2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635603

CROSS-PLATFORM SUPER-RESOLUTION FOR HUMAN CORONARY OCT IMAGING USING DEEP LEARNING

Xueshen Li[1], *Aaron Shamouil*[1], *Xinlong Hou*[1], *Brigitta C. Brott*[2], *Silvio H. Litovsky*[2], *Yuye Ling*[3], *Yu Gan*[1]

[1]Department of Biomedical Engineering, Stevens Insitute of Technology, Hoboken, USA
[2]School of Medicine, The University of Alabama at Birmingham, Birmingham,USA
[3]Department of Electronic Engineering, Shanghai Jiao Tong University

ABSTRACT

Optical coherence tomography (OCT) has emerged as a preferred imaging method for assessing plaques before stenting and understanding blood vessel responses to intervention. However, the current image resolution still limits the effective capture of crucial intravascular elements. Although deep learning-based super-resolution techniques, relying on high-resolution (HR) and low-resolution (LR) pairs, hold promise in enhancing image resolution, existing methods primarily employ HR and LR images from the same imaging platform to demonstrate the potential of deep learning. This approach is impractical in real imaging scenarios where the HR image can not be obtained from a LR imaging platform. In this paper, we present a cross-platform deep learning framework that leverages unpaired cross-platform datasets. The HR training dataset is sourced from a high-end, high-cost OCT system, while the LR training dataset originates from a low-end, low-cost OCT system. Improving a Cycle Generative Adversarial Network with a specialized focus on coronary image structure, our experiments indicate that the new network generates super-resolved images from any LR image, demonstrating image quality comparable to OCT images acquired by HR systems.

Index Terms— Optical coherence tomography, Super-resolution, Deep learning, Cross-platform

1. INTRODUCTION

Optical coherence tomography (OCT) is a non-invasive imaging modality that is capable of generating depth-resolved reflectivity profiles in real-time using infrared interferometry [1]. Recently, intravascular optical coherence tomography (IVOCT) has become a preferred imaging method for evaluating plaques before stenting, ensuring effective stent placement, and examining how blood vessels respond to interventions [2]. However, the present optical resolution of IVOCT (10-20 μm) limits its capability to detect crucial elements such as endothelial layers and plaque erosion. Although benchtop OCT systems can provide a high spatial resolution of 2μm, there is no existing method to push the resolution of IVOCT to such level due to the limitations of imaging speed and hardware cost. A representative comparison between OCT scanning of human coronary samples using low (LR) and high resolution (HR) systems is shown in Fig. 1. Detailed texture feature and layer boundary are observed in HR images.

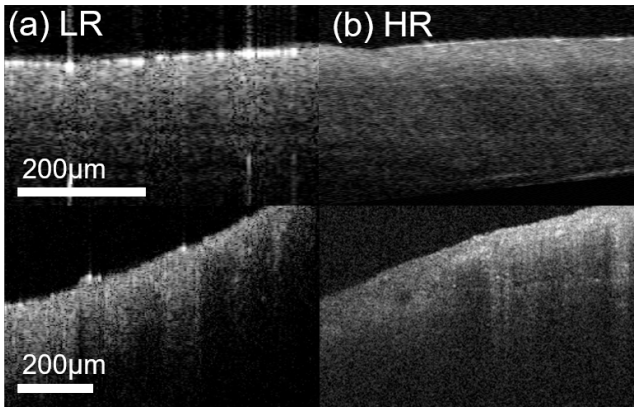

Fig. 1. An example of (a) low-resolution (LR, axial resolution of 7μm and lateral resolution of 18μm) and (b) high-resolution (HR, axial resolution of 3μm and a lateral resolution of 8μm) images of human coronary samples. The high-resolution OCT systems resolve detailed features of human coronary samples.

