

Adaptive Color Deconvolution for Histological WSI Normalization

Yushan Zheng^{a,b,c}, Zhiguo Jiang^{a,b,c,*}, Haopeng Zhang^{a,b,c}, Fengying Xie^{a,b,c}, Jun Shi^d, Chenghai Xue^e

^a*Image Processing Center, School of Astronautics, Beihang University, Beijing, 100191, China, {yszheng, jiangzg, zhanghaopeng, xfy_73@buaa.edu.cn}*

^b*Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, 100191, China*

^c*Beijing Key Laboratory of Digital Media, Beihang University, Beijing, 100191, China*

^d*School of Software, Hefei University of Technology, Hefei 230601, China*

^e*Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China*

Abstract

Background and objective: Histological slides are stained with multiple dyes to color different types of tissues. Color consistency of histological images is significant for developing reliable computer-aided diagnosis (CAD) systems. However, the color appearance of digital histological images varies across different specimen preparations, staining, and scanning situations. This variability affects the diagnosis and decreases the accuracy of CAD approaches. It is important and challenging to develop effective color normalization methods for digital histological images.

Methods: In this study, we proposed a novel color normalization approach for hematoxylin-eosin-stained whole slide images (WSIs) based on a designed adaptive color deconvolution (ACD) algorithm. Using the ACD model, the color deconvolution matrix for a specific WSI is estimated through an integrated optimization. The prior knowledge, the density of staining, and the ratio for different stains are simultaneously considered in the optimization, which delivers consistent performance of stain operation. normalization results.

Results: The proposed ACD model is evaluated based on color normal-

*Corresponding author: Zhiguo Jiang, Tel.:+86 10 82316173.

ization, stain separation, computational complexity, and effectiveness for CAD approaches on datasets of metastasis in lymph node. The results demonstrate that the proposed normalization is effective and can preserve the structure of objects in the WSI. The proposed model can be solved in 10 s. The area under receiver operating characteristic curve for metastases image classification is 0.842 and 0.914 before and after the normalization.

Conclusions: The proposed normalization method has a robust performance with hematoxylin-eosin-stained WSIs in various color appearances and is superior to other methods in both quantitative and qualitative assessments. The texture and structure of tissue in the images are well preserved in the normalization. The proposed method is time-saving, which is applicable to developing an efficient automatic CAD systems.

Keywords: Color normalization, digital pathology, stain separation, WSI, CAD

1. Introduction

Cancer diagnosis still relies on histopathology [1], which involves the microscopic examination of tissue to study the manifestations of the disease. Histological slides are stained with multiple dyes to color different types of tissues [2]. With the development of digital pathology, histological slides can be scanned rapidly using advanced micro-scanners and stored as digital whole slide images (WSIs). This enables pathologists to view slides on a screen. Based on digital WSIs, an increasing number of computer-aided-diagnosis (CAD) approaches have emerged in the last two decades [3, 4, 5, 6]. A competent WSI-based CAD system can help pathologists locate diagnostically relevant regions from the WSIs [7, 8], which can help improve the efficiency and reliability of diagnoses.

Color consistency of digital WSIs is quite important for CAD based on WSI analysis [9, 10]. In practical applications, the appearance of WSIs varies due to different specimen preparations, staining situations, and section scanners [11]. This variability affects the diagnosis and decreases the accuracy of CAD

approaches. To overcome the variability, many color normalization methods for histological images are proposed. [12]

A group of methods realized the normalization via histogram transformation in different color space, such as in Lab space [13], in HSV space [8], and in RGB space [?]. These methods were designed with reference to the normalization of natural scene images and ignores the prior knowledge that histological images are colored by multiple stains. More methods are based on the prior knowledge that the color of WSIs is the combination of several independent stains. One of the most popular methods is color deconvolution (CD) [14]. CD has proven effective as a pre-processing method for WSIs in CAD approaches [15, 16, 17, 18, 19, 20]. However, the parameters of CD are estimated by a set of pixels selected manually for each stain component. The manual interaction becomes a limitation of CD for developing CAD approaches. To relieve the dependence of interaction, the automatic approach of estimating CD parameters for specific images was proposed [21]. Then, the normalization methods based on CD were emerged [22, 23]. The transformations between the source image and the template image were established through CD and its inverse operation.

A set of these methods [23, 24, 11] established separate transformations for different stains, where pixels belonging to the stains were classified or clustered beforehand by designed algorithms. The robustness of these methods relies on the pixel classification algorithm. Moreover, the normalization via multiple transformations may cause structural artifacts at the border of different stain classes.

In contrast, a number of methods proposed establishing an integral transformation on all the pixels for normalization [25, 26, 27, 28, 29]. Instead of directly classifying pixels for different stain classes, the prior knowledge of section staining was embedded in the solving of the transformation model. The color of the normalized images is relatively smoother than that obtained by multiple-transformation-based methods, and thus the structure can be effectively preserved [25]. However, the estimation of CD parameters in these methods is generally independent of the adjustment of overall intensity and proportion of

the separated stains. [25, 29]. Especially, the proportion of stains is not considered into the solving of estimation model. This may risk a bias in the optimized CD model, for which most of the pixels would be transformed to share the appearance of single stain. Moreover, the normalization model relies on the statistical properties of separated stains of the source image and the template image, which generates additional computation.

In this study, we propose a novel color normalization method for hematoxylin-eosin-stained (H&E-stained) WSIs based on a designed adaptive color deconvolution (ACD) model. The normalization is realized through an integral color transformation for pixels between the source image to the template image. Compared to methods [24, 11], our approach does not rely on the classification of stains. The parameters for the two stains (hematoxylin and eosin) are simultaneously estimated through an integral optimization. The structural information of histological images can be effectively preserved. Different from methods [25, 26, 27], the prior knowledge of H&E-stain, the proportion of different stains, and the overall intensity of staining are simultaneously considered in the modeling of ACD, which effectively reduce the failure rate of estimation of CD parameters and thus delivers a more robust normalization performance. Furthermore, the color normalization model can be directly obtained through multiply operation between the ACD parameters of the source and template images, which is more convenient than other transformation-based methods [25, 27]. In terms of computation, both the solution and application stages of the proposed method only contain pixel-wise operation and involve no pixel interaction. It determines the proposed method is light, efficient, and applicable for developing automatic CAD programs and systems based on WSI analysis.

The proposed method was evaluated on aspects of color normalization, stain separation, computational complexity, and effectiveness for CAD approaches on four histopathological image dataset, and was compared with the state-of-the-art methods [8? , 24, 11, 26, 25]. The experimental results have shown that the proposed method is robust in color normalization for H&E-stained histological images and is superior to the compared methods in both quantitative

and qualitative assessments.

The remainder of this paper is organized as follows. Section 2 reviews the related studies. Section 3 introduces the methodology of the proposed method. The experiment is presented in Section 4. Section 5 provides necessary discussions and suggests directions for future work. Finally, Section 6 summarizes the contributions.

2. Related work

Most of the works related to our method are reviewed in this section. The histogram-transformation-based methods are first discussed. Then, the methods based on stain transformations are roughly classified and reviewed in two categories: separated-transformation-based methods, and integral-transformation-based methods.

2.1. Histogram-transformation-based methods

Wang et al. [13] introduced a linear color transform method [30] into histological images. The color is transformed to a template image with a linear projection in $l\alpha\beta$ color space. Zheng et al. [8] proposed normalizing WSIs in HSV-space, in which the saturation and value channels were stretched linearly for standardization. The method can normalize global illumination and saturation of WSIs and has proven effective in improving the performance of WSI analysis. Janowczyk et al. [?] realized stain normalization using Sparse auto-encoders (StaNoSA). Pixels in histological images were clustered based on the features generated by the sparse auto-encoders and the pixels belonging to the same cluster were transformed to the template image using a specific histogram projection. These methods derive from nature scene image processing and barely utilize the staining characteristic of histological images. As a result, the tissue area and background were occasionally confused after the normalization [23].

2.2. Separated-transformation-based methods

Based on the prior knowledge that histological images are colored by independent stains, the methods based on color deconvolution were proposed [23, 22]. Typically, Khan et al. [24] proposed a specific color deconvolution (SCD) algorithm. In this method, pixels belonging to the same stain were extracted through a classification model and an image-specific staining vector was estimated for the stain by analyzing the RGB-allocation of these pixels. Based on the staining vector, a non-linear mapping approach was designed to transfer the image color to the template image. For more stable performance, the prior knowledge of nuclei structure was considered for the pixel classification process. In [11], the nuclei were detected by using the Hough transform [31] method. Then, the pixels belonging to the nuclei and cytoplasm were accurately classified. However, the Hough transform method requires additional computation for the normalization, which may become the bottleneck of an efficient CAD approach. Moreover, the separated-transformation-based methods probably bring color discontinuity into normalized images, which may occasionally cause improper structural changes within the nuclei.

2.3. Integral-transformation-based methods

Compared to the aforementioned methods, An increasing number of methods proposed to establish an integral transformation model for all pixels in the images. These methods commonly deliver a more reasonable visual performance and help preserve the local contrast in histological images. Zhou et al. [26] proposed modifying the CD matrix [14] through an optimization for H&E-stained histological images. The variable of the optimization is the exact CD matrix. The objective function enforces the third channel (i.e., the background channel) of the deconvolution result to be zero. However, the specificity of H&E-staining is not considered in the optimization and the objective is almost equivalent to reduce the weight of the third row of the CD matrix, for which the normalization performance is limited.

