

Received December 16, 2020, accepted January 6, 2021, date of publication January 12, 2021, date of current version January 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051188

Unpaired Stain Style Transfer Using Invertible Neural Networks Based on Channel Attention and Long-Range Residual

JUNLIN LAN^{1,2}, SHAOJIN CAI^{1,2}, YUYANG XUE⁴, QINQUAN GAO^{1,2,3}, MIN DU^{1,2,6}, HEJUN ZHANG⁵, ZHIDA WU⁵, YANGLIN DENG^{1,2}, YUXIU HUANG^{1,2}, TONG TONG^{1,2,3}, AND GANG CHEN⁵

¹College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China

²Fujian Key Laboratory of Medical Instrumentation & Pharmaceutical Technology, Fuzhou 350108, China

³Imperial Vision Technology, Fuzhou 350002, China

⁴Graduate School of Science and Technology, University of Tsukuba, Tsukuba 305-8577, Japan

⁵Department of Pathology, Medical University Cancer Hospital and Fujian Cancer Hospital, Fuzhou 350014, China

⁶Fujian Provincial Key Laboratory of Eco-industrial Green Technology, Wuyi University, Wuyishan 354300, China

Corresponding authors: Tong Tong (ttravelong@gmail.com) and Gang Chen (naichengang@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61901120 and Grant 61802065, and in part by the Science and Technology Program of Fujian Province of China under Grant 2019YZ016006 and Grant 2019L3018.

ABSTRACT Hematoxylin and eosin (H&E) stained colors is a critical step in the digitized pathological diagnosis of cancer. However, differences in section preparations, staining protocols and scanner specifications may result in the variations of stain colors in pathological images, which can potentially hamper the effectiveness of pathologist's diagnosis and the robustness. To alleviate this problem, several color normalization methods have been proposed. Most previous approaches map color information between images highly dependent on a reference template. However, due to the problem that pathological images are usually unpaired, these methods cannot produce satisfactory results. In this work, we propose an unsupervised color normalization method based on channel attention and long-range residual, using a technology called invertible neural networks (INN) to transfer the stain style while preserving the tissue semantics between different hospitals or centers, resulting in a virtual stained sample in the sense that no actual stains are used. In our method, the expert does not need to choose a template image. More specifically, we have developed a new unsupervised stain style transfer framework based on INN that is different from state-of-the-art methods. Meanwhile, we propose a new generator and a discriminator to further improve the performance. Our approach outperforms state-of-the-art methods both in objective metrics and subjective evaluations, yielding an improvement of 1.0 dB in terms of PSNR. Moreover, the amount of computation of the proposed network has been reduced by 33 %. This indicates that the inference speed is almost one third faster while the performance is better.

INDEX TERMS Color normalization, stain style transfer, invertible neural networks, pathological images.

I. INTRODUCTION

Histopathological analysis has been considered to be the golden standard step for the affirmation of cancer. Traditionally, pathologist's diagnosis involve visual analysis of stained slides to study the presence and characteristics of the disease under a microscope. Staining provides valuable information for the pathologist's diagnosis in clinic. Due to the fact that

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang.

chemical dyes can selectively bind to naturally transparent in tissues, different structures can be colored differently in tissue sections [1]. Pathologists study the spatial information and morphological features to diagnose and characterize the various pathological conditions [2]. However, this tedious examination is time-consuming and prone to be subjective [3], [4]. Several previous diagnosis in histopathological grading have been shown the poor consistency among pathologists [5]–[7]. Moreover, pathologists may miss small cancer regions [8]. Until recently deep learning has open an access for automated

and accurate histopathological diagnosis with machine learning algorithms using whole-slide imaging. Whole-slide imaging is a technique where scanners digitize glass slides at a very high resolution, resulting in a image with a size in the order of 10 gigapixels, typically called as whole-slide images (WSIs). Computer aided diagnosis (CAD) using deep learning technique can not only alleviate shortcomings of human diagnosis, but also improve diagnosis by sieving out obviously benign slides and providing quantitative characterization of suspicious areas [9].

Hematoxylin and eosin (H&E) is the most widely used staining agent in histopathology. Although the same chemical dye (H&E) is used, many other factors can lead to the variations in visual appearance of the tissue sections [10]. These factors mainly exist in the specimen preparations, staining protocols and slide scanners, including variables in raw materials, reactivity to staining agents from different manufacturers, concentration of staining agents, inter-patient or inter-center variations, and difference in scanners [10]. These variations can hamper the performances of CAD systems, reducing the robustness of computer algorithm [11]. In clinic, the appearances of stain style images are different among different hospitals especially their colors, thus affecting the generalization ability of deep learning models. In order to achieve a robust diagnosis, models trained with datasets from one center should make a reliable diagnosis when it is utilized on the data of the other center. However, due to the appearance differences, the robustness of the algorithms will decrease when evaluated on the data of another centers. The adoption of standardized tissue preparations, staining protocols and scanners may reduce the variations. However, this standardization involve many human operations (e.g., manual tissue preparations and application of staining agents), resulting in the subjective variations in pathologists inevitably. Moreover, the variations become more significant due to the fact that different scanners are used among centers. This implies that eliminating all the underlying variations is infeasible in practice [12]. A more practical solution is to combine a specific image analysis task with stain normalization task in end-to-end joint training. The intrinsic stain normalization component serves as a pre-processing operation for image analysis tasks, generating images similar to target version in terms of stain and to input in terms of context. In this configuration, the task-specific network (e.g., cancer classification or tissue segmentation) is trained on target-style images drawn from target set distribution as well as arbitrary images drawn from different distribution. Although this task can handle images with different statistical properties (i.e., different staining appearances), it also needs to retrain for the task-specific network when new-style images are used as inputs. Additionally, the intrinsic stain normalization component is invisible. Thus we are not able to know the quality of these generated images.

Generally, previous automated algorithms mainly address stain inconsistency by three approaches: (i) ignoring color information using only grayscale images; (ii) color-based data augmentation method for synthesizing new images

to enhance the robustness of the model; (iii) relying on color matching strategy to normalize stain style, where all images from a given data set is mapped to a reference data set [13]–[21]. However, the methods that ignores color information to learn color-invariant features are biased towards texture-based features [22]. These methods have the main limitation of ignoring clinically relevant information captured through colors. It is uncertain whether data augmentation (i.e., (ii) above) can capture all variations that occur “in the wild” due to the linear nature of many color/stain augmentations. This may be an oversimplification of variability in tissue staining in the real world [23]. Additionally, Most color matching strategies (i.e., (iii) above) highly dependent on a template image. However, due to the problem that pathological images are usually unpaired, these methods can not produce satisfactory results.

In this work, we propose an innovative stain normalization approach using unpaired datasets based on invertible neural networks (INN). INN can be used to synthesize high-quality images [24], and to perform image classification without losing information of hidden layer [25] through addressing inverse problems [26]. In our method, experts no longer need to manually select templates by themselves bypassing capturing special characteristics of unpaired images and making an mapping between two centers. In addition, the channel attention and long-range residual schemes are integrated to further improve the performance of the proposed method. In the end, our method yields an improvement of 1dB in terms of PSNR compared to state-of-the-art methods. It’s also worth noting that our proposed framework can be implemented 33 % faster than state-of-the-art methods.

Our contributions are summarized as follows:

- 1) We develop a novel unpaired stain normalization framework via introducing the idea of INN using a technology called additive coupling.
- 2) We propose a new network structure that includes a novel generator and a new discriminator to improve the performance. The channel attention and long-range residual are used in our proposed framework to yield better performance.
- 3) The proposed method has been evaluated using public available dataset MITOS-ATYPIA14, dataset CAMELYON16 and clinical dataset from a local hospital, yielding superior performance than state-of-the-art methods while with a faster inference speed.

