Cell Image Segmentation by Integrating Pix2pixs for Each Class

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

193427019@ccalumni.meijo-u.ac.jp and kazuhotta@meijo-u.ac.jp

Abstract

This paper presents a cell image segmentation method using Generative Adversarial Network (GAN) with multiple different roles. Pix2pix is a kind of GAN can be used for image segmentation. However, the accuracy is not sufficient because generator predicts multiple classes simultaneously. Thus, we propose to use multiple GANs with different roles. Each generator and discriminator has a specific role such as segmentation of cell membrane or nucleus. Since we assign each generator and discriminator to a different role, they can learn it efficiently. We evaluate the proposed method on the segmentation problem of cell images. The proposed method improved the segmentation accuracy in comparison to conventional pix2pix.

1. Introduction

Convolutional Neural Network (CNN) [17, 18] achieved high accuracy in various kinds of image recognition problems such as image classification, object detection, pose estimation etc. In addition, semantic segmentation assigns class labels to all pixels in an input image. Semantic segmentation using CNN is also applied to cartography [20, 21], automatic driving [13, 19], medicine and cell biology [5, 6].

In recent years, many researchers pay attention to Generative Adversarial Network (GAN) [1] because it can generate realistic images in comparison with conventional methods. In addition, pix2pix [2], CycleGAN [3] and StarGAN [4] which are the extended version of GAN can train image-to-image transformation. In particular, pix2pix is effective for semantic segmentation problems. Human experts segment cell images manually now, and the criterion for segmentation varies on each expert. As a result, subjective results are obtained. If we develop an automatic segmentation method, we can obtain objective results by the same criteria. In this paper, we would like to segment three classes; membrane, nucleus and background.

Pix2pix consists of a generator and a discriminator. When pix2pix is applied to segmentation problem of multiple classes, it is difficult for only one generator to segment multiple classes simultaneously. If we assign a role such as segmentation of only cell membrane to one generator, we can train generators effectively and improve the accuracy. Therefore, we should give a different role to each generator and discriminator. Figure 1 shows the overview of our method.

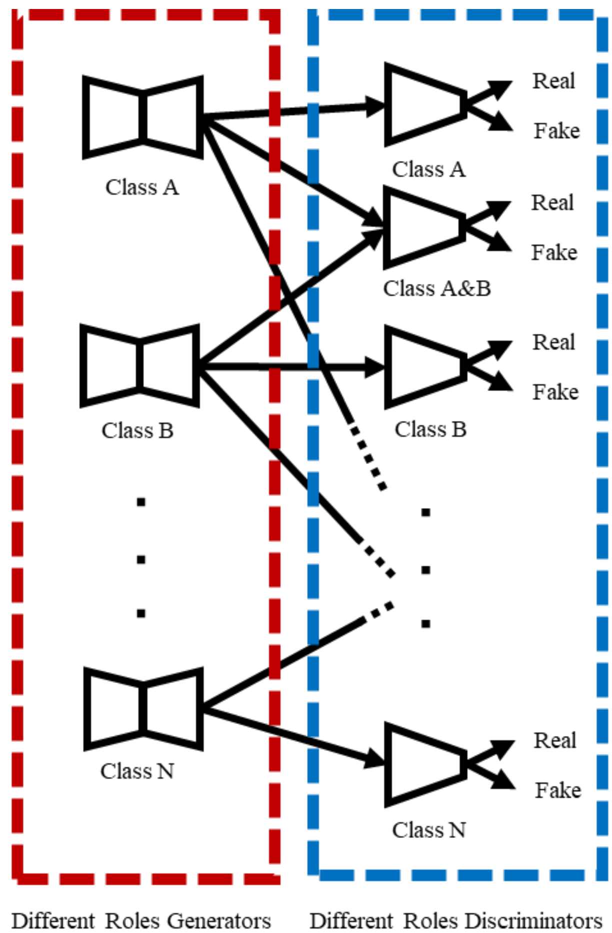


Figure 1: Overview of the proposed method.

We propose two kinds of segmentation methods by using multiple generators and discriminators with a specific different role. The first method is Dual Different Roles GAN (DDR-GAN) that two generators segment cell

