# Deep Learning based Segmentation Pipeline for Label-Free Phase-Contrast Microscopy Images

Aydin Ayanzadeh
Informatics Institute, Ayazaga Campus,
Istanbul Technical University,
Istanbul, Turkey
ayanzadeh17@itu.edu.tr

Özden Yalçın Özuysal, Devrim Pesen Okvur
Department of Molecular Biology and
Genetics, Izmir Institute of Technology,
Izmir, Turkey
{ozdenyalcin, devrimpesen}@iyte.edu.tr

Sevgi Önal
Biotechnology, Izmir
Institute of Technology,
Izmir, Turkey
sevgional1@gmail.com

Behçet Uğur Töreyin
Informatics Institute, Ayazaga Campus,
Istanbul Technical University,
Istanbul, Turkey
toreyin@itu.edu.tr

Devrim Ünay
Electrical-Electronics Engineering, Faculty of
Engineering,Izmir Demokrasi University
Izmir, Turkey
unaydevrim@gmail.com

***Abstract*—The segmentation of cells is necessary for biologists in the morphological statistics for quantitative and qualitative analysis in Phase-contrast Microscopy (PCM) images. In this paper, we address the cell segmentation problem in PCM images. Deep Neural Networks (DNNs) commonly is initialized with weights from a network pre-trained on a large annotated data set like ImageNet have superior performance than those trained from scratch on a small dataset. Here, we demonstrate how encoder-decoder type architectures such as U-Net and Feature Pyramid Network (FPN) can be improved by an alternative encoder which pre-trained on the ImageNet dataset. In particular, our experimental results confirm that the image descriptors from ResNet-18 are highly effective in accurate prediction of the cell boundary and have higher Intersection over Union (IoU) in comparison to the classical U-Net and require fewer training epochs.**

***Index Terms*—Cell Segmentation, Phase-Contrast microscopy, Deep learning**

## I. Introduction

Phase-contrast Microscopy (PCM) images have an irreplaceable role in detailed analysis of the cell structures as it allows a wide range of possibilities in studying living cells from different perspectives in a label-free manner. Segmentation of the PCM images has been challenging for the biologists due to problems specific to the examined specimen or the microscopy technique used, such as perturbations in the cell shapes, overlapping and transparent appearance of the cells in PCM images which complicates the cell segmentation with conventional computer vision approaches, effective extraction of the Region of Interest (RoI) from images. Therefore, most of the segmentation and tracking solutions available in the literature do not qualify as satisfactory in terms of robustness, completeness and accuracy.

The well-known tasks of medical image analysis like detection [1], counting [2], segmentation [3], classification [4] and tracking [5] are also integral to the automated segmentation of cells from PCM images. The literature consists of several cell segmentation solutions employing conventional algorithms such as Empirical Gradient Threshold (EGT) [6], watersheds [7] and Active Contours [8].

The recently popular Deep Neural Networks (DNNs) have significant effect on the improvement of segmentation accuracy from different perspectives such as robustness and completeness in comparison to conventional methods. Tsai et al. [5] leveraged Mask R-CNN as a solver for the segmentation of PCM images. The proposed method is powerful on the extraction of the cell shapes on adjacent and touching cells, however it is not easily applicable to PCM image datasets corresponding to different biological domains. Moreover, the proposed approach has a high level of complexity in computation to reach the determined results. Yi et al. [9], presented keypoint graph based bounding boxes to a deep learning framework for multi-resolution cell instance segmentation.

U-Net [10], with its encoder-decoder architecture, has proven highly accurate in the segmentation of medical images, even for the cases where low number of annotated images is available for training set. Accordingly, in our previous study [11], a multi-resolution model based on a U-Net like architecture is proposed, where the contracting path is eliminated from U-Net and sequential convolution is replaced with an expanding path. Experimental evaluations showed robust performance for the case of low number of annotated data in comparison to conventional methods. In a follow-up study [12], we introduced SegNet [13] as a solver for the segmentation of PCM images. Although, we obtained acceptable cell segmentation performances overall as compared to the classical approaches; the results were far from being robust in separating of the cell's boundary particularly for the adjacent or touching cells. Therefore, we aim to improve the cell segmentation of PCM images in overall and particularly for the adjacent or touching cells.

