Segmentation Guided HEp-2 Cell Classification with Adversarial Networks

Hai Xie[1], Yejun He[1], Haijun Lei[2], Jong Yih Kuo[3], Baiying Lei[4*]
[1]Guangdong Engineering Research Center of Base Station Antennas and Propagation, Shenzhen Key Lab of Antennas and Propagation
College of Electronics and Information Engineering, Shenzhen University, 518060, China
[2]Guangdong Province Key Laboratory of Popular High-performance Computers
School of Computer and Software Engineering, Shenzhen University, 518060, China
[3]Computer Science and Information Engineering
National Taipei University of Technology, Taiwan, China
[4]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging
School of Biomedical Engineering, Health Science Center, Shenzhen University, 518060, China
Email: 2166130109@email.szu.edu.cn, heyejun@126.com, lhj@szu.edu.cn, jykuo@csie.ntut.edu.tw, leiby@szu.edu.cn*

***Abstract*— The analysis of indirect immunofluorescence (IIF) on human epithelial type 2 (HEp-2) cells is of paramount importance for the autoimmune diseases diagnosis. Essentially, accurate segmentation masks can generate rich boundary information, which is beneficial for improving the performance of the classification task. For this reason, this paper proposes a novel segmentation guided HEp-2 cell classification method via generative adversarial networks (GANs), which employs the GANs as the segmentor to generate accurate masks for the subsequent classification task. Specifically, the proposed network architecture consists of three modules (i.e., the generator, discriminator and classifier). The first two modules constitute GANs model, which is trained to obtain better segmentation results via playing a min-max game. The segmentation masks and the corresponding original images are fed to the third module together to identify the category of the trained cell. Furthermore, the Xception and ResNet-50 model are used as the backbone of the segmentation and classification network, respectively. Besides, an improved classification loss function via Gaussian Mixture (GM) is proposed to optimize the classification network. The proposed architecture can learn rich boundary information and well represent the class label of HEp-2 cell images. Experimental results on the HEp-2 International Conference on Pattern Recognition (ICPR) 2016 task1 dataset demonstrate our proposed model achieves quite promising performance.**

***Index Terms*—HEp-2 cell, GANs, Segmentation, Gaussian Mixture, Classification.**

I. INTRODUCTION

The analysis of Indirect immune florescence (IIF) with human epithelial type-2 (HEp-2) has been widely used for the detection of anti-nuclear antibody (ANA) test for the diagnosis of the systemic autoimmune diseases [1]. Human experts usually discriminate the ANA by observing the intensity of the florescence (intermediate or positive). However, the large intra-class variations are caused by the inhomogeneous illumination of HEp-2 cell staining pattern, which makes the classification task quite challenging. To overcome this challenge, many algorithms were proposed for HEp-2 cell classification in the past decade. However, these methods mainly extract hand-craft features, which are limited to represent the characteristics of HEp-2 cell images [2-4].

To extract more discriminative features for HEp-2 cell images segmentation, classification and other tasks, another trend is witnessed by the deep convolutional neural network (DCNN) due to its promising performance [5-7]. These researchers paid their efforts to leveraging DCNN model to segment or classify HEp-2 cells. For instance, Lei *et al.* proposed a framework based on very deep supervised residual network and cross-modal transfer learning strategy is adapted to classify HEp-2 cell images [6]. However, these traditional DCNNs still belong to straight network architectures, which may omit the important information in the shallow layers. To improve the situation, many complex and multi-branches networks have been developed in recent years, such as Deeplab [8] and Inception [9]. These networks outperform those straightforward networks in terms of the ability of representation for image features, hence, they attracted the attention of many researchers. In addition, the recent developed generative adversarial networks (GANs) have superior performance on various vision tasks such as segmentation and classification. For example, Han *et al.* proposed Spine-GAN to perform automated segmentation and classification tasks for multiple spinal structures in MRIs [10]. Since GANs can improve the model to obtain more discriminative features, many researchers employ GANs to conduct segmentation, classification and other tasks [10-12]. Therefore, we propose a novel hybrid framework via adversarial learning for joint HEp-2 segmentation and classification tasks.

