RestainNet: a self-supervised digital re-stainer for stain normalization

Bingchao Zhao, Jiatai Lin, Changhong Liang, Zongjian Yi, Xin Chen, Bingbing Li, Weihao Qiu, Danyi Li, Li Liang, Chu Han, and Zaiyi Liu

Abstract—Color inconsistency is an inevitable challenge in computational pathology, which generally happens because of stain intensity variations or sections scanned by different scanners. It harms the pathological image analysis methods, especially the learning-based models. A series of approaches have been proposed for stain normalization. However, most of them are lack of flexibility in practice. In this paper, we formulated stain normalization as a digital re-staining process and proposed a self-supervised learning model, which is called RestainNet. Our network is regarded as a digital re-stainer which learns how to re-stain an unstained (grayscale) image. Two digital stains, Hematoxylin ($\mathcal{H}$) and Eosin ($\mathcal{E}$), were extracted from the original image by Beer-Lambert's Law. We proposed a staining loss to maintain the correctness of stain intensity during the re-staining process. Thanks to the self-supervised nature, paired training samples are no longer necessary, which demonstrates great flexibility in practical usage. Our RestainNet outperforms existing approaches and achieves state-of-the-art performance with regard to color correctness and structure preservation. We further conducted experiments on the segmentation and classification tasks and the proposed RestainNet achieved outstanding performance compared with SOTA methods. The self-supervised design allows the network learn any staining style with no extra effort.

Index Terms—Computational Pathology, Stain Normalization,

Bingchao Zhao is with the School of Medicine, South China University of Technology, Guangzhou 510006, China, and with the Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, 510080, China. (e-mail: zbcajj@qq.com)

Jiatai Lin and Zongjian Yi are with the South China University of Technology, Guangzhou 510006, China, and with the Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, 510080, China (e-mail: linjiatai_cs@163.com; jiangiegie@qq.com)

Bingbing Li is with Guangdong Provicial People's Hospital Ganzhou Hospital and Department of Pathology, Nanfang Hospital and Basic Medical College, Southern Medical University (email: libingbing19932021@126.com)

Li Liang, Weihao Qiu, Danyi Li are with Department of Pathology, Nanfang Hospital and Basic Medical College, Southern Medical University, Guangzhou 510515, Guangdong, China and Guangdong Province Key Laboratory of Molecular Tumor Pathology, Guangzhou 510515, Guangdong, China. (email: lli@smu.edu.cn; qiuweihao42@163.com; lidanyi26@163.com)

Xin Chen is with the Department of Radiology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, 510180, China. (e-mail: wolfchenxin@163.com)

Chu Han is with the Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, 510080, China. (e-mail: hanchu@gdph.org.cn)

Zaiyi Liu and Changhong Liang are with the Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, 510080, China, and with the School of Medicine, South China University of Technology, Guangzhou 510006, China. (e-mail: zyliu@163.com; liangchanghong@gdph.org.cn)
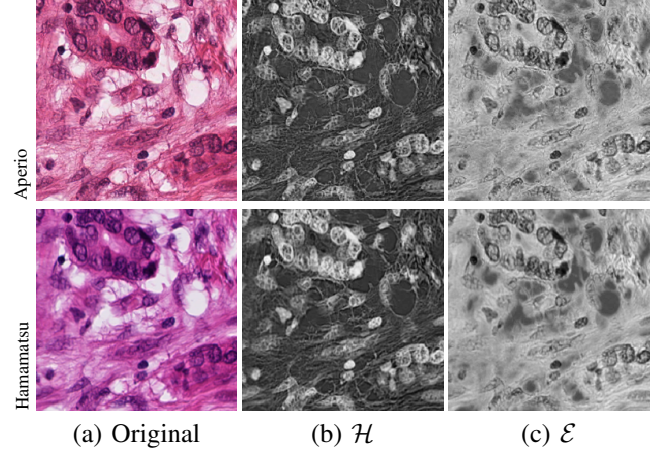
Fig. 1. Two extracted digital dyes. (a) are two images scanned by two difference scanners, Aperio and Hamamatsu. (b) and (c) are two images extracted by Beer-Lambert's Law [15]. They demonstrate consistent dyes intensity against different scanners.

Self-supervised Learning

I. INTRODUCTION

Computational pathology is now becoming more and more popular with the invention of high-precision microscope section scanners and the development of deep neural networks. It aims to quantitatively evaluate the microenvironment of diseases in the microscopic level in an objective manner, and finally achieves precision medicine, i.e., disease diagnosis [3], treatment response [4], and even gene prediction [5]. Although CNN models have demonstrated outstanding performance in various computational pathology applications, most of them still suffer from color inconsistency problem. It is introduced by the various steps during sectioning, i.e., inconsistent tissue thickness, different manufacturers of the dye, different sectioning protocols and etc [19]. In addition, scanning with different scanners may also lead to color inconsistency problem, which harms the generalization of the learning-based models, especially when applying them to the data from different institutions. That is the reason why color normalization is always an essential preprocessing step before model training [2], [7].

To alleviate the color inconsistency problem, many color normalization methods have been proposed. Traditional methods normalize pathological images in a statistical manner [24], [25]. Computational-based methods take characteristics of H&E stained images into consideration, and decompose them into different dye intensity images [13], [27], [30], [31], [33].

And then they approximate the color distribution of the source image to the distribution of a selected template image. However, the selection bias of the reference images may lead to inconsistent results. Moreover, optimization-based approaches are easily trapped in local optimum [21]. Recently, GAN-based models [35], [36] have been proposed and achieved promising results. Due to the data-driven nature, they require a large amount of paired training samples, which are hard to collect. A feasible solution is to scan the pathological section using different scanners. But we still cannot guarantee perfect groundtruth images because of the misalignment problem. Additional alignment processes may damage the structure of the original image. Moreover, it is impractical to re-collect the training data when an unseen staining style comes. Besides the stain normalization performance, the flexibility of the model extension should be also considered.

In this paper, we formulate stain normalization as a digital re-staining process. We first de-stain the pathological image by transforming the input image into a grayscale image. In order to preserve structure, L channel is extracted from Lab color space. Then we re-stain it by applying two digital dyes, $\mathcal{H}$ and $\mathcal{E}$. According to the fact that hematoxylin always stains nuclei blue and Eosin always stains the cytoplasm and extracellular matrix pink, we extract two digital dyes $\mathcal{H}$ and $\mathcal{E}$ by Beer-Lambert's Law [15]. It can be observed that $\mathcal{H}$ and $\mathcal{E}$ are insensitive to different stain styles as demonstrated in Fig. 1.

