2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Effective Deep Learning Framework for Cell Segmentation in Microscopy Images

Sherry Lin[1] and Narges Norouzi[2]

Abstract— **Cell segmentation is a common step in cell behavior analysis. Reliably and automatically segmenting cells in microscopy images remains challenging, especially in differential inference contrast microscopy images and phase-contrast microscopy images. In this paper, we propose a deep learning solution combining a Mask RCNN architecture with Shape-Aware Loss to produce cell instance segmentation. Our approach outperforms prior works in cell segmentation, achieving an IOU of 91.91% on the DIC-C2DH-HeLa dataset and an IOU of 94.93 % on the PhC-C2DH-U373 dataset. Our framework can calculate cell instance segmentation masks from both types of microscopy images without any additional post-processing.**

Clinical relevance— **The proposed approach produces accurate instance segmentation in Differential Inference Contrast and Phase-Contrast microscopy images. The segmentation results can be reliably used in cell behavior analysis and cell tracking.**

I. Introduction

Image segmentation is the process of labeling pixels in images by the object or object instances to which they belong, producing segmentation masks that mark the regions of such objects or object instances. Cell segmentation usually involves segmenting the entire cell or particular cell structures, such as the nuclei. Such a task is particularly challenging in label-free microscopy images such as Differential Inference Contrast (DIC) images and Phase-Contrast (PhC) images due to the low contrast between the cells and the background image pixels. Additionally, since the cells are not labeled with fluorescence markers, the images' resolution may be inadequate.

With the recent advancement in deep learning, Convolutional Neural Networks (CNNs) have shown remarkable effectiveness in segmentation problems. To have a more robust feature extraction framework while having a better handle over pixel localization and the information loss from upsampling, Ronneberger et al. proposed U-Net[1], an architecture that has a contracting path and an expanding path. The symmetric contracting and expanding paths provide an efficient way to capture context and local features. The model reaches an Intersection Over Union (IOU) of 92.03% and 77.56% in producing semantic segmentation for DIC-C2DH-HeLa and PhC-C2DH-U373 datasets. The model efficacy inspires other work to use U-Net and post-processing for producing segmentation masks for particular cell structures or cell instance segmentation. Al-Kofahi et al. used U-Net on their custom PhC dataset to predict cell locations and nuclei by outputting probability maps that show the likelihood of parts being cell nuclei or cytoplasm [2]. The probability maps are then used to generate seeds for the watershed to produce segmentation results. Lux and Matula (MU-Lux-CZ) also utilized U-Net on the DIC-C2DH-HeLa dataset to produce markers representing approximate locations and shapes of the cells [3]. The markers go through watershed transform in post-processing and produce cell instance segmentation masks, achieving an IOU of 86.3%. To further improve model performance, Pena et al. (CALT-US) proposed using a novel loss function that combines Youden's J statistic regularization and cross-entropy in U-Net to produce semantic segmentation, and their framework produces instance segmentation using a probabilistic post-processing [4]. They achieve high segmentation accuracy, an IOU of 87% and 92.7%, for DIC-C2DH-HeLa and PhC-C2DH-U373 datasets, respectively. Although the previous models are effective, they require post-processing to generate the final segmentation results, which can be computationally complex and inefficient.

In this work, we propose Mask RCNN (Region Based Convolutional Neural Network) deep architecture to produce cell segmentation that requires no extra post-processing steps. We use our Shape-Aware Loss, a distance-based pixel-wise weighted cross-entropy loss, to help the model better learn the segmentation boundaries. Our segmentation accuracy for both datasets outperforms all existing models for the two datasets in Cell Segmentation Benchmark[5]. We describe our approach and demonstrate our results in the later sections of this paper.

II. Dataset

We train our model using the DIC-C2DH-HeLa dataset and the PhC-C2DH-U373 dataset from the ISBI Cell Tracking Challenge [6]. The DIC-C2DH-HeLa dataset contains two time-lapse video sequences of 84 frames, for a total of 168 frames of DIC microscopy images of HeLa cells on flat glass. The PhC-C2DH-U373 dataset contains two time-lapse video sequences of 115 frames, for a total of 230 PhC microscopy images of Glioblastoma-astrocytoma U373 cells on a polyacrylamide substrate. Each frame has a corresponding instance segmentation mask, and each cell has a different instance ID. Figure 1 shows examples of raw image and ground truth segmentation masks in the two datasets.

