VSGD-Net: Virtual Staining Guided Melanocyte Detection on Histopathological Images

Kechun Liu[1*] Beibin Li[1,6] Wenjun Wu[1] Caitlin May[2]
Oliver Chang[1,3] Stevan Knezevich[4] Lisa Reisch[1] Joann Elmore[5] Linda Shapiro[1]

[1] University of Washington [2] Dermatopathology Northwest [3] VA Puget Sound
[4] Pathology Associates [5] University of California, Los Angeles [6] Microsoft Research

Abstract

Detection of melanocytes serves as a critical prerequisite in assessing melanocytic growth patterns when diagnosing melanoma and its precursor lesions on skin biopsy specimens. However, this detection is challenging due to the visual similarity of melanocytes to other cells in routine Hematoxylin and Eosin (H&E) stained images, leading to the failure of current nuclei detection methods. Stains such as Sox10 can mark melanocytes, but they require an additional step and expense and thus are not regularly used in clinical practice. To address these limitations, we introduce VSGD-Net, a novel detection network that learns melanocyte identification through virtual staining from H&E to Sox10. The method takes only routine H&E images during inference, resulting in a promising approach to support pathologists in the diagnosis of melanoma. To the best of our knowledge, this is the first study that investigates the detection problem using image synthesis features between two distinct pathology stainings. Extensive experimental results show that our proposed model outperforms state-of-the-art nuclei detection methods for melanocyte detection. The source code and pre-trained model are available at: `https://github.com/kechunl/VSGD-Net`

1. Introduction

In biomedical image analysis, the automatic detection of certain types of cells in microscopy images is of significant interest to a broad spectrum of biological research and clinical practices. Accurate identification of particular cell types helps to interpret biopsies and to diagnose the states of different diseases. For example, the diagnosis of melanoma, the most serious type of skin cancer in the United States [41], requires the assessment of the distribution disorder of melanocytes[1] under the microscopic examination of Hematoxylin and Eosin (H&E)-stained glass slides of skin biopsies by pathologists. Nevertheless, identifying melanocytic populations can be challenging on routine H&E-stained slides given the visual similarity with other cells. As a solution to this, pathologists may rely on obtaining special additional immunohistochemistry (IHC) stains, for example, Sox10 – a transcription factor expressed in melanocytic nuclei – as a specific immunomarker to highlight melanocytes (Fig. 2c). Despite this benefit, Sox10 immunostaining is not routinely obtained in clinical practice because of its high cost, especially in some low-resource regions. Hence, building computer-aided melanocyte detection methods would support the melanoma diagnosis workload and improve diagnostic accuracy.

In the last decade, benefiting from the development of deep learning techniques, researchers have leveraged deep convolutional neural networks (CNNs) with various model designs to tackle many computer vision tasks, including semantic segmentation and instance detection. As a part of instance detection, a major line of work utilizes deep convolutional neural networks (CNNs) [8, 12], U-Net [3], R-CNN [43], shape-guided CNN [37], and high-resolution networks [5] to localize general nuclei on H&E images. Similar CNN structures can also be found in specific types of cell/nuclei detection studies, such as mitotic nuclei detection [31, 42] and tumor nuclei grading [5, 27, 8, 34]. However, unlike general nuclei/cell detection, the detection of a specific class of cells is more challenging because of the inter-class visual similarity on routine H&E-stained slides. Although IHC staining can highlight certain types of cells, it is not comparable to H&E staining in terms of generalizability because of the difficult accessibility issue. Learning from only H&E-stained slides, the aforementioned CNN-based detection methods are not capable of incorporating information

*kechun@cs.washington.edu

[1] For example, melanoma in situ exhibits confluent growth of single and nested melanocytes at the epidermal base and/or extension into the mid-to-upper levels of the epidermis.

from other modalities/stainings and in this way fail to differentiate various classes of cells.

Recently, Generative Adversarial Networks (GANs) have been used for data augmentation and style transfer. In the biomedical research community, GANs also attract growing attention for virtual staining and realistic medical image synthesis to aid clinical practices. For example, researchers leverage the unsupervised CycleGAN [55] architecture and the supervised conditional GAN [13] to synthesize one modality into another, e.g. MR to CT [49, 11], H&E to IHC [48, 29]. However, there is still a gap between synthesizing convincing medical images and boosting the performance of downstream tasks. In other words, a generator cannot be trained for a specific downstream task for lack of direct feedback from another network. To make up for this issue, some studies [6, 52] cascade a segmentation net after the generator and train the network in an end-to-end style. But these methods fail to explore intermediate features from the image synthesis process, which are empirically important for the downstream tasks.

In order to aid in the pathologists' decision-making process, it is very useful to have either accurate melanocyte prediction or precise virtual staining. **Motivated by the demand, we propose *VSGD-Net*, a novel virtual-staining-guided detection architecture that provides a solution to both the detection and the virtual staining tasks simultaneously**. *VSGD-Net* boosts detection and image synthesis performance at the same time by incorporating hidden correlations between two image modalities. In Fig. 4, we illustrate our proposed model, which expands a conditional GAN to an instance detection pipeline. The generator, discriminator, and detection network are jointly trained so that the image synthesis task and the detection task can benefit from each other. We validate our approach with a carefully curated melanocyte dataset that contains biopsy images in H&E and Sox10 stainings. Moreover, we verify the significance of the intermediate features with extensive experiments. Our contributions in this work can be summarized as follows:

1. We propose *VSGD-Net* for the instance detection task. To the best of our knowledge, this work is the first to investigate the detection problem using image synthesis features between two stainings. From an information system perspective, the added modality increases information entropy and facilitates feature learning through adversarial training.

