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

Machine Learning-Based Multiplex Immunofluorescence Staining from Immunohistochemistry with Generative Adversarial Networks

Ahyoung Kim[1], Ryeogyung Kim[1] and Mark Kim[#]

[1]North London Collegiate School Jeju, Republic of Korea
[#]Advisor

ABSTRACT

Immunohistochemistry (IHC) and Immunofluorescence (IF) are the two most commonly employed methods of cell staining, which aims to enhance visibility of cellular components for medical research and diagnosis. Immunofluorescence which uses fluorescently labeled antibodies to detect specific proteins is preferred over immunohistochemistry for multiplexing and colocalization which analyze protein distributions due to its transparent nature. However access is often barred due to high costs of reagents and equipment, slide quality degradation overtime, and technical limitations such as autofluorescence, and background staining from non-specific binding or cross reactivity antibodies. To address the aforementioned issue, we propose a novel machine learning-based immunofluorescence staining approach utilizing Generative Adversarial Network (GAN). The model intakes a digital pathology image of IHC and converts it to an IF image through feature maps. The proposed model achieved state-of-art performance with PSNR value of 30.67 and SSIM value of 0.8992. Additionally, the qualitative experimental results demonstrate the efficacy of the proposed method in enhancing generating high-quality IF stained images.

Introduction

Protein and cell staining are important techniques in pathology because they enable the visualization and analysis of cellular and tissue components at a microscopic level. Staining helps pathologists identify specific proteins, cellular structures, and pathological changes, which are essential for diagnosing diseases, understanding their progression, and guiding treatment decisions. These techniques allow for the examination of tissue samples for abnormalities, such as cancerous growths or infections, and provide valuable insights into the underlying biological processes of various diseases.

Immunofluorescence (IF) and Immunohistochemistry (IHC) are both techniques used to detect specific proteins or antigens in cells and tissue samples, but they differ in their methods and applications (Tan et al. 2020). IF uses fluorescently labeled antibodies to bind to specific proteins within a sample (Im et al. 2019). When exposed to light at specific wavelengths, the fluorescent tags emit light, allowing visualization of the target proteins under a fluorescence microscope. IF is highly sensitive and can provide detailed spatial and temporal information about protein localization within cells or tissues. Immunohistochemistry (IHC) uses antibodies conjugated with chromogenic enzymes to detect proteins in tissue sections. The enzymatic reaction produces a colorimetric change at the site of the antigen-antibody binding, which is visible under a standard light microscope. IHC is widely used in clinical diagnostics, particularly for identifying cancer markers and other disease-related proteins.

IF is relatively expensive due to several factors: the high cost of fluorescent dyes and antibodies, which are necessary for strong and stable signals; the requirement for advanced fluorescence microscopes with specialized light

sources and filters, which are more expensive than standard light microscopes; the technical complexity of the procedure, which demands precise handling and expertise to prevent issues like photobleaching and non-specific binding; and the higher costs associated with maintaining the equipment and purchasing consumables, such as high-quality slides and coverslips.

To address this issue, we introduce a novel machine learning-based system for IF staining. The proposed system is developed using a Generative Adversarial Network (GAN) architecture, which takes an IHC image as input and produces an IF-stained image as output. For training, we utilized the mean squared error and perceptual loss functions for the generator, while the discriminator was trained using the binary cross-entropy loss function. The system successfully generates IF-stained images for Ki-67 protein and DAPI-stained images for regular nuclei cells.

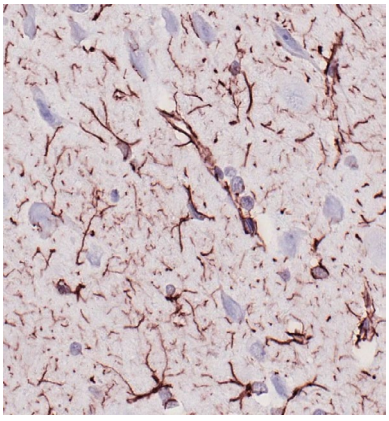
The structure of this research paper is as follows: Chapter 2 discusses IHC and IF. Chapter 3 describes the development of the proposed system, while Chapter 4 presents extensive experimental results. Finally, Chapter 5 summarizes the paper.