II. RELATED WORKS

A. STAIN NORMALIZATION

Most previous algorithms to date for stain normalization can broadly fall into three categories. **Histogram-matching based methods** that try to map the histogram distributions of source images to that of a reference template. For instance, Reinhard *et al.* proposed to map distributions of color histograms between source images and a reference template in LAB color space [16]. This method performed the mapping

on each separate channel of LAB. However, this may lead to incorrect color mapping of stain components due to the fact that dyes have independent contributions to the final colors of the image, attributed to the chemical nature that dye has its own specific reaction mode. The problem can be addressed by **Stain separation based methods**, where normalization is performed on each staining channel (i.e., hematoxylin channel and eosin channel) separately. In this type of methods, color deconvolution is applied for stain separation to decompose an image into its staining components. This is done by estimating a staining matrix that represents the RGB colors of each stain in the tissue. Khan *et al.* [15] computed the stain matrix by assigning each pixel to specific stain components. Santanu Roy *et al.* [27] performed color normalization of the source image by transferring the mean color of the target image in the source image and also to separate stain present in the source image. Li and Plataniotis [28] raised a systematical and analytical solution consisting of a circular color analysis module and a computation module based on non-negative matrix factorization. Other approaches for stain separation using non-negative matrix decomposition [29], [30] to cluster images. These methods for stain separation are supervised. Other studies based unsupervised techniques involve clustering the images using optimization algorithms such as expectation-maximization [18]. Harsha Bhat *et al.* [31] proposed an algorithm combining de-staining and wedge separation for histopathological images. The method proposed by Politecnico di Torino *et al.* [32], named SCAN (Stain Color Adaptive Normalization), is based on segmentation and clustering strategies for cellular structures detection. The above two types of methods for stain normalization have obvious drawbacks, these methods did not consider the spatial relation of the tissue, resulting in improper staining. In addition, these methods highly depend on a reference template, which may lead to normalization errors when the color distributions of a reference template is not proper.

The third category is **Generative learning based methods**, these methods exploit the powerful image representation ability of convolutional neural network (CNN) and image generation ability of Generative Adversarial Network (GAN) to normalize stain style, transferring the stain style of reference images to source images while preserving the context of source images. They are template-free, and can overcome the problem of staining errors caused by the previous two types of methods. For instance, Bentaieb and Hamarneh [33] combine a image classification task with stain style transfer task for joint training. In this configuration, the intrinsic stain style transfer component aims to generate images similar to target version in terms of stain and to input in terms of context. Shaban *et al.* [34] proposed an unpaired stain style transfer networks called StainGAN based on CycleGAN and achieved the best performance to date with the highest values in terms of PSNR and SSIM. Similar to StainGAN, Zhou *et al.* [35] proposed an enhanced cycleGAN based method with a novel auxiliary input for the generator by computing a stain color matrix for every H&E image in the training set, and

Thomas *et al.* [23] developed an improved CycleGAN for segmentation of renal histopathology.

B. GENERATIVE AND ADVERSARIAL NETWORKS (GAN)

Generative and Adversarial Networks was firstly introduced by Mirza and Osindero [36] in 2014. From then on, GAN bring significant changes in computer vision. They have already made great success in image generation [37], [38], image editing [39], representation learning [38], [40], image inpainting [41], and image-to-image translation [38], [42], [43]. The key of GAN's success is the idea of adversarial learning that forces the generated images to be indistinguishable from target images. It consists of two components, a generator which generates images from random noise, and a discriminator to distinguish generated images from real images. During training, a adversarial loss is introduced. Generator is rewarded for generating fake image to fool the discriminator, and discriminator is rewarded to correct the classifications. Both networks optimize the parameters through back-propagation algorithms and to update iteratively, aiming at the smallest value of the adversarial loss function. As we all know, GAN is difficult to train because it needs to strike a balance between different networks. Several techniques have been proposed for a stable training, such as spectral normalization [44], WGAN [45] and PGGAN [46].

C. IMAGE-TO-IMAGE TRANSLATION

Our work we refer to as stain style transfer is closer to image-to-image translation rather than style transfer, since it is not necessary for style transfer to preserve context information. Image-to-image translation is an active area which employs GAN to synthesize new images in computer vision. It is a technique that learns a function $f: x \rightarrow y$, where x ($x \in X$) and y ($y \in Y$) are two corresponding representations of an image. It can be divided into paired image-to-image translation and unpaired image-to-image translation according to the available training data.

Paired datasets consist of aligned images in domain X and corresponding domain Y . For instance, a landscape photo and a painting by Monet of the same landscape make a paired of images. With paired datasets for training, a generator is learnt to map x ($x \in X$) to y ($y \in Y$), and a discriminator is learnt to discriminate real photo(x), real painting(y) and generating painting ($f(x)$). This is called conditional GAN [36], because the discriminator shows the x in addition to $f(x)$ or y to judge its authenticity. A image-to-image translation framework called pixel2pixel has extensively introduced the idea of conditional GAN [38], such as generating photographs from sketches or from attribute and semantic layouts. However, in practice, there are usually not paired images in medical.

Unpaired datasets consist of a image collections from domain X and another image collections from domain Y . Image x randomly sampled from X and image y randomly sampled from Y are not aligned. In this case, it is a problem

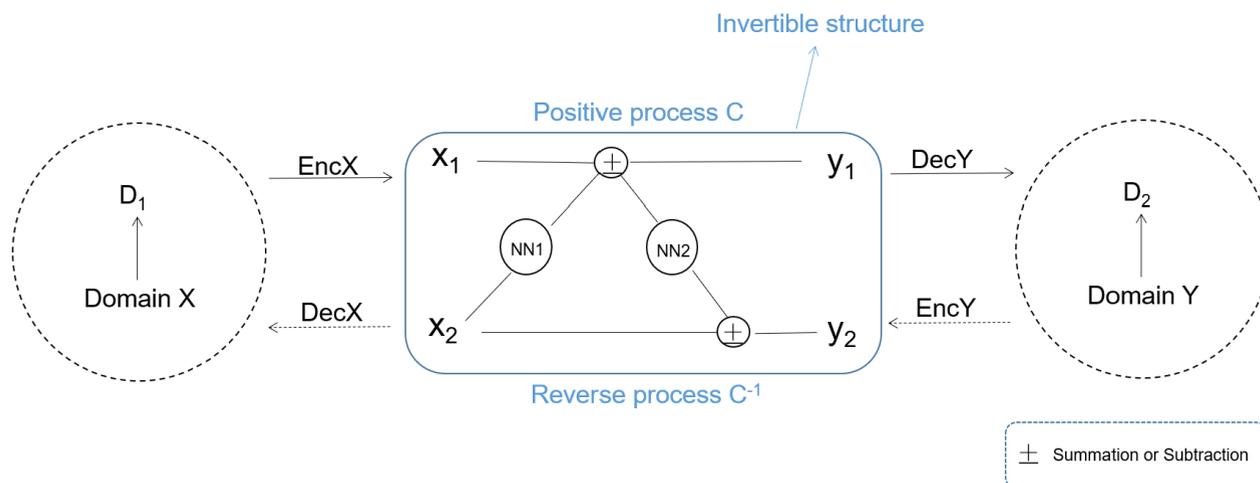


FIGURE 1. The proposed stain style transfer framework. Positive process and reverse process are constructed into invertible structure.

that keeping the content after mapping, attributing to the lack of ground truth to guide the training process. One trick for solving this problem is cycle-consistence loss, introduces another pair of generator and discriminator, which learns a reverse mapping function $g: y \rightarrow x$. A image translated by f is translated back to the source domain by g , and vice versa. That is, the cycle-consistent loss $|x - g(f(x))| + |y - f(g(y))|$, encourages both parties to map from one to another. The cycle-consistency is introduced in CycleGAN [42]. In fact, AUTOMAP [47] demonstrated that this approach can be used for reconstruction of PET, CT, and MRI images in medical applications.

D. INVERTIBLE NEURAL NETWORKS (INN)

Recent studies has shown that the input images would lost some important information with depth in convolution neural networks. [48], [49]. In this case, Multiple invertible neural networks (INN) have been introduced that are capable of learning invertible representations to alleviate information loss under certain conditions [25], [50], [51], and modeling inverse scenes [26]. Most studies related to INN, including this study, highly relies on the idea introduced in [52], later extended in [48]. INN can adjunct to cycle-consistent loss in our stain style transfer. The advantages of using INN in our work have two-fold: firstly, semantic context and texture information of image x should reappear in its corresponding image $f(x)$, and vice versa. Therefore, the two networks (F, G) should be closely related rather than completely separate. Secondly, INN can greatly reduce the amount of network calculations and parameters through parameter sharing. In our model, inspired by [53], we propose an invertible framework using additive coupling technology [52]: first we split the channels of an input feature x into $(x_1; x_2)$ and then mapping them using arbitrary functions NN_1 and NN_2 . After that y_1 and y_2 are merged. Next section will gives the details of our method.

