Cell Segmentation in Images Without Structural Fluorescent Reporters

Anonymous ECCV submission

Paper ID 9

Abstract. Computational methods for image-based profiling are under active development, but their success depends on assays that can maximize the phenotypic information captured. Fluorescent protein (FP) tags and other methods to fluorescently label proteins of interest provide a range of tools to investigate virtually any cellular process under the microscope. However, fluorescence microscopy is limited in the number of FPs that can be simultaneously imaged in the same cell. Cell segmentation methods often rely on the presence of morphological markers such as nuclei and cytoplasm. Here, we present a generalist approach to overcome the need for morphological reporters and that instead uses reporters that contain biological information for the segmentation of the nucleus and cytoplasm. Our method leverages state-of-the-art pre-trained segmentation models to segment cytoplasm and nuclei on fluorescence microscopy images. For this, we propose to fine-tune generalist networks for cell and nucleus segmentation for each individual fluorescent channel and to aggregate the respective segmentation results together. We assess our methodology performance and robustness across several fine-tuning strategies and fusion methods across different cell lines and various reporter proteins of a specific cellular signaling pathway, and for different acquisition methods. This approach will allow maximizing the extraction of relevant biological information to characterize cellular processes.

Keywords: cell segmentation, cell biology, image-based cellular assays, fluorescence microscopy

1 Introduction

Image-based cellular assays allow us to investigate cellular and population phenotypes and thus to understand biological phenomena with high precision. Single-cell sequencing computational analysis approaches have recently regained interest [12] notably due to parallel advances in sequencing technologies and data analysis techniques. Similarly, single-cell live bio-imaging analysis ideally complements sequencing-based approaches by allowing to finely measure and detect cell phenotypes and their evolution over time, both by detecting the activation of certain pathways as measured by the expression of specific fluorescent reporters, and/or by using agnostic measurements of cell evolution in the phenotypic space [3]. Cell instance segmentation is a key part of such bio-imaging

analysis pipelines as it allows us to study the cells at a single-cell level rather than at the the population level.

Deep learning models for cell instance segmentation have recently reached the quality of manual annotations [18,6,23], notably thanks to the emergence of models like U-Net [17]. Cellpose [23] is a notable U-Net based approach, which uses a multi-modal training dataset spanning several cell types and cell lines imaged under a variety of different imaging methods. It also benefits from being associated with a large community which incrementally increases the size and diversity of the dataset as well as improve the performance of the model. Cellpose approaches the problem of multiple instance segmentation by predicting spatial gradient maps from the images, from which individual cell segmentations can be inferred.

Although pre-trained models for Cellpose are available for both nucleic and cytoplasmic segmentation, they are respectively mostly trained on fluorescence microscopy images labeled with specific cellular compartments such as DAPI [9] for nuclei labeling or a fluorescent protein tagged with a Nuclear Export Signal (NES) [4] for cytoplasm labeling. Only about 15% of Cellpose training dataset contained other types of fluorescent labels [23]. While such structure-specific fluorescent reporters are generally integrated in assay designs, being able to segment cells without using them allows to maximise the number of experiment-specific reporters, freeing up the channels used for structural labeling marking. Since the total number of available channels is subject to numerous limitations – such as the bleed through effect [14] – and is typically limited to 4 the two extra channels made available can lead to assays delivering richer information about cellular processes.

In this paper we propose a new method to segment nuclei and cytoplasm without the need for specific reporters for nucleus and cytoplasm which requires very few annotations. It uses a pre-trained generalist deep learning segmentation base-model – here Cellpose – which is finetuned on each fluorescent reporter with a small set of annotated images. It augments the fluorescent signal information one can get from a single assay, similarly to multiplexing [21] without the constraints and complexity of such methods. We note that our approach is in line with a recent shift towards the "expertization" of generalist models like Cellpose [22], encouraging the prediction of a wider range of cellular image types and styles, with very small additional training from humans in the loop. However, while those recent approaches still only apply to cell images with structural fluorescent reporter proteins, we hereby propose an extension to non-structural fluorescent reporters, i.e. reporters not specifically optimized to highlight the structures we want to segment.

2 Data

The context of our work is a very flexible experimental framework, where we image different cell lines with different fluorescent markers reporting on a large variety of cellular processes. For this reason, we require a segmentation method

capable of handling any cell line observable through any set of fluorescent reporter proteins without specific or unique cellular localization.

Of note, experimental parameters such as temporal and spatial resolution can be changed, as well as acquisition conditions. Such changes can affect the type and level of noise in images and the appearance of cells. The method therefore has to be both scalable and robust to this kind of changes.

