Style Match: Reducing the Scanner-induced Domain Gap in Mitosis Detection using Style Transfer Alignment

Ben Conrad
University of Amsterdam
benjamin.conrad@student.uva.nl

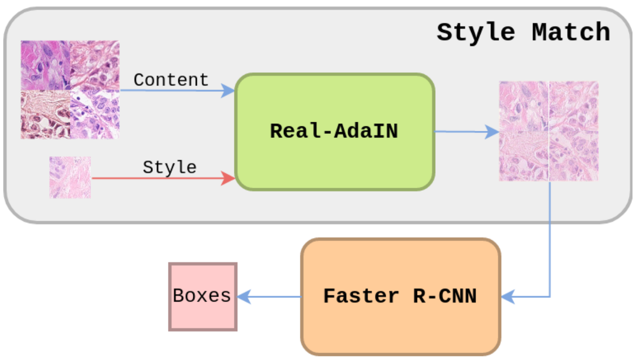


Figure 1: An illustration of style match applied to mitosis detection. Given WSI patches that we want to perform detection on, the patches are passed through the Real-AdaIN style transfer model as content images along with reference style images. The outputs of the model are augmented versions of the original patches that share a similar color and textural appearance to the style images. The detection model is trained with the augmented images, and during testing the augmentation is also applied to ensure all inputs to the detector have a similar appearance.

1. Introduction

Deep learning has led to significant advances in histopathology in recent years. Such models are frequently used as supplementary tools for pathologists to perform faster and more accurate assessments [3]. Deep learning-based models, however, are known to perform worst when applied to data that differs from the model's training set distribution [18]. In histopathology, differences in tissue preparation, staining and model of the microscopy whole slide image scanner causes scans to have different color and textural appearances that most models struggle to generalize on [3]. In many fields, this domain generalization issue can be improved by training on large and diverse datasets, however, in histopathology this is not economically feasible since medical specialists are required to annotate the data.

Since deep learning models perform best when evaluating data from the same distribution as their training set, we look at using data augmentations to transform all input images to have a visually similar appearance to the training data. To achieve this, we propose style match, a data augmentation method that uses a neural style transfer model to transform both training and test data into a shared stylistic distribution. Style match consists of two parts. First, we train an arbitrary style transfer model on unlabeled data to augment any image to have the same color and texture appearance as a given reference style image. Second, we train a model for our target task using annotated data that has first been augmented by the style transfer model. We find that style transfer models can generalize to unseen input domains, therefore, at evaluation, we also augment all inputs using the style transfer model for better domain alignment between training and test time.

In this work, we apply style match to mitotic figure detection using the MIDOG dataset [2], which focuses on domain generalization between different models of microscopy whole slide image scanners. For detection, data augmentations must preserve the original structure of the input image to not impact localization performance. Therefore, we propose the Real-AdaIN style transfer model, an improvement to AdaIN [7] which better preserves structural content and produces fewer artifacts when applied to whole slide images.

We find that style match regularly improves the generalization performance of detection models when evaluating on slides from unseen scanners and is comparable to other domain generalization augmentation approaches.

Our contributions can be summarized as follows:

- We propose Real-AdaIN, an extension to the AdaIN style transfer architecture that can accurately preserve structural content and transfer the stylistic appearance between whole slide images.
- Using the observation that style transfer models generalize well to unseen domains, we introduce style match, a data augmentation approach where a style

transfer model is applied to images during training and evaluation to align all inputs into a common stylistic distribution and improve domain generalization.

- We apply style match to the MIDOG mitotic figure detection dataset, where we find that it significantly improves detectors' domain generalization capabilities over standard training and performs comparably to other augmentation approaches.

The code and pre-trained models for this work are available at `https://github.com/bwconrad/style-match-mitosis-detection`.

2. Dataset

We first describe the training and evaluation datasets to give context to the different data domains we are considering in this work. The MItosis DOmain Generalization (MIDOG) dataset [2] consists of 200 whole slide images (WSI) of human breast cancer tissue. Scans have been digitized from 4 scanning systems (50 slides per scanner): Hamamatsu XR NanoZoomer 2.0, Hamamatsu S360, Aperio ScanScope CS2 and Leica GT450. Examples from each scanner can be seen in Figure 2. Positive and hard-negative mitotic figures annotations are provided for slides from the first 3 scanners, while no annotations are provided for the Leica GT450 slides. Unless specified otherwise, from each of the scanners, we reserve 10 slides for our test set. From now on, we will refer to the Hamamatsu XR NanoZoomer 2.0, Hamamatsu S360, Aperio ScanScope CS2 and Leica GT450 as scanners 1, 2, 3 and 4 respectively.

