UNSEG: unsupervised segmentation of cells and their nuclei in complex tissue samples

Bogdan Kochetov[1,2], Phoenix Bell[3], Paulo S. Garcia[3], Akram S. Shalaby[4], Rebecca Raphael[1,2], Benjamin Raymond[5], Brian J. Leibowitz[2, 6], Karen Schoedel[3], Rhonda M. Brand[7,8], Randall E. Brand[2,7], Jian Yu[9], Lin Zhang[9], Brenda Diergaarde[2,10], Robert E. Schoen[7], Aatur Singhi[3], and Shikhar Uttam[1,2,*]

[1]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA
[2]UPMC Hillman Cancer Center, Pittsburgh, PA, USA
[3]Department of Pathology, University of Pittsburgh, Pittsburgh, PA, USA
[4]University Hospitals Cleveland Medical Center, Case Western Reserve University, Cleveland, OH, USA
[5]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA
[6]Department of Radiation Oncology, University of Pittsburgh, Pittsburgh, PA, USA
[7]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
[8]Magee Womens Research Institute, Pittsburgh, PA, USA
[9]Department of Medicine, University of Southern California, Los Angeles, CA, USA
[10]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA
[*]shf28@pitt.edu

ABSTRACT

Multiplexed imaging technologies have made it possible to interrogate complex tumor microenvironments at sub-cellular resolution within their native spatial context. However, proper quantification of this complexity requires the ability to easily and accurately segment cells into their sub-cellular compartments. Within the supervised learning paradigm, deep learning based segmentation methods demonstrating human level performance have emerged. Here we present an unsupervised segmentation (UNSEG) method that achieves deep learning level performance without requiring any training data. UNSEG leverages a Bayesian-like framework and the specificity of nucleus and cell membrane markers to construct an *a posteriori* probability estimate of each pixel belonging to the nucleus, cell membrane, or background. It uses this estimate to segment each cell into its nuclear and cell-membrane compartments. We show that UNSEG is more internally consistent and better at generalizing to the complexity of tissue samples than current deep learning methods. This allows UNSEG to unambiguously identify the cytoplasmic compartment of a cell, which we employ to demonstrate its use in an example biological scenario. Within the UNSEG framework, we also introduce a new perturbed watershed algorithm capable of stably and accurately segmenting a cell nuclei cluster into individual cell nuclei. Perturbed watershed can also be used as a standalone algorithm that researchers can incorporate within their supervised or unsupervised learning approaches to replace classical watershed. Finally, as part of developing UNSEG, we have generated a high-quality annotated gastrointestinal tissue dataset, which we anticipate will be useful for the broader research community. Segmentation, despite its long antecedents, remains a challenging problem, particularly in the context of tissue samples. UNSEG, an easy-to-use algorithm, provides an unsupervised approach to overcome this bottleneck, and as we discuss, can help improve deep learning based segmentation methods by providing a bridge between unsupervised and supervised learning paradigms.

Introduction

Recent innovations in highly multiplexed immunofluorescence imaging[1–9] have substantially increased the range of antigens that can be spatially profiled in a tissue sample, from $3-5$ targets to ~ 60[10]. Segmentation is a required step for quantitatively associating their spatial expressions with individual cells. Since 2012, when AlexNet[11], a deep convolutional neural network (CNN), outperformed other methods in the ImageNet classification challenge, there has been a paradigm shift towards using CNN based deep learning (DL) frameworks[12] trained on curated datasets for cell and nuclei segmentation tasks[13–19]. Among them, Cellpose[18] − a DL method based on a U-Net architecture utilizing gradient flow representation of cells− and Mesmer[19] − a DL method based on ResNet50 architecture − have demonstrated human-level performance in the highly multiplexed imaging context. However, due to their dependence on stochastic gradient descent and back-propagation based optimization during the training step, it remains difficult to identify the contribution of each neuron to the eventual segmentation outcome, and as a consequence explain the source of errors in segmentation when they occur[20]. As a result, improving performance of these black-box DL models requires rewiring the input-output mapping via training on additional datasets[21]. However, in complex tissue samples with considerable heterogeneity and ambiguity in cellular organization, it is unclear whether retraining alone will consistently improve results across all samples, or if multiple DL models need to be constructed and used through a trial and error approach, with the hope that their performance will optimally generalize. Curation of annotated datasets of sufficient quality that capture the tissue microenvironment diversity also remains a critical challenge.

In contrast to DL approaches, most unsupervised cell segmentation methods[22–32] do not require training data, are explainable, and therefore where needed, can be optimized for individual images. However, to date no unsupervised segmentation method capable of approaching DL model performance exists, or has even been considered feasible. Here, we present a new unsupervised segmentation algorithm (UNSEG) capable of performing sub-cellular segmentation of tissue sample images with accuracy on par with state-of-the-art DL segmentation approaches such as Cellpose and Mesmer. UNSEG achieves this performance in two stages. At the first stage, UNSEG quantifies the intrinsic contrast provided by nucleus and cell membrane specific markers at the local and global scale, and jointly exploits it to assign each pixel to the nucleus, cell membrane, or the background class. This pixel assignment is implemented with the help of Bayesian-like framework that computes *a priori* distributions and an image contrast-based likelihood function to estimate the posterior probabilities of each pixel belonging to nucleus, cell membrane or background classes. UNSEG uses the posterior probabilities to assign the pixel to the correct compartment. At the second stage, it parses the semantic pixel assignments into topologically consistent nuclei and cell membranes. Towards this goal UNSEG introduces perturbed watershed, a new watershed algorithm we have developed, to correctly partition a nuclei cluster into individual nuclei. The final output of UNSEG are nuclei and cell segmented masks corresponding to the input image.

We have curated a labeled gastrointestinal (GI) dataset comprising diverse tissue images to benchmark UNSEG performance. In addition, we anticipate that this dataset will be useful to DL researchers and the broader research community[21]. Importantly, we note that since UNSEG does not require any training data to segment tissue images, it can be used to generate high quality segmentation of unlabeled tissue images, which is majority of the data in real-world settings, as optimized initial estimates for improving DL models within unsupervised and semi-supervised settings. UNSEG, therefore, is an easy to use method for unsupervised sub-cellular segmentation of complex tissue samples that does not require extensive setup and performs on par with state-of-the-art DL models. It also has the potential to improve the state-of-the-art in deep learning.

Results

UNSEG principle and design

Segmenting cell nuclei and membranes in tissue samples is challenging because of their complex morphology, ambiguous overlaps, and heterogeneity in spatial distribution of nuclei and cell-membrane markers within each cell. In the morphological context, although cells and their nuclei exhibit an overall convex topology, they locally deviate from it to varying degrees depending on cell types, and particularly in tumors with irregularly shaped cancer cells. Additionally, many cells in a tissue-dependent manner are clumped in clusters where their shape and overlap is difficult to parse. Cells in tissues also exhibit uneven intra-cellular distribution of marker expression. Together, these degrees of complexity make it difficult to consistently segment cell nuclei and membranes using unsupervised segmentation approaches such as classical watershed[22,23,29], shape and intensity prior[27,28,30–32], and tracking of diffused gradient flow[24,25], which have primarily been developed for segmenting cells in culture that lack tissue associated heterogeneity related to cellular morphology, expression and overlap. UNSEG framework overcomes these limitations by jointly exploiting the expression-based topology and distribution of markers specific to nucleus and cell-membrane (Figure 1).

UNSEG combines *a priori* probability of each image pixel belonging to a nucleus or cell-membrane (Figure 1a) with a contrast-based likelihood function (Figure 1b), to compute *a posteriori* semantic segmentation of image pixels into nuclei and cell-membrane (Figure 1c). UNSEG performs this segmentation both at the global level of the entire image, and at the local