2.1 Image data

We acquired data on two commonly used cancer cell lines, namely U2OS [14] and A375 [1] each containing, respectively, 3 or 4 fluorescent reporter proteins tagged with spectrally distinct fluorophores that label different proteins of a cellular signaling pathway (see Table 1). Each of the proteins localize dynamically to different cell compartments and is not labeling any particular cell component. All the images were acquired using a Nikon A1R confocal microscope as a live video microscopy imaging sequence. For each experimental condition, we acquired images on 3 xy positions every 3 hours for 72 hours, with a resolution of 512×512 pixels.

FR-P	Channel Number	Cell Line
mCerulean-P1	1	U2OS, A375
Venus-P2	2	U2OS, A375
mCherry-P3	3	U2OS, A375
miRFP670-P4	4	A375

Table 1: Description of cell line components used in the assays - Fluorescent Reporter Protein (FR-P), Channel Number, Cell Line. Fluorescent reporters are tagged to different proteins (P) integrated in each of the cell lines.

2.2 Annotated dataset

We annotated the nuclei and cell boundary of 50 images for each cell line/reporter combination. These were randomly selected from different experimental conditions and evenly distributed over time in order to capture the dynamic localization of certain proteins (e.g. due to experimental treatments) and possible variations in population size caused by cell division or cell death. Each set of 50 images was then randomly split into three subsets: Training (size: 40), Validation (5), and Test (5). The number of cells per image ranges from 20 cells to more of a 100 in some images. With an average of 50 cells per images, we have about 250 cells per validation/test set and about 2000 individual cells in the training set, an appropriate number for training and evaluation. Example images and annotations are displayed in Figure 1.

Furthermore, to test for robustness of our method to changes in assay and acquisition parameters, we also annotated 5 extra images of the U2OS cell line

acquired with a different microscope (widefield) and higher temporal resolution, resulting into noisier images. This dataset is used for evaluation purposes only.

All datasets are made public and can be found at `https://figshare.com/s/2685a87a4fd93b8393ac`.

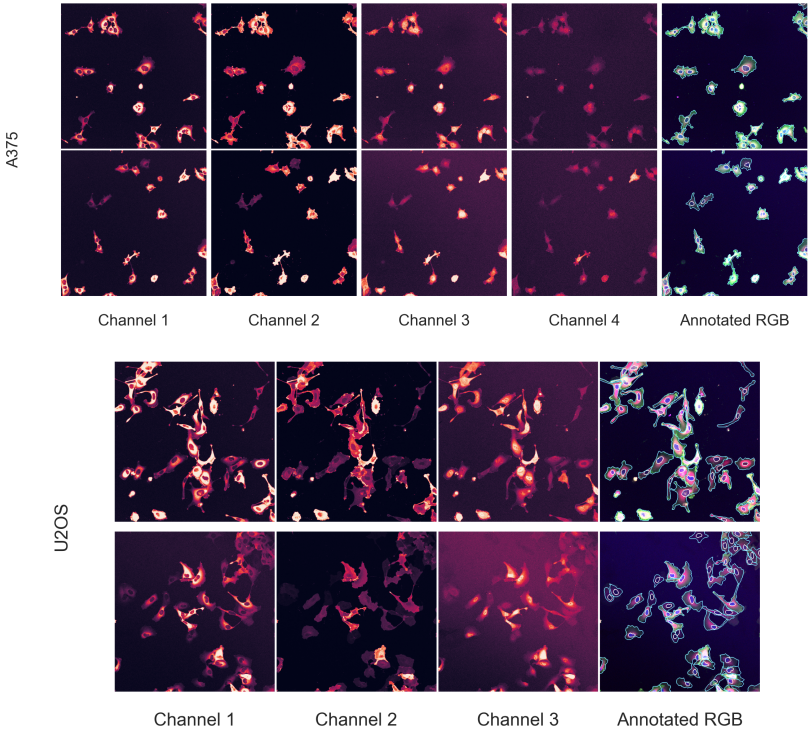

Fig. 1: Image samples from the different imaging assays showing individual channels as well as the RGB with segmentation annotations overlay.

3 Methods

3.1 Generalist Segmentation Model Backbone

Here, we use Cellpose [23] as a generalist segmentation model backbone, one of the most popular and most powerful segmentation models today. We note however, that our approach is model-agnostic and could therefore use alternative backbones [18,6] as well. Cellpose segments images using a three-fold pipeline.

First, it resizes the images so that the average cell diameter conform to the model training cells' diameter.