Since the segmentation mask can provide rich boundary information to boost the HEp-2 classification performance, we devise a novel segmentation guided HEp-2 cell classification method via GANs. Our proposed framework consists of three modules, i.e., generator, discriminator, classifier. The generator is based on Deeplabv3 plus, which has attractive segmentation ability by employing the Xception structure as the backbone [8]. The generator can encode multi-scale contextual information by using atrous convolution to obtain more fine-gained spatial information along object boundaries. The discriminator consists of some simple yet effective convolution layers and batch normalization layers, which is used to decide whether the generated image is real or fake. Finally, we select ResNet-50 [13] as the classifier to take full advantage of the global and local information to extract the discriminative features by residual connection. Besides, an improved classification loss function based on GM [14] is proposed to optimize the classification network. To our best knowledge, this is the first study that the segmentation mask of HEp-2 cell is directly fed to the classification network to extract richer feature information. We evaluate our method on the public International Conference on Pattern Recognition (ICPR) 2016 task1 dataset. Experimental results demonstrate that our proposed model is more discriminative and effective

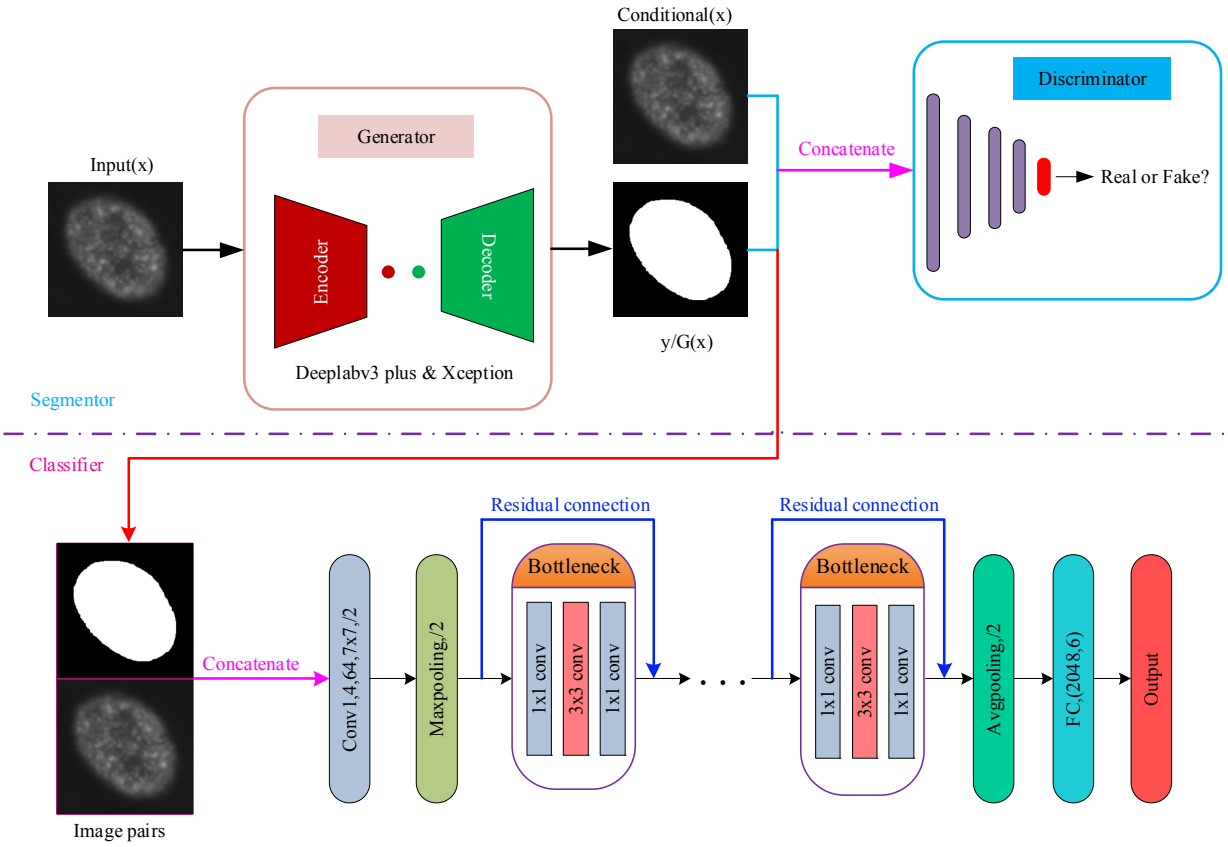

Fig.1. Architecture of the proposed model. The upper part is a segmentor, which is comprised by a GAN. The generator employs Deeplabv3 plus and Xception as the encoder-decoder module and backbone network, respectively. The training processing of GANs is similar to Pix2Pix model. The original images and the corresponding generated images are concatenated as the paired images, which are directly fed into the ResNet-50 network for the classification task.