3. Real-AdaIN

3.1. Methodology

Style match improves domain generalization by translating images into a common color and texture distribution. This translation is done using an arbitrary style transfer model [7], which takes as input a content image I_c and a style image I_s and generates a style transferred output image I_{st} that has the stylistic appearance of I_s while preserving the same structural content as I_c. In this work, we build off the AdaIN style transfer model [7] which consists of a VGG-19 [17] encoder E and a mirrored decoder D. At the bottleneck, an adaptive instance normalization layer (AdaIN) is used to afflinely transform the encoded content image feature map to align with the style image before being decoded. Given encoded content image features $E(I_c) = f_c$ and style image features $E(I_s) = f_s$, the AdaIN layer transforms the channel-wise mean and variance of f_c as follows:

$$\text{AdaIN}(f_c, f_s) = \sigma(f_s)(\frac{f_c - \mu(f_c)}{\sigma(f_c)}) + \mu(f_s)$$

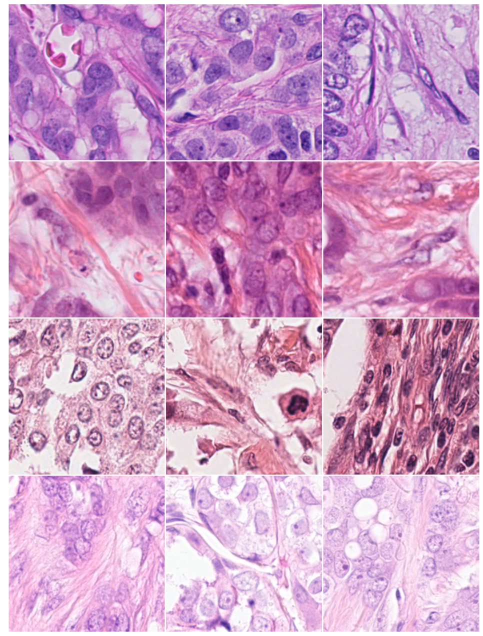


Figure 2: Sample patches from each of the scanners in the MIDOG dataset. Scanner models from top to bottom are Hamamatsu XR NanoZoomer 2.0 (scanner 1), Hamamatsu S360 (scanner 2), Aperio ScanScope CS2 (scanner 3) and Leica GT450 (scanner 4).

where μ is the mean operation and σ is the variance operation.

For detection, it is essential that the style transfer model accurately preserves the structure of the content image to avoid any misalignment of the bounding box annotations and loss of performance due to lost fine details. To account for this, we incorporate approaches from photorealistic style transfer [1], specifically bottleneck feature aggregation and skip connections. Instead of decoding using only the output of the last encoder layer, bottleneck feature aggregation (BFA) downscales and concatenates the features from each encoder layer before applying AdaIN. The aggregation allows the decoder to better preserve fine details by incorporating features from early layers. Skip connections between the encoder and decoder layers similarly helps the model with fine details by directly passing feature maps at multiple scales to the decoder. Applying skip connections by simply concatenating the encoder features to the decoder results in a diminished style transfer effect since the style information from the encoder is leaked to the decoder [1]. To avoid this, instance normalization [20] is applied to the encoder feature maps before each skip connection. We refer to the resulting AdaIN style transfer model with BFA and skip connections as Real-AdaIN.

To train Real-AdaIN, we use the content [8] and style

[5, 7] loss functions. The content loss $\mathcal{L}_c$ measures the perceptual similarity between the content image and the style transfer output by the L2 distance between the encoder feature maps. Following [7], we apply the AdaIN layer to the content features before calculating the loss. The content loss is defined as:

$$\mathcal{L}_c = \|f_{st} - \text{AdaIN}(f_c, f_s)\|_2$$

where $E(I_{st}) = f_{st}$ are the encoder features of the style transferred image.