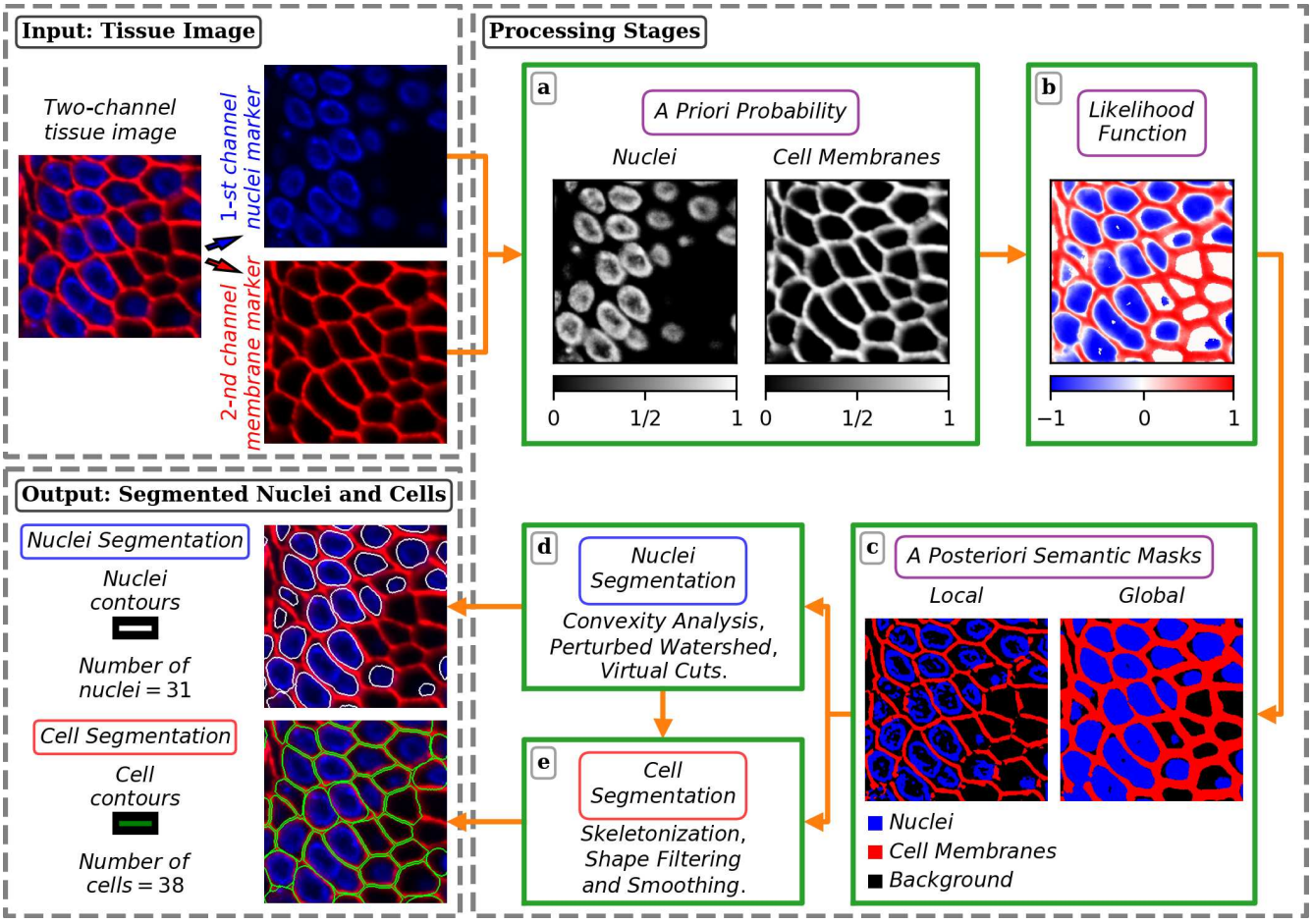

Figure 1. UNSEG framework. Input is a two-channel image comprising of nucleus (channel 1) and cell membrane (channel 2) marker expressions. **a** *A priori* spatial probability distribution of nucleus and cell membrane marker expression. **b** Likelihood map of a pixel to belong to the nucleus or cell-membrane, quantified through the visual contrast function, which mimics human perception. **c** *A posteriori* local and global semantic segmentation masks respectively capturing local morphological heterogeneity and global nucleus and cell-membrane topology. **d** Instance segmentation of nuclei from semantic segmentation masks. **e** Instance segmentation of cells based on individual nuclei masks and semantic segmentation masks. Nucleus and cell segmentation results of **d** and **e** form the UNSEG output.

level in a neighborhood around each pixel (Figure 1c). The local segmentation captures the local heterogeneity in nucleus and cellular morphology, while the global segmentation ensures that the overall topological structure of the nuclei and cell membrane compartments is preserved across the entire image. The final step of UNSEG utilizes these local and global nuclei and cell-membrane semantic masks to obtain instance segmentation of individual nuclei (Figure 1d) and cells (Figure 1e). This step includes partitioning nuclei cluster into individual nuclei based on convexity analysis, perturbed watershed and its ancillary function we refer to as virtual cuts. The latter two are briefly described below. The details of each step are described in the Methods section.

Perturbed watershed. Classical watershed based segmentation[33,34] identifies individual nuclei in a cluster as watersheds, with each watershed basin representing a nucleus in the cluster. However, heterogeneity in spatial distribution of nucleus marker can make it difficult to uniquely identify the individual basins. Cellpose overcomes this problem in the supervised context by developing a gradient flow field representation of nuclei whose ground truth is annotated by a human user[18]. This representation provided a stable and unique representation of nuclei basins. In the unsupervised context, we have developed a perturbed watershed approach (Figure 2 and Methods), where the initial watershed based segmentation (Figure 2i) of the nuclei cluster into individual nuclei is perturbed (Figures 2j - 2m) based on an adaptive distance-transform estimate (Figure 2h) computed from the global nuclei cluster (Figure 2d) and local topology of the cell-membrane network (Figure 2e). Nuclei that are correctly segmented remain stable to the perturbations, while spuriously segmented nuclei collapse to a point-like object with area not exceeding a few pixels. When applied recursively, perturbed watershed partitions the nuclei cluster into individual nuclei. An

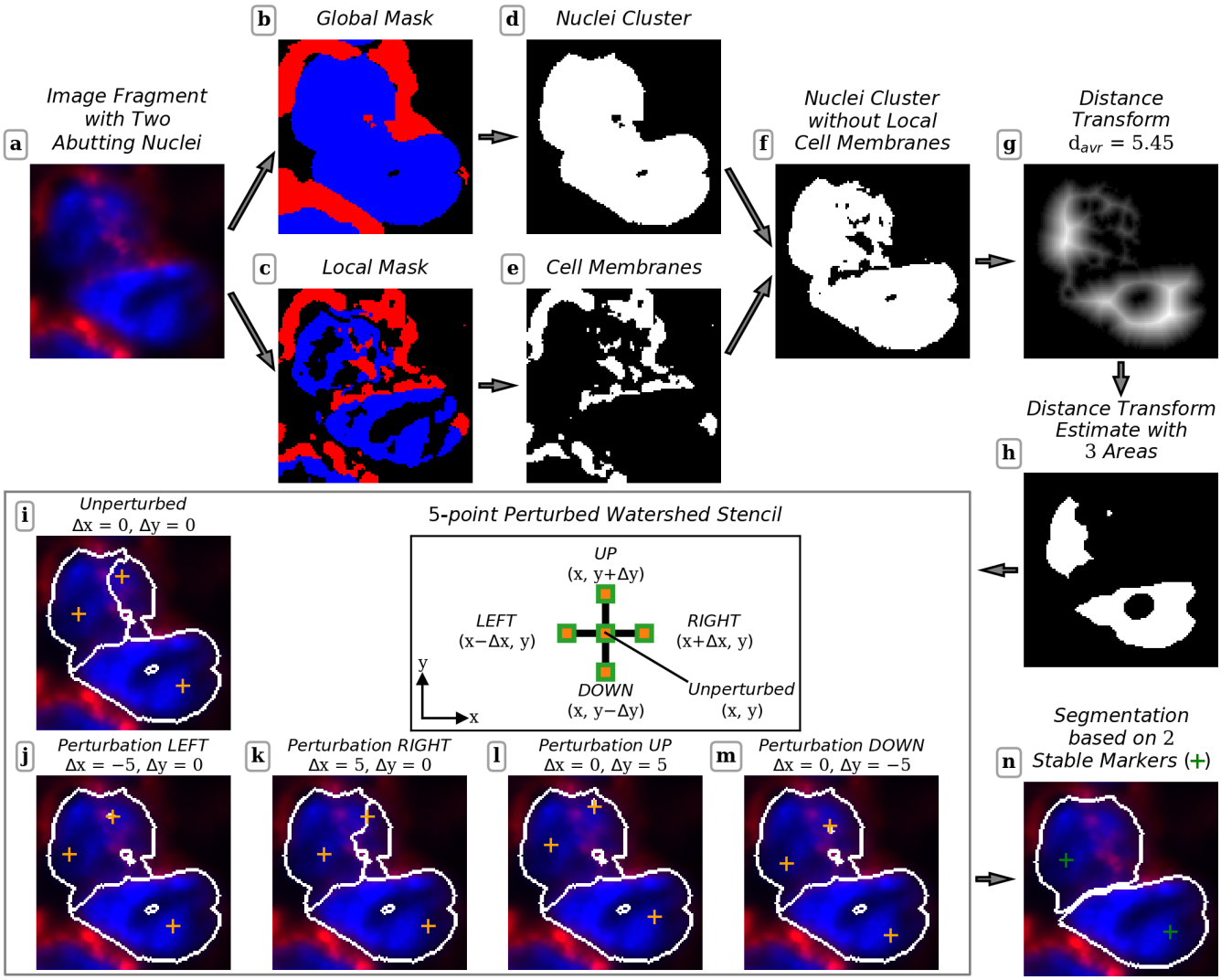

Figure 2. Perturbed watershed method. a Input image fragment with two abutting nuclei. **b-e** The posterior global and local masks of the input image from which the global nucleus cluster mask and local cell-membrane mask are extracted for downstream perturbed watershed analysis. **f** Global nuclei cluster mask with cuts corresponding to the local cell-membrane mask. **g** Distance transform of this modified global nuclei cluster mask. **h** Adaptive distance-transform estimate obtained by thresholding the distance transform by d_{avr} (see Methods). **i** Initial (unperturbed) watershed segmentation. **j-m** Perturbed watershed segmentations computed after shifting all markers from their unperturbed positions to the left ($\Delta x = -\lfloor d_{avr} \rfloor$), right ($\Delta x = \lfloor d_{avr} \rfloor$), up ($\Delta y = \lfloor d_{avr} \rfloor$), and down ($\Delta y = -\lfloor d_{avr} \rfloor$) on $\lfloor d_{avr} \rfloor = 5$ pixels, respectively. **n** Output segmentation of two-nuclei cluster based on the perturbed watershed.

example of a two-nuclei cluster is shown in Figure 2. Initial watershed partitions the cluster into three nuclei (Figure 2i), one of which shrinks to a point object on perturbation of the watershed seed point. The perturbation is performed in four directions: up, down, left, and right. In this example, the unstable nucleus collapsed for three of those perturbations (up, down, and left), indicating that the seed point is unstable and the corresponding segmentation is a spurious nucleus. Therefore, it is removed and the correct watershed based segmentation (Figure 2n) is obtained using the two remaining stable seed points and the original distance transform (Figure 2g). We note that perturbed watershed algorithm does not make any assumptions specific to the fluorescence based imaging modality. It is, in fact, agnostic to the imaging modality being used, and can be used to improve classical watershed results, wherever, the latter method is applicable.

Virtual cuts. In some cases, mostly when cell membrane marker is not present, the initial watershed segmentation step might undersegment the cluster. For such cases, we have developed the virtual cuts method that utilizes non-convex topology of the cluster to identify nuclei centroids that act as seed points for the watershed algorithm. See Methods for implementation details.

