Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images

Abhishek Vahadane*, Tingying Peng*, Amit Sethi, Shadi Albarqouni, Lichao Wang, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Irene Esposito, and Nassir Navab

***Abstract*—Staining and scanning of tissue samples for microscopic examination is fraught with undesirable color variations arising from differences in raw materials and manufacturing techniques of stain vendors, staining protocols of labs, and color responses of digital scanners. When comparing tissue samples, color normalization and stain separation of the tissue images can be helpful for both pathologists and software. Techniques that are used for natural images fail to utilize structural properties of stained tissue samples and produce undesirable color distortions. The stain concentration cannot be negative. Tissue samples are stained with only a few stains and most tissue regions are characterized by at most one effective stain. We model these physical phenomena that define the tissue structure by first decomposing images in an unsupervised manner into stain density maps that are sparse and non-negative. For a given image, we combine its stain density maps with stain color basis of a pathologist-preferred target image, thus altering only its color while preserving its structure described by the maps. Stain density correlation with ground truth and preference by pathologists were higher for images normalized using our method when compared to other alternatives. We also propose a computationally faster extension of this technique for large whole-slide images that selects an appropriate patch sample instead of using the entire image to compute the stain color basis.**

***Index Terms*—Color normalization, histopathological images, non-negative matrix factorization, sparse regularization, unsupervised stain separation.**

Manuscript received October 31, 2015; revised January 29, 2016; accepted February 05, 2016. Date of publication April 27, 2016; date of current version July 29, 2016. A. Vahadane, T. Peng and A. Sethi acknowledge the support of Heritage (Erasmus Mundus), Humboldt and IUSSTF fellowships, respectively. The work of M. Baust, I. Esposito, A. M. Schlitter, K. Steiger, and N. Navab was supported by the Collaborative Research Centre SFB 824 (Z2) "Imaging for Selection, Monitoring and Individualization of Cancer Therapies" funded by the German Science Foundation . *A. Vahadane and T. Peng contributed equally to this work and both are joint first and corresponding authors.

*A. Vahadane is with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India (e-mail: abhishek.vahadane@gmail.com).

*T. Peng, S. Albarqouni, L. Wang, and M. Baust are with the Computer Aided Medical Procedures and Augmented Reality (CAMP), Technical University of Munich (TUM), 85748 Munich, Germany (e-mail: tingying.peng@tum.de).

A. Sethi is with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India.

K. Steiger and A. M. Schlitter are with the Department of Pathology, Technical University of Munich (TUM), 81675 Munich, Germany.

I. Esposito is with Department of Pathology, Heinrich-Heine University Dusseldorf, 40225 Dusseldorf, Germany.

N. Navab is with Computer Aided Medical Procedures and Augmented Reality (CAMP), Technical University of Munich (TUM), 85748 Munich, Germany, and also with CAMP, Johns Hopkins University, Baltimore, MD 21218 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2016.2529665

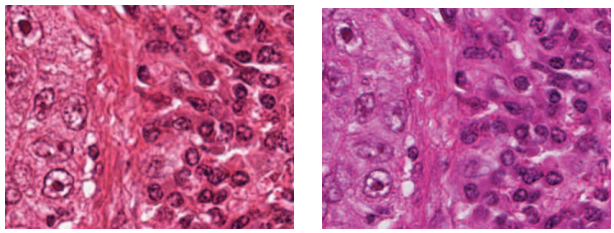
I. Introduction

In histopathology, microscopic images of tissue samples are examined to study the manifestations of disease. Reagents that bind to specific proteins, called stains, are used to increase the contrast between different structures for their manual examination. By far the most commonly used staining scheme is a combination of hematoxylin and eosin (H&E). Hematoxylin is a bluish-purple basophilic dye that binds strongly to the nuclei while eosin is a red-pink stain that binds primarily to acidophilic proteins in the cytoplasm and the stromal matrix.

Although manual examination is still the most common approach in clinical pathology, digitalization of stained tissue biopsies and microarrays (TMA) is increasingly gaining ground so as to allow easy sharing and use of computer algorithms to analyze these images [1]. Moreover, in large-scale preclinical research studies, millions of images can be generated using modern high-throughput scanning techniques (for example, the Human Protein Atlas project [2]), which is far beyond the capacity of manual analysis and has to rely on computational quantification. Many computational pathology algorithms are based on machine learning, which rely on training data sets that capture a wide range of variations to distinguish between intra-class and inter-class differences.

One common problem in analysis of tissue samples is undesirable variation in color due to differences in color responses of slide scanners (as shown in Fig. 1), raw materials and manufacturing techniques of stain vendors, and staining protocols across different pathology labs. This creates difficulty in image interpretation by software trained on a particular stain appearance [3], [4], and further adds to already existing inter- and intra-expert variance in diagnosis and labeling among pathologists [5].

Color normalization that standardizes image appearance can help both pathologists and software in comparing different tissue samples. However, many existing techniques for color normalization such as histogram specification [6] has been proposed for natural images, but not for histology. Histogram normalization works best when the source image whose histogram is to be modified captures the exact same scene as the target image, except with a color change. Otherwise, the distribution of various materials and their associated true colors can be quite different between the scenes. This is often the case in histological images even for a given staining scheme, where a source image captures a different tissue slide than a target image whose appearance is preferred by a pathologist or for an algorithm. Fortunately, only a few stains are used in any given



(a) (b)

Fig. 1. The same tissue section taken by different microscopic slide scanners. (a) Aperio scanner. (b) Hamamatsu scanner. Image source: ICPR 2014 MITOS-ATYPIA challenge—http://mitos-atypia-14.grand-challenge.org

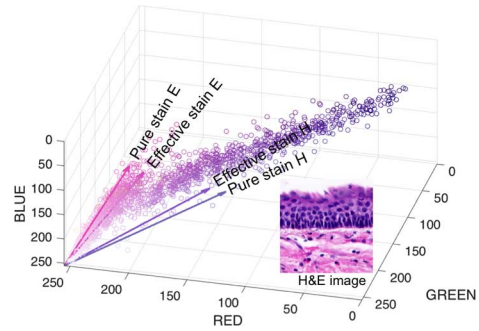


Fig. 2. Effective stain graphical representation: cell nucleus stained with effective H, which is a mixture of a large amount of pure H and a small amount of pure E; by contrast, connective tissue stained with effective E, which is a mixture of a large amount of pure E and a small amount of pure H. H&E image copyright of G.Landini, University of Birmingham, U.K., reproduced with permission. Pure stain vectors are "built-in" vectors determined experimentally (details refer to http://www.mecourse.com/landinig/software/cdeconv/cdeconv.html). Effective stain vectors are determined by our SNMF decomposition.

staining scheme. If the stain density maps can be estimated in the two images, then the stain color appearance in the source image can also be estimated and modified to match those of a preferred target image while preserving its stain density.