Deep learning (DL) algorithms have been developed to improve the resolution of OCT images [3, 4, 5, 6, 7, 8] via offline model training from HR and LR image pairs. Given an unknown LR image, the trained model can generate a HR counterpart with fine-grained details. However, there is a critical limitation on the system constrain of current DL algorithms. However, a critical constraint exists regarding the system limitations of current DL algorithms. The existing DL-based algorithms focus on training/evaluating using HR-LR image pairs from the same image platform. Specifically, the LR image is synthetically degraded (i.e., undersampled in bitwise [3], spatial domain [5], and/ or spectral domain [4, 6, 7, 8]) version of HR images. This is different from the real-

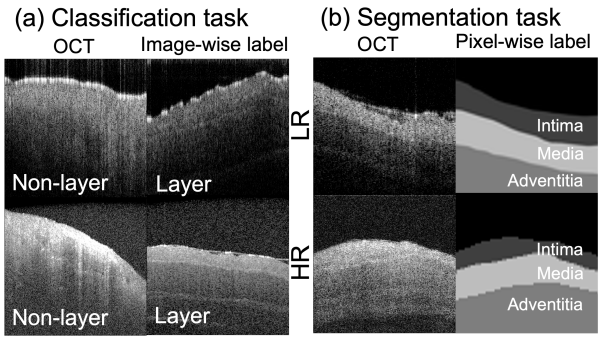

Fig. 2. Representative instances of OCT images to support (a) classification and (b) segmentation tasks. For the classification task, the image-wise labels of non-layer and layer are provided; For the segmentation task, the pixel-wise labels for the intima, media, and adventitia are provided.

world scenarios where LR imaging platform could not provide HR images due to the limitation of hardware and scanning speed. For example, from a swept source OCT imaging system with fast pullback for coronary imaging, the best resolution is 10 μm. Image resolution at 2 μm level could only be obtained in spectral domain OCT platform. Therefore, a cross-platform OCT super-resolution technology is in unmet need to bring better image quality in coronary imaging.

In this paper, we develop a deep learning framework, namely Cross-Platform Structure-Aware GAN (CPSA-GAN) to improve the quality of LR images acquired by low-cost systems. Our method is optimized for human coronary applications by considering coronary artery structures. In summary, the proposed method has the following contributions: (1) Our method functions on real-world LR and HR images acquired by low and high cost OCT systems, rather than synthetic datasets; (2) Our method does not require pixel-wisely matched OCT image pairs, which is promising to scale up and accommodate more data; (3) We incorporate structural information of human coronary samples during the process of super-resolution.

2. METHODS

2.1. OCT data acquisition

Human coronary artery samples were collected from the School of Medicine at the University of Alabama at Birmingham (UAB). We adopt a Thorlabs Ganymede commercial OCT system (axial resolution of 3μm and a lateral resolution of 8μm) for the HR system; and a Lumedica OQ Labscope OCT system (axial resolution of 7μm and lateral resolution of 18μm) for the LR system. Volumetric data (N = 30) were collected with an imaging depth of 2.56 mm and a pixel size of 2μm x 2μm.

This dataset expands upon the coronary dataset utilized in our prior research [9]. In particular, new images from a LR imaging platform were acquired. Additionally, more specimens were acquired using both LR and HR systems for evaluation purposes. Furthermore, we added more OCT coronary images with both layered and non-layered structures. A non-layered structure image corresponds to a sample where a single layer is dominating. An instance of the extended human coronary database is shown in Fig. 2. These images will be used for the classification task (for all images, Fig. 2a) and the segmentation task (for layered structure, Fig. 2b) to provide guidance in the super-resolution process.

2.2. Design of CPSA-GAN

2.2.1. Network architecture

Details of the CPSA-GAN are shown in Fig. 3. Because of the variety in coronary samples, the first layer might have dominance, or multiple layers could co-exist in OCT images. To address this, we devised a structure-aware module to classify these scenarios and incorporate this information as a constraint in the loss function. Two domains, D_H and D_L refer to HR domain and LR domain. Two generators, $G_{L\to H}$ and $G_{L\to H}$ are the image generation process from LR to HR ($G_{L\to H}$) and from HR to LR ($G_{H\to R}$). The $G_{L\to H}$ and D_H, $G_{H\to L}$ and D_L are symmetric pairs. With a unique consideration on coronary imaging structure, we design $G_{L\to H}$ and $G_{H\to R}$ with a structure that is capable of distinguishing non-layer and layer structures of human coronary structure, as well as further differentiating the intima, media, and adventitia layers. For the encoders, we adopt three convolution layers with a stride of two in a U-Net shape; for decoders, we adopt three transpose convolution layers with a stride of two. We adopt two groups of residual feature blocks (RFB) [10], eight in each group, to extract featuremaps from encoders. One group is used for image generation. The other group is combined with fully connected layers (FCL) for classification and a 1×1 convolution layer for segmentation, with an aim to provide coronary structure guidance. In addition to the generator design, we have two discriminators, D_L and D_H, which classify if the generator images is LR or HR. Here, we follow the conventional discriminator design of D_L and D_H in [11]. In testing, only the image generation module will be used for an input LR image, as shown in Fig. 3.