New dataset for benchmarking

As part of our UNSEG development, we curated 75 tiff images of tissue sections from eight organs of the extended human gastrointestinal (GI) system – appendix, colon, esophagus, gallbladder, liver, pancreas, small intestine, and stomach. The immunofluorescence images were acquired via imaging of Formalin-fixed paraffin-embedded (FFPE) tissue sections labeled using Hoechst and fluorescent-dye-conjugated Na^+K^+ATPase as respective markers for cell nuclei and membranes (See Methods). The image dimensions are 1000×1000. The images were acquired using a 0.95 numerical-aperture objective with a 40X magnification, and have a pixel pitch of 0.16 µm/pixel. The GI dataset includes images of normal tissue and tissue from malignancies related to chronic inflammation, cancer precursor lesions and cancer. These images capture a wide range of tissue organization from samples with sparsely located cells to those with very high cell density. Figure 3 shows 12 representative images from the dataset.

Expert pathologists independently annotated the 75 images resulting in ground truth with 16201 nuclei and 16217 cells. These annotations were performed manually, without any algorithmic aid, to truly reflect human performance. The detailed description of the dataset is presented in Supplementary Table 1 and Supplementary Figure 1, while the nuclei and cell annotations of the 12 representative images are shown in Supplementary Figure 2. To annotate nuclei and cells in the 75 images we developed Cellthon - a Python based graphical user interface for annotating cells and their nuclei in tissue images.

We used the GI dataset to benchmark UNSEG performance. Moreover, we anticipate that this dataset will also serve as a resource for researchers requiring annotated datasets for future algorithm development and testing[21].

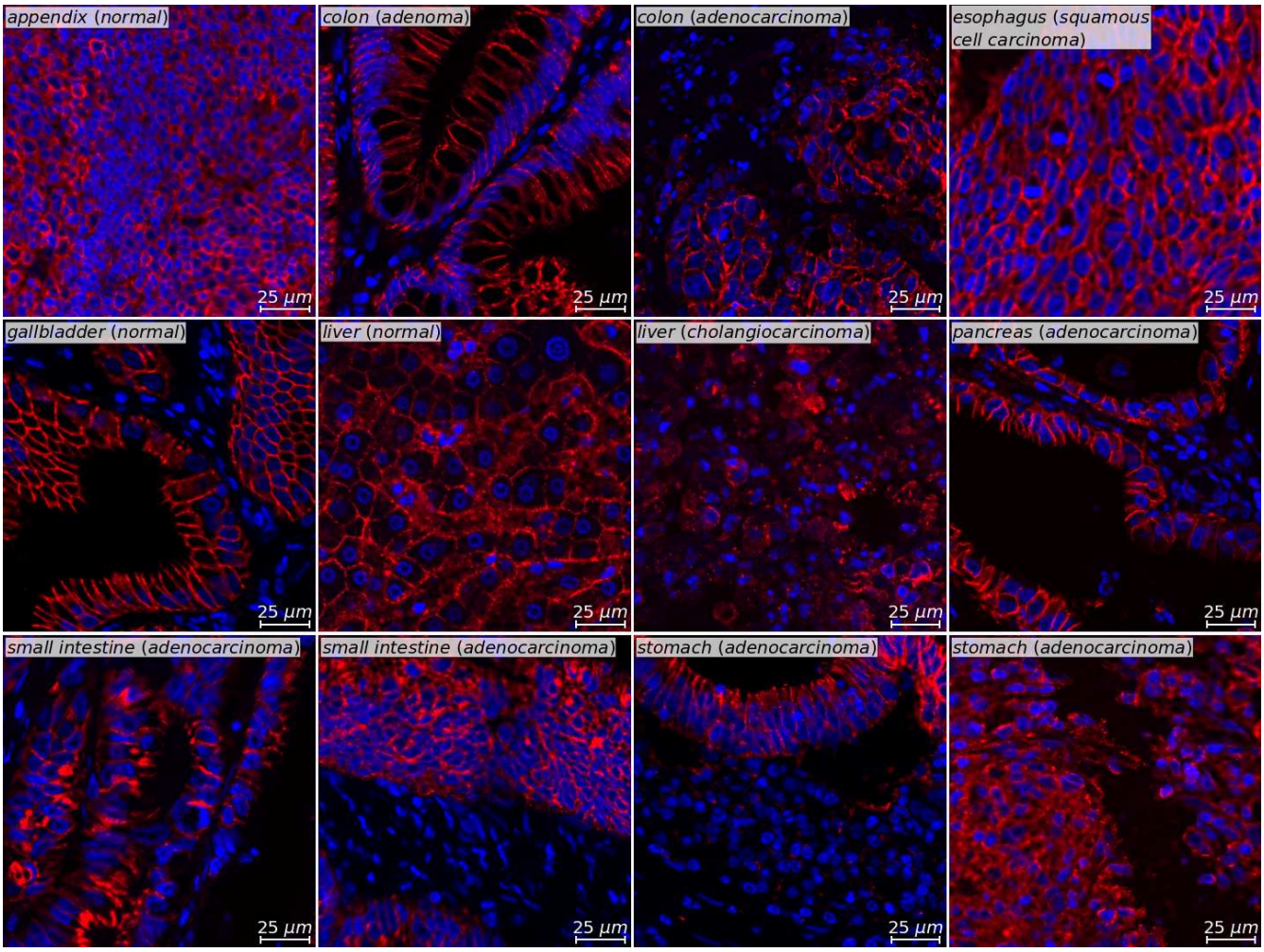

Figure 3. Gastrointestinal (GI) dataset. Twelve representative tissue images from the GI dataset drawn from different organs of the human gastrointestinal system with different pathobiology. Blue and red colors respectively indicate nuclei (Hoechst) and cell membranes (Na^+K^+ATPase) marker expression. The dimensions of each image are 1000×1000 pixels. The images were acquired using microscope with 0.95 NA, 40X objective and imaging sensor with a pixel pitch of 0.16 µm/pixel.

UNSEG benchmarking

We used the GI dataset along with publicly available datasets to benchmark segmentation performance of UNSEG with respect to Cellpose and Mesmer, two state-of-the-art deep learning models that have consistently demonstrated good performance in segmenting immunofluorescence imaging data particularly in the context of highly multiplexed imaging[18,19]. To perform the comparison with Cellpose, we used Cellpose version 2.1.0. In this version, we chose *nuclei* and *TN2* models from the Cellpose

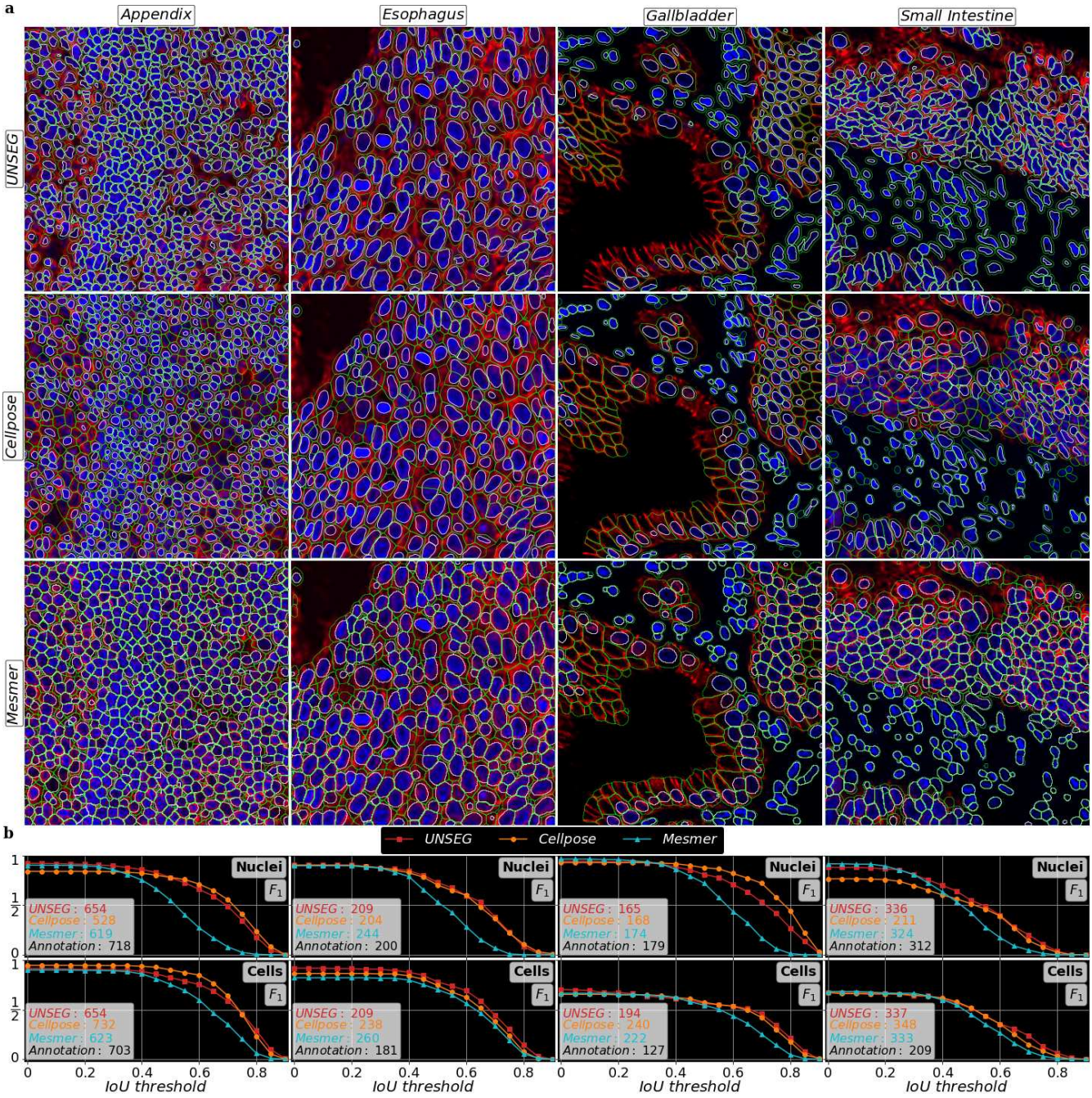

Figure 4. Comparison of UNSEG, Cellpose, and Mesmer on four example images from the GI dataset. a Columns respectively correspond to appendix, esophagus, gallbladder, and small intestine tissue images. Rows show nuclei (white boundary) and cell (green boundary) segmentation results for the four examples using UNSEG, Cellpose and Mesmer, respectively. **b** The two rows respectively show nucleus and cell segmentation accuracy of UNSEG, Cellpose, and Mesmer. Accuracy is measured using number of segmented objects (see inserts) and F_1 score curves plotted as a function of IoU ratio between the segmented and annotated labels for nuclei and cells, respectively.