2. We compare our model with previous nuclei detection and GAN-based methods in a melanocyte detection dataset. Extensive experiments show that our model achieves the state-of-the-art performance.

3. During inference time, the proposed *VSGD-Net* takes only an affordable regular H&E stain as input to identify melanocyte instances. As one of the first deep-learning-based melanocyte detection methods, the proposed model would provide reliable melanocyte results to reduce the burden on pathologists and aid in melanoma diagnosis in the future.

2. Related Work

2.1. Nuclei Detection

In recent years, deep learning-based nuclei detection methods have been widely studied. As a variant of the fully convolutional network (FCN) [22], U-Net [36] made a huge impact on the medical image research community. Many researchers extended the U-Net structure [36] into more efficient variants to identify nuclei in histopathological images, for example, R2U-Net [1], U-Net++[54], Micro-Net [35], and Triple U-Net [51]. To incorporate nuclei contour-aware modules, Zhou *et al*. presented CIA-Net [53] which contains two task-specific decoders to learn either the nuclei or the contours. Similarly, Schmidt *et al*. proposed StarDist [37] to localize nuclei via star-convex polygons. In the task of detecting nuclei of specific cells, Graham *et al*. proposed Hover-Net [8] by utilizing three downstream branches, namely segmentation, classification, and a novel Hover branch, which used the horizontal and vertical distance maps to segment attached nuclei. For better distance-map generation, Gao *et al*. presented the two-stage CHR-Net [5], which leveraged the W-Net structure [47] and high-resolution feature extractors, and achieved the new state-of-the-art performance.

Another line of approaches, e.g. Mask RCNN [9], have also achieved promising results in nuclei instance segmentation [21, 43, 44]. The feature pyramid network (FPN) backbone allows the model to extract features in multiple scales and feed into the region proposal network (RPN) to generate reasonable instance candidates in varying sizes for downstream tasks like segmentation and classification. Our proposed model, *VSGD-Net*, also takes advantage of the FPN and RPN modules to better exploit the intermediate features for nuclei detection.

2.2. Image-to-Image Translation

First proposed by Goodfellow *et al*., the Generative Adversarial Network (GAN) [7] introduces the adversarial loss to optimize the generator and the discriminator in a minimax zero-sum game. To incorporate additional constraints on the generated data, Mirza *et al*. proposed the conditional GAN (cGAN) [30], which feeds the condition to both the generator and the discriminator to guide the generation process. Successful variants of the cGAN include the LSGAN [26], the ACGAN [32], the BigGAN [2], and Pix2Pix [13]. Among these prominent variants, Pix2Pix [13] first brought the cGAN to the paired image-to-image translation task and

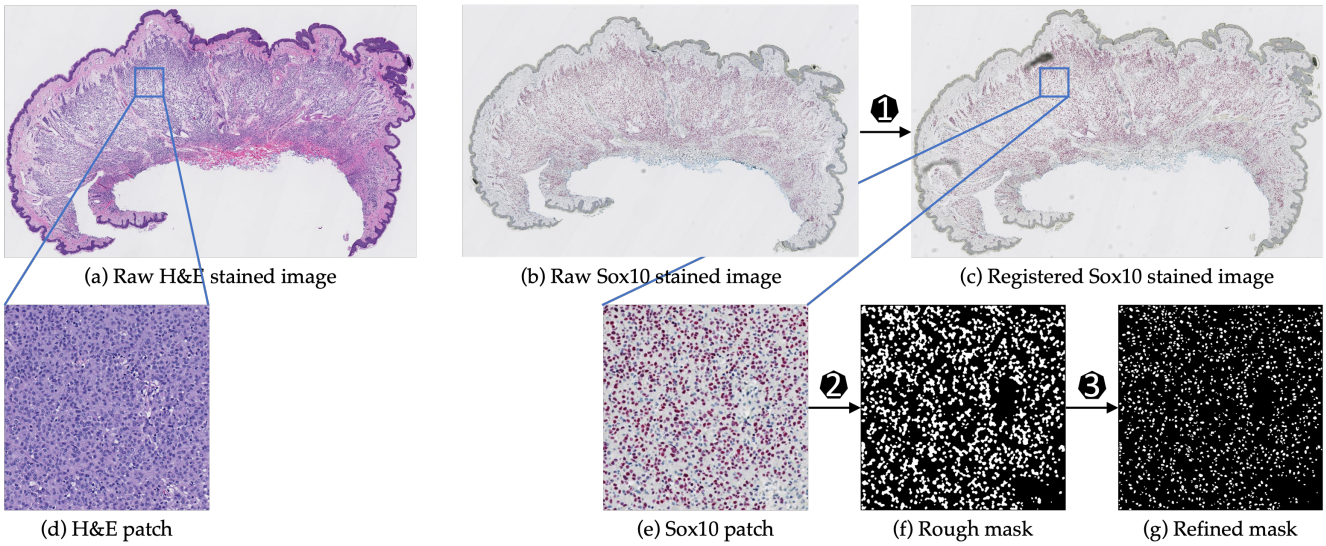


Figure 1. Preprocessing steps: First, we register raw Sox10 images (b) into aligned Sox10 images (c) using template H&E images (a) with the Histokat software[3][23]. Then, we apply a Random Forest classifier to classify pixels into melanocyte or non-melanocyte. At last, the pretrained NuSeT [50] separates touching nuclei and refine the masks.

its extension, Pix2PixHD [45] enabled high resolution image generation. To alleviate the need for paired data, Zhu *et al*. proposed CycleGAN[55] to learn the mapping between two image domains X and Y in both directions by coupling two GANs. The idea behind Cycle GAN is that ideally if we translate the image from one domain to another and back again, the reconstructed image should be the same as the input image. The CycleGAN structure has also been widely applied in stain normalization, modality conversion, and virtual staining for histopathological images. For instance, Shaban *et al*. developed Stain-GAN [39] based on the CycleGAN structure for biopsy stain normalization. Mahmood *et al*. leveraged CycleGAN to learn the mapping between histopathology images and nuclei masks to improve nuclei segmentation [25]. Xu *et al*. developed cCGAN [48] that incorporated CycleGAN with photorealism and structure similarity losses to learn virtual staining from H&E to IHC. However, the cycle consistency loss in CycleGAN only forces the reconstructed image to be similar to the original image, lacking some constraints between the two image domains, which weakens its reliability in virtual staining. To solve this, Liu *et al*. adds a pathology-consistency constraint to CycleGAN and requires the generated and source images to have the same pathological properties in both H&E and IHC stains [20].

