HIGH SCHOOL EDITION
Journal of Student Research
Volume 14 Issue 1 (2025)

Deep Learning-Inferred Multiplex ImmunoFluorescence for Lamina-Associated Polypeptide 2 and Ki-67

Jieun Won[1] and Haeran Shim[#]

[1]Skyline High School, USA
[#]Advisor

ABSTRACT

Whole Slide Imaging (WSI) is a widely used pathological technology to digitize slide images and improve their navigation. Dyeing is an important part of this process, with two major options—Immunohistochemistry (IHC) and ImmunoFluorescence (IF). IF provides noticeably higher quality and contrast, as well as multiplexing capabilities, allowing scans of multiple targets simultaneously. However, IF is considerably more expensive than IHC, putting pathologists who use WSI for applications such as cancer treatment at a disadvantage. In this paper, I propose a novel solution, generating IF images from IHC images using a style transfer network. The proposed system is based on a Generative Adversarial Network (GAN), which utilizes a generator and a discriminator. Specifically, the system is trained to stain Lamina-associated polypeptide 2 and Ki-67. Additionally, I introduce multiplex staining results to demonstrate the system's effectiveness in a broader context. The proposed system achieved a Peak Signal-to-Noise Ratio (PSNR) of 31.38 and a Structural Similarity Index Measure (SSIM) of 0.8764 on a public immunofluorescence dataset.

Introduction

Whole Slide Imaging

Whole Slide Imaging (WSI) is a technology often used in pathology to navigate slides better. It is a disruptive technology where glass slides are scanned to produce digital images (Hanna et al. 2020). It uses Whole Slide Scanners to scan the slides in parts and stitches them together to create digital files. This technology produces digital images and allows pathologists to analyze slides with ease. It is used for many purposes, such as education and training, research activity, teleconsultation, and primary diagnosis (Eccher and Girolami 2020). To visualize the slides before scanning, a process of dyeing them is needed. There are two major dye options: Immunohistochemistry (IHC) or Immunofluorescence (IF).

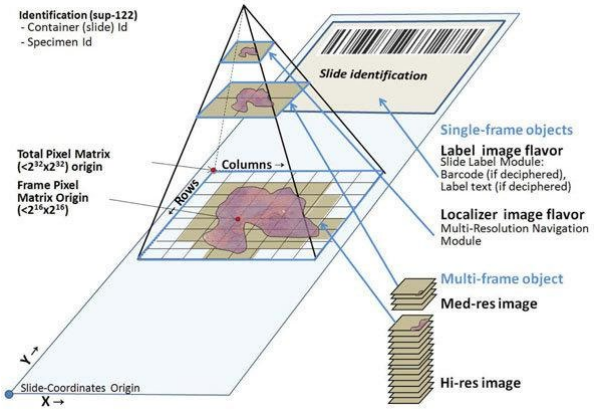
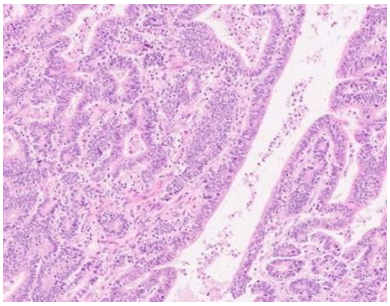


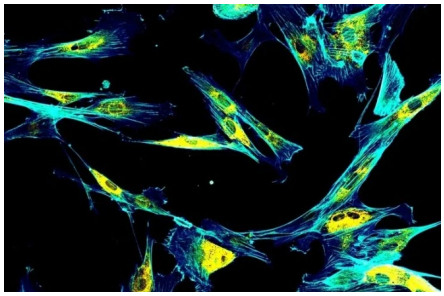
Figure 1. Process of Whole Slide Imaging (WSI) (Daniel et al. 2011)

Immunohistochemistry and Immunofluorescence

IHC and IF are both techniques used in pathology to detect specific antigens (proteins) in tissue sections. IHC utilizes antibodies conjugated to enzymes such as horseradish peroxidase or alkaline phosphatase, which catalyze a color-producing reaction. The procedure involves preparing tissue sections, retrieving masked antigens, blocking non-specific binding sites, and applying primary and secondary antibodies. The result is a visible color change at the antigen-antibody binding site under a light microscope. IHC is commonly used to identify cell types, diagnose diseases based on protein expression patterns, and determine the origin of metastatic tumors. IF, on the other hand, uses antibodies conjugated to fluorescent dyes (fluorophores) that emit light when excited by specific wavelengths. The process includes similar steps to IHC, but often uses frozen tissue sections to preserve antigenicity. IF is employed to study the spatial distribution and localization of proteins within cells and tissues, diagnose autoimmune diseases, and for various research applications.



