

SIAN: STYLE-GUIDED INSTANCE-ADAPTIVE NORMALIZATION FOR MULTI-ORGAN HISTOPATHOLOGY IMAGE SYNTHESIS

Haotian Wang, Min Xian*, Aleksandar Vakanski, Bryar Shareef

Department of Computer Science, University of Idaho, USA

ABSTRACT

Existing deep neural networks for histopathology image synthesis cannot generate image styles that align with different organs, and cannot produce accurate boundaries of clustered nuclei. To address these issues, we propose a style-guided instance-adaptive normalization (SIAN) approach to synthesize realistic color distributions and textures for histopathology images from different organs. SIAN contains four phases, semantization, stylization, instantiation, and modulation. The first two phases synthesize image semantics and styles by using semantic maps and learned image style vectors. The instantiation module integrates geometrical and topological information and generates accurate nuclei boundaries. We validate the proposed approach on a multiple-organ dataset. Extensive experimental results demonstrate that the proposed method generates more realistic histopathology images than four state-of-the-art approaches for five organs. By incorporating synthetic images from the proposed approach to model training, an instance segmentation network can achieve state-of-the-art performance.

Index Terms— Histopathology image synthesis, style manipulation, nuclei annotation

1. INTRODUCTION

Histopathology image analysis has achieved great success in automatic tissue segmentation [1, 2] and cancer grading [3]. Existing deep learning-based methods require large fully-annotated datasets during the training stage, but current annotated datasets are relatively small. For example, only tens of image patches were used in [1, 4, 5, 6]. With large annotated datasets, we could train more accurate and reliable models. However, it is expensive to annotate large datasets for histopathology images, because each image may contain more than tens of thousands of nuclei.

To overcome the challenge, image synthesis is adopted. Recent works have demonstrated that high-quality synthetic images could improve the overall performance in histopathology image analysis [7, 8, 9]. However, these methods generated images only for single cancer or cancers with shared similarity, e.g., colorectal cancer [7, 8], lymph node [9]; thus

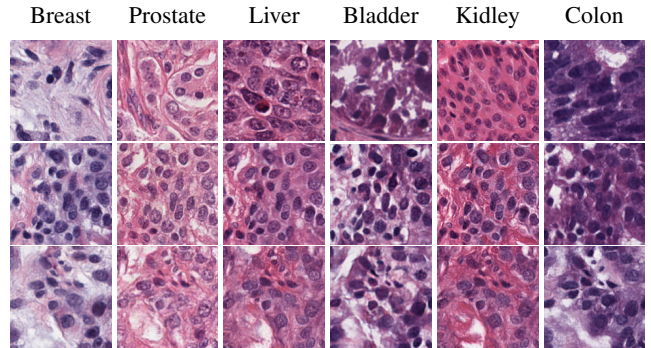


Fig. 1: Examples of image synthesis for multiple organs. The top row shows real histopathology images from six organs. The second and third rows are synthesized images generated using the proposed approach (SIAN).

their models cannot generate different image styles for different cancer types. In practice, H&E-stained images for cancers from different organs could have large color and texture variances both in foreground nuclei and background stroma (first row of Fig. 1). Therefore, it is important to have the network to generate histopathology images well in various stain distribution across multiple organs.

Recently, neural style transfer (NSF) methods have been widely exploited in many natural image synthesis tasks for manipulating image styles [10, 11, 12], they aim to learn the style from a reference image and apply it to the target image. SPADE [11] extended the AdaIN norm [10] into the spatially-adaptive manner for obtaining semantic alignments and used the encoded style vector at the beginning of a network, which enabled simultaneously style manipulation and semantic image synthesis. However, most existing histopathology image synthesis methods applied semantic layouts as the network input to learn object-level image appearance [11, 7, 8]. In histopathology image analysis, a large amount of clustered and overlapped objects may have the same semantic class label, which makes it difficult to generate accurate boundaries among clustered objects.