Thus, estimation of stain density maps and color appearances i.e., stain separation is central to many color normalization techniques including the proposed techniques presented here, which transforms the RGB channels of a histological image into an absorption or density map for each stain. This can also help in computational pathology, for example in quantification of cancer-related structure information, such as the count of certain nuclei and shape of glands [7], [8], and in custom visualization of the structure information using false colors. In fact, stain separation and color normalization are two sides of the same coin: an accurate stain separation requires a precise estimation of stain color appearance matrix (or color basis) whilst one key step of the structure-preserving color normalization (SPCN) proposed in this paper is to obtain an accurate stain density map. Other recently proposed color normalization techniques have also been based on stain separation such as [9], [10], [3], and have shown better results than previous methods. Normalizing the appearance of each stain separately is intuitive because it takes advantage of the known properties of stained tissue images.

We propose a solution for both stain separation and color normalization, which preserves biological structure information by modelling stain density maps based on the following properties:

- *Non-negativity:* It is essential to assume that a stain density and its associated optical density can either be absent (zero) or present (positive) at a given location (pixel), but it cannot be negative. Negative stain density would mean emitting light.
- *Sparsity:* We assume that the tissue behind most pixels represent one of a few types of biological material (such as nuclei or cytoplasm), characterized by their effective stain, allowing each pixel to be modeled as a sparse mixture of the constituent effective stains. Effective stain is illustrated in Fig. 2. A biological structure, such as a nucleus, can bind to more than one chemical stain, such as in case of H&E staining, but we assume that their relative proportions are fixed for a given biological structure within one image. We term this mixture of stains that is uniquely present in a type of biological structure as an *effective stain*. Thus, we estimate the effective stain (henceforth stain) basis vectors and not the pure stain basis vectors.
- *Soft-classification:* We can allow a small proportion of pixels on the boundaries (edges) of two biological materials to capture more than one stain, which also helps smoothen the optimization cost functions.

We do this by first casting the stain separation problem as a non-negative matrix factorization (NMF) to which we add a sparseness constraint, and refer to it as *sparse non-negative matrix factorization* (SNMF). An additional advantage of this formulation is that the color basis is determined in an unsupervised manner without requiring manual annotation of different areas of pure stains. Our color normalization is built on top of SNMF-based stain separation, and is called *structure-preserving color normalization* (SPCN). SPCN works by replacing the color basis of a source image with those of a pathologist-preferred target image, while still maintaining its original stain concentrations. The flexibility to select a preferred target appearance in different scenarios as opposed to a fixed target color appearance model is another advantage of our technique over others such as [9]. Information about most of the biological structure is captured in the stain concentration and hence it is retained in the proposed technique. Lastly, we propose a fast color prototype estimation scheme for large whole-slide images (WSI) that uses an adaptive patch sample instead of the entire image.

We have compared our algorithms quantitatively and qualitatively with competing techniques that are also based on stain separation [9], [11], [12]. For quantitative comparison, we generated ground truth by having pathologists annotate regions of different pure stains, and using that to compute median color basis and density maps for the stains. We validated the stain separation methods by comparing estimated color basis and stain density maps with the ground truth. We also compared the pathologists subjective grading of the NMF and SNMF. SNMF outperformed the NMF significantly, thus validating the role of sparsity in stain density estimation. In addition, we found that SNMF was robust to variations of a hyperparameter in the defined range that controls sparseness. We also compared SPCN that is based on SNMF against other color normalization techniques quantitatively and qualitatively and these results

were also favorable to our technique. We also report a 20-fold speedup on WSI using our proposed patch-based color appearance estimation scheme.

This paper extends our earlier work [13] with more thorough quantitative evaluation as well as proposal and validation of the acceleration scheme for WSI. Rest of this paper is organized as follows. In Section II we discuss background and related work behind our proposed methods. In Section III we explain our proposed algorithms in more detail. In Section IV we show results from extensive validation of our algorithms. We end with discussion and conclusion in Section V.

II. Background and Related Work

In this Section, we survey stain separation and color normalization methods, and how successively proposed techniques have moved closer to capturing the underlying biological structure of tissues, setting the stage for our proposed technique.

A. Color Normalization Without Stain Separation

An important way to overcome unwanted color variation in images of similar entities is to transform the color appearance of a source image into that of a target image preferred by an expert. Histogram specification in RGB space [6] and matching of histogram statistics (i.e., mean and standard deviation) have been used to normalize color appearance in histology after transformation of an RGB image into $l\alpha\beta$ (decorrelated) color space [14]. Both techniques inherently assume that the proportions of tissue compartments primarily stained by a particular reagent are the same across the images being normalized, which is not a valid assumption as proportion of nuclei or whitespace varies from image to image. Color appearance of slides stained to serve as calibrated samples has also been used to model the color space of an imaging system [15], [16]. Images of tissues to be examined are normalized to match the color appearance of the calibrated targets. This approach can only incorporate color variations due to different scanners but not those due to differences in staining reagents and procedures. Methods based on stain separation overcome these challenges to a large extent.

B. Soft Stain Separation Methods

Since stain separation is estimation of the density map of each stain, it is instructive to understand the relation between RGB colors and stain density at each pixel. Stained tissue attenuates light in a certain spectrum depending on the type and amount of stain it has absorbed. This relationship is captured in Beer-Lambert law as follows. Let $I \in R^{m\times n}$ be the matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images). Let $W \in R^{m\times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains, and $H \in R^{r\times n}$ be the stain density maps, where the rows of which represent the concentration of each stain. Then, I can be written as follows [11]:

$$I = I_0 \exp(-WH). \tag{1}$$

Let V be the relative optical density then,

$$V = \log \frac{I_0}{I}. \tag{2}$$

Hence, (1) becomes,

$$V = WH. \tag{3}$$

So, given an observation matrix V, the goal is to find stain color appearance matrix W and stain density map matrix H.

Color deconvolution [17] is one of the most widely used methods to separate light absorbing stains using the relationship between stain amount and light absorption given by Beer-Lambert Law [18]. The authors suggested measuring the stain color appearance from control slides with a single stain per slide, that is empirically estimating W (an example shown in Fig. 2). Stain density maps H can then be computed by multiplying the optical density with Moore-Penrose pseudoinverse of color appearance matrix W. This calibration-based method has a strong limitation that it is only valid for images using the same staining and imaging protocol whereas color variation in histology is almost inevitable as described in Section I.

Non-negative matrix factorization (NMF) [19] was used in one of the pioneering unsupervised frameworks for stain separation proposed in [12] that derived image-specific stain color basis. The non-negative constraints on the stain density and color appearance matrix capture the important property of stains that can only absorb but not emit light, which makes both their color basis and density non-negative. It, therefore, solves the following problem:

$$\min_{W,H} \ \frac{1}{2}\|V - WH\|_F^2, \ \text{such that} \ W, H \geq 0. \tag{4}$$

Estimation of both factors jointly is a non-convex optimization problem which may converge to a local optimum instead of the global optimum and give undesirable stain vectors [20].

