

Improved cell segmentation using deep learning in label-free optical microscopy images

Aydin AYANZADEH^{1,*}, Özden YALÇIN ÖZUYSAL², Devrim PESEN OKVUR²,
Sevgi ÖNAL³, Behçet Uğur TÖREYİN^{1†}, Devrim ÜNAY^{4†}

¹Informatics Institute, Istanbul Technical University, İstanbul, Turkey

²Department of Molecular Biology and Genetics, Izmir Institute of Technology, İzmir, Turkey

³Biotechnology and Bioengineering Graduate Program, Izmir Institute of Technology, İzmir, Turkey

⁴Department of Electrical and Electronics Engineering, Izmir Democracy University, İzmir, Turkey

Received: 26.05.2021

Accepted/Published Online: 09.08.2021

Final Version: 04.10.2021

Abstract: The recently popular deep neural networks (DNNs) have a significant effect on the improvement of segmentation accuracy from various perspectives, including robustness and completeness in comparison to conventional methods. We determined that the naive U-Net has some lacks in specific perspectives and there is high potential for further enhancements on the model. Therefore, we employed some modifications in different folds of the U-Net to overcome this problem. Based on the probable opportunity for improvement, we develop a novel architecture by using an alternative feature extractor in the encoder of U-Net and replacing the plain blocks with residual blocks in the decoder. This alteration makes the model superconvergent yielding improved performance results on two challenging optical microscopy image series: a phase-contrast dataset of our own (MDA-MB-231) and a brightfield dataset from a well-known challenge (DSB2018). We utilized the U-Net with pretrained ResNet-18 as the encoder for the segmentation task. Hence, following the modifications, we redesign a novel skip-connection to reduce the semantic gap between the encoder and the decoder. The proposed skip-connection increases the accuracy of the model on both datasets. The proposed segmentation approach results in Jaccard Index values of 85.0% and 89.2% on the DSB2018 and MDA-MB-231 datasets, respectively. The results reveal that our method achieves competitive results compared to the state-of-the-art approaches and surpasses the performance of baseline approaches.

Key words: Segmentation, breast cancer, convolutional neural networks, optical microscopy, phase-contrast, brightfield

1. Introduction

Phase-contrast optical microscopy (PCM) images play an indispensable role in detailed analysis of cell structures because they offer a variety of possibilities to examine living cells from different perspectives in an unlabeled manner. Visual analysis of such time-series data is essential but challenging for biologists due to the time-consuming and tiring nature of the process as well as factors such transparent and mixed light cell appearance. For this reason, most of the solutions in the literature of PCM images are unsatisfactory in terms of robustness, completeness, and accuracy in the segmentation task. Additionally, live cell microscopy imaging has the potential to assist biologists in measuring biological attributes of cells. It can be an experienced bias task for biologists to examine the properties without appropriate tools for cell segmentation.

Recently, deep learning methods have been widely used in medical informatics and gradually replaced

*Correspondence: ayanzadeh17@itu.edu.tr ; †Equal senior contribution: Behçet Uğur TÖREYİN, Devrim ÜNAY

traditional computer vision methods in medical image segmentation tasks. Additionally, deep learning methods play a pivotal role in cell segmentation and assist biologists in the identification of cell structures. Although various number of methods have been proposed for medical image segmentation; these tasks still remain challenging for biologists from different perspectives. For example, PCM image has challenged biologists during the process of scrutinizing due to low contrast, fuzzy, and overlapping boundaries and transparent appearance of the cells which encounter the cell segmentation with deep learning approach.

Although numerous methods have been proposed for medical image segmentation, it remains a challenging task due to the complexity of the data to be segmented. For example, images of the human body structure are complex, comprising numerous tissues, low contrast, fuzzy boundaries, identity variations, and small targets.

A large number of approaches have been proposed to handle the above challenges. Prior work on cell segmentation includes solutions based on traditional image processing approaches such as local contrast [1], local threshold [2], active contour [3, 4] graph partitioning [5, 6], and sparse matrix decomposition [7]. U-Net [8, 9] and its variants [10–13] are generally preferred over other deep learning-based approaches in biomedical image segmentation problems, especially in the cases where training set is small, i.e. annotated data is limited. Among various deep learning approaches employed for the analysis of microscopy images, U-Net and its variants demonstrate superior performance in comparison to other deep learning-based methods, such as SegNet [14] which utilized VGG-16 [15], and TernausNet [16] which replicated the pretrained VGG-11 in the backbone of the architecture.

U-Net++ [17] advocates redesigning the skip-connections by employing nested skip-connections from densely connected networks between the encoder and the decoder. DenseNet [18] is a specific architecture which improves the style of residual block. DenseNet concatenates all the layers of different stages. Therefore, replication of DenseNet improves the accuracy of segmentation in comparison to the plain skip-connection.

BCDUNet [19] proposed a densely connected layer with specific mechanism as an alternative to the skip-connection of U-Net by introducing the bidirectional ConvLSTM (BConvLSTM) densely connected layer which concatenates the features from the encoder to the decoder efficiently and improves the performance of the model with this alteration in comparison to the baseline approach in the field of medical segmentation. On a similar account, Mask R-CNN [20] based approaches yield competitive results in PCM image segmentation problems. However, they are computationally demanding methods and converge slowly.

Recently, Debesh et al. [21] introduced DoubleU-Net by integrating an additional combination of two U-Net architectures stacked on top of each other for capturing the higher semantic features. First U-Net, which is employed in the encoder section, utilized the VGG-19 pretrained features of the ImageNet. In addition, this model used the atrous spatial pyramid pooling (ASPP) to capture the higher number of contextual information, due to which this model can be suggested as a baseline in some modalities of medical imaging applications. In [22], an architecture with specific aggregation was introduced for biomedical image segmentation tasks. The model included two main aggregation submodules, namely, a crossing aggregation module, and a weighted aggregation module.