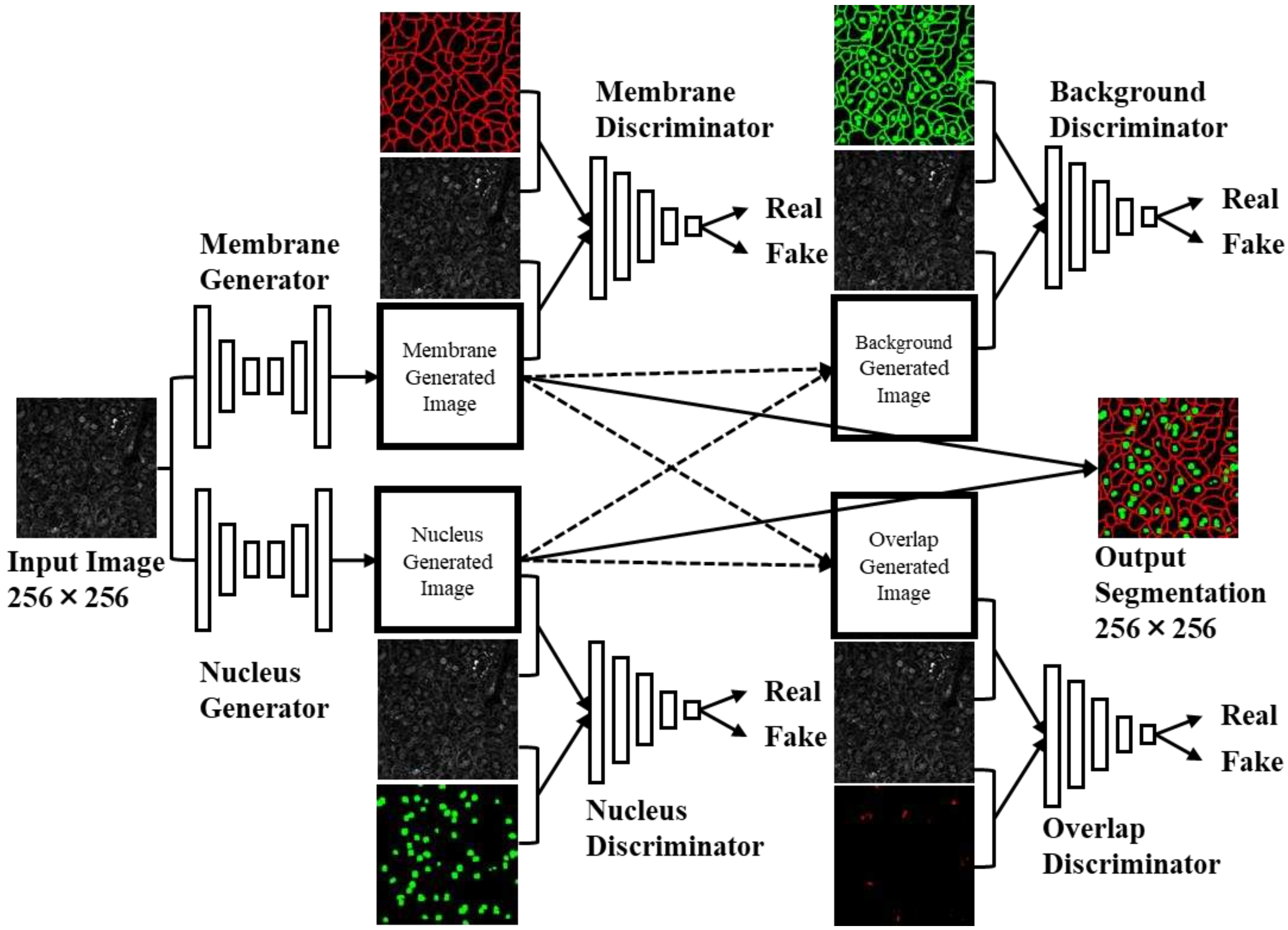


Figure 2: Proposed method: Dual Different Roles GAN (DDR-GAN).

membrane or nucleus. The second method is Triple Different Roles GAN (TDR-GAN) that three generators segment cell membrane, nucleus and background. Of course, se use multiple discriminators with a different role. Figure 2 and 3 show the overview of DDR-GAN and TDR-GAN.

In experiments, we evaluate the proposed method on a public cell image dataset [16] and our cell image dataset [15]. We use IoU (Intersection over Union) and mean IoU as evaluation measures. In the two datasets, we confirmed that the proposed method gave higher accuracy than that of the conventional pix2pix. Furthermore, we evaluate our method without some discriminators, and the effectiveness of our method is shown.

This paper is organized as follows. In section 2, we describe related works. The details of proposed method are explained in section 3. In section 4, we evaluate our proposed method on segmentation of cell images. Finally, we describe conclusion in section 5.

2. Related works

Various methods are proposed for semantic segmentation [8, 12, 13]. Pix2pix uses an encoder-decoder network as a generator in order to generate a transformed image. Then, discriminator determines whether the generated image and training image are real or fake. Pix2pix [2] is a kind of GAN can be used for segmentation, and it gave superior result on segmentation than non-adversarial standard networks [7].

Recently, there are some approaches using multiple GANs. Mixture Generative Adversarial Nets (MGAN) [9] used multiple generators. The distribution of training data is estimated by the mixture of distributions obtained by K generators. By using multiple generators, mode collapse can be avoided.

Dual Discriminator Generative Adversarial nets (D2GAN) [10] used two discriminators. Authors assign two discriminators to different roles. One discriminator is trained to give training images to high score and generated images to low score. On the other hand, the other discriminator is trained the opposite scores. Generator is trained to fool two discriminators. By using two different discriminators, we can avoid mode collapse.

