# Dual-stream Generative Network Based Staining Transfer for Biomarker in Breast Cancer

Ziyang Jin [1,2], Jiansheng Wang [2,3,*], Yan Wang [3], and Qingli Li [3]

[1] Department of Electrical Engineering, ZJU-UIUC Institute, Zhejiang University, Haining, China
[2] Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education, East China Normal University, Shanghai, China
[3] Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai, China
Email: ziyang.22@intl.zju.edu.cn (Z.J.); jswang@cee.ecnu.edu.cn (J.W.); ywang@cee.ecnu.edu.cn (Y.W.); qlli@cs.ecnu.edu.cn (Q.L.)
*Corresponding author

*Abstract*—**Pathological examination is a crucial standard in cancer diagnosis, with breast cancer being one of the leading causes of morbidity and mortality in recent years, posing a major threat to health. Enhancing pathological examination capabilities has become an important way to save lives and improve patients' quality of life. Common pathological examination methods include Hematoxylin and Eosin (H&E) and Immunohistochemistry (IHC) staining. H&E-stained images alone are often insufficient for cancer diagnosis, while IHC provides more comprehensive information for confirmed diagnosis. To address the challenges of limited IHC resources and high-cost consumption, we aim to generate virtual IHC images from H&E-stained images. In practice, it is difficult to perform multiple stains on the same tissue section, making it hard to obtain pixel-level matched data. To overcome this, we propose a dual-stream generative network that leverages pathological consistency constraints and a pathological representation network to extract pathological information and improve prediction accuracy. The network also incorporates structural similarity constraints and skip connections to enhance structural similarity. Additionally, we use stain unmixing results as annotated data, significantly reducing the workload of pathologists. We also conducted experiments to compare our model with models of similar functionality. In terms of pathological correlation, we have a lower Integrated Optical Density (IOD) and a higher Pearson-R, which are approximately 7.3% lower and 12.7% higher, respectively than the model with the best test results, so we have a higher pathological correlation. In terms of image quality, our Structural Similarity Index (SSIM) is higher than existing models, improving by approximately 8.8% compared to the model with the best test results, with higher image quality. These experiments show that our method has better stability and performance than the existing methods.**

*Keywords*—**stain transfer, generative network, breast cancer, histopathology**

## I. Introduction

In recent years, breast cancer has risen to become a severe health threat with both high incidence and mortality rates globally, particularly among women, occupying the top spot as the most frequently diagnosed cancer, surpassing lung cancer [1]. Given this dire situation, early detection and treatment are undoubtedly one of the critical pathways for breast cancer patients to improve their survival rates and quality of life. In this context, medical diagnostic technology has undoubtedly emerged as one of the vital means of timely identifying breast cancer patients and saving lives.

In the field of clinical diagnosis, Hematoxylin-Eosin (H&E) staining is one of the most crucial staining methods in histopathological examination. H&E staining uses the acid-base properties of the dye, hematoxylin staining the nucleus blue and dark purple, and eosin staining the extracellular matrix and cytoplasm pink. Other cellular structures form various colors and shades under the influence of these two stains, which helps to make the structural morphology of cells clearer and more discernible [2]. However, although H&E is a cost-effective staining method, it has limitations in terms of the composition information it provides and its lack of contrast, making it difficult to effectively distinguish specific proteins. Therefore, in such circumstances, pathologists require stains that can offer more information.

Immunohistochemistry (IHC) is a molecular-level staining technique based on the principle of specific antigen-antibody binding. It plays an indispensable role in defining malignant tumor types and providing therapeutic decisions. Specific antibodies bind to specific antigens in target tissues or cells, and then a marker is introduced through chemical reactions to make the antigen-antibody complex visible, achieving the purpose of staining. Markers used in IHC are generally related to cancer cell proliferation activities. For instance, Ki-67 protein can be

Manuscript received January 17, 2025; revised April 1, 2025; accepted May 14, 2025; published August 7, 2025.

utilized as a cellular marker in IHC examinations. It is closely related to cell proliferation [3]. In IHC, Ki-67-positive tumor cells are stained brown, while Ki-67-negative cells appear blue.

In the existing experiments, the high cost makes it difficult to promote the diagnosis of IHC [4]. However, the sole use of H&E diagnosis is related to the experience of pathologists and is prone to disputes [5]. The scarcity of such experimental resources and human resources has led to difficulty in obtaining paired datasets of different staining with high accuracy. This means that once a tissue sample has undergone a certain staining procedure, it is difficult to revert it to its pre-stained state through methods such as decolorization to perform new staining. As a result, it is nearly impossible to obtain large-scale, pixel-level matched data of different stains. Consequently, we often rely on serial tissue sections to study information about the same tissue under different staining conditions. Serial tissue sections involve cutting the same tissue block into slices with a thickness between 3 μm and 5 μm, which are then stained using different methods. However, due to this method of continuous slicing along the same axis, spatial variability within cellular and tissue structures is unavoidable. We can only ensure that adjacent slices have similarity, not pixel-level matching, which naturally leads to difficulties in locating and analyzing different slices, reducing the overall accuracy of the experiment. IHC-stained images provide selective, high-contrast imaging of cellular and tissue components. However, compared to conventional H&E staining, this high-precision IHC examination requires more time and cost. Therefore, predicting IHC expression from H&E images is of great significance in reducing the cost and time of clinical diagnosis, thereby improving the survival outcomes of cancer patients.

