

# SIAN: Style-Guided Instance-Adaptive Normalization for Multi-Organ Histopathology Image Synthesis

Haotian Wang, Min Xian\*, Aleksandar Vakanski, Bryar Shareef  
 University of Idaho, USA

haotianw, mxian, vakanski, bryarshareef@uidaho.edu

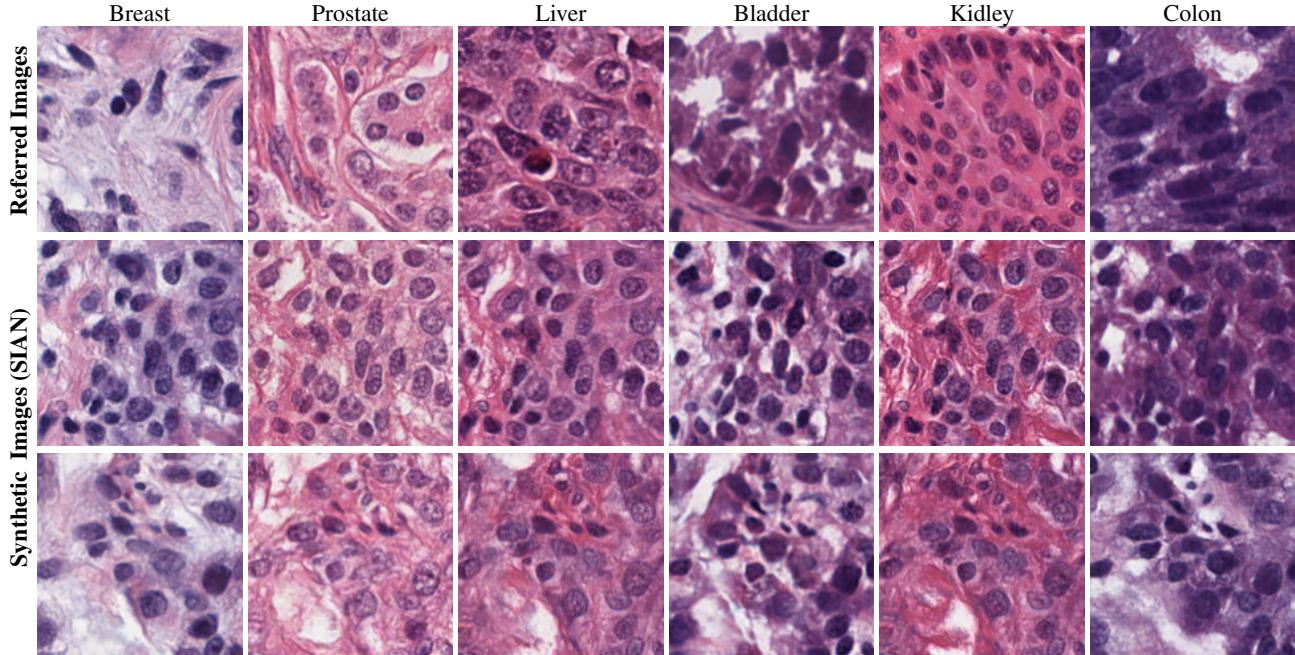


Figure 1: Examples of image synthesis for multiple organs. The top row shows real histopathology images from six organs. The second and third rows are synthesized images generated using the proposed approach (SIAN).

## Abstract

Existing deep networks for histopathology image synthesis cannot generate accurate boundaries for clustered nuclei and cannot output image styles that align with different organs. To address these issues, we propose a style-guided instance-adaptive normalization (SIAN) to synthesize realistic color distributions and textures for different organs. SIAN contains four phases, semantization, stylization, instantiation, and modulation. The four phases work together and are integrated into a generative network to embed image semantics, style, and instance-level boundaries. Experimental results demonstrate the effectiveness of all components in SIAN, and show that the proposed method outperforms the state-of-the-art conditional GANs for histopathology image synthesis using the Frechet Incep-

tion Distance (FID), structural similarity Index (SSIM), detection quality(DQ), segmentation quality(SQ), and panoptic quality(PQ). Furthermore, the performance of a segmentation network could be significantly improved by incorporating synthetic images generated using SIAN.