III. METHODS

In this study, we aim to train a novel network for unpaired stain style transfer with much better image quality and much faster inference speed compared to state-of-the-art method StainGAN. Our proposed method for stain style transfer is shown in Figure 1, where X and Y represents source domain and target domain respectively, D_1 and D_2 are discriminators in both domains, Enc represents encoding operations and Dec represents decoding operations, NN_1 and NN_2 are two functions which have same network structure. Our work has three main innovations in generator and discriminator. Firstly, we designed a new generator architecture based on UNet [54], where we introduce the channel attention and long-range residual. On this basis, we obtain an invertible process among positive generator and oppositive generator based on [53], using a technique called additive coupling [52]. After that, we develop the discriminator architecture by introducing channel attention mechanism. We will elaborate our framework in following sections.

A. OUR FRAMEWORK

At the absence of paired data, StainGAN aims to exploit the reciprocal relationship between G_1 and G_2 . However, the problem of StainGAN is that we can insert arbitrary G_1 and G_2 , missing the idea that G_1 and G_2 can be approximate inverse by design. In this work, we design G_1 and G_2 to be approximately invertible based on INN, closing to the ideal reciprocal relationship. The key benefit of inverse is that we need not have $X \rightarrow Y$ and $Y \rightarrow X$ mapping separately, only requiring a $X \rightarrow Y$ mapping, which we can run in reverse to approximate $Y \rightarrow X$. The mappings in both directions have share parameters. More specifically, we can see that training the $X \rightarrow Y$ model will also train the reverse $Y \rightarrow X$ model, which does not necessarily occur with separate models, contributing to not only preservation of

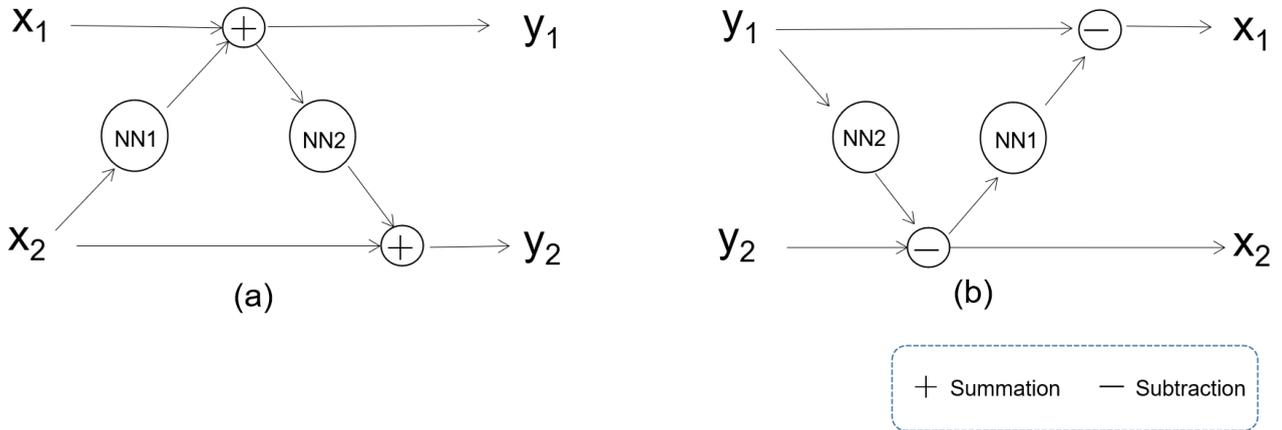


FIGURE 2. The technique called additive coupling used to construct the two generators into invertible neural networks form, positive process is illustrated in (a) and reverse process is illustrated in (b).

semantic context but also reducing the network computations and parameters.

Stain style transfer are usually between two domains. Therefore, it is impossible to be truly reversible in a low-dimensional image space. Given this situation, we divide the forward process $G_1: X \rightarrow Y$ and the reverse process $G_2: Y \rightarrow X$ into three parts respectively which include encoding, mapping and decoding. As shown in Figure 1, G_1 consists of Enc_X , C and Dec_Y , G_2 consists of Enc_Y , C^{-1} and Dec_X . Among them, Enc_X encodes image x into high-dimensional feature space \tilde{x} , and C mapping \tilde{x} to \tilde{y} which is decoded by Dec_Y to y . Similarly, Enc_Y encodes image y into high-dimensional feature space \tilde{y} , and C^{-1} mapping \tilde{y} to \tilde{x} which is decoded by Dec_X to x . C^{-1} is the inverse function of C , where C^{-1} and C together form the invertible structure. Mathematical form can be illustrated as following equations:

$$G_1 = Dec_Y \cdot C \cdot Enc_X \quad (1)$$

$$G_2 = Dec_X \cdot C^{-1} \cdot Enc_Y \quad (2)$$

In our method, the encoding and decoding parts are irreversible functions which both adopt convolutional neural networks. In this case, G_1 and G_2 should be approximately reversible. We obtain an invertible mapping of C and C^{-1} , using a trick called additive coupling [52]. As shown in Figure 2, Equation (3) and (4). Figure 2 (a) and Equation (3) corresponds to function C . Firstly, we split x into x_1 and x_2 over feature channels equally, thus transforming them into y_1 and y_2 with arbitrary function NN_1 and NN_2 respectively, then cascading y_1 and y_2 into y . The inverse function (C^{-1}) can be seen in Figure 2 (b) and Equation (4).

$$y_1 = x_1 + NN_1(x_2); \quad y_2 = x_2 + NN_2(y_1) \quad (3)$$

$$x_2 = y_2 - NN_2(y_1); \quad x_1 = y_1 - NN_1(x_2) \quad (4)$$

The proposed framework is an unpaired stain style transfer network based on StainGAN. It consists of two generators ($G_1: x \rightarrow y$ and $G_2: y \rightarrow x$) and two discriminators (D_1 for domain X and D_2 for domain Y). Each generator corresponds

to a discriminator respectively. For illustration, the forward pair (G_1 and D_2), learns to map images from domain X to domain Y . $x_i \in X$ where x_i servers as the input of G_1 , which yields generated image \hat{y} , $\hat{y} = G_1(x_i)$. D_2 distinguishes \hat{y} from y_i where $y_i \in Y$. This adversarial training process is equivalent to a min-max optimization with a loss function:

$$L_{adv}(G_1, D_2) = E_{x \sim p_{data}(x)}[\log(1 - D_2(G_1(x)))] + E_{y \sim p_{data}(y)}[\log D_2(y)] \quad (5)$$

The backward pair (G_2 and D_1), learns to map images from Y to X . y_i servers as the input of G_2 , which yields generated image \hat{x} , $\hat{x} = G_2(y_i)$. D_1 distinguishes \hat{x} from x_i . The training process can be formulated as a min-max optimization process, and the loss function is $L_{adv}(G_2, D_1)$:

$$L_{adv}(G_2, D_1) = E_{y \sim p_{data}(y)}[\log(1 - D_1(G_2(y)))] + E_{x \sim p_{data}(x)}[\log D_1(x)] \quad (6)$$

Additional, cycle-consistent loss is also introduced to enforce the two mapping functions to be cycle-consistent with each other, which reconstructs \hat{y} from \hat{x} through G_2 , $\hat{\hat{y}} = G_2(G_1(x))$, and \hat{x} from \hat{y} through G_1 , $\hat{\hat{x}} = G_1(G_2(y))$. Its mathematical form is as follows:

$$L_{cyc}(G_1, G_2) = E_{x \sim p_{data}(x)}[\|x - G_2(G_1(x))\|_1] + E_{y \sim p_{data}(y)}[\|y - G_1(G_2(y))\|_1] \quad (7)$$

Finally, the total training loss can be formulated as:

$$L(G_1, G_2, D_1, D_2) = L_{adv}(G_1, D_1) + L_{adv}(G_2, D_2) + \lambda L_{cyc}(G_1, G_2) \quad (8)$$

B. THE PROPOSED GENERATOR ARCHITECTURE

As shown in Figure 3, we designed a generator architecture called RC-Net based on UNet and split it into three parts (Enc , NN_1 and NN_2 , Dec). Enc_X and Enc_Y both use Enc while Dec_X and Dec_Y use Dec . In addition, NN_1 and NN_2 which are arbitrary functions in additive coupling both use the network structure in the dotted frame.