Recently, there has been a significant advancement in biomedical imaging by leveraging deep learning approaches. Convolutional neural networks (CNNs) tackle the limitations of traditional segmentation approaches. The modern segmentation architectures mainly have an encoder-decoder structure which is suitable for medical segmentation tasks. The main advantage of these architectures is the specific mission of skip-connection, which propagates semantics features from the encoder to the decoder. These models are capable in efficient segmentation of the region of interest (RoI) in the nucleus and PCM images even with low-annotated training set in the

variant domain in the biomedical images. Despite progressions in the field of medical segmentation, automatic segmentation still has challenging tasks in terms of completeness and robustness in some specific challenging datasets. These challenges include crowded distribution of cells and their tendency to overlap, discrimination of the cell's boundary in overlapping and adjacent regions, and inferring of cell's boundary which has blurry boundary.

To this end, we proposed the architecture which has the autoencoder structures with the backbone of ResNet18 (Conv1-Conv5), which has already learned features from ImageNet. We proved that the transfer learning from a network which is pretrained on ImageNet significantly improves the results not only on a few annotated images on MDA-MB231 dataset but also it achieves greater accuracy in comparison to the baseline approach. Additionally, to obtain a greater number of distinguishable features, we propose a specific skip-connection that contains nonlinear operations between the encoder and decoder in our proposed approach. This operation contains residual blocks which are obtained to reduce the semantic gap between the encoder and decoder in our proposed architecture. Consequently, as a postprocessing mechanism, we employed test time augmentation (TTA), which is not only an effective mechanism for measuring the level of uncertainty estimation of the architectures, but also increases the precision of the segmentation results in our utilized dataset.

To address the need for an improved segmentation performance in label-free optical microscopy images, a novel methodology with a specific autoencoder architecture is proposed in this paper. Contributions of this paper are as follows:

- We asserted that the performance of U-Net improves significantly by leveraging alternative feature extractors in the encoder of the U-Net model by applying the ResNet-18 backbone pretrained on the ImageNet dataset.
- We propose the utilization of a residual-skip-connection by replacing the plain skip-connection of U-Net's encoder-decoder structure.
- We compare our method with the state-of-the-art algorithms. The qualitative and quantitative results suggest that our method yields superior performance as compared to baseline methods on the MDA-MB-231 dataset.

The paper is organized as follows. In Section 2, we introduce the mechanism of our network and describe its architecture. We, then, provide details of the experimental validation and discussion in Sections 3 and 4, respectively. Finally, conclusions are drawn in Section 5.

2. Method

In our previous studies [23, 24], we employed our approach on the MDA-MB-231 dataset. In this study, we extend our experiments on the DSB2018 dataset and introduce a novel architecture which has the autoencoder structures with the backbones of ResNet-18 (Conv1-Conv5) [25] by excluding the dense layers from ResNet architecture. The model has the initial weights of the ImageNet [26]. Exemplary frames of MDA-MB-231 and DSB2018 dataset is shown in Figure 1. In this paper, we redesigned the structure of skip-connection between the encoder and decoder path, which reduces the semantic gap between the encoder and decoder by applying the specific residual blocks. This alternation in skip-connections makes the model capture more discriminative information. This results in a more robust segmentation over both datasets.

The input image is passed through the backbone of the pretrained ResNet-18, consisting of a consequential residual blocks as its main component. One of the modifications which we have applied in the ResNet18

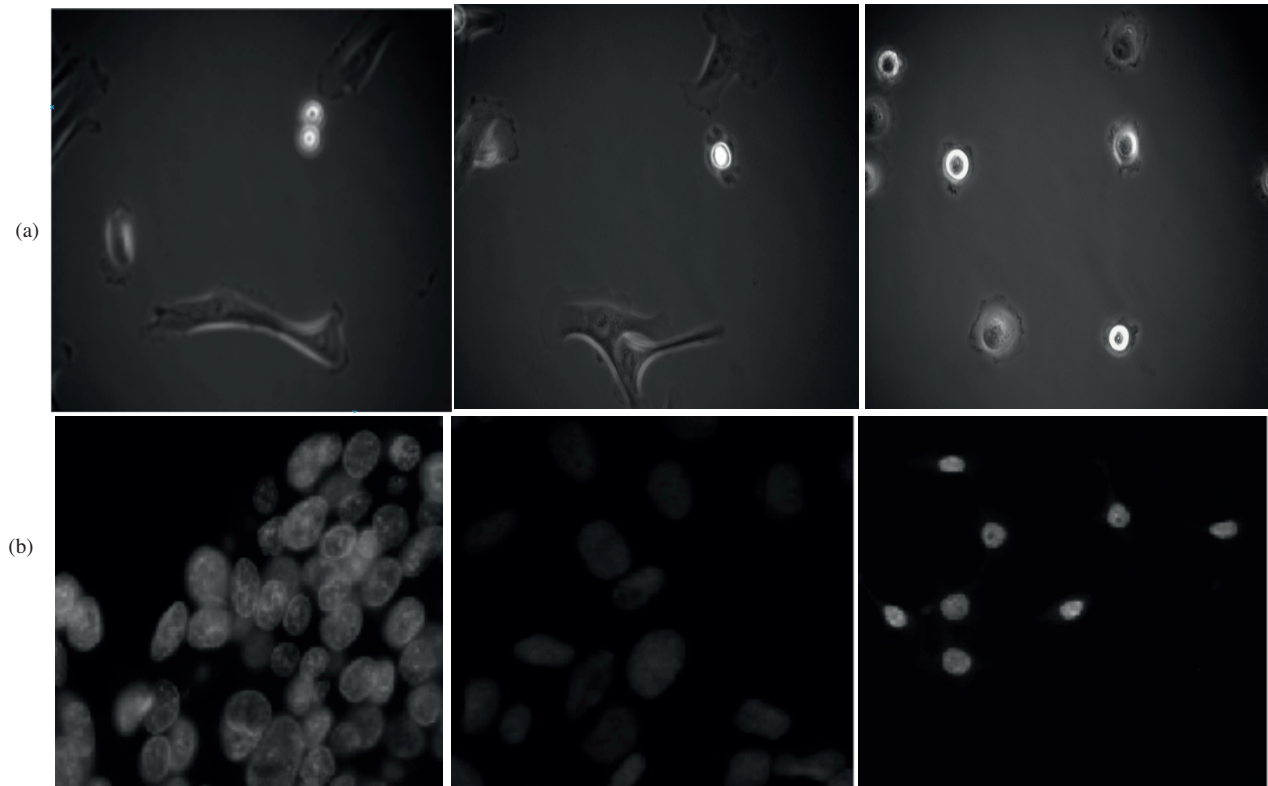


Figure 1. Representation of two types of dataset, (a) breast cancer MDA-MB-231 dataset on phase-contrast optical microscopy images. (b) Data Science Bowl 2018 datasets which contains the nuclei with fluorescence labeled cell in different illumination and appearance.

