On Data Augmentation Techniques for Deep Learning Multi-class Segmentation of Lung Confocal Immunofluorescent Images

Daiki Katsuma[1], Hiroharu Kawanaka[1], Bruce J. Aronow[2, 3, 4], V. B. Surya Prasath[2, 3, 4, 5, *]
[1] *Department of Electrical and Electronic Engineering, Mie University, Japan*
[2] *Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, OH 45229 USA*
[3] *Department of Pediatrics, University of Cincinnati, OH USA*
[4] *Department of Biomedical Informatics, College of Medicine, University of Cincinnati, OH 45267 USA*
[5] *Department of Electrical Engineering and Computer Science, University of Cincinnati, OH 45221 USA*
[*] *Email: surya.prasath@cchmc.org*

***Abstract*—Human lung development requires complex gene and cell interactions, and the lung development can be studied at gene, cellular, and molecular levels. Availability of multimodal imaging data of the lung can help us visualize protein and cell localization in various lung structures. In particular, confocal Immunofluorescence (IF) images can used in lung development modeling and we require multi-class segmentation is an important requirement. However, it is difficult to obtain a sizeable number of training images in the multi-class segmentation with recent deep learning models such as the convolutional neural network (CNN) methods. Here, in order to improve the overall accuracy of multi-class segmentation in lung confocal IF images with generative adversarial network (GAN) based approach for data augmentations. By creating high fidelity synthetic images generated via lung confocal IF images, we mix the original and synthetically generated images for increasing the accuracy of the automatic segmentation. Our experimental results on a set of confocal IF images indicate we obtain good quality synthetic images that can be of use in deep CNN based segmentation models. We add the synthesized images to the training dataset and we obtained an extra improvement of 8.9% accuracy of six classes average.**

I. Introduction

Lung cancer is one of the leading causes of cancer deaths in the world, and in particular in the United States annual mortality number is in excess of 150,000 [1]. Studying the lung structures that perform for 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, and its development and maintenance are required 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 require carefully assessing various lung imaging modalities.

The LungMAP [3] consortium is one of the research projects for the lung development, which 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. Along with the availability of various types of data, such as biological sequencing data, RNA, protein, lipid, signaling, we can also obtain multimodal 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 work, we concentrate on utilizing the confocal immunofluorescent (IF) images of lung tissues, which are stained with the specific sets of proteins, and are especially used for a process of proteomics to understand the protein localization on the sub-cellular level by image annotation [5], [6], [7]. With the availability of large-scale confocal IF images, 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 segmentations using IF images reduces manual burden among pulmonary specialists, and promises to increase the usability of such imaging data on the analysis of lung development phases.

In this work, we consider obtaining accurate multi-class segmentation of lung confocal IF images using current state-of-the-art deep learning based models [8]. One of the primary bottleneck in using deep convolutional neural network (CNN) models is the lack of availability of training or ground-truth segmentation labels. In this paper, we implement the multi-class segmentation with generative adversarial network (GAN) models for improving the training dataset and for improving overall segmentation accuracy. To classify various lung tissue classes along with region of interests with deep CNN models, we propose to utilize recent GAN models [9]. We test data augmentation techniques for the multi-class segmentation problem, since the number of training images available in lung confocal IF were very few. Our experimental results indicate that we obtain high fidelity synthetic confocal IF images which can be of beneficial use in deep CNN models.

We organized the rest of the paper as follows. Section II introduces the recent GAN models in the context of biomedical images generation. Section III contains the details of the datasets used along with methods for data augmentation. Section IV provides the results of our experiments. Section V concludes the paper along with indicating the future works.

II. Related Works

The training dataset plays an important role in deep learning networks. Image augmentation has been used for preventing model overfitting by introducing data diversity and regularization. The traditional augmentation techniques include transformations, such as rotation, translation, channel splitting, Gaussian smoothing, unsharp masking, and etc. These image transformations improve model robustness, generalization, and learn better characteristics for making image distinctions.

Generative adversarial aetworks (GANs) [10] are well-known for generating synthetic data close to the training set distribution. With their abilities to mimic data distributions and to synthesize images at yet unprecedented levels of realism, GANs have carved open new ways 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 industry, where anomaly detection is critical like electricity pilferage, fraudulent transactions in banks, identification of rare diseases, etc [11].

