# Master's thesis

# Investigation on Segmentation and Data Augmentation Method for Imbalanced Lung Confocal IF Images

## Daiki Katsuma

**Division of Electrical and Electronic Engineering**
**Graduate School of Engineering**
**Mie University**

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

Lung cancer is a leading cause of cancer death in the world, and in particular, the annual mortality number is in excess of 150,000 in the United States [1]. Studying the lung structures that perform gas exchange function is important in identifying disease mechanisms. This task looks simple; however, the human lung is a complex organ with high cellular heterogeneity. Its development and maintenance require interactive gene networks and dynamic cross-talk among multiple cell types. Furthermore, it is difficult to understand various causes of abnormal lung developmental patterns and adult lung disorders at the molecular level until we understand its normal development [2]. Quantification of normal lung development/formation and gene/cellular/molecular levels requires carefully assessing various lung imaging modalities.

The LungMAP consortium is one of the research projects for lung development. The consortium was initiated by the National Institutes of Health (NIH) in order to advance molecular, physiologic, and imaging research on the alveolar lung stage using normal, non-diseased human and mouse lung samples [3]. Along with the availability of various types of data, such as biological sequencing data, RNA, protein, lipid, signaling, we can also obtain multi-modal imaging data from the LungMAP database. In particular, lung tissue imaging data shows the protein/cell localization in different tissue structures, and these can provide a foundation upon which to build an atlas of the developing lung [4]. In this thesis, we focus on the confocal immunofluorescent (IF) images of lung tissues stained with specific sets of proteins. IF images are usually used for a proteomics process to understand the protein localization on the sub-cellular level by image annotation [5–7]. With the availability of large-scale lung confocal IF image, automatic machine learning driving image processing tasks such as segmentation of lung structures and quantification are important tasks. The realization of efficient automatic annotations and segmentation using IF images drastically reduces the manual burden among pulmonary specialists. It promises to increase the usability of such imaging data
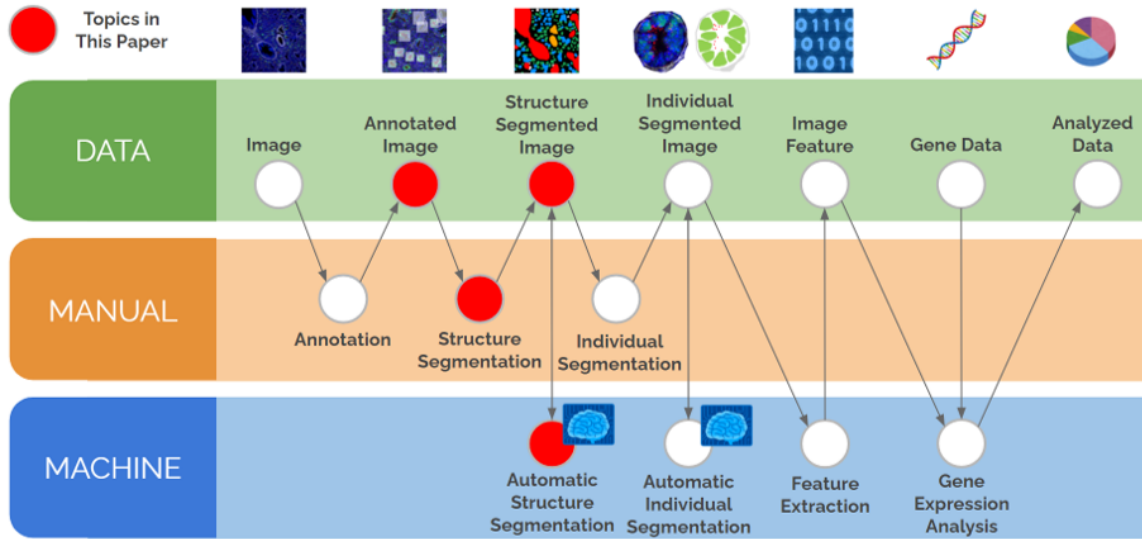
Figure 1.1: Pipeline image for the lung development analysis

on the analysis of lung development phases.

## 1.2 Research Objective

In this thesis, the author investigates obtaining accurate multi-class segmentation of lung confocal IF images using the current state-of-the-art deep learning-based model. One of the primary bottlenecks in using deep Convolutional Neural Network (CNN) models is the lack of availability of training or ground-truth segmentation labels. The author implements the multi-class segmentation with generative adversarial network (GAN) models to expand the training dataset and improve overall segmentation accuracy. To classify various lung tissue classes along with the region of interest, the author proposes to utilize recent GAN [8–10]. The author applies the method to the multi-class segmentation problem of IF images and discusses the effectiveness of created synthetic images in the segmentation.

The author organized the rest of the paper as follows. Chapter 2 introduces recent GAN models in the context of biomedical images generation. Chapter 3 mentions the experiment material for this project. Chapter 4 describes the details of our scheme, i.e., preparation of dataset and synthetic image creation with GAN. Chapter 5 provides experimental results and discusses the effectiveness of the proposed scheme from the viewpoint of multi-class segmentation. Chapter 6 concludes our work and describes the paper along with indicating future works.

# Chapter 2

# Related Works

## 2.1 Generative Adversarial Network and its Contribution

Generative Adversarial Network (GAN) is well-known for generating synthetic data close to the distribution of the training set. GAN can mimic the data distribution and synthesize images with unprecedented realism, opening up a new way to bridge the gap between supervised learning and image generation. It is worthwhile to explore effective imbalanced learning methods because imbalanced data is prevalent in many applications area in the industry, where anomaly detection is critical like electricity pilferage, fraudulent transactions in banks, identification of rare diseases, etc [11].

## 2.2 Application of GAN in Medical Field

In the field of medical imaging, many methods based GAN were applied to a variety of different modalities such as MRI, CT, OCT, chest X-Ray, Dermoscopy, Ultrasound, PET, and Microscopy to implement applications: synthesis, reconstruction, detection, denoising, registration, classification and segmentation [12]. In another research of image reconstruction in the histopathological cancer diagnosis, GAN was applied to transform Hematoxylin and Eosin (H&E) stained images into Immunohistochemistry (IHC) stained images facilitating virtual IHC staining on the same slide. This attempt will contribute to overcoming the limitation of the image analysis and cut off the staining cost [13]. For instance, X. Gong et al. discussed style consistent image generation for nuclei instance segmentation. They tried to generate H&E stained pathological images and discussed the segmentation accuracy of Mask R-CNN [14]. R. Ranjan et al. proposed a new dual deep generative method for synthesizing human cell protein images using the Generative Adversarial Network technique. Also, they evaluated the quality of generated images by the proposed method [15]. L. Hou et al. also tried to make synthetic images with GAN, and U-net-based approaches were used for segmentation [16].

These technologies, however, have not been investigated deeply for lung analysis. The main point of this thesis is how the synthesized images contribute to improving the performance of multi-class segmentation for development analysis using IF images.

# Chapter 3

# Experimental Materials

## 3.1 Materials

We collected the IF images of the mouse (Mus musculus from embryo 16.5 days to postnatal 28 days) and human (Homo sapiens from 9 months old to 4 years old) lung from the official website of LungMAP project (https://lungmap.net/). Three types of proteins are stained as red, green, and white in each IF image like Fig. 3.1. Fig. 3.1 is stained by antibodies to label specific proteins such as "Sox9", "Sftpc", and "Acta2" (for more information, see Appendix A.) In addition, these stained proteins are the markers for specific types of cells. For instance, Sox9, which is green-stained protein, marks "chondrocyte", "epithelial cell", "unclassified fibroblast", "pre-alveolar epithelial cell". The combination of stained proteins depends on each image. Furthermore, we also used the label image like Fig. 3.2, which could be classified into six tissue classes, i.e., Background (black : C1), Conductive Airway (red : C2), Distal Acinar Tubule Bud (green : C3), Proximal Acinar Tubule (blue : C4), Artery (white : C5), and Vein (orange : C6). The examples are shown in the Figure 3.3
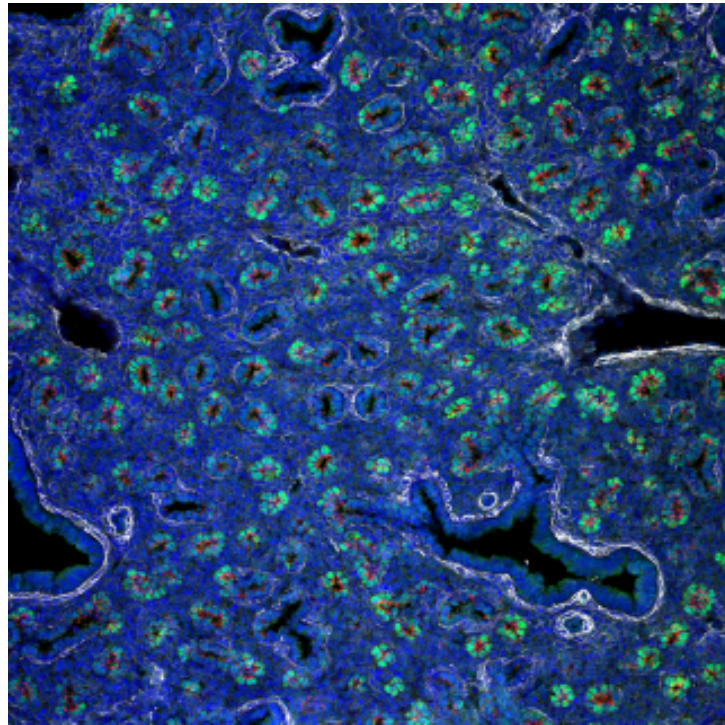
Figure 3.1: Example of lung confocal IF image
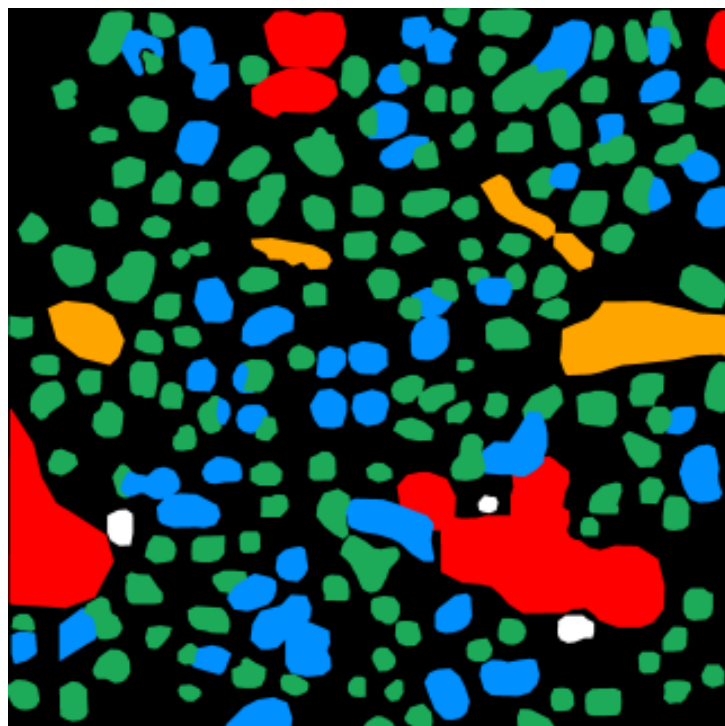


