Alleviating tiling effect by random walk sliding window in high-resolution histological whole slide image synthesis

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

Multiplex immunofluorescence (MxIF) is an advanced molecular imaging technique that can simultaneously provide biologists with multiple (i.e., more than 20) molecular markers on a single histological tissue section. Unfortunately, due to imaging restrictions, the more routinely used hematoxylin and eosin (H&E) stain is typically unavailable with MxIF on the same tissue section. As biological H&E staining is not feasible, previous efforts have been made to obtain H&E whole slide image (WSI) from MxIF via deep learning empowered virtual staining. However, the tiling effect is a long-lasting problem in high-resolution WSI-wise synthesis. The MxIF to H&E synthesis is no exception. Limited by computational resources, the cross-stain image synthesis is typically performed at the patch-level. Thus, discontinuous intensities might be visually identified along with the patch boundaries assembling all individual patches back to a WSI. In this work, we propose a deep learning based unpaired high-resolution image synthesis method to obtain virtual H&E WSIs from MxIF WSIs (each with 27 markers/stains) with reduced tiling effects. Briefly, we first extend the CycleGAN framework by adding simultaneous nuclei and mucin segmentation supervision as spatial constraints. Then, we introduce a random walk sliding window shifting strategy during the optimized inference stage, to alleviate the tiling effects. The validation results show that our spatially constrained synthesis method achieves a 56% performance gain for the downstream cell classification task. The proposed inference method reduces the tiling effects by using 50% fewer computation resources without compromising performance. The proposed random sliding window inference method is a plug-and-play module, which can be generalized for other high-resolution WSI image synthesis applications.

Keywords: Unpaired synthesis, tiling effect, spatial constraints

1. Introduction

Multiplex immunofluorescence (MxIF) is a recently developed imaging platform that allows multiple (i.e., more than 20) cellular and histological markers to be investigated on a single tissue section (Vahadane et al., 2023). The hematoxylin and eosin (H&E) staining is still considered the gold standard in pathology that has an immense amount of historical data with well-established clinical guidelines (Simonson et al., 2021). With the rich molecular features, linking the MxIF images with the more broadly used H&E images might provide new insights for clinical research. Unfortunately, the H&E images are typically unavailable with MxIF on the same tissue section due to the physical restrictions in imaging procedures (Berens et al., 2019). As the biological H&E staining is not feasible, generative adversarial networks (GANs) have been applied for obtaining virtual H&E (vH&E) whole slide image (WSI) via deep learning empowered virtual staining (Zhang et al., 2020; Nadarajan and

© 2023 Anonymous.

