Bayesian K-SVD for H&E blind color deconvolution. Applications to stain normalization, data augmentation and cancer classification.

Fernando Pérez-Bueno[a,*], Juan G. Serra[b], Miguel Vega[c], Javier Mateos[a], Rafael Molina[a], Aggelos K. Katsaggelos[b]

[a]*Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Spain*
[b]*Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA*
[c]*Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Granada, Spain*

Abstract

Stain variation between images is a main issue in the analysis of histological images. These color variations, produced by different staining protocols and scanners in each laboratory, hamper the performance of computer-aided diagnosis (CAD) systems that are usually unable to generalize to unseen color distributions. Blind color deconvolution techniques separate multi-stained images into single stained bands that can then be used to reduce the generalization error of CAD systems through stain color normalization and/or stain color augmentation. In this work, we present a Bayesian modeling and inference blind color deconvolution framework based on the K-Singular Value Decomposition algorithm. Two possible inference procedures, variational and empirical Bayes are presented. Both provide the automatic estimation of the stain color matrix, stain concentrations and all model parameters. The proposed framework is tested on stain separation, image normalization, stain color augmentation, and classification problems.

Keywords: Bayesian modelling, Histological images, Blind Color Deconvolution, Stain Normalization

1. Introduction

The analysis of Whole-Slide Images (WSI), i.e., digitalized histological slides of tissue sections, is a crucial step towards the development of Computer Aided Diagnosis (CAD) systems. The tissues in a WSI are stained with different dyes to make their structure visible under the microscope. Hematoxylin-Eosin (H&E) is the most common combination, highlighting cell nuclei in blue color and cytoplasm and connective tissue in pink, respectively

*Corresponding author
Email addresses: fpb@ugr.es (Fernando Pérez-Bueno), jgserra@northwestern.edu (Juan G. Serra), mvega@ugr.es (Miguel Vega), jmd@decsai.ugr.es (Javier Mateos), rms@decsai.ugr.es (Rafael Molina), aggk@eecs.northwestern.edu (Aggelos K. Katsaggelos)

(Fischer et al., 2008). However, the color distribution of H&E WSI is affected by the staining and scanning procedures (Tosta et al., 2019b), resulting in inter- and intra- laboratory color variations. These variations hamper the performance of CAD systems, which are usually unable to generalize to unseen color distributions. The different approaches proposed to minimize the influence of color variation on CAD systems can be categorized into three groups: Blind Color Deconvolution (BCD), Color Normalization (CN) and Stain Color Augmentation (SCA). Let us review the most important contributions in each of these groups.

1.1. Blind Color Deconvolution

BCD techniques deal with color variation by estimating the image specific stain color-vectors and stain concentrations. The pioneer approach by Ruifrok and Johnston (2001) experimentally obtained a standard color vector matrix that is still used today. More recent methods tackle inter-slide variations by using different techniques. The use of Non-Negative Matrix Factorization (NMF) was proposed by Rabinovich et al. (2004), (Vahadane et al., 2016; Xu et al., 2015) added regularization and sparsity terms which encapsulate the assumption that a type of stain is only bound to certain structures. In Tosta et al. (2019a) the sparsity parameter was estimated using a fuzzy set method. Independent Component Analysis (ICA) was utilized in Trahearn et al. (2015) and extended in (Alsubaie et al., 2016, 2017) by applying ICA in the wavelet domain where the independence condition among sources is relaxed. The use of Singular Value Descomposition (SVD) was proposed in Macenko et al. (2009) to separate H&E channels. In McCann et al. (2014), the authors take into account the interaction between dyes. The method in Macenko et al. (2009) was revised in Astola (2016), where the author states that better results are obtained by applying it in the linearly inverted RGB-space instead of the (logarithmically inverted) absorbency space. Clustering was utilized in Gavrilovic et al. (2013) using the Maxwellian chromacity plane to obtain the stain vectors. Vicory et al. (2015) used K-means and a prior on the stain vectors to prevent misclustering when the amount of each stain is not balanced. In Khan et al. (2014), images are segmented into background and pixels belonging to each stain using supervised relevant vector machines. The color-vector for each stain is then defined as the mean of the pixels in each class. The work in Zheng et al. (2019) includes the deconvolution by Ruifrok as starting point and optimizes the color-vector and concentration values using a prior knowledge based objective function. Recently, a three-step method using Gabor kernels, structure segmentation and a final deconvolution step has been presented in Salvi et al. (2020).

Several Bayesian approaches have already been presented. In Hidalgo-Gavira et al. (2018), a similarity prior on the color-vectors as well as a smoothness Simultaneous Autorregresive (SAR) prior model on each stain concentration were used. This work was extended with the use of a TV prior in Pérez-Bueno et al. (2020) and with the use of sparse general Super Gaussian priors on the high-pass filtered concentrations in Pérez-Bueno et al. (2021).

Despite the Deep Learning popularity, few works have used it for BCD. Based on Macenko et al. (2009), the work in Duggal et al. (2017) implements a stain deconvolution layer for CNNs to provide a stain separated input to CNN-classifiers. Similarly, Zheng et al. (2021) use a Capsule Network that produces multiple stain separation candidates using 1 by 1 convolution operators and finally assembles the output based on a sparse constraint.

All BCD techniques have in common that they separate color from structural information, offering a strong control on the information preprocessing. They can preserve the tissue

structure and lead to high fidelity to the original images. Often, BCD methods are presented as CN methods since the obtained results can be used for CN (by normalizing each stain separately), but this is only one of the possible solutions BCD offers to deal with color variation.

1.2. Color Normalization

CN aims to reduce the stain variation by matching the color in the images to a selected template or reference. Direct CN methods do not necessary estimate the concentrations and color vector matrix as BCD methods do. In Tosta et al. (2019b), they are classified into histogram matching, color transfer, and spectral matching. The first one adjusts image colors using histogram information (Reinhard et al., 2001), a common solution for general images. However, this is not appropriate for histological images as it ignores the local information and the unequally distribution of the stains. This work was further developed with three fuzzy normalization steps and adapted to histopathological images in Vijh et al. (2021). Color transfer methods assume that stain concentration is closely related to tissue structure and usually include region or dye segmentation. The latest color transfer methods, based on deep generative models (Janowczyk et al., 2017; Zanjani et al., 2018; Bentaieb and Hamarneh, 2018), perform CN without a previous color deconvolution by formulating the problem as a style transfer task where the style is the color distribution of a selected laboratory. Recently, other popular CNN architectures have been adapted to CN problems, such as Pix2pix (Salehi and Chalechale, 2020), disentangled representations (Xiang et al., 2020), CycleGAN (Runz et al., 2021) or Invertible Neural Networks (Lan et al., 2021). Since they require large datasets to train the networks that transform to an specific stain distribution, usually, they cannot handle intra-laboratory variations. Spectral matching methods typically perform BCD as a first step to CN. In Tosta et al. (2019b), most of the BCD methods mentioned in the previous section are reviewed as CN methods. The normalization is usually performed by replacing the stain color vectors obtained using BCD by the reference color vectors, often obtained from a template image (Vahadane et al., 2016; Vicory et al., 2015; Zheng et al., 2019). Different approaches are used to adjust the concentration intensity of both source and target images. In Macenko et al. (2009) each concentration intensity is scaled by using the 99th percentile to compute a robust estimation of the maximum. In Vicory et al. (2015) the median of the concentrations is used while in Zheng et al. (2019) the parameters normalizing the intensities are estimated jointly with the stain color vectors. Recently, Hoque et al. (2021) presented a multiscale Retinex model, that estimates and corrects the reflectance and illumination map for pixels of both stains separately.

1.3. Stain Color Augmentation

Data augmentation is a popular solution to reduce generalization error on CNN-based classifiers (Zheng et al., 2021). In contrast to BCD and CN, which aim to avoid the unseen stain distribution by eliminating the color variation, the augmentation approach aims to simulate unseen data by producing realistic variations of the available data. Although for histological images, morphological, generative (Wei et al., 2020; Zhu et al., 2017), and color augmentation techniques can be used (Tellez et al., 2019; Mpinda Ataky et al., 2020), in this study, we will focus on the latter to study the effect of color augmentation on classification in comparison to BCD and CN techniques. Color augmentation techniques do not modify the

image morphological features and only generate color variations. In Liu et al. (2017) common computer vision perturbations of brightness, contrast and hue are used. Furthermore, an specific histological stain color augmentation (SCA) technique was recently proposed in Tellez et al. (2018) where the method in Ruifrok and Johnston (2001) is applied to obtain the H&E concentration and variations of the observed data are created. In Tellez et al. (2019), several SCA and CN methods were evaluated on classification tasks with CNN. Additionally, a new CNN based CN method is proposed which is trained on SCA data.