Related Work

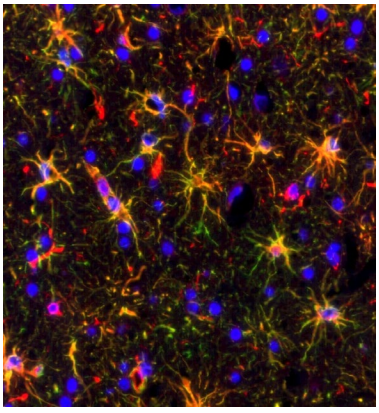
Immunohistochemistry and Immunofluorescence

Immunofluorescence (IF) and Immunohistochemistry (IHC) are both useful techniques used to detect specific proteins or antigens in cells and tissue samples. They are widely used in research and diagnostic settings in the fields of pathology, molecular biology, and cancer research.

IHC uses chromogenic (color-producing) reactions to visualize proteins in tissue sections. IHC has several advantages, including being cost-effective, as the reagents and equipment, such as light microscopes, are generally less expensive than those used in IF. IHC is widely used in clinical diagnostics, especially in pathology, where it plays a crucial role in detecting cancer markers and other diseases. Another significant benefit of IHC is the stability of its results; the chromogenic reactions produce permanent results that do not degrade over time, unlike fluorescence, making the data more reliable for long-term studies. However, IHC also has some limitations. It tends to have lower sensitivity, meaning it may not detect proteins that are present in very low quantities. Additionally, IHC is generally limited in multiplexing capabilities, as it can typically only detect one or two proteins at a time due to the potential overlap in color development. This technique also provides less detailed visualization compared to IF which makes it more difficult to determine the precise localization of proteins within cells.



(a) Immunohistochemistry



(b) Immunofluorescence

Figure 1. Sample of Immunohistochemistry and Immunofluorescence staining Image (Merck 2024)

In contrast, IF is a technique that employs fluorescent-labeled antibodies to identify specific proteins within cells or tissue sections. IF provides high sensitivity for detecting low levels of target proteins and specificity through the use of precise antibodies. It also supports multiplexing, allowing for the simultaneous detection of multiple proteins with different fluorescent dyes in the same sample. Additionally, IF provides detailed spatial and temporal information about protein localization within cells or tissues. However, IF is generally more expensive than IHC due to the cost of fluorescent dyes, specialized antibodies, and the need for advanced fluorescence microscopes. The cost is further increased by the requirement for well-trained personnel and sometimes sophisticated image analysis software to interpret results accurately.

Generative Adversarial Networks (GANs)

A Generative Adversarial Network (GAN) is a deep learning model designed for generating new data that resembles a given set of training data. GANs are popular for tasks like image generation. A GAN consists of two neural networks: the Generator and the Discriminator. These two networks are trained together in a process that can be thought of as a game where each network tries to outsmart the other.

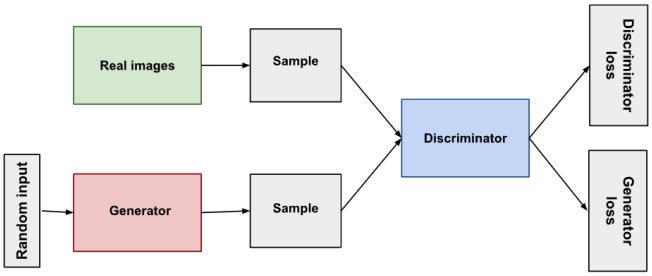


Figure 2. Architecture of GAN (Google Developers 2024)

The generator's goal is to produce data that is indistinguishable from real data. In general, it takes in a random noise vector (often sampled from a simple distribution like a Gaussian or uniform distribution) and transforms it into a data sample that should look like the real data. The discriminator's job is to differentiate between real data (from the training set) and fake data generated by the generator. It outputs a probability score indicating whether a given input is real or fake.

In this research, we approach IF staining as a generation task, where the input data consists of IHC images and the output is the generated IF image. The detailed methodology of the proposed approach will be explained in Chapter 3.

Multiplex Immunofluorescence Staining System

The proposed multiplex immunofluorescence (IF) staining system leverages a Generative Adversarial Network (GAN) composed of both generative and discriminator networks. The generator network takes immunohistochemistry (IHC) images as input and synthesizes corresponding IF stained images as output. Meanwhile, the discriminator network evaluates these synthesized images alongside real IF images to classify them as either authentic or fabricated. By

jointly training these two networks, the system not only improves the quality and realism of the generated IF images but also enhances the overall robustness and accuracy of the staining (generation) process.

