Histopathology Stain-Color Normalization Using Generative Neural Networks

Farhad G. Zanjani **Svitlana Zinger** **Peter H.N. de With**
Department of Electrical Engineering, Eindhoven University of Technology

Babak E. Bejnordi **Jeroen AWM van der Laak**
Diagnostic Image Analysis Group, Radboud University Medical Center

Abstract

Performance of designed CAD algorithms for histopathology image analysis is affected by the amount of variations in the samples such as color and intensity of stained images. Stain-color normalization is a well-studied technique for compensating such effects at the input of CAD systems. In this paper, we introduce unsupervised generative neural networks for performing stain-color normalization. For color normalization in stained hematoxylin and eosin (H&E) images, we present three methods based on three frameworks for deep generative models: variational auto-encoder (VAE), generative adversarial networks (GAN) and deep convolutional Gaussian mixture models (DCGMM). Our contribution is defining the color normalization as a learning generative model that is able to generate various color copies of the input image through a nonlinear parametric transformation. In contrast to earlier generative models proposed for stain-color normalization, our approach does not need any labels for data or any other assumptions about the H&E image content. Furthermore, our models learn a parametric transformation during training and can convert the color information of an input image to resemble any arbitrary reference image. This property is essential in time-critical CAD systems in case of changing the reference image, since our approach does not need retraining in contrast to other proposed generative models for stain-color normalization. Experiments on histopathological H&E images with high staining variations, collected from different laboratories, show that our proposed models outperform quantitatively state-of-the-art methods in the measure of color constancy with at least 10-15%, while the converted images are visually in agreement with this performance improvement.

1 Introduction

Computer-aided diagnosis systems are widely investigated for microscopic analysis of tissues to study presence, localization or grading of disease. Since the cells of an examined tissue lack sufficient contrast to be easily distinguished in bright-field microscopy, staining is a common practice in histopathology. Generally, two stains of contrasting colors are selected. The common laboratory stains are paired hematoxylin and eosin (H&E), which differentiate between nucleus and cytoplasm in color. In such a colorization procedure, several factors are involved that can be hard to be regularized between different medical centers or even within the samples of the same laboratory at different trials or time periods. Some sources of such variations originate from the method of specimen preparation, staining protocols like temperature of the adopted solutions, fixation characteristics, imaging device characteristics [5], etc. Such undesirable effects lead to color and intensity variations in histopathological images. Although human color perception can easily comprehend the staining

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

color variations among images, the design of a reliable CAD system approaching pathologists performance and robustness, is a challenging task.

Since eliminating the sources of all color variations in staining histopathology images is almost infeasible [21], two solutions have been devised. The first approach is ignoring the color information in images by transforming them into grayscale. Using texture information [22], wavelet transform [13] and local binary patterns [1] are some techniques that have been applied to grayscale histopathology images. However, bypassing the problem by eliminating the color information can loose important signal information, which is useful for medical diagnostics that are considered widely by pathologists. As a second approach, stain color normalization is broadly studied [21, 6, 15, 25, 19, 2, 16, 20, 3, 5, 17, 18, 28, 4, 31] to incorporate that signal information.

During the last decade, performance of CAD systems in computational pathology have been significantly improved, mainly by emerging new machine learning techniques, like e.g. deep learning methods. Although such complex models can learn complicated segmentation or classification tasks in the presence of color variations in samples, recent studies show that stain-color normalization can lead to a higher performance in a learning system as well as a CNN model [10, 27, 6]. For example, applying stain-color normalization for the tissue classification in H&E colorectal cancer images by a CNN, remarkably improves the accuracy up to 20% [10].

2 Related work

The previous stain-color normalization methods can be divided into three categories: *stain-color deconvolution*, *template color matching* and *multi-task learning*, which are briefly explained below after which our contributions are described.

2.1 Stain-color deconvolution

Stain-color deconvolution [26] methods are considering prior knowledge of the reference stain vector for every dye and split an input RGB image into three stain channels, each representing the actual stain color. Ruifrok *et al.* [26] introduce this prior knowledge by manually selecting pixels that represent a specific stain class and then compute the color deconvolution vector. Because of the semi-automatic nature, several studies are performed later for automated extraction of stains, by e.g. using the singular value decomposition (SVD) technique [19], probabilistic Gaussian mixture model (GMM) [20], using a prior for stain matrix estimation [21] and stain-color descriptions along with training a supervised relevance vector machine [15]. Although these solutions lead to a better stain estimation, they are solely limited to image color information, while the spatial dependency between tissue structures is still ignored [6]. This aspect causes shortcomings for stain deconvolution approaches, especially when severe staining variations occur in the data.

2.2 Template color matching

Template color matching methods proposed by Reinhard *et al.* [25] rely on aligning the statistical color distribution (e.g. mean and standard deviation) of a *source* image with a *template* (reference) image. The authors used a set of linear transformations for assigning a unimodal distribution to each channel of the CIELAB color model. Afterwards, each channel was treated independently for alignment. Since there is a dependency between the color channels due to dye contribution, this approach has drawbacks that have been addressed in [15, 6]. For solving this problem, separate transformations are performed on stain classes [20], or on tissue classes [6, 3]. For avoiding artifacts at the border of different classes under different transformations, the authors considered a weighted contribution of these transformations in the final color image. Two proposed solutions consist of estimating weights of the GMM [20] and training a naive Bayesian classifier [6]. The latter solution introduces multiple parameters and thresholds which cannot be optimally applied to a new dataset or even fail if the tissue type changes.

2.3 Multi-task learning

Recently, the capacity of deep generative models has been explored for performing stain-color normalization. This involves the application of the GAN [14] in the framework of multi-task

learning (also called *stain-style transfer* model), which has been studied by some researchers [7, 12]. According to their approach, learning the color normalization task by a GAN is integrated with another *discriminative* model (e.g. supervised classification of tumors from normal tissue). The GAN model tries to convert the image color information in order to maximize the performance of a classifier. Such approach benefits from using a generative model, but defining the problem in the principle of multi-task learning has shortcomings. Firstly, this approach to color normalization is not generic and is not able to address some histology studies, e.g. when only the slide-level labels are available or in the absence of such labels. Secondly, using discriminative learning as a supervisory signal for training a GAN does not necessarily lead to an optimal color normalization, because the objective of the color normalization module differs from that of the auxiliary classifier.

In this paper, we propose using deep generative models based on CNN for fully unsupervised learning of stain-color normalization. Our contribution is twofold. 1- We present three methods, each based on a deep generative model: GAN, VAE and Deep GMM. Since the classical deep generative models are not directly applicable to stain-color normalization, we have modified the standard versions of those models. 2- In contrast to the previous generative models which are used for stain-color normalization, our methods are fully unsupervised and do not need any labels or prior assumption on the H&E image contents.

3 Methods

In order to adapt the conventional deep generative models to stain-color normalization, we have designed our methods by considering three important properties.

1. The image structures must be preserved after color conversion and only the chromatic information can be subject to change.
2. The model does not need any labels or any assumptions about the data and should be fully unsupervised.
3. The models should learn the color transformation between any two image pairs. It means that the models do not convert the color of the input images to only a predefined reference image (or a subset of reference images). This property facilitates time-critical CAD systems in case of changing the reference image, since our approach does not need retraining but only re-adjustment of some parameters.

In the remainder of this section, these three properties are taking into account and will be guiding the explanation of our methods in detail.

3.1 GAN-based color normalization model

In contrast with the GAN-based solution for stain-color normalization from recent literature [7, 12], we presented recently a fully unsupervised GAN-based model for stain-color normalization [11]. The block diagram of the model is shown in Figure 1. The model was inspired by the *InfoGAN* [8] architecture, which in addition to the *generator* and *discriminator* networks, benefits from an *auxiliary* network (see Figure 1). Here, the auxiliary network has two functions. First, it imposes a constraint for discovering a semantic structured latent variable in the generator network. Second, after training by applying the auxiliary network on any given reference image, the model adjusts the mapping parameters of the generator to produce samples which resemble the reference image. For preserving the image structure, the *lightness* channel of the input images in the CIELAB color system is supplied to the generator and only the chromatic channels are generated by the GAN. For training the model, four loss functions were defined: generative, discriminative, mutual information and reconstruction losses. For more detail, we refer to the appendix of this paper, as this work has been accepted but still is in press [11].

