H&E Stain Normalization using U-Net

Chi-Chen Lee*, Po-Tsun Paul Kuo*†, Chi-Han Peng*

*National Yang Ming Chiao Tung University (NYCU), Taiwan

{ww888225433.ai10, pengchihan}@nycu.edu.tw

†Advantech Co., Ltd., Taiwan

paul.kuo@advantech.com.tw

Abstract—**We propose a novel hematoxylin and eosin (H&E) stain normalization method based on a modified U-Net neural network architecture. Unlike previous deep-learning methods that were often based on generative adversarial networks (GANs), we take a teacher-student approach and use paired datasets generated by a trained CycleGAN to train a U-Net to perform the stain normalization task. Through experiments, we compared our method to two recent competing methods, CycleGAN and StainNet, a lightweight approach also based on the teacher-student model. We found that our method is faster and can process larger images with better quality compared to CycleGAN. We also compared to StainNet and found that our method delivered quantitatively and qualitatively better results.**

I. INTRODUCTION

Histopathology is the examination of human tissues under a microscope to study the manifestation of diseases. The tissues under study come from a biopsy or surgical procedure, then are processed and cut into very thin layers, and then stained and examined by pathologists under microscopes to characterize the details of the cells in the tissue. Computer-aided diagnosis (CAD) systems to assist the analysis are also common. Conventionally, the tissue is stained with hematoxylin and eosin (H&E) [6]. However, the stain images can have strong variances (in terms of colors, illumination, image qualities, etc.) due to differences in the image acquisition processes, such as tissue fixation duration, compositions of the H&E-stains, or scanner settings [1]. The strong variances can hinder downstream tasks such as classification. In histopathology, *stain normalization* can mitigate the variances problem by transforming a stain image done by one kind of staining process to one done by another.

Deep learning (DL)-based methods for stain normalization are gaining popularity in recent years, showing fast improving performances when compared to traditional image processing-based approaches [1], [2]. In general, they conduct style-transfer tasks using neural networks such as Generative Adversarial Networks (GANs). Common choices for the network structures include: 1) cycle-consistent adversarial networks ("CycleGANs" [3] in short), 2) GANs with disentangled feature presentations (e.g., DRIT++ [4]), and 3) multi-task convolutional neural networks (CNNs) [20].

Although deep learning-based methods performed well in transferring the colors, there are still some limitations. GAN-based methods have complicated neural network structures. Therefore, they are more computationally expensive to train and inference. They also impose "patch consistency" problems

when the input image is too large and needs to be partitioned into smaller patches before they can processed by neural networks. On the contrary, StainNet [5] proposed a much simpler fully 1x1 convolutional network to transform colors in a pixel-to-pixel manner. StainNet's results are slightly less accurate than GAN-based methods (measured by GAN-related metrics such as FID (Fréchet Inception Distance)), but is much faster then CycleGAN-based approaches and has more consistent results across patches because this model learns pixel-wise property rather than the pixel distributions that were learned for CycleGAN.

To address the shortcomings of existing methods, we propose a novel stain normalization method that is based on a lightweight *U-Net* [7] neural network architecture. We trained our network on a bidirectionally paired dataset produced by CycleGAN. Quantitative results show that our network works well on color transfers between two types of H&E stains and can achieve better cycle consistency than the standard CycleGAN-based method. Besides, our output images obtain better FID scores than StainNet's results. Qualitative results show that our normalized images look as similar to real ones as CycleGAN's results. In contrast, StainNet's results have slight but notable color shifts. Same as StainNet, our results are more consistent across partitioned patches than CycleGAN's.

II. RELATED WORK

A. Image Processing-based Methods

Even simply converting all stain images to grayscale may be beneficial to certain downstream tasks in medicine [8], [9]. More modern approaches seek to retain the color information. Deconvolutional methods find a transform function between ground-truth source and target image pairs and use the function to transform unseen source images. Usually, the source and target images are first compressed into small "stain vectors" for the transform function to deduct from. These vectors are chosen manually [10] or by statistical algorithms such as Singular Value Decomposition (SVD) [11] or segmentation and clustering [12]. Template color matching algorithms [13] attempt to match distributions of each color channels between source and target images by various approaches such as variational Bayesian Gaussian mixture models (GMMs) [14] and histogram normalization of the colormaps [15].