Figure 3.2: Example of labelled image
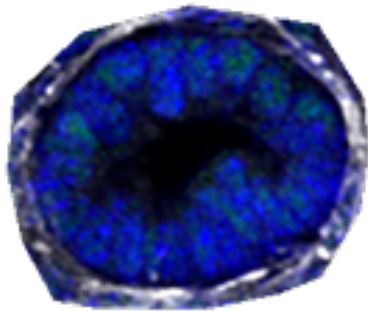
(a) Conductive Airway (C2: Red)
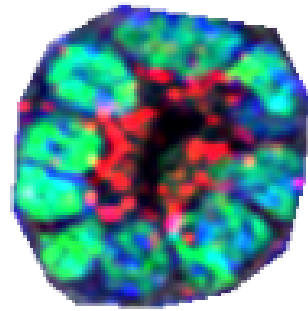


(b) Distal Acinar Tubule Bud (C3: Green)
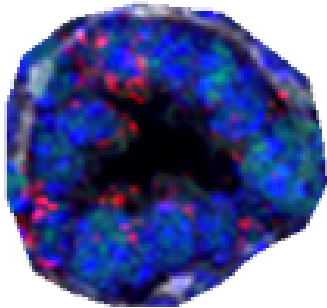


(c) Proximal Acinar Tubule (C4: Blue)
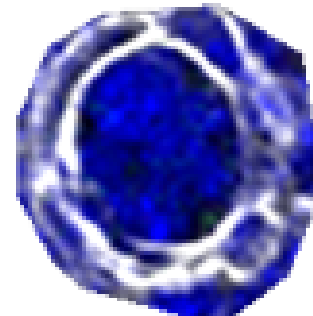


(d) Pulmonary Artery (C5: White)
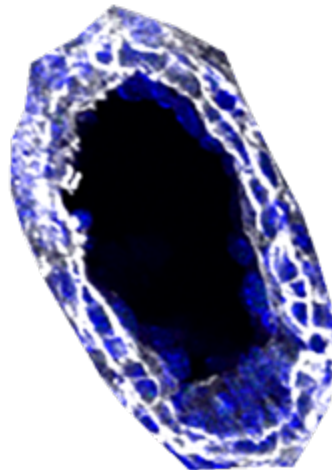


(e) Pulmonary Vein (C6: Yellow)

Figure 3.3: Examples of object

# Chapter 4

# Proposed Method

This thesis creates synthetic images with the following method, and the created images are used for expanding the learning dataset for multi-class segmentation. Figure 4.1 illustrates the scheme of our method. The method mainly consists of four processes, i.e., Preprocessing, Labeling, Image Synthesis, and Segmentation. The following sections describe the detail of each process.
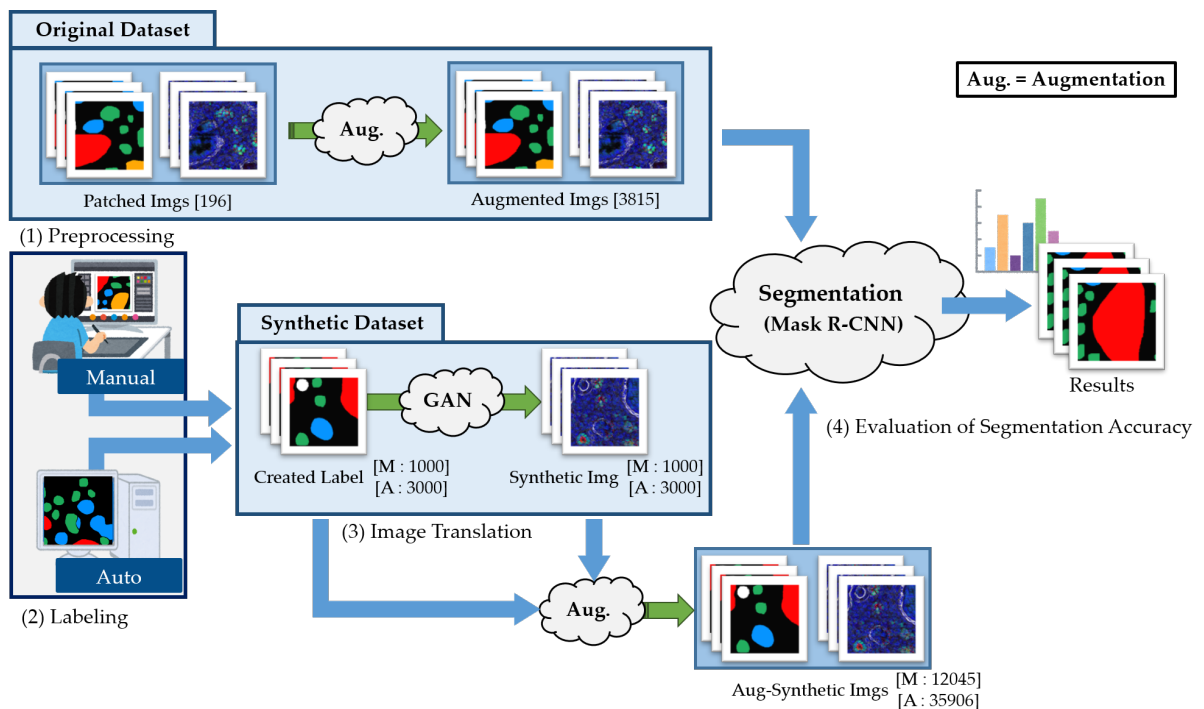


Figure 4.1: Workflow in our proposed method

## 4.1  Preprocessing

Pre-processing was applied to the IF images to prepare the baseline dataset, which is only used for the traditional augmentation. In this thesis, the original images were resized $2475 \times 2475$ [pixels] to $1024 \times 1024$ [pixels] because the size of given tissues should be matched to the patch image size for deep learning. Subsequently, the author cropped $256 \times 256$ [pixels] with 128 [pixels] stride, and applied traditional augmentation such as rotating ($0°,90°,180°,270°$), flipping, random erasing, blurring, sharpening. Consequently, 3815 images were prepared, and this dataset was used as the baseline for learning.
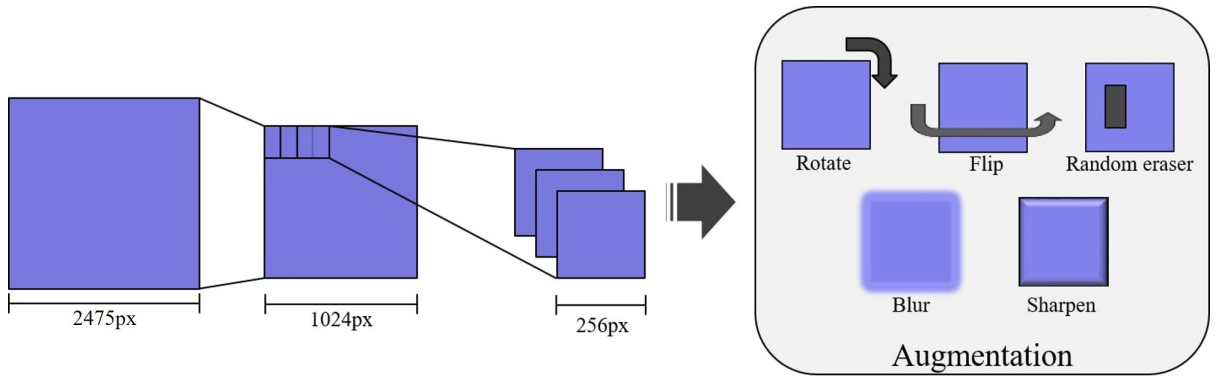


Figure 4.2: Pre-process workflow for learning models

## 4.2  Labeling

Before utilizing GAN-based image synthesis, the label image was created. Table 4.1 shows the numbers of objects in all types of creation and their average sizes before applying traditional augmentation. In the table, object C1 means background so that the number of objects and their sizes are described as "-".

### 4.2.1  Manual Labeling

In manual cration, label images that have six types of labels are manually created with a painting tool. Also, the author prepared the label images considering the positional relations of each tissue. For instance, a conductive airway, which is expressed by a red label (C2), is usually located near an artery (white label: C5). Such knowledge is very important in the labeling process because if the created label is not based on topological knowledge, a created image will not reflect features of the original image. To prevent this, the author prepared the label image based on topological information.