The contributions of this paper can be summed up as

978-1-7281-7206-4/20/$31.00 ©2020 IEEE

follows: we asserted that the performance of U-Net improves significantly by leveraging alternative feature extractors in the encoder of U-Net and FPN [14] by applying the pre-trained model of ImageNet in the network. In Section III, we give a detailed explanation of our proposed architecture. The experimental results are reported in Section IV to investigate the performance of our proposed network, which is followed by the conclusion section.

## II. Dataset Preparation

The employed dataset includes invasive breast cancer cells with mesenchymal morphology (MDA-MB-231 cell) which has been captured using an Olympus IX71 microscope. We randomly selected the specified number of frames in our study and annotated them with the supervision of experts. As pre-processing of the dataset, we adjusted the level of contrast and normalized the pixel values to the [0,1] range. The final dataset contains 600 frames of PCM images each with the dimension of $2568 \times 1912$ pixels ($0.117\mu m \times 0.117\mu m$). For manual annotation of the cell boundary on the frames ImageJ is leveraged. We randomly selected and manually annotated 45 of the frames, where 30 of them is used for training while the rest is utilized for test.

## III. Model

Deep learning is intensively greedy on the number of training data and nourishes with thousands and even millions of data in the training phase. Transfer learning (TL) is tuning the models by reusing different domains of data set or task that is already trained on a dataset with abundant number of annotations (such as the ImageNet dataset [15]). Actually, transfer learning has high impact in reducing over-fitting for limited data cases (like in this study) while speeding up the performance of the model efficiently.

We replace the backbones of U-Net with ResNet-18 (Conv1-5) [16] with excluding of the dense layers from ResNet architecture. The schematic structure of ResNet18-Unet is shown in Fig. 1. In ResNet-18, we have the two residual blocks in each resolution stages which we repeat it symmetrically to the expansive path with one residual block for each dimension level. In the lowest stage of the encoder, single residual block is employed with 1024 channels for separation of encoder and decoder as a bottleneck of the network in our presented network. The residual blocks contain of layers with 3×3 kernels where each convolutional layer is followed by a Batch Normalization (BN) layer and a Rectified Linear Unit (ReLU) activation function. In addition, the number of channels in the first block is 64 and the number of channels is regularly doubled after doubling the stride in each resolution stage.

In the decoder section, to have symmetric operation, we apply upsampling by 4 which is equal to the number of dimension reduction in the encoder part for reproducing the input images that consist of residual blocks in each resolution stage.

FPN comprises of bottom-up and top-down pathways as seen in Fig. 2. In the bottom-up pathway, we employed the ResNet18(Conv1-5) as the backbone of the model's feature encoder, and in each block the stride is doubled in each stage to reduce the spatial dimension of the pyramid level. To reduce the dimension of the last block in FPN, we utilized $1 \times 1$ convolution with channel depth of 128. In the top-down pathway as we move-up, spatial dimension of layers are increased by 2 with nearest neighbor interpolation. Therefore, we employed two sequential $3 \times 3$ convolutions to reach module T in each stage of the pyramid. At the end, feature maps are upsampled to the same dimension and concatenated.

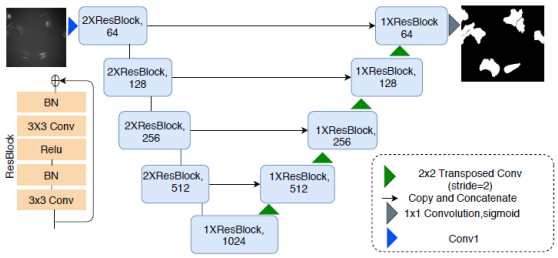


Fig. 1. Schematic structure of Res18-U-Net. Each number in the block indicates the number of channels in each stage.

