Learning with minimal effort: leveraging in silico labeling for cell and nucleus segmentation

Anonymous ECCV submission

Paper ID 5

Abstract. Deep learning provides us with powerful methods to perform nucleus or cell segmentation with unprecedented quality. However, these methods usually require large training sets of manually annotated images, which are tedious — and expensive — to generate. In this paper we propose to use In Silico Labeling (ISL) as a pretraining scheme for segmentation tasks. The strategy is to acquire label-free microscopy images (such as bright-field or phase contrast) along fluorescently labeled images (such as DAPI or CellMask™). We then train a model to predict the fluorescently labeled images from the label-free microscopy images. By comparing segmentation performance across several training set sizes, we show that such a scheme can dramatically reduce the number of required annotations.

Keywords: Segmentation · Transfer learning · Pretext task · In Silico Labeling · Fluorescence microscopy

1 Introduction

Detection and segmentation of cells and nuclei, among other cell structures, are essential steps for microscopy image analysis. Deep Learning has provided us with very powerful methods to perform these segmentation tasks. In particular, recently published neural networks, such as NucleAIzer [1], Cellpose [2] or StarDist [3], trained on hundreds of images of different modalities, give excellent results, outperforming by far traditional methods for image segmentation. However, the main drawback of state-of-the-art networks is the need for large amounts of fully annotated ground truth images, which can take a significant amount of time to create. Here, we present an alternative strategy, where we pretrain our segmentation models using In Silico Labeling (ISL) before fine-tuning them on a very small data set to perform nucleus and cell segmentation.

ISL was first introduced by [4], aiming to predict fluorescent labels from bright-field inputs. Fluorescence microscopy is the major technique employed in cellular image-based assays, as the use of fluorescence labels allows to highlight particular structures or phenotypic cell states. However, the number of fluorescent labels is limited (typically up to 4). In addition, phototoxicity and photobleaching can also represent serious drawbacks. In the same line, several variants have been proposed since [5, 6, 7, 8, 9, 10]. The principle of ISL has also been proposed for experimental ground truth generation for training cell

classifiers for the recognition of dead cells [11, 12], tumour cells [13] embryo polarization [14] or the cell cycle phase [15].

In this paper we show that models trained to generate fluorescence microscopy images with nuclear or cytoplasmic markers can be used efficiently to pretrain segmentation networks for nuclear and cell segmentation, respectively. This provides us with a powerful strategy to minimize the annotation burden for a given application, and to train models on large data sets, requiring only minimal effort in terms of manual annotation.

2 Materials and Methods

2.1 Image Acquisition

We work on two different data sets. The first dataset ($N = 960$) has been generated by the Opera system. For each position, we acquired bright-field images and DAPI, both at 4 different focal planes. DAPI is a very common fluorescent stain binding to AT-rich regions of the DNA, which can thus be used to locate the nucleus in eukaryotic cells. Additionally we have a phase-contrast image, computationally created from the 4 bright-field images by a proprietary algorithm of the Opera system.

Our second data set contains 100 images of dimensions (1024, 1024). We used Differential Interference Contrast (DIC) as label-free microscopy technique, and we marked the cytoplasmic membrane with the CellMask™marker (Cy5).

2.2 Nucleus Segmentation

Nucleus segmentation is one of the most important segmentation tasks in biology, as nuclear morphologies are indicative of cellular states, and because they are visually very different from the cytoplasm. In particular in fluorescence microscopy, segmentation of the nucleus is usually a fairly easy segmentation task, and for this reason we assumed that this might be a good first segmentation problem to investigate our ISL-based pretraining.

DAPI prediction as pretraining task The first step of our strategy for nucleus segmentation is the prediction of DAPI images from bright-field inputs.

We used a data set of 421 images of dimensions (2160, 2160), divided into 384 images for training and 37 images for testing. 5 images of dimensions (512, 512) were randomly cropped from each initial image (see Fig.1a, Fig.1b, Fig.1c). Note that we only included images containing at least one nucleus.

Inspired by the work of [4], the model is a U-net-shape model (Fig. 1d) with a densenet121 architecture [16]. It has been previously trained on ImageNet [17], hence it is referred to as 'on steroids' in the following. As input we used 3 channels, 2 being bright-field images of the same field of view with different focal planes, and the third the corresponding phase-contrast image. As output

we used only one channel, the maximum intensity projection of our DAPI images over the 4 z-stacks available in our data set.