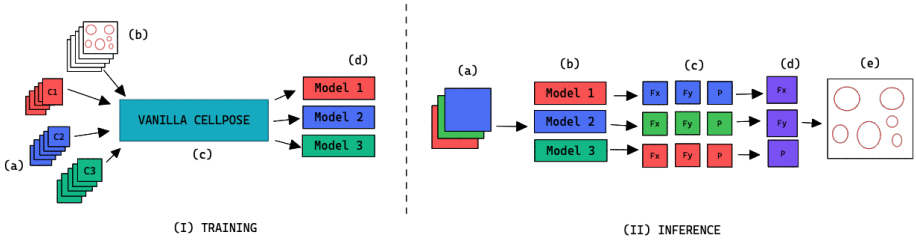

Fig. 2: Training and Inference workflow for the segmentation of cell organelles without the use of structural fluorescent reporter proteins. (I) Training: (a) Training set of multi-modal fluorescent images (3 channels represented as red blue and green). (b) Training set annotations of the organelles segmentations. (c) Out-of-the-box pre-trained Cellpose model. (d) Finetuned model for each of the individual channels. (II) Inference: (a) Multi-modal fluorescent image (3 channels). (b) Models selected from the model zoo corresponding to the image's cell line and reporter channel combination. (c) Spatial flows and probability maps outputted by the finetuned models for each of the channels. (d) Channel-wise averaging of the maps. (e) Segmentation output.

Secondly, for a cell object $o \in \{nuclei, cyto\}$, a Cellpose model M_o maps a rescaled image $\hat{I}$ intrinsic intensity space to a flow and probability space (F_X, F_Y, P). The flow maps F_X, F_Y are the derivatives (along the X and Y axes) of a spatial diffusion representation of individual cell pixels from the cell's center of mass to its extremities.

Lastly, Cellpose combines the (F_X, F_Y, P) flow and probability maps to predict instance segmentations S_o using flow analysis and thresholding on all three maps combined. First, the flows F_X and F_Y are interpolated and consolidated where the pixel-wise probabilities P are above a pre-set threshold. The instance masks are then generated by analysing the flows histogram from their peak. Cellpose overall segmentation process is summarized in the following equation:

$$M_o(\hat{I}) \rightarrow (F_X, F_Y, P) \xrightarrow[flow\ analysis]{} S_o, o \in \{nuclei, cyto\}. \quad (1)$$

3.2 Segmentation Model Finetuning Approach

Although Cellpose is optimized on a very large and generalist dataset, the fluorescence microscopy images used for training contain mostly structural reporter proteins. The trained network thus does not perform optimally on images like the ones presented in Section 2.2. It is therefore necessary to finetune Cellpose on our annotated data, such that the network can adapt to the specificities of our non-structural reporters, while still maintaining the general features learned from the heterogeneous Cellpose dataset.

To finetune Cellpose we selected a *Channel-Wise* strategy, offering the most accurate results as well as the highest level of versatility given the context of this

research: for each combination (cl, R_i) of a cell line cl and a fluorescent reporter protein $R_i \in \{R_i\}_{i=1}^{K}$, K the number of channels in the image, we finetune a Cellpose model to predict the cell segmentation. The different channel-wise predictions are fused together at inference. This method requires one model M_{o,cl,R_i} for each combination (cl, R_i).

$$M_{o,cl,R_i}(\hat{I}_k) \rightarrow (F_X^k, F_Y^k, P^k), \qquad k \in \{1, ..., K\} \tag{2}$$

Alternative strategies include a *Universal* strategy, where one finetunes a single model on all reporter channels, or a *Greyscale* strategy, where one trains a single model on channel averages for each cell line. We favor the Channel-wise strategy as it is more modular and adaptive. Indeed, in case a new reporter combination is chosen, we can still use previously trained segmentation networks. Also, combining channels before applying the segmentation network might lead to suboptimal results, as we would not make best use of the complementary information provided by different reporters.

3.3 Data Augmentation

Augmentation methods are used during training the Cellpose model to both virtually increase the size of our dataset as well as offer better generalization. They are performed iteratively from scratch on each image batch. For methods involving random distributions, the parameters are uniformly sampled from a pre-defined parameter range. Each augmentation has an application probability $p_{\text{augment}} = 0.5$, adding more variability across epochs and samples.

Classical augmentations include scaling with a distributed scaling factor $s \in [0.5, 1]$, rotation using a uniformly distributed rotation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, flipping along the X and Y axis given flipping probabilities $p_{flipX} = p_{flipY} = 0.5$.

Microscopy-related augmentations are also applied. Additive Gaussian White Noise (AGWN) mimics both the variation in the expression of the lit-up fluorescent pixels – the intensity of fluorescent expression – and random background noise [2], and is generated with mean $\mu = 0$ and standard deviation σ uniformly sampled from $[0, 0.1]$. Poisson Noise addition emulates the noise generated by the fluorescence microscope imaging and thus generalizes the trained model to a larger variance in microscope noise conditions [10][26]. Salt and Pepper noise, which randomly flips pixels to minimal or maximal brightness in the range of the microscope sensibility, mimics variation in the location and expression of the fluorescent reporters, by simulating activated or inhibited fluorescent proteins [24]. It is computed by generating two random maps, I_{salt} and I_{pepper} over the image I which follow a uniform distribution in the range $[0, 1]$. Finally, brightness augmentation increases or decreases the overall brightness of the image, thus mimicking the variance in microscope image acquisition as well as the variance in fluorescence intensity, which may differ from cell to cell and from assay to assay

[13]. The brightness transformation is applied to the image as a clipped intensity random shifting, with shift parameter Δ selected from a uniform distribution as $\Delta \sim \mathcal{U}(-0.1, 0.1)$.

