A paper

DIC IMAGE SEGMENTATION OF DENSE CELL POPULATIONS BY COMBINING DEEP LEARNING AND WATERSHED

accepted for publication in proccedings of IEEE International Symposium of Biomedical Imaging (ISBI) 2019.

DIC IMAGE SEGMENTATION OF DENSE CELL POPULATIONS BY COMBINING DEEP LEARNING AND WATERSHED

Filip Lux, Petr Matula

Masaryk University, Faculty of Informatics
Centre for Biomedical Image Analysis
Botanická 68A, 602 00 Brno

ABSTRACT

Image segmentation of dense cell populations acquired using label-free optical microscopy techniques is a challenging problem. In this paper, we propose a novel approach based on a combination of deep learning and the watershed transform to segment differential interference contrast (DIC) images with high accuracy. The main idea of our approach is to train a convolutional neural network to detect both cellular markers and cellular areas and, based on these predictions, to split the individual cells using the watershed transform. The approach was developed based on the images of dense HeLa cell populations included in the Cell Tracking Challenge database. Our approach was ranked the best in terms of segmentation, detection, as well as overall performance as evaluated on the challenge datasets.

Index Terms— Image Segmentation, Differential Interface Contrast, Convolutional Neural Networks, Watershed

1. INTRODUCTION

Cell segmentation, understood as the delineation of the cell boundary of individual cells in microscopy images, is a challenging problem in particular for dense cell populations. This problem is even more difficult when the images are obtained by label-free optical microscopy techniques such as DIC and phase contrast microscopy. Because of the absence of specific labeling, the visual properties of the cell boundaries, as well as the cellular regions, differ significantly within the image domain; it is sometimes difficult for even an expert to decide on accurate boundary positions without understanding the context. In terms of the Cell Tracking Challenge (CTC) [1] the segmentation accuracy of automatic algorithms was significantly lower on the DIC dataset (DIC-C2DH-Hela) than on datasets captured by different imaging techniques.

The work was founded by the Czech Science Foundation, project no. GA17-05048S. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated. We gratefully acknowledge the support of the NVIDIA Corporation and their donation of the Quadro P6000 GPU used for this research.

The first results on the DIC dataset, with segmentation accuracies greater than 77%, were obtained by Olaf Ronnenberger and his team by introducing his convolutional neural network architecture: U-Net [2]. In the context of the challenge deep learning has remained the most successful approach to segmenting these images until today, with the submissions by Sixta et al. (CVUT-CZ), Arbelle et al. (BGU-IL$^{(4)}$) [3] and Payer et al. (TUG-AT) [4].

Inspired by the challenging nature of the DIC dataset, we developed our own segmentation approach based on deep learning, which for the given CTC competition dataset achieves the best results in all three evaluation measures in the Cell Segmentation Benchmark (CSB)[1]: detection accuracy (DET), segmentation accuracy (SEG), and overall performance (OP_{CSB}) all defined at the challenge website[2]. The novelty of our approach lies in the combination of a convolutional neural network, which is trained for both detecting cellular markers and cellular areas, with the watershed transform to split the individual cells. In this paper, we describe our approach as well as the obtained results in details.

2. METHOD DESCRIPTION

The core of our method is made by a single convolutional neural network, which is trained to solve two different tasks: cell detection and pixel-wise classification of the input image to the foreground and the background. The network outputs are processed by mathematical morphology operations, and in particular, the watershed transform is applied to obtain the final image segmentation. During the training, we used a weighted MSE loss function to ensure that the prediction of specific image parts is accurate. The whole process is demonstrated in Fig. 1 and detailed below.

2.1. Dataset description

The DIC-C2DH-HeLa dataset consists of four time-lapse video sequences which sizes are shown in the upper part of

[1] A spin-off time-lapse segmentation benchmark of CTC.
[2] http://celltrackingchallenge.net/evaluation-methodology/