Table 4.1: Number of Objects and Average Size (256px $\times$ 256px)

| Object [# of Images] | Label Dataset | | | | | |
|---|---|---|---|---|---|---|
| | Original Image [196] | | Manual-Synthetic Image [1000] | | Auto-Synthetic Image [3000] | |
| | # of Obj. | Size | #of Obj. | Size | #of Obj. | Size |
| C1 | - | - | - | - | - | - |
| C2 | 177 | 9530 | 628 | 11515 | 2736 | 6577 |
| C3 | 1978 | 1123 | 6449 | 1384 | 8076 | 1040 |
| C4 | 767 | 2385 | 4108 | 1779 | 6868 | 1219 |
| C5 | 60 | 843 | 657 | 1071 | 1856 | 855 |
| C6 | 88 | 4739 | 295 | 5891 | 926 | 3527 |

### 4.2.2 Automatic Labeling

In automatic creation, we used Diffusion Probabilistic Model (hereinafter called "diffusion models") for automatic label creation. This model is a parameterized Markov chain trained using variational inference to produce samples matching the data after finite time. Transitions of this chain are learned to reverse a diffusion process showing in Figure 4.3. That is, a Markov chain that gradually adds noise to the data in the opposite direction of sampling until signal is destroyed. When the diffusion consists of small amounts of Gaussian noise, it is sufficient to set the sampling chain transitions to conditional Gaussians too, allowing for a particularly simple neural network parameterization [17].
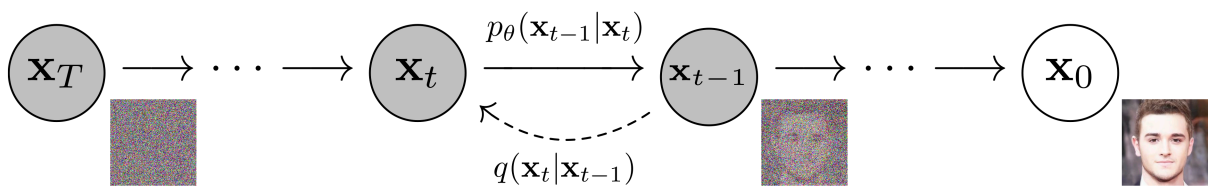


Figure 4.3: Learning process of the diffusion model

It is difficult for a generative model to have simplicity and flexibility properties. The model needs to be able to create a variety of features. The created label patterns must not be similar to the original patterns. So, GAN was unsuitable in our situations. GAN will generate the image only to deceive the discriminator even if the same pattern is generated .

On the other hand, the diffusion model can generate a wide variety of patterns if given

enough learning costs. Therefore, the author used the diffusion models for automatic label creation. The model was trained by the labels only including the few ratio tissues (C5 and C6) to generate few data and moderate the imbalanced data.

## 4.3 Image Translation Using GAN

We utilized Pix2Pix, Pix2PixHD, GauGAN for image translation [8–10]. These models can be trained with all images prepared in section 4.1 and generate a synthetic image from a given label. We focused on this operation and considered the synthetic image and the label as a pair image set for the segmentation dataset. The pair of images is applied to the segmentation dataset as a ground truth and training image.

**Pix2Pix** has a generator that uses skip connections, and its discriminator is designed to look at the patches rather than the entire image to distinguish real from synthetic images [8]. The main architecture of Pix2Pix is shown in Figure 4.4. This helps the generator to create high-level details associated with real data distribution.

**Pix2PixHD** is a method based on the Pix2Pix. This architecture stabilizes learning at high resolution and enables the generation of details. Figure 4.5 shows the generator of Pix2PixHD. Combining G1 (is called as 'global generator') and G2 (is called as local enhancer) enables to create different resolution image. On the other hand, the discriminator consists of different scale network for each generator. These architecture helps to improve the photorealism and resolution. Moreover, it embedded the instance-level object semantic information to improve the image quality [9].

**GauGAN** is a state-of-the-art image synthesis model that focuses on converting segmentation masks to photorealistic images. The architecture of GauGAN is shown in Figure 4.6. It can better preserve semantic information against common normalization layers to propose a new convolutional normalization method called spatially-adaptive denormalization (SPADE) [10]. Thus, this normalization layer tends to keep the information contained in the semantic input masks.

It is easily expected that the quality of synthetic images generated by newer generation models is higher than other models because these models have effective architectures for image translation. However, the author also aims to get any clue for an effective data augmentation approach by comparing different models and conditions.
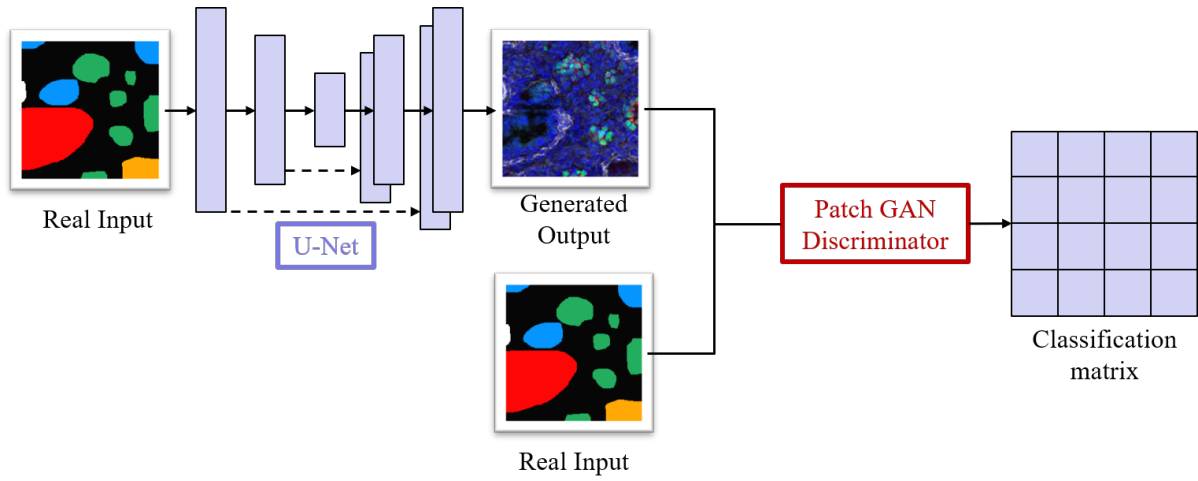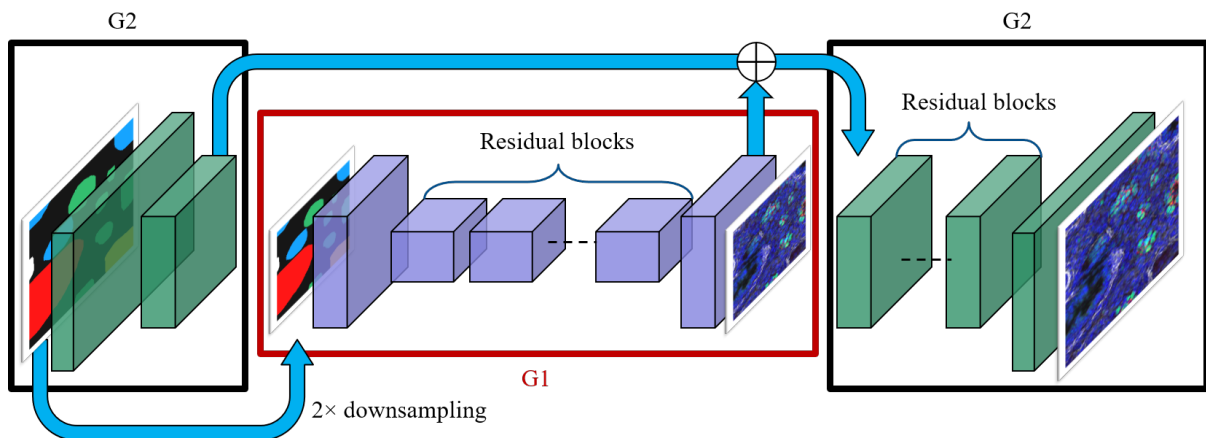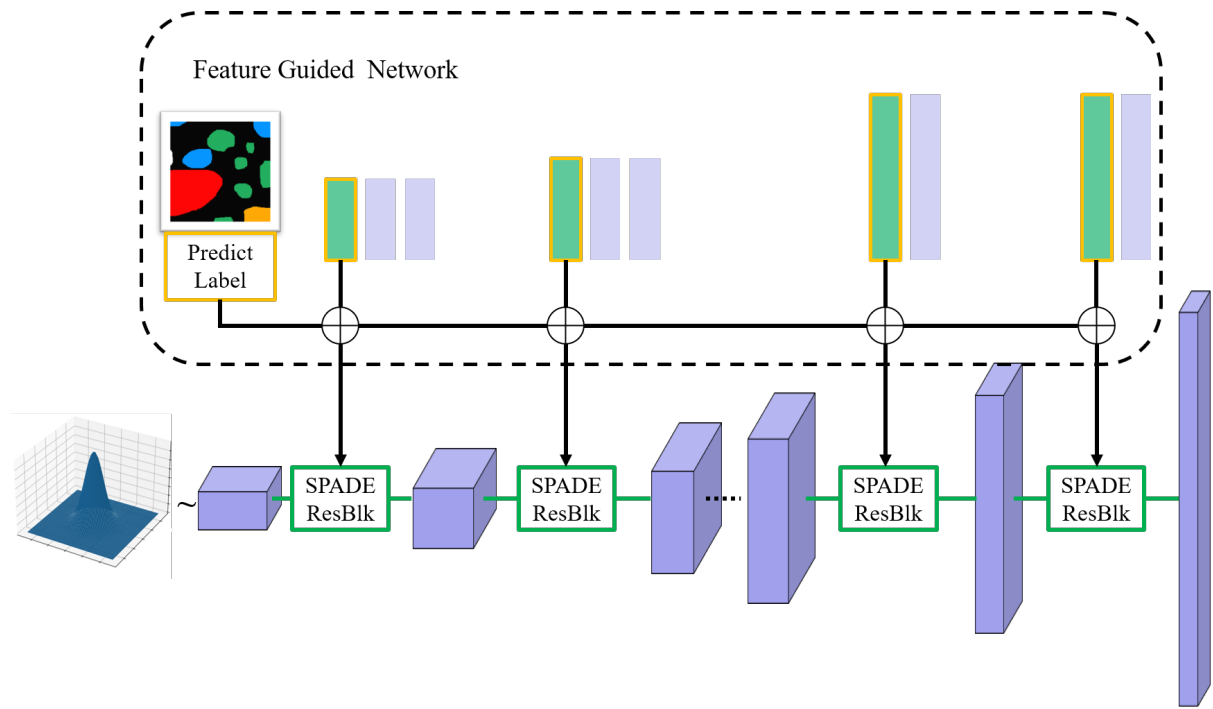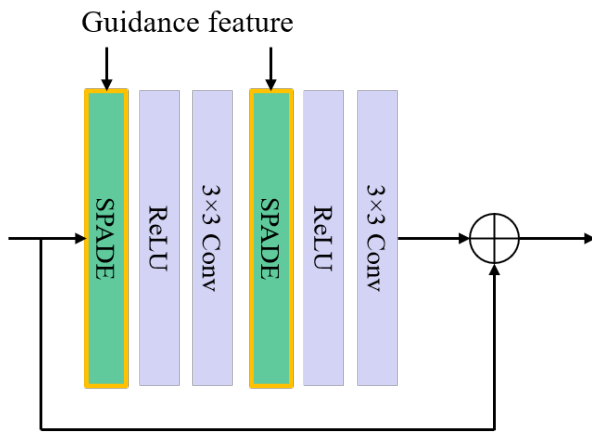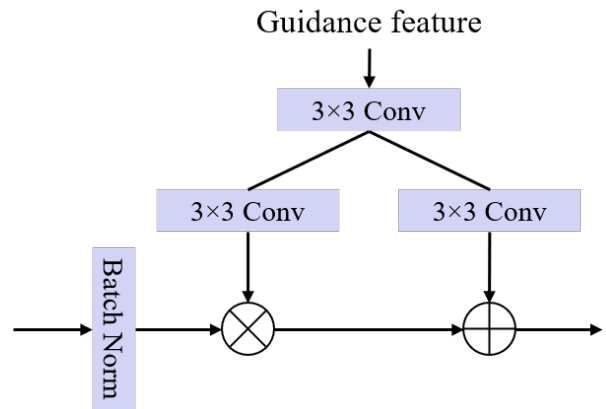
Figure 4.4: The architecture of Pix2Pix



