The Generation of Virtual Immunohistochemical Staining Images Based on an Improved Cycle-GAN

Shuting Liu[1], Xi Li[2], Aiping Zheng[3], Fan Yang[2], Yiqing Liu[1], Tian Guan[1(✉)], and Yonghong He[1(✉)]

[1] Graduate School at Shenzhen, Tsinghua University, Beijing, China
{guantian, heyh}@sz.tsinghua.edu.cn
[2] Gastroenterology Department, Peking University Shenzhen Hospital, Shenzhen, China
[3] Pathology Department, Peking University Shenzhen Hospital, Shenzhen, China

Abstract. Pathological examination is the gold standard for the diagnosis of cancer. In general, common pathological examinations include hematoxylin-eosin (H&E) staining and immunohistochemistry. H&E staining examination has the advantages of short dyeing duration and low cost, which is the most common one in the clinical practice. However, in some cases, the pathologist is hard to conduct an accurate diagnosis of cancer only according to the H&E staining images. Whereas, the immunohistochemistry examination can further provide enough evidence for the diagnosis process. Hence, the generation of virtual Ki-67 staining sections from H&E staining sections by computer assisted technology will be a good creative solution. Currently, this is still a challenge due to the lack of pixel-level paired data. In this paper, we propose a new method based on Cycle-GAN to generate Ki-67 staining images from the available H&E images, and our method is validated on a neuroendocrine tumor dataset. Massive experiment results show that the addition of skip connection and structural consistency constraint can further improve the performance of Cycle-GAN in unpaired pathological image-to-image transfer tasks. The quantification evaluation demonstrates that our proposed method achieves the state of art and reveals significant potential in clinical virtual staining.

Keywords: Virtual staining · Immunohistochemistry · Cycle-GAN

1 Introduction

As the gold standard for pathological diagnosis, pathological slices have significant application value in clinical and scientific research. The pathologists perform the pathological diagnosis and evaluation through microscopic examination of the pathological section, which is time-consuming and laborious. The digitization of pathological slices is considered to be an important turning point in the development of pathology [1]. The stained slices can be obtained with a full-scale digital image (WSI) through a professional scanner, making the transmission and storage of pathological slices safer and more convenient.

M. Guan and Z. Na (Eds.): MLICOM 2020, LNICST 342, pp. 137–147, 2021.
https://doi.org/10.1007/978-3-030-66785-6_16

Hematoxylin-eosin staining (H&E) is a common pathological examination and is widely used in clinic [2]. It has the advantages of high efficiency and low cost. The cells and tissues components can be clearly identified by the pathologists, and the cancer can be initially diagnosed and graded. However, there still need enough contrast to differentiate some low-grade cancer areas, which can be diagnosed with immunohistochemistry staining (IHC). IHC is a molecular-level staining [3]. It uses the principle of antigen-antibody binding, and the chemical reaction can bind the chromogen with labeled antibody to intracellular antigen. Meanwhile, the Ki-67 protein is a cellular marker and can be used in IHC, which is a proliferating cell-associated antigen, and can be used for qualitative and quantitative studies of cancer based on its staining results. However, Ki-67 IHC examination takes a long time and has a high cost. Hence, the generation of virtual Ki-67 staining sections from H&E staining sections by computer assisted technology will be a good creative solution.

In recent years, the deep convolution neural networks (DCNN) have received more and more attention in many aspects of medical image analysis, such as X-ray [4], CT [5], PET [6] and MRI [7, 8], which makes better use of contextual information and extracts powerful high-level features. Furthermore, it is also suitable for the analysis and mining of pathological data with large data volume. At present, the application of deep learning in the analysis of pathological slices mainly focuses on three aspects: the segmentation and detection of cells or tissues, the classification of tissue levels, and the classification of cancer grades. However, the most current DCNNs belong to supervised learning, which requires a lot of annotation information. Therefore, artificial intelligence pathological slices analysis has not been applied in clinical practice, and further research is needed.