To alleviate the above issues, we proposed a style-guided instance-adaptive normalization (SIAN) to combine image style vector with instance layout for modulating the GAN generator. The learned transformation can effectively prop-

* corresponding author. mxian@uidaho.edu

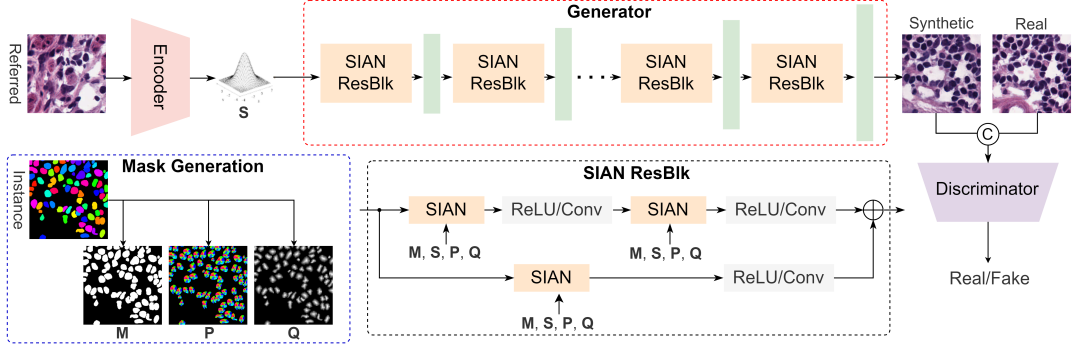


Fig. 2: Architecture of the proposed method. The encoder learns style vectors from a referred image; and the SIAN blocks integrate image style (S), semantic map (M), directional map (P), and distance map (Q) into a generator network.

agate the network to learn style factors for synthesizing histopathology images across various color distributions; the captured instance features can preserve geometrical and topological information for generating accurate densely-clustered nuclei. SIAN can generate style images from a specific organ and synthesize histopathology images with a similar style.

2. PROPOSED METHOD

2.1. Architecture and learning objectives

Fig. 2 shows the overall architecture of the proposed generator. The proposed generator has seven SIAN residual blocks (SIAN ResBlk), and each SIAN ResBlk is followed by a up-sampling layer. Each SIAN ResBlk contains two consecutive SIAN blocks, and each is followed by ReLU and convolutional layers. The skip connection has a SIAN block, a ReLU, and a convolutional layer. All input maps are down-sampled to the same height and width with the corresponding feature maps in the generator. We follow the same encoder and discriminator architectures described in [11]. The overall loss function contains five loss components, and is defined by [11]

$$\mathcal{L}_{SIAN} = \mathcal{L}_{GAN} + \lambda_1 \mathcal{L}_F + \lambda_2 \mathcal{L}_P + \lambda_3 \mathcal{L}_{KLD} + \lambda_4 \mathcal{L}_{reg} \quad (1)$$

where \mathcal{L}_{GAN} is the hinge-based conditional adversarial loss [13], \mathcal{L}_F is the feature matching loss in the multi-scale discriminator [14], \mathcal{L}_P is the perceptual loss [15] for minimizing the features between real and synthetic images, and \mathcal{L}_{KLD} is the KL divergence loss [16] for the encoder to constrain the style vector to the standard Gaussian distribution. λ_1 , λ_2 , λ_3 , and λ_4 controls the contributions of different loss terms.

2.2. Style-guided instance-adaptive normalization

We propose a new conditional normalization block, namely, the Style-guided Instance-Adaptive Normalization (SIAN) to learn instance-level features and integrate image styles for cancers from different organs. Fig. 3 shows the details of the SIAN block. The block has four phases: semantization, stylization, instantiation, and modulation. The block takes four