Generator Network

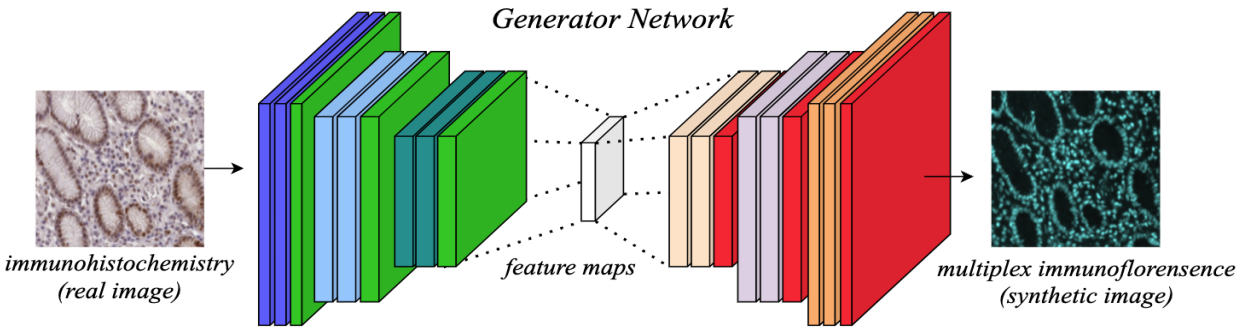


Figure 3: Architecture of the proposed IF generator network

Figure 3 illustrates the architecture of the IF generator network. The network takes a real IHC image, obtained from tissue samples, as input. The input IHC image is transformed into feature maps that mathematically represent the cells and protein structures within the original image. These feature maps are then upsampled to match the original image resolution for the reconstruction of a synthetic IF image. To measure the error of the synthesized IF image, we utilize the L2 distance function, as explained in Equation 1.

Equation 1: Generative loss function

$$L_G = \frac{1}{P}\sum_{p=1}^{P}\left(mIf(p) - \widehat{mIF}(p)\right)^2 + \gamma\frac{1}{F}\sum_{f=1}^{F}(VGG^{mIF}(f) - VGG^{\widehat{mIF}}(f))^2$$

Here, $\widehat{mIf}$ and ***mIf*** denote the synthesized IF image and its corresponding ground truth IF image, respectively. VGG$\widehat{mIF}$ and $\boldsymbol{VGG^{mIF}}$ represent the intermediate feature maps from the ImageNet-trained VGG network, obtained by inputting both $\widehat{mIf}$ and ***mIf***. Equation 1 calculates the loss of the generative function. The first part of the equation calculates the loss value of pixels in the synthetic image. This part carries more significance as it measures how closely the generated model has created the target image at a pixel level. On the other hand, the second part of the equation calculates the difference between feature maps that have been through the ImageNet (Deng et al. 2009) trained VGG (Simonyan et al. 2014) network. This perceptual loss function enhances the perceptual components of an image.

Here, the $\boldsymbol{\gamma}$ is introduced to scale down the contribution of the second part of the formula. It reflects the secondary importance of the second part, as pixel level accuracy is more critical to a generative network's performance. The value of $\boldsymbol{\gamma}$ is set to 0.8 in this study.

