SELF ADVERSARIAL ATTACK AS AN AUGMENTATION METHOD FOR IMMUNOHISTOCHEMICAL STAININGS

Jelica Vasiljević[⋆,†,‖] *Friedrich Feuerhake*[‡,✠] *Cédric Wemmert*[⋆] *Thomas Lampert*[⋆]

[⋆]ICube, University of Strasbourg, France [†]University of Belgrade, Serbia
[‖]Faculty of Science, University of Kragujevac, Serbia
[‡] Institute of Pathology, Hannover Medical School, Germany [✠] University Clinic, Freiburg, Germany

ABSTRACT

It has been shown that unpaired image-to-image translation methods constrained by cycle-consistency hide the information necessary for accurate input reconstruction as imperceptible noise. We demonstrate that, when applied to histopathology data, this hidden noise appears to be related to stain specific features and show that this is the case with two immunohistochemical stainings during translation to *Periodic acid-Schiff (PAS)*, a histochemical staining method commonly applied in renal pathology. Moreover, by perturbing this hidden information, the translation models produce different, plausible outputs. We demonstrate that this property can be used as an augmentation method which, in a case of supervised glomeruli segmentation, leads to improved performance.

Index Terms— digital pathology, image-to-image translation, cycle-consistency, self adversarial attack

1. INTRODUCTION

One of the greatest obstacles for the effective application of deep learning techniques to digital pathology is the shortage of high-quality annotated data. The annotation process itself is time consuming and expensive as expert domain knowledge is required for most complex annotations and alternative approaches such as crowd sourcing are limited by the need of specific task design and intensive training [1]. The problem is complicated by tissue appearance variability, which can occur due to different stainings, patients, procedures between different laboratories, and/or the microscope and imaging device [2]. All of this imposes a domain shift to which deep models are very sensitive [3], making their application difficult in clinical practice.

Due to their ability to produce high quality visual outputs, Generative Adversarial Networks (GANs) [4] have recently been applied to medical imaging in general and digital pathology. Finding use in histopathology to reduce intra-stain variance [5]; for virtual staining [6,7]; and for augmentation [8,9]. Virtual staining has shown that an unpaired image-to-image translation GAN is able to translate between stains. The same tissue can be (artificially) stained in multiple stainings, which is hard (or even impossible) in realty [6]. CycleGAN is the most popular and promising unpaired image-to-image translation approach [10, 11]. Nevertheless, the less obvious limitations of such methods are rarely addressed in the medical imaging literature [6]. For example, such models produce realistic translations between very different stains, which leads to the question: how is the model able to place stain related markers that are not present in the original stain? This article moves towards answering this question.

The computer vision community has recently shown with natural images that the cycle-consistency of CycleGANs renders them prone to self-adversarial attack [12]. The CycleGAN (Fig. 1) is composed of two translators: one from staining A to B, G_{AB}, and another from B to A, G_{BA}. The cycle consistency enforces that the output of G_{BA} matches the input of G_{AB}. To achieve this, each translator is forced to hide imperceptible information in its output. Our first contribution is to show that the hidden noise has a specific meaning in histopathology - it encodes stain-related markers. By perturbing this hidden noise, differently positioned stain-related markers are produced in the translated image (leaving the underlying tissue structure untouched).

This is exploited to introduce a new augmentation technique that increases the variability of stain-specific markers in histopathological data, with the goal of increasing a model's robustness when trained for non-stain-related tasks. We show that this increases the generalisation performance of a supervised deep learning approach for glomeruli segmentation, which forms this article's second contribution.

We explore the mapping between Periodic acid-Schiff (PAS), a routine staining in renal pathology that is applied for general diagnostic purposes, and two immunohistochemical stainings (CD68 for macrophages and CD34 for blood vessel endothelium), which are performed for research or specific diagnostic purposes. Separate CycleGAN models are trained to translate between PAS stained tissue patches and each of the immunohistochemical stainings.

Section 2 of this article presents adversarial attacks in stain transfer; Section 3 presents the new augmentation method and its evaluation; and Section 4 our conclusions.