by fusing the segmentation features and the corresponding original images. Also, our experiments verify that segmentation masks can provide richer boundary information for classification task, especially blurred images. Fig. 1 shows our proposed framework. Overall, our main contributions of this paper are as below:

1) We design a novel segmentation guided HEp-2 cell classification method by GANs, which combines the GANs and ResNet-50 model by utilizing the generation masks to boost the classification performance.
2) An improved classification loss function via GM is proposed to optimize the classification network.

II. Methods

The proposed framework in this paper is shown in Fig. 1. We describe it in detail as follows.

A. Architecture of the generative adversarial networks

To gain robust feature representations for HEp-2 cell classification and further disease treatment, we conduct the segmentation experiment as the primary step. The previous works show that GANs can achieve good segmentation performance by representing object boundary in a better way [10, 11]. Therefore, the GANs is selected as our segmentation model in this paper.

We define the generator and the discriminator as G and D, respectively. Supposing that x denotes the input image, and the corresponding ground-truth and generated mask are denoted as y and $G(x)$, respectively. The processing flow of the conditional GANs in this paper is similar to pix2pix model, where the generator G generates the segmented mask $G(x)$ from x and the generated mask and ground-truth are concatenated with original image respectively that are fed into discriminator to decide if the generated mask is a real mask or a synthesized mask from $G(x)$. The objective function can be defined as

$$\min_G \max_D L_{GAN} = \mathrm{E}_{x,y\sim p_{data}(x,y)}[logD(x,y)] + \mathrm{E}_{x\sim p_{data}(x)}[1 - logD(x,G(x))], \quad (1)$$

where the generator G tries to minimize this function while the discriminator D makes efforts to maximize it. Furthermore, inspired by pix2pix [15], the L_1 regularization term is utilized to give the constraint of the generated masks for the generator. The L_1 distance loss function can be indicated as

$$L_{L_1}(G) = \mathrm{E}_{x,y}[\|y - G(x)\|_1], \quad (2)$$

By integrating the above loss functions, the final objective function can be represented as

$$L_{total} = \alpha \min_G \max_D L_{GAN} + \beta L_{L1}(G), \quad (3)$$

where α and β is set as 0.5 and 10, respectively, which are empirical values. Specially, y is used as an extra condition

that is concatenated with x, which is fed into discriminator, as is shown in Eq. (1). Thus, the generator can better understand the purpose of segmentation task and generate what we want. Different from typical GANs that generates images from a random noise, our proposed structure generates image $G(x)$ by the given image x as the condition. In fact, a random noise is unable to accomplish the image-to-image task, e.g., from cell images to binary masks

B. Design of generator

Recently, the spatial pyramid pooling and encoder-decoder model are studied extensively for semantic segmentation task. The reason is that they can encode multi-scale contextual information and recover the same resolution as the input images by decoder. Therefore, we employ the Deeplabv3 plus based on Xception structure as the generator to conduct an automatic segmentation task. Similar to [8], we introduce the concept of Atrous convolution. Given each location i on the output feature map y', the weight and bias of each node k are defined as w and b, respectively. The general convolution in the input feature map x' can be denoted as:

$$y'[i] = f\left(\sum_k x'[i + r \cdot k] w(k) + b(k)\right), \quad (4)$$

where $f(\cdot)$ represents non-linear mapping of ReLU and batch normalization (BN) [16]. The r controls the stride that we sample the feature map.

For the decoder, we set two up-sample operations. Firstly, the encoder features are up-sampled in a bilinear mode by a factor of 4 and then concatenated with the low level features with the uniform resolution from the backbone network, i.e., Xception. Secondly, a 3×3 convolution is applied to refine the extracted features and followed by another up-sampling operation with a factor of 4 in a bilinear mode. In addition, we replace the plain convolution with the depth-wise separable convolution, which is a plain convolution with the kernel size of 3×3 followed by a pointwise convolution (the kernel of filters is 1×1). As a result, this architecture can decrease computational complexity and make the arbitrary resolution of output feature maps possible by using depth-wise separable convolution.