Generative Multi-Adversarial Networks (GMAN) [11]

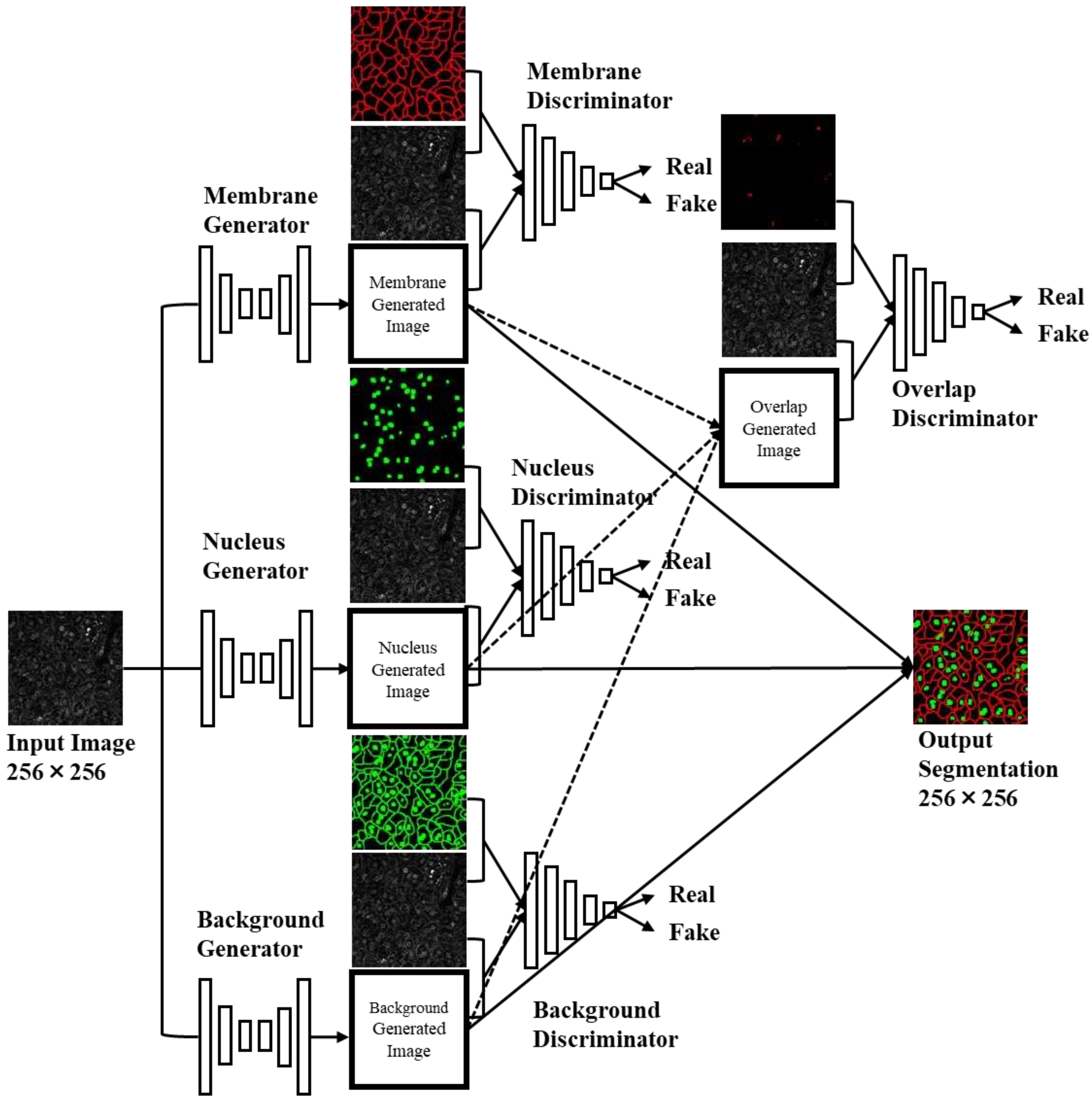


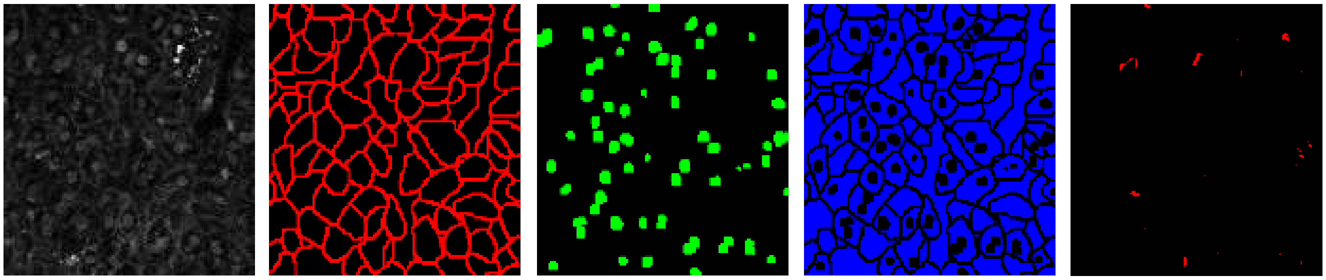
Figure 3: Proposed method: Triple Different Roles GAN (TDR-GAN).

used the multiple same discriminators. They prepare multiple discriminators and select the discriminator with the maximum score. A generator is trained for the discriminator. By using the good discriminator for training the generator, it is expected the convergence to good solution.

Conventional methods used multiple generators or discriminators. However, those methods did not give a specific role to generators or discriminators. We consider that the accuracy would improve if we use multiple GANs with different roles.