To achieve this, we proposed a self-supervised learning model, called RestainNet, which plays a role as a digital re-stainer. The network learns how to stain an unstained (grayscale) image with two digital stains $\mathcal{H}$ and $\mathcal{E}$. In order to maintain the correctness of digital stains after re-staining process, we proposed a new staining loss. GAN loss is introduced to differentiate the real or the generated images. Due to the self-supervised manner, we do not have to collect paired images for network training which is easily implemented into practical applications. Furthermore, our model is able to generate visually pleasant normalization results while preserving the structure information. Several experiments were conducted on public datasets to evaluate the effectiveness of stain normalization results. Our proposed model achieves state-of-the-art performance in all the quantitative measurements compared with existing approaches. We also conducted experiments to prove that our proposed model can benefit the following segmentation and classification tasks in practice. Furthermore, the ablation study validates the stability of the model with regard to the vibration of the digital dyes.

Our major contributions can be summarized as follows:

- We proposed a self-supervised stain normalization model, called RestainNet. Thanks to the self-supervised nature, paired training images are no longer necessary, which shows great flexibility in practical usage.
- We proposed a new staining loss to guarantee the correctness of digital dyes.
- Our RestainNet achieves state-of-the-art performance comparing with existing models in three different tasks, the color normalization task, the segmentation task and the classification task.

II. RELATED WORKS

A. Computational-based Normalization

To reduce the impact of color inconsistency, some works convert RGB images into grayscale images to protect the image texture information [14]. Although these methods obtain stable texture features that do not change with the stain style, but discard the color features that contain richer semantic information. If we regard each stain style as specific color distribution, an intuitive method is to convert all images into the same color distribution space. This conversion is usually not performed in the RGB space, because it is more difficult to normalize the type of color than the intensity. Reinhard et al. [25] use the de-correlated Lab components obtained from the RGB image to perform statistical matching with the histogram of the target image. This method may result in an image that does not conform to the actual distribution because the types and proportions of tissue components in most images are not constant.

The color of the pathological image is characterized by the dye and its concentration attached to the tissue. The RGB image can be converted to the dye concentration space through the specified mapping method, providing a more reasonable paradigm for describing the color distribution of pathological images. Among them, the widely used color deconvolution algorithm [13] decomposes the RGB image into Hematoxylin (H), Eosin (E), and Diaminobenzidine (DAB) components through the "staining matrix" obtained by statistical evaluation. The data-dependent "staining matrix" may lead to miscalculations of dye concentration when generalized to unseen data sets. Vahadane et al. [27] perform stain separation based on non-negative matrix factorization which, in a supervised way, clusters the image to two empirically selected tissue clusters. Furthermore, Janowczyk et al. [18] use Sparse AutoEncoders (StaNoSA) to independently performed normalization of each cluster to alleviate the impact of the tissue clusters imbalance among images.

These Computational-based algorithms perform stain normalization by mapping the image to the color space represented by the target image or selected distribution. But this characteristic makes them limited in actual implementation, because the normalized image may appear with tissue categories outside the selected image.

B. Learning-based Normalization

With the rapid progress of deep learning technology, convolutional neural networks have been widely applied in stain normalization tasks. Bug et al. [34] proposed a feature-aware framework that formulates the normalization task as a non-linear mapping between pixels but causes image compression. Inspired by the promising performance of the generative model (e.g., GAN [23], CycleGANs [37]), many works regard stain normalization as an image style transfer task that aims to synthesize the target image texture but preserves the semantic content of the source image. Shaban et al. [35] introduced CycleGANs [37] to realize the style transfer between images from two different scanners. Salehi et al. [36] designed an unsupervised model based on conditional generative adversarial
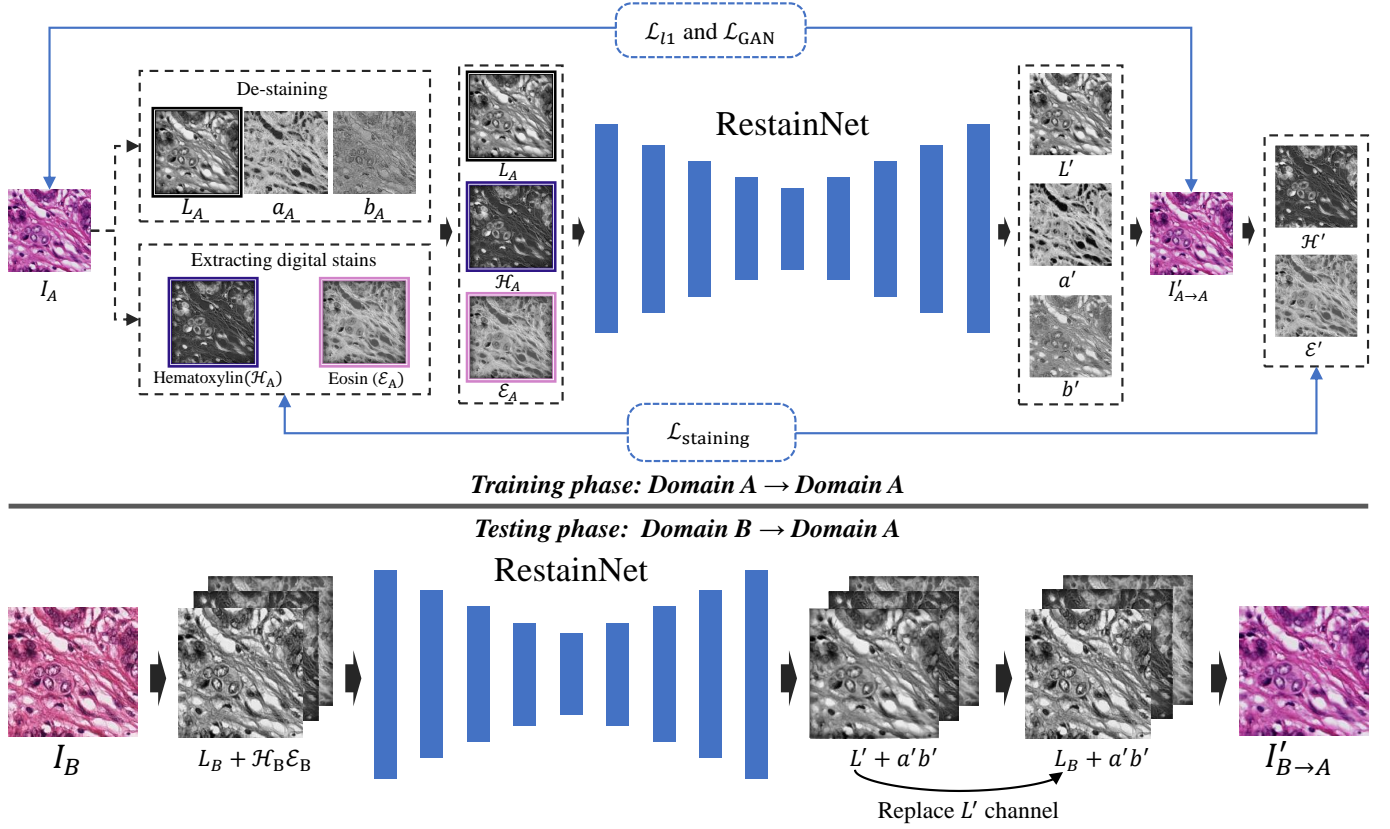
Fig. 2. Overview of RestainNet. L is the luminance channel of Lab color space. $\mathcal{H}$ and $\mathcal{E}$ are two digital dyes extracted by Beer-Lambert's Law [15]. During the training phase, RestainNet learns how to re-stain the grayscale image back to the original image in a self-supervised manner. During the testing phase, images from the source domain B can be transferred to the target domain A.