The style loss $\mathcal{L}_s$ measures the distance between feature statistics to enforce that the transformation applied by the AdaIN layer is maintained by the decoder. The style loss is defined as:

$$\mathcal{L}_s = \sum_{i=1}^{L} \|\mu(\phi_i(I_s)) - \mu(\phi_i(I_{st}))\|_2 + \sum_{i=1}^{L} \|\sigma(\phi_i(I_s)) - \sigma(\phi_i(I_{st}))\|_2$$

where ϕ_i is the output of the ith layer of the VGG-19 encoder. For our model, we use the `relu1_1`, `relu2_1`, `relu3_1`, `relu4_1` layers.

During training, the encoder is kept frozen with ImageNet [4] pre-trained weights, and the decoder is trained using the equally weighted sum of the two loss functions:

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_s$$

3.2. Experiments

3.2.1 Training Details

We train models on 256×256 random patches from WSIs in the MIDOG dataset. Slides from scanner 1 are used for content images and scanner 4 for style images. The Adam optimizer [9] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a batch size of 16 is used for training. The initial learning rate is set to $1e-4$ and decays following a cosine annealing schedule [12] for 80,000 iterations. Random horizontal flip augmentations are applied during training.

3.2.2 Architecture Ablation Study

Figure 3 shows example outputs from the AdaIN and Real-AdaIN models. Both models successfully transfer the color distribution of the style image to the content image, but AdaIN produces strong checkerboard artifacts and lacks fine details. Real-AdaIN eliminates these artifacts and produces a more natural output.

To quantify the content preservation improvements of Real-AdaIN, we follow previous work [23] and measure the structural similarity index measure (SSIM) between the content and style transferred images. Results using the scanner 1 and scanner 4 test sets for content and style images respectively can be seen in Table 1. We see that adding BFG and skip connections to AdaIN improves the SSIM by double, going from 0.362 to 0.722. The content loss also drops a significant amount, in line with the SSIM increase, while the style loss only slightly increases over AdaIN, showing that the architecture changes do not hinder the model's stylization capabilities.

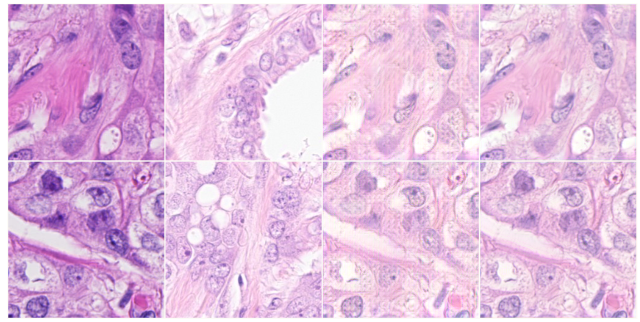


Figure 3: Example style transfer outputs. From left to right: content image, style image, output of AdaIN and output of Real-AdaIN. Real-AdaIN removes all checkerboard artifacts found in AdaIN and produces a clear and realistic output.

	SSIM	$\mathcal{L}_c$	$\mathcal{L}_s$
AdaIN	0.362	0.437	0.095
+BFG	0.594	0.304	0.096
+Skip Connections	0.722	0.245	0.106

Table 1: Quantitative results from the architecture ablation study. Adding BFG and then skip connections to AdaIN steadily improves the model's SSIM and content loss while only resulting in a small increase in the style loss.

3.2.3 Generalization to Unseen Domains

Next, we test how well Real-AdaIN generalizes when using slides from scanners that were not seen during training.

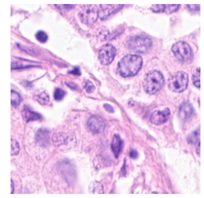
Figure 4 shows example outputs when using the same style image from scanner 4 and different content images from each of the 4 scanners. In all examples, the model generalizes very well, producing a similar visual appearance as the style image. Even when using content images from scanner 2 or 3, which the model has never seen during training, the style transfer outputs maintain the same quality. These generalization results are not surprising since the AdaIN layer first normalizes the content image's features before transforming them using the style image's statistics, allowing the input to the decoder to always be from a similar distribution independent of the content image.

This argument can also be applied in the opposite case to explain why the model does not generalize when using style images from unseen scanners, as shown in Figure 5. Here the model always produces outputs similar in appearance to scanner 4 regardless of the style image used since the decoder has not learned to adapt to different input domains during training.