In previous research, Cycle-GAN and its derivative models have been widely applied in the field of virtual staining. Zanjani *et al.* [6] and Runz *et al.* [7] developed color normalization methods based on Cycle-GAN to achieve consistent staining in histopathological images. Boyd *et al.* [8] enhanced Cycle-GAN with a Region of Interest (RoI) discriminator to improve stain transfer in whole slide images. Liu *et al.* [9] tested the GAN derivative model pix2pix on a breast cancer dataset, while Moghadam *et al.* [10] and Shen *et al.* [11] proposed multi-stain simulation models by incorporating additional modules into GANs. Veronese *et al.* combined PatchGAN and U-Net with Cycle-GAN to improve the quality of simulated images [12]. Dubey *et al.* used a Structural Cycle-GAN (SC-GAN) model to investigate glandular markers in colonic tissue [13]. Although these models have advanced image conversion, most remain unsupervised, lacking prior knowledge, and the supervised models impose weak constraints on pathological information, leading to limited prediction accuracy. In fact, the performance of Cycle-GAN in the field of pathology is not entirely satisfactory. Cycle-GAN only uses cycle consistency and adversarial loss to constrain feature loss during the transformation between two domains. Cycle consistency mainly aims to preserve the structure and content of the original image, avoiding the generation of similar images in the target domain. Adversarial loss primarily reduces the differences between the generated image and the target domain [14]. However, when dealing with more complex textures in style transfer, Cycle-GAN fails to effectively retain pathological features and other critical information from the source image.

We propose a dual-branch generative network for predicting breast cancer IHC markers from histopathological images, aiming to reduce data acquisition difficulty and improve prediction accuracy. The main innovation points of this paper are as follows.

(1) By performing coarse alignment of the data, we minimize the impact of spatial information discrepancies on training.
(2) Additionally, we simplify the acquisition of annotated data and reduce labor costs by de-mixing staining of IHC slice images. The model requires only a small amount of annotated data to effectively guide the learning of pathological features, thereby enhancing performance.
(3) Furthermore, we introduce pathological consistency and structural constraints in the pathological information learning process, combined with a supervised framework, to improve prediction accuracy.

## II. Related Work

Stain transfer networks can be divided into three main categories: traditional Generative Adversative Networks (GANs), Cycle-GAN, and stain unmixing while processing data.

### *A. Generative Adversarial Networks (GANs)*

Generative Adversarial Network (GAN) [15], as a powerful generative model, was born from Goodfellow *et al.*'s creative application of the "two-player zero-sum game" in game theory. This model optimizes the adversarial loss by constructing two adversarial networks, the generator and the discriminator, minimizing the difference between the output of the generator and the distribution of the real data while maximizing the discriminator's ability to distinguish between the real data and the generated data. This model has shown extraordinary potential and remarkable results in natural image processing. It has achieved impressive results in the areas of image generation, super-resolution reconstruction, image restoration, and image style migration. Now, this model is also used in medical image processing, especially when analyzing complex histopathological images, offering the possibility of improving the accuracy and efficiency of disease diagnosis.

### *B. Cycle-GAN*

During the training process, converting source images to the target style often relies on paired training data. In real-world situations, such paired data is often difficult to obtain. To address this issue, Zhu *et al.* [16] proposed Cycle-GAN, an unsupervised image-to-image translation method that does not require paired training samples.

Unlike traditional methods that depend on paired training data [15], Cycle-GAN can perform image translation without the need for one-to-one annotated data. Its core architecture consists of two generators and two discriminators, responsible for converting images from one domain to another and distinguishing the authenticity of the generated images. Through the combined optimization of adversarial loss and cycle consistency loss, Cycle-GAN not only generates realistic images in the target domain but also ensures that the translated images retain the content and structure of the original images. This feature makes Cycle-GAN particularly suitable for tasks in medical image processing, such as the translation between different stained pathology images, providing cross-domain image generation capabilities while preserving key pathological features.

### C. Stain Unmixing

Stain unmixing is a crucial image processing technique in digital pathology. Its purpose is to decompose stained pathology slide images into images of individual stain components, facilitating better analysis and quantification of specific biomarker expressions. In this paper, we use the Hematoxylin-Eosin-DAB (HED) decomposition method based on the Optical Density (OD) model proposed by Ruifrok and Johnston [17] to separate HER2 staining from other background stains. In our work, through linear decomposition in the optical density space, grayscale images are obtained and processed using Gaussian filtering and Otsu's thresholding method to generate corresponding masks. These masks are further processed using morphological closing operations to extract the target regions. The resulting outputs are ultimately used as our annotated data.