C. Design of classifier

As shown in Fig.1, a modified ResNet-50 network is used as the classifier. Similarly, we utilize the pre-trained parameters of ResNet-50 on the ImageNet dataset. Due to the change of the number of input channel, we modify the first convolution layer of ResNet-50. Meanwhile, we divide the output layer into the output layer and feature extracted layer so that the extracted feature can be computed to supervise the model and optimize the loss function. For the output layer, we define the output value as l_n, the true label value is $o_n \in [1, K]$, K represents the total categories of the whole dataset. The main classification function can be computed by the cross-entropy as follows:

$$L_{out} = -\frac{1}{N}\sum_{n=1}^{N} p(o_n) \log\left(q(l_n)\right), \quad (5)$$

where $p(\cdot)$ and $q(\cdot)$ represents the true label and the probability distribution of the output value, respectively, N indicates the total number of the sample. Inspired by [14], we propose an improved loss function to optimize the classification model. Given a feature f_n , we assume that μ_n and σ_n define the mean and covariance of class m for the extracted feature. The auxiliary classification loss function is denoted as:

$$L_{cla} = -\frac{1}{N}\sum_{n=1}^{N} log \frac{\theta(f_n; \mu_{y_n}, \sigma_{y_n}) p(l_n)}{\sum_{m=1}^{K} \theta(f_n; \mu_m, \sigma_m) p(m)}, \quad (6)$$

where $p(m)$ is the prior probability of class m and $\theta(\cdot)$ means the Gaussian distribution. Besides, the concept of the likelihood regularization is introduced to optimize the classification loss function. As a result, the final loss function is defined as:

$$L_{classi} = L_{out} + L_{cla} + \gamma\left(-\sum_{n=1}^{N} \log\left(f_n; \mu_{y_n}, \sigma_{y_n}\right)\right) \quad (7)$$

where γ is a positive weight value and the last part is the likelihood regularization.

III. Experiments And Results

A. Experimental setup and database

Our experiments are conducted in Pytorch framework and performed on two NVIDIA TITAN XP GPUs. For segmentation and classification tasks, we set different optimizers, e.g., we choose stochastic gradient decent with the learning rate = 0.0001, momentum = 0.9, weight decay = 1e-4 for segmentation network and Adam with the betas = (0.5, 0.999), eps = 1e-8 for classification network. Besides, the learning rate of classification model is initialized as 0.001 and multiplied by 0.1 every 50 epochs. The batch size is set as 16 and the maximum epoch is 200. The public International Conference on Pattern Recognition (ICPR) 2016 task1 dataset is chosen as our experimental dataset. As the test data of this dataset is not public, we randomly divide the dataset into training set and test dataset according to the ratio of 9:1. Thus, we obtain 12336 images for training and 1360 images for test. In addition, in order to enhance the variability of the dataset, we preprocess the data through random horizontal and vertical flips without augmenting the dataset size in the stage of training.

B. Evaluation metrics

As for the evaluation criteria, we divide them into two parts. On the one hand, we evaluate the segmentation results on test dataset via Dice coefficient and Jaccard index. On the other hand, we assess the classification performance on the test dataset in terms of Accuracy, Precision, Recall and F1-score. The Dice is a similarity metric obtained by comparing the prediction results of our model and the ground truth. The JA measures the overlap between the predicted results and ground truth and is expressed as their intersection over union. The F1-score is the same form as the Dice of segmentation metric. However, the difference of the metrics between segmentation and classification is that segmentation is evaluated on the pixels with probabilities of prediction and ground-truth and the classification is calculated on the prediction labels and the true labels. Specifically, the discriminator is not used in the process of testing. Only the generator and classifier are utilized to assess the segmentation and classification performance, respectively.