## 1. Introduction

Histopathology image analysis has achieved great success in automatic tissue segmentation [4, 36, 35, 10, 9] and cancer grading [30, 41, 28]. 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 [22, 33, 10, 26]. With large annotated datasets, we could train more accurate and reliable models. However, it is ex-

pensive 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, 3, 42, 23, 43]. However, these methods generated images only for a single cancer or cancers with shared similarity, e.g., colorectal cancer [7, 42], lymph node [43]; thus 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 Figure 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 [27, 8, 18, 14]. NSF approaches aim to learn the style from a reference image and apply it to the target image. StyleGAN [18] encoded referred style images into the latent vector and applied it to AdaIN norm [14] at different stages of a neural network for style manipulation. SPADE [27] extended the AdaIN norm 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. Most existing image synthesis methods applied semantic layouts as the network input to learn object-level image appearance [27, 7, 42, 23]. 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 those 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 propagate the network to learn style factors for synthesizing histopathology images across various color distributions and instance features for generating accurate densely-clustered nuclei. SIAN allows the user to choose a style image from a specific organ, and synthesizes histopathology images with the similar style. To the best of our knowledge, this is the first work to investigate instance-level style editing for histopathology image synthesis.

## 2. Related Work

**Conditional image synthesis.** Conditional generative adversarial networks (cGANs) engage additional information to guide image generation, and they have achieved promising results in many image synthesis tasks, e.g.,

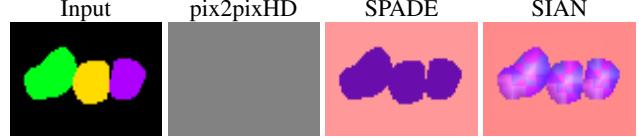


Figure 2: An illustration of results from different normalization methods. In the input mask, different colors represent different nuclei. pix2pixHD [38] used unconditional normalization (InstanceNorm [32]) that tended to wash away mask information, SPADE [27] produced semantic class features but was unable to distinguish clustered nuclei via only semantic layout. The proposed SIAN can clearly generate boundaries between nuclei.

image-to-image synthesis [16, 29], multi-modal synthesis [27, 45], high-resolution image synthesis [38, 7], image-guided image synthesis [37]. Pix2Pix GAN [16] learned mapping from an input image to an output image using the U-Net architecture. SPADE [27] adopted semantic layout for preserving the semantic class features, and improved unconditional normalization methods such as InstanceNorm [32] and BatchNorm [15].

**Histopathology image synthesis.** Histopathology image synthesis provides a promising solution for data-efficient methods without requiring extensive human resources. It has been widely studied in many applications [13, 3, 42, 7, 43, 25]. Deshpande *et al.* [7] proposed a SAFRON GAN to address high-resolutinal histopathology image generation. SAFRON could generate large-size images without requiring same size inputs. Wei *et al.* [42] introduced a GAN to generate synthetic images belonging to diagnostically less common samples to mitigate the data imbalance issue in adenomas detection.

**Style manipulation.** Style manipulation and editing has achieved great success in image synthesis and translation tasks [14, 27, 18, 46, 17, 8]. Gong *et al.* [8] converted different histopathology images to one target style for minimizing the impact of the various stain distributions. StyleGAN [18] proposed an encoder network for encoding style image into the latent vector; and one model could learn multiple image styles from a training set. This network could generate images with various styles while using the pre-trained encoder. Many recent style manipulation GANs such as StyleGAN [18], SPADE [27] and their variants [45, 29] applied the style encoding method for effective style transfer.

## 3. SIAN: Style-Guided Instance-Adaptive Normalization

Figure 2 illustrates the results of two leading image synthesis GANs, pix2pixHD[38] and SPADE [27], and our method. pix2pixHD applied unconditional normalization (InstanceNorm[32]), which tended to wash away mask in-