Besides NMF [12], other popular unsupervised decomposition methods are based on independent component analysis (ICA) [21], singular value decomposition (SVD) [9], and blind color decomposition (BCD) [11]. As per the relative root mean square error (rRMSE) between estimated color basis and ground truth in [11], BCD outperforms NMF by 20% to 40% and ICA by 3 to 5 folds. Though BCD has a good performance, it is unclear how to fix some of its hyper-parameters. For example, the authors proposed to use a piece-wise linear decomposition instead of a linear one in case of a poor stain-cluster separation but they have not provided a quantitative threshold for fisher criteria [22] to determine 'poor' clusters. SVD [9] has a closed-form solution which can be computed efficiently, although its performance deteriorates when images contain uneven proportions of each stain [23].

Addition of a sparsity constraint to NMF, as we explain in Section III-A, reduces the solution space as well as adds another biologically-concordant principle to the model. A sparse formulation for stain separation was recently reported in [24]. However, it is only an extension of color deconvolution method [17] and still requires W to be experimentally determined, which limits its applicability over large histological image dataset with considerable color variation.

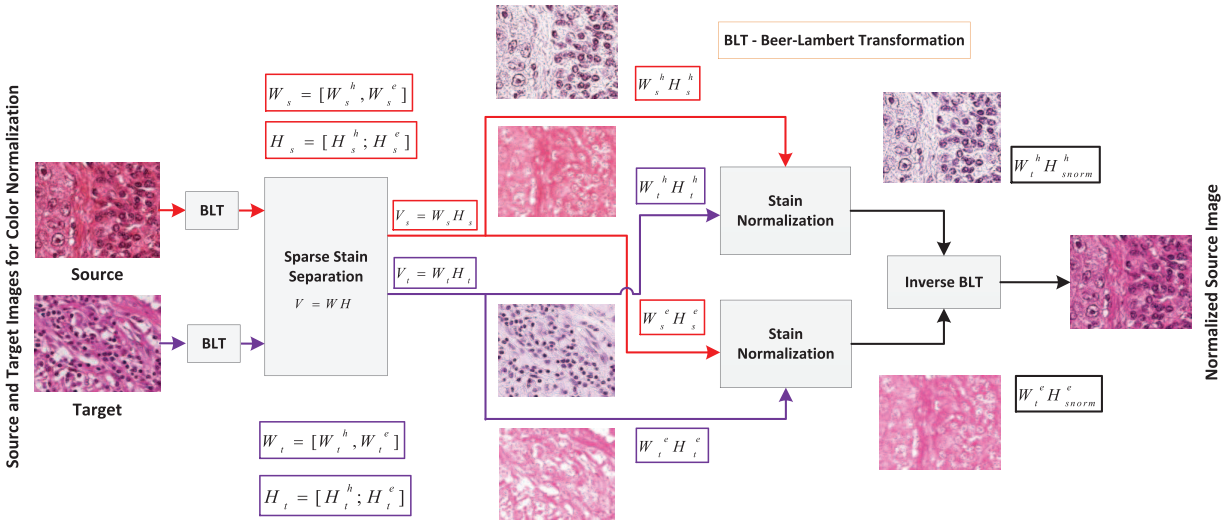


Fig. 3. Flow diagram of the proposed structured preserving color normalization (SPCN).

In this paper, we have compared the proposed SNMF-based stain separation with NMF-based [12] and SVD-based [9] methods. ICA [21] was dropped from comparison because its performance was reported to be far worse than that of NMF and BCD in [11]. BCD was dropped from comparison because neither all of the details required for its implementation nor its source code were made available in [11]. However, we indirectly compared it with SNMF based on their results relative to NMF as reported in [11].

C. Color Normalization Methods Based on Stain Separation

Supervised color normalization techniques such as [3] and [10] are limited to model the color variation of only the training dataset. Retraining could be required for different tissue types, staining schemes, and even same type of tissue from a different lab. Additionally, a useful application for the color normalization is to improve the color appearance or contrast of low-quality and faded histology slides. It would be very challenging for a training dataset to cover histology slides from all possible scenarios.

Among unsupervised methods, [9] used SVD to extract stain vectors followed by direction correction making it robust to inter-image variations. However, in the normalization process, authors used a built-in color appearance model to modify the color distribution of a given image, which is less flexible than having the ability to select a target image with color appearance preferred by an expert. Moreover, modeling stains as principal components does not guarantee their non-negativity or sparsity. Thus, the resultant component maps can be difficult to interpret biologically.

III. Methods

The proposed stain separation method based on SNMF is integral to the structure-preserving color normalization (SPCN) algorithm, as shown in Fig. 3. At its heart is addition of a sparsity constraint on NMF to capture the biological principle of discreteness of biological structures. That is, each biological structure, such as a nucleus, has finite and connected spatial extent, which is characterized by an absorption spectrum or effective stain. In this Section, we provide details of implementation of SNMF, SPCN, and a patch-based scheme to accelerate color basis estimation. The source codes of SNMF, SPCN and acceleration scheme are available for the academic use at https://github.com/abhishekvahadane/CodeRelease_ColorNormalization.git.

A. SNMF-Based Stain Separation

We first convert a given RGB image to optical density using (2) based on Beer-Lambert law. Then, we add a sparseness constraint to equation (4), i.e., we propose an improved NMF cost function for stain separation by including $l1$ sparseness regularization on stain mixing coefficients H_j, for stains indexed by $j = 1, 2, \ldots, r$, which yields,

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \quad \|W(:,j)\|_2^2 = 1 \quad (5)$$

where λ is sparsity and regularization parameter. Note that the additional constraint on W in (5) is to suppress many equivalent solutions of type $(W/\alpha, \ \alpha H)$, $\alpha > 0$.

The joint non-convex optimization in (5) is solved by alternating between W and H which optimizes one set of parameters while keeping the other fixed starting with an initialization of W by random elements from the training set (V) i.e., the RGB optical density of two randomly selected pixels corresponding to two columns of W of the histological image, as follows.

$$\text{For fixed } W, \ \hat{H} = \min_{H} \frac{1}{2}\|V - \hat{W}H\|_F^2 + \lambda\|H\|_1, \quad H \geq 0 \quad (6)$$

$$\text{For fixed } H, \ \hat{W} = \min_{W} \frac{1}{2}\|V - W\hat{H}\|_F^2, \quad W \geq 0, \quad \|W(:,j)\|_2^2 = 1. \quad (7)$$

This cost function is equivalent to a well-established dictionary learning objective [25], but with additional non-negative constraints on dictionary atoms W and coefficients H. These two alternating steps are called sparse coding for H and dictionary learning for W, and summarized as follows:

Sparse coding or estimation of H (6) with fixed $\hat{W}$ is an $l1$–regularized linear least-squares problem. A number of recent methods for solving this type of problems are based on coordinate descent with soft thresholding [26] and LARS-LASSO algorithm [27]. However, as pointed out in [28], when the columns of the dictionary are highly correlated, coordinate descent is found to be empirically slow. Hence, we use LARS with efficient Cholesky-based implementation which provides robust and accurate solution without the need for an arbitrary stopping criterion.

