



Virtualization of Tissue Staining in Digital Pathology Using an Unsupervised Deep Learning Approach

Amal Lahiani^{1,2(✉)}, Jacob Gildenblat³, Irina Klamann¹, Shadi Albarqouni²,
Nassir Navab², and Eldad Klaiman¹

¹ Pathology and Tissue Analytics, Pharma Research and Early Development,
Roche Innovation Center Munich, Penzberg, Germany
amal.lahiani@roche.com

² Computer Aided Medical Procedures, Technische Universität München,
Munich, Germany

³ DeePathology.ai, Ra'anana, Israel

Abstract. Histopathological evaluation of tissue samples is a key practice in patient diagnosis and drug development, especially in oncology. Historically, Hematoxylin and Eosin (H&E) has been used by pathologists as a gold standard staining. However, in many cases, various target specific stains, including immunohistochemistry (IHC), are needed in order to highlight specific structures in the tissue. As tissue is scarce and staining procedures are tedious, it would be beneficial to generate images of stained tissue virtually. Virtual staining could also generate in-silico multiplexing of different stains on the same tissue segment. In this paper, we present a sample application that generates FAP-CK virtual IHC images from Ki67-CD8 real IHC images using an unsupervised deep learning approach based on CycleGAN. We also propose a method to deal with tiling artifacts caused by normalization layers and we validate our approach by comparing the results of tissue analysis algorithms for virtual and real images.

Keywords: Virtual staining · Multiplexing ·
Unsupervised deep learning · Histopathology

1 Introduction

In the field of pathology, staining types determine which parts or targets in the tissue are highlighted with specific colors. Tissue staining materials and procedures can be time consuming, expensive, and typically require special expertise. These limitations usually reduce the number of examinations and stainings performed on a sample. This can limit clinicians' ability to obtain all relevant information from a patient biopsy. In many cases information exists in the stained

slide image about targets and objects not specifically targeted by the stain. For example, pathologists have the ability to identify lymphocytes in a Hematoxylin and Eosin (H&E) image [11] even without directly staining them for lymphocyte specific markers. This fact motivated the research in the direction of generating virtually stained slides from other modalities [2, 4, 7, 8, 13]. Recently, supervised deep learning based methods have been applied in the task of virtual staining generation [1, 3, 6, 14]. As supervised training methods are based on coupled pairs of aligned images, all the aforementioned methods require additional accurate registration steps between dataset image pairs.

In this work we propose to virtually generate FAP-CK stained slide images from Ki67-CD8 stained slide images. These input and output stainings were chosen for several reasons. First, information about tumor characteristics in FAP-CK could be encoded in the form of proliferation and tumor infiltrating lymphocytes in Ki67-CD8. Furthermore, Ki67-CD8 is one of the classical Immunohistochemistry (IHC) stainings used in histopathology while FAP-CK is a new duplex IHC protocol allowing to characterize tumor and to advance research in the direction of drug development. Additionally, generating virtual FAP-CK stained slide images from Ki67-CD8 allows the creation of a virtual multiplexed brightfield image, i.e. having 4 target stains on the same whole slide coordinate system, which is technically challenging using classical staining methods. In this paper, we present an unsupervised deep learning method based on Cycle-Consistent Adversarial Networks (CycleGAN) [17]. This allows avoiding the slide registration process for training datasets and facilitates dealing with variability present in sets of slide images due to different lab protocols, scanners and experiment conditions. We further present a method aimed at reducing the tiling artifact caused by tile-wise processing of large images, a common problem in image style transfer encountered when high resolution testing images can not fit into memory [16]. Finally, we validate the results of our method by comparing quantification of tumor cells and FAP in virtual slides with a real stained slide taken from the same tissue block.