## III. METHODOLOGY

### A. Network Framework

The overall framework of the proposed method is shown in Fig. 1. We propose a dual-stream generative network that leverages pathological consistency constraints and a pathological representation network to extract pathological information and improve prediction accuracy. The network also incorporates structural similarity constraints and skip connections to enhance structural similarity. Additionally, we use stain unmixing results as annotated data, significantly reducing the workload of pathologists. We intercept the stained image as a block of size 512×512 containing three color channels as our input data. We then perform the initial processing through an encoder, which captures the initial features of the image by means of a convolutional layer with a 7×7 convolutional kernel and a step size of one. Then, in order to ensure that the image is spatially coherent and as detailed as possible in terms of structural information, we chose to downsample by replacing the general pooling layer with a convolution operation of step size 2. After three consecutive uses of this 3×3 convolutional kernel with a step size of 2, we reduced the size of the original image from 512×512 to 64×64, which preserved the key features while reducing the dimensionality of the data. In the feature extraction part, we also formed a module with high efficiency by five residual convolution blocks, which are marked with orange arrows in the image.

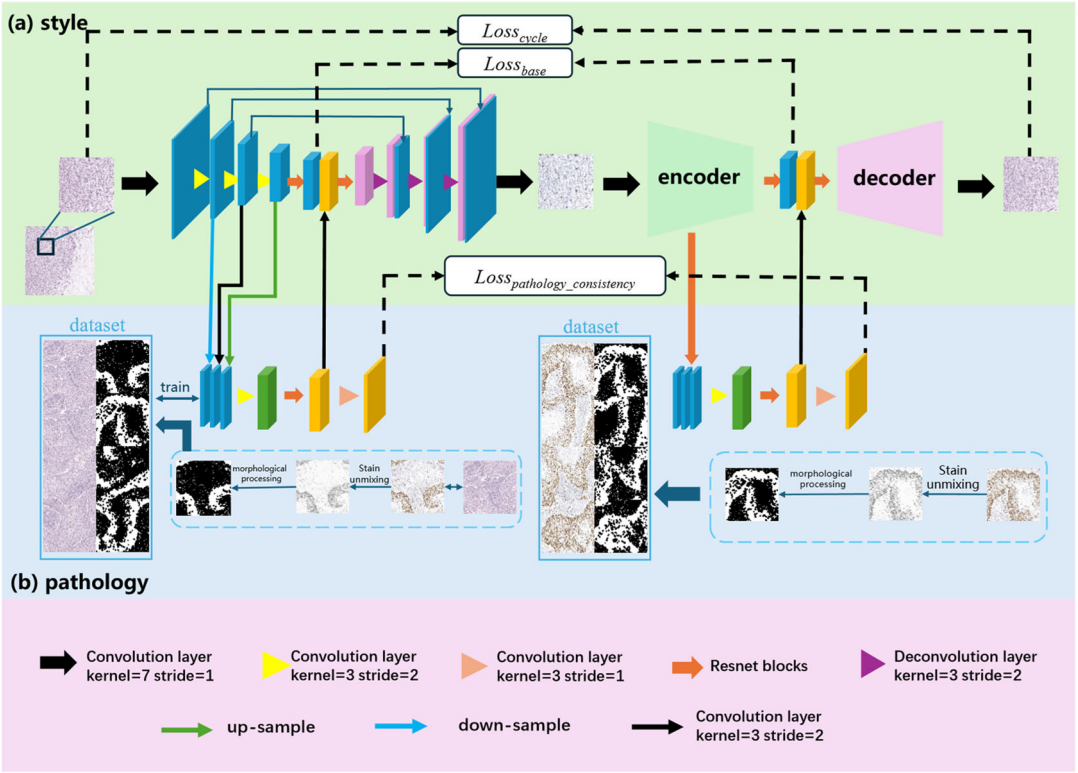


Fig. 1. Overview of the proposed method. Based on a dual-stream generative network, a pathological representation network is added. Joint training is performed using the stained de-mixing processed annotation data and the training set data. During the training process, adversarial loss, cycle-consistency loss, and pathological consistency loss are used as constraints.

In the pathology presentation network part, we select features at different resolution levels in the encoder (i.e., 256×256, 128×128, and 64×64 resolution layers) and uniformly resize these features to a pixel size of 72×72 by using B-spline interpolation and downsampling, and then process them by convolution with a step size of 2 and a kernel of 3, and again use a module of five residual convolution blocks to extract the pathology-related features in depth. of modules to deeply extract pathology-related feature information. We then generated a pathology representation map through a 3×3 convolutional layer and a Sigmoid activation function, which reflects the pathology features in the original image.

In the decoding phase, we first deeply integrate the extracted features by means of a module of five residual convolutional blocks to achieve full interaction and enhancement of the features. Subsequently, in order to recover the resolution of the feature map from the reduced 64×64 to 512×512, we justify the deconvolution operation with a step size of 2 and a convolution kernel of 3 as a way to enlarge the size of the feature map. Finally, we generated our virtual stained image by a 7×7 convolutional layer with Tanh activation function. This image maintains the structure and details of the original image while equally accurately reflecting the pathological information of the source image.