3.2 VAE-based color normalization model

A VAE [23] consists of two networks called *encoder* and *decoder*. The encoder network maps the input data x to a usually lower-dimensional latent variable z and the decoder network transforms the z back to the original data space. In contrast to a standard auto-encoder, the VAE regularized the encoder network by imposing a prior (e.g. a unit Gaussian distribution) over the latent distribution

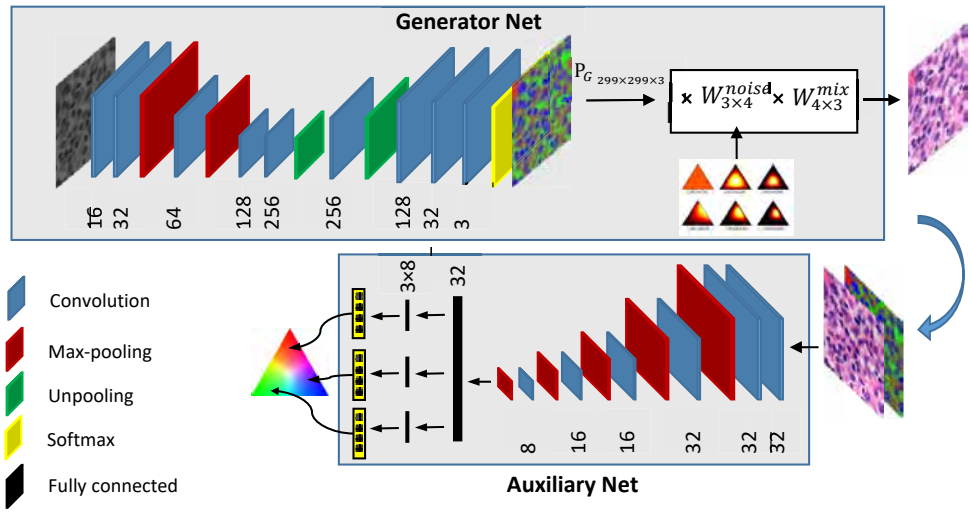


Figure 1: Schematic of GAN-based model [11] for stain-color normalization (the discriminator network was not shown here).

$p(z)$. The VAE loss includes two loss functions: generative and latent. The generative loss ($\mathcal{L}_{rec}$) is defined as the negative log-likelihood of data, which indicates how accurately the network reconstructs the input image. The latent loss ($\mathcal{L}_{latent}$) is defined by the Kullback–Leibler divergence (D_{KL}) that measures how closely the distribution of the latent variable $q(z|x)$ matches the prior distribution. Both losses are specified by

$$\mathcal{L}_{rec} = \mathbb{E}_{q(z|x)}[logp(x|z)] \quad \text{and} \quad \mathcal{L}_{latent} = D_{KL}(q(z|x)||p(z)). \tag{1}$$

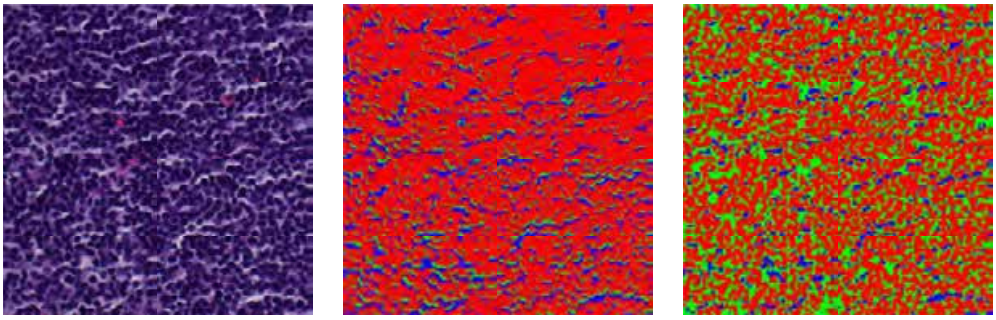
The VAE loss is simply the combination of these two losses, so that

$$\mathcal{L}_{VAE} = \mathcal{L}_{rec} + \mathcal{L}_{latent} = -\mathbb{E}_{q(z|x)}[log(\frac{p(x|z)p(z)}{q(z|x)})]. \tag{2}$$

In our proposed model, the encoder input are the H&E images in the HSD [30] color system. By applying several operations like convolutional, (un)pooling, nonlinear activation functions (ReLU) and a softmax layer at the output, the encoder transfers each pixel of input image into a K-dimensional vector (called $\boldsymbol{\gamma}$) in the unity range that adds up to unity ($Enc : \mathbb{R}^2 \rightarrow (0,1)^K$). The $\boldsymbol{\gamma} = Enc(x)$ is interpreted as a *membership* variable, which represents the tissue-class membership of each pixel. In other words, the encoder network aims to segregate the input image into tissue structure channels. In classical VAE, the latent random variable z is considered as a deterministic variable which is generated at the output of the encoder, but in our proposed model the $\boldsymbol{z}$ denotes the color distribution of tissue classes in the input image with the mean ($\boldsymbol{\mu} = [\mu_1, ..., \mu_K]$) and covariance matrix ($\boldsymbol{\Sigma}$). Since such a $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are conditional variables and dependent on the tissue class (latent variable $\boldsymbol{\gamma}$), consequently the $\boldsymbol{z}$ is treated as a latent variable. More formally, this leads to the following specification:

$$\begin{aligned}
\mu_i =& \mathbb{E}(X|argmax(\boldsymbol{\gamma}) = i) \quad \text{for} \quad i = 1, 2, ..., K. \\
\sigma_i =& \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T|argmax(\boldsymbol{\gamma}) = i] \\
\boldsymbol{z} =& \mathcal{N}(\boldsymbol{\mu}, \epsilon \cdot \boldsymbol{\Sigma})
\end{aligned}$$

Parameter K stands for a predefined number of tissue classes in the images (e.g. for H&E image is set to $K = 3$ for nuclei, the surrounding tissues and the background regions). The auxiliary variable ϵ is sampled from a prior distribution (e.g. $\epsilon \sim \mathcal{N}(0, 1)$). The meaning of the above equation is that the model segments the input image into K regions and computes the mean and standard deviation of the input pixels in those regions individually. For simplicity, we assume that the chromatic information between different tissue classes are uncorrelated, i.e. the covariance matrices given by $\Sigma = \sigma^2 I$ where I is the identity matrix, although the conclusions will hold for general covariance matrices. To summarize, the encoder is expected to segment the input images into the tissue classes (e.g. nuclei which mostly absorb hematoxylin stain, background with no stain and the other tissue structures which mainly absorb eosin stain) at its softmax layer. Afterwards, the parameters of the color distribution in the input image is computed individually per tissue class. Similar to standard



RGB H&E image conventional GMM DCGMM

Figure 2: Comparison of three-class label assignment to tissues for a conventional GMM compared to the novel DCGMM.

VAE model, by keeping the calculated $\boldsymbol{\mu}$, we only sample from the standard deviation from a prior Gaussian distribution with zero mean and unity variance. Afterwards, the decoder network transfers the sampled latent variable into the output for reconstructing the chromatic channels of the input.

3.3 Deep Gaussian mixture color normalization model

The classical GMM is a well-established machine learning technique which is applied also for stain-color normalization [20] and [3]. Applying conventional GMM for stain-color normalization has some shortcomings. Because the conventional GMM-based models clusters the pixels based on their color attributes only, while it ignores the spatial information, appearance and the shape of tissue structures, its performance deteriorates, especially when strong staining variations appear in histopathological images (see Figure 2). On the other hand, several published studies that involve the spatial properties of biological structures which are occurring in most H&E images, give better results for stain-color normalization [32, 11, 6].