2 Methodology

2.1 Dataset

We selected a subset of whole slide images (WSI) of Colorectal Carcinoma metastases in liver tissue from biopsy and surgical specimen from our internal pathology image database. All the slides were chosen following a review of tissue, staining and image quality. The training dataset includes 20 images: 10 from Ki67-CD8 stained slides and 10 from FAP-CK stained slides, each from different patients. As high resolution whole slide histology images contain billions of pixels with 20x magnification and hardware memory is limited, it is necessary to tile the image into smaller segments for analysis and when possible use lower magnification. For these reasons, slides were tiled into overlapping 512×512 images at 10x magnification. The reduced magnification allows to have enough



Fig. 1. Ki67-CD8 and FAP-CK example tiles.

contextual information in the input which is needed in order to learn a meaningful feature set in the model while at the same time facilitates dealing with the computational memory limits [12]. The tiling yielded 17025 tiles from Ki67-CD8 slides and 17812 tiles from FAP-CK slides. In order to test the performance of our method, we selected 10 pairs of Ki67-CD8/FAP-CK slides from the same tissue block. The training and testing images were taken from different patients. Figure 1 shows examples of Ki67-CD8 and FAP-CK tiles.

2.2 Network Architecture and Inference

In order to map images from one staining to the other, we used a CycleGAN [17] based method which allows to learn transforms between domain spaces without the need to use paired registered image datasets. As generator networks we used ResNet architectures [9] with 11 residual blocks. The ResNet architecture used as a generator has a receptive field of 207×207 pixels, roughly corresponding to 190×190 microns on the $10x$ magnification image of the tissue. This receptive field size allows enough contextual information from the surroundings of the target pixel in the input to find meaningful histology features for the prediction of the virtual stain on the output image. Due to the CycleGAN memory burden, we were not able to fit more than one 512×512 RGB image per batch during training per GPU (Nvidia P100). In order to distribute the training on multiple GPUs and accelerate the learning process, we implemented the stochastic synchronous ADAM algorithm [5] and used the pytorch distributed computing library which allowed us to use 12 GPUs concurrently.

In order to overcome the memory bound hardware limitations, inference of the trained network on the testing slides was also done tilewise. The tile output is then merged back in order to obtain a virtual whole slide image. This inference workflow yielded whole slide images containing tiling artifacts. As style transfer networks give better and more realistic results when instance normalization layers are used [15], these layers were introduced in the CycleGAN architecture. Instance normalization layers are applied at test time as well, making the value of any output pixel depends not only on the network parameters and the receptive field area in the input but also on the statics of the input tile image. Let's assume that:

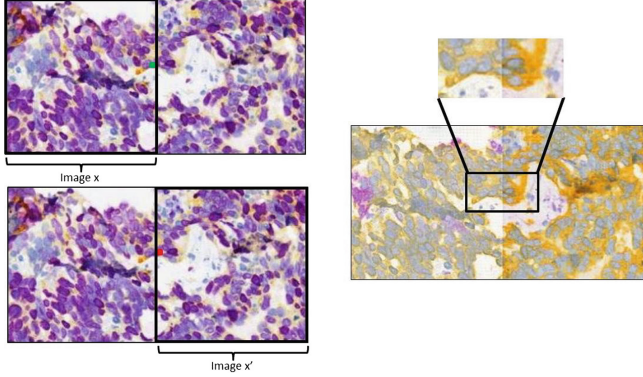


Fig. 2. (right) Tiling effect in adjacent output tiles. (left) Image x and x' correspond to 2 adjacent input tiles from a whole slide image. The green and red circles correspond to 2 adjacent pixels belonging to the same cell nucleus but to different input tiles. (Color figure online)

$$y = g(x),$$

where x , y and g correspond to an input tensor, an output tensor and an instance normalization layer respectively. Let $x \in \mathbb{R}^{T \times C \times W \times H}$ be a tensor containing a batch of T images and x_{tijk} the $tijk^{th}$ element, where j and k correspond to spatial dimensions, i corresponds to the feature channel and t corresponds to the batch index. In this case, g can be expressed as:

$$y_{tijk} = g(x_{tijk}) = \frac{x_{tijk} - \mu_{ti}}{\sqrt{\sigma_{ti}^2 + \varepsilon}}$$

where μ_{ti} and σ_{ti}^2 are the mean and variance of the input tile. If we consider 2 adjacent pixels x_{tijk} and x'_{tijk} with similar values on the edges of two adjacent input tiles x and x' having very different statistics, the instance normalization functions g and g' applied to these two pixels will be completely different. This results in a tiling artifact in the generated whole slide image as adjacent output tiles might have significantly different pixel values on their borders (Fig. 2).