1.4. Contributions

In the recent years, the field of BCD has received few contributions as CN approaches using Deep Learning usually avoid this step. However, BCD has some advantages for histological image analysis that should not be ignored. Its structure preserving properties, interpretability by doctors, and potential for classification purposes make this a field of interest for new works. The use of Bayesian models for BCD has been hardly explored and previous contributions are dependant on a similarity prior on the color-vectors. The choice of a reference color-vector matrix used for that prior, becomes a problem when working with images from different laboratories. Finally, BCD is required for the recently proposed SCA, which has been only compared to CN in Tellez et al. (2019). SCA and BCD have never been directly compared. For those reasons, in this work we propose a novel Bayesian K-SVD approach to perform BCD of histological images. K-SVD (Aharon et al., 2006) is a popular greedy algorithm for dictionary learning and sparse representation of signals. In BCD of histological images, the dictionary and the sparse representation will be the stain color vectors and the stain concentrations, respectively (Vahadane et al., 2016). However, K-SVD has two mayor drawbacks that need to be addressed for its use in BCD, the lack of uncertainty in the estimation procedure and the need to know in advance the number of non-zero components in the signal. The Bayesian K-SVD model (Serra et al., 2017) we adapt in this paper to BCD tackles these problems allowing its use for BCD of histological images. Using the obtained stain concentrations and color-vectors, our method can be utilized for CN and SCA. Our contributions are summarized as follows:

- Proposal of a new BCD framework that is able to preserve histological structures, with two possible inference approaches: variational and empirical Bayes.
- Unsupervised estimation of the stain concentrations and color properties.
- Automatic estimation of all model parameters.
- Stain specific data augmentation using the stain concentrations and color-vector matrix.
- Performance evaluation on large histological datasets with intra- an inter-laboratory variations.
- Analysis of classification performance when using normalized images or stain concentrations.

The proposed method is tested on a set of experiments designed to cover the main tasks of digital histopathology.

The paper is organized as follows. In section 2 we present the mathematical formulation of the BCD problem. In section 3, this problem is cast into the hierarchical Bayesian paradigm and inference is carried out to estimate the stain concentrations and color-vectors as well as all model parameters. Using Empirical Bayes, in section 4 we modify the inference already presented in section 3 to increase the sparsity of the obtained solution. Section 5 adapts the proposed methods to its application in massive WSI. Section 6 describes the utilized images and methods. The effectiveness of the proposed framework is experimentally assessed in Section 7, where the proposed methods are compared to classical and state-of-the-art alternatives. Finally, section 8 concludes the paper.

2. Problem Formulation

Each WSI is stored as an $M \times N \times 3$ RGB intensity image which is rearranged into the matrix $\mathbf{I} \in \mathbb{R}^{3\times Q}$, $Q = MN$, where each value $i_{cq} \in \mathbf{I}$ represents the transmitted light across the slide for pixel q and channel c. Diagnosis protocols use the contribution of each stain to this value, that is, its absorbency or *optical density* (OD). The OD corresponding to intensity i_{cq}, $y_{cq} \in \mathbf{Y}$, is defined as $y_{cq} = -\log_{10}(i_{cq}/i^0_{cq})$, where i^0_{cq} denotes the incident light. The monochromatic Beer-Lambert law establishes that a slide $\mathbf{Y}$ stained with N_s stains follows the equation

$$\mathbf{Y} = \mathbf{MC} + \mathbf{N}\,, \tag{1}$$

where $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_{N_s}] \in \mathbb{R}^{3\times N_s}$ is the normalized stains' specific color-vector matrix; $\mathbf{C} \in \mathbb{R}^{N_s\times Q}$ is the stain concentration matrix, its qth column, $\mathbf{c}_q = [c_{1,q}, \ldots, c_{n_s,q}]^{\mathrm{T}}$, represents the contribution of each stain to the qth pixel value in $\mathbf{Y}$; and, finally, $\mathbf{N} \in \mathbb{R}^{3\times Q}$ is a random matrix with i.i.d. zero-mean Gaussian components with unknown variance β^{-1}. Each column, $\mathbf{m}_s$, in matrix $\mathbf{M}$ is assumed to be a unit ℓ^2-norm stain color-vector containing the relative RGB color composition of the corresponding stain in the OD space.

Notice that each column $\mathbf{y}_q$ can be represented as a linear combination of the color vectors weighted by the corresponding concentrations, that is, $\mathbf{y}_q = \sum_{s=1}^{N_s} c_{sq}\mathbf{m}_s + \mathbf{n}_q$. Hematoxylin is a basic stain that dyes basophilic structures, namely nuclei, while eosin is an acidic stain that fixes to cytoplasm and other structures, usually referred to as eosinophilic. Although the actual color of biological structures will be influenced by both stains, they will present structure-specific color properties (Vahadane et al., 2016) (effective stains) that are the basis of differential staining. Therefore, we can assume that most pixels in the image are stained by a single effective stain (Vahadane et al., 2016), making our stain concentration matrix sparse, in other words, most of the weights, $c_{sq} \in \mathbf{c}_q$, in this linear combination are expected to be zero (or very small). We would like to find not only the sparse coefficients of these linear combinations, but at the same time, also estimate the color vectors $\mathbf{m}_s$ which result in the best, most sparse, solution. This dual estimation can clearly be understood as a dictionary learning problem. Notice that we estimate the effective stains that allow to sparsely separate biological structures.

The original problem of finding an exactly sparse solution minimizing the number of non-zero elements in each $\mathbf{c}_q$ (i.e., minimizing $\|\mathbf{c}_q\|_0, \forall q$)[1], is known to be NP-hard, see (Babacan

[1] $\|\cdot\|_0$ denotes the ℓ^0-(pseudo)norm, which counts the number of non-zero elements in a vector.

et al., 2010), for example. The true solution can be approximated with greedy methods (e.g, the popular K-SVD Aharon et al. (2006) method). Alternatively, the sparsity constraint on the concentration vectors can be relaxed by using the ℓ^1-norm instead. Formally, we can formulate this problem as

$$\begin{aligned} &\min_{\mathbf{M},\mathbf{C}} \|\mathbf{Y} - \mathbf{M}\mathbf{C}\|_{\mathrm{F}}^2 \\ &\text{s.t. } \|\mathbf{c}_q\|_1 \leq T, \ \forall q, \end{aligned} \tag{2}$$

where $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_1$ denote the Frobenius and ℓ^1-norms respectively, and T is nonnegative real parameter that determines the degree of regularization. The main advantage of this relaxation is that convex optimization techniques can be used to solve this problem (e.g., (Aharon et al., 2006; Mairal et al., 2009; Zhou et al., 2009)).

The novel Bayesian framework we propose solves the histological color deconvolution as an ℓ^1 dictionary learning problem, following the method introduced in Serra et al. (2017), automatically estimating the optimal color-vector matrix $\mathbf{M}$, the posterior distribution of $\mathbf{C}$ considering the uncertainty of the coefficients, along with all model parameters. The next section gives detailed intuition on the modelling and inference of the proposed method, albeit not a full derivation. We encourage the interested reader to consult Appendix A for further explanation.