3.4 Segmentation Model Finetuning Parameters

To finetune Cellpose, we train from the available generalist pre-trained model on our dataset using the channel-wise approach detailed in Section 3.2. The model training parameters are kept from the original Cellpose training with the following exceptions: firstly, as detailed in Section 3.3, we use non-deterministic augmentations on each of our training samples; secondly, we stop the training using early stopping on the validation set [25], with a patience of 20. Both additions are efficient regularization methods limiting over-fitting and contributing to the overall robustness of the segmentation methods with respect to changes in the imaging setting.

Furthermore, Cellpose is particularly sensitive to the cell object diameter parameter. We carefully select this parameter by assessing both the distribution of cell diameters in our annotated dataset and the performance of Cellpose models as a function of it. When using out-of-the-box Cellpose, we find that it is best to use the mean plus one standard deviation of our annotated dataset diameter distribution, overshooting the mean cell diameter and thus increasing the relative share of cells within the U-Net's receptive field. However, when using the diameter parameter in the case of our finetuned version of Cellpose, the segmentation quality does not vary with the diameter parameter as long as the parameter given between training and inference stays consistent. These findings make sense given that this parameter mainly affects the dimension of the image during the U-Net inference and therefore the scale of the cell within the U-Net's receptive field. It must therefore match the diameter of the training dataset to perform best.

3.5 Segmentation Fusion

Once the segmentations are generated for each of the image's channel using the finetuned models, a final segmentation for the image is generated by fusing the individual channel segmentation maps. We propose a method we name Flow Fusion (FF) to do so. We also consider several state-of-the-art methods. Fusion methods are described in Table 2.

The FF method uses Cellpose internal representations to aggregate the segmentation maps. Cellpose's 4-network averaging method allows to average for different models with various augmentation and tiling in order to boost the segmentation results. Similarly, FF averages the segmentation probability maps and flow maps obtained by running Cellpose on each channel individually to obtain a final aggregated segmentation map.

Given K channels in an image, our approach yields K segmentation maps $\{(F_X^k, F_Y^k, P^k)\}_{k=1}^K$ generated by the U-Net. From these representations we average the individual maps along the channel dimensions, yielding maps $(\bar{F_X}, \bar{F_Y}, \bar{P})$.

Method	Description	Reference
Flow Fusion (FF) (Ours)	Fusion of flow maps followed by flow analysis	X
Selective and Iterative Method for Performance Level Estimation (SIMPLE)	Iterative majority voting over propagated segmentations, weighted by estimated performance	[11]
Simultaneous Truth and Performance Level Estimation (STAPLE)	Statistical fusion framework using hierarchical models of rater performance	[1]
Voting (V)	Pixel-wise voting	[16]
Majority Voting (MV)	Majority label voting in images patches	[7]

Table 2: Description of the Segmentation fusion methods considered to generate aggregated segmentations from our channel-wise segmentations

These channel-wise mean representations are then transformed into instance segmentation maps using Cellpose's mask generation method.

In order to generate the best segmentations given a specific set of fluorescent reporters using our finetuned model, we evaluate the aggregation methods over the power-set of channel combinations and select the channel combinations which yield the best score on our testing set.

3.6 General Workflow

In the previous sections, we described in some detail the core of our algorithm to leverage non-structural fluorescent markers, which is also the major contribution of the paper. For completeness, we briefly present here the complete pipeline in which this strategy is embedded. This workflow includes the pre-proccessing, segmentation (divided in inference and aggregation) and post-processing. The general workflow is illustrated in Figure 2.

Preprocessing We present hereafter several pre-processing methods that we apply on our data.

We apply a normalization by percentile, i.e. scale the images' intensity for each channel between the 1st and 99th percentile of the dataset's intensity range. It allows for scaling most of pixel intensity values between 0 and 1 while keeping outliers outside of this range.

CIDRE [20] (Corrected Intensity Distributions using Regularized Energy minimization) is an illumination correction method which corrects the image by removing the background and balancing out the luminosity of objects in the image space and without the need of reference images. CIDRE improved the overall quality of segmentation on our dataset.

For widefield imaging, we apply additional denoising steps involving classical operators such as flat field correction [19] and Contrast Limited Adaptive Histogram Equalization (CLAHE) [15].

Segmentation At inference, we run a forward pass of the Cellpose finetuned model for each channel and for each object. This way we obtain K segmentations (given K reporter channels) for each organelle which are then aggregated using one of the proposed fusion methods.