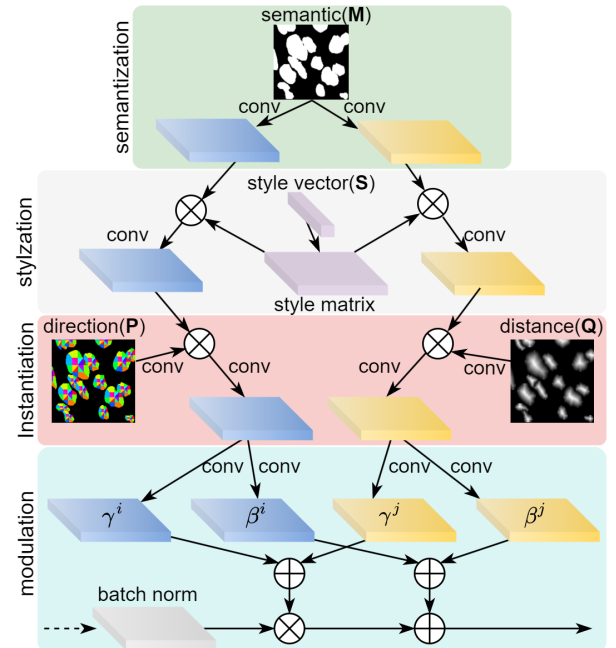


Fig. 3: SIAN normalization. SIAN block takes four inputs: semantic masks M , style vector S , directional mask P , and distance mask Q for combining different features at multiple phases in the block. \otimes denotes element-wise multiplication, and \oplus is element-wise addition.

inputs besides image feature maps, i.e., semantic mask M , style vector S , direction mask P , and distance mask Q . The semantization phase embeds image semantics from the input mask; the stylization creates a style matrix from a referred image and integrates image semantics and style. The instantiation phase uses direction and distance maps to distinguish individual nuclei. The modulation phase learns the scale and bias and integrates them into the network.

Let \mathbf{h} denote the input activation of the current layer of the proposed neural network with a batch size of N . Let H , W and C denote the height, width, and channels of an activation map in i th layer. The final modulated activation value ($n \in$

$N, c \in C, y \in H, x \in W$) is defined as

$$\gamma_{c,y,x}(\mathbf{M}, \mathbf{S}, \mathbf{P}, \mathbf{Q}) \frac{h_{n,c,y,x} - \mu_c}{\sigma_c} + \beta_{c,y,x}(\mathbf{M}, \mathbf{S}, \mathbf{P}, \mathbf{Q}) \quad (2)$$

where $h_{n,c,y,x}$ is the activation output before normalization; the modulation parameter $\gamma_{c,y,x}$ and $\beta_{c,y,x}$ are the element-wise summation of modulation parameters of two branches, i.e., $\gamma_{c,y,x}^i + \gamma_{c,y,x}^j$ and $\beta_{c,y,x}^i + \beta_{c,y,x}^j$. μ_c and σ_c are the mean and standard deviation of the activation of the channel c , respectively. In the SIAN block, the semantic layout first passes to two convolutional layers, which split the semantic information into two separate branches to learn the directional features and distance features separately. The two branches have the same architecture. In each branch, the convolutional kernel first multiplies with the reshaped style vector, which combines style factors in the block. After that, the instance layouts (direction or distance) are fed through a 1×1 convolutional layer and multiplied with the previous convolutional layer. The next convolutional layer learns the compensation of semantic, style, and instance features and then split into two convolutional layers to learn the modulation parameters (γ and β) spatially. Finally, those modulation parameters and the output of batch normalization are integrated for accurate histopathology image synthesis. All convolutional layers in SIAN use the 3×3 kernel size with 128 filters.

Instance masks are applied to generate the semantic mask, and nuclei directional and distance maps. The semantic map is used to separate nuclei and stroma, and the directional and distance maps are useful to demonstrate the boundaries and centroids between two or more touching nuclei. We employed the 2-bin direction mask [17] and Medial Axis (MA) distance mask [18] as the instance descriptors. Direction map provides important centroid and directional information of nuclei. MA distance mask shows the distance between the nucleus boundary to its skeleton while providing nuclei topological and geometrical features.

