NUCLEI INSTANCE SEGMENTATION WITH DUAL CONTOUR-ENHANCED ADVERSARIAL NETWORK

Donghao Zhang[⋆]*, Yang Song*[⋆]*, Siqi Liu*[⋆]*, Dagan Feng*[⋆]*, Yue Wang*[†]*, Weidong Cai*[⋆]

[⋆] School of Information Technologies, University of Sydney, Australia
[†] Bradley Department of Electrical and Computer Engineering,
Virginia Polytechnic Institute and State University, USA

ABSTRACT

The morphology of cancer cells is widely used by pathologists to grade stages of cancers. Accurate cancer cell segmentation is significant to obtain quantitative diagnosis. We proposed a dual contour-enhanced adversarial network to solve this challenge. The dual contour-enhanced masks and adversarial network are incorporated to improve individual cell segmentation capability. By evaluating quantitative individual cell segmentation results on 2017 MICCAI Digital Pathology Challenge, our method achieved best balance between precision and recall rate of individual cell segmentation compared to state-of-the-art cell segmentation methods.

Index Terms— Pathology Image Analysis, Generative Adversarial Network, Nuclei Segmentation

1. INTRODUCTION

Grading the stage of cancer is of significance to provide efficient treatments to patients. The tumor diagnosis is traditionally obtained by manually labelling the nuclei segmentation. Nowadays, the microscopy is capable to produce the whole-slide image. The manual morphology annotation of the cancer cells is labour-intensive and not scalable to the whole slide image because of the size. Therefore, developing an accurate and reliable cancer cell segmentation method is highly demanded. However, this task is still challenging due to the following reasons: (1) cancer cell borders are close to each other with tiny gaps; (2) there is uneven distribution of staining process using Haematoxylin and Eosin.

Cell segmentation from pathology images usually applies morphology based [1, 2] or patch-based pixel level classification [3] methods. With patch-level processing, the pre-trained convolutional neural network (CNN) is fine-tuned or other hand-crafted features are used to determine the class of each pixel. More recently, U-shaped neural network (U-Net) [4] is proposed, and it often outperforms the previous sliding-window based CNN models. Compared with patch-based segmentation techniques, U-Net is capable of learning from less data and is particularly useful for the challenging bio-medical segmentation tasks that usually do not have sufficient high-quality data. In addition, another new attempt of cell segmentation is to combine CNN with graph-based method [5]. Besides, deep contour-aware network [6] and gland instance deep multichannel networks [7] combine boundary and texture information into the designed network architecture.

Besides the network architectures [4, 6, 3] used for pathology images, many new neural network architectures for the semantic segmentation [8, 9, 10, 11] have also been proposed. Fully Convolutional Networks (FCNs) [8] introduce upsampling of feature maps using deconvolution and achieve breakthrough compared to previous semantic segmentation approaches using the hand-crafted features such as scale-invariant feature transform (SIFT). After the pioneering work of FCNs, many image-to-image network architecture for the semantic segmentation such as SegNet [9], Linknet [10] and pyramid scene parsing network (PSPNet) [11] were proposed. PSPNet introduces dilated convolutions, auxiliary loss and Spatial Pyramid Pooling to the ResNet [12]. The major contribution of LinkNet is to use convolution kernel with the size 1×1 and channel reduction scheme in the decoder block, which makes it highly efficient.

Generative adversarial networks (GANs) [13] are traditionally used to estimate the generative models by the adversarial networks which are the generator and discriminator. By introducing image-to-image translation, Pix2Pix [14] showed that the GAN loss can improve the blurred quality of image-to-image translation generated by U-Net [4]. Inspired by these existing studies, in this work, we incorporate the GAN loss to enhance the cell segmentation performance. We also propose to generate contour-enhanced ground truth to limit individual cell segmentation interference from cell boundary touching each other. Based on this, a dual contour-enhanced networks (DCANet) are designed to fully utilize two types of ground truth provided by the 2017 MICCAI Digital Pathology Challenge dataset and speedup the optimization process by sharing weights of Convolution layers. Our overall cell segmentation framework consists of data augmentation, adversarial network training and post-processing. Our evaluation results show that our method outperformed state-