3. Proposed Method

This section describes the details of the proposed method. We explain the details of networks in section 3.1. Proposed method 1; DDR-GAN (Dual Different Roles GAN) is explained in section 3.2. We explain the proposed method 2; TDR-GAN (Triple Different Roles GAN) in section 3.3.



(a) Input (b) Membrane (c)Nucleus (d) Background (e) Overlap

Figure 4: Examples of fluorescence images of the liver of transgenic mice that expressed fluorescent markers on the cell membrane and nucleus.

3.1. Network Details

Pix2pix consists of a generator and a discriminator. When we apply it to a segmentation problem, the generator must classify multiple classes at the same time. This decreases the segmentation accuracy. Therefore, we prepare a generator for each class to segment the own class, and the results by generators are integrated to get the segmentation result with high quality. In addition, by preparing a discriminator for each class, we can judge whether the generated image is real or fake from various viewpoints. Therefore, we train each class independently by a generator, and the probability of each class is generated from each generator. Each probability map is input to the discriminator of the corresponding class to judge whether it is real or fake. Moreover, it is possible to create a discriminator that trains only the part that we want to judge. The details are explained in the proposed method.

Generators and discriminators in the proposed method have the same architecture. Thus, we use U-Net [8] as generators. CNN with 5 convolution layers is used as discriminators. Full scratch learning is used for all generators and discriminators.

3.2. Dual Different Roles GAN

Figure 2 shows the DDR-GAN. DDR-GAN generates the probability maps of membrane and nucleus by two generators, and the classification result of each class is fed into each own discriminator.

From the classification results of membrane and nucleus, we regard non-membrane and non-nucleus as background class. As a result, we could obtain background segmentation results. The segmentation result of background class is fed into the discriminator for background.

Ground truth labels have the overlap between cell membrane and nucleus because human expert assigned each class label to images independently. In the case, human expert gave a priority to cell membrane. Thus, we also gave a priority to cell membrane for overlapping pixels. To do it effectively, we used the discriminator for overlapping pixels between cell membrane and nucleus.

The loss function is updated from three discriminators for a generator. For example, the loss function is updated as follows for the generator of cell membrane.

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

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

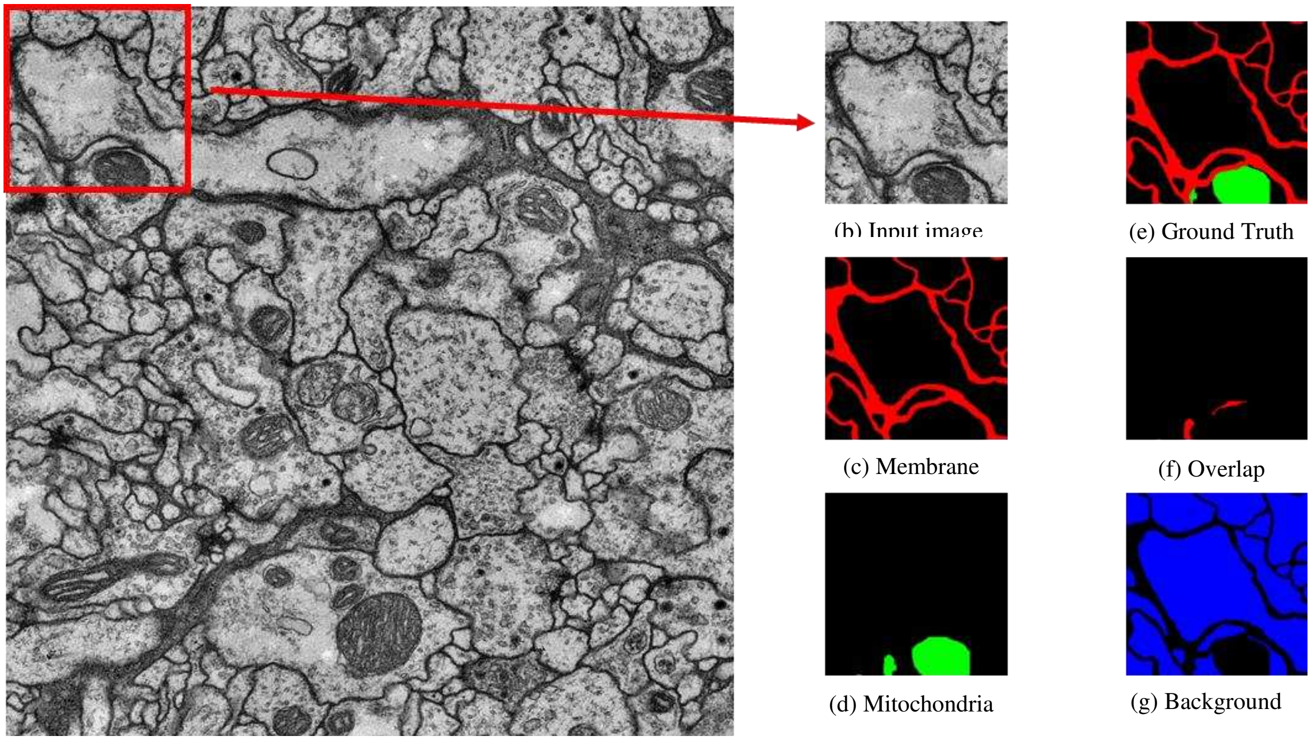
$$G^*_{mem} = arg\ min_{G_{mem}} max_{D_{mem}} L_{cGAN}(G_{mem}, D_{mem}) + arg\ min_{I_{back}} max_{D_{back}} L_{cGAN}(I_{back}, D_{back}) + arg\ min_{I_{overlap}} max_{D_{overlap}} L_{cGAN}(I_{overlap}, D_{overlap}) + \lambda \mathcal{L}_{L1}(G_{mem}) \quad (3)$$