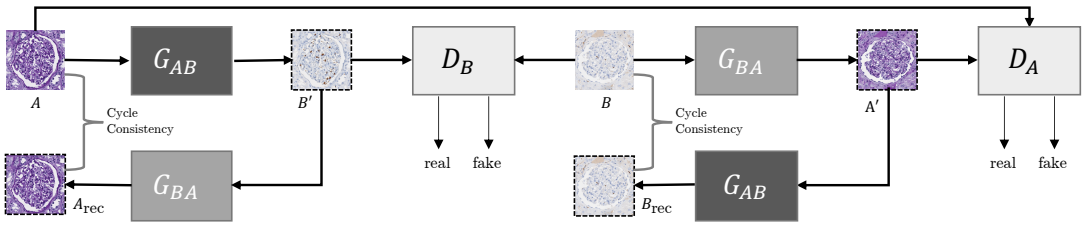

Fig. 1: CycleGAN approach (with PAS and CD68 staining examples). Framed images are translated, i.e. 'fake'.

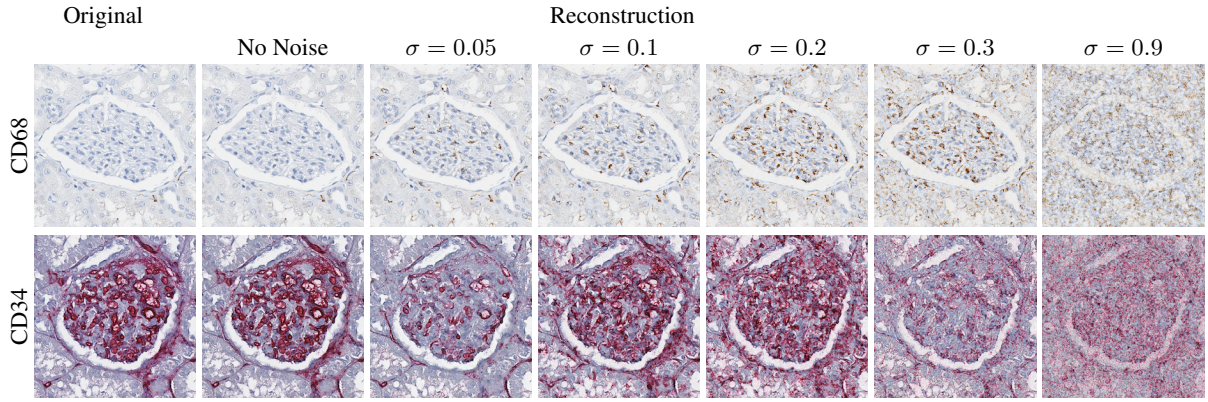

Fig. 2: Generating variation by adding noise, the images are reconstructions of CD68/CD34 $\rightarrow$ PAS + $\mathcal{N}(0, \sigma)$ $\rightarrow$ CD68/CD34.

2. STAIN TRANSFER SELF ADVERSARIAL ATTACK

Given samples of two histopathological stains $a \sim A$ and $b \sim B$, the goal is to learn two mappings (translators) $G_{AB} : a \sim A \rightarrow b \sim B$ and $G_{BA} : b \sim B \rightarrow a \sim A$. In order to do so, two adversarial discriminators D_A and D_B are jointly trained to distinguish between translated and real samples, i.e. D_A aims to distinguish between real samples $a \sim A$ and B translated to A ($a' = G_{BA}(b), b \sim B$), while D_B performs the equivalent task for $b \sim B$ and $b' = G_{AB}(a), a \sim A$. In addition to the adversarial loss [4, 10], the learning process is regularised by a cycle-consistency loss $\mathcal{L}_{cyc}$ that forces the generators to be consistent with each other [10], such that

$$\mathcal{L}_{cyc}(G_{AB}, G_{BA}) = \mathbb{E}_{a\sim A}[\|G_{BA}(G_{AB}(a)) - a\|_1] + \mathbb{E}_{b\sim B}[\|G_{AB}(G_{BA}(b)) - b\|_1]. \quad (1)$$