networks (cGANs) [23] by coloring grayscale images to the target stain style. A multi-task network equipped with normalized branches and task-specific branches [21] is proposed to improve the generalization ability of target tasks on multiple data sets. Liang et al. [22] trained a multi-task network that combined clinical diagnosis information to synthesize the structure-maintained normalized image.

The learning-based method trains the huge parameters using a well-designed learning strategy to transfers the source image to the target style. With a large amount of training data, the deep learning model can fit the color distribution of a specified dataset, and generate images with richer colors than the traditional mathematical method that refers to a single target image. The model synthesis images with richer semantic information, but the large number of high-quality images required for training restricts its practicality. Moreover, existing methods transfer styles through image texture similarity but rarely use prior knowledge, which may cause normalization errors when feeding images with different tissue types or proportions from the training images.

III. Methodology

In this section, we introduce the proposed model RestainNet for stain normalization. Section III-A shows the problem formulation. Section III-B demonstrates the details of the complete model. Section III-C shows the loss function of the model.

A. Problem Formulation

In this paper, we proposed a simple and flexible model for stain normalization. Generally, we formulate stain normalization as a digital re-staining process, as demonstrated in Fig. 2. Given an input image, we first de-stain it by extracting the grayscale image. To preserve the structure of the original image, L channel is extracted from Lab color space. Then two digital dyes, Hematoxylin ($\mathcal{H}$) and Eosin ($\mathcal{E}$), are extracted to represent the dye intensity maps of each pixel.

Given a target domain $\mathcal{D}_A$ and a source domain $\mathcal{D}_B$, let us denote two images from two different domains as $I_A \in \mathcal{D}_A$ and $I_B \in \mathcal{D}_B$. $f_\theta(\cdot)$ denotes the RestainNet with the parameter set θ.

- In *training phase*, we aim to learn the internal characteristics of the target domain $\mathcal{D}_A$ in a self-supervised manner. Given the grayscale image L_A and two digital dyes $\mathcal{H}_A$ and $\mathcal{E}_A$, RestainNet generates the re-stained image $I'_{A \to A}$ by $f_\theta(L_A, \mathcal{H}_A, \mathcal{E}_A)$ which satisfies $I'_{A \to A} \approx I_A$.
- In *testing phase*, RestainNet allows input images from any other source domain, e.g. domain $\mathcal{D}_B$, and re-stains it to the target domain $\mathcal{D}_A$, i.e., $f_\theta(L_B, \mathcal{H}_B, \mathcal{E}_B) = I'_{B \to A}$. Finally, we replace $(L'_A, \mathcal{H}_A, \mathcal{E}_A)$ with $(L_B, \mathcal{H}_A, \mathcal{E}_A)$ to save the complete texture information. In this paper, the results of the proposed model are obtained in this way.

B. RestainNet

Network architecture: RestainNet intends to learn to stain the de-stained image with the guidance of the H and E digital dyes through a self-supervised learning training phase. The re-stained images should maintain the original resolution with the input, to achieve a structure-preserving and practical normalizationing For this reason, the re-stainer can be generalized to a structure with such properties, and not being restricted to a fixed network. The U-net [17] structure maintains the invariance of the output resolution through the same number of upsampling and downsampling phases and employees skip links to protect the high-frequency information, which can serve as an eligible re-stainer. In addition, the autoencoder-decoder structure with a limited number of encoding can also achieve the above requirements, otherwise, the information loss caused by excessive downsampling can result in the inability to generate a complete image. We find that an autoencoder with two downsamplings and two upsampling can achieve the above requirements. In this paper, a U-net structure is adopted, although these two structures have shown similar performance in the experiment.

Digital stains extraction: Hematoxylin, and Eosin are two stains that make pathological section pigment for better observation. During the staining process, hematoxylin is used for staining the cells into blue and eosin is used for staining the cytoplasm into pink. Therefore, the color appearance of the H&E stained pathological section is determined by the specific dye attached to the tissue.

We want to extract two digital dyes for our proposed RestainNet. According to Beer-Lambert's Law [15], the transmission of light through a material is described by the incident light $\mathcal{I}_0$, the dye intensity A and the absorption factor c as follows:

$$\mathcal{I}_r = \mathcal{I}_0 exp(-Ac) \tag{1}$$

where $\mathcal{I}_r$ is the received light passing through the section. After a simple mathematical transformation of Equation 1, we have:

$$OD = -\log \frac{\mathcal{I}_r}{\mathcal{I}_0} = Ac \tag{2}$$

Note that OD is the abbreviation of optical density. The dye intensity A can be obtained by decomposing the optical intensity OD due to the linearity. Through experimental statistics, we can get a normalized pure stain OD matrix M:

$$M = \begin{bmatrix} 0.65 & 0.7 & 0.29 \\ 0.07 & 0.99 & 0.11 \\ 0.27 & 0.57 & 0.78 \end{bmatrix}$$

On this basis, we can further calculate the optical density y of each pixel by the following equation:

$$y = AM \Rightarrow A = yM^{-1} \tag{3}$$

C. Loss function

We first introduce the adversarial loss to evaluate the re-stained image.

$$\mathcal{L}_{GAN} = E_{I\sim p_{data}(I)}[\log D(I)] + E_{I'\sim p_z(I')}[log(1-D(G(I')))]$$

where I is the input and I' is the generated image, model C maximize the input while G minimizes the $log(1-D(G(I')))$.