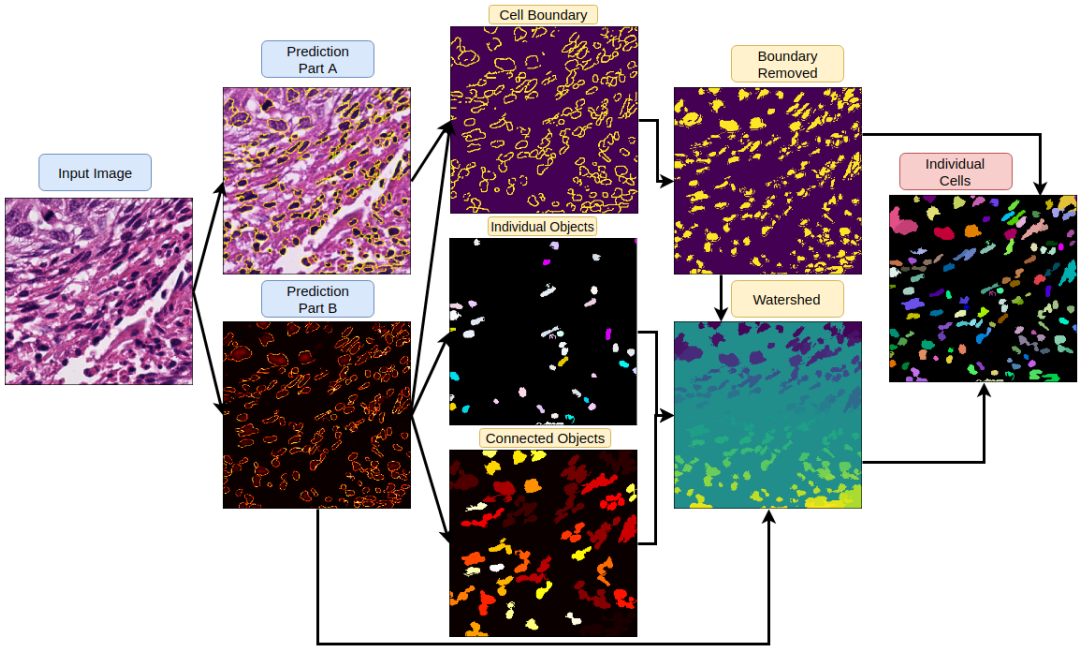

Fig. 1. An example of the overall framework from prediction results generated by the dual contour-enhanced adversarial network to the individual cells.

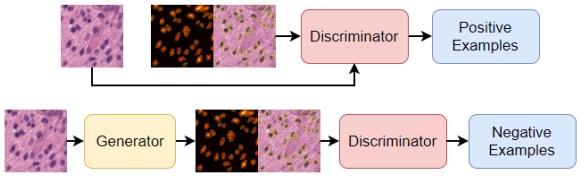

Fig. 2. Training of the dual contour-enhanced adversarial network.

of-the-art network architectures on the Challenge dataset.

2. METHODS

2.1. Overall Framework

In our framework, the generator of adversarial networks maps the test image to learned outputs including cell image masked with yellow boundary (Prediction Part A) and boundary distance-transformed image (Prediction Part B) shown in Fig. 1. Then cell boundary image is obtained by the combination of boundary information from two dual contour-enhanced outputs. Also, by setting an area threshold, small individual cell objects are extracted from Prediction Part B. Similarly, another image storing connected objects is generated by removing small individual cell objects from binary segmentation of Prediction Part B. Next, watershed image [15] finds watershed basins using the boundary removed image as the basin masked image and Prediction Part B as its input image. The final individual cells image is obtained by fusing boundary removed image and watershed image.

2.2. Dual Contour-Enhanced Mask Generation

The dual contour information is designed to improve the capability of separating clustered cell objects and fully utilize two types of ground truths. The distance-transformed map is better than the binary map by forcing networks to learn more morphology information of cancer cells. The distance-transformed map is robust to small noises. There are two types of ground truths originally provided by the Digital Pathology dataset: (1) cell image with boundary masked as yellow; (2) a masked image tile whose array element stores a value between 0 and N representing the individual nuclei object. The proposed boundary distance transform and boundary highlighting are applied to Type II ground truth inspired by boundary-aware instance segmentation [16]. Unlike boundary-aware instance segmentation, the pixels on cell boundary are also highlighted to further separate close cells. The proposed distance transform of every pixel p is defined as:

$$D(p) = \begin{cases} \min\lceil d(p,q) \rceil & \text{for } p \in \text{pixels inside Q} \\ 255 & \text{for } p \in \text{pixels on Q} \\ 0 & \text{for } p \in \text{pixels outside Q} \end{cases} \quad (1)$$

where Q denotes the pixel set on the individual cell object boundary, $d(p,q)$ represents the Euclidean distance between p and q and $\lceil . \rceil$ is nearest integer operation. The dual-contour enhanced masks are defined as contour 1 provided by Type