Figure 4.5: The architecture of Pix2PixHD

(a) Generator network in GauGAN



(b) SPADE Residual Brock



(c) Detail of SPADE filter

Figure 4.6: The architecture of GauGAN

## 4.4 Evaluation of Segmentation Accuracy

In this scheme, Mask R-CNN, which was one of the well-known segmentation models in the field of medical image segmentation, was used as a segmentation model [18]. The author implemented Mask R-CNN and tuned the parameters of the model by changing the number of synthetic images.

Before the evaluation experiments, the author prepared 1000 synthetic images for segmentation detailed in Table 4.1. Traditional augmentation was also applied to the synthetic images. Consequently, 3815 original images, 12046 manual-synthetic and 35894 automatic-synthetic images were used for the experiments. The augmented data conditions are shown in Table 4.2.

In the learning process, the author conducted the 4-fold cross-validation to divide only the original datasets. Divided one was used as a validation dataset, and the others were used as train datasets detailed in Table 4.2. The datasets did not make the class ratio uniform because each dataset was created from each image. To compare the experimental conditions, the author conducted the learning model in three steps. First, manual-synthetic images add to the training dataset gradually. Second, each automatic-synthetic images also add to the training dataset. Finally, all synthetic images used for training.

In the test process, the test dataset was prepared before image patching and augmentation, and its condition is shown in Table 4.3. The segmentation model predicts an image whose size is $256 \times 256$ [pixels]. As shown in Figure 4.7, image prediction is conducted with over-wrapped every 64 [pixels] for deciding by a majority results. Moreover, predicted images deleted edge prediction (32 [pixel]) because the prediction edge tends to be misrecognized. Finally, the final prediction determine the most possibly class in each pixels.

The author evaluated segmentation performance with the dice coefficient score for class $c$ ($D_c$), which was defined by

$$D_c = \frac{2\sum_x \sum_y |\mathbf{C}_{output_{xy}} \cdot \mathbf{C}_{gt_{xy}} \cdot \mathbf{U}_c|}{\sum_x \sum_y |\mathbf{C}_{output_{xy}} \cdot \mathbf{U}_c| + \sum_x \sum_y |\mathbf{C}_{gt_{xy}} \cdot \mathbf{U}_c|} \tag{4.1}$$

where $\mathbf{C}_{output_{xy}}$ was predicted class vector at any pixel $(x, y)$ in the image, and $\mathbf{C}_{gt_{xy}}$ was ground truth class vector as a one-hot vector. In the above equation, $\mathbf{U}_c$ means the one-hot vector for class $c$, and it is defined by the following equation.

$$\mathbf{U}_c = (U_1, ..., U_N) \text{ where } U_i = \begin{cases} 0 & (i \neq c) \\ 1 & (i = c), \end{cases} c = 1, 2, ..., N. \tag{4.2}$$

Table 4.2: Number of objects and pixels in Train Dataset (256px × 256px)

| Original Dataset [3815] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Obj. [#] | Dataset-1 [914] | | Dataset-2 [992] | | Dataset-3 [929] | | Dataset-4 [980] |
| | # of Obj. | Pixels | # of Obj, | Pixels | # of Obj. | Pixels | # of Obj, | Pixels |
| C1 | - | 28.8M | - | 34.5M | - | 36.4M | - | 33.0M |
| C2 | 1140 | 13.3M | 807 | 4.7M | 603 | 4.6M | 886 | 9.3M |
| C3 | 7029 | 7.6M | 11386 | 12.3M | 10617 | 11.3M | 9402 | 10.8M |
| C4 | 3538 | 7.5M | 4055 | 12.4M | 3649 | 6.5M | 3621 | 8.1M |
| C5 | 385 | 256k | 260 | 233k | 221 | 263k | 281 | 182k |
| C6 | 352 | 2.1M | 332 | 840k | 459 | 178k | 648 | 2.8M |

| Synthetic Dataset [47940] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Obj. [#] | Manual [12046] | | Automatic-1 [11920] | | Automatic-2 [12017] | | Automatic-3 [11957] |
| | # of Obj. | Pixels | # of Obj, | Pixels | # of Obj. | Pixels | # of Obj, | Pixels |
| C1 | – | 492.1M | - | 631.7M | - | 637.3M | - | 632.2M |
| C2 | 7683 | 83.1M | 11181 | 69.1M | 10878 | 68.4M | 10884 | 69.1M |
| C3 | 76901 | 102.6M | 30519 | 30.5M | 32060 | 31.9M | 32301 | 33.0M |
| C4 | 49074 | 83.5M | 27349 | 31.9M | 27310 | 31.8M | 26164 | 31.2M |
| C5 | 7838 | 8.1M | 7208 | 5.9M | 7343 | 6.1M | 7188 | 5.9M |
| C6 | 3606 | 19.9M | 3522 | 12.1M | 3719 | 12.8M | 3935 | 12.9M |

Table 4.3: Number of objects and pixels in Test Dataset (1024px × 1024px)

| Obj. | Dataset-1 | | Dataset-2 | | Dataset-3 | | Dataset-4 | |
|------|-----------|--------|-----------|--------|-----------|--------|-----------|--------|
| | # of Obj. | Pixels | # of Obj, | Pixels | # of Obj. | Pixels | # of Obj, | Pixels |
| C1 | - | 473k | - | 530k | - | 614k | - | 535k |
| C2 | 7 | 261k | 10 | 103k | 4 | 84k | 5 | 144k |
| C3 | 91 | 131k | 130 | 200k | 136 | 199k | 113 | 186k |
| C4 | 43 | 150k | 37 | 200k | 43 | 114k | 41 | 138k |
| C5 | 7 | 6.6k | 4 | 5.0k | 3 | 3.8k | 4 | 30k |
| C6 | 2 | 28k | 2 | 9.5k | 4 | 34k | 4 | 43k |

In the equation (4.2), $c$ means the class, $N$ is the number of classes. In the testing process, the only original images were used for the evaluation. The distribution of objects in each testing dataset is shown in Table 4.3.
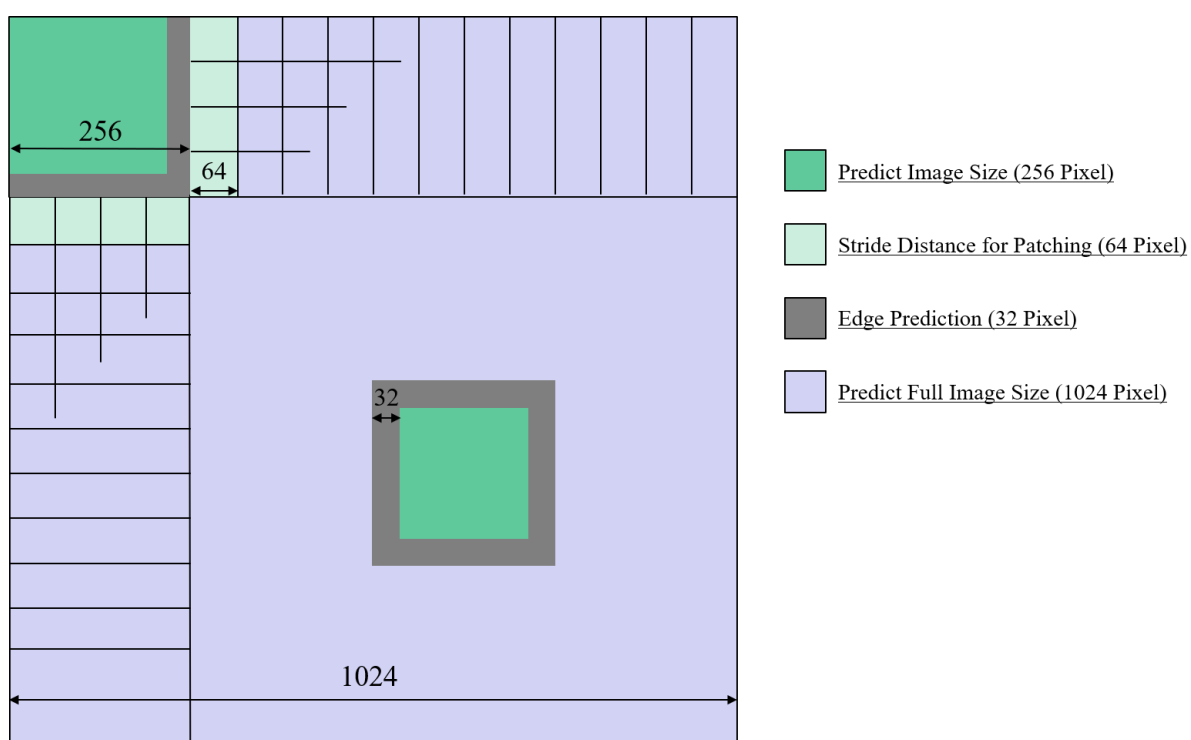
**Predict Image Size (256 Pixel)**

**Stride Distance for Patching (64 Pixel)**

**Edge Prediction (32 Pixel)**

**Predict Full Image Size (1024 Pixel)**

Figure 4.7: Each area of the image prediction

# Chapter 5

# Experimental Results and Discussion

The GAN models used in this paper were trained with the original dataset and applied for the image translation from the label images. The each hyper parameter for learning was shown in Table 5.1. Some columns explain omitted forms, i.e., 'batch' means batch size, 'lrate' means learning rate, and 'opt' means optimizer. Almost parameter were not changed in the public domain system described in [8–10,18]. Also, the batch of GauGAN and Mask R-CNN is restricted by the machine performance the author used.