C. Experimental results

1) The results of segmentation

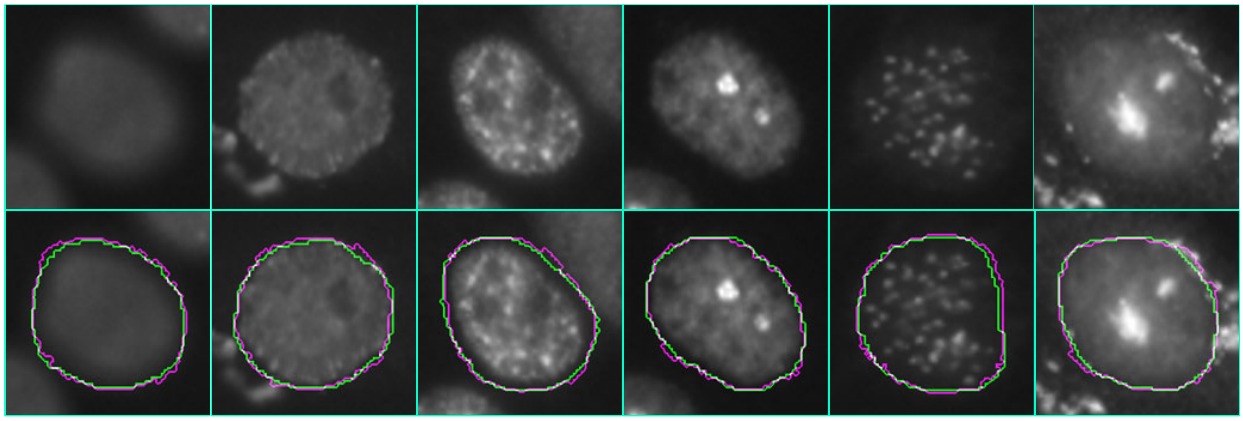

Fig.2. Segmentation results on ICPR 2016 dataset. From left to right, every column indicates the class of Homogeneous, Speckled, Nucleolar, Centromere, Nuclear membrane, Golgi, respectively.

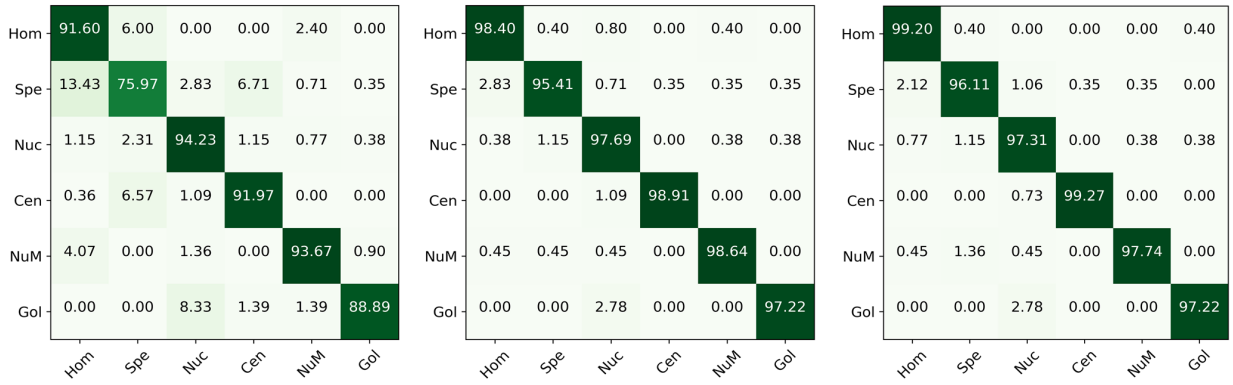

Fig. 3. The confusion matrix of different methods on six classes of the ICPR 2016 task1 dataset. From left to right, the confusion matrix of ResNet-50, ResNet-50-T, the proposed method. The Hom, Spe, Nuc, Cen, Num, Gol represent Homogeneous, Speckled, Nucleolar, Centromere, Nuclear membrane, Golgi, respectively.

Our segmentation results are shown in Fig.2. The first row is the original input images, the second row is the segmentation results, in which the purple lines represent the segmentation contour and the green lines indicate the ground truth contour. We can see that our segmentation result is almost similar to the original and ground truth images. Specially, our method is very effective for those cell images with some disturbed regions, such as the pattern of Speckled, Nucleolar and Golgi, as shown in Fig.2. In addition, we also quantify the segmentation results according to the segmentation metrics. We obtain the result that the mean Dice coefficient is 96.40% and the Jaccard index is 93.07%. These results show that our proposed segmentor with GANs is effective, which can help boost the classification performance as well.