2.2.2. Loss function

The loss function of CPSA-GAN consists of following components: adversarial loss L_{adv}, cycle-consistency loss L_{cycle}, embedding loss L_{emb}, coronary segmentation loss L_{seg}, and coronary classification loss L_{cla}.

$$\begin{aligned} &L(G_{L\to H}, G_{H\to L}, D_H, D_L, G^{seg}_{L\to H}, G^{seg}_{H\to L}, G^{cla}_{L\to H}, G^{cla}_{H\to L}) \\ &= L_{adv}(G_{L\to H}, D_H) + L_{adv}(G_{H\to L}, D_L) \\ &+ \alpha L_{cycle}(G_{L\to H}, G_{H\to L}) + \beta L_{emb}(G_{L\to H}, G_{H\to L}) \\ &+ \gamma L_{seg}(G^{seg}_{L\to H}, G^{seg}_{H\to L}) + \delta L_{cla}(G^{cla}_{L\to H}, G^{cla}_{H\to L}) \end{aligned} \quad (1)$$

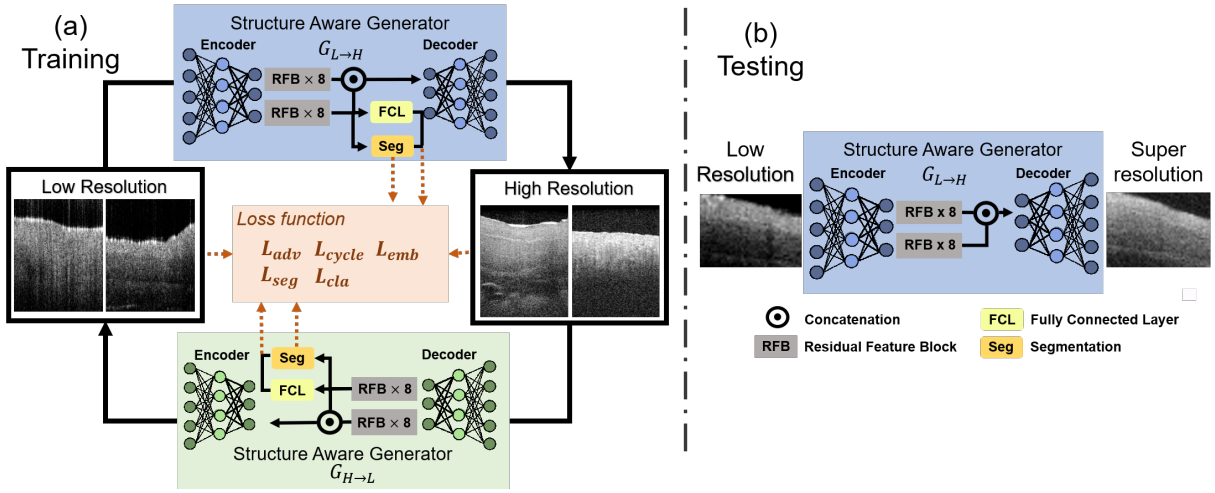

Fig. 3. Scheme of the proposed CPSA-GAN. The CPSA-GAN consists of two structure-aware generators: $G_{L\to H}$ and $G_{H\to L}$; and two discriminators. (a): During the training process, the CPSA-GAN is trained using LR and HR OCT images. The generators are aware of the human coronary structures via classification and segmentation subtasks. We use 1x1 convolution layer to merge the concatenated features from the two RFB groups. (b): During the testing phase, SR OCT images can be generated using LR images as inputs.