Since the model generates images in Lab color space. L1 loss is applied on Lab color space as follows:

$$\mathcal{L}_{l1} = \Sigma(||L - L'||_1 + ||a - a'||_1 + ||b - b'||_1) \tag{4}$$

The existing learning-based normalization algorithms usually evaluate the generated images in the RGB color space, but ignore the unique optical characteristics of pathological images. To preserve the consistency of the digital stain intensity during the re-staining process, we proposed a new loss, called staining loss. From the re-stained image I' generated by RestainNet, we extract its digital dyes $\mathcal{H}'$ and $\mathcal{E}'$ and compare them with the original ones, $\mathcal{H}$ and $\mathcal{E}$. And the staining loss is calculated by measuring the L1 distance of them.

$$\mathcal{L}_{staining} = \Sigma(||\mathcal{H} - \mathcal{H}'||_1 + ||\mathcal{E} - \mathcal{E}'||_1) \tag{5}$$

The final objective function is defined as:

$$\mathcal{L} = \lambda_{GAN}\mathcal{L}_{GAN} + \lambda_{l1}\mathcal{L}_{l1} + \lambda_{staining}\mathcal{L}_{staining} \tag{6}$$

where $\lambda_{\{GAN,l1,staining\}}$ are the hyperparameters.

D. Implementation and Training Details

We build our model under Ubuntu18.04 using Pytorch1.8 architecture. The optimizer is Adam [1]. The momentum parameter β is 0.5. The learning rate is 0.0002 with an exponential decay of 0.6. The network runs on a workstation with an NVIDIA GeForce RTX 3090. We select the best performing model on the validation set as our final model. Hyperparameters λ_{GAN}, λ_{l1}, λ_{stain} are set to 0.1, 1, 1, respectively.

IV. EXPERIMENTS

A. Dataset and Evaluation Metrics

We evaluate our proposed method in three datasets, including a stain normalization dataset (MITOS-ATYPIA14 challenge [1]), a patch-level classification dataset (Zhao et al [20]) and a segmentation dataset (MICCAI'16 GlaS challenge dataset [29]).

MITOS-ATYPIA14 challenge contains images with three magnifications of 10X, 20X, and 40X, which were respectively scanned by two different scanners Aperio Scanscope XT and Hamamatsu Nanozoomer 2.0-HT. Our paper uses images under 20X magnification, which contains 424 images, including 300 images for the training set and 124 images for the testing set. All the images were cropped into $256{\times}256$ without overlapping. We randomly selected 600 patches from the training set for validating our network. The validation set was not involved in network training. Due to the self-supervised nature, our model does not require paired images for training. So groundtruth images are only used for validating the performance of the proposed model.

Zhao [20] dataset, a colorectal cancer classification dataset, is committed to achieving tissue segmentation of pathological

---

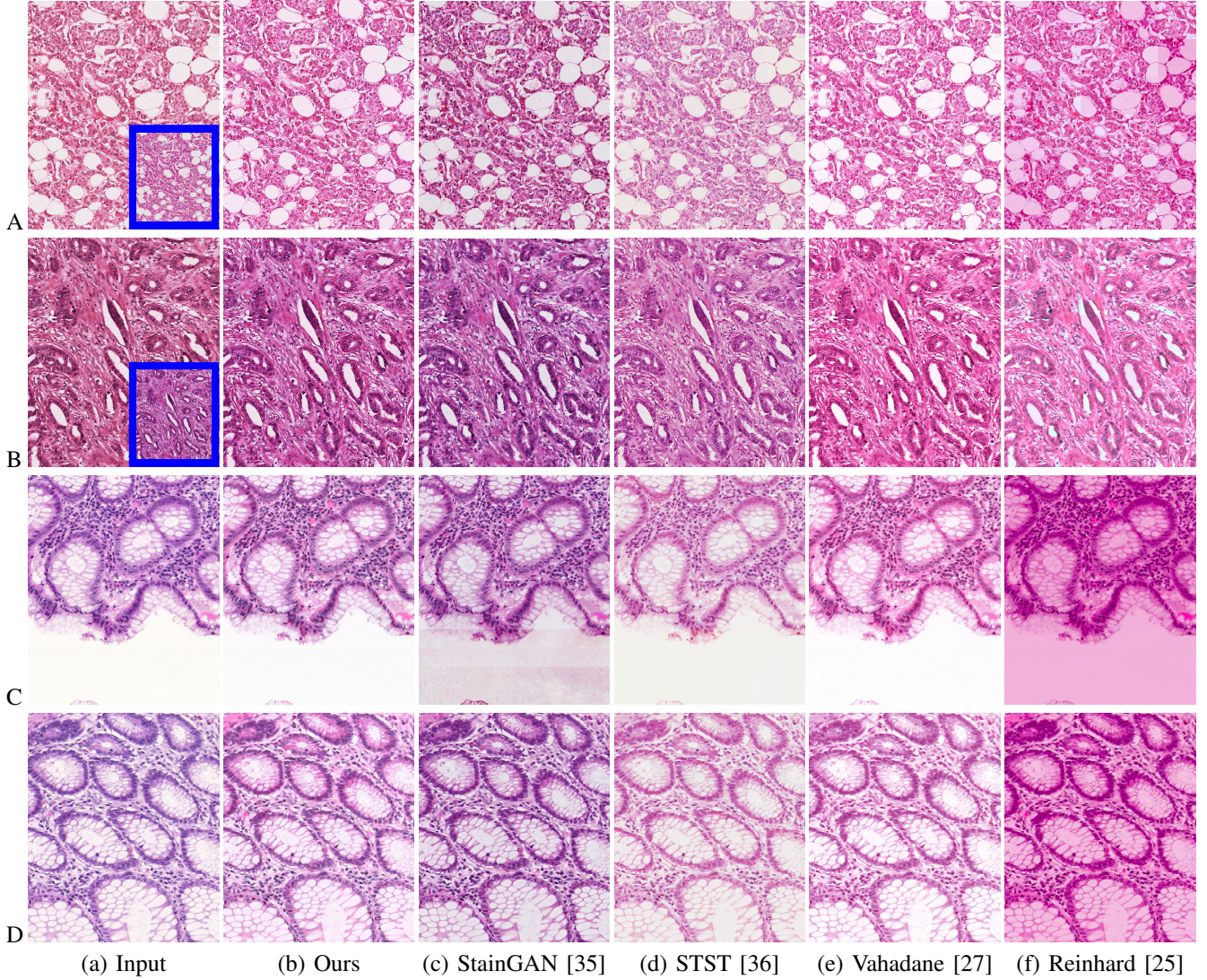[1] https://mitos-atypia-14.grand-challenge.org/Home/

Fig. 3. Qualitative comparison with different methods. (a) is the input images scanned by Aperio (source domain). Images in blue box were scanned by Hamamatsu (target domain). (b)-(f) are the results generated from different methods. Since the input images in row C and D are from MICCAI'16 GlaS challenge dataset, there is no reference image from the target domain.

| (a) Input | (b) Ours | (c) StainGAN [35] | (d) STST [36] | (e) Vahadane [27] | (f) Reinhard [25] |

images through the patches classification task and consists of multiple centers and multiple tissues pathological images. The entire data set contains a total of nine tissues, which are blank areas and eight tissue types: adipose, debris, lymphocyte aggregates, mucus, muscle, normal mucosa, stroma, and tumor epithelium. The complete dataset consists of four subsets, which are training set (283.1k patches), test set1 (16.3k patches), test set (22.5k patches), TSR evaluation set (126 blocks). In this paper, we only use the training set and the test set1 and separate the data from the TCGA as the validation set.

