

UNPAIRED H&E TO PR STAIN TRANSFER WITH SELF-SUPERVISED AUXILIARY SEGMENTATION

Yiyang Lin^{1*}, Yifeng Wang^{2*}, Yang Chen^{1*}, Zirui Zhu², Zijie Fang¹, Miaorun Lin³, Zexin Li¹, Yongbing Zhang^{2†}

¹Tsinghua University ²Harbin Institute of Technology, Shenzhen
³Beijing University of Posts and Telecommunications

ABSTRACT

Histochemical staining is a critical step in the diagnosis of cancer, where hematoxylin-eosin (H&E) stain is used most commonly in clinical practice. However, the H&E images often cannot be used for making accurate diagnoses. To this end, pathologists must perform immunohistochemical (IHC) stain, which is time-consuming and costly. In the field of computer-aided diagnosis, existing models can virtually generate IHC staining images, but they often require pixel-aligned data and annotations from pathologists, which are difficult to be obtained. To address this problem, we propose a self-supervised PR (a typical type of IHC) virtual staining model utilizing unpaired data without pathologists' annotations for the first time. Based on the observation that PR images are easy to be segmented, we introduce segmentation as the proxy task to make the virtual staining more accurate. Experimental results show that our model can generate PR images with the highest accuracy. Moreover, our model achieves the desired results on an external dataset.

Index Terms— PR virtual staining, self-supervised, auxiliary segmentation task

1. INTRODUCTION

Cancer is the second leading cause of death today, and pathological analysis plays a crucial role in cancer diagnosis, treatment, and prognosis. As the central part of the pathological analysis, histochemical staining enables different parts of the tissue to exhibit different colors. As the most common staining, hematoxylin-eosin (H&E) can stain the nucleus blue or dark-purple and stain extracellular matrix and cytoplasm pink, respectively, to facilitate the pathologists' observation of cellular tissue structures. However, H&E staining does not always provide enough contrast to distinguish normal cells from cancer cells during tissue analysis and cannot accurately analyze oncogene expression. Clinically, immunohistochemical (IHC) stain can distinguish normal cells from

oncogene-positive cells by antigen-antibody binding so that pathologists can make more accurate diagnoses. For example, during breast cancer tissue analysis, pathologists often use IHC staining to analyze the expression of genes such as ER, PR, Ki67, and HER2 to determine the type of mutation in the patient to guide the proper treatment. Specially, PR is a progesterone-related gene, which is usually used to reflect the dependence of cancer cell growth on progesterone, and is the main focus of our paper.

Compared with H&E staining, IHC staining is a more time-consuming and costly procedure with a very rigorous staining process. As a result, only about 1% of cancer patients in resource-poor countries and regions can cover the cost of IHC examination, severely hindering its use in clinical practice [1]. Therefore, developing virtual staining technologies to reduce the cost of staining for pathologists and patients will significantly advance the use of IHC staining in clinical practice.

However, from the viewpoint of computer vision, the virtual generation of IHC images from H&E images is a challenging task due to the fact that IHC images contain many features that are not included in H&E images, such as negative and positive oncogene expression.

In recent years, deep learning has been widely used in image processing. Deep neural networks can theoretically fit any mapping from the input domain to the output domain, providing an ideal solution for virtual staining techniques. With the rapid development of deep neural networks, some work has been successfully applied in virtual staining [2], and the current models are mainly divided into two categories: supervised and unsupervised models. Supervised models, such as Rivenson et al. [3] and Zhang et al. [4], use pixel-by-pixel aligned data to train staining models and achieve high-precision virtual staining. However, existing supervised models (e.g., pix2pix [5]) cannot be used in many cases due to the difficulty of obtaining pixel-by-pixel paired data in clinical practice. Unsupervised models, such as Li et al. [6] and Lo et al. [7], overcome the problem of relying on paired data in supervised models, but these unsupervised models lack effective supervised information, resulting in inaccurate stain-

[†]Corresponding Author

*Equal Contribution

ing in many cases [8]. Liu et al. [1] proposed a model to add pathologists’ annotations as additional supervised information. However, this model relies heavily on expert labeled data, which is also difficult to obtain in clinical practice.

