Cell Segmentation by Image-to-Image Translation using Multiple Different Discriminators

Sota Kato and Kazuhiro Hotta
Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan

Keywords: Image to Image Translation, Semantic Segmentation, Cell Segmentation.

Abstract: This paper presents a cell image segmentation method by improving the pix2pix. Pix2pix improves the accuracy by competing a generator and a discriminator. The relationship of generator and discriminator is likened as follows. A generator is a fraudster who creates a fake image to fool the discriminator. A discriminator is a police officer who checks the fake image created by the generator. If we increase the number of police officers and different police officers are used, they have different roles and various viewpoints are used to check the fake image. In experiments, we evaluate our method on segmentation problem of cell images. We compared our method with conventional pix2pix using one discriminator. As a result, the accuracy will be improved. Thus, we propose to use multiple different discriminators to improve the segmentation accuracy of pix2pix. We confirmed that our proposed method outperformed conventional pix2pix and pix2pix using multiple same discriminators.

1 INTRODUCTION

In recent years, researches on generative adversarial networks (GAN) have been paid attention (Goodfellow, 2014). Pix2pix (Isola and Zhu, 2017) and CycleGAN (Zhu and Park, 2017) which can train image-to-image transformation are effective for many tasks. In recent research, segmentation methods using GAN has been proposed and a cleaner segmentation images can be generated. Pix2pix is also effective for segmentation problem that assigns class labels to all pixels in an image, and it has been applied to medical and cell biology (Ehsani and Mottaghi, 2018). In particular, cell image segmentation tends to be subjective because it has been done manually, but we may get objective results by deep learning (Ji and Li, 2015).

In this paper, we focus on the problem of automatic cell image segmentation using pix2pix. Pix2pix consists of a generator and a discriminator. In general, the relationship is considered as follows. The generator is like a fraudster who creates a fake image to fool the discriminator. The discriminator is a police officer who checks the fake image created by the generator. Generator and discriminator are improved by competing each other.

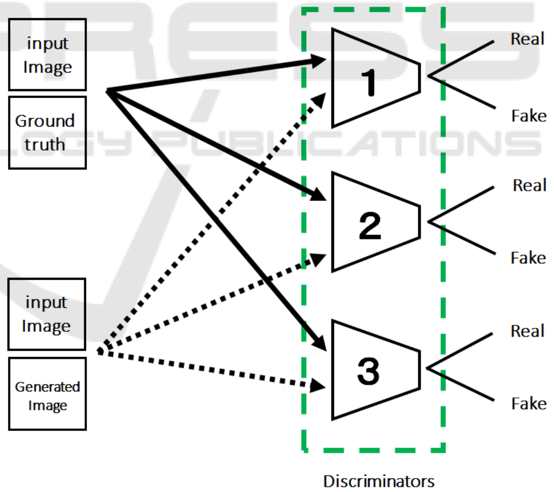


Figure 1: Concept of the proposed method.

If there are multiple police officers for finding out fake images, it is more effective in comparison with only one police officer. In recent years, the similar idea for improving the accuracy of the generated image by increasing the number of discriminators in GAN has been proposed (Durugkar and Gemp, 2017). Although they used the same discriminators, multiple different discriminators should be used because different judgment criteria are required to find out

Kato, S. and Hotta, K.
Cell Segmentation by Image-to-Image Translation using Multiple Different Discriminators.
DOI: 10.5220/0009170103300335
In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 4: BIOSIGNALS*, pages 330-335
ISBN: 978-989-758-398-8; ISSN: 2184-4305