The model was trained on 1000 epochs with an ADAM optimizer, a learning rate of 0.1 and L1 as loss function.

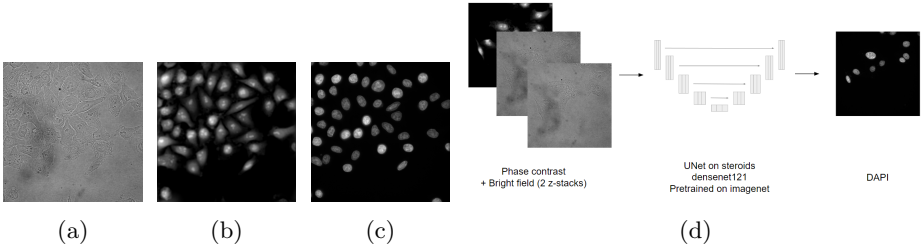


Fig. 1: (a-c) Images from the same field-of-view, for a given z-stack (a) Bright-field image (b) Phase contrast image (c) Fluorescent DAPI image (d) U-net model used to predict DAPI image from both bright-field and phase contrast images.

Transfer Learning for Nucleus Segmentation. In a first step, we aimed at investigating how pretraining on fluorescent markers impacts semantic segmentation. For this, we turned to nucleus segmentation.

In order to generate the ground truth, we applied Cellpose [2], a widely used segmentation technique in bioimaging, based on a U-net-shaped network, trained on massive amounts of heterogeneous data. We applied Cellpose to the DAPI channel and corrected the segmentation results manually. As segmentation of nuclei from DAPI images with high resolution is a fairly simple task, the results were overall excellent, as expected (Fig.2a, Fig.2b).

Next, we used training sets with different sizes $N \in \{1, 10, 50, 100, 200, 500\}$, composed of images of dimension (2160, 2160) and evaluated the accuracy for each N. Testing is always performed on the same withheld 190 images. 5 images of dimensions (512, 512) were randomly cropped from each initial image.

To investigate whether our pretraining scheme is useful for segmentation, we compare two different models. The first model is composed of the U-net 'on steroids' followed by a sigmoid activation function in order to output, for each pixel, its probability of belonging to a nucleus (Fig.2c). The second model has the same U-net architecture but is pretrained on DAPI images, and has an activation function displayed in (1) that takes the different range into account (centered on 0.5), (Fig.2d).

$$f(x) = \frac{1}{1 + \exp(x - 0.5)} \tag{1}$$

The number of epochs during training depends on the size of the training set. Both models are trained with an ADAM optimizer, a learning rate of 0.01 and Jaccard loss as loss function.

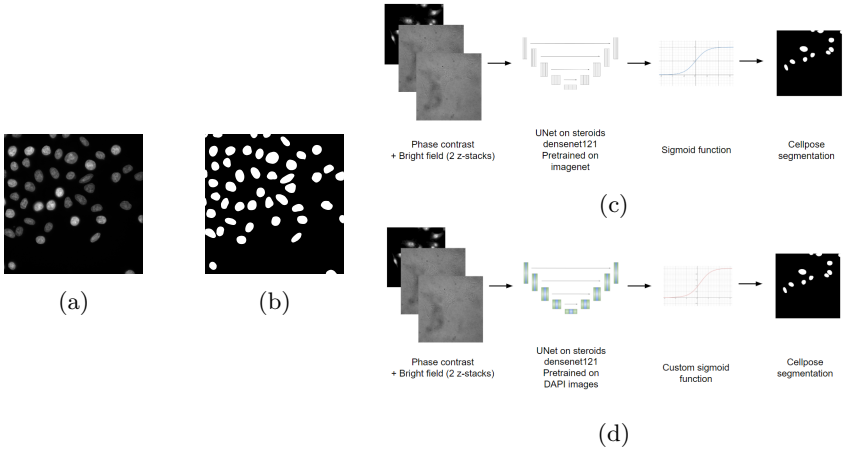


Fig. 2: (a) Fluorescent DAPI image (b) Corresponding nucleus semantic segmentation image generated by Cellpose. (c - d) Models compared to predict nucleus semantic segmentation image: (c) U-net 'on steroids' which has not been trained on DAPI images. (d) U-net 'on steroids' pretrained on DAPI images. Note the difference in the activation functions as well.