*Sherry Lin (email: sherrylin42@gmail.com) and Narges Norouzi (email: nanorouz@ucsc.edu) are with the University of California, Santa Cruz, Santa Cruz, CA 95064 USA

For each dataset, we shuffle the images and split them into a training set and a validation set. For DIC-C2DH-HeLa dataset, the training set consists of 134 images, and the validation set consists of 34 images. For PhC-C2DH-U373 dataset, the training set consists of 184 images, and the validation set consists of 46 images.

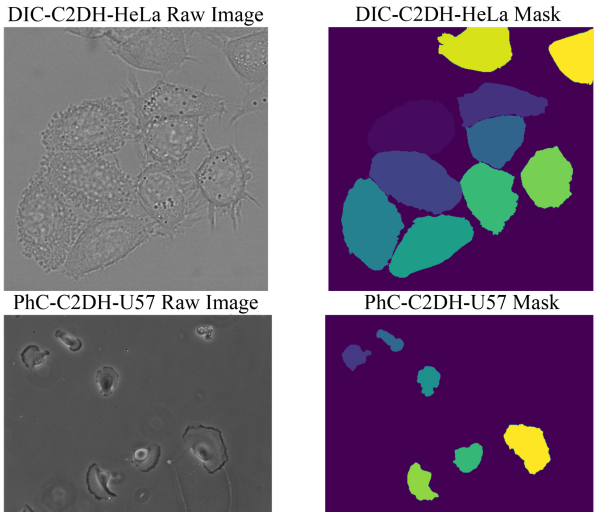

Fig. 1. DIC-C2DH-HeLa and PhC-C2DH-U373 Raw Images and Ground Truth Segmentation Masks

III. Methodology

We use Mask RCNN as our base architecture to directly produce instance segmentation. We apply Distance Transform to compute a weight map and incorporate the weight map in our Shape-Aware Loss. The model produces final instance segmentation masks, and no additional post-processing is required in our method, differentiating our method from other segmentation paradigms in the Cell Tracking Challenge[6].

A. Shape-Aware Loss

Binary cross-entropy (BCE) loss is a typical loss function used in image segmentation. However, when the proportion of foreground pixels to background pixels is significantly imbalanced, the model may not be able to segment the under-represented pixel class. Additionally, in dense environments where segmentation masks are very close to one another, the model may not correctly segment the boundaries.

Focal Loss is proposed to handle situations where the proportion of foreground to background pixels is highly imbalanced [7]. Focal Loss modifies BCE loss to increase relative loss when pixels are incorrectly labeled, pushing the model to update weights to prioritize correctly classifying the incorrect pixels. Focal Loss shows excellent efficacy in improving model results on images with significantly imbalanced pixel classes. However, because it does not explicitly emphasize mask boundaries or the small gaps between cells, Focal Loss may not push the model to yield accurate segmentation masks in dense environments, especially when mask boundaries are complex.

Using weight maps to capture the shapes of segmentation masks in pixel-wise weighted cross-entropy loss could help the model learn complex cell boundaries. The effectiveness of this approach is demonstrated by Ronneberger et al. in U-Net [1], and their weight map is computed using a pixel's distance both to the nearest cell boundary and to the second-nearest cell boundary. We compute a weight map based on each pixel's distance to the nearest cell boundary to reduce computational complexity. The pixel-wise weighted cross-entropy loss is defined as:

$$L = -\sum_{t}^{n} w_t \times log(p_t) \tag{2}$$

where w_t is the value of pixel t on the weight map and p_t is the softmax at pixel t.

The weight map intendeds to capture pixel importance from two aspects: pixel class (i.e., background or foreground) and pixel distance to the closest boundary. We compute the weight map for each ground truth segmentation mask using

$$w_t = w_c + \alpha \times exp(\frac{-D_t^2}{2\sigma^2}) \tag{3}$$

where $w_c \in \mathbb{R}^+$ is a parameter to adjust the imbalance between the number of foreground and background pixels. The parameter is set to 1 when pixel t is a background pixel, and is set to the ratio between the total number of background pixels and total number foreground pixels when pixel t is a foreground pixel. $D_t \in \mathbb{R}^+$ is the distance from pixel t to its nearest cell boundary. $\alpha \in \mathbb{Z}$ is a parameter that controls the relative weight between the pixel class-based term and the pixel distance-based term. $\sigma \in \mathbb{Z}$ is a parameter that controls the gradual decrease of the distance-based term in the w_t. We illustrate weight maps with different σ in Figure 2 and weight maps with different α in Figure 3. We chose these hyper-parameters experimentally and set $\alpha = 10$ and $\sigma = 5$.