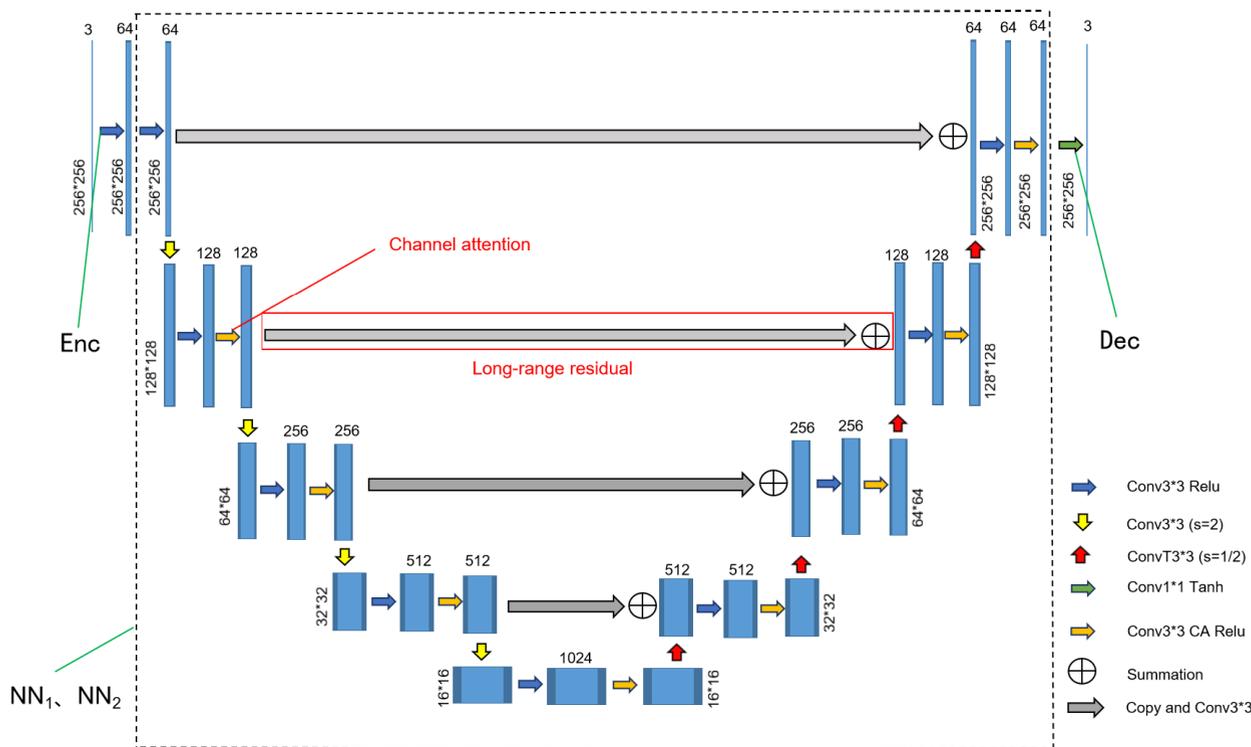


FIGURE 3. We propose the RC-Net and split it into three parts which include encoding function, mapping function and decoding function.

RC-Net consists of a coding phase and a decoding phase. The channel sizes and numbers are written in the side and upside of every layer respectively. The coding phase has four downsampling and the decoding phase includes four upsampling. All downsampling or upsampling followed by two stride-1 convolutional layers with 3×3 kernels where the second layer uses the channel attention mechanism. Shortcuts in the form of summation are used to add the features from the coding phase to the decoding phase in all corresponding layers.

Compared to UNet, RC-Net has two main differences which result in a better performance in image quality: 1) To assist the decoding process, skip connections copy feature maps from the encoder to the decoder in UNet. Differently, as shown in Figure 3, the copied feature maps combined with decoding feature maps through summation, instead of concatenation used in UNet. A 3×3 convolutional layer with a stride of 1 is used for processing copied feature maps before the summation operation. It has two advantages to use summation [55]. First, the summation does not increase the number of feature maps, thus it can reduce the number of trainable parameters in the following layer; Second, the summation operation can be regarded as a long-range residual connection, which can facilitate model training. Moreover, we aim to convert the stain style while keeping the image content unchanged, which is exactly in line with our residual idea. It will be helpful that the encoding features are combined with the residual features to form the decoding feature.

2) We introduce the channel attention mechanism into our generator architecture based on the assumption that different feature channels contribute differently to the results. Channel attention mechanism [56] gives different weights to different feature channels, resulting in more accurate features in the following layers.

C. DISCRIMINATOR WITH CHANNEL ATTENTION MECHANISM

Although L_1 loss which is a pixel-wise loss contribute to low-frequency information reconstruction, they are usually unable to generate high-frequency information. It will generate blurry results on image generation tasks if only L_1 loss is used [57]. For capturing both low-frequency and high-frequency information, a patch-level classifier called PatchGAN which is good at capturing high-frequency features [38] was additional used as the discriminator. The combination of PatchGAN and L_1 loss guides the generators to generate high quality images with high-frequency and low-frequency details. In addition, PatchGAN outputs a 30×30 matrix after several convolution layers firstly, where each element has a 70×70 receptive field, and then provides a value by averaging the matrix.

Given that the discriminator is based on the style of the image to distinguish true or false, the style feature channels should be given greater weights and the content feature channel should be given less weights. Therefore, we introduce channel attention mechanism used in [56] to

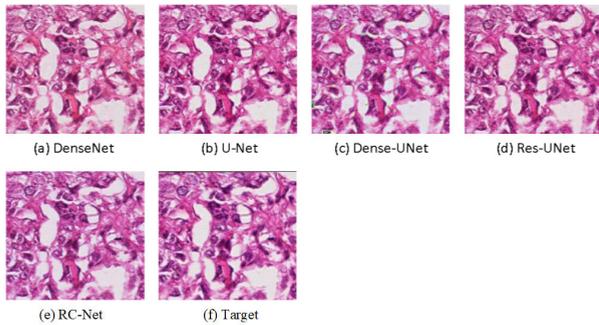


FIGURE 4. Visual comparisons of results using RC-Net and other different generators.

PatchGAN, after that, we adopt this network as the discriminator architecture. The structure of each layer of the discriminator is shown in Table 1, where k means the kernel size, s means the stride, BN represents batch normalization and LeakyRelu(0.2) means LeakyRelu function with a 0.2 slope to all negative values. We add a channel attention mechanism in front of the last convolutional layer, assigning different weights to different channel features that directly affect the discriminant value.

TABLE 1. The components of each layer of the discriminator architecture.

Layer	Component
1	Conv($k=4,s=2$),LeakyRelu(0.2)
2	Conv($k=4,s=2$),BN,LeakyRelu(0.2)
3	Conv($k=4,s=2$),BN,LeakyRelu(0.2)
4	Conv($k=4,s=1$),BN,LeakyRelu(0.2)
5	Channel attention , Conv($k=1,s=1$)

IV. EXPERIMENTS AND RESULTS

We make fair and extensive comparisons as follows: 1) Analysis of the effect of channel attention mechanism and long-range residual on the proposed generator. 2) Analysis of the advantages of RC-Net and comparisons of results using different generators. 3) Analysis and comparisons of the impact of adding invertible neural network on the image quality, network parameters and calculations. 4) Experimental comparisons to show that the introduction of channel attention mechanism in the discriminator can improve the discriminatory ability of the discriminator, thereby improving the image generation ability of the generator. 5) Comparisons with state-of-the-art methods. More detail descriptions about training, dataset, metrics and results will be given in the following subsections.

A. DATA AND EXPERIMENTAL SETUP

1) DATASET

The dataset comes from the MITOS-ATYPIA14 challenge.¹ It consists of 424 pairs of pathological images scanned by 2 different scanners (A and H). More specifically, a slide stained with hematoxylin and eosin scanned by 2 different

scanners will form a pairs of images. This dataset was split into two parts, which is training dataset includes 300 pairs and test dataset includes 124 pairs. Images from scanner H were resized to the identical size (1539×1376) of images from scanner A. We randomly extracted 9000 unpaired training patches and 620 paired test patches from training dataset and test dataset respectively. All sizes of the patches are 256×256 . We regarded the patches from scanner H as ground truth.