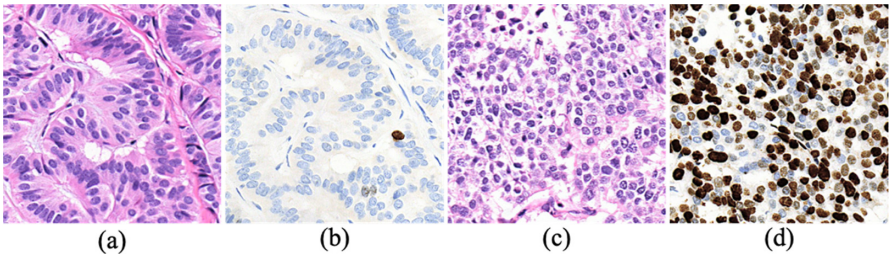

Fig. 1. Examples of H&E and Ki-67 staining image. (a) and (c) are the H&E staining patches; (b) and (d) are the corresponding Ki-67 staining patches.

Generating virtual Ki-67 staining images from H&E staining images is a challenging task due to three reasons: first, most of the DCNN-based image synthesis methods require a large number of registered image pairs, like Pix2pix, which is infeasible to obtain paired H&E/Ki-67 staining images in clinical practice; second, the image appearance between two different image modalities can be significantly different. For instance, hematoxylin principally stains cell nuclei blue or dark-purple, and eosin stains the extracellular matrix and cytoplasm pink, however, in Ki-67 staining image, Ki-67-positive tumor cells will be stained to be brown, and the Ki-67-negative ones

will be colored blue, as shown in Fig. 1; third, the field of view between two different modalities can be different, and some voxels in one modality might not have correspondences in the other modality. At present, most of the available rough paired pathological images stained from two adjacent slices in the same part. In addition, currently there are few rigidly paired HE and IHC images, because the processes of de-staining and re-staining will destroy the original tissue structure of the slice.

In this paper, we explored the potential of deep learning in unpaired image-to-image transformation in the field of histopathological analysis. We propose a new method to generate virtual Ki-67 staining images from H&E staining images. First, we employ the combination of structural similarity constraint and mean squared error constraint as cycle loss to improve the synthesis quality of virtual staining images. Second, the skip connections are added between the encoder and decoder of generators, which provide more texture information under different resolutions for the stain transformation process. In addition, the experimental results demonstrate that our method can achieve better virtual Ki-67 staining image synthesis results both qualitatively and quantitatively compared with Cycle-GAN.

2 Related Work

The generation of virtual Ki-67 staining images is similar to some natural image related tasks like style transfer, image synthesis, image super-resolution. For these tasks, generative adversarial networks (GANs) [9], has gained more and more attention from the researchers. The GAN contains two sub-networks, one is generator and the other is discriminator. Under the specific constraint, the generator is employed to generate fake similar data which then is used to fool the discriminator; the discriminator is trained by both real data and fake similar data, which is employed to differentiate the fake similar data from the real data. The generator and the discriminator compete against each other. Once the discriminator cannot distinguish the authenticity of the image, it means that the generator has learned to model the distribution of the input data appropriately. Not only natural image processing, but GAN is also a pretty hopeful approach for pathological image analysis.

At present, the common GANs include conditional generation adversarial Network (CGAN), Pix2pix and cyclic adversarial network (Cycle-GAN), etc. CGAN is based on the GAN network and adds conditional constraints to the generator and discriminator. The condition can be labels or other modal information [10]. The Pix2pix network can be used to translate pixel-level paired images, which can convert the image's expression while ensuring semantics [11]. The Cycle-GAN can be used to do style transfer between unpaired images, such as horses to zebras, and apples to oranges [12].

In general, most available clinical data is lack of the semantic annotations, and different modality data are not rigidly registered. Hence, Cycle-GAN shows significant potential in clinical pathological image analysis. Currently, stain normalization is the most common uses of Cycle-GAN in pathological image processing. Stain normalization can be also regarded as a basic simple form of staining transfer, which mainly focuses on reducing the staining variance among slides caused by different staining protocols and scanners.

3 Method

Generally, the design idea of GAN is one-way from the perspective of information flow. In the absence of unpaired images, GAN may cause different H&E staining images to be mapped to the same KI-67 staining image. Compared with GAN, Cycle-GAN is one of the state-of-the-art unpaired image synthesis algorithms, which introduces two cycle consistency losses. It contains two generators $G_{HE\text{-}to\text{-}Ki67}$ and $G_{Ki67\text{-}to\text{-}HE}$ and two associated adversarial discriminators D_{HE} and D_{Ki67}.