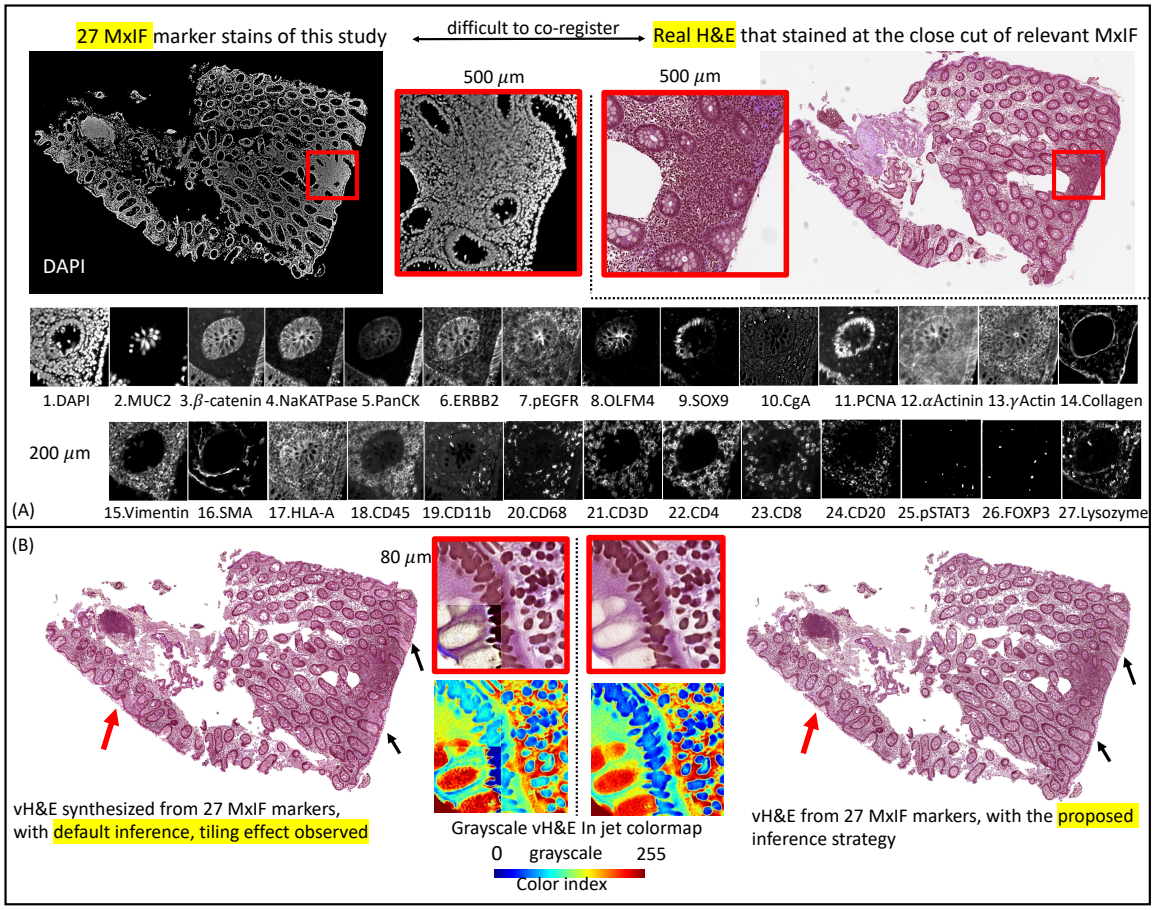

Figure 1: This figure shows the problem setting of the unpaired "set to image" synthesis and the tiling effects in WSI synthesis. (A) We propose to use all 27 markers (set) to synthesize the virtual H&E images (vH&E). (B) The left panel shows that the default patch-wise inference method causes the tiling. The right panel shows that our new method alleviates solve the tiling effect.

Doyle, 2020; Bayramoglu et al., 2017). However, the prior arts did not explicitly model the uniquely rich structural information in MxIF. Some research works also demonstrated MxIF inter-modality set synthesis (i.e., more than 10 stains) (Saurav et al., 2022; Bao et al., 2021), but their models were trained with paired MxIF data, and H&E was not involved.

In this work, we propose a novel unpaired "set to image" synthesis model (MxIF to H&E) by utilizing 27 MxIF channels (Figure 1A). Briefly, we extend the CycleGAN (Zhu et al., 2017) framework with a multi-channel synthesis network with extra semantic segmentation branches to ensure anatomical consistency across modalities.

Another long-lasting issue for deep learning based synthesis in histology is that patch-wise training is inevitable for the high-resolution WSIs (de Bel et al., 2018). As shown in Figure 1B, discontinuous intensities are visually identified along the patch boundaries when

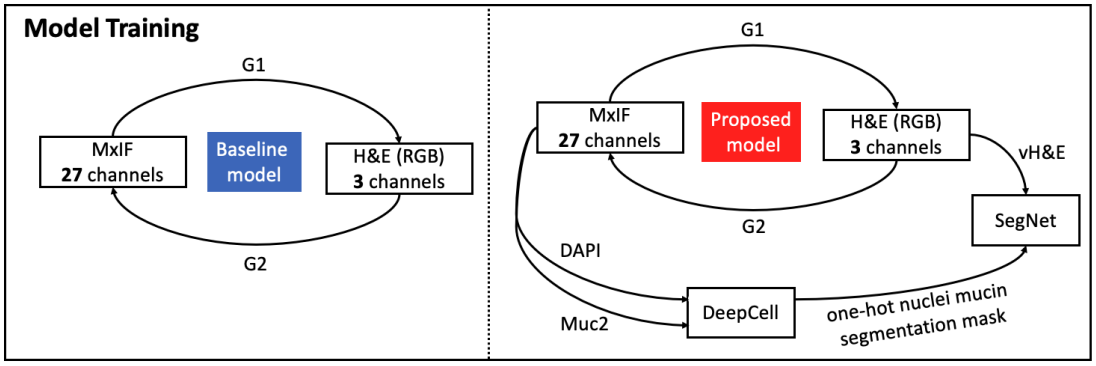

Figure 2: This figure presents the baseline and proposed models for the MxIF and vH&E unpaired synthesis. The baseline models take all 27 channels as a set to synthesize the H&E image. A structural consistent branch (SegNet) is introduced that takes vH&E as input to conduct the nuclei (DAPI) and mucin (Muc2) segmentation.

assembling the individual patches back to a WSI. To address this problem, Lahiani et al. refined the CycleGAN to reconstruct H&E WSIs from FAP-CK images and reduce the tiling artifact by adding the encoder level consistent loss (Lahiani et al., 2020). However, given the domain shifts on H&E RGB images and 27 multi-channel MxIF grayscale images, it is challenging in practise to impose constraints between the encoder of MxIF to H&E and the encoder for HE to MxIF. In this paper, we introduce a random walk sliding window shifting strategy during the optimized inference stage to alleviate the tiling effects that is generalizable.