## IV. Experiment Results

### A. Training Methodology

We applied data augmentation to the annotated training set where the main transformations utilized comprised of random elastic transformation, rotation and translation. To have robust prediction, we applied Test Time Augmentation (TTA) involving horizontal, vertical, diagonal flips and rotation in the prediction step. Moreover, Spatial Dropout is utilized with the rate of 0.5 at the last layer of the FPN network. Batch size is considered 8 for the network and weights of the backbone in the encoder are pre-trained in the ImageNet [15] and the rest of the network layers in the decoder are initialized with the LeCun uniform weight initialization. Adam with the initial learning rate of $10^{-4}$ is considered as the optimization function and loss function is controlled in the validation set to be stopped for avoiding the method from over-fitting. All of the experiments have been realized on a single NVIDIA TitanX graphics card with 12 GB memory. The training set was resized to $518 \times 518$ to reduce memory consumption. For loss function, we combined the Binary Cross Entropy (BCE) indicated in (1) with the Dice loss (DL) shown in (2) as cost function of the network as the combination achieved superior performance in overlap measures empirically. We integrated the local Cross Entropy with the global Dice loss for our loss function. In BCE, $\beta$ is utilized to create a harmonic balance between BCE and DL; we set it to 1 empirically in

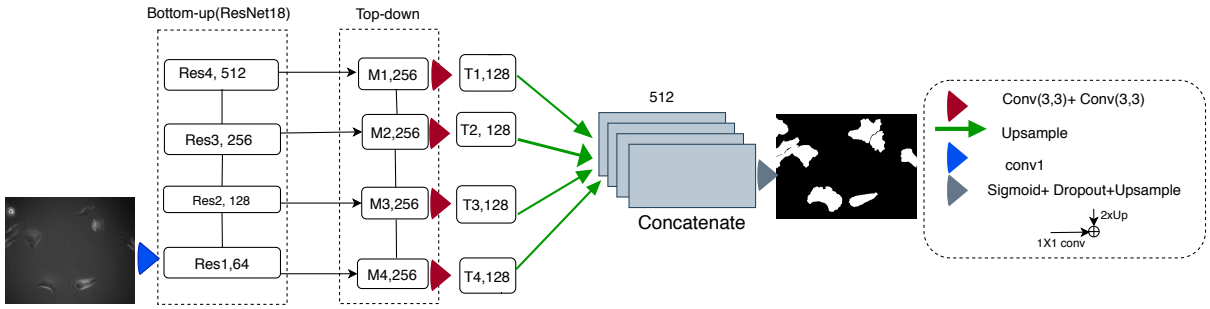


Fig. 2. Schematic structure of FPN with the backbone as modified ResNet18 pre-trained on ImageNet. Conv stands for Convolution in the above figure. The number in each block represents the number of channel which is employed in each stage.

our presented loss.

$$BCE\left(X,\hat{Y}\right) = -\beta\left(X\log\left(\hat{Y}\right) + (1-X)\log\left(1-\hat{Y}\right)\right) \quad (1)$$

$$DL(X,\hat{Y}) = 1 - \frac{2|X\cdot\hat{Y}|}{|X|+|\hat{Y}|} \quad (2)$$

$$L(X,Y) = DL(X,\hat{Y}) + BCE(X,\hat{Y}) \quad (3)$$

### B. Evaluation Metrics

In order to show the efficacy of the algorithm, evaluation of the images from different perspectives is necessary. The evaluation criteria in the segmentation of cells comprise of the following metrics: Jaccard index or IoU(Intersection over Union), Precision, Recall, and Dice Coefficient, which are indicated in (4), (5) and (6) respectively.

$$Jaccard(X,Y) = \frac{|X\bigcap Y|}{|X\bigcup Y|} = \frac{|X\bigcap Y|}{|X|+|Y|-|X\bigcap Y|} \quad (4)$$

$$Precision = \frac{n_{tp}}{n_{fp}+n_{tp}}, Recall = \frac{n_{tp}}{n_{fn}+n_{tp}} \quad (5)$$

$$Dice(X,Y) = \frac{2|X\cdot Y|}{|X|+|Y|} \quad (6)$$

### C. Results

We compare the performance of our proposed approach with various baselines such as EGT [6], U-Net [10], LinkNet [17], TipNet [11], and PHANTAST [18]. As can be seen from Fig. 3, classical methods like PHANTAST and EGT are not satisfactory in the extraction of the cell border and the RoI. The proposed method outperforms the state-of-the-art methods and has noticeably robust performance in the extraction of cell boundaries. Moreover, it is resilient in the prediction of the specific cell structure. The qualitative results of the utilized methods are shown in Fig. 3 (h-i). We achieved Jaccard indexes of **0.8791** and **0.8710** with ResNet18-U-Net and ResNet18-FPN, respectively. This shows that our proposed approach not only improves segmentation accuracy, but also it is effective in the separation of overlapping cells without the use of any post-processing procedure. Moreover, our proposed approach has surpassed the results of U-Net and other baselines in completeness, robustness and other factors of efficiency in the segmentation task on the MDA-MB-231 dataset.

