its medial axis skeleton⁷⁷ (MAT). At this
189 point, every object must be spatially separated. Crucially for 2D-to-3D segmentation, the skeleton of 2D
190 slices coincides with the medial axis skeleton of the corresponding 3D object. Resolution permitting, the 2D
191 skeletal slices of each unique 3D object remain spatially proximal after 3D stacking, enabling identification
192 by spatial proximity. Then, reversing the erosion will yield the 3D instance segmentation. To implement a
193 reversible erosion process we note that the medial-axis skeletons are attractors of distance transforms^{77,78},
194 Φ and are the ‘ridges’ in the distance transform^{78,79}. Thus uniform erosion is performed by iteratively advecting
195 the foreground coordinates with stepsize η , in the direction of the local gradient, $\nabla\Phi$. Finally, as the distance
196 transform is continuous and real-valued, the 3D distance transform gradients can be approximately
197 reconstructed from its 2D slice counterparts. The general 2D-to-3D segmentation algorithm (Fig. 1d) is thus:

- 198 1. Generate all 2D segmentations in orthogonal x-y, x-z, y-z views.
- 199 2. Choose a distance transform specifying the 2D medial-axis skeleton and apply this to compute the 2D
200 gradients for the 2D segmentations in x-y, x-z, y-z views.
- 201 3. Reconstruct the 3D gradients of the distance transform $\nabla\Phi_{xyz}^{3D}$ from the 2D gradients, using an averaging
202 function, F

205 where $\nabla\Phi_{xyz}^{3D} \approx F(\nabla\Phi_{xy}^{2D} + \nabla\Phi_{xz}^{2D} + \nabla\Phi_{yz}^{2D})$

206 Use F to also reconstruct the 3D foreground binary, B

207 4. Identify all (x, y, z) foreground coordinates in B .

208 $\text{Foreground} := \{(x_1^{t=0}, y_1^{t=0}, z_1^{t=0}), \dots, (x_n^{t=0}, y_n^{t=0}, z_n^{t=0})\} | B(x_n^{t=0}, y_n^{t=0}, z_n^{t=0}) = 1$

210 5. Apply gradient descent in 3D to iteratively propagate all foreground coordinate points for a fixed number
211 of total iterations, T , to uncover its 3D skeleton attractor

212
$$(x_n^t, y_n^t, z_n^t) \leftarrow (x_n^{t-1}, y_n^{t-1}, z_n^{t-1}) - \eta \nabla\Phi_{xyz}^{3D}(x_n^{t-1}, y_n^{t-1}, z_n^{t-1})$$

214 6. Group all coordinates at final advected positions by spatial proximity and assign to each a unique positive
215 integer object id.

216
$$L^{t=T}(x_n^T, y_n^T, z_n^T) = id \in \mathbb{Z}^+$$

217 7. Transfer the labels back from the final to the initial coordinates to obtain the 3D instance segmentation.

218
$$3D \text{ segmentation} := L^{t=0}(x_n^{t=0}, y_n^{t=0}, z_n^{t=0}) = L^{t=T}(x_n^T, y_n^T, z_n^T) = id \in \mathbb{Z}^+$$

219 We assume throughout that gradients are unit length normalized, $\nabla\Phi \leftarrow \frac{\nabla\Phi}{|\nabla\Phi|}$. In order to operationalize this
220 conceptual algorithm effectively for real datasets and models that predict imperfect 2D segmentations we
221 developed a comprehensive toolbox, u-Segment3D.

223 ***u-Segment3D is a toolbox to create consensus 3D segmentations from 2D segmentations***

224 u-Segment3D aims to execute a robust, consistent and scalable framework of generating consensus 3D
225 segmentations given an input volume and 2D segmentations or model outputs from orthogonal views, (Fig.
226 1e, Suppl. Movie 1). To achieve this, u-Segment3D implements robust algorithms for each key component
227 of the outlined general 2D-to-3D segmentation algorithm. To retain flexibility to account for imperfect 2D
228 segmentations, algorithms expose tunable hyperparameters.

229 First is implementing multiple distance transforms to allow choices between speed, accuracy, and
230 compatibility with 2D model outputs, (Extended Data Fig. 1). There is no unique definition nor computation
231 for an object's medial-axis skeleton^{77,78}. u-Segment3D considers a single centroid lying on/near the medial-
232 axis as valid and usable for 2D-to-3D segmentation. Our implemented distance transforms are categorized
233 into two classes; 'explicit' (Extended Data Fig. 1b) or 'implicit' (Extended Data Fig. 1c) by how their attractor
234 is specified. Explicit transforms define explicit attractor coordinates which are incorporated as boundary
235 conditions in computation (Methods). This ensures gradients are 0 in the attractor and enables stable
236 convergence via gradient descent (Suppl. Movie 2). u-Segment3D implements single 'point' and multi 'point
237 set' source attractors. The single point is the internal medial centroid, whose placement is adjustable by
238 percentage thresholding of the cells' Euclidean distance transform (EDT) (Extended Data Fig. 1d, Methods),
239 extending the definition in Cellpose⁶. The point set is its 2D skeleton (Methods). To compute the distance
240 transforms, u-Segment3D considers two different partial differential equations (PDEs); the Eikonal equation
241 which gives the geodesic solution and the Poisson equation which gives the heat diffusion solution, as used
242 in Cellpose⁶. The Eikonal equation is faster to solve using the fast marching method⁸⁰ but generate less
243 smooth gradients. The Poisson equation is solved exactly using LU decomposition (Methods).

244 Implicit transforms specify the medial skeleton implicitly as ridges. Consequently, convergence to the attractor
245 is unstable^{7,78} (Suppl. Movie 2) but is more efficient, requiring solve only the PDEs without additional
246 constraints. The Eikonal equation can then be solved using EDT which is also an intermediary output of many
247 2D segmentation models^{3,7}. Irrespective of the chosen distance transform, it is imperative for 2D-to-3D
248 segmentation that the distance transform is computed in the cell to faithfully capture its shape. Iterative
249 solutions implemented by Cellpose to solve the Poisson equation are not applicable. When diffusion is
250 restricted in elongated and torturous structures, the gradients collapse to zero (Extended Data Figure 1e). u-
251 Segment3D's exact solution remains robust, never zero even in very long structures, as evidence by raising
252 the distance transform by an exponent p , Φ^p to avoid floating-point introduced non-unique gradients.

253 Second is implementing a content-based averaging function, F to fuse 2D image stacks, (Extended Data Fig.
254 2). 2D slice-by-slice segmentation may miss or under- or over-segment a cell across slices. Inspired by
255 multiview image fusion^{81,82}, u-Segment3D fuses multiple image stacks using linear inverse local variance

256 weighting (Extended Data Fig. 2a, Methods). Using EDT as example, segmentation errors across slices
257 cause non-continuity such that erroneous pixels have high local variance. Using inverse weighting the value
258 of pixels from images with high-variance are down-weighted in the final fusion (Extended Data Fig. 2b).
259 Increasing the size of the local pixel neighborhood enables correcting larger errors. For a 1x1x1 pixel
260 neighborhood, F is the mean average fusion of Cellpose⁶ and therefore no error correction. With a 5x5x5
261 pixel neighborhood, binary thresholding on the fused EDT perfectly recovers the foreground nuclei without
262 artefacts (Extended Data Fig. 2c).

263 Third, is robustly implementing gradient descent in 2D and 3D, (Extended Data Fig. 3). For downstream
264 spatial proximity clustering gradient descent must propagate points of the same attractor together whilst
265 retaining spatial compactness (Extended Data Fig. 3b). We verify our implementation using a synthetic 2D
266 image of two objects, a circle within a doughnut (Extended Data Fig. 3a). Though simple, the object gradients
267 is complex with features typical of more nuanced morphologies such as local sinks and separating flows of
268 opposite orientation. Running 100 iterations, whereas Cellpose implementation has orphaned foreground
269 points of the ring and has an erroneous line attractor for the circle, u-Segment3D propagates points stably
270 converging towards their two point attractors (Extended Data Fig. 3c) and perfect reconstruction of the original
271 objects.

272 Last is implementing robust spatial proximity clustering using image-based connected component analysis,
273 (Extended Data Fig. 4). Too many or too few clusters directly translates to over- and under- segmentation.
274 With heterogeneity in cell shape, points will not converge to their attractors at the same time. Running gradient
275 descent to ensure convergence for all cells is limiting in 3D. Consequently clustering must generalize to
276 uniform point densities and irregular-shaped attractors. Adaptive local histogram thresholding used by
277 Cellpose⁶ is sensitive to point density. Density-based clustering are sensitive to the minimum number of
278 points or radius used to define a cluster. u-Segment3D instead exploits the fact that foreground coordinates
279 are on an image grid (Extended Data Fig. 4a). The final advected coordinates are rasterized (if floating-point)
280 using flooring (step i). A count of the number of coordinates in each voxel is tabulated (step ii) and smoothed
281 with a Gaussian filter of σ to build an approximate kernel density heatmap, ρ (step iii). ρ is sparse, enabling
282 clusters represented by regional hotspots to be identified using a global threshold, $mean(\rho) + k \cdot std(\rho)$
283 where k can be used for adjustment. Connected component analysis labels all spatially contiguous regions
284 with unique ids (step iv). The final segmentation is generated by indexing into this labeled image at the final
285 advected coordinates of foreground voxels, and transferring the labeling to initial coordinates. ρ enables
286 probabilistic cluster identification. By increasing σ u-Segment3D can ‘fuzzy’ link erroneously multiple clusters,
287 equivalent to merging segmentations in the final 3D. We validated our implementation, by reconstructing the
288 2D cell segmentation as we propagate foreground coordinates along the gradients of the geodesic centroid
289 distance transform (Extended Data Fig. 4b i,ii). As expected, initially (iteration 0), the segmentation is identical
290 to applying connected component analysis to the foreground binary. As iterations increase, and attractors
291 are found, detected cell numbers converge on the true number (Extended Data Fig. 4b iii, top). Correspondingly,
292 segmentation quality, measure by the intersection-over-union (IoU) and F1 score,
293 increases to 1 (Extended Data Fig. 4b iii, bottom). These observations translate also to elongated, touching
294 cells (Extended Data Fig. 4c). Moreover, only our clustering recovers the number of clusters present in the
295 final coordinates propagated by either Cellpose or u-Segment3D’s gradient descent for the synthetic image
296 of a circle within a doughnut (Extended Data Fig. 3d, 4d). In contrast Cellpose’s clustering artificially breaks
297 up what should be single clusters (Extended Data Fig. 3d). This critically impacts Cellpose 3D segmentation
298 of low signal-to-noise ratio cells (Extended Data Fig. 4e,f). Whereas, Cellpose 3D grossly oversegments and
299 fractures individual cells, u-Segment3D’s gradient descent and connected component clustering recovers
300 complete cell segmentations when applied to reparse the same predicted foreground binary and 3D gradients
301 (Extended Data Fig. 4e,f, Suppl. Movie 3).

302 To maximize the utility of pretrained 2D models, u-Segment3D further implements preprocessing and
303 postprocessing modules. The image to segment may not reflect the quality, acquisition, noise distribution and
304 modality of the training dataset that a model was trained on. Preprocessing can help transform input images
305 to improve performance^{29,64}. However it is dataset and model-specific. Nevertheless, the following general
306 order of processing implemented by u-Segment3D works well in practice: intensity normalization, none or
307 any combination of denoising, deconvolution and ridge feature enhancement, and uneven illumination
308 correction with optional gamma correction (Methods). Postprocessing follows the order of filtering out
309 implausible segmentations based on size and consistency with the reconstructed 3D gradients (as in

310 Cellpose⁶), then optionally, spatial-connectivity aware label diffusion to refine segmentations to better adhere
311 to cell boundaries within a guide image, and guided filtering to recover missing or intricate subcellular details
312 to the individual cell segmentations. No postprocessing requires further 3D training (Methods).

313 The rest of the paper explores in detail each module of u-Segment3D, highlighting salient parameters and
314 specific modifications for application to real datasets.

315

316 **3D smoothing of reconstructed 3D gradients from 2D and suppressed gradient descent are 317 essential for 2D-to-3D segmentation**

318 To understand how the different components of the 2D-to-3D algorithm may impact 3D segmentation, we
319 first empirically investigated 1D-to-2D segmentation of cell morphologies from the Cellpose⁶ and Omnipose⁷
320 training datasets, which can be intuitively visualized (Extended Data Fig. 5). We first examined the
321 approximation of 2D gradients using 1D gradients (Extended Data Fig. 5a). To compute 1D gradients, we
322 consider each disconnected 1D region as a unique ‘cell’. Then, for each cell, we computed the distance of
323 its coordinate to the slice centroid, took central differences and unit length normalized the vectors. The 2D
324 gradients are reconstructed by stacking the x- and y- direction gradients and smoothing with a 2D Gaussian
325 filter, width σ . We then performed gradient descent to recover the 2D segmentation and compute the gradient
326 descent trajectory. Across single cells representing spherical, convex, branched and vessel morphologies,
327 Gaussian filtering was essential to recover the original 2D segmentation (Extended Data Fig. 5b). With no
328 smoothing, 1D gradients have insufficient 2D context. The reconstructed 2D gradients can fail to specify a
329 single fixed-point attractor. Consequently trajectories do not connect all foreground pixels to a single point,
330 unlike 2D computed gradients. 2D Gaussian filtering however restores lost 2D correlations. Increasing σ ,
331 trajectories are regularized ultimately converge to single points. The smoothing is conformalizing the initial
332 shape, shifting its 2D centroid towards the centroid of its convex hull. For concave structures, this attractor
333 may lie outside the cell. To examine the implications of this for gradient descent, we considered full image
334 segmentations. For a Cellpose exemplar (90 cells), across many cells, the reconstructed 2D gradients
335 smoothed by $\sigma = 1$ contains more than one attractor (Extended Data Fig. 6a). Consequently, after 50
336 iterations of gradient descent we oversegment (143 cells). Nevertheless, the reconstruction is good ($F1 =$
337 0.77, $IoU=0.91$), with the fragmentation splitting off largely small cell fragments. many cells are correctly
338 segmented. Expectedly, increasing σ regularizes the reconstructed gradients. For $\sigma = 5$, segmentation (93
339 cells, $F1=0.94$, $IoU=0.93$) is on-par from ideal 2D gradients (93 cells, $F1=0.98$, $IoU=1.00$). Beyond $\sigma > 5$,
340 gradients interact across neighboring cells, decreasing IoU and F1 performance (Extended Data Fig. 6a iv).
341 Consequently, the number of predicted cells drops. Thus σ should be less than the smallest separation
342 distance between the medial axis skeletons of any two cells. For long and thin tubular structures in the
343 Omnipose exemplar (86 cells), increasing σ shifts the attractor centroid into neighboring cells. Thus
344 increasing σ improved F1 but also decreased IoU, with an optimal balance at $\sigma = 3$ ($F1=0.49$, $IoU=0.60$)
345 (Extended Data Fig. 6b iv). As motivated, gradient descent is only used to separate adjacent cells, not to be
346 run to convergence. We hypothesize improved segmentations with fewer iterations. To implement this without
347 changing total iterations, we use a variable stepsize⁷, $\eta = \frac{1}{1+\tau t}$ which decays temporally with increasing
348 iteration number t and τ adjusts the decay rate. Applying the suppressed gradient descent⁷ with $\sigma = 1$ now
349 perfectly reconstructed the 2D segmentation (86 cells, $F1=1.00$, $IoU=1.00$) (Extended Data Fig. 6c).

350 To test if insights from 1D-to-2D translate to 2D-to-3D we conducted the analogous reconstruction experiment
351 for single 3D cells (Extended Data Fig. 7a). Similar to 1D-to-2D, the 2D geodesic centroid distance transform
352 (Methods) was computed slice-by-slice in orthogonal xy, xz, yz stacks, treating spatially contiguous 2D
353 regions as unique ‘cells’. The 3D gradients was then reconstructed by averaging (F with 1x1x1 pixel
354 neighborhood). 3D cells were selected to represent a spectrum of distinct morphologies from pseudo-
355 spherical, to pseudo-convex and branched, and with different types of surface protrusions (Extended Data
356 Fig. 7b). Applying suppressed gradient descent ($\tau = 0.1$) for 200 iterations, we found similar results as 1D-
357 to-2D, with 3D cell examples of pseudo-spherical, pseudo-convex and branched morphologies with different
358 types of surface protrusions, (Extended Data Fig. 7b.). Gaussian smoothing aids regularization and
359 increasing σ ensures convergence to a single cell, even for the highly branched cell with filopodia ($\sigma = 15$).
360 The same cell was fragmented into several regions at branch junctions at lower $\sigma = 1$. As expected,
361 increasing $\tau = 0.5$ recovers perfect construction of the branched cell at the lower $\sigma = 1$.

362 In summary, Gaussian filtering of the reconstructed 3D gradients from 2D is key for 3D segmentation, but not
363 implemented by Cellpose 2D. σ should be smaller than the expected minimum distance between 3D cell
364 skeletons. To enable 3D segmentation under this limiting condition for heterogeneous morphologies,
365 suppressed gradient descent is also essential. We next tested how different distance transforms impact 2D-
366 to-3D segmentation in whole datasets.

367

368 ***u-Segment3D reconstructs the ideal 3D segmentation from ideal orthogonal 2D slice-by-slice 369 instance segmentations***

370 We assembled 10 published 3D datasets with dense segmentation labels and 1 additional zebrafish
371 macrophage dataset (Suppl. Table 1). This latter dataset was curated in-house by combining connected
372 component analysis and u-Segment3D generated segmentations⁸³. DeepVesselNet¹⁰ is a dataset of
373 simulated binary vasculature networks. We applied connected component analysis to identify disconnected
374 subnetworks as unique ‘cells’. The total number of cells across all datasets was 73,777. For each reference
375 cell segmentation, we extracted 8 morphological features (Fig. 2a, Methods), chosen to assess cell size (total
376 number of voxels), the extent of elongation (stretch factor) and the topological complexity (# of skeleton
377 nodes). To visualize in 2D the morphological diversity and variation in cell numbers across datasets, we
378 applied UMAP⁸⁴ to the normalized features (Fig. 2b, Methods). Two plant datasets: Arabidopsis (CAM)
379 (24,439) and Ovules (37,027) contribute the majority of the cells (83%) and dominates the UMAP. Random
380 sampling 6 UMAP regions, the assembled datasets captures commonly found 3D morphological archetypes
381 encountered in tissue including thin, complex vessel-like networks (Region 1), pseudo-spherical (Regions 2-
382 4), irregular (Region 5), and tubular or branched (Region 6). Using the per dataset median UMAP coordinate,
383 and colored UMAP by stretch factor, # skeleton nodes and volume, we broadly group the 11 datasets by the
384 three super-morphological archetypes they best represent: complex networks (DeepVesselNet),
385 irregular/branched (Zebrafish macrophages/Platynereis ISH nuclei/MedMNIST3D/Lateral Root Primordia),
386 and convex (*C. Elegans* embryo/mouse organoid/mouse skull nuclei/Platynereis nuclei/Arabidopsis
387 (CAM)/Ovules).

388 For all images in each dataset, we reconstruct the reference 3D segmentation from their ideal 2D slice-by-
389 slice segmentations (Fig. 2c). Scanning the reference 3D segmentation in xy, xz, yz views slice-by-slice, we
390 treat each 2D contiguous region in a 2D slice as a unique ‘cell’. For each 2D ‘cell’, the 2D gradients is
391 computed and used to reconstruct the 3D gradients, to generate the reconstructed 3D segmentation using
392 3D gradient descent and connected component analysis. Under this experiment setup, the foreground will
393 always be correct. This allows us to unambiguously assess 3D reconstruction using 2D gradients. Three
394 different 2D distance transforms were tested: Poisson diffusion centroid as example of an explicit transform
395 and used in Cellpose⁶; Euclidean distance transform as example of an implicit transform and used within
396 models like Omnipose⁷, StarDist³; and geodesic centroid as a second example of an implicit transform, but
397 computed differently (Methods). For all datasets, total gradient descent iterations was fixed at 250, and
398 reference segmentations were resized to be isotropic voxels with nearest-neighbor interpolation (Suppl.
399 Table. 1). Drawing analogy to neural network optimization, we further incorporate momentum into suppressed
400 gradient descent to expedite convergence (Methods). Temporal decay τ was the only parameter we adjusted
401 for each transform and dataset (Suppl. Table 2). For all, postprocessing was only to remove cells < 15 voxels.
402 Reconstructed 3D segmentations were evaluated using the average precision (AP) curve (Methods). The AP
403 curve reports the average fraction of cells matched between reference and predicted segmentations as the
404 overlap cutoff (IoU) for a valid match is increased from 0.5 to 1.0 (perfect overlap) (Methods). We use the
405 notation AP_{0.5} to denote AP with IoU cutoff henceforth. For perfect reconstruction, AP=1 at all IoU. In practice,
406 due to numerical accuracy, AP always drops to 0 above an IoU cutoff.

407 We first analyzed the dataset from each of the three super-morphological archetypes, with the most number
408 of cells: for convex (Ovules, Fig. 2d), for irregular (Lateral Root Primordia (LRP), Fig. 2e) and for networks
409 (DeepVesselNet, Fig. 2f) (Suppl. Movie 5). AP curves are plotted by color for transform: magenta (geodesic),
410 cyan (diffusion) and navy (EDT) and by marker for provided dataset split: circle (train), diamond (test) and
411 square (validation). Impressively, we find near-perfect reconstruction across all distance transforms,
412 morphotypes and data splits, qualitatively and quantitatively: Ovules, AP_{0.5} ≈ 1, LRP, AP_{0.5} ≥ 0.8, and
413 DeepVesselNet, AP_{0.5} ≥ 0.8. As expected, increased τ was required for thinner, branching cells: using EDT,
414 $\tau = 0.5$ for Ovules, and LRP, $\tau = 2.0$ for DeepVesselNet. Findings were also reflected in the other 8 datasets

(Extended Data Fig. 8), with $AP_{0.5} \geq 0.75$. Moreover IoU was high, with the curve decaying prominently at $\text{IoU} \geq 0.85$ and for many, $\text{IoU} \geq 0.95$. $\text{IoU} > 0.8$ masks are near-indistinguishable from the reference by eye^{7,85}. Gradient descent 2D-to-3D aggregation requires a spatial contiguous path in 3D. Consequently generated 3D segmentations naturally enforce the ideal of a single spatial component per object. This was not checked and enforced in the reference segmentation. Since we fix the foreground, the performance gap from an ideal $AP_{0.5} = 1$ largely reflect inconsistent labeling in the reference 3D segmentation. For example, balanced dataset splits should exhibit the same performance. However, in LRP, the AP curve of all three transforms on the validation (val) split were notably worse. In 6/11 datasets (Ovule/Arabidopsis(CAM)/*C. Elegans*/mouse organoid/Platynereis nuclei/vesselMNIST3D), the best distance transform achieved perfect $AP_{0.5} = 1.00$. Notably these are largely convex-shaped datasets or empirically observed to have images with unambiguous cell edges and minimal background. The other datasets are noisier and more morphologically complex, and thus harder to annotate and proofread in 3D. LRP is known to contain spurious labels^{7,44}. For zebrafish macrophages which had the lowest maximum reconstruction performance across all datasets for a transform ($AP_{0.5} = 0.8$ with Poisson), final segmentations were not rigorously proofread. Pretrained Cellpose and u-Segment3D segmentations were replaced by connected component segmentations based on automated hard-coded rules⁸³. Consequently there are small and multi-component ‘cells’. For DeepVesselNet, the errors are over-estimated quantitatively. The average number of subnetworks is 3, thus our results reflect on average 1 misidentified small subnetwork. Qualitatively, there is no noticeable difference in coverage (Fig. 2f), thus errors are likely data resolution-related, for example joining two subnetworks separated by a small gap (Fig. 2f, white arrow), or size filtering removed a small subnetwork or a segment connecting two subnetworks is too small and therefore unsupported during gradient computation.

Importantly we observed bias of different transforms for different morphotypes. There was minimal differences between the two explicit transforms with point-source attractors, Poisson and geodesic. However, both outperformed EDT on convex morphologies, most evidently in the LRP val (Fig. 2e), mouse skull nuclei test (Extended Data Fig 8d), Platynereis ISH nuclei test (Extended Data Fig. 8e) and zebrafish macrophages (Extended Data Fig 8h) datasets. This is primarily due to the increased stability of explicit transforms. EDT was superior for thin and complex vasculature networks (Fig. 2f, DeepVesselNet), by minimizing the distance all points needed to propagate. Overall the quantitative difference was small (<0.5 difference in $AP_{0.5}$) and not as dramatic as suggested by Omnipose⁷. This is because under gradient descent the medial axis 3D skeleton is always an intermediate structure when converging towards a centroid attractor (Suppl. Movie 4). Qualitatively, we visualized both the diffusion and EDT reconstruction on exemplars from Ovules, LRP and DeepVesselNet. Despite similar F1 and IoU, only the EDT fully reconstructed all branching cells. Diffusion fragmented the cell with the longest branch (Fig. 2e, white arrows) into two ‘cells’. Importantly the fragments are standalone and not erroneously part of or included parts of neighbor cells.

In summary, u-Segment3D empirically achieves near-perfect, consistent 3D segmentations from 2D slice-by-slice segmentation from orthogonal views. In the best case, we have perfect reconstruction. In the worst case, a subset of branching cells will be decomposed into a few standalone segments to be subsequently stitched. At the expense of speed, our results show the optimal distance transform applicable for all morphotypes to be explicit transforms with the 2D object skeleton defined as the attractor. This is why we have additionally implemented these in u-Segment3D (Methods). Using the 3D reconstruction results as the best upper bound of segmentation performance we next assessed the application of u-Segment3D to pretrained 2D segmentation models on the same datasets.

458 **u-Segment3D generates consensus 3D segmentation from orthogonal 2D slice-by-slice instance 459 segmentations of any 2D method**

460 2D segmentation models either (i) already predict a suitable distance transform or the 2D gradients directly
461 as an output e.g. Cellpose⁶, or (ii) provides the 2D segmentation. u-Segment3D accounts for both cases (Fig.
462 3a). In the former, predicted 2D gradients can be used to directly generate the 3D segmentation (the direct
463 method). In the latter, a chosen 2D distance transform is used to compute the 2D gradients from the 2D
464 segmentations, (the indirect method). We demonstrate the pros and cons of both methods using pretrained
465 Cellpose models. Unlike with our ideal 2D segmentations, now the reconstructed 3D foreground binary plays
466 an additional crucial factor in performance. If the foreground does not provide a contiguous path for gradient
467 descent, the resulting segmentation will be fragmented, even with correct gradients. For pretrained Cellpose

468 2D models we empirically found two model parameters to crucially determine performance: (i) the diameter,
469 which determines what-sized objects are segmented and (ii) the cell probability threshold used to determine
470 foreground.

471 Cellpose 2D models already enable ‘optimal’ diameter prediction based on a pretrained regression model.
472 However, this assumes one size fits all. An image can contain objects of different scales we wish to segment
473 e.g. cell body vs cell nuclei, cells within an embryo vs the embryo shape. Moreover, a trained model is not
474 guaranteed to generalize on out-of-sample datasets or be consistent across sequential 2D slices. When we
475 examined Cellpose predicted cell probability and gradients on cross-sections LRP, we found seemingly
476 similar results over a broad diameter range (Extended Data Fig.9a-c). To set diameter objectively without
477 training, we developed an automatic tuning method based on examining the model’s self-confidence. Our
478 method runs Cellpose over a test diameter range to compute a ‘contrast score’ per diameter using the
479 predicted gradients and cell probability using local pixel variance (Extended Data Fig. 10a, Methods). The
480 resulting function uncovers all salient object scales as local maxima. Our contrast function serves as a tuning
481 guide. Cellpos models are trained using a mean diameter of 30 pixels and documented to perform best for
482 diameter=15-45 pixels. Based on the peak of the contrast function images are resized accordingly prior to
483 Cellpose input. In batch operation, u-Segment3D automatically selects the optimal diameter as that which
484 has maximum contrast. If multiple peaks are present, we can bias the selected optimal parameter to favor
485 other maxima by adjusting the size of the considered pixel neighborhood used to compute contrast score
486 (Extended Data Fig. 10) or by constraining the diameter range. In 3D, the cross-sectional appearance of an
487 object can have different aspect ratios and size, even if the image is resized to be isotropic voxels. Therefore
488 we apply our tuning to set the optimal diameter in each of xy, xz, and yz views using a representative 2D
489 slice (Methods). As validation, the Cellpose predicted diameter matches the predicted maxima of our method
490 (Extended Data Fig. 9a-c). Moreover, the direct method 3D segmentation using our method ($AP_{0.5}=0.28$) is
491 comparable vs using Cellpose’s method ($AP_{0.5}=0.23$), if not better (Extended Data Fig. 9d).

492 For thresholding cell probability, Cellpose does not provide automated means. u-Segment3D uses multi-
493 threshold Otsu to statistically determine a finite number of thresholds (Methods). We can then use flooring to
494 round thresholds to the nearest decimal point, or choose a lower threshold to strike a balance between
495 segmentation accuracy and ensuring contiguous space for gradient descent. This works excellently for both
496 2D Cellpose and for 3D reconstructed cell probabilities. Given the problems we found with Cellpose’s gradient
497 descent (Extended Data Fig. 3), and spatial clustering (Extended Data Fig. 4), henceforth we always use u-
498 Segment3D’s equivalent to generate segmentations. Thus Cellpose 2D refers only to predicted gradients and
499 cell probability outputs. Cellpose 2D segmentation refers to that after applying u-Segment3D’s statistical
500 binary thresholding, suppressed gradient descent with momentum and connected component analysis to
501 Cellpose 2D outputs.

502 Using our tuning and parsing of Cellpose, with image preprocessing and segmentation postprocessing by
503 size and gradient-consistency, we compared the direct and indirect method of u-Segment3D on 9/11 datasets
504 (see Suppl. Table. 3 for parameter details). We excluded Zebrafish macrophages whose labels derive from
505 u-Segment3D and VesselMNIST3D which only contains binary masks. We also considered two pretrained
506 Cellpose models, ‘cyto’ and ‘cyto2’, both generalist models but have been trained on different datasets to
507 assess if 2D performance translates to better 3D segmentation. To minimize data leakage, we applied models
508 to only the validation or test splits when available.

509 On Ovules val split ($n=2,840$ cells, $m=2$ images, Fig. 3b i), we found excellent performance with both models
510 (cyto: magenta line, square marker, cyto2: purple line, diamond marker) using the direct method (Fig. 3b ii,
511 $AP_{50} \approx 0.80$ for both). This was expected as the cells are convex and image quality is good, with well-defined
512 cell edges. This is also evidenced by good performance of running Cellpose 3D mode on the same
513 preprocessed input. Cellpose 3D does not have automatic diameter tuning and allows only one diameter
514 across all views. To attempt as fair a comparison as possible as an end-user, without modifying the source
515 code, we considered the oversegmentation tendency of Cellpose 3D and used the maximum of u-Segment3D
516 inferred diameters. We also ran Cellpose 3D twice, the first to obtain 3D cell probabilities to compute the
517 equivalent Otsu thresholds and the second to obtain final segmentations (Methods). We also used the same
518 postprocessing parameters. As expected from algorithm design, for the same model, u-Segment3D
519 consistently outperforms Cellpose 3D (cyto: orange line, square marker, cyto2: brown line, diamond marker).
520 Most impressively however is that u-Segment3D even boosted the ‘cyto’ model to be on par with ‘cyto2’ from
521 $AP_{0.5} = 0.7$ to $AP_{0.5} = 0.8$. The same was true for the test split ($n=10,328$ cells, $m=7$ images, Extended Data

522 Fig. 11). Again u-Segment3D boosted ‘cyto’ to be on par with ‘cyto2’ from $AP_{0.5} = 0.65$ to $AP_{0.5} = 0.7$.
523 Compared to the best 3D construction with ideal 2D segmentations (black line, circle marker, ($AP_{50} = 1.0$)
524 however, there is a noticeable gap of 0.2. Interestingly, the indirect method with either the geodesic (magenta
525 colored) or diffusion (cyan colored) distance transforms for both models was better quantitatively than the
526 direct method (Fig. 3b iii, $AP_{50} > 0.80$). This is likely due to better cell boundary delineation from aggregating
527 on the hard-thresholded 2D segmentations. As compromise however total number of cells predicted is
528 decreased (reference=1686, direct=1697, indirect=1529). Lastly, we asked if the 2D slices of the direct 3D
529 segmentation (black lines) still retain good 2D segmentation of the image, by comparing to the native 2D
530 slice-by-slice segmentations in xy, xz, and yz views (magenta lines) aggregated by the indirect method (Fig.
531 3b iv). We found that not only is the 2D segmentation preserved but also consistently improved in xy view for
532 both models ($AP_{0.5} = 0.45$ to 0.55). This demonstrate u-Segment3D exploits complementary information from
533 orthogonal predictions.

534 LRP is much more challenging, containing not only a mixture of both compact and elongated/branching cells,
535 but cell edges are also weakly-defined (Fig. 3c). Unsurprisingly, direct u-Segment3D segmentation with both
536 models on the val split was substantially lower ($AP_{50} \approx 0.30$ for both) than that from ideal 2D segmentations
537 ($AP_{50} \approx 0.90$). This time, u-Segment3D is significantly better than Cellpose 3D ($AP_{50} \approx 0.05$ for both) in both
538 the val and test split (Extended Data Fig. 11, improving Cellpose 3D cyto ($AP_{50} = 0.18$ to 0.37) and cyto2
539 ($AP_{50} = 0.19$ to 0.40)). Interestingly the AP_{50} we measured for u-Segment3D and pretrained Cellpose were
540 on-par reported of a Cellpose 2D (plant-cp) and Omnipose 3D (plant-omni) model trained specifically on LRP
541 by Omnipose⁷. We thus performed a like-for-like evaluation using their pretrained model weights (Methods).
542 Unexpectedly, plant-cp with indirect u-Segment3D and any distance transform performed best for both val
543 ($AP_{0.5}=0.50$, Extended Data Fig. 11c) and test ($AP_{0.5}=0.50-0.56$) splits. Amazingly, 3D trained plant-omni and
544 plant-cp (Cellpose 3D mode) performed only on-par with pretrained cyto2 and direct u-Segment3D in both
545 splits. Close inspection revealed whilst plant-omni looked excellent in 3D, in cross-sectional views, it can be
546 seen its segmentation is not complete, with many missing internal pixels. We also find plant-omni
547 oversegments despite our additional size filtering (Methods). These results highlight the robustness of u-
548 Segment3D and verifies we can translate better 2D models into better 3D segmentations, on-par with natively
549 3D trained models. Again, direct and indirect u-Segment3D segmentations were on-par in AP_{50} , but indirect
550 is better IoU-wise, with a slightly slower drop-off (Fig. 3c iii). Again, u-Segment3D demonstrates the ability to
551 exploit complementary information from orthogonal views. Impressively, by minimally sacrificing yz IoU, it
552 consistently increases both xy and xz performance for both models (Fig. 3c iv).