Actually, without expert annotations, achieving accurate virtual staining of IHC based on unpaired data is a challenging task. First, no studies have addressed this problem before. Second, existing unsupervised models have difficulty in extracting the features necessary for generating IHC images from H&E images. In this paper, to address these problems, we propose a model to accurately transfer H&E images to IHC images by using a self-supervised auxiliary segmentation task to help extract the features necessary for generating IHC images. The self-supervised segmentation proxy task is based on the observation that the positive areas in IHC images are easy to be segmented. Our main contributions are as follows:

- (1) We propose a stain transfer model for the virtual generation of IHC images from H&E images, and apply it to PR virtual staining for the first time. This model needs neither paired data nor expert annotations;
- (2) We propose a self-supervised image style transfer model. With the observation of the easy segmentation property of PR images, we introduce a segmentation proxy task to improve the accuracy of style transfer. Moreover, the model can be extended to the cases where the image features are easy to extract, like the positive areas of PR images;
- (3) We enhance the model’s focus on morphological information by introducing the auxiliary segmentation task. Moreover, the gap between real H&E images and virtual H&E images can be reduced, and the features extracted in the virtual H&E images can be closer to those extracted in the real ones. Ultimately, the accuracy of IHC virtual staining can be improved.

2. METHOD

In the virtual staining of H&E to IHC, it is hard to accurately stain the positive areas directly utilizing unsupervised models, since it is challenging to find the positive areas based on the color information provided by the H&E images. Instead, the segmentation task can focus on the morphological information of H&E images and find out their positive areas, which would be beneficial for improving the accuracy of the virtual staining. In addition, since the color differences between positive areas (brown) and negative areas (blue) of IHC images are evident, we can easily obtain positive areas from IHC-stained images. The positive areas can provide accurate labeling information for the segmentation task. Motivated by this, we propose an IHC virtual staining model assisted with a self-supervised segmentation task. We design the task based on the observation that the positive areas in IHC images can be easily segmented, which provides supervised information to the model effectively and makes the generated IHC images

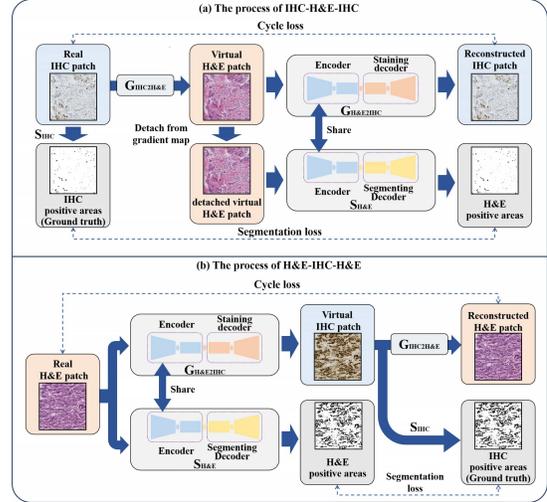


Fig. 1. The overall structure of our model. (a): The process of IHC-H&E-IHC; (b): The process of H&E-IHC-H&E.

more accurate.

Similar to the traditional unsupervised generative adversarial network [9], our model employs two generators, $G_{IHC2H\&E}$ and $G_{H\&E2IHC}$. $G_{IHC2H\&E}$ generates H&E images based on the IHC images, and $G_{H\&E2IHC}$ generates IHC images based on the H&E stained images. Then, we also introduce an auxiliary segmenter $S_{H\&E}$, which segments out the positive areas of the H&E images, serving as a proxy task to construct self-supervised constraints. In addition, we use a threshold segmenter S_{IHC} to segment out the positive areas of IHC, which serve as ground truth to constrain $S_{H\&E}$.

$G_{IHC2H\&E}$ consists of an encoder and a staining decoder. The encoder is used to extract the features of the input images, and the staining decoder generates the virtual staining results based on the extracted features. $G_{H\&E2IHC}$ also consists of an encoder and a staining decoder, with the similar goal to that of $G_{IHC2H\&E}$. Moreover, $S_{H\&E}$ consists of an encoder and a segmenting decoder, where the function of the encoder is similar to that of $G_{H\&E2IHC}$, and the segmenting decoder generates the positive areas corresponding to the H&E images based on the extracted features. Since both of $G_{H\&E2IHC}$ and $S_{H\&E}$ need to focus on the positive areas of H&E images, they share the encoder. In addition, S_{IHC} can segment the positive areas of IHC images based on their Red channel values using a predefined threshold (In this paper, we set the threshold to be 120). Overall, as shown in Fig.1, our model consists of IHC-H&E-IHC and H&E-IHC-H&E processes, which will be described as follows.