To benefit from GANs, some studies utilize the synthesized data to enhance the performance on downstream tasks such as detection and segmentation. A R-CNN-based detector is cascaded after the generator to learn nuclei segmentation [18, 6] and disease localization [52, 19]. However, these models fail to exploit the informative hidden features during the generation, and the feedback from the downstream tasks may only yield minor improvements in the image synthesis process. To this end, we propose *VSGD-Net* that can jointly optimize the image synthesis and cell-type-of-interest detection via the shared intermediate features. The improvements in both the synthesis and the detection tasks are validated through our comprehensive experiments.

3. Methodology

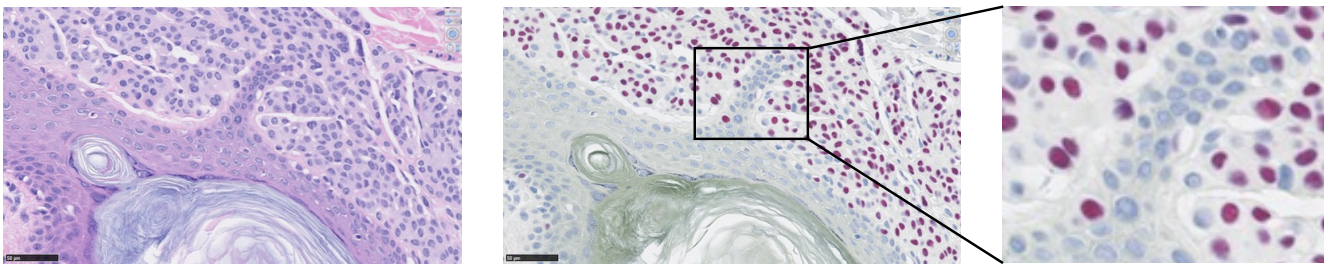
In this section, we explain the data preprocessing (Fig. 1), the design of our proposed *VSGD-Net*, and the training procedure; the following section will examine various methods and ablate *VSGD-Net*'s components to show its performances and design decisions.

3.1. Dataset

The skin biopsy dataset used in this study consists of skin tissue from paraffin-embedded blocks of 15 cases, which were chosen at random among historical cases from a private dermatopathology laboratory, including three cases for each MPATH-Dx diagnostic category[4, 33][4]. The tissue from each skin biopsy case is cut into multiple (4-6) thin slices for microscopic examination, resulting in 75 slices in 20x magnification. We stain each WSI with H&E first (see Fig. 2a). We then carefully destain the exact tissue sections and re-stain them in Sox10. The Sox10 stain highlights the nuclei of melanocytes in red, while the nuclei of

[3] https://histoapp.mevis.fraunhofer.de/

[4] Classes 1-5: Benign mildly atypical nevi, Moderate dysplastic nevi, Melanoma in situ, Invasive melanoma T1a, and Invasive melanoma T1b.



(a) H&E Staining (b) Sox10 Staining – melanocytes are red (c) Crop from Sox10

Figure 2. Sample H&E stained image and Sox10 stained image. The Sox10 stain highlights the nuclei of melanocytes in red, while the nuclei of other cells appear in blue.

other cells appear in blue, which provides the ground truth label of melanocytes and non-melanocytes (see Fig. 2b).

To generate ground truth labels for melanocyte detection, we introduce a pseudo-automatic procedure. We trained a Random Forest classifier on 100 manually labeled melanocytes in Sox10 to generate coarse melanocyte masks. Then, we applied a pretrained nuclei detection model, NuSeT [50], to separate touching nuclei and refine the masks. We find that this procedure yields accurate melanocyte masks, which can serve as ground truth labels in this study (see Fig. 3).

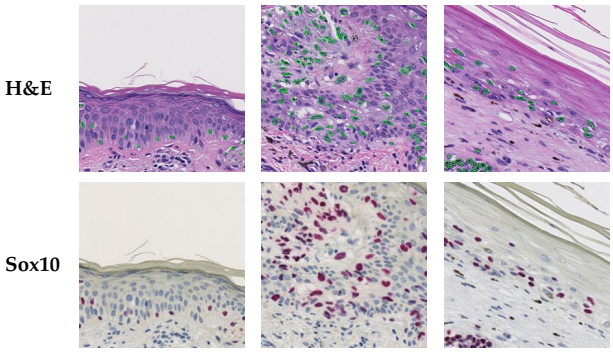


Figure 3. Nuclei groundtruth: The top row shows the H&E images with melanocytes marked with green boundaries. The bottom row shows their corresponding Sox10-stained images.

To fit images into memory as well as keep adequate information, we cropped the registered paired images into 256x256 patches with 10x magnification. The background patches were excluded, leaving a total of 25,314 patches to use. We reserved 9652 paired image patches from 5 patients for the testing set and the rest for training and validation, where data from patients in the testing set never appeared in the training and validation sets. Both the training and testing sets contain the full range of MPATH-Dx diagnostic classes for a fair evaluation.

