2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Self-Attention Based Virtual Staining for Bright-field Images of Label-free Human Carotid Atherosclerotic Plaque Tissue Section*

Guanghao Zhang, Hui Hui, Bin Ning, Di Dong, Jie Tian, *Fellow, IEEE* and Wen He

***Abstract*— Histological analysis of carotid atherosclerotic plaque tissue specimens is a widely used method for studying the diagnosis of ischemic heart disease and stroke. Understanding the physiological and pathological mechanisms of carotid atherosclerotic plaque is of great significance for the effective prevention and treatment of plaque formation and rupture. In this work, we adapted a self-attention generative adversarial model to virtually stain label-free human carotid atherosclerotic plaque tissue sections into corresponding H&E stained sections. The self-attention mechanism and multi-layer structure are introduced into the residual steps of the generator and in the discriminator. Our method achieved the best performance (SSIM, PSNR, and LPIPS of 0.53, 20.29, and 0.30, respectively) in comparison with other state-of-the-art methods.**

***Clinical Relevance*—The proposed approach allows for the virtual staining of unlabeled human carotid plaque tissue images. It identifies the histopathological features of atherosclerotic plaques in the same tissue sample, which could facilitate the development of personalized prevention and other interventional treatments for carotid atherosclerosis.**

I. Introduction

Atherosclerosis, a multifactorial disease of the arterial wall, is a major precursor of ischemic heart disease and stroke. Ischemic stroke is attributed to thrombosis and cerebral ischemia and is often associated with chronic atherosclerotic plaques that accumulate in the subendothelial layer (intima) of the carotid artery [1]. Understanding the physiological and pathological mechanisms of carotid atherosclerotic plaque is crucial for the effective prevention and treatment of plaque formation and rupture. Histological analysis is adopted in investigating the cellular and molecular features of stained atheroma tissue sections [2, 3]. However, histological staining of carotid atherosclerotic plaque is a complex and laborious process. In addition, histological staining inevitably introduces human and laboratory variations that pose challenges to the accurate analysis of histopathological images [4]. The time-consuming histological staining procedures also create obstacles for fast pathological diagnosis.

Recently, efforts have been focused on virtually histological staining images via deep-learning-based methods. One study uses conditional generative adversarial networks (cGAN) to transfer unstained hyperspectral lung histology images to their corresponding hematoxylin and eosin (H&E) stained images [5]. Rivenson et al also employed the GAN model to virtually stain the autofluorescence images to H&E images [6]. Another study used cGAN to perform virtual staining of bright-field microscopic images of unlabeled rat carotid artery tissue sections to assess the condition of carotid artery intimal hyperplasia [7].

In this study, we propose a self-attention GAN model that can virtually stain label-free carotid atherosclerotic plaque tissue sections into corresponding H&E stained sections, as summarized in Fig. 1. We demonstrate the generation from an unlabeled tissue section to corresponding H&E virtual stained sections. Our virtual staining images can also observe the typical characteristics of human carotid artery plaque, such as neovascularization, necrotic core, and cholesterol crystals.

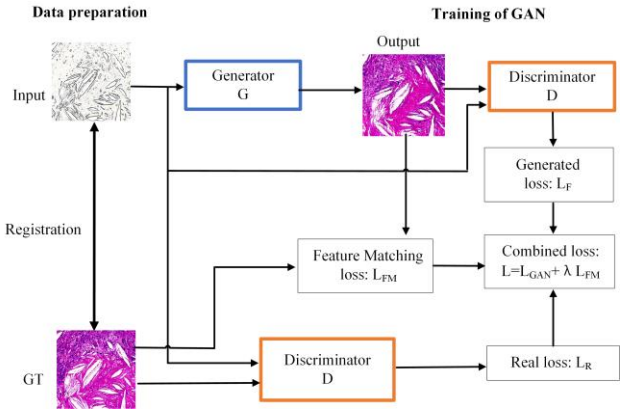

Figure 1. Schematic diagram of our virtual staining framework for unstained tissue using self-attention GAN model. The schematic diagram outlines the data preparation and training framework that can be used to generate H&E virtual stains.

II. Methods

A. Self-attention

Wang et al. [8] proposed the non-local model that could obtain the global receptive field.

*Research supported by the National Key Research and Development Program of China under Grant 2017YFA0700401, 2016YFC0103803; the National Natural Science Foundation of China under Grant 81730050, 81827808, 62027901, 81671851, 81527805; CAS Youth Innovation Promotion Association under Grant 2018167 and CAS Key Technology Talent Program. The Project of High-Level Talents Team Introduction in Zhuhai City (Zhuhai HLHPTP201703). G. Zhang, H. Hui, and B. Ning contributed equally to this work. Corresponding authors: J. Tian and W. He.