Referring to CD [14], the items in the CD matrix are not independent variables but decided by a stain color appearance (SCA) matrix. In contrast, the SCA matrix is more reasonable than CD matrix to be the variables of normalization models. Li et al. [32, 33] regarded the SCA matrix as the model variables and applied non-negative matrix factorization (NMF) to solving the corresponding model. Furthermore, sparse constraint was considered in NMF model for the assumption that most of the pixels contain one type of stain [28, 25, 27]. Typically, Vahadane [25] designed a normalization approach for WSIs based on sparse NMF (SNMF). The sparse constraint enhanced the recognition ability of the model for independent stains. However, the SNMF model did not consider the proportion of the stains in the solution. The color of image was occasionally recognized to derive from the same stain, which caused color artifacts in the normalization results.

Recently, Zanjani [34] et al. proposed a parametric model based on generative adversarial networks (GAN), where the generator was trained to convert the color of source image to the template image. It contributes to a stable performance of color normalization. While, the deep structure with constitutional operations in the generator brings high computational cost into the normalization, which may affect the efficiency of the entire CAD system.

Building on these methods, we propose an adaptive color deconvolution (ACD) model for stain separation and color normalization of histological images. Besides the prior knowledge of H&E-stained images that utilized in recent studies [25, 27, 26], the consideration for proportion of different stains, and the overall intensity of staining are embedded in the model. Taking the SCA matrix as the variables, the ACD model is solved through an integral optimization. The proposed method inherits the advantage of the integral-transformation-based methods [25, 27]. The structural information of histological images is well preserved. Furthermore, because of the staining constraint and integral optimization, the color artifacts are effectively reduced.

Category	Model Consideration		
	Staining specificity	Structure preservation	Color consistency
Histogram trans.	No	No (W)	Yes (M)
Separated trans.	Yes	Yes (M)	Yes (S)
Integral trans.	Yes	Yes (S)	Yes (W)

3. Method

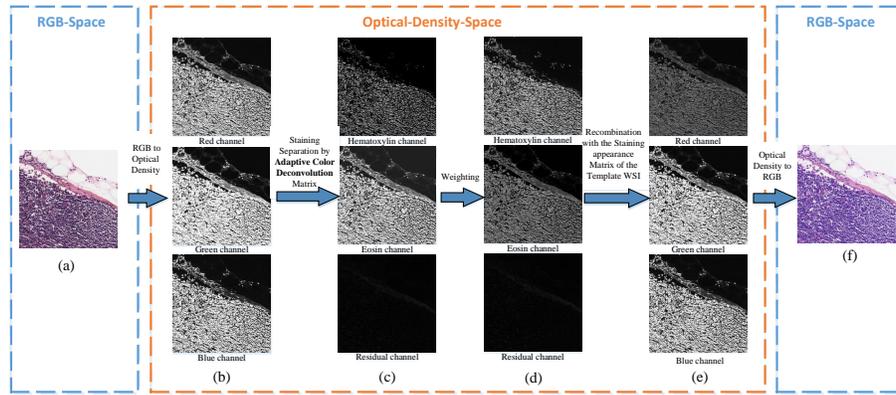


Figure 1: Flowchart of the proposed normalization method, where (a) denotes the original WSI (A featured sub-region of the WSI is chosen to display.), (b) is the visualization of R/G/B channel in the optical-density-space, (c) shows the density of hematoxylin and the eosin stains separated by an adaptive color deconvolution matrix, (d) displays the re-weighted stains, (e) shows the R/G/B channel recombined with the stain parameters of the template WSI, and (f) shows the result of the normalization.

3.1. Overview

Figure 1 presents the flowchart of the proposed normalization. For a certain WSI, a group of pixels are sampled from the tissue region and converted into optical density (OD) space. The normalized H&E components are obtained based on an ACD matrix and a stain-weight matrix for the WSI. Finally, the H and E components are recombined with the SCA matrix of a template WSI,

achieving the color normalization. The approach to obtain the ACD matrix and the stain-weight matrix are the essential of our method. In this section, the ACD model is first introduced and then the normalization method based on ACD is described.

3.2. Color deconvolution

The theory of color deconvolution (CD) [14] is the basis of ACD. CD is proposed based on Beer-Lambert law. The stains are separated in the optical density space. Letting $\mathbf{x}_i \in \mathbb{R}^{3 \times 1}$ denote the value in RGB color space for the i -th pixel in a WSI, CD can be briefly represented with the following equations

$$\begin{aligned}\mathbf{o}_i &= -\ln(\mathbf{x}_i/I_{max}) \\ \mathbf{s}_i &= \mathbf{D} \cdot \mathbf{o}_i\end{aligned}\tag{1}$$

where $\mathbf{o}_i \in \mathbb{R}^{3 \times 1}$ denotes the optical density (OD) of RGB channels, $\mathbf{D} \in \mathbb{R}^{3 \times 3}$ is the so-called color deconvolution matrix, and $\mathbf{s}_i \in \mathbb{R}^{3 \times 1}$ is the output that contains stain densities. I_{max} denotes the intensity of background, i.e. the value of pixel when no stained tissue is present. The exact value of I_{max} varies with different section scanners. Generally I_{max} approximates the maximum of digital image intensity (255 for 8-bit data format). For H&E-stained WSIs, the separated densities of stains can be represented as $\mathbf{s}_i = (h_i, e_i, d_i)^T$, where h_i and e_i are the values for hematoxylin and eosin stains, respectively, and d_i represents the residual of the separation. The deconvolution matrix \mathbf{D} is determined by a SCA matrix \mathbf{M} with an inverse operation $\mathbf{D} = \mathbf{M}^{-1}$. Further, \mathbf{M} can be manually measured using a designed experiment [14].

3.3. Adaptive color deconvolution

The ACD parameters are obtained by optimization. The variables, objective and solving of the optimization are presented in the following sub-sections.

3.3.1. Variables

Considering that the deconvolution matrix \mathbf{D} is determined by the SCA matrix \mathbf{M} , we propose directly optimizing \mathbf{M} and then calculating \mathbf{D} . Specifically, the SCA matrix can be decomposed as $\mathbf{M} = (\mathbf{m}_h, \mathbf{m}_e, \mathbf{m}_d)$, where $\mathbf{m}_j \in$

$\mathbb{R}^{3 \times 1} (j = h, e, d)$ is defined as SCA vector. In general, \mathbf{m}_j is a unit vector, which describes the contributions of R, G, and B channels to the j -th stain. To ensure $\mathbf{m}_j \equiv 1$, we propose representing \mathbf{m}_j using two degree variables as

$$\mathbf{m}_j = (\cos \alpha_j \sin \beta_j, \cos \alpha_j \cos \beta_j, \sin \alpha_j)^T, j = h, e, d.$$

Then, the SCA matrix \mathbf{M} can be represented by six independent degree variables. For convenience, the six degree variables are represented by a collection

$$\varphi = \{\alpha_h, \beta_h, \alpha_e, \beta_e, \alpha_d, \beta_d\},$$

the SCA matrix decided by φ is represented as $\mathbf{M}(\varphi)$, and the corresponding CD matrix is $\mathbf{D}(\varphi)$.

Since \mathbf{M} is composed of *unit* vectors, the entire intensity of \mathbf{s}_i is bounded by the value of \mathbf{o}_i . To improve the capacity of the model for normalizing stain intensities, a stain-weight matrix $\mathbf{W} = \text{diag}(w_h, w_e, 1)$ is defined to modify the CD algorithm (Eq. 1):

$$\begin{aligned} \mathbf{o}_i &= -\ln(\mathbf{x}_i / I_{max}) \\ \mathbf{s}_i &= \mathbf{W} \cdot \mathbf{D} \cdot \mathbf{o}_i. \end{aligned} \tag{2}$$

\mathbf{W} is also regarded as the variable of ACD model and simultaneously obtained with φ in the optimization.

3.3.2. Objective

An objective function about variables φ and \mathbf{W} are defined. By resolving the function, the optimized set of variables $\hat{\varphi}$ and $\hat{\mathbf{W}}$ are obtained, and then the adaptive matrices $\mathbf{M}(\hat{\varphi})$ and $\mathbf{D}(\hat{\varphi})$ for the WSI are determined. For brevity, $\mathbf{M}(\hat{\varphi})$ and $\mathbf{D}(\hat{\varphi})$ are also represented as $\hat{\mathbf{M}}$ and $\hat{\mathbf{D}}$ in this study.

The objective function for ACD is designed primarily on the basis of the following prior knowledge: (1) There are two types of stains in H&E-stained WSIs. Therefore, the third channel of the separated result (d_i) should be zero in ideal situation. (2) H&E staining has high specificity. Hematoxylin mainly stains nuclei and eosin mainly stains the cytoplasm and stroma. Therefore,

majority of pixels in images alternatively contain H or E stain. Based on the prior knowledge, the objective function is defined as

$$L_p(\varphi) = \frac{1}{N} \sum_{i=1}^N d_i^2(\varphi) + \lambda_p \frac{1}{N} \sum_{i=1}^N \frac{2h_i(\varphi)e_i(\varphi)}{h_i^2(\varphi) + e_i^2(\varphi)}, \quad (3)$$

where the first item of the function minimizes the residual of the separation, the second item enforces the value of a pixel being assigned to the same stain (H or E) after the separation, λ_p is the weight of the two items, and N is the number of pixels used for the optimization.