Dictionary learning or estimating W is done by using parameter free block-coordinate descent with warm restart [29] which does not require learning rate tuning and still guarantees convergence to a global optimum for convex optimization in equation (7). We use publicly available SPArse Modelling Software (SPAMS) [28] for sparse coding and dictionary learning. It should be noted that although SPAMS has been optimized to solve equation (5), the computational expense for such an iterative solver is too costly for the application to large size whole slide images.

B. Structure-Preserving Color Normalization (SPCN)

To normalize the color appearance of a source image s to that of a target image t, first we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$, and V_t into $W_t H_t$ using proposed SNMF (5). Then, a scaled version of the density map of source H_s is combined with color appearance of the target W_t instead of that of the source W_s to generate the normalized source image. This preserves the structure in terms of the stain density H, and only changes the appearance in terms of W, and can be described as follows:

$$H_s^{norm}(j,:) = \frac{H_s(j,:)}{H_s^{RM}(j,:)} H_t^{RM}(j,:), \quad j = 1, \ldots, r. \tag{8}$$

$$V_s^{norm} = W_t H_s^{norm} \tag{9}$$

$$I_s^{norm} = I_0 \exp(-V_s^{norm}) \tag{10}$$

where $H_i^{RM} = RM(H_i) \in R^{r \times 1}$, $i = (s,t)$ and $RM(\bullet)$ compute robust pseudo maximum of each row vector at 99%.

Unlike a non-linear mapping between statistics of H_s and H_t in [3], here we only multiply H_s by a scalar and hence keep the relative stain density maps of source image intact. This is similar to [9] where the maps of mixing coefficients of *principal* components were preserved for the source image. In this way, once accurate stain separation is done, our color normalization technique only changes the stain color appearance (basis) while preserving the source's structures. This structure-preserving property has been validated in Section IV-B for its utility. As far as we know, this is the first time *structure-preservation* has been simultaneously considered in both stain separation and color normalization. Previous attempts at stain separation (for example [9]) do not start with structural properties of stains such as sparseness and non-negativity and hence cannot always guarantee the structural invariance after color normalization.

C. Smart Patch-Based Acceleration Scheme for WSI

As explained in Sections III-A and III-B, a majority of computational time of SPCN is spent in the iterative optimization of SNMF, which slows its performance on WSI, especially when a computer RAM is limited with respect to the size of the WSI. Therefore, we propose a novel acceleration scheme for estimation of global color appearance matrix W of a WSI based on smart patch sampling and patch-wise stain separation. The patches have the same resolution as the original WSI to preserve local structures, which could have been lost by using downsampling – a trivial alternative. A schematic plot of the flow diagram of this scheme is shown in Fig. 4, and its details are described below.

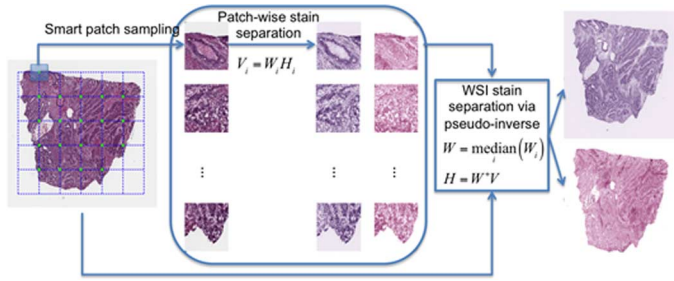


Fig. 4. Schematic diagram of fast WSI stain separation and global color matrix estimation.

We sample patches centered at corner points on grid as shown by green dots in Fig. 4, and discard those that lie in whitespace by comparing their luminosity against a threshold (we use 0.9 for all of our testing). The luminosity is the L value in the $L*a*b$ color space. Then we estimate the color basis matrices W_i for each of the sampled patches indexed by i using SNMF. Stain color columns in the W_i are sorted by ranking the blue channel intensities such that first column corresponds to hematoxylin and second corresponds to eosin. Finally, we take element-wise median of these matrices to make color estimation more robust to artifacts such as folding, blurring and holes. We normalize this median matrix to have unit vector columns, and denote the final color matrix thus obtained as W. The choice of grid and patch size will be addressed in Sections IV and V.

The stain separation for WSI is obtained through color deconvolution [17]:

$$H = W^+ V, \quad H \geq 0 \tag{11}$$

where $W^+ = (W^T W)^{-1} W^T$ is the Moore-Penrose pseudo-inverse matrix of W. Note that this operation can also be done separately for sub-images of WSI by using a single color appearance matrix W obtained for the entire image as described above i.e., obtaining H through pseudo-inverse will hold for any sub-image and hence can be parallelized.

After stain separation of source and target WSIs, $V_s = W_s H_s$ and $V_t = W_t H_t$, respectively, we change the color appearance of the source WSI to that of the target WSI while preserving original source stain concentration to obtain normalized source WSI (using (8)–(10)).

IV. Experiments and Results

The proposed SNMF-based stain separation and SPCN were extensively tested qualitatively and quantitatively. All of the datasets used for testing were publicly-available benchmark

data, which allow an easy and fair comparison with other competing methods.

A. Stain Separation Performance of SNMF

SNMF was quantitatively validated against NMF [12] and SVD [9] based stain separation on the samples from four different tissue types: stomach, prostate, colon, and bladder, available at The Cancer Genome Atlas (TCGA) data portal[1]. Blind color deconvolution (BCD) [11] does not provide source code for public access and thus we only quote their results and make a comparison relative to NMF.

1) Ground Truth Generation: While multiple ideas have been proposed for generating ground truth for stain density maps and their color appearance models, such as scanning slides with pure stains separately [12], [17] or scanning a single tissue section using chemical de-staining [30], these methods lead to errors because of stain leakage and co-location such as base-level presence of eosin even in areas covered primarily by hematoxylin. We found the ideas proposed in [11] to be closest to capturing the actual manifestation of the stain appearances in the context of real tissue structure. Following this approach, experts on our team marked a subset of tissue regions that *should* be stained by hematoxylin (i.e., cell nuclei) separately from a subset that *should* be stained by eosin (i.e., stromal matrix) in n randomly selected non-overlapping patches of size 1000×1000 for each tissue type (about the size of a microscopic field of view (FOV), each FOV covering physical area of $0.25\ \text{mm} \times 0.25\ \text{mm} = 0.0625\ \text{mm}^2$. The median of the selected pixels of each stain from each FOV in RGB space formed the reference color bases (I_k^{median}) for that stain and FOV (median in RGB space aid to filter out non-stain related noise such as small dusts or holes) and median of all n FOVs formed the ground truth color bases (I_{gt}^{median}) for that particular tissue. The W_k and W_{gt} corresponding to I_k^{median} and I_{gt}^{median} are computed from (1) followed by normalization of each column vector to unity. The variance of W_k across FOVs can be attributed to variation of stain quality over the tissue sections as well as intra-pathologist variation. Given the color appearance matrix, the stain density can be generated using color deconvolution method [17] as:

$$H_{gt} = (W_{gt}^T W_{gt})^{-1} W_{gt}^T V. \tag{12}$$

We also use the same comparison metrics as those in [11] to evaluate our stain separation.