architecture is to reduce the stride in Conv1 from 2 to 1. Applying the 7×7 layer in the ResNet leads to loss of important information during the downsampling operation, especially for medical datasets. Therefore, after applying of Conv1 in ResNet18, the dimension of the network is not halved. In the backbones of ResNet-18, we have the two residual blocks in each resolution stage, which we have the residual block which is symmetrical to the decoding path with two residual blocks for each dimension level. These residual blocks help the encoder extract the important features from the input image, which are then transferred to the decoder. The residual blocks contain 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 feature map is regularly doubled after doubling the stride in each resolution stage. In the decoder section, to have symmetric operation, we apply upsampling by 4 times 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. Each decoder network starts with a 4×4 transpose convolution that doubles the spatial dimensions of the incoming feature maps. The schematic structure of the proposed network is shown in Figure 2. Moreover, as an extension, we applied the residual blocks as an alternative of plain skip-connection, which increases the performance of the network on the MDA-MB231 and DSB2018 datasets. The proposed skip-connections aim to reduce the semantic gap between the encoder and decoder part, the residual pathway is introduced to transfer all the low-level semantic information from the encoder to the decoder parts. The skip-connections are flourished with 4 residual blocks which are gradually increased from 1 to 4 by reduction of

spatial dimension.

In the decoder fold, there is the features map which comes from pretrained encoder transferred via the skip-connection to each dimensional stages. Each stage contains two 3×3 convolution layers, where each layer is followed by a batch normalization layer and a ReLU function sequentially. Finally, the last decoder block passes through a 1×1 convolution layer, which is followed by a sigmoid activation function for generation of determined binary mask as a prediction of the network.

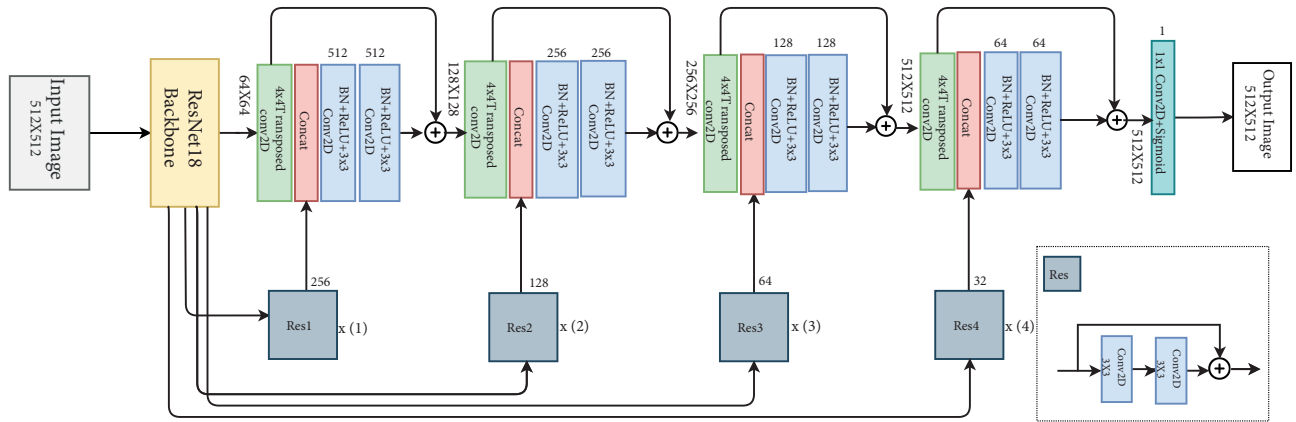


Figure 2. The architecture of the proposed model contains the backbones of the ResNet-18 in the encoder side. The skip-connections provide access to low-level semantic information via the residual blocks. To match image dimensions, the number of skip-connections is set to four and residual blocks are concatenated recursively. Skip-connections are forwarded with residual block to the decoder which sequentially increased by increasing of dimension. The residual pathway includes residual blocks each of which contains successive 3×3 convolution with the same number of output channels for each block.

2.1. Loss function

In this study, we integrate a hybrid loss function, which contains binary cross-entropy, \mathcal{L}_{BCE} , and dice loss, \mathcal{L}_{Dice} . Dice loss is added up to the proposed function which has robust performance on the boundary regions. The employed loss function is shown in (1)–(3), where p_i is predicted binary segmentation and g_i the ground truth binary volume which is part of g .

$$BCE(p, g) = -(p \log(g) + (1 - p) \log(1 - g)) \tag{1}$$

$$DL(p, g) = 1 - \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{2}$$

$$\mathcal{L}_T = \mathcal{L}_{BCE} + \mathcal{L}_{Dice} \tag{3}$$

2.2. Redesigning skip-connection

We proposed a residual skip-connection, which is a robust alternative for the skip-connection of the U-Net, which concatenates the encoder features to the decoder features. Therefore, we also gradually reduce the number of convolutional blocks used along the residual pathway. About the number of residual blocks along the skip-connection, we gradually reduce the number of residual blocks from 4 to 1 subblocks from high-resolution stages

to the low-resolution stage in the architecture, which we applied on ResUNet18 architecture as an extension. The model contains 4 residual blocks each of which contains 3×3 with the same number of output channels for each block which are followed by BN and ReLU in each block and followed with residual connections for further subblock in each stage. Moreover, about the number of feature maps in the proposed pathway, we applied the set of (32, 64, 128, 256) filters in the each subblock of the four residual blocks in each stage respectively. The schematic structure of the proposed residual pathway is shown in Figure 2. The identity mapping tries to learn an identity function since the input is directly passed to the output. It also leads to better gradients during the back-propagation. This pathway is employed on ResNet-18-U-Net, which is entitled ResNet18-UNet+RP and efficiently increases the performance of the architecture in comparison to applying the plain skip-connection in ResNet-18-U-Net in MDA-MB-231 and DSB2018 dataset.