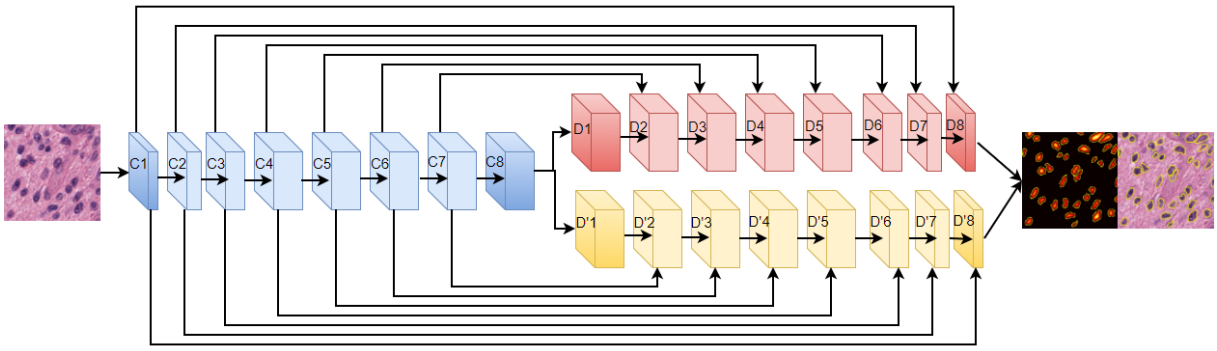

Fig. 3. The network architecture of generator with blue blocks representing convolution layers and red and yellow blocks denoting deconvolution layers.

I ground truth and contour 2 generated by the proposed distance transform method.

2.3. Network Training and Architectures

The proposed dual U-shaped adversarial network as shown in Fig. 2 is designed to map the input image to the learned dual contour-enhanced outputs. Two deconvolution block series correspond to different functions. To be more specific, boundary distance transformed predictions provide the base image for watershed algorithm and boundary highlighted cell images enhance boundary information for separating cells close to each other. The accelerated optimization process is achieved by sharing the weights of Convolution layers. The generator and discriminator are trained iteratively to optimize the objective of adversarial network [14] by combining the GAN loss and $L1$ loss in the following equation:

$$G^* = arg \min_G \max_D \mathcal{L}_{cGAN} + \lambda \mathcal{L}_{L1}(G) \qquad (2)$$

where $cGAN$, $L1$, G and D represent conditional GAN, $L1$ loss, generator and discriminator.

The generator architecture of DCANet is shown in Fig. 3 with 8 encoder and 16 decoder layers respectively. $C2 \ldots C7$ represent an encoder layer including convolution (Conv) and instance batch normalisation (InstBatchNorm) layers with $LeakyReLU$. $C1$ and $C8$ are Conv without InstBatchNorm. $D2 \ldots D7$ represent deconvolution (Deconv), InstBatchNorm and dropout layers with $ReLU$. $C1$ and $C8$ denote Deconv with $Tanh$. $D1$ and $D8$ represent Deconv and BatchNorm with $ReLU$. The discriminator architecture of first 6 layers is the same as the encoder of G and final layer maps output to one dimensional value. In addition, the cell images and dual contour-enhanced masks are defined as positive examples. Similarly, the negative examples are defined as the cell images and prediction results generated by the generator.

3. EXPERIMENTAL RESULTS

3.1. Dataset

2017 MICCAI Digital Pathology Challenge dataset includes four types of cancer cells from non small cell lung cancer (NSCLC), head and neck squamous cell carcinoma (HNSCC), glioblastoma multiforme (GBM), and lower grade glioma (LGG) tumors. The image tiles are extracted from the whole slide image. The size of image tile is either 500×500 or 600×600. Nuceli of the cancer cell of image tile is manually labelled. The whole challenge dataset consists of 32 images with ground truth provided and 8 image tiles with no ground truth provided. 16 and 16 image tiles are chosen as training and testing dataset respectively. It is a known problem that biomedical related image processing tasks do not have enough labelled ground truth compared to semantic segmentation. Cropped images with 256×256 pixels were randomly sampled from original and transformed training images. Transformation operation includes rotation through $+45°$ to $-45°$, mirroring, dropout and Gaussian filtering. After the data augmentation, the number of cropped training images are 2464. For the testing purpose, the test image is rescaled into a fine size which can be processed by the Conv layers of network successfully such as 512×512.