3. Bayesian Model and Inference

Our Bayesian model for solving the dictionary learning problem in (2) relies on defining suitable probability distributions on the observations $\mathbf{Y}$ and on the set of unknowns $\{\beta, \mathbf{M}, \mathbf{C}\}$. The observation model in (1) described above corresponds to the isometric Gaussian distribution on $\mathbf{Y}$ given by

$$\mathrm{p}(\mathbf{Y}|\beta, \mathbf{M}, \mathbf{C}) \propto \beta^{\frac{3Q}{2}} \exp\big(-\frac{\beta}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{C}\|_{\mathrm{F}}^2\big). \tag{3}$$

The obvious choice for the noise precision, β, since it is a positive-valued variable, is a gamma distribution, thus, $\mathrm{p}(\beta) = \Gamma(\beta|a^\beta, b^\beta)$ with $a^\beta, b^\beta > 0$. Our modelling for the stain vector matrix $\mathbf{M}$ focuses on imposing unit norm for each column $\mathbf{m}_s$; for this purpose we use a flat prior on the columns of $\mathbf{M}$ such that $\mathrm{p}(\mathbf{m}_s) = \text{const.}$, if $\|\mathbf{m}_s\| = 1$, 0 otherwise, and assume independent column vectors. Finally, notice that the sparsity constraint on the coefficient vectors $\mathbf{c}_q$ in (2) is equivalent to imposing a zero-mean Laplace distribution with scale parameter $\lambda_q > 0$, $\mathrm{p}(\mathbf{c}_q) \propto \exp\big(-\sqrt{\lambda_q}\|\mathbf{c}_q\|_1\big)$. The Laplace prior is more peaked than the normal distribution with longer tails, which is also interesting for structure preserving (Babacan et al., 2012). Unfortunately, the non-conjugacy of this distribution with the likelihood in (3) makes inference intractable. We circumvent this problem by using a two-tiered hierarchical prior on $\mathbf{c}_q$ instead. First, we impose a zero-mean normal distribution with diagonal covariance matrix $\mathbf{\Gamma}_q = \mathrm{diag}(\boldsymbol{\gamma}_q)$, i.e., $\mathbf{c}_q \sim \mathcal{N}(\mathbf{c}_q|\mathbf{0}_{N_s}, \mathbf{\Gamma}_q)$. And secondly, we use the Gamma hyperpriors on the positive-valued γ_{sq} given by $\gamma_{sq} \sim \Gamma(1, \lambda_q/2)$ and assume independence yet again, so that $\mathrm{p}(\boldsymbol{\gamma}_q) = \prod_s \mathrm{p}(\gamma_{sq})$. This two-tier prior can be further expanded with a third prior on the scale parameters λ_q, however, although it gives more flexibility to the model, in practice does not turn into noticeable estimation improvement. The idea behind

this hierarchical prior is to sample the covariance matrix of the normal distribution $p(\mathbf{c}_q|\boldsymbol{\gamma}_q)$ from a Gamma distribution with shape 1 and variable scale (an exponential distribution). The samples produced using this scheme follow a Laplace distribution, which can be shown by marginalization of $\boldsymbol{\gamma}_q$, i.e., $\int \mathrm{p}(\mathbf{c}_q|\boldsymbol{\gamma}_q)\mathrm{p}(\boldsymbol{\gamma}_q|\lambda_q)d\boldsymbol{\gamma}_q \sim \mathrm{Laplace}(\mathbf{c}_q|\lambda_q)$.

In order to estimate the whole set of unknowns, $\boldsymbol{\Theta}$, that includes the noise precision, the color-vector matrix and the coefficient matrix along with the corresponding hyperparameters, $\boldsymbol{\Theta} = \{\beta, \mathbf{M}, \mathbf{C}, \boldsymbol{\Gamma}, \boldsymbol{\lambda}\}$ with $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma}_q\}_{\mathrm{q}=1}^{Q}$ and $\boldsymbol{\lambda} = \{\lambda_q\}_{\mathrm{q}=1}^{Q}$, we make use of Bayesian inference. The exact calculation of the true posterior $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y}) = \mathrm{p}(\mathbf{Y}, \boldsymbol{\Theta})/\mathrm{p}(\mathbf{Y})$, with joint distribution

$$\mathrm{p}(\mathbf{Y}, \boldsymbol{\Theta}) = \mathrm{p}(\mathbf{Y}|\beta, \mathbf{M}, \mathbf{C})\mathrm{p}(\beta)\mathrm{p}(\mathbf{M})\mathrm{p}(\mathbf{C}|\boldsymbol{\Gamma})\mathrm{p}(\boldsymbol{\Gamma}|\boldsymbol{\lambda}), \tag{4}$$

cannot be done analytically since it requires the marginal $\mathrm{p}(\mathbf{Y}) = \int \mathrm{p}(\mathbf{Y}, \boldsymbol{\Theta})d\boldsymbol{\Theta}$ which is intractable. We use variational inference to approximate the true posterior, which requires the assumption of simplifications on the form of the posterior. These simplifications should render the inference tractable, while at the same time ensure that the model is flexible enough to closely approximate the true posterior distribution. Concretely, we will assume that our approximate posterior $\mathbf{q}(\boldsymbol{\Theta})$ factorizes as $\mathrm{q}(\boldsymbol{\Theta}) = \mathrm{q}(\beta)\mathrm{q}(\boldsymbol{\Gamma})\,\mathrm{q}(\boldsymbol{\lambda})\,\mathrm{q}(\mathbf{C})\prod_s \mathrm{q}(\mathbf{m}_s)$, which is referred to as mean-field factorization in the literature, see (Bishop, 2006). Notice, however, that we do not make any assumption on the individual distributions of each random variable; this will be determined by the inference procedure. The optimal solution is found by minimizing the Kullback-Leibler divergence between the approximate $\mathrm{q}(\boldsymbol{\Theta})$ and true posterior $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y})$. This optimization has a well-known optimum given by

$$\log \mathrm{q}(\boldsymbol{\theta}_i) = \langle \log\, \mathrm{p}(\mathbf{Y}, \boldsymbol{\Theta})\rangle_{\boldsymbol{\Theta}\setminus\boldsymbol{\theta}_i} + \text{const.}, \tag{5}$$

where $\langle\cdot\rangle_{\boldsymbol{\Theta}\setminus\boldsymbol{\theta}_i}$ denotes the expectation taken w.r.t. all approximating variables $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}$, with $j \neq i$. In the case of degenerate distributions $\mathrm{q}(\boldsymbol{\theta}_i)$, this calculation simplifies to finding the maximum w.r.t. $\boldsymbol{\theta}_i$ of the same expectation as in (5). Notice that this implies that we will not find a proper distribution for $\boldsymbol{\theta}_i$, but only its mean with zero variance. We will assume degenerate posterior distributions on $\mathbf{M}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\lambda}$, which will simplify the calculation of the expectations w.r.t. these random variables since $\langle f(\boldsymbol{\theta}_i)\rangle_{\boldsymbol{\theta}_i} = f(\hat{\boldsymbol{\theta}}_i)$, where $\hat{\boldsymbol{\theta}}_i := \langle\boldsymbol{\theta}_i\rangle_{\boldsymbol{\theta}_i}$. In contrast, we will obtain full distributions for the noise precision and the stain concentration sparse vectors.

After careful derivation using (5) on $\mathbf{C}$, (Serra et al., 2017) for the details, we find that each $\mathbf{c}_q$ follows a Gaussian distribution with mean and covariance matrix given by

$$\hat{\mathbf{c}}_q = \hat{\beta}\boldsymbol{\Sigma}_{\mathbf{c}_q}\hat{\mathbf{M}}^{\mathrm{T}}\mathbf{y}_q, \tag{6}$$

$$\boldsymbol{\Sigma}_{\mathbf{c}_q} = \left(\hat{\beta}\hat{\mathbf{M}}^{\mathrm{T}}\hat{\mathbf{M}} + \hat{\boldsymbol{\Gamma}}_q^{-1}\right)^{-1}. \tag{7}$$

We can now find the optimal estimations for the associated hyperparameters of the hierarchical prior $\boldsymbol{\gamma}_q$ and λ_q by maximization of the right-hand side of (5) as described above, obtaining

$$\hat{\gamma}_{sq} = -\frac{1}{2\hat{\lambda}_q} + \sqrt{\frac{1}{4\hat{\lambda}_q^2} + \frac{\hat{c}_{sq}^2 + \boldsymbol{\Sigma}_{\mathbf{c}_q}(s,s)}{\hat{\lambda}_q}}, \tag{8}$$

$$\hat{\lambda}_q = \frac{2N_s}{\sum_{s=1}^{N_s}\hat{\gamma}_{sq}}. \tag{9}$$

It is interesting to study the effect of the uncertainty on the estimates of c_{sq} given by $\mathbf{\Sigma}_{\mathbf{c}_q}(s,s)$ in (8). As our uncertainty in the estimation grows, so will γ_{sq}, which models the variance of c_{sq}, and, therefore, it will increase the uncertainty on this parameter.