In addition to the Haematoxylin counterstain (common to all the stainings studied herein) that highlights cell nuclei, CD68 marks a protein exclusively produced by macrophages, and CD34 stains a protein specific to the endothelial cells of blood vessels. PAS, as a chemical reaction staining glycolysated proteins in general, can highlight some parts of macrophages (co-located but not overlapping with CD68), the basal lamina of blood vessels (co-located with CD34), and other structures not highlighted by either CD68 nor CD34 that contain glycolysated proteins. During translation from PAS to CD68, the model could choose not to produce macrophages (which would be a valid CD68 sample) but D_{CD68} would easily discriminate real/fake images based on this absence, and therefore the model is biased to deduce their position from information present in PAS. Conversely, i.e. CD68 $\rightarrow$ PAS, the model should induce the presence of glycolysated proteins, for which CD68 is not specific. As such, the translation process is a many-to-many mapping (equivalent arguments can be made for PAS $\leftrightarrow$ CD34).

The cycle-consistency constraint Eq. (1), Fig. 1 forces compositions of translations ($A \rightarrow B \rightarrow A$) to accurately reconstruct the input. Taking CD68 $\rightarrow$ PAS $\rightarrow$ CD68 for example, macrophages in the reconstructed image should be in the same locations as those in the original, which implies that the intermediate PAS image contains additional information defining these macrophage positions. Bashkirova et al. [12] recently showed that information necessary for perfect reconstruction takes the form of imperceptible low amplitude, high frequency noise in order to fool the discriminator, and recent literature [12, 13] names this a self-adversarial attack. Since PAS does not contain information specific to macrophages/blood vessels this is likely to be the case.

2.1. Dataset

Tissue samples were collected from a cohort of 10 patients who underwent tumor nephrectomy due to renal carcinoma. The kidney tissue was selected as distant as possible from the tumors to display largely normal renal glomeruli, some samples included variable degrees of pathological changes

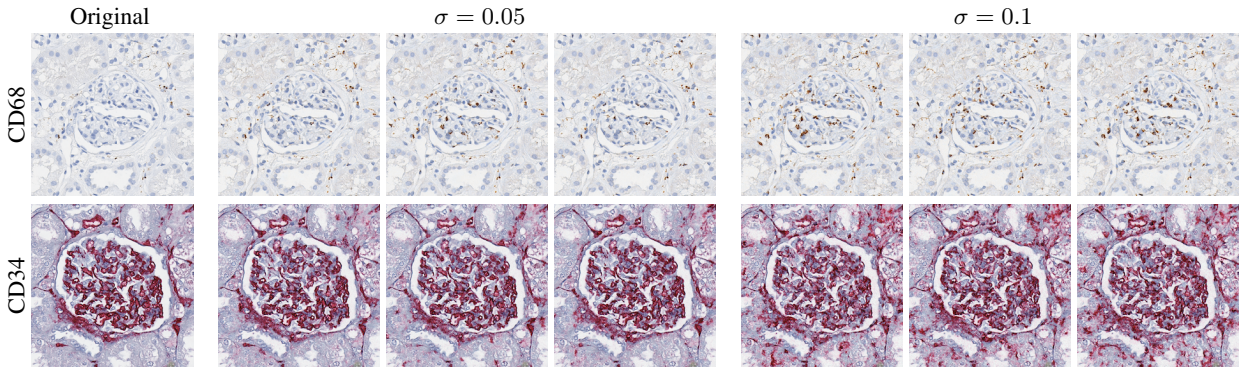

Fig. 3: Effects of additive Gaussian noise with the same standard deviation, the images are reconstructions of CD68/CD34 $\rightarrow$ PAS + $\mathcal{N}(0, \sigma)$ $\rightarrow$ CD68/CD34

