VSGD-Net: Virtual Staining Guided Melanocyte Detection on Histopathological Images (Supplementary material)

A. Implementation details

A.1. Training configuration

We used ImageNet pretrained ResNet-50 from torchvision as the backbone of our encoder. The ResNet-34 backbone in CHR-Net, is pre-trained with ImageNet and loaded from torchvision as well for fair comparisons. All experiments were implemented with Pytorch and performed on four NVIDIA RTX 2080Ti GPUs. Random flipping was applied during training. We used the Adam optimizer to train the G and D with an initial learning rate (lr) of 0.0002, and the SGD optimizer to train the detection branch with an initial lr of 0.01. The model was trained for 300 epochs and the detection lr was decayed by 0.5 every 30 epochs after the first 50, while the GAN lr was decayed to 0 with a constant gap in every epoch after the first 100.

A.2. GAN loss

The GAN loss is comprised of the minimax loss and the feature similarity loss (Section 3.3.1). The feature similarity loss is the L1 term of the features of real and fake data given by the discriminator and the pretrained VGG19. The features of the pretrained VGG19 model are the outputs of layers 1, 6, 11, 20, 29. The features of the discriminator are from the outputs of each of the convolutional blocks (feat_{0,1,2,3,4} in Fig. 1).

A.3. Model architecture

Fig.1 demonstrates the architecture of the discriminator in *VSGD-Net*, and Fig. 2 demonstrates the architecture of the attention block between the encoder and decoder in the generator.

The discriminator in *VSGD-Net* has two identical CNNs (see Fig. 1) for coarse-level and fine-level image translation. The input to the coarse-level discriminator is downscaled by 2 from the input to the fine-level discriminator. Each discriminator CNN outputs a value from 0 to 1 to compute the minimax loss, and the features from each convolutional block are extracted to compute the feature similarity loss (Section 3.3.2).

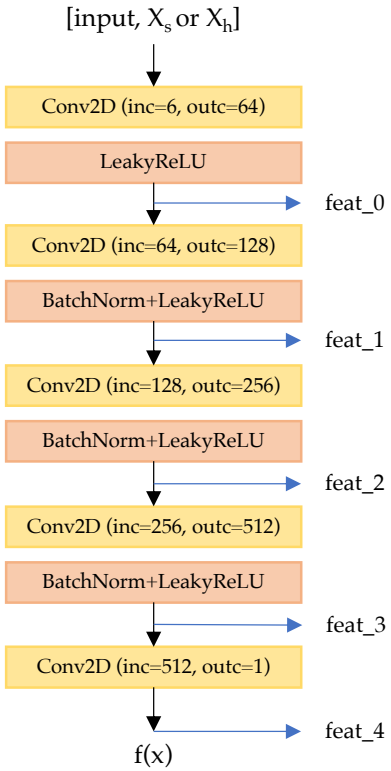


Figure 1. Architecture for a single discriminator CNN.

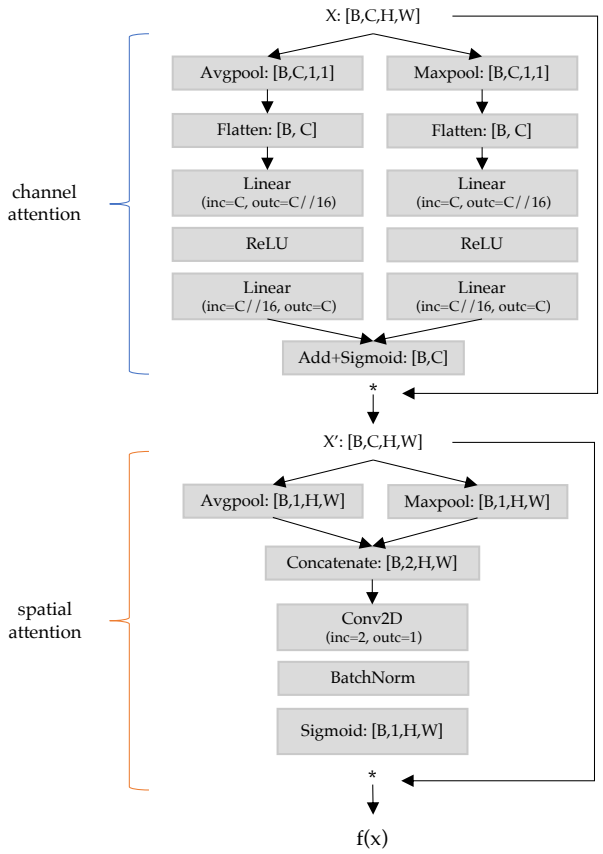


Figure 2. Architecture for the attention block. The input features X are first passed to learn the attention weights for each channel. Then the channel-weighted features are passed to learn the spatial attention. The output features are the weighted sum of the spatial attention map and the channel-weighted features.