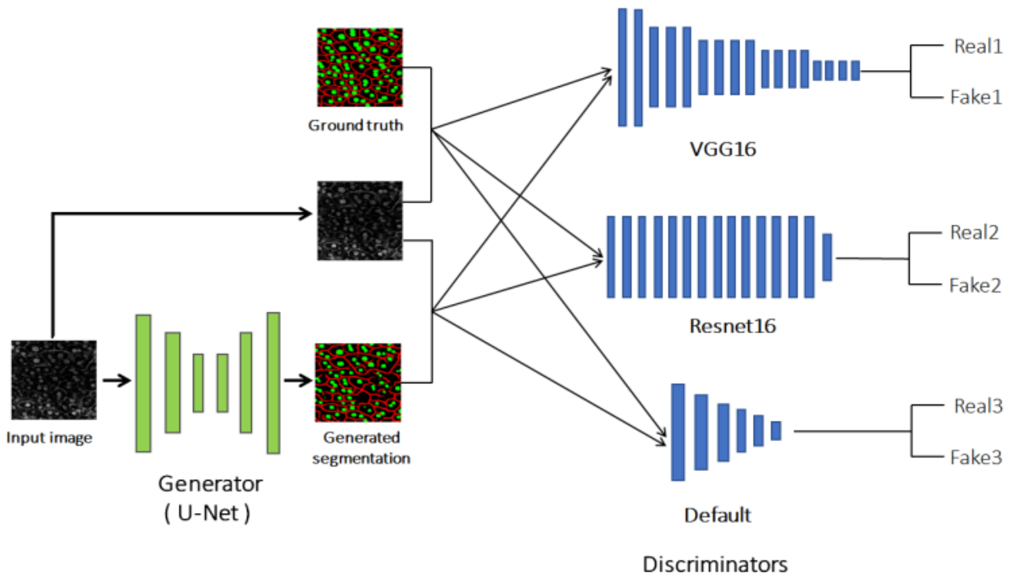


Figure 2: Overview of the proposed method.

fake images well. Therefore, we propose the pix2pix using multiple different discriminators as shown in Figure 1.

In experiments, we evaluate our method on segmentation problem of cell images. We classify cell images into three categories; cell membrane, cell nucleus and background. We compared our method with conventional pix2pix using one discriminator. In addition, we also evaluated the pix2pix with multiple same discriminators, the effectiveness of our method is demonstrated.

The structure of this paper is as follows. Section 2 describes related works. Section 3 explains the proposed method. Section 4 describes the dataset and evaluation method. We show the experimental results in section 5. Finally, a summary and future works are described in section 6.

2 RELATED WORKS

2.1 Generative Adversarial Network

Generative Adverbial Network (GAN) consists of a generator and a discriminator. Generator generates an image from noise and a discriminator judges whether the generated image is true or not. Since the input is noise, it cannot be used for training image-to-image transformation.

2.2 Conditional GAN

Conditional GAN (cGAN) has been proposed use class label information in GAN to generate the image of specific class. This combines label information with the noise vector z at the input of the generator. Label information is also added to the discriminator. Thus, cGAN can generate images of the specific class.

2.3 Pix2pix

Pix2pix can learn image-to-image transformation by using the U-net as a generator. Loss function of pix2pix is similar with conditional-GAN shown in equation (1). Pix2pix added L1 regularization loss between ground truth and generated image in equation (2) to the loss function as shown in equation (3). CycleGAN is extends this approach. It does not require the corresponding image pairs.

$$L_{cGAN}(G,D) = E_{x,y\sim P_{data}(x,y)}[\log D(x,y)] + E_{x,y\sim P_{data}(x),z\sim P_z(z)}[\log(1-D(x,G(x,z))] \quad (1)$$

$$L_{L1}(G) = E_{x,y\sim P_{data}(x,y),z\sim P_z(z)}[\|y-G(x,z)\|_1] \quad (2)$$

$$G^* = argmin_G max_D L_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G) \quad (3)$$

2.4 Generative Multi-adversarial Networks

Recently, similar method to our method has been proposed. However, they used multiple same discriminators to improve the accuracy of DCGAN. We consider that multiple different discriminators should be used because different viewpoints are required to not be fooled by a generator. In our method, multiple different discriminators are used to improve the accuracy of image-to-image transformation.

3 PROPOSED METHOD

DCGAN is likened as a fraudster and a police officer. Generator creates a fake image and discriminator judges whether it is real or fake. In pix2pix, U-net is used as a generator and normal CNN is used as a discriminator (Ronneberger and Fischer, 2015). We consider that discriminator with normal CNN (a police officer) is weak in comparison with generator with U-net (a fraudster). Discriminator may be fooled easily by generator. If there are some police officers, it is not fooled easily. Thus, we use multiple discriminators in pix2pix.

As described previously, Generative Multi-Adversarial Networks used multiple same discriminators in DCGAN. However, similar judgment may be obtained if multiple same discriminators are used. If multiple different discriminators are used, various viewpoints are used to judge whether generated image is real or fake. In addition, it is expected that each discriminator has a different role. As a result, better image-to-image transformation can be trained.

Figure 2 shows the overview of the proposed method. In the proposed method, U-net is used as a generator. This is the same as the conventional pix2pix. However, we use three different discriminators. The first discriminator is VGG16 (Simonyan and Zisserman, 2015). We use the same number of filters in the convolution layers as VGG16. However, we did not use fully connected layers, and we use global average pooling at the last convolutional layer.

