

FAST, SELF SUPERVISED, FULLY CONVOLUTIONAL COLOR NORMALIZATION OF H&E STAINED IMAGES

Abhijeet Patil^{*1} Mohd. Talha^{*1} Aniket Bhatia¹ Nikhil Cherian Kurian¹
Sammed Mangale¹ Sunil Patel² Amit Sethi¹

¹Department of Electrical Engineering, IIT Bombay, India, ² Nvidia, Mumbai, India

ABSTRACT

Performance of deep learning algorithms decreases drastically if the data distributions of the training and testing sets are different. Due to variations in staining protocols, reagent brands, and habits of technicians, color variation in digital histopathology images is quite common. Color variation causes problems for the deployment of deep learning-based solutions for automatic diagnosis system in histopathology. Previously proposed color normalization methods consider a small patch as a reference for normalization, which creates artifacts on out-of-distribution source images. These methods are also slow as most of the computation is performed on CPUs instead of the GPUs. We propose a color normalization technique, which is fast during its self-supervised training as well as inference. Our method is based on a lightweight fully-convolutional neural network and can be easily attached to a deep learning-based pipeline as a pre-processing block. For classification and segmentation tasks on CAMELYON17 and MoNuSeg datasets respectively, the proposed method is faster and gives a greater increase in accuracy than the state of the art methods.

Index Terms— Color normalization, self supervised learning, computational pathology

1. INTRODUCTION

Deep learning-based algorithms have been used for a variety of diagnostic tasks for histopathology images of tissue samples stained using hematoxylin and eosin (H&E) [1, 2, 3]. Most of these algorithms yield expected results when training and testing data have a similar color appearance. However, the performance of these algorithms drops drastically when tested with images from other labs [4], primarily due to variations in stain appearance that result from differences in scanner sensor responses, H&E staining protocols, reagents, and habits of technicians. Such variations in colors are not an issue for expert pathologists, as they are trained to prioritize morphological features over strict color definitions of H&E stained slides. However, stain colors of training data create bias for a deep learning algorithm, which can cause problems at the time of deployment.

Color normalization is used to transform input samples in computational pathology to a pre-defined color space to aid deep learning algorithms [4, 5, 6, 7, 8, 9]. Most of the practical color normalization approaches depend on matrix factorization for color deconvolution. Such approaches take the reference color matrix from a small patch of the H&E stained image, which causes problems when the source image is from significantly different tissue region (anatomically speaking) than the target patch [9]. Additionally, in order to estimate a robust factorization for the entire whole slide image (WSI), such methods often become computationally expensive. On the other hand, more recent color normalization techniques have turned to deep learning, particularly generative adversarial networks [10, 11, 12, 13], although such methods are computationally quite expensive to train, and cannot be easily integrated with downstream deep learning pipelines.

Our innovation lies in departing from matrix factorization as well as GANs. Instead, we use self-supervised learning, where a lightweight neural network is trained to estimate the color shift needed in each channel to match a pre-determined target stain in appearance. Compared to the state-of-the-art color normalization algorithms, the advantages of our algorithm are that it:

- Trains fast (Section 3),
- Tests faster (Table 2),
- Has a more positive impact on the accuracy of downstream tasks, such as classification on CAMELYON17 dataset [14] and segmentation on MoNuSeg dataset [15] (Table 1),
- Has fewer post-normalization image artifacts (Figure 1), and
- Integrates easily with deep learning pipelines (Figure 2).

2. RELATED WORK

Most of the color normalization methods compute a stain color matrix $W \in \mathbb{R}^{3 \times 2}$ and stain density maps H of source and target images and try to project the color matrix of the

^{*} These authors contributed equally.