B. Deep Learning-based Methods

Many stain normalization methods are based on generative adversarial networks (GANs). In short, GANs are suitable for conducting *image-to-image translation* [17] tasks in which images of one group (i.e., a "style") can be transformed into one of another group. Note that GAN-based style transfer is a kind of *unsupervised learning*, meaning that images in the two groups don't need to be paired. This is a major advantage over many traditional methods in which paired images are needed. StainGAN [18] trained a cycle-consistent adversarial network ("CycleGAN" [3]) to transfer H&E stain images from one scanner style to another (i.e., Hamamatsu Nanozoomer 2.0-HT to Aperio Scanscope XT). Similarly, Runz et al. [19] investigated the potential and limitations of using CycleGAN for transfering H&E stain images. [1] and [2] used GANs with disentangled feature presentations (e.g., DRIT++ [4]) to solve the stain normalization problem. Marini et al. [20] proposed a novel convolutional neural network (CNN)-based approach that aims to tackle highly heterogeneous images.

In StainNet [5], the authors proposed a lightweight fully 1x1 convolutional network to conduct color normalization in a pixel-to-pixel manner. Their method is much faster than GAN-based methods but performed slightly worse measured by GAN-related metrics such as FID (Fréchet Inception Distance) scores. They took a teacher-student model in which their neural network is trained on *paired* images synthesized by StainGAN. Our U-Net-based method is most similar to theirs.

III. METHOD

A. Dataset

The HE-Staining Variation (HEV) dataset [19], produced by the Institute of Pathology at Heidelberg University, was used to train and evaluate our model. The dataset offers serial sections of a follicular thyroid carcinoma, stained with different HE-staining protocols. We work on two styles of images. The first is done by a standard H&E stain process (of the Institute of Pathology at Heidelberg University), denoted as "HE". The second are images intentionally stained too-long, denoted as "Long HE". Note that the dataset is *unpaired* - meaning that we don't have the exact Long HE counterpart of a HE image and vice versa. Each histopathological whole-slide image (WSI) is preprocessed into tiles of 256×256 pixels. We use 10,000 tiles of each style for training CycleGAN, 10,000 tiles of each style for inferencing CycleGAN to generate the training data for StainNet and U-Net, and 5000 HE image tiles for evaluating the performance of methods.

B. Method Overview

An overview of our method is shown in Fig. 1. We first trained a CycleGAN using the HE and Long HE images in the HEV dataset. Next, inspired by the teacher-student model in deep learning, we used CycleGAN to generate synthetic (fake) Long HE images from real HE images and vice versa. In the end, we have paired HE and Long HE images to train a U-Net neural network. As shown in top of Fig. 1, we also enlarged our HE to Long HE training dataset by combining
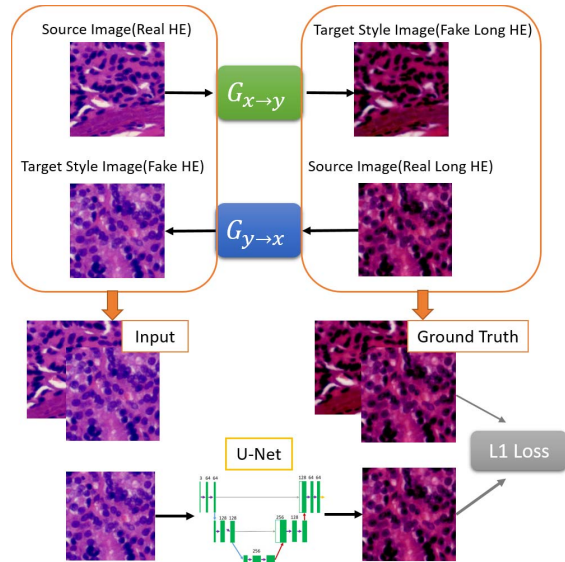

Fig. 1. Overview of our method. The $G_{x \to y}$ and $G_{y \to x}$ denote the generators of a trained CycleGAN that can synthesize a "fake" Long HE image from a real HE image and vice versa. We use the generated HE and Long HE image pairs, and the image pairs from the original HEV dataset (not shown here), to train a modified U-Net neural network that can be used to synthesize Long HE images from any HE images.

real and fake HE images as input and combining real and fake Long HE images as ground truth to improve quantitative performance of our model.