553 DeepVesselNet, comprised of thin, complex vasculature networks represents the largest challenge for 2D-
554 to-3D segmentation, (Fig. 3d). During application we found both Cellpose models predict segmentations
555 uniformly larger than the actual vessel radii in 2D slices. Hence we additionally uniformly eroded aggregated
556 3D segmentations to obtain the final segmentation (Suppl. Table 3). Nevertheless there was a clear difference
557 between the two models. Using direct segmentation, ‘cyto’ ($AP_{0.5} = 0.5$, IoU drop-off≈0.75) noticeably
558 outperforms ‘cyto2’ ($AP_{0.5} = 0.5$, IoU drop-off≈0.75) (Fig. 3d ii). Without suppressed gradient descent,
559 Cellpose 3D grossly oversegments ($AP_{0.5} = 0$). Again, direct and indirect u-Segment3D segmentations were
560 on-par in AP_{50} . However the indirect method is far superior in IoU, with drop-off extending to 0.95 with similar
561 AP curves across all distance transforms (Fig. 3d iii). Again, comparing 2D segmentation performance, the
562 direct 3D aggregated cyto outperforms individual 2D segmentations in $AP_{0.5}$ but exhibits faster IoU drop-off
563 (Fig. 3d iv). The direct aggregated cyto2 was significantly worse than its 2D counterpart. This is likely due to
564 the 3D erosion postprocessing removing too many small 2D segmentations in slices. Since, the background
565 appeared homogeneous in this dataset we additionally tested 2D binary Otsu thresholding. This yielded the
566 highest $AP_{0.5}$ 2D segmentations in all orthogonal views (Fig. 3d iv, green lines). Applying u-Segment3D, we
567 consequently also achieve the highest $AP_{0.5}$ 3D segmentation (Fig. 3d iii, green line).

568 Altogether, these three datasets, representing the three super-morphotypes of convex, irregular/branched
569 and vessel-like, demonstrate the robust implementation and applicability of u-Segment3D to real datasets.
570 Moreover u-Segment3D can be applied to any 2D segmentation method using the direct or indirect methods
571 with similar $AP_{0.5}$ performance. Thus we applied only the direct method on remainder datasets (Extended
572 Data Fig. 12). Except for Arabidopsis (CAM) (best $AP_{0.5} = 0.4$), whose image quality was similar to LRP and
573 densely packed, all others had $AP_{0.5} \geq 0.6$. For mouse organoids, pretrained cyto2 with u-Segment3D
574 ($AP_{0.5}=0.93$) nearly matched the ideal 2D segmentation ($AP_{0.5}=1.0$).

575 In summary, u-Segment3D reliable 3D segmentation from 2D. Crucially the results are consistent with
576 theoretical expectations. The better the 2D segmentation, the better the resulting 3D segmentation. However
577 the real potential of u-Segment3D lies in its flexibility to enable segmentation of diverse unseen and out-of-
578 distribution datasets, which do not have reference segmentations or when annotation may be ambiguous, or
579 time-consuming to acquire.

580

581 ***u-Segment3D reconstructs consensus 3D segmentation from 2D slice-by-slice instance 582 segmentations from one orthogonal view for anisotropic 3D data***

583 Due to the microscope or culture conditions, 3D cell imaging cannot always be acquired isotropically or be
584 interpolated to be near-isotropic later in analysis, such that image quality is similar in xy, xz, and yz. In these
585 cases applying pretrained models like Cellpose, trained on the equivalent of in-focus 'xy' slices, to xz and yz
586 views may yield worse segmentations. If segmenting a timelapse, we may wish to save time. In all cases, u-
587 Segment3D can be applied without all three views. For one view, this is conceptually similar to the stitching
588 across xy slices. Looking top-to-bottom through an epidermal organoid culture⁸⁶ (Methods), cells are initially
589 spherical, becoming increasingly irregular and elongated (Fig. 4a). Even when interpolated to isotropic voxels,
590 cells are still flat, and stretched in appearance (Fig. 4b). Consequently we applied Cellpose 2D to segment
591 only xy slices, using the optimal predicted diameter of each slice. u-Segment3D then aggregated the 2D
592 segmentations into 3D. We compared the 3D segmentation from Cellpose predicted optimal diameters (Fig.
593 4c,d) and from our contrast score diameters (Fig. 4e,f). Qualitatively, both look similar. Without ground-truth,
594 and ambiguity in manual labelling without a nuclear marker, we assessed the segmentation consistency
595 between consecutive xy slices, slice i and slice $i + 1$ with $AP_{0.5}$. This revealed $AP_{0.5}$ variation is correlated
596 with morphology. We find a systematic drop in $AP_{0.5}$ as cell morphology changed from spherical to more
597 elongated. Overall, our contrast score determination appears more stable, with a higher mean $AP_{0.5} = 0.59$.
598 We plotted the predicted mean cell diameter per slice (green line) with the measured cell diameter of the
599 resultant segmentation (black line) for each method (Fig. 4g,h). Whilst Cellpose better predicts the absolute
600 diameter per slice, their correlation across xy slices was only moderate (Pearson's $R = 0.47$). In contrast, u-
601 Segment3D's contrast-score method exhibits strong correlation (Pearson's $R = 0.89$). This consistency likely
602 translated to the improved slice-to-slice $AP_{0.5}$.

603 A second example is a video of MDA231 human breast carcinoma cells embedded in a collagen matrix from
604 the single cell tracking challenge⁸⁷ (Fig. 4j). These cells have small area and thin, protrusive morphologies
605 imaged with a noisy background. The 3D image had only 30 z slices, each cell spanning <5 slices. Again,
606 applying pretrained Cellpose with automatic contrast-score diameter determination on xy slices only, we
607 successfully generated consistent 3D cell segmentation. Visual inspection confirm the same cell is
608 consistently segmented across slices. Applying our strategy to every timepoint, we also observed consistent
609 segmentation of cells across time (Suppl. Movie 6).

610 Our last example is the segmentation of cells on the surface of a developing drosophila embryo from the
611 single-cell tracking challenge⁸⁷. Due to the curved surface, cell dynamics are better visualized using
612 cartographic surface projections¹⁴. Using u-Segment3D we segmented the embryo surface (Methods) and
613 applied u-Unwrap3D¹⁵ to extend the cartographic projection to mapping a surface proximal subvolume (Fig.
614 4k). We then attempted to segment the 3D cells using xy slices only. Amazingly, despite the unequal pixel-
615 wise metric distortion due to unwrapping, u-Segment3D still produced consistent 3D cell segmentations. This
616 enabled us to uniquely visualize the migration (black arrows) of cells toward the ventral midline from the side
617 in relation to cells underneath the embryo surface (Fig. 4k, Suppl. Movie 7).

618

619 ***u-Segment3D can refine 3D segmentations and recover subcellular detail not captured by 620 pretrained models***

621 In 3D, cells demonstrate a rich spectrum of protrusive, subcellular surface morphologies. Biologically, these
622 protrusions can be classified into recurring morphological motifs⁷⁵ such as blebs, lamellipodia, filopodia, and
623 villi. These motifs are integral to cell function. Microvilli are used by T-cells to efficiently scan target cells for
624 antigen⁸⁸. Blebs and lamellipodia are key for 3D migration^{89,90} but have also recently been found to drive
625 proliferation and survival in cancer cells^{91,92}. These protrusions are incompletely represented in existing 2D

626 cell training datasets. Importantly, neural networks exhibit spectral bias⁹³, a phenomenon of learning low-
627 frequency modes faster and more robustly than high-frequency modes. This means that they can learn strong
628 shape priors to predict a complete shape from hollow, partial or noisy versions. However, the recovered
629 shape lacks intricate details such as protrusions. Rectifying this bias requires revising the architecture and
630 additional training on fine-grained higher quality masks⁹⁴.

631 Active contour⁵⁵ refinement is slow, struggle with concavities⁹⁵, and cannot handle multiple cells. Random
632 walker approaches are faster⁹⁶ but requires careful design of image forces, fine-tuning and iterations for
633 convergence.

634 u-Segment3D proposes a two-stage solution to be applied after filtering out implausible cells by size and
635 gradient consistency (Fig. 5a, i-iii). The first stage (Fig. 5a, iv) is label diffusion based on semi-supervised
636 learning⁹⁷ to smooth and improve adherence to the cell boundaries within a guide image, whilst enforcing
637 spatial connectivity. Each cell in the input segmentation are unique ‘sources’. Each voxel simultaneously
638 diffuses their ‘source’ to neighbor voxels for T iterations based on an affinity graph combining the local
639 intensity differences in the guide image, and spatial proximity (Methods). The final segmentation is generated
640 by assigning each voxel to the source with highest contribution (Methods). We can control the extent diffusion
641 refines the input segmentation using a ‘clamping’ factor such that if ‘clamped’, diffusion can only modify voxels
642 assigned to background only. We observe improved boundary matching for $T < 50$ iterations. The guide image
643 can be the intensity normalized raw image or any image enhancing the desired features to capture in the
644 segmentation. The second stage uses a guide image to transfer all intricate details in the local spatial
645 neighborhood around the diffusion-refined cell in one pass using linear-time guided filtering⁹⁸ (Methods).
646 Conceptually, this filter is analogous to an interpolation between the binary cell mask and the intensities in
647 the corresponding spatial region of the guide image. The neighborhood size may be fixed for all cells or set
648 as a proportion of cell diameter. For guided filtering segmentations, we find a good image is $G = \alpha \cdot I_{norm} +$
649 $(1 - \alpha)I_{ridge}$, a weighted sum of the normalized input image, I_{norm} and its ridge filter-enhanced counterpart,
650 I_{ridge} , which exaggerates subcellular protrusions.

651 Applying this workflow, we recovered the majority of missing surface protrusions for cell tightly packed as an
652 aggregate whilst simultaneously retaining the benefits of the shape prior from Cellpose (Fig. 5a, Suppl. Movie
653 8). This meant we should be able to segment individual cells imaged with high-resolution lightsheet
654 microscopy even when membrane staining is inhomogeneous or sparse, situations which challenge
655 thresholding-based techniques⁷⁵. We tested this on single cells with different morphological motifs. Knowing
656 there is only one cell, we directly threshold the 3D reconstructed cell probability (Fig. 5b-d i). The result
657 captures well the global morphology but cell protrusions only approximately (ii). After guided filtering, all
658 protrusions are recovered (iii), with comparable fidelity to that of binary thresholding (iv). However, the
659 segmentation is now much better suited for surface analysis, as measured by the genus number, g of the
660 extracted surface mesh. The postrefined mesh has consistently lower genus than that of thresholding which
661 were as high as $g = 91$ for filopodia. We can further recover protrusive features on touching cells in a field-
662 of-view as shown for T-cells (Fig. 5e) and zebrafish macrophages (Fig. 5f). Lastly, as a non-cell
663 demonstration, we tested the segmentation of zebrafish vasculature undergoing angiogenesis (Fig. 5g). The
664 combination of using pretrained Cellpose 2D as prior and guided filtering recovered the extensive, thin
665 sprouting vessels, despite the noisy background and inhomogeneous staining (Suppl. Table 4, Suppl. Movie
666 9)).

667

668 ***u-Segment3D can multiprocess to generate consensus 3D segmentations of tissue***

669 3D tissues readily contain 10,000’s of cells even for thin tissue of millimeter-width, sub-50 micrometer thick²⁵.
670 The time for gradient descent increases with iteration number and the number of foreground pixels (related
671 to image size). Postprocessing increases with the number of segmented cells. To allow segmentations to be
672 computed in a reasonable time, we also implemented a multiprocessing variant of 2D-to-3D segmentation in
673 u-Segment3D to take advantage of the wide availability of CPU-based cluster computing (Methods). Fig. 6a
674 illustrates the key steps, (i) the pretrained 2D model runs fast GPU inference^{6,99} on 2D slices from orthogonal
675 views; (ii, iii) gradient descent is applied in parallel to local spatially-overlapped subvolumes to generate
676 global image coordinates. This critically ensures that border cells across subvolumes retain the same global
677 attractor, avoiding the need to post-stitch; (iv) an existing parallelized connected component analysis

678 developed for large connectomics datasets is applied to generate the full image 3D segmentation; and (v)
679 postprocessing of segmentations is applied in parallel to individual segmented cells. The segmentation of a
680 metastatic melanoma CyCIF multiplexed tissue sample using fused nuclear and membrane signals, imaged
681 with an equivalent isotropic voxel size 280 nm resolution (Methods), and size 194 x 5440 x 4792 pixels took
682 \approx 2h for preprocessing and running Cellpose slice-by-slice in xy, xz, yz, \approx 2h to generate the initial 3D
683 segmentation from 250 gradient descent iterations and using subvolume sizes 128 x 256 x 256 with 25%
684 spatial overlap, \approx 1h for size filtering and gradient consistency checking, \approx 2h for label diffusion refinement,
685 a total of 7h to yield the final segmentation with 43,779 cells (Suppl. Movie 10), on a CPU cluster with 32
686 physical cores, 72 threads, 1.5TB RAM and a single A100 GPU (40GB). Notably, the gradient descent alone
687 would be $>$ 20x slower. Importantly, we obtained segmentations with no stitching artifacts and agreed well
688 with the fused cell nuclei and membrane markers when we visualize zoom-ins of the mid-slices from each of
689 the three orthogonal views (Fig. 6a). Functionally, these segmentations enabled us to improve the accuracy
690 of 3D cell phenotyping and to show mature and precursor T cells in metastatic melanoma engage in an
691 unexpectedly diverse array of juxtracrine and membrane-membrane interactions²⁵. Further, the extraction of
692 3D cell morphological features such as sphericity (Fig. 6a, region α), enabled us to reveal looser
693 “neighbourhood” associations¹⁰⁰ whose morphologies reveal functional states.

694 Axially-swept lightsheet microscopy¹⁰¹ can image thick cleared tissue volumes at subcellular resolution over
695 thick sections up to 2mm. This enabled us to visualize single cells within micrometastases in lung tissue.
696 Unexpectedly, despite the weak fluorescence of the injected cancer cells and absence of membrane markers
697 for cells, we could still segment all salient nuclei and micrometastases (Fig. 6c, Suppl. Movie 11). Indeed,
698 our initial micrometastases 3D segmentation combining both channels contained many extraneous, spurious
699 segmented cells, presumably because Cellpose attempts to also infer the shape of non-cancer cells
700 (Extended Data Fig. 13a). Thanks to the consensus segmentation of u-Segment3D, we could nevertheless
701 use the segmentations to measure the mean fluorescence intensity to identify well-segmented micrometastases
702 (Extended Data Fig. 13b).

703 Our final application was to segment cellular structures in brain tissue acquired using a recently developed
704 technique, CATS⁷⁶ to label the extracellular compartment, and STED microscopy with tissue expansion. The
705 result reproduces the detail within electron microscopy images whilst preserving the 3D tissue (Fig. 6d).
706 However it is difficult to visualize any 3D structure. We applied Cellpose and u-Segment3D with the aim of
707 an exploratory tool that ‘scans’ the volume to generate consensus 3D segmentations of the larger
708 extracellular space. However these spaces are heterogeneous, different in size, and morphotype. This is
709 challenging for Cellpose which is biased towards generating segmentations of the same scale. Applying u-
710 Segment3D as previously was observed to fragment the thick, dominant, branching dendrites (data not
711 shown). It is only thanks to the ability to fine-tune every step of the 2D-to-3D segmentation process in u-
712 Segment3D that we could overwrite this bias and largely preserve the multiscale tissue architecture in the
713 final segmentation, in 3D and in 2D cross-sections (Fig. 6d-f, Suppl. Movie 12). In particular, we fine-tuned
714 the content-based averaging, the filtering σ to smooth 3D gradients and that used for connected component
715 analysis.

716 Discussion

717 Here we have presented a formalism and general algorithm based on distance transforms and gradient
718 descent to generate optimal, consensus 3D segmentations from 2D segmented volume stacks. Our
719 formalism unifies existing works in 2D-to-3D segmentation and shows near-perfect segmentations are
720 achievable. Conceptually our work reformulates the widespread ad hoc procedure of stitching discrete label
721 segmentations into a continuous domain problem with controllable and easy fine-tuning. Meanwhile our work
722 shows that the initial 2D-to-3D segmentation proposal of Cellpose, based on a point-source centroid distance
723 transform is only a specific instance of a broader class of medial-axis distance transforms.

724 This led us to develop a general toolbox, u-Segment3D to robustly implement 2D-to-3D segmentation in
725 practice for any 2D segmentation method. Through extensive validation on public datasets, we showed u-
726 Segment3D consistently translates the performance of 2D models to 3D segmentation. The better the 2D
727 model, the better the 3D segmentation. Moreover u-Segment3D provides fine-tuning and postprocessing
728 method for further improving 3D segmentations. We also implemented multiprocessing to enable scalable
729

730 2D-to-3D segmentation on CPU clusters. Further speed improvements could be made such as implementing
731 a multi-scale scheme to run u-Segment3D, which we leave for future work.

732 With the successes of foundation models such as ChatGPT for natural language processing and Segment
733 Anything model in segmentation, there is a prevalent notion that everything should be learnt from data, that
734 more data is better and models should be ‘turnkey’, working directly out-of-the-box or if not, be ‘fine-tuned’.
735 In the quest for generality we must not neglect the value of grounded formalism and robust design. Our
736 analyses provide multiple cautionary tales. First, parsing the outputs of neural network models is just as
737 important as training. By identifying and rectifying the spatial proximity clustering of Cellpose, we significantly
738 reduced over-segmentation and boosted performance on noisy datasets. Second, considering extreme
739 morphotypes and the simpler 1D-to-2D segmentation problem showed the critical importance of smoothing
740 reconstructed gradients and the role of suppressed gradient descent to enable 2D-to-3D segmentation to be
741 applied to branched and vasculature networks. Third, running Cellpose with optimal diameters in different
742 views is necessary to capture general 3D shape, irrespective of voxel anisotropy. Because Cellpose was
743 trained using a fixed size diameter, we could exploit Cellpose and its strong cell shape prior to ‘scan’ and
744 infer all salient diameters of objects in the image. This in-turn enabled us to set optimal diameters in
745 orthogonal views for 3D segmentation training-free. Lastly, by recognizing the spectral bias of neural
746 networks and annotation bias, we developed simple label diffusion and guided filter postprocessing to recover
747 intricate surface morphologies of 3D cells. This enabled us to extend neural network methods to segment
748 high-resolution single cells where classical methods are still state-of-the-art.

749 In sum, our experiments question the proposition value of directly training 3D segmentation models. Using
750 only pretrained Cellpose models equipped with automated parameter tuning, here we demonstrate an
751 unprecedented capacity to 3D segment cells from diverse microenvironments, from single cells through to
752 entire tissues, in-vitro, in-vivo and in-situ, and acquired from different modalities and with different resolutions.
753 With widespread availability of diverse generalist and specialized 2D segmentation models, u-Segment3D
754 paves a way towards accessible 3D segmentation, translating time-consuming annotation and training
755 towards more impactful time spent on analyzing the acquired 3D datasets to provide biological insights.

756 757 **Author Contributions**

758 Conception: AJ, FYZ, GD (using pretrained 2D for 3D segmentation), FYZ (u-Segment3D); Investigation and
759 Analysis: FYZ, CY, ZS. Data generation: CY (CyCIF tissue), SD (Zebrafish macrophages), ZM, MTI, KD, SM
760 (lung micrometases), BN (epidermal organoid), EJ, MD (T cell co-culture), GMG, BJC (COR-L23 single cell
761 with ruffles), AW, KD (Septin cleared tissue), RF (HBEC cell aggregate); Supervision: GD; Funding
762 acquisition: GD, PKS, KD; Writing – Original Draft: FYZ, GD; Writing – Review and Editing: all authors.

763 764 **Acknowledgments**

765 Funding for this work in the Danuser lab was provided by the grants R35 GM136428 (NIH), and
766 U54CA268072 (NIH). GMG is a HHMI Hanna H. Gray Fellow. EJ was funded by Wellcome Trust (grant
767 #224040/Z/21/Z). We thank Qiongjing Zou for putting the Python library into a continuous integration coding
768 framework. Imaging was performed using the Oxford-Zeiss Centre of Excellence in Biomedical Imaging and
769 the Kennedy Trust for Rheumatology Research (Grant #202117 and 202103, respectively). The authors
770 would like to acknowledge the Quantitative Light Microscopy Core, a Shared Resource of the Harold C.
771 Simmons Cancer Center, supported in part by an NCI Cancer Center Support Grant, 1P30 CA142543-01.
772 This research was supported in part by the computational resources provided by the BioHPC supercomputing
773 facility located in the Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, TX. URL:
774 <https://portal.biohpc.swmed.edu>. The research within this work complies with all relevant ethical regulations
775 as reviewed and approved by the University of Texas Southwestern Medical Center. Zebrafish husbandry
776 and experiments described here have been approved and conducted under the oversight of the Institutional
777 Animal Care and Use Committee (IACUC) at UT Southwestern under protocol number 101805.

778 779 **Data Availability**

780 All data used in this study except for the imaging of COR-L23 ruffles, epithelial organoid, zebrafish vasculature, T cell
781 coculture and HBEC cell aggregate are publically available from their original sources as documented in Suppl. Table
782 1 and in the Dataset section of Methods. All others will be made available on request to the corresponding author.
783
784 **Code Availability**
785 u-Segment3D will be available at <https://github.com/DanuserLab/u-Segment3D>. This Python library will also include our
786 code to automate parameter tuning of Cellpose models.
787
788
789
790
791
792

793 Methods

794 Datasets

796 Validation Datasets

797 10 independent public datasets with reference 3D segmentations and 1 dataset collected in-house with 3D
798 segmentation constructed with the aid of u-Segment3D were used to evaluate the ability of u-Segment3D to
799 reconstruct 3D segmentation from its ideal slice-by-slice 2D segmentations (Fig. 2) from orthogonal xy, xz, yz
800 views. 9 of the public datasets with both images and reference segmentations were used to assess the
801 performance of u-Segment3D with pretrained Cellpose models (Fig. 3). Details of all datasets are given in
802 Suppl. Table 1.

803

804 Demonstration Datasets

805 The following datasets were collected largely in-house and 3D segmented using either the pretrained
806 Cellpose cyto, cyto2 or nuclei 2D segmentation models, whichever qualitatively appeared to perform best
807 with u-Segment3D (if not stated otherwise). Parameter details are provided in Suppl. Table 4.

808

809 3D epidermal organoid culture (Fig. 4a-i)

810 *Cell culture.* Human keratinocyte Ker-CT cells (ATCC #CRL-4048) were a kind gift from Dr. Jerry Shay (UT
811 Southwestern Medical Center). A keratinocyte cell line stably expressing mNeonGreen-tagged keratin 5 (K5) to label
812 intermediate filaments was created as previously described⁸⁶.

813 *Epidermal organoid culture.* We adapted an epidermal organoid culture model from existing protocols^{86,101,102}.
814 Polycarbonate filters with 0.4μm pore size (0.47cm² area, Nunc #140620) were placed in larger tissue culture dishes
815 using sterile forceps. On day 0, a cell suspension of 5x10⁵ keratinocytes expressing K5-mNeonGreen in 400μL of K-
816 SFM was added to each filter well, and additional K-SFM was added to the culture dish to reach the level of the filter.
817 On day 10, culture medium was aspirated from above the filter to place the cultures at an air-liquid interface. At the
818 same time, medium in the culture dish was changed from K-SFM to differentiation medium (see Cell Culture above). On
819 day 13, an additional 0.5 mM calcium chloride was added to the differentiation medium in the culture dish. Mature
820 epidermal organoids were processed for imaging on day 20, after 10 days of differentiation at the air-liquid interface.
821 Throughout the procedure, culture media were refreshed every two to three days.

822 *Epidermal organoid imaging.* Mature epidermal organoids were transferred to a clean dish, washed three times with
823 PBS, then fixed in 4% paraformaldehyde (Electron Microscopy Sciences #15713) for 1 hour at room temperature. Filters
824 with the organoids were cut out of the plastic housing using an 8 mm punch biopsy tool and inverted onto glass-bottom
825 plates. Throughout imaging, PBS was added one drop at a time as needed to keep each organoid damp without flooding
826 the dish. Organoids were imaged using a Zeiss LSM880 inverted laser scanning confocal microscope equipped with a
827 tunable near-infrared laser for multiphoton excitation and a non-descanned detector optimized for deep tissue imaging.
828 Images were acquired using an Achroplan 40x/0.8NA water-immersion objective resulting in an effective planar pixel
829 size of 0.21 μm, and z-stack volumes with 1 μm step size.

830

831 Single cell tracking challenge datasets^{87,103} (Fig. 4j,k)

832 *MDA231 human breast carcinoma cells (Fluo-C3DL-MDA231) (Fig. 4j).* Cells infected with a pMSCV vector
833 including the GFP sequence, embedded in a collagen matrix captured with an Olympus FluoView F1000
834 microscope with Plan 20x/7 objective lens, sampling rate of 80 min and voxel size 6.0 x 1.242 x 1.242 μm.

835

836 *Drosophila Melanogaster embryo (Fluo-N3DL-DRO) (Fig. 4k).* Developing embryo imaged on a SIMView
837 light-sheet microscope¹⁰⁴ with a sampling rate of 30s, 16x/0.8 (water) objective lens and voxel size 2.03 x
838 0.406 x 0.406 μm. We used Cell01 from the test dataset containing 50 timepoints. We used the pretrained
839 Cellpose cyto model and u-Segment3D to segment the surface for each timepoint (Suppl. Table 3). Using
840 the binary segmentation, we unwrapped a proximal surface depth using u-Unwrap3D¹⁵.

841

842 *Human bronchial epithelial (HBEC) cell aggregate (Fig. 5a)*

843 Transformed HBEC cells expressing eGFP-Kras^{V12} were cultured and imaged using meSPIM and
844 published previously¹⁰⁵.

845

846 *Single dendritic cell with lamellipodia (Fig. 5b)*

847 Conditionally immortalized hematopoietic precursors to dendritic cells expressing Lifeact-GFP were
848 cultured, imaged and published previously⁷⁵.

849

850 *Single HBEC cell with filopodia (Fig. 5c)*

851 HBEC immortalized with Cdk4 and hTERT expression and transformed with p53 knockdown, Kras^{V12} and
852 cMyc expression cultured, imaged and published previously⁷⁵.

853

854 *Single COR-L23 cell with ruffles (Fig. 5d)*

855 *Culture.* COR-L23 cells (Human Caucasian lung large cell carcinoma) were resuspended in 2mg/mL bovine
856 collagen (Advanced BioMatrix 5005) and incubated for 48 hours in RPMI 1640 medium (Gibco 11875093)
857 supplemented with 10% fetal bovine serum (PEAK SERUM PS-FB2) and 1% antibiotic-antimycotic (Gibco
858 14240062). Cells were incubated in a humidified incubator at 37°C and 5% carbon dioxide.

859

860 *Imaging.* Images were acquired with our home-built microscope system that generates equivalents to
861 dithered lattice light-sheet through field synthesis¹⁰⁶. Briefly, the system employs a 25X NA 1.1 water
862 immersion objective (Nikon, CFI75 Apo, MRD77220) for detection, and a 28.6X NA 0.7 water immersion
863 (Special Optics 54-10-7) for illumination. With a 500 mm tube lens, the voxel size of the raw data is 0.104 μm
864 × 0.104 μm × 0.300 μm. The volumetric imaging was performed by scanning the sample along the detection
865 axis. We used a Gaussian light-sheet and optimized the light-sheet properties so the confocal length was
866 enough to cover the cell size without sacrificing too much the axial resolution¹⁰⁷. Typically, the light-sheets
867 are about 20 μm long and 1 μm thick. In each acquisition, we optimized the laser power and the exposure
868 time to achieve fast acquisition without introducing too much photo-bleaching. Usually, the time interval for
869 our volumetric acquisition is chosen to be either 5 or 10 s.

870

871 *T-cell coculture (Fig. 5e)*

872 Blasted human CD8⁺ T cells, were produced by activating naïve T cells isolated from PBMCs using anti-
873 CD3/CD28 Dynabeads for 2 days, then rested for 5 days after removing the beads. Cells were frozen on day
874 5 of resting and thawed 48 hours before use. T cells were grown in complete RPMI-1640 (10% FCS, 1%
875 Pen/Strep, 1% Glutamine, 1% HEPES) + 50 U/ml of IL-2. For migration-based imaging of multiple T cells, 8-
876 well glass-bottom IBIDI chambers were coated with 1μg/ml of hICAM-1-6xHis linker and hCXCL11
877 (Peprotech) for 1 h at room temperature, washed, then coated with 1% BSA. 0.5x10⁶ blasted human CD8⁺ T
878 cells were labelled with CellMask DeepRed diluted to a 1x working solution in imaging buffer (i.e., colourless
879 RPMI with 1% added Pen/Strep, 1% glutamine, 1% HEPES) for 30 minutes at 37°C. Cells and the glass
880 slides were washed and resuspended in pre-warmed (37°C) imaging buffer. 0.1-0.2x10⁶ Cells were gently
881 added to the coated glass slides and left to settle for 30-minutes before imaging. Cells were imaged using
882 the Lattice Lightsheet microscope 7 (LLSM7) from Zeiss using the 641nm laser at 4% power with 4ms of
883 exposure. A large field of view was used for imaging multiple cells at once, with a complete volume taken
every second. Deconvolution was performed using the Zeiss software.

884

885 *Zebrafish macrophages (Fig. 5f)*

886 Zebrafish larvae with fluorescent macrophages, labelled with Tg(mpeg1:EGFP) were imaged as published
887 previously⁸³ (Suppl. Table 1).

888
889 *Zebrafish vasculature (Fig. 5g)*

890 Zebrafish (*Danio rerio*) embryos, larvae and adults were kept at 28.5°C and were handled according to
891 established protocols^{108,109}. All zebrafish experiments were performed at the larval stage and therefore the
892 sex of the organism was not yet determined. To visualize the growing vasculature at around 34 h post
893 fertilization (hpf), zebrafish larvae expressing the vascular marker *Tg(kdrl:Hsa.HRAS-mCherry)*¹¹⁰ in a casper
894 background¹¹¹ were used. To immobilize the zebrafish larvae for imaging, they were anesthetized with 200
895 mg/l Tricaine (Sigma Aldrich, E10521)¹¹² and mounted in 0.1% low melting agarose (Sigma Aldrich, A9414)
896 inside fluorinated ethylene propylene (FEP) tubes (Pro Liquid GmbH, Art: 2001048_E; inner diameter 0.8
897 mm; outer diameter 1.2 mm), coated with 3% methyl cellulose (Sigma Aldrich, M0387)¹¹³. The mounted
898 zebrafish larvae were imaged on a custom multi-scale light-sheet microscope with axially-swept light-sheet
899 microscopy⁸³.

900
901 Multiplexed CyCIF tissue

902 A primary melanoma sample from the archives of the Department of Pathology at Brigham and Women's
903 Hospital was selected. The protocol was adapted from Nirmal et al.¹¹⁴. Briefly, a fresh 35-micron thick FFPE
904 tissue section was obtained from the block and de-paraffinized using a Leica Bond. The region in Fig. 6b was
905 selected and annotated from a serial H&E section by board-certified pathologists as a vertical growth phase.
906 The 35 µm thick section underwent 18 rounds of cyclical immunofluorescence (CyCIF)¹¹⁵ over a region
907 spanning 1.4 mm by 1.4 mm and sampled at 140 nm laterally and 280 nm axially. Image acquisition was
908 conducted on a Zeiss LSM980 Airyscan 2 with a 40x/1.3NA oil immersion lens yielding a 53-plex 3D dataset²⁵.
909 A custom MATLAB script was used to register subsequent cycles to the first cycle, which was stitched in ZEN
910 3.9 (Zeiss). The quality of image registration was assessed with Hoechst across multiple cycles in Imaris
911 (Bitplane). For segmentation, multiple channel markers were combined to create fused nuclei and
912 cytoplasmic channels. Hoechst and lamin B1 were combined for nuclei. MHC-II, CD31, and CD3E were
913 combined as a cytoplasm marker to cover all cells including tumor, blood vessels, and T cells.

914
915 Cleared tissue lung micrometastases

916 *Cancer growth.* Lung tissue containing a metastatic tumor was provided by the Morrison lab at UT
917 Southwestern Medical Center, USA. Mice were injected with Y1.7-GFP-luciferase cells¹¹⁶ and grown as
918 previously described¹¹⁷.