The optimal $\mathbf{m}_s \in \mathbf{M}$ can be found assuming column independence and degenerate approximate posteriors $\mathrm{q}(\mathbf{m}_s)$ on a point on the unit sphere, $\|\mathbf{m}_s\| = 1$. Following the inference procedure described above, we have

$$\hat{\mathbf{m}}_s \propto \left[\mathbf{Y} - \sum_{i \neq s} \hat{\mathbf{m}}_i \hat{\mathbf{c}}_{i,:}\right] \hat{\mathbf{c}}_{s,:}^{\mathrm{T}} - \sum_{i \neq s} \sum_{q} \sigma_{isq} \hat{\mathbf{m}}_i, \tag{10}$$

where σ_{isq} denotes $\mathbf{\Sigma}_{\mathbf{c}_q}(i,s)$ and defines the influence of the uncertainty of the estimation of the coefficient vectors. The actual estimate of $\mathbf{m}_s$ is obtained by normalizing (10).

Finally, applying (5) for β results in a gamma-distributed posterior with mean given by

$$\hat{\beta} = \frac{3Q + 2a^{\beta}}{\|\mathbf{Y} - \hat{\mathbf{M}}\hat{\mathbf{C}}\|_{\mathrm{F}}^2 + \sum_{q=1}^{Q} \mathrm{tr}(\hat{\mathbf{M}}^{\mathrm{T}}\hat{\mathbf{M}}\mathbf{\Sigma}_{\mathbf{c}_q}) + 2b^{\beta}}\,. \tag{11}$$

Once more, note here how the uncertainty in the estimation of the coefficient vectors $\mathbf{c}_q$ given by $\mathbf{\Sigma}_{\mathbf{c}_q}$ impacts the estimation of the noise precision, resulting in lower precision (higher variance) when this uncertainties grow.

The procedure to obtain the estimated $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ using the above presented modelling and inference is summarized in **Algorithm 1**.

Algorithm 1 Pseudocode for BKSVD BCD algorithm

Input: Observed image $\mathbf{I}$, initial normalized $\underline{\mathbf{M}}$, no. stains N_s.
Output: Estimated stain color-vector matrix, $\hat{\mathbf{M}}$, and concentrations, $\hat{\mathbf{C}}$,

Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$ and set $\hat{\mathbf{m}}_s = \underline{\mathbf{m}}_s$, $\mathbf{\Sigma}_{\mathbf{c}_s} = \mathbf{0}$, $\hat{\mathbf{C}} = \underline{\mathbf{M}}^{+}\mathbf{Y}$, with $\underline{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$ and $\mathbf{\Gamma} = \mathbf{1}$
while $\hat{\mathbf{C}}$ has not converged **do**
  for q in $1, \ldots, Q$ **do**
    Update $\hat{\lambda}_q$ using (9)
    Update $\hat{\gamma}_{sq}$ using (8), for all s in $1, \ldots, N_s$
    Update $\mathbf{\Sigma}_{\mathbf{c}_q}$ and $\hat{\mathbf{c}}_q$ using (7) and (6), respectively
  end for
  for s in $1, \ldots, N_s$ **do**
    Update $\hat{\mathbf{m}}_s$ using (10)
  end for
  Update $\hat{\beta}$ using (11)
end while
return $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$

4. Sequential Inference for Sparse Bayesian Models

The previous section introduced a mathematically sound inference procedure. However, the sparse values in the columns $\mathbf{c}_q$ are not guaranteed to be zero. Since most of the pixels

in the image should be stained only by one stain, higher sparsity is desired. To increase the sparsity of the obtained solution we use Empirical Bayes (Tipping and Faul, 2003; Babacan et al., 2010; Serra et al., 2017) to obtain a new inference procedure. This approach was first presented in (Tipping and Faul, 2003) for Sparse Bayesian Learning (SBL) and later in (Babacan et al., 2010) and (Serra et al., 2017) for recovery of sparse signals. In this paper, we introduce the necessary adaptation for the application to histological blind color deconvolution.

In particular, for each $\mathbf{c}_q$, we use a constructive approach for identifying the locations where it takes non-zero values, i.e., its support. At these non-zero locations, we use Maximum *A Posteriori* (MAP) estimation to obtain the values of the hyperparameters. Therefore, sparsity makes the effective problem dimensions to be drastically reduced. The estimated values of the columns $\mathbf{c}_q$ in its support are obtained using (6).

The main idea behind this inference scheme consists on replacing the variational inference of hyperparameters $\boldsymbol{\gamma}_q$ with direct maximization of the (log) marginal likelihood

$$\mathcal{L}(\boldsymbol{\gamma}_q) = \log\left[\mathrm{p}(\boldsymbol{\gamma}_q|\hat{\lambda}_q)\int \mathrm{p}(\mathbf{y}_q|\mathbf{c}_q,\hat{\beta})\mathrm{p}(\mathbf{c}_q|\boldsymbol{\gamma}_q)d\mathbf{c}_q\right], \tag{12}$$

where $\mathrm{p}(\mathbf{y}_q|\mathbf{c}_q,\hat{\beta}) = \mathcal{N}(\mathbf{y}_q|\hat{\mathbf{M}}\mathbf{c}_q,\hat{\beta}^{-1}\mathbf{I})$, following the observation model, and $\hat{\mathbf{M}}$, $\hat{\beta}$ and $\hat{\lambda}_q$ are estimated as shown in Sec. 3. The marginal likelihood $\mathcal{L}(\boldsymbol{\gamma}_q)$ has interesting properties that allow for a highly efficient maximization thereof. Concretely, its functional form allows us to separate the contribution of a single γ_{sq} so that $\mathcal{L}(\boldsymbol{\gamma}_q) = \mathcal{L}(\{\gamma_{iq}\}_{i\neq s}) + l(\gamma_{sq})$. A closed form solution of the maximization of $\mathcal{L}(\boldsymbol{\gamma}_q)$, when only its s-th component is changed, can be found by holding the other hyperparameters fixed, taking its derivative with respect to γ_{sq} and setting it equal to zero. Note that this derivative will be different from zero only for $l(\gamma_{sq})$. Analysis of $l(\gamma_{sq})$ (see Appendix A) shows that the marginal likelihood has a unique maximum w.r.t. γ_{sq} and allows us to efficiently estimate the increase in log-likelihood that changing this parameter will introduce.

The Empirical Bayesian K-SVD (EBKSVD) in **Algorithm 2** is initialized by including only one color vector, the one that produces the highest increase in log-likelihood, and the corresponding γ_{sq}; the remaining $\{\gamma_{iq}\}_{i\neq s}$ are set to 0. At each iteration of the algorithm we will be able to add a new color vector, and its corresponding γ_{sq}, to our current model if the previous value of the γ_{sq} that produces the greatest increase of $\mathcal{L}(\boldsymbol{\gamma}_q)$ was zero; we will remove the element from the model if the optimal value of γ_{sq} is 0; or, finally, reestimate (update) γ_{sq} if $\mathbf{m}_s$ was already part of the model. In all three cases we are able to make incremental changes to the model structure while guaranteeing an increase of log-likelihood. Finally, the updates for $\mathbf{c}_q$ and $\boldsymbol{\Sigma}_q$ will be done using only the γ_{sq} included in the model, which will guarantee the sparsity of this inference method. See details in Appendix A.

To conclude this section, let us briefly compare the variational and empirical approaches. As previously discussed, the variational inference in section 3 achieves a softer sparsity, where the concentrations will include residual non-zero values. The combination of residual and non-sparse values might influence the final estimation of $\hat{\mathbf{M}}$. The empirical approach reduces this effect by calculating only the values where $\mathbf{c}_q$ takes non-zero values. Empirical Bayes is usually used to reduce the computational burden of Bayesian methods, as the calculation of the covariance matrix in (7) require to calculate the inverse matrix at each step and might be expensive for big matrices. Note that this is not the case for BCD of histological images,

Algorithm 2 Pseudocode for Empirical BKSVD BCD algorithm

Input: Observed image $\mathbf{I}$, initial normalized $\underline{\mathbf{M}}$, no. stains N_s.
Output: Estimated stain color-vector matrix, $\hat{\mathbf{M}}$, and concentrations, $\hat{\mathbf{C}}$,

Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$ and set $\hat{\mathbf{m}}_s = \underline{\mathbf{m}_s}$, $\boldsymbol{\Sigma}_{\mathbf{c}_s} = \mathbf{0}$, $\hat{\mathbf{C}} = \underline{\mathbf{M}}^{+}\mathbf{Y}$, with $\underline{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$, $\boldsymbol{\Gamma} = \mathbf{0}$, and $\boldsymbol{\lambda} = \mathbf{0}$
while $\hat{\mathbf{C}}$ has not converged **do**
  for q in $1, \ldots, Q$ **do**
    Choose a $s \in \{1, \ldots, n_s\}$ (or equivalently choose a γ_{sq})
    Find the optimal value of $\hat{\gamma}_{sq}$ using (A.9)
    Update $\boldsymbol{\Sigma}_{\mathbf{c}_q}$ and $\hat{\mathbf{c}}_q$ using (7) and (6), respectively
    Update g_{sq} and h_{sq} using (A.12) and (A.14), respectively, for all s in $1, \ldots, N_s$
    Update $\hat{\lambda}_q$ using (9)
  end for
  for s in $1, \ldots, N_s$ **do**
    Update $\hat{\mathbf{m}}_S$ using (10)
  end for
  Update $\hat{\beta}$ using (11)
end while
return $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$

where the number of stains is usually $N_s = 2$ and the inversion of 2×2 matrices is not costly. Although the computational saving is reduced, the additional sparsity induced by the empirical method is useful in the estimation of $\mathbf{M}$, as we will make clear in the following sections.