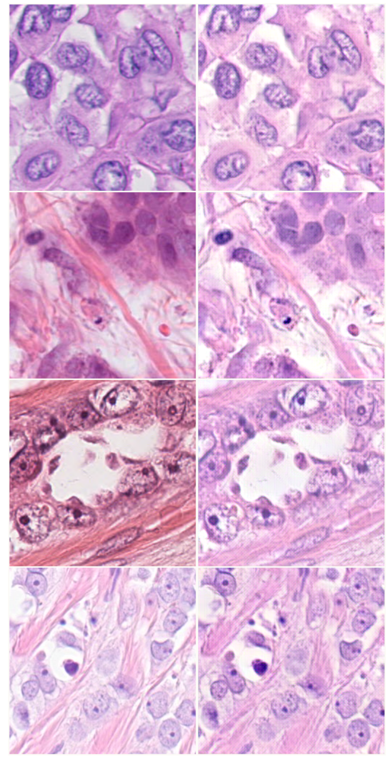
3.2.4 Scanner Classification

After observing the strong generalization capabilities of style transfer models, we perform preliminary experiments with a classification network to see how style transfer can be used to translate images between scanner domains. Specifically, we trained a ResNet-18 [6] on the MIDOG dataset to classify which of the 4 scanners a slide patch was captured from. The network is trained for 5 epochs on 256×256 random patches using Adam [9] with a learning rate of $1e-4$ and batch size of 8. For this experiment, we take 25 slides for training and 25 for testing from each scanner.

The network easily solves the classification task, achieving 100% test set accuracy with an average confidence of 0.988 when evaluating the original images. We next use the Real-AdaIN model trained with style images from scanner 4 to augment the test set images before evaluation. Applying the augmentation now causes the network to classify all images as from scanner 4 while only slightly lowering the model's average confidence to 0.968. This result suggests that, from the point of view of a deep neural network, style transfer can be used to change the domain in which a data sample belongs to.



(a) Style



(b) Content (c) Output

Figure 4: Examples when changing the content image scanner of the Real-AdaIN style transfer model. The top to bottom order of 4b and 4c is scanner 1, 2, 3 and 4. The model is trained with content images from scanner 1 and style images from scanner 4, and is evaluated with content images from each of the 4 scanners. Real-AdaIN produces realistic results even on unseen scanner types that it has not encountered during training.

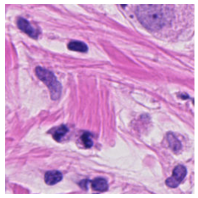
4. Style Match

Based on the experimental results in Section 3.2, we propose style match, a data augmentation approach where input images at both training and test time are transformed using a style transfer model into a common color and texture distribution. By applying the augmentation during training and testing, the target task model is only given data that comes from a similar stylistic data distribution, which decreases the domain gap when evaluating samples from an unseen domain. An illustration of the approach is shown in 1. Style match is a general framework which can use any arbitrary style transfer model to train any target task. Here we will explore applying style match to train a mitotic figure detector using the Real-AdaIN model.

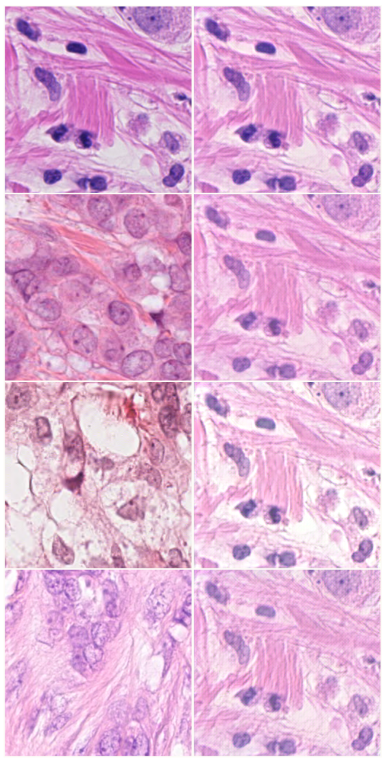
4.1. Mitosis Detection using Style Match

Style match is simply a form of data augmentation that does not change the standard detection training procedure, aside from prepossessing all input slides using the Real-AdaIN model. At each iteration, two batches are loaded: one with samples that we want to apply the detection model on, which will be the content images, and another with samples used as the style images. The two batches are processed through the Real-AdaIN model, and its output is then passed to the detection model to generate bounding box results.