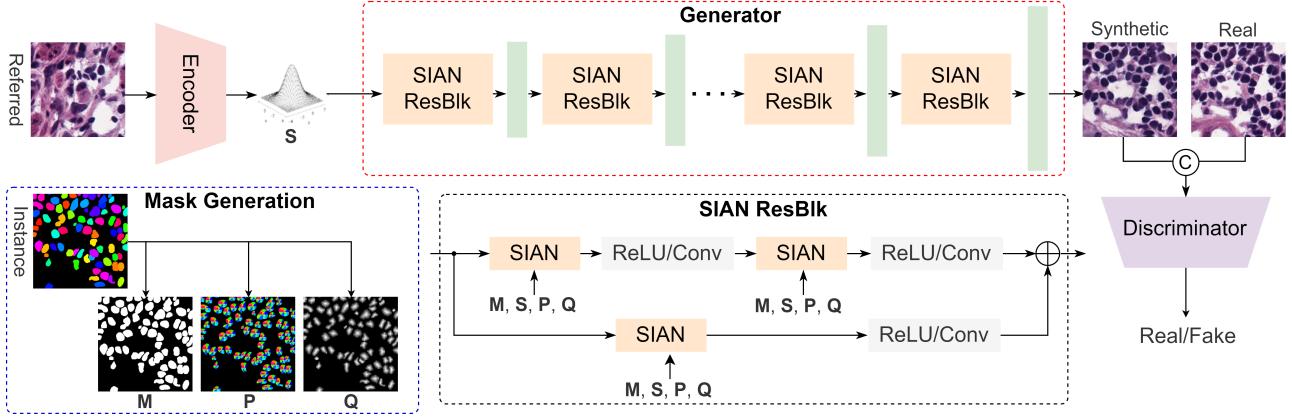


Figure 3: Architecture of the proposed method. The encoder learns style vectors from a referred image; and the SIAN blocks integrate image style ( $S$ ), semantic map ( $M$ ), directional map ( $P$ ), and distance map ( $Q$ ) into a generator network.

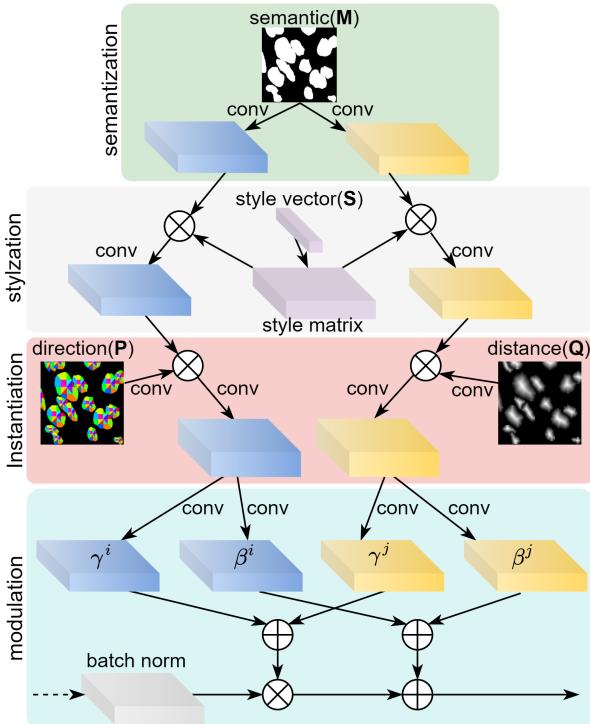


Figure 4: SIAN normalization. SIAN block takes four inputs: semantic mask  $M$ , style vector  $S$ , directional mask  $P$ , and distance mask  $Q$  for combining different features at multiple phases in the block.  $\otimes$  denotes element-wise multiplication, and  $\oplus$  is element-wise addition.

formation in the normalization. SPADE utilized semantic class maps to modulate the layer, but was unable to distinguish clustered nuclei. We adopt instance segmentation layouts for preserving topological and geometrical information

to generate clustered nuclei (Figure 2). Moreover, SPADE only used the style vector at the beginning of the generator, thus the style vector may not be effectively exploited. StyleGAN [18] and its variants demonstrate that inserting the encoded style vector into multiple layers of the network could produce promising style editing performance. Inspired by the design, we integrate the encoded style vector into every normalization layer for strengthening the impact of style vectors and generating more realistic histopathology images with different stain distributions.

Let  $\mathbf{X} \in \mathbb{R}_{\geq 0}^{M \times N}$  be a instance segmentation mask with  $n$  nucleus instances inside, where  $M$  and  $N$  are the image height and width. Each element in  $\mathbf{X}$  takes value from 0 to  $n$ , where 0 represents background stroma, and 1 to  $n$  are different individual nuclei. The goal is to learn a mapping function from mask  $\mathbf{X}$  to the realistic histopathology image.

### 3.1. Architecture and learning objective

Figure 3 shows the overall architecture of the proposed generator. We employed SPADE[27] as our baseline architecture, and all the SPADE blocks and SPADE ResNet blocks are replaced with our SIAN blocks and SIAN residual blocks, respectively. In total, our generator has seven SIAN residual blocks (SIAN ResBlk) and the following up-sampling layer; and each SIAN ResBlk contains two consecutive SIAN blocks followed by ReLU and convolutional layers. The skip connection has a SIAN block, a ReLU, and a convolutional layer. All input factors are down-sampled to the same height and width with the corresponding feature maps in the generator. We follow the same encoder and discriminator architecture described in [27].