2.3. Test-time augmentation

TTA employs a specific mechanism that includes augmentation, prediction, reaugmentation, and merging steps. In TTA, augmentation is specified on the testset. The number of augmentation for each testset can be variable based on the level of uncertainty estimation on the dataset. Therefore, the model predicts the new-generated testset besides the original testset in the prediction image. Afterward, the generated image from original testsets should reaugment after the prediction. In the last step, all of the generated images for each testset are merging based on the number of voting for each image by summing and averaging all predicted images with probability maps. The schematic diagram of the TTA is represented in Figure 3. In this paper, we address the problem of cell segmentation with the help of TTA. We employed this mechanism to decrease the uncertainty on the regions which are wrongly predicted on the testset. We achieve better outcome quantitatively and qualitatively compared to the model whenever does not contain the TTA in the prediction step.

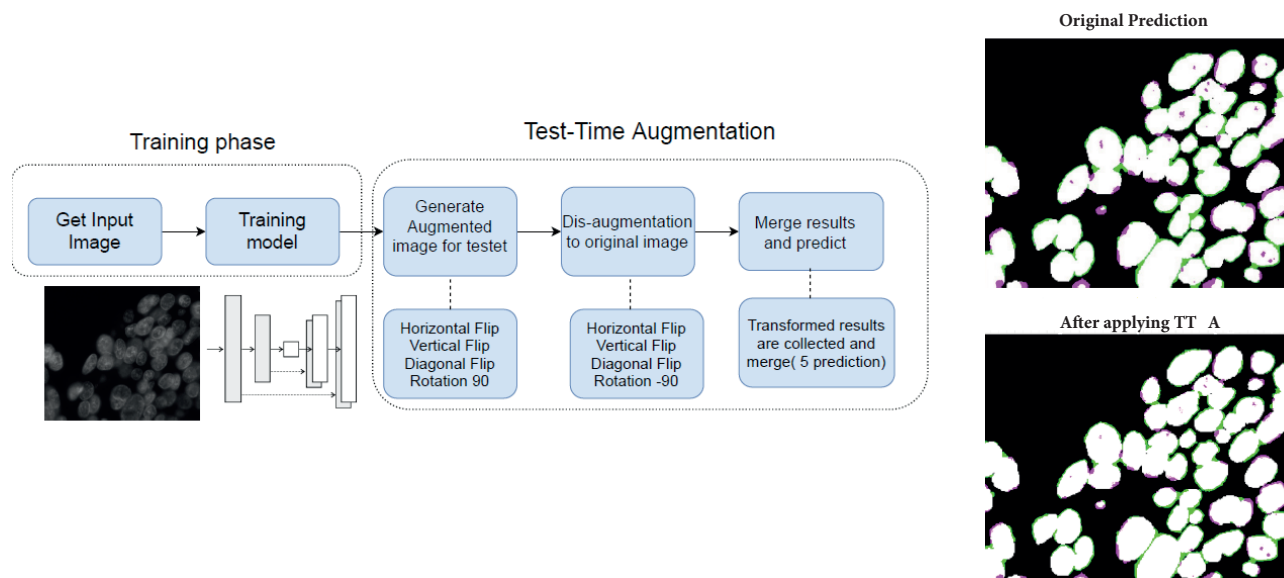


Figure 3. Schematic diagram of TTA which illustrates the process of TTA in the prediction step in our proposed approach.

3. Experimental validation

In this section, we present datasets, evaluation metrics, implementation details, and experimental results to validate the proposed method.

3.1. Dataset

MDA-MB-231 [27] corresponds to the invasive breast cancer cells with mesenchymal morphology captured using an Olympus IX71 microscope. The final dataset contains 600 frames of PCM images, each with a dimension of 2568×1912 pixels ($0.117\mu\text{m} \times 0.117\mu\text{m}$). For manual annotation of the cell boundary on the frames, the Fiji distribution of ImageJ [28] is leveraged. We manually annotated 45 of the frames, 30 of which are used for training while the rest is utilized for the test set.

Data Science Bowl 2018 [29] (DSB2018) is one of the challenging datasets that contains bright-field and fluorescence images of nuclei cells acquired under different illumination and appearance. The dataset contains 670 frames, we employed 509 frames for training, and 161 frames for the validation of the set. Moreover, to facilitate fair comparisons with existing approaches, we follow the evaluation protocol outlined for each method.

3.2. Evaluation metrics

For performance evaluation of the approaches, the well-known metrics will be proposed for measuring the efficiency of the results. The formula of precision, recall, intersection over union (IoU) and dice index are presented in (4)–(7), respectively, where P denotes the predicted result, G denotes the ground-truth, and n_{tp} , n_{fp} , and n_{fn} indicate the numbers of true positives, false positives, and false negatives, respectively.

$$Precision = \frac{n_{tp}}{n_{fp} + n_{tp}} \quad (4)$$

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

$$JS(P, G) = \frac{|P \cap G|}{|P \cup G|} = \frac{|P \cap G|}{|P| + |G| - |P \cap G|} \quad (6)$$

$$Dice(P, G) = \frac{2|P \cdot G|}{|P| + |G|} \quad (7)$$

3.3. Implementation details

We applied the data augmentation on the training set whose main transformations include rotation, elastic transformation, and horizontal and vertical flip. To have robust prediction on the testset, we employed the TTA mechanism which comprises the rotation at 90 degree, vertical, horizontal and diagonal flip in the prediction phase. The backbones of the proposed model which is utilized in the encoder part has the pretrained weights of ImageNet and set the other parts with a standard Gaussian distribution. Moreover, the backbone's layers prevent their weights from being fixed during the training step. All models are trained on single NVIDIA TitanX graphics card which has a 12 GB memory. We used the batch size of 8 in our implementation. There was a trade-off in accuracy when we decreased the batch size; because smaller batch size can lead to overfitting. The training of the two datasets all used the ADAM [30] optimizer, the initial learning rate was 0.001, and

the gradient descent algorithm was implemented. Due to limitation of the memory, we resized the dataset to 512×512 pixels and 256×256 pixels on MDA-MB231 and DSB2018, respectively. The proposed method is implemented in Keras with Tensorflow backend. To evaluate the performance of the model efficiently, we employed the 3-fold cross-validation on the networks and reported the best fold performances in the results section to all models on our dataset.