3.1 Overview

Based on the Cycle-GAN basic concept, our framework, as depicted in Fig. 2, transfer the H&E staining domain to the Ki-67 staining domain, without the need of paired images from both domains. It consists of two generator and discriminator pairs: the first pair ($G_{HE\text{-}to\text{-}Ki67}$ and D_{Ki67}) and the second pair ($G_{Ki67\text{-}to\text{-}HE}$ and D_{HE}). The generator $G_{HE\text{-}to\text{-}Ki67}$ tries to map images from domain H&E to domain Ki-67,

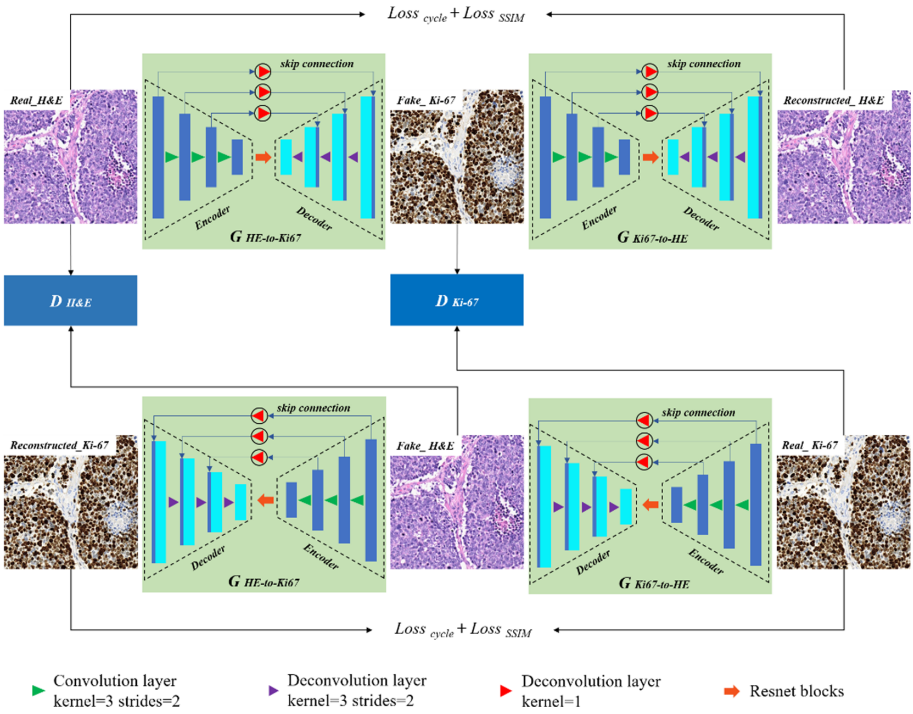

Fig. 2. An overview of our framework. Real H&E gets fake Ki-67 through $G_{HE\text{-}to\text{-}Ki67}$, and then gets reconstructed H&E through $G_{Ki67\text{-}to\text{-}HE}$. Similarly, real Ki-67 gets fake H&E through $G_{Ki67\text{-}to\text{-}HE}$, and then gets reconstructed Ki-67 through $G_{HE\text{-}to\text{-}Ki67}$. There are two discriminators, among which $D_{H\&E}$ can discriminate the authenticity of H&E; D_{Ki67} can discriminate the authenticity of Ki-67.

$G_{HE\text{-}to\text{-}Ki67}$: $S_{HE} \rightarrow S_{Ki67}$; while the generator $G_{Ki67\text{-}to\text{-}HE}$ tries to map images from domain Ki-67 to domain H&E, $G_{Ki67\text{-}to\text{-}HE}$: $S_{Ki67} \rightarrow S_{HE}$. The discriminator D_{Ki67} tries to verify if images come from the real domain Ki-67 or the fake generated ones. Similarly, the discriminator D_{HE} tries to verify if images come from the real domain H&E or the fake generated ones.