MICCAI'16 GlaS challenge dataset [29] is published for the study of gland segmentation of colorectal cancer. It contains 165 $\mathcal{H\&E}$ stained pathological images and corresponding gland segmentation masks. All images are divided into three subsets: Training Part (85 images), Test Part A (60 images), and Test Part B (20 images). In the experiment of this paper, all the images are cropped into 256×256 pixels patches, we use 30% of the training part as the validation set and merge the two test parts.

We used multiple quantitative measurements to evaluate the performance of our network, including: Feature Similarity (FSIM) Index [12], Peak Signal-to-Noise Ratio (PSNR), Erreur Relative Globale Adimensionnelle de Synthèse (ER-GAS),Universal Quality Index (UQI) [11], Structural Similarity Index (SSIM) [10], Mean Squared Error (MSE),Relative Average Spectral Error (RASE), Root Mean Square Error (RMSE) [8], Multi-scale Structural Similarity Index (MS-SSIM) [9].

B. Comparisons with Existing Approaches

We use the same setting with the existing methods for a fair comparison. Images scanned by the Hamamatsu Nanozoomer 2.0-HT scanner are regarded as the target domain, and images scanned by the Aperio Scanscope XT scanner are regarded as the source domain.

Table I demonstrates the quantitative comparisons with two computational-based stain normalization methods [25], [27]

TABLE I
QUANTITATIVE COMPARISONS.

Method	STST [36]	Reinhard [25]	Vahadane [27]	StainGAN [35]	Ours (w/o staining loss)	Ours
FSIM	0.878 ± 0.052	0.872 ± 0.060	0.884 ± 0.053	0.904 ± 0.023	0.901 ± 0.033	**0.917 ± 0.019**
PSNR	18.627 ± 2.731	19.384 ± 2.772	18.571 ± 2.864	20.689 ± 3.422	20.505 ± 3.570	**21.550 ± 3.665**
ERGAS	7.166 ± 2.75E3	7.03E3 ± 2.55E3	7.39E3 ±2.80E3	5.91E3 ± 2.98E3	6.37E3 ±3.52E3	**5.70E3 ± 2.83E3**
UQI	0.951 ± 0.043	0.964 ± 0.038	0.96 ± 0.045	0.966 ± 0.056	0.93 ± 0.065	**0.973 ± 0.042**
SSIM	0.842 ± 0.102	0.843 ± 0.05	0.863 ± 0.079	0.916 ± 0.018	0.873 ± 0.066	**0.916 ± 0.018**
MSE	1.19E3 ± 1.25E3	0.95E3 ± 0.77E3	1.23E3 ± 1.30E3	0.85E3 ± 0.82E3	0.87E3 ± 1.105E3	**0.76E3 ± 0.77E3**
MS-SSIM	0.924 ± 0.066	0.941 ± 0.061	0.947 ± 0.061	0.954 ± 0.024	0.953 ± 0.031	**0.976 ± 0.006**
RASE	1.03E3 ± 3.92E2	1.00E3 ± 3.65E2	1.05E3 ± 4.01E2	1.02E2 ± 4.02E2	0.917E3 ± 4.12E2	**8.12E2 ± 4.10E2**
RMSE	31.728 ± 13.415	28.938 ± 10.582	32.096 ± 13.982	25.764 ± 14.12	26.900 ± 14.818	**23.770 ± 14.02**

and two GAN-based models, StainGAN [35] and STST [36]. Reinhard et al [25] match the histograms after converting the RGB image to the Lab color space. Vahadane et al [27] perform stain separation base on non-negative matrix factorization to improve the accuracy of the dye concentration matrix obtained. Since the proposed model performs stain normalization in Lab color space, an unchanged L channel helps to preserve image structure. Moreover, two digital dyes $\mathcal{H}$ and $\mathcal{E}$ demonstrate great consistency of dye concentration with different scanners. Therefore, our model demonstrates superiority in both structure-preserving and stain style transfer accuracy. We also evaluate the effectiveness of our proposed staining loss by comparing our model without this loss in Table I. With the proposed staining loss, RestainNet gets overall improvements in all the quantitative measurements.

Fig. 3 shows the qualitative comparisons with the existing methods mentioned in Table I. The computational-based methods (e.g. (e) and (f)) have good color consistency and are less affected by the source image, but the imbalance of tissue clusters still misleads the method to perform a global stain style transfer. For example in (f), the blank area is incorrectly stained by the algorithm to reduce the histogram distance between the source image and the target image. In comparison, GAN-based methods (c) and (d) demonstrate great flexibility in global-local stain style transfer. However, StainGAN (c) derived from CycleGAN introduces artifacts in the results, which is the common problem of CycleGAN-based models. STST (d) derived from cGAN is not able to preserve the correct stain style. Since our RestainNet uses the L channel from the Lab color space as the model input, it can greatly preserve the structural information. Defining the digital dyes in every pixel can ensure the correctness of the stain style. Furthermore, self-supervised learning alleviates the requirements of paired training data.

C. Stain Normalization in the Segmentation Task

In this experiment, we directly applied the stain normalization algorithms on a MICCAI'16 GlaS challenge dataset [29] as the preprocessing step. To be fair, the training sets with different stain normalization models were trained with the same network architecture U-net [17] with the same epochs. All learning-based stain normalization methods were trained on the MITOS-ATYPLA14 dataset in this experiment. The quantitative results of the segmentation under different stain normalization methods are shown in Table II.

TABLE II
QUANTITATIVE COMPARISONS OF DIFFERENT NORMALIZATION METHODS IN THE SEGMENTATION TASK. ORI IS THE ORIGINAL IMAGES WITHOUT STAIN NORMALIZATION PREPROCESSING.