The contribution of this work is four-fold: **(1)** We propose a novel unpaired "set to image" synthesis model (MxIF to H&E) that utilizes the rich structural information in 27 MxIF channels by aggregating simultaneous nuclei and mucin segmentation supervision as spatial constraints; **(2)** We introduce a sliding window shifting strategy during the optimized inference stage so as to alleviate the tiling effects; **(3)** The proposed random sliding window inference method is a plug-and-play module, which can be generalized and used for other high-resolution WSI image synthesis applications; **(4)** We evaluated the synthesis model in a quantitative manner by employing epithelial cell classification as a downstream task.

2. Methods

2.1. Unpaired set to image synthesis

We used CycleGAN as the backbone of the study. Given two synthesis domains, the generators G_1 and G_2 synthesize fake images to fool the relevant discriminators D_1 and D_2 to enhance the reconstruction in both ways. As Figure 2 shows, the G_1 of the base model

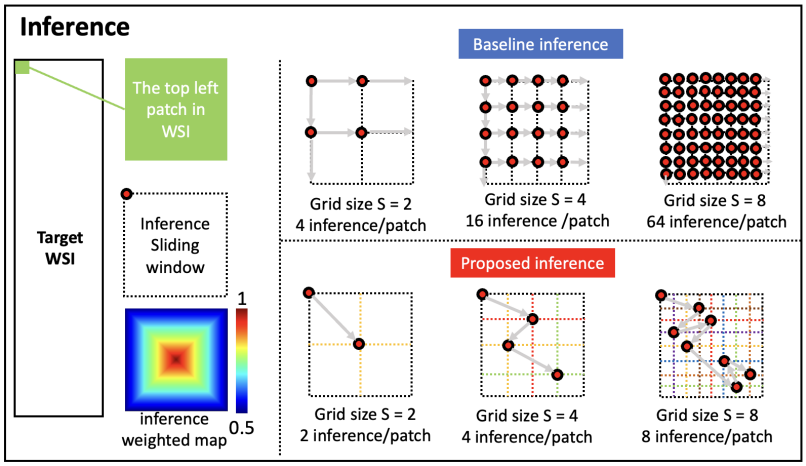

Figure 3: The baseline inference setup and proposed random sliding window inference way. We fix the size of the inference sliding window, and the weighted map is used to average the different inferences. We start the inference from the left top corner of the WSI, then we split the target patch by a grid with size as S, the baseline inference will cost S^2 inference/patch with a pre-defined step, and the proposed inference will take S inference/patch.

has multiple channels for the MxIF markers (denoted as X), G_2 takes H&E (denoted as Y) to synthesize MxIF, so the input of G_2 is a regular three channels for RGB images. The output of G_1 is three channels, while the output of G_2 is 27 channels. Since the G_1 and G_2 have incompatible output channels, the total loss we use for the base model is defined in equation 1 without identity loss (Zhu et al., 2017):

$$L_{base} = L_{GAN}(G_1, D_2, X, Y) + L_{GAN}(G_2, D_1, X, Y) + w_{cycle} L_{cycle}(G_1, G_2) \tag{1}$$

where L_{GAN} is the adversarial loss and L_{cycle} is the cycle consistency loss, w_{cycle} is a constant value that equals 10 references from the original CycleGAN article (Zhu et al., 2017). Because the pattern and shape of the nuclei are the critical factors for H&E interpretation, to ensure the structure consistency for each nucleus, and inspired by the Synseg-net (Huo et al., 2018b,a), we added a Res-net based segmentation network (SegNet) that took the vH&E as input (Figure 2). The target reference truth labels were generated from the pre-trained model DeepCell using the DAPI stains and the Muc2 stains (Bannon et al., 2021; Greenwald et al., 2022). We converted the DeepCell instance labels to one-hot semantic labels (denoted as Z) to ease the segmentation task. Then, the final loss for the proposed MxIF and H&E refining CycleGAN model is defined in equation 2:

$$L_{propose} = L_{base} + L_{DICE}(G_1(X), Z) \tag{2}$$

where L_{DICE} is the DICE loss to penalize the wrong prediction (Shen et al., 2018).