2.3 Cell Segmentation

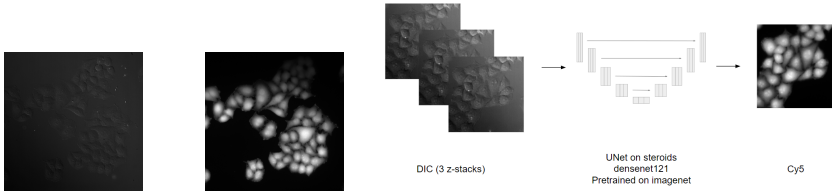
We next turned to the application of our pretraining scheme to cell segmentation, a more difficult multiple instance segmentation scenario.

CellMask™Prediction as Pretraining Task In our pretraining strategy, the first step of cell segmentation is the prediction of CellMask™(Fig.3b) images from DIC microscopy as inputs (Fig.3a).

We used a data set of 100 images of dimensions (1024, 1024), divided into 90 images for training and 10 images for testing. 5 images of dimensions (512, 512) were randomly cropped for each initial image.

For comparison, we again used the U-net 'on steroids' (Fig.3c). The model was trained on 1000 epochs with an ADAM optimizer, a learning rate of 0.1 and L1 as loss function.

Transfer Learning for Cell Segmentation. Segmentation of cells is usually more difficult than nuclear segmentation, because cells tend to touch each other,



(a) (b) (c)

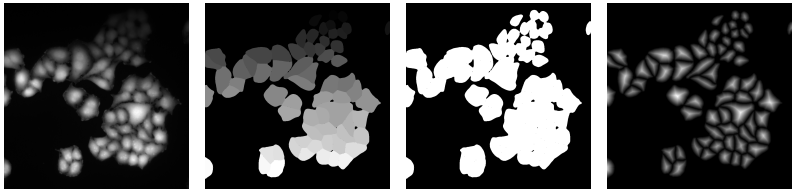
Fig. 3: (a-b) Images from the same field-of-view, for a given z-stack (a) DIC image (b) Fluorescent Cy5 image. (c) U-net model used to predict Cy5 image from DIC images.

and the precise detection of the contact line can be challenging. Indeed, we need to turn to multiple instance segmentation, where object properties are predicted together with pixel labels.

Again, we used Cellpose [2] with manual correction to generate this instance segmentation ground truth images from associated CellMask™images (Fig.4a, Fig.4b).

As for nuclear segmentation, we used training sets of different sizes $N \in \{1, 10, 50, 80\}$ of dimensions (1024, 1024) and evaluated the accuracy for each of them. Testing is always performed on the same 17 images. 5 images of dimensions (512, 512) were randomly cropped from each initial image.

To tackle the issue of instance segmentation, we implemented a model predicting both a cell semantic segmentation image (Fig.4c) and a distance map, i.e. an image where pixels values get higher as they are closer to the center of the cell, the background remaining black (Fig.4d), as proposed in [18], [19].



(a) (b) (c) (d)

Fig. 4: (a) Fluorescent Cy5 image (CellMask™). (b) Corresponding cell instance segmentation image generated by Cellpose. (c) Cell semantic segmentation image generated from Cellpose output. (d) Distance map generated from Cellpose output.

Like in the previous section we compare two models to investigate whether transfer learning from an ISL model can significantly improve the accuracy of our segmentation. The first model is the U-net 'on steroids', outputting 2 channels

(Fig.5a). The second model has the same U-net architecture but is pretrained on CellMask™images, thus outputting only 1 channel. Hence we add two Conv2d layers at the end to upscale to 2 channels (Fig.5b).