2.1. The IHC-H&E-IHC and H&E-IHC-H&E processes

Similar to UGATIT, as shown in Fig.1(a), in each training iteration, we first feed a real IHC image into $G_{IHC2H\&E}$ to generate a virtual H&E image, and a reconstructed IHC image is

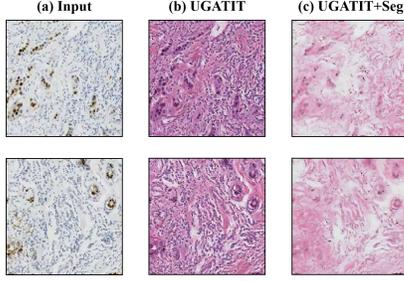


Fig. 2. Real IHC images and virtual H&E results under different conditions.

generated after feeding the virtual H&E image to $G_{H\&E2IHC}$. By employing cycle loss, the reconstructed IHC image is constrained to be consistent with the real IHC image. In the process of generating the reconstructed IHC image, we simultaneously input the virtual H&E image to $S_{H\&E}$ and obtain the positive areas of the virtual H&E image. Moreover, the label obtained by inputting the real IHC image to S_{IHC} is used to constrain $S_{H\&E}$ by using segmentation loss. In addition, as shown in Fig.1(a), we detach $S_{H\&E}$ from the current gradient map so that the segmentation task does not affect the generation of the virtual H&E image. Thus, the virtual H&E image can be much closer to the real H&E image, as the detailed description in the following sections.

Similar to Fig.1(a), in the process of generating the reconstructed H&E image, we input the real H&E image into $S_{H\&E}$ and use the label obtained by inputting the virtual IHC image into S_{IHC} as ground truth to constrain $S_{H\&E}$, as shown in Fig.1(b). The total loss formulation of the generator and the discriminator are as follows (We retain most of the loss functions of UGATIT):

$$L_G = \lambda_1 \times l_{adv}^G + \lambda_2 \times l_{cam}^G + \lambda_3 \times l_{cyc} + \lambda_4 \times l_{idt} + \lambda_5 \times l_{seg}. \quad (1)$$

$$L_D = \lambda_1 \times l_{adv}^D + \lambda_2 \times l_{cam}^D. \quad (2)$$

2.2. The impact of segmentation on virtual H&E images

In the process of IHC-H&E-IHC, since the virtual H&E image is the intermediate result between the input IHC image and the reconstructed IHC image, there exist differences between such virtual H&E image and the real H&E image. Such phenomena can be verified by the following observations.

As shown in Fig.2(a)&(b), the model without the auxiliary segmenter $S_{H\&E}$ (UGATIT) maps the brown areas in IHC to dark purple areas in H&E, and maps the blue areas in IHC to light purple areas in H&E, which constitutes many color differences that exist rarely in real H&E images. The color differences show that UGATIT mainly learns the color mapping, and the real mapping of H&E images and IHC images is not obtained. Moreover, this phenomenon reflects that the

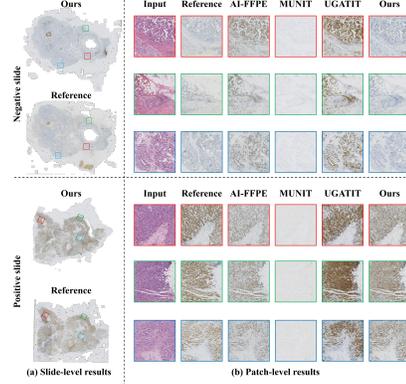


Fig. 3. The slide-level and patch-level results of our model and competing models.

process of IHC-H&E-IHC mainly focuses on the color information of the pathological images, and thus cannot achieve accurate staining for the real H&E images. Compared with Fig.2(b), the color differences between the negative and positive areas in Fig.2(c) are much smaller, reflecting that the pathology image segmentation task pays more attention to the morphological features.