And at the same time, in order to improve the performance as well as stability of the network, we performed convolutional operations with a convolutional kernel of 3×3 after each convolutional layer, instance normalization as well as LeakyReLU activation layer. In addition, the jump-linking mechanism between the encoder and decoder not only facilitates the sharing of low-level information, but also provides a direct and effective path for the recovery of image details, thus ensuring that the virtual stained images we generate are realistic and rich in pathological information.

*B. Loss Function*

The overall objective function of our model includes four types of losses in total. In addition to the general adversarial loss $\mathcal{L}_{adv}$ and cycle consistency loss $\mathcal{L}_{cycle}$, we also added the circular structural consistency loss, pathological consistency loss $\mathcal{L}_{pathology}$, and basic spatial alignment loss $\mathcal{L}_{base}$.

$$\mathcal{L}(G_1,G_2,D_X,D_Y,P_X,P_Y)=\mathcal{L}_{adv}(G_1,D_Y)+\mathcal{L}_{adv}(G_2,D_X)+\lambda\mathcal{L}_{cycle}(G_1,G_2)+\beta\mathcal{L}_{pathology}(P_X,P_Y)+\gamma\mathcal{L}_{base}(G_1,G_2) \quad (1)$$

where $\lambda,\beta,\gamma$ are hyperparameters that reflect the importance of each item and are tuned to the specific task and dataset to achieve optimal performance. $G_1$ is a generator of $X \to Y$ and $G_2$ is a generator of $Y \to X$. $D_X$ and $D_Y$ are respectively the discriminators of $X$ data and $Y$ data. $P_X$ and $P_Y$ are separately pathological representation networks for input $X$ data and $Y$ data. From this basis, we target to solve the following optimization problem:

$$G_1^*,G_2^*=arg\min_{G_1,G_2,P_X,P_Y}\max_{D_X D_Y}\mathcal{L}(G_1,G_2,D_X,D_Y,P_X,P_Y) \quad (2)$$

Based on this method, we adjust the generators $G_1$ and $G_2$ to make them compete with each other while also cooperating, thereby continuously optimizing our data conversion effect. This enables our generators to accurately convert data styles while preserving the key features and information of the data.

The core objective of our network lies in generating images that embody the unique staining characteristics of the target domain, rather than simply performing pixel-level RGB value matching. To achieve this, we need to match the distributions between the two domains. For this purpose, we employ two generators, $G_1$ and $G_2$ which are responsible for the style transfer between $X$ and $Y$, and vice versa, respectively. An adversarial loss is utilized for optimization. During the transformation process of $G_1$: $X \to Y$, $G_1$ undergoes adversarial training with the discriminator $D_Y$, aiming to enhance the realism of the generated images, matching their distribution to that of the target domain $Y$, while preserving crucial features and information. The adversarial loss is defined as:

$$\mathcal{L}_{adv}(G_1,D_Y)=\mathbb{E}_x\left[\log D_Y(x)\right]+\mathbb{E}_x\left[\log(1-D_Y(G_1(x)))\right] \quad (3)$$

In this process, $G_1$ generates images $G_1(x)$ from the input $x$, aiming to make these generated images as indistinguishable as possible from the real images in domain $Y$. Meanwhile, task of $D_Y$ is to accurately distinguish which images are generated by $G_1$ and which are authentic images from domain $Y$. The goal of $G_1$ is to minimize the risk of its generated images being identified by $D_Y$, while $D_Y$ strives to maximize its ability to discern images produced by $G_1$. Correspondingly, the adversarial loss for the mapping function $G_2$ : $Y \to X$ and its corresponding discriminator is defined as:

$$\mathcal{L}_{adv}(G_2,D_X)=\mathbb{E}_y\left[\log D_Y(y)\right]+\mathbb{E}_y\left[\log(1-D_X(G_2(y)))\right] \quad (4)$$

Structural Similarity Index (SSIM) [18] as a Comprehensive Image Quality Assessment Tool. SSIM comprehensively evaluates image quality by considering three key dimensions: luminance, contrast, and structure, providing a relatively comprehensive similarity measure for each pixel. In order to further enhance the preservation of tissue structural details in images, we have defined a cyclic structural consistency loss function based on the SSIM index, which is incorporated into the original cyclic consistency loss framework [18]. The definition of SSIM is as follows:

$$ssim(a,b)=\frac{(2\mu_a\mu_b+c_1)(2\sigma_{ab}+c_2)}{(\mu_a^2+\mu_b^2+c_1)(\sigma_a^2+\sigma_b^2+c_2)} \quad (5)$$

where $\mu_a$ and $\mu_b$ represent the average values within an $N \times N$ sliding window centered on a specific pixel, while $\sigma_a$ and $\sigma_b$ represent the standard deviations within these windows. $\sigma_{ab}$ denotes the covariance between the two. $c_1$ and $c_2$ act as regularization parameters, ensuring the stability of calculations under low variance conditions, thereby leading to a more accurate similarity assessment. Based on this, our new cycle consistency loss can be expressed as:

$$\mathcal{L}_{cycle} = \left\{\mathbb{E}_x\left[\left\|G_2\left(G_1(x)\right)-x\right\|_1\right]+\mathbb{E}_y\left[\left\|G_1\left(G_2(y)\right)-y\right\|_1\right]\right\} + \left\{\mathbb{E}_x\left[1-ssim\left(G_2\left(G_1(x)\right),x\right)\right]+\mathbb{E}_y\left[1-ssim\left(G_1\left(G_2(y)\right),y\right)\right]\right\} \quad (6)$$