3.2. Model Architecture

Fig. 4 illustrates the *VSGD-Net* architecture. We built the generator G based on an adapted UNet [36] structure with ResNet-50 [10] being the encoder. The encoder learns the high-dimensional feature representations of input H&E images in multiple scales, and the decoder translates them into target Sox10 stained images. Given the 2^5x downsampling in the encoder, the decoder comprises 5 deconvolution layers. To better focus on melanocytes without expanding the model architecture, we incorporated attention blocks in the skip connections between the encoder and the decoder. The attention blocks leverage the design of CBAM [46], which contains a 3-layer MLP channel attention block and a convolutional spatial attention block to learn the attention maps in different dimensions (see Fig. 5).

While the generator G learns the virtual staining process, the discriminator D attempts to differentiate real and synthesized Sox10 images. Inspired by Pix2PixHD [45], we adopted a multi-scale architecture that has 2 identical CNN networks as discriminators: the two discriminators work at coarse and fine levels separately, where the input to the coarse-level discriminator is downsampled by a factor of 2 from the input to the fine-level discriminator. Similar to PatchGAN [13], each discriminator evaluates the realism of every fixed-sized patch in the image instead of directly evaluating the realism of the whole image. With the minimax loss introduced in [7], this multi-scale design guides G to synthesize images with globally consistent patterns as well as finer details. The architectural details of the attention block and the discriminator are explained in the supplementary material.

Similar to Mask R-CNN [9], our detection branch consists of a feature pyramid network (FPN), a region proposal network (RPN), and the downstream heads. Learning to generate Sox10 images, the decoder layers have higher correlations with the Sox10 images than the encoder layers; moreover, Sox10 staining can highlight melanocytes in a red chromogenic color, which is consistent with the detection goal. In light of this, we place the detection branch in

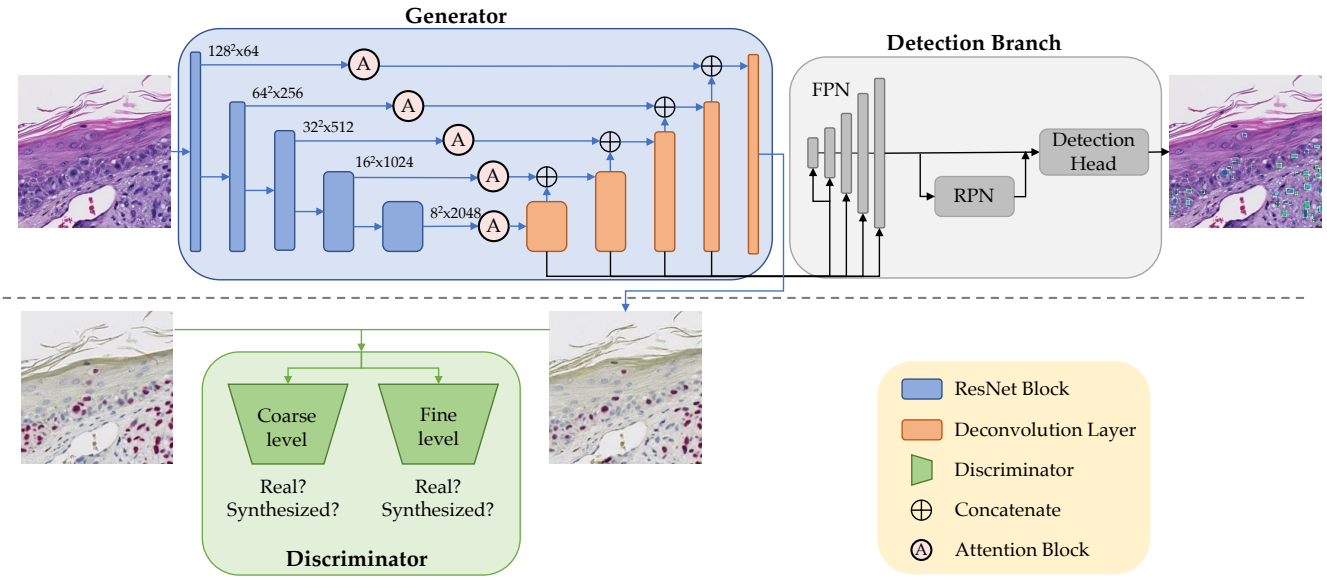


Figure 4. Our *VSGD-Net* framework: H&E images are virtually stained to Sox10. The jointly trained detection branch utilizes the intermediate features in the generator to detect melanocytes and provides feedback to the generator to enhance synthesis quality. The inference phase only uses the upper part of the architecture.

the decoder of G instead of the encoder, which is proven to be effective in the ablation study.

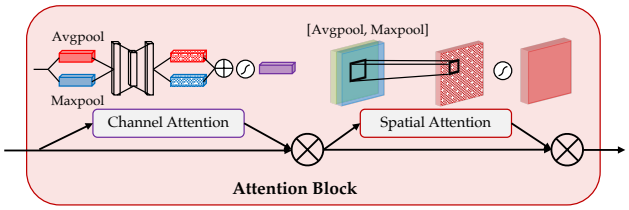


Figure 5. Attention block: Channel attention and spatial attention are consecutively computed to refine the features.

3.3. Training Process

In our end-to-end model, the virtually stained images and the detected instances are predicted from the shared intermediate features. To incorporate the feedback from both the image synthesis and the instance detection, we train G, D, and the detection branch jointly to learn from both the GAN loss L_{GAN} and the detection loss L_{DET}.