A subset of the K reporter channels segmentation are then selected to be aggregated together using our fusion method. The set of reporter channels to aggregate together are selected at training. Variation in the channels to aggregate may arise if the image acquisition method changes, or if the assay parameter may alter the quality or reliability of a specific reporter channel. We have built a pipeline which allows us to train new models for previously unseen cell-line/reporter combination or select models and reporter channels to fuse when they are available in our model zoo.

4 Results

4.1 Evaluation Metrics

We use as binary segmentation metric the Jaccard similarity [8], which measures the ratio of intersection-over-union (IoU) of object and background binary segmentations. It varies between 0 for no overlap and 1 for perfect overlap.

For instance segmentation evaluation, we use the Precision, Recall and F1-score. A predicted segmentation is considered as a true positive if the IoU between this segmentation and a ground truth segmentation is above a threshold here set at 0.5. Intuitively, Precision represents the correctness of our predictions, Recall represents their completeness and F1-score is the harmonic mean of the two.

4.2 Evaluation

Segmentation Fusion Table 3 conveys the performance of the different segmentation fusion methods tested in this work. The benchmarking is shown on the aggregation of all channels for each image, on both nuclei and cytoplasm. The results clearly indicate a better performance when using the Flow Fusion (FF) method introduced in Section 3.5: FF always appears as the best fusion method among the ones benchmarked here, in some cases with a large margin. We believe that this result is owed to the fact that the diffusion maps are optimized representations combining information about the pixel-level probability and about object-level shape properties. This makes them particularly useful for fusion and conveys them an advantage over raw image fusion and fusion of segmentation maps. We therefore set this method as the default fusion method, and present our method's evaluation using this method only in the remaining of the paper.

Cell Line	Evaluation set	FF (Ours)	SIMPLE	STAPLE	V	MV
U2OS	**Cytoplasm**, Channels 1, 2, 3	**0.8392**	0.8178	0.7858	0.8182	0.8182
	Nuclei, Channels 1, 2, 3	**0.9169**	0.816	0.816	0.816	0.816
A375	**Cytoplasm**, Channels 1, 2, 3, 4	**0.8955**	0.8587	0.83	0.814	0.8935
	Nuclei, Channels 1, 2, 3, 4	**0.9542**	0.9415	0.864	0.8601	0.8273

Table 3: Comparison of the performance of the different channel fusion methods on the testing set images, assessed using the F1-score.

Cell Line	Channel(s)	Jaccard	F1-Score	Precision	Recall
U2OS Cyto	**Vanilla Cellpose**	0.7492	0.8120	0.8182	0.8114
	Channel 1	**0.7848**	**0.857**	0.8568	**0.8601**
	Channel 2	0.7475	0.7877	0.7867	0.7915
	Channel 3	0.7673	0.8446	0.8381	0.8569
	Channel 1, 2, 3	0.7729	0.8392	0.8301	0.8526
	Channel 2, 3	0.766	0.8465	**0.8735**	0.8226
A375 Cyto	**Vanilla Cellpose**	0.7482	0.7905	0.7908	0.8064
	Channel 1	0.8244	0.8909	0.8918	0.9036
	Channel 2	0.8425	0.8764	0.8854	0.8764
	Channel 3	0.8364	0.8798	0.8804	0.8884
	Channel 4	0.8167	0.8672	0.9014	0.8378
	Channel 1, 2, 3, 4	**0.8439**	0.8955	0.9015	0.8982
	Channel 1, 2	0.8388	**0.8996**	0.9025	**0.9088**
	Channel 3, 4	0.8352	0.884	**0.9048**	0.869
U2OS Nuclei	**Vanilla Cellpose**	0.0570	0.1236	0.2068	0.0910
	Channel 1	**0.6911**	**0.9388**	**0.9289**	**0.9506**
	Channel 2	0.4358	0.7089	0.8251	0.63
	Channel 3	0.4774	0.7675	0.8315	0.7146
	Channel 1, 2, 3	0.6406	0.9261	0.9272	0.9262
	Channel 2, 3	0.5519	0.8516	0.8954	0.8146
A375 Nuclei	**Vanilla Cellpose**	0.1637	0.289	0.3067	0.3125
	Channel 1	**0.743**	**0.9684**	**0.9627**	**0.9742**
	Channel 2	0.6396	0.9106	0.9107	0.9111
	Channel 3	0.5502	0.8352	0.83	0.8468
	Channel 4	0.4302	0.6911	0.8737	0.633
	Channel 1, 2, 3, 4	0.6659	0.9542	0.9433	0.9661
	Channel 2, 3, 4	0.6045	0.9266	0.9064	0.9496

Table 4: Evaluation scores for the cytoplasm and nuclei segmentation on U2OS and A375 cell lines, assessing Vanilla Cellpose segmentations, individually fine-tuned channels, all channels aggregated and the best channel combination (underlined) as well as its complementary channel combination.