919 *Lung tissue staining and clearing.* Lung tissue was fixed in 4% PFA at 4°C for less than 24 h and then washed
920 three times with PBS with 0.02% sodium azide for 2 h per wash. The tissue was sliced into 2 mm thick
921 sections. Tissue slices (~2mm) were permeabilized and blocked in blocking buffer (0.5% NP40, 10% DMSO,
922 0.5% Triton X-100, 5% donkey serum, 1X PBS) overnight at room temperature (RT). Primary and secondary
923 antibody stock solutions were prepared in the desired concentration. Antibodies were centrifuged (MyFuge
924 mini centrifuge) for 2 s before aliquoting in blocking buffer. Antibodies stock solutions were homogenized for
925 at least 1 h in a shaker at 4°C before staining. Tissues were incubated in anti-GFP (1:100) for 72 h at room
926 temperature in a tube revolver rotator. After incubation, samples were washed with wash buffer (0.5% NP40,
927 10% DMSO, 1X PBS) three times for 2 h each and then left rotating in wash buffer overnight. Tissues were
928 immersed in the secondary antibody AF488 (1:250) solution for 72 h at RT. Then, secondary antibody was
929 removed with wash buffer for at least two days changing the solution: first day 3x every 2 h and on the second
930 day refreshed just one time. Finally, tissues were stained for nuclei with TO-PRO-3 647 (2 drops/mL) in PBS
931 for 24 h at room temperature. Nuclear dye was washed out with wash buffer three times for 2 h each and
932 then left rotating in wash buffer overnight. Samples were washed two times for 5 min each in PBS to remove
933 the nuclear dye. Lung tissue was cleared using a modified iDISCO+ protocol. Lungs were dehydrated in a
934 methanol gradient (25%/50%/75%/100%). Final clearing was achieved with repeated fresh Benzyl Alcohol
935 and Benzyl Benzoate (BABB, 1:2) with 5g of activated aluminium oxide incubations. The samples were

washed with BABB 1:2 3x, then left standing in fresh BABB 1:2 for 15 min. Sample BABB 1:2 was refreshed and left overnight. The sample BABB 1:2 was refreshed again shortly before imaging.

Lung tissue imaging. Lung tissue slices were imaged on a ctASLMv2¹¹⁸ microscope chamber controlled by navigate¹¹⁹. Nuclei were imaged using the TO-PRO-3 647 via illumination with a LuxX 642 nm, 140 mW at 100% laser power and a Semrock BLP01-647R-25 filter in the detection path. Cancer cells were imaged via illumination with a LuxX 488-150, 150 mW at 100% laser power and a Semrock FF01-515/30-32 bandpass filter in the detection path. Images were acquired with a Hamamatsu ORCA-Flash 4.0 v3 with 200 ms integration time in lightsheet readout mode.

coCATS labelled volume

We used a coCATS⁷⁶ imaging volume recorded with z-STED at near-isotropic resolution in neuropil of an organotypic hippocampal brain slice published in Michalska et al.⁷⁶ (c.f. Fig. 3). This volume was downloaded already denoised with Noise2Void.

UMAP to map morphological diversity of different cell datasets

Morphological features. Eight features were extracted for each cell based on their 3D reference segmentations.

1. *Volume* - the total number of voxels occupied by the segmented volume, calculated by binary indexing.
2. *Convexity* - the ratio of total volume to total volume occupied by the convex hull. Convex hull was computed with Python Scipy, `scipy.spatial.ConvexHull` on the 3D coordinates of the segmentation volume.
3. *Major length* - The longest axial length of an ellipse fitted to the cell. Computed with Python Scikit-Image, `skimage.measure.regionprops`
4. *Minor length* - computed from Python Scikit-Image, `skimage.measure.regionprops`
5. *1 – minor length/major length* - Measure of the extent of elongation with value 0-1. When spherical, minor length = major length and the measure is 0. When very elongated, minor length << major length and the measure is 1.
6. *# skeleton segments* - Number of straight line segments the 3D binary skeleton is partitioned.
7. *# skeleton nodes* - Number of branch point nodes, where a node is defined as at least three line segments meeting at a point.
8. *Mean skeleton segment length* - Mean number of voxels in each straight line segment of the 3D binary skeleton.

The 3D binary skeleton was computed using Python Scikit-Image, `skimage.morphology.skeletonize`. The decomposition of the skeleton into nodes and segments was computed using the Python sknw library (<https://github.com/Image-Py/sknw>). Non-dimensionless measurements such as volume were not converted to metric units as only the number of raw voxels is relevant for segmentation.

UMAP parameters. The 8 features were power transformed to be more Gaussian-like using the Yeo-Johnson method¹²⁰ (Python Scikit-learn, `sklearn.preprocessing.power_transform`). Then z-score normalization was applied to create normalized features. Uniform Manifold Approximation and Projection (UMAP) (using the Python umap-learn library) was used to project the 8 features after normalization to 2 dimensions for visualization (`n_neighbors=15`, `random_state=0`, `spread=1`, `metric='Euclidean'`). The median UMAP coordinate for each dataset was computed by taking the median of the 2D UMAP coordinates of individual cells comprising the respective dataset. The heatmap coloring of the UMAP uses the normalized feature value and the coolwarm colorscheme clipping values to be in the range [-2,2].

983 u-Segment3D is a toolbox that aims to provide methods that require no further training to aggregate 2D slice-
984 by-slice segmentations into consensus 3D segmentations. It is provided as a Python library
985 (<https://github.com/DanuserLab/u-unwrap3D>). The methods within are broadly categorized into modules
986 based on their purpose; module 1: image preprocessing; module 2: general 2D-to-3D aggregation using
987 suppressed gradient descent with choice of different 3D distance transforms; and module 3: postprocessing
988 to improve the concordance of segmentation to that of a guide image. Postprocessing helps achieve a tighter
989 segmentation and recover missing local high-frequency surface protrusions.

990

991 **Module 1: Preprocessing**

992 Described below are the image preprocessing functions included in u-Segment3D to combat the primary
993 problems of intensity normalization, image feature enhancement and uneven illumination that can greatly
994 affect pretrained segmentation models, like Cellpose. Generally, the order of operation or the
995 inclusion/exclusion of a step is dependent on the input data. We have found the basic workflow of i) rescaling
996 to isotropic voxels and resizing for the desired segmentation scale, ii) uneven illumination correction, adaptive
997 histogram equalization or gamma correction, iii) deconvolution, and iv) intensity normalization applied to the
998 3D raw image, to be a good starting point for Cellpose models. For Omnipose⁷ models we only use intensity
999 normalization. Any other preprocessing led to worse performance. When nuclei and cytoplasm channels
000 present, we find Cellpose cell segmentation was better if both channels are jointly used as input.

001

002 *Rescaling to isotropic voxels and resizing to the desired segmentation scale.* Pretrained segmentation models
003 work best when input images contain objects types and object sizes reflective of the original training dataset.
004 If images are upscaled to be bigger, segmentation models may be biased towards segmenting physically
005 smaller objects. Correspondingly if images are downscaled to be smaller, larger objects become enhanced
006 and easier to segment as smaller objects become oversmoothed. Cellpose models are trained at a fixed
007 diameter of 30 pixels and with isotropic xy images. We find empirically, the u-Segment3D tuning performs
008 best for each orthoview if the input image volume is first rescaled to isotropic voxels and resized using linear
009 interpolation so the desired feature to segment such as cell / vessel results in a peak around 30 pixels (c.f.
010 Extended Data Fig. 9). The rescale and resize is implemented as one function using Python Scipy,
011 `scipy.ndimage.zoom` function with a Python Dask tiled accelerated variant for large volumes.

012

013 *Contrast enhancing intensity normalization.* Image intensities are normalized such that 0 is set to the p_{lower}
014 percentile and 1 is the p_{upper} percentile of the image intensity. By default, $p_{lower} = 2$ and $p_{upper} = 99.8$. This
015 contrast enhances the image by clipping out sporadic high intensities caused by camera shot noise and
016 zeroing small, but non-zero background intensities common to fluorescent microscopy.

017

018 *Image deconvolution.* For 2D fluorescent microscopy images or anisotropic 3D images, we use blind
019 deconvolution with the unsupervised Wiener-Hunt approach¹²¹ (slice-by-slice for 3D) where the
020 hyperparameters are automatically estimated using a Gibbs sampler (implemented using Python Scikit-
021 image, `skimage.restoration.unsupervised_wiener`). The initial point-spread function is specified as a 15x15
022 pixel sum normalized Gaussian ($\sigma = 1$) squared kernel. For 3D lightsheet imaging we use Wiener-Hunt
023 deconvolution, with our previously published experimental PSF⁷⁵ used as a synthetic PSF.

024

025 *Model-free uneven illumination correction.* The raw image intensity of 2D or 3D images, I_{raw}^{ch} is corrected for
026 uneven illumination ratiometrically, $I_{correct}^{ch} = \frac{I_{raw}^{ch}}{\overline{I_{raw}^{ch}}} \frac{I_{raw}^{ch}}{I_{bg}^{ch}}$ where $\overline{I_{raw}^{ch}}$ the mean image intensity of the input image
027 and I_{bg}^{ch} is an estimate of the uneven background illumination. I_{bg}^{ch} is estimated by downsampling the image
028 by a factor of ds , isotropic Gaussian smoothing of σ then resizing back to the dimensions of the input image.
029 For 2D images, the downsampling factor does not need to be used and σ is specified as a fraction of the
030 image dimension, typically 1/4 or 1/8 is a good starting point. For 3D images a default $\sigma = 5$ is used, and a

031 $ds = 8$ or 16 . If segmentation is worse, we decrease ds by factor of 2. If $ds = 1$, Gaussian smoothing is
032 applied at the original image resolution. The resultant enhanced image should have even illumination whilst
033 minimal artefactual enhancement of border background intensities. A more sophisticated background
034 correction is the N4 bias correction available in SimpleITK, originally developed for MRI image and has been
035 successfully applied to 3D cleared-tissue imaging²⁹.

036
037 *Adaptive histogram equalization (AHE)*. Contrast limited AHE or CLAHE (Python Scikit-image,
038 skimage.exposure.equalize_adapthist) can also be used as an alternative to our model-free uneven
039 correction. The image is divided into non-overlapping tiles and the pixel intensity is histogram equalized within
040 each tile. Whilst this obtains good results, we find the method is computationally more memory-intensive and
041 slower for large 3D volumes if the size of individual tiles is required to be small, thus increasing the overall
042 number of tiles. But, there is less artefact for originally low-valued intensities compared to our fast ratiometric
043 method.

044
045 *Gamma correction*. Transforms the input image, I_{in} pixelwise, raising the intensity to a power γ (float between
046 0-1) according so that the output image, $I_{out} = I_{in}^\gamma$ after scaling the image pixel intensity linearly to the range
047 0 to 1. Used to nonlinearly amplify low-intensity pixels to create a more uniform illumination for segmentation
048 that is computationally inexpensive.

049
050 *Vessel-like feature enhancement*: Neurites, tubes, vessels, edges of cell surface protrusions all represent
051 ridge-like structures that are both thin and long or exhibit high curvature and tortuous morphologies that are
052 often weakly stained or visualized by raw image intensities. Ridge image filters uses the eigenvalues of the
053 Hessian matrix of image intensities to enhance these ridge-like structures assuming the intensity changes
054 perpendicular to but not along the structure. Many ridge filters have been developed. u-Segment3D uses the
055 Meijering¹²² filter (Python Scikit-image, skimage.filters.meijering) which enhances ridge image features of by
056 pooling the filter responses over a list of multiple Gaussian σ . We observe empirically good performance for
057 a diverse range of objects including vessels and cells, without requiring any other hyperparameter tuning
058 unlike Frangi filtering¹²³.

059
060 *Semi-automated diameter tuning for pretrained Cellpose models*. The tuning process is illustrated in
061 Extended Data Fig. 10a. Given a 2D image, the cellpose outputs, the non-normalized cell probability, p and
062 predicted 2D gradients in x- ($\nabla_x \Phi$) and y- ($\nabla_y \Phi$) directions are computed. p is clipped to a range of [-88.72,
063 88.72] to avoid overflow for float32 and normalized to a value in the range [0,1], $p \leftarrow \frac{1}{1+e^{-p}}$. These ouputs
064 are used to compute the pixelwise contrast score, $w \cdot \{\sigma_N(\nabla_x \Phi) + \sigma_N(\nabla_y \Phi)\}$. w is a pixelwise weight. We
065 set $w = p$ but observe no difference if $p = 1$ for cellpose models. $\sigma(\cdot)$ is the local standard deviation at each
066 pixel, computed over the local pixel neighborhood of width $P \times P$ pixels. The mean score over all pixels,
067 $\frac{1}{N} \sum w \cdot \{\sigma_N(\nabla_x \Phi) + \sigma_N(\nabla_y \Phi)\}$ is computed over a range of equisampled diameters e.g. 15 to 120 at 2.5
068 increments. A centered moving average of window using symmetric padding at edges, of default 5 is used
069 for smoothing. The result is a plot of score vs diameter. Prominent peaks in this plot highlight potential
070 segmentations at different size scales. The more possibilities, the more peaks. Users may use this in turn to
071 inform the diameter range to search. For automatic operation, the diameter that maximizes the contrast score
072 is used as the optimal diameter for Cellpose. The neighborhood size functions acts like an attention
073 mechanism (Extended Data Fig. 10c). The larger the size, the more the segmentation result corresponding
074 to larger objects is favored. If there is no larger salient segmentation, the optimal diameter selection will be
075 unchanged.

076
077 *Semi-automated cell probability thresholding for pretrained Cellpose models*. We observe for out-of-
078 distribution images and noisy input images, pretrained Cellpose 2D models can perform well using an
079 appropriate threshold for cell probability combined with u-Segment3D's gradient descent and spatial

connected component analysis to parse segmentations (Fig. 4j). The choice of threshold is particularly important. If the threshold is too high, there is no continuous path for the gradient descent. This results in over-segmentation. It is therefore better to veer on the side of caution and use a lower threshold to get a more connected foreground binary. However if this threshold is too low, it will not be in concordance with the predicted gradient field such that all voxels with predicted zero gradients will be extraneously and erroneously segmented. To automate the threshold, u-Segment3D applies multi-class Otsu thresholding to the normalized cell probability ($p \in [0,1]$) output of Cellpose, $p \leftarrow \frac{1}{1+e^{-p}}$. u-Segment3D further performs morphological closing to infill small holes. If only one object is known to be present, further morphological operations such as extracting the largest connected component and binary infilling can be conducted. The default Otsu thresholding is 2-class. If the segmentation partially captures the cells, we use 3-class Otsu and the lower of the two thresholds. Vice versa, if too much area is segmented, we use 3-class Otsu and the higher of the two thresholds. Optionally, we cast the threshold to the nearest decimal point, rounding down ($threshold \leftarrow \lfloor threshold * 10 \rfloor / 10$ where $\lfloor \cdot \rfloor$ is the floor operator).

Module 2: Gradient descent and distance transforms to assemble 2D slice-by-slice segmentation stacks into a 3D consensus segmentation

Methods in this module are used to implement the core 2D-to-3D segmentation algorithm outlined in Fig. 1d. If 2D segmentations are not provided in the form as a normalized cell probability (0-1) and 2D gradients in the manner of Cellpose⁶, then a 2D distance transform must be used to generate the necessary 2D gradients for consensus 3D segmentation.

2D Distance transforms

u-Segment3D categorizes the distance transforms according to whether the limit or attractor of propagating points using gradient descent over an infinite number of steps is implicitly or explicitly defined (Extended Data Fig.1). Explicitly defined transforms are further categorized by the type of attractor: a single fixed point source or comprises a set of points.

u-Segment3D implements distance transforms, Φ that solve the Eikonal equation ($\|\nabla\Phi\|^2 = 1$, which gives the shortest geodesic solution) or Poisson's equation ($\nabla^2\Phi = -1$, which gives a smooth harmonic solution), for the cell interior using numerically stable methods. The Eikonal equation finds the shortest time of propagation for a point. Poisson's equation can also be viewed as solving the shortest time of propagation but with the additional constraint of minimizing curvature, yielding smoother solutions.

Implicit attractor distance transforms

With only the boundary condition $\Phi = 0$, the Eikonal and Poisson equation conceptually propagates a wave inwards symmetrically from the cell boundaries. The limit solution is the definition of the medial axis skeleton, the locus of the centers of all inscribed maximal spheres of the object where these spheres touch the boundary at more than one point^{77,124,125}.

Euclidean distance transform (EDT in text). Solves the Eikonal equation using fast image morphological operations. u-Segment3D uses the memory and speed optimized implementation in the Python `edt` package released by the Seung Lab (<https://github.com/seung-lab/euclidean-distance-transform-3d>).

Poisson distance transform. Solves the Poisson equation using LU decomposition (Python Scipy, `scipy.sparse.linalg.spsolve`) for each cell in an image. It is solved in parallel in u-Segment3D using the Python Dask library.

Explicit attractor distance transforms

126 The implicit attractor solves the equations everywhere in the cell interior. The explicit attractor variants
127 modifies the equations to have different source terms (right hand side of equation) in different parts of the
128 cell interior. For the Eikonal equation, $\Phi = 0$ at the cell boundary and outside, non-source points obey
129 $\|\nabla\Phi\|^2 = 1$ and source points act as obstacles with vanishing speed, so that $\|\nabla\Phi\|^2 = 0$. For the Poisson
130 equation, $\Phi = 0$ at the cell boundary and outside, non-source points obey the Laplace equation, $\nabla^2\Phi = 0$ and
131 source points obey $\nabla^2\Phi = -1$.

132
133 (i) Point sources. A single interior point is designated as a point source. u-Segment3D finds the interior point
134 with Euclidean distance transform value greater than the percentile threshold (default: 10th percentile) nearest
135 the median coordinate of all points.

136 *Eikonal equation solution (Geodesic centroid distance)*. At the interior point, $\|\nabla\Phi\|^2 = 0$. The modified
137 equations are solved using the Fast Marching Method (FMM)⁸⁰, with the constraint enforced using a masked
138 array by the Python scikit-fmm library. Central first order differences are used to compute the unit normalized
139 2D gradient.

140
141 *Poisson equation solution (Poisson or diffusion centroid distance)*. Only at the interior point, $\nabla^2\Phi = -1$. The modified
142 equations are solved using LU decomposition as before. To apply power transformation with
143 exponent $p > 0$, the minimum is first subtracted from Φ to ensure positivity, $\Phi^p := (\Phi - \Phi_{\min})^p$. Central first
144 order differences are used to compute the respective unit normalized 2D gradient.

145
146 (ii) Point set sources. Any number of interior points are designated as point sources. u-Segment3D computes
147 the 2D medial axis skeleton as the point set. The binary skeleton is computed from the binary cell image by
148 iteratively removing border pixels over multiple image passes¹²⁶ (Python Scikit-image,
149 skimage.morphology.skeletonize). This raw result can often produce skeletons that have extraneous
150 branches that may be too close to a neighboring, contacted cell. To improve the skeleton quality, the binary
151 image is Gaussian filtered with $\sigma = 3$ pixels, rebinarized by mean value thresholding and reskeletonized.

152
153 *Eikonal equation solution (Geodesic centroid distance)*. For all points part of the skeleton, $\|\nabla\Phi\|^2 = 0$. The modified
154 equations are then solved using the Fast Marching Method (FMM)⁸⁰ as above with central first order
155 differences for computing the unit normalized 2D gradient. The gradients for all points part of the skeleton is
156 set to zero to enforce the limiting behavior under gradient descent.

157
158 *Poisson equation solution (Poisson or diffusion centroid distance)*. For all points part of the skeleton, $\nabla^2\Phi = -1$. The modified
159 equations are solved using LU decomposition as above with central first order differences
160 for computing the unit normalized 2D gradient. The gradients for all points part of the skeleton is set to zero
161 to enforce the limiting behavior under gradient descent.

162 163 Content-based averaging function, F

164 u-Segment fuses 3D volume images, I^i from $i = 1, \dots, N$ multiple views using a content-based average
165 function, F , with pixelwise weighting of the contribution from each view i given by the inverse local-variance,
166 $\sigma_{\mathcal{N}}^i$ evaluated over an isotropic neighborhood, \mathcal{N} of width P pixels

$$167 I_{fuse} = \frac{\sum_{i=1}^N \frac{1}{\sigma_{\mathcal{N}}^i + \alpha} I^i}{\sum_{i=1}^N \frac{1}{\sigma_{\mathcal{N}}^i + \alpha} + \varepsilon}$$

168 and α functions like a pseudo count. If α is small, $\sigma_{\mathcal{N}}^i$ dominates. If α is large, $\sigma_{\mathcal{N}}^i$ has little effect and all views
169 are equally weighted. F is equivalent to the simple mean as used by Cellpose⁶. ε is a small value (10^{-20}) to
170 prevent infinity. If the neighborhood is of width $P = 1$ pixels, F is also equivalent to the simple mean
171 (Extended Data Fig. 2a). Compared to potentially more accurate approaches such as solving the multi-view
172 reconstruction problem⁸², entropy-based averaging¹²⁷ or using Gaussian filters⁸¹, the above can be
173 implemented more efficiently with uniform filters.

174 175 Fusing normalized 2D cell probabilities (0-1) from orthoviews and binary thresholding

176 Stacked normalized 2D cell probabilities (0-1) are fused using the content-based averaging function, F above
177 with for neighborhood, $P = 1$ (default) pixels in concordance with the fusion of the 2D gradients below. For
178 Cellpose models, the raw cell probability output, p are first clipped to the range [-88.72, 88.72] to prevent
179 underflow/overflow in float32 and transformed, $p \leftarrow \frac{1}{1+e^{-p}}$. For methods yielding only 2D segmentations,
180 either i) fuse using the binary then apply appropriate Gaussian filter, ii) use the intermediate cell probability
181 image, which is always available for deep learning methods, or iii) generate a proxy cell probability image
182 e.g. using a rescaled Euclidean distance transform.

183 184 Fusing 2D gradients from orthoviews

185 Stacked 2D gradients from xy, xz, yz are pre-filtered with an isotropic Gaussian of $\sigma_{pre} = 1$. The fused 3D
186 gradients combines three separate fusion: fusing x-component from xy and xz views, y-component from xy
187 and yz views and z-component from xz and yz views. The 3D gradients is then post-filtered with $\sigma_{post} = 1$
188 (default) and unit length normalized. The greater σ_{post} is, the more the regularization effect, reducing the
189 number of attractors and preventing oversegmentation. This is helpful to segment larger and more branching
190 structure than represented by the majority of cells using pretrained Cellpose models. However it can also
191 merge smaller cells. For fusion, we use $\alpha = 0.5$ and in general $P = 1$ to maximize segmentation recall and
192 perform postprocessing to remove erroneous segmentations. Larger P improves segmentation precision but
193 may lose cells with lower contrast. These settings are generally not modified from the default. Preventing
194 oversegmentation can be more controllably carried out by adjusting the temporal decay parameter in the
195 gradient descent below first.

196 197 Gradient descent

198 Given the reconstructed 3D gradients, $\nabla\Phi$, gradient descent is applied to the set of all foreground image
199 coordinates, $\{x_n, y_n, z_n\}$. The iterative update equation for 3D gradient descent with momentum for iteration
200 number, $t = 0, \dots, T$, where $T = 250$ is the total number of iterations implemented by u-Segment3D is

$$201 \quad (x_n^t, y_n^t, z_n^t) \leftarrow (x_n^{t-1}, y_n^{t-1}, z_n^{t-1}) - \eta \frac{(\delta \cdot \nabla\Phi(x_n^{t-1}, y_n^{t-1}, z_n^{t-1}) + \mu \cdot \nabla\Phi(x_n^{t-2}, y_n^{t-2}, z_n^{t-2}))}{\delta + \mu}$$

202 Where $\nabla\Phi$ is the gradient map, μ is the momentum parameter governing the extent the past gradient is
203 considered, ranging from 0-1 (default $\mu = 0.95$), and $\delta > \mu$ is the weighting of the current gradient and the
204 step-size. $\mu = 0$ recovers the standard gradient descent. Nearest interpolation is used for efficiency so that
205 (x_n^t, y_n^t, z_n^t) is always integer valued. η defines the step-size and is defined as a function of the iteration
206 number,

$$207 \quad \eta = \frac{\delta}{1 + t \cdot \tau}$$

208 $\tau \in \mathbb{R}^+$ is a floating point number that controls the step-size decay⁷. The greater τ is, the less the points are
209 propagated. When $\tau = 0$, the step-size is constant $\eta = \delta$.

210 211 Parallelized variant of gradient descent on subvolumes

212 The volume was divided into subvolumes of (256, 512, 512) with 25% overlap. Within each subvolume we
213 run gradient descent with momentum for 250 iterations, momenta, $\mu = 0.98$, step size $\delta = 1$ to propagate the
214 position of foreground pixels towards its final attractor in the 3D gradient map.

215

216 **Image-based connected component analysis for identifying the unique number of cell centers for** 217 **instance segmentation**

218 The method is depicted in Extended Data Fig. 4 for a 2D image and described here for a 3D image. Step (i),
219 the final ($t = T$) gradient descent advected foreground coordinate positions, $\{(x_n^{t=T}, y_n^{t=T}, z_n^{t=T})\}$ is rasterized
220 onto the image grid by flooring, i.e. $\{(\lfloor x_n^{t=T} \rfloor, \lfloor y_n^{t=T} \rfloor, \lfloor z_n^{t=T} \rfloor)\}$ and clipping values to be within the bounds of the
221 $L \times M \times N$ image volume i.e. $0 \leq \lfloor x_n^{t=T} \rfloor \leq L - 1$, $0 \leq \lfloor y_n^{t=T} \rfloor \leq M - 1$, $0 \leq \lfloor z_n^{t=T} \rfloor \leq N - 1$. Step (ii), the
222 number of points at each voxel position is tabulated, each point contributing +1 count. Step (iii), the counts
223 image is Gaussian filtered with $\sigma = 1$ as a fast approximation to the Gaussian kernel density estimate to
224 produce a density heatmap, $\rho(x, y)$ for 2D and $p(x, y, z)$ for 3D. This step functions to account for uncertainty
225 and spatially connect up points into a cluster in a soft manner assignment. The greater the Gaussian filter σ
226 the more nearby points will be grouped into the same hotspot. This can be helpful when segmenting
227 branching structures. (iv) The density heatmap is sparse and therefore can be segmented using a mean
228 threshold with an optional tunable offset specified as a constant multiplicative factor, k of the standard
229 deviation (std) of ρ , $threshold = mean(\rho) + k \cdot std(\rho)$. For real images, we set $k = 0$ and have never needed
230 to change this. Image connected component analysis is applied to the segmented binary to create the distinct
231 spatial cluster segmentation at $t = T$, $L^{t=T}(x, y)$ for 2D and $L^{t=T}(x, y, z)$ for 3D. Each foreground coordinate
232 positions, $\{(x_n^{t=T}, y_n^{t=T}, z_n^{t=T})\}$ is then assigned the cell id by lookup and the final cell segmentation is
233 computed by generating the segmentation at their initial voxel positions, $\{(x_n^{t=0}, y_n^{t=0}, z_n^{t=0})\}$. For connected
234 component u-Segment3D uses the optimized, parallel implementation developed by the Seung Lab
235 (<https://github.com/seung-lab/connected-components-3d>).

236

237 **Module 3: Postprocessing the 3D consensus segmentation**

238 Described below are the implemented postprocessing methods that can be applied to the initial 3D
239 segmentation generated by gradient descent and connected component analysis (module 2). The
240 recommended sequential u-Segment3D workflow is: i) removal of implausible predicted cells involving ia)
241 removal of predicted cells below a user-specified size limit (in voxels), ib) the removal of cells which are
242 inconsistent with that implied by the predicted 3D gradients and ic) the removal of cells that are too statistically
243 large ($volume > mean(volumes) + k \cdot std(volumes)$ where k is a multiplicative factor, default $k = 5$); ii)
244 labelspreading to smooth, enforce the spatial connectivity constraint of segmentation and propagate
245 segmentation to better adhere to the desired features given by a guide image; iii) guided filter refinement to
246 transfer missing local cellular structures to the segmentation.

247 The guided image does not need to be the same as the raw. Generally it is a version of the raw whereby
248 desired cellular features are enhanced.

249 **(i) Removal of implausible predicted cells.**

250 *(ia) Removal of predicted cells that are too small.* Volume of individual cells are computed as number of
251 voxels and the respective ids are removed by setting to 0 if volume less than the user-specified threshold
252 (default 200). Additionally each cell is checked whether they comprise multiple spatially disconnected
253 components. If so, only the largest component is retained as each segmented cell should be spatially
254 contiguous.

255 *(ib) Removal of predicted cells that are inconsistent with the predicted gradients.* The reconstructed 3D
256 gradients, $\nabla \Phi_{3D \text{ segmentation}}$ are computed from xy, xz, yz views of the assembled consensus 3D
257 segmentation. The mean absolute error with the predicted 3D gradients is computed per cell, $MAE_{cell} =$
258 $mean(|\nabla \Phi_{3D \text{ segmentation}} - \nabla \Phi_{3D}|)_cell$. If $MAE_{cell} >$ user-defined threshold (default 0.8 for $\sigma_{post} = 1$). If the
259 post Gaussian filter σ_{post} used when fusing gradients from orthoviews is > 1 , the threshold may need to be
260 relaxed i.e. threshold > 0.8 .

(ic) *Removal of predicted cells that are statistically too large.* Ratiometric uneven illumination correction may over enhance background at the borders of the image, which may result in the segmentation of very large regions. Also in dense tissue, when staining is inhomogeneous and weak, multiple closely packed cells may be segmented as one in the initial 2D segmentation. Assuming cell volumes are approximately normally distributed, we filter out improbably large cell segmentation we use mean and standard deviation (std) of all cell volumes to set a cutoff retaining all cells with volume smaller than $\text{mean}(\text{volume}) + k \cdot \text{std}(\text{cell volumes})$, where $k = 5$ by default.

(ii) Labelspredding to smooth and propagate cell segmentation with spatial connectivity constraint.

Labelspredding⁹⁷ is a semi-supervised learning method developed to infer the label of objects in a dataset given the labels to a partial subset of the objects by diffusing on a constructed affinity graph between objects. u-Segment3D adapts this algorithm for cell segmentation. To be computationally scalable, for each cell mask, M_i , a subvolume, V_i , is cropped with the size of its rectilinear bounding box padded isotropically by a default 25 voxels. Every label in V_i is one-hot encoded to form a label vector $L \in \mathbb{R}^{N \times p}$ where N is the total number of voxels and p the number of unique cell ids, including background. We then construct an affinity matrix, A between voxels as a weighted sum ($\alpha = 0.25$) of an affinity matrix based on the intensity difference in the guide image, I between 8-connected voxel neighbors, $A_{\text{intensity}}$ and one based on the connectivity alone, $A_{\text{laplacian}}$:

$$A = \alpha A_{intensity} + (1 - \alpha) A_{laplacian} \text{ and}$$

$$A_{intensity}(i, j) = \begin{cases} e^{-D_{intensity}^2 / (2\mu(D_{intensity})^2)} & i \neq j \\ 1 & i = j \end{cases}$$

$$A_{laplacian}(i, j) = \begin{cases} e^{-D_{laplacian}^2 / (2\mu(D_{laplacian})^2)} & i \neq j \\ 1 & i = j \end{cases}$$

$D_{intensity}$ is the pairwise absolute difference in intensity values between two neighboring voxels i and j . $D_{laplacian}$ is the graph Laplacian with a value of 1 if a voxel i is a neighbor of voxel j , and 0 otherwise. $\mu(D)$ denotes the mean value of the entries of the matrix D . The iterative labelspread propagation is then

$$\begin{aligned} z &\in \mathbb{R}^{N \times p} \\ z^{t=0} &= \mathbf{0} \\ z^{t+1} &\leftarrow (1 - \gamma)A z^t + (\gamma)I_p \end{aligned}$$

where t is the iteration number, $\mathbf{0}$, the empty vector and γ is a ‘clamping’ factor that controls the extent the original labeling is preserved. The final z is normalized using the softmax operation, and argmax is used to obtain the final cell ids. The refined cell mask, M_i^{refine} for cell id i is all voxels where the final z is assigned to the same cell id i . Parallel multiprocessing is used to efficiently apply the refinement to all individual cells. It is recommended to set the parameters per dataset, depending on the extent of correction required. We typically start with a conservative $\alpha = 0.25$, $\gamma = 0.75$, and run the propagation for 25 iterations. The guide image, I is usually the normalized input image (after any preprocessing) to the 2D segmentation but can be any processed image that enhances the desired cell features. For additional speed, particularly for tissue, we also typically treat each cell mask, M_i as binary instead of multi-label.

(iii) Guided image filtering to recover missing high-frequency features and subcellular protrusions.

The guided filter⁹⁸, a local linear filter that can be implemented in linear time, is used to efficiently transfer the structures of a guidance image, I to the input image to be filtered, P . Setting I to be the ridge-filtered input image to enhance high-frequency cellular protrusion and vessel features, and P to be the binary mask of cell i , the resulting filtered output Q is a ‘feathered’ binary, being refined to appear an alpha matte near the object boundaries. The radius of the boundary that is refined is controlled by a radius parameter, $r = 35$ voxels (by default), and the extent of structure transfer by a regularization parameter, $\epsilon = 1 \times 10^{-4}$. We find the binary mask can be rough. The stronger the features are enhanced in I the more prominent the transferred structure. Q is then re-binarized using multi-threshold Otsu. Typically, we use the two-class binary Otsu. As for labelspreading, guided filtering is applied to a cropped subvolume, V_i , with the size of its rectilinear bounding box padded isotropically by a default 25 voxels. For computational efficiency, for touching cells, we perform the guided filter segmentation independently for each cell and mask out spatial regions occupied by another cell id. More accurately, we can obtain the guided filter response for all cell ids in the subvolume to define the valid region. Parallel multiprocessing is used to perform the guided filter refinement to all individual cells. Long protrusions with a length longer than r cannot be recovered using guided filtering however, the guided

313 filter result may assist the application of subsequent matching algorithms or serve as a improved seed image
314 for watershed algorithms.

316 Semi-automatic tuning of diameter parameter in Cellpose models

317 The process is illustrated in Extended Data Fig. 10a. for 3D and described below.

318 *Determining the optimal diameter* for 2D image. Given a pixel neighborhood size with isotropic width, P pixels,
319 we conduct a parameter screen of diameter = $[d_{low}, d_{high}]$ (typically $d_{low} = 10$, $d_{high} = 120$) at equal
320 increments of 2.5 or 5. For each diameter, a contrast score is computed taking into account the ‘sharpness’
321 of the Cellpose model predicted 2D x- and y- gradients ($\nabla_x \Phi$ and $\nabla_y \Phi$ respectively) and optionally the
322 normalized cell probability map, p (0-1).