With instance normalization it is possible to use the same running mean and variance values for all the tiles at inference time in a way similar to inference with batch normalization. This approach indeed yielded output images without the tiling artifact. However, the resulting output at inference was of lower quality and had very faint colors. This effect can be explained by the fact that the training datasets are quite variable and containing both tissue and background, this makes the running mean and variance locally irrelevant. We propose a solution to the tiling artifact problem by using overlapping tiles during inference. Our solution is based on using a smaller input size of 128×128 instead of 512×512 and on using a sliding window for the instance normalization function statistics, allowing to have a smooth transition in the statistic values when deploying on 2

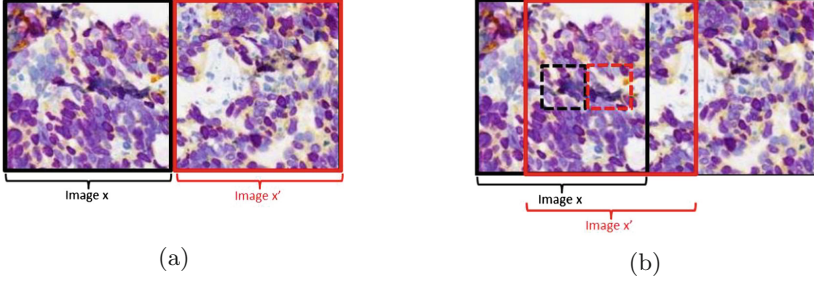


Fig. 3. Image (a) corresponds to the inference performed on 2 adjacent slides using the classical method. Image (b) corresponds to the new inference approach. The solid and dotted line squares correspond to the sliding window considered and the effective tile used for inference respectively.

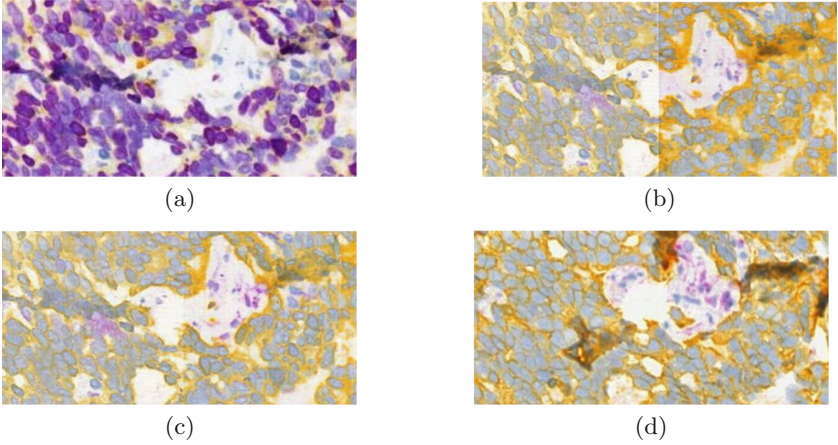


Fig. 4. Effect of the new inference approach. (a) corresponds to 2 adjacent tiles from the input Ki67-CD8 image. (b) and (c) show to the corresponding image area in the output virtual FAP-CK image using the classical inference method and the proposed solution respectively. (d) shows a corresponding area from a real FAP-CK image from the same tissue block.

adjacent slides (Fig. 3). Using this approach, we manage to substantially reduce the tiling artifact (Fig. 4).

3 Results and Validation

Visual assessment of the generated images shows that the results are visually similar to the real staining of a slide from the same tissue block (Fig. 5).

We notice that in several cases FAP (purple connective tissue in FAP-CK images) expression is different between the real and virtual images (Fig. 6). The localisation of FAP is generally successful however the patterns and the amounts