The number of epochs during training depends on the size of the training set. Both models are trained with an ADAM optimizer, a learning rate of 0.01 and the loss function presented in (2). MSELoss stands for the usual mean square error, while BCEWithLogitsLoss combines a sigmoid layer with the binary cross entropy loss. y represents the output of our model, with the two channels y_d and y_m standing for the distance and semantic segmentation image, respectively. The factor α is used to balance the weights of the different losses during training. It has been set as $\alpha = 2000$, 2000 being the initial ratio between MSELoss and BCEWithLogitsLoss. This has been inspired by the loss function used in Cellpose [2], which also uses a loss function computed as the sum of two loss functions, one for each output channel.

$$\text{Loss}(y) = \text{Loss}((y_d, y_m)) = \text{MSELoss}(y_d) + \alpha \cdot \text{BCEWithLogitsLoss}(y_m) \quad (2)$$

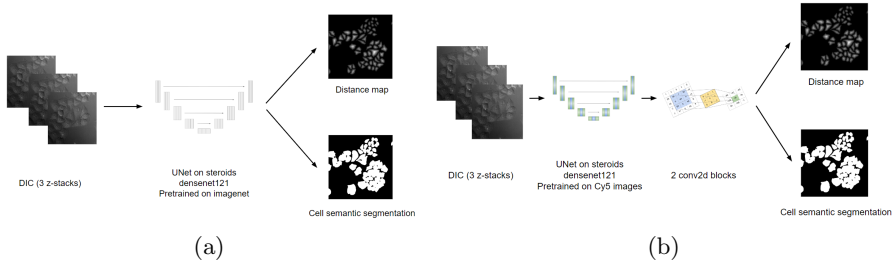


Fig. 5: Models compared to predict cell instance segmentation: (a) U-net 'on steroids' which has not been trained on Cy5 images. (b) U-net model pretrained on Cy5 images.

Finally, we apply a post-processing step to get the final results. For this, we apply the h-maxima transformation of the predicted distance map, with $h = 10$. The h-maxima transformation is defined as the reconstruction by dilation of $f-h$ under f: $HMAX_h(f) = R_f^{\delta}(f-h)$, and removes insignificant local maxima (i.e. local maxima with a local contrast lower than h). The local maxima of $HMAX$ then serve as seed for the watershed algorithm (Fig.6), providing the final result.

3 Results

3.1 Nucleus Segmentation

DAPI prediction yields very good results, with a Pearson correlation coefficient (PCC) of 0.95±0.08.

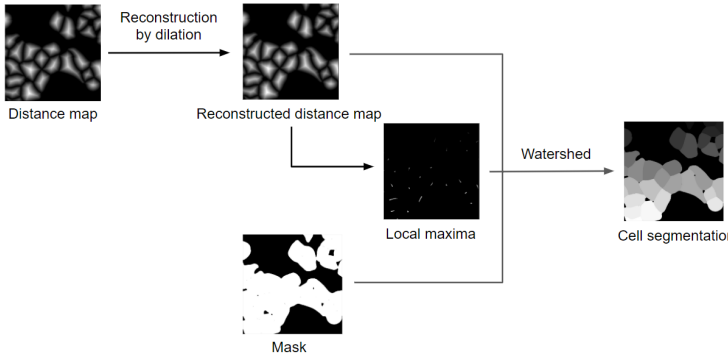
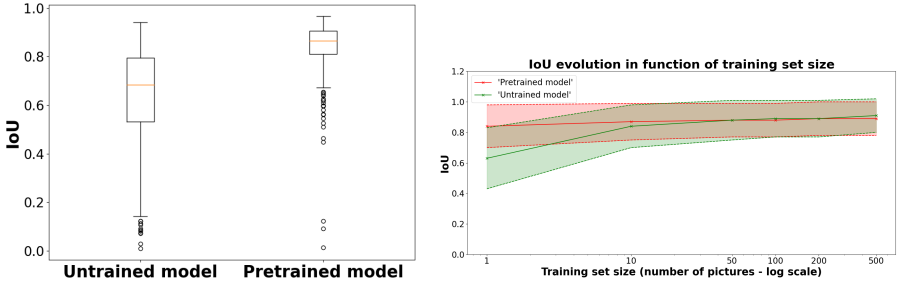


Fig. 6: Pipeline to get instance segmentation image from both distance map and semantic segmentation image. H-maxima transform followed by watershed algorithm enable to segment cells instance-wise.