In the medical image field, methods based GANs were applied to a variety of different modalities such as MRI (Magnetic Resonance Imaging), CT (Computed Tomography), OCT (Optical Coherence Tomography), chest X-Ray, Dermoscopy, Ultrasound, PET (Positron Emission Tomography) and Microscopy for implement applications: synthesis, reconstruction, detection, denoising, registration, classification and segmentation [12]. In another research of image reconstruction in the Histopathological cancer diagnosis, GAN 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 overcome the limitation of the image analysis and cut off the staining cost [13].

This study aims to synthesize the lung confocal IF images to expand the training datasets, and we investigate the synthesized IF image is effective or not by applying to the multi-class segmentation [8]. For the first stage, we focused on the evaluation of the synthesized image quality. We generated synthetic confocal IF images from the ground truth images to compare the synthetic and original images.

III. Proposed Method

A. Materials

We controlled the IF confocal 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 Lung MAP web site. Three types of proteins are stained as red, green, and white in each IF confocal image like Figure 1. The combination of stained proteins depends on each image. Furthermore, we also used the label image like Figure 2, which classify 6 class lung tissues, such as background (black), conductive airway (red), distal acinar tubule bud (green), proximal acinar tubule (blue), artery (white), and vein (orange).

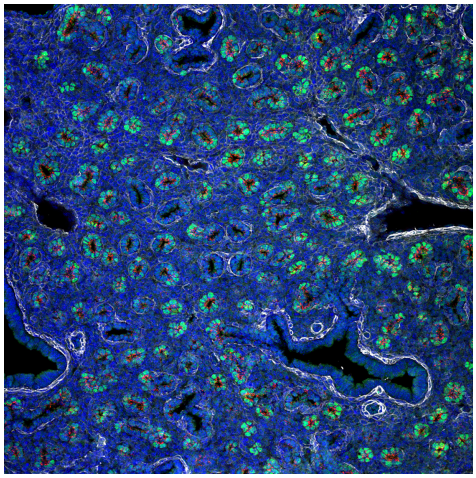


Fig. 1. Example of lung immunofluorescent confocal image. Three types of proteins are stained as red, green, and white. In this example, red, green, and white pixels correspond to the proteins named Sftpc, Sox9, Acta2, respectively.

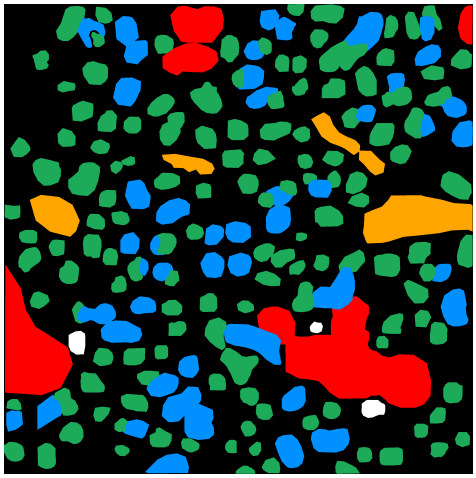


Fig. 2. Example of Label Image. This lable classify 6 class lung tissues, such as background (black), conductive airway (red), distal acinar tubule bud (green), proximal acinar tubule (blue), artery (white), and vein (orange), respectively.

Our mainstream application is for the generation of the synthetic images to increase the performance and accuracy of segmentation models [8]. Therefore, we use the synthetic images for expanding the dataset, and this work we investigate the degree of similarity between the synthetically created and original confocal IF images with three different evaluation methods.