The overall loss function contains five major loss com-

ponents, and is defined by

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

where  $\mathcal{L}_{GAN}$  is the hinge-based conditional adversarial loss[24, 44],  $\mathcal{L}_F$  is the feature matching loss to minimize the difference between synthetic and real images in the multi-scale discriminator[39],  $\mathcal{L}_P$  is the perceptual loss[17] in the pre-trained VGG19 [31] for minimizing the features between real and synthetic images,  $\mathcal{L}_{KLD}$  is the KL divergence loss[20] for the encoder to constrain the style vector to the standard Gaussian distribution. Inspired by [29], we introduced a new loss term ( $\mathcal{L}_{reg}$ ) to change the encoded style vectors to their average style vectors in a mini-batch for increasing the fidelity of synthetic images. It is given by

$$\mathcal{L}_{reg} = \sum_{i=1}^D |z_i - \bar{z}_i| \quad (2)$$

where the encoded style vectors  $z = (z_i)$  and its averaged style vectors  $\bar{z} = (\bar{z}_i)$  are  $D$  dimensional vectors. The design is similar to the truncation trick in BigGAN [2], and we note that using the mean absolute loss to constrain the encoded style vectors could improve the fidelity of synthetic images. The contribution of loss terms are  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$ .

### 3.2. 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. Figure 4 shows the details of the SIAN block. The block has four phases: semantization, stylization, instantiation, and modulation. The block takes four 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  $h$  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

$$\gamma_{c,y,x}(M, S, P, Q) \frac{h_{n,c,y,x} - \mu_c}{\sigma_c} + \beta_{c,y,x}(M, S, P, Q) \quad (3)$$

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,

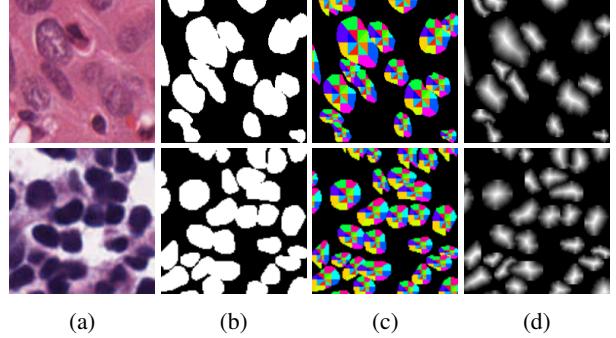


Figure 5: Examples of clustered nuclei. a) Real histopathology image patches; b) semantic masks with nuclei in white and stroma in black; c) 2-bin directional masks, which split each nucleus into 16 directional regions (different colors represent different regions); and d) the MA distance masks.

i.e.,  $\gamma_{c,y,x}^i + \gamma_{c,y,x}^j$  and  $\beta_{c,y,x}^i + \beta_{c,y,x}^j$ .  $\mu_c$  and  $\sigma_c$  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  $1 \times 1$  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 ( $\gamma$  and  $\beta$ ) spatially. Finally, those modulation parameters and the output of batch normalization are integrated for accurate histopathology image synthesis.

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[5] and Medial Axis (MA) distance mask[36] as the instance descriptors. Figure 5 shows two examples. 2-bin direction mask uses two parts of 8 directional regions (8 class insider, 8 class outsider) to obtain directional features, and each class label points to the nucleus centroid. These contain a total number of (16 + 1) classes in the feature map. As shown in Figure 5(c), the inside 8 classes in our direction maps can provide important centroid information of nuclei, which could improve the traditional eight-directional class [11]. MA distance mask shows the distance between the nucleus boundary to its skeleton while providing nuclei topologi-

cal and geometrical features. It can be observed from Figure 5 that the semantic masks lack the ability to show clear boundaries on clustered nuclei, while direction masks and distance masks provide useful important information, e.g., nuclei connectedness, distance, and centroids.