Using the Jaccard index (or IoU, Intersection Over Union) as metric, the U-net 'on steroids' gives 0.64±0.2 after training on 1 single image. In comparison, the model pretrained on DAPI reaches 0.84±0.1, improving the previous score by 31.3% (Fig.7a). This improvement decreases as the size of the training set increases, being 4.8% (respectively 1.1%, 0.0%, 0.0%, -1.1%) after training on 10 (respectively 50, 100, 200, 500) (Fig.7b).

Results from both models trained on 1 single image are displayed in Fig.8.



(a) (b)

Fig. 7: Nucleus segmentation results. (a) Intersection Over Union (IoU) score for untrained and pretrained models, after training on 1 image. (b) Evolution of IoU average score for both models for different training set sizes.

3.2 Cell Segmentation

CellMask™prediction also yields very good results, with a PCC of 0.97±0.02.

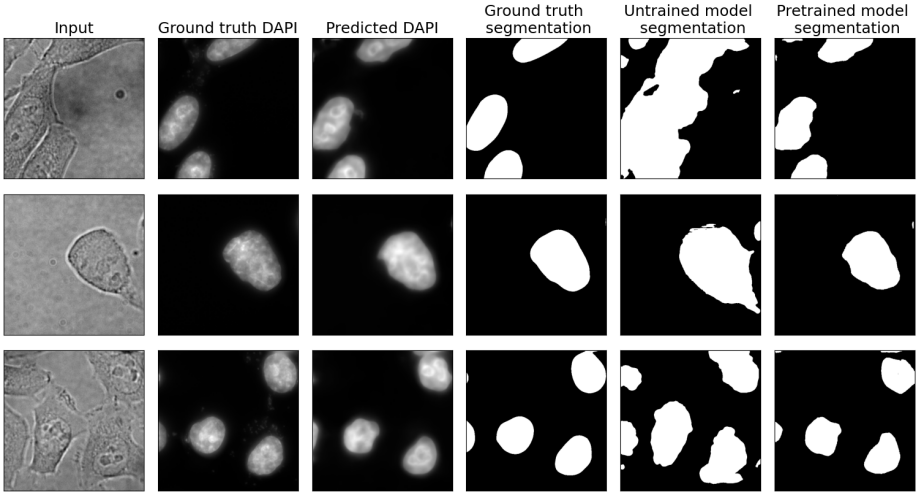


Fig. 8: Input bright-field images, DAPI images, DAPI predictions generated by U-net 'on steroids', ground truth instance segmentation generated by Cellpose, untrained U-net 'on steroids' segmentation prediction, pretrained U-net 'on steroids' segmentation prediction. Segmentation is performed after training on 1 image for both models.

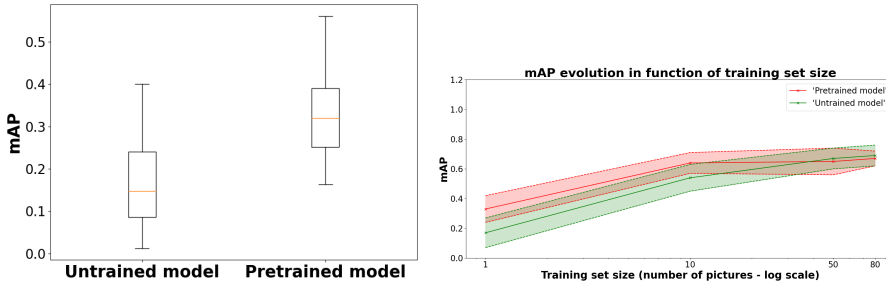
Using the mean average precision (mAP) as metric, the U-net 'on steroids' gives 0.17±0.1 after training on 1 single image. In comparison, the model pretrained on CellMask™reaches 0.33±0.09, improving the previous score by 94.1% (Fig.9a). As in the previous section this improvement decreases as the size of the training set increases, being 18.5% (respectively -3.0%, -2.9%) after training on 10 (respectively 50, 80) (Fig.9b).

Results from both models trained on 1 single image are displayed in Fig.10.

4 Discussion

The results presented in the previous sections show that pretraining with in silico labeling as pretext task significantly improves the performance of a segmentation model trained on a very small data set. Indeed, the accuracy raises by 31.3% and 94.1% for nucleus semantic segmentation and cell instance segmentation, respectively, after training on 1 single image, using a model pretrained in an ISL setting.