G. Zhang is with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100190, China (e-mail: zhangguanghao2019@mails.ucas.ac.cn).

B. Ning and W. He are with the Department of Ultrasound, Beijing Tiantan Hospital, Capital Medical University, Beijing, 100070, China (e-mail: nbb0104@sina.com, hewen@bjtth.org).

H. Hui and D. Dong are with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Institute of Automation, Beijing 100190, China (e-mail: hui.hui@ia.ac.cn, di.dong@ia.ac.cn).

J. Tian is with Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, 100083, China (corresponding author, phone: 0086-10-82618465; fax: 0086-10-82618465; e-mail: tian@ieee.org).

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_i) \quad (1)$$

where x is the input feature map, i represents the spatial location of the output, f is the function to calculate the similarity between i and j, and g is the function to calculate the represented feature map at j. In Fig. 2(b), the functions f(x), g(x), and h(x) represent query, key, and val mapping. B, C, W, and H are the batchsize, the number of channels, the width of the feature map, and the height of the feature map, respectively. Non-local operations directly capture remote dependencies by calculating the relationship between any two locations. The operation need not be limited to adjacent areas. It is equivalent to the construction of a convolution kernel as large as the size of the feature map to obtain more information.

B. *Self-Attention-Based Pix2pix model*

The architecture of the proposed model is shown in Fig. 2. We adapt the non-local model as a self-attention generative adversarial network to the image-to-image translation Pix2pix framework [9, 10], enabling the generator to efficiently model relationships between widely separated spatial regions. In this model, the self-attention mechanism is introduced into the residual steps of the generator, as shown in Fig. 2(a). In the discriminator, a multi-layer structure is introduced. More specifically, three layer discriminators with identical network structure opera operates at different image scales. The layer at the coarsest scale has the largest receptive field to guide the generator generating globally consistent images.

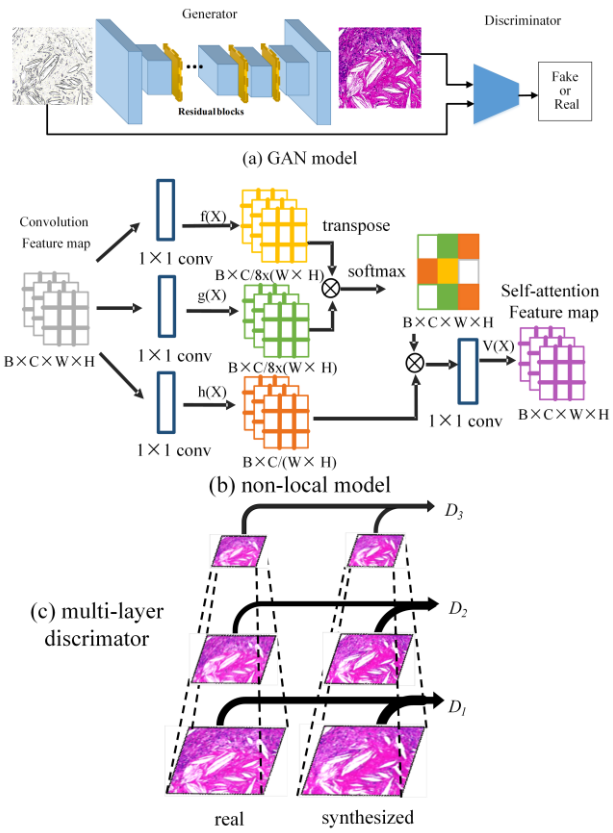

Figure 2. The architecture of our H&E virtual staining model. (a), the architecture of the non-local model. (b), schematic diagram of the location of adding the non-local model to the generator. (c), the architecture of the multi-layer discriminator.

In our non-local model, the discriminator and loss function are different with the self-attention GAN.