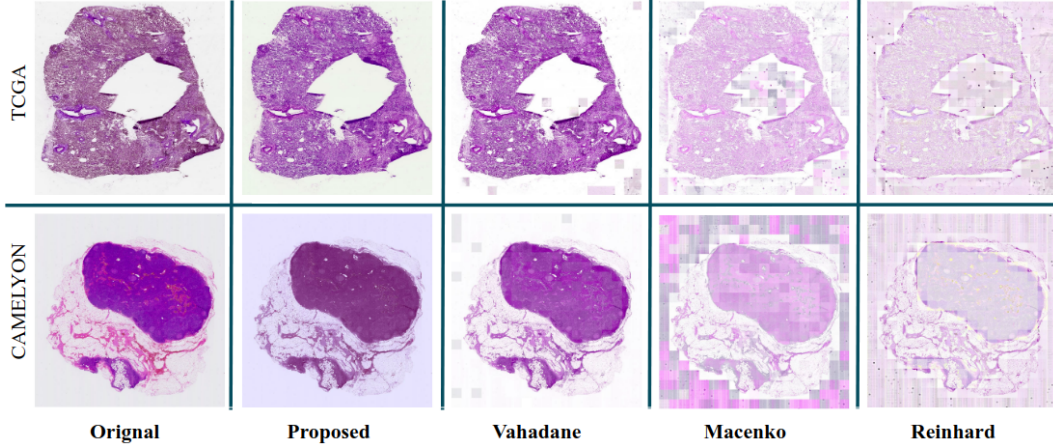


Fig. 1. Visual comparison between the proposed ColorNormNet and other popular color normalization methods on representative WSIs from TCGA (top) and CAMELYON (bottom) datasets. Notice the image artifacts introduced by the other techniques.

source image to that of the target image. Matrix factorization is a popular technique used for color deconvolution. Mecencko et al. [7] had framed color deconvolution as singular value decomposition (SVD) problem to guarantee the solution for an optimization problem. Reinhard et al. [6] used methods to match histograms for color densities in source and target images. Vahadane et al. [8] added sparsity to non-negative matrix factorization (NMF) to achieve better results. These approaches are patch-based. That is, the reference color matrix is taken from a small patch of the H&E stained image. This causes problems when the source image is from significantly different tissue region than the target patch [9]. The workaround in these methods against getting a biased estimate of the stain color matrix W of source image is to sample the matrix from the entire image, which is computationally expensive. The proposed method does not compute the stain color matrix.

Deep learning-based approaches have recently been employed for color normalization, especially those using generative adversarial networks (GANs). For example, Aicha et al. [16] proposed an end-to-end deep learning method to learn stain normalization along with specific diagnosis tasks, where stain normalization is learned adversarially. Farhad et al. [10] and Shaban et al. [11] use info-GAN [17] and cycleGAN [18] respectively in their color normalization techniques. Niyun et al. [12] uses a color matrix of the source image as auxiliary information for a generator based on cycleGAN for better learning. GANs are hard to train and require higher compute at inference time. Moreover, GANs can change microscopic morphological features which may important for diagnosis. Dwarikanath et al. [13] tries to resolve this problem by adding auxiliary feature from CNN trained for nucleus segmentation task, while SAASN [19] uses self-attention to preserve the local context, which makes their training process even more complicated. Additionally, Stanosa et al.’s model [20] learns

to pick a target template for normalization of H&E image according to the tissue type of the source image.

3. PROPOSED METHOD

The proposed method alleviates several problems in the existing factorization-based and deep learning-based color normalization methods based on the following features:

- It uses a self-supervised fully convolutional neural network that is lightweight (2,352 trainable parameters) and easy to train. Training the proposed CNN architecture takes only half an hour on Nvidia 1080Ti GPU, and inference takes about a minute and a half per gigapixel. By comparison, GAN-based color normalization model can increase the inference time drastically as they have millions of parameters. Also, for template-based methods computation is performed on CPUs which makes inference time slower.
- In the proposed method, WSI or set of WSIs with target color appearance acts as target domain, which can include a variety of tissue types. This eliminates problems caused by template-based methods such as the introduction of artifacts.
- We can change the target color appearance by training proposed architecture on patches from new target WSIs easily, which is difficult in GAN based color normalization methods.

The overall block diagram is shown in Figure 2. We next describe the self-supervision, the architecture, and the training details.