'model zoo' to respectively segment nuclei and cells. Our choice was based on them giving the best segmentation results for the GI dataset in comparison to all other Cellpose models. We used Cellpose size calibration procedure to estimate the cell diameter for each of the 75 images in our dataset. We also chose the most recent Mesmer model (DeepCell 0.12.6), and set the model parameter `image_mpp` to the pixel pitch in microns per pixel for our imaging dataset. Benchmarking was performed by computing the F_1 score as a function of intersection over union (IoU) threshold[35]. The IoU metric quantifies the degree of overlap between algorithm prediction and the annotated ground truth. It is bounded between 0 and 1, with one indicating perfect overlap. By thresholding the IoU metric over its range, we obtain F_1 accuracy curve for each method (see Methods).

Figure 4a shows UNSEG, Cellpose, and Mesmer segmentation results applied to four representative examples from our 75 image GI dataset. Visual comparison shows similar performance between the different methods. One difference between UNSEG and the other two methods is that, although, UNSEG does implement boundary smoothing, it does not enforce strict shape constraints. As a consequence, the shape of UNSEG-based nucleus and cell segmentation is more irregular, but also more realistic and less synthetic appearing than Cellpose and Mesmer, where the segmented shape and the partitioning of the cell clusters strongly correlates with the nucleus and cell ground truth for the training data.

The F_1 curves for the four examples (Figure 4b) demonstrate that UNSEG performance is similar to that of the deep learning methods trained on about a million cells. The ground truth annotations for these four examples are shown in Supplementary Figure 2.

The similarity in their performance on the four example images generalizes to the entire 75 image dataset. The results are shown in Figure 5. The first row depicts the median F_1 curves corresponding to nuclei and cell segmentation by the three methods. The curves indicate that the three methods have similar segmentation performance. For cell segmentation, the median UNSEG performance is slightly below the other two methods, which is partly due to the conservative nature of UNSEG cell segmentation in resolving cell boundary ambiguity in cases where the tissue section capture partial cell membranes without their respective nuclei. In these cases, UNSEG does not always include their segmentation masks in the final results. Nevertheless, if we look at the pairwise 95% F_1 confidence interval comparison between UNSEG performance, with Cellpose and Mesmer – the second and third rows of Figure 5 respectively – we clearly see their complete overlap, indicating their overall identical

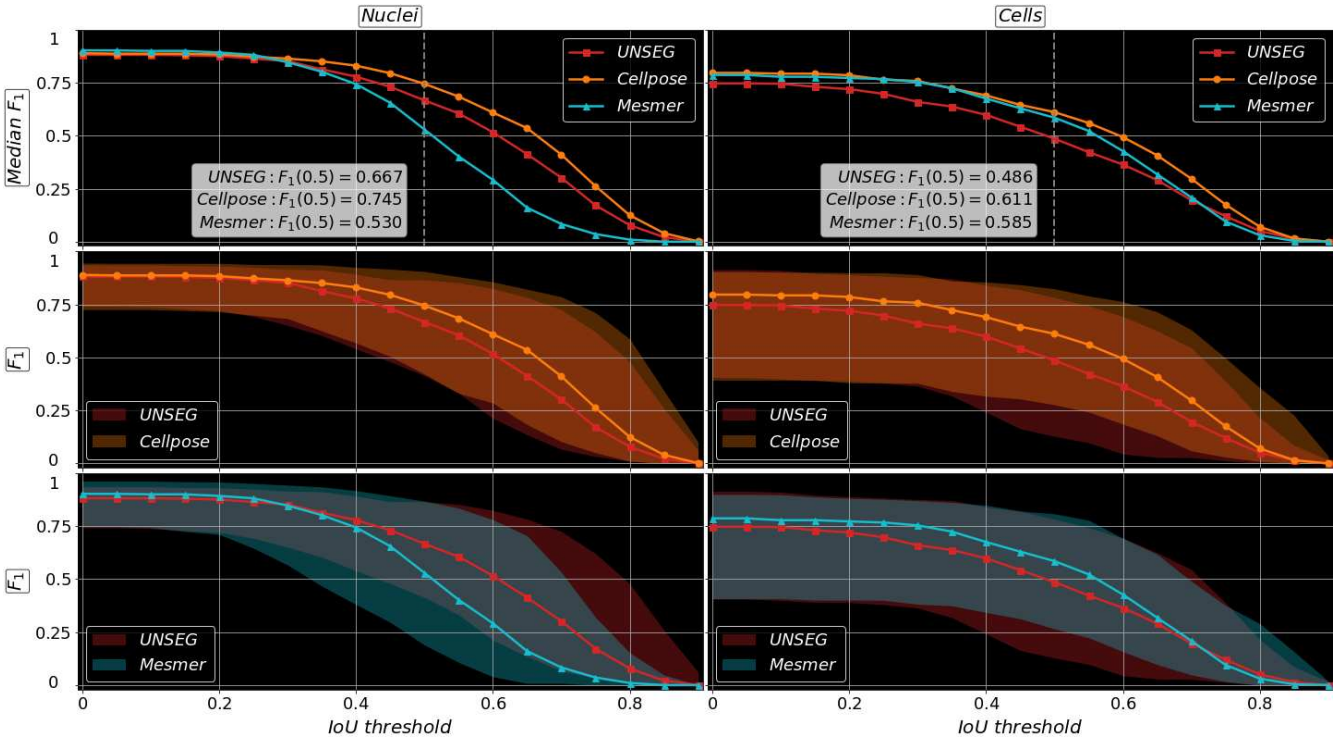

Figure 5. Performance comparison of UNSEG, Cellpose, and Mesmer for the entire GI dataset. First row compares median F_1 score performance curves for the three methods as a function of IoU threshold for nucleus and cell segmentation of images in the GI dataset. The insert contains median F_1 score values at the IoU threshold of 0.5 for three algorithms. The second and third rows respectively show pairwise comparison between UNSEG and Cellpose, and UNSEG and Mesmer. The comparison includes median F_1 score curves along with their 95% confidence intervals. Their complete overlap indicates similar performance of all three methods.

performance. A more detailed version of Figure 5 is presented in Supplementary Figure 3. To segment nuclei and cells in images from the GI dataset we used default values for all UNSEG parameters (see Methods and Supplementary Table 2).

Furthermore, we also benchmarked segmentation performance of UNSEG with respect to Cellpose and Mesmer using publicly available, multiplexed imaging tissue datasets acquired using CODEX, Vectra, and Zeiss imaging platforms[36]. These datasets include cell annotations. Supplementary Figure 4 shows the cell segmentation performance of UNSEG, Cellpose, and Mesmer on CODEX dataset comprising of ten 400×400 images. Supplementary Figure 4a depicts an example image of lymph node from the CODEX dataset along with its ground truth cell annotation, the cell segmentations predicted by UNSEG, Cellpose, and Mesmer, and their corresponding F_1 score based performance curves. Due to the high cell density, lymph node samples are typically difficult to segment. This example provides a clear visual and quantitative demonstration of UNSEG performing segmentation on par with Cellpose and MESMER. Supplementary Figure 4b further shows that the quality UNSEG performance extends to the entire CODEX dataset.

Similarly, Supplementary Figures 5 and 6 compare the performance of UNSEG cell segmentation with that of Cellpose and Mesmer for Vectra and Zeiss datasets[36] respectively. The Vectra dataset includes 131 images of size 400×400, while the Zeiss dataset consists of nineteen images of size 800×800. Although, UNSEG performs stable and high quality segmentation, faithfully capturing cell shapes, its F_1 score based performance is upper bounded by Cellpose and MESMER. This is partly due to annotated ground truth being manually curated using algorithms available through ImageJ 1.52p. The resulting over segmentation in annotated ground truth tends to favor Cellpose and MESMER performance. Second, these datasets do not include appropriately imaged cell-membrane markers across all images. Instead we primarily relied on cytoplasmic pan-cytokeratin marker, which slightly degraded UNSEG performance. Nevertheless, UNSEG results, more closely resemble human annotation of cells than the tessellation-like partitioning achieved by deep learning methods.

UNSEG characteristics and use case

UNSEG employs an integrated approach to segmenting nuclei and cells that, by design, emphasizes internal consistency between each cell nucleus and its membrane. As a consequence, UNSEG guarantees that no segmented nucleus can be located beyond the boundaries of its cell. This drawback is often found in both Cellpose and Mesmer, where nucleus and cell-segmentation are performed independently. Figure 6a depicts a colon tissue section illustrating the internal inconsistency in nucleus and cell masks estimated by Cellpose and Mesmer for a pair of examples highlighted with dashed boxes. In the case of Cellpose two cells are sharing the same nucleus, while in Mesmer, for region marked as 1, two nuclei are included within the same cell. UNSEG avoids such discrepancies due to its joint segmentation of cell nuclei and membrane. This joint processing ensures that UNSEG can unambiguously identify the cytoplasmic compartment of cells. The internal consistency among sub-cellular compartments is of particular importance in biological studies where correct sub-cellular localization of signaling pathway components is essential to study intra-cellular signaling. For example, tumor protein *P*53 can be sequestered in the cytoplasm, or localized in the nucleus depending on DNA damage, and other exogenous and endogenous stresses. However, in unstressed cells, it is expressed at low levels and localizes in both cytoplasm and the nucleus[37]. As another example, histone methyltransferase *EZH*2 localizes in the nucleus, where it regulates gene expression through its canonical histone lysine methyltransferase activity[38]. Supplementary Figure 7 depicts an example of such a real use case, where UNSEG is used in a multiplexed imaging context to segment cells and their nuclei based on Hoechst and Na^+K^+ATPase. The UNSEG-based segmentation is used to localize intra-cellular *P*53 and *EZH*2 expression in a region of healthy colon tissue with densely located cells. The internal consistency of UNSEG segmentation ensures that the user is correctly able to evaluate *P*53 expression in the nucleus and the cytoplasm, while ensuring that the canonical activity of *EZH*2 in the healthy tissue is not associated with the cytoplasm.