(a)



(b)

Figure 2. (a): Immunohistochemistry (IHC) (Woolley 2024) and (b): Immunofluorescence (IF) (Cuffari 2021)

IF is considerably more costly than IHC due to several characteristics. Firstly, the reagents and antibodies used in IF, particularly fluorophore-conjugated antibodies, are generally more expensive than the enzyme-conjugated antibodies used in IHC. Additionally, IF requires advanced fluorescence microscopes and imaging systems, which are significantly pricier than the light microscopes used for IHC. The technical expertise required to perform and interpret IF, coupled with the need for specialized training to handle and maintain sophisticated equipment, further drives up costs. Consumables such as specialized slides and mounting media used in IF are also more expensive, and the

protocols can be more time-consuming due to additional steps needed to avoid photobleaching. The maintenance and calibration of fluorescence microscopy equipment, along with precise environmental controls to maintain fluorescence stability, contribute to the ongoing expenses. While IF offers higher sensitivity, specificity, and the ability to multiplex, these advantages come at a significantly higher cost, posing a barrier for routine diagnostic use in pathology labs, especially in resource-limited settings.

Proposed System

To address this issue, I propose a novel solution for generating IF images from IHC images using a style transfer network. The proposed system is based on a Generative Adversarial Network (GAN), which consists of a generator and a discriminator. The network is specifically trained to simulate the staining of Lamina-associated polypeptide 2 and Ki-67. Additionally, I present results from multiplex staining to showcase the system's effectiveness and versatility in a broader range of applications.

Proposed ImmunoFluorescence Staining System

This chapter explains the detailed steps in developing the proposed Immunofluorescence (IF) staining system. The system is composed of two modules: the style transfer network and the discriminator network. The style transfer network takes an IHC image as input and generates the corresponding IF-stained image as output. The discriminator network then classifies these generated images as either real or fake by comparing them to original images. These two networks are trained in an adversarial manner to enhance the accuracy of the system.

Immunofluorescence Generation Network

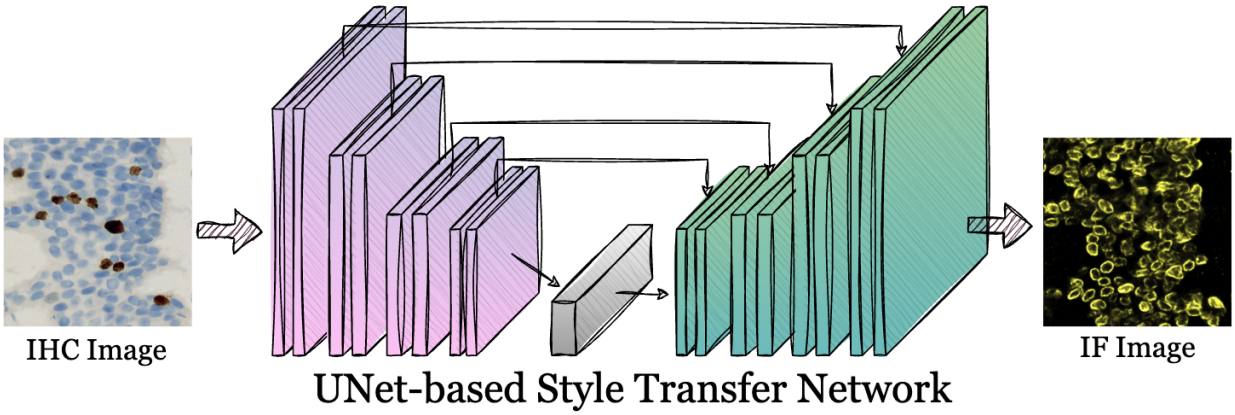


Figure 3. Architecture of the proposed immunofluorescence generation network

To obtain high quality images without the drawbacks of physically using IF, I propose the method of using Generative Adversarial Network (GAN) to convert Immunohistochemistry (IHC) images into IF ones. The first step of this method is the generator (Figure 3), which generates the IF dyed images. It consists of two parts, the encoder and the decoder. The encoder is a CNN that down-samples the input (IHC image) to focus on the important parts, generating feature maps (mathematical representations, vector matrices of the important features). The decoder is another CNN that up-samples its input to match the original size as well as converting the IHC into IF.