3. EXPERIMENTAL RESULTS

3.1. Dataset, metrics, and setting

Dataset. The experiments are conducted on the multi-organ nuclei segmentation dataset (MoNuSeg) [4] which has 44 H&E stained histopathology image patches. Both the training and test sets contain images from six organs including breast, liver, kidney, prostate, bladder, and colon; while the training set includes stomach as the seventh organ, and the testing set has brain images.

Evaluation metrics. We employed five metrics to evaluate the method performance for image synthesis, e.g., FID [19], SSIM [20], DQ [21], SQ [21], and PQ [21]. We used two metrics FID and SSIM to measure the distribution distance and structural similarity between real images and synthetic images, respectively; and used DQ, SQ, and PQ are utilized

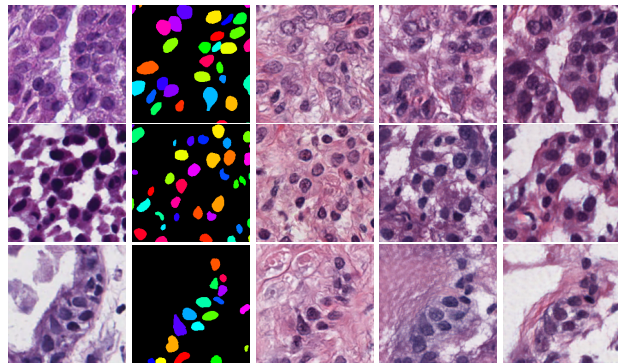


Fig. 4: Visual comparison of histopathology image synthesis for the MoNuSeg test set. (a) Image patches, (b) nuclei masks, (c-e) results of Sharp-GAN, SPADE, and ours, respectively

to assess the nuclei segmentation performance. Specifically, we run a pre-trained segmentation model (SegNet [22]) which is trained on real images, and then test and evaluate using the synthetic images. In addition, we show the visual comparison of our synthetic images compared to other methods.

Implementation details. The input image size of our approach is 256×256 . We used random flip, rotation and median blur for data augmentation. We use the ADAM optimizer with the total training epochs of 50 and batch size of 8 to train the network. The experiments are conducted on a NVIDIA RTX 8000 GPU.

During inference, for the style encoder, we take an arbitrary histopathology image as input and output the encoded style vectors. Then, the encoded style vectors together with an arbitrary instance mask are input into the trained generator network to produce histopathology images.

3.2. Image quality assessment

The proposed method is compared with four state-of-the-art image synthesis models: pix2pix GAN [23], Sharp-GAN [24], pix2pixHD [25], and SPADE [11] using FID, SSIM, DQ, SQ, and PQ metrics. The quantitative results of different approaches on the MoNuSeg test set are shown in Table 1. The proposed method outperforms the state-of-the-art methods both in image reconstruction quality using SSIM and FID, and segmentation quality using PQ, SQ and DQ. In addition, we integrated instantiation phase (INST) only, and the SIAN block with the style vectors (STYLE), all the evaluation metrics are improved from the baseline SPADE. Overall, we can conclude that our SIAN achieved the best quantitative performance among other methods.

Fig. 4 compares the proposed method, Sharp-GAN [24], and SPADE [11] using three examples. We noted that Sharp-GAN cannot recover the texture and color distributions of nuclei and stroma in real images, especially in the first and second row. I.e., the synthetic nuclei have different appearances from real nuclei, and their background stroma lacks meaningful texture and color. SPADE achieved better per-

Methods	FID↓	SSIM↑	DQ↑	SQ↑	PQ↑
pix2pix [23]	170.1	0.467	0.687	0.717	0.493
Sharp-GAN [24]	155.2	0.483	0.721	0.745	0.538
pix2pixHD [25]	186.3	0.479	0.750	0.753	0.566
SPADE [11]	134.6	0.488	0.705	0.738	0.552
INST	125.4	0.491	0.748	0.768	0.575
STYLE	116.5	0.506	0.743	0.769	0.571
SIAN	115.7	0.515	0.757	0.761	0.586

Table 1: Overall performance on MoNuSeg datasets with reconstruction metrics and segmentation metrics.