The loss function of GAN loss is defined as follows:

$$L_{\mathrm{GAN}}(G, D) = E[\log D(\mathrm{s}, \mathrm{x})] + E[\log(1 - D(\mathrm{s}, \mathrm{G}(\mathrm{s}))] \quad (2)$$

where s represents the input unstained section, and x represents the histologically stained section. In the proposed pix2pix model, the loss function of GAN loss is defined as the sum of the GAN loss at each discriminator layer k:

$$\min_G \max_D \sum_{k=1,2,3} L_{\mathrm{GAN}}(G, D_{\mathrm{k}}) \quad (3)$$

To extract features from multiple layers of the discriminator and match these intermediate representations from the real and the synthesized image, the feature matching loss is defined as:

$$L_{FM}(G, D_{\mathrm{k}}) = E_{(\mathrm{s},\mathrm{x})} \sum_{\mathrm{i}=1}^{\mathrm{T}} \frac{1}{N_{\mathrm{i}}} \left[||\mathrm{D}_{\mathrm{k}}^{(\mathrm{i})}(\mathrm{s}, \mathrm{x}) - \mathrm{D}_{\mathrm{k}}^{(\mathrm{i})}(\mathrm{s}, \mathrm{G}(\mathrm{x}))||_1 \right]. \quad (4)$$

where T is the total number of layers and N_{i} denotes the number of elements in i^{th} layer. The discriminators D_1, D_2, and D_3 are trained to differentiate real and synthesized images at the 3 different scales, respectively. The full objective function is the combination of the total GAN loss and feature matching loss:

$$\min_G \left((\max_D \sum_{k=1,2,3} L_{\mathrm{GAN}}(G, D_{\mathrm{k}})) + \lambda \sum_{k=1,2,3} L_{FM}(G, D_{\mathrm{k}}) \right) \quad (5)$$

where λ controls the importance of the two terms.

III. Experiments and Results

A. *Sample acquisition and Image pre-processing*

The samples were obtained from the patients with advanced carotid artery stenosis who underwent carotid endarterectomy (CEA) at Beijing Tiantan Hospital, China. The surgically resected atherosclerotic plaques were fixed with formalin overnight, and embedded in paraffin. The paraffin-embedded samples were further cut into 10μm thick slices and placed on thin glass slides for bright-field microscopic imaging with a digital slide scanner (Zeiss Axio Scan.Z1, Germany) equipped with a ×10/0.45 NA objective (0.44 μm/pixel) [11].

After histological staining process, the whole-slide tissue sections were used for co-registration. The registered whole-slice images were cropped into patches of 512 × 512 pixels. For quantitative evaluation, the training set and test set were divided by the ratio of 8:2. In total, 7039 and 2514 image pairs were obtained for the training and the test set, respectively.

B. Evaluation Metrics

The outputs of the network were measured by three metrics: structural similarity index (SSIM) [12], peak signal-to-noise ratio (PSNR), and Learned Perceptual Image Patch Similarity (LPIPS) [13]. SSIM is a reference-based quality assessment indicator that compares the pixel intensity between the reference ground truth image and the output image. Peak signal-to-noise ratio (PSNR) indicates the quality of image generation. PSNR is the most common and widely used image evaluation index. The larger the PSNR value, the less distortion, and higher image quality. PSNR and SSIM have commonly used evaluation indicators for image generation and restoration, but the two indicators pay more attention to the fidelity of the image rather than the visual quality. LPIPS introduces a new indicator of perceptual similarity judgment, which is consistent with humans. LPIPS pays more attention to whether the visual characteristics of images are similar. Therefore, the smaller the LPIPS, the closer the generated image is to ground truth. In this study, we used the trained Alexnet [14] to extract image features.

C. Virtual staining of human carotid artery plaque tissue sections

The proposed method is compared with pix2pix [10], UNIT [15] and cycleGAN [16]. Note that we trained our networks with the same datasets for a fair comparative study. In Fig. 3, these generated images showed that our model can convert the bright-field images of unstained human carotid artery plaque tissue sections to the desired H&E-stained sections. The first column in Fig. 3 show that the neovascularization is presented in H&E virtual and histological staining, respectively. The presence of neovascularization indicated that the plaque is vulnerable and highly permeable, which can further promote plaque progression and induce plaque rupture and bleeding. In the middle column, the results show that our model can infer unlabeled bright-field images of necrotic cores in H&E staining. In the last column, we can see that the staining effect of cholesterol crystals under our model is similar to ground truth.

D. Quantitative and Visual comparison with state-of-the-art methods

The quantitative evaluation for the proposed method and other state-of-the-art methods are summarized in Table I. Our method achieves the best performance on SSIM, PSNR, and LPIPS. Specifically, our method achieves the lowest LPIPS, which demonstrated that our results are much visually closer to the ground truth than other state-of-the-art methods as LPIPS is a visual quality indicator of human visual characteristics.