Equation 1. L1 loss function

$$J_{G-L1}(\Phi_{STNet}) = \sum_{x}^{X}\sum_{y}^{Y}|I(x,y) - \hat{I}(x,y)|$$

In Equation 1, G is the generator's loss. L1 means it is the first part of the loss. STNet is the parameter of the generator. Uppercase X and Y are the sizes of the image. Lowercase x and y show the location of the specific pixel the equation is on. I is the real IF image while $\hat{I}$ is the generated IF image.

Equation 2. VGG loss function

$$J_{G-vgg}(\Phi_{STNet}) = \sum_{x}^{X}\sum_{y}^{Y}\sum_{c}^{C}[z^{real}(x,y,c) - z^{fake}(x,y,c)]^2$$

In Equation 2, G is the generator's loss. vgg means it is the second part of the loss using Visual Geometry Group (VGG). STNet is the parameter of the generator. Uppercase X, Y and Z are the sizes of the feature maps produced by VGG. Lowercase x, y and z show the location of the specific pixel the equation is on. z^{real} is the real IF image's feature map produced by VGG while z^{fake} is the generated IF image's feature map produced by VGG.

Equation 3. Generation loss function

$$J_G(\Phi_{STNet}) = J_{G-L1} + \eta J_{G-vgg}$$

In Equation 3, G is the generator's loss. STNet is the parameter of the generator. L1 is the first part of the loss while vgg is the second part of the loss.

To obtain the loss of the generator, or how much average error it has, let's consider two aspects. One is the mathematical , literal loss that compares the values of the pixels of the generated IF image to the real IF image (Equation 1). The second is perceptual loss which compares qualities such as contrast between the cells and background. To do this, we use VGG that is trained on ImageNet (large visual dataset with over 14 million images) and can work comparably to a human in terms of classifying images. We run both images (generated and real) through it, and compare the feature maps the VGG produces (Equation 2). By combining the two losses (Equation 3), we can compare the images effectively.

Discriminator Network

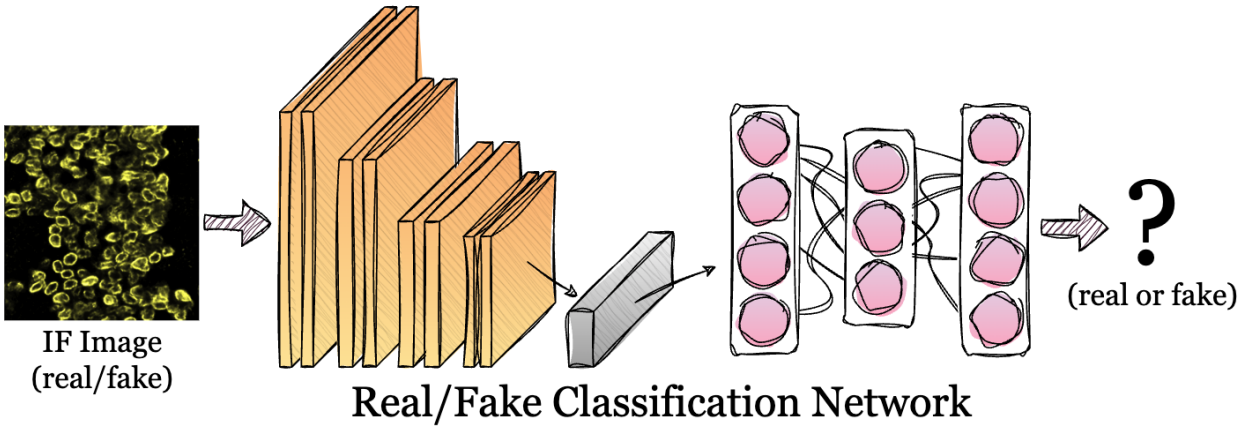


Figure 4. Architecture of the proposed discriminator network

The second part is the key of GAN: the discriminator (Figure 4). It works with the generator to improve both systems at the same time. The discriminator gets the generated IF image from the generator and tries to figure out if it was generated or if it was taken authentically. In this competing system, the generator works to generate better images so it won't be detected by the discriminator, and the discriminator works to better distinguish real and generated images, improving both systems.

Equation 4: Binary cross entropy loss function

$$J_D(\Phi_{RFCNet}) = -[p \times log_e(\hat{p}) + (1-p) \times log_e(1-\hat{p})]$$

To obtain the loss of the discriminator, a different function is used. It is the binary cross entropy loss function (Equation 4). Here, 0 means generated and 1 means real. y is the real answer, while $\hat{y}$ is what the discriminator predicts (given in a percentage it thinks the image could be real). If the real answer is generated, or 0, the first part of the equation cancels out ($0 \times log_e\hat{y}$). Otherwise when y is real, or 1, the second part cancels out ($(1-1) \times log_e(1-\hat{y})$). So one of the two parts is left. Then, the natural log function is used to get the loss of the difference between the expected output and the generated output ($1 - \hat{y}$ or $\hat{y} - 0$).