3.4. Results

In what follows, we first conduct qualitative and quantitative results on the DSB2018 and the MDA-MB-231 datasets. We then justify the effectiveness of the use of TTA on the two refereed datasets. The datasets consist of challenging phase-contrast and brightfield optical microscopy image series.

The qualitative results of the proposed method are presented in Figure 4. As the results show in Table 1, we achieved Jaccard indexes of $87.9\% \pm 1.7$, $87.1\% \pm 2.3$ and $89.2\% \pm 1.3$ with ResNet18-UNet, ResNet18-FPN, and ResNet18-UNet+RP on MDA-MB-231 dataset, respectively. The results of this comparison without implementation of TTA are represented in Table 2. Moreover, to better analyze the reliability of the applied methods, we add up the standard deviation to each evaluation metric which delegates the average scores among the testset in the applied dataset. Therefore, it proves that our proposed approach not only improves segmentation accuracy but also it is effective in the separation of overlapping cells in some of the adjacent or overlapping cells. As the experimental results presented in the tables suggest, our proposed approach outperforms the traditional U-Net and the state-of-the-art architectures in terms of segmentation performance. Moreover, when our method is equipped with the TTA, it achieves a better average performance than all other alternatives. According to the qualitative results, classical methods like PHANTAST and EGT are not satisfactory in the extraction of the cell border and the RoI in MDA-MB-231 dataset. Our proposed method outperforms the state-of-the-art and achieves a noticeably more robust performance in the extraction of cell boundaries.

Table 1. Quantitative results of the different architectures applied on the MDA-MB-231 dataset after applying TTA. EGT and PHANTAST methods are not proceeded by TTA.

Networks	JS(%)	DC(%)	Precision(%)	Recall(%)
EGT[2]	42.2 ± 3.8	59.4 ± 3.1	45.7 ± 3.8	84.7 ± 2.4
PHANTAST [1]	54.6 ± 4.4	70.6 ± 3.5	68.2 ± 4.1	73.3 ± 3.6
U-Net [9]	86.4 ± 1.9	92.7 ± 1.7	93.3 ± 1.9	92.4 ± 1.8
BCDUNet [19]	82.2 ± 2.4	90.2 ± 2.1	89.2 ± 2.8	91.2 ± 1.6
Tip-Net [23]	85.4 ± 2.7	92.1 ± 2.2	93.8 ± 2.9	90.5 ± 1.9
ResNet18-U-Net [24]	87.9 ± 1.7	93.5 ± 1.3	95.0 ± 1.8	92.1 ± 1.5
ResNet18-UNet+RP	89.2 ± 1.3	94.3 ± 1.1	95.8 ± 1.3	92.8 ± 1.6
ResNet18-FPN [24]	87.1 ± 2.3	93.1 ± 2.0	93.0 ± 2.6	94.1 ± 1.9
Double-UNet [21]	86.9 ± 2.1	92.3 ± 2.3	93.5 ± 1.7	89.8 ± 2.8
DDANet [31]	87.5 ± 2.4	93.9 ± 1.9	94.1 ± 2.0	92.6 ± 2.2

In this experiment, we conduct the qualitative comparisons on DSB2018 dataset in Figure 5. As can be seen in the quantitative results reported in Table 3, we achieved Jaccard indexes of $81.1\% \pm 2.2$, $83.1\% \pm 1.8$, and $85.0\% \pm 2.1$ with ResNet18-U-Net, ResNet18-FPN, and ResNet18-UNet+RP, respectively. The results of this comparison without applying TTA on DSB2018 dataset are tabulated in Table 4. Due to high