## 5.1 Manual Labeling Augmentation

### 5.1.1 Created Synthetic IF Images by GANs

In the experiment, the author created all synthetic images by using each GAN model. Figure. 5.1 shows examples of created synthetic images by GAN models and label images. These results show that all GAN models could create synthetic IF images from the given label images, and the created images reflect features of the original images well.

In the case of Pix2Pix, the created images lost focus compared to others, i.e., GauGAN and Pix2PixHD. Pix2Pix is an old model compared to others, and its architecture does not support high resolution. This is why the results were obtained. However, the created IF images look like they have enough capability to be used as a dataset not only for segmentation but also for image classification.

### 5.1.2 Evaluation of Segmentation Accuracy with Manual Datasets

Table 5.2 and Figure 5.2 show the segmentation accuracy of Mask R-CNN. This table summarizes the dice coefficient score for each tissue when the number of added synthetic images are 0 (i.e., only original dataset), 6000. 9000, and all manual-synthetic images, respectively. The rightmost column means the average of segmentation performance in each condition. The obtained results indicate that the accuracy of six class segmentation

Table 5.1: Each model's hyper parameter for learning

| Model | batch | lrate | momentum | epoch | opt | loss |
|---|---|---|---|---|---|---|
| Pix2Pix | 8 | 0.002 | beta1:0.9 beta2: 0.999 | 990 | adam | BCE |
| Pix2PixHD | 8 | 0.0002 | beta1: 0.5 beta2: - | 200 | adam | BCE |
| GauGAN | 1 | 0.0002 | beta1: 0.0 beta2: 0.9 | 50 | adam | CE |
| Mask R-CNN | 2 | 0.001 | beta1: 0.9 beta2: - | 50 | SGD | CE, BCE |

Table 5.2: Comparison with segmentation accuracy using manual datasets

| GAN | Training Set | C1 | C2 | C3 | C4 | C5 | C6 | Average |
|---|---|---|---|---|---|---|---|---|
| | Original | 0.837 | 0.717 | 0.809 | 0.625 | 0.104 | 0.642 | 0.622 |
| Pix2Pix | 6000 | 0.862 | 0.792 | 0.806 | 0.737 | 0.527 | 0.784 | 0.752 |
| | 9000 | 0.869 | 0.772 | 0.817 | 0.684 | 0.594 | 0.843 | 0.763 |
| | All | 0.866 | 0.809 | 0.811 | 0.710 | 0.652 | 0.789 | **0.773** |
| Pix2PixHD | 6000 | 0.872 | 0.781 | 0.815 | 0.716 | 0.625 | 0.707 | 0.753 |
| | 9000 | 0.870 | 0.783 | 0.809 | 0.685 | 0.629 | 0.762 | 0.756 |
| | All | 0.872 | 0.779 | 0.816 | 0.689 | 0.626 | 0.805 | **0.765** |
| GauGAN | 6000 | 0.871 | 0.796 | 0.809 | 0.720 | 0.614 | 0.785 | 0.766 |
| | 9000 | 0.863 | 0.807 | 0.804 | 0.704 | 0.550 | 0.798 | 0.754 |
| | All | 0.867 | 0.833 | 0.804 | 0.708 | 0.632 | 0.765 | **0.768** |

using Mask R-CNN was improved in all cases. In particular, the segmentation accuracy was increased by 15.1% when 1000 additional synthetic images were used for learning of Pix2Pix. As mentioned before, the synthetic images created by Pix2Pix lost focus compared to others, but these images also contributed to improving the segmentation accuracy of Mask R-CNN. In the field of medical imaging, it is generally difficult to collect plenty of data for machine learning. In particular, deep learning requires several thousands of images for accurate recognition. By using this approach, the author does not need to collect plenty of data and only prepare label images considering features of the baseline dataset.

Figure 5.3 and Table 5.3 show examples of failures in the segmentation process. In the figure, these images mean the original IF images, ground truth, and segmentation result from the first-tier image. In these cases, Mask R-CNN could not segment red (conductive airway: C2) and orange (vein: C6) objects appropriately. Both objects are a

Table 5.3: Confusion matrices in fail cases

(a) Fail Case 1

|  | C1 | C2 | C3 | C4 | C5 | C6 | Sum of Pred |
|---|---|---|---|---|---|---|---|
| **C1** | 433213 | 86296 | 26443 | 31059 | 2035 | 4077 | 563123 |
| **C2** | 3871 | 183311 | 0 | 10901 | 193 | 0 | 198276 |
| **C3** | 16979 | 0 | 101199 | 9200 | 26 | 0 | 127404 |
| **C4** | 18149 | 2586 | 3195 | 98642 | 0 | 0 | 122572 |
| **C5** | 284 | 245 | 0 | 0 | 4330 | 0 | 4759 |
| **C6** | 565 | 8082 | 0 | 0 | 0 | 23975 | 32442 |
| **Sum of GT** | 472961 | 280520 | 130837 | 149802 | 6584 | 27872 | 1048576 |

(b) Fail Case 2

|  | C1 | C2 | C3 | C4 | C5 | C6 | Sum of Pred |
|---|---|---|---|---|---|---|---|
| **C1** | 456470 | 20981 | 26839 | 27755 | 349 | 12607 | 545001 |
| **C2** | 20945 | 116653 | 0 | 4557 | 152 | 0 | 142307 |
| **C3** | 20883 | 0 | 151074 | 17362 | 0 | 0 | 189319 |
| **C4** | 24583 | 6280 | 7605 | 88683 | 621 | 0 | 127772 |
| **C5** | 881 | 0 | 0 | 0 | 1918 | 0 | 2799 |
| **C6** | 11100 | 128 | 0 | 0 | 0 | 30150 | 41378 |
| **Sum of GT** | 534862 | 144042 | 185518 | 138357 | 3040 | 42527 | 1048576 |

type of muscle fibers and their feature, e.g., shapes and distributions of pixels, is similar to each other. Also, in Table 5.3, the author can indicate the same considerations above. The table shows that Mask R-CNN still failed to determine the object's boundary ideally. This is one of the reasons for such experimental results being obtained. It is, however, a heavy task for us to prepare thousands of label images manually. The author needs to use the automatic process to create the labels. The following section tried to extend the research and show the conducted experiment's results.

## 5.2 Automatic Labeling Augmentation

### 5.2.1 Automatic Labeling with Diffusion Model

In the automatic creation, the diffusion models create the label images, and the detail of parameters is shown in Table 5.4. The model was learned by labels including white (pulmonary artery: C5) or yellow (pulmonary vein: C6) in the original and manual label datasets. The generation results are shown in Figure 5.4. The diffusion model created the label images to moderate imbalanced data. The author also applied the GAN models to translate the synthetic images from these created labels. The synthesized datasets were divided into three ones randomly for observing the average effects.

Table 5.4: Diffusion model's parameter

| | |
|---|---|
| Batch size | 4 |
| Learning rate | $2 \times 10^{-5}$ |
| Number of steps | 500k |
| Gradient accumulation steps | 2 |
| Exponential moving average decay | 0.995 |

### 5.2.2 Evaluation of Segmentation Accuracy with Automatic Datasets

This experiment was conducted to add the automatic-synthetic dataset to the manual-synthetic dataset. Table 5.5 and Figure 5.5 show the segmentation accuracy of added automatic synthetic datasets. This table summarizes the dice coefficient score for each tissue when the added datasets are Manual (i.e., the best accuracy in Table 5.2), Auto-1, Auto-2, Auto-3, and all synthetic images, respectively. The obtained results indicate that Pix2Pix's accuracy was also the best, improving 1.2% from the accuracy using Manual dataset. In the Pix2Pix's rows, the automatic-synthetic datasets effectively improved segmentation accuracy. However, other models' accuracy were not improved using automatic-synthetic datasets; nevertheless, the author considered the synthetic image quality was better than Pix2Pix's. This reason is that the Mask R-CNN cannot extract detailed features from the IF images but only use outlines such as general shape and color. High-resolution synthetic images might make it complicated for Mask R-CNN. In addition, each accuracy in 'All' rows indicates the worst in each model's accuracy. The author considers that the number of additional images exceeded the saturation point of improving accuracy.

In the next trials, the author refines the number of additional images to determine

the limitation of the Mask R-CNN and our methods. In another method, other segmentation models are necessary to extract high-resolution synthetic features. Also, if the model changes, the author needs to pursue the experimental conditions for other model performances.

In terms of the data, the given image was imbalanced, e.g., the number of C5 and C6 objects is small compared to others. Hence, an architecture for imbalanced data should also be involved in the improved method. In addition, the error of 1% corresponds to 10 pixels in the case of LungMAP IF images. If high-resolution images are used for image analysis, this small difference may cause misrecognition. At the moment, nobody knows the reasons and bottleneck of the proposed method. We have to investigate what contributed to image segmentation.