Adversarial Discriminator Network

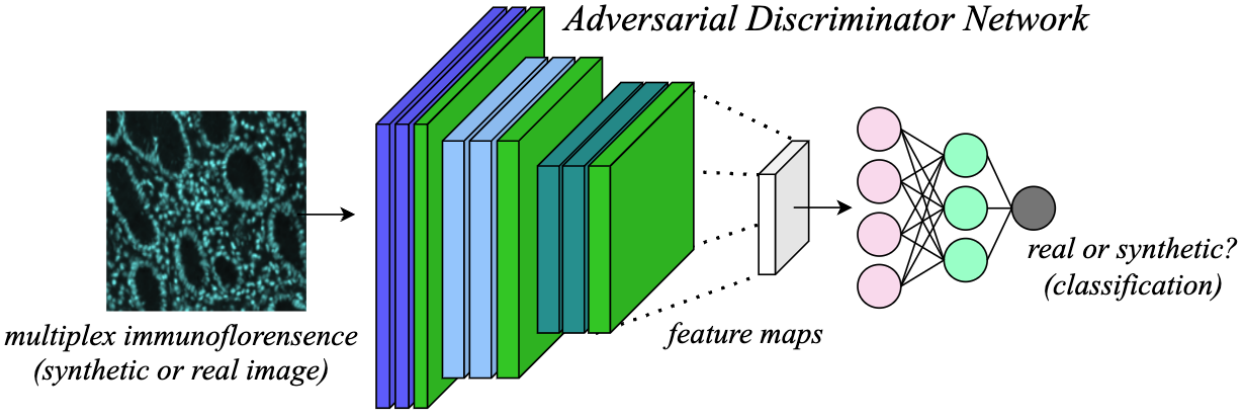


Figure 4. Architecture of the proposed adversarial discriminator network

The second module, discriminator network, takes the reconstructed IF image as an input. The input goes through a classification network which differentiates whether the image is real or synthetic. The network first processes the image through several convolutional layers that extract the key features. It then constructs a feature map which mathematically represents the input. After calculations in the connecting layers, a probability is calculated to distinguish the input's class.

Equation 2: Discriminator loss function

$$L_D = -\frac{1}{N}\sum_{i=1}^{N} y_i \times log_e(\hat{y}_i) + (1 - y_i) \times log_e(1 - \hat{y}_i)$$

Here, $\hat{y}$ and $\boldsymbol{y}$ denote the prediction probability and its ground truth probability, respectively. Equation 2 calculates the loss of the binary classification network. Where N represents the number of total samples in the function. The overall loss $\boldsymbol{L_D}$ is the average of individual loss samples and provides the model's overall performance of the dataset. The first part of the equation calculates the probability of guessing that the "image was generated by A.I." It multiplies the guessed value by the corresponding actual label. The second part calculates the probability of guessing that the input is a "real image." It ensures that the model is penalized when it incorrectly predicts a real image as being A.I. generated. The multiplication helps in balancing the false positives and false negatives, making the process effective.

Experimental Results

Dataset

The dataset was obtained by staining the same tumor sections with both immunohistochemistry (mIHC) and immunofluorescence (mIF), and constitutes the first public dataset to employ both methods of staining on the same cell specimens. Samples included are in 512x512 pixels and cover a range of specimens including those from Head and Neck Squamous cell carcinoma, Bladder cancer and Lung cancer. From a total of 1624 samples, 80% will be reserved for training the network and the remaining 20% will be dedicated to testing the network's accuracy.

PSNR and SSIM

Equation 3: Peak Signal-to-Noise Ratio (PSNR)

$$MSE = \frac{1}{P}\sum_{p}[\,\widehat{mIF}(p) - mIF(p)\,]^2$$

$$PSNR = 10 \times log_{10}(\frac{255^2}{MSE})$$

Peak Signal to Noise Ratio (PSNR) evaluates the quality of a compressed or reconstructed image by calculating the error between corresponding pixels in the original data and the reconstruction (Hore and Ziou 201). It calculates the Mean Square Error (MSE) value for each pixel, and higher values of the MSE result in lower PSNR values. Assuming the usage of an 8-bit representation, as is in most cases, the upper limit value of PSNR is 48dB, defined by the maximum pixel value of 255. PSNR values of over 30dB are associated with high image quality by quantitative analysis.

Equation 4: Structural Similarity Index Measure (SSIM)

$$SSIM = lum(\widehat{mIF}, mIF) \times con(\widehat{mIF}, mIF) \times str(\widehat{mIF}, mIF)$$

$$lum(\widehat{mIF}, mIF) = \frac{2\mu_{\widehat{mIF}}\mu_{mIF} + C_1}{\mu^2_{\widehat{mIF}} + \mu^2_{mIF} + C_1}$$

$$str(\widehat{mIF}, mIF) = \frac{\sigma_{\widehat{mIF}mIF} + C_3}{\sigma_{\widehat{mIF}}\sigma_{mIF} + C_3}$$

$$con(\widehat{mIF}, mIF) = \frac{2\sigma_{\widehat{mIF}}\sigma_{mIF} + C_2}{\sigma^2_{\widehat{mIF}} + \sigma^2_{mIF} + C_2}$$

Structural Similarity Index Measure (SSIM) is used to evaluate the quality of the image's structure, particularly the similarity between the original image (Hore and Ziou 201). It focuses on the structural information such as luminance, contrast, andstructure. Luminance measures the similarity in the brightness between the original and distorted image. Contrast, evaluates the similarity in the pixel intensity. Structure evaluates the structural similarity in the images. Overall, these components are multiplied to calculate SSIM.