2.2. Random sliding window for inference

(de Bel et al., 2018) proposed using overlap patches to generate inference of the adjacent tile to prevent artifacts at the edges of tiles. To create the final inference result of the overlap sections, (de Bel et al., 2018) claimed to calculate the sum of the weighted pixel values and divide by the sum of the weights. Inspired by (de Bel et al., 2018), we further define the pixel close to the center of the sliding window would contribute 100%, and the pixel in the inference sliding window boundary would contribute 50%, as shown in Figure 3. To avoid confusion, we only make the sliding window move in the direction of right or down. The region of the sliding window outside the target patch would be weighted average by relevant adjacent patches. We define the inference sliding window's size as W_s. The issue of (de Bel et al., 2018) is computationally expensive. For instance, if the pre-defined moving step of the sliding window is $step = \frac{1}{2}W_s$, then the target patch is split as a grid with a size $S = 2$. The base model would traverse all possible $S^2 = 4$ nodes. Similarly, if $step = \frac{1}{4}W_s$, then the grid size is $S = 4$, and the base model would use $S^4 = 16$ inferences per patch.

In our proposed approach, for the first patch of the whole WSI (the top left one), if the grid size S is pre-defined, then we downgrade the seeking of the potential inference point as a random walk problem, which means the inference sliding window start moving from the top left corner and it should not visit any inner grid lines twice. Then, we randomly select a potential S node as a set as

$$R = \{(x_0, y_0), ..., (x_{s-1}, y_{s-1})\} \tag{3}$$

where (x, y) is a node in the grid, (x_0, y_0) is always the top left corner of the WSI. Obviously, the size of R is S. For any two nodes (x_i, y_i) and (x_j, y_j) in R, $x_i \neq x_j, y_i \neq y_j$. Including the randomness on sliding window has been proven effective for model training, especially on 3D medical images (Tang et al., 2021, 2022). Very few, if any, studies have been focused on utilizing the random shifting window for medical image inference. Once the R is finalized, we keep moving sliding window by step $= W_s$ horizontally and vertically from any node in R. The main motivation behind the random design is to ensure that multiple sliding windows would cover each pixel of the tissue in the WSI under a statistical fashion. The proposed way reduces the inference computation complexity from $O(S^2)$ to $O(S)$ per patch.

3. DATA AND EXPERIMENTAL DESIGN

3.1. Data and preprocessing

The data were collected from in-house patients with Crohn's disease (CD), a challenging inflammatory bowel disease (IBD), which is characterized by chronic relapsing and remitting bowel inflammation and causes tremendous intestinal inury(Baumgart and Sandborn, 2012). 28 biopsies (14 ascending colon (AC) tissue samples, 13 terminal ileum (TI) samples) were collected from 16 CD patients, and 4 healthy controls with IRB approval ($***$). The MxIF markers were stained and acquired from three batches with 20$\times$ magnification. The markers are DAPI, Muc2, γActin, CD45, CD11B, Collagen, CD20, PCNA, β-catenin, pSTAT3, pEGFR, CgA, CD4, CD3d, HLA-A, PanCK, OLFM4, CD8, αActinin, CD68, NaKATPase, Vimentin, Sox9, FOXP3, Lysozyme, SMA, and ERBB2. Each MxIF marker images was preprocessed by DAPI-based spatial alignment, autofluorescence correction, and