TABLE I. QUANTITATIVE EVALUATION FOR THE PROPOSED NETWORK AND THE STATE-OF-THE-ART METHODS

Methods	**H&E**		
	SSIM ↑	***PSNR*** ↑	***LPIPS*** ↓
UNIT	0.34±0.61	16.64±0.88	0.41±0.41
CycleGAN	0.45±0.27	17.25±0.19	0.38±0.30
Pix2pix	0.46±0.19	17.27±0.38	0.43±0.52
Ours	**0.53±0.06**	**20.29±1.06**	**0.30±0.14**

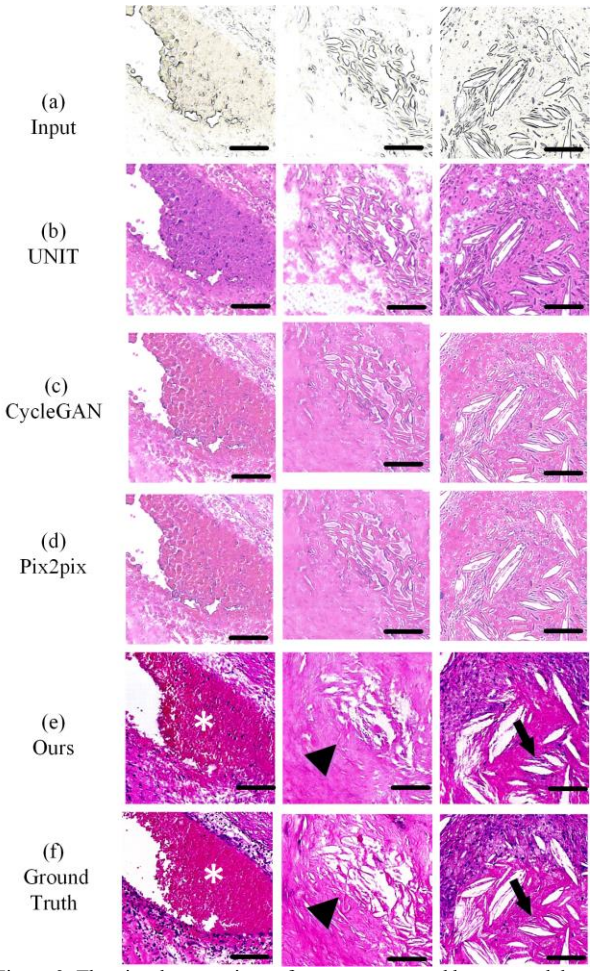

Figure 3. The visual comparison of outputs generated by our model and state-of-the-art methods. (a) The original unstained tissue section is used as the input of the neural network, (b) (c) (d) The three H&E virtual stained by the pix2pix, cycleGAN, and UNIT, (e) The outputs generated by our model, (f) The histologically stained sections. Scale bar, 100 μm. The histopathological features of atherosclerotic plaque, such as neovascularization (marked by white asterisk), necrotic nuclei (marked by black arrowhead), and cholesterol crystals (marked by black arrow), in e and f.

E. Ablation study

To study the role of the non-local model and multi-layer discriminator, ablation experiments were performed with or without a non-local model and multi-layer discriminator, detailed in Table II.

TABLE II. ABLATION STUDY WITH OR WITHOUT THE NON-LOCAL MODEL AND MULTI-LAYER DISCRIMINATOR

	H&E		
	SSIM ↑	***PSNR*** ↑	***LPIPS*** ↓
Pix2pix	0.46±0.19	17.27±0.38	0.43±0.52
Pix2pix+multi-layer discriminator	0.49±0.14	17.34±0.53	0.38±0.18
Ours	**0.53±0.06**	**20.29±1.06**	**0.30±0.14**

In the experiments, we observed that many repeated patterns often appear in the generated images without the multi-layer discriminators. From the evaluation index, our model achieves the best performance with the non-local model and the multi-layer discriminator.

F. Implementation

The network was implemented using Python and Pytorch libraries. All calculations used to train the network were performed on a system equipped with NVIDIA GeForce TITAN RTX with 24 GB memory. The parameter λ in Equation 5 was set to 10. The momentum parameters were set to 0.5, and the batch size was set to 16. The initial learning rate of the generator and the discriminator was set to 0.0002. A total of 100 epochs were trained. It took 39 hours to train the model and 104 ms to predict one image.

IV. Discussion

Histological staining analysis is performed as a "golden standard" in diagnostic pathology. It is widely used to identify carotid artery tissue constituents such as neovascularization, necrotic core, and cholesterol crystals. However, the tissue staining process is laborious, and the image quality of tissue staining images is variable. This staining process has led to a barrier in the development of standard and rapid histological image analysis systems. Therefore, virtually histological staining images by deep-learning-based methods is highly required.