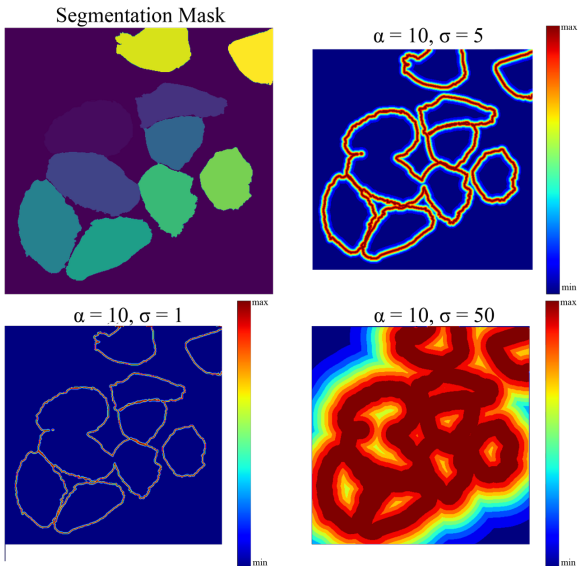

Fig. 2. Examples of weight maps with different σ values.

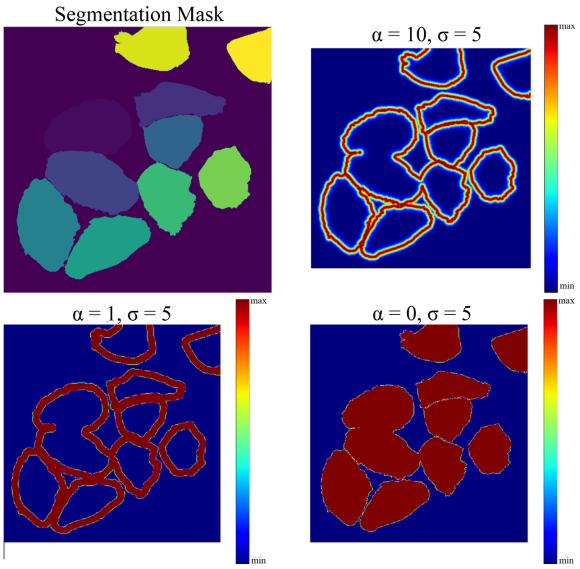

Fig. 3. Examples of a weight maps with different α values.

B. Deep Learning Architecture

We use Mask RCNN architecture [8] to produce instance segmentation masks. The architecture uses a Region Proposal Network (RPN) to propose regions that are likely to contain an object. After the RPN, the model has two branches, one for predicting object class labels, object class scores, object bounding boxes, and the other for predicting object segmentation masks. To use the model to predict cell segmentation masks, we change the number of object classes in both branches to 2(i.e., cell and background) to reflect the nature of our task.

To increase the training efficiency, we initialize the model with Mask RCNN pre-trained weights, use Kaiming normal weight initialization [9] to initialize weights in the modified box branch and mask branch, and use Adam optimizer to optimize mini-batch gradient descent. In our experiment, we set the learning rate in the Adam optimizer to 0.0005. We also apply a learning rate scheduler to decay the learning rate by 0.1 every three epochs to avoid overshooting the optima.

Additionally, to explore Focal Loss and Shape-Aware Loss's efficacy compared to the BCE loss, we train three models with U-Net architecture on each dataset. We initialize the weights with Xavier normal initialization [10] and use Adam optimizer with the learning rate set to 0.001.

IV. Experiments

A. Model Training

We train three Mask RCNN models with three different loss functions for each dataset: BCE loss, Focal Loss, and Shape-Aware Loss. We use random horizontal flip to augment the images and their masks during the training phase to increase robustness. We use early stopping in the training process to avoid overfitting. Additionally, we train three U-Net models using three different loss functions to compare the performances with those of the Mask RCNN models.