323

$$\text{Contrast score}(d) = \frac{1}{N} \sum w \cdot \{\sigma_N(\nabla_x \Phi) + \sigma_N(\nabla_y \Phi)\}$$

324 Where N is the total number of image pixels, w is a pixelwise weighting set to be p and $\sigma_N(I)$ is the pixelwise
325 local standard deviation of the image I evaluated over the isotropic local neighborhood of width P pixels. p is
326 computed from the unnormalized raw cell probabilities after clipping to range [-88.72, 88.72] (to prevent
327 overflow or underflow in float32) by applying the transformation, $p \leftarrow \frac{1}{1+e^{-p}}$. The result is a contrast score
328 function of d . A centralized moving average of 5 (if diameter increment is 2.5) or 3 (if diameter increment is
329 5) is applied to smooth the contrast score function. The diameter d that maximizes the contrast score is used
330 as the optimal diameter, d_{opt} in the Cellpose model. We generally observe no difference in d_{opt} between
331 $w = 1$ or $w = p$ for Cellpose models.

332

333 *Determining the optimal diameter for 3D volume.* If cell size varies slice-by-slice, the optimal diameter
334 determination for 2D is applied slice-by-slice (Fig. 4). This becomes computationally limiting as the number
335 of slices increase. Instead, we find good performance, if we set the optimal diameter for a representative 2D
336 slice in each orthoview. This representative 2D slice can be set automatically to (i) the most in-focus slice as
337 determined by the highest mean sobel magnitude, (ii) the slice with highest mean intensity, (iii) the mid-slice,
338 or (iv) be user-defined.

340 Other tested segmentation methods

341 **Cellpose 3D mode with pretrained models.** We ran pretrained cellpose models in 3D mode by setting
342 do_3D = True. Since this mode is prone to oversegmentation and Cellpose 3D only allows to have one
343 diameter, we used the largest diameter inferred by our contrast score function. Models are run twice. The
344 first time is to obtain the raw, unnormalized cell probability image, which was then used to determine the
345 binarization threshold. We then ran a second time using the determined threshold to generate a 3D
346 segmentation. We had to additionally remove all cells with volume < 2500 voxels to get the maximum average
347 AP.

348

349 **Omnipose 3D.** We ran the pretrained *plant_omni* model following the example in the documentation
350 (https://omnipose.readthedocs.io/examples/mono_channel_3D.html). This model operates on the raw image
351 downsampled by a factor 1/3 in all dimensions and does not reinterpolate the raw image to isotropic resolution.
352 No other preprocessing was used. We found the raw output to predict many small objects leading to an
353 artificially low computed AP relative to qualitative assessment. Therefore we additionally removed all objects
354 with volume < 2500 voxels to get the maximum average AP. If volumes > 2500 voxels were removed, this
355 affected the AP computed for lateral primordial images containing smaller cells.

357 **Cellpose 3D mode with Omnipose trained ‘plant_cp’ model.** We ran the 2D pretrained Cellpose *plant_cp*
358 model using the same function call as the example in the Omnipose documentation for *plant_omni* but with
359 `omni=False` and `do_3D=True`. As for Omnipose3D we found many small predicted object and additionally
360 postprocessed the output segmentation by removing all objects with volume < 2500 voxels.
361

362 Evaluation of segmentation quality

363 Segmentation quality in single images

364 For single 2D and 3D images, we find the optimal matching between predicted and reference cell
365 segmentations. Given a total number of M predicted cells, and N reference cells, we iterate and find for each
366 predicted cell i , its K -nearest reference cells according to the distance between their centroids. For each of
367 the K -nearest reference cells, we compute the intersection-over-union (IoU) metric (0-1) (see below). This
368 produces a $\text{IoU}(i,j) \in \mathbb{R}^{M \times K}$ matrix. We convert this to a distance cost matrix, $\text{dist}(i,j) = 1 - \text{IoU}(i,j) \in$
369 $\mathbb{R}^{M \times K}$. The optimal matching between predicted and reference cells is then found by solving the linear sum
370 assignment using a modified Jonker-Volgenant algorithm with no initialization¹²⁸ (Python Scipy,
371 `scipy.optimize.linear_sum_assignment`) and retaining only the pairings that spatially overlap ($\text{IoU}(i,j) > 0$).
372 The segmentation quality for an image is then assessed by (i) the mean IoU, to measure the spatial overlap
373 of matched predicted and reference cells and (ii) the F1 score (see below), the harmonic mean of precision
374 and recall to measure how accurately the segmentation detects the correct number of reference cells.
375

376 **Intersection-over-union (IoU).** Also called the Jaccard index, is defined as the total number of pixels in
377 the intersection divided by the total number of pixels in the union of two binary segmentation masks A and
378 B , $\text{IoU}(A,B) = \frac{|A \cap B|}{|A \cup B|}$.

379 **F1 score.** Predicted cells that are validly matched to a reference cell ($\text{IoU}>0$) define the true positives, TP.
380 Predicted cells that are not matched are false positives, FP, and the reference cells with no valid matches
381 are false negatives, FN. The precision is the number of matched cells divided by the total number of
382 predicted cells, precision = $\frac{\text{TP}}{\text{TP} + \text{FP}}$. The recall is the number of matched cells divided by the total number of
383 reference cells, recall = $\frac{\text{TP}}{\text{TP} + \text{FN}}$. F1 score is the harmonic mean of precision and recall, $\text{F1} = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$.
384

385 Average precision curve

386 We evaluate the quality of cell segmentation using average precision consistent with popular segmentation
387 models such as StarDist³, Cellpose⁶ and Omnipose⁷. Each predicted cell label mask is matched to the
388 reference cell label mask that is most similar, as defined by IoU. The predictions in an image are evaluated
389 at various levels of IoU. At a lower IoU, fewer pixels in a predicted cell have to match a corresponding
390 reference cell for a valid match. The valid matches define the true positives, TP, the cells with no valid
391 matches are false positives, FP, and the reference cells with no valid matches are false negatives, FN at that
392 IoU threshold. Using these values, the standard average precision metric (AP) for each image is:

$$393 \quad \text{AP} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

394 The average precision (AP) curve is reported for a dataset by averaging over the average precision metric
395 for each image in the dataset. Optimal matching of predicted and reference cells is too computationally
396 demanding in 3D even when restricting the search to nearest neighbors. We use the same approximate
397 matching implementation in Cellpose which is derived from the fast matching functions in StarDist. It is
398 important to note this matching is not invariant to cell id permutation. To compute the correct AP, we first
399 relabel all cells sequentially after performing an indirect stable sort based on their (x,y,z) centroid for both
400 reference and predicted cell segmentation independently. In line with Cellpose, the AP curve is reported for
401 11 IoU thresholds equisampling the range [0.5,1.0]. Many datasets e.g. Ovules do not rigorously label every
402 cell in the image but only the cells of the primary, single connected component object in the field of view.

403 However pretrained Cellpose models would predict all cells in the field-of-view. For fair evaluation, for these
404 datasets (all except for Embedseg skull nuclei, *Platynereis* nuclei and *Platynereis* ISH nuclei), we use the
405 reference segmentation to define the foreground connected component cluster to evaluate AP and include
406 all predicted cells within spatial connected components that have at least 25% overlap with a reference
407 connected component cluster. For DeepVesselNet, this is at least 1% overlap due to the thinness.

408

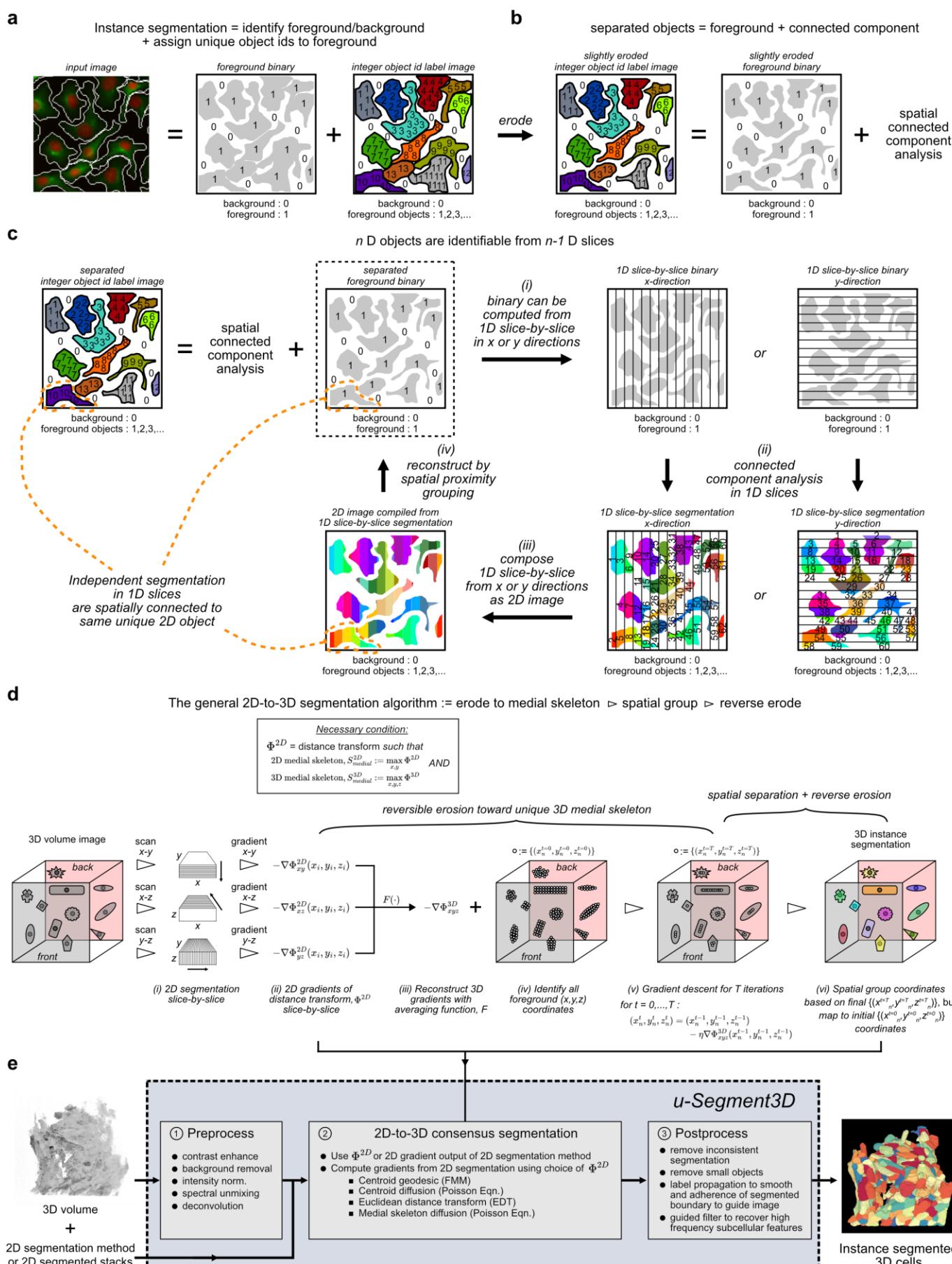
409 **Visualization**

410 We use the Fiji ImageJ¹²⁹ 3D viewer plugin to render 3D intensity and segmentation image volumes. For
411 visualization of intensity in Fig. 5, we acquired a snapshot of the rendering and applied an inverse lookup
412 table. Surface meshes in Fig. 5 were extracted using u-Unwrap3D¹⁵ and visualized using MeshLab¹³⁰.
413 Rotating surface mesh movies were created using ChimeraX¹³¹.

414

415

Figures



416

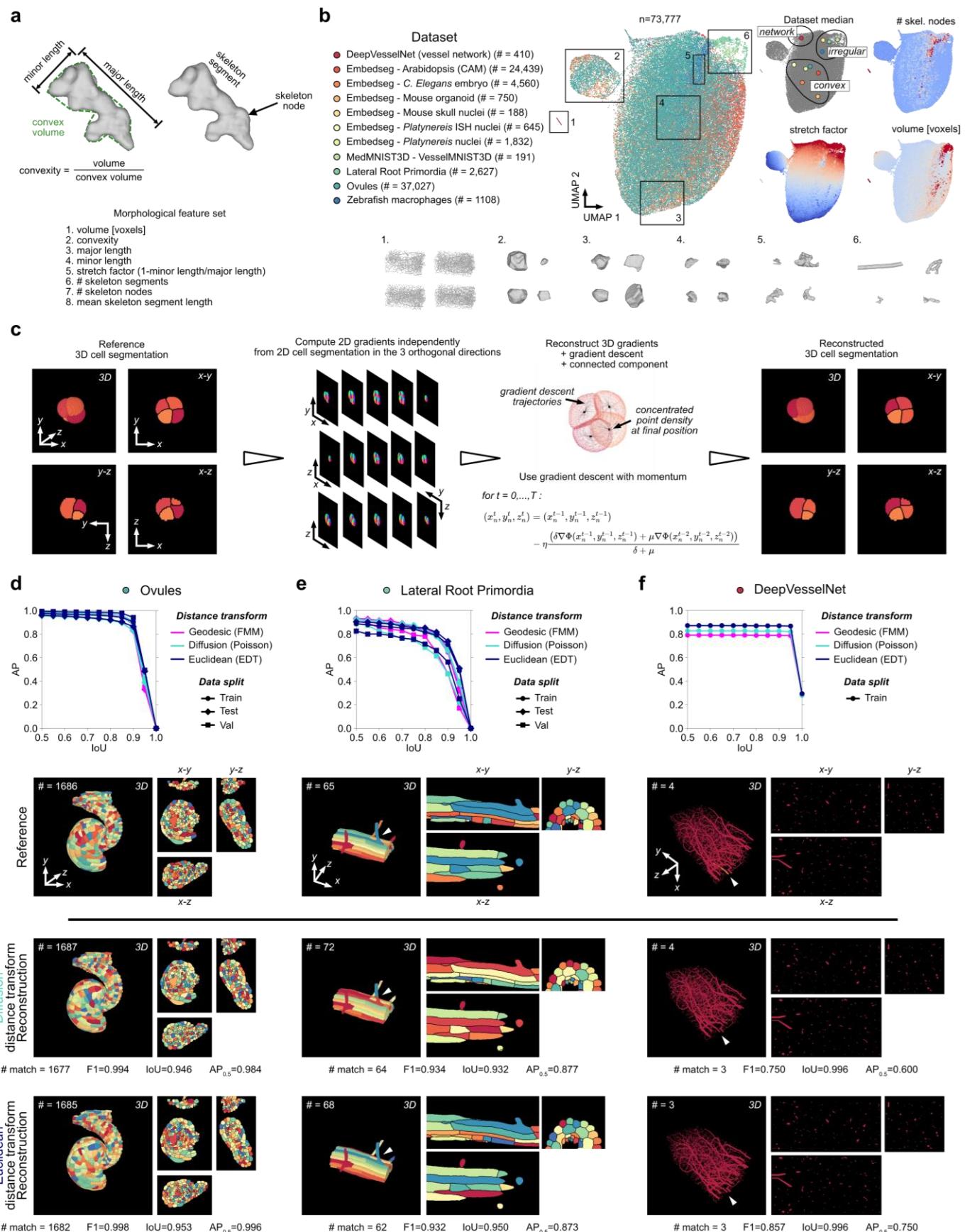
Figure 1. u-Segment3D is a toolbox for generating consensus instance 3D segmentation from 2D segmentation methods. a) Computational representation of the 2D segmentation of densely packed cells as two images, a foreground binary mask and a labelled image where each unique cell is assigned a unique integer id. **b)** Equivalent representation of the eroded segmentation such that individual cells are now spatially

417
418
419
420

421 separated using a single foreground binary mask which can be parsed using connected component analysis
422 to recover individual cell ids. **c)** Schematic of the factorization of 2D instance cell segmentation to orthogonal
423 1D slices in x- or y- directions and subsequent perfect reconstruction from 1d instance segmentation by
424 stacking and 2D spatial proximal grouping. **d)** Schematic of the minimal set of algorithmic steps to
425 operationalize the conceptual framework in c) for 2D to generate the consensus 3D segmentation when cells
426 may be densely packed. **e)** u-Segment3D is a toolbox to enable the application of the algorithmic steps in d)
427 to real datasets with additional preprocessing methods to adapt any pretrained 2D segmentation model or
428 2D method and postprocessing methods to improve and recover missing local features in the reconstructed
429 3D segmentation such as subcellular protrusions.

430

431



432

433

434

435

436

437

438

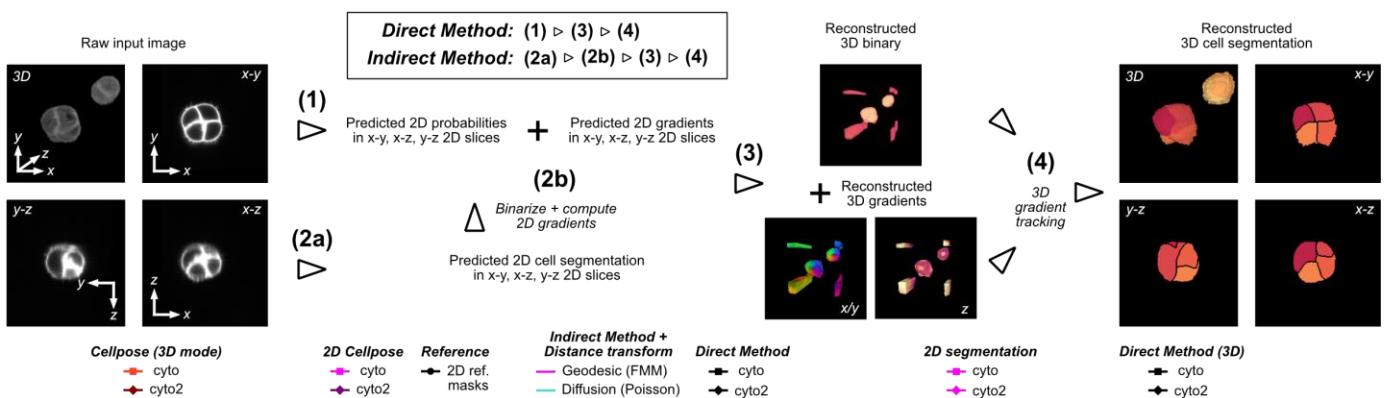
439

Figure 2. u-Segment3D reconstructs optimal 3D segmentation from synthetic ideal 2D segmentation labels from orthogonal x-y, x-z, y-z views. a) Illustration of the 8 computed geometrical and topological features to describe shape complexity. **b)** UMAP embedding of individual cells from 11 real datasets which together represent the spectrum of morphological complexity from convex-spherical, branching to networks. The zebrafish macrophages dataset was internally curated with the aid of u-Segment3D. All others are public (Methods). Left: colormap of individual dataset and total number of uniquely labelled cells in each dataset. Middle: UMAP, each point is a cell, color-coded by their origin dataset. Right: Median UMAP coordinate of

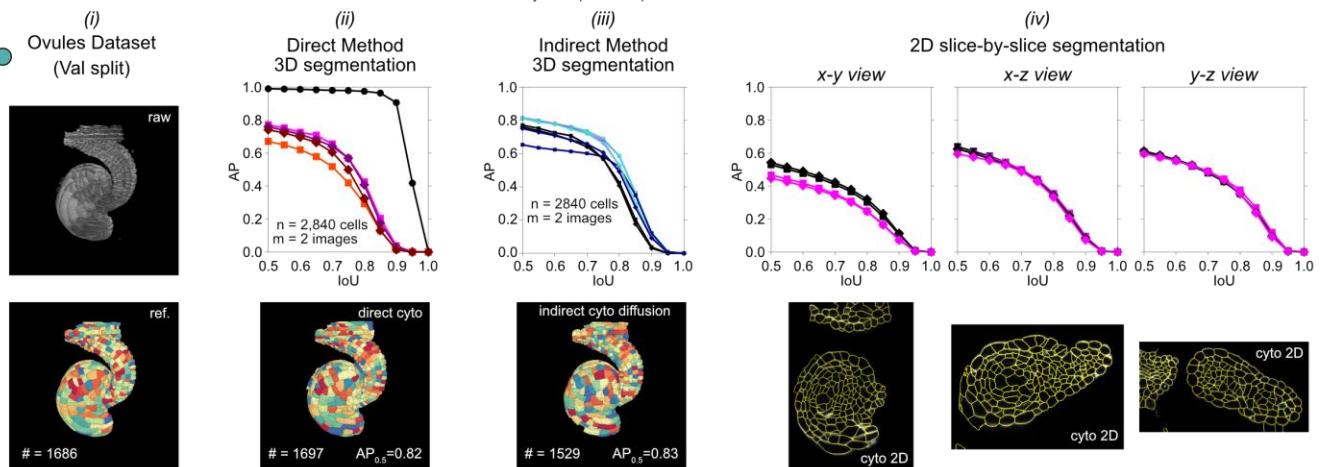
440 each dataset (top left) and heatmap of three features representing the extent of branching (total number of
441 skeleton nodes, top right), the extent of elongation (stretch factor = 1 – minor length/major length) and their
442 image size (total number of voxels). **c)** Illustration of the experimental workflow to compute 2D slice-by-slice
443 distance transforms in orthogonal directions given the reference 3D cell segmentation labels and then
444 applying u-Segment3D to reconstruct the 3D segmentation from 2D stacks for the 11 real datasets in b). **d)**
445 Reconstruction performance measured by the average precision curve (Methods) for the Ovules dataset
446 using three different 2D distance transforms. From top to bottom: average precision vs intersection over union
447 (IoU) curve; 3D rendering of reference, point-based diffusion distance transform reconstructed vs skeleton-
448 based Euclidean distance transform reconstructed 3D cell segmentation and their respective midsllices in the
449 three orthogonal views. **e), f)** Same as d) for the Lateral Root Primordia dataset containing examples of
450 branching morphology and DeepVesselNet representing entire complex, thin network morphologies.
451 Individual cells are uniquely colored but are not color matched with respect to the reference segmentation.

452

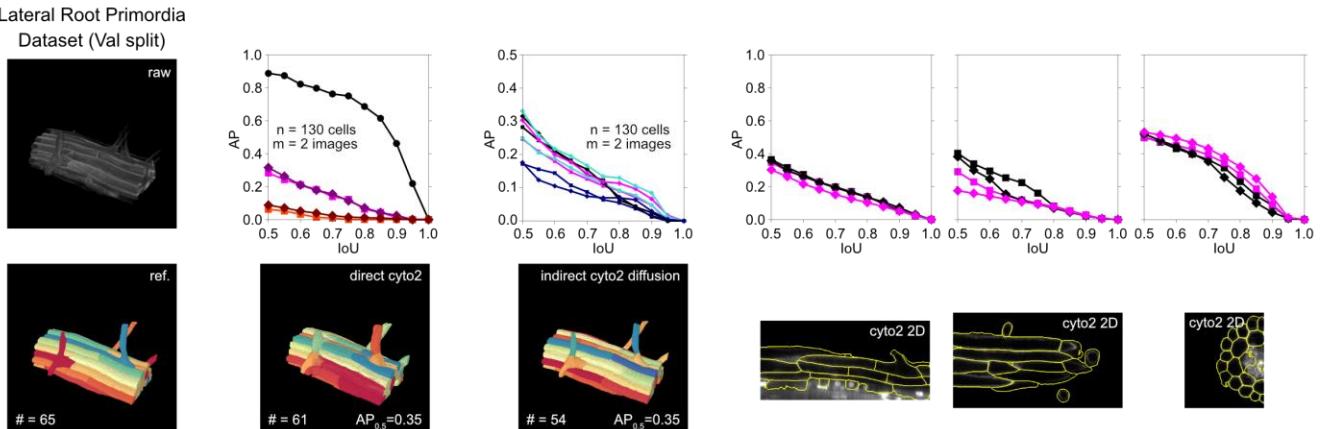
a Constructing consistent 3D cell segmentation from 2D segmentation



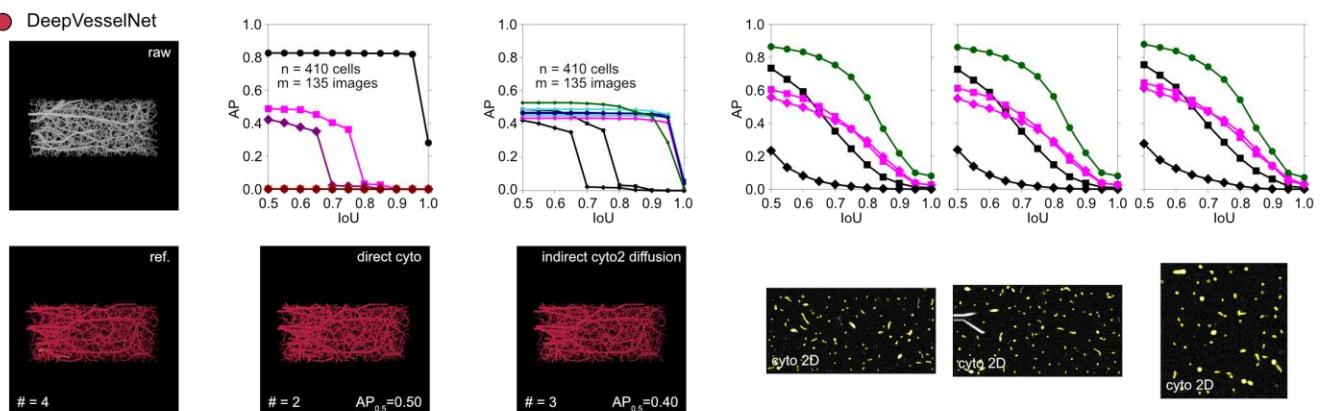
b



c



d



453

454

455

456

457

458

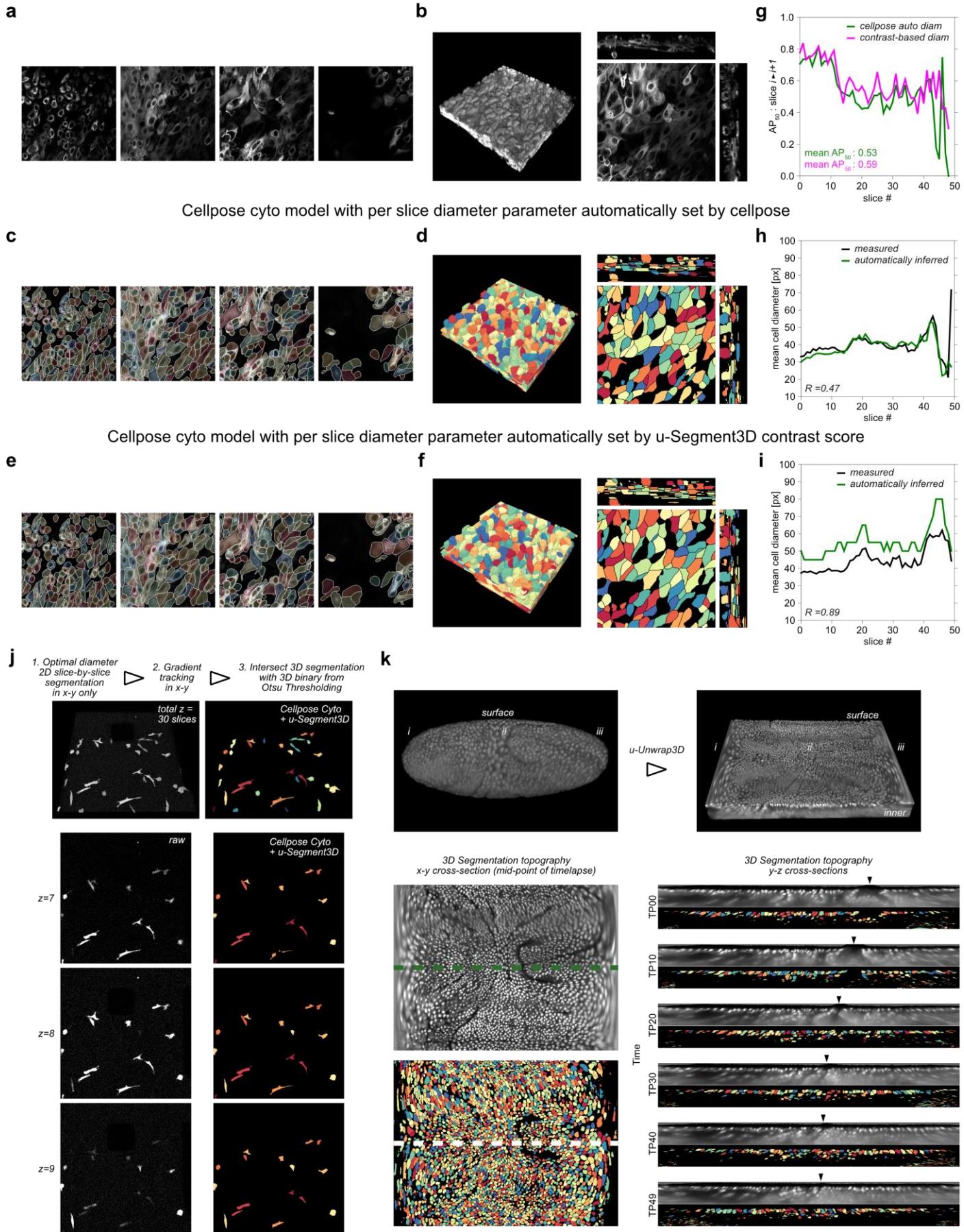
Figure 3. u-Segment3D segmentation of 3D real datasets using pretrained cellpose2D applied to orthogonal x-y, x-z, y-z views. a) Illustration of the two workflows that can be implemented by u-Segment3D to generate 3D cell segmentation. The direct method (steps 1,3,4) performs the generation utilising the 2D segmentation method's predicted distance transform or spatial gradients and cell probability maps from orthogonal views. The indirect method (steps 2a, 2b, 3, 4) first converts the stack of 2D cell segmentation

459 labelled images from orthogonal views using a chosen 2D distance transform to generate the necessary
460 spatial gradients, as in Fig. 2c. **b)** 3D cell segmentation performance of the Ovules dataset (validation (val)
461 split, n=2840 cells, m=2 volumes) using pretrained cellpose 2D with u-Segment3D. (i) 3D rendering of raw
462 image (top) and reference 3D labels (bottom). (ii) Average precision (AP) curve for the direct method using
463 pretrained Cellpose 2D cyto or cyto2 models relative to the AP curve of the best reconstruction from synthetic
464 2D segmentation in Fig. 2d (top). 3D rendering of the segmentation using the best cellpose model for the
465 direct method (bottom). (iii) Average precision (AP) curve for the indirect method using the 2D segmentation
466 of pretrained Cellpose 2D cyto or cyto2 models and different 2D distance transforms relative to the AP curve
467 of the corresponding direct method constructed 3D segmentation (top). 3D rendering of the segmentation
468 using the best cellpose model for the indirect method (bottom). (iv) Average precision (AP) curve of the 2D
469 segmentation accuracy averaged across all 2D slices for each Cellpose 2D model (magenta lines) in each
470 orthogonal view, x-y, x-z, y-z from left-to-right relative to the 2D segmentation accuracy of the corresponding
471 direct method 3D segmentation (black lines). **c), d)** Same as b) for the Lateral Root Primordia (validation (val)
472 split, n=130 cells, m=2 volumes) and DeepVesselNet (n=410 network components, m=135 images) dataset
473 containing examples of branching morphology and DeepVesselNet representing entire complex, thin network
474 morphologies. For d) we additionally evaluated the performance of binary Otsu thresholding as a baseline
475 2D segmentation method (green line).