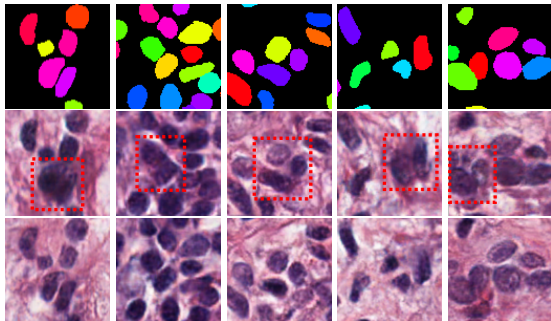


Fig. 5: Synthesis for clustered nuclei. First row: instance masks (different colors represent different nuclei). Second row: results of SPADE [11]. Third row: results of SIAN.

formance compared to Sharp-GAN, but the generated images were not realistic. Our SIAN generates more realistic images than SPADE and Sharp-GAN. SPADE used the semantic layout as input, while our method used the instance layouts. As shown in Figure 5, SPADE tends to generate blur and incorrect nuclei in the clustered region. Our approach produces more accurate boundaries for clustered nuclei.

3.3. Multi-organ image synthesis

To evaluate synthetic images across multiple organs, we compare the generation performance of four methods for each organ using FID scores. The results are shown in Table 2. The proposed method outperforms other state-of-the-art methods in kidney, prostate, bladder, colon, and brain images, and achieved the second-best results for synthesizing breast and lung images. Fig. 1 shows the results of SIAN across multiple organs, the color and texture distribution of foreground nuclei and background stroma are close to the reals.

Meth.	Bre.	Kid.	Pro.	Bla.	Col.	Lun.	Bra.
[23]	211.0	197.3	237.5	246.9	239.3	187.5	304.4
[24]	198.0	190.4	235.3	211.4	208.6	195.4	260.3
[25]	260.0	222.1	260.1	278.6	221.8	252.8	278.6
[11]	212.4	187.9	200.4	212.5	222.3	207.2	245.5
SIAN	200.4	173.9	190.2	196.9	204.1	194.6	239.6

Table 2: Performance comparison of image synthesis for multiple organs using the FID score.

Training Set	DQ	SQ	PQ
MoNuSeg training set	0.704	0.737	0.521
MoNuSeg training set*	0.732	0.739	0.538
+pix2pixHD	0.742	0.737	0.544
+Sharp-GAN	0.740	0.739	0.547
+SPADE	0.743	0.738	0.549
+SIAN	0.748	0.742	0.555

Table 3: Performance of SegNet using different training sets.

* denotes the training set augmented using traditional augmentation techniques, e.g., flip, rotate, blur. '+method' denotes applying synthetic augmentation* with 5,000 synthetic images generated from 'method' to the training set.

3.4. Nuclei segmentation using synthetic images

In this experiment, we evaluate the effectiveness of synthetic augmentation for training segmentation networks. We train SegNet [22] with different input configurations (as shown in Table 3). In experiments, nucleus-like polygons are generated as the synthetic nuclei instance masks [26], in total 5,000 synthetic instance masks are generated and applied to produce corresponding semantic, directional, and distance masks. Then, the pre-trained SIAN is used to apply seven different style vectors encoded from seven different organs (around 700 synthetic images per organ) and generate realistic histopathology images. Finally, we test and evaluate the segmentation performance with the MoNuSeg test set using DQ, SQ, PQ metrics. We compared the proposed approach to other methods on synthetic augmentation. Synthetic images generated from other methods follow their design. As shown in Table 3, with synthetic training images from pix2pixHD, Sharp-GAN, and SPADE, the performance of nuclei segmentation could be significantly improved. The proposed SIAN help generate the best segmentation performance.