The recent development of deep generative models has invoked some extensions to the classical GMM [33, 29]. Two proposed approaches are (1) constructing a stack of multiple GMM layers on top of each other in a hierarchical architecture [29], and (2) using auto-encoder neural networks while applying the GMM on their low-dimensional representations [33]. Here, we introduce an alternative approach, called deep convolutional GMM (DCGMM), that represents a fully-convolutional CNN of which the parameters are jointly optimized with the parameters of the GMM, in an end-to-end learning. For further explaining this alternative method, we first detail the needed terminology for GMM.

The GMM can be presented as a linear superposition of K Gaussian mixtures in terms of discrete *latent* ($\boldsymbol{z}$) variable vector, in the form of

$$p(x) = \Sigma_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k). \tag{3}$$

The K-dimensional binary random variable $\boldsymbol{z}$ has one-hot encoding ($z_k \in \{0, 1\}$; $\Sigma_{k=1}^{K} z_k = 1$), which represents tissue classes in our study. In Eq. (3), the mixing coefficient π_k must satisfy $0 \leq \pi_k \leq 1$ together with $\Sigma_{k=1}^{K} \pi_k = 1$, in order to fulfill a valid probability definition [9]. Here, $\mathcal{N}$ stands for a multivariate normal distribution with mean μ_k and covariance matrix Σ_k. If we consider π_k as prior probability of class z_k, its posterior probability, called *responsibility*, can be written [9] as follows:

$$\gamma(z_k) = p(z_k = 1|x) = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\Sigma_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}. \tag{4}$$

According to Eq. (3), the (natural) log-likelihood function is

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \Sigma_{n=1}^{N} \ln\{\Sigma_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)\}, \tag{5}$$

where N is the total number of pixels in the input image. Given the GMM, the objective is to maximize the likelihood function (Eq. (5)) with respect to the parameters (μ_k, Σ_k, π_k). The common approach for this is using the expectation maximization (EM) algorithm by iteratively evaluating the responsibilities (Eq. (4)) and re-estimating the parameters.

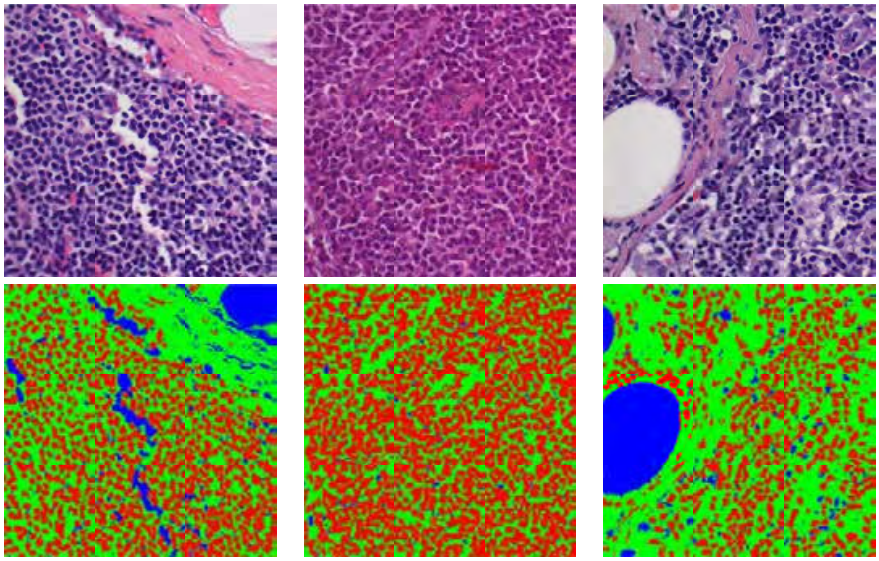


Figure 3: Three examples of tissue-class assignment by the DCGMM. (red) nuclei; (green) surrounding tissue; (blue) background.

The DCGMM fits a GMM to the pixel-color distribution, while being conditioned on tissue classes. For processing the image and detecting the tissue classes, the high image-representation capability of the CNN has been exploited. To do so, estimating the responsibility coefficients (see Figure 3) is performed by a CNN, similar to what the encoder network in VAE performs, as we explained earlier. Instead of using iterative EM algorithm, all parameters of the DCGMM are jointly optimized by minimizing the negative log-likelihood (maximizing Eq. (5)) with the gradient descent algorithm. The calculation of the required partial derivatives of negative log-likelihood with respect to its parameters ($\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) for performing a gradient descent optimization can be found in [24, p. 45]. For a better understanding, one can consider that the *E-step* in an EM-based algorithm is replaced by a CNN. Afterwards, by having a current estimate for the responsibility parameters, we can calculate the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of multivariate Gaussian distributions over the channels of an input image, similar to *M-step*. Therefore, we can compute Eq. (5) and consequently use gradient descent and the back-propagation algorithm for training the DCGMM.

After training the model, by applying DCGMM on any given test images, we have the parameters of multivariate Gaussian distributions. The color conversion can be performed by aligning the Gaussian distributions in a source image to its counterpart in a template image. Such a conversion consists of *shifting* the mean, *whitening* and *coloring* transformation. The coloring transformation refers to applying the SVD technique to the whitened Gaussian distribution and scale it to obtain the same covariance matrix as the template image.

4 Experimental Results

We focus on inter-laboratory variations of the H&E staining, as it is a major concern in large-scale application of CAD in pathology. We evaluate the performance of our three proposed generative neural networks in comparison to that of five competing state-of-the-art algorithms: linear appearance normalization by Macenko *et al.* [19], statistical color properties alignment by Reinhard *et al.* [25], nonlinear mapping for stain normalization by Khan *et al.* [15], sparse non-negative matrix factorization by Vahadane*et al.* [2] and WSI color standardization by Bejnordi *et al.* [6].

4.1 Dataset

For better comparison with recent studies, we use the same dataset as has been introduced in [6]. The dataset contains 625 images (each 1388×1040 pixels) from 125 digitized H&E stained WSIs of lymph nodes from 3 patients. The tissues are sectioned and sent to five Dutch pathology laboratories,

Table 1: Standard Deviation (SD) and Coefficient of Variation (CV) of NMI for all five laboratories for hematoxylin and eosin dyes.

Method	Hematoxylin												Eosin	
	Lab 1		Lab 2		Lab 3		Lab 4		Lab 5		**Average**			
	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV
Original	0.033	0.065	0.031	0.060	0.037	0.078	0.029	0.049	0.028	0.051	0.032	0.060	0.0563	0.0748
Macenko[19]	0.029	0.052	0.026	0.046	0.020	0.037	0.025	0.044	0.020	0.035	0.024	0.043	0.0362	0.0439
Reinhard[25]	0.032	0.058	0.025	0.044	0.020	0.035	0.030	0.052	0.029	0.049	0.027	0.047	0.0386	0.0494
Khan[15]	0.066	0.156	0.067	0.155	0.085	0.158	0.054	0.110	0.049	0.093	0.064	0.135	0.0434	0.0555
Vahadane[2]	0.036	0.065	0.032	0.058	0.024	0.046	0.023	0.042	0.020	0.038	0.027	0.050	0.034	0.041
Bejnordi[6]	**0.016**	**0.029**	**0.015**	**0.027**	0.018	**0.034**	0.029	0.055	0.024	0.044	0.021	0.038	0.0191	0.0220
VAE	0.029	0.052	0.022	0.043	0.021	0.064	0.028	0.050	0.025	0.055	0.025	0.0528	0.026	0.036
GAN	0.024	0.053	0.019	0.043	0.020	0.043	0.027	0.057	0.024	0.053	0.022	0.050	0.0195	0.0218
DCGMM	0.022	0.045	0.017	0.034	**0.017**	0.036	**0.014**	**0.030**	**0.017**	**0.035**	**0.017**	**0.036**	**0.0188**	**0.0209**