In the encoder part of each generator, the staining images with the size of 288 × 288 × 3 is cropped from whole slice image as input. The encoder starts with a convolution with a kernel size of 7 × 7 and stride of 1. In order to maintain the spatial continuity of features, there does not exist any pooling operation, which is instead by the convolution with a stride of 2. It is worth to note that each convolution layer, as shown in Fig. 2, is a series of operations, i.e. convolution with a kernel of 3 × 3, instance normalization, and 'Leaky ReLU' activation layer. Thus, the input image is down-sampled from 288 × 288 to 36 × 36 after three convolutions with the stride of 2. The following part architecture is feature extractor which consists of ten residual convolution blocks, as shown by red arrow in Fig. 2. In the decoder part, it aims to recover the resolution of the feature map from 36 × 36 to 288 × 288, deconvolution with the kernel of 3 × 3 and strider of 2 is adopted to up-sample the feature map. Finally, the virtual staining image is obtained by a 7 × 7 convolution and a 'tanh' activation function. In addition, the skip connections are added between the encoder and decoder at the same resolution, thus more low-level features are integrated into the finally recovered feature map, and the fusion of features of different scales is also achieved in this process.

3.2 Loss Function

The overall optimization function used to train the designed framework includes a combination of adversarial loss, cycle consistency loss and structural cycle consistency loss based on the structural similarity index (SSIM), which is as followed:

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda \mathcal{L}_{cycle} + \beta \mathcal{L}_{ssim}$$

where $\mathcal{L}_{adv}$ is the adversarial loss; $\mathcal{L}_{cycle}$ is the cycle consistency loss, and λ is a regularization parameter for consistency loss; $\mathcal{L}_{ssim}$ is the structural cycle consistency loss, and β is the associated regularization parameter.

Adversarial Loss, is employed to match the distribution of the generated images to that of the target domain, and match the distribution of the generated target domain back to the source domain as $\mathcal{L}_{\text{adv}} = \mathcal{L}_{\text{GAN}}^{\text{H\&}E} + \mathcal{L}_{\text{GAN}}^{\text{Ki67}}$, where $\mathcal{L}_{\text{GAN}}^{\text{H\&}E}$ and $\mathcal{L}_{\text{GAN}}^{\text{Ki67}}$ are defined as

$$\begin{aligned}\mathcal{L}_{GAN}^{H\&E} = {} & \mathbb{E}_{S_{H\&E} \sim p(S_{H\&E})}[\log D_{H\&E}(S_{H\&E})] \\ & + \mathbb{E}_{S_{Ki67} \sim p(S_{Ki67})}[\log (1 - D_{H\&E}(G_{Ki67-to-HE}(S_{Ki67})))]\end{aligned}$$

and,

$$\mathcal{L}_{GAN}^{Ki67} = \mathbb{E}_{S_{Ki67} \sim p(S_{Ki67})}[\log D_{Ki67}(S_{Ki67})] + \mathbb{E}_{S_{H\&E} \sim p(S_{H\&E})}[\log (1 - D_{Ki67}(G_{HE-to-Ki67}(S_{H\&E})))]$$

Cycle Consistency Loss. To alleviate the lack of ground truth images for the fake images generated in a particular domain, the image is mapped back to its source domain using the reverse mapping function. This loss component ensures that the reconstructed images preserve similar structure as in the source domain. In addition, this loss goes in both directions forward and backward cycles to assure stability, which is given as

$$\mathcal{L}_{cycle} = \mathbb{E}_{S_{Ki67} \sim p(S_{Ki67})}\left[\left\|G_{HE-to-Ki67}(G_{Ki67-to-HE}(S_{Ki67})) - S_{Ki67}\right\|_1\right] + \mathbb{E}_{S_{H\&E} \sim p(S_{H\&E})}\left[\left\|G_{Ki67-to-HE}(G_{HE-to-Ki67}(S_{H\&E})) - S_{H\&E}\right\|_1\right]$$

Structural Cycle Consistency Loss. In some cases, the reconstructed images are likely to have a distinct color distribution than any of the sub-domains. Therefore, minimizing the L1 distance between the source and the reconstructed images alone is not an effective way to ensure cycle consistency. We introduce the structural cycle consistency loss to our model to regulate the structural changes between the input and output images. This loss is calculated based on the SSIM which has been used for assessing the image quality in many related studies [13]. The SSIM is defined as followed,