The fact that pretraining on DAPI images helps to generate a nucleus semantic segmentation was actually expected since the two outputs (DAPI and binary segmentation maps) are very close to each other. On the other hand, cell instance segmentation is a much more complex problem, and our results clearly indicate that also in this situation, pretraining of fluorescent markers significantly improves segmentation accuracy for small datasets. We also observe that



(a) (b)

Fig. 9: Cell segmentation results (a) mAP score for untrained and pretrained models, after training on 1 image. (b) Evolution of mAP average score for both models for different training set sizes.

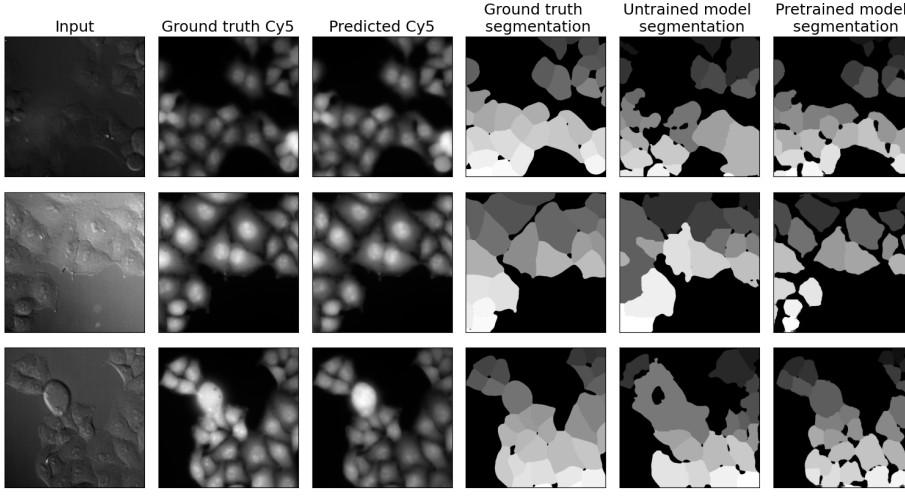


Fig. 10: Input DIC images, Cy5 images, Cy5 predictions generated by U-net 'on steroids', ground truth instance segmentation generated by Cellpose, untrained U-net 'on steroids' segmentation prediction, pretrained U-net 'on steroids' segmentation prediction. Segmentation is performed after training on 1 image for both models.

transfer learning is useful if we work on a very small data set (1 to 10 images), but that for both nucleus and cytoplasmic segmentation, the accuracy difference disappears if the models are trained on more than 10 images.

From a practical point of view, this idea provides an interesting alternative to manual annotation, in particular in the context of High Content Screening, where it is fairly easy to generate large amounts of data that contain both label-free and fluorescently labeled microscopy images. In this case, we can train efficient models for fluorescence prediction, and use these models in a pre-training scheme to reduce the manual annotation burden. Finally, we showed here that this pre-training scheme is effective for segmentation of nuclei and cells, but we also believe that this could be effective for any other type of cell structures as soon as you can get the associated fluorescent images available. Furthermore, it will be interesting to investigate to which extent the pre-training scheme provides good starting points for generalist networks, applicable to a wide variety of modalities.

5 Conclusion

In this paper, we demonstrated that pretraining on the prediction of relevant fluorescent markers can be very useful to segment nuclei or cells. We showed that a model trained to predict some fluorescent structures from label-free microscopy can learn to segment these structures from a very small data set, down to 1 single image. We believe that this can be of great help for applications where fluorescent data are easily available, if one wants to avoid tedious manual annotation to build large ground truth datasets for the training of neural networks. With only a few images, it is possible to fine-tune a pretrained model achieving performances matching those obtained by ImageNet-pretrained state-of-the-art networks fine-tuned on a much larger set of images. Our pre-training scheme can thus help biologists to save time and money without sacrificing any accuracy.

References

[1] Reka Hollandi et al. “nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer”. In: *Cell Systems* 10.5 (May 2020), 453–458.e6. ISSN: 24054712. DOI: 10.1016/j.cels.2020.04.003.

[2] Carsen Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. In: *Nature Methods* 18.1 (Jan. 2021), pp. 100–106. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01018-x. URL: https://doi.org/10.1038/s41592-020-01018-x.