Besides the features considered above, the balance of the two stains and the overall density of staining are equally important for the normalization of WSIs, which are also considered in the model. First, an energy function to control the ratio of H and E components is defined:

$$L_b(\varphi) = \left[(1 - \eta) \frac{1}{N} \sum_{i=1}^N h_i(\varphi) - \eta \frac{1}{N} \sum_{i=1}^N e_i(\varphi) \right]^2, \quad (4)$$

where $\eta \in (0, 1)$ is defined as the balance parameter. Similarly, a function to control the overall density of staining is defined:

$$L_s(\varphi) = \left[\gamma - \frac{1}{N} \sum_{i=1}^N (h_i(\varphi) + e_i(\varphi)) \right]^2, \quad (5)$$

where γ controls the desired density of staining.

Then, the objective function is modified as

$$L(\varphi) = L_p(\varphi) + \lambda_b L_b(\varphi) + \lambda_s L_s(\varphi), \quad (6)$$

where λ_b and λ_s are the weights.

3.3.3. Solution

The objective is a function of variables φ and \mathbf{W} , and thus the optimization is described as

$$(\hat{\varphi}, \hat{\mathbf{W}}) = \arg \min_{(\varphi, \mathbf{W})} L(\varphi, \mathbf{W})$$

$L(\varphi, \mathbf{W})$ is continuous and differentiable for variables φ and \mathbf{W} . Therefore, we utilized a gradient descent algorithm to solve it. The derivatives of the objective function on the variables of the model are given in the appendix.

In the optimization, only the pixels located on the tissue area are used. The background (devoid of stain) area in WSIs is approximately white, and the optical density of background pixels is close to zero. Therefore, the background pixels can be easily filtered by a threshold [23, 11, 25]. Specifically, the pixels within $\mathbf{o}_i < T_{back}$ are recognized as background. T_{back} was tuned in the interval of [0.2, 0.5] and determined as 0.28 for the most robust normalization performance in the statistical assessment. Then, a tissue mask for the WSI can be obtained. The pixels used in the optimization are randomly sampled from the WSI based on the tissue mask.

3.4. Color normalization

After the Optimization, the adaptive variables $\hat{\mathbf{D}}$, $\hat{\mathbf{W}}$ for ACD are determined. With $\hat{\mathbf{D}}$, the stain components of a WSI can be separated. Next, a reconstruction result can be obtained by recombining the separated stain components with the SCA matrix of a template WSI $\bar{\mathbf{M}}$. Finally, the normalization is completed by converting the reconstruction result from OD space to RGB color space. Specifically, for the i -th pixel \mathbf{x}_i of the WSI, the normalization can be achieved by equations

$$\begin{aligned}\mathbf{o}_i &= -\ln(\mathbf{x}_i/I_{max}), \\ \bar{\mathbf{o}}_i &= \bar{\mathbf{M}} \cdot \hat{\mathbf{W}}\hat{\mathbf{D}} \cdot \mathbf{o}_i, \\ \bar{\mathbf{x}}_i &= \exp(-\bar{\mathbf{o}}_i) \cdot I_{max},\end{aligned}\tag{7}$$

where $\bar{\mathbf{x}}_i$ is the normalized result for \mathbf{x}_i . Because, the matrices $\bar{\mathbf{M}}$, $\hat{\mathbf{W}}$ and $\hat{\mathbf{D}}$ are constant after the ACD, the three matrices can be regarded as a transform matrix. Defining the transform matrix as $\mathbf{T} = \bar{\mathbf{M}}\hat{\mathbf{W}}\hat{\mathbf{D}}$, the normalization in density space can be simplified as $\bar{\mathbf{o}}_i = \mathbf{T} \cdot \mathbf{o}_i$.

4. Experiments and results

4.1. Setup

Four dataset, Camelyon-16, Camelyon-17, Motic-cervix, and Motic-lung, were used in the experiments. Camelyon-16 and Camelyon-17 were obtained

from the Camelyon challenge¹ for cancer metastasis detection in the lymph node [35, 36] and Motic-cervix and Motic-lung were supplied by Motic². The profiles are listed as follows.

- *Camelyon-16* contains 400 H&E-stained lymph node WSIs, in which 270 WSIs are used for training and the remainder are used for testing. Regions with cancer in these WSIs are annotated by pathologists. All the annotations for Camelyon-16 are available.
- *Camelyon-17* contains 1000 WSIs from 5 medical centers, in which 500 WSIs are used for training and the remainder are used for testing. The annotations of testing WSIs are not yet available.
- *Motic-cervix* contains 47 WSIs from 47 patient with cervical cancer (including adenocarcinoma and squamous carcinoma), in which regions with cancer are annotated by pathologists.
- *Motic-lung* contains 39 WSIs from 39 patient with lung cancer (including adenocarcinoma and squamous carcinoma), in which regions with cancer are annotated by pathologists.

The quantitative and qualitative assessments were processed on the Camelyon-17 dataset, since it consists of WSIs from 5 medical centers and contains rich color variations. The Camelyon-16 dataset is used to evaluate the normalization performance for the CAD method, because the labels for both the training and testing set are already published. The experiments were also conducted on Motic-cervix and Motic-lung datasets to evaluate the robustness of the proposed method.

The normalized median intensity (NMI) measure [37] is used to quantitatively assess the consistency of normalization, NMI is defined as

$$NMI(I) = \frac{Med(u_i)}{P_{95}(u_i)}, \quad (8)$$

¹<https://camelyon17.grand-challenge.org/>

²Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen 361101, China

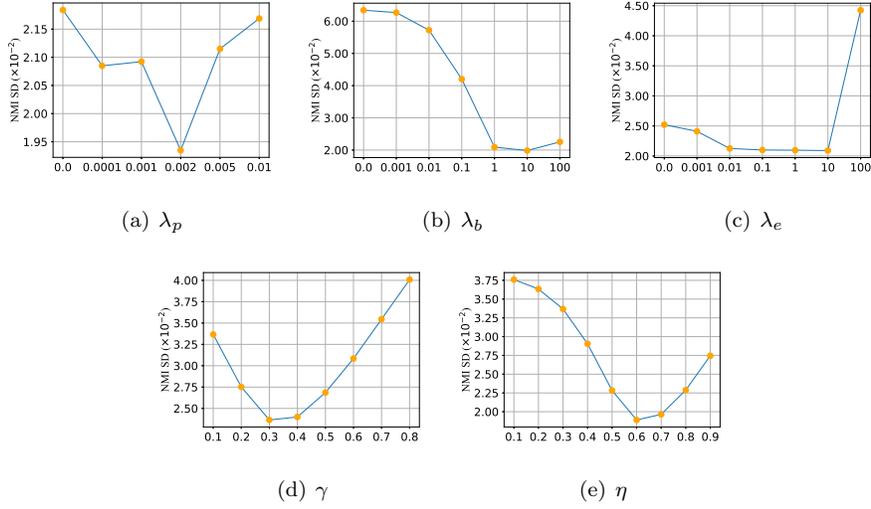


Figure 2: NMI SDs for hyper-parameters of ACD model.

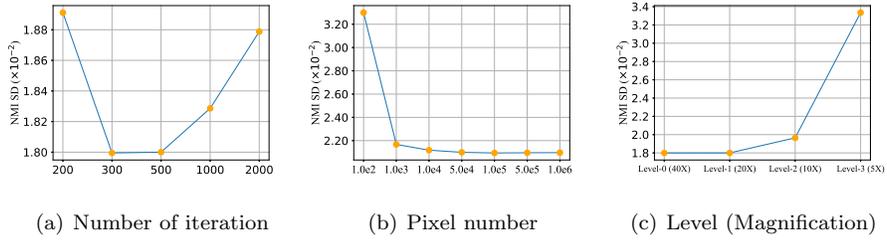


Figure 3: NMI SDs for different training settings of ACD model.

where I denotes a WSI, u_i denotes the mean value of R, G and B channels of the i -th pixel in the WSI. $Med()$ denotes the median value, and $P_{95}()$ denotes the 95th percentile [11]. The standard deviation of the NMI values (NMI SD) and coefficient of the variation (i.e., standard deviation divided by mean) of the NMI values (NMI CV) for all testing WSIs were calculated and used as the metrics. The lower the values of NMI SD and NMI CV, the more consistent the normalization. To avoid the impact of extensive background regions in the WSI and limit the amount of calculation, sub-images were sampled from the tissue region of a WSI to substitute the WSI, and the NMI for the WSI was calculated based on all the pixels in the sub-images. Specifically, 20 sub-

images of 2048×2048 pixels were sampled and the percentage of tissue pixels in each sub-image was controlled at more than 70% (according to the tissue mask defined in section 3.3.3).

The gradient descent algorithm was empirically determined as AdaGrad [38] for its desirable convergency. The model is trained with a mini-batch of 500 samples. The variable φ is initialized based on the SCA matrix suggested in [14] and, \mathbf{W} was initialized as an unitary matrix.

The algorithm was implemented in python with tensorflow [39] and was processed on a computer with an Intel Core i7-7700k CPU of 4.2 GHz and a RAM of 32GB. All the experiments were conducted on the same computer.

In this section, the the hyper-parameters of ACD model are first determined. Then, the normalization performance of the proposed method is evaluated and compared with the state-of-the-art methods.