5. Application of Bayesian K-SVD for WSI analysis

Bayesian methods are usually computationally expensive as they require to take into account the uncertainties of the coefficients at each element in the image. While previous applications as denoising or inpainting (Serra et al., 2017) were carried out on small 256×256 grayscale images (64Kpixels), its application to blind color deconvolution problem is hindered by the massive size of WSI images. WSIs are RGB images in the Gigapixels order which makes their processing challenging. Therefore, it is extremely necessary to introduce additional adaptations that make the BKSVD and EBKSVD more suitable for WSI images.

First, during training and reconstruction of the histological images, the highest computational cost is the computation of the sparse representation of the concentrations for each pixel. However, the reduced amount of stains suggests that it is not required to use all WSI pixels to learn $\hat{\mathbf{M}}$. We here reformulate (1) as

$$\mathbf{Y}_B = \mathbf{M}\mathbf{C}_B + \mathbf{N}_B\,, \tag{13}$$

where $\mathbf{Y}_B$ is a representative subset of the pixels in $\mathbf{Y}$ and $\mathbf{C}_B$ its associated concentration matrix. To find the representative set of pixels, we first look at those that can be discarded. Large background areas are typically removed upon patching for most applications. However background pixels can also appear on lumens or tissue borders. Since those low stained

pixels do not provide information on the stain's color, following Vahadane et al. (2016), we can remove them for the estimation of $\mathbf{M}$. The optical density of those pixels is close to zero making it easy to filter them. The removal of low stained pixels accelerates the procedure of estimating $\hat{\mathbf{M}}$ and eliminates the influence of background pixels.

Despite considering only tissue pixels, usually there are still too many pixels for practical application of the algorithms. WSI images often include several resolutions. While using the smaller images obtained at lower magnifications could be tempting, we should avoid them in the estimation of $\mathbf{M}$. Pixels values at lower resolutions, when interpolated linearly, are a weighted average of a set of pixels at a higher resolution. Note, however, that this average takes place in the RGB space. Then, obtaining the OD image requires the use of the non-linear logarithmic transformation. As the logarithm is a concave function, for a single pixel at a lower resolution, we have for non-negative weights $\{\tau_q\}$ in a neighborhood that add up to one

$$\mathbf{y} = -\log\left(\sum_q \tau_q \frac{\mathbf{i}_q}{i^0}\right) \leq -\sum_q \tau_q \log\left(\frac{\mathbf{i}_q}{i^0}\right) = -\sum_q \tau_q(\mathbf{M}\mathbf{c}_q) \tag{14}$$

where $\mathbf{i}_q = [\mathbf{i}_1, ..., \mathbf{i}_N]$ are the high resolution pixels contributing to the averaged pixel. Although the linearity in the RGB space is not preserved in OD, we can expect the assumption that most pixels are stained by a single stain to be less satisfied as resolution decreases. Therefore, as the proposed methods are based on the sparsity assumption, it is preferred to extract a subset of pixels from the WSI at the higher magnification available, typically $40\times$.

Therefore we need find another method to reduce the amount of pixels to be considered. Patching is the most common way of dealing with the massive size of WSIs during preprocessing or classification. This approach allows to take into account local tissue structures, which are important for WSI interpretation. However, it is not a suitable solution for obtaining $\hat{\mathbf{M}}$ as local tissue structures may not correctly represent both stains. Note that the proposed framework assumes that each pixel stain's concentrations are independent, thus eliminating spatial constrains. This modelling allows us to select individual pixels in the image, independently of their neighbours. Therefore, we can obtain a representative subset $\mathbf{Y}_B$ within the image using an uniform random sampling of the stained pixels. This allows, on the one side, to accurately sample the whole WSI the image and, on the other, dramatically to reduce the number of pixels used to estimate the stain-color matrix $\mathbf{M}$.

For a given subset $\mathbf{Y}_B$, the color vector matrix $\hat{\mathbf{M}}$ can be estimated using BKSVD or EBKSVD in Alg. 1 and Alg. 2, respectively. To avoid overfitting to a given subset $\mathbf{Y}_B$, once the chosen method converges, a new batch of pixels $\mathbf{Y}_B$ is selected and the estimation procedure is repeated until the matrix $\hat{\mathbf{M}}$ converges. Notice that we do not use complete epochs as our objective is to ensure that the obtained $\hat{\mathbf{M}}$ faithfully represents the colors in the WSI without using all pixels in the image.

Once the color vectors of the image are estimated using $\mathbf{Y}_B$, we still need to obtain the stain concentrations $\mathbf{C}$ for the whole image. We could consider to execute Alg. 1 or Alg. 2 for the whole image keeping $\hat{\mathbf{M}}$ fixed. However, this still requires to iterate in order to estimate the model parameters and concentrations at each pixel, which is time prohibitive for the whole image. Then, assuming that $\hat{\mathbf{M}}$ is an accurate estimation of $\mathbf{M}$, the final values of the concentrations, $\hat{\mathbf{C}}$, for the whole image will be computed as $\hat{\mathbf{C}} = \hat{\mathbf{M}}^{+}\mathbf{Y}$ (Ruifrok and Johnston, 2001; Alsubaie et al., 2017), with $\hat{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\hat{\mathbf{M}}$.

Note that, for a fixed $\hat{\mathbf{M}}$, this is also the minimum squared error estimator of $\mathbf{C}$ from (1).

Finally, the described multibatch procedure is summarized in **Algorithm 3**.

Algorithm 3 Multibatch Bayesian KSVD

Input: Observed image $\mathbf{I}$, initial normalized $\underline{\mathbf{M}}$, no. stains N_s, batch size B.
Output: Estimated stain color-vector matrix, $\hat{\mathbf{M}}$, and concentrations, $\hat{\mathbf{C}}$,

1: Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$
2: Remove low stained pixels from $\mathbf{Y}$
3: **while** $\hat{\mathbf{M}}$ has not converged **do**
4: Sample a batch $\mathbf{Y}_B$ of B stained pixels from $\mathbf{Y}$.
5: Estimate $\hat{\mathbf{M}}$ using BKSVD or EBKSVD
6: **end while**
7: **return** $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}} = \hat{\mathbf{M}}^{+}\mathbf{Y}$

6. Materials and methods

To assess its quality, the proposed BKSVD and EBKSVD were compared to the following methods frequently used in the literature: the classical non-blind CD method by Ruifrok and Johnston (2001) and the BCD methods by Macenko et al. (2009), Vahadane et al. (2016), Alsubaie et al. (2017), Hidalgo-Gavira et al. (2020), Pérez-Bueno et al. (2020), and Zheng et al. (2019). They will be denoted by RUI, MAC, VAH, ALS, HID, PER, and ZHE, respectively. All experiments in the following sections were conducted using the multibatch Bayesian K-SVD[2] in Alg. 3 with $N_s = 2$. As initial color-vector matrix, we used the standard H&E vectors proposed by Ruifrok and Johnston (2001). The proposed method was run until the criterion $\|\mathbf{M}^{(n)} - \mathbf{M}^{(n-1)}\|_F^2 < 5 \times 10^{-3}$ was met. Algorithms 1 and 2 were run until the criterion $\| \langle \mathbf{c}_s \rangle^{(n)} - \langle \mathbf{c}_s \rangle^{(n-1)} \|^2 / \| \langle \mathbf{c}_s \rangle^{(n)} \|^2 < 10^{-4}$ was met by both stains. All model parameters are automatically estimated. Using the obtained $\hat{\mathbf{M}}$ and $\hat{\mathbf{C}}$ it is possible to perform CN and SCA. Further details are provided in the following experimental section.