Segmentation Table 4 displays the performance of our method through the different evaluation metrics. In those tables we compare the score of individual channels, aggregated channels, best scoring channel combination aggregation and its complementary channel combination aggregation. We also include Vanilla Cellpose results, generated using pre-trained Cellpose without fine-tuning on each channel individually followed by aggregation by flow fusion (as detailed in 3.5). Our method consisting in fine-tuning and FF aggregation out-performs this Vanilla Cellpose both for cytoplasm and nuclei segmentation.

The results of the segmentation using the Cellpose fine-tuning method conveyed in the tables show that fine-tuning is an essential step when dealing with datasets which do not contain cytoplasmic and nucleic structural fluorescent reporters. With fine-tuning, we obtain excellent results for both cell lines on individual channels and even better in some instances with our channel fusion approach. We show that combining the some of the different channel-wise segmentations into an aggregated segmentation outperforms the individual channel segmentations in some cases. It is also noticeable that complementary channel or channel aggregations to the best performing one still significantly outperforms vanilla Cellpose. Examples of our segmentation are displayed in Figure 3.

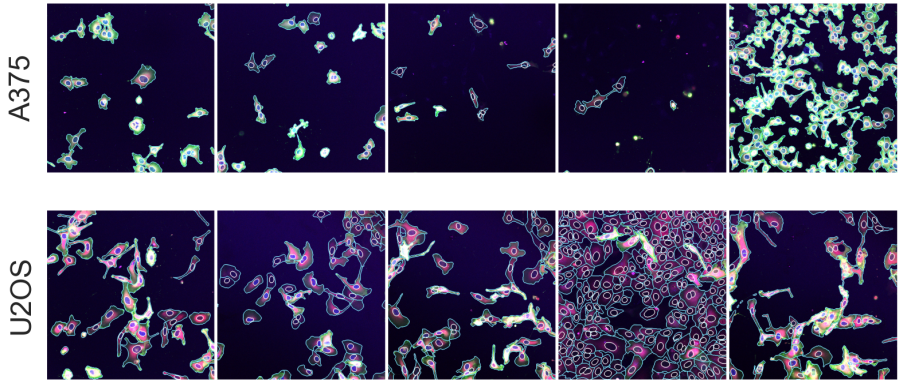

Fig. 3: Example of segmentations using the proposed method on the Testing set images for both the A375 and U2OS cell lines.

Impact of microscopy-related data augmentation on model generalization We test the impact of the microscopy-related data augmentation introduced in Section 3.3 on model generalization to other imaging acquisition procedures. To do so, we compare the performances of two models trained on U2OS cells with acquisition setup $A1$, resp. with and without augmentations, when applied to 5 images of the same cell line acquired with a setup $A2$ leading

to noisier images. $A1$ was acquired using a confocal microscope at a 512×512 resolution, whereas $A2$ was acquired at the same resolution but with a widefield microscope and with a higher temporal resolution. Results are presented in table 5 and demonstrate that models trained with microscopy-related augmentations generalize better to different experimental conditions and acquisition procedures.

Object	Method	Jaccard	F1-score	Precision	Recall
Cytoplasm	**Vanilla Cellpose**	0.3521	0.5029	**0.9062**	0.3493
	Without augmentations best channels(s) : 3	0.6157	0.7288	0.6858	**0.7824**
	With augmentations best channel(s) : 3	**0.6955**	**0.797**	0.8174	0.7797
Nuclei	**Vanilla Cellpose**	0.0768	0.0867	0.2600	0.0526
	Without augmentations best channel(s) : (1,3)	0.4413	0.7134	0.9086	0.5882
	With augmentations best channel(s) : (1,3)	**0.4548**	**0.7233**	**0.9276**	**0.5941**

Table 5: Segmentation scores for a model trained with and without microscopy-specific data augmentations on U2OS cells acquired with a confocal microscope when evaluated on U2OS cells acquired with a widefield microscope ($A2$).

5 Discussion

Our work aimed at providing a reliable, re-usable and scalable method for the segmentation of cell images without structural markers, with manageable annotation effort. Our results show that fine-tuning and aggregation of diffusion maps outperform generalist models when evaluated on datasets with no structural fluorescent reporters. The channel-wise finetuning approach presented in this paper leads to the development of a model to segment both nuclei and cytoplasm with very few annotated examples, benefiting largely from the generalist pre-training of the Cellpose backbone model. As can be seen in the A375 cytoplasm segmentation results presented in Table 4, the aggregated channel-wise approach benefits from the complementarity of the different reporter channels in outlining individual cell objects albeit their non-structural roles. Each channel covers incomplete - yet overlapping - areas of the individual cells, which can be regrouped into integral individual cell using our method. Moreover, due to the singular phenotypic nature of individual cells, over time the signaling pathways or morphological characteristics highlighted by the non-structural reporters may translocate and dynamically lit-up parts of the cell. Hence using several fluorescent reporters in concert allows our approach to segment cells across reporters, some of which may not appear at all in some channels. Our approach therefore allows to segment cells across reporters that may be complementary at the pixel scale and at the cell instance scale.