### 3.3. Implementation details

The 30 training images are tiled with overlap to generate 2,994 image patches, and flipping and rotation are applied to augment data during model training. The 14 test images are split into 224 non-overlapping image patches. The input image size of the proposed approach is  $256 \times 256$ . We use the ADAM [19] optimizer ( $\beta_1 = 0$ ,  $\beta_2 = 0.999$ ) with the total training epochs of 50 and batch size of 8 to train the network. The learning rate is set to  $1 \times 10^{-4}$  and  $4 \times 10^{-4}$  for the generator and discriminator, respectively. The contributions of loss terms are set by experiments:  $\lambda_1=10$ ,  $\lambda_2=10$ ,  $\lambda_3=0.05$ , and  $\lambda_4=0.5$ . Spectral Norm (L2-norm) [24] is used as the extra normalization layer in the both generator and discriminator[27]. We evaluate the model every five epochs and select the model with the best performance for the inference. The experiments are conducted on a single 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.

## 4. Experimental Results

### 4.1. Dataset and evaluation metrics

The experiments are conducted on the MoNuSeg dataset[22] which has 44 H&E stained histopathology image patches from eight organs. The image size is  $1000 \times 1000$ . The original WSI images were scanned under  $40 \times$  magnification, and the nuclei included both benign and diseased appearances. 30 images with 21,623 fully annotated nuclei are used for training, and 14 images with 7,223 fully annotated nuclei are in the test set. 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.

We employed five metrics to evaluate the method performance for image synthesis, e.g., Frechet Inception Distance (FID) [12], structural similarity Index (SSIM) [40], detection quality(DQ) [21], segmentation quality(SQ) [21], and panoptic quality(PQ) [21]. Previous image synthesis works[27, 45] applied two approaches to evaluate the performance: 1) reconstruction performance, and 2) segmentation results. We used two metrics FID and SSIM to measure the distribution distance and structural similarity between

Methods	FID $\downarrow$	SSIM $\uparrow$	DQ $\uparrow$	SQ $\uparrow$	PQ $\uparrow$
pix2pix [16]	170.3	0.468	0.651	0.718	0.493
Sharp-GAN [3]	156.2	0.482	0.699	0.744	0.538
pix2pixHD [38]	186.3	0.479	0.750	0.753	0.566
SPADE [27]	134.6	0.488	0.681	0.738	0.533
SIAN	<b>115.7</b>	<b>0.515</b>	<b>0.757</b>	<b>0.761</b>	<b>0.586</b>

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

real images and synthetic images, respectively. The segmentation results are evaluated by using the instance-level object detection and segmentation performance [10, 34]. The detection quality(DQ) [21], segmentation quality(SQ) [21], and panoptic quality(PQ) [21] are utilized to assess the nuclei segmentation performance. Specifically, we run a pre-trained segmentation model (SegNet[1]) 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 the state-of-the-art methods.

### 4.2. Overall performance

The proposed method is compared with four state-of-the-art image synthesis models: pix2pix GAN [16], Sharp-GAN [3], pix2pixHD [38], and SPADE [27]. For a fair comparison, we train all methods using their official implementations provided by the authors and use the best model for testing. We employ FID [12], SSIM [40], DQ [21], SQ [21], and PQ [21] to measure the generation performance. 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.

Figure 6 compares the proposed method, Sharp-GAN[3], and SPADE[27] using five examples. Sharp-GAN generates images conditioned on masks only, and SPADE and the proposed method used the test images as the referred images to generate the style vectors. 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 appearance from real nuclei, and their background stroma lacks meaningful texture and color. SPADE achieved better performance compared to Sharp-GAN, but the generated images were not realistic. The proposed method 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 7, SPADE tends to generate blur and incorrect nuclei in the clustered region. The proposed approach produces more accurate boundaries for clustered nuclei.