A key difference between style match and most data augmentations is that it is used both during training and evaluation. This ensures that all inputs the detector evaluates are from a similar distribution as the data it was originally trained with. For histopathology slides, this means all slides



(a) Content



(b) Style (c) Output

Figure 5: Examples when changing the style image scanner of the Real-AdaIN style transfer model. The top to bottom order of 5a and 5c is scanner 1, 2, 3 and 4. The model is trained with content images from scanner 1 and style images from scanner 4, and is evaluated with style images from each of the 4 scanners. Real-AdaIN does not generalize to unseen style scanners, producing outputs with a similar appearance to scanner 4 regardless of the style image input.

have the visual appearance of being taken from the same model of scanner and using the same staining procedure. Style match is not limited to a single scanner style and can translate slides into multiple style distributions given that the style transfer model is trained with style images from multiple scanner types. This approach is appealing as it adds more variation to the training data, which generally increases performance in most tasks [16].

4.2. Experiments

4.2.1 Training & Evaluation Details

We train Faster R-CNN detectors [15] with ResNet-50 [6] feature pyramid [10] backbones. The Torchvision implementation [14] is adopted with its default parameters. We train using SGD with a momentum of 0.9 and a batch size of 4. Training is done for 40,000 iterations with an initial learning rate of $1e-3$ and reduced to $1e-4$ after 20,000 iterations. The model is trained on 256×256 patches from the MIDOG dataset. Since the majority of random patches do not include any annotations, we sample 50% of patches with at least one bounding box annotation while the rest are randomly sampled. Random horizontal flips and $90°$ rotations are also applied during training.

The Real-AdaIN model is trained using the same settings as described in Section 3.2.1 except that slides from all 4 scanners are used as content and style images. When training and evaluating using style match, by default, we use slides from scanner 2 for style images.

At evaluation, entire WSIs are processed in a sliding window fashion. The detector is independently applied to 256×256 windows with neighbouring windows overlapping by 32 pixels. Non-maximum suppression with a threshold of 0.2 and a confidence threshold of 0.8 is used to filter the final results.

4.2.2 Comparison with Standard Training

To see how style match helps generalization to novel scanner domains, we train detection models with slides from a single scanner and evaluate their performance on each of the dataset's 3 scanners (MIDOG dataset does not contain annotations for scanner 4). We investigate 3 scenarios: 1) training and evaluating on the original images, 2) training on the original images and evaluating on style matched images using style images from the detector's training set scanner (e.g. for a detector trained on scanner 1 slides, we also use style images from scanner 1), 3) training and evaluating on style matched images.

Table 2 shows the F1@0.5 scores of the experiments for scenarios 1, 2, and 3. When training and evaluating slides from the same scanner (highlighted in red), standard training always achieves the best performance. Applying style match when there is no domain shift lowers the detector's F1 score, which is expected since the style transfer model loses some image quality as shown by the SSIM results in Section 3.2.2. When we introduce a domain shift by evaluating on an unseen scanner, standard detectors consistently degrade in performance by up to half compared to when there is no domain shift. Training with style match similarly observes a performance dip, however, in 5/6 domain shift experiments, this reduction is signifi-

		Test		
		Scanner 1 (XR)	Scanner 1 (S360)	Scanner 3 (CS2)
Train	Scanner 1 (XR)	0.538/0.513/0.517	0.373/0.433/0.542	0.500/0.498/0.582
	Scanner 2 (S360)	0.341/0.114/0.359	0.733/0.654/0.678	0.539/0.350/0.543
	Scanner 3 (CS2)	0.340/0.421/0.394	0.661/0.605/0.545	0.638/0.560/0.591

Table 2: Test set F1@0.5 scores of Faster R-CNN models on the MIDOG dataset. Each model is trained on slides from a single scanner type (rows) and is evaluated on slides from each of the 3 scanners (columns). Each cell contains the results from the following 3 scenarios: 1) training and evaluating on the original images, 2) training on the original images and evaluating on style matched images using style images from the detector's training set scanner, 3) training and evaluating on style matched images. Red indicates the best results for that test scanner and blue indicates the best result when training on a different scanner type than the test set.

cantly smaller than with standard training.