TABLE I
Cell Segmentation Experiment Results and Comparison

Architecture	Loss Function	DIC-C2DH-HeLa	PhC-C2DH-U373
Mask RCNN	Shape-Aware Loss	**0.919***	**0.949***
Mask RCNN	BCE Loss	0.905	0.948
Mask RCNN	Focal Loss	0.918	0.920
MU-Lux-CZ [3]	Weighted MSE Loss	0.863	-
CALT-US [4]	Custom Loss Function	0.870	0.927
U-Net	BCE Loss	0.663	0.566
U-Net	Focal Loss	0.670	0.597
U-Net	Shape-Aware Loss	0.704	0.820

B. Results and Discussion

We evaluate our model results on the respective validation sets with IOU, which is Segmentation Accuracy as specified in the Cell Tracking Challenge Evaluation Methodology [11].IOU is a metric based on the Jaccard Similarity Index and is described as:

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

where A denotes the predicted segmentation mask and B denotes the ground truth segmentation mask. The Mask RCNN experiment results are shown in Table I.

The best results, achieved previously by Pena et al. (CALT-US), reach an IOU of 0.870 and 0.927 for DIC-C2DH-HeLa and PhC-C2DH-U37, respectively [4], and we achieve an IOU of 0.919 for the DIC-C2DH-HeLa dataset and 0.949 for the PhC-C2DH-U37 dataset. Based on the results, we conclude that Shape-Aware Loss significantly outperforms BCE Loss and Focal Loss for the cell segmentation task.

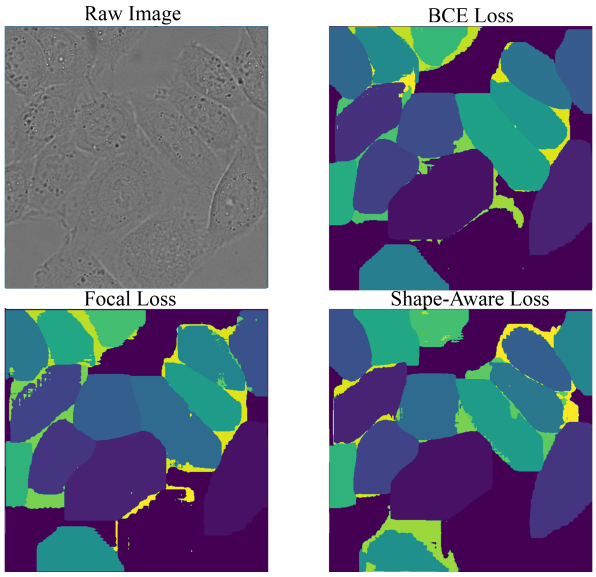

Fig. 4. Instance segmentation results for a Sample from DIC-C2DH-HeLa dataset.

Figure 4 shows that the cells in DIC-C2DH-HeLa are densely populated but that Mask RCNN effectively produces instance segmentation for cells in the image. Comparatively,

results from the model using Shape-Aware Loss have more distinctive cell boundaries. Figure 5 illustrates segmentation results for the PhC-C2DH-U37 dataset. The two left-most cells are segmented as one cell in both BCE loss and Focal Loss, but they are correctly segmented as separate cells in Shape-Aware Loss. Figure 6 shows the segmentation results from different loss functions in the U-Net experiments. The Shape-Aware Loss produces more accurate cell segmentation boundaries, despite the complexity in cell features and the shape of the cells.

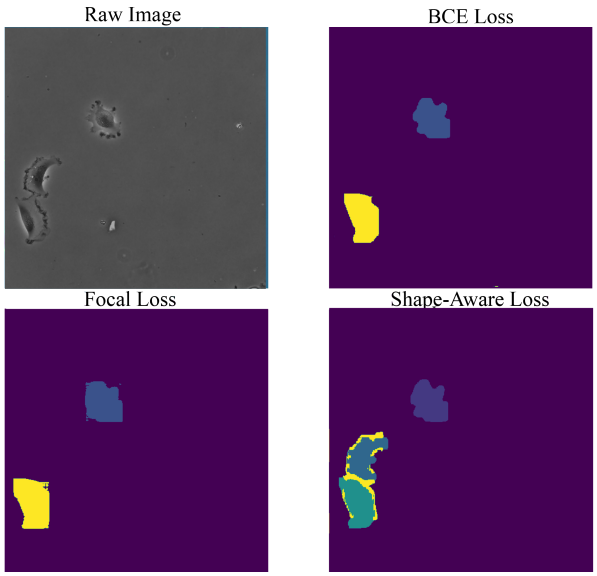

Fig. 5. Instance segmentation results for a sample from PhC-C2DH-U373 dataset.

