

## Article

# A Virtual Staining Method Based on Self-Supervised GAN for Fourier Ptychographic Microscopy Colorful Imaging

Yan Wang <sup>†</sup>, Nan Guan <sup>\*,†</sup>, Jie Li and Xiaoli Wang

Electronics Information Engineering College, Changchun University, Changchun 130022, China

\* Correspondence: 210401118@mails.ccu.edu.cn

† These authors contributed equally to this work.

**Abstract:** Fourier ptychographic microscopy (FPM) is a computational imaging technology that has endless vitality and application potential in digital pathology. Colored pathological image analysis is the foundation of clinical diagnosis, basic research, and most biomedical problems. However, the current colorful FPM reconstruction methods are time-inefficient, resulting in poor image quality due to optical interference and reconstruction errors. This paper combines coloring and FPM to propose a self-supervised generative adversarial network (GAN) for FPM color reconstruction. We design a generator based on the efficient channel residual (ECR) block to adaptively obtain efficient cross-channel interaction information in a lightweight manner, and we introduce content-consistency loss to learn the high-frequency information of the image and improve the image quality of the staining. Furthermore, the effectiveness of our proposed method is demonstrated through objective indicators and visual evaluations.

**Keywords:** Fourier ptychographic microscopy; self-supervised GAN; virtual staining; digital pathology



**Citation:** Wang, Y.; Guan, N.; Li, J.; Wang, X. A Virtual Staining Method Based on Self-Supervised GAN for Fourier Ptychographic Microscopy Colorful Imaging. *Appl. Sci.* **2024**, *14*, 1662. <https://doi.org/10.3390/app14041662>

Academic Editor: Nektarios A. Valous

Received: 14 January 2024

Revised: 6 February 2024

Accepted: 16 February 2024

Published: 19 February 2024



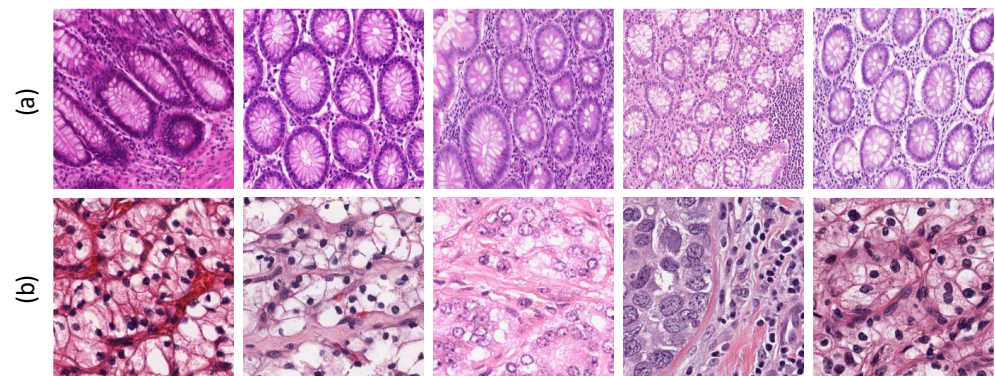
**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The use of light microscopy for minimally invasive high-resolution imaging is fundamental to human understanding of biological systems and processes. In clinical pathology, analyzing high-resolution color pathological images from optical microscopes remains a widely used standard method for diagnosing diseases ranging from cancer to blood-borne infections. To perform quantitative and accelerated pathological image analysis, the whole slide imaging (WSI) [1] system is gradually replacing traditional optical microscopes. Fourier ptychographic microscopy (FPM) [2,3], which has been popularized in recent years, is considered a high-throughput imaging technology that does not require mechanical scanning and has more advantages over conventional whole slide imaging systems in digital pathology. First of all, FPM does not require mechanical scanning to achieve both high resolution (HR) and a large field of view, and it can simply add LED arrays to other microscopes to set up experimental platforms. Secondly, the digital refocusing process is utilized by FPM to correct sample defocusing after acquisition [2,4]. Furthermore, the phase information recovered from the FPM provides more localized scattering details of the sample, which enhances the accuracy and reliability of pathological diagnosis [5,6].

However, the application of FPM in pathology still faces some challenges [6]. On the one hand, the time efficiency of FPM acquisition and reconstruction is relatively poor. FPM usually requires continuous acquisition with red, green, and blue separate illumination, composing hundreds of low-resolution (LR) images for better color reconstruction, leading to slower acquisition and reconstruction speeds. On the other hand, FPM color reconstruction images are usually of poor quality due to coherent artifacts caused by light interference and reconstruction errors. Furthermore, the GlaS dataset in Figure 1 is pathological sections of the differentiation tissue of colon cancer glands stained with hematoxylin and eosin

(H&E), and the MoNuSeg dataset is composed of various H&E stained histological images across multiple patients and organs obtained from different hospitals. It can be observed in Figure 1 that the color reconstruction results of the same batch of pathological sections will have significant visual differences due to the influence of factors such as collection personnel, technology, and the external environment. In biological and clinical medicine applications, the color information of the sample after staining with reagents helps the observer quickly locate the regions of interest and interpret the relevant information about tissues and cells. Therefore, we use a virtual staining method for the colorful reconstruction of FPM single-channel grayscale images.



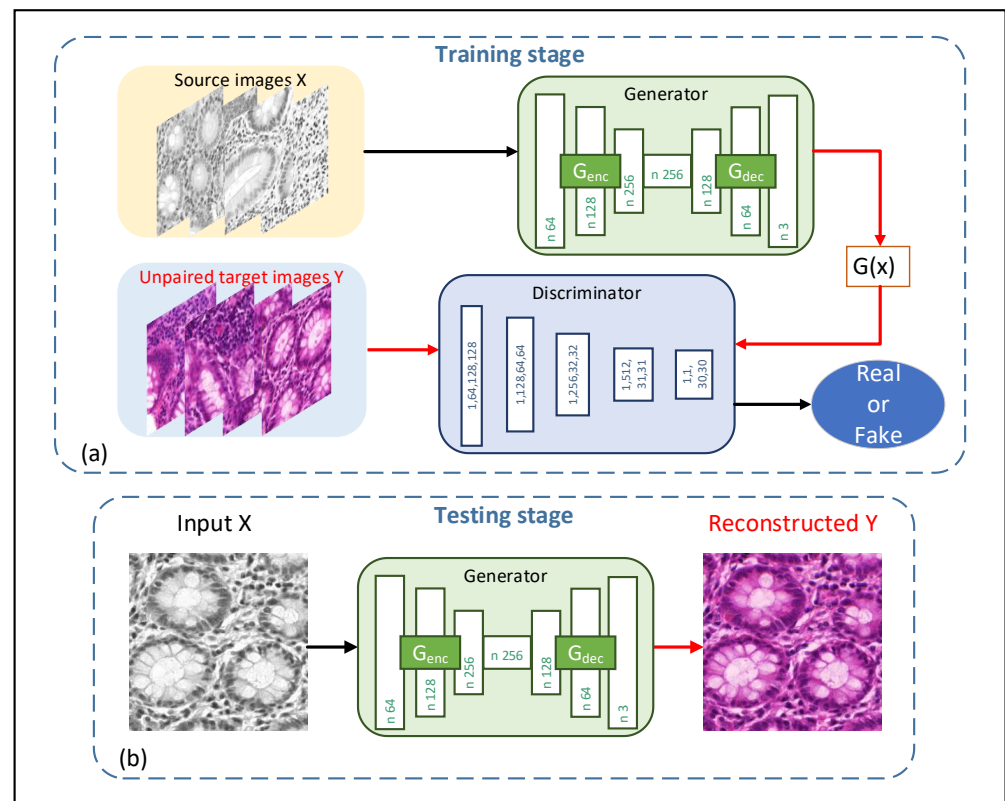
**Figure 1.** Different H&E color styles on the two datasets. (a) GlaS dataset, (b) MoNuSeg dataset.