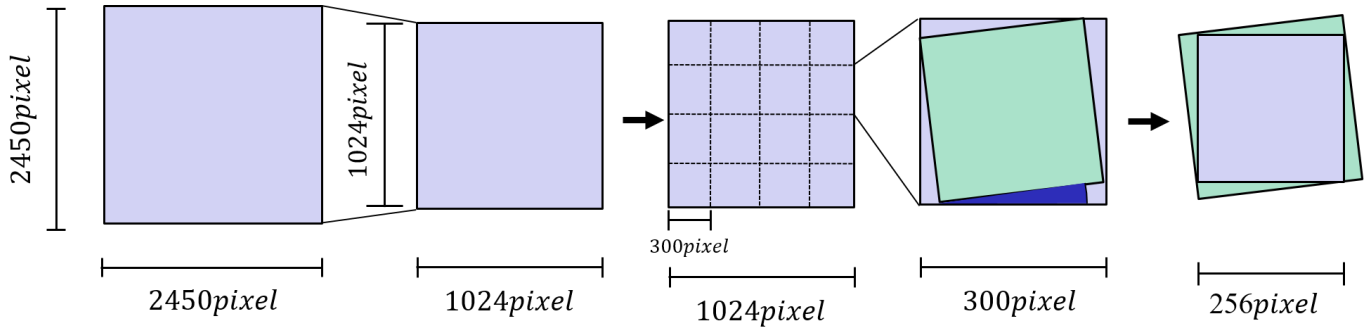


Fig. 3. Original images were resized 2450×2450 [pixel] to 1024×1024 [pixel], cropped 300×300 [pixel], rotated ($0° \pm 20°, 90° \pm 20°, 180° \pm 20°, 270° \pm 20°$), and cropped 256 [pixel]. Consequently, we prepared about $10,000$ images for Pix2Pix learning.

B. Image synthesis with Pix2Pix

We adapted the Pix2Pix [9] which was developed based on a GAN structure.

GAN consists of two main structures, which are Generator and Discriminator. The generator tries to make an image like the original image, and Discriminator judges whether the generated image is real or fake. By enhancing each network structure with iterated learning, GAN acquires a network to create quality synthesized-images. Pix2Pix is a kind of GAN. GAN uses skip connections and a discrimination technique with N $\times$ N patch images to improve its performance.

As a main characteristic of Pix2Pix, Pix2Pix can generate a synthetic images from a variety of materials such as labels to images, black and white to RGB images, Aerial to Map, and so on. We focused on the example of labels to images. This pair of images is directly applied to the segmentation dataset as a ground truth and training image.

C. Similarity metrics

We utilize mutual information (MI), structural similarity index and Pearson Correlation Coefficient (CC) for measuring the similarity between the synthetic images and original images, see Figure 4 (a), (b) along with a multi-class segmentation result [8]. For each pair of synthetic and original images, we compute the three metrics for their RGB images. The metrics are separately calculated for each channel and then averaged across channels. We used these evaluation methods as this reference[14].

1. Mutual information (MI)

Mutual information of the two random variables is a measure of the mutual dependence between the original and synthetic image pixels.

$$MI(x,y) = \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y) log \frac{P_{X,Y}(x,y)}{P_X(x) P_Y(y)} \quad (1)$$

2. Correlation Coefficient (CC)

Correlation is a measure of the strength and direction of the linear relation between the original and synthetic image pixels.

TABLE I
OUR GAN PARAMETER SET USED HERE FOR GENERATING LUNG CONFOCAL IF IMAGES.

Patch size	Epoch	Batch size	Learning rate[%]
64*64	1000	8	$2 * 10^{-3}$

$$CC(x,y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2)$$

3. Structural similarity (SSIM) [15]

Structural similarity (SSIM) defines the structural information in an image as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. Since luminance and contrast can vary across a scene, this methods use the local luminance and contrast for definition.