On that matter, we note that channel 1 is the dominant channel throughout our segmentation evaluation. This is most certainly due to this fluorescent protein reporter not changing localization from the cytoplasm and showing the nuclei excluded from the cytoplasm, as can be seen in Figure 1. Channel 1's reporter is hence a reliable reporter for segmentation, the individual use of which is favored by our pipeline. Would that channel be removed from the cell lines in favor of a new reporter with a dynamic or inconsistent localization, the fusion of the remaining reporters channel would produce better segmentation than those channel's individual segmentation. Indeed, we observe in Table 4 that F1-scores are consistently better for segmentation using fusion over channels 2, 3 (and 4 for the A375 cell line) over models using these channels separately.

It must however be noted that one must be careful with the use of non-structural reporters when using the approach presented in this paper. Indeed, some reporters under the influence of compound introduced in the assay regimen, may differ in phenotypes from the images the models were pre-trained with. Although it may be noted that this also holds for structural reporters, as the performance of a generalist network in the presence of morphological phenotypes has not yet been investigated. We can for example imagine a compound which would make a protein translocate from the cytoplasm to the nuclei, thus making a finetuned model that learned nuclei as "negatives" incompetent at the segmentation task. Although our method encourages the creation of a pre-trained model zoo for each cell line/reporter combination to be used on any new assay answering to these characteristics, we recommend to wisely choose a combination of channels to aggregate which takes into account the evolution of the reporters with the introduced compound, and ideally to make sure that at least one channel would be reliable and coherent with the behaviour of the training data, otherwise re-finetuning would be needed. It is also interesting to note, that given the results presented in Table 4, often single channels perform best, or at least satisfyingly well. It is therefore possible to use of the channel-wise pre-trained model on any new assay as long as it has one channel in common with a previous assay having undergone our finetuning method.

Our approach is limited by the same limitations as Cellpose. Although it works as an "expertization" of generalist models, the same way [22] encourages a specialization of generalist models, it is still constrained by the same limitations. For example, it does not handle occluded or overlapping cells. It is also susceptible to merging or splitting of individual cell instances which could only be corrected with a robust post-processing step. It is also limited in terms of building cell objects, as – like Cellpose – it detects nuclei and cytoplasm independently, thus yielding standalone cytoplasm and nuclei.

In theory, some of those limitations can be avoided by using a different backbone than Cellpose, one which would handle such issues. Other notable recent deep learning segmentation methods for cellular biology are StarDist [18] and NucleAIzer [6] which make use of the U-Net architecture as well but with different representation for their images, Mask-RCNN [5] which segments ROIs in Images with convolutional deep learning methods. We have made the choice of

using Cellpose in this paper as it yielded the best out-of-the-box results on our data, and offered to be the most generalist with its community driven, ever-expanding training dataset. However, it is possible to apply our methodology with different Deep Learning architectures using our Channel-wise finetuning approach, although changes to the fusion method would be required.

6 Conclusion

In this work, we have proposed a method to train and infer segmentations of nuclei and cytoplasm on images without structural fluorescent reporters and for a variety of cell line/reporter configurations. We demonstrated that our approach could be used for any assay while freeing up bandwidth space for two experiment-specific reporter proteins, where the use of structural reporters take space. We have shown that, in absence of the cytoplasmic and nucleic reporters, state-of-the-art segmentation performance can be achieved and replicated on various assays from a single finetuned pre-trained model on any cell line/reporter combination.

Our method is easily adaptable to fit a generalist image processing pipeline to be applied on various assays by sharing a common segmentation model zoo trained once on a cell line/reporter combination. Such a zoo of fine-tuned models will greatly support microscopy based cellular assays, and High Content Screening.

References

1. Asman, A.J., Dagley, A.S., Landman, B.A.: Statistical label fusion with hierarchical performance models. In: Ourselin, S., Styner, M.A. (eds.) SPIE Proceedings. SPIE (Mar 2014). https://doi.org/10.1117/12.2043182
2. Boulanger, J., Kervrann, C., Bouthemy, P., Elbau, P., Sibarita, J.B., Salamero, J.: Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. IEEE Transactions on Medical Imaging **29**(2), 442–454 (Feb 2010). https://doi.org/10.1109/tmi.2009.2033991
3. Chandrasekaran, S.N., Ceulemans, H., Boyd, J.D., Carpenter, A.E.: Image-based profiling for drug discovery: due for a machine-learning upgrade? Nature Reviews Drug Discovery **20**(2), 145–159 (Dec 2020). https://doi.org/10.1038/s41573-020-00117-w
4. la Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., Brunak, S.: Analysis and prediction of leucine-rich nuclear export signals. Protein Engineering, Design and Selection **17**(6), 527–536 (Jun 2004). https://doi.org/10.1093/protein/gzh062
5. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
6. Hollandi, R., Szkalisity, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., Kovacs, M., Migh, E., Goodman, A., Balassa, T., Koos, K., Wang, W., Caicedo, J.C., Bara, N., Kovacs, F., Paavolainen, L., Danka, T., Kriston, A., Carpenter, A.E., Smith, K., Horvath, P.: nucleAIzer: A parameter-free deep learning framework for nucleus segmentation using image style transfer. Cell Systems **10**(5), 453–458.e6 (May 2020). https://doi.org/10.1016/j.cels.2020.04.003