Recently, significant advances have been made in the field of deep learning, paving the way for the application of neural networks in a variety of image-processing tasks, including virtual tissue staining. To better obtain the color information of FPM-stained tissue sections, image translation can be applied to convert the recovered monochromatic FPM image to the style of regular incoherent microscopy. The mathematical mapping between input and output [7] can be found accurately by this deep learning technology based on the deep convolutional neural network (CNN), and the coherence artifacts of FPM recovery can also be reduced [8], thus improving the quality of colored FPM pathology images.

Since its inception, the unsupervised learning image-to-image translation model CycleGAN [9] has shown its powerful transformation ability between two image domains, which makes it a popular method of virtual staining in the field of biomedical imaging. However, the cyclic consistency [9] assumption dictates that the relationship between the two domains must be bijective [10], which is often not ideal, as it may limit the diversity of images. To overcome this limitation, contrastive learning has recently achieved optimal progress [11,12] in the field of self-supervised representation learning. Subsequently, contrastive unpaired translation (CUT) [13] introduced contrastive learning to maximize the learning of mutual information between the two image domains for unpaired image translation.

In this paper, we propose a self-supervised generative adversarial network (GAN) based on contrastive learning for the virtual staining of single-channel images of FPM. While CUT has demonstrated the efficiency of contrastive learning, it may not efficiently capture domain gaps, color, and the high-frequency information of the images in our application. To further take advantage of contrastive learning while avoiding the disadvantages of cyclic consistency, our method trains a GAN with efficient channel residual (ECR) block and content-consistency loss to learn the cross-channel and high-frequency information of images more effectively, as well as to reduce coherent artifacts, and improve the image quality. The network training and testing processes are illustrated in Figure 2. This method provides a better foundation for subsequent research on medical cell recognition, segmentation, classification, and other related fields. The main contributions of this paper are listed below:

- We adopt the unpaired learning architecture of CUT to treat the virtual staining of the FPM pathology images as a transformation from a single-channel image to a color image. Since CUT cannot accurately identify the light-colored part of the image and will produce unnecessary artifacts, we design the ECR block in the generator to extract cross-channel interaction information and improve the network performance.
- CUT cannot effectively capture the edges and details of the image. We introduce the content-consistency loss based on multi-scale structural similarity (MS-SSIM) to avoid feature distortion between input and output images and enhance the high-frequency information of images.



**Figure 2.** Training and testing process framework for the proposed method. (a) Train one generator-discriminator pair using two unpaired datasets (X and Y). (b) The generator converts a FPM single-channel grayscale image into a three-channel virtual stained image with improved image quality.

We conducted comparative experiments to demonstrate the qualitative and quantitative advantages of our method in single-channel FPM virtual staining tasks compared to other methods. In addition, our method theoretically reduces the collection time of 2/3 of color FPM, standardizes the staining process, and improves the objectivity of disease diagnosis. Moreover, we do not need to obtain paired data, as simply using unpaired input domain X and output domain Y images can achieve good staining results, which effectively solves the problem of difficult matching data acquisition, especially in digital pathology.

## 2. Related Work

### 2.1. Color-Based FPM Imaging

The combination of virtual staining and FPM can improve imaging efficiency while providing further convenience for the detection and diagnosis of pathological sections. A classical coloring method [14] is used to restore the HR image at three wavelengths with a monochromatic camera and then synthesize the HR full-color image. The Wavelength-Multiplexed FPM (WMFPM) [15] scheme was proposed by Dong et al. with a monochrome camera and multi-wavelength simultaneous illuminations. A color transfer-based FPM

virtual staining method called CFPM [16] was reported by Gaod et al. The CFPM scheme sacrifices only about 0.4% imaging accuracy, but the efficiency is improved by three times.

Different deep learning methods represented by CNN have also been widely used in FPM. Zhang et al. [17] proposed color FPM based on a 3D CNN, using the trained CNN to establish the mapping relationship between monochrome HR images and color LR images to generate color HR images. Wang et al. [18] proposed the application of cyclic consistency to generate the adversarial network CycleGAN in FPM imaging, enabling virtual bright field and fluorescence staining of single-channel images. Zhang et al. [19] designed a CNN model for reconstructing FPM HR color pathology images and phase maps.

## 2.2. Image-to-Image Translation

The application of deep learning in virtual staining is mainly reflected in image-to-image translation. The introduction of this model effectively promotes the development of virtual staining. GANs [20] have been widely used in a series of image applications, achieving significant results in the field of image-to-image translation in particular. Generally speaking, image-to-image translation can be categorized into two methods, i.e., supervised learning and unsupervised learning methods. The widespread application of supervised methods is limited because it requires a large amount of paired data for training.

In recent years, unsupervised methods have become a hot topic of focus for researchers. UNIT [21] proposed a shared latent space assumption that generates high-quality translations by utilizing shared information between two domains. CycleGAN proposed the cyclic consistency assumption, which learns the mapping relationship between two domains through two generator networks and two adversarial discriminator networks to improve the quality and consistency of image conversion. This assumption requires the transformation to be reversible and consistent, resulting in strong limitations. Recently, to alleviate the problems caused by cyclic consistency, some methods [22,23] have adopted different constraints from different aspects in an attempt to break this cycle.

## 2.3. Contrastive Learning

Contrastive learning facilitates deep learning by training a model to maximize the similarity between similar samples and minimize the similarity between dissimilar samples in the field of self-supervised learning. This method is very suitable for image translation models, especially for image data whose output shape is the same as its input. The introduction of contrastive learning in image translation can preserve the content of input [23] and reduce mode collapse [24–26]. CUT introduced noise contrast estimation into the image translation tasks and achieved better performance than that based on cyclic consistency through learning the correspondence between input image patches and generated image patches for the first time.

## 2.4. Attention Mechanism

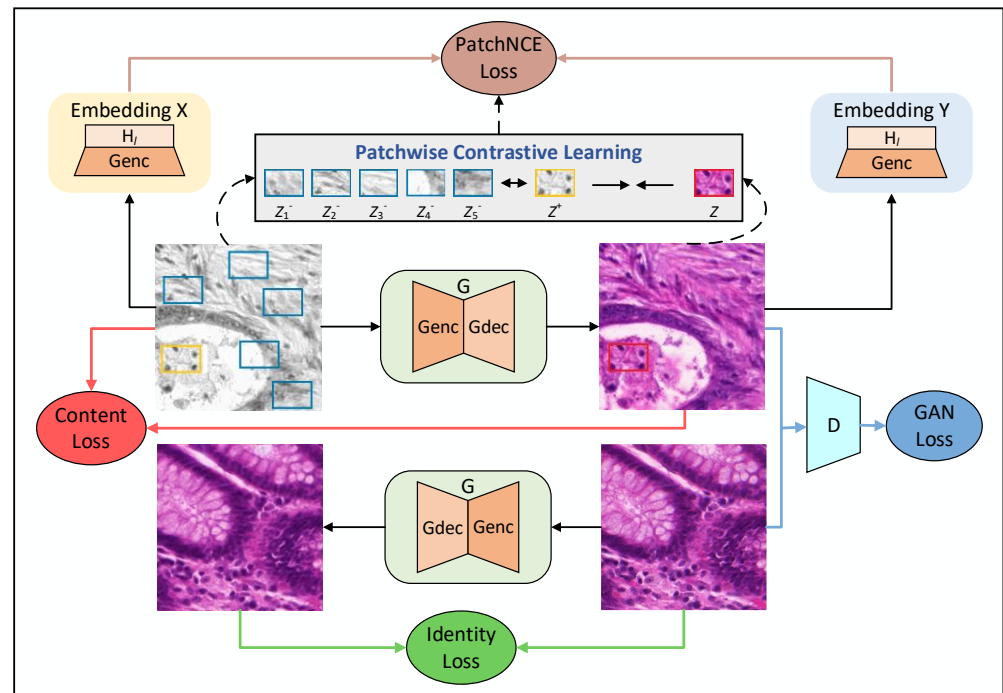
Inspired by the fact that humans can naturally and efficiently find significant areas in complex scenes, attention mechanisms have been introduced in computer vision to simulate this process to focus on important features and rationally assign weights. The attention mechanism can overcome the acceptance field limitation of CNN by capturing long-distance dependencies in images. There have also been several unprecedented efforts to integrate GAN with attention mechanisms in recent years. Zhang [27] obtained remote relationships in the images by self-attention. Spatial attention GAN (SPA-GAN) utilized attention maps from the discriminator to guide the generator's focus on discriminative areas in the image [28]. Hu [29] designed a query-selected attention (QS-Attn) module to intentionally select important anchors for comparative learning. Torbunov [30] equipped the generator of CycleGAN with a Vision Transformer (ViT) to improve non-local pattern learning and network performance.