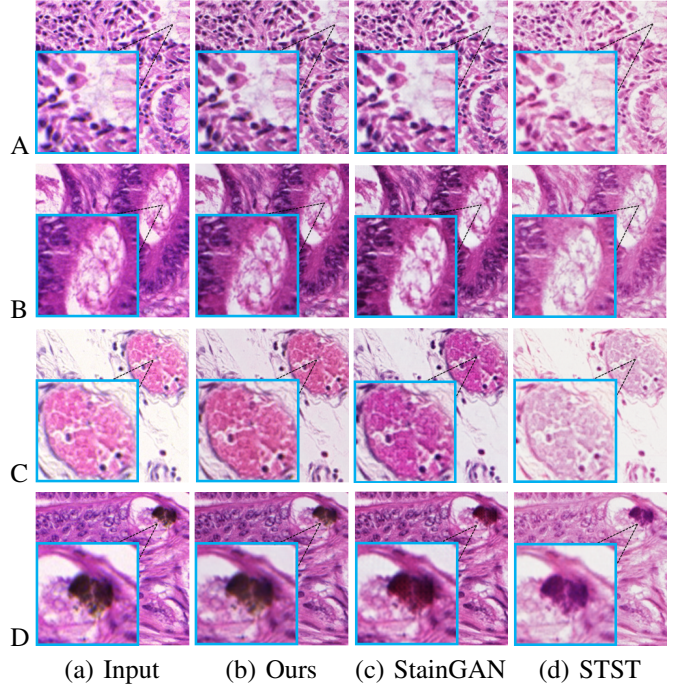| Method	Dice	ACC	Pre	Recall
(1) Ori	0.842	0.844	0.745	0.773
(2) Reinhard [25]	0.856	0.873	0.872	0.824
(3) Macenko [30]	0.861	0.876	0.878	0.833
(4) Vahadane [27]	0.856	0.872	0.874	0.840
(5) StainGAN [35]	0.872	0.880	0.859	0.852
(6) STST [36]	0.859	0.866	0.851	0.853
(7) **Ours**	**0.878**	**0.883**	**0.883**	**0.861**

(a) Input   (b) Ours   (c) StainGAN   (d) STST

Fig. 4. Comparisons of stain normalization in MICCAI'16 GlaS challenge dataset. Our proposed RestainNet successfully preserves the structural information with richer and sharper textures.

We can easily observe that all the stain normalization methods can reduce the stain inconsistency of the data, thereby help improve the segmentation task, shown in (2) to (7). (4) uses the non-negative matrix factorization method to obtain the dye concentration matrix to generate a more detailed normalized image, so obtains a better segmentation effect than (2) and (3). On the contrary, (4) also consumes a longer time in stain normalization. In this task, the segmentation

Method	F1-score	Accuracy	Precision
(1) Ori	0.815	0.914	0.818
(2) Reinhard [25]	0.903	0.930	0.906
(3) Macenko [30]	0.895	0.923	0.896
(4) StainGAN [35]	0.905	0.931	0.911
(5) STST [36]	0.879	0.912	0.886
(6) **Ours**	**0.910**	**0.935**	**0.919**

model maps pixels into differentiated representations based on the significant color and morphological differences between tissues to achieve the purpose of gland segmentation. Therefore, to ensure the efficiency of the segmentation model, the normalized image should have a broader color distribution to maintain the distinguishability of the histological tissue clusters. (6) to (7) implement the conversion between different styles of pathological images through a learning-based way, which can get normalized images with richer colors because they have learned a large number of different types of tissues in the training phase. Although the same dataset is employed, the motivations in (5) to (7) are different, resulting in a visible difference in stain style in the synthesized images. (5) uses semi-supervised to achieve the transfer between two image styles in the training set and shows satisfactory performance in the MITOS-ATYPLA14 database, but when predicting on the unfamiliar MICCAI'16 GlaS dataset, the performance weakened. (6) also employs an semi-supervised method to color the grayscale images, so it may be difficult to handle the tissue distribution that does not exist in the training data.

To further compare the performance of the learning-based normalization methods, we show normalized results of MIC-CAI'16 GlaS dataset in Fig. 4. We can find that the image generated by our proposed model in (b) has richer and sharper textures than the other two models, as shown in Fig. 4-A and Fig. 4-B. This is because RestainNet refers to the dye concentration components rather than just the texture information, which reduces the influence of the background area and forces the network to be sensitive to dyes. Thanks to the dye concentration matrix, the proposed network is also sensitive to the color of the tissue clusters. We can find that (d) shows insufficient tissue recognition ability and colors all tissues pink. Since the blood cells in Fig. 4-C are not combine with the dyes, they should show the original red color of hemoglobin, while (c) and (d) transfer it to purple. In Fig. 4-D, the contamination should maintain the original color, but (c) and (d) transfer it as the normal tissue. These all highlight the benefits of the dye concentration matrix to our model.

D. Stain Normalization in the Classification Task

In this experiment, we conduct a classification task on Zhao dataset [20] to examine the influence of various normalization algorithms on the classification task. The classification task in this experiment is more challenging than the segmentation task described in Sec. IV-C for the following two reasons. First, in the Zhao dataset, the data obtained by multiple centers

brings serious inconsistencies in stain styles, which requires the normalization algorithm to have a better generalization. Secondly, there are night types of tissues with different dye adsorption capabilities in the Zhao dataset, which poses the normalization algorithm with a great challenge to identify inconsistent stained areas. Similar to Section IV-C, we first use the normalization algorithm for preprocessing then construct the standard ResNet34 network to classify all the tissues. Except that the stain normalization method is different, the structure, the training process, and the testing process of the model are the same. The quantitative results are described in Table III.

Since the data were collected from different institutions, there are significant stain styles differences due to different scanners and pathological section production specifications, which harm the performance of the classification model. As shown in Table III, we can observe different degrees of performance improvement of the classification task with different stain normalization methods. Our proposed model brings the most significant benefit for the classification task among the five approaches, which demonstrates the effectiveness of our proposed method.

E. Model Stability

We have already demonstrated the superiority of our proposed model in the previous experiments. In this part, we want to evaluate the stability of the proposed RestainNet, to see how the input $\mathcal{H}$ and $\mathcal{E}$ really affect the network outputs. Because even $\mathcal{H}$ and $\mathcal{E}$ components extracted by the Beer-Lambert's Law have demonstrated a certain degree of tolerance in stain style inconsistency. It may still be various when using different staining protocols or scanners. As an crucial part of the network input, we discuss how $\mathcal{H}$ and $\mathcal{E}$ work in the network. The this end, we add some vibrations to $\mathcal{H}$ and $\mathcal{E}$ by manually introducing a set of multiplication coefficients $\{0.6, 0.9, 1.0, 1.1, 1.2\}$.

Fig. 5 demonstrates the results with the input $\mathcal{H}$ and $\mathcal{E}$ multiplied by coefficients. We can find that the staining style of the results from Fig. 5 (b)-(f) is actually preserved very well even with different degrees of vibrations. This observation proves that our proposed model has successfully learn the staining style of the source domain. $\mathcal{H}$ and $\mathcal{E}$ deliver the hints to decide whether a pixel should be stained by hematoxylin or eosin, not the staining style. This property is meaningful for the robustness and the stability of the model.