The second discriminator is Resnet16 (He and Zhang, 2016). Due to the memory of GPU, we did not use deeper Resnet. The third discriminator is the original discriminator used in pix2pix. It is CNN with 6 layers, and we call it “Default” discriminator. All discriminators use global average pooling (Lin and Chen, 2014) in the last layer. We train all discriminators from full scratch.

The final loss function is the same as conventional pi2pix, and each discriminator is updated in the order of equations (4) to (6).

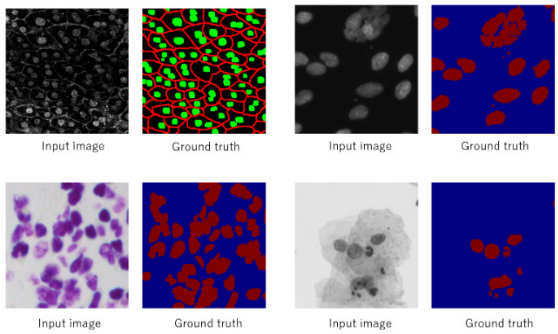


Figure 3: Example of dataset.

$$G^* = argmin_G max_{D_{VGG16}} L_{cGAN}(G, D_{VGG16}) + \lambda \mathcal{L}_{L1}(G) \quad (4)$$

$$G^* = argmin_G max_{D_{Resnet}} L_{cGAN}(G, D_{Resnet}) + \lambda \mathcal{L}_{L1}(G) \quad (5)$$

$$G^* = argmin_G max_{D_{Default}} L_{cGAN}(G, D_{Default}) + \lambda \mathcal{L}_{L1}(G) \quad (6)$$

4 DATASET AND EVALUATION METHOD

4.1 Dataset

We used 50 cell images with ground truth attached by Kyoto University. They were emitted with a fluorescent marker on the cell membrane and nucleus of the mouse liver. Images sizes are 256×256 pixels. 40 images were used for training, five images are used for validation and the remaining five images are for test.

We also evaluate our method on different datasets. We used nuclei segmentation datasets, 2018 Data Science Bowl in Kaggle. This dataset contains a large number of nuclei images. The images were captured under various conditions. We select three kinds of cell images from the dataset and make three datasets. Images sizes are 256 x 256 pixels and the number of images in each dataset is 50. 40 images were used for training, 5 images are used for validation and the remaining 5 images are for test.

Examples of cell image and ground truth are shown in Figure 3. The ground truth image includes three labels; cell nucleus (green), cell membrane (red) and background (black). Other 3 datasets include cell nucleus (red) and background (blue).

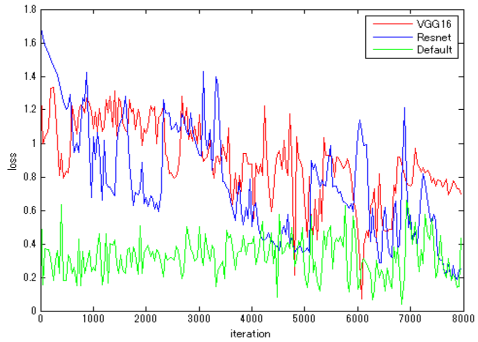


Figure 4: Loss of three discriminators in our method.

4.2 Evaluation Measures

The segmentation accuracy of each class is evaluated by Interactive over Union (IoU). IoU computes the overlapping rate between the predicted result and ground truth. Since the number of pixels in each class is different, we used mean IoU (mIoU) as the final evaluation measure.