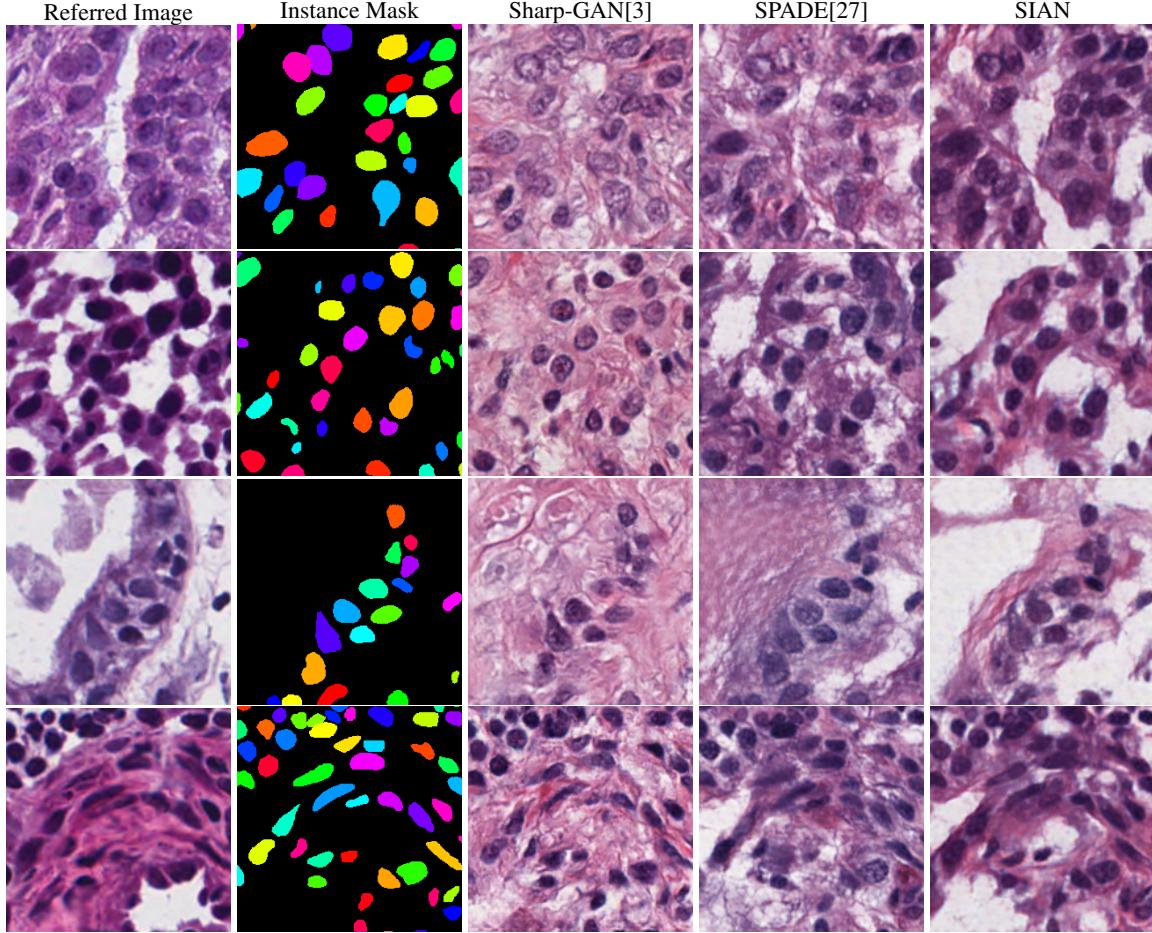


Figure 6: Visual comparison of histopathology image synthesis for the MoNuSeg test set.

Methods	Breast	Kidney	Prostate	Bladder	Colon	Lung	Brain
pix2pix [16]	211.0	197.3	237.5	246.9	239.3	<b>187.5</b>	304.4
Sharp-GAN [3]	<b>198.0</b>	190.4	235.3	211.4	208.6	195.4	260.3
pix2pixHD [38]	260.0	222.1	260.1	278.6	221.8	252.8	278.6
SPADE [27]	212.4	187.9	200.4	212.5	222.3	207.2	245.5
SIAN	200.4	<b>173.9</b>	<b>190.2</b>	<b>196.9</b>	<b>204.1</b>	194.6	<b>239.6</b>

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

### 4.3. 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. FID calculates the distribution similarity between real and synthetic images, and a lower FID score suggests that a synthetic image is more similar to the real image. The FID scores 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. It is worth noting that, although there are no referred brain images in the training set, the proposed

approach still outperforms the other four methods.

### 4.4. Ablation study

Three major components exist in the proposed approach: 1) the instantiation phase (INST), 2) the SIAN block with the style vectors (STYLE), 3) the style regularization loss (Style Loss) ( $\mathcal{L}_{reg}$ ). In the experiments, SPADE is used as the baseline model, and SIAN is the final approach that contains all three components. As shown in Table 4, involving the instance layouts in the normalization could improve the PQ score by 10.1% when compared to the baseline. The STYLE improves FID of the baseline and the INST by 10.8% and 5.3%, respectively. The Style Loss increases the