As briefly mentioned earlier, UNSEG does not impose a strict shape constraint on the nucleus segmentation mask by allowing it to be locally non-convex. Consequently, in complex tissue sections it is, on average, better at preserving true nuclei shape than Cellpose and Mesmer, which either are usually more rounded, and in regions of the tissue with high cell density, appear like Voronoi partitions of the tissue region. Figure 6b shows an example pancreas tissue with elongated cells that deviate from round shapes. As can be seen, the ability of UNSEG to combine knowledge of global tissue architecture and local topology, with a relaxed shape constraint allows it to better capture elongated nuclei morphology when compared to Cellpose and Mesmer. This ability is highly relevant in the context of the use case mentioned above, where users, such as cancer biologists are studying the tumor microenvironment that might include a diversity of cell shapes associated with cancer, immune, and stromal cell populations.

Runtime complexity of UNSEG is a function of number of cells and not the image size. Specifically, UNSEG runtime complexity scales approximately linearly with respect to the number of segmented cells in the image. This translates to linear scaling with respect to image area, if the spatial distribution of cells is approximately uniform. However, for sparsely populated images UNSEG run time will be significantly sub-linear. Figure 6c shows linear dependence with respect to the number of segmented cells and the image area, under the assumption of uniform cell distribution. The results were generated using a colon

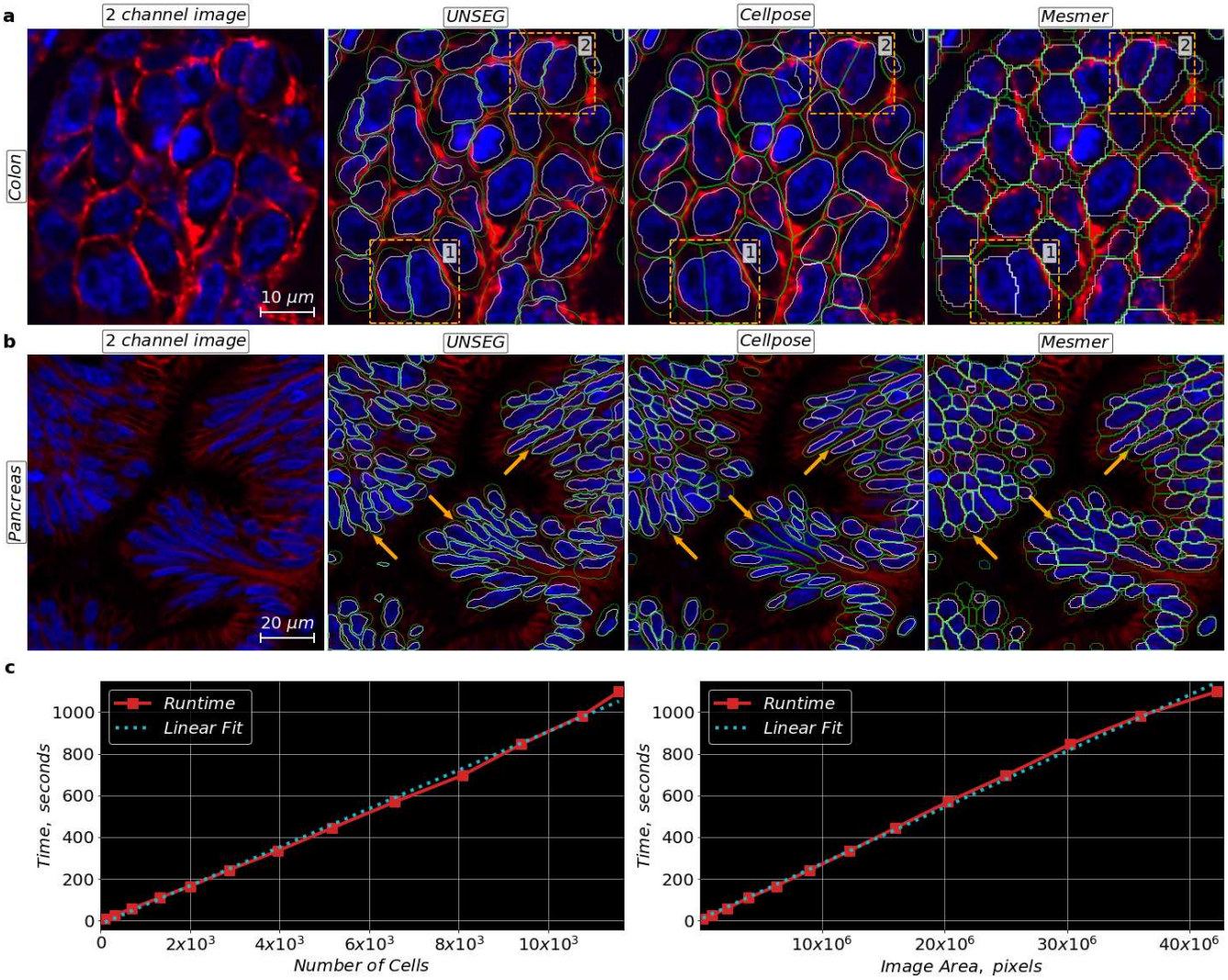

Figure 6. Characteristics of UNSEG method. a UNSEG demonstrates internal consistency between nucleus and cell boundaries. The dashed box 1 contains two cells, where both Cellpose and Mesmer have mismatch between nuclei and cell boundaries. The dashed box 2 contains another set of two cells, where Cellpose has mismatch between nuclei and cell boundaries. **b** UNSEG is better at capturing complex nuclei shapes in comparison to Cellpose and Mesmber, as exemplified by the arrows indicating examples of nuclei with complex shapes. **c** Runtime complexity of UNSEG as a function of number of cells and image area, assuming uniform cell distribution for the latter.

tissue microarray (TMA) spot with approximately uniform cell distribution. The segmentation results for the whole TMA spot are presented in Supplementary Figure 8.

Discussion

The importance of segmenting cells and their nuclei has gained renewed prominence due to the advent of multiplexed imaging technologies that have significantly enhanced the depth of information that can potentially be extracted from samples in a cell specific manner. However, tissue sections have complex cell organizations and unlike computer vision tasks, segmenting individual cells even by human experts is a difficult challenge, resulting in inter-observer discordance. Such discordance usually grows as the number of cells requiring annotation grows. This, in turn, affects ground truth quality used to train supervised learning models, and is a bottleneck for generating high quality training data. The unsupervised approach provides a complementary paradigm to segmenting complex tissue images without requiring training data. Unsupervised methods are also more adaptable to individual images of varying complexity. However, until now, no method within the unsupervised paradigm had demonstrated performance approaching supervised learning methods, particularly those based on deep learning. As a

consequence, none of its advantages were relevant. UNSEG for the first time demonstrates that unsupervised cell and nuclei segmentation can achieve accuracy at par with the current state of the art methods in deep learning. It also introduces the perturbed watershed algorithm, a new standalone algorithm with considerably improved ability to correctly segment nuclei clusters than the classical watershed algorithm. Perturbed watershed is applicable in all cases where the classical version can be used.

Segmentation fundamentally involves learning features and image representations that help the algorithm identify individual cells. Deep learning models extract these features and representations in a supervised manner. Interestingly, UNSEG performance reveals an intriguing hypothesis: utilizing intrinsic features of the cellular topology embedded within the spatial context of the tissue in a single image is equivalent to training on one million cells[19]. Thus, the performance of UNSEG suggests that feature representation in current state-of-the-art deep learning methods has considerable potential for improvement in the context of complex tissue samples. One such potential avenue could be incorporating UNSEG itself in different learning frameworks. For example, in a self-supervised learning framework UNSEG could be used to optimally initialize joint learning of neural network parameters and k-means based segmentation of cells and nuclei[39]. Another application could be in a semi-supervised setting, where a small portion of imaging data is annotated, while the majority of remaining images are unlabeled. This is usually the case in most practical settings involved segmentation of complex tissue samples. Here, UNSEG could be used to provide pseudo-labeling estimate of cell and nuclei segmentation for the unlabeled images, which can then be used to refine the DL network trained on labeled data[40]. Finally, UNSEG could be used in the setting of learning with noisy labels, where the UNSEG generated segmentation masks are noisy labels on which robust DL models can be trained[41].

UNSEG performs sub-cellular segmentation based on nucleus and cell membrane compartment markers. However, its framework does not impose any constraint on the number of markers that can be used. For example, in multi-nucleated cells, UNSEG can be modified to incorporate an additional marker specific to the nuclear membrane to coherently segment multiple overlapping nuclei belonging to the same cell. Supplementary Figure 9 depicts an example of a multi-nucleated cell, with Lamin A/C (shown in green) marking the nucleus membranes. As depicted in this figure, the modification of UNSEG utilizes the specificity of the extra marker to segment the nuclei and associate them with the same cell. Another adaptation of the UNSEG framework can be used to directly capture spatially localized cell-specific expression of markers imaged using multiplexed imaging without performing sub-cellular segmentation (see Supplementary Figure 10).