7. Huo, J., Wang, G., Wu, Q.M.J., Thangarajah, A.: Label fusion for multi-atlas segmentation based on majority voting. In: Lecture Notes in Computer Science, pp. 100–106. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-20801-5$_1$1
8. Jaccard, P.: The distribution of the flora in the alpine zone. New Phytologist **11**(2), 37–50 (Feb 1912). https://doi.org/10.1111/j.1469-8137.1912.tb05611.x
9. Kapuscinski, J.: DAPI: a DNA-specific fluorescent probe. Biotechnic &amp Histochemistry **70**(5), 220–233 (Jan 1995). https://doi.org/10.3109/10520299509108199
10. Laine, R.F., Jacquemet, G., Krull, A.: Imaging in focus: an introduction to denoising bioimages in the era of deep learning. The International Journal of Biochemistry & Cell Biology **140**, 106077 (2021)
11. Langerak, T.R., van der Heide, U.A., Kotte, A.N.T.J., Viergever, M.A., van Vulpen, M., Pluim, J.P.W.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Transactions on Medical Imaging **29**(12), 2000–2008 (Dec 2010). https://doi.org/10.1109/tmi.2010.2057442
12. Martins, B.M., Locke, J.C.: Microbial individuality: how single-cell heterogeneity enables population level strategies. Current Opinion in Microbiology **24**, 104–112 (2015). https://doi.org/https://doi.org/10.1016/j.mib.2015.01.003
13. Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F., Zoccoli, A.: Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. Scientific Reports **11**(1) (Nov 2021). https://doi.org/10.1038/s41598-021-01929-5
14. Pawley, J.B. (ed.): Handbook Of Biological Confocal Microscopy. Springer US (2006). https://doi.org/10.1007/978-0-387-45524-2
15. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing **39**(3), 355–368 (Sep 1987). https://doi.org/10.1016/s0734-189x(87)80186-x
16. Rohlfing, T., Maurer, C.R.: Multi-classifier framework for atlas-based image segmentation. Pattern Recognition Letters **26**(13), 2070–2079 (Oct 2005). https://doi.org/10.1016/j.patrec.2005.03.017
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
18. Schmidt, U., Weigert, M., Broaddus, C., Myers, G.: Cell detection with star-convex polygons. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, pp. 265–273. Springer International Publishing (2018). https://doi.org/10.1007/978-3-030-00934-2$_3$0
19. Seibert, J.A., Boone, J.M., Lindfors, K.K.: Flat-field correction technique for digital detectors. In: III, J.T.D., Boone, J.M. (eds.) SPIE Proceedings. SPIE (Jul 1998). https://doi.org/10.1117/12.317034
20. Smith, K., Li, Y., Piccinini, F., Csucs, G., Balazs, C., Bevilacqua, A., Horvath, P.: CIDRE: an illumination-correction method for optical microscopy. Nature Methods **12**(5), 404–406 (Mar 2015). https://doi.org/10.1038/nmeth.3323
21. Stack, E.C., Wang, C., Roman, K.A., Hoyt, C.C.: Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis. Methods **70**(1), 46–58 (Nov 2014). https://doi.org/10.1016/j.ymeth.2014.08.016
22. Stringer, C., Pachitariu, M.: Cellpose 2.0: how to train your own model. bioRxiv (2022). https://doi.org/10.1101/2022.04.01.486764

23. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature Methods **18**(1), 100–106 (dec 2020). https://doi.org/10.1038/s41592-020-01018-x
24. Tofighi, M., Kose, K., Cetin, A.E.: Denoising images corrupted by impulsive noise using projections onto the epigraph set of the total variation function (PES-TV). Signal, Image and Video Processing **9**(S1), 41–48 (Oct 2015). https://doi.org/10.1007/s11760-015-0827-8
25. Yao, Y., Rosasco, L., Caponnetto, A.: On early stopping in gradient descent learning. Constructive Approximation **26**(2), 289–315 (Apr 2007). https://doi.org/10.1007/s00365-006-0663-2
26. Zhang, Y., Zhu, Y., Nichols, E., Wang, Q., Zhang, S., Smith, C., Howard, S.: A poisson-gaussian denoising dataset with real fluorescence microscopy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11710–11718 (2019)