4. CONCLUSION

In this paper, we propose the style-guided instance-adaptive normalization (SIAN) for multi-organ histopathology image synthesis, which integrates instance layouts and encodes style vectors into a generative network. SIAN synthesizes histopathology images with styles that align with the image styles of multiple organs. SIAN utilizes the directional and distance masks from the nuclei instance maps and generates clear boundaries for densely-clustered nuclei. With the integration of the stylization phase, SIAN allows style editing for synthesizing images of multiple organs. In addition, SIAN demonstrates its effectiveness in augmenting the training set and improving the overall performance of a deep learning model for nuclei segmentation.

5. REFERENCES

- [1] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, et al., “Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images,” *Medical Image Analysis*, vol. 58, pp. 101563, 2019.
- [2] Haotian Wang, Min Xian, and Aleksandar Vakanski, “Bending loss regularized network for nuclei segmentation in histopathology images,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1–5.
- [3] Muhammad Shaban, Ruqayya Awan, Muhammad Moazam Fraz, et al., “Context-aware convolutional neural network for grading of colorectal cancer histology images,” *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2395–2405, 2020.
- [4] Neeraj Kumar, Ruchika Verma, Deepak Anand, et al., “A multi-organ nucleus segmentation challenge,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020.
- [5] Quoc Dang Vu, Simon Graham, Tahsin Kurc, Minh Nguyen Nhat To, et al., “Methods for segmentation and classification of digital microscopy tissue images,” *Frontiers in bioengineering and biotechnology*, p. 53, 2019.
- [6] Peter Naylor, Marick Laé, Fabien Reyat, and Thomas Walter, “Segmentation of nuclei in histopathology images by deep regression of the distance map,” *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.
- [7] Srijay Deshpande, Fayyaz Minhas, Simon Graham, and Nasir Rajpoot, “Safron: Stitching across the frontier network for generating colorectal cancer histology images,” *Medical Image Analysis*, vol. 77, pp. 102337, 2022.
- [8] Jerry Wei, Arief Suriawinata, Louis Vaickus, et al., “Generative image translation for data augmentation in colorectal histopathology images,” *Proceedings of machine learning research*, vol. 116, pp. 10, 2019.
- [9] Yuan Xue, Jiarong Ye, Qianying Zhou, et al., “Selective synthetic augmentation with histogan for improved histopathology image classification,” *Medical image analysis*, vol. 67, pp. 101816, 2021.
- [10] Xun Huang and Serge Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1501–1510.
- [11] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, et al., “Semantic image synthesis with spatially-adaptive normalization,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2337–2346.
- [12] Tero Karras, Samuli Laine, and Timo Aila, “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.
- [13] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida, “Spectral normalization for generative adversarial networks,” in *International Conference on Learning Representations*, 2018.
- [14] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, et al., “High-resolution image synthesis and semantic manipulation with conditional gans,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.
- [15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European conference on computer vision*. Springer, 2016, pp. 694–711.
- [16] Diederik P. Kingma and Max Welling, “Auto-encoding variational bayes,” in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun, Eds., 2014.
- [17] Liang-Chieh Chen, Alexander Hermans, George Papandreou, et al., “Masklab: Instance segmentation by refining object detection with semantic and direction features,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4013–4022.
- [18] Haotian Wang, Min Xian, and Aleksandar Vakanski, “Ta-net: Topology-aware network for gland segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1556–1564.
- [19] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, et al., “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” *Advances in neural information processing systems*, vol. 30, 2017.
- [20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [21] Alexander Kirillov, Kaiming He, Ross Girshick, et al., “Panoptic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9404–9413.
- [22] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [23] 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.
- [24] Sujata Butte, Haotian Wang, Min Xian, and Aleksandar Vakanski, “Sharp-gan: Sharpness loss regularized gan for histopathology image synthesis,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–5.
- [25] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, et al., “High-resolution image synthesis and semantic manipulation with conditional gans,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.
- [26] Le Hou, Ayush Agarwal, Dimitris Samaras, et al., “Robust histopathology image analysis: To label or to synthesize?,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8533–8542.