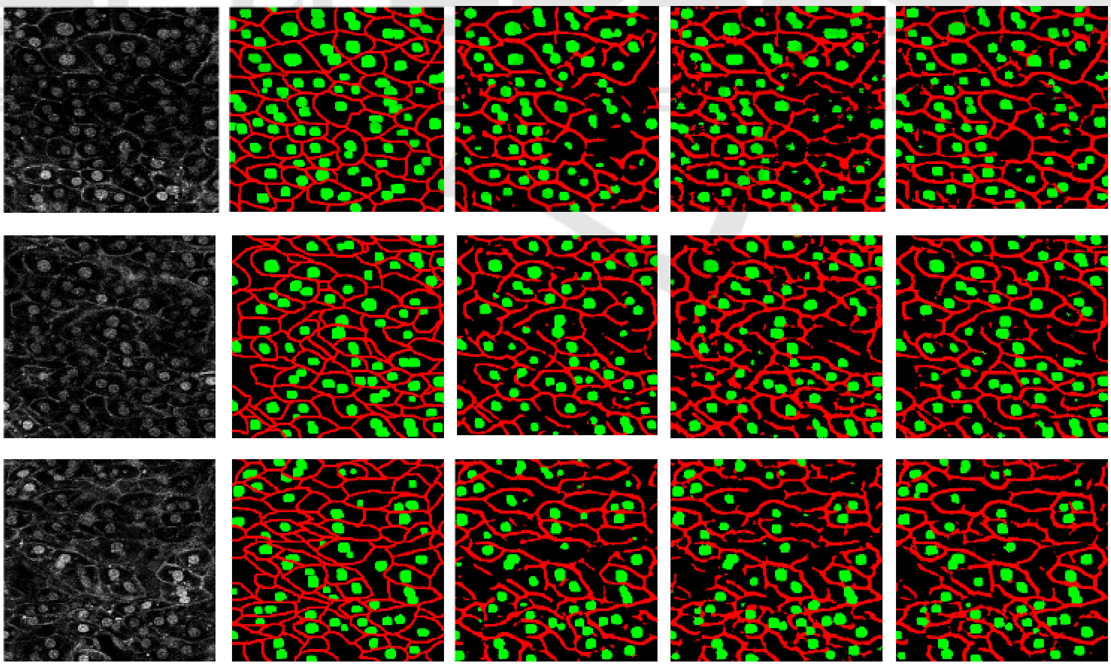
4.3 Comparison Methods

In the experiments, we evaluate the proposed method, original pix2pix, pix2pix with three same discriminators used in the original pix2pix, pix2pix with three same discriminators (VGG16) or (Resnet16). We evaluated all methods three times because the accuracy changes by the random number. We used average accuracy of three times evaluations.

5 EXPERIMENTS

Table 1 shows IoU of each class and mIoU. The pix2pix using three same discriminators was able to improve mIoU in comparison with the original pix2pix using only one discriminator. Furthermore, mIoU was improved by using multiple different discriminators in comparison with multiple same discriminators. This demonstrated the effectiveness of our method.

The IoU of the cell membrane was the best when three different discriminators were used. The IoU of the cell nucleus was the best when three default discriminators were used. The accuracy of background was the best when three Resnet16 were used. These results show that each discriminator has good class. We considered that mIoU was improved by using discriminators with different viewpoints.



(a) input (b) Ground truth (c) U-Net (d) pix2pix (e) Our method

Figure 5: Comparison results (a) Input image. (b) Ground truth. (c) U-Net. (d) Conventional pix2pix. (e) Our method.

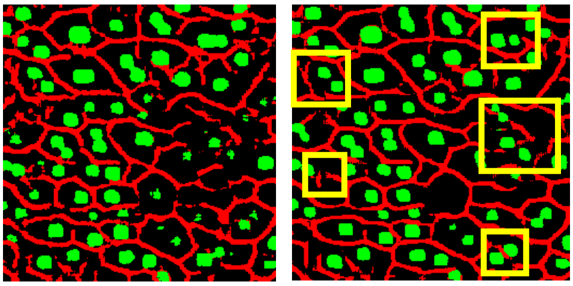


Figure 6: Enlarged results. The left column shows the result of conventional pix2pix and right column shows that of the proposed method.

Table 1: Comparison between conventional u-net, pix2pix and our proposed method.

	membane	nucleus	background	mIoU
U_Net	41.02	64.51	71.11	58.66
pix2pix	44.91	59.90	71.16	58.88
Our method	45.94	70.12	71.42	62.49

mIoU : mean IoU

Table 2: Evaluation result while changing discriminators table type styles.

	membane	nucleus	background	mIoU
Default x 3	47.53	63.34	71.55	60.81
VGG16 x 3	43.57	67.88	72.31	61.25
Resnet16 x 3	47.02	66.88	72.40	62.10
Our method	45.94	70.12	71.42	62.49

mIoU : mean IoU

Table 2 shows the results of only the conventional pix2pix and the proposed method. The proposed method improved 10.2% for the cell nucleus, 1 % for the cell membrane. Totally, mIoU was improved 3.8% in comparison with the original pix2pix.

Figure 4 shows the graph of the loss of three discriminators used in the proposed method. The loss of each discriminator is different. Default is nearly flat but VGG 16 and Resnet16 were changing their roles.

Figure 5 shows the result of cell segmentation, and Figure 6 shows the enlarged images of Figure 5. Focusing on the yellow frames in Figure 6, the proposed method segments cell nucleus that pix2pix could not segment well.