B. Additional results

B.1. P-R curve

To further show the effectiveness of our proposed model, we draw the P-R curve of *VSGD-Net* and compare it with other baselines, namely StarDist[2], CHR-Net[1], and the GAN-based Segmentation. In our extensive experiments, StarDist achieves the highest precision score of 0.745, while the GAN-based segmentation method achieves the highest recall score of 0.719, and CHR-Net achieves the second high F1 score 0.645 (Section 4.3). Hence we compare our method with these three baselines in the P-R curve. As discussed in Section 4.2, it is significant for a potential real-world application to have both high recall and high precision. We show the precision score at recall=0.6 to highlight the overwhelming performance of *VSGD-Net*. As Fig. B.1 shows, our *VSGD-Net* achieves satisfact precision with high recall, which additionally validates our framework's strength.

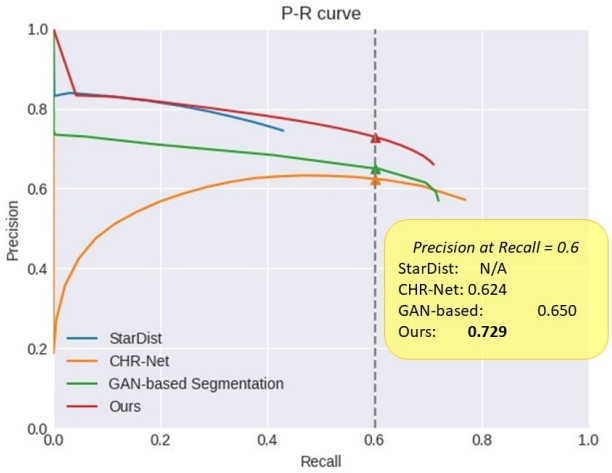
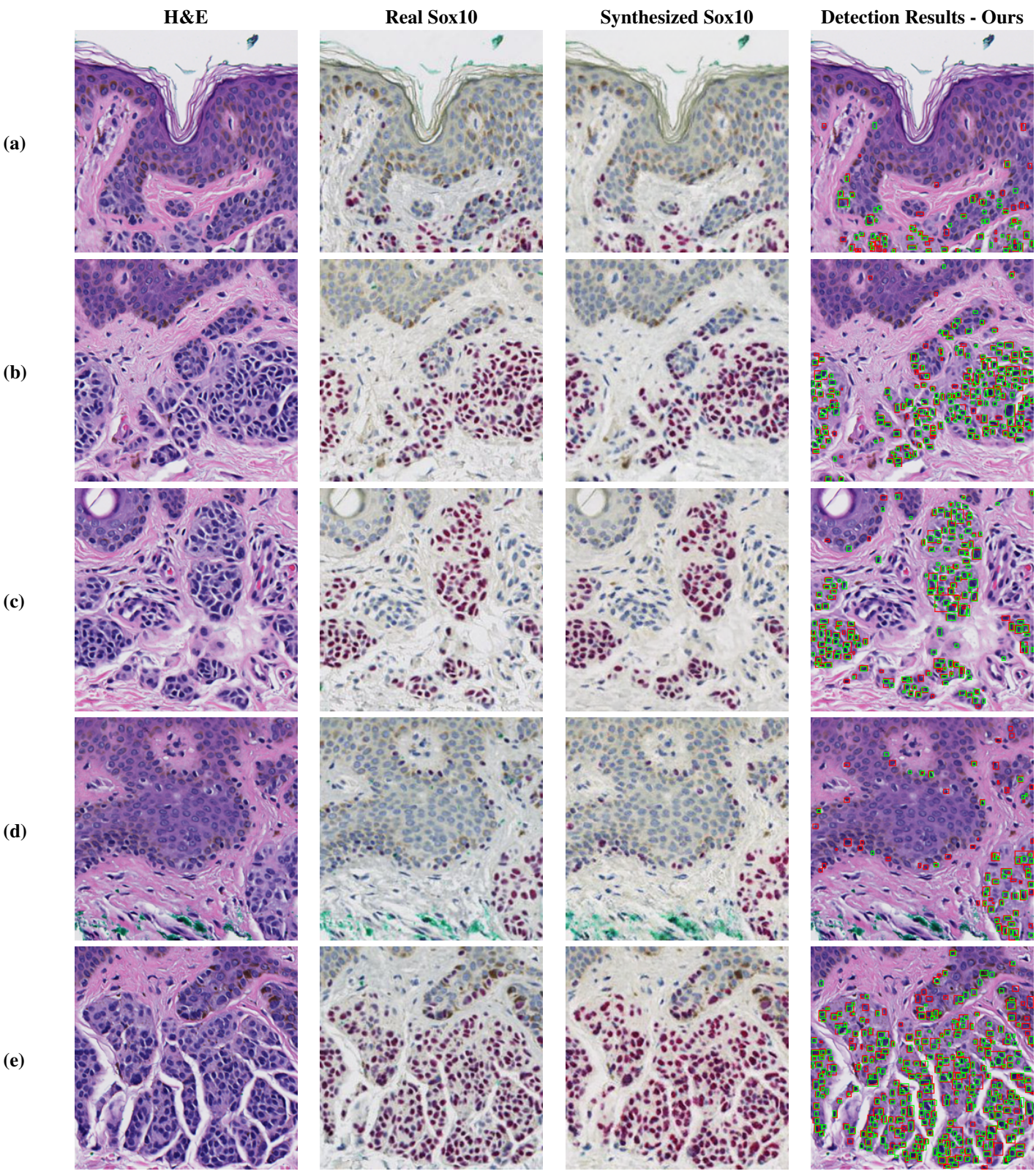