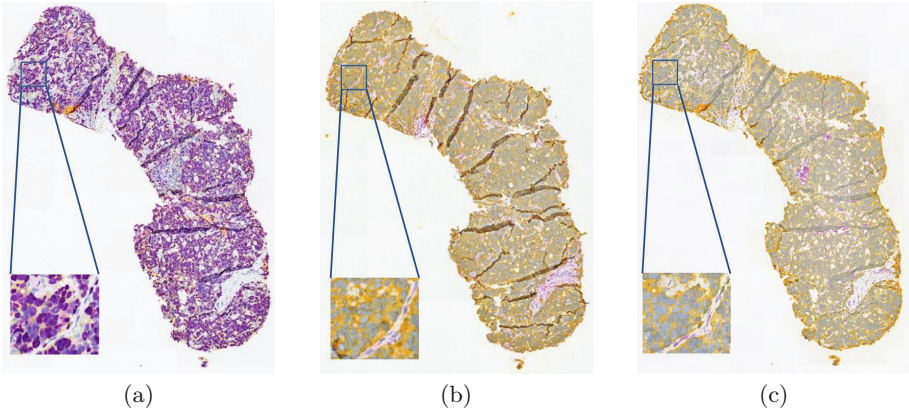


Fig. 5. (a), (b) and (c) correspond to an input Ki67-CD8 image, the image of a real stained FAP-CK slide from the same tissue block and the virtual FAP-CK slide image.

are not always matching. One explanation for this effect can be that FAP is associated with tumor growth and increased angiogenesis rather than with anatomical or phenotypic features [10]. If these functional features do not elicit a change visible in the input Ki67-CD8 staining, the model cannot correctly learn the mapping. The model success in localizing FAP, even if not flawlessly, suggests there are visual features in the input images that indicate the expression of FAP in the tissue. Identifying these visual features might lead to better FAP stain virtualisation as well as to new insights into tumor microenvironment anatomies. It is worthwhile to note that slide staining is a complex process with many variable and it is not uncommon to see variability in images of slides from the same tissue block stained with the same staining protocol due to variations in the tissue preparation, staining or imaging processes. For this reason, even with the visible difference in FAP expression in our virtual slides compared to the real stained slides, the results could still prove practically useful in some cases. Obtaining a virtual image which is slightly different from the real one should be acceptable as long as the pathological interpretations of both of them are similar.

We validate our results on a dataset of 10 testing paired images of slides from the same tissue block using an automatic algorithm for CK⁺ and FAP cells detection. The algorithm was developed and validated using real FAP-CK images. The results include CK⁺ cell densities and FAP densities in real and virtual whole slide images. We verify that the difference between results on real and virtual stained slides is not one sided, meaning that our mapping algorithm does not consistently over-generate or under-generate CK or FAP.

In order to visualize the difference between these densities we compute the absolute relative difference between the results obtained in the real and virtual slides (Fig. 7). Analysis of the results shows a median absolute relative difference of 8% with 0.016 variance between CK densities in real and virtual slides. This was also confirmed by our expert pathologist who evaluated real and virtual

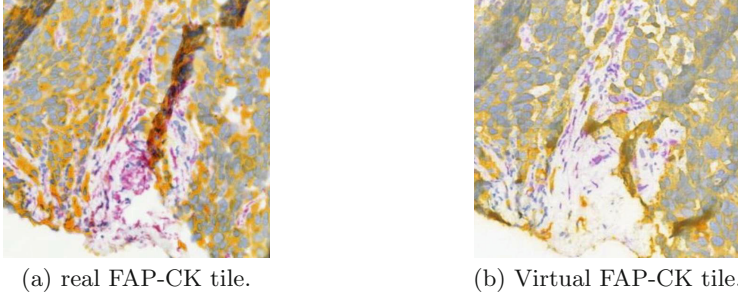


Fig. 6. FAP expression differences in real and virtual images.

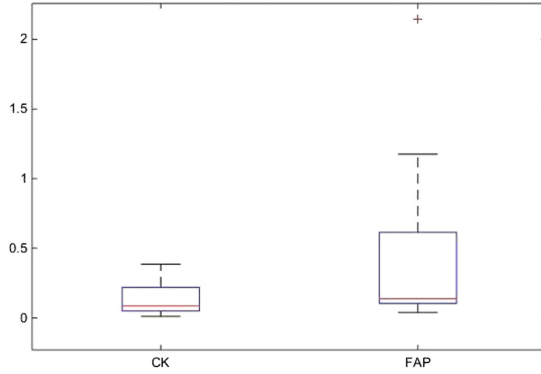


Fig. 7. Boxplot representation of the absolute relative difference between real and corresponding virtual slides for CK⁺ cell densities (left side) and FAP cell densities (right side).

paired slides and reported high correlation in CK expression. For FAP, we report a median of 14% with a variance of 0.466, reflecting a substantially higher variability than for CK. Our expert pathologist also confirmed this observation and mentioned that FAP features are completely not visible for pathologists in Ki67-CD8 staining. This observation is very interesting for our research and allows us to discover the limitations of simulation methods when biological constraints are present.