2) TRAINING DETAILS

The number of training iterations for all experiments was approximate 50000 and the number of training epochs for all experiments was approximate 10. We adopted Adam optimizer to update the parameters with a learning rate of 0.0002, λ in Equation (4) was set to 10. This value was optimally chosen to control the range of gradient values in back propagation during each training iteration since the weight values were big in the network. And least-squares loss [58] was introduced to replace the original negative log likelihood objective in L_{GAN} and a image buffer includes previous 50 images rather than one images was used to update the parameters of the discriminators.

3) EVALUATION METRICS

Results were compared to the ground truth with two similarity metrics: Peak Signal-to-Noise Ratio (PSNR), Wasserstein Distance (WD) and Structural Similarity index (SSIM). WD measures differences in brightness, contrast, and colors by calculating the shortest distance between two histogram distributions averaged across the RGB channels [59], and the lower value indicates the better results. Furthermore, to illustrate the role of INN, we additionally use floating point calculations and parameter amounts to measure the inference speed and memory size of the networks respectively when performing INN related experiments. Since we used the reversible convolutional neural network, the parameter sharing makes the two generators become inverse function each other, combining with the cycle-consistent to analyze and solve stain style transfer which belongs to reverse task. It should be noted that the reduction is nearly half but not half, because the encoding and decoding parts do not use the additive coupling. This comparison was done to show that the use of INN not only improves image quality, but also reduces the amount of network calculations and parameters.

B. THE PROPOSED GENERATOR WITH CHANNEL ATTENTION AND LONG-RANGE RESIDUAL

As shown in Table 2, Model1, Model2, and Model3 are all our proposed methods, where Model1 means the UNet structure, as shown in Figure 3, but lacks channel attention and uses skip connection instead of summation. Model2 based on Model1 uses summation instead of skip connection, and Model3 means RC-Net which is the proposed generator in the paper. Using long-range residual idea, the PSNR value increases from 22.21 to 22.47, the SSIM value increases from 0.812 to 0.825, and the WD value decreases from 2.96 to

¹<https://mitos-atypia-14.grand-challenge.org>

TABLE 2. The impact of long-range residual and channel attention mechanism. Model1 is a baseline method using UNet. Model2 is based on UNet and long-range residual. Model3 is the proposed RC-Net based on long-range residual and channel attention.

Method	PSNR(dB)	SSIM	WD
Original images	19.61	0.741	10.92
Model1	22.21	0.812	2.96
Model2	22.47	0.825	2.32
Model3	22.52	0.826	2.28

TABLE 3. The improvement brought by invertible neural networks on different evaluation metrics.

Method	PSNR(dB)	SSIM	WD	GFLOPs	PARAMS(M)
Original images	19.61	0.741	10.92	\	\
StainGAN	21.73	0.804	3.56	56.93	11.4
StainGAN+INN	22.00	0.818	3.47	35.19	6.1
RC-Net	22.52	0.826	2.28	75.20	47.29
RC-Net+INN	22.59	0.831	1.44	38.96	23.67

2.32. Futher, by adding channel attention mechanism, PSNR value increases from 22.47 to 22.52, the SSIM value increases from 0.825 to 0.826, and the WD value decreases from 2.32 to 2.28. Above results illustrate the benefit of using long-distance residual idea and channel attention mechanism in stain style transfer.

C. COMPARISONS WITH DIFFERENT GENERATORS

Several networks was used as the generator to make a comprehensive comparison with RC-Net, which includes UNet [54], Res-UNet [60], DenseNet [61] and Dense-UNet [61].

As shown in Table 3, they can not achiever higher PSNR value, higher SSIM value and lower WD value than RC-Net. And the visual comparisons are shown in Figure 4, demonstrating the superiority of RC-Net in producing much better images which have closer colors, contrast and texture compared to ground truth.

D. IMPROVEMENT OF USING INVERTIBLE NEURAL NETWORKS

The performance of using invertible neural networks is illustrated in Table 3, where INN means reconstructing the generators into the form of invertible neural networks using additive coupling technology. By adding INN, the StainGAN and our proposed method can be improved 0.27dB and 0.07dB respectively in terms of PSNR. This was resulted from the parameter sharing of positive and opposite generators by additive coupling technology. The parameter sharing instead of separation training makes the two generators become inverse function each other, combining with the cycle-consistent to analyze and solve stain style transfer which belongs to reverse task. Another benefit of using additive coupling technology is the significant reduction in the amount of network parameters and calculations. After using the additive coupling, the amount of parameters and calculation of StainGAN and ours are reduced by nearly half, which greatly speeds up the training and inference process of the network. It should be noted that the reduction is nearly half but not half, because the encoding and decoding parts do not

TABLE 4. The improvement of evaluation metrics attributing to adding channel attention mechanism in discriminator.

Method	PSNR(dB)	SSIM	WD
Original images	19.61	0.741	10.92
StainGAN	21.70	0.801	3.56
StainGAN(DCA)	22.16	0.806	3.48
INN	22.59	0.826	2.16
INN(DCA)	22.71	0.831	1.01

use the additive coupling. The computation and parameters are reduced due to the use of the channel split operation in additive coupling, as shown in Figure 2. In medical diagnosis, the speed of diagnosis is particularly important to the clinic, which can greatly reduce the burden on doctors and improve work efficiency. In addition, the reduction of network parameters also can reduce storage space.

E. ADVANTAGE OF THE PROPOSED DISCRIMINATOR

In this section, we introduced channel attention into the discriminators which uses PatchGAN structure. As illustrated in Table 1, we added a channel attention mechanism before the last convolutional layer. After that, we analyzed the effect of adding channel attention mechanism to the discriminators on StainGAN and our method. Results are shown in Table 4, where DCA means the discriminators which were introduced the channel attention mechanism. After introducing the channel attention mechanism to discriminators, StainGAN and our method achieved an improvement of 0.46 and 0.12 respectively in terms of PSNR, demonstrating that the channel attention mechanism can improve the discriminatory ability of the discriminators in stain style transfer.

F. COMPARISONS BETWEEN THE PROPOSED METHOD AND STATE-OF-THE-ART METHOD

We aim to transform the stain style of images from domain A to domian H while keeping the semantic context unchange, where domain A and H consists of images scanned by scanner A and H respectively. To alleviate the problem that keeping the semantic context unchanged when training images are unpaired, cycle-consistent loss [42] was introduced to map images between domain A and domain H.

Structure Preserving Color Normalization (SPCN) [29] is a method based on stand separation. SPCN scheme can change color of one image (source) to match that of another (target) while reliably keeping source structural information intact. A key step in SPCN scheme is an accurate stain separation of both source and target images based on sparse regularized NMF. StainGAN [34] is one of state-of-the-art methods. We have made three main diffenences compared to StainGAN. Firstly, we proposed a new generator called RC-Net, while StainGAN adopted the structure proposed by Johnson et al. [62] as generator. Compared to the generators of StainGAN, RC-Net has more downsampling layers and upsampling layers, which can generate a higher level of semantic information. Moreover, the introductions of long-range residual instead of skip connection contributing to faster training and better optimization. Further, the addition

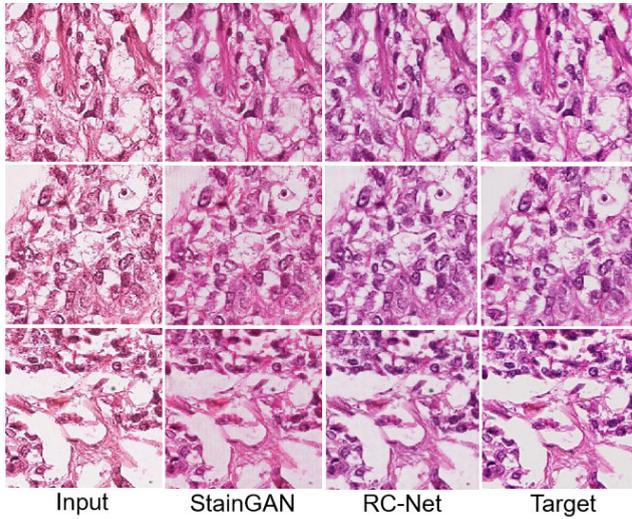


FIGURE 5. Visual comparison of results produced by RC-Net and StainGAN.

TABLE 5. Comparisons between the proposed method and state-of-the-art method on stain style transfer.