The L_{adv} and L_{cycle} follows the same definition in a CycleGAN structure [11]. The L_{emb} is the differences between the embeddings of two generators using L1-norm. The α, β, γ, and δ are hyper-parameters. $G^{seg}_{L\to H}$ and $G^{seg}_{H\to L}$, $G^{cla}_{L\to H}$ and $G^{cla}_{H\to L}$ are segmentation and classification subtasks performed by $G_{L\to H}$ and $G_{H\to L}$. $G^{seg}_{L\to H}$ and $G^{seg}_{H\to L}$ predict the pixel-wise segmentation of the feature map acquired by RFB. The training of $G^{seg}_{L\to H}$ and $G^{seg}_{H\to L}$ are supervised by the human coronary structural database using cross-entropy loss term. $G^{cla}_{L\to H}$ and $G^{cla}_{H\to L}$ classify coronary images with layer and non-layer structures, which is supervised by binary cross-entropy loss. We aim to solve the following minmax optimization problem:

$$G^*_{L\to H}, G^*_{H\to L} = \arg\min\max L(G_{L\to H}, G_{H\to L}, D_H, D_L, G^{seg}_{L\to H}, G^{seg}_{H\to L}, G^{cla}_{L\to H}, G^{cla}_{H\to L}) \quad (2)$$

3. EXPERIMENTAL DESIGN AND RESULTS

3.1. Experimental dataset

We performed three-fold cross-fold validation on our dataset. The OCT images were cropped and resized to 512×512. Next, image patches were randomly flipped from left to right for data augmentation. In total, 208 OCT images were acquired. A customized image processing toolkit was developed to ensure that the image angle and region was approximately the same when comparing OCT volumes from both platforms.

3.2. Evaluation setup and metrics

3.2.1. Network training details

The pixel values of OCT images were scaled to [0, 1]. The batch size was 12. The learning rate was initialized as 10^{-4}, followed by a linearly decaying decay for every two epochs. α, β, γ, and δ were empirically set to 10, 5, 5, and 5 respectively. In total, the networks were trained with 3,000 epochs to ensure convergence. The experiments were carried out in parallel on two RTX A6000 GPUs.

3.2.2. Quantitative evaluation metrics

We measure the similarity of pairs of super-resolved (SR) OCT images and HR OCT images using reference-free metrics including Fréchet inception distance (FID) [12] and Perceptual hash value (PHV) [13]. The FID is defined as:

$$FID = |\mu(SR) - \mu(HR)|^2 - Tr(\sum SR + \sum HR - 2\sqrt{\sum SR \sum HR}) \quad (3)$$

where $\mu(SR)$ and $\mu(HR)$ are the magnitudes of the SR and HR OCT images; Tr is the trace of the matrix; $\sum SR$ and $\sum HR$ are the covariance matrix of the SR and HR images. The PHV is defined as:

$$PHV = \frac{1}{N}\sum U[|avg(F_i(SR)) - avg(F_i(HR))| - T] \quad (4)$$

where N is the total number of extracted features; F_i represents the featuremap extracted from i-th layer of ResNet-101; avg is the average pooling operation that turns 3-D features