$$SSIM(x,y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

IV. EXPERIMENTAL RESULTS

A. Creating a dataset for Pix2Pix

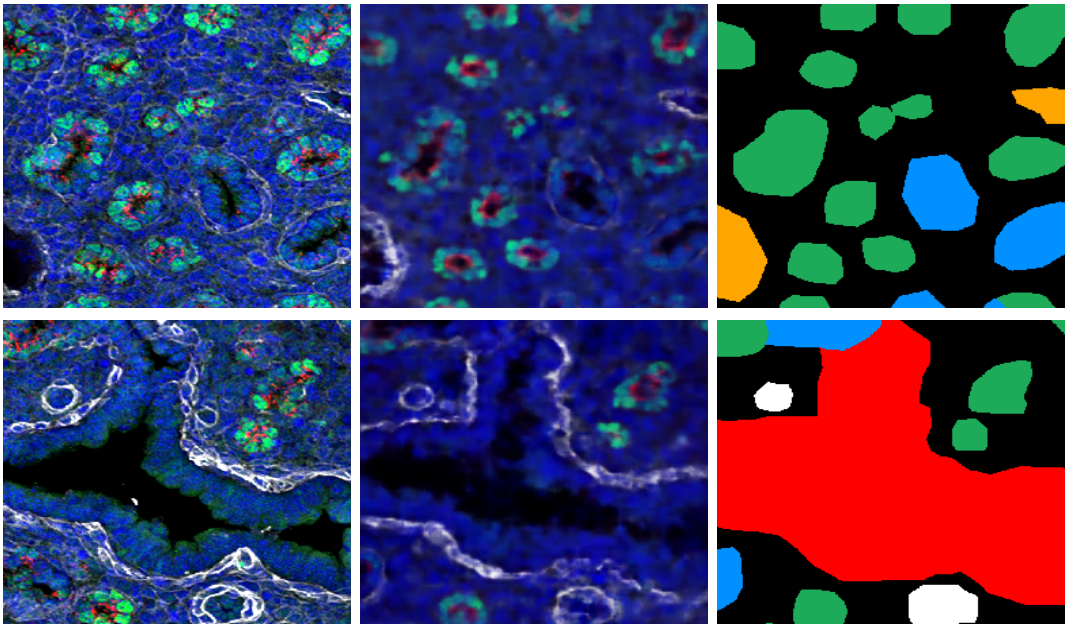
We conducted the image pre-processing step for Pix2Pix learning as given in Figure 3. We utilized four original images, and they were resized 2450×2450 [pixel] to 1024×1024 [pixel], cropped 300×300 [pixel], rotated ($0° \pm 20°, 90° \pm 20°, 180° \pm 20°, 270° \pm 20°$), and cropped 256 [pixel]. Consequently, we prepared about $10,000$ images for Pix2Pix learning.

B. Training and Implementation

We implemented the adapted Pix2pix GAN model using Keras and Tensorflow. Subsequently, we set the main parameters as shown in the Table I.

C. Similarity evaluation

As shown in the Table II and Figure 5, our evaluation indicates that we obtain higher than per pixel evaluation with CC and SSIM values. Further, SSIM can be trusted because the values are densely distributed. We considered that Pix2Pix would not focus on the pixel, but focus on the label shape when



(a) Original (b) Synthetic (c) Segmentation

Fig. 4. Examples of our GAN model for synthetic generation of images and their corresponding multiclass segmentation results. 4.(a) original, 4.(b) synthesized images, and their 4.(c) segmentation labels.

TABLE II
SIMILARITY METRICS OBTAINED WITH OUR GAN BASED AUGMENTATION OF LUNG CONFOCAL IF IMAGES.

Method	MI	Correlation	SSIM
Value	0.714 ± 0.018	0.829 ± 0.128	0.322 ± 0.148

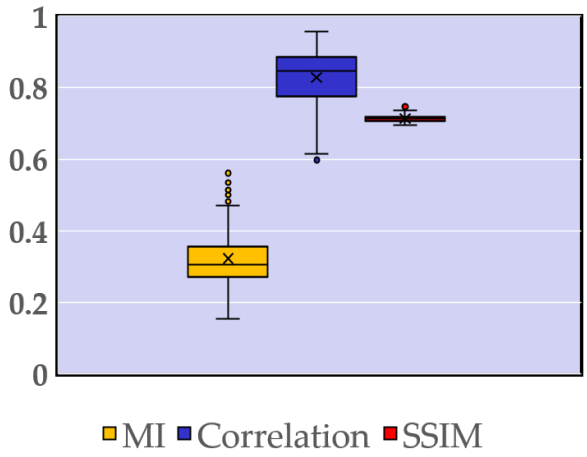


Fig. 5. Box plot showing the similarity metrics for our lung confocal IF images generated by GAN and real images.

Pix2Pix generated the synthetic images. Therefore, Pix2Pix cannot generate the synthetic image well when Pix2Pix does not retain some specific characteristics, for example, when the label image contain bigger regions, like a black and red labels based region in Figure 4 (bottom row).