There are three inputs in the equation, x is the input image, y is the ground truth. G(.) represents the segmentation result which is the output of a generator, and $I(.)$ represents the segmentation result of background or overlapping regions which is obtained by integrating the results of each generator, and D(.) represents the result by a discriminator whether the input is real or fake. G_{mem} is a generator of cell membranes, D_{mem} is that of cell membrane, D_{back} is that of background, and $D_{overlap}$ is the discriminator for overlapping parts. I_{back} and $I_{overlap}$ represent the segmentation results of background and overlapping parts obtained from the segmentation results of cell membrane and cell nucleus.

The generator was trained using the loss function in Eq.(2) and Eq.(3) which is the combination of three equations in Eq.(1) for the cell membrane, background and overlapping part. The same loss function was used for the generator of cell nucleus.

3.3. Triple Different Roles GAN

Figure 3 shows the second proposed method (TDR-GAN). Since the dataset has three classes, the



(b) Input image (e) Ground Truth

(c) Membrane (f) Overlap

(d) Mitochondria (g) Background

(a) Original size image

Figure5: Examples of ssTEM images

proposed method using three generators may be better than that with two generators. Thus, TDR-GAN generates three probability maps of membrane, nucleus and background by three generators. The classification results obtained by three generators are fed into each own discriminator. Similarly as the DDR-GAN, the discriminator for overlapping pixels is used. TDR-GAN updates the loss function from two discriminators for one generator.

For the generator of cell membrane, TDR-GAN trained as follows.

$$G^*_{mem} = arg\ min_{G_{mem}} max_{D_{mem}} L_{cGAN}(G_{mem}, D_{mem}) + arg\ min_{I_{overlap}} max_{D_{overlap}} L_{cGAN}(I_{overlap}, D_{overlap}) + \lambda\mathcal{L}_{L1}(G_{mem}) \tag{4}$$

The loss function is similar with that of DDR-GAN except for that we use two equations (1) for cell membrane and overlapping part. The generators for cell nucleus and background also use the same loss function.

4. Experiments

This section shows evaluation results by the proposed method. We explain the dataset used in experiments in section 4.1. The second dataset is described in section 4.2. Experimental results are shown in section 4.3.

4.1. Cell Dataset

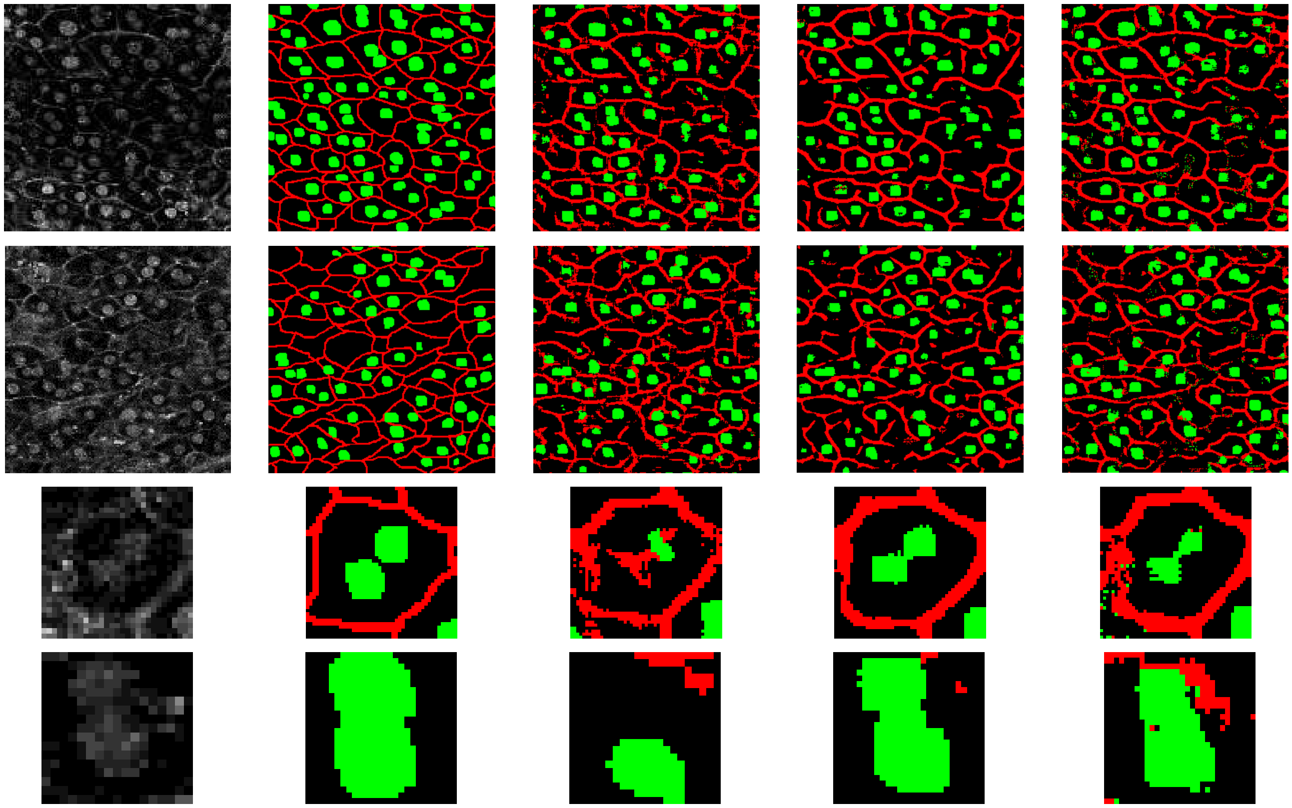
We evaluate our method on the fluorescence images of the liver of transgenic mice that expressed fluorescent markers on the cell membrane and nucleus. There are 50 images in total and the size of the image is 256 x 256 pixels. These images consist of three classes; membrane, nucleus and background. We divided those images into 30 training, 10 validation and 10 testing images [15].