### 3. Proposed Method

#### 3.1. Network Architecture

Figure 3 shows the overall architecture of our proposed method, which has been improved based on the CUT model. The core concept is to use only one generator (G) and discriminator (D) pair to map features in one direction, which greatly improves memory efficiency and training speed. We utilize G to convert the single-channel FPM grayscale images (input domain X) into high-resolution FPM color images (output domain Y). The true color image of the target domain Y and the generated color image of the input domain X are distinguished by the trained D.



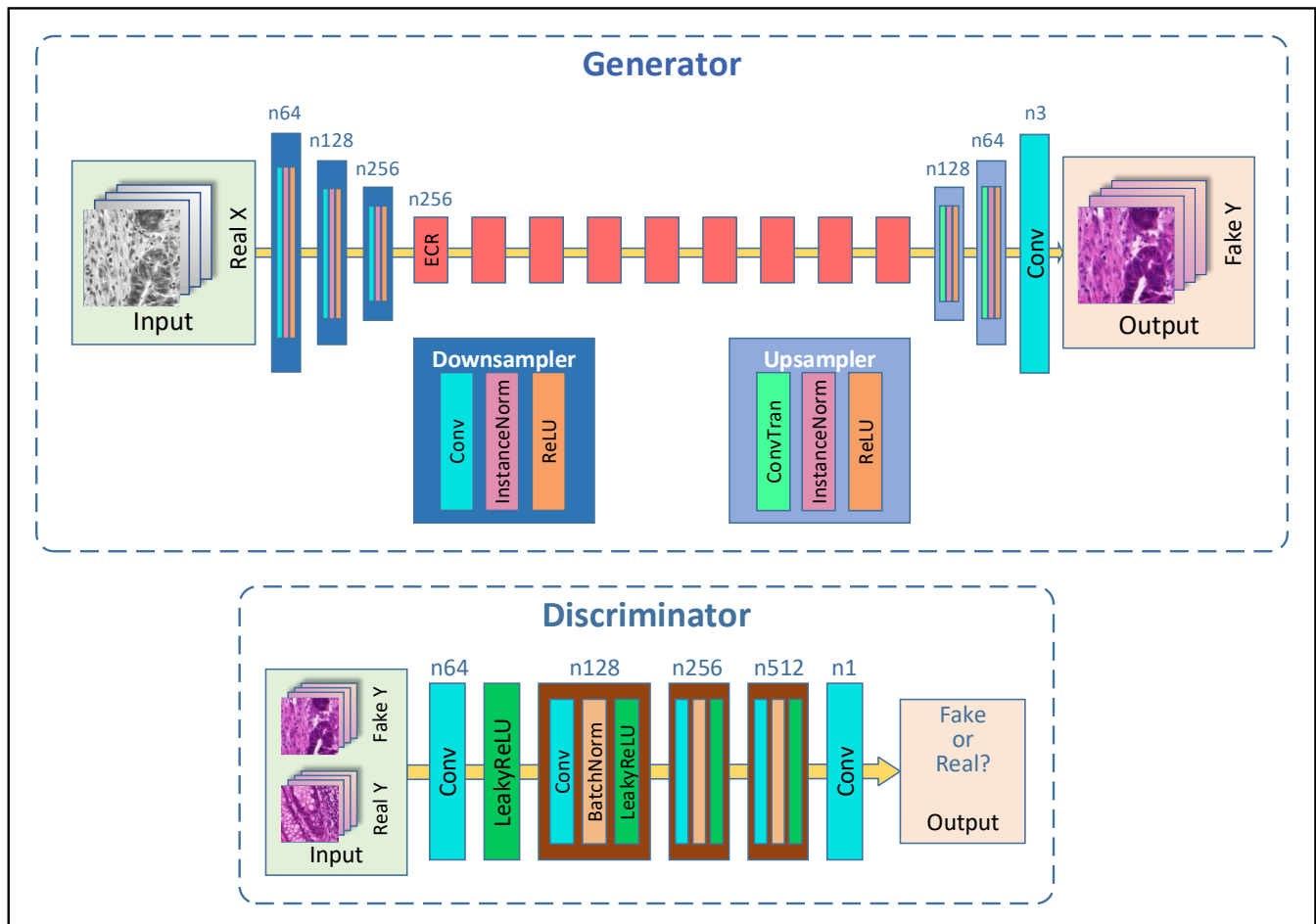
**Figure 3.** The overall architecture of the network. The generator G is broken down into two components: an encoder  $G_{enc}$  and a decoder  $G_{dec}$ . We use  $G_{enc}$  and the MLP ( $H_f$ ) as the embedding of domain X and domain Y. We expect the PatchNCE loss to make the red patch in the fake image generated by generator G more consistent with the yellow patch in the real input image while minimizing the similarity with other blue patches.

The generator and discriminator structures of this paper are shown in Figure 4. Generator G includes three downsampling modules, nine ECR blocks, two upsampling modules, and one convolutional layer. The downsampling module extracts low-frequency features of the image through convolution operation and enters it as input into the ECR. The ECR block is formed from a residual block and a channel attention block through the residual connection, which adaptively combines similar features of the image to provide higher-level semantic information. The upsampling module reconstructs the details and resolution of the image through transposed convolution operations. We utilize PatchGAN [31] as the discriminator D, and the receptive field of PatchGAN is set to  $70 \times 70$ , which makes PatchGAN faster and still guides the generator to produce realistic results.

#### 3.2. Efficient Channel Residual Block

In this paper, we expect the generator to be able to generate FPM color images similar to the domain Y. The ResNet-based [32] generator is widely used in the field of image-to-image translation. However, CUT failed to accurately identify the unstained part of the pathological section when staining the single-channel FPM input image, resulting in unnecessary artifacts and producing low-quality color images. The weight of each channel

in a deep neural network can be recalibrated by using channel attention. This allows for a more adaptive selection process for determining the object of focus in different feature maps, as each channel can represent a different object. However, most of the methods proposed in recent years have focused on designing more complex attention modules. It undoubtedly increases the complexity of the model while delivering better performance. For this purpose, we designed an ECR block to replace the residual block in the CUT.