To test the performance and robustness of our algorithm in different scenarios related to digital histopathology (i.e., stain separation quality, color normalization, and stain color augmentation for cancer classification), we have selected data containing a variety of histopathological images from several types of tissue and laboratories. In this section we describe the details of the databases used in this paper.

6.1. Warwick Stain Separation Benchmark (WSSB)

WSSB dataset (Alsubaie et al., 2017) contains 24 H&E stained images of different tissues (breast, colon, and lung) from different laboratories which have been captured with different microscopes. For each image, its ground truth stain color-vector matrix, $\mathbf{M}_{GT}$, was manually obtained by expert pathologists as follows. The experts selected a set of pixels for each stain, based on biological structures: nuclei for hematoxylin and cytoplasm for eosin. Then, the median value of each set of pixels with a single stain was used as a measure of the

[2]The code used in the experiments will be made available at https://github.com/vipgugr upon acceptance of the paper.

Table 1: CAMELYON17 dataset labeling structure

		Stage label			
Subset	WSI total	Negative	ITC	Micro	Macro
Whole training set	500	318	36	59	87
Annotated	50	0	16	17	17
Not annotated	450	318	20	42	70

corresponding stain color-vector. Ground truth concentrations were obtained in Alsubaie et al. (2017) from the ground-truth color-vector matrix as

$$\mathbf{C}_{GT} = \mathbf{M}_{GT}^{+}\mathbf{Y}. \tag{15}$$

From those ground-truth concentrations and color-vectors, a separate RGB image for each stain is obtained. This database will be used for BCD evaluation.

6.2. CAMELYON17

This database is part of the CAMELYON17 challenge (Bándi et al., 2019) for breast cancer metastasis detection in the lymph node sections. We will use it in CN and classification experiments including the use of SCA.

CAMELYON17 contains a total of 1000 WSIs from 5 medical centers. Only the training set, which contains 500 WSIs, was used since the annotations for the test WSIs are not available yet. The dataset comprises 20 patients per center and 5 slides per patient. Cancer regions were annotated by pathologists only on 50 WSIs, but the stage label: negative, isolated tumor cells (ITC), micrometastasis (Micro), macrometastasis (Macro), is available for all the slides in the training set. See Table 1 for details.

Following Zheng et al. (2019) the experiments on this dataset were performed using non-overlapping 224×224 pixel patches, with at least a 70% of tissue, sampled from each WSI.

7. Experimental results

We have carried out a set of experiments to evaluate the performance of the proposed framework on the most common histological color deconvolution related tasks: stain separation, image normalization, and CNN-based classification, where we include the use of SCA.

First, we evaluate the influence of the pixel batch size on the proposed methods. Then we assess the quality of the concentration and color-vector matrices obtained by the BCD algorithms. In a third experiment, we analyze the quality of the CN obtained by the algorithms when the color-vectors are substituted by those of a reference-image, keeping the concentration values. Finally, the deconvolved, normalized, and SCA images are evaluated on a histological classification scenario.

7.1. Influence of the batch size in the color vector estimation

The use of pixel sampling introduced in Section 5 requires to assess the influence of the pixel batch size on the similarity of the obtained color vector matrix $\hat{\mathbf{M}}_P$ (obtained using P

Table 2: Mean values required to estimate $\hat{\mathbf{M}}_P$ and $\hat{\mathbf{M}}_{all}$ using Alg. 3

BKSVD	batch size in pixels										full images		
	50	100	300	500	1000	2000	4000	10^4	$2{\cdot}10^4$	10^5	$2{\cdot}10^5$	$4{\cdot}10^6$	$16{\cdot}10^6$
no. batches	9.8	10	7.3	7.2	5.3	4.7	4.2	4.1	4.3	4.3	1	1	1
no. total iter.	97.34	80.07	43.87	41.53	31.20	29.07	27.87	27.34	27.6	27.67	18	13	17
time/iter. (s)	0.09	0.11	0.14	0.15	0.17	0.20	0.24	0.34	0.52	1.88	2.78	34.08	268.80
total time	8.68	8.86	6.00	6.05	5.37	5.86	6.8	9.31	14.47	52.09	50.06	443.06	4569.60
EBKSVD	batch size in pixels										full images		
	50	100	300	500	1000	2000	4000	10^4	$2{\cdot}10^4$	10^5	$2{\cdot}10^5$	$4{\cdot}10^6$	$16{\cdot}10^6$
no. batches	9.3	8.7	7.8	8	5.3	5.5	5.4	5.2	4.8	3.7	1	1	1
no. total iter.	154.47	135.27	109.33	98.73	85.00	79.13	76.2	67.8	49.4	43.00	34	8	15
time/iter. (s)	0.10	0.13	0.17	0.21	0.27	0.43	0.77	2.02	3.99	19.75	38.23	431.01	2338.30
total time	15.95	18.03	18.65	20.33	22.83	33.91	58.78	136.99	197.27	849.37	1300.15	3448.11	35074.50

pixels) to the $\hat{\mathbf{M}}_{all}$ obtained using all non-white pixels and the execution time required for the estimation. Unfortunately, it is not possible to use complete WSIs in this experiment due to the computational burden, therefore we use three different images of typical sizes 500×500, 2000×2000 and 4000×4000 pixels and batch sizes from 50 to $1.6 \cdot 10^7$ pixels. Algorithm 3, using both BKSVD and EBKSVD, was run 5 times for each different batch size up to $2 \cdot 10^4$ pixels and only once for bigger ones.

Table 2 summarizes the mean number of batches, iterations, time per iteration and total time required by Alg. 3 on the three images tested when a different batch sizes are used. Analogous figures for Alg. 1 and Alg. 2 using the whole image are also reported. For both BKSVD and EBKSVD, the number of batches and the total number of iterations required to estimate $\hat{\mathbf{M}}_P$ decrease with the size of the batch P. The time per iteration grows with P reaching unaffordable values for higher values of P, which supports the idea of working with smaller batches. EBKSVD consumes more time, both with a larger number of iterations required to converge and a higher time per iteration. The times required by Alg. 1 and Alg. 2 are usually higher than those needed by Alg. 3, even when a large batch size is used. Although the tested images are far from the Gigapixel size of a WSI, the total time required to estimate the color vector matrix using the full images shows the importance of the adaptation introduced in section 5 for the use of Bayesian methods on the BCD problem for histological images.

Furthermore, the comparison plotted in Figure 1, depicting the time and convergence ratio for the different images and batch size, shows that the execution time grows linearly with P while the difference between $\hat{\mathbf{M}}_P$ and $\hat{\mathbf{M}}_{all}$ quickly converges to zero. Note that using only a batch size of 50 pixels we achieve a difference in norm of less than 0.05 in most cases. The EBKSVD method, plotted in dashed lines, requires a lower amount of pixels to converge but also requires more time since it needs to find the location of non-zero elements in each step. Using a batch size of 1000 pixels ensures an accurate estimation, with low variance and an affordable computational burden. Note that the time needed by EBKSVD grows significantly faster for batch sizes above 1000 pixels. The BKSVD method is significantly faster but requires more pixels to reach the same output as using the whole image. According to the three images tested and the above mentioned criteria, a batch size of 4000 pixels is the best choice for this method.

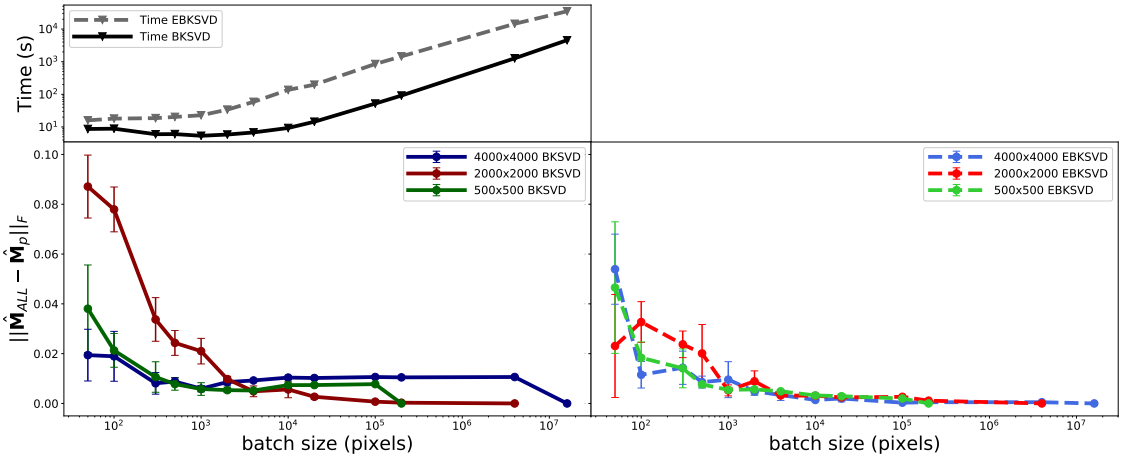

Figure 1: **Top:** Mean time required to obtain the dictionary for the proposed methods. **Bottom:** Difference between the obtained dictionary $\hat{\mathbf{M}}_P$ with a batch size P and the dictionary $\hat{\mathbf{M}}_{all}$ obtained using all pixels for the BKSVD (left) and EBKSVD (right) methods.