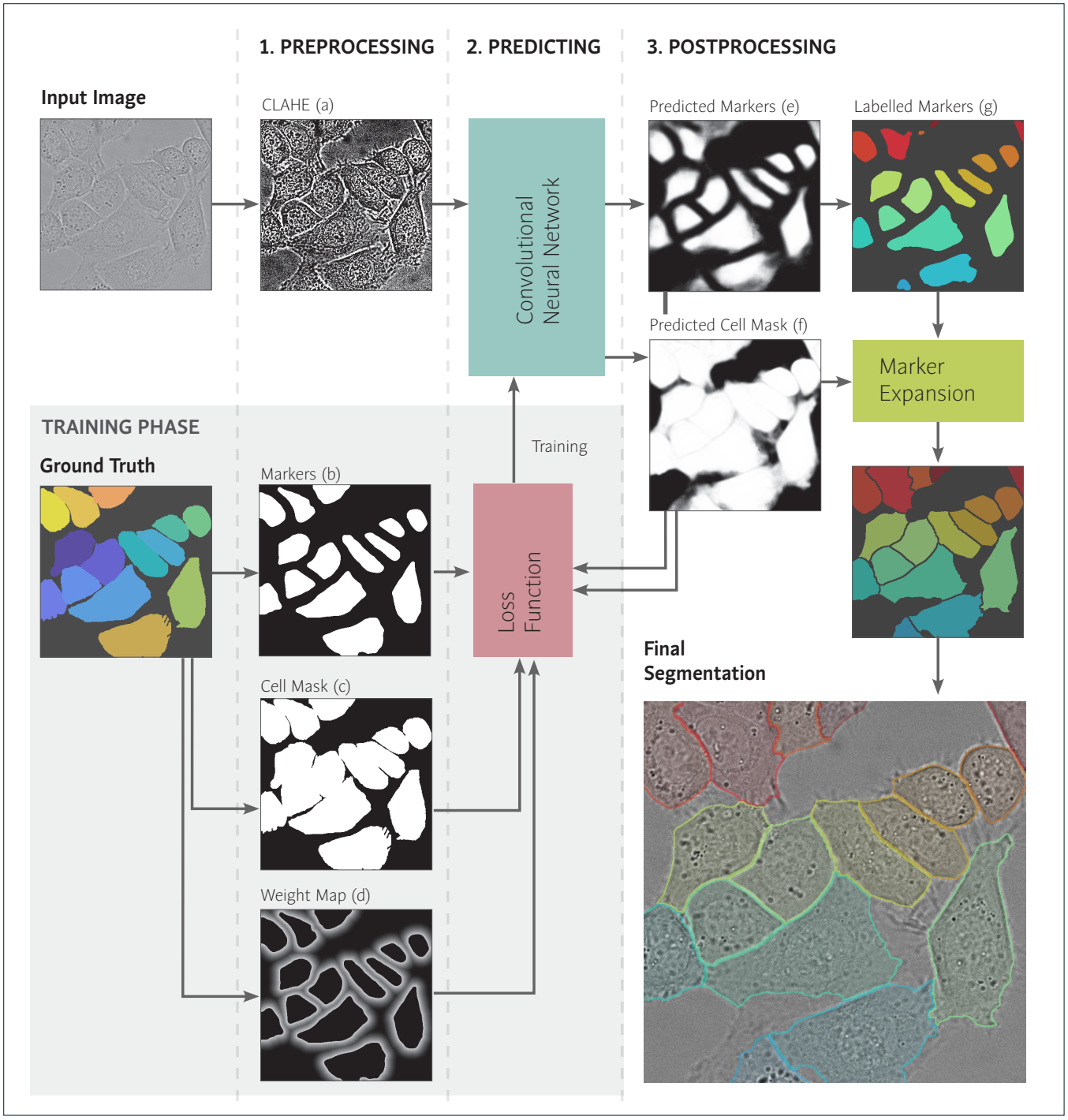

Fig. 1: **Schema of the segmentation process.** The method consists of three steps: **1. Data Preprocessing**, where an input image is normalized by Contrast limited adaptive histogram equalization (CLAHE): (a) in the training procedure, we extract three images from the given ground-truth image: Markers (b), the Cell Mask (c), and the Weight Map (d). **2. Predicting** using a convolutional neural network. The network produces Predicted Markers (e) and the Predicted Cell Mask (f). These images are used in the computation of the loss function during the network training phase. **3 Image Postprocessing**. The Predicted markers (e) are filtered, labeled (g) and expanded by the watershed transform to the Final Segmentation.