Figure 3. Precision-Recall Curve. The dashed vertical line represents the precision score at recall (P@R) equals to 0.6 The proposed *VSGD-Net* (i.e., the red curve) outperformed all other methods.

B.2. Qualitative results

We present more qualitative results from our *VSGD-Net* in Table. 1 and Table. 2. In Table. 1, *VSGD-Net*'s predicted instances have a high coincidence with the groundtruth instances. Table. 2 show some failure cases. In Table. 2 (a), a lot of melanocytes are missed because the biopsy artifacts in the middle of the image are too challenging for the model to synthesize the corresponding Sox10. In Table. 2 (b), *VSGD-Net* over-predicts some melanocytes. The false positive instances are located in melanocytic nests, which have the specific growth pattern of clustered melanocytes, making it harder to distinguish the non-melanocytes from melanocytes.

We also compare the image synthesis results and the detection results of *VSGD-Net* and the GAN-based segmentation model in Table. 3. The generator, discriminator and the detection branch are the same in the models, and the only difference is that the GAN-based segmentation model leverages a cascaded structure of directly feeding the synthesized output to the detection instead of using the intermediate features. The comparison demonstrates the effectiveness of our design that the shared features can improve the detection and the image synthesis at the same time. The synthesized Sox10 images from *VSGD-Net* are closer to real Sox10 images than the GAN-based segmentation method, even when biopsy artifacts exist (see Table. 3 (e)). Detection results in *VSGD-Net* also have fewer false positives than the cascaded model.


H&E
Real Sox10
Synthesized Sox10
Detection Results - Ours
(a)
(b)
(c)
(d)
(e)

(f)