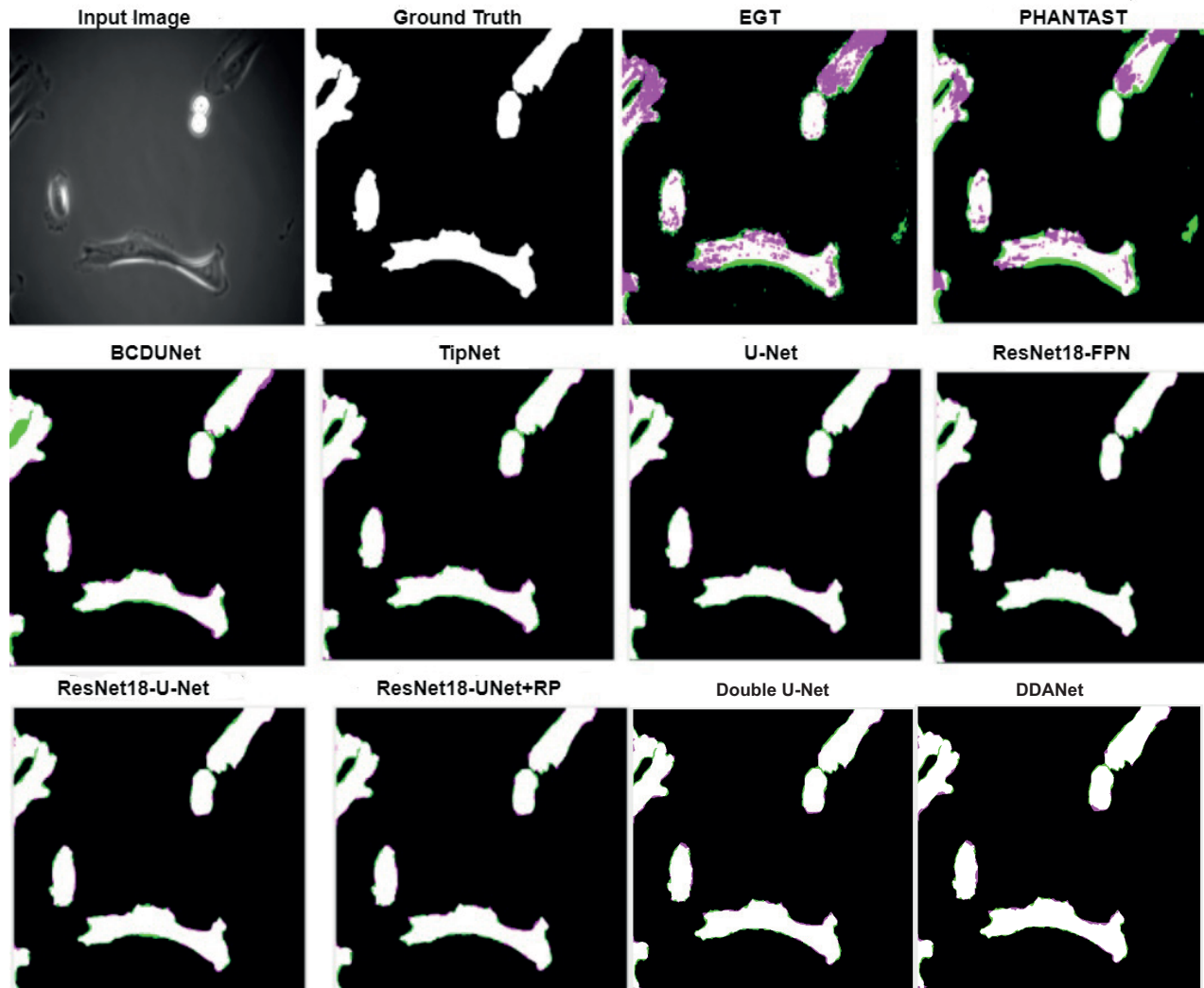


Figure 4. Visualization of the segmentation results obtained from the MDA-MB-231 dataset. Magenta marks represent the false negatives and green marks indicate the false positives.

perturbation in some frames of DSB2018 dataset, the 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 DSB2018 dataset.

Both qualitative and quantitative results suggest that our proposed approach not only improves segmentation accuracy but also it is effective in the separation of overlapping cells as compared to the baseline methods and even state-of-the-art methods, such as BCDUNet [19], DDANet [31], and Double-UNet [21], on MDA-MB-231 dataset, and achieves competitive results on the DSB2018 dataset. 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. We have conducted an analytical ablation study on the impact of residual pathway on the performance of model's accuracy. For this, we demonstrate the superiority of the proposed skip-connection on the MDA-MB-231 and DSB2018 datasets. To allow for a fair comparison, we evaluated the performances of BCDUNet, U-Net, and SegNet using the codes released by the

Table 2. Quantitative results of the architectures applied on MDA-MB-231 dataset without applying TTA as a postprediction technique.

Networks	JS(%)	DC(%)	Precision(%)	Recall(%)
U-Net [9]	84.4 ± 2.0	90.8 ± 1.9	92.5 ± 2.3	89.1 ± 1.8
BCDUNet [19]	81.3 ± 2.6	89.7 ± 2.2	88.6 ± 2.9	90.7 ± 2.0
Tip-Net[23]	83.2 ± 3.1	90.8 ± 2.5	91.6 ± 3.3	90.1 ± 2.1
ResNet18-U-Net	86.6 ± 2.2	92.8 ± 1.6	93.8 ± 2.1	91.8 ± 1.7
ResNet18-UNet+RP	87.2 ± 1.8	93.1 ± 1.5	94.5 ± 1.7	91.8 ± 1.6
ResNet18-FPN	85.2 ± 2.6	92.0 ± 2.3	91.9 ± 2.7	92.1 ± 2.2
Double-UNet[21]	86.3 ± 2.6	92.1 ± 2.4	91.8 ± 1.8	88.2 ± 3.1
DDANet[31]	86.7 ± 2.8	92.5 ± 2.3	93.6 ± 2.1	91.1 ± 2.6