Dataset	Train	Valid.	Test	
Sequence	S1	S2	S3	S4
Frames	84	84	115	115
Detection GT	84	84	-	-
Segmentation GT	9	9	-	-
Manual seg.	84	-	-	-

Table 1: DIC-C2DH-HeLa dataset size (in frames)

Table 1. We set the whole sequence $S1$ as a training dataset and the sequence $S2$ as a validation dataset. There was an insufficient number of segmentation ground truth images for neural network training available in the first sequence; therefore, we manually segmented the remainder of the $S1$ to extend the training dataset (bottom line of Table 1).

Except for additional annotations, we enlarged the set of available training data by data augmentation to improve the network generalization ability. We used *Random Distortion* [5] to simulate different variations in cell shapes. Because cells have no fixed orientation, we also used geometrical transformations such as *rotation*, *mirroring* and *scaling*. Augmentation was applied on the fly right before the training. For all these operations, the data outside the image domain were mirrored.

2.2. Pre-Processing

Before the neural network training, each training sample is transformed into four different grayscale images: (a) Normalized Input, (b) Markers, (c) Cell mask, and (d) Weight map; see Fig. 1.

(a) Each input image is normalized by *Contrast limited adaptive histogram equalization* (CLAHE) [6]. It enables us to process images with different illumination levels. (b) A marker of each cell is defined as a cell GT mask eroded by a disc structuring element of radius 24 pixels. (c) The Cell Mask distinguishes cells and the background. (d) The Weight Map W defines the importance of each pixel for a good prediction; this map is defined by a map of real values greater than one, where a higher value means higher pixel importance. We assign a high weight to the areas nearby the markers. Let Φ be a set of all markers ϕ in the given image. The weight $w_i \in W$ of pixel p_i is computed by the formula:

$$w_i = 1 + a \sum_{\phi \in \Phi} \max(b - ||p_i, \phi||, 0)^2, \qquad (1)$$

where $||p_i, \phi||$ denotes the distance between pixel p_i and marker ϕ, i.e., the minimal Euclidean distance between p_i and any pixel in ϕ. The parameter $b \in \mathbb{R}^+$ relates to the width of area around each marker with the higher weight. By setting the parameter $a \in \mathbb{R}^+$ we can can regulate the map magnitude. In our experiments, we set a to 0.004 and b to 40.

2.3. Predicting

For the prediction of cell markers and masks, we used a convolutional neural network under the U-Net topology [2]. In contrast to the original U-Net network, our network produces two different predictions; the first defines the location of the cells markers, and the second distinguishes input image pixels to foreground and background. The network is trained to produce these two outputs simultaneously. During the training, network outputs are weighted by the same weight map W. As the loss function, we used the *Weighted Mean Square Error* inspired by the original U-Net network. Let $m_i, c_i \in \{0, 1\}$ denote the expected values of pixel p_i given by the markers and cell mask, and $\widehat{m}_i, \widehat{c}_i \in \langle 0, 1 \rangle$ denote the predicted values. The loss function is given by the following formula:

$$L = \frac{1}{2} \frac{\sum_i w_i((\widehat{m}_i - m_i)^2 + (\widehat{c}_i - c_i)^2)}{\sum_i w_i}, \qquad (2)$$

where w_i is the weight of the pixel i. We normalize the result by the sum of all weights. The range of L is from 0 to 1.

The optimization method that we used to find proper network weights is called *Adam* [7]. This method has only three hyperparameters for performance tuning. The most important hyperparameter is the learning rate, which we in our experiments set to 0.001.

2.4. Post-Processing

The neural network is not predicting the final segmentation directly, and we apply additional post-processing steps. The network outputs are two grayscale images with values in the range from 0 to 1. The first image represents the predicted cell markers (e), and the second image is the predicted cell mask (f). The predicted marker image is thresholded to a binary image, and then a morphological opening operation using a disk structuring element of diameter 12 pixels is applied to remove small objects and weak connections between markers. Next, every connected component is labeled (g). To get final segmentation, we combined the labeled markers (g) with Predicted Cell Mask (f) by a marker-controlled watershed transform [8]. All the pixels predicted as a background are labeled as a background also in the final segmentation. An example of the result is shown in Figure 1.

3. EXPERIMENTS

3.1. Network training

We trained the neural network from scratch for 600 epochs with 25 mini-batches of size 8 samples in one epoch. The training process ran on a machine with the GPU NVIDIA Quadro P6000, and one epoch took approximately 18 seconds. We also stored the model at every tenth epoch to pick the model with the best performance.