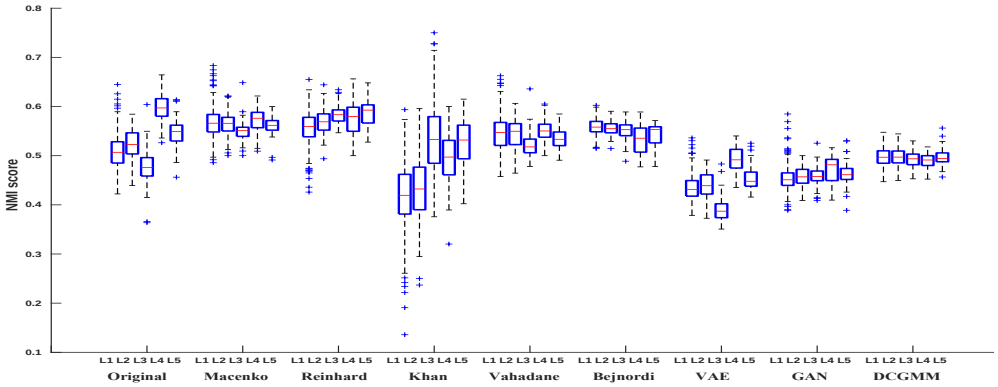


Figure 4: Boxplot of NMI scores for hematoxylin regions.

each using their own routine staining protocols (more details can be found in [6]). Our model is trained on randomly cropped patches (299×299 pixels for VAE and GAN networks and 576×576 pixels in DCGMM experiments) and evaluated on the full-size images, since all three networks have a fully convolutional architecture. The evaluation is performed by using leave-one-out cross-validation based on the laboratories where the data was collected.

4.2 Results

The normalized median intensity (NMI) measure [3, 6] is used to evaluate the color constancy of the normalized images. Quantitative analysis is based on independently evaluating the color constancy in the regions that show mostly absorbed hematoxylin or eosin. Since nuclei mostly absorb hematoxylin, they first are detected automatically by using a fast radial symmetry transform and a marker-controlled watershed algorithm [6]. For evaluation of the eosin analysis, several regions are manually annotated for 25 images. The results of the different methods for assessing hematoxylin and eosin regions are shown in Table 1 and illustrated in Figure 4. As we can observe from the Table 1, our proposed models show better results in eosin regions when compared with most of the previous methods. For the hematoxylin regions, our proposed models show acceptable color constancy and the DCGMM method outperforms all the state-of-the-art methods, measured as average performance on all laboratories. The VAE-based model did not show high performance in comparison with our other two deep generative models. This can originate from leaking the color information between the source image at the input of the VAE-based model to its output. In other words, the encoder could not perfectly disentangle the chromatic information from the intensity of the input image. Among our proposed methods, the DCGMM converges very rapidly after a few epochs on the training set. Among previous methods, Bejnordi *et al.* method [6] shows a higher performance and it is close to the performance of deep generative models. However, we should consider that while our models are generic and lack any assumption about H&E image contents, their method is based on some assumption like e.g. fitting elliptical shapes to nuclei by the Hough transform and applying several thresholds within the algorithms that can be easily violated if the tissue type changes. The visual illustration of stain-color conversions between a source and reference image, resulting from the different methods, is shown in Figure 5. It can be observed that the quantitative measure has high correlation with the visual perception of the results.

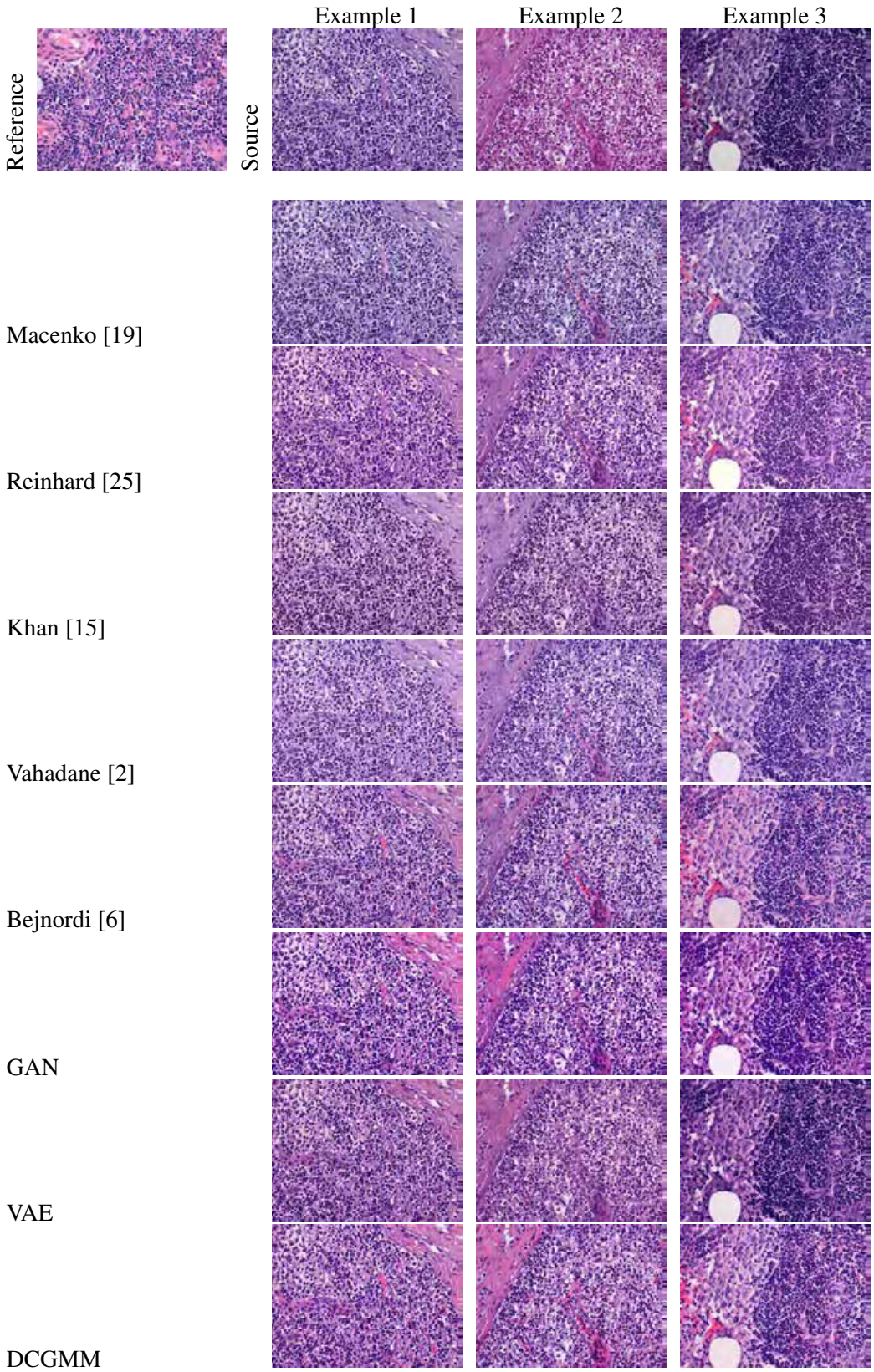


Figure 5: Performance of different methods on stain color-normalization between three H&E images from three different labs and a reference image.

5 Conclusion

In this paper, we have introduced three deep generative models for performing stain-color normalization in histopathological H&E images. For designing such a model, three main properties were considered: (1) the converted image should preserve the image appearance, excluding the chromatic information; (2) the model should train fully unsupervised; (3) the model should be able to convert the color information between any two unseen image pairs and not only convert the color of input images to the color of a predefined reference image. Since the classical deep generative models cannot fulfill all aforementioned considerations, we have introduced the GAN-based, VAE-based and the DCGMM methods. The experiments have shown that the DCGMM outperforms state-of-the-art methods in the measure of color constancy with at least 10-15% (or more) on a challenging dataset which contains severe staining variations between laboratories. For the near future, we aim to study the possibility of applying the generative models to a unified image dataset, consisting of more samples from different tissue types, collected from different organs.