Applying style match only during evaluation and not during training is the worst performing approach in 6/9 experiments, often significantly worse than the other scenarios. The exception to this is when training on scanner 3 and evaluating on scanner 1 where it performs the best. Since style match cannot perfectly replicate the appearance of the target style scanner, a domain gap exists between real and artificially augmented images, making this approach less consistent than applying style match during both training and evaluation.

We can see that the effectiveness of style match varies drastically depending on the model's training set. When using scanner 1, style match significantly outperforms standard detectors in domain shift experiments by up to 0.169 points, while only receiving a performance hit of 0.022 points when also testing on scanner 1. On the contrary, when training on scanner 2, we see that models generalize much more poorly and style match only sees marginal improvements over standard training. Since we use style images from scanner 2 in all our experiments, we hypothesize that the choice of content and style datasets plays a significant role in the effectiveness of style match and leave it as an open question for future research.

Method	Scanner 1 (XR)	Scanner 2 (S360)	Scanner 3 (CS2)
Baseline	0.485	0.241	0.409
Stain Norm. [13]	0.502	0.474	0.533
STRAP [21]	0.349	0.322	0.353
FDA [22]	0.400	0.379	0.488
Style Match	0.463	0.382	0.468

Table 3: Test set F1@0.5 scores of Faster R-CNN models on the MIDOG dataset. Each model is trained with slides from scanner 1.

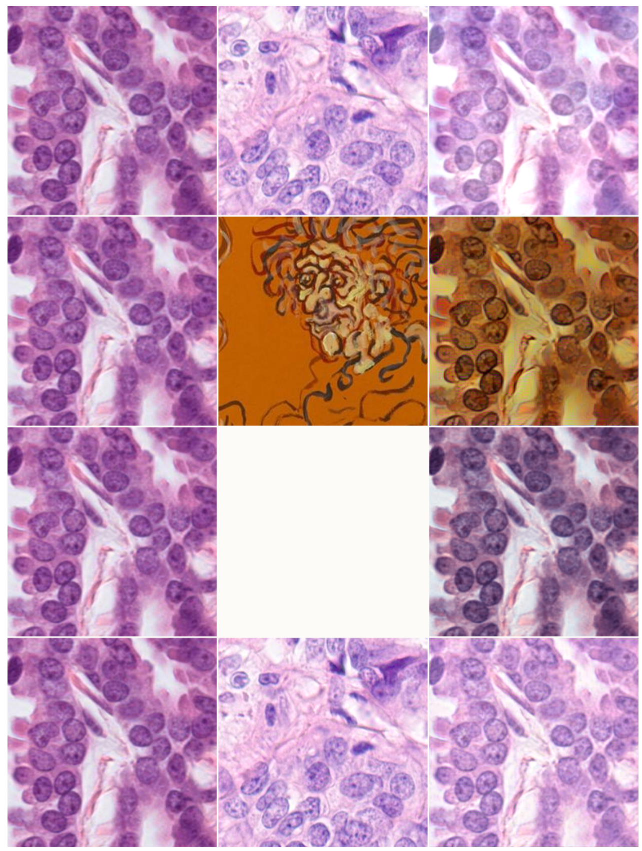


Figure 6: Qualitative comparison between augmentation approaches. From top to bottom: FDA [22], STRAP [21], stain normalization [13] and style match. From left to right: original image, style/reference image and output. Stain normalization does not use any additional reference image.

4.2.3 Comparison with Previous Approaches

We compare the performance of style match against other commonly used and recently proposed augmentation strategies for improving domain generalization in histopathology tasks. These methods are the following:

- **STRAP** [21] augments slides using a style transfer model, however, different from style match, uses an AdaIN model trained with content images from MS-COCO [11] and style images from the WikiArt dataset [19]. For a fair comparison, we train a Real-AdaIN network following the settings used in Section 3.2.1, using the MS-COCO and WikiArt content and style datasets and using WikiArt style images when training and evaluating detectors.

- **Fourier Domain Adaptation (FDA)** [22] aligns the appearance of a source image with a target image by replacing the source image's low frequency Fourier components with those of the target. When training a detector with FDA, we take target images from scanner 2 and use $\beta = 0.01$.

- **Stain Normalization** [13] reduces the visual variation of H&E stained histopathology slides by using SVD to find 2 stain vectors and projecting slide images onto the stain vector plane. We use the Macenko [13] stain normalization method to preprocess the dataset before training and evaluation.