4.2. Parameters setting

The hyper-parameters of ACD model and the settings in the model solving are validated in this section. The experiments were conducted on the training set of Camelyon-17 dataset. For each WSI, 100,000 pixels were randomly sampled from the tissue area and used to solve the ACD model. The NMI SD for the normalized WSIs under $40\times$ lenses is used as the metric.

4.2.1. Hyper-parameters

There are five hyper-parameters $\lambda_p, \lambda_b, \lambda_s, \gamma, \eta$ involved in our model. These parameters were adjusted in large ranges. The curve of NMI SDs for different setting of the hyper-parameters are presented in Fig. 2. A low NMI SD indicates a good normalization performance. Note that the other parameters were fixed when adjusting a specific parameter.

$\lambda_p, \lambda_b,$ and λ_e are weights of different items involved in the cost function. The results for different settings of the three parameters are shown in Fig. 2 (a-c). Obviously, when the three parameters were set to zero, the performance of normalization deteriorated. It demonstrated that the items defined based

on the prior knowledge (Eq. 3), the balance (Eq. 4), and the density (Eq. 5) of staining are all necessary for a good normalization performance. When a weight was set too large, the normalization performance also decreased. This is because the cost function focused too much on the item controlled by the large weight and ignored the effect of the other items. It indicates that a reliable normalization model cannot be obtained under a single constraint. Finally, the three hyper-parameters were selected for relatively low NMI SDs. Specifically, the three parameters were set as $\lambda_p = 0.002$, $\lambda_b = 10$, and $\lambda_e = 1$ in the following experiments.

In the ACD model, γ constrains the ratio of the staining components, and η constrains the staining density, respectively. The setting of the two parameters influences the visual performance. The normalization performance for the different settings of the two parameters are visualized in Fig. 4. According to the statistical metrics (Fig. 2(d,e)), we suggested $\gamma \in (0.25, 0.4)$ and $\eta \in (0.55, 0.7)$ for a consistent normalization. In the following experiments, γ is set to 0.3 and η is set to 0.6 for a relative low NMI SD.

4.2.2. Optimization settings

The normalization performance is also influenced by the settings of the optimization, including the number of pixels, the magnification of pixel sampling, and the number of interaction in the optimization. The curves of NMI SD for different settings of these factors are given in Fig. 3. It can be seen from Fig. 3 (a) that the normalization is stable when the training step is set between 300 and 500. To limit the calculation amount of optimization, the step is set to 300 in the following experiments. In this experiment, the number of pixels used in the optimization is ranged from 10^2 to 10^6 . Fig. 3 (b) shows that the model trained with 1,000 pixels can achieve a desirable normalization consistency with a NMI SD of 0.0217. This indicates that the proposed model does not rely on massive number of pixels. The metric improves from 0.0217 to 0.0210 when the training pixels are increased from 1,000 to 100,000 and changes little when further increasing the training pixels. Hence, the number of training pixels is

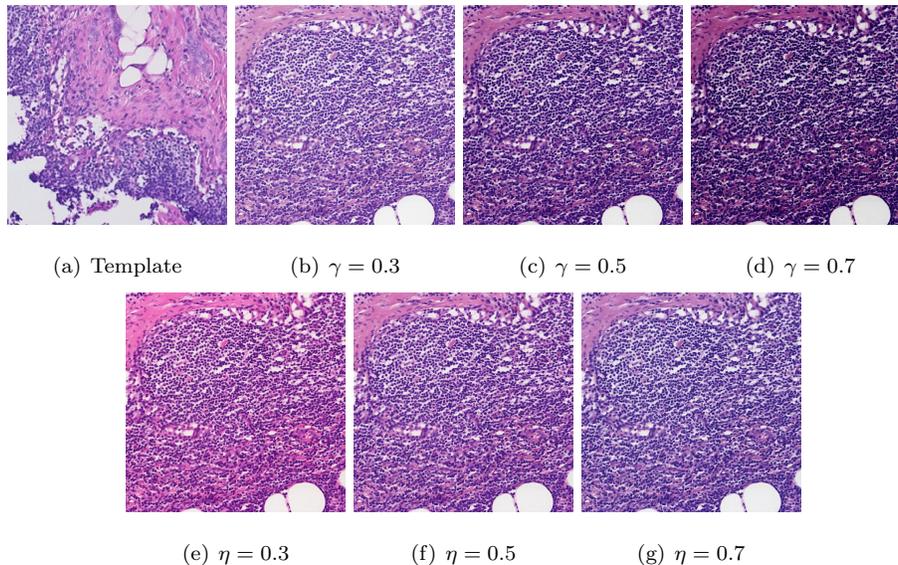


Figure 4: Visual performance of the normalized image varied with the control parameters γ and η , where (a) is a region from the template WSI, (b-d) display the results for different γ , and (e-g) present the results for different η .

set 100,000 in our experiments. The magnification of pixel sampling is also important for the optimization. According to Fig. 3 (c), the proposed model has a certain robustness to decrease in magnification. For reasonable normalization results, pixels used in the optimization are sampled from WSIs under $20\times$ lenses.

4.3. Comparison with the state-of-the-art

4.3.1. Methods for comparison

The color normalization methods developed from different aspects of the histological slides are compared. Specifically, two methods introduced from nature scene image processing proposed by Zheng et al. [8] and Janowczyk et al. [?], the separated-transformation-based methods proposed by Khan et al. [24] and Bejnordi et al. [11], and other two integral-transformation-based methods developed by Vahadane et al.[25] and Zhou et al.[26] are involved in the comparison. The methodologies for these approaches are introduced in the

related works (Section 2).

4.3.2. Quantitative comparison

The testing set of CAMLEYON17 is used in this experiment. The NMI SD and NMI CV calculated based on all testing WSIs are used as metrics. The results of the proposed method are obtained under the hyper-parameters determined in the training set.

The results of the compared methods are presented on Table 2. It is shown that the proposed method achieves the best performance in NMI SD and NMI CV assessment. To intuitively present the allocation of NMI values for the normalized images, the violin plots [40] for different methods are utilized (Fig. 5). The NMI values of the proposed method are the most clustered. It indicates that the normalization performance of the proposed method is consistent.

The stability for staining separation of the compared methods was also evaluated. The normalization results were separated using CD with the parameters of the template WSIs. The NMI metrics for independent staining components are presented in Table 3. Correspondingly, violin plots of NMIs for independent stains are given in Fig. 6. Janowczyk et al. [?] achieves the best NMI SD in the hematoxylin stain component, but the metrics for eosin are inferior to other methods. ACD and Bejnordi et al. [11] achieve an equally good quantitative performance for eosin component. While, the performance of ACD for hematoxylin is better than Bejnordi et al. [11]. Overall, ACD is the most consistent for stain separation among all the compared methods.

4.3.3. Qualitative comparison

Fig. 7 visualizes the normalization performance of the compared methods for five challenging WSIs in Camelyon-17 dataset. The method Zheng et al. [8] is designed to eliminate the variances of illumination and saturation, and has barely standardized the staining color of these images. In contrast, the methods Janowczyk et al. [?], Khan et al. [24], and Bejnordi et al [11] have successfully transformed the color to the template WSI. However, various arti-

Table 2: The comparisons of NMI SD and NMI CV for different normalization methods.

Method	NMI SD	NMI CV
Original	0.140	0.216
Zheng et al. [8]	0.077	0.117
Janowczyk et al.[?]	0.027	0.037
Khan et al. [24]	0.049	0.067
Bejnordi et al. [11]	0.028	0.045
Vahadane et al. [25]	0.042	0.062
Zhou et al. [26]	0.054	0.095
The proposed	0.025	0.034

Table 3: The comparisons of NMI SD and NMI CV for different normalization methods, where NMI_h and NMI_e represent the NMI for hematoxylin and eosin stains, respectively.

Method	NMI_h SD	NMI_h CV	NMI_e SD	NMI_e CV
Original	0.167	0.584	0.146	0.440
Zheng et al. [8]	0.092	0.321	0.081	0.242
Janowczyk et al. [?]	0.017	0.071	0.184	0.392
Khan et al. [24]	0.055	0.203	0.089	0.214
Bejnordi et al. [11]	0.042	0.117	0.028	0.070
Vahadane et al. [25]	0.043	0.109	0.036	0.103
Zhou et al. [26]	0.160	0.362	0.142	0.363
The proposed	0.030	0.068	0.027	0.086

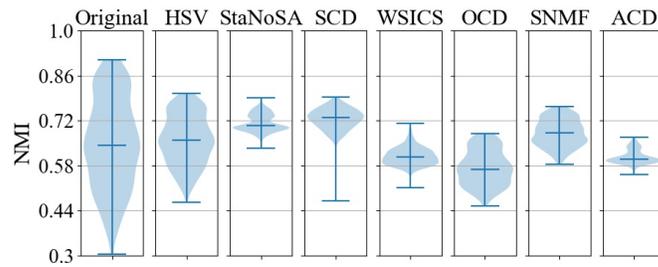


Figure 5: Violin plots of NMIs for the compared methods, where the blue shadow presents the allocation of NMIs for each plot, and the maximum, median, and minimum values for each plot are signed with bars.