TABLE I
QUANTITATIVE RESULTS OF THE SEGMENTATION METHODS.

| Methods | IoU | Dice | Precision | Recall |
|---|---|---|---|---|
| EGT [6] | 0.4225 | 0.5940 | 0.4574 | 0.8470 |
| PHANTAST [18] | 0.5465 | 0.7068 | 0.6823 | 0.7330 |
| LinkNet [17] | 0.8593 | 0.9235 | 0.9454 | 0.9025 |
| U-Net [10] | 0.8646 | 0.9274 | 0.9335 | 0.9214 |
| Tip-Net [19] | 0.8542 | 0.9214 | 0.9382 | 0.9052 |
| ResNet18-UNet | **0.8791** | **0.9357** | **0.9508** | **0.9211** |
| ResNet18-FPN | **0.8710** | **0.9311** | **0.9419** | **0.9206** |

## V. CONCLUSION

In this paper, we have a proposed hybrid deep neural networks based method for cell segmentation in PCM images and demonstrated its superior performance in comparison to the conventional approaches. To reduce the disparity in the encoder feature and the features that propagate in the decoder of the U-Net architecture. We proposed an alternative feature extractor by applying the modified ResNet18 in the encoder of U-Net and replaced the plain blocks with residual blocks in the decoder of U-Net architecture. Furthermore, we utilized the ResNet18 in the backbone of FPN which has higher performance in comparison of conventional methods. This alteration made the model effective in taking low and high-level semantics and the detail of the images into account. In addition, employing transfer learning in the presented approach increased the training convergence, and improved the resiliency of the results in the prediction step, reduced training time and also contributed to the prevention of the model from over-fitting.

## ACKNOWLEDGMENT

The data used in this study is collected under the Marie Curie IRG grant (no: FP7 PIRG08-GA-2010-27697).

Aydin Ayanzadeh's work is supported, in part, by Vodafone Turkey, under project no. ITUVF20180901P04 within the context of ITU Vodafone Future Lab RD program.

This work is in part funded by İTÜ BAP MGA-2017-40964.