Method	PSNR(dB)	SSIM	WD	GFLOPs
Original images	19.61	0.741	10.92	\
SPCN	20.40	0.673	1.97	\
StainGAN	21.70	0.801	3.56	56.93
Our method	22.71	0.831	1.01	38.96

of channel attention mechanism avoids the problem that each channel contributes the same to the results, assigning adaptive weights to different channels, contributing to a more reasonable representation. Therefore, the introduction of long-range residual and channel attention mechanism is helpful to produce more detailed texture and color information. It shows that the great superiority of the proposed RC-Net in stain style transfer. Figure 5 provides their visual comparisons. Secondly, we used additive coupling technology to construct G_1 and G_2 as reversible functions, after that, G_1 and G_2 shared the same parameters (NN_1 and NN_2). This technology not only achieved lower values in terms of WD, higher values in terms of PSNR and SSIM but also reduced the amount of parameters and calculation of our networks by nearly half. Finally, we introduced channel attention into the discriminators which uses PatchGAN structure. After adding this scheme, our method has been further improved in terms of PSNR, SSIM and WD. The visual comparisons can be seen in Figure 6. PSNR, SSIM and WD metrics are illustrated in Table 5. The visual comparisons demonstrate that the results produced by the proposed method are much more closer to the ground truth than the results using StainGAN or SPCN in terms of color, texture and context. Finally, the PSNR and SSIM values of the proposed method are higher than those using SPCN, while the WD is lower, indicating better results. Our method can achieve an improvement of 1.01dB in terms of PSNR, an improvement of 0.03 in terms of SSIM and a depress of 2.55 in terms of WD than StainGAN, while increasing the inference speed by one third.

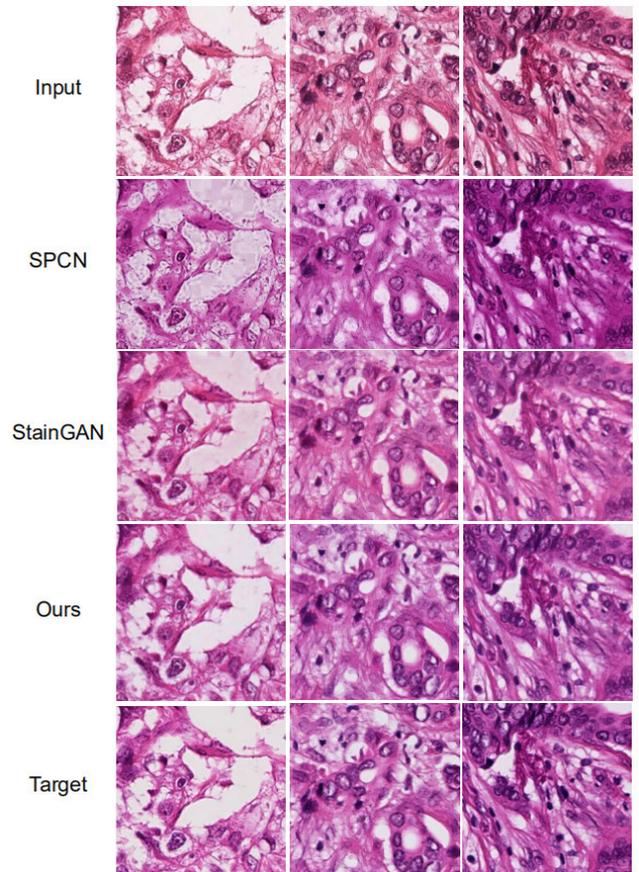


FIGURE 6. Comparisons between the proposed method and state-of-the-art method on stain style transfer.

G. ASSESSMENT ABOUT THE COLOR NORMALIZATION EFFECT ON COMPUTER-AIDED DIAGNOSIS

We used the public available CAMELYON16 dataset² to verify the impact of our color standardization method on the performance of Computer-aided Diagnosis(CAD). As this dataset were collected from 2 independent medical centers, the staining style variations will hamper the segmentation performance without color normalization. In camelyon16 dataset, only 110 negative WSIs have mask level label, of which 70 were from one medical center and 40 from another. We cut WSIs from two medical centers into 256×256 patches and randomly selected 80000 patches for experiment. 40000 patches were used for training and 40000 patches were used to verify the segmentation performance.

The U-Net model was used as the segmented prediction network. The batch size was set to 16 and Adam optimizer was used with a learning rate of 0.0001. All models were trained for 300000 iterations. We trained a model on data from one center, and different color standardization methods were used to transform the data from another center. The mean Intersection over Union(mIoU) and Pixel

²<https://camelyon16.grand-challenge.org/>

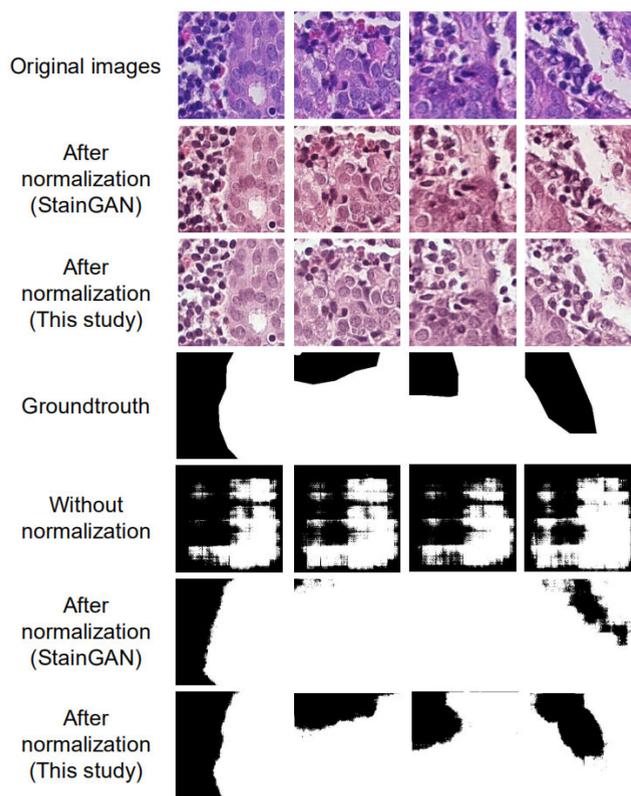


FIGURE 7. Comparisons of visual segmentation results among the proposed method, StainGAN and before color standardization.

Accuracy(PA) of two centers with or without color normalization was used to evaluate our normalization method.

We compared the StainGAN methods and our proposed model in the pre-processing step of tumor segmentation task. The mIoU and PA are list in Tabel 6 and the segmentation results are shown in Figure 7. As shown in Table 6, 40000 patches without color standardization were used for testing, mIoU was only 0.38, PA was only 0.54. Figure 7 shows the comparisons of visual results on tumor segmentation among the proposed method, StainGAN and before color standardization. Compared with the image without color standardization, the segmentation results are significantly improved after color normalization. Moreover, our method also has great advantages in tumor segmentation compared with StainGAN. Above results have demonstrates that our standardized method is of great help to improve the robustness of the CAD systems for tumor diagnosis.

V. STAIN RESTORATION

In this section, we use clinical data from a local hospital to verify the feasibility of the proposed method in stain restoration.

A. DATA AND EXPERIMENTAL SETUP

1) RESEARCH MOTIVATION

The stain of the pathological slices in hospital will gradually fade over time, affecting the appearance and the color in

TABLE 6. Influence of different color non-standard methods on segmentation performance.

Method	mIoU	PA
Without normalization	0.38	0.54
StainGAN	0.67	0.88
This study	0.68	0.91

TABLE 7. Comparisons between the proposed method and state-of-the-art method on stain restoration.

Method	SC	WD	GFLOPs
StainGAN	0.935	12.74	56.93
This study	0.946	3.45	38.96

pathological images. It can neither be used for pathological diagnosis nor as training data for artificial intelligence algorithms due to its faded color. To alleviate this problem, we used the proposed method to restore the stain of pathological images, eliminating the cumbersome operations such as reparation of specimens and rescanning.

2) DATASET

The datasets are the gastric cancer pathological images from Fujian Provincial Cancer Hospital. Faded pathological slices scanned by a scanner are served as original input images and freshly prepared pathological slices scanned by a scanner are served as target images. It is noted that these datasets are unpaired. 20,000 unpaired patches were extracted from source domain and target domain for training. During testing, 2000 unpaired patches were extracted. The size of these patches are all 256×256 .