[3] Uwe Schmidt et al. “Cell Detection with Star-Convex Polygons”. In: *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II*. 2018, pp. 265–273. DOI: 10.1007/978-3-030-00934-2_30.

[4] Eric M. Christiansen et al. "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images". In: *Cell* 173.3 (2018), 792–803.e19. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2018.03.040. URL: https://www.sciencedirect.com/science/article/pii/S0092867418303647.

[5] Chawin Ounkomol et al. "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy". In: *Nature Methods* 15.11 (Nov. 2018), pp. 917–920. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0111-2. URL: https://doi.org/10.1038/s41592-018-0111-2.

[6] Yair Rivenson et al. "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning". In: *Nature Biomedical Engineering* 3.6 (June 2019), pp. 466–477. ISSN: 2157-846X. DOI: 10.1038/s41551-019-0362-y. URL: https://doi.org/10.1038/s41551-019-0362-y.

[7] Yair Rivenson et al. "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning". In: *Light: Science & Applications* 8.1 (Feb. 2019), p. 23. ISSN: 2047-7538. DOI: 10.1038/s41377-019-0129-y. URL: https://doi.org/10.1038/s41377-019-0129-y.

[8] Dan Li et al. "Deep Learning for Virtual Histological Staining of Bright-Field Microscopic Images of Unlabeled Carotid Artery Tissue". In: *Molecular Imaging and Biology* 22.5 (Oct. 2020), pp. 1301–1309. ISSN: 1860-2002. DOI: 10.1007/s11307-020-01508-6. URL: https://doi.org/10.1007/s11307-020-01508-6.

[9] Yi Liu et al. "Global Pixel Transformers for Virtual Staining of Microscopy Images". In: *IEEE Transactions on Medical Imaging* PP (Jan. 2020), pp. 1–1. DOI: 10.1109/TMI.2020.2968504.

[10] Yoav N. Nygate et al. "Holographic virtual staining of individual biological cells". In: *Proceedings of the National Academy of Sciences* 117.17 (2020), pp. 9223–9231. DOI: 10.1073/pnas.1919569117. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1919569117. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1919569117.

[11] J. Boyd et al. "Experimentally-Generated Ground Truth for Detecting Cell Types in an Image-Based Immunotherapy Screen". In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Apr. 2020, pp. 886–890. ISBN: 1945-8452. DOI: 10.1109/ISBI45749.2020.9098696.

[12] Chenfei Hu et al. "Live-dead assay on unlabeled cells using phase imaging with computational specificity". In: *Nature Communications* 13.1 (Feb. 2022), p. 713. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28214-x. URL: https://doi.org/10.1038/s41467-022-28214-x.

[13] Jingfang K Zhang et al. "Automatic Colorectal Cancer Screening Using Deep Learning in Spatial Light Interference Microscopy Data". en. In: *Cells* 11.4 (Feb. 2022).

[14] Cheng Shen et al. "Stain-free detection of embryo polarization using deep learning". In: *Scientific Reports* 12.1 (Feb. 2022), p. 2404. ISSN: 2045-2322.

DOI: 10.1038/s41598-022-05990-6. URL: https://doi.org/10.1038/s41598-022-05990-6.

[15] Yuchen R. He et al. "Cell Cycle Stage Classification Using Phase Imaging with Computational Specificity". In: *ACS Photonics* 9.4 (2022), pp. 1264–1273. DOI: 10.1021/acsphotonics.1c01779. eprint: https://doi.org/10.1021/acsphotonics.1c01779. URL: https://doi.org/10.1021/acsphotonics.1c01779.

[16] Pavel Yakubovskiy. *Segmentation Models Pytorch*. https://github.com/qubvel/segmentation_models.pytorch. 2020.

[17] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[18] Peter Naylor et al. "Nuclei Segmentation in Histopathology Images Using Deep Neural Networks". In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (2017), IEEE, EMB, IEEE Signal Proc Soc. ISSN: 19458452. DOI: 10.1109/ISBI.2017.7950669.

[19] Peter Naylor et al. "Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map". In: *IEEE Transactions on Medical Imaging* 38.2 (2019), pp. 448–459. DOI: 10.1109/TMI.2018.2865709.