**Figure 4.** The specific structure of generator and discriminator networks. The generator adopts an encoder–decoder architecture with nine ECR blocks. The discriminator utilizes the classic PatchGAN network, which has five convolutional layers.

As shown in Figure 5, we introduce ECANet [33] into the residual block of the generator. Concretely speaking, the ECR block consists of two parts by the shortcut connection. One part is the original residual block, which includes two convolutional layers and one RELU layer, and the other is the efficient channel attention (ECA) block, which includes a global average pooling layer, a one-dimensional convolutional layer with an adaptive convolutional kernel, and a sigmoid layer. These two parts each have a shortcut connection between the input and output.

ResNet can truly deepen the depth of convolutional neural networks without reducing network accuracy, and its powerful representation ability can significantly improve network performance in many computer vision fields. The ECA block improves network performance by avoiding reducing channel dimensions to learn effective channel attention while obtaining cross-channel interaction information in an extremely lightweight manner. We add identity maps to the ECR block, which not only ensures that there is no gradient vanishing problem during backpropagation but also shifts the network to learning residual

functions, ensuring a smoother decision function for the entire network and improving generalization performance.

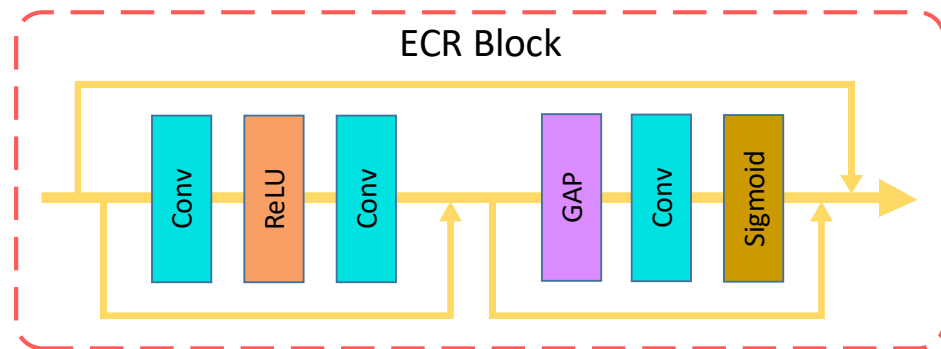


Figure 5. The structure of the ECR block.

### 3.3. Loss Functions

#### 3.3.1. PatchNCE Loss

As with the setting of CUT, we enable the network to maximize the mutual information between the input and output corresponding patches by using a noise contrastive estimation framework [34]. The core idea of contrast learning is to relate the two signals of a “query” and its “positive” example, not associated with other examples in the dataset (called “negatives”). Firstly, the probability of choosing “positive” instead of other “negatives” is represented by calculating the cross-entropy loss [35] as follows:

$$\ell(v, v^+, v^-) = -\log \left[ \frac{\exp(v \cdot v^+ / \tau)}{\exp(v \cdot v^+ / \tau) + \sum_{n=1}^N \exp(v \cdot v_n^- / \tau)} \right] \tag{1}$$

where  $v, v^+ \in R^K$ , and  $v^- \in R^{N \times K}$  indicate mapping query, positive, and  $N$  negatives to  $K$ -dimensional vectors, respectively.  $v_n^- \in R^K$  denotes the  $n$ -th negative.  $\tau$  is the parameter used to scale the distance between the query and the other examples, with the default value set to 0.07 in our experiments.

The  $G_{enc}$  and  $H_l$  are used to share weights to extract features from domain  $X$  and domain  $Y$ . The  $L$  layers are selected first and pass the feature mappings through  $H_l$ , as used in SimCLR [12]. A feature stack  $\{Z_l\}_L = \{H_l(G_{enc}^l(x))\}_L$  is then generated, where the output of the  $l$ -th selection layer is represented by  $G_{enc}^l$ . Similarly, we encode the output image  $G(x)$  belonging to domain  $Y$  into  $\{\hat{Z}_l\}_L = \{H_l(G_{enc}^l(G(x)))\}_L$ . Our goal is to match the input and output patches corresponding to a specific position. The PatchNCE Loss [13] for the one-way mapping in this article is represented as follows:

$$L_{PatchNCE}(G, H, X) = E_{x \sim X} \sum_{l=1}^L \sum_{s=1}^{S_l} \ell(\hat{Z}_l^s, Z_l^s, Z_l^{s \setminus s}) \tag{2}$$

where the number of spatial positions in each layer is represented by  $S_l$ .  $Z_l^s \in R^{C_l}$  represents the corresponding feature (“positive”),  $Z_l^{s \setminus s} \in R^{(S_l-1) \times C_l}$  represents other features (“negatives”), where the number of channels in each layer is represented by  $C_l$ .

#### 3.3.2. Content-Consistency Loss

Although CUT has produced competitive results with less computational costs for training, we believe that the model’s performance is limited due to certain design options. CUT may not be able to capture domain gaps efficiently, thus missing edges and details

of the image data, as one embedding is shared in two distinct image domains (domain  $X$  and domain  $Y$ ). For the above problems, the content-consistency loss is designed in the loss function for joint training. The content-consistency loss derived from the multiscale structural similarity index (MS-SSIM) enhances the high-frequency information of the image [36]. The definition of content-consistency loss is as below:

$$L_{\text{Content}}(G, X) = E_{x \sim X}[1 - \text{msSSIM}(G(x), x)] \quad (3)$$

where the content change between the generated color image  $G(x)$  and the input single-channel image  $x \in X$  is reduced as much as possible. We aim to avoid color inversion and feature distortion between input and output images through content-consistency loss [37].

### 3.3.3. Adversarial Loss

We use the adversarial loss [20] to make the output image  $G(x)$  of the generator  $G$  as visually identical as possible to the color image of the target domain  $Y$ . For the one-way mapping with the generator/discriminator ( $G/D$ ) pair, the GAN loss is represented as follows:

$$L_{\text{GAN}}(G, D, X, Y) = E_{y \sim Y}[\log D(y)] + E_{x \sim X}[\log(1 - D(G(x)))] \quad (4)$$

where  $G$  tries to generate images  $G(x)$  that look similar to images from domain  $Y$ , while  $D$  aims to distinguish between translated samples  $G(x)$  and real samples  $y$ .

### 3.3.4. Identity Loss

We add an identity loss [10] to avoid unnecessary changes from generator  $G$ . The main idea of identity loss is to take the image  $y$  of domain  $Y$  as the input of  $G$ , and the output image  $G(y)$  should be as consistent with the input image  $y$  as possible, without producing other content.

$$L_{\text{Identity}}(G, Y) = E_{y \sim Y}[\|G(y) - y\|_1] \quad (5)$$

The input and output mappings are encouraged to preserve color and brightness composition by such an identity loss.

## 3.4. Overall Objective

Our ultimate goal is to make the patches of the network input and output images learn the same correspondence, thereby generating more realistic color images. The total loss is estimated by:

$$L(G, D, H) = \lambda_{\text{GAN}}L_{\text{GAN}}(G, D, X, Y) + \lambda_{\text{NCE}}L_{\text{PatchNCE}}(G, H, X) + \lambda_{\text{CCL}}L_{\text{Content}}(G, X) + \lambda_{\text{Idt}}L_{\text{Identity}}(G, Y) \quad (6)$$

where  $\lambda_{\text{GAN}}$ ,  $\lambda_{\text{NCE}}$ ,  $\lambda_{\text{CCL}}$  and  $\lambda_{\text{Idt}}$  are the weight coefficients of adversarial loss, patchNCE loss, content-consistency loss and identity loss, respectively. The weight coefficients of losses are set to  $\lambda_{\text{GAN}} = 1$ ,  $\lambda_{\text{NCE}} = 10$ ,  $\lambda_{\text{CCL}} = 1$ , and  $\lambda_{\text{Idt}} = 1$  by default.