3) EVALUATION METRICS

Due to the unpaired of datasets, we used two different image similarity metrics which are structural comparison and wasserstein distance to evaluate the quality of the generated images. Structural comparison is used to compare the structure between current images and target images. Its mathematical form is illustrated in Equation (9), where σ_{xy} represents the covariance of x and y , σ_x represents the variance of x , σ_y represents the variance of y , and c is a constant. The higher value of $s(x, y)$, the better image quality. In addition, we also use floating point calculations to measure the inference speed.

$$s(x, y) = \frac{\sigma_{xy} + c}{\sigma_x \sigma_y + c} \tag{9}$$

4) TRAINING DETAILS

In all experiments, we trained 20000 unpaired patches for approximate 100000 iterations with a batch size of 1. Other training details are similar to previous section.

B. EXPERIMENTAL RESULTS

As shown in Table 7, where SC respresents structural comparison, and WD represents wasserstein distance. Using the proposed method, compared with StainGAN, structural comparison has increased from 0.935 to 0.945, and wasserstein distance has reduced from 12.74 to 3.45. In addition, the

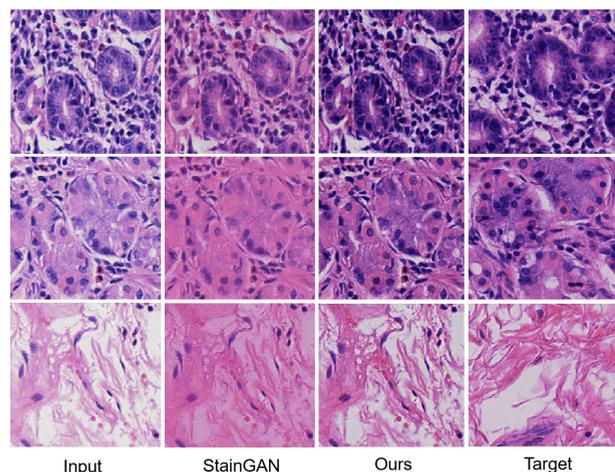


FIGURE 8. Comparisons of visual results on stain restoration between the proposed method and StainGAN.

floating point calculation amount has reduced from 56.93 to 38.96. Figure 8 shows the comparisons of visual results on stain restoration between the proposed method and StainGAN. It shows that the proposed method can not only keep the content information unchanged, but also get closer to the target images in colors, brightness and contrast compared to StainGAN. Moreover, the amount of calculation is also reduced by 33%, which means that the proposed method can speed up the stain restoration task by 33%. Above results demonstrate that our method has great advantages in stain restoration compared with StainGAN.

VI. CONCLUSION

In our study, a novel framework was proposed to transform the stain style between two pathological image collections. The measure of objective indicators and subjective evaluation of the generated images demonstrate the great superiority of the proposed method in stain style transfer compared to state-of-the-art methods. The main contributions of this work are summarized as follows: i) The generators adopts RC-Net structure, which increases the number of downsampling layers and upsampling layers to capture high-level semantic information. This can result in detailed texture and context in the constructed images. In addition, the skip connection operation is replaced with a long-range residual, which combines the copied feature maps and the decoding feature maps through summation, instead of concatenation in UNet. A 3×3 convolutional layer with a stride of 1 is used for processing copied feature maps before the summation operation. Long-range residual can not only reduce the number of trainable parameters in the following layers, but also facilitate model training. After that, we introduce channel attention mechanism into our generator architectures based on the assumption that different feature channels contribute differently to the results. Channel attention mechanism adaptively assigns different weights to different feature channels, resulting in more accurate features in the following layers. In addition,

to alleviate the problem that is single encoding and loss of edge or other information, we replaced the max pooling operation in UNet with a 3×3 convolutional layer with a stride of 2, and replaced the upsampling operation in the decoding path with a 3×3 deconvolutional layer with a stride of $\frac{1}{2}$. ii) We used additive coupling technology to construct the positive generator and opposite generator into invertible architecture based on invertible neural networks, which can not only result in more accurate images, but also accelerate the inference speed of the network by nearly twice. iii) Given that the discriminator is based on the style of the image to distinguish true or false, the style feature channels should be given greater weights and the content feature channel should be given less weights. Therefore, we introduced channel attention mechanism into the discriminator, which has been shown that it can further improve the discriminatory ability of the discriminators in stain style transfer.

We performed two different virtual coloring tasks on two datasets respectively. Firstly, the stain style transfer operation between the digital pathological images from scanner A and scanner H was performed on the datasets from the MITOS-ATYPIA14 challenge. Secondly, we use CAMELYON16 dataset to assessment about the color normalization effect on computer-aided diagnosis. Moreover, the stain restoration operation was performed on the digital pathological images of gastric cancer from a local cancer hospital, which was used to restore image colors that have faded over time. In these three tasks, the performance of proposed method has greatly outperformed state-of-the-art methods, which demonstrates the great advantages of our proposed method. In future, we will further investigate the reversible structure in low-dimensional image space instead of the high-dimensional feature space, and design a loss that is more suitable than the L_1 loss for stain style transfer.

REFERENCES

- [1] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: Whole-slide imaging and beyond," *Annu. Rev. Pathol., Mech. Disease*, vol. 8, no. 1, pp. 331–359, Jan. 2013.
- [2] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [3] S. M. Ismail, A. B. Colclough, J. S. Dinnen, D. Eakins, D. M. Evans, E. Gradwell, J. P. Osullivan, J. M. Summerell, and R. G. Newcombe, "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," *Brit. Med. J.*, vol. 298, no. 6675, pp. 707–710, 1989.
- [4] A. Andron, C. Magnani, P. G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M. Botta, and B. Terracini, "Malignant mesothelioma of the pleura: Interobserver variability," *J. Clin. Pathol.*, vol. 48, no. 9, pp. 856–860, 1995.
- [5] J. I. Epstein, W. C. Allsbrook, Jr., M. B. Amin, and L. L. Egevad, "Update on the Gleason grading system for prostate cancer: Results of an international consensus conference of urologic pathologists," *Adv. Anatomic Pathol.*, vol. 13, no. 1, pp. 57–59, 2006.
- [6] M. H. Stoler and M. Schiffman, "Interobserver reproducibility of cervical cytologic and histologic interpretations: Realistic estimates from the ASCUS-LSIL Triage Study," *Jama*, vol. 285, no. 11, pp. 1500–1505, 2001.
- [7] C. A. Roberts, P. D. Beitsch, C. E. Litz, D. S. Hilton, G. E. Ewing, E. Clifford, W. Taylor, M. R. Hapke, A. Babaian, I. Khalid, J. D. Hall, G. Lindberg, K. Molberg, and H. Saboorian, "Interpretive disparity among pathologists in breast sentinel lymph node evaluation," *Amer. J. Surg.*, vol. 186, no. 4, pp. 324–329, Oct. 2003.