References

[1] Ruiz A., Sertel O., Ujaldon M., Catalyurek U., Saltz J., and Gurcan M. Pathological image analysis using the gpu: Stroma classification for neuroblastoma. In *Bioinformatics and Biomedicine, 2007. BIBM 2007. IEEE International Conference on*, pages 78–88. IEEE, 2007.

[2] A. Sethi S. Albarqouni L. Wang M. Baust K. Steiger A.M. Schlitter I. Esposito A. Vahadane, T. Peng and N. Navab. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE transactions on medical imaging*, 35(8):1962–1971, Aug. 2016.

[3] Anant Madabhushi Ajay Basavanhally. Em-based segmentation-driven color standardization of digitized histopathology. *Proc.SPIE*, 12:8676 – 8676 – 12, 2013.

[4] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, and N.M. Rajpoot. Stain deconvolution using statistical analysis of multi-resolution stain colour representation. *PloS one*, 12(1), 2017.

[5] P. A. Bautista, N. Hashimoto, and Y. Yagi. Color standardization in whole slide imaging using a color calibration slide. *Pathol. Inf.*, 5, 2014.

[6] B.E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Höller, A. Homeyer, N. Karssemeijer, and J. AWM van der Laak. Stain specific standardization of whole-slide histopathological images. *IEEE transactions on medical imaging*, 35(2):404–415, 2016.

[7] A. BenTaieb and G. Hamarneh. Adversarial stain transfer for histopathology image analysis. *IEEE Transactions on Medical Imaging*, (99), Dec 2017.

[8] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *In Advances in Neural Information Processing Systems*, pages 2172–2180, 2016.

[9] M. Bishop Christopher. *Pattern Recognition and Machine learning*. Springer-Verlag New York, 2016.

[10] F. Ciompi, O. Geessink, B.E. Bejnordi, G.S. de Souza, A. Baidoshvili, G. Litjens, B. van Ginneken, I. Nagtegaal, and J. van der Laak. The importance of stain normalization in colorectal tissue classification with convolutional networks. *arXiv preprint arXiv:1702.05931*, 2017.

[11] Zanjani F. G., Zinger S., Bejnordi B. E., van der Laak J. A.W.M., and Peter H.N. de With. Stain normalization of histopathology images using generative adversarial networks. In *In Proc. IEEE 15th Int. Symp. on Biomedical Imaging (ISBI), Washington D.C.*, 2018 (in press).

[12] Cho H., Lim S., Choi G., and Min H. Neural stain-style transfer learning using gan for histopathological images. *CoRR*, abs/1710.08543, 2017.

[13] Qureshi H., Sertel O., Rajpoot N., Wilson R., and Gurcan M. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 196–204. Springer, 2008.

[14] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., and Bengio Y. Generative adversarial nets. *In Advances in neural information processing systems*, pages 2672–2680, 2014.

[15] 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*, 61(6):1729–1738, 2014.

[16] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young, and M. D. Wang. Automatic batch-invariant color segmentation of histological cancer images. *in Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, pages 657–660, 2011.

[17] X. Li and K.N. Plataniotis. A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics. *IEEE Transactions on Biomedical Engineering*, 62(7):1862–1873, 2015.

[18] X. Li and K.N. Plataniotis. Circular mixture modeling of color distribution for blind stain separation in pathology images. *IEEE journal of biomedical and health informatics*, 21(1):150–161, 2017.

[19] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, , and N. E. Thomas. A method for normalizing histology slides for quantitative analysis. *in Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, pages 1107–1110, 2009.

[20] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, and P. Quirke. Color normalization in digital histopathology images. *in Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endosc.*, pages 100–111, 2009.

[21] M. Niethammer, Borland D., J. Marron, J. Woosley, and N.E. Thomas. Appearance normalization of histology slides. *in Machine Learning in Medical Imaging. New York: Springer*, pages 58–66, 2010.

[22] Hamilton P., Allen DC., Watt P., Patterson C., and Biggart J. Classification of normal colorectal mucosa and adenocarcinoma by morphometry. *Histopathology*, 11(9):901–911, 1987.

[23] Kingma D. P. and M. Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations (ICLR)*, 2014.

[24] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. Technical University of Denmark, nov 2012. Version 20121115.

[25] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Comput. Graph. Appl.*, 21(5):34–41, 2001.

[26] A. C. Ruifrok and D. A. Johnston. Quantification of histochemical staining by color deconvolution. *Analyt. Quant. Cytol. Histol. Int. Acad. Cytol. Am. Soc. Cytol.*, 23(4):291–299, 2001.

[27] A. Sethi, L. Sha, A.R. Vahadane, R.J. Deaton, N. Kumar, V. Macias, and P.H. Gann. Empirical comparison of color normalization methods for epithelial-stromal classification in h and e images. *Journal of pathology informatics*, 7, 2016.

[28] L. Shaa, D. Schonfelda, and A. Sethi. Color normalization of histology slides using graph regularized sparse nmf. *Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology*, 2017.

[29] Aaron van den Oord and Benjamin Schrauwen. Factoring variations in natural images with deep gaussian mixture models. In *Advances in Neural Information Processing Systems 27*, pages 3518–3526. Curran Associates, Inc., 2014.

[30] Jeroen A.W.M. van der Laak, Martin M.M. Pahlplatz, Antonius G.J.M. Hanselaar, and Peter C.M. de Wilde. Hue-saturation-density (hsd) model for stain recognition in digital images from transmitted light microscopy. *Cytometry*, 39(4):275–284, 2000.

[31] Y. Wang, S. Changa, L. Wu, S. Tsai, and Y. Sun. A color-based approach for automated segmentation in tumor tissue classification. *in Proc. Conf. IEEE Eng. Med. Biol. Soc.*, pages 6577–6580, 2007.

[32] M. D. Zarella, C. Yeoh, D. E. Breen, and F. U. Garcia. An alternative reference space for h&e color normalization. *PloS one*, 12(3), 2017.

[33] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International Conference on Learning Representations*, 2018.

Appendix: Zanjani F. G., Zinger S., Bejnordi B. E., van der Laak J. A.W.M., and Peter H.N. de With. Stain normalization of histopathology images using generative adversarial networks. In In Proc. IEEE 15th Int. Symp. on Biomedical Imaging (ISBI), Washington D.C., 2018 (in press).

STAIN NORMALIZATION OF HISTOPATHOLOGY IMAGES USING GENERATIVE ADVERSARIAL NETWORKS

Farhad Ghazvinian Zanjani[⋆] *Svitlana Zinger*[⋆] *Babak Ehteshami Bejnordi*[†]
Jeroen AWM van der Laak[†] *Peter H. N. de With*[⋆]

[⋆] Department of Electrical Eng., Eindhoven University of Technology, Eindhoven, The Netherlands.
[†] Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands.

ABSTRACT

Computational histopathology involves CAD for microscopic analysis of stained histopathological slides to study presence, localization or grading of disease. An important stage in a CAD system, stain color normalization, has been broadly studied. The existing approaches are mainly defined in the context of stain deconvolution and template matching. In this paper, we propose a novel approach to this problem by introducing a parametric, fully unsupervised generative model. Our model is based on end-to-end machine learning in the framework of generative adversarial networks. It can learn a nonlinear transformation of a set of latent variables, which are forced to have a prior Dirichlet distribution and control the color of staining hematoxylin and eosin (H&E) images. By replacing the latent variables of a source image with those extracted from a template image in the trained model, it can generate a new color copy of the source image while preserving the important tissue structures resembling the chromatic information of the template image. Our proposed method can instantly be applied to new unseen images, which is different from previous methods that need to compute some statistical properties on input test data. This is potentially problematic when the test sample sizes are limited. Experiments on H&E images from different laboratories show that the proposed model outperforms most state-of-the-art methods.