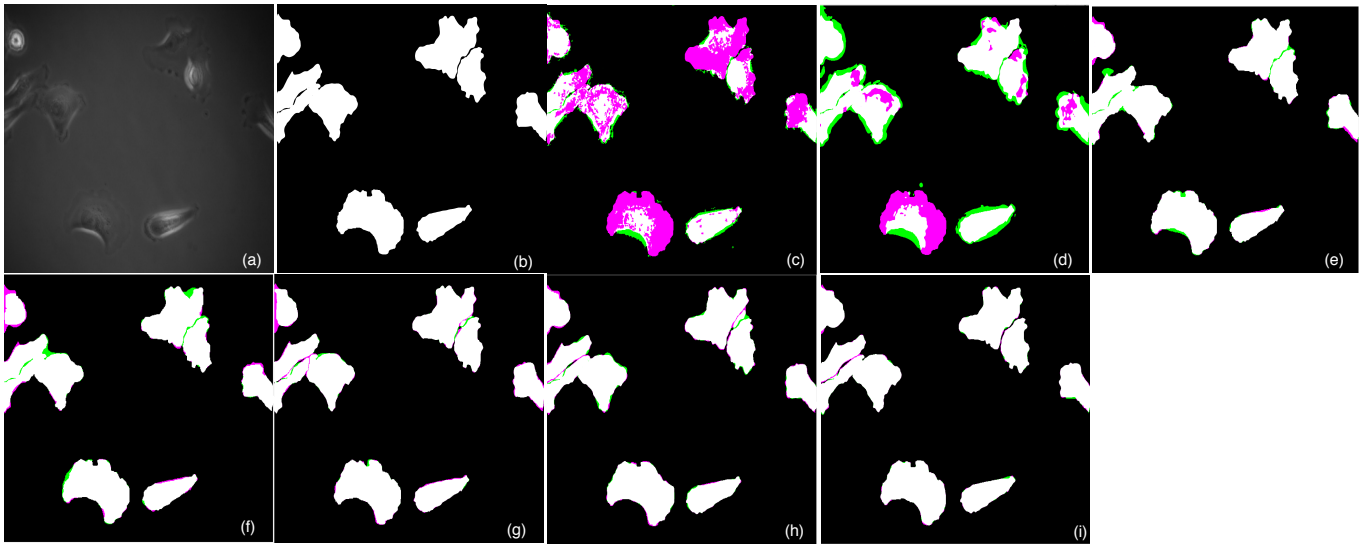


Fig. 3. Segmentation results for the MDA-MB-231 dataset. Pink marks represent the false negatives and green marks indicate the false positives. (a-b) is the input image and ground truth of the determined frames. (c-d) is the qualitative results of EGT and PHANTAST subsequently. (e-f) is the visualization of the TipNet and LinkNet without post-processing method. (g-h) is the results of U-Net and FPN. (i) is the qualitative results of our proposed method.

This work has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 119E578.

## REFERENCES

[1] K. Lomanov, J. M. del Rincón, P. Miller, and H. Gribben, "Cell detection with deep convolutional networks trained with minimal annotations," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 943–947.

[2] R. Flight, G. Landini, I. Styles, R. Shelton, M. Milward, and P. Cooper, "Automated noninvasive epithelial cell counting in phase contrast microscopy images with automated parameter selection," *Journal of microscopy*, vol. 271, no. 3, pp. 345–354, 2018.

[3] L. Zhang, "An efficient approach for cell segmentation in phase contrast microscopy images," *arXiv preprint arXiv:1904.00328*, 2019.

[4] D. Bhaskar, D. Lee, H. Knútsdóttir, C. Tan, M. Zhang, P. Dean, C. Roskelley, and L. Edelstein-Keshet, "A methodology for morphological feature extraction and unsupervised cell classification," *bioRxiv*, p. 623793, 2019.

[5] H.-F. Tsai, J. Gajda, T. F. Sloan, A. Rares, and A. Q. Shen, "Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning," *SoftwareX*, vol. 9, pp. 230 – 237, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352711018301882

[6] J. Chalfoun, M. Majurski, A. Peskin, C. Breen, P. Bajcsy, and M. Brady, "Empirical gradient threshold technique for automated segmentation across image modalities and cell lines," *Journal of microscopy*, vol. 260, no. 1, pp. 86–99, 2015.

[7] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 583–598, 1991.

[8] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," *Signal processing*, vol. 71, no. 2, pp. 203–213, 1998.

[9] J. Yi, P. Wu, Q. Huang, H. Qu, B. Liu, D. J. Hoeppner, and D. N. Metaxas, "Multi-scale cell instance segmentation with keypoint graph based bounding boxes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 369–377.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[11] A. Ayanzadeh, H. O. Yağar, Ö. Y. Özuysal, D. P. Okvur, B. U. Töreyin, D. Ünay, and S. Önal, "Cell segmentation of 2d phase-contrast microscopy images with deep learning method," in *2019 Medical Technologies Congress (TIPTEKNO)*. IEEE, 2019, pp. 1–4.

[12] R. C. Binici, U. Şahin, A. Ayanzadeh, B. U.Töreyin, S. Önal, D. P. Okvur, Y. Özuysal, and D. Ünay, "Automated segmentation of cells in phase contrast optical microscopy time series images," in *2019 Medical Technologies Congress (TIPTEKNO)*, Oct 2019, pp. 1–4.

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[14] S. S. Seferbekov, V. Iglovikov, A. Buslaev, and A. Shvets, "Feature pyramid network for multi-class land segmentation." in *CVPR Workshops*, 2018, pp. 272–275.

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.

[18] N. Jaccard, L. D. Griffin, A. Keser, R. J. Macown, A. Super, F. S. Veraitch, and N. Szita, "Automated method for the rapid and precise estimation of adherent cell culture characteristics from phase contrast microscopy images," *Biotechnology and bioengineering*, vol. 111, no. 3, pp. 504–517, 2014.

[19] V. Iglovikov and A. Shvets, "Ternausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation," *CoRR*, vol. abs/1801.05746, 2018. [Online]. Available: http://arxiv.org/abs/1801.05746