- [8] P. J. van Diest, C. H. M. van Deurzen, and G. Cserni, "Pathology issues related to SN procedures and increased detection of micrometastases and isolated tumor cells," *Breast Disease*, vol. 31, no. 2, pp. 65–81, Nov. 2010.
- [9] G. Litjens, P. Bandi, and B. E. Bejnordi, "1399 H&E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset," *GigaScience*, vol. 7, no. 6, 2018, Art. no. giv065.
- [10] D. L. Page, "Theory and practice of histological techniques," *Human Pathol.*, vol. 14, no. 10, pp. 925–926, 1983.
- [11] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. de Souza, A. Baidoshvili, G. Litjens, B. van Ginneken, I. Nagtegaal, and J. van der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 160–163.
- [12] M. Niethammer, D. Borland, J. Marron, J. Woosley, and N. E. Thomas, "Appearance normalization of histology slides," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Berlin, Germany: Springer, 2010, pp. 58–66.
- [13] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohece, and P. Quirke, "Colour normalisation in digital histopathology images," in *Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endoscopy*, vol. 100, 2009, pp. 100–111.
- [14] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro*, Jun. 2009, pp. 1107–1110.
- [15] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, Jun. 2014.
- [16] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 4, pp. 34–41, Jul./Aug. 2001.
- [17] B. E. Bejnordi, N. Timofeeva, I. Otteholler, N. Karssemeijer, and J. A. W. M. V. Der Laak, "Quantitative analysis of stain variability in histology slides and an algorithm for standardization," *Proc. SPIE*, vol. 9041, Mar. 2014, Art. no. 904108.
- [18] A. Basavanthally and A. Madabhushi, "EM-based segmentation-driven color standardization of digitized histopathology," *Proc. SPIE*, vol. 8676, Mar. 2013, Art. no. 86760G.
- [19] P. Bautista, N. Hashimoto, and Y. Yagi, "Color standardization in whole slide imaging using a color calibration slide," *J. Pathol. Informat.*, vol. 5, no. 1, p. 4, 2014.
- [20] A. C. Ruijrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, 2001.
- [21] A. Janowczyk, A. Basavanthally, and A. Madabhushi, "Stain normalization using sparse AutoEncoders (StaNoSA): Application to digital pathology," *Computerized Med. Imag. Graph.*, vol. 57, pp. 50–61, Apr. 2017.
- [22] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Med. Image Anal.*, vol. 33, pp. 170–175, Oct. 2016.
- [23] T. de Bel, M. Hermsen, J. Kers, J. van der Laak, "Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology," in *Proc. 2nd Int. Conf. Med. Imag. Deep Learn., Process. Mach. Learn. Res.*, 2019, pp. 151–163.
- [24] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10215–10224.
- [25] J.-H. Jacobsen, A. Smeulders, and E. Oyallon, "I-RevNet: Deep invertible networks," 2018, *arXiv:1802.07088*. [Online]. Available: <http://arxiv.org/abs/1802.07088>
- [26] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, and U. Köthe, "Analyzing inverse problems with invertible neural networks," 2018, *arXiv:1808.04730*. [Online]. Available: <http://arxiv.org/abs/1808.04730>
- [27] S. Roy, A. K. Jain, S. Lal, and J. Kini, "A study about color normalization methods for histopathology images," *Micron*, vol. 114, pp. 42–61, Nov. 2018.
- [28] X. Li and K. N. Plataniotis, "Circular mixture modeling of color distribution for blind stain separation in pathology images," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 150–161, Jan. 2017.
- [29] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1962–1971, Aug. 2016.
- [30] D. Onder, S. Zengin, and S. Sarioglu, "A review on color normalization and color deconvolution methods in histopathology," *Appl. Immunohistochem. Mol. Morphol.*, vol. 22, no. 10, pp. 713–719, 2014.
- [31] H. Bhat, A. Kanakatte, R. Nayak, and J. Gubbi, "A hybrid approach for nucleus stain separation in histopathological images," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 1218–1221.
- [32] M. Salvi, N. Michielli, and F. Molinari, "Stain color adaptive normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105506.
- [33] A. Bentaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 792–802, Mar. 2018.
- [34] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "StainGAN: Stain style transfer for digital histological images," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 953–956.
- [35] N. Zhou, D. Cai, X. Han, and J. Yao, "Enhanced cycle-consistent generative adversarial network for color normalization of H&E stained images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer*, 2019, pp. 694–702.
- [36] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [37] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," 2015, *arXiv:1506.05751*. [Online]. Available: <http://arxiv.org/abs/1506.05751>
- [38] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
- [39] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2016, pp. 597–613.
- [40] M. F. Mathieu, J. J. Zhao, J. Zhao, A. Ramesh, P. Sprechmann, and Y. LeCun, "Disentangling factors of variation in deep representation using adversarial training," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 5040–5048.
- [41] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.
- [42] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
- [43] T. F. A. van der Ouderaa and D. E. Worrall, "Reversible GANs for memory-efficient Image-To-Image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4720–4728.
- [44] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: <http://arxiv.org/abs/1802.05957>
- [45] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
- [46] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: <http://arxiv.org/abs/1710.10196>
- [47] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, Mar. 2018.
- [48] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," 2016, *arXiv:1605.08803*. [Online]. Available: <http://arxiv.org/abs/1605.08803>
- [49] A. Mahendran and A. Vedaldi, "Visualizing deep convolutional neural networks using natural pre-images," *Int. J. Comput. Vis.*, vol. 120, no. 3, pp. 233–255, Dec. 2016.
- [50] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, "Improved variational inference with inverse autoregressive flow," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4743–4751.
- [51] M. Cisse, P. Bojanowski, E. Grave, Y. Dauphin, and N. Usunier, "Parseval networks: Improving robustness to adversarial examples," 2017, *arXiv:1704.08847*. [Online]. Available: <http://arxiv.org/abs/1704.08847>
- [52] L. Dinh, D. Krueger, and Y. Bengio, "Nice: Non-linear independent components estimation," 2015, *arXiv:1410.8516*. [Online]. Available: <https://arxiv.org/abs/1410.8516>

- [53] A. N. Gomez, M. Ren, R. Urtasun, and R. B. Grosse, "The reversible residual network: Backpropagation without storing activations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 2214–2224.
- [54] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [55] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
- [56] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*. [Online]. Available: <http://arxiv.org/abs/1807.06521>
- [57] A. Boesen Lindbo Larsen, S. Kaae Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," 2015, *arXiv:1512.09300*. [Online]. Available: <http://arxiv.org/abs/1512.09300>
- [58] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.
- [59] H. Ling and K. Okada, "An efficient Earth Mover's distance algorithm for robust histogram comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 840–853, May 2007.
- [60] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Edu. (ITME)*, Oct. 2018, pp. 327–331.
- [61] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [62] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.



JUNLIN LAN received the B.S. degree from the School of Physics and Information Engineering, Fuzhou University, in 2018, where he is currently pursuing the Ph.D. degree. His research interests include image recognition and processing, and computer aided diagnosis.



SHAOJIN CAI received the B.S. degree from the School of Optoelectronics and Information Engineering, Fujian Normal University, in 2017, and the M.S. degree from the School of Physics and Information Engineering, Fuzhou University, in 2020. His research interests include medical image analysis and reconstruction.



YUYANG XUE received the M.D. degree from the School of Computer Science, University of Southampton, U.K., in 2017, and the M.D. degree from the School of Computer Science, Fuzhou University, in 2016. He is currently pursuing the Ph.D. degree with the Graduate School of Science and Technology, University of Tsukuba. His research interests include computer vision and image processing.



QINQUAN GAO received the B.S. degree in automation from Xiamen University, China, in 2008, the M.S. degree in systems engineering, in 2010, and the Ph.D. degree from Imperial College London, in 2014. He is currently an Associate Professor with Fuzhou University, working on model compressing, machine learning, biomedical image processing, and computer vision.



MIN DU received the Ph.D. degree in electrical engineering from Fuzhou University, Fuzhou, China, in 2005. Since 2007, she has been an Associate Director of the Fujian Key Laboratory of Medical Instrumentation and Pharmaceutical Technology. She is currently a Professor and a Ph.D. Supervisor with Fuzhou University. Her research interests include smart instrument and photoelectric.



HEJUN ZHANG received the M.A. degree in medicine from Fujian Medical University, in 2009. He currently works as a Doctor with the Department of Pathology, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital. His research interest includes diagnostic pathology of tumors.



ZHIDA WU received the B.S. degree from the Fujian University of Traditional Chinese Medicine, in 2018. He is currently pursuing the master's degree with Fujian Medical University. His research interests include diagnosis and treatment of gastric cancer.



YANGLIN DENG received the B.S. degree from the School of Information Engineering, Jimei University, in 2019. He is currently pursuing the master's degree with the School of Physics and Information Engineering, Fuzhou University. His research interests include medical image processing and few-shot learning.



YUXIU HUANG received the B.S. degree from the School of Physics and Information Engineering, Fuzhou University, in 2019, where she is currently pursuing the master's degree. Her research interests include medical image analysis and bioinformatics.



TONG TONG received the Ph.D. degree from Imperial College London, in 2015. He was a Research Fellow with the MGH/Harvard Medical School, in 2016. He is currently a Full Professor with the College of Physics and Information Engineering, Fuzhou University. His research interests include machine learning, medical image analysis, and computer aided diagnosis.



GANG CHEN is currently the Director of the Department of Pathology, Fujian Provincial Cancer Hospital. He is also the Director of the Molecular Pathology Laboratory of the Fujian Cancer Institute. He has engaged in tumor pathological diagnosis and research for 20 years. His research interests include lymphoma, obstetrics and gynecology and soft tissue pathology, especially good at the diagnosis and differential diagnosis of lymphoma.

...