Experimental Results

Dataset and Evaluation Metric

The dataset used in this study consists of tumor sections that were initially stained using the multiplex immunofluorescence (IF) assay and then restained using the multiplex immunohistochemistry (IHC) method. The dataset is composed 1,239 samples designated for training (80% of the dataset) and the remaining 20% reserved for testing.

To examine the performance of the proposed approach, four Convolutional Neural Networks (CNN) with different depths were tested. I used two evaluation metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) (Hore and Ziou 2010). PSNR is the ratio between signal and corrupting noise in an image. Rates between 30 and 50 are considered good quality. SSIM is the measure of perceptive similarity between two images, which in this case was used between the generated IF image and the real IF image. -1 is perfectly dissimilar while 1 is perfectly similar.

Results

Table 1. Four CNN architectures evaluated (PSNR and SSIM)

	PSNR	**SSIM**
VGG-19 (Simonyan et al. 2014)	24.52	0.7148
MobileNetV2 (Sandler et al. 2018)	26.71	0.7453
ConvNext (Liu et al. 2022)	29.45	0.8449
ResNet-152 (He et al. 2016)	31.87	0.8764

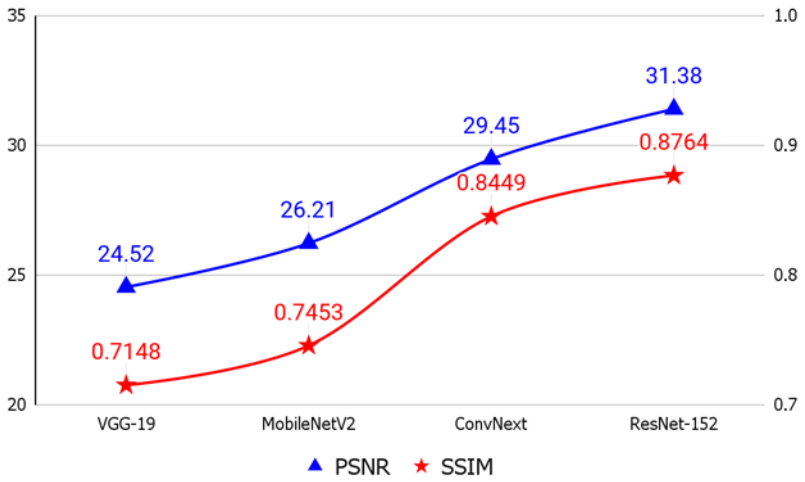
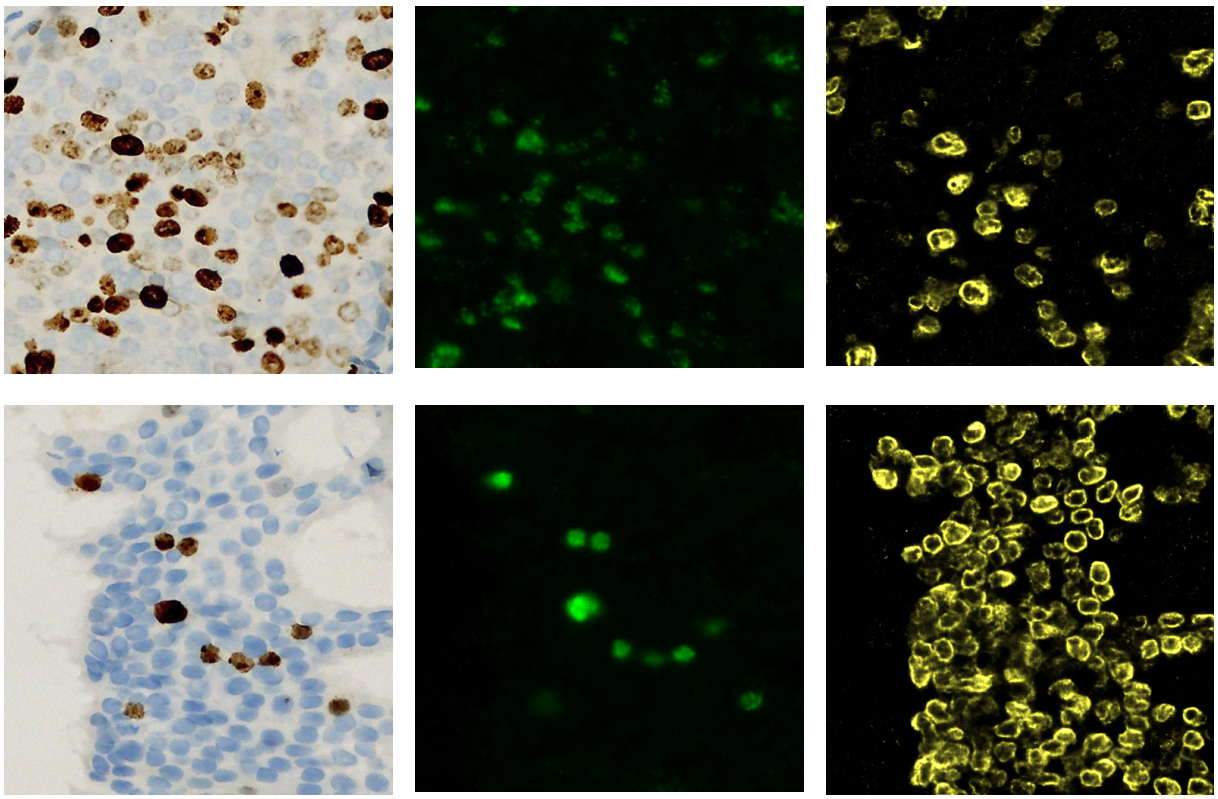