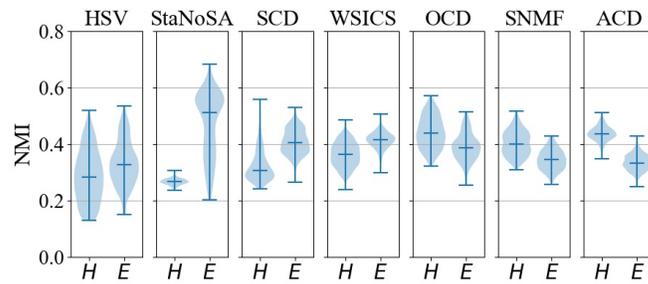


Figure 6: Violin plots of NMIs for independent stains, where H represents the hematoxylin stain, E represents the eosin stain, the blue shadow presents the allocation of NMIs for each plot, and the maximum, median, and minimum values for each plot are signed with bars

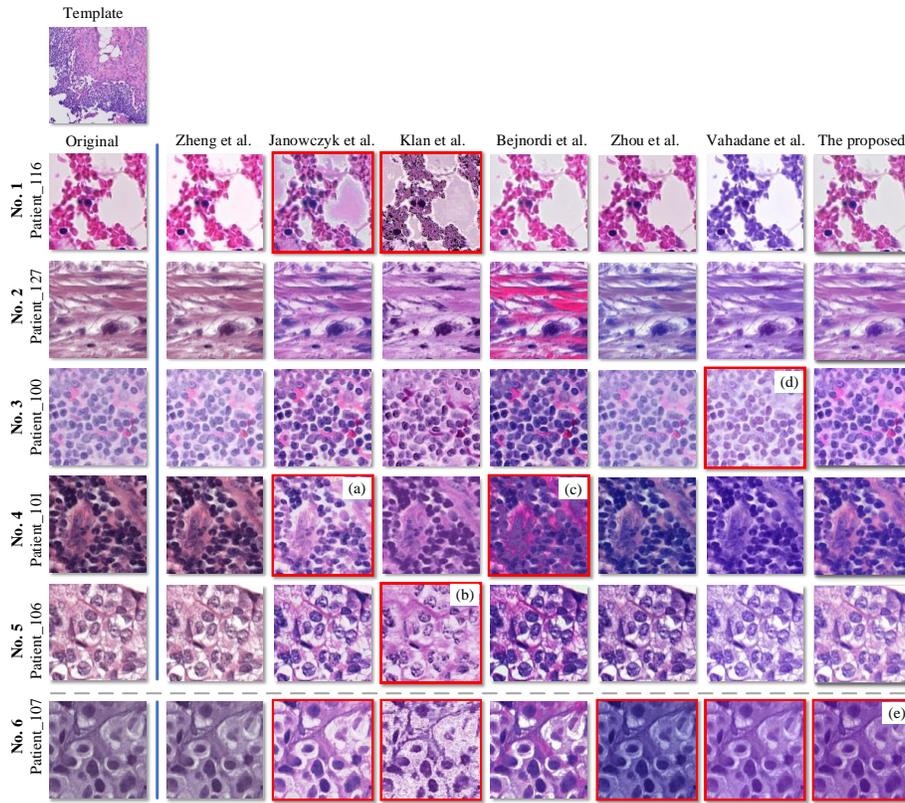


Figure 7: Visual performance of ROIs cropped from five challenging WSIs for the compared methods, in which the original ROIs are represented in the first column, the names of these WSIs are given on the left, the normalization results of the compared methods are displayed on the right, and the results that have serious artifacts are framed with red boxes.

facts appeared in the images normalized by the compared methods. To evaluate the robustness of normalization, three pathologists were invited to inspect the normalized results and the WSIs containing structure or color artifacts were marked. The statistical results of the assessment are given in Table and the results that contain typical artifacts were framed by red boxes in Fig. 7.

Specifically, the eosin stain and the background are occasionally confused in the results of Janowczyk et al. [?]. In Fig. 7(a), A certain amount of eosin stain surrounding the nuclei is eliminated, which may change the structure of cell in the image. For the same WSI, the result obtained by Bejnordi et al.

Table 4: Statistical results in the visual assessment, where the number of WSIs that the three pathologists marked containing structural or color artifacts are presented, and the average number is given at the last column.

Method	<i>Path. #1</i>	<i>Path. #2</i>	<i>Path. #3</i>	Average
Zheng et al. [8]	0.092	0.321	0.081	0.242
Janowczyk et al. [?]]	0.017	0.071	0.184	0.392
Khan et al. [24]	0.055	0.203	0.089	0.214
Bejnordi et al. [11]	0.042	0.117	0.028	0.070
Vahadane et al. [25]	0.043	0.109	0.036	0.103
Zhou et al. [26]	0.160	0.362	0.142	0.363
The proposed	0.030	0.068	0.027	0.086

[11] (Fig. 7(c)) exhibits ringing artifacts around nuclei. It is mainly because that the pixels at the edge of nucleus regions are mis-classified as eosin stain, which causes abrupt changes on the edge of these pixels. The results of Khan et al. [24] also has the similar problem. In Fig. 7(b), the area of nuclei apparently decreases, which will affect the performance of CAD approaches developed on nuclei statistics. Compared to the methods above, the results obtained by integral-transformation-based methods, Zhou et al. [26], Vahadane et al. [25], and ACD, have reserved the structure of tissue in the images, and avoided the artifacts discussed above. This property is significant to develop a reliable CAD approach. On the other hand, these methods may produce color artifacts. For instance, the result in Fig. 7(d) has unreasonable color appearance. **That is because the staining components are not correctly separated by Vahadane et al. [25].**

Overall, the proposed method can obtain both structure-preserving and color-consistent normalization results for WSIs in various appearance. The visual performance is robust and superior to the compared methods.

4.3.4. Time complexity

The time complexity of color normalization methods is equally important in application. Especially for an automatic CAD approach based on WSIs, the

Table 5: The comparison of time complexity and the average running time for the estimation of model parameters.

Method	Time complexity		Running time (second)
	Pre-processing	Model solution	
Zheng et al. [8]	-	$\mathcal{O}(n \log_2 n)$	62.1
Janowczyk et al. [?]]	-	$\mathcal{O}(n)$	93.1
Khan et al. [24]	$\mathcal{O}(n)$		1420
Bejnordi et al. [11]	$\mathcal{O}(n^3)$		370
Vahadane et al. [25]	-		65.4
Zhou et al. [26]	-		7.29
The proposed	-		8.23

running time of the normalization module should not become a bottleneck for the entire system. For the compared methods, the difference in computational complexity mainly derives from the estimation of model parameters. The time complexity for the pixel number n is given in Table 5. The methods proposed by Khan et al. [24] and Bejnordi et al. [11] depend on pixel classification, which is individually considered as pre-processing in Table 5. Khan et al. applied a pixel-wise classifier and the computation is linearly related to the pixel number n . Bejnordi et al. utilized Hough transform to detect nuclei, for which the time complexity is $\mathcal{O}(n^3)$. And in the stage of parameter estimation, the method of Zheng et al. [8] involved sorting algorithm of pixel values, for which the complexity is $\mathcal{O}(n \log_2 n)$. Other methods estimated the parameters based on pixel-wise operation, and the complexity is $\mathcal{O}(n)$.

The average times required for model parameter estimation are given in Table 5, where all the methods are implemented on CPU. Furthermore, the joint results of the quantitative metric and running time of the compared methods are presented in Fig. 8. For the proposed ACD model, the average time to calculate the adaptive parameters is 8.23 s, which is relatively short compared to a WSI-based CAD approach.

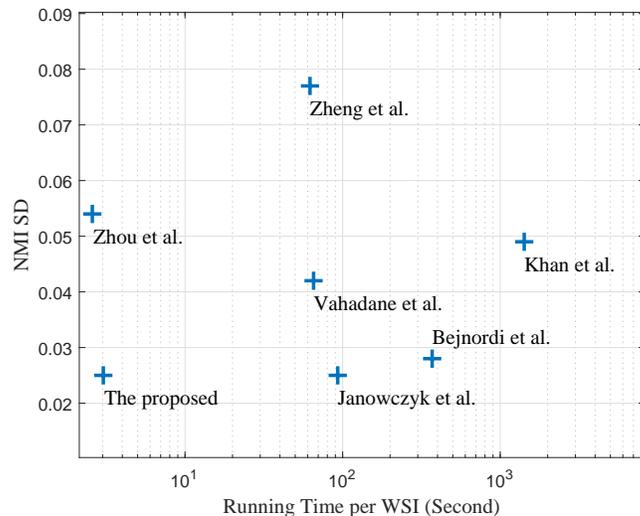


Figure 8: Joint plot of NMI SD and average running time for computing the model parameters for a WSI, where the time is counted in second and presented in logarithmic coordinate.

Janowczyk et al. [?] and Bejnordi et al. [11], which are competitive in the evaluation of NMI SD, respectively take 65 s and 370 s in average. Janowczyk et al. [?] utilized sparse auto-encoders to cluster pixels in different textures, and Bejnordi et al. [11] applied Hough transform to detect the nuclei for pixel classification. These operations extended the running time of the two methods. In contrast, the proposed method estimates the stain parameters using an integrated optimization and involves no additional operation to identify different types of pixels. Therefore, The proposed method is much faster than the two methods. Zhou et al. [26] is simpler than our model, thereby taking less time. However, the stability of the model cannot surpass that of ours. Overall, the proposed method is both effective and efficient for histological image normalization.

4.3.5. Effectiveness in CAD

In recent years, the convolutional neural network (CNN) has become the main algorithm for medical image analysis [41]. An increasing number of stud-

ies have applied CNNs to histological image analysis [42, 35, 43, 44]. The recent studies [9, 10] have shown that the normalization of histological images can improve the performance of CNN-based CAD methods. We conducted experiments to evaluate the improvement of the normalization methods for CNN model. Camelyon-16 dataset was used in this experiment.