2) The results of classification

The comparison results with other classification methods are shown in Table I. The ResNet-50-T (w/o GANs) represents the result of ResNet-50 model being pre-trained on ImageNet dataset and the segmentation masks are not used. The ResNet-50-T (w/o GM) indicates the result of segmentation guided HEp-2 cell classification without GM loss optimization. And the w/ indicates the corresponding module is used in the experiments. We can see that our proposed method achieves the best performance and the classification accuracy is 98.16%. Moreover, the Precision, Recall and F1-score are 98.24%, 98.06% and 98.15%, respectively, which demonstrates that the segmentation result and the optimized loss function can contribute to improving the classification performance. From Table I, we also display the comparison with the existing methods, which indicates our proposed method is better than the compared methods in terms of accuracy.

TABLE I. THE CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS ON THE PROPOSED TEST DATASET OF ICPR 2016 DATASET

Method	Accuracy	Precision	Recall	F1-score
Gao et al. [5]	96.76	NA	NA	NA
Prasath et al. [4]	94.26	NA	NA	NA
VGG-16	83.09	85.13	83.29	83.41
ResNet-50	89.12	89.93	89.39	89.56
ResNet-50-T (w/o GANs)	97.72	97.66	97.71	97.68
ResNet-50-T (w/o GM)	97.94	97.85	97.73	97.78
Proposed (w/ GANs and GM)	**98.16**	**98.24**	**98.06**	**98.15**

In addition, Fig.3 displays the confusion matrix of the baseline model and the proposed method. We can see that our method achieve the best performance and the cell images of all the staining pattern being predicted correctly are the most. The ROC curves of different methods are shown in Fig.4, we can see that the performance of the proposed method is the best. For the reason that the proposed segmentor with GANs can provide fine contour for the classification task and the addition of GM for loss

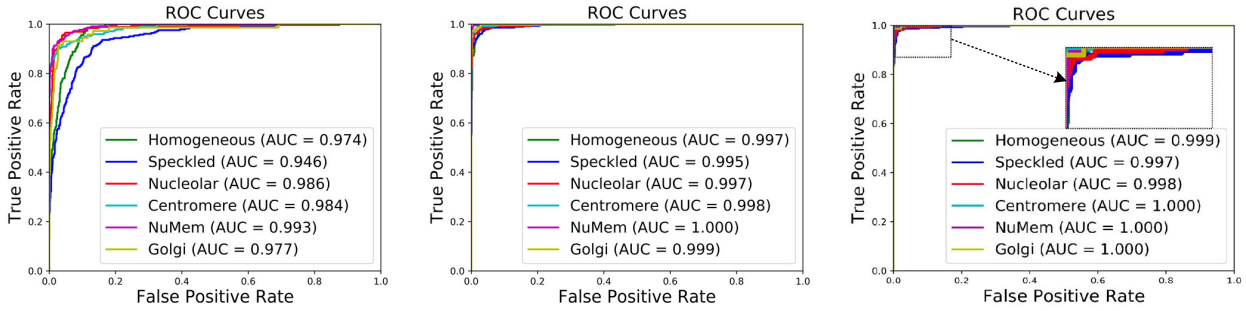

Fig. 4. The ROC curve of different methods on six staining patterns of ICPR 2016 dataset. From left to right, the ROC curves of ResNet-50, ResNet-50-T, the proposed method.

function can also optimize the proposed network to learn the distribution of characteristics so that the proposed model can converge fast and achieve promising performance.

To display the results more intuitively, we conduct the t-SNE visualization of the classification performance, as shown in Fig. 5. We can observe that most cells are correctly separated and only few cells are falsely classified to other categories. Moreover, the number of the Homogeneous and Centromere being predicted correctly are largest in the proposed method, the reason is that these two class cells are with more distinct boundaries and uniform illumination than other classes. The number of the Speckled is prediction correctly least for all classification networks except our model, the reason is that the most inhomogeneous illumination of these cells, as well as some bright spots inside and around the cells and some non-cellular interference areas that can greatly interfere with the prediction results.