UNSEG is an easy to use method for sub-cellular segmentation of complex tissue images using multiplexed imaging technologies. It requires minimal time to setup and is accessible to researchers with varying computational backgrounds. It has three primary parameters (minimal area, convexity threshold, and dilation radius; see Methods and Supplementary Table 2.) that can be adjusted by the user to optimize segmentation performance for individual images including relatively large images like shown in Supplementary Figure 8. It is a flexible framework that can be extended to include additional markers to enhance cell segmentation and to extract localized expression of individual markers across the tissue sample. Finally, we reemphasize that unlike segmentation of objects in computer vision based situational awareness tasks, segmenting cells and their nuclei, particularly in the context of tissue samples, often results in subjective ground truth. By being able to capture intrinsic, marker-specific topological structure of cell compartments, UNSEG offers opportunities to further improve current-state-of-the-art deep learning methods. To aid in this task, we have also generated a GI dataset of 75 tissue images from eight organs of the human gastrointestinal system, along with their corresponding nuclei and cell annotations independently generated by expert pathologists.

Methods

GI dataset generation

Formalin-fixed paraffin-embedded (FFPE) tissue microarray (TMA) slides were obtained from Pantomics (Pantomics, DID381) which contain thirty-eight 2mm cores of normal and diseased tissues the human digestive system. The slides went through cyclic immunofluorescence antigen retrieval protocol[8]. They were stained with 1:200 dilution of Anti-Sodium Potassium ATPase antibody (Abcam ab198367, clone EP1845Y) overnight at 4°C in the dark, followed by staining with Hoechst 33342 (Cell Signaling Technology, 4082S) for 10 minutes at room temperature in the dark. TMA images were acquired using a 0.95 NA, 40X objective on a Nikon Ti2E microscope. Seventy five, 1000×1000 high quality regions were identified and extracted from the TMA images and saved as tiff images. Expert pathologists independently annotated these images. The annotations were done using Cellthon, a python based cell annotation graphical user interface (GUI) we created using Tkinter toolkit[42]. Together the images and annotations comprise the GI dataset.

UNSEG algorithm

Input image

The input to our algorithm is a two channel image. An example is illustrated in the "input" panel of Figure 1 and Supplementary Figure 11, as well as in Figure 3 and Supplementary Figure 8. Channel one, depicted in blue, and channel two shown in red, are

respectively associated with nuclei and cell-membrane marker expressions. Both channels are independently scaled to 0 and 1, such that $I_i : \Omega \rightarrow [0,1]$. Here I_i is the normalized image intensity for i-th channel, Ω is the image domain, and $i = 1,2$ is the indexing representing the two channels.

The algorithm performs the segmentation in four processing stages detailed below and illustrated in Figure 1 and Supplementary Figure 11.

Processing stage 1: Computing *a priori* nuclei and cell-membrane masks

In Stage 1 we compute *a priori* estimates of image foreground for each channel. The estimates are computed at the global and local scale as described below.

A priori probability

Each channel, $I_i(x,y)$, $i = 1,2$, is first pre-processed using a combination of a Gaussian filter[43] and multi-level Otsu[43–45]. The standard deviation of the Gaussian filter kernel, σ is a parameter of the algorithm that allows the user to control the degree of smoothing. This and other algorithm parameters are summurized in Supplementary Table 2. Our default setting is $\sigma = 3$. A three-level Otsu is next applied to the smoothed image and the lowest level is selected as the threshold to obtain the initial estimate of the channel foreground.

We use the initial, per-channel foreground estimate to compute the cumulative distribution function (CDF), $\mathscr{F}_i$ of I_i using intensity values, $I_i(x,y)$, of pixels (x,y) within this estimate. Two examples of CDFs are presented in Supplementary Figure 11. Using the monotonically non-decreasing property of CDF we map I_i to its cumulative probabilistic representation P_i^e, where $P_i^e(x,y) = F_i(I_i(x,y))$. We define $P_i^e(x,y)$ to be the *a priori* probability of the pixel being the nucleus $(i = 1)$ or cell-membrane $(i = 2)$. We note that this definition quantifies the intuition that stronger the marker intensity at a particular pixel, the higher its *a priori* probability. Examples of *a priori* probabilities for nuclei (P_1^e) and cell membranes, (P_2^e) are presented in Figure 1 and Supplementary Figure 11.

A priori global mask

We compute the *a priori* global mask $M_i^g(x,y)$, $i = 1,2$ using P_i^e and a simple filter called local mean suppression filter (LMSF) that we have developed. The foreground pixels (x,y) where $M_i^g(x,y) = 1$ are designed to be a superset of the pixels belonging to the true nucleus $(i = 1)$ and cell membrane $(i = 2)$ compartments of cells in $I_i(x,y)$, $i = 1,2$. M_i^g, therefore, ensures that no pixels belonging to the cells are missed.

LMSF is designed to identify the valleys (or space) that exist between nuclei (or cell membranes) of closely located cells that nevertheless have some spill over marker expression, and are therefore, difficult to identify as background. We define LMSF as,

$$\hat{I}_i(x,y) = \begin{cases} 0, & \textbf{if } \dfrac{I_i(x,y)}{\bar{I}_i(x,y)} < t_0 \\ I_i(x,y), & \textbf{otherwise} \end{cases}, \quad \text{where } \bar{I}_i(x,y) = \frac{1}{(2n_0+1)^2} \sum_{\xi=x-n_0}^{x+n_0} \sum_{\eta=y-n_0}^{y+n_0} I_i(\xi,\eta). \tag{1}$$

The above definition states that for a given pixel $(x,y) \in \Omega$, LMSF replaces the original intensity value with 0 only if the ratio of the pixel intensity to the average intensity, computed locally around the pixel neighborhood, is below the threshold parameter t_0. The size of the kernel defining the neighborhood over which the local mean intensity is computed is parameterized by n_0. We set $t_0 = 0.5$. Consequently, all pixels with intensity value less than half the mean intensity in their respective neighborhoods are replaced with zeros, allowing us to identify valleys between cells. By varying n_0 we can identify valleys and gaps of different widths. UNSEG performs LMSF filtering for $n_0 = 5,10,20,40$. If $\hat{I}_i(x,y) = 0$ for any value of n_0, then the final pixel value is set to 0 and assigned to be background in the global mask, $M_i^g(x,y)$. Thus, LMSF allows us to capture valleys of different widths. The values of n_0 are user defined and can be optimized according to complexity of individual images.

We refine the global mask $M_i^g(x,y)$ by reassigning those pixels currently in the foreground that have *a priori* probability $P_i^e(x,y) < p_i$, $i = 1,2$ to the background. This refinement is particularly useful for images with highly heterogeneous tissue with varying marker expression. The threshold value p_i should be small and by default is set to 0.01.

An example of *a priori* global mask is presented in Supplementary Figure 11.

A priori local mask

Complementing $M_i^g(x,y)$, we next compute $M_i^l(x,y)$, the *a priori* local mask corresponding to image $I_i(x,y)$. $M_i^l(x,y)$ captures the local peculiarities of the compartments – nuclei or cell membrane – associated with their local structure and morphology.

First, $I_i(x,y)$ is filtered by applying a single iteration of gradient adaptive smoothing (GAS)[34,46],

$$\tilde{I}_i(x,y) = \frac{1}{N_i(x,y)} \sum_{\xi=-1}^{1} \sum_{\eta=-1}^{1} I_i(x+\xi, y+\eta) w_i(x+\xi, y+\eta), \quad \text{where} \quad N_i(x,y) = \sum_{\xi=-1}^{1} \sum_{\eta=-1}^{1} w_i(x+\xi, y+\eta),$$
$$w_i(x,y) = \exp\left[-\frac{d_i^2(x,y)}{2k_0^2}\right], \quad d_i(x,y) = \sqrt{\left[\frac{\partial I_i(x,y)}{\partial x}\right]^2 + \left[\frac{\partial I_i(x,y)}{\partial y}\right]^2}. \tag{2}$$

This GAS filtered image, $\tilde{I}_i(x,y)$ smooths the original image, $I_i(x,y)$, while preserving the local variations within and around cell nuclei and membranes. The local neighborhood is defined via a 3×3 kernel, w_i, that also performs variation preserving smoothing. Here, variation is quantified via computation of local gradient and the degree of smoothing is controlled by k_0, which is an algorithmic parameter. Its default setting is 1.

To obtain $M_i^l(x,y)$, a two-level, local Otsu is applied to $\tilde{I}_i(x,y)$ based on disk kernel whose radius r_0 is an algorithmic parameter. Its default setting is 5 pixels. The Otsu output faithfully captures the local structure but is also noisy, particularly in image regions where no tissue samples are present and the gradients are being computed on the background noise. As $M_i^g(x,y)$ can accurately identify such background, the output of the local Otsu is restricted to where $M_i^g(x,y) = 1$, resulting in local foreground mask $M_i^l(x,y)$.

An example of *a priori* local mask is presented in Supplementary Figure 11.

Processing stage 2: computing *a posteriori* nuclei and cell-membrane masks

The *a priori* global and local binary masks are computed independently for both channels. As a result, non-negligible probability exists for a pixel to be classified as being both in the nucleus and cell membrane. This is particularly true in tissue regions with crowded cells, or when the nature of the tissue section is such that cell membrane is laying over the nucleus. This processing stage reconciles these overlaps and generates *a posteriori* global and local nuclei and cell-membrane masks.

Contrast based likelihood function

Human visual perception of cell membrane and nuclei is based on inherent contrast between the two channels. Usually this contrast is visualized via imbuing the individual intensity based channels with colors. Here, we adapt this notion to compute a visual contrast function based on nucleus and cell membrane marker specific expression to quantify the likelihood of pixel belonging to either the nucleus or cell-membrane. The first step computes the contrast function for each pixel in the *a priori* local mask as follows,

$$L_0(x,y) = \begin{cases} \dfrac{I_2(x,y) - I_1(x,y)}{I_2(x,y) + I_1(x,y)}, & \textbf{if } I_1(x,y) > i_1 \textbf{ or } I_2(x,y) > i_2 \\ 0, & \textbf{otherwise} \end{cases},$$

where $i_i = \min_{(x,y)\in\Omega_i} I_i(x,y)$, $\Omega_i = \left\{(x,y) \in \Omega \mid M_i^l(x,y) = 1\right\}$, $i = 1,2$. The second step ensures that this function is consistent with the *a priori* global mask for each channel, resulting in the contrast based likelihood function,

$$\mathbf{L}(x,y) = \begin{cases} L_0(x,y), & \textbf{if } L_0(x,y) < 0 \textbf{ and } M_1^g(x,y) = 1 \textbf{ or } L_0(x,y) > 0 \textbf{ and } M_2^g(x,y) = 1 \\ 0, & \textbf{otherwise} \end{cases}, \tag{3}$$

$\mathbf{L}(x,y)$ is bounded between $[-1,1]$, with the contrast of -1 indicating the strong likelihood that the pixel (x,y) belongs to the nucleus, while 1 indicating the pixel most likely belongs to the cell membrane. Two examples of likelihood function are presented in Figure 1 and Supplementary Figure 11.