SSIM ranges between zero to one, where one indicates perfect similarity and values closer to zero indicate significant difference with the original image. Overall, it provides a perpetually relevant assessment compared to other methods, focusing on aspects that are more relevant to human visual perception.

Evaluation Result

To evaluate the proposed approach, we chose four different Convolutional Neural Network (CNN) architectures which display comparable performance in many computer vision tasks.

Table 1. PSNR and SSIM evaluation

	PSNR	**SSIM**
HRNet-W48	27.72	0.8011

Xception	27.89	0.8045
ConvNext	29.85	0.8847
ResNet-50	30.67	0.8992

The 4 employed CNN architectures had different numbers of layers in their neural network, thus different depths, and were studied to observe the effect of depth on network performance and select the optimum architecture for mIF stained image generation. All 4 architectures were tested with the identical experimental protocol involving the measuring of PSNR and SSIM values of images generated by each architecture. Table 1 summarizes the experimental results, and reveals the PSNR and SSIM values of the ResNet-50 architecture to be the highest of those of the 4 architectures, with 30.67 and 0.8992 for its PSNR and SSIM values respectively. Given that the PSNR and SSIM values of a respective 27.72 and 0.8011 yielded by the lowest performing architecture, the HRNet-W48, is still considered to confer high image quality, the ResNet-50 architecture selected for the approach is likely to yield accurate image processings.

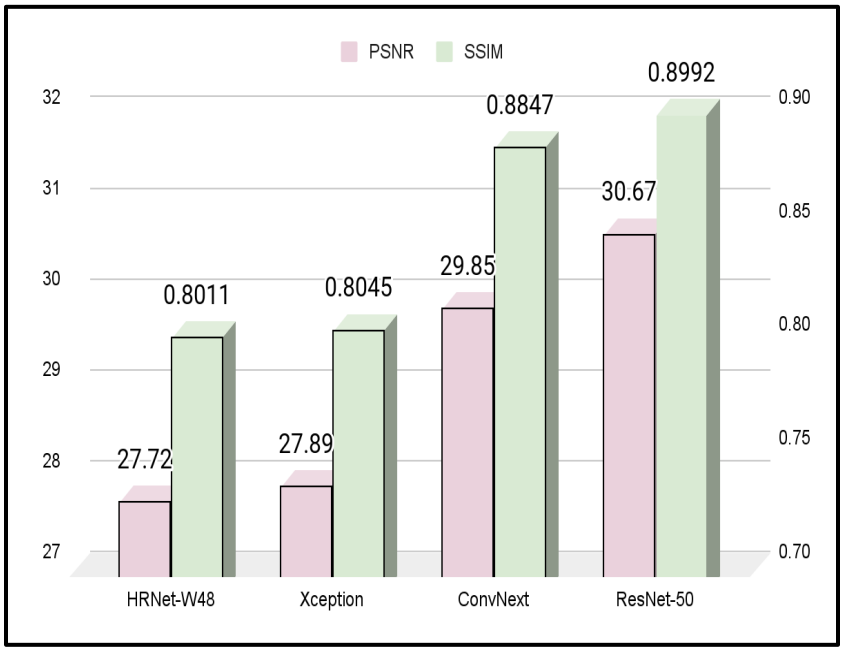
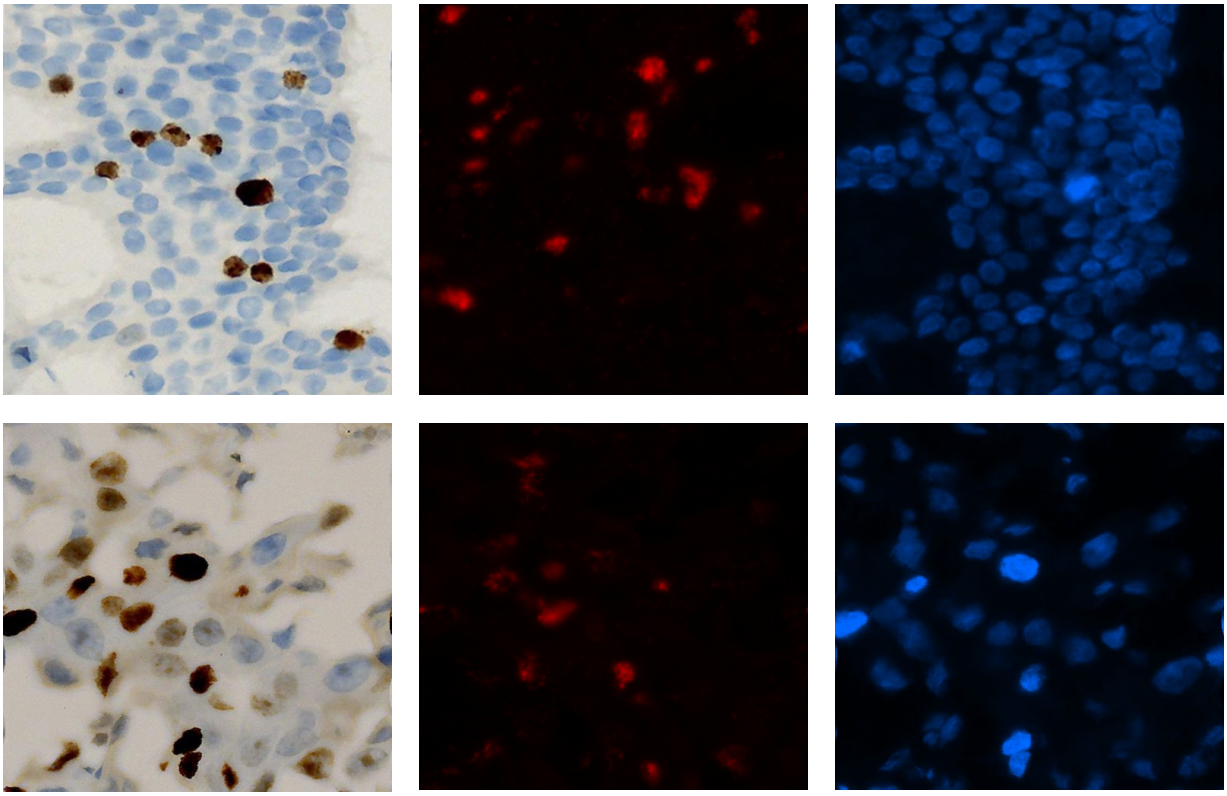