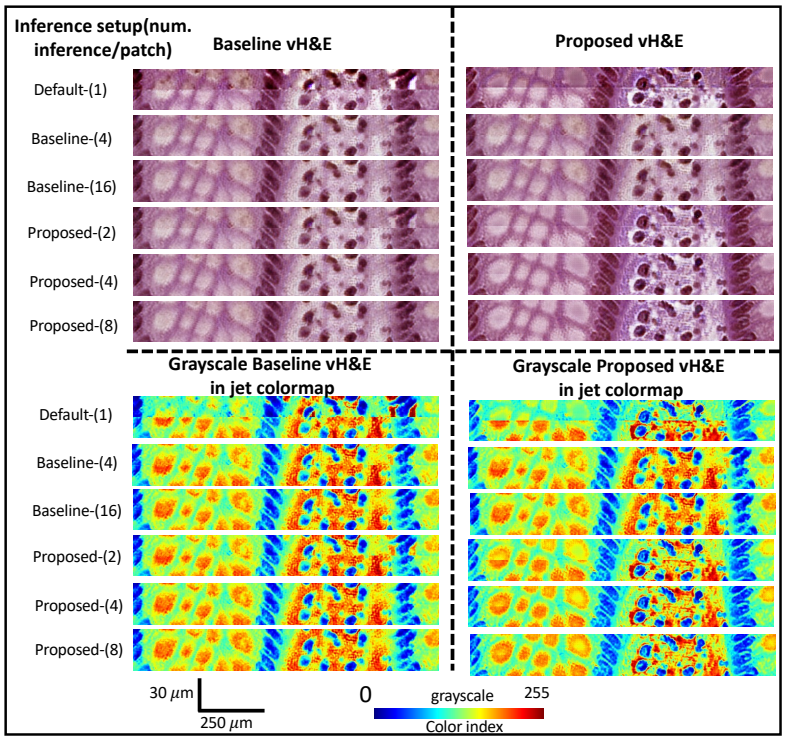

Figure 4: The qualitative results of titling effect removal on a random selective ROIs. Both vH&E in raw intensity and grayscale vH&E in jet colormap are shown.

background masked-out (McKinley et al., 2017). 28 H&E WSI images were used from the close cut of MxIF tissue sections, where the MxIF is about 5 μm deeper than H&E.

3.2. Experimental Design

We set up the training testing 8 : 2 split-ratio and considered sample diversity, where 22 samples were used for training while 6 were tested. We crop all WSIs into 1024×1024 patches (0.324 μm/pixel) while only keeping the ones with 80% and more foreground tissues. Finally, 265 patches were remained for MxIF and 181 for H&E. The experiments were performed on a workstation with an NVIDIA RTX A6000 GPU.

Tiling effect evaluation. We infer the testing samples using six ways. The first way is default inference only uses one inference. The second to fourth approaches use the proposed method with different grid sizes ($S = 2, 4, 8$) for each target patch. Then the proposed method would use 2, 4, and 8 inferences per patch. The fifth and sixth inference methods use the baseline when $S = 2$ and 4, which causes 4 and 16 inferences per patch. For computation efficiency, we do not perform baseline when $S = 8$. For the ablation study, we validate the proposed method when $S = 4$, where $3! = 6$ cases are available.

Table 1: Results for the tiling effect removal mean (standard deviation) Euclidian distance values from different inference methods. The best strategy is highlighted with a significant difference from other ways by the Wilcoxon signed-rank test ($p < 0.05$).

		vH&E base model	**vH&E proposed model**
Validation of different setup (num. of inference per patch)	Default-(1)	460.6 (408.4)	489.2 (398.2)
	Baseline-(4)	383.2 (237.1)	413.4 (243.9)
	Baseline-(16)	356.2 (200.2)	386.8 (213.1)
	Proposed-(2)	397.9 (252.9)	425.7 (254.5)
	Proposed-(4)	366.0 (206.0)	394.7 (217.7)
	Proposed-(8)	**349.7 (193.4)**	**379.2 (207.6)**
		vH&E base model	**vH&E proposed model**
Ablation study	Proposed-(4) case1	366.0 (206.0)	394.7 (217.7)
	Proposed-(4) case2	365.7 (206.2)	394.9 (217.8)
	Proposed-(4) case3	365.6 (205.8)	394.6 (217.5)
	Proposed-(4) case4	366.1 (206.1)	395.3 (217.8)
	Proposed-(4) case5	366.6 (206.3)	394.8 (217.8)
	Proposed-(4) case6	366.6 (206.4)	395.3 (217.9)

To evaluate the tilting effect removal performance, we find all lines of longitude and latitude if the coordinates of lines can be divisible by 128. Then we collect the pixels of two sides of the lines with one width or height, convert the RGB values to CIELAB space, and calculate the Euclidean distance of relevant two adjacent stripe pixels along the lines. In summary, there are 1521 lines across all six testing samples are collected for validation.

vH&E for downstream task. To further quantitatively evaluate the overall synthesis performance, we performed downstream epithelial cell segmentation experiments using vH&E. To generate the reference label, we manually select the threshold to find definite positive signals of (NaKATPase+ PanCK+) and exclude Vimentin+. Then we integrate such MxIF biological rule-based threshold knowledge to annotate the DeepCell instance segmentation label and convert the annotation to semantic labels as the training and testing annotations. We crop the WSI into 256×256 patches (0.5 μm/pixel). We run the pretrained HoVerNet model for the CoNIC challenge (Graham et al., 2019, 2021) on the 319 vH&E patches and get the generated epithelial labels. We utilize the masked DICE coefficient score for epithelial cell classification label consistency from MxIF and vH&E.