A posteriori global mask

We combine the *a priori* probability with the contrast based likelihood function to compute the *a posteriori* global mask $\mathbf{M}^g(x,y)$, such that $\mathbf{M}^g : \Omega \to \{0,1,2\}$, where the labels 0, 1, and 2 correspond to the background, nuclei, and cell membranes, respectively. However, before performing this combination, we enhance $P_i^e(x,y)$ as follows,

$$P_i^s(x,y) = \begin{cases} 1, & \textbf{if } M_i^l(x,y) = 1 \\ P_i^e(x,y), & \textbf{otherwise} \end{cases}, \tag{4}$$

where $i = 1,2$. This enhancement, saturates $P_i^e(x,y)$ – that is, sets $P_i^e(x,y) = 1$ – where the *a priori* local mask is 1. It ensures graceful performance of our algorithm in the global context, when computing *a posteriori* global mask $\mathbf{M}^g(x,y)$. We then

compute the *a posteriori* global probability $P_i^g(x,y)$, via $P_i^s(x,y)$-weighted convex combination of the likelihood and *a priori* belief,

$$P_1^g(x,y) = \begin{cases} P_1^s(x,y) + (1 - P_1^s(x,y))\ |\mathbf{L}(x,y)|, & \textbf{if } \mathbf{L}(x,y) < 0 \\ 0, & \textbf{otherwise} \end{cases},$$
$$P_2^g(x,y) = \begin{cases} P_2^s(x,y) + (1 - P_2^s(x,y))\ |\mathbf{L}(x,y)|, & \textbf{if } \mathbf{L}(x,y) > 0 \\ 0, & \textbf{otherwise} \end{cases}. \quad (5)$$

The final posterior global mask is obtained by either applying *k-means* clustering, with $k = 3$, or *arg max* operation[34] on $P_i^g(x,y)$, $i = 1,2$ (Eq. 5) to compute $\mathbf{M}^g(x,y)$. The default setting is *k-means* clustering. We note that *k-means* (or *arg max*) is performed under the constraint that pixel $(x,y) \in \Omega$ is assigned to the common background if both global probabilities have zeros values, i.e. $P_i^g(x,y) = 0$, $i = 1,2$. Examples of the *a posteriori* global mask are presented in Figure 1 and Supplementary Figure 11.

A posteriori local mask

We define the *a posteriori* local mask, $\mathbf{M}^l : \Omega \to \{0,1,2\}$, simply by restricting the *a priori* probability $P_i^e(x,y)$ to the local mask $M_i^l(x,y)$,

$$P_i^l(x,y) = \begin{cases} P_i^e(x,y), & \textbf{if } M_i^l(x,y) = 1 \\ 0, & \textbf{otherwise} \end{cases}, \quad (6)$$

where $i = 1,2$. This restriction allows us to optimally capture the local *a posteriori* structure of the nuclei and cell membranes in a self-consistent manner.

Similar to computing the global mask, we either apply *k-means* clustering (default setting) or *arg max* operation on $P_i^l(x,y)$, $i = 1,2$ (Eq. 6) to obtain the *a posteriori* local mask $\mathbf{M}^l(x,y)$. As mentioned above for the global mask, the same constraint for the common background is also applied here. Examples are presented in Figure 1 and Supplementary Figure 11.

Processing stage 3: nuclei segmentation

The *a posteriori* global and local masks provide a semantic segmentation of image pixels comprising the tissue into nuclei and cell membranes. This, and the following processing stages are designed to obtain every instance of individual nucleus and its cell from the semantic segmentation of the tissue. Specifically, in this stage, we first segment all nuclei, and use them as a basis to identify their cells in the next stage. These steps ensure that the nucleus and cell masks are internally consistent with the latter always bounding the former.

To segment nuclei we process the *a posteriori* global mask for the nuclei, $\mathbf{M}_{nuc}^g(x,y) := \mathbf{M}^g(x,y)|_{\text{label}=1}$ with help from the *a posteriori* local mask for the cell membrane, $\mathbf{M}_{cell}^l(x,y) := \mathbf{M}^l(x,y)|_{\text{label}=2}$. Particular examples of these two masks are presented in Supplementary Figure 11.

Convexity analysis

Nuclei segmentation begins with convex analysis of every connected component of $\mathbf{M}_{nuc}^g(x,y)$. As a part of this analysis, we compute area and the steepest concave point (SCP)[28] of every component. SCP is a boundary point of the component with the largest deviation from its convex hull. The area parameter allows us to filter out exceedingly small objects that are not nuclei, while SCP helps us determine if the component is nuclei cluster (NC) or not. The component is kept for further analysis only if the area of the component exceeds a_0. Otherwise it is removed. Each component that passes the area threshold, is either classified as an NC or non-NC depending on whether SCP is above or below the threshold d_0. Both a_0 – default set to 20 pixels – and d_0 – default value is 4 pixels – are the primary algorithm parameters (Supplementary Table 2). The non-NC components are statistically analyzed to obtain the initial segmentation masks for all individual nuclei, along with a small component (SC) list comprising of small convex objects that we are less confident about being nuclei.

Convexity analysis of $\mathbf{M}_{nuc}^g(x,y)$, is illustrated in Supplementary Figure 11.

Perturbed watershed and virtual cuts

We process the NC components using perturbed watershed (PW) and virtual cut (VC) algorithms that we have developed. Their goal is to partition the NC mask into individual nuclei.

PW steps are illustrated in Figure 2. Briefly, the NC component mask (Figure 2d) is first modified by $\mathbf{M}_{cell}^l$ (Figure 2e). Specifically, cuts are introduced in the NC component mask where the local cell membrane is indicated in the $\mathbf{M}_{cell}^l$ spatially-corresponding to the NC component (Figure 2f). We next apply distance transform (DT) on the modified NC component and use the resulting DT image (Figure 2g) to compute d_{avr} – the average of all non-zero DT values in the DT image. d_{avr} is used

to threshold the distance transform to identify n sub-regions with large DT values indicative of interior of the sub-regions – putative nuclei – making up the NC splitting (Figure 2h). Within every sub-region we identify a pixel with the maximal distance transform value as the watershed seed point for that sub-region. We perform watershed segmentation of NC based on these n seed points to obtain our initial estimate of the nuclei comprising the NC (Figure 2i). If these estimates are correct, then perturbing the markers does not affect segmentation of the NC. However, if the estimates are incorrect, then sub-region estimates are not stable on perturbation. We exploit this perturbation-based stability to identify the correct segmentation of the NC mask. Specifically, we perturb the marker location and recompute the watershed based segmentation. The perturbations are implemented by shifting each watershed marker location sequentially in the horizontal and vertical directions by $\pm\lfloor d_{avr}\rfloor$, resulting in four perturbations: $(x_j \pm \lfloor d_{avr}\rfloor, y_j)$ and $(x_j, y_j \pm \lfloor d_{avr}\rfloor)$ with $j = 1,\ldots,n$ (Figures 2j–2m). Here, $\lfloor\cdot\rfloor$ stands for the floor function. If during any of the four scenarios, the size of any of the n putative nuclei collapses to a point object with an area size bounded to a few pixels (Figures 2j, 2l, and 2m), we deem them as unstable and remove their corresponding seed points from the list of n seed points, and recompute the watershed based segmentation with the remaining seed points (Figures 2n). If the segmentation results remain stable for all four shifts, then the estimate is considered correct. To ensure that each of the segmented sub-regions are indeed nuclei and not smaller NCs, we recursively perform convexity analysis and PW on each sub-region. An example of this recursion is illustrated in Supplementary Figure 12.

The above recursive segmentation of an NC mask can sometimes result in a specific pathological situation, where the convex analysis identifies a sub-region as an NC, but PW does not segment it into sub-regions. For this specific scenario, we have developed the virtual cuts (VC) approach, where a virtual cut is defined through the SCP of the NC component mask to identify virtual sub-regions. We use 'virtual' to emphasize that this cut and the resulting sub-regions are only used to identify their respective watershed seed points based on which we perform the actual segmentation. The hypothesis driving the VC method is based on the idea of PW method: although the locations of the respective watershed markers identified using virtual cuts might not exactly coincide with their true locations, they do represent a perturbed version of the true location. Thus, they yield stable and accurate segmentation into the two sub-regions. These sub-regions follow the same recursive logic of the PW method detailed above. VC method is illustrated in Supplementary Figure 11.

Finally, we process the small components in the SC list in a context dependent manner, with small isolated SCs included in the final nuclei segmentation result. Multiple examples of nuclei segmentation are presented in Figures 1, 4, 6 and Supplementary Figures 7, 8, and 11. The contours of the nuclei are outlined in white.

Processing stage 4: cell segmentation

We segment cells via joint use of *a posteriori* global mask for the cell membranes $\mathbf{M}^g_{cell}(x,y) := \mathbf{M}^g(x,y)|_{\text{label}=2}$ and the segmented nuclei.