3.2. Analysis of Cell Segmentation

The proposed method was quantitatively compared with state-of-the-art Pix2Pix [14], fast neural style (FnsNet) using residual block between Conv and Deconv layers [17], Linknet [10] and SegNet [9] methods using the precision, recall and F1-score metrics. For a fair comparison, the same data augmentation techniques were applied to all comparing methods. True Positive, False Positive, False Negative are defined as the total cell detection number with more than 50% of the overlapping area with ground truth objects, the total object number not in the ground truth but in the prediction results, the number of cell objects not being detected from ground truth respec-

tively. As shown in Table 1, our method ranked first in terms of average precision with 0.7022 and F1-score with 0.6969 among all compared methods. The average recall value of the proposed framework is still competitive ranked second with 0.7021 compared with 0.7245 achieved by the FnsNet.

Table 1. The quantitative individual cell detection results.

Network	Precision	Recall	F1
SegNet [9]	0.3251	0.6173	0.4064
LinkNet [10]	0.3431	0.5951	0.4161
Pix2Pix [14]	0.4681	0.6615	0.5467
FnsNet [17]	0.5855	0.7245	0.6240
Proposed	0.7022	0.7021	0.6969

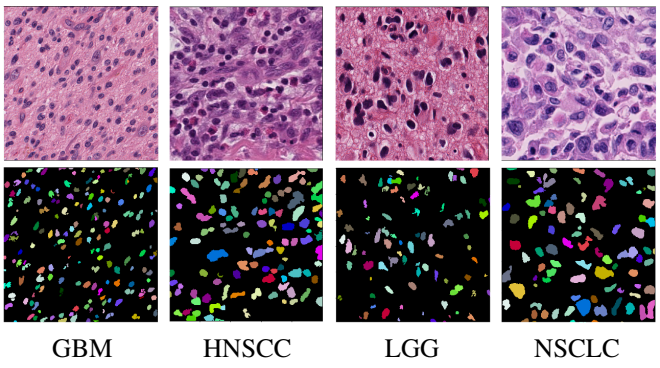

Fig. 4. Examples of pathology images and corresponding instance cell segmentation results by the proposed method.

Testing examples of four types of cancer cells are illustrated in Fig. 4. Observing from pathology image example, the individual cell segmentation is challenging because of the large amount of touching cells, morphology variations and incomplete cell on the boundary of image. Even when the shape of cancer cells are irregular, the majority of them are positively recognized. On the other hand, the proposed framework is more likely to make false detection when several challenging conditions presented at the same time. For example, some cells can be connected and the texture and boundary information are blurred at the same time.

4. CONCLUSION

In this paper, we presented a new dual contour-enhanced adversarial network framework to generate individual cell segmentation. By combining distance transformed boundary information with adversarial network, the proposed network outperformed state-of-the-art individual cell segmentation method in terms of F-scroes evaluated on the 2017 MICCAI Digital Pathology Challenge dataset.

5. REFERENCES

[1] A. Paul and D. P. Mukherjee, "Gland segmentation from histology images using informative morphological scale space," in *ICIP*, 2016, pp. 4121–4125.

[2] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *ISBI*, 2008, pp. 284–287.

[3] W. Li, S. Manivannan, S. Akbar, J. Zhang, E. Trucco, and S. J. McKenna, "Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks," in *ISBI*, 2016, pp. 1405–1408.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[5] L. Zhang, M. Sonka, L. Lu, R. M. Summers, and J. Yao, "Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei," in *ISBI*, 2017, pp. 406–409.

[6] H. Chen, X. Qi, L. Yu, and P. Heng, "Dcan: Deep contour-aware networks for accurate gland segmentation," in *CVPR*, 2016, pp. 2487–2496.

[7] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, and E. Chang, "Gland instance segmentation using deep multichannel neural networks," *IEEE Transactions on Biomedical Engineering*, 2017.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[9] 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*, 2017.

[10] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," *arXiv preprint arXiv:1707.03718*, 2017.

[11] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *CVPR*, 2017.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014, pp. 2672–2680.

[14] P. Isola, J. Y. Zhu, Tinghui Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017.

[15] P. Neubert and P. Protzel, "Compact watershed and preemptive slic: On improving trade-offs of superpixel segmentation algorithms," in *ICPR*, 2014, pp. 996–1001.

[16] Z. Hayder, X. He, and M. Salzmann, "Boundary-aware instance segmentation," in *CVPR*, 2017.

[17] J Johnson, A Alahi, and L Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016, pp. 694–711.