Table 5.5: Comparison with segmentation accuracy using automatic datasets

| GAN | Training Set | C1 | C2 | C3 | C4 | C5 | C6 | Average |
|---|---|---|---|---|---|---|---|---|
| Pix2Pix | Manual | 0.866 | 0.809 | 0.811 | 0.710 | 0.652 | 0.789 | 0.773 |
| | Auto-1 | 0.872 | 0.803 | 0.819 | 0.709 | 0.621 | 0.819 | 0.774 |
| | Auto-2 | 0.873 | 0.791 | 0.810 | 0.714 | 0.670 | 0.802 | 0.777 |
| | Auto-3 | 0.871 | 0.810 | 0.814 | 0.725 | 0.656 | 0.835 | **0.785** |
| | All | 0.869 | 0.797 | 0.800 | 0.666 | 0.631 | 0.738 | 0.750 |
| Pix2PixHD | Manual | 0.872 | 0.779 | 0.816 | 0.689 | 0.626 | 0.805 | **0.765** |
| | Auto-1 | 0.869 | 0.786 | 0.809 | 0.700 | 0.583 | 0.650 | 0.733 |
| | Auto-2 | 0.872 | 0.814 | 0.801 | 0.707 | 0.613 | 0.735 | 0.757 |
| | Auto-3 | 0.871 | 0.804 | 0.810 | 0.708 | 0.643 | 0.613 | 0.742 |
| | All | 0.858 | 0.759 | 0.799 | 0.667 | 0.507 | 0.613 | 0.700 |
| GauGAN | Manual | 0.867 | 0.833 | 0.804 | 0.708 | 0.632 | 0.765 | **0.768** |
| | Auto-1 | 0.868 | 0.818 | 0.792 | 0.689 | 0.637 | 0.758 | 0.760 |
| | Auto-2 | 0.854 | 0.739 | 0.791 | 0.663 | 0.637 | 0.763 | 0.741 |
| | Auto-3 | 0.871 | 0.825 | 0.796 | 0.698 | 0.644 | 0.737 | 0.762 |
| | All | 0.850 | 0.708 | 0.800 | 0.613 | 0.700 | 0.732 | 0.734 |

(a) GauGAN      (b) Pix2PixHD      (c) Pix2Pix      (d) Label Image
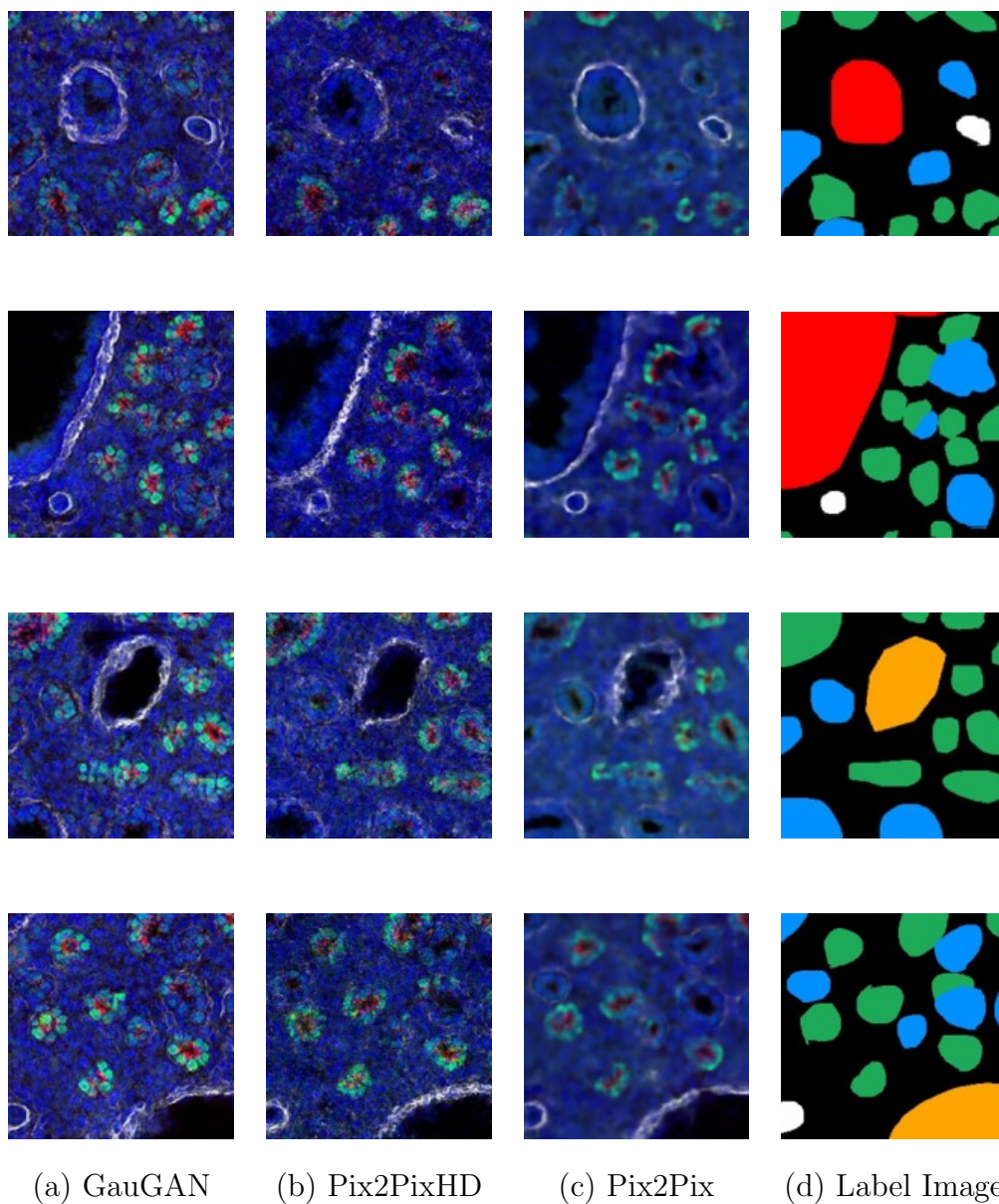
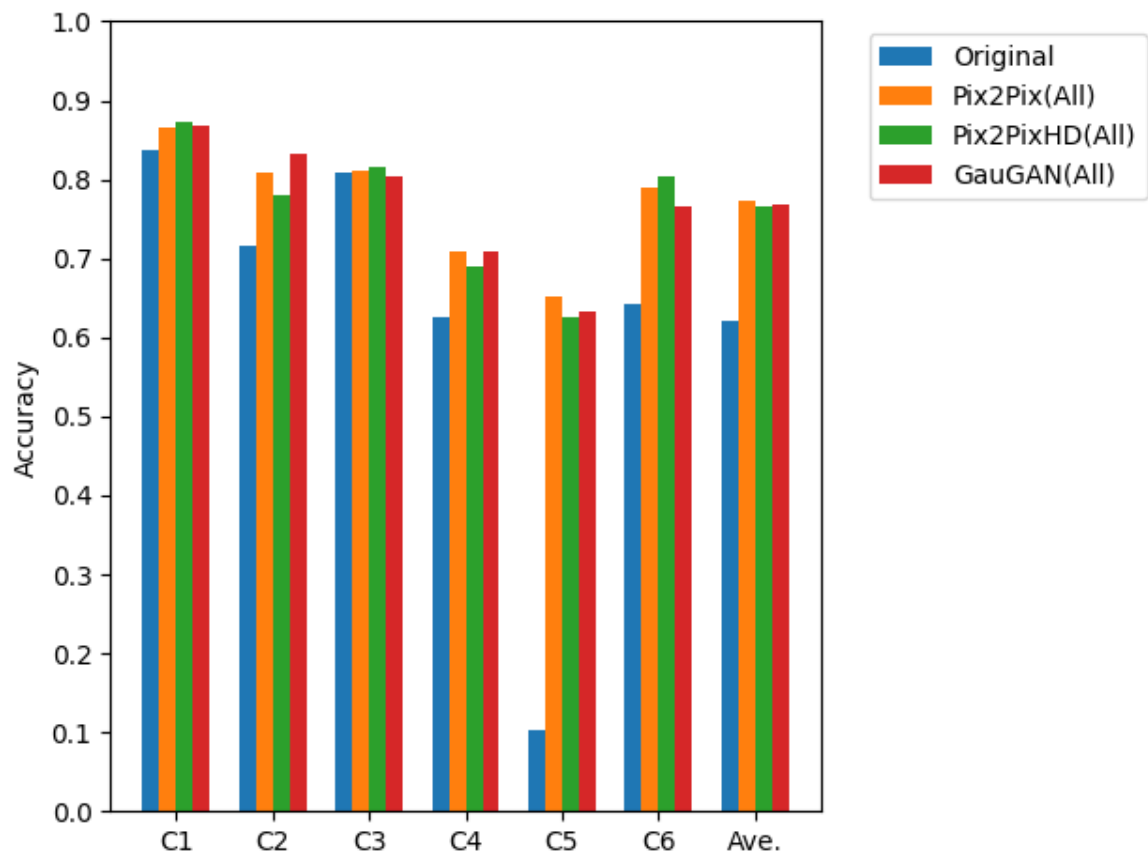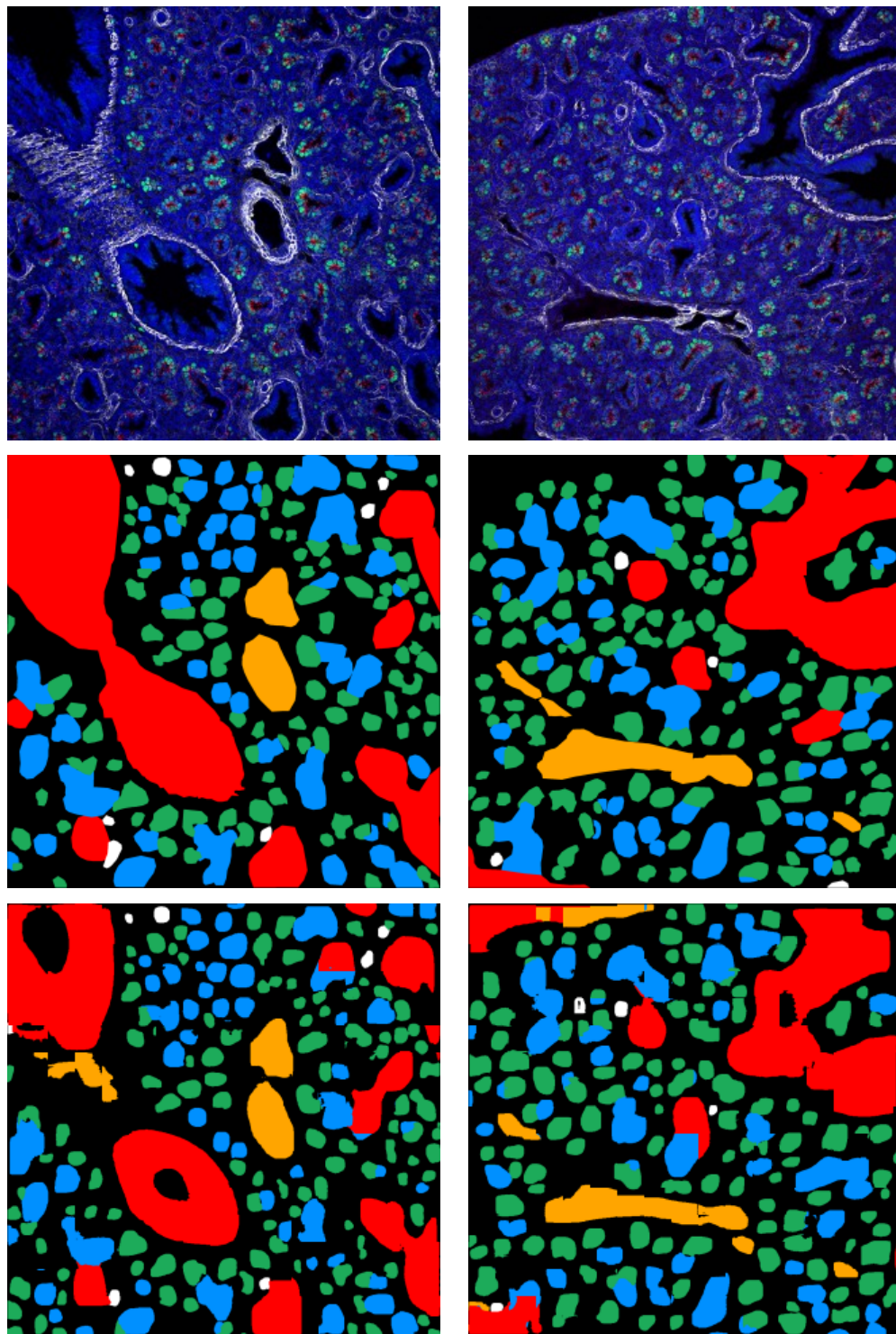Figure 5.1: Examples of synthetic images the author created