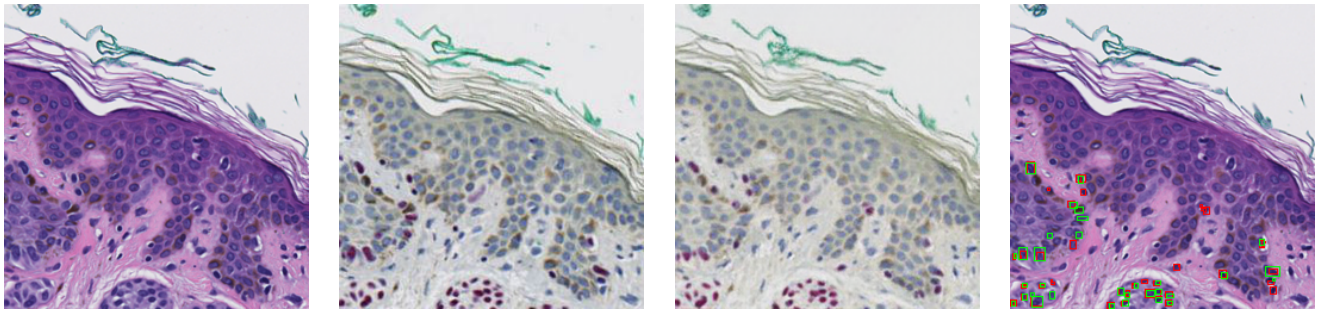


Table 1: Qualitative results. The green bounding boxes denote the groundtruth melanocytes while the red bounding boxes denote the predicted melanocytes. (**Zoom in for best view**)

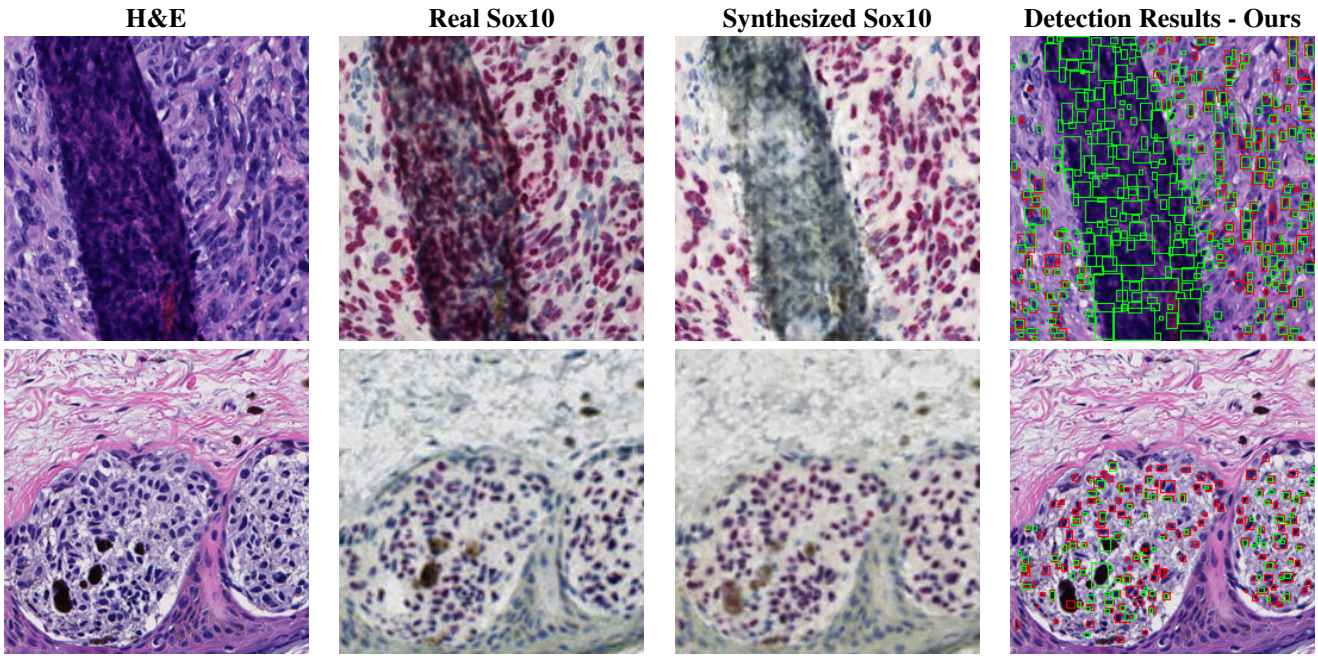


Table 2: Failure cases. The green bounding boxes denote the groundtruth melanocytes while the red bounding boxes denote the predicted melanocytes. (**Zoom in for best view**)