2) Comparison of Stain Color Appearance W: Relative root mean square error (rRMSE) has been used for statistical comparison of different stain separation methods on their ability to estimate stain color basis close to the ground truth basis W_{gt}. This is a measure of how close the estimated color appearance W is to the ground truth W_{gt}, [11], as

$$rRMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \frac{\|(W_k - W_{gt})^T (W_k - W_{gt})\|}{\sqrt{\|(W_k)^T (W_k)\| \cdot \|(W_{gt})^T (W_{gt})\|}}} \tag{13}$$

[1]https://tcga-data.nci.nih.gov/tcga/

TABLE I
COMPARISON OF RRMSE OF STAIN COLOR BASIS MATRICES (W) BASED ON PATHOLOGIST'S ANNOTATIONS, NMF [12], SVD [9], AND THE PROPOSED SNMF WITH RESPECT TO GROUND TRUTH W_{gt}

Tissue	n samples	Pathologist	NMF	SVD	SNMF
Stomach	71	0.048	0.204	0.094	0.059
Prostate	70	0.034	0.200	0.074	0.049
Colon	70	0.046	0.140	0.061	0.062
Bladder	84	0.042	0.163	0.057	0.041

where $\|\cdot\|$ denotes the matrix trace. Table I summaries the results for 70 to 84 randomly selected fields of view (FOVs) for each of the four tissue types. SNMF outperformed both NMF [12] and SVD [9]. Our rRMSE improved about 3 folds as compared to that of NMF for all four tissue types and about 20–30% over SVD for three tissue types except colon where the performance of both methods was comparable. It is worth mentioning that our improvement over NMF was much larger compared to that of BCD method, since in [11] authors reported only a 20% to 40% increase over NMF. The pathologist performed better than all computational methods, as expected, because of the ground truth generation process that was based on pathologist's color models to begin with.

3) Comparison of Stain Density Maps H: Although stain color appearance is vital for color normalization, comparing that alone is not sufficient to quantify the accuracy of individual stain components, because the error in hematoxylin and eosin color appearance estimation can propagate to the error in estimating density maps H. We assess the accuracy of the estimated stain density maps H by computing Pearson correlation index between them and a map obtained using ground truth. As shown in the box plots of Fig. 5, while all of the three methods performed well in estimating hematoxylin density with average correlation > 0.95 in almost all cases, SNMF improved eosin quantification significantly over the other techniques. Wilcoxon signed rank tests were statistically significant at $p < 0.01$ for all cases except colon for which the performance was approximately similar for all techniques. An example of stain separation for each tissue type using the methods compared is included in the supplementary material[2].

4) Sensitivity Analysis of Sparse Coefficient λ: We analyzed sensitivity of SNMF performance to different values of the hyper parameter λ and found a wide range around 0.1, the value set for comparison to other techniques, in which the performance was largely the same. We varied λ from 0 to 0.2 in steps of 0.02 and computed rRMSE of the estimated stain color matrix and the Pearson correlation of the estimated stain density maps. Fig. 6 shows the sensitivity analysis of proposed SNMF across λ. Only eosin detection results are presented here as hematoxylin detection was close to 1 regardless of the value of λ and tissue type.

Setting $\lambda = 0$ reduces SNMF to NMF. As we increased λ from 0 to 0.02, we observed a rapid reduction in rRMSE as well as a sharp increase in the correlation index, which suggests a key role of sparse regularization in improving stain separation accuracy. The performance remained largely flat around 0.1

[2]Supplementary material are available in the supplementary files /multimedia tab.

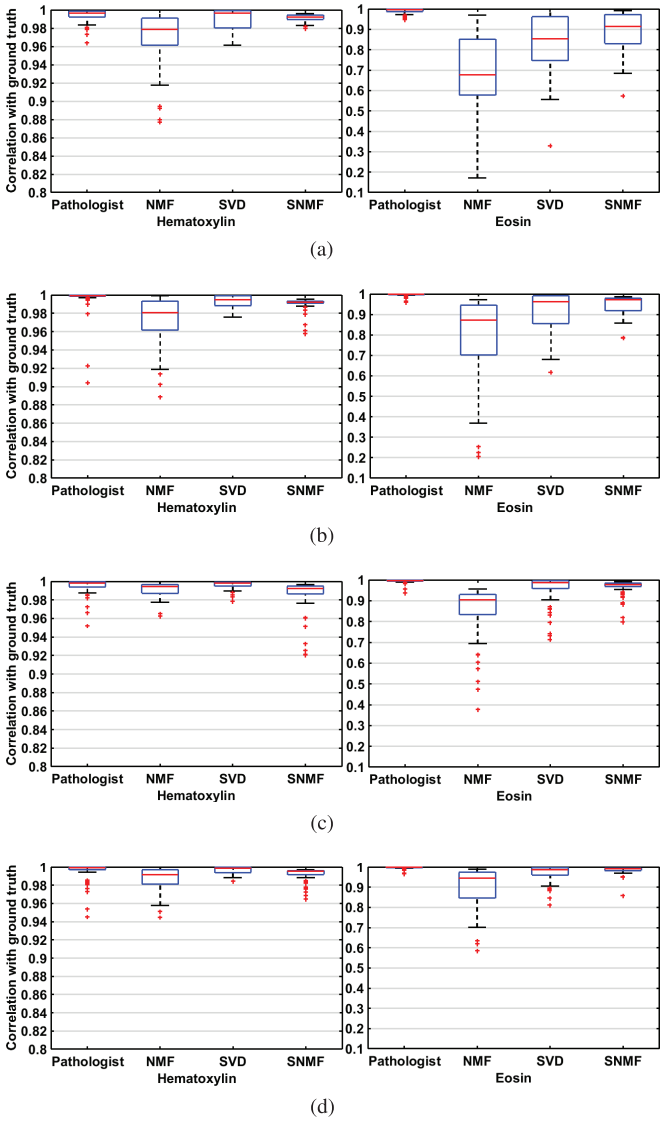


Fig. 5. Density map comparison of pathologist, NMF [12], SVD [9], and proposed SNMF against ground truth H_{gt}. (a) Stomach. (b) Prostate. (c) Colon. (d) Bladder.

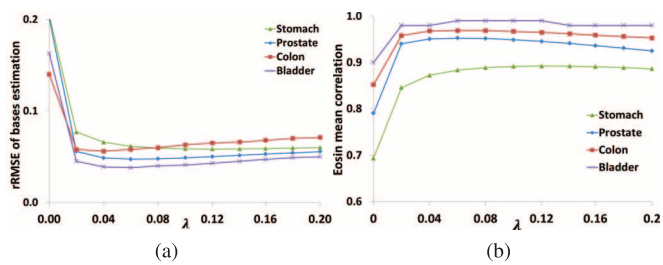


Fig. 6. Sensitivity of SNMF to hyper-parameter λ. Note that NMF is equivalent to $\lambda = 0$. (a) rRMSE of W. (b) Pearson correlation of eosin.

(chosen default value). However, larger values of λ led to slight decrease in performance because increase in sparsity most likely erased local structures such as some stain co-location or higher densities.