Figure 5. Graph of the tested results

To evaluate the performance of the proposed approach, I utilized four convolutional neural network architectures: VGG-19 (Simonyan et al. 2014), MobileNetV2 (Sandler et al. 2018), ConvNext (Liu et al. 2022), and ResNet-152 (He et al. 2016), each with different layer depth settings. I trained these architectures with an identical experimental setup and measured PSNR and SSIM to compare their performance.

As shown in Figure 5 and Table 1, ResNet-152 outperformed the other architectures, achieving a PSNR of 31.38 and an SSIM of 0.8764. In contrast, VGG-19 produced the least accurate results, likely due to its shallower layer depth.



(a) (b) (c)

Figure 6. IHC images and generated IF images using the proposed method

To visually examine the results of the proposed method, I generated images using ResNET-152. Figure 6 (b) and (c) demonstrate the generated images of Ki-67 and LAP-2 (Lamina Associated Polypeptide) respectively. Ki-67 is an important biomarker for cancer, specifically tumor growth. LAP-2 is a protein found in the cell membrane and shows upregulation in cancer.

The images (Figure 6) are accurate and have high quality, especially the LAP-2 Figure 6 (c) where the contrast between the target and background is exceptionally clear. With its quality, the proposed system will be of great use to pathologists.

Conclusion

In this paper, I proposed a Generative Adversarial Network (GAN) consisting of a generator and a discriminator to generate high-quality Immunofluorescence (IF) images from Immunohistochemistry (IHC) images. The generator creates the IF images, while the discriminator works to improve the system by distinguishing between real and generated images, leading to simultaneous enhancement of both networks. To determine the most effective architecture for this method, four Convolutional Neural Networks (CNNs)—VGG-19, MobileNetV2, ConvNext, and ResNet-152—were evaluated. Among these, ResNet-152 demonstrated the highest performance, achieving a Peak Signal-to-Noise Ratio (PSNR) of 31.38 and a Structural Similarity Index Measure (SSIM) of 0.8764. Visual quality experiments confirmed that the generated IF images were of high quality and accuracy. The proposed method offers a viable alternative to expensive IF staining which reduces the need for specialized lab technicians and costly reagents. In future research, I aim to further enhance the PSNR and SSIM rates and explore the application of this method to a broader range of biomarkers and tissue types.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Daniel, C., Macary, F., García Rojo, M., Klossa, J., Laurinavičius, A., Beckwith, B. A., & Della Mea, V. (2011). Recent advances in standards for collaborative Digital Anatomic Pathology. Diagnostic pathology, 6, 1-10.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Hore, A., & Ziou, D. (2010, August). Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition (pp. 2366-2369). IEE

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11976-11986).

Native Antigen Company. (2019, Jan 28). "*Visualising viral infection with immunofluorescence microscopy*": Native Antigen Company https://thenativeantigencompany.com/visualising-viruses-with-immunofluorescence-microscopy/

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520). https://doi.org/10.48550/arXiv.1801.04381

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Woolley, D. (2024, Jul 25). "*An Introduction to Immunohistochemistry (IHC)*": Leica Microsystems. https://www.leicabiosystems.com/us/educational-resources/articles/an-introduction-to-immunohistochemistry-ihc/