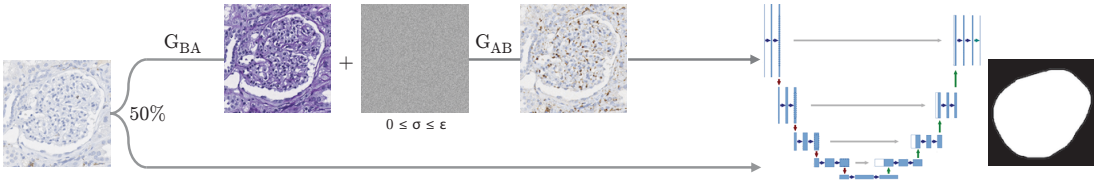

Fig. 4: Proposed augmentation approach.

such as full or partial replacement of the functional tissue by fibrotic changes ("scerosis") reflecting normal age-related changes or the renal consequences of general cardiovascular comorbidity (e.g. cardial arrhythmia, hypertension, arteriosclerosis). The paraffin-embedded samples were cut into $3\,\mu\text{m}$ thick sections and stained with either PAS or immunohistochemistry markers CD34 and CD68 using an automated staining instrument (Ventana Benchmark Ultra). Whole slide images (WSIs) were acquired using an Aperio AT2 scanner at $40\times$ magnification (a resolution of $0.253\,\mu\text{m/pixel}$). All glomeruli (healthy, partially sclerotic, and completely sclerotic) in each WSI were annotated and validated by pathology experts using Cytomine [14]. The dataset was divided into 4 training, 2 validation, and 4 test patients.

For CycleGAN training, 5000 random 508×508 pixel patches were extracted from the training patients and scaled to the range $[-1, 1]$. The model's architecture (9 ResNet blocks) and training details were taken from the original article [10].

2.2. Results and Analysis

Figure 2 shows that translation output (i.e. reconstructed input, B_{rec}) variance is directly proportional to the level of additive noise and Fig. 3 shows that different translations result from varying noise of the same standard deviation.

As such, they give evidence to support that when translating between immunohistochemical and histochemical stains, imperceptible noise is present in the intermediate translation and this contains information about stain-related markers (this is related to macrophages marked in brown, and blood vessel endothelium marked in red in CD68 and CD34 respectively). Thus, changing the encoded noise changes the reconstruction of stain related markers. This noise can be perturbed by introducing additive zero-mean Gaussian noise to the intermediate translation [12]. The amount of stain related characteristics can be controlled through the Gaussian's standard deviation. The physical accuracy of the resulting stain-related markers remains an open question, but the fact that they are positioned in plausible locations opens the possibility of exploiting them to reduce a model's sensitivity to such stain related markers.

It should be noted that the amount of additive noise is stain dependent: a standard deviation, σ, of 0.3 produces realistic CD68, but a noisy CD34, output. As the translation process hides non-overlapping inter-stain information, the intermediate stain likely determines which information is encoded.

3. SELF ADVERSARIAL ATTACK AUGMENTATION

CyleGANs are unsupervised and unpaired, therefore training them does not require additional annotation effort but does require additional stain samples. PAS is a routine stain so these should be readily available. The fact that intermediate representations contain imperceptible noise related to stain features can be used to increase the variance of existing datasets by randomly perturbing the noise. CycleGAN is incapable of performing geometrical changes [10, 11], so cannot change the morphological structures in these images, e.g. it will not remove glomeruli. Thus, it is safe to use as an augmentation technique in supervised problems related to morphologically