The classification of image patches is the major action of CNN in CAD systems for WSI analysis. Hence, a CNN model for histological image patch classification is established as the benchmark of this experiment. Specifically, 110,000 training images in size of 224×224 pixels were sampled from the 270 training WSIs, in which 55,000 patches contained metastases (positive samples) while the others were normal (negative samples). In addition, 65,000 patches were sampled from the testing WSIs to evaluate the performance of classification. A ResNeXt [45] model with 7.2×10^6 parameters was trained on the training patches for each normalization method. The optimizer was momentum SGD with default setting of tensorflow. The receiver operating characteristic (ROC) curves and the corresponding area under curve (AUC) values for the testing set were calculated after 60,000 steps of training. The results are shown in Fig. 9.

Overall, the proposed method achieves the most evident outcome, which increases the AUC from 0.842 (the bench mark) to 0.914. All of the compared methods are effective in improving the accuracy of the cancer image classification. It indicates that the color normalization of histological images is necessary for a good-quality CAD system.

4.4. Application performance on Motic datasets

To further evaluate the application performance of the proposed method on WSIs containing other lesions, we also conducted experiments on Motic-cervix and Motic-lung datasets. In this experiment, the parameters of ACD model were the same as that determined on Camelyon dataset. The metrics for quantitative assessment was calculated before and after the normalization. Furthermore, the performance for improving the classification accuracy of CNN model (cancer

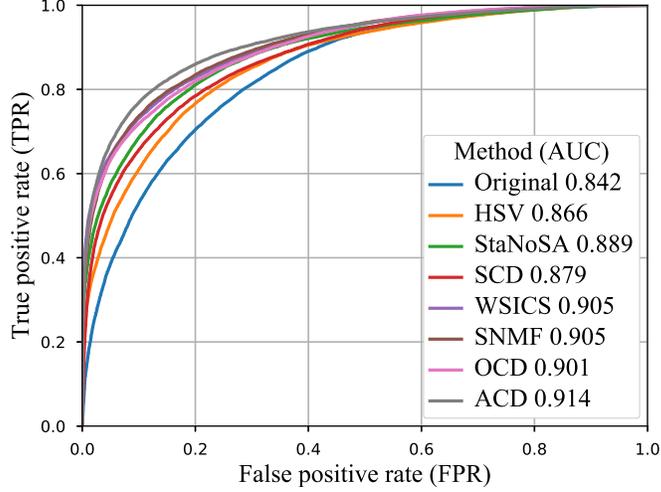


Figure 9: Comparison of ROC curve and AUC in application of breast histological image classification using different normalization methods.

versus non-cancer patches) was also evaluated, where half of the WSIs in each dataset were regarded as training set and the rest for testing. The paradigm of the evaluation is the same as that on Camelyon dataset. The quantitative results are given in Tables 6 and 7. The visual performance of normalization is presented in Figures 10 and 11.

5. Discussion

The ACD model can also be used for color transformation between any two WSIs based on Eq. 7. The only modification is the transform matrix \mathbf{T} . For example, the color transform from WSI p to WSI q can be achieved with a matrix

$$\mathbf{T}_{pq} = (\hat{\mathbf{W}}_q \hat{\mathbf{D}}_q)^{-1} \hat{\mathbf{W}}_p \hat{\mathbf{D}}_p, \quad (9)$$

where $\hat{\mathbf{W}}_p$ and $\hat{\mathbf{D}}_p$ are the adaptive variables for WSI p , and $\hat{\mathbf{W}}_q$ and $\hat{\mathbf{D}}_q$ are the adaptive variables for WSI q . Note that the adaptive variables $\hat{\mathbf{W}}$ and $\hat{\mathbf{D}}$ for a WSI needs to be solved only once after the digitalization of the WSI, and

Table 6: Normalization results on Motic-cervix dataset

Metrics	NMI		NMI _h		NMI _e		Classification
	SD	CV	SD	CV	SD	CV	AUC
Original	0.071	0.087	0.107	0.551	0.188	0.683	0.831
Normalized	0.062	0.077	0.105	0.542	0.138	0.409	0.894

Table 7: Normalization results on Motic-lung dataset

Metrics	NMI		NMI _h		NMI _e		Classification
	SD	CV	SD	CV	SD	CV	AUC
Original	0.112	0.139	0.142	0.711	0.183	0.879	0.886
Normalized	0.089	0.108	0.154	0.834	0.164	0.658	0.911

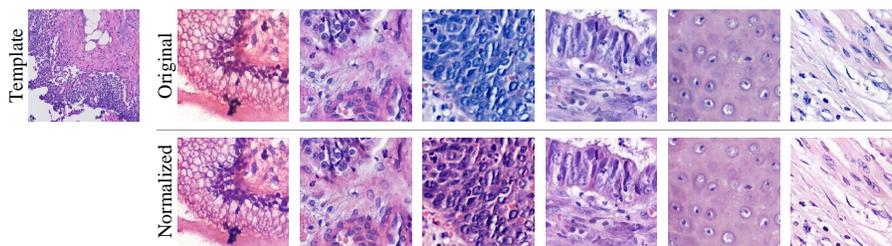


Figure 10: Normalized regions from different WSIs in Motic-cervix dataset.

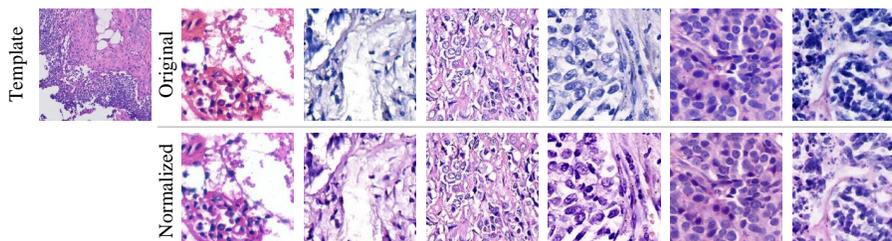


Figure 11: Normalized regions from different WSIs in Motic-lung dataset.

can be stored along with the WSI. When the color transform between any two WSIs is required, the transform matrix can be immediately obtained using Eq. 9. Therefore, it is very convenient to develop online color transformation appli-

cations for digital pathology systems (e.g. MoticGallery³) using the proposed model.

Based on the transform matrix \mathbf{T} , the normalization of the proposed method is a pixel-wise transformation. It can be completed efficiently by parallel computing on CPU or GPU. For further acceleration, a look-up-table (LUT) from the original pixel values to the normalized pixel values can be established for the transformation. Then, the normalization of the WSI can be efficiently achieved through the LUT.

According to the visual assessment of pathologists, the proposed model has the lowest failure rate.

In the evaluation for different number of pixels used in the solving of ACD model, the results show that the thousands of pixels can deliver a consistent performance of normalization (Fig. 2(b)). The main reason is that the pixels used in the optimization are sampled from the tissue area (defined in section 3.3.3), and the staining in Camelyon dataset is relatively homogeneous in the WSI. Hence, thousands of pixels can cover the staining condition of the WSI, which makes the ACD model successfully estimate the stain appearance matrix. for bad quality stain area. For more robust normalization performance, we set the number of pixels to 100,000 in the experiments. We also tuned the number in the experiment on Motic dataset and found that 100,000 pixels were sufficient to the ACD model.

The normalization methods proposed by Khan et al. [24] and Bejnordi et al. [11] are two typical methods based on pixel classification. In these methods, the pixels in the WSI are classified to different classes of stains (including a class of background). Then, pixels belonging to the same class are normalized based on a specific color transform model. Once pixels sharing the similar color in the original images are classified into different classes, the color of these pixels may be quite different after the normalization. It will generate apparent color discontinuity in the normalized WSIs. This type of color discontinuity mainly

³<https://med.motic.com.MoticGallery/>

appears on the boundary of eosin and hematoxylin stains, which may cause structural artifacts of nuclei (see Fig. 7 (b-c)).

The optimization in Zhou et al. [26] is similar with our model. However, Zhou et al. used the deconvolution matrix \mathbf{D} as the variable of the optimization, and objective function did not consider the distinction, the balance, and the overall density of the H and E stains. It makes the optimization mainly adjust the third row of matrix \mathbf{D} , and pay little attention to the first two rows in \mathbf{D} (i.e. the deconvolution parameters for H and E stains). Therefore, the color appearance was not sufficiently transformed to the template WSI.

The model of Vahadane et al. [25] used staining appearance matrix \mathbf{M} as the variable. However, the optimization of Vahadane et al. was solved by alternating between the staining matrix \mathbf{M} and the stain density of pixels (\mathbf{s}_i). Specifically, \mathbf{M} was optimized through a dictionary learning process based on fixed density of training pixels, and then the density of these pixels were estimated through sparse coding process based on fixed \mathbf{M} . The two steps are alternatively processed until the model converges. In contrast, our model is solved through an integrated optimization, in which only the matrix \mathbf{M} and the weighted matrix \mathbf{W} are the variables to be optimized. And the density of training pixels can be directly calculated through matrix product operation (Eq. 2) in each interaction, which is efficient than that in Vahadane et al. [25]. It contributes to a faster solution of our model.