3.3.1 GAN Loss

The generator G and the multi-scale discrimator D are optimized following the minimax loss [7]:

$$\min_G \max_D \sum_{i=1,2} \left(\log(D_i(X_s)) + \log(1 - D_i(G(X_h))) \right)$$

where D_1 and D_2 are the coarse- and fine-level discriminators, and X_s and X_h are the Sox10 and H&E images.

Besides the minimax loss, we add a feature similarity loss L_{feat} to improve the similarity between the generated and the real images. The calculation of L_{feat} involves multiple layers in D and a pretrained VGG19 model, and is given by the following equation:

$$L_{feat} = \sum_{i=1}^{N} ||D_i(X_s) - D_i(G(X_h))||_1 + \sum_{j=1}^{M} ||VGG_j(X_s) - VGG_j(G(X_h))||_1$$

where N and M denote the layers to extract features. The details of feature similarity loss is provided in the supplementary material.

3.3.2 Detection Loss

The detection loss L_{DET} is separated into L_{rpn}, L_{box_c}, L_{box_r}, and L_{seg}. L_{rpn} the total loss of the candidate classification and the coarse bounding box regression in the RPN, given by the summation of binary cross entropy of the candidate classification and L1 loss on the coarse bounding box regression in the RPN. It forces the RPN to learn the location of anchor boxes and whether the anchor boxes contain objects. L_{box_c}, L_{box_r}, and L_{seg} are the losses for the instance classification, the final bounding box regression, and the segmentation in the downstream heads, which are given by the binary cross entropy of the instance classification, the binary cross entropy of the mask prediction, and the L1 loss

of bounding box coordinates. The total loss is defined as:

$$L_{DET} = L_{rpn} + L_{box_c} + L_{box_r} + L_{seg}$$

3.3.3 Overall Losses and Training

In our *VSGD-Net*, the shared intermediate features are learned to characterize features of melanocytes and boost the Sox10 image synthesis at the same time. To facilitate such multi-task learning, we combine L_{GAN} with L_{DET} and backpropagate them to the encoder inside G. The final total loss is defined below,

$$\min_G \big(\max_D \sum_{i=1,2} \big(\log(D_i(X_s)) + \log(1 - D_i(G(X_h))) \big) + \lambda * L_{feat} + L_{DET} \big) \tag{1}$$

4. Experiments and Results

4.1. Experimental Design and Baseline Methods

To comprehensively evaluate the performance of our proposed *VSGD-Net*, we compared *VSGD-Net* with two lines of methods. The first group is specialized in nuclei detection, including Radial Line Scanning (RLS)[24], Mask R-CNN[9], U-Net[36], StarDist[37], HoverNet[8], the new state-of-the-art CHR-Net[5], and a "nuclei classification" method we designed. RLS was specifically proposed to study melanocyte detection. It leverages a feature-based approach based on the "halo region" assumption that melanocytes appear with a brighter region surrounding the nuclei under H&E staining. Furthermore, to investigate the local texture around nuclei, we designed the "nuclei classification" method, which first applies a fine-tuned ensemble model [38] to detect nuclei and then trains the open-source ESPNetv2[28] to classify cropped nuclei patches.

The second group of methods consists of GAN-based approaches, including StainGAN [39], PC-StainGAN [20], and a self-implemented GAN-based segmentation model similar to [6]. The segmentation model, whose G and D are the same as *VSGD-Net*, directly feeds the synthesized image to the segmentation net and is trained end-to-end. For the other GAN models that do not incorporate any downstream modules, we tested their performances in a two-stage manner, using the random forest and the NuSeT model in our groundtruth-generating step (Section 3.1).

In our experiments, the ResNet-50 backbone in Mask R-CNN and the ResNet-34 backbone in CHR-Net are pre-trained with ImageNet for fair comparisons. We empirically set $\lambda = 10$ in Eq. 1. We report precision (P), recall (R), F_1-score, and Jaccard index on the test set in our experiments. More training details are explained in the supplementary material.

4.2. Main Results

In clinical practice, pathologists diagnose and grade melanoma based on the distribution of melanocytes, hence it is important to have both high precision and recall. With high precision but low recall, malignant melanocytes may be missed, leading to under-diagnosis of melanoma. On the other hand, a case may be over-diagnosed with high recall but low precision. Thus the F_1-score and Jaccard index are the most significant metrics. More detailed analyses, such as precision-recall curve and P@R metrics, are included in the Appendix for reference.

Table 1. Comparison with nuclei detection methods.

Method	**P**	**R**	F_1	**Jaccard**
RLS [24]	0.443	0.570	0.499	0.332
Nuclei Classification	0.693	0.506	0.585	0.413
Mask R-CNN [9]	0.735	0.514	0.605	0.434
U-Net [36]	0.630	0.639	0.635	0.465
StarDist[37]	**0.745**	0.426	0.542	0.372
HoverNet[8]	0.729	0.499	0.592	0.421
CHR-Net [5]	0.607	0.688	0.645	0.476
Ours	0.660	**0.710**	**0.684**	**0.520**

Table 2. Comparison with GAN-based methods.

Method	**P**	**R**	F_1	**Jaccard**
StainGAN [39]	0.476	0.299	0.367	0.225
PC-StainGAN [20]	0.591	0.343	0.434	0.277
GAN-based Segmentation	0.569	**0.719**	0.636	0.466
Ours	**0.660**	0.710	**0.684**	**0.520**

As shown in Table 1, *VSGD-Net* achieves the best F_1-score and Jaccard index. RLS, although it heuristically utilizes the "halo region" characteristics of melanocytes, demands a huge workload in hyperparameter tuning and lacks generalizability in this way. Both "Nuclei Classification" and Mask R-CNN show high precision but low recall, because they only predict instances with high confidence scores under the instance-level learning schema. Given the shape similarity between melanocytes and other cells, StarDist and HoverNet fail to utilize the shape representation and the distance map of nuclei. Benefiting from the skip connections, U-Net reaches a decent result. Furthermore, the CHR-Net leverages a double U-Net structure and high-resolution feature extractors to achieve a 1% improvement over U-Net, which is consistent with the previous findings[5]. However, without learning from Sox10 staining, U-Net and CHR-Net still underperform *VSGD-Net*.