Self-supervision: In self-supervision, we use label that comes for “free” with data. There are several self-supervised approaches to learn representation of unlabeled data [21]. In

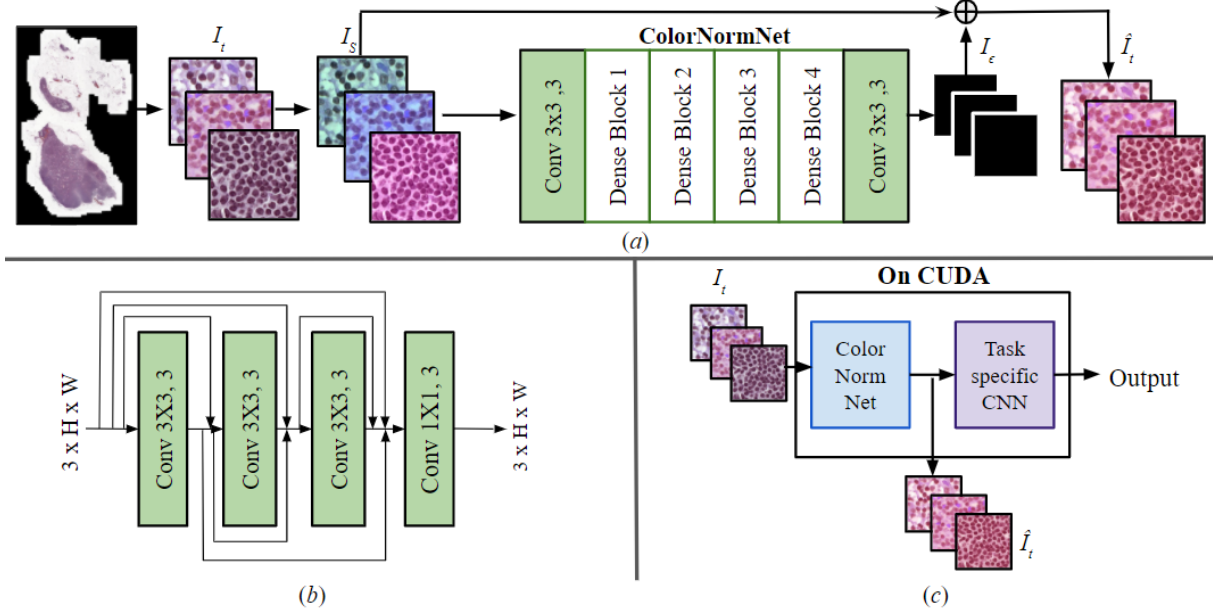


Fig. 2. Working of the proposed method color normalization network (ColorNormNet), including (a) the training pipeline, (b) the architecture of Dense Block used in ColorNormNet, and (c) how the trained ColorNormNet can be attached to task-specific CNN as a pre-processing block.

our setting, we train a deep learning model to predict the offset which is added synthetically to an unaltered patch sampled from WSI, W_{target} . Target domain $D_{target} = \{I_1, I_2, \dots, I_N\}$ is a set of images, such that I_n is the n^{th} patch in W_{target} . We perturb the R and B channel of a target histopathology tissue image $I_t \in D_{target}$ by adding a random offset to it to get a synthetic source image, I_s as shown in following equation.

$$I_s^R = I_t^R + \epsilon_1; I_s^B = I_t^B + \epsilon_2; I_s^G = I_t^G \quad (1)$$

where ϵ_1 and ϵ_2 are random numbers drawn independently from uniform distribution $Unif(-0.2, 0.2)$. We do not add offset to G channel of I_t because stain variation generally do not effect intensity of G channel of image. Figure 2 shows sample target image and synthetically generated source image corresponding to it.

Model Architecture: Our model uses a modified version of DenseNet blocks. We have used four DenseNet blocks with only three convolutional filters at each layer. We hypothesize that learning the color offset I_ϵ does not require to learn too many features as long as the prediction is obtained using a large enough receptive field. To increase the receptive field, we have used dilated convolutional layers in Dense Block 2 and Dense Block 3. Each convolutional layer in Dense Block is followed by leaky ReLU and a batch normalization layer. We have not used any pooling operation in our architecture. The size of feature map remains equal to size of the input image throughout the model. As we have used only three filters in every convolutional layer, the number of trainable parameters is also very small. Our model, which we call *ColorNorm-*

Net, contains only 2,352 trainable parameters.