In the experiment of image classification, the results based on Bejnordi et al. [11], Janowczyk et al. [?], Khan et al. [24], Vahadane et al. [25] and Zhou et al. [26] cannot surpass the proposed model. The main reason is that the normalized images of these methods contain various artifacts. Bejnordi et al. [11], Janowczyk et al. [?] and Khan et al. [24] may change the texture or structure of tissue. The results of Vahadane et al. [25] and Zhou et al. [26] have color distortion. These artifacts are detrimental for the CNN-model to distinguish patterns in histological images. In comparison, our method performs more robust than these methods in color normalization, and therefore achieves better performance in the classification of histological images.

6. Conclusion

In this study, we proposed an efficient color normalization method for H&E-stained histological WSIs based on an adaptive color deconvolution (ACD) model. The ACD model does not rely on pixel classification algorithm and is solved through end-to-end learning. The texture and structure of tissue in the images are well preserved. The prior knowledge, the staining density, and the staining ratio of different stains are simultaneously considered in the modeling of ACD, which delivers a statistically consistent performance of normalization. The entire flowchart of normalization involves only pixel-wise operation. The average time of solving the ACD model is less than 10 second for a WSI, which is applicable to automatic CAD systems. The future work will focus on developing automatic WSI analysis methods based on the proposed method. Future work 1) using pixel classification or clustering approach to estimate the ratio of H and E. 2) more general stain separation theory for stains do not satisfy Beer-Lamber law.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61771031, 61371134, 61471016, and 61501009).

Appendix

The derivatives of the objective function (Eq.6) of the ACD model on the variables are given in this section.

For a variable $\theta \in \{\alpha_h, \beta_h, \alpha_e, \beta_e, \alpha_d, \beta_d, w_h, w_e\}$, the partial derivatives of the variable can be calculated based on Eq.6 in the body of the paper as

$$\frac{\partial L}{\partial \theta} = \frac{\partial L_p}{\partial \theta} + \lambda_b \frac{\partial L_p}{\partial \theta} + \lambda_s \frac{\partial L_s}{\partial \theta}, \quad (10)$$

where the items of $\partial L_p/\partial\theta$, $\partial L_b/\partial\theta$, and $\partial L_s/\partial\theta$ are given as follows

$$\begin{aligned}
\frac{\partial L_p}{\partial\theta} &= \frac{\partial}{\partial\theta} \left[\frac{1}{N} \sum_{i=1}^N d_i^2 + \lambda_p \frac{1}{N} \sum_{i=1}^N \frac{2h_i e_i}{h_i^2 + e_i^2} \right] \\
&= \frac{2}{N} \sum_{i=1}^N d_i \frac{\partial d_i}{\partial\theta} \\
&\quad + \frac{2\lambda_p}{N} \sum_{i=1}^N \left[\frac{(e_i \frac{\partial h_i}{\partial\theta} + h_i \frac{\partial e_i}{\partial\theta})(h_i^2 + e_i^2)}{(h_i^2 + e_i^2)^2} \right. \\
&\quad \left. - \frac{h_i e_i (2h_i \frac{\partial h_i}{\partial\theta} + 2e_i \frac{\partial e_i}{\partial\theta})}{(h_i^2 + e_i^2)^2} \right] \\
&= -\frac{4\lambda_p}{N} \sum_{i=1}^N \frac{h_i^2 e_i}{(h_i^2 + e_i^2)^2} \frac{\partial h_i}{\partial\theta} \\
&\quad - \frac{4\lambda_p}{N} \sum_{i=1}^N \frac{h_i e_i^2}{(h_i^2 + e_i^2)^2} \frac{\partial e_i}{\partial\theta} \\
&\quad + \frac{2}{N} \sum_{i=1}^N d_i \frac{\partial d_i}{\partial\theta}
\end{aligned} \tag{11}$$

$$\begin{aligned}
\frac{\partial L_b}{\partial\theta} &= \frac{\partial}{\partial\theta} \left[(1-\eta) \frac{1}{N} \sum_{i=1}^N h_i - \eta \frac{1}{N} \sum_{i=1}^N e_i \right]^2 \\
&= 2\sqrt{L_b} \left[\frac{(1-\eta)}{N} \sum_{i=1}^N \frac{\partial h_i}{\partial\theta} - \frac{\eta}{N} \sum_{i=1}^N \frac{\partial e_i}{\partial\theta} \right] \\
&= 2\sqrt{L_b} \frac{(1-\eta)}{N} \sum_{i=1}^N \frac{\partial h_i}{\partial\theta} - 2\sqrt{L_b} \frac{\eta}{N} \sum_{i=1}^N \frac{\partial e_i}{\partial\theta}
\end{aligned} \tag{12}$$

$$\begin{aligned}
\frac{\partial L_s}{\partial\theta} &= \frac{\partial}{\partial\theta} \left[\gamma - \frac{1}{N} \sum_{i=1}^N (h_i + e_i) \right]^2 \\
&= -2\sqrt{L_s} \frac{1}{N} \sum_{i=1}^N \left(\frac{\partial h_i}{\partial\theta} + \frac{\partial e_i}{\partial\theta} \right) \\
&= -2\sqrt{L_s} \frac{1}{N} \sum_{i=1}^N \frac{\partial h_i}{\partial\theta} - 2\sqrt{L_s} \frac{1}{N} \sum_{i=1}^N \frac{\partial e_i}{\partial\theta}
\end{aligned} \tag{13}$$

From the definition of stains $\mathbf{s}_i = (h_i, e_i, d_i)^\top$, the partial derivatives of h_i, e_i, d_i on θ can be represented by a vector

$$\frac{\partial \mathbf{s}_i}{\partial\theta} = \left(\frac{\partial h_i}{\partial\theta}, \frac{\partial e_i}{\partial\theta}, \frac{\partial d_i}{\partial\theta} \right)^\top.$$

Then, the calculation of $\partial L/\partial\theta$ can be written as

$$\frac{\partial L}{\partial\theta} = \frac{1}{N} \sum_{i=1}^N \mathbf{c}^T \cdot \frac{\partial \mathbf{s}_i}{\partial\theta}, \quad (14)$$

where \mathbf{c} is a vector that consists of coefficients summarized from Equations 10-13:

$$\mathbf{c} = \begin{pmatrix} -4\lambda_p \frac{h_i^2 e_i}{(h_i^2 + e_i^2)^2} + 2(1-\eta)\lambda_b \sqrt{L_b} - 2\lambda_s \sqrt{L_s} \\ -4\lambda_p \frac{h_i e_i^2}{(h_i^2 + e_i^2)^2} - 2\eta\lambda_b \sqrt{L_b} - 2\lambda_s \sqrt{L_s} \\ 2d_i \end{pmatrix}$$

Next, the calculation of $\partial \mathbf{s}_i/\partial\theta$ is presented. From Eq.2 in the body of the paper, it is

$$\frac{\partial \mathbf{s}_i}{\partial\theta} = \frac{\partial}{\partial\theta} (\mathbf{W}\mathbf{D}\mathbf{o}_i).$$

Here, \mathbf{W} consists of the weighting variables w_h and w_e , and \mathbf{D} is a function of degree variables φ . Thus, the partial derivatives of \mathbf{s}_i on $\theta \in \{w_h, w_e\}$ and $\theta \in \varphi$ are discussed separately.

The partial derivatives of \mathbf{s}_i on w_h and w_e are

$$\frac{\partial \mathbf{s}_i}{\partial w_h} = \text{diag}(1, 0, 0)\mathbf{D}\mathbf{o}_i, \quad \frac{\partial \mathbf{s}_i}{\partial w_e} = \text{diag}(0, 1, 0)\mathbf{D}\mathbf{o}_i. \quad (15)$$

And, for variables $\theta \in \varphi$,

$$\begin{aligned} \frac{\partial \mathbf{s}_i}{\partial\theta} &= \frac{\partial}{\partial\theta} (\mathbf{W}\mathbf{D}\mathbf{o}_i) = \mathbf{W} \frac{\partial \mathbf{D}}{\partial\theta} \mathbf{o}_i \\ &= \mathbf{W} \frac{\partial \mathbf{M}^{-1}}{\partial\theta} \mathbf{o}_i \\ &= \mathbf{W}\mathbf{M}^{-1} \frac{\partial \mathbf{M}}{\partial\theta} \mathbf{M}^{-1} \mathbf{o}_i \\ &= \mathbf{W}\mathbf{D} \frac{\partial \mathbf{M}}{\partial\theta} \mathbf{D}\mathbf{o}_i. \end{aligned}$$

Specifically, the derivatives of \mathbf{M} on each degree variables are

$$\begin{aligned} \frac{\partial \mathbf{M}}{\partial \alpha_h} &= \left(\frac{\partial \mathbf{m}_h}{\partial \alpha_h}, \mathbf{0}, \mathbf{0} \right), & \frac{\partial \mathbf{M}}{\partial \beta_h} &= \left(\frac{\partial \mathbf{m}_h}{\partial \beta_h}, \mathbf{0}, \mathbf{0} \right), \\ \frac{\partial \mathbf{M}}{\partial \alpha_e} &= \left(\mathbf{0}, \frac{\partial \mathbf{m}_e}{\partial \alpha_e}, \mathbf{0} \right), & \frac{\partial \mathbf{M}}{\partial \beta_e} &= \left(\mathbf{0}, \frac{\partial \mathbf{m}_e}{\partial \beta_e}, \mathbf{0} \right), \\ \frac{\partial \mathbf{M}}{\partial \alpha_d} &= \left(\mathbf{0}, \mathbf{0}, \frac{\partial \mathbf{m}_d}{\partial \alpha_d} \right), & \frac{\partial \mathbf{M}}{\partial \beta_d} &= \left(\mathbf{0}, \mathbf{0}, \frac{\partial \mathbf{m}_d}{\partial \beta_d} \right), \end{aligned} \quad (16)$$

and

$$\begin{aligned}\frac{\partial \mathbf{m}_j}{\partial \alpha_j} &= (-\sin \alpha_j \sin \beta_j, -\sin \alpha_j \cos \beta_j, \cos \alpha_j)^T, \\ \frac{\partial \mathbf{m}_j}{\partial \beta_j} &= (\cos \alpha_j \cos \beta_j, -\cos \alpha_j \sin \beta_j, 0), \\ j &= h, e, d\end{aligned}$$