## 4. Experiments

### 4.1. Datasets

We need to construct a large-scale dataset containing a training set and a test set to train the neural network. However, directly collecting high-resolution color images from FPM has a series of disadvantages: (1) the amount of raw data collected is too large to build datasets of sufficient size; (2) the ground truth of the dataset is obtained through traditional reconstruction methods, so the effect of network reconstruction is limited by traditional methods; (3) the dataset is bound to a specific system, so it is difficult to flexibly apply to other systems.



To avoid the above problems, we directly used the publicly available dataset GlaS@MICCAI' 2015 [38] as the ground truth (the image resolution was  $775 \times 522$ ) and generated the corresponding single-channel high-resolution input data through a Fourier ptychographic microscopy imaging model simulation. Our simulation dataset includes 1363 single-channel FPM images and 1362 unpaired color FPM images. Additionally, 1208 single-channel grayscale images and 1279 color images were used for network training, while the rest were used for testing.

#### 4.2. Training Settings

Our proposed model was optimized and improved based on CUT. To better compare with the baseline, we trained our model primarily using the settings specified in CUT. More specifically, the initial learning rate was set to 0.0002, and due to GPU memory limitations, the batch size was set to 1. Because the Adam [39] optimizer has the advantages of an adaptive adjusted learning rate, integrated gradient and momentum, robustness, and parallel computation-friendliness, it was selected in this paper to adaptively update the model parameters. The size of all training images was resized to  $256 \times 256$  pixels and 200 epochs were trained. Cropping the image to  $256 \times 256$  was done because the size is large enough to include the main details in the image while maintaining computational efficiency, making it easy to input them into the neural network for training and processing. The learning rate began to decay linearly when the network was trained to half of the total epochs. The model presented in this paper was deployed and trained on the PyTorch deep learning framework and a single NVIDIA GPU (RTX 3080, 10 GB).

#### 4.3. Comparison Results with Unsupervised Deep Learning Methods

Here, we provide a comprehensive qualitative and quantitative evaluation of our approach by comparing it to recent state-of-the-art unsupervised methods, such as CUT [14], CycleGAN [10], UNIT [21], GCGAN [22], and DCLGAN [23]. These methods are retrained in our datasets and evaluated using ground truth values.

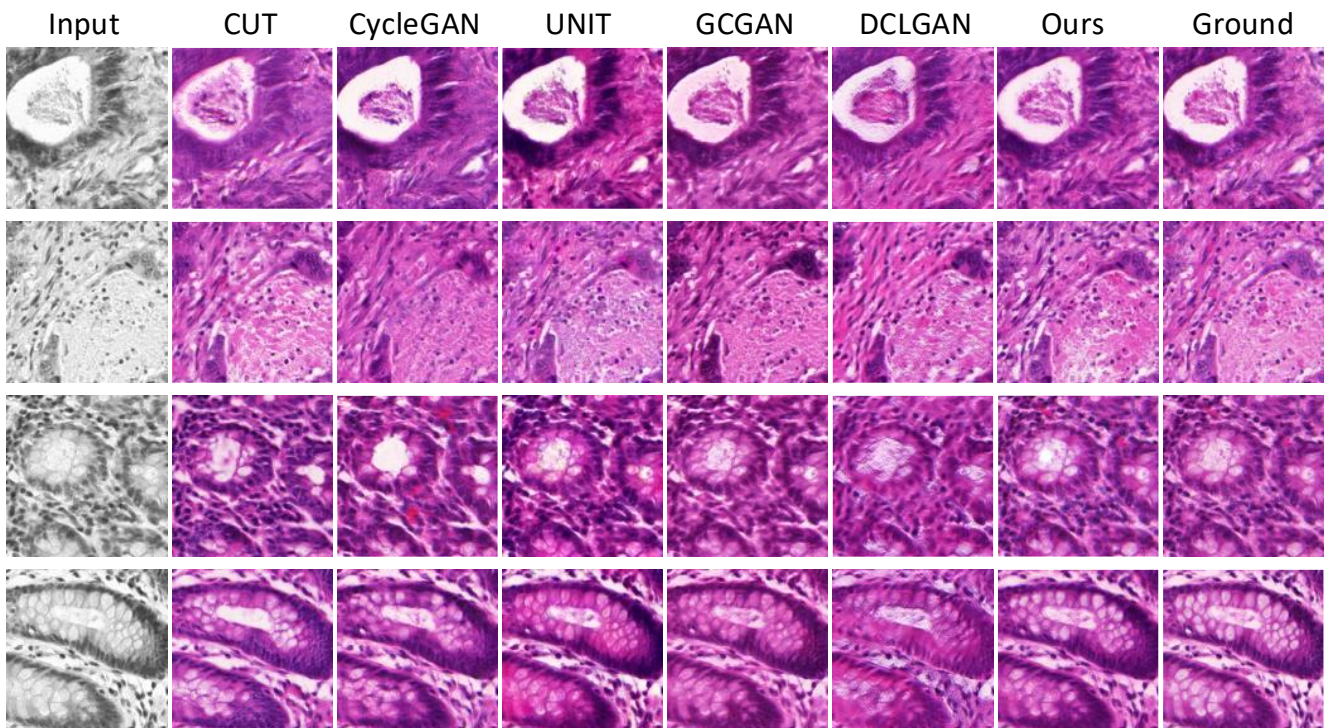
##### 4.3.1. Visual Comparisons

Figure 6 illustrates the results of the qualitative comparison of the different contrasting methods. Our method successfully stains the unpaired single-channel FPM input image into a three-channel color FPM target domain image through image translation. The method proposed in this article can preserve the structure and content of the input image to the maximum extent while performing good virtual coloring on the input image, achieving the highest color image quality.

Specifically, CUT cannot accurately identify the unstained part of the image. CycleGAN and UNIT can recognize the white part of pathological sections well, but lose the texture details inside. DCLGAN introduces artifacts and has low image quality. In addition, all baselines have a certain difference in color compared to the ground truth value. Our proposed method has the highest visual similarity, which proves the effectiveness of the network.

##### 4.3.2. Quantitative Examinations

We mainly choose frechet inception distance (FID) [40], LPIPS [41], structural similarity (SSIM), and peak signal-to-noise ratio (PSNR) as the evaluation indicators in this paper to measure the quality of the generated color images. It is worth noting that FID and LPIPS show high correspondence with human visual perception. Lower FID and LPIPS mean generated images are more realistic. On the contrary, larger values of SSIM and PSNR indicate better image quality.



**Figure 6.** Qualitative comparison between the best five methods.

Table 1 shows the evaluation metrics for the different comparison experiments. It can be observed that our method significantly outperforms the other baselines in the FID, LPIPS, SSIM, and PSNR metrics. The specific indicators strongly demonstrate the effectiveness of our proposed method.

**Table 1.** Quantitative results of different approaches.

Method	FID↓	LPIPS↓	SSIM↑	PSNR↑
CUT [14]	58.7077	0.1832	0.9078	20.1739
CycleGAN [10]	63.1189	0.1688	0.9095	19.4760
UNIT [21]	81.0382	0.1709	0.8884	19.1821
GCGAN [22]	76.1018	0.1649	0.8934	18.3975
DCLGAN [23]	95.7042	0.2131	0.8722	17.5598
Ours	49.3686	0.1334	0.9241	20.4459

#### 4.4. Comparative Experiments under Noisy Conditions

In the process of actual data acquisition by FPM, it is easy to be affected by the equipment, light source, dust, and other environments producing a certain degree of noise, and thus affecting the color image quality of the FPM acquisition and reconstruction. We added Gaussian noise with mean 0 and standard deviation  $6 \times 10^{-4}$  as the main interference condition for the single-channel input data of the network to simulate the potential noise effects during the actual acquisition process, thus further verifying the effectiveness of our proposed algorithm.