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}$$

where μ_x, μ_y are the mean of a fixed window ($N \times N$) centered as the pixel, σ_x, σ_y are the standard derivations, σ_{xy} is the covariance. c_1 and c_2 are stabilizing factors that prevent the denominator from being zero. Hence, the structural cycle consistency loss can be formulated as:

$$\mathcal{L}_{ssim} = 2 - SSIM(G_{HE-to-Ki67}(G_{Ki67-to-HE}(S_{Ki67})), S_{Ki67}) - SSIM(G_{Ki67-to-HE}(G_{HE-to-Ki67}(S_{H\&E})), S_{H\&E})$$

4 Dataset and Implementation

4.1 Dataset

In this work, our model is validated on a neuroendocrine tumor dataset which is collected from Peking University Shenzhen Hospital. During data collection, two slices were consecutively cut at the same site of each patient for pathological examination,

one of which was stained with H&E dyes, and the other was processed with Ki-67 antibody. These slices were saved in the archives of the Department of Pathology, Peking University Shenzhen Hospital. The head of the Department of Pathology approved the use of the samples in this study. The samples were anonymous, and all patient-related data and unique identifiers were deleted. These procedures are carried out under the supervision and approval of the Ethics Committee of Peking University Shenzhen Hospital.

Our experiment dataset consists of 180 pairs H&E and Ki-67 staining images with the size of 6000 × 6000, which are basically similar in tissue structure but not pixel-level matched. All of the images were scanned at 40× with Slide Scanner Systems. We divide the total of the 180 pairs images into training and test set with 150 pairs and 30 pairs respectively. For the testing sets, the consecutive slide images are registered to each other using an alignment technique that constructs a coarse pose transformation matrix to perform an initial alignment of the tissue, then dynamically warps the Ki-67 to the respective H&E staining image. Due to memory limitations, all high-resolution training images were split into 288 × 288 tiles with 144 overlap at 20× magnification factor. After tiling, our training dataset contains about 60000 H&E 288 × 288 RGB tiles and 60000 Ki-67 288 × 288 RGB tiles.

4.2 Implementation Details

All the parameters in the convolutional layers are initialized according to Xavier et al.'s work [14]. The parameter values of $\lambda = 5$, $\beta = 5$ are chosen for the model, and Adam optimizer [15] is utilized to minimize the loss. Meanwhile, the batch size is set to 4, and the learning rate is set as 0.0002 initially and decreases using exponential decay with the decay rate of 0.9 and the decay epoch of 2. The network is trained on a computer with Intel Core i7-6850 k CPU, 128 GB RAM, and three NVidia GTX 1080-Ti GPUs.

5 Results and Evaluation

5.1 Improvement of Details Expression

We evaluate and compare our method to several variants of Cycle-GAN architecture by measuring four different evaluation metrics, i.e. SSIM [13], multi-scale structural similarity index (MS-SSIM) [16], Peak Signal to Noise Ratio (PSNR) and Mean Absolute Error (MAE), between the reconstructed H&E staining images and source H&E staining images.

Four evaluation metrics value of different methods that are calculated at different epochs are presented in Fig. 3. Compared with the other three Cycle-GAN variants, our proposed method shows a significant improvement. Meanwhile, the variant, Cycle-GAN with skip connection, and the variant, Cycle-GAN with structural cycle consistency loss, also achieve a relatively equal improvement, which means that both the skip connection and structural cycle consistency loss can further improve the quality of generated images on detail expression. In addition, more intuitive results are shown in Fig. 4.