In our framework, our stain transformation process is divided into two steps: firstly, the step of removing the original stain, and secondly, the step of restaining. During this process, we need to ensure that after stain removal, image $x$ can be mapped to a shared base space state with its corresponding image $y$, ideally achieving $G_{1_{en}}(x) = o;\ G_{2_{en}}(y) = o$, thereby enabling both images to share the same abstract representations of anatomical structures and pathological features. To achieve this goal, we introduce a base space consistency alignment loss function between the encoders of $G_1$ and $G_2$. This function of $\mathcal{L}_{base}$ is based on the calculation of L1-norm, aiming to optimize the consistency of feature representations between the two generators at the base space level. The specific formula for this loss $\mathcal{L}_{base}$ is:

$$\mathcal{L}_{base} = \mathbb{E}_x\left[\left\|G_{2en}\left(G_1(x)\right)-G_{1en}(x)\right\|_1\right]+\mathbb{E}_y\left[\left\|G_{1en}\left(G_2(x)\right)-G_{2en}(y)\right\|_1\right] \quad (7)$$

where $G_1$ and $G_2$ are the generators for the $X \to Y$ and $Y \to X$ transformations, respectively. $G_{1_{en}}$ and $G_{2_{en}}$ represent the encoder parts of these two generators. The notation $\| \|_1$ denotes the *L1* distance.

In the analysis of multi-stained histopathological images, the issue of case consistency is a core concern. With this objective in mind, we have incorporated a pathological representation network into both generators to obtain pathological representation heatmaps of the input images. To further enhance the pathological consistency between the input images and the generated images, we have additionally introduced a loss function to reduce the differences in pathological representation heatmaps between the input image *x* and the generated image.

In our experimental process, the data we used were images of tissue sections that had undergone continuous slicing and were stained using both H&E staining and immunohistochemical HER2 staining techniques. These two staining methods yielded images with high structural consistency. Furthermore, we employed staining unmixing techniques on the images obtained from immunohistochemical HER2 staining and used them as a reference for annotating the H&E-stained images. Similarly, we also used the unmixed images from immunohistochemical HER2 staining as a reference for their own annotations. Based on this, we obtained the annotation databases $K_X$ and $K_Y$ derived from staining unmixing for training the pathological representation network $P_X\ and\ P_Y$. The definition of pathological consistency loss is as follows:

$$\mathcal{L}_{pathology} = \mathbb{E}_x\left[\left\|P_X(x)-P_Y\left(G_1(x)\right)\right\|_1\right]+\mathbb{E}_y\left[\left\|P_Y(y)-P_X\left(G_2(y)\right)\right\|_1\right] + \omega_1 * \mathrm{E}_{k\_x}\left[\left\|P_X(k\_x)-label(k\_x)\right\|_1\right]+\omega_2 * \mathrm{E}_{k\_y}\left[\left\|P_Y(k\_y)-label(k\_y)\right\|_1\right] \quad (8)$$

where $P_X\ and\ P_Y$ represent the pathological representation networks for domains $X\ and\ Y$ in generators $G_1$ and $G_2$, respectively. $k\_x$ and $k\_y$ are samples from the annotation databases $K_X$ and $K_Y$ where the label (*) represents the annotation for the continuous section of sample $*$. $\omega_1$ and $\omega_2$ are the weights in the supervised learning process, which are set to 2.0 in the experiment.

### C. Implementation Details

Our model was implemented on a computer equipped with an NVIDIA GeForce GTX 3090 GPU using the Python-based open-source deep learning library PyTorch. The input image size for the model is 1024×1024, and the parameters of the convolutional layers were initialized according to the method described in [19]. Additionally, according to [20], we set the hyperparameters of the overall loss function, with $\lambda$ = 10, $\beta$ = 10, and $\gamma$ = 5. The Adam optimizer was used to minimize the overall loss. The entire model is trained end-to-end using backpropagation. Given computational constraints, the batch size for the training dataset was set to 1, the batch size for the expert knowledge database was set to 4, and the initial learning rate was set to 0.0002, with an exponential decay rate of 0.9 and a decay period of 2 epochs.

## IV. Experiment and Result

### A. Datasets and Evaluation Metrics

In terms of datasets, we choose the publicly available Multi-IHC Stain Translation (MIST) dataset [21]. In the MIST dataset, we use the biomarker HER2, which contains 5642 paired samples from 64 WSIs. We selected 1000 pairs of samples and divided them equally into 500 samples each of the labeled data sets of H&E and HER2. Each part accounted for about 8.8% of the total sample. The final labeled data sets of H&E and HER2 were obtained by means of staining and unmixing. At the same time, 1000 samples were used as the test set, which accounts for about 17.7% of the total sample, and the remaining 3642 samples were used as the training set. It accounts for 64.55% of the dataset.