However, observing from the images in Fig.2(c), directly adding the auxiliary segmenter into the model brings another problem: The color of virtual H&E images is not close to the real ones, where the virtual ones are mostly pink but the real ones are mostly purple. This is because the segmenter constrained by the supervised information focuses more on morphological features and less on color features. Therefore, the quality of the virtual H&E images is not ideal in this case. To solve this problem, as shown in Fig.1(a), we detach $S_{H\&E}$ from the current gradient map. And since $S_{H\&E}$ and $G_{H\&E2IHC}$ share the encoder which can extract both the color features and morphological features well, the virtual H&E images are much closer to the real H&E images.

3. EXPERIMENTS

We have evaluated our proposed model over a breast cancer dataset. The results show our model can transfer H&E images into PR (a typical type of IHC) images with unpaired patches efficiently, indicating that the self-supervised auxiliary segmentation task improves the accuracy greatly.

In the breast cancer dataset (containing 33 pairs of consecutive slides, 28 for training and 5 for testing), the slides are essentially similar in tissue morphological structure but are not pixel-level aligned. Before being input to the model, each image is preprocessed by two data augmentation strategies: random horizontal flip and random vertical flip.

Virtual PR staining marks of different methods

		A	B	C	D			A	B	C	D			A	B	C	D			A	B	C	D	
Reference marks	A	6	0	0	0	A	6	0	0	0	A	6	0	0	0	A	6	0	0	0	0	0	0	
	B	0	0	11	0	B	11	0	0	0	B	0	1	3	7	B	0	6	5	0	D	0	0	0
	C	0	0	5	0	C	5	0	0	0	C	0	0	0	5	C	0	0	4	1	D	0	0	1
	D	0	0	10	3	D	10	0	3	0	D	0	0	3	10	D	0	0	1	12				

(a) AI-FFPE (b) MUNIT (c) UGATIT (d) Ours

Fig. 4. Confusion matrices for the judgment results of Fiji IHC Toolbox. A-paracancerous tissue area, B-negative area, C-weakly positive area, D-positive area.

3.1. Comparison Results

Figure 3 exhibits the virtual generation of PR stained images from H&E stained images. In Fig.3(a), the slide-level results of a positive slide and a negative slide are provided. It can be seen that compared with consecutive layers (considered as reference), our model achieves high accuracy. In Fig.3(b), the patch-level results of some key parts of slides are provided. It can be observed that the traditional unsupervised image generation models only match the color of H&E and PR domains. For example, UGATIT [9] matches the brown areas of PR domain with the dark purple areas of H&E domain, and matches the blue areas of PR domain with the light purple areas of H&E domain together. This phenomenon is correct in positive images but incorrect in negative images. AI-FFPE [10] has the same problem as UGATIT. Moreover, MUNIT [11] maps all the tissue areas into paracancerous tissue areas (the color of paracancerous tissue areas is usually gray). These phenomena are because traditional unsupervised models focus relatively on the color features of pathological images and have insufficient ability to extract morphological features. In contrast, our model can stain all the positive areas brown and stain all the negative areas blue, which keeps strong consistency with reference.

In addition, we invite pathologists to split each slide into several ROIs (for each H&E slide, the ROIs corresponding to IHC slides generated by different models and the reference are approximately the same) and employ Fiji IHC Toolbox to evaluate the results generated by all the models. The confusion matrices of the judged results are shown in Fig.4. It can be seen that our model achieves the highest accuracy compared to all the unsupervised competing models.

Moreover, the evaluation metric Contrast-Structure Similarity (CSS) can express how much morphological information is preserved from the original images, so we use CSS to measure the models' ability to extract content. Table 1 shows that our model achieves the best among these models, as the proposed self-supervised auxiliary segmenter can improve our model's focus on morphological information.

Table 1. The CSS of different models (higher is better).

Models	AI-FFPE	MUNIT	UGATIT	Ours
CSS	0.629±0.169	0.384±0.207	0.613±0.174	0.672±0.158

Virtual PR staining marks of different methods

		A	B	C	D			A	B	C	D			A	B	C	D			
Reference marks	A	6	0	0	0	A	6	0	0	0	A	6	0	0	0	B	0	6	5	0
	B	0	1	3	7	B	4	3	4	0	C	0	4	1	0	C	0	0	4	1
	C	0	0	0	5	C	0	4	1	0	D	1	2	0	10	D	0	0	1	12
	D	0	0	3	10	D	1	2	0	10										

(a) UGATIT (b) UGATIT+Seg (c) UGATIT+Seg+detach

Fig. 5. Confusion matrices for the judgment results of Fiji IHC Toolbox. A-paracancerous tissue area, B-negative area, C-weakly positive area, D-positive area.