Note that both EBKSVD and BKSVD provide an accurate estimation with a batch size of 1000 and 4000 pixels, respectively, for all the image sizes tested. The results plotted in Fig. 1 suggest that these batch sizes will also provide an accurate estimation of $\mathbf{M}$ for larger images in a similar time, making Alg. 3 an scalable solution for obtaining $\hat{\mathbf{M}}$ in WSIs. As a consequence, for the rest of the experiments in this paper, the batch size was fixed to 1000 pixels for EBKSVD and 4000 pixels for BKSVD.

7.2. BCD Stain Separation

Table 3: PSNR and SSIM for the different methods on the WSSB dataset (Alsubaie et al., 2017).

PSNR Image	Stain	RUI	MAC	VAH	ALS	HID	PER	ZHE	EBKSVD	BKSVD
Colon	H	22.27	23.91	25.83	21.11	28.57	28.62	17.89	32.12	**34.08**
	E	20.70	21.55	26.29	21.94	27.58	27.60	14.76	31.11	**33.32**
Breast	H	15.27	26.24	25.46	24.60	28.81	29.14	15.31	31.69	**32.20**
	E	17.66	23.62	27.68	25.92	26.60	26.76	14.99	28.81	**29.43**
Lung	H	22.47	19.52	25.87	20.62	32.91	**33.10**	19.51	33.06	32.67
	E	22.05	18.09	25.53	23.95	30.77	31.02	16.23	**31.87**	30.61
Mean	H	20.00	23.22	25.72	22.11	30.10	30.29	17.57	32.29	**32.98**
	E	20.14	21.08	26.50	23.94	28.32	28.46	15.33	30.60	**31.12**
SSIM Image	Stain	RUI	MAC	VAH	ALS	HID	PER	ZHE	EBKSVD	BKSVD
Colon	H	0.8141	0.8095	0.8851	0.7241	0.9542	0.9544	0.7894	0.9733	**0.9826**
	E	0.7456	0.6365	0.8904	0.8540	0.9139	0.9161	0.4625	0.9422	**0.9646**
Breast	H	0.6215	0.9552	0.9239	0.8068	0.9528	0.9560	0.6488	**0.9845**	0.9801
	E	0.7644	0.9336	0.9550	0.9380	0.9464	0.9492	0.7150	**0.9717**	0.9632
Lung	H	0.7987	0.7389	0.8912	0.5551	0.9763	0.9757	0.8116	0.9759	**0.9764**
	E	0.7734	0.5088	0.8195	0.8939	0.9306	0.9353	0.5390	**0.9670**	0.9461
Mean	H	0.7448	0.8345	0.9100	0.6953	0.9611	0.9621	0.7500	0.9779	**0.9797**
	E	0.7611	0.6930	0.8883	0.8953	0.9303	0.9336	0.5722	**0.9603**	0.9580

To evaluate the fidelity of the H&E separation obtained by the different BCD methods, we use the WSSB database (introduced in Section 6). A ground truth separation from WSSB

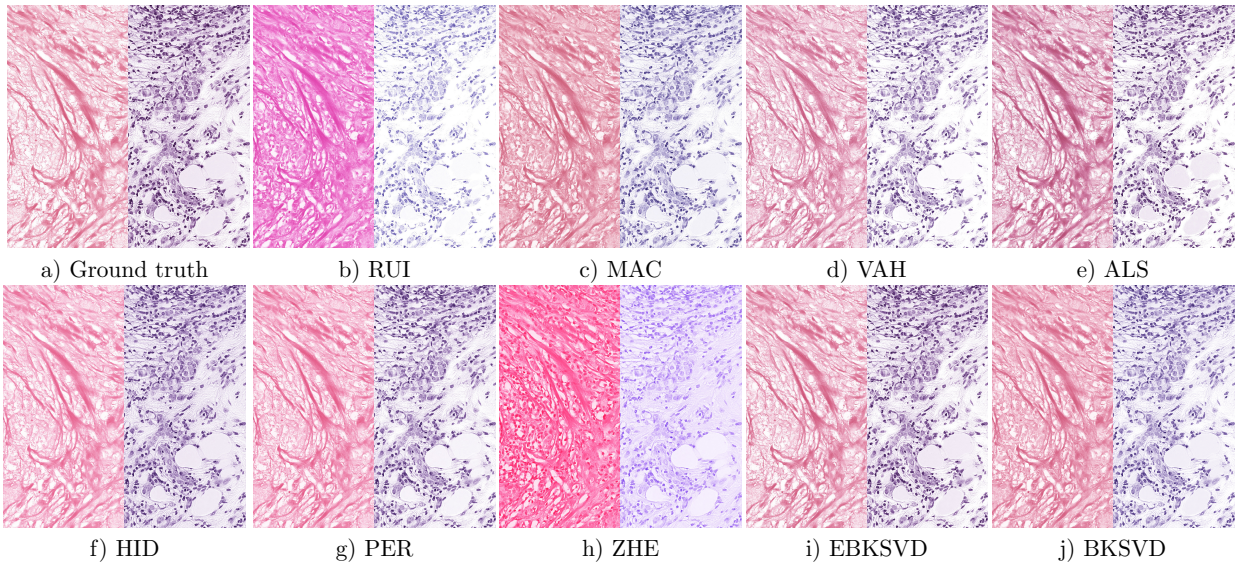

Figure 2: a) Ground truth separated E-only (left) and H-only (right) images from a Breast image of the WSSB dataset in Alsubaie et al. (2017) and results for the b)-h)competing and i-j) proposed methods.

is shown in Figure 2(a). Figures 2(b)–2(j) contain the separated images obtained by different BCD methods. RUI obtains highly contrasted images, but the fixed color vectors are far from those of the ground truth in Figure 2(a). Some nuclei are moved from the H to the E channel. MAC results are closer to the ground truth but the eosin channel still presents residual information from the nuclei. ALS creates artifacts in the flat zones of the H channel and over-saturates the colors. HID obtains colors slightly more saturated than the ground truth and smooths some details. ZHE colors seem unreal and it tends to mix the information of both channels with nuclei clearly appearing in the E channel and cytoplasm in the H channel. The proposed EBKSVD and BKSVD, VAH, as well as PER produce colors very similar to the ground truth separation in Figure 2(a). VAH obtains very similar colors with high differentiation between bands but some information is lost in the H channel, apparently moved to the E channel (see, for instance, the right side of the H channel and the center-left side of the E channel in Figure 2(d)). PER obtains a very good stain separation, although the E color is slightly more reddish than the ground truth. This is due to the prior on the color matrix. It imposes similarity to a reference color vector matrix manually selected for each tissue type. The proposed EBKSVD and BKSVD produce sharp edges, and automatically estimate the color vector matrix without manually selecting a reference. EBKSVD obtains a better mean estimation for the eosin and hematoxylin channels, while BKSVD obtains a slightly darker eosin and a bluish hematoxylin color. Both methods obtain richer details, and a stain separation closer to the ground truth than the competing methods.

The quantitative comparison, based on the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM), is presented in Table 3. The proposed BKSVD outperforms the rest of methods obtaining a higher mean PSNR (+2.69dB in H and +2.66dB in E) and a higher SSIM than the closest competitor (PER). The proposed EBKSVD obtains the second best mean performance just behind BKSVD, and is able to obtain better values for some tissue types (i.e., lung tissue). For SSIM, both BKSVD and EBKSVD methods are close and the best choice depends on the tissue type.