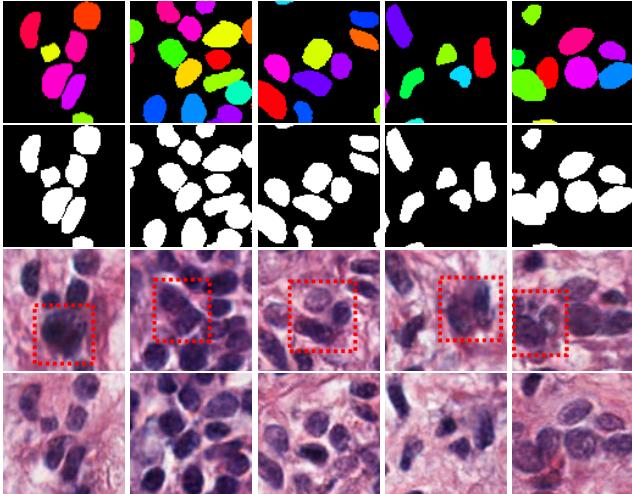


Figure 7: Synthesis for clustered nuclei. First row: instance masks (different colors represent different nuclei). Second row: semantic masks. Third row: results of SPADE[27]. Fourth row: results of SIAN.

PQ value of the baseline by 4.9%. Overall, the final SIAN improves the FID, SSIM, and PQ of the baseline by 16.1%, 5.5%, and 6.3%.

#### 4.5. Nuclei segmentation using synthetic images

In this experiment, we evaluate the effectiveness of the proposed approach by using synthetic images to augment the training set for training segmentation networks. Image augmentation using synthetic images has been widely applied in medical image analysis [6] and histopathology image analysis[3, 43, 42, 7]. In this work, we train SegNet[1] with four different configurations for evaluating synthetic augmentation, including 1) training with only real images, 2) training with both real and augmented images using traditional augmentation methods such as image rotations, image flips, and median blur (real images\*), 3) training with real and synthetic images (synthetic aug.), and 4) training with real and synthetic images plus traditional augmentation methods (synthetic aug.\*). To conduct synthetic image augmentation, nucleus-like polygons are generated as the synthetic nuclei instance masks[13], 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 using the MoNuSeg test set. Table 3 demonstrates that, without adding more real training images, the performance of SegNet could be significantly improved by adding synthetic images during the training.

Methods	FID↓	SSIM↑	PQ↑
Baseline	134.6	0.488	0.523
INST	127.9	0.497	0.576
STYLE	121.4	0.509	0.582
Style Loss	130.1	0.493	0.549
SIAN	<b>115.7</b>	<b>0.515</b>	<b>0.586</b>

Table 3: The effectiveness of SIAN components.

Training Set	DQ	SQ	PQ
real images	0.704	0.737	0.521
real images*	0.725	0.738	0.533
synthetic augmentation	0.742	0.740	0.547
synthetic augmentation*	<b>0.748</b>	<b>0.742</b>	<b>0.555</b>

Table 4: Performance of SegNet using different training sets. \* denotes the training set augmented using the traditional augmentation techniques

## 5. Conclusion

In this paper, we propose a novel approach for histopathology image synthesis, i.e., style-guided instance-adaptive normalization (SIAN), which integrates instance layouts and style vectors into a deep image generator network. SIAN could synthesize histopathology images that align with image styles of multiple organs. SIAN utilizes the directional and distance masks from nuclei instance map and could generate 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 training set and improving the overall performance of a deep learning model for nuclei segmentation.

## References

- [1] 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*, 39(12):2481–2495, 2017.
- [2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.
- [3] 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)*, pages 1–5. IEEE, 2022.
- [4] Hao Chen, Xiaojuan Qi, Lequan Yu, and Pheng-Ann Heng. Dcan: deep contour-aware networks for accurate gland segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2487–2496, 2016.
- [5] Liang-Chieh Chen, Alexander Hermans, George Papandreou, Florian Schroff, Peng Wang, and Hartwig Adam. Masklab: Instance segmentation by refining object detection