4. RESULTS

Tiling effect evaluation. Figure 4 shows the qualitative results of titling effect removal on a random selective region of interest (ROI). Both vH&E in raw intensity and grayscale vH&E in jet colormap are shown. The quantitative results are listed in Table 1, where we found that for each of the two vH&E syntheses, the proposed method with grid size $S = 8$ outperformed the best baseline setup with 50% less computation usage. Table 4 also presents the ablation study, where no significant difference is found.

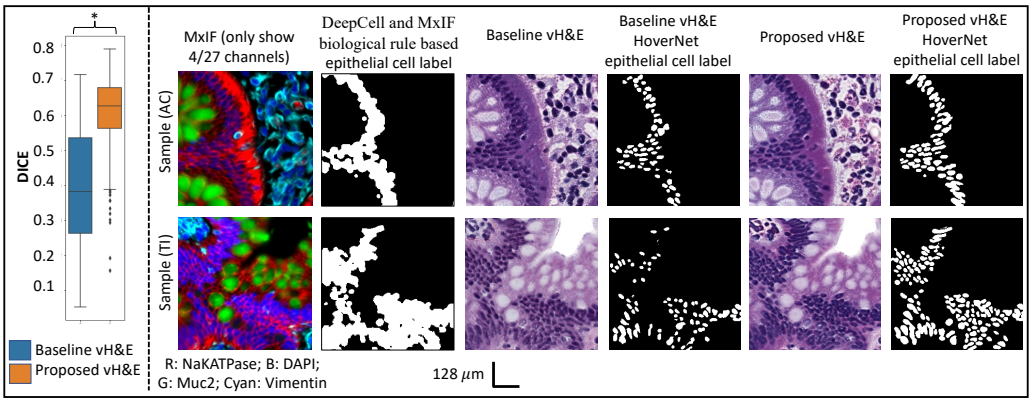

Figure 5: The performance of the downstream epithelial cell segmentation task using vH&E synthesis. The left panel shows the epithelial cell segmentation performance with significance found by the Wilcoxon signed-rank test ($p < 0.05$, marked as *). The right panel shows the qualitative results between the DeepCell and MxIF rule-based label versus the vH&E HoverNet label.

vH&E for downstream task. Figure 5 shows qualitative and quantitative results for the two vH&E syntheses on epithelial cell classification consistency between DeepCell and MxIF biological rule-based label and H&E HoverNet label. Two sample ROIs from testing AC and TI samples are randomly selected. The results show that the proposed CycleGAN with structure constraints outperforms the baseline by 56% improvement.

5. Discussion and Conclusion

Direct evaluation on image similarity is essential. However, we do not have paired MxIF and H&E data for such direct comparison. A downstream epithelial cell classification task would be helpful for evaluating the quality of synthesis only, which is a main limitation of our work. DeepCell nuclei segmentation is prone to over-segmenting the nuclei, which could introduce bias in the downstream analysis when we compare the results between the MxIF DeepCell label and the HoverNet label. A refined MxIF label could potentially improve validation accuracy that requires further investigation.

In this paper, we present a novel unpaired "set to image" synthesis model (MxIF to vH&E) that fully utilizes the rich structural information in 27 MxIF channels by adding simultaneous nuclei and mucin segmentation supervision as spatial constraints. A random sliding window shifting strategy is introduced during the optimized inference stage towards alleviate the tiling effects. Our simple inference method is a plug-and-play module that can be generalized and used for other high-resolution WSI image synthesis applications. Our model our proposed vH&E synthesis method achieves superior performance compared with the baseline approach.

References

Dylan Bannon, Erick Moen, Morgan Schwartz, Enrico Borba, Takamasa Kudo, Noah Greenwald, Vibha Vijayakumar, Brian Chang, Edward Pao, Erik Osterman, et al. Deepcell kiosk: scaling deep learning–enabled cellular image analysis with kubernetes. *Nature methods*, 18(1):43–45, 2021.