Stain		Baseline			Noise Augmented		
		F_1	Precision	Recall	F_1	Precision	Recall
CD68	10% - 53	0.739 (0.018)	0.754 (0.047)	**0.728** (0.034)	**0.767** (0.036)	**0.832** (0.053)	0.713 (0.044)
	30% - 159	0.812 (0.026)	0.839 (0.038)	0.788 (0.038)	**0.828** (0.026)	**0.848** (0.065)	**0.812** (0.017)
	60% - 317	0.831 (0.024)	0.812 (0.037)	**0.852** (0.014)	**0.856** (0.017)	**0.888** (0.026)	0.826 (0.021)
	100% - 529	0.853 (0.018)	0.849 (0.024)	**0.858** (0.020)	**0.878** (0.007)	**0.899** (0.023)	**0.858** (0.010)
CD34	10% - 57	0.837 (0.017)	0.770 (0.033)	**0.919** (0.009)	**0.839** (0.035)	**0.778** (0.061)	0.913 (0.008)
	30% - 170	0.877 (0.012)	0.841 (0.030)	**0.917** (0.012)	**0.890** (0.011)	**0.867** (0.023)	0.916 (0.009)
	60% - 341	0.882 (0.008)	0.840 (0.015)	**0.927** (0.005)	**0.901** (0.007)	**0.884** (0.019)	0.919 (0.010)
	100% - 568	0.888 (0.015)	0.849 (0.033)	**0.931** (0.010)	**0.903** (0.006)	**0.888** (0.014)	0.919 (0.009)

Table 1: Quantitative results, standard deviations are in parentheses, # of glomeruli training patches follow the data percentages.

consistent structures, in this case glomeruli segmentation.

The proposed augmentation process is described in Fig. 4. Let us denote PAS as A and an immunohistochemical stain as B. During supervised training of a model on B (e.g. for glomeruli segmentation), each sample b_i is first translated to PAS, A', using the trained CycleGAN generator G_{BA}, with a probability of 50%. Next, zero-mean Gaussian noise with standard deviation σ is added to the intermediate translation, which is translated back to B using G_{AB}, where $\sigma \in (0, \epsilon_{\text{stain}}]$ with uniform probability. The value ϵ_{stain} is determined for each staining separately. As such, the input is altered by the arbitrary appearance of stain related markers and the supervised model is forced to be less sensitive to their appearance.

The U-Net [15] gives state-of-the-art performance in glomeruli segmentation [16] and is adopted herein. The architecture and training details are the same as in [16].

3.1. Dataset

The U-Net training set comprised all glomeruli from the 4 training patients - 529 for CD68 and 568 for CD34 - and 3685 and 3958 tissue patches respectively (to account for the variance of non-glomeruli tissue). The validation sets (2 patients) were composed of 524 and 598 glomeruli patches, and 3650 and 4168 negative patches for CD68 and CD34 respectively. Patches are standardised to $[0, 1]$ and normalised by the mean and standard deviation of the training set. To evaluate the augmentation's effect with few data samples, each training set is split into 5 folds containing 10%, 30%, and 60% of each class taken at random. A separate random 10% subset of the training data is extracted to choose ϵ_{stain}. All models are trained for 250 epochs, the best performing model on the validation partition is kept, and tested on the 4 held-out test patients. The average F_1-score and standard deviation is reported.

3.2. Choosing the Level of Noise

As with all augmentation techniques, a parameter value must be chosen. In this case it is the noise level ϵ_{stain}. Since the problem being addressed is supervised, ϵ_{stain} can be optimised experimentally, however, it could be chosen by manually validating the reconstructions. A grid search was conducted on a separate dataset partition containing a random 10% subset of each class. The range $\epsilon_{\text{stain}} \in [0.01, 0.05, 0.1, 0.3, 0.5, 0.9]$ was tested by averaging 3 repetitions. It was found that adding noise in the range that produces realistic output improves upon the baseline ($\epsilon_{\text{CD68}} \leq 0.3$ and $\epsilon_{\text{CD34}} \leq 0.1$), confirming that the parameter can be chosen manually. Nevertheless, the best value should be determined for each stain to maximise F_1 score and these were found to be $\epsilon_{\text{stain}} = 0.05$.

3.3. Results

Table 1 presents the baseline and noise augmented results with varying amounts of data. The proposed augmentation improves F_1 scores unanimously due to increased precision. Recall does not improve since no new task-specific information is added, e.g. glomeruli shape or positional variance. Since stain related markers are not indicative of glomeruli in general, the model should largly ignore them. However, fibrotic and sclerotic glomeruli are present, to which the model can wrongly associate a specific pattern or marker. For example, fibrotic changes are associated with CD68 positive macrophages [17] and a loss of CD34 positive vascular structures. Overemphasising immunohistochemical variations via augmentation biases the model to other properties, decreasing recall but disproportionately increasing precision.