C. Neural Network Architecture

As shown in Fig. 2, our U-Net architecture is symmetric and consists of a contraction and an expansion sections. The contraction section consists of two contraction blocks. Each contraction block takes an input and passes it to two 3x3 convolution layers followed by a 2x2 max pooling. We double the number of feature channels and halve the size of feature maps after each contraction block. The expansion section also consists of two expansion blocks. Each block passes the input to two 3x3 CNN layers followed by a 2x2 up convolution layer that halves the number of feature channels. We combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the corresponding contraction block. At the final layer a 1x1 convolution layer is used to map each 64-component feature vector to a 3-channel vector.

D. Implementation

We use CycleGAN to generate the training dataset for both U-Net and StainNet. The CycleGAN model was trained using Adam optimizer for 60 epochs. For U-Net, the model was trained with SGD optimizer for 300 epochs, and we use cosine annealing scheduler to decrease the learning rate from 0.01 to 0. The L1 loss function is used to minimize the error which is the average of the all the absolute differences between the output image and the ground truth. We use a NVIDIA RTX 3060 Ti machine.
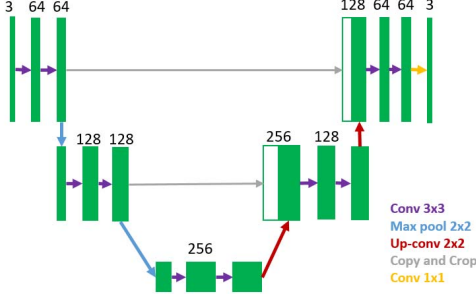
Fig. 2. Our U-Net neural network architecture.

TABLE I
Times (in seconds) to generate 5000 256x256 Long HE stain images using
CycleGAN, StainNet, and our U-Net based method.

	CycleGAN	StainNet	Our
Time(s)	271.18	123.22	205.05

IV. RESULTS AND ANALYSIS

We first compare the inference speeds of our method versus a CycleGAN-based stain normalization method and StainNet. As shown in Table I, our method is 24.39% faster than CycleGAN, but is 66.4% slower than StainNet.

To quantitatively evaluate our method, we measure the distances between distributions of our generated images and real images using Fréchet Inception Distance (FID), which is a standard way to evaluate the quality of results of generative models. In addition, to measure cycle consistency, we transfer HE stain images to the Long HE domain and transfer them back, and then calculate the similarity between each original HE image and its reconstructed one. Following the procedures in the previous papers [18], [19], we use Structural Similarity Index Measure (SSIM) and Peak Signal to Noise Ratio (PSNR) to compare images. These metrics are explained as follows:

Fréchet Inception Distance (FID) is the squared Wasserstein distance between two multidimensional Gaussian distributions, $N(\mu, \Sigma)$ and $N(\mu_w, \Sigma_w)$:

$$FID = ||\mu - \mu_w||_2^2 + tr(\Sigma + \Sigma_w - 2(\Sigma^{\frac{1}{2}} \cdot \Sigma_w \cdot \Sigma^{\frac{1}{2}})^{\frac{1}{2}}),$$

where μ and Σ, and μ_w and Σ_w, are the mean vector and covariance matrix of some neural network internal representations of the generated images and real images, respectively. tr is the trace of a matrix. The Inception v3 [21] trained on the ImageNet is commonly used as the internal representation.

Structural Similarity Index (SSIM) is used to measure the similarity between two images. It is the mean of the following score calculated over many windows, x and y, of the given image pair:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c1)(\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)},$$

where μ_x and μ_y, and σ_x^2 and σ_y^2, are the average and variance of x and y, respectively. σ_{xy} is the covariance of x and y. The range of SSIM is from 0 (least similar) to 1 (most similar).

TABLE II
FID scores between real Long HE images and generated Long HE images.

	CycleGAN	StainNet	Our
FID	16.97	24.83	18.07

TABLE III
Cycle Consistency scores between real HE images and reconstructed HE images.