476

477

x-y aggregation only to construct 3D cell segmentation of thin tissue



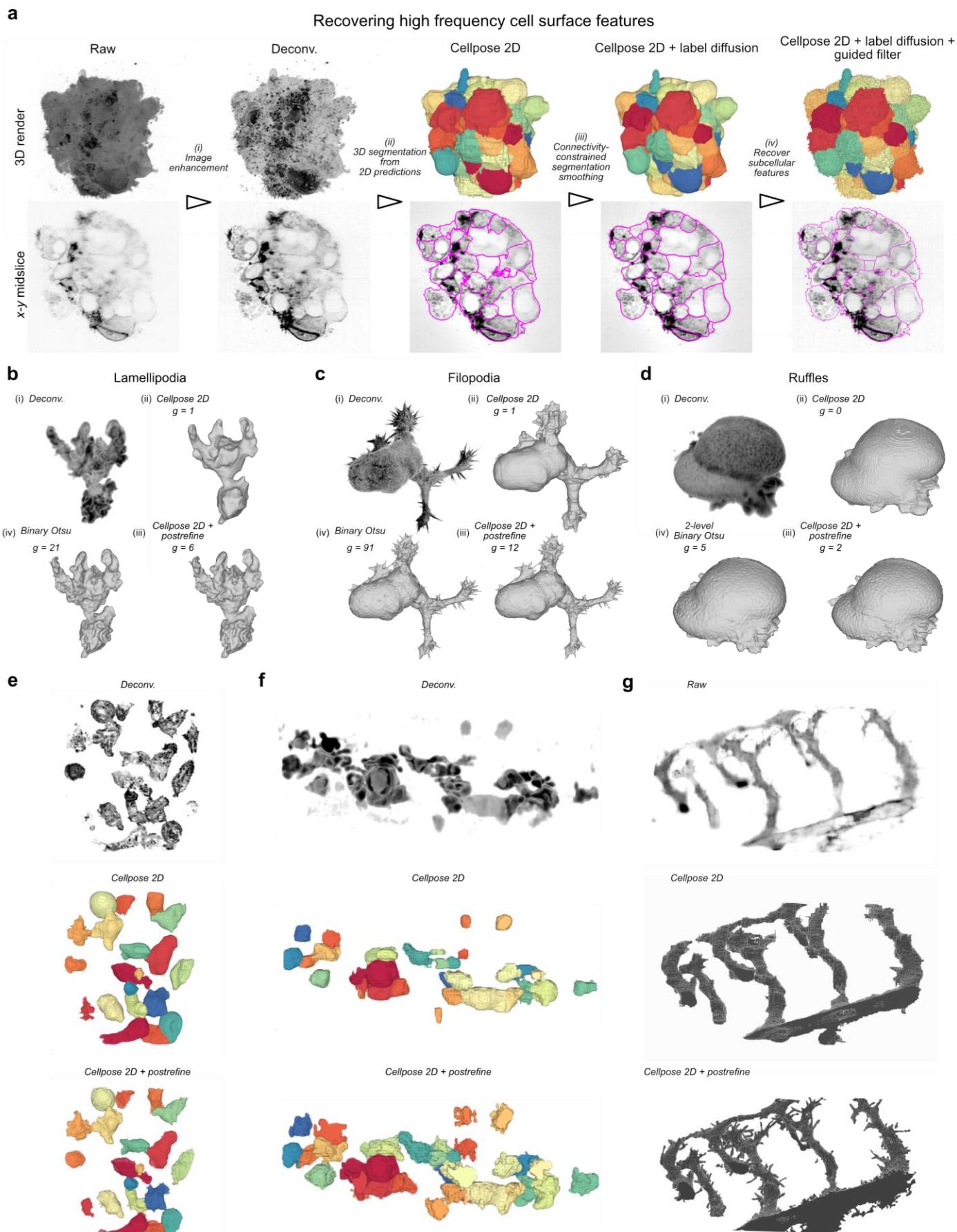
478
480
481
482
483

Figure 4. u-Segment3D segmentation of anisotropic 3D volumes using only x-y 2D stacks. a) Four equi-sampled x-y image slices from top-to-bottom of the air-liquid interface keratinocyte culture. **b)** 3D render of the air-liquid interface keratinocyte culture with axial interpolation to isotropic voxel resolution (left) and corresponding mid-section orthoslices (right). **c)** Cellpose 2D cell segmentations using the 'cyto' model and diameter automatically determined per-slice by cellpose. Cells are individually colored and overlaid onto the

484 four x-y image slices in a). White boundaries delineate individual cell boundaries within a slice. **d)** 3D render
485 of the u-Segment3D consensus segmentation of the x-y 2D segmentation stacks in c) (left) and
486 corresponding mid-section orthoslices (right). **e)** Cellpose 2D cell segmentations using the 'cyto' model and
487 diameter automatically determined per-slice by u-Segment3D contrast score. Cells are individually colored
488 and overlaid onto the four x-y image slices in a). White boundaries delineate individual cell boundaries within
489 a slice. **f)** 3D render of the u-Segment3D consensus segmentation of the x-y 2D segmentation stacks in e)
490 (left) and corresponding mid-section orthoslices (right). **g)** 2D cell segmentation consistency measured by
491 the average precision at IoU cutoff = 0.5 between success z-slices as a function of z-slice id for per-slice
492 Cellpose model diameter auto-determined by Cellpose (green line) or u-Segment3D contrast score (magenta
493 line). **h)** Mean cell diameter inferred by Cellpose (green line) and measured after obtaining the corresponding
494 2D cell segmentation (black line) for each xy-slice. **i)** Mean cell diameter inferred by peak position in the u-
495 Segment3D contrast score (green line) and measured after obtaining the corresponding 2D cell segmentation
496 (black line) for each xy-slice. **j)** Segmentation of MDA231 human breast carcinoma cells from the 3D Cell
497 Tracking Challenge using u-Segment3D to aggregate Cellpose 2D xy-slice segmentations with optimal
498 diameter selection by contrast score. 3D render of raw and 3D cell segmentation (top) and in consecutive 2D
499 xy-slices (bottom). **k)** u-Segment3D per-frame segmentation of unwrapped proximal surface topography
500 volumes of drosophila. i) Unwrapping of the proximal surface using u-Unwrap3D. ii) Cellpose 2D and u-
501 Segment3D contrast score diameter segmentation of the surface x-y slice at timepoint (TP) 25. iii) Mid y-z
502 cross-section snapshots of the raw (top) and segmented (bottom) topography volumes for 6 equi-spaced
503 timepoints. R denotes the Pearson's R in panels g)-i).

504

505



506

507

508

509

510

511

512

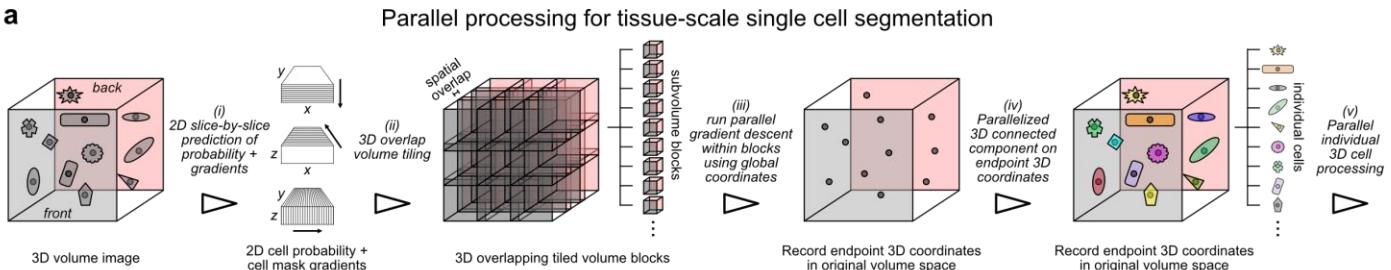
Figure 5. u-Segment3D postprocessing recovers missing high-frequency, high-curvature subcellular features. a) General u-Segment3D workflow with postprocessing to segment individual cells and recover subcellular features of each cell. 3D render (top) and x-y midslice (bottom) of the output at each step. **b)** Binary segmentation and recovery of lamellipodial features on a dendritic cell using u-Segment3d postprocessing. 3D rendering of the (i) deconvolved input, (ii) initial 3D segmentation from aggregated cellpose 2D cell probability map (after step ii of a)), (iii) final postprocessed 3D segmentation (after step iv of

513 a)) and comparison with the segmentation from binary Otsu thresholding on the 3D image intensity. g = genus
514 of extracted surface mesh. **c), d)** Binary segmentation and recovery of filopodial and ruffle features on a
515 HBEC and COR-L23 cell using u-Segment3d postprocessing. **e)** Single cell 3D segmentation of T-cells using
516 cellpose 2D with u-Segment3D postprocessing. 3D render of deconvolved image volume (top), initial 3D
517 segmentation from aggregated cellpose 2D (middle) and final 3D segmentation with recovered subcellular
518 protrusions (bottom). **f)** Single cell 3D segmentation of zebrafish macrophages using cellpose 2D with u-
519 Segment3D postprocessing. 3D render of deconvolved image volume (top), initial 3D segmentation from
520 aggregated cellpose 2D (middle) and final 3D segmentation with recovered subcellular protrusions (bottom).
521 **g)** Binary 3D segmentation of developing zebrafish vasculature using cellpose 2D with u-Segment3D
522 postprocessing. 3D render of raw image volume (top), initial 3D segmentation from aggregated cellpose 2D
523 cell probability maps (middle) and final 3D segmentation with recovered sprouting vessels (bottom).

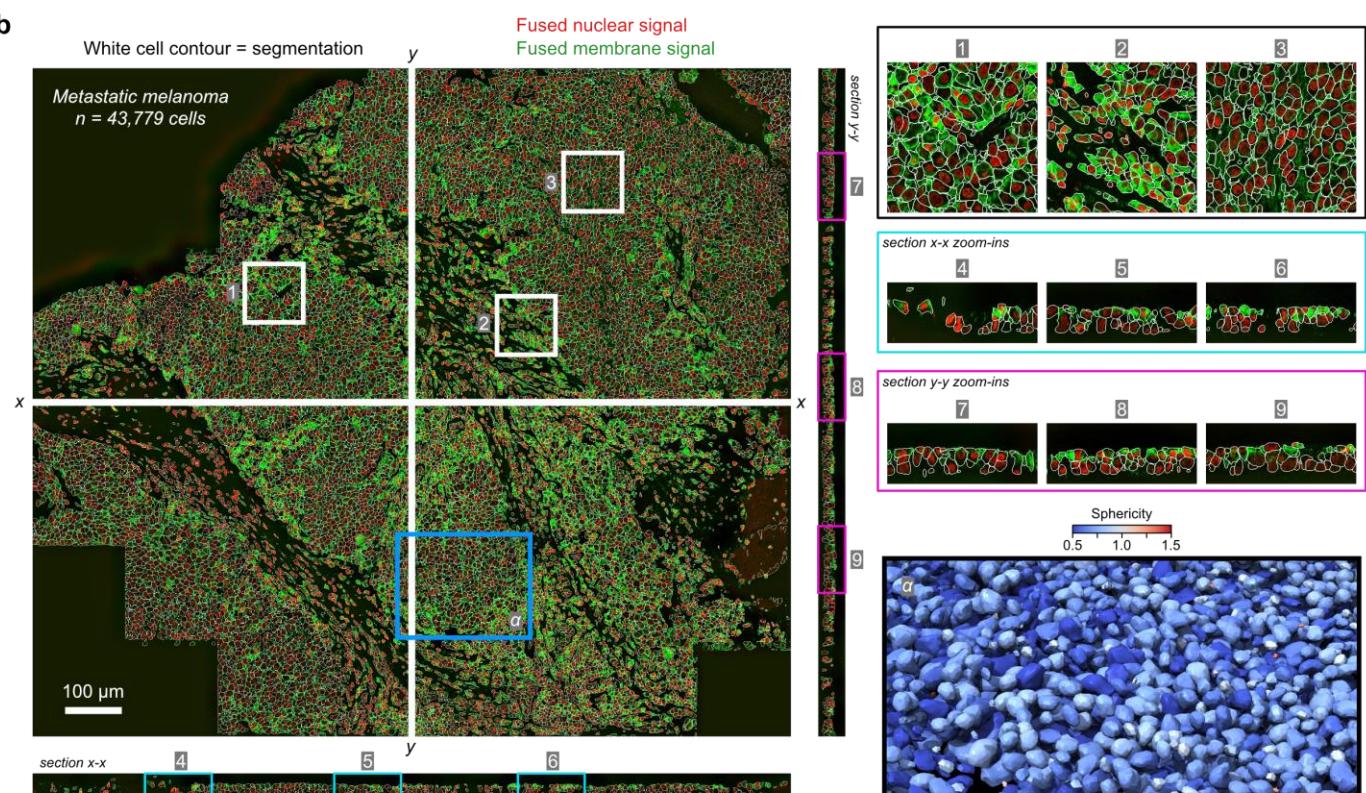
524

525

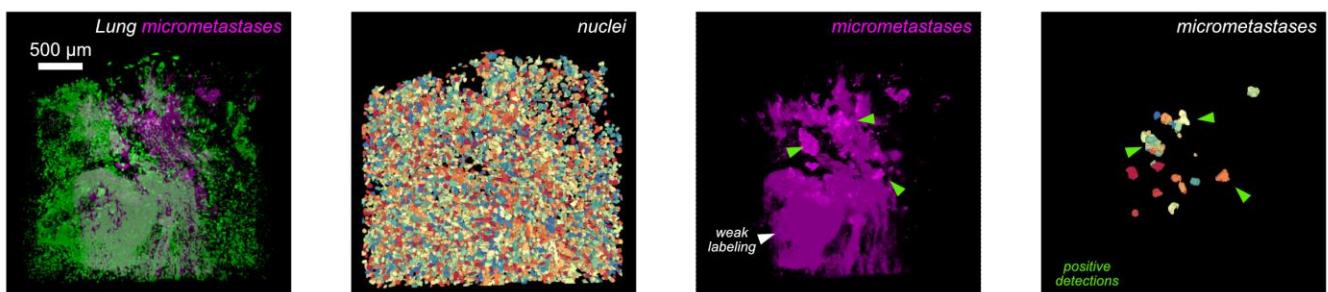
a



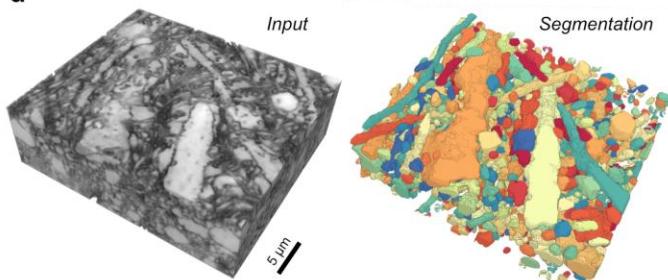
b



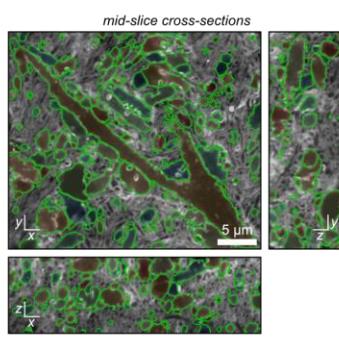
c



d



e



f

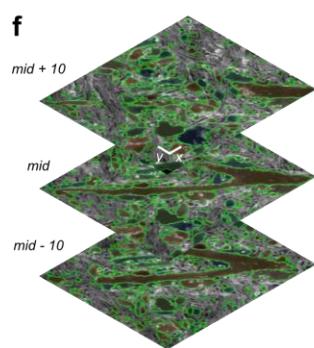


Figure 6. u-Segment3D uses parallel computing for tissue-scale segmentation. a) Schematic of the parallelized gradient descent tracking in overlapped subvolume tiles used by u-Segment3D to facilitate single cell 3D segmentation in tissue. b) x-y, x-z, y-z midslice cross-sections of the fused nuclear (red) and membrane (green) signal channels from multiple biomarkers (Methods) for a CyCIF multiplexed patient biopsy of metastatic melanoma with white boundaries to delineate the individual cells in each view (left). Zoom-ins of 3 subregions in x-y (black box, regions 1-3), x-z (cyan box, regions 4-6), y-z (magenta box, 7-9)

533 cross-sections (top right). Zoom-in of the extracted 3D cell meshes within the blue rectangle (subregion α) of
534 xy view, heatmap colored by sphericity (sphere = 1). **c)** 3D Segmentation of individual lung nuclei (green)
535 and cancer micrometases (magenta) in cleared tissue. Left-to-right: merged input volume image, individual
536 segmented nuclei from nuclei channel, micrometastases only image showing weak, non-specific staining
537 (white arrow) compared to specific positive staining (green arrow), and final u-Segment3D micrometastases
538 3D segmentation post-filtered by mean cell intensity. **d)** 3D render of the input coCATs volume (left) and u-
539 Segment3D aggregated cellpose2D 3D segmentation of salient tissue architecture (right). **e)** Mid-slice cross-
540 sections in x-y, x-z, y-z with individual segmentation boundaries outlined in green and its area individual color
541 overlaid with the input image. **f)** Mid \pm 10 z-slice x-y cross-section with individual segmentation boundaries
542 outlined in green and its area individual color overlaid with the input image.

543

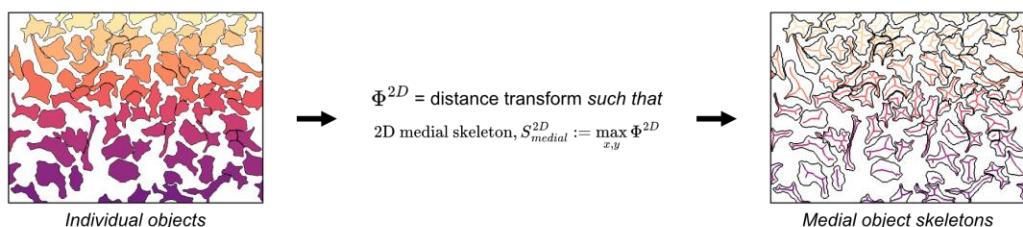
544

545

Supplementary Figures

a

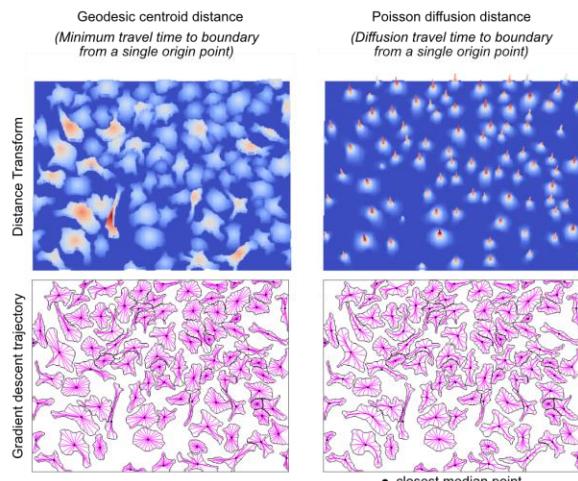
Using distance transforms to implicitly describe individual medial object skeletons



b

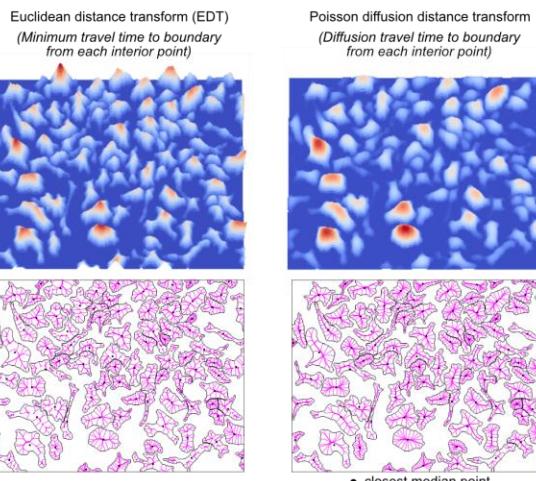
'Explicit' attractors

(i) 'Point' sources

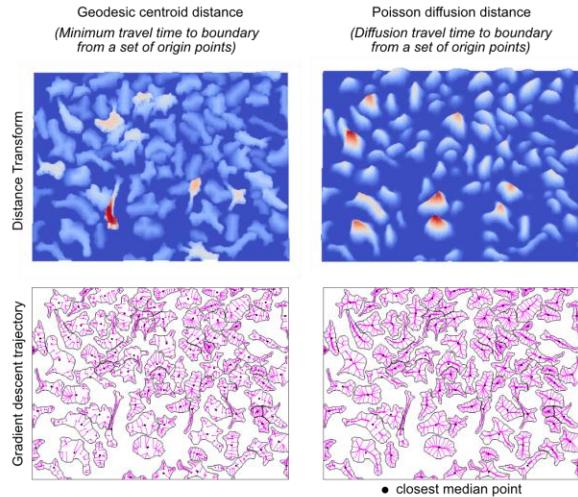


c

'Implicit' attractors

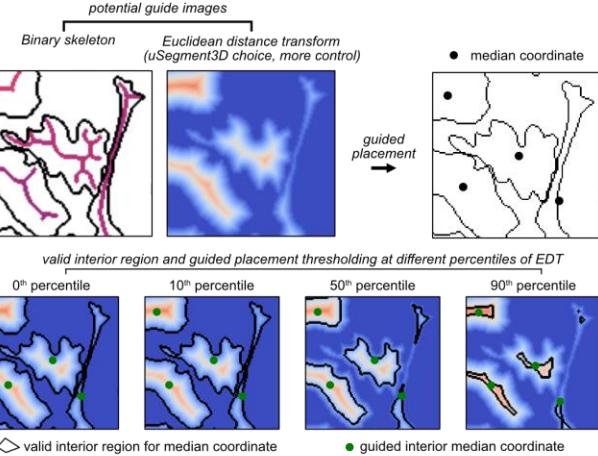


(ii) 'Point Set' sources



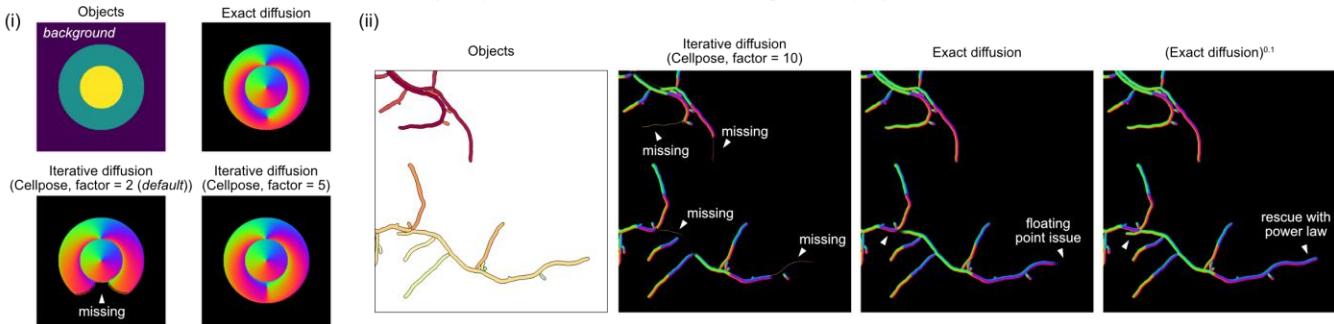
d

Controlling the location of the origin point using guide image



e

Distance transform computation must be valid everywhere inside shape
 (example: Poisson heat diffusion for 'elongated' shapes)



546

Extended Data Figure 1. 2D distance transforms for 2D-to-3D segmentation in u-Segment3D. a)

Schematic illustration of the necessary property of 2D distance transforms for 3D segmentation should have maxima that sample the 2D medial skeleton of individual shapes. b) Example of the first- (geodesic, left) and second-order (diffusion, right) shortest distance 'explicit' attractors distance transforms implemented in u-Segment3D where the limit of performing gradient descent is explicitly specified as either (i) a single point-

547

548

549

550

551

552 source on the 2D medial skeleton or (ii) a set of points as the source along the 2D medial skeleton. **c)** Example
553 of the first- (geodesic, left) and second-order (diffusion, right) shortest distance ‘implicit’ skeletal-based
554 attractor distance transforms implemented in u-Segment3D where performing gradient descent may not
555 necessarily converge stably to the limit. For each example in **b), c)**, distance transform is represented as a
556 relative 3D height map, colored blue (lowest) to red (highest) (top) and trajectory (magenta) of equi-sampled
557 boundary points under gradient descent (bottom). Black point = closest internal shape point to the median
558 shape coordinate. **d)** Illustration of using percentile-based thresholding of the Euclidean distance transform
559 of individual cells as a soft constraint to find the medial centroid 2D coordinate for convex and concave
560 shapes used in computing the 2D point-source distance transforms in c). **e)** The computed unit-normalized
561 2D gradients for (i) circle in doughnut synthetic shape and (ii) elongated touching bacterial shapes using the
562 iterative simulated diffusion in Cellpose compared to the exact diffusion with boundary conditions solution
563 obtained by LU factorization (Methods) in u-Segment3D. 2D gradients are colored according to direction. The
564 number of simulated diffusion steps in Cellpose equals the factor multiplied by the number of pixels occupied
565 per shape.

566

567

a

Content-based average function, F

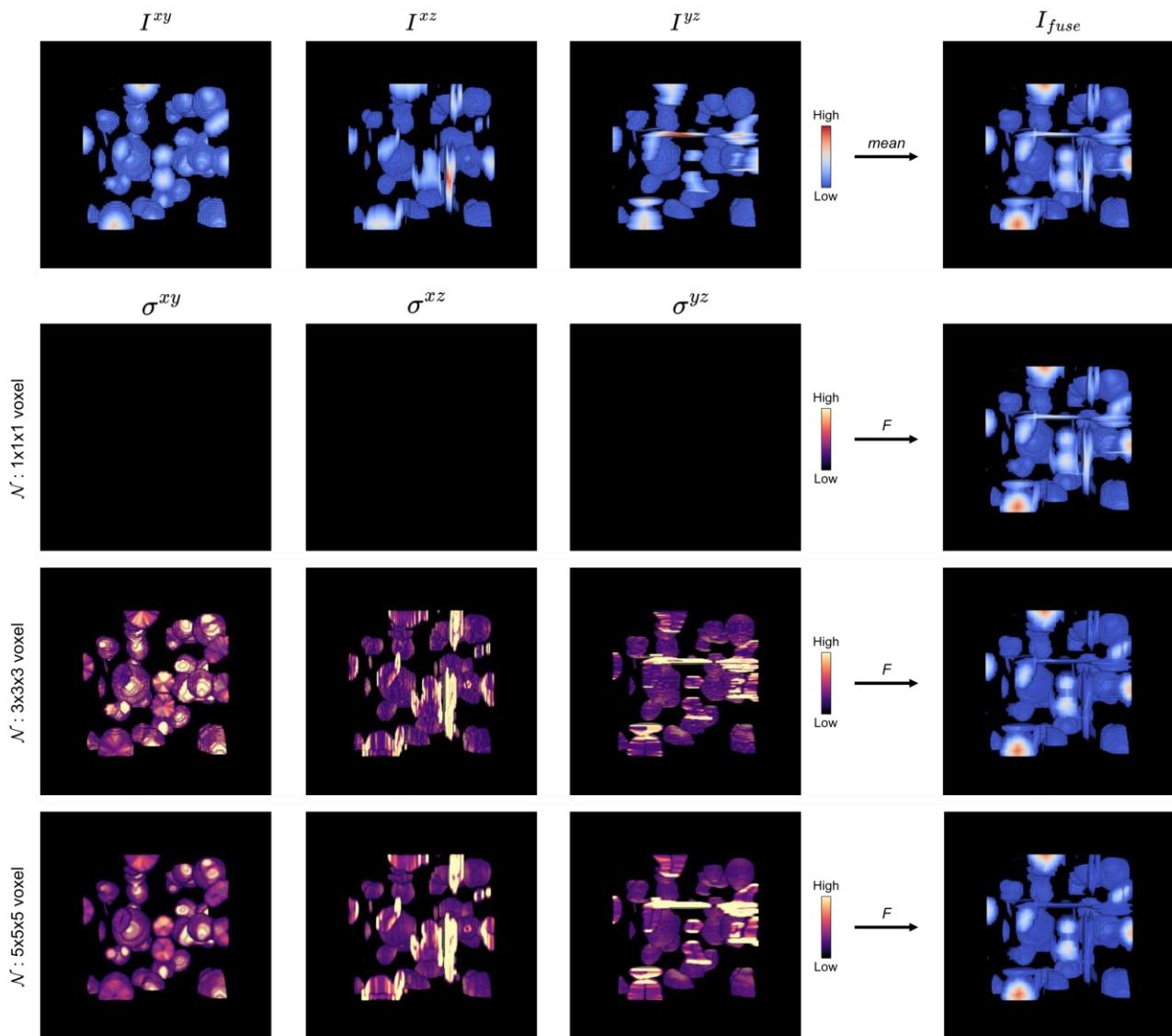
where:

$$I_{\text{fuse}} = \frac{\sum_{i=1}^N \frac{1}{\sigma_N^i + \alpha} I^i}{\sum_{i=1}^N \frac{1}{\sigma_N^i + \alpha} + \varepsilon}$$

I^i : image, i corresponding to data from view i
 σ_N^i : is the standard deviation of values of image, i within a local neighborhood of width P pixels
 α : pseudo value
 ε : small value to prevent infinity

b

Example: Combining 2D distance transform 2D segmentation from x-y, x-z, y-z views



c

Binary Otsu thresholding of the fused image



568

569

570

571

572

573

574

575

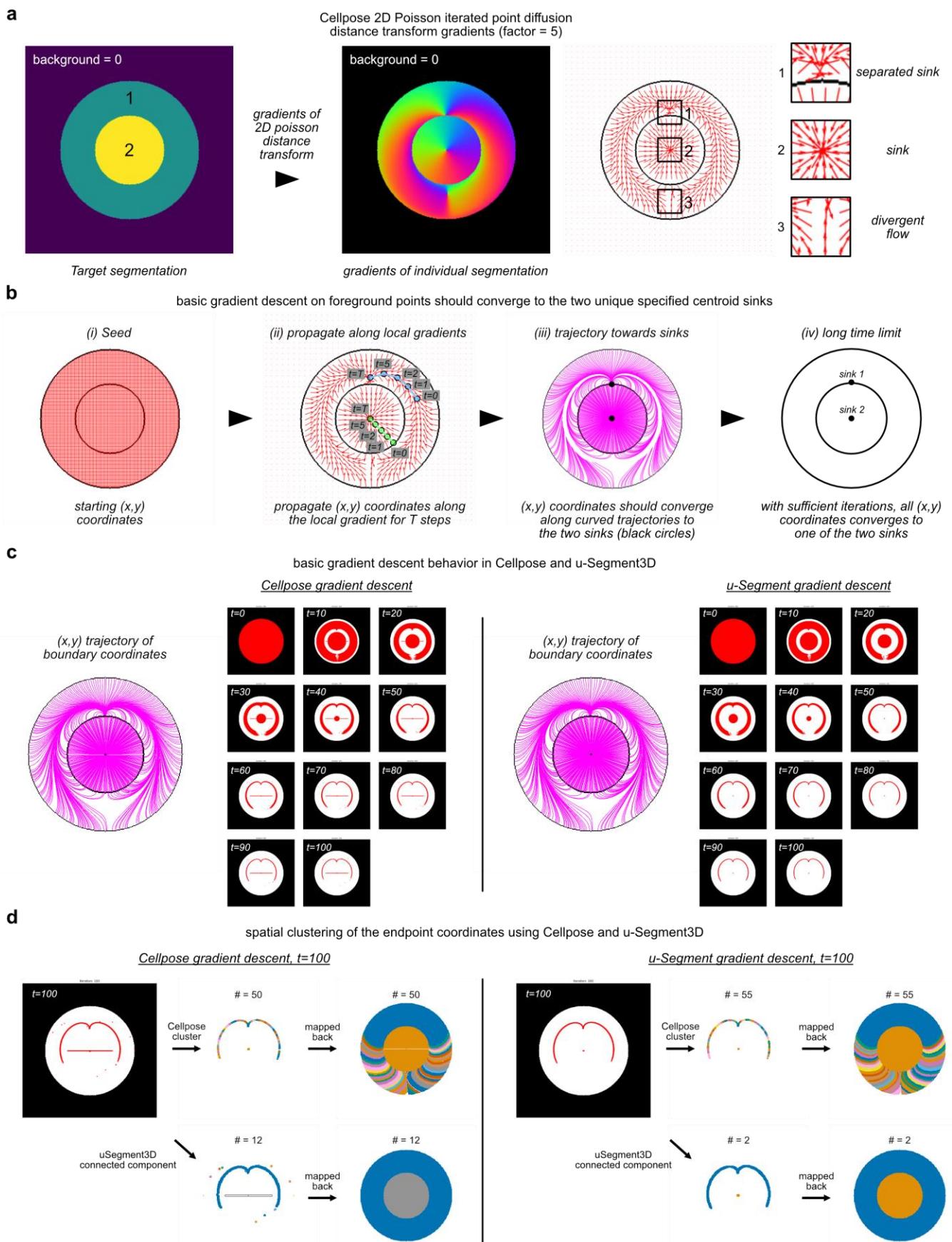
576

Extended Data Figure 2. Content-based average function for combining data from multiple views for 2D-to-3D segmentation. **a)** Mathematical definition of the content-based average function as the inverse local variance weighted mean of input image values in an isotropic neighborhood of width P pixels. **b)** Example of applying the average function defined by a) to fuse the stacked 2D distance transforms after 2D slice-by-slice segmentation of a volume of spherical cells from xy, xz and yz views. 1st row: Left-to-right, the Euclidean distance transform colored blue (low) to red (high) from the three orthoviews and fused distance transform from pixelwise mean. 2nd row: Left-to-right, the per-pixel local variance weight image σ for each orthoview for a neighborhood of width $P = 1$ pixel and resultant fused distance transform using the content-

577 based averaging. 3rd row: Left-to-right, the per-pixel local variance weight image σ for each orthoview for a
578 neighborhood of width $P = 3$ pixel and resultant fused distance transform using the content-based averaging.
579 4th row: Left-to-right, the per-pixel local variance weight image σ for each orthoview for a neighborhood of
580 width $P = 5$ pixel and resultant fused distance transform using the content-based averaging. **c)** Result of
581 applying binary Otsu thresholding on the fused distance transform based on pixelwise mean, and content-
582 based averaging using neighborhood of width $P = 1, 3, 5$ pixels from left-to-right.

583

584



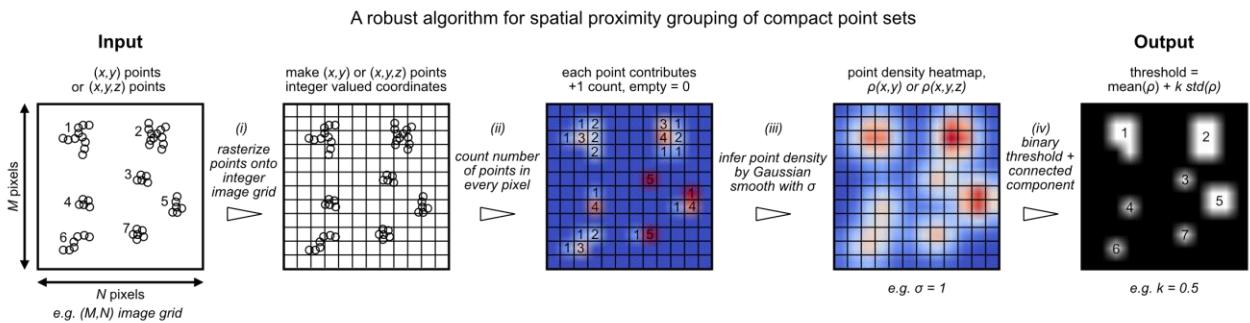
585
586 **Extended Data Figure 3. Illustration of 2D Gradient descent for reversible 2D shape erosion. a)**
587 Cellpose computed 2D gradients for a synthetic shape of a doughnut (labelled 1) surrounding a circle
588 (labelled 2) (left) with the 2D gradients visualized as red arrows (right) and zoomed in at three regions with
589 different types of flow behavior. The number of simulated steps equals the factor = 5 multiplied by the number
590 of pixels occupied per shape. **b)** Schematic of the expected behavior when gradient descent is iteratively
591 applied to propagate the initial foreground (x,y) coordinates with the limit being convergence to 2 black centroids. **c)** Observed point trajectories (magenta lines) and snapshots of
592 coordinate positions (red points in images) running gradient descent in cellpose (left) vs u-Segment3D (right)
593

594 for 100 iterations. **d)** Recovered cell shapes based on applying Cellpose (top) or u-Segment3D (bottom)
595 spatial proximity clustering on the final coordinate positions after 100 iterations of Cellpose (left) or u-
596 Segment3D (right) gradient descent.