Training details: We optimize parameters of a DenseNet [22] inspired fully convolutional neural network $f(I_s; \theta_f)$ to predict offset tensor I_ϵ such that $\hat{I}_t = I_s + I_\epsilon$. We train ColorNormNet by optimizing following loss function

$$L(I_t, \hat{I}_t) = \|\hat{I}_t - I_t\|_1 + \lambda \|\hat{I}_t - I_t\|_2^2$$

where λ is a hyper-parameter which balances $L1$ and $L2$ loss terms. We used $\lambda = 0.1$ for training ColorNormNet. We train this model with the patches from a target WSI. We observed that even 1000 patches are sufficient to train a color normalization network. Training the proposed CNN architecture takes about half an hour (800-1000 iterations) to converge on a Nvidia 1080Ti GPU with a batch size of 128 and patches of size 256×256 , and Adam optimizer with learning rate 0.001. Figure 2 shows training and inference pipeline along with architecture of CNN model. We can also train a model with arbitrary patch size as the proposed model is fully convolutional.

4. EXPERIMENTS AND RESULTS

We trained ColorNormNet on patches extracted from a randomly selected WSIs from the CAMELYON17 dataset [14]. We compared the inference times for ColorNormNet with four state of the art methods – by Vahadane et al. [8], Reinhard et al. [6], Macenko et al. [7], and Ramakrishna et al. [9]. We extracted the patches from WSI with more than one gigapixel in size and observed the inference time for each method.

Table 1. Comparison of classification and segmentation results for different color normalization methods.

Method	Classification Results (AUC)						Segmentation Results		
	Center 1	Center 2	Center 3	Center 4	Center 5	Average	Dice	AJI	PQ
W/O Norm.	0.83	0.83	0.67	0.87	0.80	0.80	0.78	0.52	0.55
Reinhard	0.83	0.82	0.80	0.87	0.81	0.83	0.81	0.53	0.54
Macenko	0.77	0.83	0.84	0.86	0.76	0.81	0.66	0.39	0.46
Vahadane	0.81	0.82	0.86	0.86	0.88	0.84	0.74	0.40	0.49
ColorNormNet	0.90	0.99	0.79	0.98	0.93	0.91	0.80	0.53	0.56

Table 2. Comparison of computation times for different color normalization methods to normalize WSIs of size one gigapixel

Method	Time (seconds)
Vahadane	2642
Reinhard	260
Mecenko	2139
Goutham-CPU	360
Goutham-GPU	120
ColorNormNet-CPU (our)	485
ColorNormNet-GPU (our)	87

Not only is ColorNormNet faster than the other methods, but it also does not produce any artifacts, as it estimates a single offset for the entire image. ColorNormNet also generalizes on the dataset from other centers and different organs. We have normalized WSIs from the TCGA dataset and observed that ColorNormNet yields better normalization on it. Figure 1 shows color normalization results on a sample WSI each from CAMELYON17 lymph node and TCGA lung adenocarcinoma datasets. A comparison of test time per gigapixel for different techniques, including ours, is shown in Table 2.

We used CAMELYON17 [14] and MoNuSeg [15] to evaluate ColorNormNet. CAMELYON17 contains a total of 500 WSIs, out of which 400 WSIs are of normal tissue while remaining 100 WSIs contain tumor region within them. This dataset is well-suited for the evaluation since the data is taken from five different centers and there is a significant staining variation amongst them. Out of 100 positive WSIs, only 50 WSIs (10 WSIs from each center) are annotated with the tumor regions. We used these 50 positive WSIs in our training and testing data along with the available negative WSIs. We split data from each center into training, validation, and testing with about 50%, 30%, and 20% slides in each split respectively. We got following train-validation-test split for the centers 1: 33(4)-22(4)-19(2), center 2: 33(4)-20(3)-15(3), center 3: 41(5)-24(2)-19(3), center 4: 31(4)-19(3)-20(3), and center 5: 32(3)-20(4)-19(3) (Numbers in parentheses indicate the number of positive WSIs in split). We trained a ResNet50 [23] architecture for the classification task. We used batch size of 32 with a learning rate of 0.001 and weight decay of 0.01 for

the training. The classification model was trained and validated with data from one center and tested with test data from each center separately. We report the average of AUC obtained on test data of five data centers in Table 1. ColorNormNet achieves better average AUC across the centers by a large margin.