Figure 5.2: Accuracy of each best model's score with manual datasets

(a) Case 1

(b) Fail Case 2

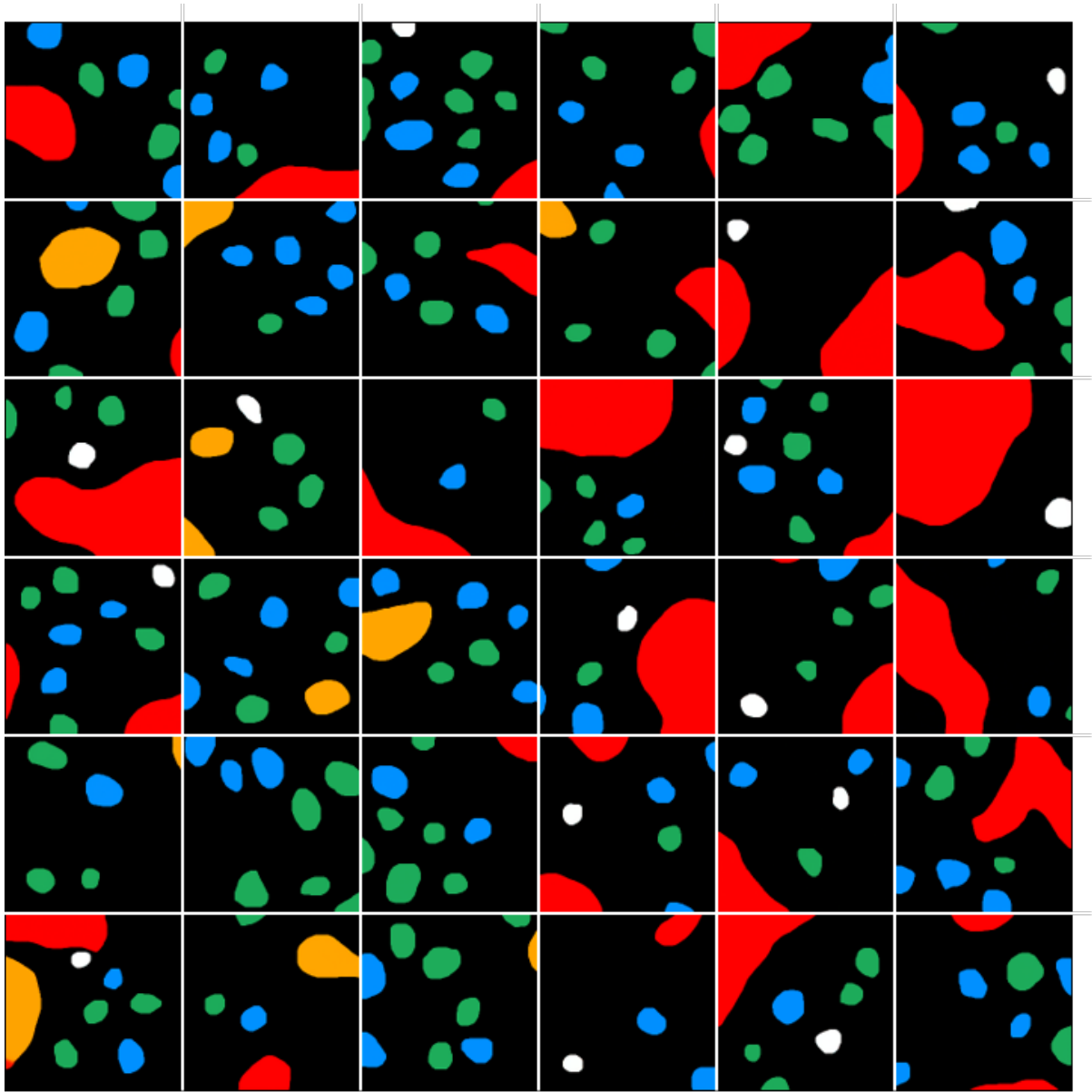Figure 5.3: Examples of failures in segmentation with our approach

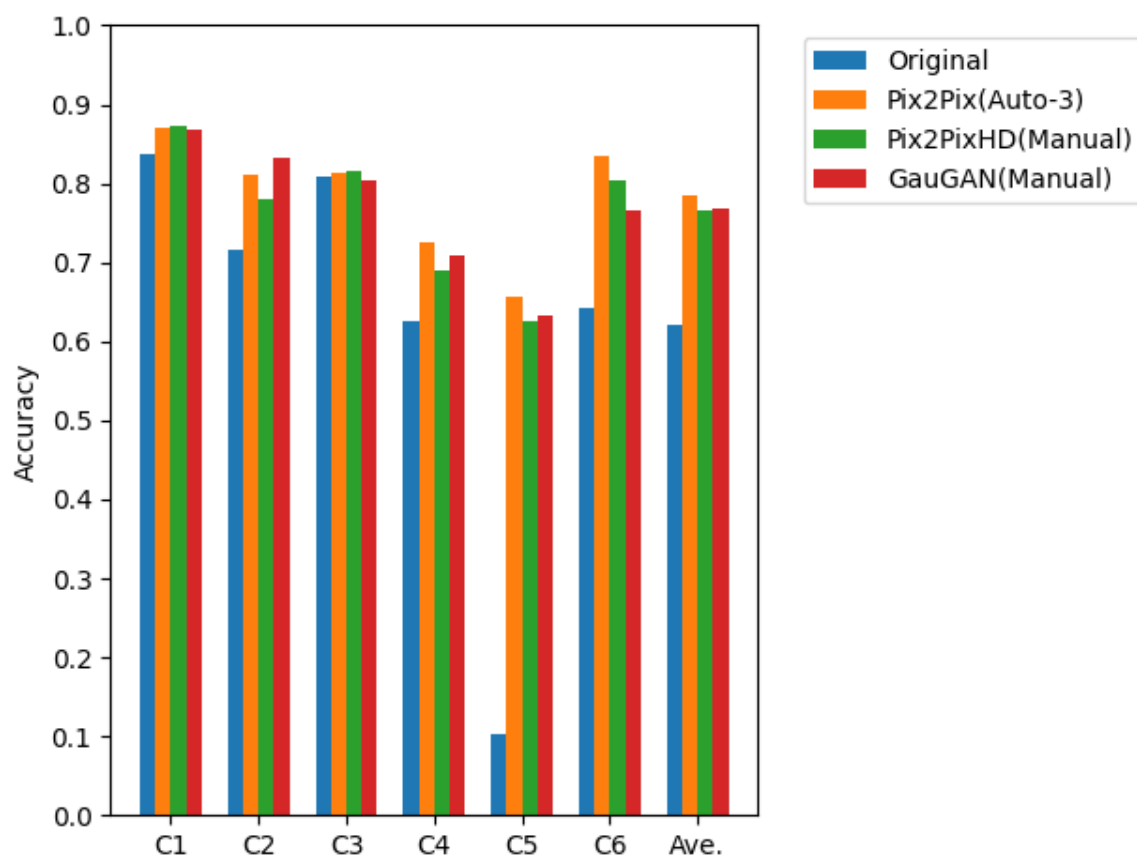Figure 5.4: The created label samples using diffusion model

Figure 5.5: Accuracy of each model's best score with automatic datasets

# Chapter 6

# Conclusion

## 6.1 Conclusion

In this paper, the author investigated obtaining accurate multi-class segmentation of lung confocal IF images using the current state-of-the-art deep learning-based model. One of the primary bottlenecks in using deep convolutional neural network (CNN) models was the lack of availability of training or ground-truth segmentation labels. The author implemented Mask R-CNN with GAN models to expand the training dataset and improve overall segmentation accuracy. To classify various lung tissue classes along with the region of interest, the author proposed to utilize recent GAN based-models, such as Pix2Pix, Pix2PixHD, and GauGAN. The author applied the method to the multi-class segmentation problem of lung IF images, and the accuracy of six-class segmentation using Mask R-CNN was improved in all cases. In particular, the segmentation accuracy increased by 16.3% when the segmentation model trained the manual-dataset and the automatic-dataset by synthesized Pix2Pix.