Figure 6 shows the qualitative comparisons of *VSGD-Net*, CHR-Net, and GAN-based segmentation. The pre-

dictions in *VSGD-Net* have a high coincidence with the ground truth, while CHR-Net over-predicts the melanocytes on the bottom-left of the image, and GAN-based segmentation over-predicts the melanocytes on the top of the image. More qualitative visualizations are provided in the supplementary material.

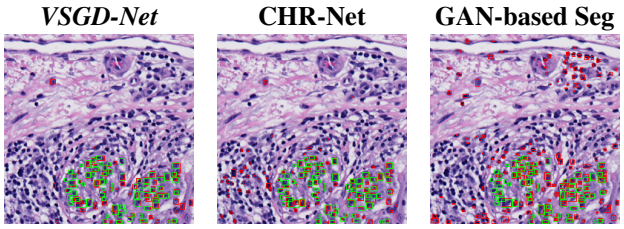


Figure 6. The green and red bounding boxes denote the groundtruth and the predicted instances. (Zoom in for best view)

Table 2 and Figure 7 demonstrate the performance of GAN-based methods. StainGAN [39] and PC-StainGAN [20] were designed based on unsupervised CycleGAN [55]. Without any additional supervision, StainGAN fails to learn the distribution gap between the two stainings. Although PC-StainGAN adds a pathology constraint to the CycleGAN, it still lacks supervision on the conversion between H&E and Sox10. On the other hand, the GAN-based segmentation method has supervision on the synthesized images, but its detection performance is bounded by the image synthesis quality due to its architecture.

4.3. Image Synthesis Evaluation

Although image synthesis is only auxiliary in our *VSGD-Net* framework, we still evaluate its quality to show that the virtual staining is improved by the shared intermediate features. To measure the reliability of the virtual staining, we calculate the average Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM). Larger numbers in PSNR and SSIM indicate better image quality and higher similarity with the groundtruth. As Table 3 shows, our *VSGD-Net* achieves the highest PSNR and a comparable SSIM to PC-StainGAN. By assessing the mean squared error of the synthesized images, higher PSNR indicates more reliable results with regard to the virtual staining task.

Table 3. Synthesized image quality assessment.

Method	**PSNR**(dB)	**SSIM**
StainGAN [39]	19.010	0.577
PC-StainGAN [20]	19.344	**0.618**
GAN-based Segmentation	19.583	0.569
Ours	**19.815**	0.611

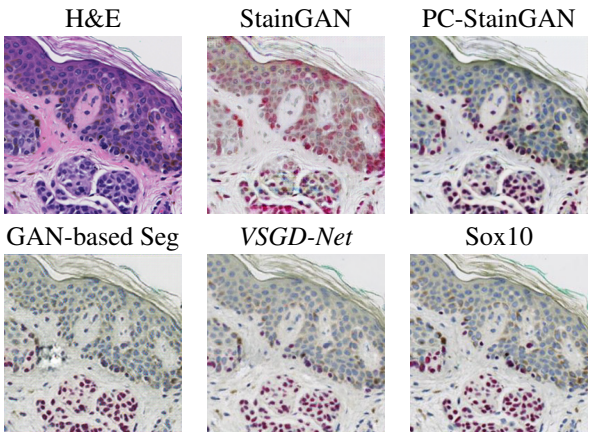


Figure 7. Synthesized Sox10 images.

4.4. Ablation Study

In Table 4, we ablated each key component in *VSGD-Net*, namely the image synthesis features, the location of the detection branch and the attention module's presence. To verify the efficacy of the image synthesis features, we replaced the generator of *VSGD-Net* with the generator in Pix2PixHD[45], which has fewer convolution layers, no skip connections, and no attention module. As Row 1 of Table 4 shows, despite the weakness of the Pix2PixHD generator, it still achieves comparable results and outperforms other baselines with the key component of boosting detection with image synthesis features. We assumed the features in the decoders have higher correlations with Sox10 staining and melanocytes, and the attention module refines the intermediate features. Such assumptions are verified by the notable performance gains in Table 4 row 5.

Table 4. Ablation results.

Generator	**Features From**	**Atten.**	F_1	**Jaccard**
Pix2pixHD	Decoder	-	0.654	0.486
Ours	Encoder	✗	0.641	0.472
Ours	Decoder	✗	0.674	0.508
Ours	Encoder	✓	0.660	0.492
Ours	Decoder	✓	**0.684**	**0.520**

4.5. Discussion

The *VSGD-Net* successfully detects melanocytes using the features from image synthesis between H&E and Sox10 stainings. Considering the large quantity of melanocytes (*e.g.* total number range from 3,780 to 830,750 per WSI) on a single segment of a skin biopsy, it is not feasible to label melanocytes manually for training. While the pseudo ground truth labels are not perfect, it is sufficient to provide highly accurate annotation given how Sox10 staining works in skin biopsies. One limitation is that we utilize a simple U-Net with a ResNet-50 backbone as our generator. With more recent works studying GANs on histopathology

images [17, 40], we believe the synthesis features can be further improved by state-of-the-art GAN models. Another consideration is that we only evaluate *VSGD-Net* on the melanocyte dataset. Although researchers have publicized some multi-modality medical imaging datasets for image synthesis study, such as CT-MRI [14], PET-MRI [15], and H&E-Trichrome staining [16], these datasets do not have any annotations on lesions or cell-type-of-interest. In the future, researchers can add pathologists' annotations or leverage self-supervised learning to overcome these issues.