4 Conclusions

We propose to use an unsupervised deep learning method based on CycleGAN in order to virtually generate FAP-CK from Ki67-CD8 tissue stained images. Instance normalization used with the CycleGAN architecture helps the network learn a more realistic mapping between the stainings but introduces a tiling artifact in the merged testing set whole slide images. We significantly reduce this artifact by using a new inference approach based on overlapping tiles to

create smooth transitions in the instance normalization layers. We validate our method using a test dataset of Ki67-CD8 input images for which a real FAP-CK images from the same tissue blocks were generated.

In the next steps, we plan to use additional input stainings and compare the results to our current results in order to define biological constraints to the success of stain virtualization. We also plan to replace the instance normalization module in the CycleGAN with an improved normalization element that reduces the titling effect during inference.

References

1. Bayramoglu, N., Kaakinen, M., Eklund, L., Heikkilä, J.: Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks. In: ICCV Workshops, pp. 64–71 (2017)
2. Bini, J.M., Spain, J., Nehal, K.S., Rajadhyaksha, M., Hazelwood, V., DiMarzio, C.A.: Confocal mosaicing microscopy of human skin ex vivo: spectral analysis for digital staining to simulate histology-like appearance. *J. Biomed. Opt.* **16**(7), 076008 (2011)
3. Burlingame, E.A., Margolin, A., Gray, J.W., Chang, Y.H.: Shift: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks. In: Medical Imaging 2018: Digital Pathology, vol. 10581, p. 1058105. SPIE (2018)
4. Can, A., Bello, M.O., Gerdes, M.J., Li, Q.: System and methods for mapping fluorescent images into a bright field color space. US Patent 8,269,827 18 Sept 2012
5. Chen, J., Pan, X., Monga, R., Bengio, S., Jozefowicz, R.: Revisiting distributed synchronous SGD. arXiv preprint [arXiv:1604.00981](https://arxiv.org/abs/1604.00981) (2016)
6. Christiansen, E.M., et al.: In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**(3), 792–803 (2018)
7. Gareau, D.S.: Feasibility of digitally stained multimodal confocal mosaics to simulate histopathology. *J. Biomed. Opt.* **14**(3), 034050 (2009)
8. Giacomelli, M.G., et al.: Virtual hematoxylin and eosin transillumination microscopy using epi-fluorescence imaging. *PLoS ONE* **11**(8), e0159337 (2016)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE CVPR, pp. 770–778 (2016)
10. Kelly, T., Huang, Y., Simms, A.E., Mazur, A.: Fibroblast activation protein- α : a key modulator of the microenvironment in multiple pathologies. *Int. Rev. Cell Mol. Biol.* **297**, 83–116 (2012)
11. Kuse, M., Sharma, T., Gupta, S.: Classification scheme for lymphocyte segmentation in H&E stained histology images. In: Ünay, D., Çataltepe, Z., Aksoy, S. (eds.) ICPR 2010. LNCS, vol. 6388, pp. 235–243. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17711-8_24
12. Lahiani, A., Gildenblat, J., Klamann, I., Navab, N., Klaiman, E.: Generalising multistain immunohistochemistry tissue segmentation using end-to-end colour deconvolution deep neural networks. *IET Image Process.* **13**(7), 1066–1073 (2019)
13. Lahiani, A., Klaiman, E., Grimm, O.: Enabling histopathological annotations on immunofluorescent images through virtualization of hematoxylin and eosin. *JPI* **9**, 1 (2018)

14. Rivenson, Y., Wang, H., Wei, Z., Zhang, Y., Gunaydin, H., Ozcan, A.: Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue. arXiv preprint [arXiv:1803.11293](https://arxiv.org/abs/1803.11293) (2018)
15. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arxiv 2016. arXiv preprint [arXiv:1607.08022](https://arxiv.org/abs/1607.08022) (2016)
16. Zhu, J.Y.: Freezing instance normalization during test phase? (2017). <https://github.com/junyanz/CycleGAN/issues/32>. Accessed 01 Oct 2018
17. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)