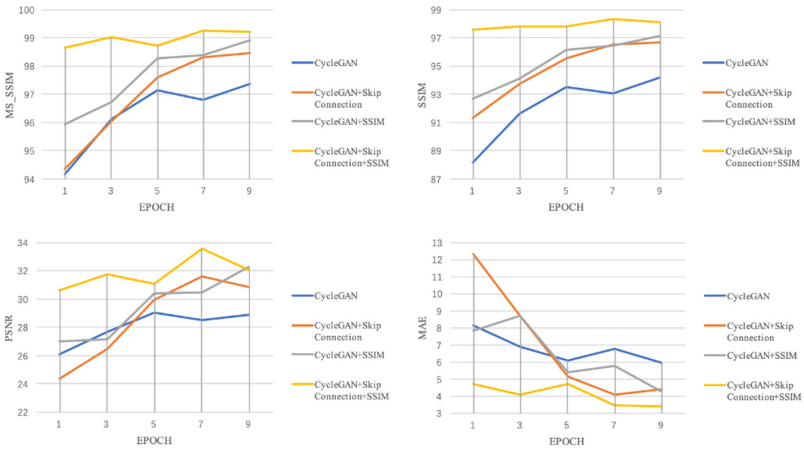

Fig. 3. Four evaluation metrics value of different methods at different epochs. CycleGAN + skip connection: Cycle-GAN with skip connection; CycleGAN + SSIM: CycleGAN with structural cycle consistency loss; CycleGAN + skip connection + SSIM: CycleGAN with skip connection and structural cycle consistency loss (our proposed).

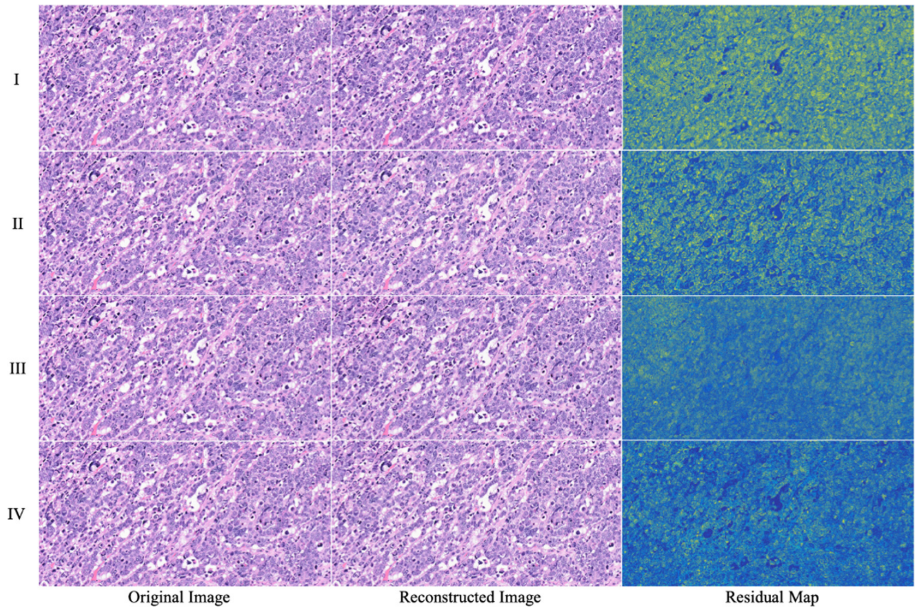

Fig. 4. The columns from left to right correspond to original image, reconstructed image, and residual map; the rows from I to IV correspond to CycleGAN, CycleGAN + skip connection, CycleGAN + SSIM, and CycleGAN + skip connection + SSIM (our proposed).

5.2 Generation of Virtual Ki-67 Staining Image

Since our original data is not pixel-level paired, the spatial structure and pixel-level evaluation strategy are not suitable for this task. Several mainly correlated features are selected from the second layer of resnet18, then the average value of each channel feature is calculated, so that the generated staining image can be expressed by a channel-level vector. Then the perceptual hash algorithm [17] is employed to calculate the correlation between the generated staining image and the referenced staining image. The visual comparison is shown in Fig. 5, where the image generated by our proposed method is obviously superior to the others' results. Furthermore, the quantification evaluation result is presented in Table 1. Our proposed method achieves the state of art with the P-hash performance of 0.7767.

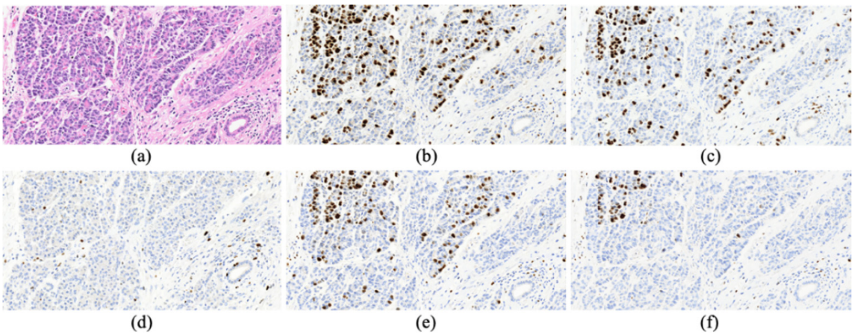

Fig. 5. Experiment results with different methods. (a) original H&E image; (d) referenced Ki-67 image; (b) generated Ki-67 image by CycleGAN; (c) generated Ki-67 from CycleGAN + skip connection; (e) generated Ki-67 from CycleGAN + SSIM; (f) generated Ki-67 from CycleGAN + skip connection + SIM (our proposed).