5) Qualitative Comparison by Expert Grading: We also evaluated SNMF-based stain separation qualitatively using

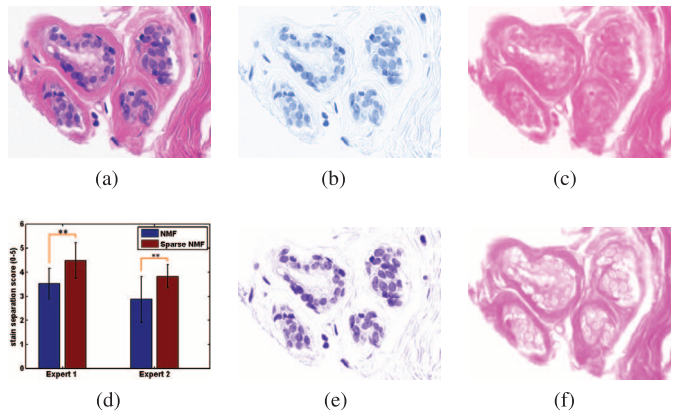


Fig. 7. Application of NMF and SNMF based stain separation on a sample image. (a) H&E image. (b) NMF H stain. (c) NMF E stain. (d) Expert scores. (e) SNMF H stain. (f) SNMF E stain.

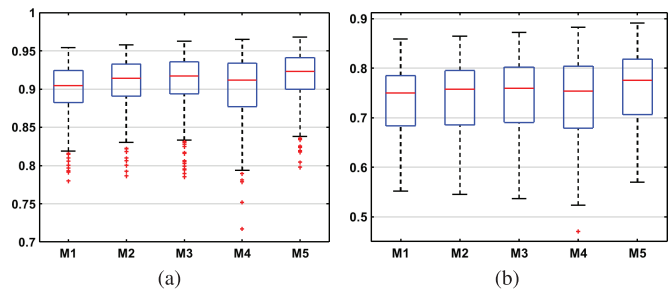


Fig. 8. Boxplot showing similarity metrics with respect to ground truth stain density map for different normalization methods: M1 (Reinhard), M2 (Macenko), M3 (Khan), M4 (NMF-CN), and M5 (SNMF-SPCN). (a) Pearson Correlation. (b) QSSIM.

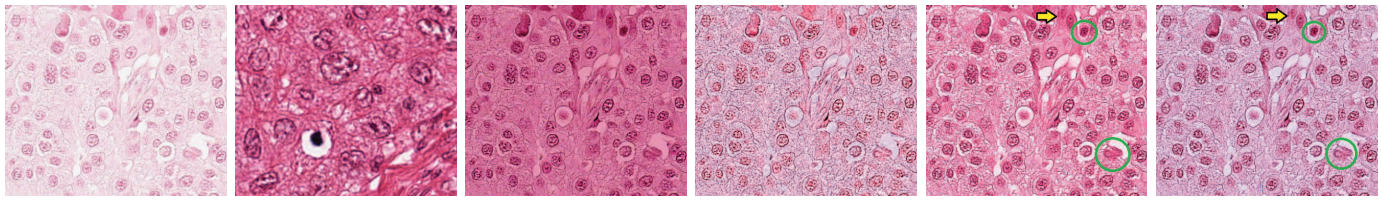
expert grading on 50 H&E breast cancer image dataset made available in [31]. Two experts independently evaluated the quality of stain separation using scores in the range of 0–5 (0 poorest, 5 perfect). Fig. 7(d) shows bar plot of mean $\pm$ standard deviation for the expert scores. It shows that on average SNMF was scored 4.48 and 3.82 respectively by the two pathologists, and deemed suitable for clinical use. In comparison, NMF only achieved average scores of 3.52 and 2.86 respectively by the two experts, which are significantly lower than SNMF ($p < 0.001$ using Wilcoxon signed-rank test). An example of stain separation results on sample image is shown in Fig. 7, which illustrates that SNMF achieved a much clearer separation with less distracting background stains (e.g., less stroma in H channel and less nuclei in E channel), demonstrating the crucial role of sparse regularization.

B. SPCN Validation

We compared the proposed SPCN with three state-of-the-art histological color normalization techniques: Reinhard *et al.* [14], Macenko *et al.* [9][3] and Khan *et al.* [3]. We start with quantitative validation first.

1) Quantitative Results: Validation of color normalization in histology images is not an easy task, especially in a quantitative manner. The results presented in previous studies ([3], [9], [10])

[3]In the original implementation, authors used built-in reference color appearance as W_t while here we compute it from the target image for a fair comparison.



(a) (b) (c) (d) (e) (f)

Fig. 9. Visual comparison of different normalization techniques. See Section IV-B.2 for detailed explanation. (a) Source. (b) Target. (c) Reinhard. (d) Macenko. (e) Khan. (f) Proposed SPCN.

were mostly qualitative or indirect validation (for example, improvement in the tumor segmentation performance). Here we use a direct benchmark for color variation using the ICPR2014 contest dataset[4]. In this dataset, each breast cancer section was acquired using two different scanners, which results in the color variation between them (as shown in Fig. 1).

We randomly sampled 300 images taken using an Aperio scanner as source images, and 300 images from Hamamatsu scanner as target images. We applied different color normalization techniques to change source color appearance into target color appearance and compared normalized source to the corresponding target image of the same tissue section (ground truth). Since there is a minor misalignment between matching Aperio and Hamamatsu images due to imperfect calibration and different spatial resolutions of the scanners, we *rigidly* registered Hamamatsu image to the corresponding Aperio image to form the ground truth. We chose multichannel[5] Pearson correlation and quaternion structural similarity index (QSSIM) [32] as similarity metrics between the normalized source and ground truth. Boxplots in Fig. 8 demonstrate SPCN superior performance over other competing methods ($p < 0.001$ by Wilcoxon signed-rank test). An example is shown in Fig. 10 and also available in high resolution in supplementary material.

2) Qualitative Validation: The proposed color normalization method can also be used to improve contrast of low-quality histology images, such as faded stain. Fig. 9(c) show that Reinhard *et al.* [14] is unable to preserve local structure and fills lumen with color when it should remain white after normalization. Fig. 9(d) show that SVD-based normalization by Macenko *et al.* [9] is largely monochrome, noisy and is also unable to preserve the lumen structure, possibly due to the unrealistic negative stain components at these regions. Normalization method by Khan *et al.* [3] in Fig. 9(e) does a better job of keeping the structure intact and showing a differentiated second stain, but in some area (highlighted in green circles), it is unable to reproduce the less abundant stain (hematoxylin) faithfully. Compared to all these methods SPCN reproduces the local structure, and is able to differentially show the two stains in Fig. 9(f), even in those areas where the method by Khan *et al.* had difficulty (green circles). High resolution images are available in supplementary material for better visual comparison.