We also evaluated ColorNormNet on nucleus segmentation task. We used MoNuSeg [15] for evaluation. MoNuSeg dataset contains data from nine different organs with structural and stain variation across different images. It contains a total of 30 annotated images. We split these images in training, validation and testing datasets with 20, 5 and 5 images in each split respectively. We trained HoVer-Net [24] five times using the normalized images with different methods and averaged the Dice score, aggregated Jaccard index (AJI), and panoptic quality (PQ) across the five runs for each normalization method. Table 1 shows that ColorNormNet gives better AJI and PQ than other normalization methods.

5. CONCLUSION

Color normalization is an often-used image pre-processing method for deep learning based computational pathology tasks. We have proposed a color normalization method based on a fully convolutional neural network, which is trained in a self-supervised manner. The proposed ColorNormNet is computationally fast during both training and inference, and it yields a neural network that can be attached as pre-processing block to task specific CNNs. We have also validated ColorNormNet on tumor classification and nucleus segmentation task and shown that the accuracy of these downstream tasks improve more using ColorNormNet as compared to other color normalization methods.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by [14, 15]. Ethical approval was not required as confirmed by the license attached with the open access data.

7. ACKNOWLEDGMENTS

No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

8. REFERENCES

- [1] Aditya Golatkar, Deepak Anand, and Amit Sethi, "Classification of breast cancer histology using deep learning," in *Image Analysis and Recognition*, 2018.
- [2] Wouter Bulten, Hans Pinckaers, and Geert Litjens, "Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study," *The Lancet Oncology*, Feb 2020.
- [3] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, and Aristotelis Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," in *Nature Medicine*, 2018.
- [4] Amit Sethi, Lingdao Sha, Abhishek Ramnath Vahadane, Ryan J Deaton, Neeraj Kumar, Virgilia Macias, and Peter H Gann, "Empirical comparison of color normalization methods for epithelial-stromal classification in h and e images," *Journal of pathology informatics*, 2016.
- [5] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Transactions on Biomedical Engineering*, 2014.
- [6] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, 2001.
- [7] Marc Macenko, Marc Niethammer, J. S. Marron, Charles Schmitt, and Nancy E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *ISBI*, 2009.
- [8] A. Vahadane, T. Peng, A. Sethi, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging*, 2016.
- [9] Goutham Ramakrishnan, Deepak Anand, and Amit Sethi, "Fast gpu-enabled color normalization for digital pathology," 2019.
- [10] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. W. M. van der Laak, and P. H. N. de With, "Stain normalization of histopathology images using generative adversarial networks," in *ISBI*, 2018.
- [11] M Tarek Shaban, Christoph Baur, Nassir Navab, and Shadi Albarqouni, "Staingan: Stain style transfer for digital histological images," 2018.
- [12] Zhou, Niyun, Cai, De, , Yao, and Jianhua, "Enhanced cycle-consistent generative adversarial network for color normalization of h&e stained images," in *MICCAI*, 2019.
- [13] Dwarikanath Mahapatra, Behzad Bozorgtabar, Jean-Philippe Thiran, and Ling Shao, "Structure preserving stain normalization of histopathology images using self-supervised semantic guidance," 2020.
- [14] P. Bándi, O. Geessink, Q. Manson, M. Van Dijk, and G. Litjens, "From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge," *IEEE Transactions on Medical Imaging*, 2019.
- [15] N. Kumar, R. Verma, D. Anand, , and A. Sethi, "A multi-organ nucleus segmentation challenge," *IEEE Transactions on Medical Imaging*, 2020.
- [16] A. Bentaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE Transactions on Medical Imaging*, 2016.
- [17] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," 2016.
- [18] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017.
- [19] Aman Shrivastava, Will Adorno, Lubaina Ehsan, S. Asad Ali, Sean R. Moore, Beatrice C. Amadi, Paul Kelly, Sana Syed, and Donald E. Brown, "Self-attentive adversarial stain normalization," 2019.
- [20] Janowczyk Andrew, Basavanahally Ajay, and Madabhushi Anant, "Stain normalization using sparse autoencoders (stanosa): Application to digital pathology," in *Computerized Medical Imaging and Graphics*, 2017.
- [21] Muzammal Naseer, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Fatih Porikli, "A self-supervised approach for adversarial robustness," in *CVPR*, 2020.
- [22] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger, "Densely connected convolutional networks," 2016.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," 2015.
- [24] Simon Graham, Quoc Dang Vu, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," 2018.