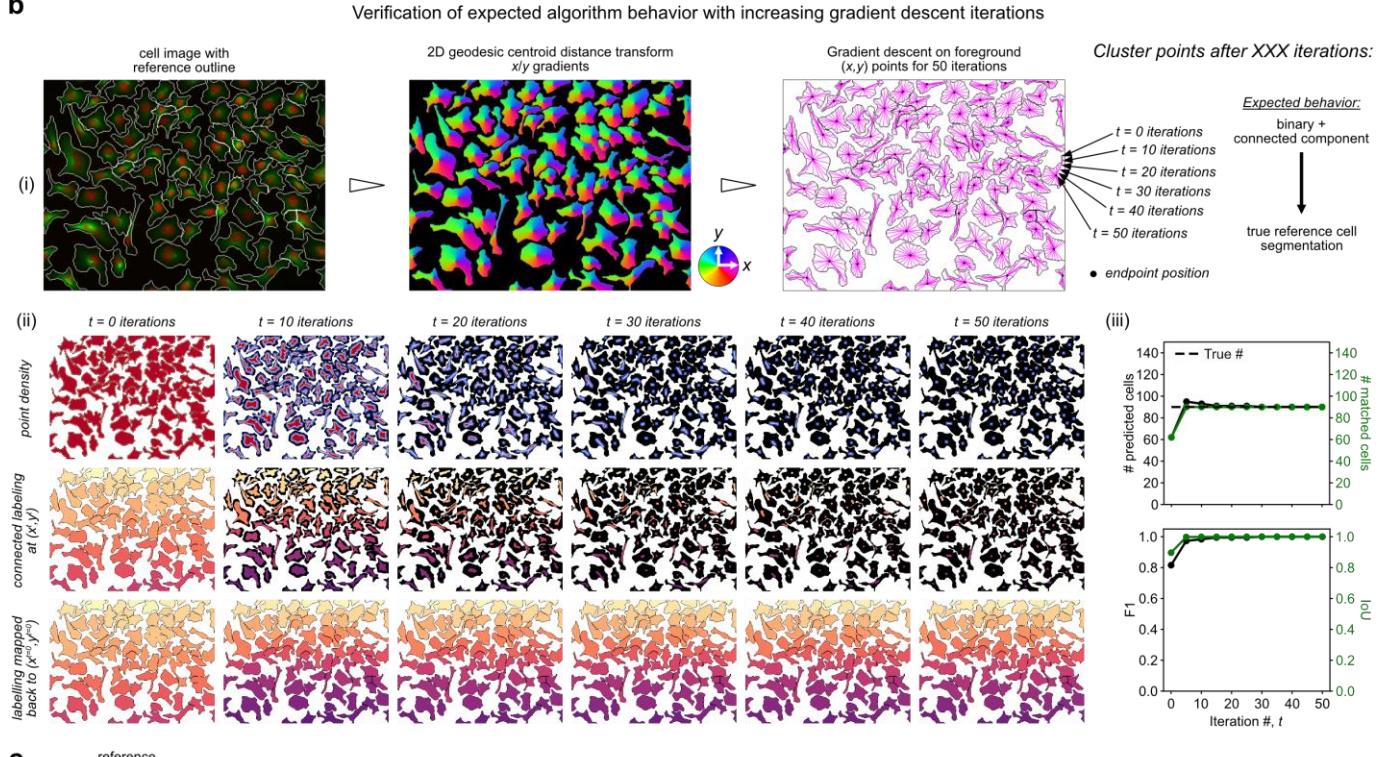
597

598

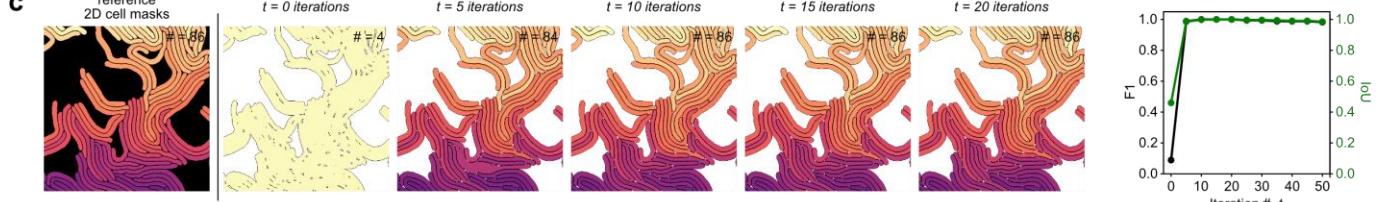
a



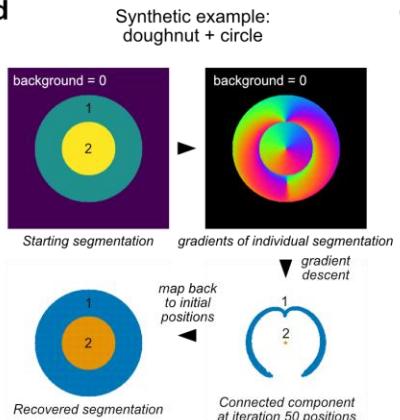
b



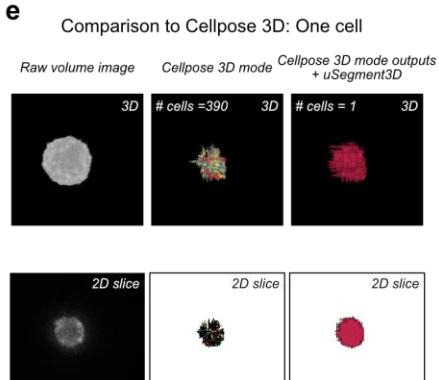
c



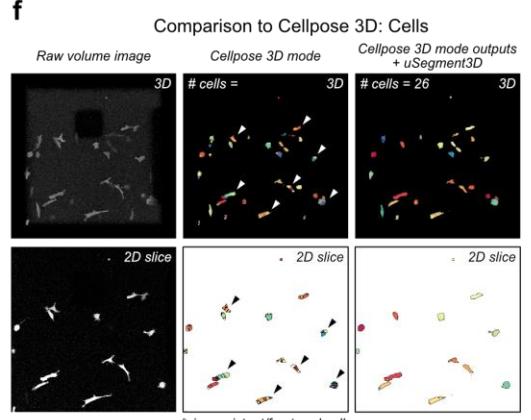
d



e



f



599

600

601

602

603

604

605

606

607

608

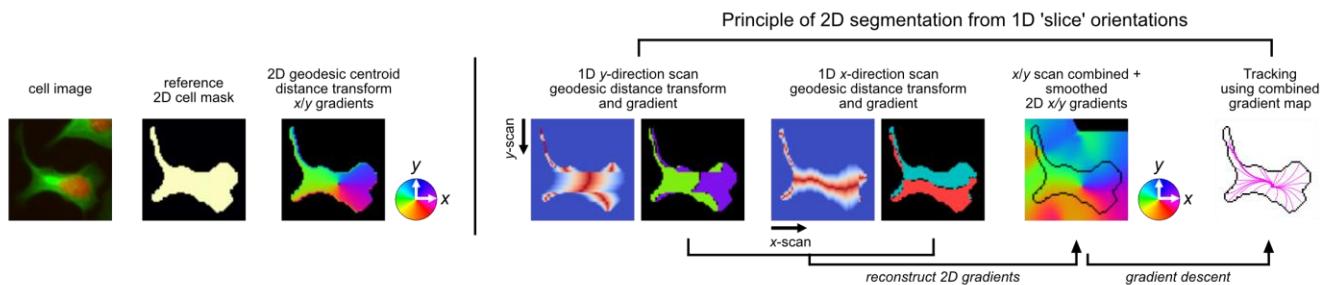
Extended Data Figure 4. Using image-based spatial connected components to robustly identify distinct spatially compact point sets. a) 2D Illustration of the image-based connected components spatial clustering approach in u-Segment3D involving left-to-right, rasterization of floating-point coordinates onto a discrete image pixel grid, building a count of the number of points in each pixel, approximate Gaussian kernel density estimation using a Gaussian filter of σ , binary thresholding on the mean density and subsequent connected component analysis to identify distinct spatial clusters. b) Verification of stable algorithm behavior by application to foreground (x,y) coordinates after propagation of XXX iterations with gradient descent. 2D gradients are computed using the 2D geodesic centroidal point distance transform. (i) Schematic of the process: cell image with reference outline, 2D geodesic centroidal distance transform x/y gradients, gradient descent on foreground (x,y) points for 50 iterations, resulting cluster points after XXX iterations. (ii) Time series of point density, connected labeling at (x,y), and labeling mapped back to (x^v, y^v) over 50 iterations. (iii) Two plots showing convergence: # predicted cells vs Iteration #, t, and F1 score vs Iteration #, t. c) Comparison of results for a complex reference mask (2D cell masks) over 20 iterations. The left column shows the reference mask and initial segmentation. Subsequent columns show the segmentation at t=0, t=5, t=10, t=15, and t=20 iterations. A plot on the right shows the F1 score increasing from ~0.4 to 1.0 over 50 iterations. d) Synthetic example showing a donut and a circle. The starting segmentation is a donut with two regions labeled 1 and 2. The gradients of individual segmentation are shown as a color map. The recovered segmentation is a donut with a hole, where regions 1 and 2 have been merged. e) Comparison to Cellpose 3D for a single cell. f) Comparison to Cellpose 3D for multiple cells. In all cases, the proposed method (Cellpose 3D mode outputs + uSegment3D) produces more accurate and complete segmentations compared to Cellpose 3D alone, especially for complex shapes and multiple cells."/>

608 experiment setup for an example 2D cell segmentation with densely touching cells (top panel). (ii) The point
609 density, connected component labelling at current (x,y) coordinates and labeling mapped back to initial (x,y)
610 coordinates (top-to-bottom) after t=0, 10, 20, 30, 40, 50 iterations of gradient descent (left-to-right). (iii) Plot
611 of number (#) of distinct predicted cells (left, black-colored y-axis and line) and # of matched cells with
612 reference cell segmentation (right, green-colored y-axis and line) with iteration #, t (top). Dashed black
613 horizontal line indicates the true cell number. Plot of F1 score of matching with reference cells (left, black-
614 colored y-axis and line) and the mean intersection-of-union (IoU) of matched cells with reference (right, green-
615 colored y-axis and line) with iteration #, t (bottom). **c)** Reference 2D cell segmentation of elongated touching
616 bacteria (left), identified unique cells by spatial connected component at gradient descent propagated
617 coordinates after $t = 0, 5, 10, 15, 20$ iterations (middle), and Plot of F1 score of matching with reference cells
618 (left, black-colored y-axis and line) and the mean intersection-of-union (IoU) of matched cells with reference
619 (right, green-colored y-axis and line) with iteration #, t . **d)** Image-based connected component applied to
620 recover a doughnut (region 1) surrounding a circle (region 2) after 50 iterations of gradient descent. **e)**
621 Comparison of using image-based connected component to construct robust segmentations from Cellpose
622 3D mode outputs for an isolated noisy single cell from the 3D cell tracking challenge. Left-to-right: 3D render
623 of raw volume, Cellpose 3D mode segmentation with diameter=50 and cellprob_threshold=0, reparsed 3D
624 segmentation using u-Segment3D gradient descent and connected component analysis (top row) and
625 corresponding mid x-y slice (bottom row). **f)** Comparison of using image-based connected component to
626 construct robust segmentations from Cellpose 3D mode outputs for a noisy image of multiple cells of
627 elongated morphologies from the 3D cell tracking challenge. Left-to-right: 3D render of raw volume, Cellpose
628 3D mode segmentation with diameter=15 and cellprob_threshold=-1.2, reparsed 3D segmentation using u-
629 Segment3D gradient descent and connected component analysis (top row) and corresponding mid x-y slice
630 (bottom row). Black arrowheads highlight examples of fractured single cells due to unstable spatial clustering
631 in Cellpose.

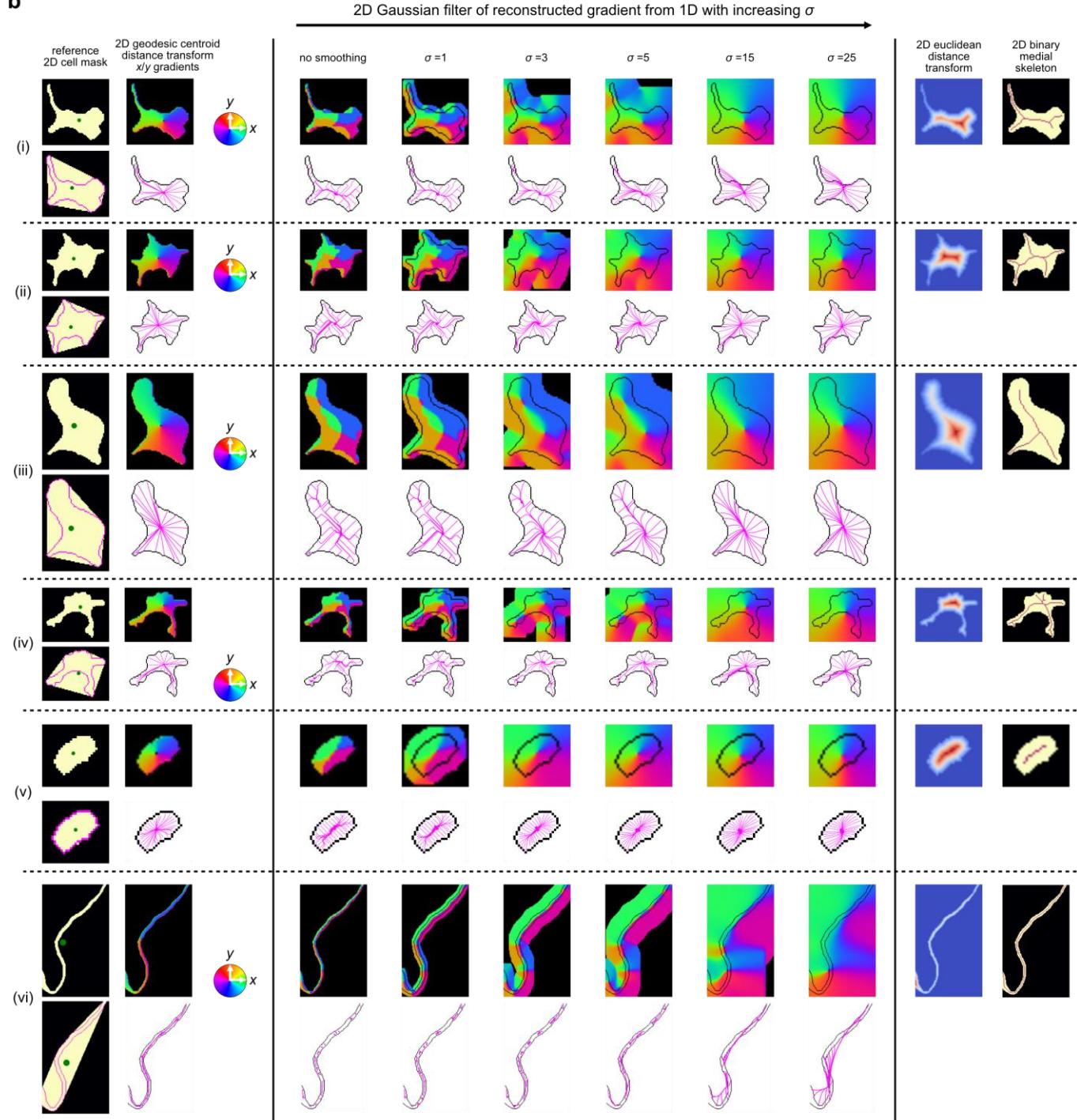
632

633

a



b



634
635
636
637
638
639
640
641
642

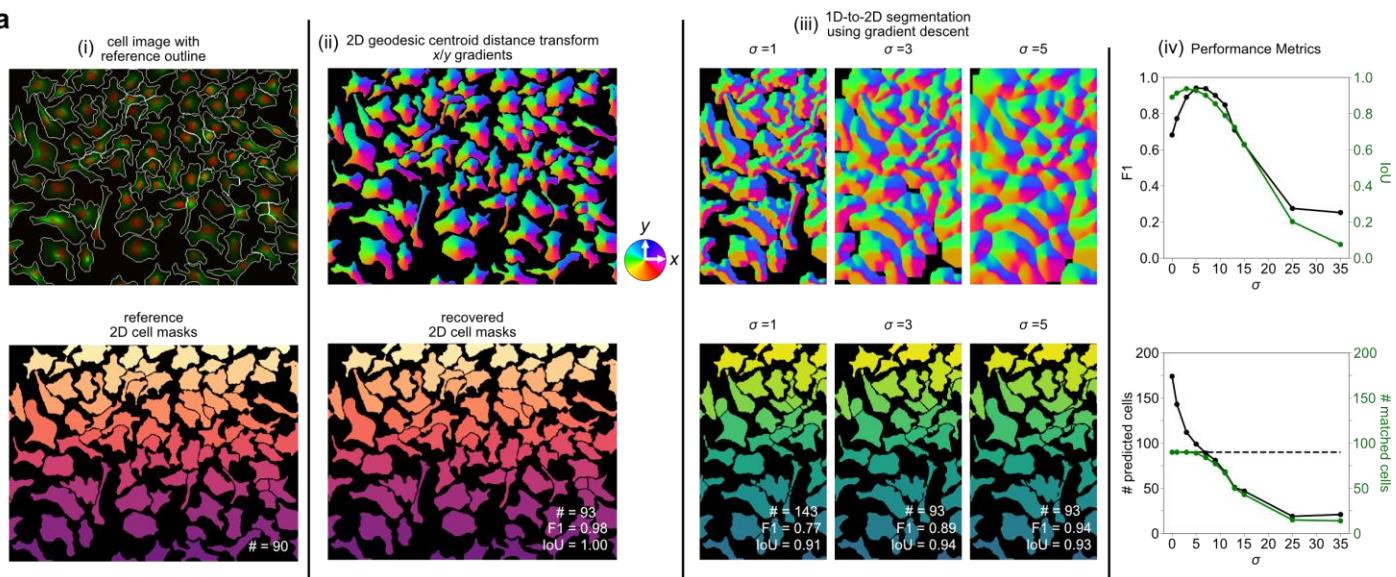
Extended Data Figure 5. 1D-to-2D segmentation for single cells from synthetic ideal 1D segmentation stacks. **a)** Schematic of the workflow to investigate 1D-to-2D segmentation. Reference cell segmentation and ideal 2D geodesic distance transform to reconstruct (left). Reconstruction of 2D gradient field from computed 1D gradients (right). Left-to-right: Determination of the y-direction and x-direction gradient by computing the distance transform of each pixel in 1D slices to the respective slice centroid and taking the gradient; combining the x- and y- gradient into a 2D gradient field and smoothing with a Gaussian σ ; performing gradient descent on the smoothed 2D gradients to propagate all interior points to a unique centroid (magenta line trajectories). **b)** Gradient descent behavior using Gaussian filter of increasing σ on the

643 initial reconstructed 2D gradients for individual cell examples representing the diversity of morphologies.
644 Example of approximately star-convex shapes: (i) Montage of 4 images depicting reference cell shape and
645 its convex hull image with green point representing their centroid coordinate and the cell's exact 2D geodesic
646 centroid distance transform with associated gradient descent trajectory to reconstruct (left panels). Observed
647 gradient descent trajectory (magenta line) when reconstructed 2D gradients is isotropically smoothed with
648 Gaussian filter of increasing σ (middle panels). Comparison of the gradient descent trajectory with the implicit
649 medial skeleton specified by the 2D Euclidean distance transform and explicit medial skeleton from
650 morphological operations (right panels). (ii), (iii) are further examples of approximate star-convex shapes as
651 in (i). (iv) represents a branching shape whose branch lengths are comparable to the cell body, (v) a convex
652 shape, and (vi) a thin vessel-like shape with panels depicted as for (i).

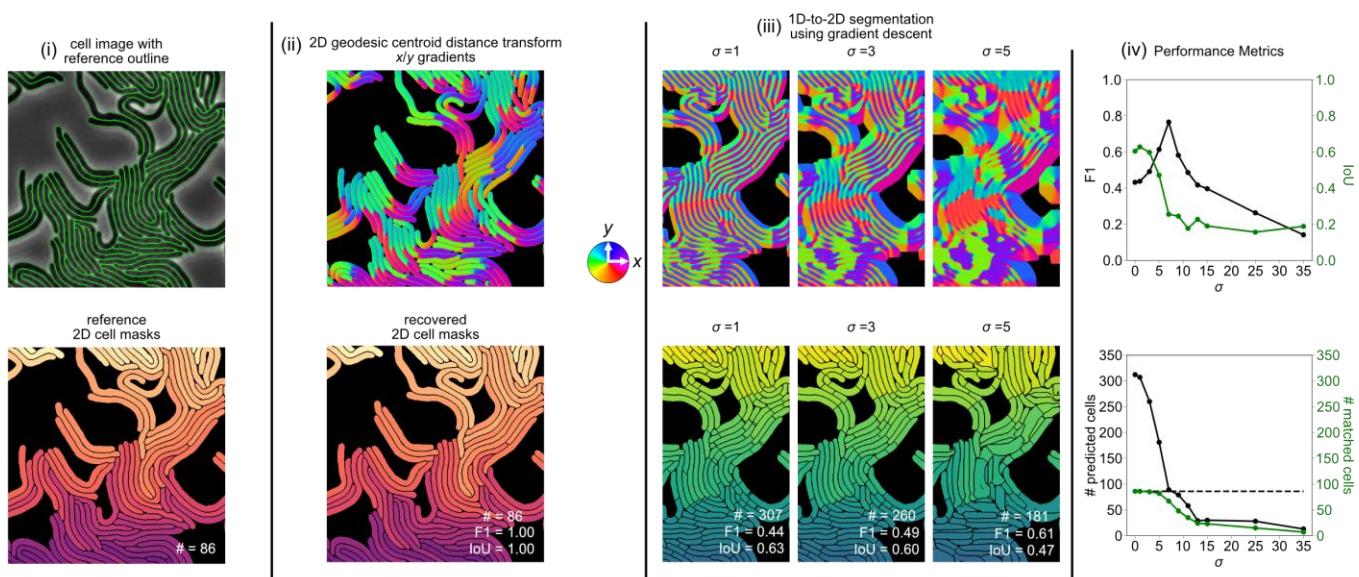
653

654

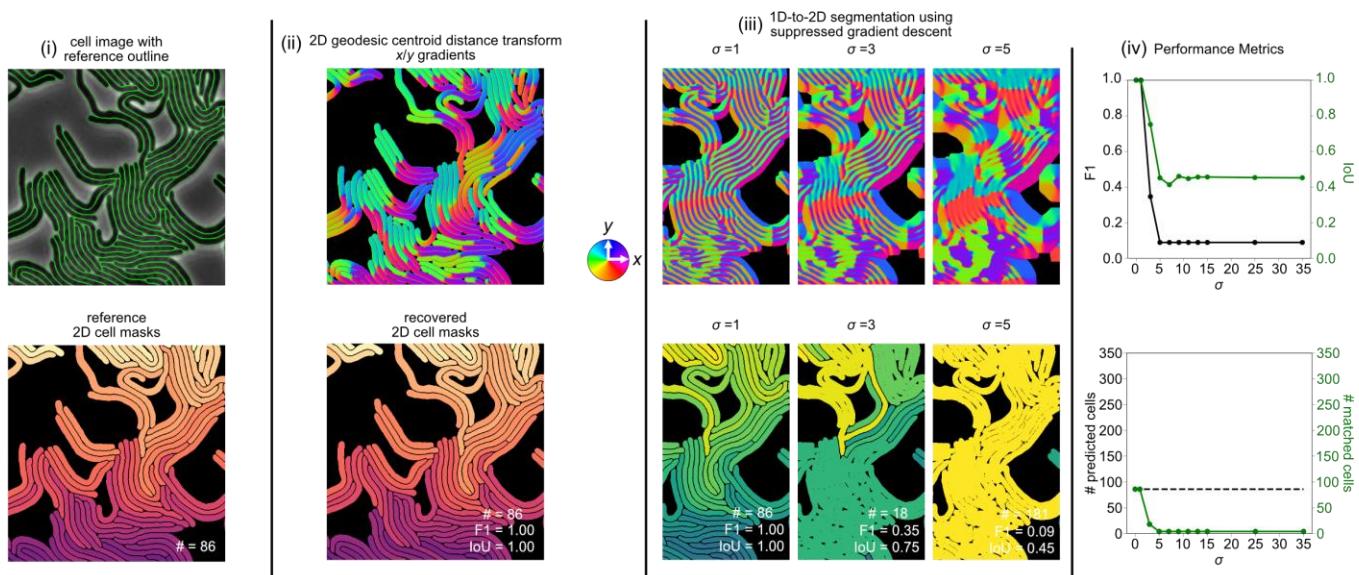
a



b



c



655
656
657
658
659
660
661
662

Extended Data Figure 6. 1D-to-2D segmentation for dense clustered cells from synthetic ideal 1D segmentation stacks. a) Representative example of a dense 2D cell culture with diverse morphologies with (i) reference cell boundaries overlaid and delineated in white (top) and uniquely colored segmented masks (bottom). (ii) Exact unit-normalized 2D gradients geodesic centroid distance transform colored by direction (top) and the recovered cell masks using connected component analysis after 100 iterations of gradient descent with step-size one pixel. (iii) Reconstructed 2D gradient from 1D after Gaussian filtering with increasing σ left-to-right (top) and corresponding recovered cell masks using connected component analysis

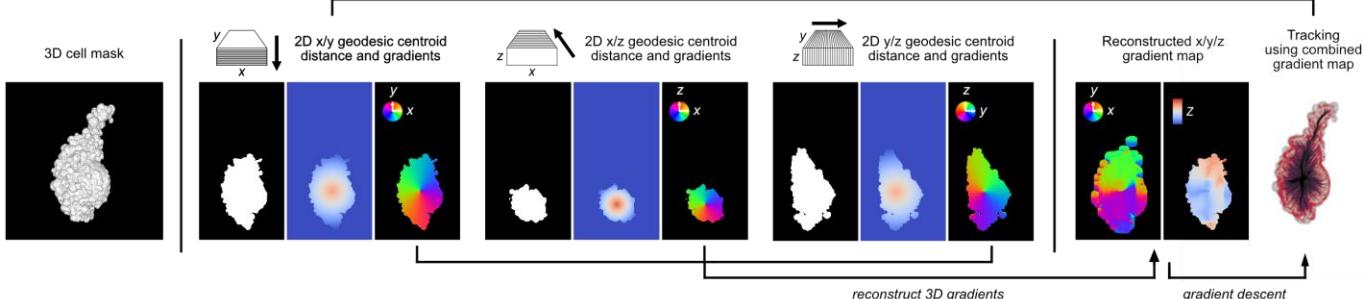
663 after 100 iterations of gradient descent. (iv) Recovery performance of the 1D-to-2D segmentation with
664 increasing σ . Plot of F1 score of matching with reference cells (left, black-colored y-axis and line) and the
665 mean intersection-of-union (IoU) of matched cells with reference (right, green-colored y-axis and line) (top).
666 Plot of the number (#) of distinct predicted cells (left, black-colored y-axis and line) and # of those that could
667 be matched with reference cells (right, green-colored y-axis and line) (bottom). Dashed black horizontal line
668 indicates the true cell number. Evaluation was performed after 100 iterations of gradient descent. **b)**
669 Representative example of a dense 2D cell culture where each cell has highly elongated, vessel-like
670 morphology. Panels (i)-(iv) as in **a)**. **c)** Same example image and panels (i)-(iv) as in **b)** but using suppressed
671 gradient descent where the step-size $\eta = \frac{1}{1+\tau \cdot t}$, is attenuated with gradient descent iteration number, t and
672 $\tau = 1$ (Methods).

673

674

a

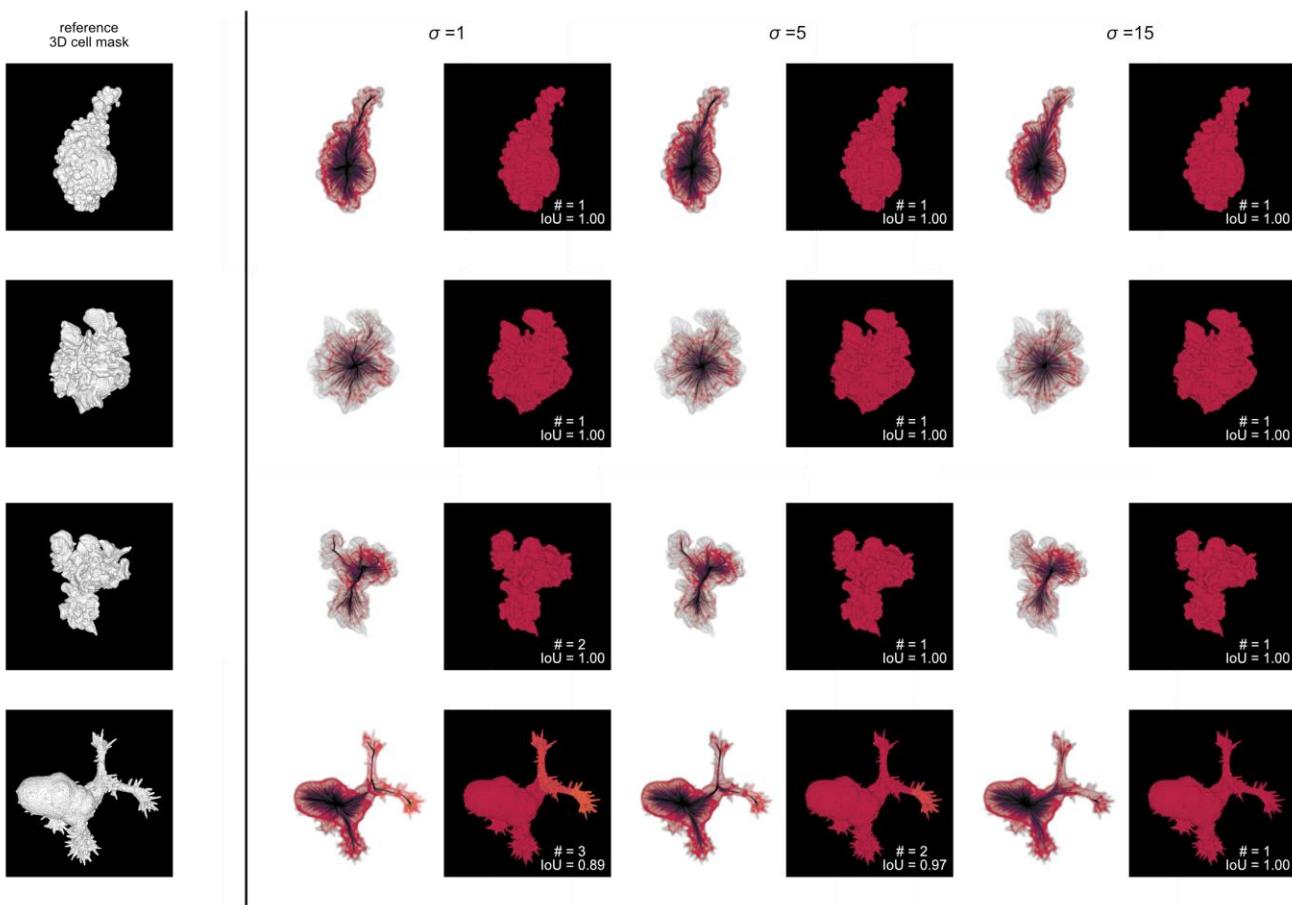
Principle of 3D segmentation from 2D 'slice' orientations



b

Using suppressed gradient descent ($\tau=0.1$), 200 iterations

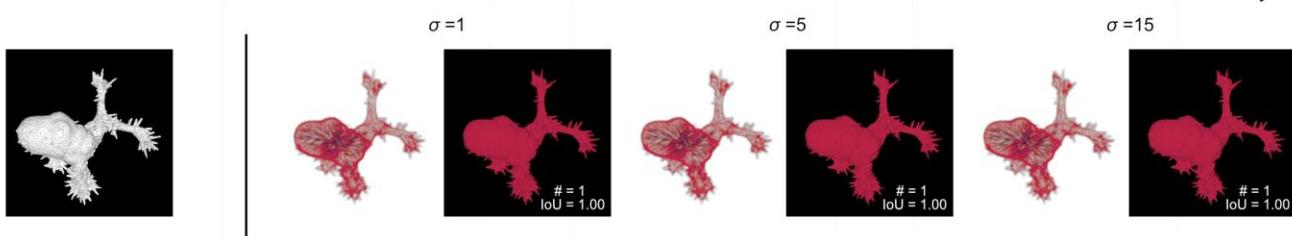
3D Gaussian filter of reconstructed gradient from 2D with increasing σ



C

Using suppressed gradient descent ($\tau=0.5$), 200 iterations

3D Gaussian filter of reconstructed gradient from 2D with increasing σ



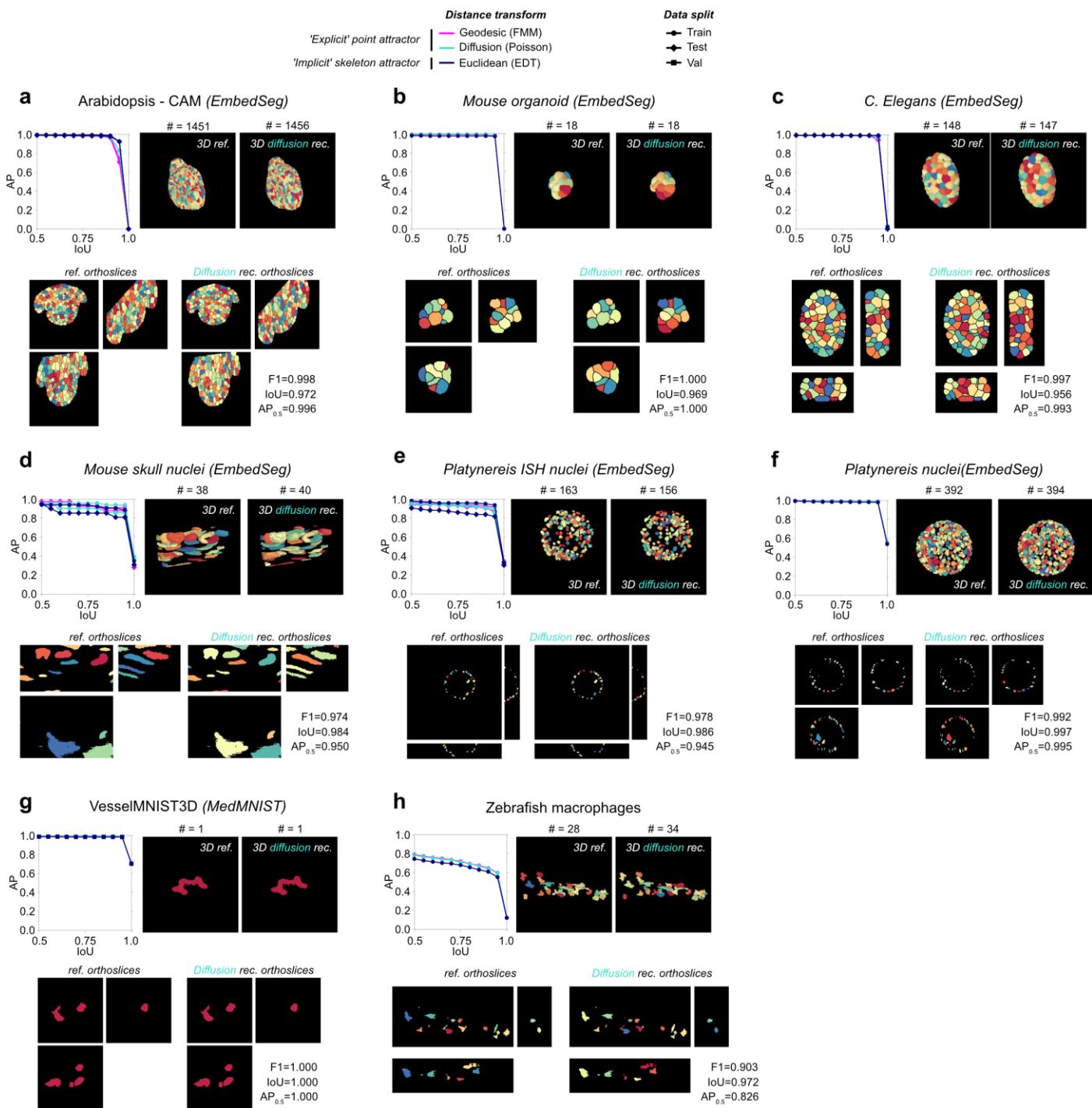
675
676
677
678
679
680
681
682
683

Extended Data Figure 7. 2D-to-3D segmentation reconstruction for single cells from synthetic ideal 2D segmentation stacks. **a)** Illustration of the reconstruction experiment given a single 3D cell segmentation (left), by generating 2D gradients slice-by-slice in xy, xz, yz views, treating each disconnected spatial component as a unique 2D cell (middle) and performing gradient descent on the reconstructed 3D xyz gradient followed by connected component analysis on the final advected 3D coordinates (right). **b)** Reconstruction examples for cells with blebs (1st row), lamellipodia (2nd, 3rd rows) and filopodia (4th row). In each row, left-to-right: reference binary 3D cell segmentation, the 3D gradient descent trajectory (left) and reconstructed 3D segmentation (right) for Gaussian filtering of the 3D reconstructed gradients with $\sigma = 1, 5, 15$,

684 using suppressed gradient descent ($\tau = 0.1$) with momentum (0.95). **c**). Reconstruction example of the same
685 cells with filopodia in b) for post- Gaussian filtering with $\sigma = 1,5,15$ (left-to-right), using suppressed gradient
686 descent with greater decay ($\tau = 0.5$) with momentum (0.95).