Figure 4 shows the examples of cell images. Figure 4 (a) is a cell image, (b) is the ground truth of membrane attached by human expert. (c) and (d) show the ground truth of nucleus and background. (e) shows overlapping pixels between cell membrane and nucleus.

We use intersection over union (IoU) and mean IoU (mIoU) as evaluation measures. They are computed as

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

where TP, FP, and FN denote the true positive, false positive and false negative counts, respectively.



(a) Input image (b) Ground Truth (c) pix2pix [2] (d) DDR-GAN (e) TDR-GAN

Figure6: Segmentation results on cell dataset.

Table 1: Accuracy of proposed method on cell dataset.

method	Membrane[%]	Nucleus[%]	Background[%]	mIoU[%]
1. Comparison of DDR-GAN				
w/o Background & Overlap	36.61	58.74	70.54	55.30
w/o Background	37.13	59.62	70.43	55.73
w/o Overlap	36.28	61.18	70.38	55.95
Proposed method (Full model)	37.92	60.12	70.58	56.21
2. Comparison of TDR-GAN				
w/o Overlap	38.60	59.26	69.53	55.80
Proposed method (Full model)	38.83	58.77	70.06	55.89

Table 2: Segmentation results (IoU and mIoU) in comparison with the conventional method on cell dataset.

Method	Membrane[%]	Nucleus[%]	Background[%]	mIoU[%]
pix2pix	36.67	57.99	67.98	54.21
DDR-GAN	37.92	**60.12**	**70.58**	**56.21**
TDR-GAN	**38.83**	58.77	70.06	55.89

4.2. ssTEM Dataset

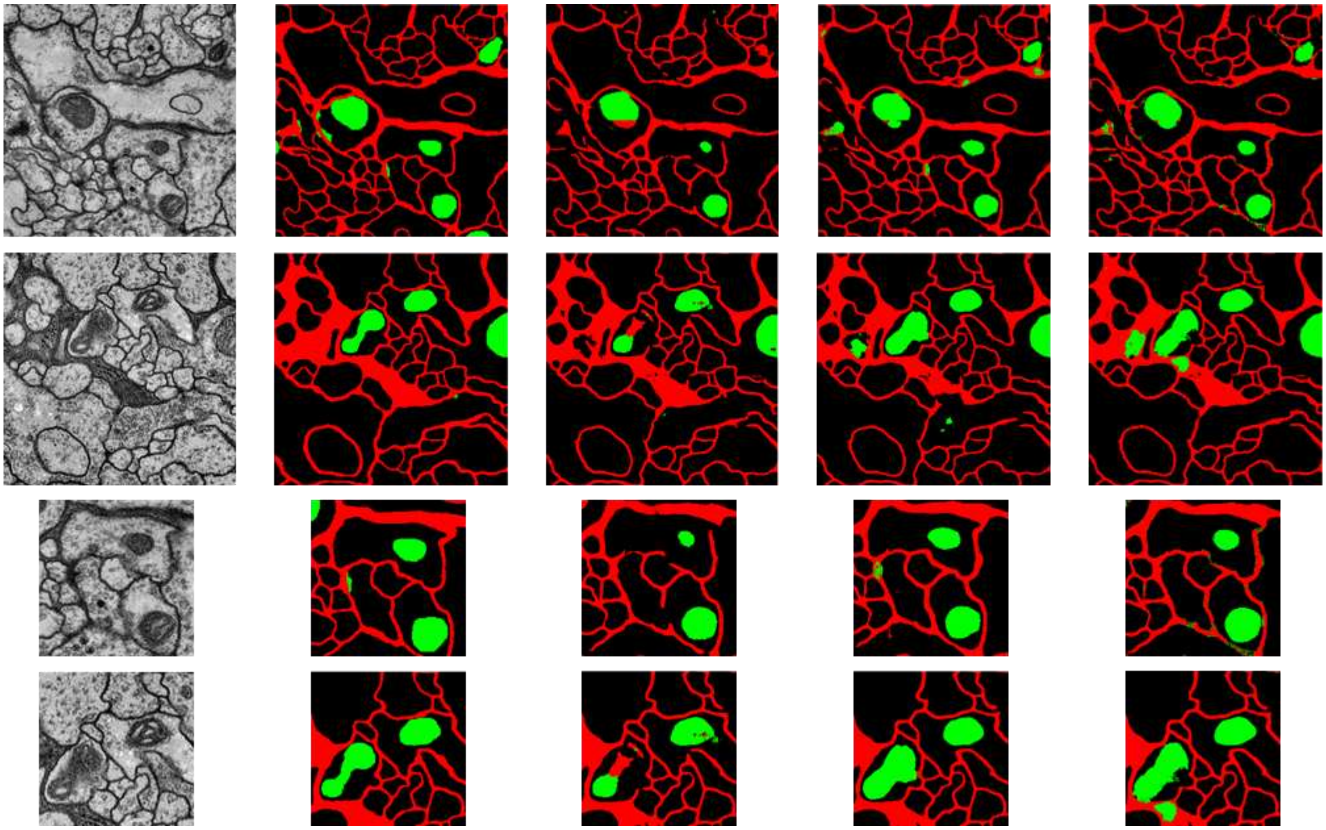
This dataset shows neural tissue from a Drosophila larva ventral nerve cord and was acquired using serial section Transmission Electron Microscopy at HHMI Janelia Research Campus [16]. There are 20 images of 1024 x 1024 pixels and ground truth. In this experiment, semantic segmentation is performed in three classes; membrane, mitochondria and others. We augmented 20 images to 320 images by cropping an image to 16 images on 256 x 256 pixels. We divided those images into 256 training, 32 validation and 32 test images.