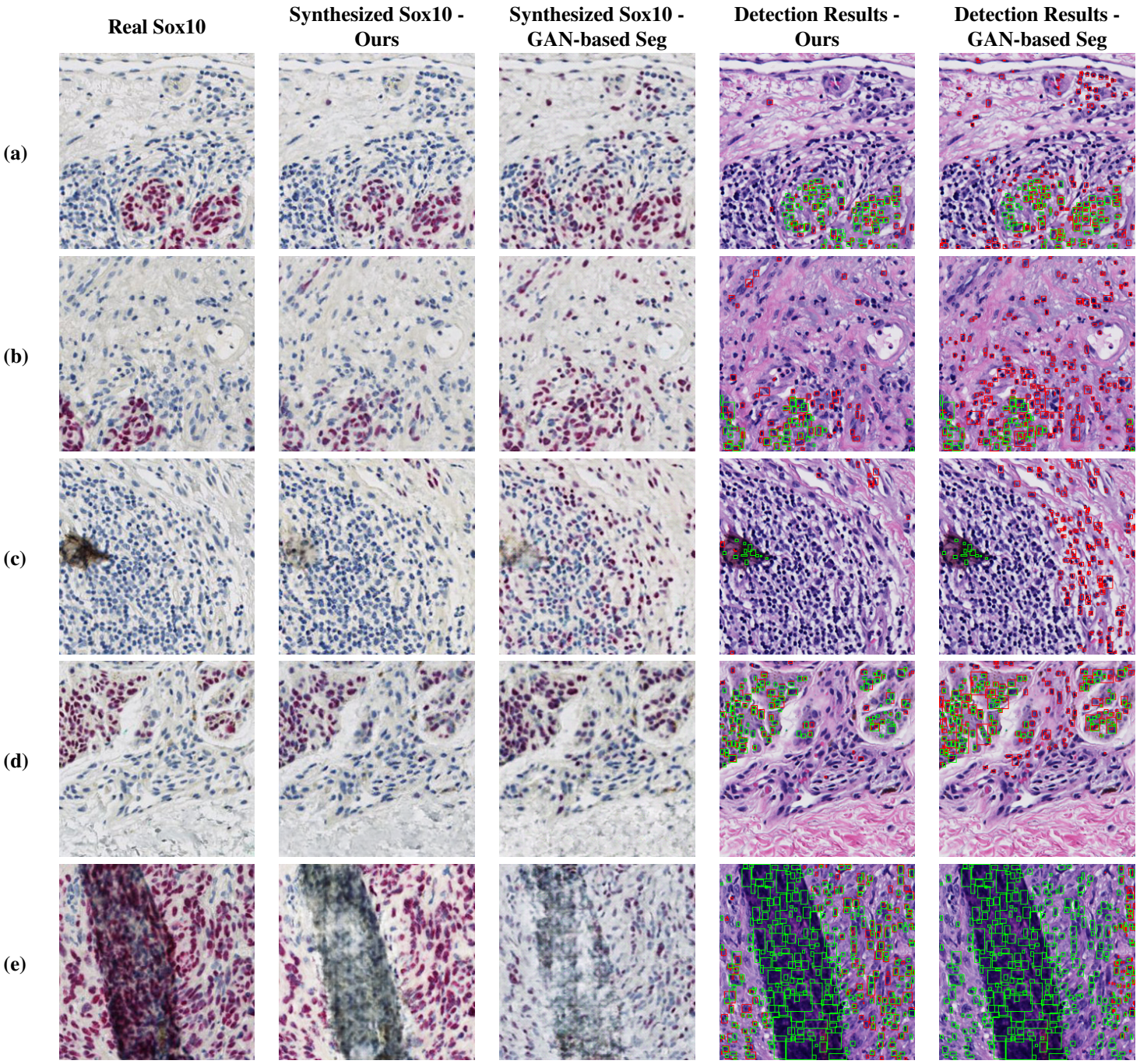


Table 3: Comparison with GAN-based segmentation. In the last two columns, the green bounding boxes denote the groundtruth melanocytes, while the red bounding boxes denote the predicted melanocytes of the corresponding model. The architectures of *VSGD-Net* and GAN-based segmentation are the same except *VSGD-Net* uses shared image synthesis features to learn detection, while the GAN-based segmentation uses the synthesized output to learn the detection. These comparison results demonstrate the benefits of using shared features to boost the detection and image synthesis performance. (**Zoom in for best view**)

References

[1] 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.

[2] 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.