A qualitative comparison and quantitative analysis under the same noise condition are shown in Figure 7 and Table 2, respectively. With the same noise added, our method has an optimal visual performance compared to contrast experiments. Our method also performs the best on the four indicators in Table 2.



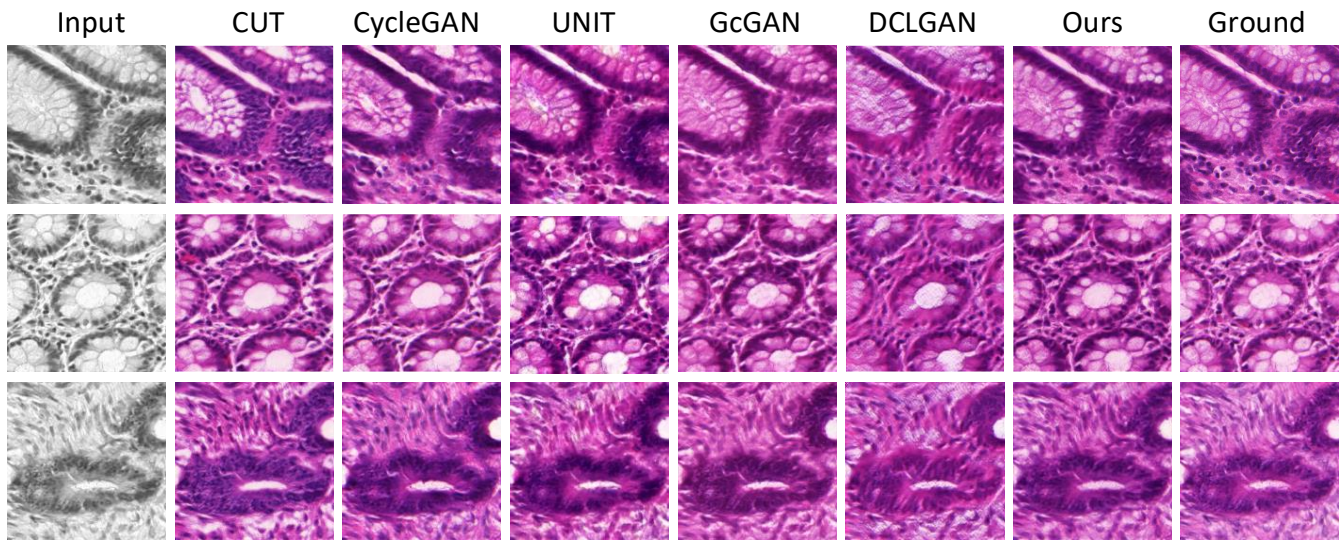


Figure 7. Qualitative comparison under the same noise condition.

Table 2. Quantitative results under the same noise condition.

Method	FID↓	LPIPS↓	SSIM↑	PSNR↑
CUT [14]	67.6780	0.1832	0.9066	20.0794
CycleGAN [10]	72.8258	0.1649	0.9093	19.4901
UNIT [21]	90.3650	0.1699	0.8884	19.3234
GCGAN [22]	83.3363	0.1692	0.8875	18.2488
DCLGAN [23]	104.5146	0.2109	0.8716	17.6097
Ours	56.4709	0.1327	0.9220	20.3231

Different degrees of noise may be generated during the actual acquisition process, and to further verify the robustness and generalization ability of our proposed method, we conducted three sets of experiments. Gaussian noise with a mean of 0 and standard deviation of  $6 \times 10^{-4}$ ,  $6 \times 10^{-3}$ , and  $6 \times 10^{-2}$  was added to the input images of the network. Figure 8 shows the visual analysis results of our method under different noise conditions, and Table 3 shows the evaluation index of stained images under three kinds of noise interference. As can be seen from the specific metrics in Table 3, the PSNR metric performs well as the Gaussian noise increases, and the FID, LPIPS, and SSIM metrics produce some degree of deterioration, but the degree of change is small and essentially flat. The visual analysis and quantitative results demonstrate the effectiveness of our proposed method against noise.

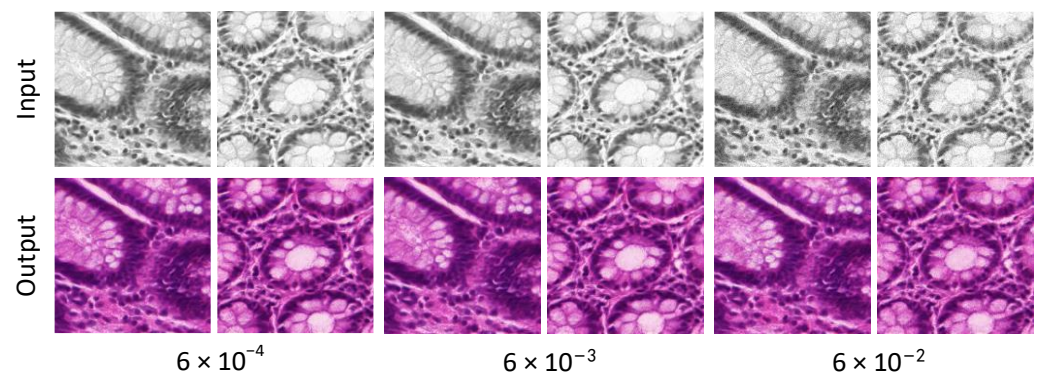


Figure 8. Qualitative comparison of our method under the different noise conditions.

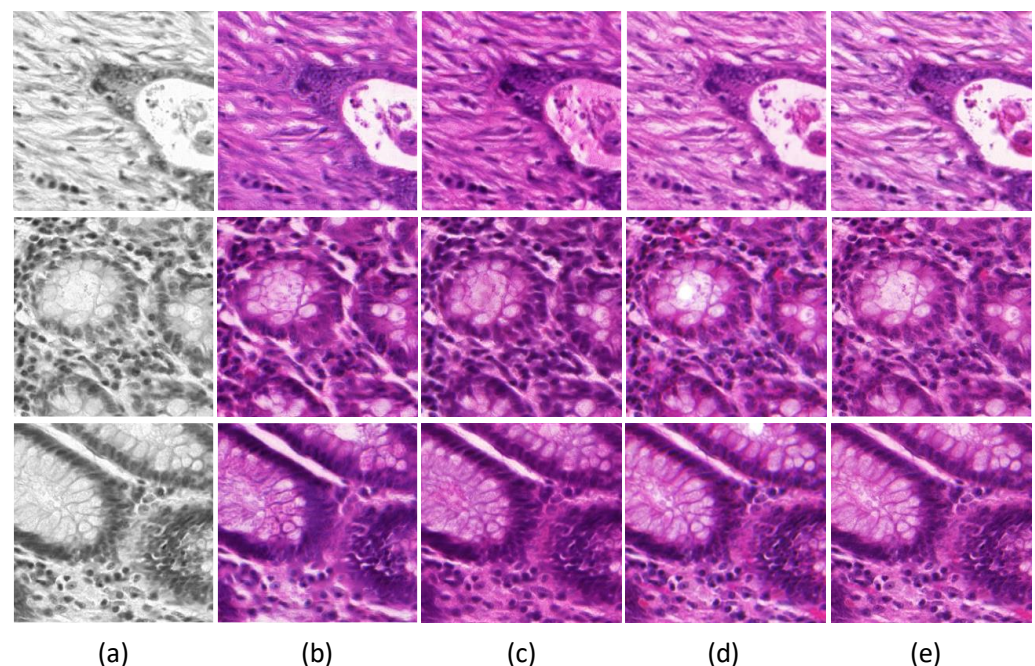
**Table 3.** Quantitative results of our method under the different noise conditions.