Figure 5 shows the examples of cell images. Figure 5 (a) is a cell image, (b) is the ground truth of membrane attached by human expert. (c) and (d) show the ground truth of nucleus and background. (e) shows overlapping pixels between cell membrane and nucleus.

We found that the mitochondrial segmentation is not successful when the proposed method is applied to this dataset. Therefore, in order to make an effective method for this dataset, we first learned with the conventional pix2pix and fine-tuned the generator to the generator of mitochondria and cell membrane.

4.3. Experimental Results

To show the effectiveness of our proposed method using multiple GANs with different roles, we evaluate the accuracy without some discriminators. Table 1 shows the accuracy on cell dataset. DDR-GAN using all discriminators (Full model) gave the best mIoU. When we



(a) Input image (b) Ground Truth (c) pix2pix [2] (d) DDR-GAN (e) TDR-GAN

Figure7: Segmentation results on ssTEM dataset.

Table 3: Accuracy of proposed method on ssTEM dataset.

method	Membrane[%]	Mitochondria[%]	Background[%]	mIoU[%]
1. Comparison of DDR-GAN				
w/o Background & Overlap	73.00	81.50	89.39	81.30
w/o Background	68.22	75.82	88.32	77.45
w/o Overlap	69.02	81.98	88.66	79.89
Proposed method (Full model)	74.35	80.56	89.59	81.50
2. Comparison of TDR-GAN				
w/o Overlap	71.56	80.26	90.26	80.69
Proposed method (Full model)	73.34	80.45	89.30	81.03

Table 4: Segmentation results (IoU and mIoU) in comparison with the conventional method on ssTEM dataset.

Method	Membrane[%]	Mitochondria[%]	Background[%]	mIoU[%]
pix2pix	73.81	78.36	**89.66**	80.61
DDR-GAN	**74.35**	**80.56**	89.59	**81.50**
TDR-GAN	73.34	80.45	89.30	81.03

remove the discriminators for background and overlapping pixels, the accuracy was the worst. This shows the effectiveness of those discriminators.

Table 1 also shows the accuracy of TDR-GAN without discriminator for overlapping pixels or background decreased. These results demonstrated the effectiveness of multiple GANs with multiple different roles.

We compare the proposed method with the original pix2pix. Table 2 shows the comparison result. From Table 2, the proposed two methods outperformed the conventional pix2pix. DDR-GAN was better than TDR-GAN. This is because three discriminators (e.g cell membrane, background and overlapping pixels) are used for one generator in DDR-GAN though TDR-GAN used two discriminators (e.g. cell membrane and overlapping pixels) for one generator. Multiple discriminators were effective to train each generator.

When we compare the original pix2pix with our method without discriminators for background and overlapping pixels in Table 1, our method outperformed the original pix2pix. This demonstrated the effectiveness of task specific generators and multiple discriminators.

From Table 2, we confirmed that the DDR-GAN improved 1.25 % on cell membrane, 2.13% on cell nucleus, 2.6% on background and 2.00% on mean IoU in comparison with the conventional pix2pix.