	CycleGAN	StainNet	Our
SSIM	0.9554	0.9637	0.9672
PSNR	36.07	34.37	36.27

Peak Signal to Noise Ratio(PSNR) also measures the similarity of an original image versus a generated image:

$$PSNR = 20log_{10}\frac{MAX_f}{\sqrt{(MSE)}},$$

where MAX_f is the maximal value of the original image and MSE is the Mean Squared Error of the two images. Higher PSNR value mean more similarity and vice versa.

In Table II, we show the FID scores between real Long HE images and fake Long HE images generated by the three methods. As expected, our method's FID score is slightly worse (higher) than CycleGAN's because our method is trained on the data generated by CycleGAN. However, our method has significantly better FID score than StainNet (also a teacher-student model), meaning that our results are statistically more similar to ground truth. In Table III, we show cycle consistency scores of the three methods. Our method outperforms other two methods in terms of SSIM and PSNR scores, which shows that our network retains the source image information better. We illustrate qualitative comparison of the methods in Fig. 3 left. We can see that fake Long HE images generated by CycleGAN and our method are more visually similar to real ones than StainNet's results (which have some color shifts). In Fig. 3 right, we show that our method, similar to StainNet, maintains better consistency between different partitioned patches of the input image than CycleGAN.

V. CONCLUSION

In this paper, we explore the novel idea of using U-Net to perform H&E stain normalization tasks. Through experiments, we show that our method is significantly faster than a competing CycleGAN-based approach while still delivering similar results measured quantitatively and qualitatively. Compared to StainNet, our method is slower but delivers results with better qualities. For future work, we aim to improve the speed of our method while still maintaining the qualities and investigate the improvement of the downstream tasks (such as cancer cell detection) by this faster stain normalization towards larger digital pathological images. .

REFERENCES

[1] Sophia J. Wagner, Nadieh Khalili, Raghav Sharma, Melanie Boxberg, Carsten Marr, Walter de Back, and Tingying Peng. "Structure-Preserving
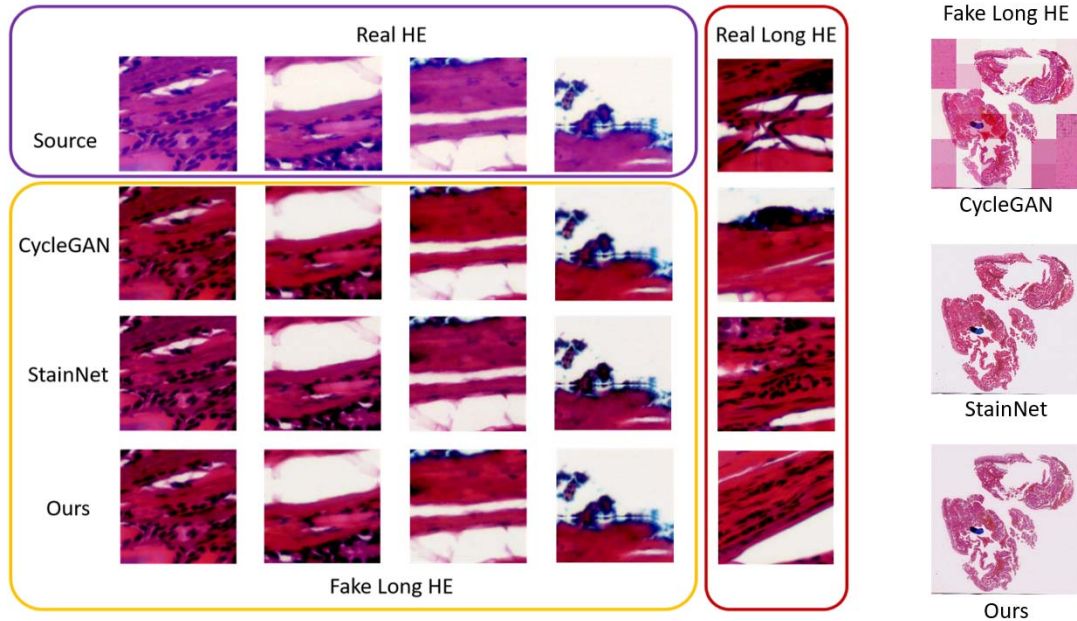
Fig. 3. Left: Qualitative results of various stain normalization methods, which transfer from HE to Long HE. The right column shows some real Long HE. Right: large inputs (3120x3168) are partitioned into smaller (512x512) patches before they can be processed by the stain normalization methods.