3.2. Ablation Study

Due to the fact that the UGATIT model mainly focuses on the color features and does not fully utilize the morphological features, the model mainly learns the color mapping between H&E and IHC while ignoring the morphological mapping, and thus cannot achieve IHC virtual staining accurately. In contrast, by utilizing the self-supervised constraints brought by $S_{H\&E}$, our model is able to learn not only color information but also morphological information, achieving better staining transfer from H&E images to IHC ones.

We separate $S_{H\&E}$ from the current gradient map to ensure the virtual H&E images are much closer to the real H&E images. Therefore, $S_{H\&E}$ obtains images that are much closer to reality and enables a stronger ability of extracting features from real H&E images. Moreover, since $G_{H\&E2IHC}$ and $S_{H\&E}$ shares an encoder, the results of IHC virtual staining can be more accurate. As shown in Fig.5(c), we can see by introducing and detaching $S_{H\&E}$, the accuracy of IHC generation improves a lot.

4. CONCLUSIONS

In this paper, we propose an H&E to PR stain transfer model with self-supervised auxiliary segmentation, which can be trained with unpaired patches. The ground truth of the self-supervised auxiliary segmentation task is obtained by employing the property that PR images are easily segmented by the thresholding method. We introduce a detaching operation in the training process to ensure the high quality of virtual H&E required for generating PR results accurately. Experiments on PR images show that our model is superior to traditional unsupervised models, and our approach can also achieve pleasing results on the external dataset.

Using the model we devise, in clinical practice, pathologists can save a lot of time, and patients can save much money. Meanwhile, in scientific studies, our model can provide a huge amount of data for researchers. In the future, we hope to study more elegant self-supervised constraints in histopathology image style transfer and extend our model to broader clinical applications.

5. REFERENCES

- [1] Shuting Liu, Baochang Zhang, Yiqing Liu, Anjia Han, Huijuan Shi, Tian Guan, and Yonghong He, “Unpaired stain transfer using pathology-consistent constrained generative adversarial networks,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 1977–1989, 2021.
- [2] Maximilian E Tschuchnig, Gertie J Oostingh, and Michael Gadermayr, “Generative adversarial networks in digital pathology: a survey on trends and future potential,” *Patterns*, vol. 1, no. 6, pp. 100089, 2020.
- [3] Yair Rivenson, Hongda Wang, Zhensong Wei, Kevin de Haan, Yibo Zhang, Yichen Wu, Harun Günaydin, Jonathan E Zuckerman, Thomas Chong, Anthony E Sisk, et al., “Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning,” *Nature biomedical engineering*, vol. 3, no. 6, pp. 466–477, 2019.
- [4] Yijie Zhang, Kevin de Haan, Yair Rivenson, Jingxi Li, Apostolos Delis, and Aydogan Ozcan, “Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue,” *Light: Science & Applications*, vol. 9, no. 1, pp. 1–13, 2020.
- [5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [6] Xinyang Li, Guoxun Zhang, Hui Qiao, Feng Bao, Yue Deng, Jiamin Wu, Yangfan He, Jingping Yun, Xing Lin, Hao Xie, et al., “Unsupervised content-preserving transformation for optical microscopy,” *Light: Science & Applications*, vol. 10, no. 1, pp. 1–11, 2021.
- [7] Ying-Chih Lo, I-Fang Chung, Shin-Ning Guo, Mei-Chin Wen, and Chia-Feng Juang, “Cycle-consistent gan-based stain translation of renal pathology images with glomerulus detection application,” *Applied Soft Computing*, vol. 98, pp. 106822, 2021.
- [8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [9] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee, “U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation,” *arXiv preprint arXiv:1907.10830*, 2019.
- [10] Kutsev Bengisu Ozyoruk, Sermet Can, Guliz Irem Gokceler, Kayhan Basak, Derya Demir, Gurdeniz Serin, Uguray Payam Hacisalihoglu, Berkan Darbaz, Ming Y Lu, Tiffany Y Chen, et al., “Deep learning-based frozen section to ffpe translation,” *arXiv preprint arXiv:2107.11786*, 2021.
- [11] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz, “Multimodal unsupervised image-to-image translation,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.