3) Histogram Interpretation of Structure-Preservation of SPCN: We plotted 3D histogram of RGB image to visualize color normalization process. As shown in Fig. 10(a) and (b), the source and target images have different color distribution (the volume of the ball indicates histogram density). Unlike histogram specification (Fig. 10(c)), SPCN does not seek an identical match to the histogram of target (Fig. 10(d)). For example, SPCN retains equal amount of white lumens as that in the source image. This is reflected by the comparable size of white balls in the two corresponding histograms. By contrast, histogram specification tries to achieve an equal amount of white lumens as in the target image and therefore creates artificial structures.

C. Validation of Acceleration Scheme for WSIS

We validated the proposed acceleration scheme on 27 H&E stained colon tissue WSIs of size 20000×20000 covering a physical area of $5\ \text{mm} \times 5\ \text{mm} = 25\ \text{mm}^2$ available at TCGA data portal[6]. An example sub-image is shown in schematic plot of Fig. 4. Ground truth stain density was obtained through expert stain separation as described in Section IV-A1. We performed the stain separation using the proposed patch-based scheme and compared it to a direct stain separation on WSI.

Fig. 11 shows box plots of Pearson correlation between ground truth and methods (NMF and SNMF). Similar to small-size images, SNMF outperformed NMF in WSI stain separation, particularly for the estimation of eosin component ($p < 0.001$ by Wilcoxon signed-rank test). The rRMSE measures of W estimation for slow and fast scheme were 0.066 and 0.068 for SNMF; 0.115 and 0.097 for NMF respectively. The proposed acceleration scheme achieved approximately 20-fold acceleration (in our implementation we use 6×6 grid, about 20 patches were sampled in the decomposition and each of size 1000×1000) over a direct decomposition on WSI for both NMF and SNMF methods while maintaining a consistent or improved performance. It took an average of 82 seconds on a four core processor for patch-based scheme, compared to an average 25 minutes for the direct stain separation for each of the 27 H&E stained breast WSIs. There are several factors that contribute to this acceleration: (1) whitespaces were excluded based on the proposed sampling scheme while computing SNMF or NMF step, (2) Several patches can be processed in parallel on different CPU cores for a theoretical speed up of $(\text{sizeofWSI} \times \text{numofCPUcores})/(\text{numofpatches} \times \text{sizeofpatch})$, and (3) Inputs and temporary variables corresponding to a small patch can be accommodated in the RAM, thus reducing time-consuming read-write calls to the hard disk.

[4]http://mitos-atypia-14.grand-challenge.org/

[5]Multichannel measure is the average of three RGB channel-wise measures

[6]http://tcga-data.nic.nih.gov/tcga/

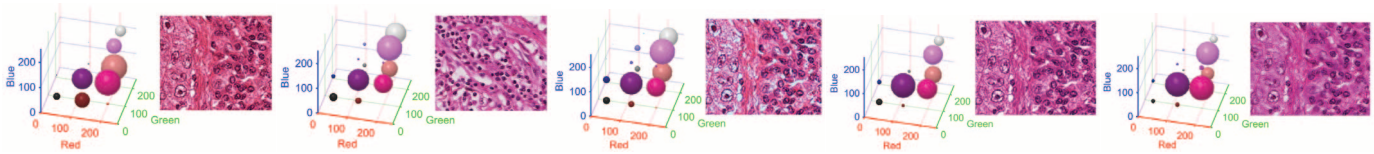


Fig. 10. 3D histogram visualization of color normalization on a sample source and target image from the data used in Section IV-B.1 . (a) Source. (b) Target. (c) Histogram Specification. (d) Proposed SPCN. (e) Ground Truth.

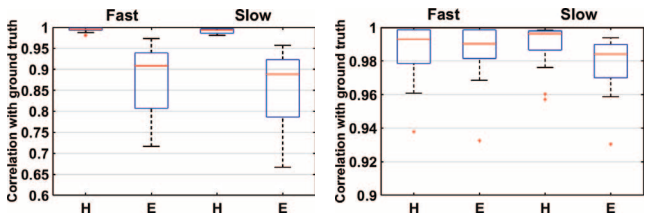


Fig. 11. Box plots of quality metrics between separated stain and ground truth for 27 H&E stained colon tissue WSIs for hematoxylin (H) and eosin (E) suggest consistent results between our fast patch-based scheme and the direct decomposition of WSI (slow approach). (a) NMF. (b) SNMF.

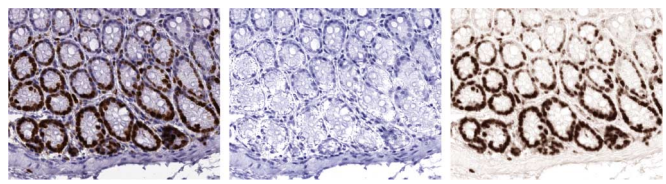


Fig. 12. Proposed SNMF on IHC (H&DAB) stained image of head and neck tissue: (left) input image, (middle) Hematoxylin, (right) Diaminobenzidine.

V. Discussion and Conclusions

Color normalization is an essential step to remove undesirable color variation in histopathological images and is useful for analyzing disease and its progression on large datasets from different pathology labs. A potential danger of such a technique that concerns pathologists is that tissue structures presented in the original image could be altered after normalization. In this paper, we proposed a novel structure-preserving color normalization (SPCN) scheme, which changes color of one image (source) to match that of another (target) while reliably keeping source structural information intact. A key step in our normalization scheme is an accurate stain separation of both source and target images based on sparse regularized NMF. Extensive qualitative and quantitative validation demonstrated superior performance of the proposed SNMF stain separation and SPCN techniques. Additionally, we proposed a smart patch-based scheme to accelerate computation on WSI. It is worth mentioning that the proposed estimation of global color appearance based on smart patch sampling is a general strategy, which can be easily combined with other stain separation and color normalization techniques to accelerate their implementation on WSI. Also, the proposed algorithms were insensitive to the different random initialization of W when initialized by the elements from the training set.

Our work is an improvement over pioneering method of Macenko et.al. [9]. We assumed certain structural properties of histopathological images that are biologically more plausible. First, we assumed non-negativity. That is, a biological material is either absent (zero light absorption) or present (positive light absorption), and thus optical density cannot be negative. Second, we assumed sparsity. That is, a given location (pixel) is likely to be occupied by one or the other type of biological material, but not a mix of the two. Based on the two assumptions, we modeled stain or structure density maps as sparse and non-negative. By contrast, [9] computes singular value decomposition, which can lead to dense and even unrealistic negative stain amount and cannot guarantee the preservation of underlying biological structures (as shown in Fig. 9(d)).