TABLE III
BEFORE AND AFTER COMPARISON OF EACH CLASS AND WHOLE ACCURACY

[%]	C1	C2	C3	C4	C5	C6	Ave
Before	87.5	81.5	81.8	69.6	29.0	75.5	70.8
After	89.1	86.6	83.4	76.5	62.4	79.7	79.7

D. Segmentation accuracy

In this paper, we used U-net for segmentation because U-net was a well-known segmentation model and showed the best segmentation in the experiments without the mimic dataset. Table III shows the segmentation result of U-net. We confirmed the accuracy of six class segmentation improved 8.9% by adding 500 mimic data, which was translated with Pix2Pix and self-build data. In the case of C5, the number of the given image was not enough compared to other classes originally. The classification performance for C5 was drastically improved from 29.0% to 62.4% by adding the mimic dataset.

V. CONCLUSION

In the work, we described the image synthesis for the lung confocal IF images to improve the automatic segmentation accuracy. Firstly, we generated the synthetic images by implementing and learning a GAN based model.

We evaluated the similarity between the original and synthetic images with three evaluation metrics. The obtained result showed that the values of Correlation and SSIM were higher than MI. The variance of SSIM was small compared to others, , $i.e.$, SSIM would be a creditable metric. Finally, we

applied the acquired mimic dataset to the U-net, we confirmed that the segmentation accuracies for all classes were improved.

For advanced discussion, we will investigate another model to create synthetic images for this problem. The segmentation performance depends on the given dataset, and many approaches have been proposed to improve the performance of segmentation/classification. We will also investigate methods to make an effective dataset for segmentation.

ACKNOWLEDGMENT

The results shown here are in whole or part based upon data generated by the LungMAP Consortium and downloaded from (www.lungmap.net). The LungMAP consortium and the LungMAP Data Coordinating Center (1U24HL148865) are funded by the National Heart, Lung, and Blood Institute (NHLBI).

REFERENCES

[1] Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, et al. SEERCancer Statistics Review, 1975–2012, National Cancer Institute. Bethesda,MD. Available from: https://seer.cancer.gov/index.html.

[2] M. Heriges, E. E. Morrisey. "Lung development: orchestrating the generation and regeneration of a complex organ", The Company of Biologists, 141, 502-513, 2014.

[3] LungMAP Official Web Site, https://lungmap.net, [Accessed 15 March 2020].

[4] M. E. Ardini-Poleske et al. "LungMAP: The molecular atlas of lung development program", Am J Physiol Lung Cell Mol Physiol, 313, L733–L740, 2017.

[5] S. Isaka, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, S. Tsuruoka. Development of a web based image annotation tool for lung immunofluorescent confocal images. 4th International Symposium on Affective Science and Engineering, and the 29th Modern Artificial Intelligence and Cognitive Science Conference (ISASE-MAICS), Spokane, WA, USA, 5pp, Jun 2018.

[6] M. E. Ardini-Poleske et al. "Initiating multiomics approach to understand neonatal chronic lung disease: The LungMAP experience", pp. 45-59, 2020.

[7] L. Barbe et al. "Toward a confocal subcellular atlas of the human proteome", Molecular Cellular Proteomics, 7.3, pp.499-508, 2007.

[8] S. Isaka, H. Kawanaka, B. J. Aronow, V. B. S. Prasath. "Multi-class segmentation of lung immunofluorescence confocal images using deep learning", IEEE BIBM, 2019.

[9] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros. "Image-to-image translation with conditional adversarial networks", IEEE Conference on Computer Vision and Pattern Recognition, pp.1125-1134, 2017.

[10] I. Goodfellow et al. "Generative adversarial nets", In NIPS, 2014.

[11] Y. Zhang. "Deep Generative model for multiclass imbalanced learning", Open Access Master's Theses. Paper 1277, 2018.

[12] S. Kazeminia et al. "GANs for medical image analysis", IEEE Computer Vision and Pattern Recognition, 2018.

[13] Z. Xu, Carlos F. Moro, B. Bozoky, Q. Zhang. "GAN-based virtual restaining: A promising solution for whole slide image analysis", ArXiv, 2019.

[14] Q. D. Vu, K. Kim, K. T. Kwak. "Quantitative assessment of colorectal cancer via conditional generative adversarial networks", MICCAI, 2019.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.