F. Stain Intensity Preservation

The results in Table I show the excellent performance of our model in protecting the structure of the image. Not only that, but the normalization algorithm should also preserve the dye component to maintain the optical characteristics. To this end, we compare the gray histogram of the normalized image and the input to qualitatively evaluate the ability of the algorithm to protect the dye components. As shown in Fig. 6, the first row is the RGB image, and the second and third rows are the gray histograms of $\mathcal{H}$ and $\mathcal{E}$, respectively.
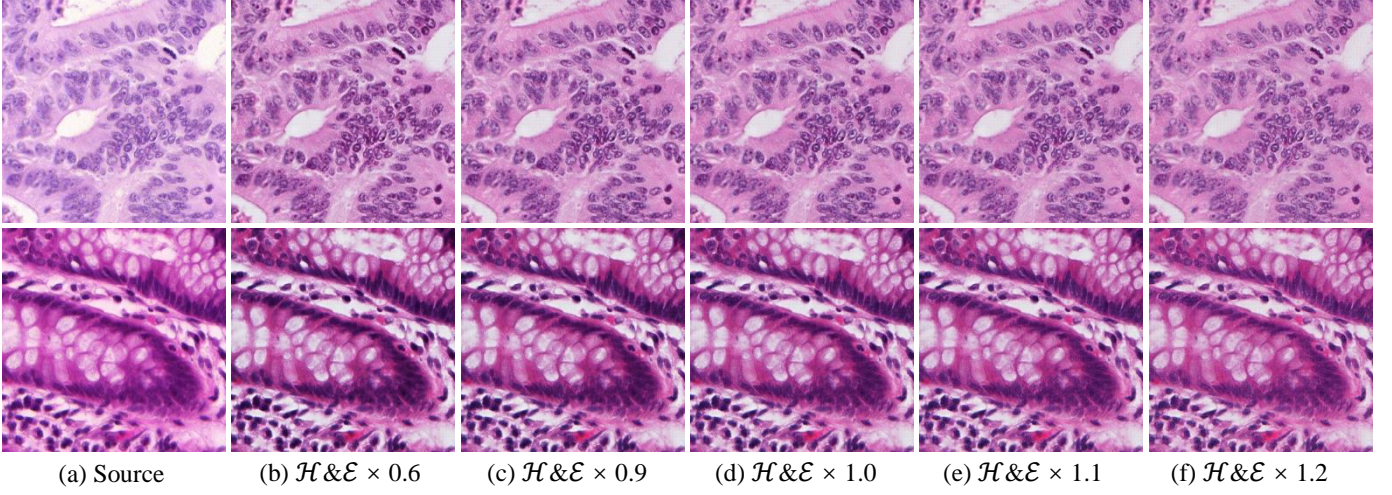
Fig. 5. (a) is the input images without stain normalization. (b) to (f) indicate the outputs of the RestainNet of with different vibration of $\mathcal{H}$ and $\mathcal{E}$ components by different multiplication coefficients 0.6, 0.9, 1.0, 1.1, and 1.2, respectively.
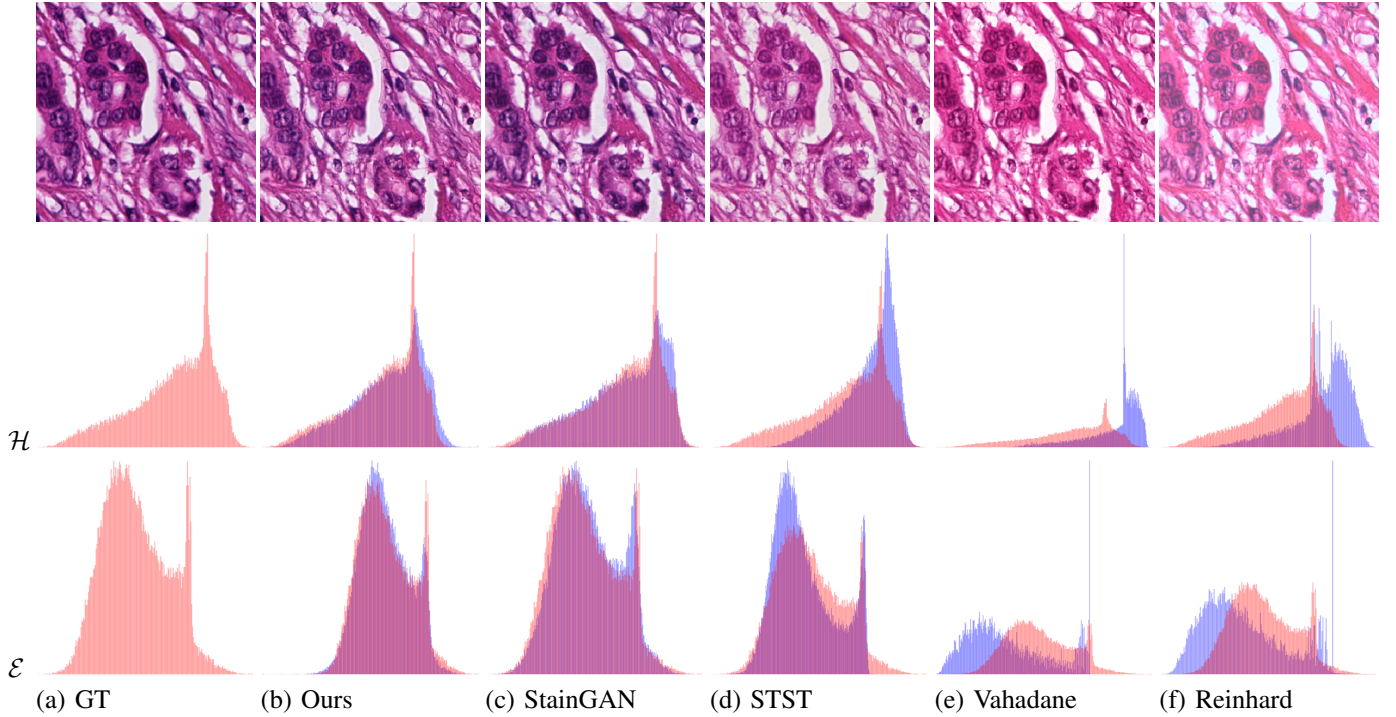

Fig. 6. Dye intensity preservation. GT are the images from the target domain (scanned by Hamamatsu scanner). $\mathcal{H}$ and $\mathcal{E}$ are the histograms of dye intensity. We overlay the histogram of the generated image (blue) on the ground truth (pink). Dye intensity is well preserved by our model.