In addition, we can see that ambiguous cell membranes can be more clearly classified. These results also demonstrated the effectiveness of usage of multiple different discriminators.

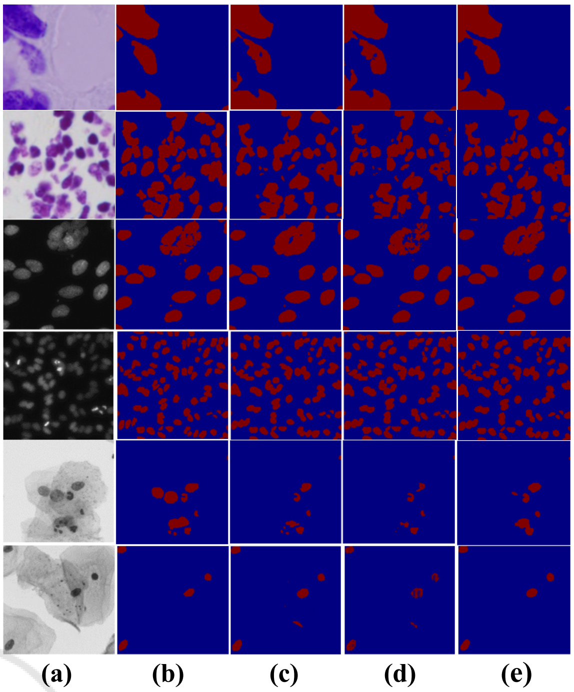


Figure 7: Comparison results. (a) Input image. (b) Ground truth. (c) U-net. (d) Conventional pix2pix. (e) Our method.

We also evaluated our method on three different nuclei segmentation datasets. Figure 7 shows the result of cell segmentation. Figure shows that our approach worked well for all datasets even if input images are much different. The proposed method is better than conventional U-net and pix2pix. Experimental results demonstrated the effectiveness of the proposed method.

6 CONCLUSIONS

In this paper, we improved the pix2pix using multiple different discriminators for the segmentation of cell images. By the experiments on cell images, mIoU was improved 3.8% in comparison with conventional pix2pix.

However, the accuracy of the cell membrane was not much improved by the proposed method though the accuracy of cell nucleus was much improved. Here we used multiple different discriminators but multiple different generators may improve the accuracy. This is a subject for future works.

ACKNOWLEDGEMENTS

This work is partially supported by MEXT/JSPS KAKENHI Grant Number 18H04746 "Resonance Bio" and 18K111382.

REFERENCES

I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D.W.Farley, S.Ozair, A.Courville, Y. Bengio, “Generative Adversarial Networks,” In Proc. Advances in Neural Information Processing Systems, 2014.

P. Isola, JY. Zhu, T. Zhou, A.A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” In Proc. International Conference on Computer Vision and Pattern Recognition, pp.1125-1134, 2017.

J.Y. Zhu, T. Park, P. Isola, Alexei A, “Unpaired Image - to - Image Translation using Cycle-Consistent Adversarial Networks,” In Proc. International Conference on Computer Vision, pp.2223-2232, 2017.

K. Ehsani, R. Mottaghi, A. Farhadi, “SeGAN: Segmenting and Generating the Invisible”, In Proc. Conference on Computer Vision and Pattern Recognition, pp.6144-6153, 2018.

Z. Zhang, L. Yang, Y. Zheng, “Translating and Segmenting Multimodal Medical Volumes With Cycle- and Shape-Consistency Generative Adversarial Network”, In Proc. Conference on Computer Vision and Pattern Recognition, pp.9242-9251, 2018.

I. Durugkar, I. Gemp, and S. Mahadevan,”Generative Multi-Adversarial Networks,” In Proc. International Conference on Learning Representations, 2017.

K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” In Proc. International Conference on Learning Representations, 2015.

K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learninig for Image Recognition,” In Proc. International Conference on Computer Vision and Pattern Recognition, pp.770-778, 2016.

O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” In Proc. Medical Image Computing and Computer-Assisted Intervention, pp.234-241, 2015.

M. Havaei, A. Davy, D. W. Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, H.Larochelle, “Brain tumor segmentation with Deep Neural Networks,” Medical Image Analysis, vol. 35, pp.18-31 , 2017.

X. Ji , Y. Li , J. Cheng , Y. Yu , M. Wang, “Cell image segmentation based on an improved watershed algorithm,” In Proc. International Congress on Image and Signal Processing, pp.433-437, 2015.

M. Lin, Q. Chen, and S. Yan, “Network in network,” In Proc. International Conference on Learning Representations, 2014.