Shunxing Bao, Yucheng Tang, Ho Hin Lee, Riqiang Gao, Sophie Chiron, Ilwoo Lyu, Lori A Coburn, Keith T Wilson, Joseph T Roland, Bennett A Landman, et al. Random multi-channel image synthesis for multiplexed immunofluorescence imaging. In *MICCAI Workshop on Computational Pathology*, pages 36–46. PMLR, 2021.

Daniel C Baumgart and William J Sandborn. Crohn's disease. *The Lancet*, 380(9853): 1590–1605, 2012.

Neslihan Bayramoglu, Mika Kaakinen, Lauri Eklund, and Janne Heikkila. Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 64–71, 2017.

Michael E Berens, Anup Sood, Jill S Barnholtz-Sloan, John F Graf, Sanghee Cho, Seungchan Kim, Jeffrey Kiefer, Sara A Byron, Rebecca F Halperin, Sara Nasser, et al. Multiscale, multimodal analysis of tumor heterogeneity in idh1 mutant vs wild-type diffuse gliomas. *PLoS One*, 14(12):e0219724, 2019.

Thomas de Bel, Meyke Hermsen, Jesper Kers, Jeroen van der Laak, and Geert Litjens. Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. In *International Conference on Medical Imaging with Deep Learning–Full Paper Track*, 2018.

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.

Simon Graham, Mostafa Jahanifar, Quoc Dang Vu, Giorgos Hadjigeorghiou, Thomas Leech, David Snead, Shan E Ahmed Raza, Fayyaz Minhas, and Nasir Rajpoot. Conic: Colon nuclei identification and counting challenge 2022. *arXiv preprint arXiv:2111.14485*, 2021.

Noah F Greenwald, Geneva Miller, Erick Moen, Alex Kong, Adam Kagel, Thomas Dougherty, Christine Camacho Fullaway, Brianna J McIntosh, Ke Xuan Leow, Morgan Sarah Schwartz, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology*, 40(4):555–565, 2022.

Yuankai Huo, Zhoubing Xu, Shunxing Bao, Albert Assad, Richard G Abramson, and Bennett A Landman. Adversarial synthesis learning enables segmentation without target modality ground truth. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 1217–1220. IEEE, 2018a.

Yuankai Huo, Zhoubing Xu, Hyeonsoo Moon, Shunxing Bao, Albert Assad, Tamara K Moyo, Michael R Savona, Richard G Abramson, and Bennett A Landman. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE transactions on medical imaging*, 38(4):1016–1025, 2018b.

Amal Lahiani, Irina Klaman, Nassir Navab, Shadi Albarqouni, and Eldad Klaiman. Seamless virtual whole slide image synthesis and validation using perceptual embedding consistency. *IEEE Journal of Biomedical and Health Informatics*, 25(2):403–411, 2020.

Eliot T McKinley, Yunxia Sui, Yousef Al-Kofahi, Bryan A Millis, Matthew J Tyska, Joseph T Roland, Alberto Santamaria-Pang, Christina L Ohland, Christian Jobin, Jeffrey L Franklin, et al. Optimized multiplex immunofluorescence single-cell analysis reveals tuft cell heterogeneity. *JCI insight*, 2(11), 2017.

Gouthamrajan Nadarajan and Scott Doyle. Realistic cross-domain microscopy via conditional generative adversarial networks: converting immunofluorescence to hematoxylin and eosin. In *Medical Imaging 2020: Digital Pathology*, volume 11320, pages 187–197. SPIE, 2020.

Jillur Rahman Saurav, Mohammad Sadegh Nasr, Paul Koomey, Michael Robben, Manfred Huber, Jon Weidanz, Bríd Ryan, Eytan Ruppin, Peng Jiang, and Jacob M Luber. A ssim guided cgan architecture for clinically driven generative image synthesis of multiplexed spatial proteomics channels. *arXiv preprint arXiv:2205.10373*, 2022.

Chen Shen, Holger R Roth, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. On the influence of dice loss function in multi-class organ segmentation of abdominal ct using 3d fully convolutional networks. *arXiv preprint arXiv:1801.05912*, 2018.

Paul D Simonson, Xiaobing Ren, and Jonathan R Fromm. Creating virtual hematoxylin and eosin images using samples imaged on a commercial codex platform. *Journal of Pathology Informatics*, 12(1):52, 2021.

Yucheng Tang, Riqiang Gao, Ho Hin Lee, Shizhong Han, Yunqiang Chen, Dashan Gao, Vishwesh Nath, Camilo Bermudez, Michael R Savona, Richard G Abramson, et al. High-resolution 3d abdominal segmentation with random patch network fusion. *Medical image analysis*, 69:101894, 2021.

Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20730–20740, 2022.

Abhishek Vahadane, Shreya Sharma, Devraj Mandal, Madan Dabbeeru, Josephine Jakthong, Miguel Garcia-Guzman, Shantanu Majumdar, and Chung-Wein Lee. Development of an automated combined positive score prediction pipeline using artificial intelligence on multiplexed immunofluorescence images. *Computers in Biology and Medicine*, 152: 106337, 2023.

Yijie Zhang, Kevin de Haan, Yair Rivenson, Jingxi Li, Apostolos Delis, and Aydogan Ozcan. Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light: Science & Applications*, 9(1):1–13, 2020.

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.

Appendix A. Appendix A. Tiling effect examples

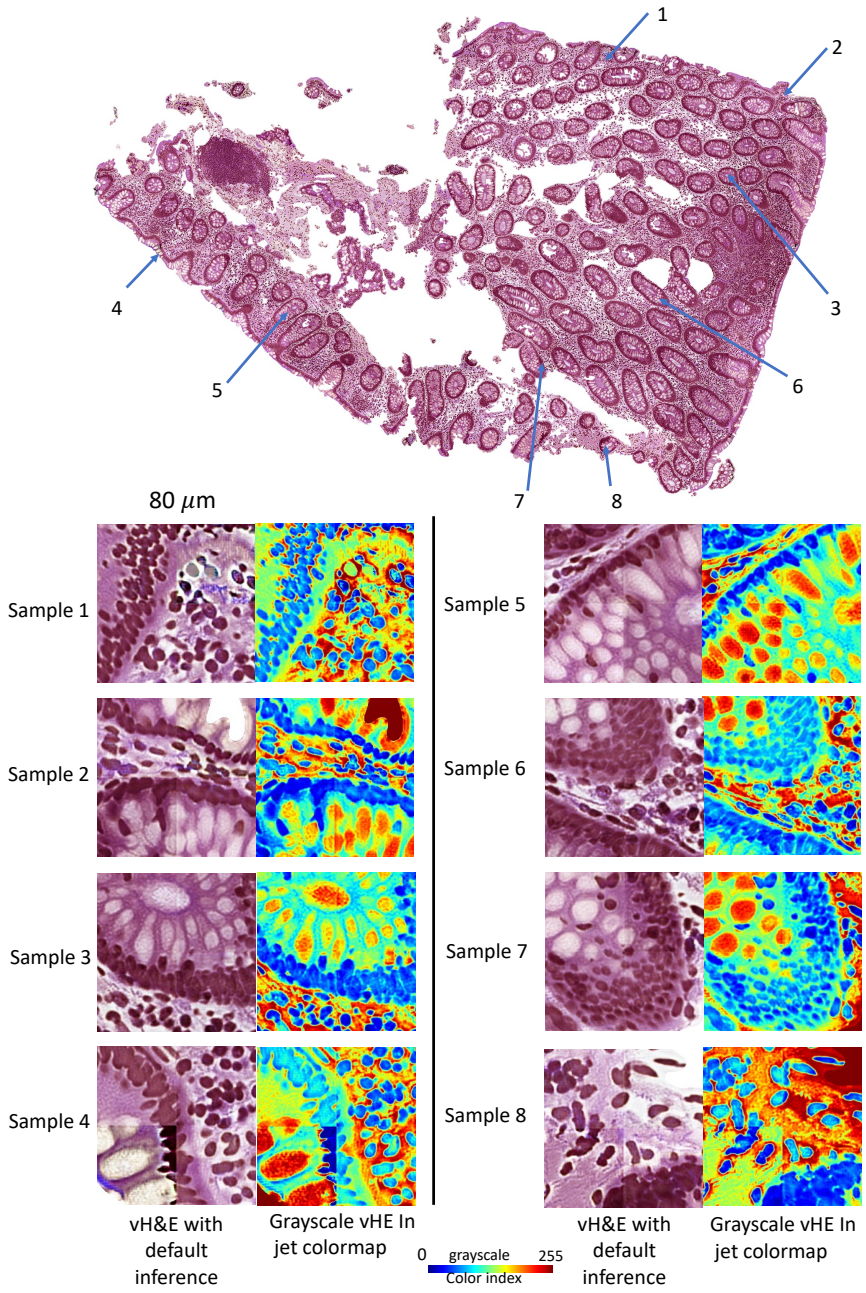

Figure 6: A few bad examples of the tiling effect after stiching the patch back into WSI using default inference method.