References

- [1] S. L. Robbins, V. Kumar, A. K. Abbas, N. Fausto, J. C. Aster, Robbins and Cotran pathologic basis of disease, Saunders/Elsevier, 2010.
- [2] F. Ghaznavi, A. Evans, A. Madabhushi, M. Feldman, Digital imaging in pathology: Whole-slide imaging and beyond, Annual Review of Pathology-mechanisms of Disease 8 (1) (2013) 331–359.
- [3] J. S. Duncan, N. Ayache, Medical image analysis: Progress over two decades and the challenges ahead, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1) (2000) 85–106. doi:10.1109/34.824822.
- [4] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, B. Yener, Histopathological image analysis: a review, IEEE Reviews in Biomedical Engineering 2 (2009) 147–171.
- [5] S. Zhang, D. Metaxas, Large-scale medical image analytics: Recent methodologies, applications and future directions, Medical Image Analysis 33 (2016) 98–101.
- [6] Z. Li, X. Zhang, H. Müller, S. Zhang, Large-scale retrieval for medical image analytics: A comprehensive review., Medical Image Analysis 43 (2018) 66–84.
- [7] B. E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, J. A. van der Laak, Automated detection of dcis in whole-slide h&e stained breast histopathology images, IEEE Transactions on Medical Imaging 35 (9) (2016) 2141–2150. doi:10.1109/tmi.2016.2550620.

- [8] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, Y. Zhao, Histopathological whole slide image analysis using context-based cbir, *IEEE Transactions on Medical Imaging* 37 (7) (2018) 1641–1652. doi:10.1109/TMI.2018.2796130.
- [9] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. De Souza, A. Baidoshvili, G. J. S. Litjens, B. Van Ginneken, I. D. Nagtegaal, J. A. W. M. V. Der Laak, The importance of stain normalization in colorectal tissue classification with convolutional networks, in: *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2017, pp. 160–163.
- [10] B. E. Bejnordi, J. Lin, B. Glass, M. Mullooly, G. L. Gierach, M. E. Sherman, N. Karssemeijer, J. V. D. Laak, A. H. Beck, Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images, in: *IEEE International Symposium on Biomedical Imaging*, 2017, pp. 929–932.
- [11] B. E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Höller, A. Homeyer, N. Karssemeijer, J. A. V. D. Laak, Stain specific standardization of whole-slide histopathological images, *IEEE Transactions on Medical Imaging* 35 (2) (2016) 404–415.
- [12] D. Onder, S. Zengin, S. Sarioglu, A review on color normalization and color deconvolution methods in histopathology, *Applied Immunohistochemistry & Molecular Morphology* 22 (10) (2014) 713–719.
- [13] Y. Y. Wang, S. C. Chang, L. W. Wu, S. T. Tsai, Y. N. Sun, A color-based approach for automated segmentation in tumor tissue classification., in: *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, p. 6577.
- [14] A. C. Ruifrok, D. A. Johnston, Quantification of histochemical staining by color deconvolution, *Analytical and quantitative cytology and histology* 23 (4) (2001) 291–299.

- [15] Y. Ma, S. Jun, Z. Jiang, F. Hao, Plsa-based pathological image retrieval for breast cancer with color deconvolution, *Proceedings of SPIE Medical Imaging* 8920 (8) (2013) 89200L–89200L–7.
- [16] Y. Zheng, Z. Jiang, J. Shi, Y. Ma, Retrieval of pathology image for breast cancer using plsa model based on texture and pathological features, in: *IEEE International Conference on Image Processing*, 2014, pp. 2304–2308.
- [17] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, Y. Zhao, Breast histopathological image retrieval based on latent dirichlet allocation, *IEEE Journal of Biomedical and Health Informatics* 21 (4) (2017) 1114–1123. doi:10.1109/jbhi.2016.2611615.
- [18] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, Y. Zhao, Size-scalable content-based histopathological image retrieval from database that consists of wsis, *IEEE journal of biomedical and health informatics* 22 (4) (2018) 1278–1287. doi:10.1109/jbhi.2017.2723014.
- [19] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, Y. Zhao, Proposing regions from histopathological whole slide image for retrieval using selective search, in: *IEEE International Symposium of Biomedical imaging*, 2017.
- [20] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, Y. Zhao, J. Shi, Generating region proposals for histopathological whole slide image retrieval, *Computer Methods and Programs in Biomedicine* 159 (2018) 1–10. doi:{10.1016/j.cmpb.2018.02.020}.
- [21] A. Rabinovich, S. Agarwal, C. A. Laris, J. H. Price, S. J. Belongie, Un-supervised color decomposition of histologically stained tissue samples, in: *Advances in Neural Information Processing Systems*, 2003, pp. 667–674.
- [22] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, N. E. Thomas, A method for normalizing histology slides for quantitative analysis, in: *IEEE International Conference on*

- Symposium on Biomedical Imaging: From Nano To Macro, 2009, pp. 1107–1110.
- [23] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, P. Quirke, Colour normalisation in digital histopathology images, in: Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop), 2009, pp. 100–111.
- [24] A. M. Khan, N. M. Rajpoot, D. Treanor, D. R. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, *IEEE Transactions on Biomedical Engineering* 61 (6) (2014) 1729–1738.
- [25] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, N. Navab, Structure-preserving color normalization and sparse stain separation for histological images, *IEEE Transactions on Medical Imaging* 35 (8) (2016) 1962–1971.
- [26] N. Zhou, Y. Gao, Optimized color decomposition of localized whole slide images and convolutional neural network for intermediate prostate cancer classification, in: *Proceeding of SPIE Medical Imaging*, Vol. 10140, 2017, p. 101400W.
- [27] L. Sha, D. Schonfeld, Sethi, Color normalization of histology slides using graph regularized sparse nmf, in: *Proceeding of SPIE Medical Imaging*, Vol. 10140, 2017, p. 1014010.
- [28] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N. N. Shih, H. Gilmore, A. Madabhushi, Sparse non-negative matrix factorization (snmf) based color unmixing for breast histopathological image analysis., *Computerized Medical Imaging and Graphics* 46 Part 1 (2015) 20–29.
- [29] N. Hidalgotavira, J. Mateos, M. Vega, R. Molina, A. K. Katsaggelos, Fully automated blind color deconvolution of histopathological images., in:

Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2018, pp. 183–191.

- [30] E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images, *IEEE Computer Graphics and Applications* 21 (5) (2001) 34–41.
- [31] J. Illingworth, J. Kittler, A survey of the hough transform, *Graphical Models and Image Processing* 44 (1) (1988) 87–116.
- [32] X. Li, 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) (2015) 1862–1873. doi:10.1109/TBME.2015.2405791.
- [33] X. Li, 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) (2017) 150–161. doi:10.1109/JBHI.2015.2503720.
- [34] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, P. H. N. de With, Stain normalization of histopathology images using generative adversarial networks, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 573–577. doi:10.1109/ISBI.2018.8363641.
- [35] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. J. S. Litjens, J. A. W. M. V. Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, *JAMA* 318 (22) (2017) 2199–2210.
- [36] G. J. S. Litjens, P. Bandi, B. E. Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. J. M. V. De Loo, R. Vogels, et al.,

- 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset, *GigaScience* 7 (6) (2018).
- [37] A. Basavanthally, A. Madabhushi, Em-based segmentation-driven color standardization of digitized histopathology, in: *Proceeding of SPIE Medical Imaging*, Vol. 8676, 2013, p. 86760G.
- [38] J. C. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *Journal of Machine Learning Research* 12 (2011) 2121–2159.
- [39] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, *arXiv abs/1603.04467* (2016).
- [40] J. L. Hintze, R. D. Nelson, Violin plots: A box plot-density trace synergism, *The American Statistician* 52 (2) (1998) 181–184.
- [41] G. Litjens, T. Kooi, B. E. Bejnordi, S. Aaa, F. Ciompi, M. Ghafoorian, V. D. L. Jawm, G. B. Van, C. I. Sánchez, A survey on deep learning in medical image analysis, *Medical Image Analysis* 42 (9) (2017) 60–88.
- [42] Y. Xu, Z. Jia, L. B. Wang, Y. Ai, F. Zhang, M. Lai, E. I. Chang, Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features, *Bmc Bioinformatics* 18 (1) (2017) 281.
- [43] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, Y. Zhao, Feature extraction from histopathological images based on nucleus-guided convolu-

tional neural network for breast lesion classification, *Pattern Recognition* 71 (2017) 14–25. doi:10.1016/j.patcog.2017.05.010.

[44] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, *Neurocomputing* 191 (2016) 214–223. doi:10.1016/j.neucom.2016.01.034.

[45] S. Xie, R. B. Girshick, P. Dollar, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.