We begin by initializing the segmented cell mask as the segmented nuclei mask. The mask is then expanded till its boundary coincides with that of the closest cell membrane around it. It is possible that the cell membrane marker used for cell segmentation is not expressed by all cells. Therefore, for cells without any cell-membrane marker expression, the nucleus mask is morphologically dilated a small amount u_0 (1-10 pixels) to obtain an estimate of the cell membrane. u_0 with its 9 pixels default value is the third primary algorithm parameter (Supplementary Table 2). In the opposite scenario, where due to the nature of the tissue section, a cell is present with a membrane but without a nucleus, we utilize $\mathbf{M}^g_{cell}$. Specifically, the skeleton of $\mathbf{M}^g_{cell}$ is computed and subtracted from $\mathbf{M}^g_{cell}$ itself. This operation naturally reveals the cell membrane contour within $\mathbf{M}^g_{cell}$, which we identify via computing the Euler number of its connected component. When the Euler number is zero and the area of the connected component exceeds half of the average area of nuclei, the connected component is identified as the segmented cell. Examples of cell segmentation are presented in Figures 1, 4, 6 and Supplementary Figures 4-8 and 11, where the contours of the segmented cells are outlined in green.

Performance evaluation

To evaluate UNSEG performance and compare it with Cellpose and Mesmer results, we compared the predicted segmentation with the expert annotated ground truth. The comparison was made utilizing intersection over union (IoU) metric[35] to compute true positive (TP), false positive (FP) and false negative (FN) values for IoU threshold values ranging from 0 to 1. The threshold value indicates how much of an overlap between the predicted segmentation and ground truth is considered a match. TP, FP, and FN are next used to compute the F_1-score (or Dice coefficient) metric,

$$F_1 = \frac{2\,TP}{2\,TP + FP + FN}. \tag{7}$$

Data availability

The annotated dataset consisting of 75 fluorescence images of normal and pathologically changed tissue sections of eight organs of the human GI system will be made publicly available.

Code availability

The Python implementation of the UNSEG is available at https://github.com/uttamLab/UNSEG.git. Both Linux and Windows versions of Cellthon will be made publicly available at https://github.com/uttamLab/cellthon.git.

Author contributions

S.U conceived the idea. B.K and S.U. designed the overall UNSEG algorithm and planned the key steps. B.K. wrote the UNSEG code and performed the analysis. R.R performed the immunofluorescence labeling and acquired the imaging data. B.R, B.K, R.R, and S.U. identified the 75 images for GI dataset. P.B., P.S.G, A.S.S, and A.S. provided expert annotation of the GI dataset images. B.J.L, L.S. R.M.B, R.E.B, J.Y., L.Z., B.D., R.E.S. and A.S. helped in tissue sample acquisition, assay optimization, and data generation. B.K. and S.U wrote the manuscript. All authors reviewed and edited the manuscript before submission.

Funding

This project was supported in part by the National Institutes of Health through Grant Number UL1TR001857.

References

1. Wählby, C., Erlandsson, F., Bengtsson, E. & Zetterberg, A. Sequential immunofluorescence staining and image analysis for detection of large numbers of antigens in individual cell nuclei. *Cytometry* **47**, 32–41, DOI: https://doi.org/10.1002/cyto.10026 (2002). https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.10026.

2. Schubert, W. *et al.* Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat. Biotechnol.* **24**, 1270–1278 (2006).

3. Zrazhevskiy, P. & Gao, X. Quantum dot imaging platform for single-cell molecular profiling. *Nat. Commun.* **4**, 1619 (2013).

4. Gerdes, M. J. *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl. Acad. Sci.* **110**, 11982–11987, DOI: 10.1073/pnas.1300136110 (2013). https://www.pnas.org/doi/pdf/10.1073/pnas.1300136110.

5. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. methods* **11**, 417–422 (2014).

6. Lin, J.-R., Fallahi-Sichani, M. & Sorger, P. K. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat. Commun.* **6**, 8390 (2015).

7. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042, DOI: 10.1126/science.aar7042 (2018). https://www.science.org/doi/pdf/10.1126/science.aar7042.

8. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-cycif and conventional optical microscopes. *eLife* **7**, e31657, DOI: 10.7554/eLife.31657 (2018).

9. Radtke, A. J. *et al.* Ibex: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. *Proc. Natl. Acad. Sci.* **117**, 33455–33465, DOI: 10.1073/pnas.2018488117 (2020). https://www.pnas.org/doi/pdf/10.1073/pnas.2018488117.

10. Hickey, J. W. *et al.* Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* **19**, 284 – 295 (2022).

11. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L. & Weinberger, K. (eds.) *Advances in Neural Information Processing Systems*, vol. 25 (Curran Associates, Inc., 2012).

12. Goodfellow, I. J., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, Cambridge, MA, USA, 2016). http://www.deeplearningbook.org.

13. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *ArXiv* **abs/1505.04597** (2015).

14. Al-Kofahi, Y., Zaltsman, A. B., Graves, R., Marshall, W. A. & Rusu, M. A deep learning-based algorithm for 2-d cell segmentation in microscopy images. *BMC Bioinforma.* **19** (2018).

15. Schmidt, U., Weigert, M., Broaddus, C. & Myers, E. W. Cell detection with star-convex polygons. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2018).

16. Hollandi, R. *et al.* A deep learning framework for nucleus segmentation using image style transfer. *bioRxiv* (2019).

17. Yang, L. *et al.* Nuset: A deep learning tool for reliably separating and analyzing crowded cells. *PLoS Comput. Biol.* **16** (2019).

18. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2020).

19. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. biotechnology* **40**, 555 – 565 (2021).

20. Blazek, P. J. & Lin, M. M. Explainable neural networks that simulate reasoning. *Nat. Comput. Sci.* **1**, 607 – 618 (2021).

21. Stringer, C. & Pachitariu, M. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634 – 1641 (2022).

22. Lin, G. *et al.* A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytom. Part A* **56A** (2003).

23. Lin, G. *et al.* Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation of nuclei. *Cytom. Part A* **63A** (2005).

24. Li, G. *et al.* 3d cell nuclei segmentation based on gradient flow tracking. *BMC Cell Biol.* **8**, 40, 1–10 (2007).

25. Li, G. *et al.* Segmentation of touching cell nuclei using gradient flow tracking. *J. Microsc.* **231**, 47–58 (2008).

26. Coelho, L. P., Shariff, A. & Murphy, R. F. Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. *2009 IEEE Int. Symp. on Biomed. Imaging: From Nano to Macro* 518–521 (2009).

27. Lou, X., Koethe, U., Wittbrodt, J. & Hamprecht, F. A. Learning to segment dense cell nuclei with shape prior. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 1012–1018, DOI: 10.1109/CVPR.2012.6247778 (2012).

28. Qi, J. *et al.* Drosophila eye nuclei segmentation based on graph cut and convex shape prior. In *2013 IEEE International Conference on Image Processing*, 670–674, DOI: 10.1109/ICIP.2013.6738138 (2013).

29. Stoeger, T., Battich, N., Herrmann, M. D., Yakimovich, Y. & Pelkmans, L. Computer vision for image-based transcriptomics. *Methods* **85**, 44–53, DOI: https://doi.org/10.1016/j.ymeth.2015.05.016 (2015). Inferring Gene Regulatory Interactions from Quantitative High-Throughput Measurements.

30. Isack, H. N., Gorelick, L., Ng, K., Veksler, O. & Boykov, Y. K-convexity shape priors for segmentation. In *European Conference on Computer Vision* (2018).

31. Kostrykin, L., Schnörr, C. & Rohr, K. Globally optimal segmentation of cell nuclei in fluorescence microscopy images using shape and intensity information. *Med. Image Analysis* **58**, 101536, DOI: https://doi.org/10.1016/j.media.2019.101536 (2019).

32. Winter, M. R. *et al.* Separating touching cells using pixel replicated elliptical shape models. *IEEE Transactions on Med. Imaging* **38**, 883–893 (2019).

33. Gonzalez, R. C. & Woods, R. E. *Digital image processing* (Prentice Hall, Upper Saddle River, N.J., 2008).

34. Toennies, K. D. *Guide to medical image analysis* (Springer, 2017).

35. Caicedo, J. C. *et al.* Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry* **95**, 952 – 965 (2018).

36. Aleynick, N. *et al.* Cross-platform dataset of multiplex fluorescent cellular object image annotations. *Sci. Data* **10** (2023).

37. Maki, C. G. *p53 Localization*, 117–126 (Springer US, Boston, MA, 2010).

38. Huang, J. *et al.* The noncanonical role of ezh2 in cancer. *Cancer Sci.* **112**, 1376–1382, DOI: https://doi.org/10.1111/cas.14840 (2021). https://onlinelibrary.wiley.com/doi/pdf/10.1111/cas.14840.

39. Caron, M., Bojanowski, P., Joulin, A. & Douze, M. Deep clustering for unsupervised learning of visual features (2019). 1807.05520.

40. Feng, S., Wu, Z., Zhao, Y. & Vela, P. A. Image-based trajectory tracking through unknown environments without absolute positioning. *IEEE/ASME Transactions on Mechatronics* **27**, 2098–2106, DOI: 10.1109/tmech.2022.3175819 (2022).

41. Zheltonozhskii, E., Baskin, C., Mendelson, A., Bronstein, A. M. & Litany, O. Contrast to divide: Self-supervised pre-training for learning with noisy labels. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, DOI: 10.1109/wacv51458.2022.00046 (IEEE, 2022).

42. Van Rossum, G. *The Python Library Reference, release 3.8.2* (Python Software Foundation, 2020).

43. Chityala, R. & Pudipeddi, S. *Image Processing and Acquisition using Python* (Chapman and Hall/CRC, 2014).

44. Otsu, N. A threshold selection method from gray level histograms. *IEEE Transactions on Syst. Man, Cybern.* **9**, 62–66 (1979).

45. Liao, P.-S., Chen, T.-S. & Chung, P. C. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.* **17**, 713–727 (2001).

46. Saint-Marc, P., Chen, J.-S. & Medioni, G. Adaptive smoothing: a general tool for early vision. *IEEE Transactions on Pattern Analysis Mach. Intell.* **13**, 514–529, DOI: 10.1109/34.87339 (1991).