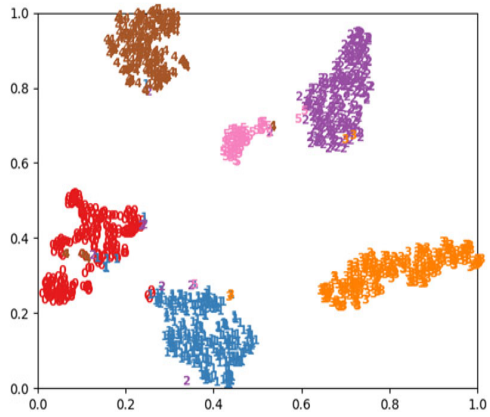

Fig. 5. t-SNE visualization of the classification performance on six staining patterns, where 0, 1, 2, 3, 4, 5 represent Homogeneous, Speckled, Nucleolar, Centromere, Nuclear membrane, Golgi, respectively.

IV. Conclusions

In this paper, we propose a novel segmentation guided HEp-2 cell images classification framework. It contains three modules, the generator, discriminator and classifier module. The first two modules form the GANs for segmentation task and the third module is used for classification task. The deeplabv3 plus and Xception model are employed to build the generator to generate multi-scale feature information and refine the object boundaries. Besides, a discriminator is leveraged to boost the generator to produce segmentation masks that are closer to the ground truth. Accordingly, the generated masks are concatenated with the corresponding original images, which is fed to the ResNet-50 network to complete the classification task. In addition, an improved classification loss function based on GM is proposed to optimize the classification network. Experimental results show that the proposed framework is effective.

Acknowledgment

This work was supported partly by National Natural Science Foundation of China (Nos.61871274, 61801305 and 81571758), Shenzhen Science and Technology Program under Grants (Nos. JCYJ 20170302150411789, JCYJ 20170302142515949, GCZX 2017040715180580, GJHZ 20180418190529516, and JSGG 20180507183215520), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Peacock Plan (Nos. KQTD2016053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (Nos. CYJ20170413152804728, JCYJ20180507184647636, JCYJ20170818142347251 and JCYJ20170302153337765).

References

[1] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking HEp-2 cells classification methods," *IEEE transactions on medical imaging,* vol. 32, pp. 1878-1889, 2013.

[2] A. B. L. Larsen, J. S. Vestergaard, and R. Larsen, "HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling," *IEEE transactions on medical imaging,* vol. 33, pp. 1573-1580, 2014.

[3] A.-D. Khamael, J. Banks, I. Tomeo-Reyes, and V. Chandran, "Automatic segmentation of HEp-2 cell Fluorescence microscope images using level set method via geometric active contours," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 81-83.

[4] V. S. Prasath, Y. M. Kassim, Z. A. Oraibi, J.-B. Guiriec, A. Hafiane, G. Seetharaman, and K. Palaniappan, "HEp-2 cell classification and segmentation using motif texture patterns and spatial features with random forests," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 90-95.

[5] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 cell image classification with deep convolutional neural networks," *IEEE journal of biomedical and health informatics,* vol. 21, pp. 416-428, 2016.

[6] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, and B. Lei, "A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning," *Pattern Recognition,* vol. 79, pp. 290-302, 2018.

[7] H. Xie, H. Lei, Y. He, and B. Lei, "Deeply supervised full convolution network for HEp-2 specimen image segmentation," *Neurocomputing,* vol. 351, pp. 77-86, 2019.

[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801-818.
[9] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.
[10] Z. Han, B. Wei, A. Mercado, S. Leung, and S. Li, "Spine-GAN: Semantic segmentation of multiple spinal structures," *Medical image analysis,* vol. 50, pp. 23-35, 2018.
[11] D. Mahapatra, Z. Ge, S. Sedai, and R. Chakravorty, "Joint registration and segmentation of xray images using generative adversarial networks," in *International Workshop on Machine Learning in Medical Imaging*, 2018, pp. 73-80.
[12] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing,* vol. 321, pp. 321-331, 2018.
[13] 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.
[14] W. Wan, Y. Zhong, T. Li, and J. Chen, "Rethinking feature distribution for loss functions in image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9117-9126.
[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125-1134.
[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," [Online]. Available: https://arxiv.org/abs/1502.03167, 2015.