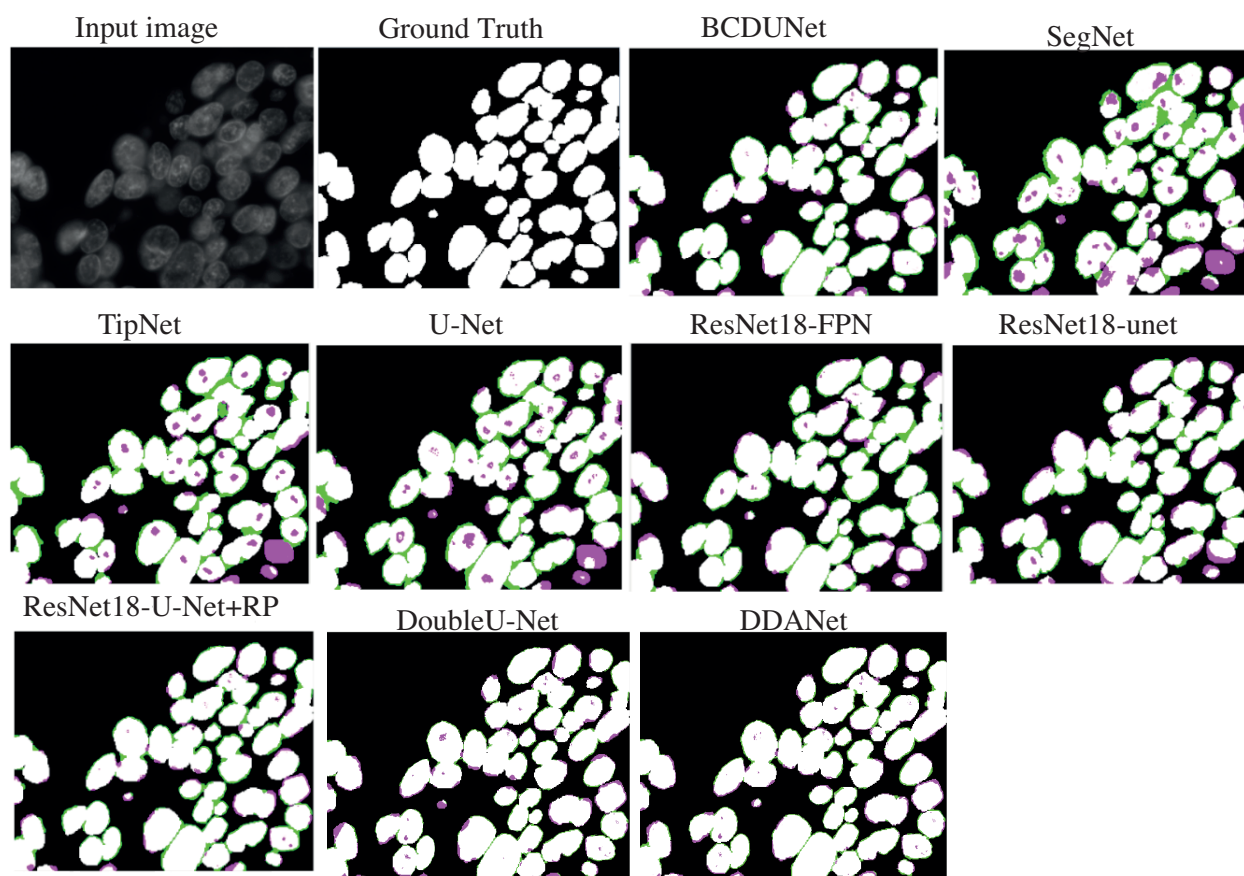


Figure 5. Visual examples of segmentation results obtained from the DSB2018 dataset.

corresponding authors under the same evaluation protocol. The ResNet18-FPN, ResNet-18-U-Net, and Tip-Net are the methods we had proposed in our previous study for the segmentation of PCM image series.

Table 3. Qualitative results of the different architectures applied on the DSB2018 dataset after applying TTA.

Networks	JS(%)	DC(%)	Precision(%)	Recall(%)
SegNet[14]	72.3 ± 2.9	83.9 ± 3.8	81.5 ± 4.5	86.5 ± 3.3
BCDUNet [19]	86.7 ± 1.6	92.9 ± 1.9	93.9 ± 3.5	91.8 ± 2.2
Tip-Net[23]	74.1 ± 2.8	85.1 ± 2.5	87.5 ± 2.9	86.9 ± 2.4
U-Net[9]	79.2 ± 2.2	88.4 ± 2.1	87.3 ± 2.5	89.5 ± 1.9
DDANet[31]	86.9 ± 2.0	93.2 ± 1.6	94.6 ± 1.7	92.8 ± 1.5
Double-UNet[21]	85.8 ± 2.1	90.3 ± 2.7	96.6 ± 1.1	68.3 ± 4.4
ResNet18-U-Net	83.1 ± 1.8	90.7 ± 2.2	92.3 ± 2.3	90.5 ± 1.6
ResNet18-FPN	81.1 ± 2.2	89.6 ± 2.4	91.0 ± 2.7	88.2 ± 1.9
ResNet18-U-Net+RP	85.0 ± 2.1	91.9 ± 1.9	93.0 ± 2.4	90.8 ± 1.6

Table 4. Qualitative results of the different architectures applied on the DSB2018 dataset without applying TTA.

Networks	JS(%)	DC(%)	Precision(%)	Recall(%)
SegNet[14]	69.2 ± 3.2	81.8 ± 4.1	80.1 ± 4.7	83.5 ± 3.6
BCDUNet[19]	85.5 ± 1.7	92.1 ± 2.0	92.5 ± 2.7	91.8 ± 2.5
Tip-Net[23]	73.5 ± 2.9	84.7 ± 2.7	85.9 ± 3.1	83.6 ± 2.6
U-Net[9]	76.4 ± 2.3	86.6 ± 2.3	87.0 ± 2.6	86.2 ± 2.1
DDANet[31]	86.2 ± 2.2	92.6 ± 1.8	93.3 ± 1.9	92.4 ± 1.6
DoubleU-Net[21]	84.0 ± 2.3	88.1 ± 2.9	94.9 ± 1.2	66.1 ± 4.9
ResNet-18-U-Net	82.1 ± 2.0	90.1 ± 2.3	91.5 ± 2.5	88.8 ± 1.7
ResNet18-FPN	79.6 ± 2.6	88.6 ± 2.5	89.7 ± 2.8	87.5 ± 2.2
ResNet18-UNet+RP	84.8 ± 2.0	91.7 ± 2.1	92.7 ± 2.5	90.8 ± 1.8

4. Discussion

We have evaluated the segmentation performances of our proposed methods on image series collected by two different optical microscopy techniques, phase-contrast, and brightfield, which are entitled MDA-MB231 and DSB2018 datasets, respectively. We realized the performance evaluations on these two datasets in a similar fashion; for instance, by employing TTA as a postprocessing step to evaluate its effect on the performance. To this end, the use of TTA increased the Jaccard index from around 1% to 3%. For instance, the Jaccard index of the proposed method is increased from 87.2% to 89.2% on MDA-MB-231 dataset. Regarding the points mentioned earlier, the proposed approach outperformed all the methods on almost all evaluation metrics with respect to the best-performing SOTA in the MDA-MB-231 dataset. Our approach produces high-quality segmentation in both low-annotated and challenging datasets in MDA-MB-231 and DSB2018, respectively. This fact suggests that our method can be employed as a strong baseline for comparison on cell segmentation tasks, especially for datasets with low-annotated frames. Regarding the loss function, we applied the Dice + Binary Cross Entropy which resulted in higher segmentation accuracies in comparison to the use of a single loss function and even with respect to the Jaccard + Binary Cross Entropy.