To verify whether our virtual staining results meet clinical requirements, we chose to evaluate the pathological relevance of the results using Integrated Optical Density (IOD), Mean Integrated Optical Density (mIOD), and Pearson's correlation coefficient (Pearson-R). Additionally, we assessed the quality of the final generated images using the Structural Similarity Index (SSIM) [18].

IOD is used to quantify and count the number of positive signals in the image [22], while mIOD represents the intensity of positive signals within a region [23]. We subtract the mIOD and IOD values from Ground Truth (GT). The closer the final value is to 0, the better the result. Pearson's correlation coefficient (Pearson-R) is employed to evaluate the pathological relevance between the staining results and protein expression levels across the entire region [14]. SSIM evaluates the quality of the image from multiple dimensions, including brightness, contrast, and structure. Based on these metrics, we can comprehensively assess the generated virtual staining results to determine whether they meet the requirements for clinical application.

### B. Experiment Result

The experimental results are shown in Fig. 2. We visually compared the virtual HER2-stained images

generated by different methods and observed that, under the pathology consistency constraint, the virtual HER2-stained images produced by our model more closely resembled the reference HER2-stained images in consecutive sections, reducing many noticeable recognition errors. This demonstrates that our pathology consistency constraint plays a significantly positive role in extracting pathological features, contributing to the improved accuracy of virtual HER2-stained images.

To demonstrate the superiority of our model, we compare its performance with the most advanced methods. The superiority of this method is proved by the final empirical results. It can be seen from Table I that the mIOD and IOD simulated by our model have the smallest deviation compared with the Ground Truth. At the same time, the Pearson-R score is much higher than that of the other two models, which reflects the highest pathological consistency of our model. In addition, our model has a higher SSIM. This shows that the virtual colored image generated by our model is closer to the original result. It can be seen that our model has higher performance than the existing models.

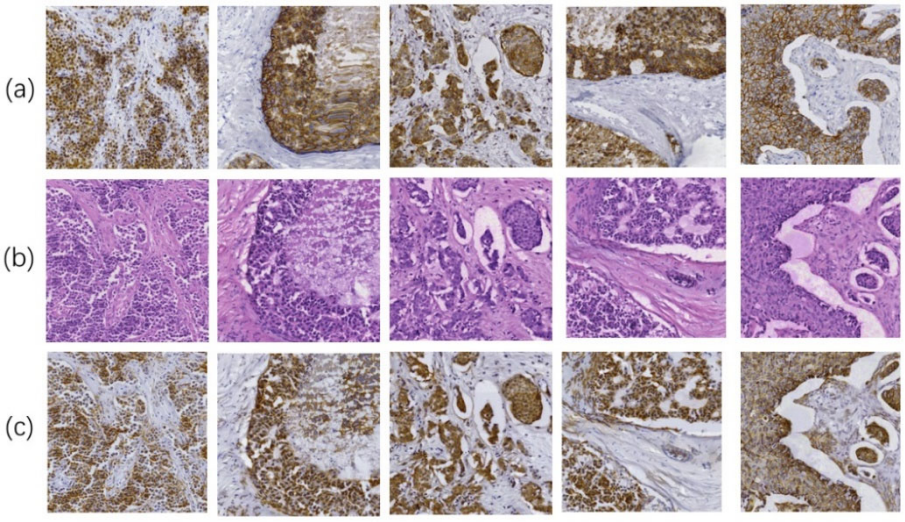


Fig. 2. Results of virtual HER2 staining images with different methods on MIST datasets. These from top to bottom are (a) Ground Truth, (b) source H&E-stained image and (c) generated HER2-stained image by our model.

TABLE I. QUANTITATIVE EVALUATION RESULTS OF VIRTUAL HER2 STAINING IMAGE ON MIST DATASET

| Methods | Pathological Relevance | | | Image Quality |
|---|---|---|---|---|
| | mIOD ↓ | IOD$_{\times 10^7}$ ↓ | Pearson-R ↑ | SSIM ↑ |
| ASP [21] | 0.5308 | 2.3781 | 0.5343 | 0.3837 |
| PSPStain [24] | 0.2014 | 1.0012 | 0.7543 | 0.4345 |
| Ours | 0.1873 | 0.9808 | 0.8638 | 0.4731 |

*C. Ablation Study*

To verify the significance of the pathology consistency constraint, we conducted an ablation study as shown in Fig. 3, by using a model that combines the standard SSIM constraint, skip connections, and Cycle-GAN as a comparison method. The same training tasks were performed under identical training datasets and hardware conditions. Cycle-GAN+SSIM model presents the model without skip connections, which means the model without complete pathological consistency. Meanwhile, Cycle-GAN+UNet [23] model presents the model without SSIM constraint. In this way, we prove the importance of each module in the model by removing some of the modules in the model. The experimental results show that with the addition of pathological representation network and structural similarity constraint module, the proposed method has higher accuracy in predicting IHC-marker expression.