***Index Terms*—** Computational histopathology, stain normalization, generative adversarial networks.

1. INTRODUCTION

Histopathology involves a manual staining procedure for preparing tissues prior to microscopic imaging for cancer diagnosis. This non-quantified procedure may cause a considerable variation in color characteristics of tissue samples. However, computer-aided diagnosis (CAD) is affected by the variations in color and intensity of the images. To compensate these effects in a CAD system, stain color normalization is a common practice. Recent studies show that statistical normalization of data in general [1, 2] and additional color normalization of H&E stained histopathology images can increase the computational efficiency of CAD models and leads to a higher performance [3–5]. Various stain normalization methods have been proposed for stain color normalization in histopathology images [5–19]. Previously published studies often follow one the of two main concepts: stain color deconvolution and template matching. In this section, we briefly explain these two approaches, their properties and drawbacks, and then we describe our contributions.

1. *Stain deconvolution* [20] and its variants are used extensively in histopathology image analysis [7]. By considering prior knowledge of the reference stain vector for every dye, which is present in whole-slide images (WSIs), this approach splits an input RGB image into three stain channels, each representing the actual color of the stain used. Ruifrok *et al.* [20] introduced this prior knowledge by manually selecting pixels which represent a specific stain class and then computing the color deconvolution vector. Because of some drawbacks of this semi-automatic procedure, several studies have been done later for automatic extraction of stain by using the singular value decomposition technique [10], probabilistic Gaussian mixture model (GMM) [12], using a prior for stain matrix estimation [6] and stain color descriptions along with training a supervised relevance vector machine [7]. Although these solutions all aimed at better estimation of stain vectors, the estimation of the stain vectors has been solely restricted to limited analysis of color information in the image contents, while the spatial dependency among tissue structures has been ignored [5]. Such ignorance causes some shortcomings for approaches based on stain deconvolution when severe staining variations occur in the data.

2. *Template color matching* has been proposed by *Reinhard et al.* [8] relies on aligning the statistical color properties (e.g. mean and standard deviation) of the source image with a template image. The authors used a set of linear transforms for assigned unimodal distribution to each channel of the Lab color model. Then, each channel of Lab color coordinates was treated independently for alignment, however, such dependencies are valid for dye contribution to the final color appearance [5]. The drawbacks of such an approach have been mentioned in [5, 7]. For addressing this problem, sep-

arate transformations are performed on stain classes [5, 12] or on tissue classes [13]. For avoiding artifacts at the border of different classes under different transformations, a weighted contribution of these transformations in the final color image has been considered. Two proposed solutions are estimated weights of the GMM [12] and training a naive Bayesian classifier [5]. Although the mentioned studies try to reduce the artifacts in normalized color images, using a template matching approach has inherently two major drawbacks. Firstly, the color transformation between the source image and the template image is split into two transformation classes (one for background with no staining and one for all tissue structure in the foreground [12]), or three transformation classes (background, elliptical hematoxylin-stained cell nuclei and other eosin-stained tissue structures in H&E staining). Hypothesis of solely presenting a limited (e.g. up to three) number of classes for structures present in images can be easily violated if the tissue type changes, and defining more tissue classes needs prior shape knowledge and designing appropriate discriminative shape descriptors for classification. Secondly, the approaches for performing inference on new samples with unknown chromatic characteristics start with computing statistical properties and then fitting a model to the data, which can be prone to failure if the number of test samples is limited. Addressing the first issue, our model implicitly learns to extract different image structures that have the same chromatic characteristics. It lacks any threshold, ground truth or any other assumption about the shape and color of structures present in images. Having less assumptions about the image contents enables the method to be generalized and applicable to different histopathology images corresponding with different tissue types. Regarding the second issue, since our model is based on artificial neural networks, it can be applied instantly on test images with unknown chromatic distribution. As our model does not rely on extraction of statistical properties from the source images to align with a template image, the size of test data does not affect the performance of method.

In this paper, we propose a new generic model for color normalization based on generative adversarial networks (GANs). Our contribution to the problem is two-fold. At first, we introduce an end-to-end learning model based on CNNs which can learn both the image-content structures and the relation of these structures to their color attributes. Unlike the previous studies, there are neither any hard constraints on the number of classes for image structures nor any explicit assumptions about the image contents, such as defining prior shape for cells and etc. Second, since our model benefits from the deep CNNs for nonlinear approximation of image data distribution over chromatic space, it can align the color distribution between source and template image better and not only based on some statistical properties like mean, standard deviation [8], [12] or the first Eigenvector of the covariance matrix in chromatic plane [5]. In the remainder of this paper, we explain our model in detail. Afterwards, the performed experiments and obtained results, compared with other state-of-the-art methods, are presented. At the end, we provide a discussion and conclusions.

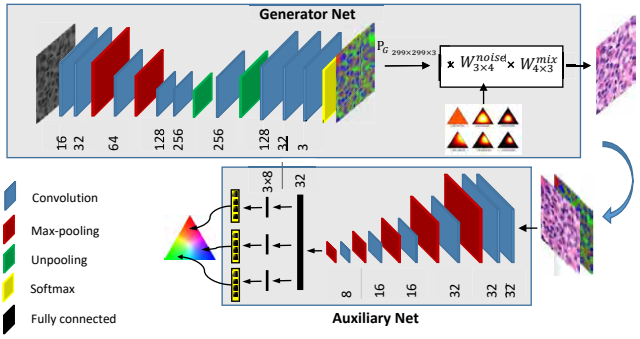


Fig. 1: Schematic of the generator $\mathcal{G}$ and the auxiliary $\mathcal{Q}$ networks (the discriminator network $\mathcal{D}$ was not shown here).

2. PROPOSED METHOD

Our model is based on GANs [21], using a minmax game. The goal in GANs is learning a generative distribution $P_G(x)$ that matches the real data distribution $P_{data}(x)$. The GANs include a generator network $\mathcal{G}$ that generates samples $\mathcal{G}(z)$ from generator distribution P_G, given a noise variable z drawn from $P_{noise}(Z)$. The generator is trained by playing against an adversarial discriminator network $\mathcal{D}$ that tries to distinguish between samples from true data distribution P_{data} and the generator distribution P_G [21]. The minmax game is given by the value function as below:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} V(\mathcal{D},\mathcal{G}) = E_{x\sim P_{data}}[log\mathcal{D}(x)] + E_{z\sim noise}[log(1-\mathcal{D}(\mathcal{G}(z)))]. \quad (1)$$

While the standard GANs formulation does not have any restriction on the contribution of noise variables in generating the samples, it can cause the individual dimension of z to be highly entangled and not correspondence with semantic features of the data [22] (e.g. with a realistic colorizing of the H&E images in our case study). *InfoGAN* [22] tries discovering a set of semantic structured latent variables $\mathcal{C} = \{c_1, c_2, ..., c_L\}$ that along with the noise variable $\mathbf{z}$, are given to the generator model $\mathcal{G}(\mathbf{z}, \mathcal{C})$. Discovering latent factors $\mathcal{C}$ in an unsupervised way is performed by adding an information-theoretic regularization term to the minmax value function of the GANs [22]:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} V_I(\mathcal{D},\mathcal{G}) = V(\mathcal{D},\mathcal{G}) - \lambda I(\mathcal{C};\mathcal{G}(\mathbf{z},\mathcal{C})). \quad (2)$$

Here, $I(.)$ represents mutual information and λ is a contribution weight (in our experiments $\lambda = 1$). Because computing $I(\mathcal{C};\mathcal{G}(\mathbf{z},\mathcal{C}))$ needs access to the posterior $P(\mathcal{C}|x)$, which is hard to maximize, the authors suggest using an auxiliary distribution $Q(\mathcal{C}|x)$ as a lower bound to approximate $P(\mathcal{C}|x)$. Estimating $Q(\mathcal{C}|x)$ is performed by training a neural network called *auxiliary* network ($\mathcal{Q}$). We used the same concept of InfoGAN for learning the chromatic distribution of H&E images and consequently generating different colorized versions

of images with adjusting the distribution parameters. Our model is slightly different from InfoGAN for two reasons. Firstly, our model aims to be used as a color-normalizing generative model which must preserve the structures present in the source image. Therefore, instead of noise, the *lightness* channel of the CIEL*a*b color coordinates of source images are supplied to the generator $\mathcal{G}$. The CIEL*a*b color space has been chosen because of its higher performance for reconstructing histopathological images [23]. Secondly, in our model, the structured latent variables $\mathcal{C}$ have the role of color system matrix for representing color for image structures. Figure 1 shows the schematic of generator and auxiliary networks. The model consists of three CNNs: generative network ($\mathcal{G}$), discriminator network ($\mathcal{D}$) and auxiliary network ($\mathcal{Q}$), which are trained simultaneously. A detailed description of the each network is provided below.