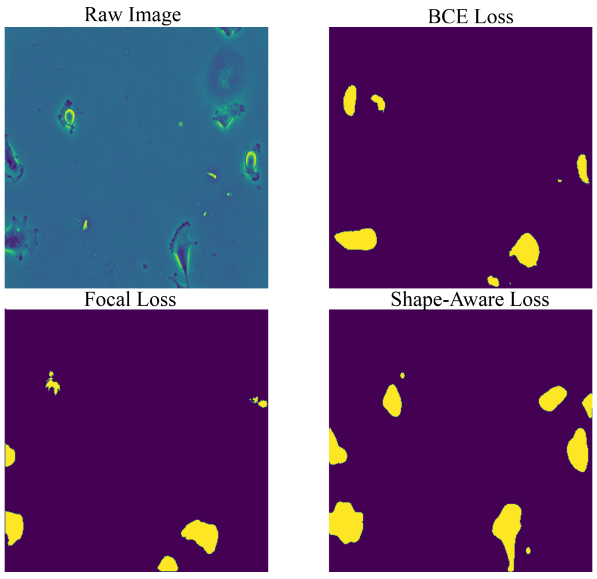

Fig. 6. Cell segmentation results using U-Net Architecture based on Different Loss Function

V. Conclusion and Future Work

This paper proposes a framework for cell segmentation in DIC and PhC microscopy images using a combination of Mask RCNN and Shape-Aware Loss. Such a framework can predict cell instance segmentation masks from images without additional post-processing. Since the Shape-Aware Loss only considers a pixel's distance to the closest cell boundary, computing weight maps using Distance Transform is a manageable and straightforward process. Additionally, we show that the simple Shape-Aware Loss effectively improves segmentation performance in both Mask RCNN and U-Net models.

In the future, we will work on proposing a unified framework for analyzing different types of microscopy images so that effective cell segmentation can be done irrespective of the microscopy technique. To achieve a unified solution, we are interested in learning whether using an ensemble framework could capture a variety of different appearance features. Since our current models give accurate instance segmentation results, we are also looking into developing a reliable cell tracker framework on top of our existing segmentation framework.

References

[1] 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*, ser. Lecture Notes in Computer Science, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[2] Y. Al-Kofahi, A. Zaltsman, R. Graves, W. Marshall, and M. Rusu, "A deep learning-based algorithm for 2-D cell segmentation in microscopy images," *BMC Bioinformatics*, vol. 19, no. 1, p. 365, Oct. 2018. [Online]. Available: https://doi.org/10.1186/s12859-018-2375-z

[3] F. Lux and P. Matula, "DIC Image Segmentation of Dense Cell Populations by Combining Deep Learning and Watershed," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 236–239, iSSN: 1945-8452.

[4] F. A. G. Peña, P. D. M. Fernandez, P. T. Tarr, T. I. Ren, E. M. Meyerowitz, and A. Cunha, "J Regularization Improves Imbalanced Multiclass Segmentation," *arXiv:1910.09783 [cs]*, Oct. 2019, arXiv: 1910.09783. [Online]. Available: http://arxiv.org/abs/1910.09783

[5] "Cell Segmentation Benchmark – Cell Tracking Challenge." [Online]. Available: http://celltrackingchallenge.net/latest-csb-results/

[6] "Cell Tracking Challenge." [Online]. Available: http://celltrackingchallenge.net/

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *arXiv:1708.02002 [cs]*, Feb. 2018, arXiv: 1708.02002. [Online]. Available: http://arxiv.org/abs/1708.02002

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *arXiv:1703.06870 [cs]*, Jan. 2018, arXiv: 1703.06870. [Online]. Available: http://arxiv.org/abs/1703.06870

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *arXiv:1502.01852 [cs]*, Feb. 2015, arXiv: 1502.01852. [Online]. Available: http://arxiv.org/abs/1502.01852

[10] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," vol. 9, pp. 249–256, 2010.

[11] "Evaluation Methodology – Cell Tracking Challenge." [Online]. Available: http://celltrackingchallenge.net/evaluation-methodology/