5. Conclusion and future work

In this study we presented autoencoder networks by employing the specific ResNet-18 in the backbone of the U-Net architecture and redesigned the skip-connection by employing a residual skip-connection on the network. We used transfer learning to expedite training and employed test-time-augmentation mechanisms to help the network focus on uncertain regions of the images. Our network can thus be used to automate the segmentation of datasets that are generally considered too small for deep learning techniques.

The future direction of this research has several branches. In future work, we have a plan to solidify our approach. However, we would like to conduct experiments to determine the best hyperparameter set for the model in more detail in the future. Moreover, we would like to extend our experiments by deploying the more pretrained backbone for training our model. In addition, we would like to evaluate the performance of proposed method for different modality of datasets. As stated earlier, we are going to extend our method on time series data by establishing lineage relationships that provide detailed information about cell behavior over time.

Acknowledgments

This work has been supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant 119E578. The data used in this study is collected under the Marie Curie IRG grant (no: FP7 PIRG08-GA-2010-27697). Aydın Ayanzadeh's work is supported, in part, by Vodafone Turkey, under project no. ITUVF20180901P04 within the context of ITU Vodafone Future Lab R&D program.

Author contributions

AA designed the model and the computational framework, carried out the experiment, analyzed the data, and drafted the manuscript with support from BUT, and DÜ, DPO, ÖYÖ, BUT, and DÜ conceived the project idea. DÜ supervised the project. DPO, ÖYÖ, and SÖ designed and performed the biological experiments and collected the image data. All the authors discussed the results and contributed to the final manuscript.

References

- [1] Jaccard N, Griffin LD, Keser A, Macown RJ, Super A et al. Automated method for the rapid and precise estimation of adherent cell culture characteristics from phase contrast microscopy images. *Biotechnology and Bioengineering* 2014; 11 (3): 504-517.
- [2] Chalfoun J, Majurski M, Peskin A, Breen C, Bajcsy P et al. Empirical gradient threshold technique for automated segmentation across image modalities and cell lines. *Journal of Microscopy* 2015; 260 (1): 86-99.
- [3] Ambühl ME, Brepant C, Meister JJ, Verkhovsky AB, Sbalzarini IF. High-resolution cell outline segmentation and tracking from phase-contrast microscopy images. *Journal of Microscopy* 2012; 245 (2): 161-170.
- [4] Seroussi I, Veikherman D, Ofer N, Yehudai-Resheff S, Keren K. Segmentation and tracking of live cells in phase-contrast images using directional gradient vector flow for snakes. *Journal of Microscopy* 2012; 247 (2): 137-146.
- [5] Chowdhury S, Kandhavelu M, Yli-Harja O, Ribeiro AS. Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE). *BMC Bioinformatics* 2013; 14 (10): 1-3.
- [6] Essa E, Xie X, Errington RJ, White N. A multi-stage random forest classifier for phase contrast cell segmentation. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Milan, Italy; 2015. pp. 3865-3868.

- [7] Su H, Yin Z, Huh S, Kanade T. Cell segmentation in phase contrast microscopy images via semi-supervised classification over optics-related features. *Medical Image Analysis* 2013; 17 (7): 746-765
- [8] Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods* 2019; 16 (1): 67-70.
- [9] 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*; Munich, Germany; 2015. pp. 234-241. Springer, Cham.
- [10] Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*; Stanford, CA, USA; 2016. pp. 565-571.
- [11] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- [12] Leclerc S, Smistad E, Grenier T, Lartizien C, Ostvik A et al. RU-Net: A refining segmentation network for 2D echocardiography. In *2019 IEEE International Ultrasonics Symposium (IUS) 2019*, pp. 1160-1163. IEEE.
- [13] Alom MZ, Yakopcic C, Taha TM, Asari VK. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). In: *NAECON 2018-IEEE National Aerospace and Electronics Conference*; Dayton, OH, USA; 2018. pp. 228-233.
- [14] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017; 39 (12): 2481-2495.
- [15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014.
- [16] Igloukov V, Shvets A. Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*. 2018.
- [17] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3-11. Springer, Cham, 2018.
- [18] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Honolulu, HI, USA; 2017. pp. 4700-4708.
- [19] Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S. Bi-directional ConvLSTM U-Net with densely connected convolutions. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops 2019*.
- [20] Tsai HF, Gajda J, Sloan TF, Rares A, Shen AQ. Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. *SoftwareX* 2019; 9: 230-237.
- [21] Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD. Doubleu-net: A deep convolutional neural network for medical image segmentation. In: *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) 2020*, pp. 558-564. IEEE.
- [22] Cao X, Lin Y. Caggnet: Crossing aggregation network for medical image segmentation. In: *2020 25th International Conference on Pattern Recognition (ICPR)*; Milan, Italy; 2021. pp. 1744-1750. IEEE.
- [23] Ayanzadeh A, Yağar HO, Özuysal ÖY, Okvur DP, Töreyn BU et al. Cell segmentation of 2D phase-contrast microscopy images with deep learning method. In: *2019 Medical Technologies Congress (TIPTEKNO) 2019*, pp. 1-4. IEEE.
- [24] Ayanzadeh A, Özuysal ÖY, Okvur DP, Önal S, Töreyn BU et al. Deep learning based segmentation pipeline for label-free phase-contrast microscopy images. In: *2020 28th Signal Processing and Communications Applications Conference (SIU) 2020*, pp. 1-4. IEEE.
- [25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Las Vegas, NV, USA; 2016. pp. 770-778.

- [26] Russakovsky O, Deng J, Su H, Krause J, Satheesh S et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*. 2015;115 (3):211-52.
- [27] Binici RC, Şahin U, Ayanzadeh A, Töreyn BU, Önal S et al. Automated segmentation of cells in phase contrast optical microscopy time series images. In 2019 Medical Technologies Congress (TIPTEKNO) 2019; pp. 1-4. IEEE.
- [28] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 2012; 9 (7): 671-675.
- [29] Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nature Methods* 2019; 16 (12): 1247-1253.
- [30] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.
- [31] Tomar NK, Jha D, Ali S, Johansen HD, Johansen D et al. DDANet: Dual decoder attention network for automatic polyp segmentation. arXiv preprint arXiv:2012.15245. 2020.