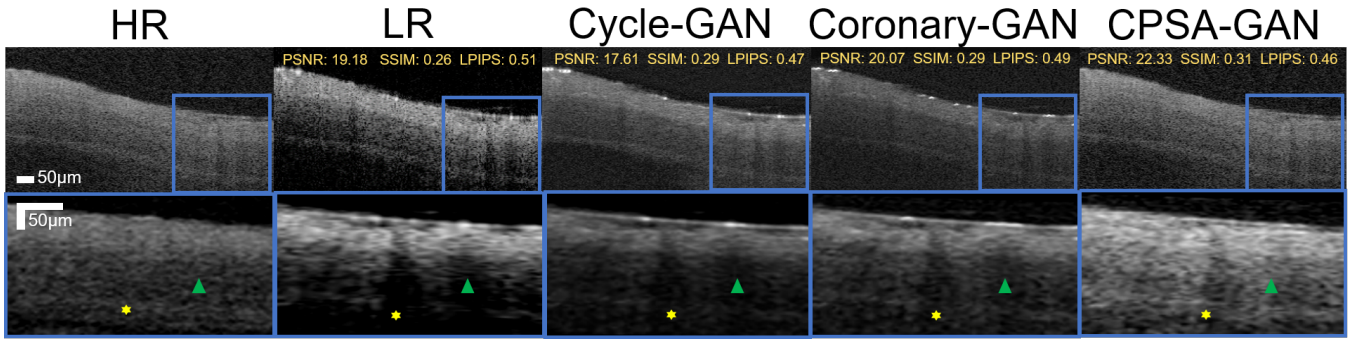

Fig. 4. The visual inspection of representative OCT images acquired from (a) LR and (b) HR systems. The SR images generated by (c) Cycle-GAN, (d) Coronary-GAN, and (e) CPSA-GAN are attached. The blue rectangle demonstrates a zoomed-in region of interest (ROI). The PNSR, SSIM, and LPIPS scores of ROIs are attached. The green triangle indicates the media region and the yellow hexagon indicates the adventitia region.

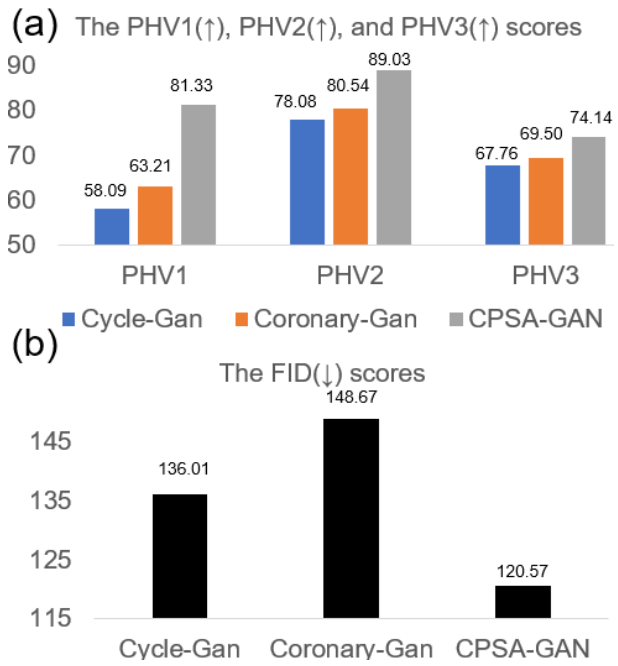

Fig. 5. The (a) PHV and (b) FID scores of our method, Coronary-GAN, and Cycle-GAN. We calculate PHV1, PHV2, and PHV3 scores from different levels of ResNet.

into 1-D features; U is the unit step function; and T is a preset threshold. We use the three variations of PHV scores ($i = 1$, PHV1), ($i = 2$, PHV2), and ($i = 3$, PHV3) which are extracted from i-th layer of ResNet-101.

3.2.3. Comparison with existing methods

We compare CPSA-GAN with existing methods including Cycle-GAN [11] and Coronary-GAN [9]. Coronary-GAN [9] is a Cycle-GAN based framework, originally used for translating OCT images to virtual histology. Minor modifications are made in input and retraining for testing.

3.3. Results

A representative SR OCT image and its corresponding LR OCT image are shown in Fig. 4. The SR image generated by CPSA-GAN is capable of revealing the intima, media (yellow hexagon symbols), and adventitia layers (green rectangle symbols). On the contrary, the media and adventitia are compromised in LR images, as well as the SR images generated by Cycle-GAN and Coronary-GAN. We register HR image with LR on pixel wise. The SR image generated by CPSA-GAN has better performance, with higher peak signal-to-noise ratio (PSNR), higher structural similarity index measure (SSIM), and lower learned perceptual image patch similarity (LPIPS) [14] scores, when compared to Cycle-GAN and Coronary-GAN.

Additionally, we calculated PHV_1, PHV_2, PHV_3, and FID scores from our cross-validation experiments over the entire dataset. The results are reported in Fig.5. Comparing to Cycle-GAN and Coronary-GAN, our CPSA-GAN generates SR OCT images that are more similar to HR OCT images, which is confirmed by higher PHV scores and lower FID scores.