Generator network learns how to generate a colorized H&E image in its CIEL*a*b space by giving the image lightness channel and a set of structured latent variables drawn randomly from a prior distribution. The $\mathcal{G}$ net consists of several convolutional, Rectified Linear Unit (RELU) nonlinearity functions, max pooling and batch normalization operators in its hidden layers. The given lightness channel after passing through the previously mentioned computational layers maps to a latent k-simplex probability subspace (e.g. $k = 3$) by using a softmax layer. Each point in this probability simplex represents a k-dimensional probability vector $P_{G_{1\times3}}$ that softly clusters the pixels of the input image into k clusters. Afterwards, the produced $P_{G_{1\times3}}$ vectors in this latent space are passed on to the second part of $\mathcal{G}$ net, which are transformed linearly to the output for generating the full-color images. Analogous to the *color system matrix* of Ruifrok *et al.* [20], we call this linear transformation the color system transform. This transformation consists of two matrix multiplications: $\mathbf{W}_{noise_{3\times4}}$ and $\mathbf{W}_{mixture_{4\times3}}$. Since the elements of each i-th row of the W_{noise} matrix represent Dirichlet distributions ($i = 1, 2, 3$) with random parameters ($\alpha^i = [\alpha_1, \alpha_2, ..., \alpha_4]$), this transformation can be considered as a stochastic process on $P_{G_{1\times3}}$ vectors. To avoid swapping colors between image structures in each training iteration, We put a constraint ($argmax(\alpha^i) = i$) on randomly drawn α^i parameters. This constraint forces the rows of the color system matrix to be sampled from three isolated regions of the probability simplex, so it leads to assigning consistent colors to structures.

We use Dirichlet distribution as a prior for the color system matrix. If we consider the contribution of used staining dyes in colorizing each pixel as a multinomial distribution over dyes, then using a different amount of dye or any variation in the staining procedure alters the parameters of the multinomial distribution. The Dirichlet distribution prior allows for more flexible modeling of data, when considering it as a distribution over possible parameter vectors of the multinomial distribution. In fact, it can be seen as a "distribution over distribution" [24]. From the experiments we observe that using such a prior distribution for $\mathbf{W}_{noise}$ can lead to a good approximation of the posterior in a generative model. The $\mathbf{W}_{mixture}$ transformation is used for shifting and scaling of network output to produce CIEL*a*b space of images. The elements of $\mathbf{W}_{mixture}$ matrix and all parameters of $\mathcal{G}$ net are learned by gradient descent optimization. As standard GANs framework, the $\mathcal{G}$ net does not have its own loss function and its parameters are jointly optimized with the $\mathcal{D}$ net, by using minmax loss that is defined at the output of the $\mathcal{D}$ net.

Auxiliary network ($\mathcal{Q}$) has been defined in the context of the *InfoGAN* model [22], for learning a disentangled latent space by the GANs. The $\mathcal{Q}$ net has a reverse functionality compared with the second part of $\mathcal{G}$ net. It receives the generated images along with their P_G vectors from the $\mathcal{G}$ net and estimates the elements of $\mathbf{W}_{noise}$ at its output. Its loss function has been defined to maximize the mutual information between the output and the elements of the $\mathbf{W}_{noise}$ matrix.

Discriminator network ($\mathcal{D}$) minimizes its loss function and tries to distinguish the generated color images from the original ones. The architecture of $\mathcal{D}$ net is very similar to $\mathcal{Q}$ net and consists of similar convolutional hidden layers, but different from $\mathcal{G}$ net, instead of the last max pooling and the fully connected layers, it has a global average pooling layer and a single sigmoid neuron as its output. The probability at the output of $\mathcal{D}$ net is learned to be maximal (e.g. equal to unity) when given a real image and is minimal (e.g. zero) when supplied with a fake generated image from the $\mathcal{G}$ net.

Reconstruction loop: For training of the $\mathcal{G}$, $\mathcal{D}$ and $\mathcal{Q}$ networks, apart from value function in Equation 2, we perform a reconstructing loop as supervisory signal through $\mathcal{G}$ and $\mathcal{D}$ nets. This introduces another loss function which measures the reconstructing quality of a given real image. Our experiments shows that this extra procedure helps for speeding up the convergence of networks. To do so, we first apply $\mathcal{G}$ net to the lightness channel of an input train sample. It produces P_G vectors at its hidden layers. We feed the same real image along with the obtained P_G vectors as input to the $\mathcal{Q}$ net. Consequently, the $\mathbf{W}_{noise}$ matrix replaced with the estimated values at the output of $\mathcal{Q}$ net. Now, the $\mathcal{G}$ should generate the same full-color copy of the given image. The $L2$ loss between input and the generated sample is used for measuring the reconstruction quality. This process is almost similar to what happens by an auto-encoder network.

The number of kernels (of size 3×3) per convolutional layer is shown in Figure 1. The ADAM optimizer with fixed learning rate equal to $10e - 4$ has been used for sequential minimization of the losses of all these three networks. In inference mode, the $\mathcal{Q}$ and $\mathcal{G}$ networks are first applied to the template image, so the color system matrix is obtained. Afterwards, any given source image to the $\mathcal{G}$ net along with the obtained color system matrix from the template image, converts the color of the source image to resemble the target image.

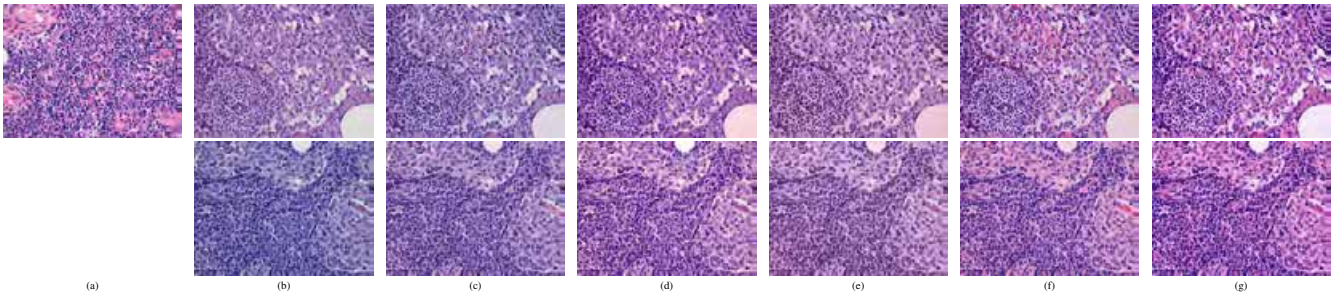


Fig. 2: Performance of different stain color-normalization methods on two H&E images from two different labs. (a) template image (b) original images, (c) Macenko *et al.* [10], (d) Reinhard *et al.* [8], (e) Khan *et al.* [7], (f) Bejnordi *et al.* [5], (g) ours.

Table 1: Standard Deviation (SD) and Coefficient of Variation (CV) of NMI for all five laboratories for hematoxylin dye.