Noise	FID↓	LPIPS↓	SSIM↑	PSNR↑
$6 \times 10^{-4}$	56.4709	0.1327	0.9220	20.3231
$6 \times 10^{-3}$	56.4854	0.1343	0.9214	20.3286
$6 \times 10^{-2}$	56.5298	0.1378	0.9203	20.4288

#### 4.5. Ablation Experiment

To analyze the validity of our approach in depth, we conducted several experiments to study each of our contributions individually. The first ablation experiment removed content-consistency loss and identity loss while using the patchNCE loss in CUT as a learnable and domain-specific identity loss to prevent unnecessary changes by the generator. Another ablation experiment replaced the proposed ECR block with the residue block from CUT in the generator.

The experiments are summarized as described in Figure 9 and Table 4. The removal of content-consistency loss and identity loss results in the loss of high-frequency information in the output image, and the visual effect of coloring deteriorates. The ability of our method to accurately recognize the blank portion of a pathology section is diminished with the removal of the ECR block, resulting in poor image quality. Our proposed method shows superior performance in all metrics and visual comparisons.

**Figure 9.** The qualitative comparison results between the two ablation experiments and the proposed method. (a) Input, (b) Ours w/o Loss, (c) Ours w/o ECR, (d) Ours, (e) Ground Truth.**Table 4.** Quantitative results for ablations.

Method	FID↓	LPIPS↓	SSIM↑	PSNR↑
Ours w/o Loss	67.8005	0.1866	0.9086	20.3208
Ours w/o ECR	57.3747	0.1372	0.9180	19.8297
Ours	49.3686	0.1334	0.9241	20.4459

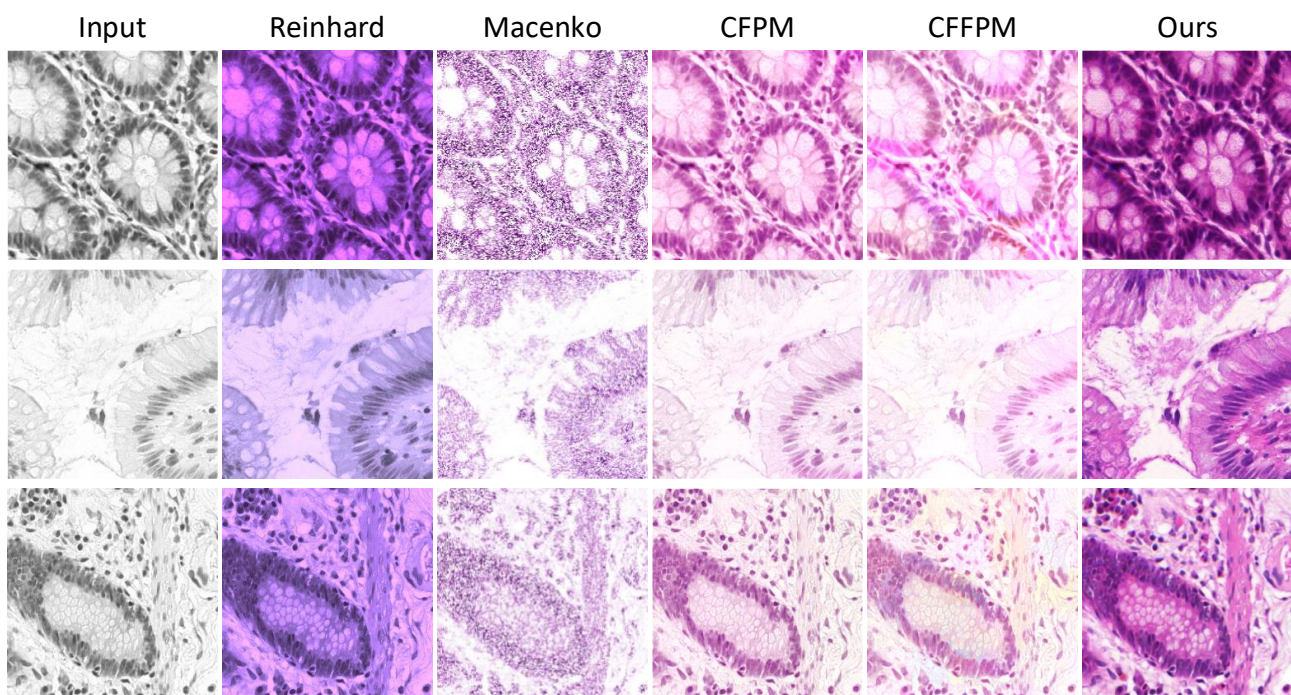
#### 4.6. Comparison Results with Classical Virtual Staining Methods

We compared our model with four other classical and recent virtual coloring methods to further validate the effectiveness and time efficiency of the network in grayscale image



coloring. Reinhard [42] and Macenko [43] are classic methods for image color transfer, while CFPM [16] and CFFPM [44] are recently proposed methods for FPM virtual staining based on color transfer. The above method was applied to our dataset for training to achieve virtual coloring of FPM single-channel grayscale images. The experimental results show that the proposed method has advantages in terms of color and time.

We randomly selected three images for display in Figure 10, and through visual analysis, it can be seen that the method proposed in this article has achieved the best results in terms of color and content. In our data, the Reinhard and Macenko methods performed poorly, with a wide difference in color styles and an inability to accurately stain cells at the appropriate locations. CFPM and CFFPM had relatively good results in distinguishing and coloring areas with different shades of color, but the overall color was relatively light and did not meet the expected goals. Moreover, CFFPM introduced significant artifacts. In conclusion, the deep learning method proposed in this paper had the best experimental results.



**Figure 10.** Qualitative comparison with classical virtual staining methods.

Specific evaluation indicators are shown in Table 5. We choose the number of iterations and the inference time as indicators to measure the time efficiency. Our proposed approach has the least time cost and the best performance. CFFPM requires iteration and takes the longest time. The Reinhard method is also fast in time, but ineffective.

**Table 5.** Comparison of time efficiency between different staining methods.

Virtual Staining Method	Iterations↓	Inference Time↓
Reinhard [40]	0	1.354 s
Macenko [41]	0	17.130 s
CFPM [16]	0	50.426 s
CFFPM [42]	5	799.343 s
Ours	0	0.313 s

## 5. Conclusions

We propose an effective self-supervised deep learning method based on contrastive learning for virtual staining of single-channel FPM pathological sections. The proposed



method designs an ECR-based generator and a content-consistency loss for joint training to obtain efficient cross-channel interaction information and improve the quality of stained images in an extremely lightweight manner. The experimental results show that this method maximizes the high-frequency information of the input image while effectively staining the FPM images with the best visual similarity. This virtual staining method can avoid the differences in staining results due to the stochasticity of personnel and technology, which can standardize staining, and it has various potential applications in digital pathology in the future.