4. CONCLUSION

This article studies CycleGAN self-adversarial attacks in translating immunohistochemical stainings to PAS. It presents evidence that imperceptible noise induced by cycle consistency relates to immunohistochemical markers. Perturbing this hidden information causes these markers to appear in different, plausible locations although their physical meaning remains an open question. This finding is used in an augmentation method to increase segmentation accuracy by reducing false positive rates and therefore increasing F_1 scores. We also found that the translations result in rich and realistic images, which may provide cellular information and future work will take this direction by investigating their physical meaning, in addition to analysing different reference stains.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Hannover Medical School (Date 12/07/2015, No. 2968-2015).

6. ACKNOWLEDGMENTS

This work was supported by: ERACoSysMed and e:Med initiatives by the German Ministry of Research and Education (BMBF); SysMIFTA (project management PTJ, FKZ 031L-0085A; Agence National de la Recherche, ANR, project number ANR-15—CMED-0004); SYSIMIT (project management DLR, FKZ 01ZX1608A); and the French Government through co-tutelle PhD funding. We thank Nvidia Corporation for donating a Quadro P6000 GPU and the *Centre de Calcul de l'Université de Strasbourg* for access to the GPUs used for this research. We also thank the MHH team for providing high-quality images and annotations, specifically Nicole Kroenke for excellent technical assistance, Nadine Schaadt for image management and quality control, and Valery Volk and Jessica Schmitz for annotations under the supervision of domain experts.

7. REFERENCES

[1] A. Grote et al., "Crowdsourcing of histological image labeling and object delineation by medical students," *IEEE Trans Med Imaging*, vol. 38, pp. 1284–1294, 2018.

[2] P. Leo et al., "Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images," *J Med Imaging*, vol. 3, no. 4, 2016.

[3] G. Csurka, "A comprehensive survey on domain adaptation for visual applications," in *Domain adaptation in computer vision applications*, chapter 1, pp. 1–35. Springer, 2017.

[4] I. Goodfellow et al., "Generative adversarial nets," in *NIPS*, 2014, pp. 2672–2680.

[5] D. Tellez et al., "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *Med Image Anal*, vol. 58, pp. 101544, 2019.

[6] C. Mercan et al., "Virtual staining for mitosis detection in breast histopathology," in *ISBI*, 2020.

[7] A. Lahiani et al., "Virtualization of tissue staining in digital pathology using an unsupervised deep learning approach," *ECDP*, vol. 11435, 2019.

[8] J. Vasiljević et al., "Achieving histopathological stain invariance by unsupervised domain augmentation using generative adversarial networks," *Under Review*.

[9] A. Quiros et al., "Pathologygan: Learning deep representations of cancer tissue," in *PMLR*, 2020, vol. 121, pp. 669–695.

[10] J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2242–2251.

[11] M. Gadermayr et al., "Which way round? A study on the performance of stain-translation for segmenting arbitrarily dyed histological images," 2018, pp. 165–173.

[12] D. Bashkirova et al., "Adversarial self-defense for cycle-consistent GANs," in *NeurIPS*, 2019, pp. 635–645.

[13] C. Chu et al., "CycleGAN, a master of steganography," in *NIPS workshop on Machine Deception*, 2017.

[14] R. Marée et al., "Collaborative analysis of multi-gigapixel imaging data using cytomine," *Bioinformatics*, vol. 32, no. 9, pp. 1395–1401, 2016.

[15] O. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[16] T. Lampert et al., "Strategies for training stain invariant CNNs," in *ISBI*, 2019, pp. 905–909.

[17] M. Adler et al., "Principles of cell circuits for tissue repair and fibrosis," *iScience*, vol. 23, no. 2, pp. 100841, 2020.