Multi-domain Stain Color Augmentation Using Style-Transfer with Disentangled Representations", in Medical Image Computing and Computer Assisted Intervention (MICCAI), 2021.

[2] Atefeh Ziaei Moghadam, Hamed Azarnoush, Seyyed Ali Seyyedsalehi, and Mohammad Havaei. "Stain transfer using Generative Adversarial Networks and disentangled features", in Computers in Biology and Medicine, 2022.

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.

[4] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Kumar Singh, and Ming-Hsuan Yang. "Diverse Image-to-Image Translation via Disentangled Representations", in European Conference on Computer Vision (ECCV), 2018.

[5] Hongtao Kang, Die Luo, Weihua Feng, Shaoqun Zeng, Tingwei Quan, Junbo Hu, and Xiuli Liu. "StainNet: A Fast and Robust Stain Normalization Network", in Frontiers in Medicine, 2021.

[6] John K C Chan. "The wonderful colors of the Hematoxylin-Eosin stain in diagnostic surgical pathology", in International Journal of Surgical Pathology, 2014.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation", in Medical Image Computing and Computer Assisted Intervention (MICCAI), 2015.

[8] P.W. Hamilton, P.H. Bartels, D. Thompson, N.H. Anderson, R. Montironi, and J.M. Sloan. "Automated Location of Dysplastic Fields in Colorectal Histology using Image Texture Analysis", in Journal of Pathology, 1997.

[9] Antonio Ruiz, Olcay Sertel, Manuel Ujaldon, Umit Catalyurek, Joel Saltz, and Metin Gurcan. "Pathological Image Analysis Using the GPU: Stroma Classification for Neuroblastoma", in IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2007.

[10] Arnout Ruifrok and Dennis A Johnston. "Quantification of Histochemical Staining by Color Deconvolution", in Analytical and Quantitative Cytology and Histology, 2001.

[11] Marc Macenko, Marc Niethammer, J. S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E. Thomas. "A Method for Normalizing Histology Slides for Quantitative Analysis", in IEEE International Symposium on Biomedical Imaging, 2009.

[12] Massimo Salvi, Nicola Michielli, and Filippo Molinari. "Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology", in Computer Methods and Programs in Biomedicine, 2020.

[13] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. "Color Transfer between Images", in IEEE Engineering in Medicine and Biology Magazine, 2001.

[14] Derek Magee, Darren Treanor, Doreen Crellin, Michael Shires, Katherine Smith, Kevin Mohee, and Philip Quirke. "Colour Normalisation in Digital Histopathology Images", in Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy, 2009.

[15] Sonal Kothari, John H. Phan, Richard A. Moffitt, Todd H. Stokes, Shelby E. Hassberger, Qaiser Chaudry, Andrew N. Young, adn May D. Wang. "Automatic Batch-Invariant Color Segmentation of Histological Cancer Images", in IEEE Int. Symposium on Biomedical Imaging, 2011.

[16] Andrew Janowczyk, Ajay Basavanhally, Anant Madabhushi. "Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to Digital Pathology", in CMIG, 2017.

[17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-Image Translation with Conditional Adversarial Networks", in Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[18] M. Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqoun. "Staingan: Stain Style Transfer for Digital Histological Images", in IEEE International Symposium on Biomedical Imaging (ISBI), 2017.

[19] Marlen Runz, Daniel Rusche, Stefan Schmidt, Martin R. Weihrauch, Jürgen Hesser, and Cleo-Aron Weis. "Normalization of H&E-Stained Histological Images using Cycle Consistent Generative Adversarial Networks", in Diagnostic Pathology, 2021.

[20] Niccol'o Marini, Manfredo Atzori, Sebastian Ot´alora, Stephane Marchand-Maillet, and Henning M¨uller. "H&E-Adversarial Network: a Convolutional Neural Network to Learn Stain-Invariant Features through Hematoxylin Eosin Regression", in IEEE International Conference on Computer Vision (ICCV), 2017.

[21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going Deeper with Convolutions" in Conference on Computer Vision and Pattern Recognition (CVPR), 2015.