- with semantic and direction features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4013–4022, 2018.
- [6] Yizhou Chen, Xu-Hua Yang, Zihan Wei, Ali Asghar Heidari, Nenggan Zheng, Zhicheng Li, Huiling Chen, Haigen Hu, Qianwei Zhou, and Qiu Guan. Generative adversarial networks in medical image augmentation: a review. *Computers in Biology and Medicine*, page 105382, 2022.
  - [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*, 77:102337, 2022.
  - [8] Xuan Gong, Shuyan Chen, Baochang Zhang, and David Doermann. Style consistent image generation for nuclei instance segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3994–4003, 2021.
  - [9] Simon Graham, Hao Chen, Jevgenij Gamper, Qi Dou, Pheng-Ann Heng, David Snead, Yee Wah Tsang, and Nasir Rajpoot. Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical image analysis*, 52:199–211, 2019.
  - [10] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019.
  - [11] Hongliang He, Zhongyi Huang, Yao Ding, Guoli Song, Lin Wang, Qian Ren, Pengxu Wei, Zhiqiang Gao, and Jie Chen. Cdnet: Centripetal direction network for nuclear instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4026–4035, 2021.
  - [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. *Advances in neural information processing systems*, 30, 2017.
  - [13] Le Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M Kurc, Rajarsi R Gupta, and Joel H Saltz. Robust histopathology image analysis: To label or to synthesize? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8533–8542, 2019.
  - [14] 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*, pages 1501–1510, 2017.
  - [15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.
  - [16] 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*, pages 1125–1134, 2017.
  - [17] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
  - [18] 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*, pages 4401–4410, 2019.
  - [19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.
  - [20] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.
  - [21] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9404–9413, 2019.
  - [22] Neeraj Kumar, Ruchika Verma, Deepak Anand, et al. A multi-organ nucleus segmentation challenge. *IEEE Transactions on Medical Imaging*, 39(5):1380–1391, 2020.
  - [23] Wenyuan Li, Jiayun Li, Jennifer Polson, Zichen Wang, William Speier, and Corey Arnold. High resolution histopathology image generation and segmentation through adversarial training. *Medical Image Analysis*, 75:102251, 2022.
  - [24] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.
  - [25] David Morrison, David Harris-Birtill, and Peter D Caie. Generative deep learning in digital pathology workflows. *The American Journal of Pathology*, 191(10):1717–1723, 2021.
  - [26] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2):448–459, 2018.
  - [27] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2337–2346, 2019.
  - [28] Eman Rezk, Zainab Awan, Fahad Islam, Ali Jaoua, Somaya Al Maadeed, Nan Zhang, Gautam Das, and Nasir Rajpoot. Conceptual data sampling for breast cancer histology image classification. *Computers in biology and medicine*, 89:59–67, 2017.
  - [29] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2287–2296, 2021.
  - [30] Muhammad Shaban, Ruqayya Awan, Muhammad Moazam Fraz, Ayesha Azam, Yee-Wah Tsang, David Snead, and Nasir M Rajpoot. Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE transactions on medical imaging*, 39(7):2395–2405, 2020.

- [31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [32] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.
- [33] Quoc Dang Vu, Simon Graham, Tahsin Kurc, Minh Nguyen Nhat To, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram, Jayashree Kalpathy-Cramer, Tianhao Zhao, et al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology*, page 53, 2019.
- [34] Haotian Wang, Aleksandar Vakanski, Changfa Shi, and Min Xian. Bend-net: Bending loss regularized multitask learning network for nuclei segmentation in histopathology images. *arXiv preprint arXiv:2109.15283*, 2021.
- [35] 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)*, pages 1–5. IEEE, 2020.
- [36] 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*, pages 1556–1564, 2022.
- [37] Pei Wang, Yijun Li, Krishna Kumar Singh, Jingwan Lu, and Nuno Vasconcelos. Imagine: Image synthesis by image-guided model inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3681–3690, 2021.
- [38] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.
- [39] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.
- [40] 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*, 13(4):600–612, 2004.
- [41] Jerry Wei, Arief Suriawinata, Bing Ren, Xiaoying Liu, Mikhail Lisovsky, Louis Vaickus, Charles Brown, Michael Baker, Mustafa Nasir-Moin, Naofumi Tomita, et al. Learn like a pathologist: curriculum learning by annotator agreement for histopathology image classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2473–2483, 2021.
- [42] Jerry Wei, Arief Suriawinata, Louis Vaickus, Bing Ren, Xiaoying Liu, Jason Wei, and Saeed Hassanpour. Generative image translation for data augmentation in colorectal histopathology images. *Proceedings of machine learning research*, 116:10, 2019.
- [43] Yuan Xue, Jiarong Ye, Qianying Zhou, L Rodney Long, Sameer Antani, Zhiyun Xue, Carl Cornwell, Richard Zaino, Keith C Cheng, and Xiaolei Huang. Selective synthetic augmentation with histogan for improved histopathology image classification. *Medical image analysis*, 67:101816, 2021.
- [44] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7354–7363. PMLR, 09–15 Jun 2019.
- [45] Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5104–5113, 2020.
- [46] Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5104–5113, 2020.