As mentioned in the Section I, stain leakage is a major factor for stain co-localization in histology images. It is caused by more than one target protein being present in the same area. For instance, in H&E stained specimen, hematoxylin can leak into structures that should be ideally colored by eosin only for some infinitesimal and unintended binding and vice versa. This is also a reason that the reference color of eosin-specific structure (e.g., stroma) in eosin-alone staining is different from that of in H&E combination. Our SNMF utilizes a sparse constraint to decompose an image into different tissue structures represented by effective stains, rather than the physical amount of hematoxylin and eosin like NMF (an example shown in Fig. 7). We introduced the term *effective stain* to describe stain combination (usually mostly an intended and some amount of unintended stains) uniqueness to different biological structures. This clear separation of effective stains can lead to better tissue quantification as it would help nuclear segmentation due to less distraction from residual background (unintended) stains. In the presented color normalization evaluation, SNMF-SPCN attained a higher accuracy than NMF-CN. A likely biological interpretation is that the level of stain leakage is dependent on tissue preparation and staining procedure and hence is image-specific, which also contributes to inter-image color variation. NMF does not take this factor into account during the normalization process. On the other hand, SNMF-based color normalization, by contrast, is based on a clear separation of the structures, and therefore overcomes this problem and acts like a repainting of the source structures.

Our earlier work [13] was developed in parallel with Xu *et al.*'s [33] and both use sparse non-negative matrix factorization (SNMF) for for stain separation, with a minor difference that the latter also penalizes $l2$-norm of the factors. However, there are other major differences between the two independent works. While the main objective of performing stain separation in this work is color normalization, Xu *et al.*'s [33] objective

was nuclear segmentation in breast tissue images. Moreover, we propose a patch-based acceleration scheme to process whole slide images more efficiently, which is very general and useful tool in practice. Additionally, we evaluated our methods on a more extensive set of tissues and also performed sensitivity analysis of sparse regularization parameter to ascertain the desired range of interest across different tissue types.

All data used in the main paper were H&E stained histological images, which is mainly due to two reasons: 1) H&E is the most widely used histological staining in pathology labs, whose popularity is based on its comparative simplicity and ability to demonstrate on large number of different tissue structures [34], and 2) we preferred to evaluate our proposed algorithms on open source data, most of which is H&E stained, to allow a fair comparison with other competing methods. Nevertheless, we have performed some preliminary testing on the immunohistochemistry (IHC) staining such as H&DAB (DAB stands for Diaminobenzidine). Results in Fig. 12 show that SNMF can achieve a good structure separation of H&DAB images, yet the resulted stain density map of H and DAB does not linearly correlate to the actual amount of DAB stain due to refraction. Also, the brown DAB is not a true absorber of light, but a scatterer, and has a very broad and featureless spectrum [35]. Thus, Beer-Lambert law of absorption is not valid for brown DAB. Besides, immunohistochemical DAB stains can selectively bind to cancer target structures and hence only occupy a small region of the whole image, in which case the current simple grid patch sampling scheme may miss it. One possible solution for efficient exploration of WSI for locating cancer region is the dynamic sampling approach presented in [36] where small patches were dynamically sampled using a score function and Voronoi diagram. A more thorough evaluation on IHC stained histology needs to be done. Another direction to explore based on this study is to determine how color normalization impacts downstream tasks in computational pathology such as mitosis detection, epithelial-stromal classification or clinical end-point prediction.

REFERENCES

[1] M. N. Gurcan *et al.*, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.

[2] M. Uhlen *et al.*, "Towards a knowledge-based human protein atlas," *Nature Biotechnol.*, vol. 28, no. 12, pp. 1248–1250, 2010.

[3] 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 Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, Jun. 2014.

[4] L. Peter *et al.*, "Leveraging random forests for interactive exploration of large histological images," in *Proc. MICCAI*, 2014, pp. 1–8.

[5] H. Llewellyn, "Observer variation, dysplasia grading, and HPV typing a review," *Am. J. Clin. Pathol., Pathol. Patterns Reviews*, vol. 114, no. Suppl 1, pp. S21–S35, 2000.

[6] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1989.

[7] M. Veta *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Med. Image Anal.*, vol. 20, no. 1, pp. 237–248, 2015.

[8] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.

[9] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *Proc. ISBI*, 2009, vol. 9, pp. 1107–1110.

[10] D. Magee *et al.*, "Colour normalisation in digital histopathology images," in *Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endosc. (MICCAI Workshop)*, 2009, pp. 100–111.

[11] M. Gavrilovic *et al.*, "Blind color decomposition of histological images," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 983–994, Jun. 2013.

[12] A. Rabinovich, S. Agarwal, C. Laris, J. H. Price, and S. J. Belongie, "Unsupervised color decomposition of histologically stained tissue samples," in *Adv. Neural Inf. Process. Syst.*, 2003.

[13] A. Vahadane *et al.*, "Structure-preserved color normalization for histological images," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 1012–1015.

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

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

[16] P. A. Bautista and Y. Yagi, "Staining correction in digital pathology by utilizing a dye amount table," *J. Digital Imag.*, pp. 1–12, 2015.

[17] 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.

[18] W. W. Parson, *Modern Optical Spectroscopy*. New York: Springer, 2007.

[19] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[20] J. Newberg and R. F. Murphy, "A framework for the automated analysis of subcellular patterns in human protein atlas images," *J. Proteome Res.*, vol. 7, no. 6, pp. 2300–2308, 2008.

[21] N. Trahearn, D. Snead, I. Cree, and N. Rajpoot, "Multi-class stain separation using independent component analysis," in *SPIE Med. Imag.*, 2015, pp. 94200J–94200J.

[22] A. R. Webb, *Statistical Pattern Recognition*. New York: Wiley, 2003.

[23] M. Niethammer, D. Borland, J. S. Marron, J. Woosley, and N. E. Thomas, "Appearance normalization of histology slides," in *Mach. Learn. Med. Imag.*. : Springer, 2010, pp. 58–66.

[24] T. Chen and C. Chefd'hotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in *Machine Learning in Medical Imaging*. New York: Springer, 2014, pp. 17–24.

[25] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[26] T. T. Wu and K. Lange, "Coordinate descent algorithms for lasso penalized regression," *Ann. Appl. Stat.*, pp. 224–244, 2008.

[27] B. Efron *et al.*, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.

[28] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.

[29] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena, 1999.

[30] M. T. McCann, J. Majumdar, C. Peng, C. A. Castro, and J. Kovacevic, "Algorithm and benchmark dataset for stain separation in histology images," in *Proc. IEEE Int. Conf. Image Process.*, 2014.

[31] E. D. Gelasca, J. Byun, B. Obara, and B. S. Manjunath, "Evaluation and benchmark for biological image segmentation," in *Int. Conf. Image Process.*, 2008, pp. 1816–1819.

[32] A. Kolaman and O. Yadid-Pecht, "Quaternion structural similarity: A new quality index for color images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1526–1536, Apr. 2012.

[33] J. Xu *et al.*, "Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis," *Comput. Med. Imag. Graph.*, 2015.

[34] M. Gamble and I. Wilson, "The hematoxylins and eosin," in *Theory and Pratice of Histological Techniques*. New York: Churchill Livingstone, 2008, vol. 2008, pp. 121–35.

[35] C. M. van der Loos, "Multiple immunoenzyme staining: Methods and visualizations for the observation with spectral imaging," *J. Histochem. Cytochem.*, vol. 56, no. 4, pp. 313–328, 2008.

[36] A. Veillard, N. Loménie, and D. Racoceanu, "An exploration scheme for large images: Application to breast cancer grading," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 3472–3475.