Table 1. Perceptual hash based on resenet18 of different methods.

Methods	P-hash
CycleGAN	0.7366 ± 0.2015
CycleGAN + Skip connection	0.7544 ± 0.2097
CycleGAN + SSIM	0.7545 ± 0.1939
CycleGAN + Skip connection + SSIM	0.7767 ± 0.1870

6 Conclusion

In the paper, we presented a method for the virtual staining task based on an improved Cycle-GAN. Our experiments revealed that our method significantly outperforms the state of the art. The visual appearance of different methods can be seen in Fig. 5. It clearly shows that the generated staining images are very similar to the referenced

staining image. However, due to the lack of pixel-level paired data, we cannot guarantee that each pixel in the H&E staining images is correctly mapped to the real Ki-67 staining domain, and there is still more room for improvement, for example, the generated Ki-67 staining image should keep same pathological presentation with the source H&E staining image. Hence, there are still certain challenges in clinical application and popularization.

Acknowledgment. This research was made possible with the financial support from National Science Foundation of China (NSFC) (61875102, 81871395, 61675113), Science and Technology Research Program of Shenzhen City (JCYJ20170816161836562, JCYJ20170817111912585, JCYJ20160427183803458, JCYJ20170412171856582, JCY20180508152528735), Oversea cooperation foundation, Graduate School at Shenzhen, Tsinghua University (HW2018007).

References

1. Weinstein, R.S., et al.: Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. Human Pathol. **40**(8), 1057–1069 (2009)
2. Soares, C.T., Frederigue-Junior, U., de Luca, L.A.: Anatomopathological analysis of sentinel and nonsentinel lymph nodes in breast cancer: hematoxylin-eosin versus immunohistochemistry. Int. J. Surg. Pathol. **15**(4), 358–368 (2007)
3. Sheikh, R.A., et al.: Correlation of Ki-67, p53, and Adnab-9 immunohistochemical staining and ploidy with clinical and histopathologic features of severely dysplastic colorectal adenomas. Dig. Dis. Sci. **48**(1), 223–229 (2003). https://doi.org/10.1023/A:1021727608133
4. Wang, Y., Sun, L.L., Jin, Q.: Enhanced diagnosis of pneumothorax with an improved real-time augmentation for imbalanced chest x-rays data based on DCNN. In: IEEE/ACM Transactions on Computational Biology and Bioinformatics (2019)
5. Tang, Z., et al.: An augmentation strategy for medical image processing based on statistical shape model and 3D thin plate spline for deep learning. IEEE Access **7**, 133111–133121 (2019)
6. Yang, J., et al.: Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain 18F-FDG PET. Phys. Med. Biol. **64**(7), 075019 (2019)
7. Tang, Z., Wang, M., Song, Z.: Rotationally resliced 3D prostate segmentation of MR images using Bhattacharyya similarity and active band theory. Physica Med. **54**, 56–65 (2018)
8. Zhang, B., et al.: Cerebrovascular segmentation from TOF-MRA using model-and data-driven method via sparse labels. Neurocomputing **380**, 162–179 (2020)
9. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural. Inf. Process. Syst. **27**, 2672–2680 (2014)
10. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
11. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
12. Zhu, J.-Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. (2017)
13. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

14. Liu, L., et al.: Understanding the Difficulty of Training Transformers. arXiv preprint arXiv: 2004.08249 (2020)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980 (2014)
16. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, vol. 2. IEEE (2003)
17. Weng, L., Preneel, B.: A secure perceptual hash algorithm for image content authentication. In: De Decker, B., Lapon, J., Naessens, V., Uhl, A. (eds.) CMS 2011. LNCS, vol. 7025, pp. 108–121. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24712-5_9