Figure 6 shows segmentation results. From left to right shows the input image, ground truth, the result of pix2pix, that of DDR-GAN and that of TDR-GAN. The images in the lower row are the enlarged view of the upper row.

When we focus on the enlarged images, the segmentation of cell membrane by our method is better than conventional pix2pix. Cell nucleus is also detected more correctly by our method.

We evaluate the accuracy without some discriminators on the ssTEM dataset. Table 3 shows the accuracy. DDR-GAN using all discriminators (Full model) also gave the best mIoU on ssTEM dataset. In addition, when we remove the discriminators for background and overlapping pixels, the accuracy was the second best. When we remove the discriminator for background, the accuracy was the worst. Ablation study demonstrated the effectiveness of our proposed method.

Table 3 also shows the accuracy of TDR-GAN without discriminator for overlapping pixels or background. The accuracy was dropped. These results demonstrated the effectiveness of multiple GANs with multiple different roles on the ssTEM dataset.

We compare the proposed method with the original pix2pix. Table 4 shows the comparison result. From Table 4, the proposed methods outperformed the conventional pix2pix. We confirmed that the DDR-GAN improved 0.54 % on cell membrane, 2.20% on mitochondria and 0.89% on mean IoU in comparison with the conventional pix2pix.

Figure 7 shows segmentation results. From left to right shows the input image, ground truth, the result of pix2pix, that of DDR-GAN and that of TDR-GAN. The images in the bottom row are the enlarged view of the top row. When we focus on the enlarged images, the segmentation of cell membrane by our method is better than conventional pix2pix. Mitochondria is also segmented more correctly by our method.These results demonstrated the effectiveness of our method using multiple GANs with different roles.

5. Conclusion

In this paper, we proposed a semantic segmentation method using multiple GANs with different roles. In experiments on cell images, our method was improved mIoU in comparison with the conventional pix2pix.

However, our method is specialized for cell images, it is difficult to apply our method to the other dataset which includes many classes. We need to separate the categories for each GAN automatically. Gating network [14] can be used for this process. This is a subject for future works.

References

[1] I.J. Goodfellow, J.P. Abadie, M.Mirza, B. Xu, D.W. Farley, S. Ozair, A. Courville, and Y. Bengio.s, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014.

[2] P. Isola, JY. Zhu, T. Zhou, and A.A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5967-5976, 2017.

[3] J. Zhu, T. Park, P. Isola, and A.A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *IEEE International Conference on Computer Vision*, pp. 2223-2232, 2017.

[4] Y. Choi, M. Choi, M. Kim, J.W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789-8797, 2018.

[5] M. Havaei, A. Davy, D.W. Farley, A. Biard, A. Courville, Y.Begio, C. Pal, P.M. Jodoin, and H. Larochelle, "Brain tumor segmentation with Deep Neural Networks," *Medical Image Analysis*, vol. 35 pp. 18-31, 2017.

[6] O.X. Ji, Y. Li, J. Cheng, Y. Yu, and M. Wang, "Cell image segmentation based on an improved watershed algorithm," *International Congress on Image and Signal Processing*, pp. 433-437, 2015.

[7] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *Advances in Neural Information Processing Systems Workshops*, 2016.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241, 2015.

[9] Q. Hoang, T.D. Nguyen, T. Le, and D. Phung, "MGAN: Training Generative Adversarial Nets with Multiple Generators," *International Conference on Learning Representations*, 2018.

[10] T. Nguyen, T. Le, H. Vu, and D. Phung, "Dual Discriminator Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, pp. 2667-2677, 2017.

[11] I. Durugkar, I. Gemp, and S. Mahadevan, "Generative Multi-Adversarial Networks," *International Conference on Learning Representations*, 2017.

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789-8797, 2015.

[13] H. Zhao, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789-8797, 2017.

[14] Y. Hiramatsu, K. Hotta, A. Imanishi, M. Matsuda, and K. Terai, "Cell Image Segmentation by Integrating Multiple CNNs," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2205-2211, 2018.

[15] A. Imanishi, T. Murata, M. Sato, K. Hotta, I. Imayoshi, M. Matsuda, and K. Terai, "A Novel Morphological Marker for the Analysis of Molecular Activities at the Single-cell Level," *Cell Structure and Function*, Vol43, No.2, pp.129-140, 2018

[16] S. Gerhard, J. Funke, J. Martel, A. Cardona, and R. Fetter. "Segmented anisotropic ssTEM dataset of neural tissue," *figshare*, http://dx.doi.org/10.6084/m9.figshare.856713

[17] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning Applied to Document Recognition", *Proceedings of the IEEE*, vol.86, Issue.11, pp.2278-2324, 1998.

[18] A. Krizhevsky, I. Sutskever, G. and E. Hinton, "ImageNet classification with deep convolutional neural networks",

Proceedings of the International Conference on Neural Information Processing Systems, vol.1, pp.1097-1105, 2012.

[19] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", *Pattern Analysis and Machine Intelligence*, vol.39, pp.2481-2495, 2017.

[20] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote Sensing Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol.55, pp.645-657, 2017

[21] P. Ghamisi, and J.A. Benediktsson, "Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization", *IEEE Geoscience and Remote Sensing Letters*, vol.12, pp.309-313, 2015