Examples of each method can be seen in Figure 6.

The F1@0.5 scores when training on scanner 1 slides are shown in Table 3. Style match has mixed results, achieving similar performance to FDA but falling behind stain normalization. Compared to FDA, style match better preserves its performance when the training and test sets are the same. However, stain normalization actually improves the F1 score over the baseline with no augmentations. Even though visually, style match is better at transforming slides into a shared style distribution, stain normalization augments slides such that there is more contrast between the salient and background regions, which helps improve detection results.

STRAP differentiates itself from the other approaches by increasing the variation in the training set so the model becomes less biased towards the color distribution of a slide. While STRAP has been shown to help in classification tasks [21], the augmentation severely impacts detection performance and suggests that normalizing data into a common distribution is a better approach for detection domain generalization.

5. Limitations & Future Work

We showed with style match that using style transfer to augment input images during training and evaluation is a promising direction for domain generalization in histopathology, however, some limitations and unexplored problems have been discovered during our research.

An issue currently with style match is that the style transfer output changes depending on the choice of style image, which adds some randomness to the method. This variation is possibly beneficial during training by adding more diversity to the data. However, this also introduces variation in the test results (in our experiments, we do not shuffle the data during evaluation for consistent evaluation). Some preliminary experiments have been conducted into finding a single global style feature vector, such as a static randomly sampled vector or the average feature vector across the entire style dataset, but all of these cause significant artifacts. We believe finding optimal content/style slide pairs would benefit style match during both training and evaluation and leave it for future work.

Another issue that was not explored was the choice of content and style datasets used during style transfer training. Similar to how we believe optimal content/style pairs would improve detection performance, finding optimal content and style training sets may similarly improve the quality of the style transfer model and its performance when applied to inputs from novel domains. Style match is also not limited to using raw slides but can also use stain normalized data as its style dataset to gain the benefits from both methods.

6. Conclusion

In this work, we proposed style match, a training and test time augmentation method which reduces the domain gap between histopathology slides taken from different scanner types. Style match uses a style transfer model to transform input images into a color and texture distribution matching that of a reference style image. Since style transfer models can generalize to unseen content image domains, style match can be effectively applied to slides from any scanner type, which allows the data seen during evaluation to match that of the model's training set. Using our proposed Real-AdaIN style transfer model, we applied style match to mitosis detection and found that it can significantly improve standard training and achieves similar performance to other augmentation methods in domain generalization experiments. While traditional methods like stain normalization do achieve better performance, we believe that style match is a promising alternative direction to domain generalization in histopathology, which we hope will inspire future work in the field.

References

[1] Jie An, Haoyi Xiong, Jun Huan, and Jiebo Luo. Ultrafast photorealistic style transfer via neural architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10443–10450, 2020.

[2] Marc Aubreville, Christof Bertram, Mitko Veta, Robert Klopfleisch, Nikolas Stathonikos, Katharina Breininger, Natalie ter Hoeve, Francesco Ciompi, and Andreas Maier. Quantifying the scanner-induced domain gap in mitosis detection. *arXiv preprint arXiv:2103.16515*, 2021.

[3] Marc Aubreville, Christof A Bertram, Christian Marzahl, Corinne Gurtner, Martina Dettwiler, Anja Schmidt, Florian Bartenschlager, Sophie Merz, Marco Fragoso, Olivia Kershaw, et al. Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Scientific Reports*, 10(1):1–11, 2020.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[5] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

[8] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[9] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[10] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[12] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[13] Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE international symposium on biomedical imaging: from nano to macro*, pages 1107–1110. IEEE, 2009.

[14] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.

[16] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[18] Karin Stacke, Gabriel Eilertsen, Jonas Unger, and Claes Lundström. Measuring domain shift for deep learning in histopathology. *IEEE journal of biomedical and health informatics*, 25(2):325–336, 2020.

[19] Wei Ren Tan, Chee Seng Chan, Hernán E Aguirre, and Kiyoshi Tanaka. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE international conference on image processing (ICIP)*, pages 3703–3707. IEEE, 2016.

[20] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[21] Rikiya Yamashita, Jin Long, Snikitha Banda, Jeanne Shen, and Daniel L Rubin. Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation. *IEEE Transactions on Medical Imaging*, 40(12):3945–3954, 2021.

[22] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020.

[23] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9036–9045, 2019.