5. Conclusion

In this study, we introduce a novel virtual staining guided detection network, *VSGD-Net*, and investigate cell-type-of-interest detection with the boost of image synthesis features between two distinct stainings on the skin biopsy specimen. During inference, the model can produce promising results from only the routine H&E staining. Extensive experiments validate the effectiveness of our method on a corresponding dataset of melanocytes in H&E and Sox10 stained images. We anticipate that the proposed method can adapt to a broad category of different tissue types and diseases.

Acknowledgment: The research reported in this study was supported by Grants R01CA151306, R01CA200690, and U01CA231782 from the National Cancer Institute of the National Institutes of Health, 622600 from the Melanoma Research Alliance, and W81XWH-20-1-0797 and W81XWH-20-1-0798 from the US Department of Defense. The funders had no role in the design and conduct of the study; collection; management; analysis; and interpretation of the data, preparation, review, or approval of the manuscript, nor the decision to submit the manuscript for publication.

References

[1] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv:1802.06955 [cs]*, May 2018. arXiv: 1802.06955.

[2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[3] Kemeng Chen, Ning Zhang, Linda Powers, and Janet Roveda. Cell Nuclei Detection and Segmentation for Computational Pathology Using Deep Learning. In *2019 Spring Simulation Conference (SpringSim)*, pages 1–6, Apr. 2019.

[4] Joann G Elmore, Raymond L Barnhill, David E Elder, Gary M Longton, Margaret S Pepe, Lisa M Reisch, Patricia A Carney, Linda J Titus, Heidi D Nelson, Tracy Onega, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *Bmj*, 357, 2017.

[5] Zeyu Gao, Jiangbo Shi, Xianli Zhang, Yang Li, Haichuan Zhang, Jialun Wu, Chunbao Wang, Deyu Meng, and Chen Li. Nuclei Grading of Clear Cell Renal Cell Carcinoma in Histopathological Image by Composite High-Resolution Network. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture Notes in Computer Science, pages 132–142, Cham, 2021. Springer International Publishing.

[6] Xuan Gong, Shuyan Chen, Baochang Zhang, and David Doermann. Style Consistent Image Generation for Nuclei Instance Segmentation. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3993–4002, Waikoloa, HI, USA, Jan. 2021. IEEE.

[7] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. In *Advances in neural information processing systems*, volume Advances in neural information processing systems, June 2014. arXiv: 1406.2661.

[8] 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.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. *arXiv:1703.06870 [cs]*, Mar. 2017. arXiv: 1703.06870.

[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)*, Las Vegas, NV, USA, June 2016. IEEE.

[11] Yuta Hiasa, Yoshito Otake, Masaki Takao, Takumi Matsuoka, Kazuma Takashima, Aaron Carass, Jerry L Prince, Nobuhiko Sugano, and Yoshinobu Sato. Cross-modality image synthesis from unpaired data using cyclegan. In *International workshop on simulation and synthesis in medical imaging*, pages 31–41. Springer, 2018.

[12] Henning Höfener, André Homeyer, Nick Weiss, Jesper Molin, Claes F. Lundström, and Horst K. Hahn. Deep learning nuclei detection: A simple approach can deliver state-of-the-art results. *Computerized Medical Imaging and Graphics*, 70:43–52, Dec. 2018.

[13] 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.

[14] Cheng-Bin Jin, Hakil Kim, Mingjie Liu, Wonmo Jung, Seongsu Joo, Eunsik Park, Young Saem Ahn, In Ho Han, Jae Il Lee, and Xuenan Cui. Deep ct to mr synthesis using paired and unpaired data. *Sensors*, 19(10):2361, 2019.

[15] Florian Knoll, Martin Holler, Thomas Koesters, Ricardo Otazo, Kristian Bredies, and Daniel K Sodickson. Joint mr-pet reconstruction using a multi-channel image regularizer. *IEEE transactions on medical imaging*, 36(1):1–16, 2016.

[16] Joshua J Levy, Nasim Azizgolshani, Michael J Andersen, Arief Suriawinata, Xiaoying Liu, Mikhail Lisovsky, Bing Ren, Carly A Bobak, Brock C Christensen, and Louis J Vaickus. A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies. *Modern Pathology*, 34(4):808–822, 2021.

[17] Hanwen Liang, Konstantinos N Plataniotis, and Xingyu Li. Stain style transfer of histopathology images via structure-preserved generative learning. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 153–162. Springer, 2020.

[18] Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren O'Donnell, Heng Huang, Mei Chen, and Weidong Cai. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4243–4252, 2020.

[19] Lanlan Liu, Michael Muelly, Jia Deng, Tomas Pfister, and Li-Jia Li. Generative modeling for small-data object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6073–6081, 2019.

[20] 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*, 40(8):1977–1989, 2021.

[21] Yiming Liu, Pengcheng Zhang, Qingche Song, Andi Li, Peng Zhang, and Zhiguo Gui. Automatic segmentation of cervical nuclei based on deep learning and a conditional random field. *IEEE Access*, 6:53709–53721, 2018.

[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[23] Johannes Lotz, Nick Weiss, Jeroen van der Laak, and StefanHeldmann. High-resolution Image Registration of Consecutive and Re-stained Sections in Histopathology. *arXiv:2106.13150 [cs, eess]*, June 2021. arXiv: 2106.13150.

[24] Cheng Lu, Muhammad Mahmood, Naresh Jha, and Mrinal Mandal. Detection of melanocytes in skin histopathological images using radial line scanning. *Pattern Recognition*, 46(2):509–518, Feb. 2013.