**Author Contributions:** Conceptualization, Y.W., N.G. and J.L.; methodology, Y.W. and N.G.; software, N.G.; validation, N.G.; investigation, N.G.; writing—original draft preparation, N.G.; writing—review and editing, Y.W., N.G. and X.W.; supervision, Y.W., J.L. and X.W.; project administration, Y.W. and N.G. contributed equally to this paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Development Plan Projects of Jilin Province [NO: YDZJ202301ZYTS180].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the raw data used for training and evaluating the performance of the models are publicly available datasets that can be obtained from [36].

**Acknowledgments:** We would like to thank Hao Wang and Xinbo Wang for their support and contributions to the work of this article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Treanor, D. Virtual Slides: An Introduction. *Diagn. Histopathol.* **2009**, *15*, 99–103. [[CrossRef](#)]
2. Zheng, G.; Horstmeyer, R.; Yang, C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nat. Photon* **2013**, *7*, 739–745. [[CrossRef](#)] [[PubMed](#)]
3. Zheng, G.; Shen, C.; Jiang, S.; Song, P.; Yang, C. Concept, implementations and applications of Fourier ptychography. *Nat. Rev. Phys.* **2021**, *3*, 207–223. [[CrossRef](#)]
4. Chung, J.; Lu, H.; Ou, X.; Zhou, H.; Yang, C. Wide-field Fourier ptychographic microscopy using laser illumination source. *Biomed. Opt. Express* **2016**, *7*, 4787–4802. [[CrossRef](#)] [[PubMed](#)]
5. Ou, X.; Horstmeyer, R.; Yang, C.; Zheng, G. Quantitative phase imaging via Fourier ptychographic microscopy. *Opt. Lett.* **2013**, *38*, 4845–4848. [[CrossRef](#)] [[PubMed](#)]
6. Horstmeyer, R.; Ou, X.; Zheng, G.; Willems, P.; Yang, C. Digital pathology with Fourier ptychography. *Comput. Med. Imaging Graph.* **2015**, *42*, 38–43. [[CrossRef](#)] [[PubMed](#)]
7. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
8. Guo, C.; Jiang, S.; Yang, L.; Song, P.; Wang, T.; Shao, X.; Zhang, Z.; Murphy, M.; Zheng, G. Deep learning-enabled whole slide imaging (DeepWSI): Oil-immersion quality using dry objectives, longer depth of field, higher system throughput, and better functionality. *Opt. Express* **2021**, *29*, 39669–39684. [[CrossRef](#)]
9. Zhu, J.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
10. Li, C.; Liu, H.; Chen, C.; Pu, Y.; Chen, L.; Henao, R.; Carin, L. Alice: Towards understanding adversarial learning for joint distribution matching. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
11. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
12. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning (ICML), Vienna, Austria, 14–16 July 2020.
13. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive learning for unpaired image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020.
14. Zhou, Y.; Wu, J.; Bian, Z.; Suo, J.; Zheng, G.; Dai, Q. Fourier ptychographic microscopy using wavelength multiplexing. *J. Biomed. Opt.* **2017**, *22*, 066006. [[CrossRef](#)]
15. Dong, S.; Shiradkar, R.; Nanda, P.; Zheng, G. Spectral multiplexing and coherent-state decomposition in Fourier ptychographic imaging. *Biomed. Opt. Express* **2014**, *5*, 1757–1767. [[CrossRef](#)]

16. Gao, Y.; Chen, J.; Wang, A.; Pan, A.; Ma, C.; Yao, B. High-throughput fast full-color digital pathology based on Fourier ptychographic microscopy via color transfer. *Sci. China Phys. Mech. Astron.* **2021**, *64*, 114211. [[CrossRef](#)]
17. Zhang, M.; Liang, Y. Color Fourier stacked microscopy based on three-dimensional convolutional neural networks. *J. Opt.* **2020**, *40*, 2011001.
18. Wang, R.; Song, P.; Jiang, S.; Yan, C.; Zhu, J.; Guo, C.; Bian, Z.; Wang, T.; Zheng, G. Virtual brightfield and fluorescence staining for Fourier ptychography via unsupervised deep learning. *Opt. Lett.* **2020**, *45*, 5405–5408. [[CrossRef](#)] [[PubMed](#)]
19. Zhang, J.; Li, J.; Sun, H.; Jiang, S.; Zhang, Y.; Chen, Y.; Zhang, J.; Xu, T. Edge-enabled anti-noise telepathology imaging reconstruction technology in harsh environments. *IEEE Netw.* **2022**, *36*, 92–99. [[CrossRef](#)]
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014.
21. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
22. Fu, H.; Gong, M.; Wang, C.; Batmanghelich, K.; Zhang, K.; Tao, D. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
23. Han, J.; Shoeiby, M.; Petersson, M.; Armin, M.A. Dual contrastive learning for unsupervised image-to-image translation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021.
24. Jeong, J.; Shin, J. Training GANs with stronger augmentations via contrastive discriminator. In Proceedings of the International Conference on Learning Representations (ICLR), Vienna, Austria, 4–8 May 2021.
25. Kang, M.; Park, J. ContraGAN: Contrastive learning for conditional image generation. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 6–12 December 2020.
26. Liu, R.; Ge, Y.; Choi, C.L.; Wang, X.; Li, H. DivCo: Diverse conditional image synthesis via contrastive generative adversarial network. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
27. Zhang, H.; Goodfellow, I.; Metaxes, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning (ICML), Long Beach, CA, USA, 10–15 June 2019.
28. Emami, H.; Aliabadi, M.M.; Dong, M.; Chinnam, R.B. SPA-GAN: Spatial attention gan for image-to-image translation. *IEEE Trans. Multimed.* **2020**, *23*, 391–401. [[CrossRef](#)]
29. Hu, X.; Zhou, X.; Huang, Q.; Shi, Z.; Sun, L.; Li, Q. QS-Attn: Query-selected attention for contrastive learning in I2I translation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.
30. Torbunov, D.; Huang, Y.; Yu, H.; Huang, G.; Yoo, S.; Lin, M.; Viren, B.; Ren, Y. UVCGAN: UNet Vision Transformer cycle-consistent GAN for unpaired image-to-image translation. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023.
31. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, HI, USA, 21–26 July 2017.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.
33. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
34. Oord, A.v.d.; Li, Y.; Vinyals, O. Representation learning with contrastive predictive coding. *arXiv* **2018**, arXiv:1807.03748.
35. Michael, G.; Aapo, H. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In Proceedings of the Thirtieth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010.
36. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57. [[CrossRef](#)]
37. Ouyang, W.; Aristov, A.; Lelek, M.; Hao, X.; Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **2018**, *36*, 460–468. [[CrossRef](#)]
38. Sirinukunwattana, K.; Pluim, J.P.; Chen, H.; Qi, X.; Heng, P.A.; Guo, Y.B.; Wang, L.Y.; Matuszewski, B.; Bruni, E.; Sanchez, U.; et al. Gland segmentation in colon histology images: The glas challenge contest. *Med. Image Anal.* **2017**, *35*, 489–502. [[CrossRef](#)]
39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2019**, arXiv:1412.6980.
40. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 8–14 December 2017.
41. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.

42. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput. Graph. Appl.* **2001**, *21*, 34–41. [[CrossRef](#)]
43. Macenko, M.; Niethammer, M.; Marron, J.S.; Borland, D.; Woosley, J.T.; Guan, X.; Schmitt, C.; Thomas, N.E. A method for normalizing histology slides for quantitative analysis. In Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009.
44. Chen, J.; Wang, A.; Pan, A.; Zheng, G.; Ma, C.; Yao, B. High-throughput fast full-color Fourier ptychographic microscopy via color transfer and spatial filtering. *Photonics Res.* **2022**, *10*, 2410–2421. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.