Figure 5. Graphical representation of CNN architecture PSNR and SSIM evaluation

Visual Experiement



(a) (b) (c)

Figure 6. (a): IHC image, (b): Ki-67 protein and (cc): DAPI-stained image

Figure 6 demonstrates the visual experiment of the proposed system. Figure 6(a) shows the input IHC image, Figure 6(b) displays the IF-stained Ki-67 protein, and Figure 6(c) presents the DAPI-stained image. The proposed system efficiently generates IF images without requiring costly IF staining reagents or specialized equipment. We expect this approach to offer a cost-effective and accessible alternative for producing high-quality IF images.

Conclusion

In this research, we propose a machine learning-based system for Immunofluorescence (IF) staining using Generative Adversarial Networks (GANs). We evaluated the prposed approach with four state-of-the-art convolutional neural network architectures. The proposed system achieved a Peak Signal-to-Noise Ratio (PSNR) of 30.67 and a Structural Similarity Index (SSIM) of 0.8992 on a public dataset which demonstrates remarkable performance. Additionally, we conducted visual experiments to assess the quality of the generated IF images and performed mask alignment experiments to further evaluate the system's effectiveness. Future work will focus on applying the proposed system to various proteins and cell types to expand its utility.

Acknowledgments

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

References

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

Google Developers. (2024, Aug 31). "*Overview of GAN architecture*": Google Developers. https://developers.google.com/machine-learning/gan/gan_structure

Harms, P. W., Frankel, T. L., Moutafi, M., Rao, A., Rimm, D. L., Taube, J. M., ... & Pantanowitz, L. (2023). Multiplex immunohistochemistry and immunofluorescence: a practical update for pathologists. Modern Pathology, 36(7), 100197.

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

Im, K., Mareninov, S., Diaz, M. F. P., & Yong, W. H. (2019). An introduction to performing immunofluorescence staining (pp. 299-311). Springer New York.

Merck. (2024, Aug 31). "Immunohistochemistry": Merck. https://www.sigmaaldrich.com/KR/ko/applications/protein-biology/immunohistochemistry

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

Tan, W. C. C., Nerurkar, S. N., Cai, H. Y., Ng, H. H. M., Wu, D., Wee, Y. T. F., ... & Lim, T. K. H. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Communications, 40(4), 135-153.