In this study, we adapted a self-attention GAN model that virtually stain label-free human carotid atherosclerotic plaque tissue sections into corresponding H&E stained sections. In terms of quantitative evaluation metrics, such as SSIM, PSNR, and LPIPS, our method achieved the best performance in comparison with other state-of-the-art methods. The deep-learning-based method can provide digital staining of label-free tissue sections, which can bypass the lengthy and laborious tissue preparation process and can mitigate human-to-human variations for standard histological staining of tissue samples.

Our study has some limitations. First, the dataset for training the network is relatively small, and the generated staining components of atherosclerotic plaque are relative small. In future work, the number of human carotid atherosclerotic plaque tissue sections will be increased and multiple type of stains will be added to train the model. Furthermore, the generalization of the proposed method will be verified on the datasets from other hospitals. In addition, this virtual staining method can be enhanced by combing the unstained images acquired using other advanced label-free imaging techniques, for example, Two-photon and photoacoustic microscopy. Finally, the efficiency of model training needs to be further improved to increase clinical practicability.

V. Conclusion

In conclusion, we adapted a self-attention GAN model that virtually stain label-free carotid atherosclerotic plaque tissue sections into corresponding H&E stained sections. In comparison with other state-of-the-art methods, our method achieved the best performance for evaluation metrics of virtual staining and histological staining. This deep-learning-based method can provide digital staining of label-free tissue sections, which can bypass the lengthy and laborious tissue preparation process and can mitigate human-to-human variations for standard histological staining of tissue samples. The virtual staining method will provide strong support to the applications of histological analysis of atherosclerotic plaque.

Acknowledgment

We would acknowledge the instrumental and technical support of multimodal biomedical imaging experimental platform, Institute of Automation, Chinese Academy of Sciences.

References

[1] G. W. Stone, A. Maehara, A. J. Lansky, B. de Bruyne, E. Cristea, G. S. Mintz, et al., "A prospective natural-history study of coronary atherosclerosis," *N Engl J Med*, vol. 364, pp. 226-35, Jan 20 2011.

[2] W. Tong, H. Hui, W. Shang, Y. Zhang, F. Tian, Q. Ma, et al., "Highly sensitive magnetic particle imaging of vulnerable atherosclerotic plaque with active myeloperoxidase-targeted nanoparticles," *Theranostics*, vol. 11, pp. 506-521, 2021.

[3] W. Wang, Y. Zhang, H. Hui, W. Tong, Z. Wei, Z. Li, et al., "The effect of endothelial progenitor cell transplantation on neointimal hyperplasia and reendothelialisation after balloon catheter injury in rat carotid arteries," *Stem Cell Res Ther*, vol. 12, p. 99, Feb 3 2021.

[4] Y. Rivenson, K. de Haan, W. D. Wallace, and A. Ozcan, "Emerging Advances to Transform Histopathology Using Virtual Staining," *BME Frontiers*, vol. 2020, pp. 1-11, 2020.

[5] N. Bayramoglu, M. Kaakinen, L. Eklund, and J. Heikkila, "Towards Virtual H&E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW 2017)*, pp. 64-71, 2017.

[6] Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nat Biomed Eng*, vol. 3, pp. 466-477, Jun 2019.

[7] D. Li, H. Hui, Y. Q. Zhang, W. Tong, F. Tian, X. Yang, et al., "Deep Learning for Virtual Histological Staining of Bright-Field Microscopic Images of Unlabeled Carotid Artery Tissue," *Molecular Imaging and Biology*, vol. 22, pp. 1301-1309, Oct 2020.

[8] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local Neural Networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7794-7803, 2018.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative Adversarial Nets," *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, vol. 27, pp. 2672-2680, 2014.

[10] P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp. 5967-5976, 2017.

[11] M. Seeger, A. Karlas, D. Soliman, J. Pelisek, and V. Ntziachristos, "Multimodal optoacoustic and multiphoton microscopy of human carotid atheroma," *Photoacoustics*, vol. 4, pp. 102-111, Sep 2016.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, Apr 2004.

[13] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," *2018 IEEE/Cvf Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586-595, 2018.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the Acm*, vol. 60, pp. 84-90, Jun 2017.

[15] M. Y. Liu, T. Breuel, and J. Kautz, "Unsupervised Image-to-Image Translation Networks," *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, vol. 30, 2017.

[16] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242-2251, 2017.