According to Table II, through the analysis of the quantitative results, we found that in terms of pathological information extraction, our model, compared to the Cycle-GAN model constrained solely by SSIM, had a smaller difference when subtracted from the GT image, and the pathological information was closer to the ground truth. Furthermore, in the prediction of positive IHC-stained regions, the Pearson correlation coefficient of our model reached 0.8638, significantly higher than other comparison models. This indicates that the skip connection and pathological consistency constraint we introduce effectively improves the model's ability to extract pathological features. Additionally, in terms of image quality, our model also outperformed the comparison models which were constrained solely by UNet, better restoring the tissue structure of the original images. In summary, the pathological consistency constraint not only improved the accuracy of Cycle-GAN in predicting IHC images from H&E images but also enhanced the integrity of structural restoration.

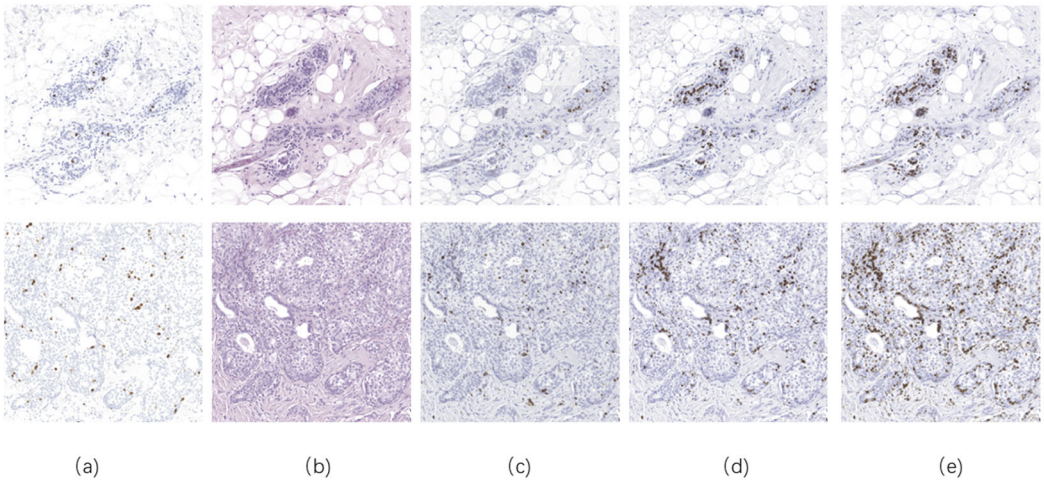


Fig. 3. Ablation study results of different experiments. The columns from left to right correspond to (a) Ground Truth, (b) input image, (c) our model results, (d)Cycle-GAN+SSIM results, and (e) Cycle-GAN+UNet results.

TABLE II. QUANTITATIVE EVALUATIONS USING TWO PATHOLOGY-RELATED METRICS AND ONE IMAGE QUALITY METRICS FOR THE ABLATION STUDIES

| Methods | Pathological Relevance | | | Image Quality |
|---|---|---|---|---|
| | mIOD ↓ | $IOD_{\times 10^7}$ ↓ | Pearson-R ↑ | SSIM ↑ |
| Cycle-GAN+SSIM | 0.2308 | 1.3781 | 0.7734 | 0.4678 |
| Cycle-GAN+UNet | 0.2283 | 1.4102 | 0.7228 | 0.4573 |
| Ours | 0.1873 | 0.9808 | 0.8638 | 0.4731 |

## V. CONCLUSION

In this study, we proposed a dual-stream generative network that can effectively predict the virtual staining of pathological images from H&E images to IHC images. This model links H&E images from consecutive slices with IHC images through stain unmixing, allowing the unmixing results of IHC images to serve as initial annotation data, thereby significantly reducing the workload of pathologists. Additionally, the model incorporates a pathological consistency loss function to constrain the learning of pathological features, requiring only approximately 10% of the training data to be annotated, which considerably shortens the data preparation phase. Through extensive experiments, we validated that this method is highly effective at the current stage, outperforming state-of-the-art methods. In conclusion, the proposed method exhibits a good performance in predicting IHC biomarker.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Ziyang Jin conducted the experiments, performed data analysis, and contributed to writing the paper; Jiansheng Wang designed the research framework and participated in drafting the manuscript; Yan Wang and Qingli Li provided guidance on the experimental methodology and participated in critical discussions throughout the study. All authors had approved the final version.

## FUNDING

This work was supported by the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science—MOE, ECNU (Grant No. KLATASDS2406).

## REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021. https://doi.org/10.3322/caac.21660

[2] C. Teixeira Soares, U. Frederigue-Junior, and L. A. de Luca, "Anatomopathological analysis of sentinel and nonsentinel lymph nodes in breast cancer: Hematoxylin-Eosin versus immunohistochemistry," *International Journal of Surgical Pathology*, vol. 15, no. 4, pp. 358–365, 2007. https://doi.org/10.1177/1066896907302124

[3] R. A. Sheikh, B. H. Min, S. Yasmeen, R. Teplitz, H. Tesluk, B. H. Ruebner, M. Tobi, J. Hatfield, S. Fligiel, and M. J. Lawson, "Correlation of Ki-67, p53, and Adnab-9 immunohistochemical staining and ploidy with clinical and histopathologic features of severely dysplastic colorectal adenomas," *Dig. Dis. Sci.*, vol. 48, no. 1, pp. 223–229, 2003. https://doi.org/10.1023/a:1021727608133

[4] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rüschoff, and M. Claassen, "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," *Sci. Rep.*, vol. 8, no. 1, 12054, 2018. https://doi.org/10.1038/s41598-018-30535-1

[5] J. Eaden, K. Abrams, H. McKay, H. Denley, and J. Mayberry, "Inter-observer variation between general and specialist gastrointestinal pathologists when grading dysplasia in ulcerative colitis," *J. Pathol.*, vol. 194, no. 2, pp. 152–157, 2001. https://doi.org/10.1002/path.876

[6] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. W. M. Laak, and P. H. N. With, "Stain normalization of histopathology images using generative adversarial networks," in *Proc. 2018 IEEE 15th*

*International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 573–577.

[7] M. Runz, D. Rusche, S. Schmidt, M. R. Weihrauch, J. Hesser, and C. A. Weis, "Normalization of HE-stained histological images using cycle consistent generative adversarial networks," *Diagn. Pathol.*, vol. 16, no. 1, p. 71, 2021. https://doi.org/10.1186/s13000-021-01126-y

[8] J. Boyd, I. Villa, M.-C. Mathieu, E. Deutsch, N. Paragios, M. Vakalopoulou, and S. Christodoulidis, "Region-guided cycle GANs for stain transfer in whole slide images," arXiv Print, arXiv:2208.12847, 2022.

[9] S. Liu, C. Zhu, F. Xu, X. Jia, Z. Shi, and M. Jin, "BCI: Breast cancer immunohistochemical image generation through pyramid Pix2pix," arXiv Print, arXiv:2204.11425, 2022.

[10] A. Z. Moghadam, H. Azarnoush, S. A. Seyyedsalehi, and M. Havaei, "Stain transfer using generative adversarial networks and disentangled features," *Comput. Biol. Med.*, vol. 142, 105219, 2022. https://doi.org/10.1016/j.compbiomed.2022.105219

[11] Y. Shen and J. Ke, "StainDiff: Transfer stain styles of histology images with denoising diffusion probabilistic models and self-ensemble," in *Lecture Notes in Computer Science*, 2023, pp. 567–577.

[12] L. Veronese, I. Poles, E. D. Arnese, and M. D. Santambrogio, "Stain transfer using CycleGAN for histopathological images," in *Proc. IEEE EUROCON 2023—20th International Conference on Smart Technologies*, 2023, pp. 1–5.

[13] S. Dubey, T. Kataria, B. Knudsen, and S. Y. Elhabian, "Structural cycle GAN for virtual immunohistochemistry staining of gland markers in the colon," arXiv Print, arXiv:2308.13182, 2023.

[14] S. Liu, B. Zhang, Y. Liu, A. Han, H. Shi, T. Guan, and Y. He, "Unpaired stain transfer using pathology-consistent constrained generative adversarial networks," *IEEE Trans. Med. Imaging*, vol. 40, no. 8, pp. 1977–1989, 2021. https://doi.org/10.1109/TMI.2021.3069874

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," arXiv Print, arXiv:1406.2661, 2014.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," arXiv Print, arXiv:1703.10593, 2017.

[17] A. C. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, 2001.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004. https://doi.org/10.1109/tip.2003.819861

[19] L. Liu, X. Liu, J. Gao, W. Chen, and J. Han, "Understanding the difficulty of training transformers," arXiv Print, arXiv:2004.08249, 2020.

[20] A. Lahiani, N. Navab, S. Albarqouni, and E. Klaiman, "Perceptual embedding consistency for seamless reconstruction of tilewise style transfer," in *Proc. MICCAI 2019*, 2019, pp. 568–576.

[21] F. Li, Z. Hu, W. Chen, and A. Kak, "Adaptive supervised PatchNCE loss for learning H&E-to-IHC stain translation with inconsistent ground truth image pairs," arXiv Print, arXiv:2303.06193, 2023.

[22] C. J. Wang, A. Zhou, H. Holmqvist, Y. Zhang, G. Li, and X.-F. Sun, "Survivin expression quantified by image pro-plus compared with visual assessment," *Appl. Immunohistochem. Mol. Morphol.*, vol. 17, no. 6, pp. 540–545, 2009.

[23] R. Zhang, Y. Cao, Y. Li, Z. Liu, J. Wang, J. He, C. Zhang, X. Sui, P. Zhang, L. Cui, and S. Li, "MVFStain: Multiple virtual functional stain histopathology images generation based on specific domain mapping," *Med. Image Anal.*, vol. 80, 102520, 2022. https://doi.org/10.1016/j.media.2022.102520

[24] F. Chen, R. Zhang, B. Zheng, Y. Sun, J. He, and W. Qin, "Pathological semantics-preserving learning for H&E-to-IHC virtual staining," arXiv Print, arXiv:2407.03655, 2024.