In general, the method involved in this paper can maintain the texture structure of the image, but the stain style of the generated image is different. We observe that the computational-based normalization methods (e.g. (c) and (d)) are template dependent which can only generate the selected color style, and their grayscale histograms are the most different from (a). (d) uses the same training data as (b) to color the grayscale images, but discard the color information of the input image, so it is difficult to distinguish the tissue components, resulting in inefficient normalization. (c) consumes more computing resources to construct a more powerful generative network and synthesize a normalized image highly similar to (a). However,

(c) does not obtain prior knowledge of the digital dyes, which leads to the error coloring of the pathological tissue, harming the consistency of the histogram with (a). Thanks to the re-stain mechanism and customized staining loss, our model can synthesize the normalized image that best preserving the dye components, thereby promoting the performance of the subsequent models (e.g. segmentation or classification model), as described in the Sec IV-D and Sec IV-C.

V. CONCLUSION

In this work, we formulated stain normalization in a new perspective which regarded it as a digital staining process. To
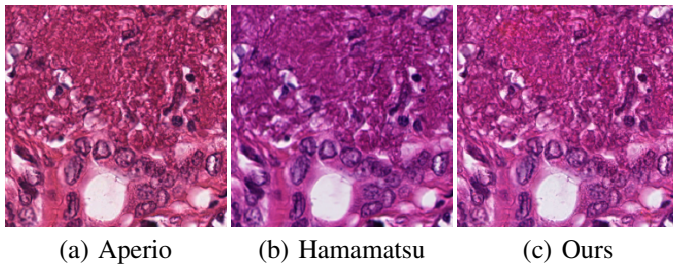
(a) Aperio   (b) Hamamatsu   (c) Ours

Fig. 7. A challenging case of the over-stained images.

achieve this goal, we present a self-supervised model Restain-Net. Besides superior quantitative and qualitative results, our model demonstrates the greatest flexibility in practical usage comparing with existing approaches. First of all, RestainNet does not require paired data for network training thanks to the self-supervised learning nature. Second, since RestainNet learns the color distribution of one specific domain (scanner) in a self-supervised manner. It allows transferring input pathology images from any other domains to the target domain without additional training cost.

However, there is still one challenge we have to tackle in future works. Since our model learns the color distribution of the specific domain. It mainly focuses on solving the color inconsistency problem introduced by different scanners. When color inconsistency is introduced by human error during sectioning, it will become a challenge of our model. Fig. 7 demonstrates an extreme case which is an over-stained section scanned by two different scanners. Our RestainNet achieves state-of-the-art performance comparing with existing models in three different tasks, the color normalization task, the segmentation task and the classification task.

REFERENCES

[1] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[2] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, vol. 58, p. 101563, 2019.
[3] K.-H. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. Ré, D. L. Rubin, and M. Snyder, "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features," *Nature communications*, vol. 7, no. 1, pp. 1–10, 2016.
[4] F. Zhang, S. Yao, Z. Li, C. Liang, and Z. Liu, "Predicting treatment response to neoadjuvant chemoradiotherapy in local advanced rectal cancer by biopsy digital pathology image features," *Clinical and Translational Medicine*, vol. 10, no. 23, 2020.
[5] B. He, L. Bergenstrhle, L. Stenbeck, A. Abid, and J. Zou, "Integrating spatial gene expression and breast tumour morphology via deep learning," *Nature Biomedical Engineering*, pp. 1–8, 2020.
[6] B. Zhao, X. Chen, Z. Li, Z. Yu, and C. Han, "Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Medical Image Analysis*, vol. 65, p. 101786, 2020.
[7] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. De Souza, A. Baidoshvili, G. Litjens, B. Van Ginneken, I. Nagtegaal, and V. D. L. Jeroen, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *IEEE International Symposium on Biomedical Imaging*, 2017, pp. 160–163.
[8] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.
[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc IEEE Asilomar Conference on Signals*, 2003.
[10] Wang, Zhou, Bovik, A. Conrad, Sheikh, H. Rahim, Simoncelli, and P. Eero, "Image quality assessment: From error visibility to structural similarity." *IEEE Transactions on Image Processing*, 2004.
[11] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
[12] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 20, no. 8, pp. 2378–2386, 2011.
[13] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution." *Anal Quat Cytol Histol*, vol. 23, 2001.
[14] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE reviews in biomedical engineering*, vol. 2, pp. 147–171, 2009.
[15] W. W. Parson, *Modern optical spectroscopy*. Springer, 2007, vol. 2.
[16] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," 2016.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.
[18] A. Janowczyk, A. Basavanhally, and A. Madabhushi, "Stain normalization using sparse autoencoders (stanosa): application to digital pathology," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 50–61, 2017.
[19] S. Roy, A. kumar Jain, S. Lal, and J. Kini, "A study about color normalization methods for histopathology images," *Micron*, vol. 114, pp. 42–61, 2018.
[20] Z. Ke, C. Zl, Y. D. Su, E. Yw, F. Xw, A. Zx, W. G. Lin, H. Yhb, B. Cl, and B. Zl, "Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer."
[21] A. BenTaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 792–802, 2017.
[22] H. Liang, K. N. Plataniotis, and X. Li, "Stain style transfer of histopathology images via structure-preserved generative learning," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2020, pp. 153–162.
[23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
[24] A. K. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.
[25] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.
[26] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
[27] 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 transactions on medical imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.
[28] A. Rabinovich, S. Agarwal, C. Laris, J. H. Price, and S. J. Belongie, "Unsupervised color decomposition of histologically stained tissue samples," in *Advances in neural information processing systems*, 2004, pp. 667–674.
[29] K. Sirinukunwattana, J. Pluim, C. Hao, X. Qi, and N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Medical image analysis*, vol. 35, pp. 489–502, 2016.
[30] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, and N. E. Thomas, "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, June 28 - July 1, 2009*, 2009.
[31] M. Gavrilovic, J. C. Azar, J. Lindblad, and C. Wahlby, "Blind color decomposition of histological images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 983–994, 2013.
[32] 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 Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1729–1738, 2014.
[33] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, and P. Quirke, "Colour normalisation in digital histopathology images," in *Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, vol. 100. Citeseer, 2009, pp. 100–111.

[34] D. Bug, S. Schneider, A. Grote, E. Oswald, F. Feuerhake, J. Schüler, and D. Merhof, "Context-based normalization of histological stains using deep convolutional features," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 135–142.

[35] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "Staingan: Stain style transfer for digital histological images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 2019.

[36] P. Salehi and A. Chalechale, "Pix2pix-based stain-to-stain translation: A solution for robust stain normalization in histopathology images analysis," *arXiv preprint arXiv:2002.00647*, 2020.

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