3.2. Evaluation

In our method, we use a different definition of the weight map compared to the original U-Net. To examine an effect of this decision, we trained three models using different setting of the weight map: the first one (M_{w}) uses our weight map W, second one ($M_{\text{unet\_w}}$) uses the original U-Net weight map, and as a baseline we trained model ($M_{\text{no\_w}}$) with no pixel weighting. All other settings were the same for all these experiments. The performance of each model was measured after every tenth epoch on a validation sequence $S2$ by SEG and DET measures. The results are shown in Figure 2. Model M_{w} performs best in DET measure in the majority of training steps. In SEG measure the performance of the models M_{w} and $M_{\text{unet\_w}}$ is similar.

As the best performing model, we picked model M_{w} trained for 290 epochs, which performance equals to 0.8511 in the SEG measure and 0.9502 in the DET measure. All these results we achieved on the validation data.

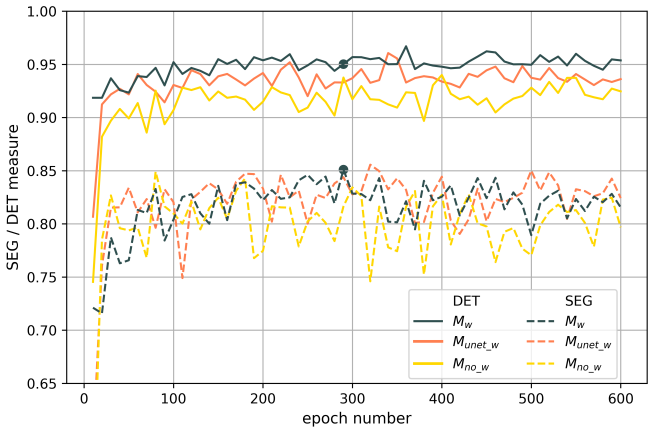

Fig. 2: Model comparision during the training

Our algorithm was submitted and officially evaluated in the CSB to obtain a fair comparison with other DIC image segmentation methods on the testing data. For DIC-C2DH-HeLa challenge dataset, our method achieved the best results in all three competition measures. The results of the four best competition participants are listed in Table 2.

participant	DIC-C2DH-HeLa SEG	DET	OP_{CSB}
MU-Lu-CZ	**0.843**	**0.944**	**0.894**
TUG-AT	0.814	0.918	0.866
BGU-IL[(4)]	0.793	-	0.839
CVUT-CZ	0.792	0.906	0.849

Table 2: Cell Segmentation Benchmark results for DIC-C2DH-HeLa challenge dataset

4. DISCUSSION AND FUTURE WORK

Our segmentation approach combines methods from the fields of deep learning and mathematical morphology to obtain more accurate image segmentation results. It exploits the ability of neural networks to detect objects in the image, and it uses deterministic marker expansion algorithms to find object shapes. Because of the data augmentation, only tens of labeled GT images were sufficient for the network training.

The method outperformed other participants in the international cell segmentation challenge on a difficult dataset. In the future, we will focus on generalizing the method also for another biomedical image data types.

5. REFERENCES

[1] V. Ulman, M. Maška, K. Magnusson, O. Ronneberger, C. Haubold, N. Harder, Pa. Matula, Pe. Matula, D. Svoboda, M. Radojevic, I. Smal, K. Rohr, J. Jaldén, H. Blau, O. Dzyubachyk, B. Lelieveldt, P. Xiao, Y. Li, S. Cho, and C. Ortiz-de Solorzano, "An objective comparison of cell tracking algorithms," *Nature Methods*, vol. 14, Oct. 2017.

[2] 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, vol. 9351 of *LNCS*, pp. 234–241, Springer.

[3] A. Arbelle and T. R. Raviv, "Microscopy cell segmentation via convolutional LSTM networks," 2018, Accepted to ISBI 2019, (available on arXiv:1805.11247v2 [cs.CV]).

[4] Ch. Payer, D. Štern, T. Neff, H. Bischof, and M. Urschler, "Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 3–11. Springer International Publishing, 2018.

[5] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, Washington, DC, USA, 2003, ICDAR '03, pp. 958–, IEEE Computer Society.

[6] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics Gems*, pp. 474–485, 1994.

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[8] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communication and Image Representation*, vol. 1, no. 1, pp. 21–46, sep 1990.