## 6.2 Future Works

In the discussion, the author described four contents in future works. First, it is necessary to attempt how are the best conditions, e.g., pursuing the saturation point of Mask R-CNN learning. Second, the author suspects that the Mask R-CNN cannot obtain detailed features because the best accuracy was obtained by synthesized Pix2Pix, which is the vaguest of these generation results. Third, the given data is imbalanced, in particular, the number of C5 and C6 objects are small compared to others. Robust architectures for handling imbalanced data should also be involved to improve the classification results. Finally, to discover the detailed reasons and bottlenecks, the author has to investigate what contributed to image segmentation. These future works are necessary to proceed with the LungMAP project and analyze lung development.

# Acknowledgement

# Publication List

## Journal Papers

(1) Daiki Katsuma, Hiroharu Kawanaka, V. B. Surya Prasath, and Bruce J. Aronow, "Data Augmentation Using Generative Adversarial Networks for Multi-class Segmentation of Lung Confocal IF Images," *Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII)*, Vol. 26, No. 2, 2022 (in Print)

## International Conferences

(1) Daiki Katsuma, Shu Isaka, Hiroharu Kawanaka, Bruce J. Aronow, and V. B. Surya Prasath, "On Data Augmentation Techniques for Deep Learning Multi-class Segmentation of Lung Confocal Immunofluorescent Images", *Proc. of Joint 2020 9th International Conference on Informatics, Electronics & Vision (ICIEV) & 2nd International Conference on Imaging, Vision & 4th Pattern Recognition (icIVPR)*, PID-152,2020.

(2) Daiki Katsuma, Hiroharu Kawanaka, V. B. Surya Prasath, and Bruce J. Aronow, A Study on Data Augmentation Using GAN for Lung Image Segmentation, *Proc. of the 10th International Symposium for Sustainability by Engineering at Mie University (Research Area C)*, pp. 19-20, 2020

(3) Daiki Katsuma, Hiroharu Kawanaka, Bruce J. Aronow, and V. B. Surya Prasath, The Effects of Augmentation Using GAN for Confocal Immunofluorescence Image Segmentation, *Proc. of Work-in-Progress session in 10th International Conference on Informatics, Elec-tronics & Vision (ICIEV2021)*, PID-88, 2021 (WIP Award)

(4) Daiki Katsuma, Hiroharu Kawanaka, V. B. Surya Prasath, and Bruce J. Aronow, "Generative Adversarial Network driven Synthetic Augmentation of Confocal Immunofluorescence Image Segmentation," *The 2021 5th IEEE International Conference on Cybernetics (CYBCONF)*, 2M4-CC-04, 2021

(5) Daiki Katsuma, Hiroharu Kawanaka, V. B. Surya Prasath, and Bruce J. Aronow, "Effects of Augmentation with GAN and Diffusion Probabilistic Models for Immunofluorescence Image Segmentation," *Proc. of the 11th International Symposium for Sustainability by Engineering at Mie University (Research Area C)*, pp. 63-64, 2021

# Domestic Conferences

(1) 勝間大喜, 川中普晴, V. B. Surya Prasath, Bruce J. Arronow, ”肺組織の免疫蛍光染色画像を対象とした Pix2Pix によるデータオーギュメンテーションに関する一検討 (A Study on Data Augmentation Method Using Pix2Pix for Lung Immnunofluorescence-Confocal Images),” 令和二年度電気・電子・情報学会東海支部連合大会講演論文集, E6-4, 2020

(2) 勝間大喜, 川中普晴, V. B. Surya Prasath, Bruce J. Arronow, ”免疫蛍光染色画像解析のための GAN による データセットの拡張方法に関する一考察,” 2020 年度日本生体医工学会東海支部大会抄録集, pp. 10, 2020

(3) 勝間大喜, 川中普晴, V. B. Surya Prasath, Bruce J. Arronow, ”肺の免疫蛍光染色画像を用いた拡散過程モデルと GAN を組み合わせたデータオーギュメンテーション法に関する一考察 (A Study on Data Augmentation Method Using Diffusion Probability Model and GAN for Lung Immunofluorescence-Confocal Images),” 令和三年度電気・電子・情報学会東海支部連合大会講演論文集, E4-4, 2021

# Reference

[1] N. Howlader, A. Noone, M. Krapcho, J. Garshell, and D. Miller, *et al.*, "Seercancer statistics review," 1975-2012, national Cancer Institute. Bethesda,MD. Available from: https://seer.cancer.gov/index.html. [accessed July 16, 2021].

[2] M. Herriges and E. Morrisey, "Lung development: orchestrating the generation and regeneration of a complex organ," *Development*, vol. 141, pp. 502–513, 2014.

[3] LungMAP, "Official web site," retrieved September 01, 2021, from: https://lungmap.net/.

[4] M. Ardini-Poleske, R. Clark, C. Ansong, J. Carson, R. Corley, G. Deutsch, J. Hagood, N. Kaminski, T. Mariani, S. Potter, G. Pryhuber, D. Warburton, J. Whitsett, S. Palmer, and N. Ambalavanan, "Lungmap: The molecular atlas of lung development program," *American Journal of Physiology - Lung Cellular and Molecular Physiology*, vol. 313, p. ajplung.00139.2017, 08 2017.

[5] S. Isaka, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, and S. Tsuruoka, "Development of a web based image annotation tool for lung immunofluorescent confocal images," *International Symposium on Affective Science and Engineering*, vol. ISASE2018, pp. 1–5, 2018.

[6] M. E. Ardini-Poleske, T. J. Mariani, G. S. Pryhuber, R. S. Misra, and The LungMAP Consortium, "Chapter 4 - initiating multiomics approach to understand neonatal chronic lung disease: the lungmap experience," in *Updates on Neonatal Chronic Lung Disease*, S. G. Kallapur and G. S. Pryhuber, Eds. Elsevier, 2020, pp. 45–59. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978032368353100004X

[7] L. Barbe, E. Lundberg, P. Oksvold, A. Stenius, E. Lewin, E. Björling, A. Asplund, F. Pontén, H. Brismar, M. Uhlén, and H. Andersson-Svahn, "Toward a confocal subcellular atlas of the human proteome*," *Molecular & Cellular Proteomics*, vol. 7, no. 3, pp. 499–508, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S153594762031210X

[8] 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 (CVPR)*, July 2017, https://github.com/tdeboissiere/DeepLearningImplementations/tree/master/pix2pix.

[9] T. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8798–8807, 2018, https://github.com/NVIDIA/pix2pixHD.

[10] T. Park, M.-Y. Liu, T. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2332–2341, 2019, https://github.com/NVlabs/SPADE.

[11] Y. Zhang, "Deep generative model for multi-class imbalanced learning," 2018, open Access Master's Theses. Paper 1277. [Online]. Available: https://digitalcommons.uri.edu/theses/1277

[12] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "Gans for medical image analysis," *Artificial Intelligence in Medicine*, vol. 109, p. 101938, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365719311510

[13] Z. Xu, C. Fernández Moro, B. Bozóky, and Q. Zhang, "GAN-based Virtual Restaining: A Promising Solution for Whole Slide Image Analysis," *ArXiv*, vol. abs/1901.04059, Jan 2019.

[14] X. Gong, S. Chen, B. Zhang, and D. Doermann, "Style consistent image generation for nuclei instance segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 3994–4003.

[15] R. Ranjan, S. Inoue, and T. Shibata, "Synthesizing cell protein data for human protein cell profiling using dual deep generative modeling," in *2020 Joint 9th International Conference on Informatics, Electronics Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, 2020, pp. 1–6.

[16] L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, "Unsupervised histopathology image synthesis," *arXiv preprint arXiv:1712.05021*, 2017.

[17] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *ArXiv*, vol. abs/2006.11239, 2020.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988, https://github.com/matterport/Mask_RCNN.

# AppendixA

# Detailed Immunofluorescence Method

**Purpose**

  Purpose Immunofluorescence on slides of $7\mu$m frozen sections of C57BL6 E16.5 lungs for SOX9, SFTPC, and ACTA2.

**Day 1**

1. For Frozen tissue, rinse 2X in PBS, then 5 min in 4% PFA/PBS, then rinse 1X in PBS.
2. Briefly equilibrate slides in Antigen Retrieval Buffer.
3. Antigen retrieval, pH 6.0 (times will vary according to microwave).
4. 10 mM sodium citrate, pH 6.0, and heat in a microwave at 96oC.
5. Microwave according to instructions on microwave.
6. Cool on countertop, 15 min.
7. Rinse with dH2O.
8. 1X PBS, 5 min.
9. Block in 4% Donkey serum/PBS ‑ T, 2 hours at RT.
10. For Rabbit anti-SOX9 (AB‑5535, Lot# 2167153, Millipore) dilute 1:100, For goat anti-SFTPC (SC‑7706, Santa Cruz) dilute 1:100, for mouse anti-ACTA2 ($\alpha$ Smooth muscle actin, A5228, Sigma) dilute 1:2000 in blocking buffer. Spin down in $\mu$fuge for 10 minutes and incubate on tissue overnight @ 4oC.

**Day 2**

1. Rinse slides in PBS ‑ T 3X, 5 min.
2. Apply secondary antibody , Donkey Alexa Fluor 488 anti-rabbit IgG (A21206, Lot# 1608521, for anti-SOX9), Donkey Alexa Fluor 568 anti-goat IgG (A11057, Lot# 1485187, for anti-SFTPC) at 1:200, Donkey Alexa Fluor 647 anti-mouse IgG A31571, Lot# 1549801, ACTA2) at 1:200, in blocking buffer. Spin down in $\mu$fuge for 10 min, apply to tissue and incubated at room temperature for 1 hour.
3. Rinse in PBS-T 3X, 5 min.

4. Dilute DAPI 1:2000 and apply to slides for 10 min.

5. Wash in PBS-T 3X, 5 min.

6. Rinse slides in 0.1M PB, 3X, 5 min.

7. Add 1 drop of Prolong Gold anti-fade mounting medium (P36930).

8. Coverslip with Gold Seal Coverslip (Cat# 3422 Electron Microscopy Sciences, 22 X 22 mm).

9. Allow Prolong Gold to cure overnight at room temp in light sealed box.

**Tissue Used**

LMM.14.24.4.46, E16.5, C57BL6, Gender: Male, Crown-to-rump: 16.0mm, and Weight: 0.592g