Method	**Lab 1**		**Lab 2**		**Lab 3**		**Lab 4**		**Lab 5**		**Average**	
	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV
Original	0.033	0.065	0.031	0.060	0.037	0.078	0.029	0.049	0.028	0.051	**0.032**	**0.060**
Macenko [10]	0.029	0.052	0.026	0.046	0.020	0.037	0.025	0.044	0.020	0.035	**0.024**	**0.043**
Reinhard [8]	0.032	0.058	0.025	0.044	0.020	0.035	0.030	0.052	0.029	0.049	**0.027**	**0.047**
Khan [7]	0.066	0.156	0.067	0.155	0.085	0.158	0.054	0.110	0.049	0.093	**0.064**	**0.135**
Bejnordi [5]	0.016	0.029	0.015	0.027	0.018	0.034	0.029	0.055	0.024	0.044	**0.021**	**0.038**
Ours	0.024	0.053	0.019	0.043	0.020	0.043	0.027	0.057	0.024	0.053	**0.022**	**0.050**

Table 2: Standard Deviation (SD) and Coefficient of Variation (CV) of NMI on 25 images from five labs for eosin dye.

Method	**Original**	**Macenko [10]**	**Reinhard [8]**	**Khan [7]**	**Bejnordi [5]**	**Ours**
NMI SD	0.0563	0.0362	0.0386	0.0434	**0.0191**	0.0195
NMI CV	0.0748	0.0439	0.0494	0.0555	0.0220	**0.0218**

3. EMPIRICAL EVALUATION

Histopathology image dataset: We focus on inter-laboratory variations of the H&E staining in the lymph node dataset, as this is a major concern in large-scale application of CAD in pathology. For better comparison with recent studies, we used a similar dataset as introduced in [5]. The dataset contains 625 images (each 1388×1040 pixels) from 125 digitized H&E stained WSIs of lymph nodes from 3 patients, collected from five different Dutch pathology laboratories, each using their own routine staining protocols. More details about this dataset can be found in [5]. Our model is trained on 299×299 randomly cropped patches and evaluated on full size images by using leave-one-out cross-validation, based on the laboratories where the samples were collected.

Results: The performance of our method is compared to that of four previously published algorithms: linear appearance normalization by Macenko *et al.* [10], statistical color properties alignment by Reinhard *et al.* [8], nonlinear mapping for stain normalization by Khan *et al.* [7] and whole-slide image color standardizer by Bejnordi *et al.* [5]. The normalized median intensity (NMI) measure [13], [5] is used to evaluate color constancy of normalized images. Quantitative analysis of results is based on color constancy of nuclear staining and eosin staining, independently. To evaluate the color constancy of the nuclear staining, nuclei were first detected automatically. Similar to [5], the fast radial symmetry transform [25] and marker-controlled watershed [26] algorithms are used for nuclei detection. Eosin-stained regions are manually annotated for 25 images. The results for hematoxylin regions and eosin regions are shown in Table 1 and Table 2, respectively. The results indicate that our proposed model outperforms many previous methods for stain color normalization. Figure 2 illustrates normalizing of two example images from two different labs by using different methods. As can be observed, our method has acceptable qualitative performance in normalizing different image structures including the white background to resemble the template image.

4. DISCUSSIONS AND CONCLUSIONS

In this paper we proposed a new model based on end-to-end machine learning methods for stain color normalization. Our generative model can learn the chromatic space of H&E images and normalizes them. The color normalized image preserves the structures of the source image while forced to have a high mutual chromatic information with a template. In contrast to most of the previous methods which prior computing of statistical properties from source and template images are essential in inference time, our model can be applied instantly on unseen given images. This can be crucial if the number of test samples are small. Moreover, our proposed framework has minimal assumptions about the number, shape, color and other image attributes of H&E images. This leads to a generic model that can be applied to histopathological images from different organs containing different tissue structures and potentially to other stainings such as immunohistochemistry staining as well. The experiments showed that our proposed model can outperform many previous methods qualitatively and quantitatively.

5. REFERENCES

[1] N.M. Nawi, W.H. Atomi, and M.Z. Rehman, "The effect of data pre-processing on optimized training of artificial neural networks," *Procedia Technology*, vol. 11, pp. 32–39, 2013.

[2] T. Jayalakshmi and A. Santhakumaran, "Statistical normalization and back propagation for classification," *International Journal of Computer Theory and Engineering*, vol. 3, no. 1, 2011.

[3] F. Ciompi, O. Geessink, B.E. Bejnordi, G.S. de Souza, A. Baidoshvili, G. Litjens, B. van Ginneken, I. Nagtegaal, and J. van der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," *arXiv preprint arXiv:1702.05931*, 2017.

[4] A. Sethi, L. Sha, A.R. Vahadane, R.J. Deaton, N. Kumar, V. Macias, and P.H. Gann, "Empirical comparison of color normalization methods for epithelial-stromal classification in h and e images," *Journal of pathology informatics*, vol. 7, 2016.

[5] B.E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Hller, A. Homeyer, N. Karssemeijer, and J. AWM van der Laak, "Stain specific standardization of whole-slide histopathological images," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 404–415, 2016.

[6] M. Niethammer, Borland D., J. Marron, J. Woosley, and N.E. Thomas, "Appearance normalization of histology slides," *in Machine Learning in Medical Imaging. New York: Springer*, pp. 58–66, 2010.

[7] 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*, vol. 61, no. 6, pp. 1729–1738, 2014.

[8] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, 2001.

[9] A. K. Jain, "Fundamentals of digital image processing," Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[10] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, , and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," *in Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, pp. 1107–1110, 2009.

[11] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young, and M. D. Wang, "Automatic batch-invariant color segmentation of histological cancer images," *in Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, pp. 657–660, 2011.

[12] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, and P. Quirke, "Color normalization in digital histopathology images," *in Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endosc.*, pp. 100–111, 2009.

[13] Anant Madabhushi Ajay Basavanhally, "Em-based segmentation-driven color standardization of digitized histopathology," *Proc.SPIE*, vol. 12, pp. 8676 – 8676 – 12, 2013.

[14] P. A. Bautista, N. Hashimoto, and Y. Yagi, "Color standardization in whole slide imaging using a color calibration slide," *Pathol. Inf.*, vol. 5, 2014.

[15] X. Li and K.N. Plataniotis, "A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1862–1873, 2015.

[16] X. Li and K.N. Plataniotis, "Circular mixture modeling of color distribution for blind stain separation in pathology images," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 150–161, 2017.

[17] L. Shaa, D. Schonfelda, and A. Sethi, "Color normalization of histology slides using graph regularized sparse nmf," *Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology*, 2017.

[18] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, and N.M. Rajpoot, "Stain deconvolution using statistical analysis of multi-resolution stain colour representation," *PloS one*, vol. 12, no. 1, 2017.

[19] Y. Wang, S. Changa, L. Wu, S. Tsai, and Y. Sun, "A color-based approach for automated segmentation in tumor tissue classification," *in Proc. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 6577–6580, 2007.

[20] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analyt. Quant. Cytol. Histol. Int. Acad. Cytol. Am. Soc. Cytol.*, vol. 23, no. 4, pp. 291–299, 2001.

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Yoshua Bengio, "Generative adversarial nets," *In Advances in neural information processing systems*, pp. 2672–2680, 2014.

[22] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *In Advances in Neural Information Processing Systems*, pp. 2172–2180, 2016.

[23] G. Bueno, O. Dniz, J. Salido, M.M. Fernndez, N. Vllez, and M.G. Rojo, "Colour model analysis for histopathology image processing," *In Color medical image analysis, Springer Netherlands*, pp. 165–180, 2013.

[24] L. Du, H. Lang, Y.L. Tian, C.C. Tan, J. Wu, and H. Ling, "Covert video classification by codebook growing pattern," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 11–18, 2016.

[25] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959973, 2003.

[26] R. Moshavegh, B. E. Bejnordi, A. Mehnert, K. Sujathan, P. Malm, and E. Bengtsson, "Automated segmentation of free-lying cell nuclei in pap smears for malignancy-associated change analysis," *in Proc. Annu. Int. Conf. IEEE EMBC*, p. 53725375, 2012.