[25] Faisal Mahmood, Daniel Borders, Richard J Chen, Gregory N McKay, Kevan J Salimian, Alexander Baras, and Nicholas J Durr. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE transactions on medical imaging*, 39(11):3257–3267, 2019.

[26] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

[27] André Victória Matias, Allan Cerentini, Luiz Antonio Buschetto Macarini, João Gustavo Atkinson Amorim, Felipe Perozzo Daltoé, and Aldo von Wangenheim. Segmentation, detection, and classification of cell nuclei on oral cytology samples stained with papanicolaou. *SN Computer Science*, 2(4):1–15, 2021.

[28] Sachin Mehta, Mohammad Rastegari, Linda Shapiro, and Hannaneh Hajishirzi. ESPNetv2: A Light-Weight, Power Efficient, and General Purpose Convolutional Neural Network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9182–9192, Long Beach, CA, USA, June 2019. IEEE.

[29] Caner Mercan, Germonda Reijnen-Mooij, David Tellez Martin, J. Lotz, Nick Weiss, M. V. Gerven, and F. Ciompi. Virtual Staining for Mitosis Detection in Breast Histopathology. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020.

[30] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[31] Shima Nofallah, Sachin Mehta, Ezgi Mercan, Stevan Knezevich, Caitlin J May, Donald Weaver, Daniela Witten, Joann G Elmore, and Linda Shapiro. Machine learning techniques for mitoses classification. *Computerized Medical Imaging and Graphics*, 87:101832, 2021.

[32] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pages 2642–2651. PMLR, 2017.

[33] Michael W Piepkorn, Raymond L Barnhill, David E Elder, Stevan R Knezevich, Patricia A Carney, Lisa M Reisch, and Joann G Elmore. The mpath-dx reporting schema for melanocytic proliferations and melanoma. *Journal of the American Academy of Dermatology*, 70(1), 2014.

[34] Hui Qu, Gregory Riedlinger, Pengxiang Wu, Qiaoying Huang, Jingru Yi, Subhajyoti De, and Dimitris Metaxas. Joint segmentation and fine-grained classification of nuclei in histopathology images. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 900–904. IEEE, 2019.

[35] Shan E Ahmed Raza, Linda Cheung, Muhammad Shaban, Simon Graham, David Epstein, Stella Pelengaris, Michael Khan, and Nasir M Rajpoot. Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical image analysis*, 52:160–173, 2019.

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, May 2015. arXiv: 1505.04597.

[37] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 265–273. Springer, 2018.

[38] Selim Seferbekov. DSB2018 [ods.ai] topcoders 1st place solution, Feb. 2022. original-date: 2018-04-10T20:06:11Z.

[39] M. Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni. Staingan: Stain Style Transfer for Digital Histological Images. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019.

[40] Faezehsadat Shahidi. Breast cancer histopathology image super-resolution using wide-attention gan with improved

wasserstein gradient penalty and perceptual loss. *IEEE Access*, 9:32795–32809, 2021.

[41] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.

[42] Anabia Sohail, Asifullah Khan, Noorul Wahab, Aneela Zameer, and Saranjam Khan. A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images. *Scientific Reports*, 11(1):6215, Mar. 2021. Number: 1 Publisher: Nature Publishing Group.

[43] Aarno Oskar Vuola, Saad Ullah Akram, and Juho Kannala. Mask-RCNN and U-Net Ensembled for Nuclei Segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 208–212, Apr. 2019. ISSN: 1945-8452.

[44] Shidan Wang, Ruichen Rong, Donghan M Yang, Junya Fujimoto, Shirley Yan, Ling Cai, Lin Yang, Danni Luo, Carmen Behrens, Edwin R Parra, et al. Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer research*, 80(10):2056–2066, 2020.

[45] 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 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, Salt Lake City, UT, USA, June 2018. IEEE.

[46] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[47] Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*, 2017.

[48] Zhaoyang Xu, Carlos Fernández Moro, Béla Bozóky, and Qianni Zhang. GAN-based Virtual Re-Staining: A Promising Solution for Whole Slide Image Analysis. *arXiv:1901.04059 [cs]*, Jan. 2019. arXiv: 1901.04059.

[49] Heran Yang, Jian Sun, Aaron Carass, Can Zhao, Junghoon Lee, Zongben Xu, and Jerry Prince. Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 174–182. Springer, 2018.

[50] Linfeng Yang, Rajarshi P. Ghosh, J. Matthew Franklin, Simon Chen, Chenyu You, Raja R. Narayan, Marc L. Melcher, and Jan T. Liphardt. NuSeT: A deep learning tool for reliably separating and analyzing crowded cells. *PLOS Computational Biology*, 16(9):e1008193, Sept. 2020. Publisher: Public Library of Science.

[51] Bingchao Zhao, Xin Chen, Zhi Li, Zhiwen Yu, Su Yao, Lixu Yan, Yuqian Wang, Zaiyi Liu, Changhong Liang, and Chu Han. Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. *Medical Image Analysis*, 65:101786, 2020.

[52] Jianfeng Zhao, Dengwang Li, Zahra Kassam, Joanne Howey, Jaron Chong, Bo Chen, and Shuo Li. Tripartite-GAN: Synthesizing liver contrast-enhanced MRI to improve tumor detection. *Medical Image Analysis*, 63:101667, July 2020.

[53] Yanning Zhou, O. F. Onder, Q. Dou, E. Tsougenis, Hao Chen, and P. Heng. CIA-Net: Robust Nuclei Instance Segmentation with Contour-aware Information Aggregation. *IPMI*, 2019.

[54] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[55] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.