687

688



689
690 **Extended Data Figure 8. Reconstruction performance of 3D cell segmentation from synthetic ideal**
691 **2D segmentation stacks for real datasets.** Reconstruction performance measured by the mean average

692 precision curve (Methods) using three different 2D distance transforms for all datasets not included in Fig.

693 2d-f. **a)** Arabidopsis-CAM, **b)** mouse organoid, **c)** *C. Elegans*, **d)** mouse skull nuclei, **e)** *Platynereis* ISH nuclei,

694 **f)** *Platynereis* nuclei, **g)** vesselMNIST3D and **h)** zebrafish macrophages. For each dataset, top row, left-to-right:

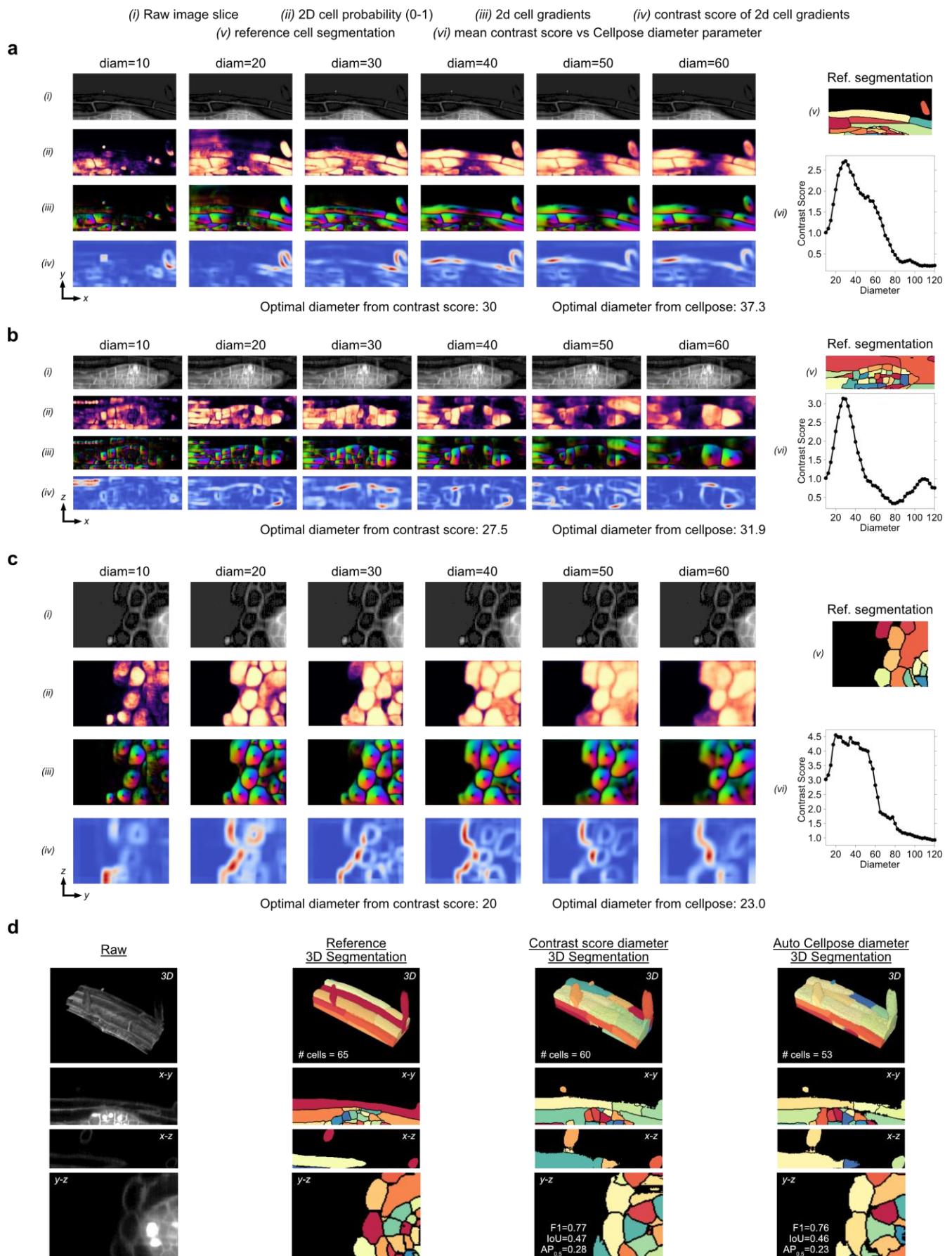
695 average precision vs intersection over union (IoU) curve; 3D rendering of reference segmentation, point-based

696 centroid diffusion distance transform reconstructed 3D cell segmentation. Bottom row, left-to-right, the

697 respective midplane orthoslices in the three orthogonal views. All available data splits is used for each dataset

698 except VesselMNIST3D which we use only the validation split (Methods). See Suppl. Table 1 for number of

699 objects and images in each split.



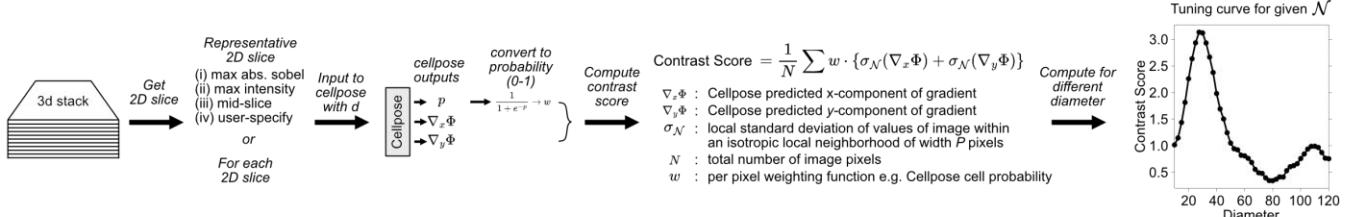
Extended Data Figure 9. The diameter parameter in pretrained Cellpose models should be individually set in orthogonal views. a) Cellpose ‘cyt2’ model outputs and per-pixel u-Segment3D contrast score on the most in-focus (Methods) 2D xy-slice of a Lateral Primordia as diameter (diam) is increased. (i) Raw input image slice, the same for all values of diameter. (ii) Normalized (0-1) Cellpose 2D pixel probability map colored black=0 to yellow=1. (iii) Unit-normalized Cellpose 2D predicted gradients colored by direction. (iv) Contrast score of predicted cellpose 2D gradient (Methods, Extended Data Figure 10a). (v) Corresponding reference cell segmentation for the raw 2D input. (vi) Mean contrast score averaged over the image for each

710 value of diam. The diameter with maximum contrast score is taken by u-Segment3D as the optimal diameter.
711 **b)** Same as in **a)** for the most in-focus 2D xz-slice and **c)** the most in-focus 2D yz-slice. **d)** From left-to-right,
712 3D render of the raw 3D volume, reference 3D segmentation, u-Segment3D consensus 3D segmentation
713 using the direct method and diameter in each orthoview set by contrast score or by the Cellpose ‘cyto2’ model
714 (top), and corresponding mid-plane orthoslices in all three views (below).

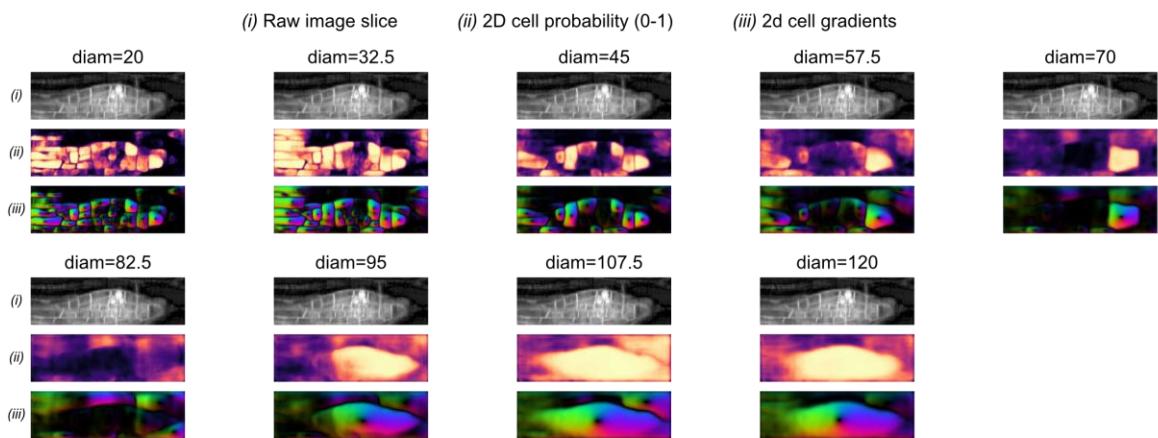
715

716

a Contrast score for tuning the diameter, d hyperparameter in Cellpose models



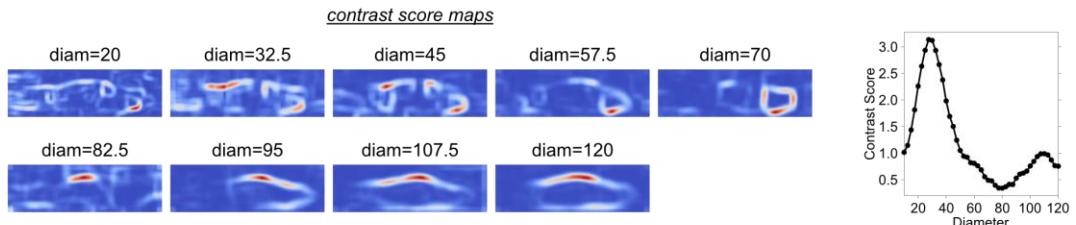
b



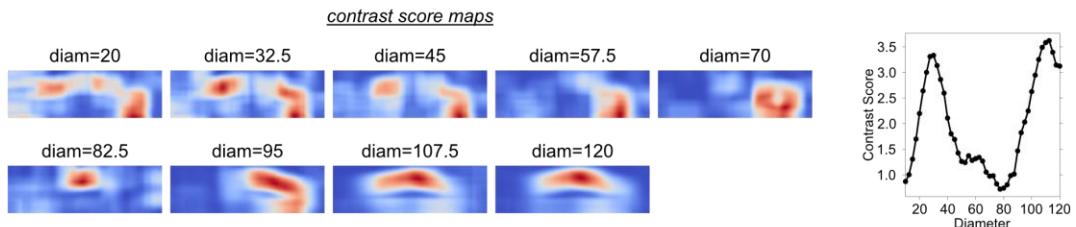
c

Biassing the contrast score maxima by adjusting local neighborhood size

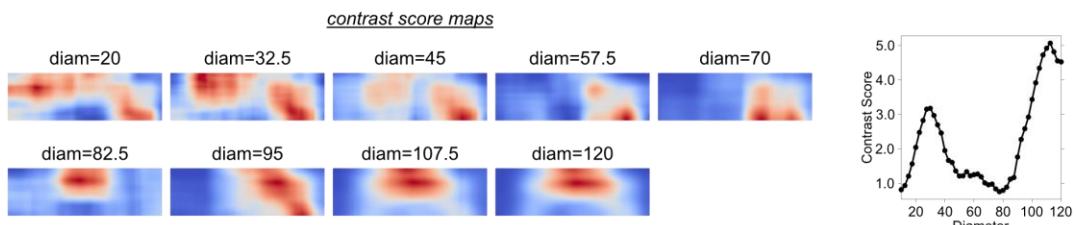
\mathcal{N} : 21 x 21 x 21 pixels neighborhood



\mathcal{N} : 61 x 61 x 61 pixels neighborhood

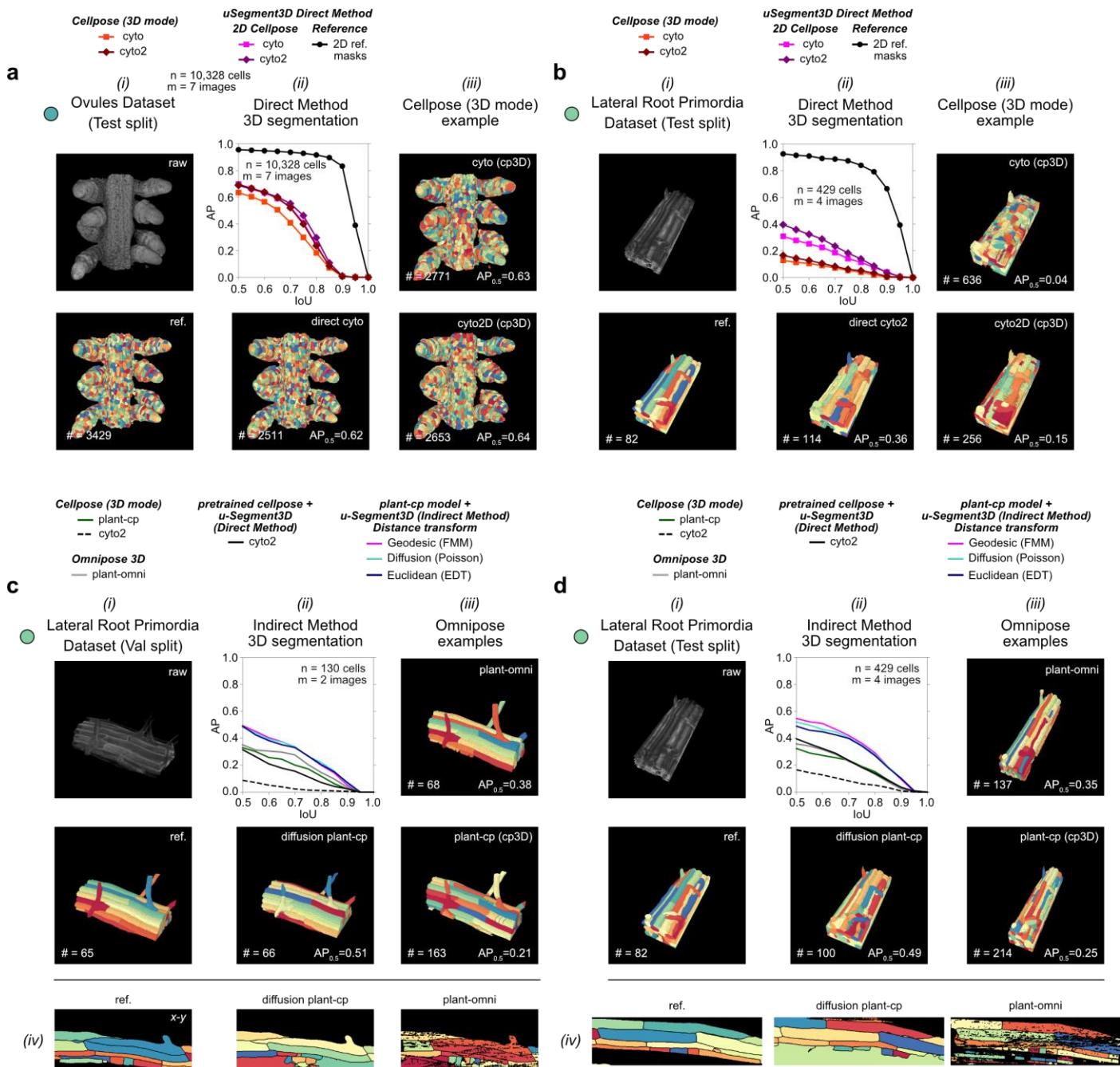


\mathcal{N} : 101 x 101 x 101 pixels neighborhood



717
718
719
720
721
722
723
724
725
726

Extended Data Figure 10. Semi-automatic determination of the diameter parameter in pretrained Cellpose models using local variance. a) Schematic illustration of the computation and definition of a contrast score function, evaluated based on a user-specified local neighborhood of width P pixels to evaluate Cellpose outputs for a 2D image when the diameter parameter is set to d . **b)** (i) Raw input image slice, (ii) normalized (0-1) Cellpose 2D pixel probability map colored black=0 to yellow=1, and (iii) unit-normalized Cellpose 2D predicted gradients colored by direction for 9 equisampled diameters in the range $d = [20, 120]$. **c)** Contrast score maps (colored blue-to-red for low-to-high values) for the same d as in b) (left) and resulting contrast score function and optimal diameter inferred (right), given a specified neighbourhood of width $P = 21, 61, 101$ pixels (top-to-bottom).



727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

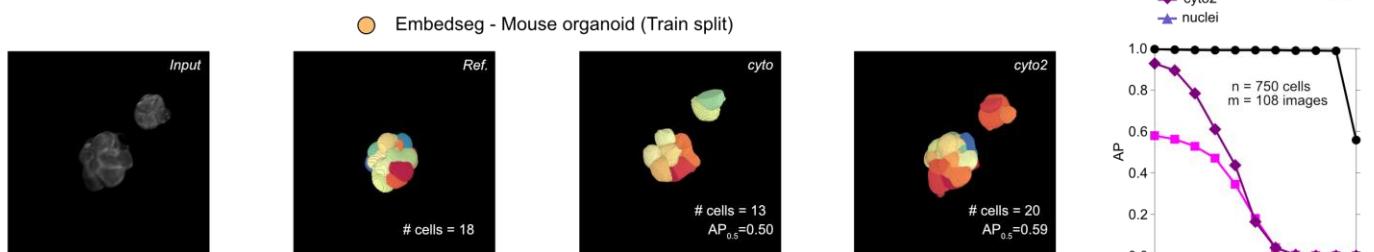
742

743

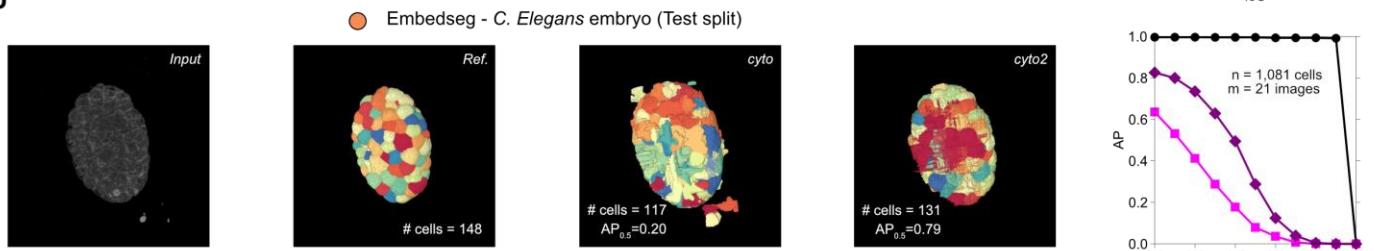
Extended Data Figure 11. Performance of u-Segment3D using pretrained vs specialized plant Cellpose segmentation models. **a)** Performance of pretrained Cellpose models with u-Segment3D vs Cellpose 3D mode on test split of Ovules. (i) example volume and corresponding reference segmentation; (ii) AP curves of all models, with best 3D segmentation reconstructed from ideal 2D slices (black line) (top) and segmentation of best pretrained model with u-Segment3D (bottom); (iii) Cellpose 3D mode segmentations using cyto (top) or cyto2 (bottom) models. **b)** Performance of pretrained Cellpose models with u-Segment3D vs Cellpose 3D mode on test split of Lateral Root Primordia (LRP) dataset. (i)-(iii) similar to a). **c)** Performance on val split using pretrained Cellpose models with u-Segment3D or Cellpose 3D mode, plant-cp: a specialized Cellpose 2D model trained on LRP with u-Segment3D or Cellpose 3D mode, and plant-omni: a specialized Omnipose 3D model trained on LRP natively in 3D. (i)-(ii) similar to a). (iii) Native 3D segmentation using plant-omni (top) or 2D-to-3D segmentation using Cellpose 3D mode with plant-cp (bottom). (iv) mid xy slice of the reference, u-Segment3D diffusion centroid transform aggregated plant-cp and 3D plant-omni segmentation (left-to-right). **d)** Same as c) for the test split of LRP.

uSegmented 3D cell segmentation from pretrained Cellpose 2D models (Direct Method)

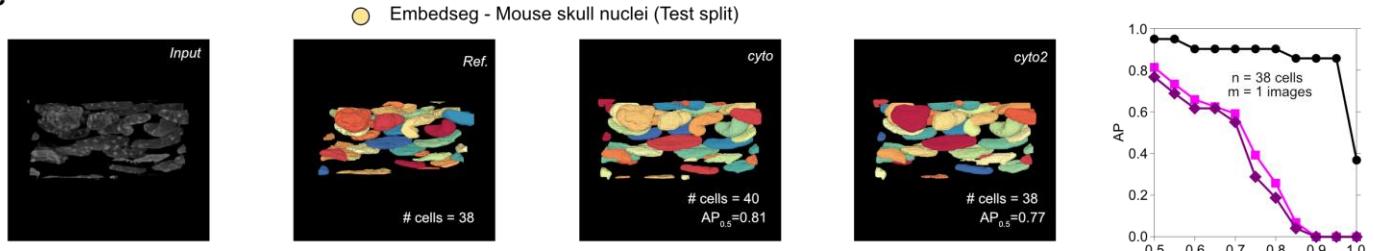
a



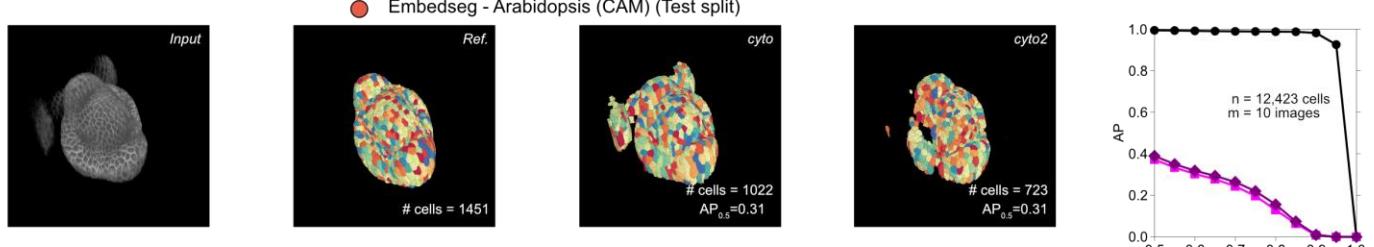
b



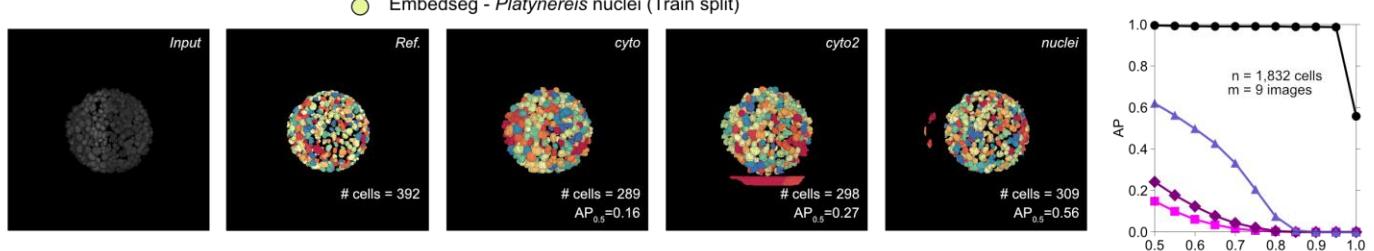
c



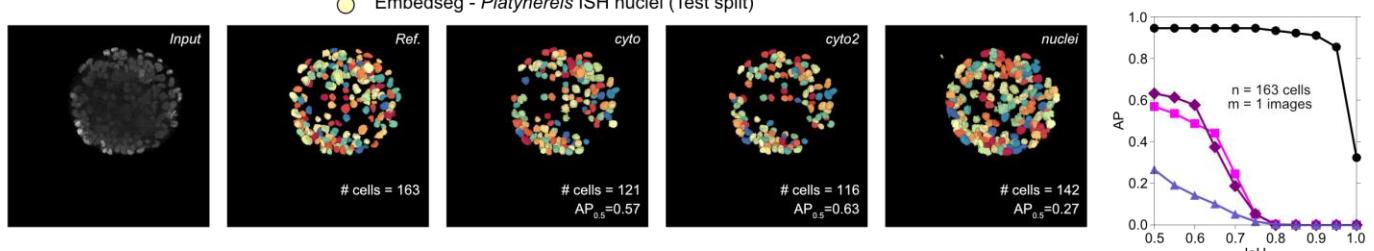
d



e



f



744

745

746

747

748

749

750

751

752

Extended Data Figure 12. Performance of 2D-to-3D segmentation for real datasets using u-Segment3D and pretrained cellpose2D model outputs. a) 3D cell segmentation performance of the mouse organoid (in the Embedseg paper) for the train data split, n=740 cells, m=108 volumes) using pretrained Cellpose 2D models with u-Segment3D and the direct method illustrated in Fig.3a. Left-to-right: 3D rendering of the raw volume, reference 3D segmentation, generated 3D segmentations for each Cellpose 2D model and the combined average precision (AP) curve coplotted with the AP curve of the best reconstruction with synthetic reference 2D segmentations in the three orthoviews (black line with circles). The same as a) for b) *C. elegans* embryo (test data split, n =1,081 cells, m = 21 images), c) mouse skull nuclei (test data split,

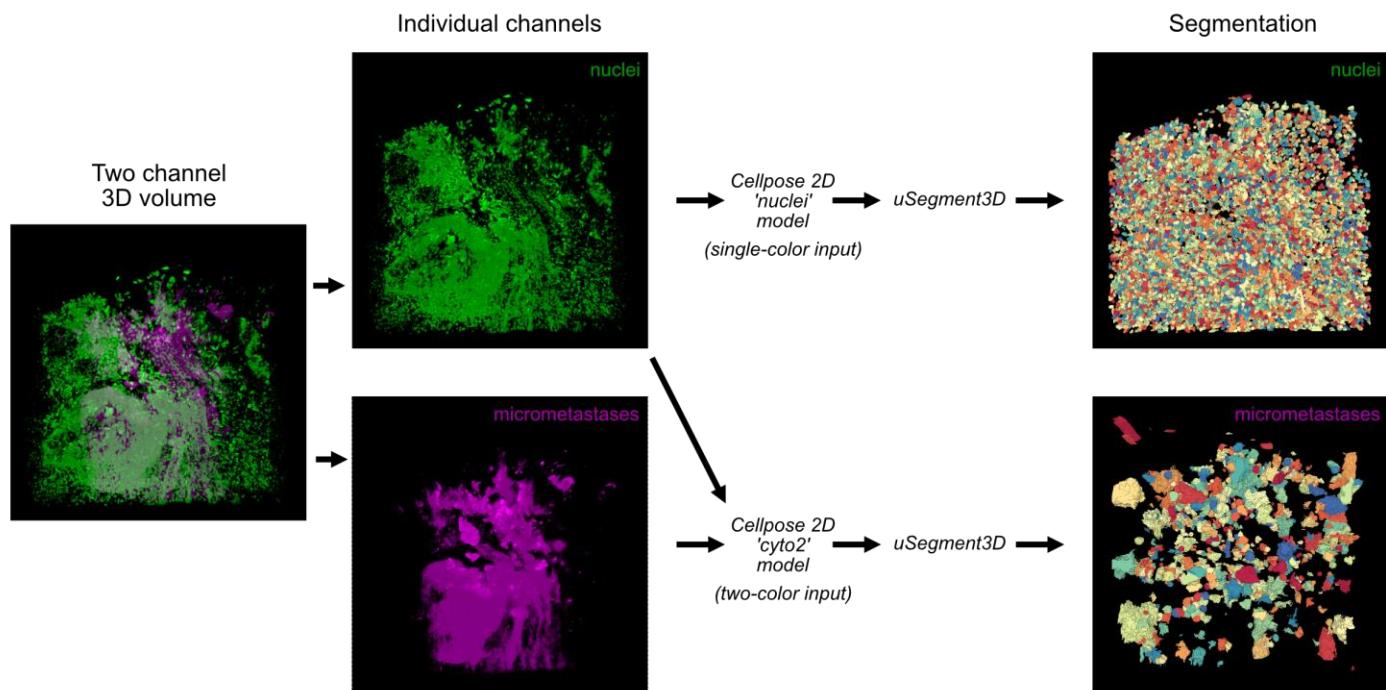
753 n=38 cells, m=1 image), **d)** Arabidopsis (CAM) (test data split, n=12,424 cells, m=10 images), **e)** Platynereis
754 nuclei (train data split, n=1,832 cells, m=9 images), **f)** Platynereis ISH nuclei (test data split, n=163 cells, m=1
755 image). For the Platynereis nuclei which are approximately spherical, we additionally evaluated the
756 performance of the Cellpose ‘nuclei’ 2D model (light purple line with triangles).