4. CONCLUSIONS

We have developed a cross-platform super-resolution deep learning framework to improve the resolution of LR OCT images acquired by low-cost systems. The CPSA-GAN is optimized for human coronary samples by considering the layered structure. The experimental results indicate that the CPSAG generates super-resolved images that are comparable to HR OCT images acquired by high-cost systems. The CPSA-GAN does not need pixel-wisely matched synthetic image pairs for training, which is promising to scale up and accommodate more data. In the future, we plan to apply CPSA-GAN to the real time coronary imaging during the diagnosis and treatment of coronary artery disease.

5. COMPLIANCE WITH ETHICAL STANDARDS

The human coronary autopsy specimens were de-identified and not considered as human subjects, in compliance with the UAB at Birmingham's Institutional Review Board (IRB).

6. ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation (CRII-2222739, CAREER-2239810) and New Jersey Health Foundation.

7. REFERENCES

[1] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al., "Optical coherence tomography," *science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[2] William Wijns, Junya Shite, Michael R Jones, Stephen W-L Lee, Matthew J Price, Franco Fabbiocchi, Emanuele Barbato, Takashi Akasaka, Hiram Bezerra, and David Holmes, "Optical coherence tomography imaging during percutaneous coronary intervention impacts physician decision-making: Ilumien i study," *European heart journal*, vol. 36, no. 47, pp. 3346–3355, 2015.

[3] Qiangjiang Hao, Kang Zhou, Jianlong Yang, Yan Hu, Zhengjie Chai, Yuhui Ma, Gangjun Liu, Yitian Zhao, Shenghua Gao, and Jiang Liu, "High signal-to-noise ratio reconstruction of low bit-depth optical coherence tomography using deep learning," *Journal of Biomedical Optics*, vol. 25, no. 12, pp. 123702, 2020.

[4] Antonia Lichtenegger, Matthias Salas, Alexander Sing, Marcus Duelk, Roxane Licandro, Johanna Gesperger, Bernhard Baumann, Wolfgang Drexler, and Rainer A. Leitgeb, "Reconstruction of visible light optical coherence tomography images retrieved from discontinuous spectral data using a conditional generative adversarial network," *Biomedical optics express*, vol. 12, no. 11, pp. 6780–6795, 2021.

[5] Bin Qiu, Yunfei You, Zhiyu Huang, Xiangxi Meng, Zhe Jiang, Chuanqing Zhou, Gangjun Liu, Kun Yang, Qiushi Ren, and Yanye Lu, "N2NSR-OCT: Simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning," *Journal of biophotonics*, vol. 14, no. 1, pp. e202000282, 2021.

[6] Yijie Zhang, Tairan Liu, Manmohan Singh, Ege Çetintaş, Yilin Luo, Yair Rivenson, Kirill V. Larin, and Aydogan Ozcan, "Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data," *Light: Science & Applications*, vol. 10, no. 1, pp. 155, 2021.

[7] Xueshen Li, Shengting Cao, Hongshan Liu, Xinwen Yao, Brigitta C. Brott, Silvio H. Litovsky, Xiaoyu Song, Yuye Ling, and Yu Gan, "Multi-scale reconstruction of undersampled spectral-spatial oct data for coronary imaging using deep learning," *IEEE transactions on bio-medical engineering*, vol. 69, no. 12, pp. 3667–3677, 2022.

[8] Woojin Lee, Hyeong Soo Nam, Jae Yeon Seok, Wang-Yuhl Oh, Jin Won Kim, and Hongki Yoo, "Deep learning-based image enhancement in optical coherence tomography by exploiting interference fringe," *Communications Biology*, vol. 6, no. 1, pp. 464, 2023.

[9] Xueshen Li, Hongshan Liu, Xiaoyu Song, Brigitta C. Brott, Silvio H. Litovsky, and Yu Gan, "Structural constrained virtual histology staining for human coronary imaging using deep learning," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 2023, pp. 1–5.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[11] Jun-Yan Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.

[12] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2017, NIPS'17, pp. 6629–6640, Curran Associates Inc.

[13] 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.

[14] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.