757

758

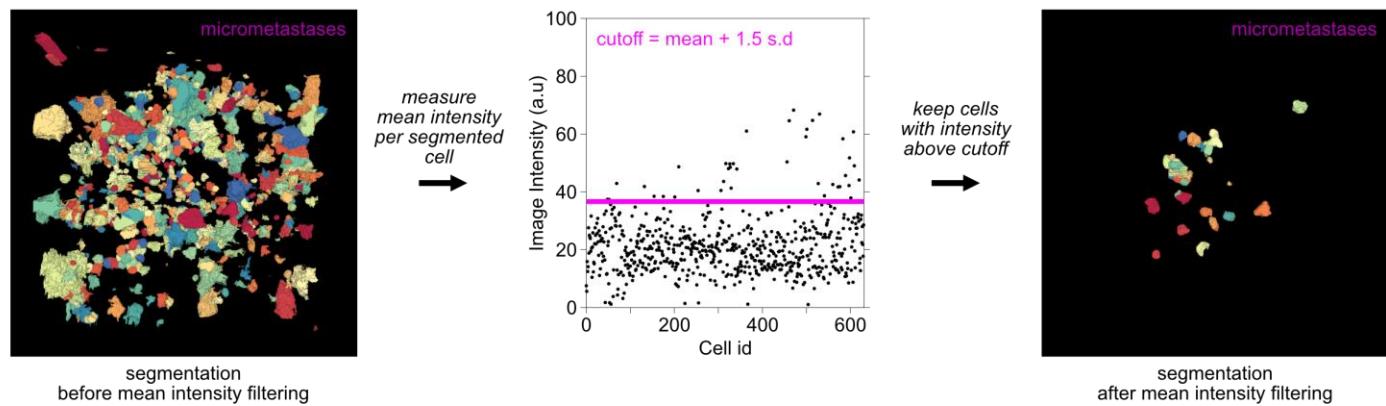
a

Workflow for nuclei and micrometastases cell segmentation



b

Filtering micrometastases cell segmentation after measuring image intensity statistics



759

Extended Data Figure 13: u-Segment3d generated consensus 3D segmentations can be filtered with image statistics to improve specificity for weakly labeled cells. a) Schematic of the workflow and 3D rendering of the resulting segmentation using the nuclei only stained channel (green) to segment all lung nuclei and using both nuclei (green) and micrometastases stain (magenta) channels to segment micrometastases with Cellpose and u-Segment3D. **b)** Schematic of the procedure to leverage the micrometastases stain to post-filter out mis-segmented micrometastases (left) by measuring the mean intensity in each segmented cell (middle) and performing a global statistically determined cutoff, keeping the subset of cell segmentations with mean intensity above the cutoff (right).

760
761
762
763
764
765
766
767
768

769 **Supplementary Tables**

770 **Supplementary Table 1. Summary of datasets used for validation of u-Segment3D.** Unless otherwise
771 noted, all images were first resampled to isotropic voxel resolution by appropriate downsampling of xy slices
772 and then resized isotropically by the indicated rescaling factor.

773

774 *Plants*

Name	Description	Pixel Size (Z,Y,X) [μm^3]	Used Microscope	Isotropic Rescaling factor	Total #images (Train/Test/Val)	Total #cells
Lateral Primordia ⁸	<i>Arabidopsis thaliana</i> lateral root of the line sC111 were used at 5 day post germination	(0.25, 0.1625, 0.1625)	Multi-view Selective Plane Illumination Microscopy	0.5	21 / 4 / 2	2,068 / 429 / 130
Ovules ^{8,132}	<i>Arabidopsis thaliana</i> ovules stained with SR2200 and TO-PRO-3 iodide	(0.235, 0.075, 0.075) *	Confocal laser scanning microscopy	1	22 / 7 / 2	23,860 / 10,328 / 2,839

775 * We find (0.235, 0.15, 0.15) to be more isotropic and expected appearance.

776 Lateral primordia was downloaded from <https://osf.io/2rszy/>

777 Ovules was downloaded from <https://osf.io/w38uf/>

778

779 *Embedseg*

Name	Description	Pixel Size (Z,Y,X) [μm^3]	Used Microscope	Isotropic Rescaling factor	# images (Train/Test/Val)	Total # cells
Arabidopsis-Cells-CAM-small ^{133,134}	<i>Arabidopsis Thaliana</i> YFP membrane labelled and imaged between 24 and 28 days after germination	(0.26, 0.22, 0.22)	Confocal Microscopy	1	11 / 10 / -	12,016 / 12,423 / -
C.elegans-Cells-HK ⁵⁶	C.elegans embryos membrane labeled.	(0.25, 0.25, 0.25)	Confocal Microscopy	1	54 / 21 / -	3,479 / 1,081 / -
Mouse-Organoid-Cells-CBG ¹⁷	Mouse Embryonic Stem Cells, R1 cell line, labeled membrane	(1.0, 0.1733, 0.1733)	Selective Plane Illumination Microscopy	0.5	108 / - / -	750 / - / -
Mouse-Skull-Nuclei-CBG ¹⁷	Nuclei of the skull region of developing mouse embryos, labeled with DAPI	(0.200, 0.073, 0.073) *	Inverted Zeiss LSM 880 Microscope	0.5	2 / 1 / -	150 / 38 / -
Platyneris-Nuclei-CBG ¹⁷	Nuclei of whole-mount <i>Platyneris dumerilli</i> specimens at stages between 0 to 16 hours post fertilization, injected with a fluorescent nuclear tracer	(2.031, 0.406, 0.406)	Simultaneous Multi-view Light-Sheet Microscopy	1	9 / - / -	1832 / - / -

Platyneris-ISH-Nuclei-CBG ¹⁷	Nuclei of whole-mount <i>Platynereis dumerilli</i> specimens at stage of 16 hours post fertilization, labeled with DAPI	(0.45, 0.45, 0.45)	Laser Scanning Confocal Microscopy	1	2 / 1 / -	486 / 159 / -
---	---	--------------------	------------------------------------	---	-----------	---------------

* For Mouse-Skull-Nuclei-CBG we found that pretrained Cellpose 2D models and u-Segment3D performed better without resizing to isotropic voxels and applying isotropic rescaling factor.

All of these 3D datasets were downloaded from <https://github.com/juglab/EmbedSeg/releases/tag/v0.1.0>

VesselMNIST3D

Name	Description	Pixel Size (Z,Y,X)	Used Microscope	Isotropic Rescaling factor	# images (Train/Test/Val)	Total # cells
VesselMNIST3D ^{135,136}	Derived from the IntrA dataset. Vessel segments were generated from 103 3D meshes of entire brain vessels.	(1,1,1)	Time-of-Flight Magnetic Resonance Angiography (TOF-MRA)	9.14	1,335/382/192 (*191)	1,335 / 382 / 191^

¹³⁶We only used the val split as all images were single-component and morphological properties similar across data splits.

¹³⁵* 1 of the images in the val split was found to be blank.

The dataset was downloaded as .npz from <https://zenodo.org/records/10519652>.

DeepVesselNet

Name	Description	Pixel Size (Z,Y,X)	Used Microscope	Isotropic Rescaling factor	# images (Train/Test/Val)	Total # cells
DeepVesselNet ^{10,137,138}	Synthetic generated dataset with images of 325 x 304 x 600. Vessel intensities were randomly chosen in the interval [128, 255] and non-vessel intensities from the interval [0 – 100]. Gaussian noise was then randomly applied.	(1,1,1)	-	1	136 (135*) / - / -	410* / - / -

792	* We could not download and unzip 1 of the raw images in the archive.														
793	The dataset was downloaded from https://github.com/gieseckow/deepvesselnet/wiki/Datasets														
794															
795	<i>Zebrafish macrophages</i>														
	<table border="1"><thead><tr><th>Name</th><th>Description</th><th>Pixel Size (Z,Y,X) [μm^3]</th><th>Used Microscope</th><th>Isotropic Rescaling factor</th><th># images (Train/Test/Val)</th><th>Total # cells</th></tr></thead><tbody><tr><td>Zebrafish-macrophages⁸³</td><td>Human U-2 OS osteosarcoma cancer cells, labelled with pVimentin-PsmOrange, were injected into zebrafish larvae with fluorescent macrophages, labelled with Tg(mpeg1:EGFP). Zebrafish were selected for imaging two hours after injection.</td><td>(0.4, 0.117, 0.117)</td><td>Light-Sheet microscopy</td><td>1</td><td>50 / - / -</td><td>1,108 / - / -</td></tr></tbody></table>	Name	Description	Pixel Size (Z,Y,X) [μm^3]	Used Microscope	Isotropic Rescaling factor	# images (Train/Test/Val)	Total # cells	Zebrafish-macrophages ⁸³	Human U-2 OS osteosarcoma cancer cells, labelled with pVimentin-PsmOrange, were injected into zebrafish larvae with fluorescent macrophages, labelled with Tg(mpeg1:EGFP). Zebrafish were selected for imaging two hours after injection.	(0.4, 0.117, 0.117)	Light-Sheet microscopy	1	50 / - / -	1,108 / - / -
Name	Description	Pixel Size (Z,Y,X) [μm^3]	Used Microscope	Isotropic Rescaling factor	# images (Train/Test/Val)	Total # cells									
Zebrafish-macrophages ⁸³	Human U-2 OS osteosarcoma cancer cells, labelled with pVimentin-PsmOrange, were injected into zebrafish larvae with fluorescent macrophages, labelled with Tg(mpeg1:EGFP). Zebrafish were selected for imaging two hours after injection.	(0.4, 0.117, 0.117)	Light-Sheet microscopy	1	50 / - / -	1,108 / - / -									
796	This dataset was curated in-house with the assistance of Cellpose 'cyto2' 2D model and u-Segment3D.														
797															
798	Supplementary Table 2. u-Segment3D settings and parameters for testing the reconstruction of 3D segmentation from ideal 2D segmentations with public datasets.														
799															
800															
801	Supplementary Table 3. u-Segment3D settings and parameters for consensus 3D segmentation from the output of pretrained Cellpose 2D predictions with public datasets.														
802															
803															
804	Supplementary Table 4. Cellpose2D and u-Segment3D settings and parameters for consensus 3D segmentation on additional demonstration datasets.														
805															
806															

- 807 **Supplementary Movies**
- 808 **Supplementary Movie 1. u-Segment3D enables consensus 3D segmentation from 2D segmented**
809 **stacks**
- 810 **Supplementary Movie 2. Gradient descent dynamics of foreground cell coordinates using different**
811 **2D transforms.**
- 812 **Supplementary Movie 3. Comparison of the spatial proximity clustering used by Cellpose 3D mode**
813 **and u-Segment3D's image-based connected component analysis on noisy cell tracking challenge**
814 **datasets.**
- 815 **Supplementary Movie 4. Gradient descent dynamics of foreground cell coordinates of 3D single cells**
816 **under 2D reconstructed 3D gradients.**
- 817 **Supplementary Movie 5. Gradient descent dynamics during 3D reconstruction of ovules, lateral root**
818 **primordial and vasculature from ideal 2D segmented stacks.**
- 819 **Supplementary Movie 6. Segmentation of a movie of thin MDA231 human breast carcinoma cells embedded**
820 **in collagen from the 3D cell tracking challenge using u-Segment3D to aggregate only 2D xy segmentations only.**
- 821 **Supplementary Movie 7. Segmentation of unwrapped surface cells of a Drosophila embryo over time using**
822 **u-Segment3D to aggregate only 2D xy segmentations only.**
- 823 **Supplementary Movie 8. u-Segment3D postprocessing enables recovery of missing surface protrusions in**
824 **the 3D segmentation of a HBEC cell aggregate.**
- 825 **Supplementary Movie 9. u-Segment3D postprocessing enables the segmentation of vessel sprouting in**
826 **zebrafish.**
- 827 **Supplementary Movie 10. u-Segment3D segmentation of all cells in a $\approx 35\mu\text{m} \times 1.5\text{mm} \times 1.5\text{mm}$**
828 **CYCIF multiplexed tissue section of metastatic melanoma.**
- 829 **Supplementary Movie 11. u-Segment3D segmentation enabled detection of weakly fluorescent lung**
830 **micrometastases in cleared tissues.**
- 831 **Supplementary Movie 12. u-Segment3D segmentation of heterogeneous cell structures in brain**
832 **tissue labelled using coCATs.**
- 833
- 834
- 835
- 836
- 837
- 838
- 839
- 840
- 841
- 842

843

References

- 844 1 Ma, J. *et al.* Segment anything in medical images. *Nature Communications* **15**, 654 (2024).
- 845 2 Ma, J. *et al.* The multi-modality cell segmentation challenge: Towards universal solutions. *arXiv preprint arXiv:2308.05864* (2023).
- 846 3 Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II* **11**. 265–273 (Springer).
- 847 4 Graham, S. *et al.* CoNIC Challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting. *Medical image analysis* **92**, 103047 (2024).
- 848 5 Müller, A. *et al.* Modular segmentation, spatial analysis and visualization of volume electron microscopy datasets. *Nature Protocols*, 1–31 (2024).
- 849 6 Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods* **18**, 100–106 (2021).
- 850 7 Cutler, K. J. *et al.* Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nature methods* **19**, 1438–1448 (2022).
- 851 8 Wolny, A. *et al.* Accurate and versatile 3D segmentation of plant tissues at cellular resolution. *Elife* **9**, e57613 (2020).
- 852 9 Huang, Y. *et al.* Segment anything model for medical images? *Medical Image Analysis* **92**, 103061 (2024).
- 853 10 Tetteh, G. *et al.* Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *Frontiers in Neuroscience* **14**, 1285 (2020).
- 854 11 Segal, D. *et al.* In vivo 3D profiling of site-specific human cancer cell morphotypes in zebrafish. *Journal of Cell Biology* **221**, e202109100 (2022).
- 855 12 Lin, J.-R. *et al.* Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell* **186**, 363–381. e319 (2023).
- 856 13 Haghghi, M., Caicedo, J. C., Cimini, B. A., Carpenter, A. E. & Singh, S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nature methods* **19**, 1550–1557 (2022).
- 857 14 Heemskerk, I. & Streichan, S. J. Tissue cartography: compressing bio-image data by dimensional reduction. *Nature methods* **12**, 1139–1142 (2015).
- 858 15 Zhou, F. Y. *et al.* Surface-guided computing to analyze subcellular morphology and membrane-associated signals in 3D. *bioRxiv*, 2023.2004. 2012.536640 (2023).
- 859 16 Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**, 1–11 (2006).
- 860 17 Lalit, M., Tomancak, P. & Jug, F. Embedseg: Embedding-based instance segmentation for biomedical microscopy data. *Medical image analysis* **81**, 102523 (2022).
- 861 18 Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods* **15**, 1090–1097 (2018).
- 862 19 Archit, A. *et al.* Segment anything for microscopy. *bioRxiv*, 2023.2008. 2021.554208 (2023).
- 863 20 Israel, U. *et al.* A Foundation Model for Cell Segmentation. *bioRxiv* (2023).
- 864 21 Welf, E. S. *et al.* Quantitative Multiscale Cell Imaging in Controlled 3D Microenvironments. *Dev Cell* **36**, 462–475, doi:10.1016/j.devcel.2016.01.022 (2016).
- 865 22 Weems, A. D. *et al.* Blebs Promote Cell Survival by Assembling Oncogenic Signaling Hubs. *bioRxiv*, 2021.2004.2023.441200, doi:10.1101/2021.04.23.441200 (2021).
- 866 23 Drost, J. & Clevers, H. Organoids in cancer research. *Nature Reviews Cancer* **18**, 407–418 (2018).
- 867 24 Fatehullah, A., Tan, S. H. & Barker, N. Organoids as an in vitro model of human development and disease. *Nature cell biology* **18**, 246–254 (2016).
- 868 25 Yapp, C. *et al.* Multiplexed 3D Analysis of Cell Plasticity and Immune Niches in Melanoma. *bioRxiv* (2023).
- 869 26 Dutta, A. & Zisserman, A. in *Proceedings of the 27th ACM international conference on multimedia*. 2276–2279.
- 870 27 Torralba, A., Russell, B. C. & Yuen, J. Labelme: Online image annotation and applications. *Proceedings of the IEEE* **98**, 1467–1484 (2010).
- 871 28 Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology* **40**, 555–565 (2022).
- 872 29 van Ineveld, R. L. *et al.* Revealing the spatio-phenotypic patterning of cells in healthy and tumor tissues with mLSR-3D and STAPL-3D. *Nature biotechnology* **39**, 1239–1245 (2021).
- 873 30 Maitin-Shepard, J. Neuroglancer. *github. com/google/neuroglancer*, Retrieved, 04-30 (2021).

- 899 31 Boergens, K. M. *et al.* webKnossos: efficient online 3D data annotation for connectomics. *nature methods* **14**,
900 691-694 (2017).
- 901 32 Hanslovsky, P. *et al.* (September, 2019).
- 902 33 Yushkevich, P. A., Gao, Y. & Gerig, G. in *2016 38th annual international conference of the IEEE engineering in*
903 *medicine and biology society (EMBC)*. 3342-3345 (IEEE).
- 904 34 Januszewski, M. *et al.* High-precision automated reconstruction of neurons with flood-filling networks.
905 *Nature methods* **15**, 605-610 (2018).
- 906 35 Tasnadi, E. A. *et al.* 3D-Cell-Annotator: an open-source active surface tool for single-cell segmentation in 3D
907 microscopy images. *Bioinformatics* **36**, 2948-2949 (2020).
- 908 36 Schmidt, M., Motta, A., Sievers, M. & Helmstaedter, M. RoboEM: automated 3D flight tracing for synaptic-
909 resolution connectomics. *Nature Methods*, 1-6 (2024).
- 910 37 Shapson-Coe, A. *et al.* A connectomic study of a petascale fragment of human cerebral cortex. *BioRxiv*,
911 2021.2005. 2029.446289 (2021).
- 912 38 Dorkenwald, S. *et al.* Binary and analog variation of synapses between cortical pyramidal neurons. *Elife* **11**,
913 e76120 (2022).
- 914 39 Consortium, M. *et al.* Functional connectomics spanning multiple areas of mouse visual cortex. *BioRxiv*,
915 2021.2007. 2028.454025 (2021).
- 916 40 Zheng, Z. *et al.* A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell*
917 **174**, 730-743. e722 (2018).
- 918 41 Chen, J. *et al.* The Allen Cell and Structure Segmenter: a new open source toolkit for segmenting 3D
919 intracellular structures in fluorescence microscopy images. *BioRxiv*, 491035 (2018).
- 920 42 Vicar, T. *et al.* Cell segmentation methods for label-free contrast microscopy: review and comprehensive
921 comparison. *BMC bioinformatics* **20**, 1-25 (2019).
- 922 43 Wiesmann, V. *et al.* Using simulated fluorescence cell micrographs for the evaluation of cell image
923 segmentation algorithms. *BMC bioinformatics* **18**, 1-12 (2017).
- 924 44 Wang, A. *et al.* A novel deep learning-based 3D cell segmentation framework for future image-based disease
925 detection. *Scientific reports* **12**, 342 (2022).
- 926 45 Li, G. *et al.* 3D cell nuclei segmentation based on gradient flow tracking. *BMC cell biology* **8**, 1-10 (2007).
- 927 46 Marzec, M., Piórkowski, A. & Gertych, A. Efficient automatic 3D segmentation of cell nuclei for high-content
928 screening. *BMC bioinformatics* **23**, 203 (2022).
- 929 47 Wu, T.-C., Wang, X., Li, L., Bu, Y. & Umulis, D. M. Automatic wavelet-based 3D nuclei segmentation and
930 analysis for multicellular embryo quantification. *Scientific reports* **11**, 9847 (2021).
- 931 48 McDole, K. *et al.* In toto imaging and reconstruction of post-implantation mouse development at the single-
932 cell level. *Cell* **175**, 859-876. e833 (2018).
- 933 49 Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. in *Proceedings of the IEEE/CVF winter*
934 *conference on applications of computer vision*. 3666-3673.
- 935 50 Wen, C. *et al.* 3DeeCellTracker, a deep learning-based pipeline for segmenting and tracking cells in 3D time
936 lapse images. *Elife* **10**, e59187 (2021).
- 937 51 Dunn, K. W. *et al.* DeepSynth: Three-dimensional nuclear segmentation of biological images using neural
938 networks trained with synthetic data. *Scientific reports* **9**, 18295 (2019).
- 939 52 Chen, H. & Murphy, R. F. 3DCellComposer-A Versatile Pipeline Utilizing 2D Cell Segmentation Methods for 3D
940 Cell Segmentation. *bioRxiv*, 2024.2003. 2008.584082 (2024).
- 941 53 Kar, A. *et al.* Benchmarking of deep learning algorithms for 3D instance segmentation of confocal image
942 datasets. *PLoS computational biology* **18**, e1009879 (2022).
- 943 54 Eschweiler, D., Rethwisch, M., Koppers, S. & Stegmaier, J. in *2021 IEEE 18th International Symposium on*
944 *Biomedical Imaging (ISBI)*. 792-796 (IEEE).
- 945 55 Smith, M. B. *et al.* Active mesh and neural network pipeline for cell aggregate segmentation. *Biophysical*
946 *Journal* **122**, 1586-1599 (2023).
- 947 56 Cao, J. *et al.* Establishment of a morphological atlas of the *Caenorhabditis elegans* embryo using deep-
948 learning-based 4D segmentation. *Nature communications* **11**, 6254 (2020).
- 949 57 Hardo, G., Noka, M. & Bakshi, S. Synthetic Micrographs of Bacteria (SyMBac) allows accurate segmentation
950 of bacterial cells using deep neural networks. *BMC biology* **20**, 263 (2022).
- 951 58 Wu, L. *et al.* Nisnet3d: Three-dimensional nuclear synthesis and instance segmentation for fluorescence
952 microscopy images. *Scientific Reports* **13**, 9533 (2023).
- 953 59 Martinez, N., Sapiro, G., Tannenbaum, A., Hollmann, T. J. & Nadeem, S. Impartial: Partial annotations for cell
954 instance segmentation. *bioRxiv*, 2021.2001. 2020.427458 (2021).

- 955 60 Eschweiler, D., Rethwisch, M., Jarchow, M., Koppers, S. & Stegmaier, J. 3D fluorescence microscopy data
956 synthesis for segmentation and benchmarking. *Plos one* **16**, e0260509 (2021).
957 61 Eschweiler, D. *et al.* Denoising diffusion probabilistic models for generation of realistic fully-annotated
958 microscopy image datasets. *PLOS Computational Biology* **20**, e1011890 (2024).
959 62 Chai, B., Efstathiou, C., Yue, H. & Draviam, V. M. Opportunities and challenges for deep learning in cell
960 dynamics research. *Trends in Cell Biology* (2023).
961 63 Kirillov, A. *et al.* in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015-4026.
962 64 Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation.
963 *bioRxiv*, 2024.2002. 2010.579780 (2024).
964 65 Carreira, J. & Zisserman, A. in *proceedings of the IEEE Conference on Computer Vision and Pattern
965 Recognition*. 6299-6308.
966 66 Mazurowski, M. A. *et al.* Segment anything model for medical image analysis: an experimental study.
967 *Medical Image Analysis* **89**, 102918 (2023).
968 67 Wang, Y., Zhao, Y. & Petzold, L. An empirical study on the robustness of the segment anything model (sam).
969 *arXiv preprint arXiv:2305.06422* (2023).
970 68 Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nature methods* **19**, 1634-1641
971 (2022).
972 69 Berg, S. *et al.* Ilastik: interactive machine learning for (bio) image analysis. *Nature methods* **16**, 1226-1232
973 (2019).
974 70 Bui, N.-T., Hoang, D.-H., Tran, M.-T. & Le, N. Sam3d: Segment anything model in volumetric medical images.
975 *arXiv preprint arXiv:2309.03493* (2023).
976 71 Jeong, W.-K. *et al.* Scalable and interactive segmentation and visualization of neural processes in EM
977 datasets. *IEEE transactions on visualization and computer graphics* **15**, 1505-1514 (2009).
978 72 Funke, J. *et al.* Large scale image segmentation with structured loss based deep learning for connectome
979 reconstruction. *IEEE transactions on pattern analysis and machine intelligence* **41**, 1669-1680 (2018).
980 73 Sonneck, J., Zhao, S. & Chen, J. in *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
981 3894-3902.
982 74 Liu, Y., Jin, Y., Azizi, E. & Blumberg, A. J. Cellstitch: 3D cellular anisotropic image segmentation via optimal
983 transport. *BMC bioinformatics* **24**, 480 (2023).
984 75 Driscoll, M. K. *et al.* Robust and automated detection of subcellular morphological motifs in 3D microscopy
985 images. *Nat Methods* **16**, 1037-1044, doi:10.1038/s41592-019-0539-z (2019).
986 76 Michalska, J. M. *et al.* Imaging brain tissue architecture across millimeter to nanometer scales. *Nature
987 Biotechnology*, 1-14 (2023).
988 77 Lam, L., Lee, S.-W. & Suen, C. Y. Thinning methodologies-a comprehensive survey. *IEEE Transactions on
989 Pattern Analysis & Machine Intelligence* **14**, 869-885 (1992).
990 78 Tagliasacchi, A., Delame, T., Spagnuolo, M., Amenta, N. & Telea, A. in *Computer Graphics Forum*. 573-597
991 (Wiley Online Library).
992 79 Jain, A. K. *Fundamentals of digital image processing*. (Prentice-Hall, Inc., 1989).
993 80 Sethian, J. A. A fast marching level set method for monotonically advancing fronts. *proceedings of the
994 National Academy of Sciences* **93**, 1591-1595 (1996).
995 81 Preibisch, S., Rohlfing, T., Hasak, M. P. & Tomancak, P. in *Medical Imaging 2008: Image Processing*. 128-135
996 (SPIE).
997 82 Preibisch, S. *et al.* Efficient Bayesian-based multiview deconvolution. *Nature methods* **11**, 645-648 (2014).
998 83 Daetwyler, S. *et al.* Imaging of cellular dynamics *in vitro* and *in situ*: from a whole
999 organism to sub-cellular imaging with self-driving, multi-scale microscopy. *bioRxiv*, 2024.2002.2028.582579,
000 doi:10.1101/2024.02.28.582579 (2024).
001 84 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension
002 Reduction. *arXiv:1802.03426* (2018). <<https://ui.adsabs.harvard.edu/abs/2018arXiv180203426M>>.
003 85 Laine, R. F., Arganda-Carreras, I., Henriques, R. & Jacquemet, G. Avoiding a replication crisis in deep-learning-
004 based bioimage analysis. *Nature methods* **18**, 1136-1144 (2021).
005 86 Nanes, B. A. *et al.* Keratin isoform shifts modulate motility signals during wound healing. *bioRxiv*,
006 doi:10.1101/2023.05.04.538989 (2023).
007 87 Maška, M. *et al.* The Cell Tracking Challenge: 10 years of objective benchmarking. *Nature Methods* **20**, 1010-
008 1020 (2023).
009 88 Cai, E. *et al.* Visualizing dynamic microvillar search and stabilization during ligand detection by T cells. *Science*
010 **356**, eaal3118 (2017).

- 011 89 Petrie, R. J. & Yamada, K. M. At the leading edge of three-dimensional cell migration. *Journal of cell science*
012 125, 5917-5926 (2012).
- 013 90 García-Arcos, J. M., Jha, A., Waterman, C. M. & Piel, M. Blebology: principles of bleb-based migration. *Trends
014 in Cell Biology* (2024).
- 015 91 Weems, A. D. et al. Blebs promote cell survival by assembling oncogenic signalling hubs. *Nature* 615, 517-525
016 (2023).
- 017 92 Mohan, A. S. et al. Enhanced Dendritic Actin Network Formation in Extended Lamellipodia Drives
018 Proliferation in Growth-Challenged Rac1(P29S) Melanoma Cells. *Dev Cell* 49, 444-460 e449,
019 doi:10.1016/j.devcel.2019.04.007 (2019).
- 020 93 Rahaman, N. et al. in *International conference on machine learning*. 5301-5310 (PMLR).
- 021 94 Ke, L. et al. Segment anything in high quality. *Advances in Neural Information Processing Systems* 36 (2024).
- 022 95 Xu, C. & Prince, J. L. Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing* 7, 359-
023 369 (1998).
- 024 96 Lutton, J. E., Collier, S. & Bretschneider, T. A curvature-enhanced random walker segmentation method for
025 detailed capture of 3D cell surface membranes. *IEEE Transactions on Medical Imaging* 40, 514-526 (2020).
- 026 97 Zhou, D., Bousquet, O., Lal, T., Weston, J. & Schölkopf, B. Learning with local and global consistency.
Advances in neural information processing systems 16 (2003).
- 027 98 He, K., Sun, J. & Tang, X. Guided image filtering. *IEEE transactions on pattern analysis and machine
intelligence* 35, 1397-1409 (2012).
- 028 99 Stringer, C. & Pachitariu, M. Transformers do not outperform Cellpose. *bioRxiv*, 2024.2004. 2006.587952
031 (2024).
- 032 100 Gaglia, G. et al. Lymphocyte networks are dynamic cellular communities in the immunoregulatory landscape
033 of lung adenocarcinoma. *Cancer Cell* 41, 871-886. e810 (2023).
- 034 101 Smits, J. P. H. et al. Immortalized N/TERT keratinocytes as an alternative cell source in 3D human epidermal
035 models. *Sci Rep* 7, 11838, doi:10.1038/s41598-017-12041-y (2017).
- 036 102 Rikken, G., Niehues, H. & van den Bogaard, E. H. Organotypic 3D Skin Models: Human Epidermal Equivalent
037 Cultures from Primary Keratinocytes and Immortalized Keratinocyte Cell Lines. *Methods in molecular biology*
038 2154, 45-61, doi:10.1007/978-1-0716-0648-3_5 (2020).
- 039 103 Ulman, V. et al. An objective comparison of cell-tracking algorithms. *Nature methods* 14, 1141-1152 (2017).
- 040 104 Amat, F. et al. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data.
Nature methods 11, 951-958 (2014).
- 041 105 Welf, E. S. et al. Quantitative multiscale cell imaging in controlled 3D microenvironments. *Developmental cell*
043 36, 462-475 (2016).
- 044 106 Chang, B.-J. et al. Universal light-sheet generation with field synthesis. *Nature methods* 16, 235-238 (2019).
- 045 107 Chang, B.-J., Dean, K. M. & Fiolka, R. Systematic and quantitative comparison of lattice and Gaussian light-
046 sheets. *Optics express* 28, 27052-27077 (2020).
- 047 108 Brand, M., Granato, M., Nüsslein-Volhard, C., Nusslein-Volhard, C. & Dahm, R. Zebrafish: A practical
048 approach. *Keeping and raising zebrafish* 7, 39 (2002).
- 049 109 Westerfield, M. The Zebrafish Book; A guide for the laboratory use of zebrafish (*Danio rerio*). (No Title)
050 (2007).
- 051 110 Chi, N. C. et al. Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. *Genes &
052 development* 22, 734-739 (2008).
- 053 111 White, R. M. et al. Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell stem cell* 2,
054 183-189 (2008).
- 055 112 Kaufmann, A., Mickoleit, M., Weber, M. & Huisken, J. Multilayer mounting enables long-term imaging of
056 zebrafish development in a light sheet microscope. *Development* 139, 3242-3247 (2012).
- 057 113 Daetwyler, S., Günther, U., Modes, C. D., Harrington, K. & Huisken, J. Multi-sample SPIM image acquisition,
058 processing and analysis of vascular growth in zebrafish. *Development* 146, dev173757 (2019).
- 059 114 Nirmal, A. J. et al. The spatial landscape of progression and immunoediting in primary melanoma at single-
060 cell resolution. *Cancer Discovery* 12, 1518-1541 (2022).
- 061 115 Lin, J.-R. et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF
062 and conventional optical microscopes. *elife* 7 (2018).
- 063 116 Quintana, E. et al. Human melanoma metastasis in NSG mice correlates with clinical outcome in patients.
Science translational medicine 4, 159ra149-159ra149 (2012).
- 064 117 Quintana, E. et al. Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is
066 reversible and not hierarchically organized. *Cancer cell* 18, 510-523 (2010).

- 067 118 Dean, K. M. *et al.* Isotropic imaging across spatial scales with axially swept light-sheet microscopy. *Nature protocols* **17**, 2025-2053 (2022).
- 068 119 Marin, Z. *et al.* navigate: an open-source platform for smart light-sheet microscopy. *bioRxiv*, 2024.2002. 2009.579083 (2024).
- 069 120 Yeo, I. K. & Johnson, R. A. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954-959 (2000).
- 070 121 Orieux, F., Giovannelli, J.-F. & Rodet, T. Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution. *JOSA A* **27**, 1593-1607 (2010).
- 071 122 Meijering, E. *et al.* Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry Part A: the journal of the International Society for Analytical Cytology* **58**, 167-176 (2004).
- 072 123 Frangi, A. F., Niessen, W. J., Vincken, K. L. & Viergever, M. A. in *Medical Image Computing and Computer-Assisted Intervention—MICCAI’98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings* 1. 130-137 (Springer).
- 073 124 Blum, H. A transformation for extracting new descriptions of shape. *Models for the perception of speech and visual form*, 362-380 (1967).
- 074 125 Tam, R. & Heidrich, W. in *IEEE Visualization, 2003. VIS 2003*. 481-488 (IEEE).
- 075 126 Zhang, T. Y. & Suen, C. Y. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM* **27**, 236-239 (1984).
- 076 127 Goshtasby, A. A. Fusion of multi-exposure images. *Image and Vision Computing* **23**, 611-618 (2005).
- 077 128 Crouse, D. F. On implementing 2D rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems* **52**, 1679-1696 (2016).
- 078 129 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).
- 079 130 Cignoni, P. *et al.* in *Eurographics Italian chapter conference*. 129-136 (Salerno, Italy).
- 080 131 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science* **30**, 70-82 (2021).
- 081 132 Tofanelli, R., Vijayan, A., Scholz, S. & Schnitz, K. Protocol for rapid clearing and staining of fixed Arabidopsis ovules for improved imaging by confocal laser scanning microscopy. *Plant Methods* **15**, 1-13 (2019).
- 082 133 Willis, L. *et al.* Cell size and growth regulation in the *Arabidopsis thaliana* apical stem cell niche. *Proceedings of the National Academy of Sciences* **113**, E8238-E8246 (2016).
- 083 134 Refahi, Y. *et al.* A multiscale analysis of early flower development in *Arabidopsis* provides an integrated view of molecular regulation and growth control. *Developmental Cell* **56**, 540-556. e548 (2021).
- 084 135 Yang, J. *et al.* MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data* **10**, 41 (2023).
- 085 136 Yang, X., Xia, D., Kin, T. & Igashiki, T. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2656-2666.
- 086 137 Schneider, M., Reichold, J., Weber, B., Székely, G. & Hirsch, S. Tissue metabolism driven arterial tree generation. *Medical image analysis* **16**, 1397-1414 (2012).
- 087 138 Schneider, M., Hirsch, S., Weber, B., Székely, G. & Menze, B. H. in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14–18, 2014, Proceedings, Part II* **17**. 89-96 (Springer).
- 088 139 108