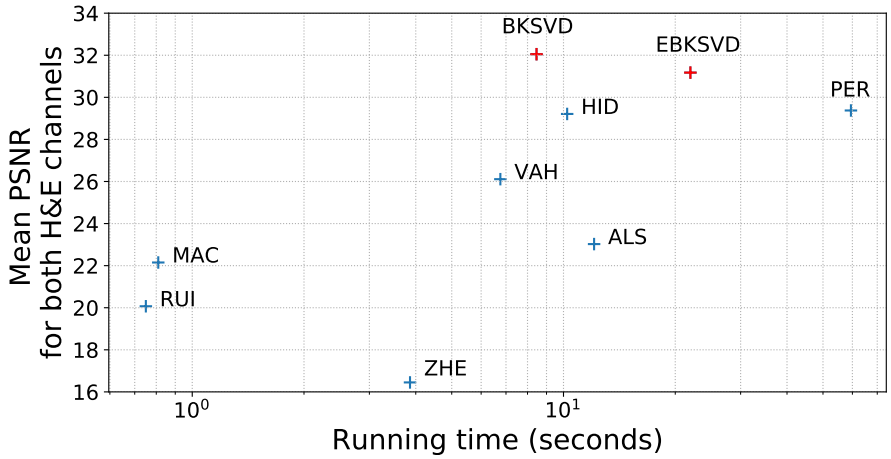

Figure 3: Mean PSNR in dB vs running time in seconds for deconvolving a 2000×2000 image.

The obtained results indicate that the proposed EBKSVD and BKSVD correctly separate the structural information in the image for all tested tissue types. BKSVD obtains the best estimation in mean, mainly due to its higher performance in the colon images. Since colon images are obtained at a lower magnification ($20\times$), this suggests that BKSVD performs better than EBKSVD when a lower magnification is used, that is when a lower sparsity is expected. This is consistent with the results obtained in Serra et al. (2017) where the performance of the EBKSVD is affected by a lower sparsity.

In both cases, the high quality stain separation obtained by the proposed methods guarantees the fidelity to the tissue in CN and SCA transformations detailed in the following sections.

7.2.1. Time Comparison

One important issue with BCD methods is that the required time to perform deconvolution needs to be low enough for practical use. Figure 3 shows the time needed by each BCD method vs. PSNR for the WSSB dataset. The RUI method is the fastest since no color estimation is performed. The computational time increases with the complexity of the method. The proposed BKSVD method outperforms the rest obtaining a significantly higher PSNR while requiring a similar time to HID and VAH methods. EBKSVD obtains the second highest mean PSNR but requires a higher computational time to obtain the sparser solution. Note that the proposed EBKSVD and BKSVD methods are scalable, requiring a similar time for larger images (see Section 7.1).

7.3. Color Normalization

This section compares the color distribution in the original data and the CN obtained by the competing methods. CN is the most extended procedure to deal with stain color variations because CNN based CAD systems usually work with the observed RGB image. CN aims to reduce the impact of color variations on those systems. With the use of BCD, CN can be easily achieved as an additional step, as the stain color information and stain concentrations are separated and can be modified independently. CN based on BCD ensures

Table 4: NMI values for the centers in CAMELYON17.

	Center 0		Center 1		Center 2		Center 3		Center 4		All centers	
Method	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV	SD	CV
Original	0.0403	0.0527	0.0464	0.0667	0.0574	0.0792	0.0601	0.0867	0.0377	0.0441	0.0774	0.1036
MAC	0.0474	0.0734	0.0585	0.1035	0.0855	0.1559	0.0812	0.1489	0.0577	0.0771	0.1032	0.1689
VAH	0.0535	0.0868	0.0658	0.1236	0.0929	0.1787	0.0818	0.1582	0.0638	0.0892	0.1058	0.1823
ALS	0.0512	0.0855	0.0632	0.1303	0.0641	0.1267	0.0841	0.1740	0.0554	0.0821	0.0993	0.1806
HID	0.0413	0.0637	0.0363	0.0576	0.0587	0.0868	0.0463	0.0718	0.0478	0.0636	0.0635	0.0948
PER	0.0405	0.0626	0.0359	0.0570	0.0561	0.0832	0.0454	0.0706	0.0471	0.0628	0.0629	0.0941
ZHE	0.0345	0.0434	0.0277	0.0365	0.0449	0.0608	0.0428	0.0566	0.0311	0.0375	0.0489	0.0632
EBKSVD	0.0243	0.0313	0.0331	0.0440	**0.0292**	**0.0379**	0.0327	0.0436	**0.0252**	**0.0323**	0.0320	0.0418
BKSVD	**0.0202**	**0.0258**	**0.0239**	**0.0317**	0.0304	0.0398	**0.0280**	**0.0372**	0.0258	0.0329	**0.0290**	**0.0378**

fidelity to the image structures, while reducing color variations. Following Vahadane et al. (2016) we normalize an input image to a reference image using

$$\hat{\mathbf{Y}}^{norm} = \sum_{s=1}^{n_s} \hat{\mathbf{m}}_s^{ref} \hat{\mathbf{c}}_{s,:}^{norm}, \tag{16}$$

where

$$\hat{\mathbf{c}}_{s,:}^{norm} = \hat{\mathbf{c}}_{s,:} \frac{P_{99}(\hat{\mathbf{c}}_{s,:}^{ref})}{P_{99}(\hat{\mathbf{c}}_{s,:})}, \tag{17}$$

and $\hat{\mathbf{m}}_s{}^{ref}$ and $\hat{\mathbf{c}}_s{}^{ref}$ are the color vectors and concentrations obtained from the reference image. $P_{99}(\cdot)$ represents the pseudo-maximum at 99%. Note that the color vectors $\hat{\mathbf{m}}_s$ are replaced by $\hat{\mathbf{m}}_s^{ref}$ corresponding to the reference image, and the dynamic range of $\hat{\mathbf{c}}_s$ is corrected to be the same as that of $\hat{\mathbf{c}}_s^{ref}$. Therefore, $\hat{\mathbf{Y}}^{norm}$ is the normalized OD image and the normalized RGB image is obtained as $\hat{\mathbf{I}}^{norm} = \exp(-\hat{\mathbf{Y}}^{norm})$.

To measure the quality of the CN, we used the normalized median intensity (NMI) (Basavanhally and Madabhushi, 2013), defined as

$$NMI(\mathbf{I}) = median(\mathbf{u})/P_{95}(\mathbf{u}), \tag{18}$$

where $\mathbf{I}$ denotes a WSI and $\mathbf{u}$ is a vector where each component u_i is the mean value of the R, G, and B channels at the ith pixel, (Bejnordi et al., 2016). The NMI value was obtained for each image in a given dataset, and the standard deviation (NMI SD) and coefficient of the variation (NMI CV), i.e., NMI SD divided by the mean, were used as metrics. Lower values of NMI SD and NMI CV indicate a more consistent normalization.

CN tests are carried out on the CAMELYON17 dataset, introduced in Sect. 6, which includes images from 5 different centers. Following Zheng et al. (2019), 500 patches of size 224×224 pixels were sampled from each WSI in the dataset for CN and classification purposes. To avoid the influence of large background regions, only patches with at least 70% tissue were considered. The patch size is motivated for its use in the classification experiments in section 7.4 and does not affect the measurement of the normalization quality.

The results of the proposed and competing CN algorithms for each center and the whole dataset are reported numerically in Table 4 and graphically in Figure 4 where the NMI information for each center and method is plotted as a violin plot. MAC, VAH and ALS transform the images in each center to a similar distribution, but with a larger inter and

intra-center variance than the original images' distribution. Bayesian methods HID and PER strongly reduce the intra-center differences, but are not able to completely reduce inter-center differences. They have a similar behavior as they share the same similarity prior on the color vector matrix. ZHE significantly reduces intra-center differences but does not completely eliminate inter-center variance. The proposed methods outperform all competitors. Figure 4 and Table 4 show that BKSVD obtains the most consistent normalization, with the lowest intra-center variance and the most similar median values for all the centers in the dataset. EBKSVD closely follows, obtaining the best values for two out of five centers, but with slightly more variation than BKSVD, as can be seen in Figure 4(h) and 4(i).

The CN results were also compared in terms of fidelity to the original observed image using PSNR and SSIM. Although it is important to keep the structure of the original image, notice that fidelity and CN could be conflicting goals as the best fidelity is obtained by not modifying the image. PSNR and SSIM values are shown in Table 5. ZHE obtains the highest fidelity, followed by the proposed BKSVD and EBKSVD. Except for ZHE, that was optimized for its use in CN, the results obtained by the other methods are consistent with those presented in Section 7.2. The better the fidelity to the H&E GT, the better the fidelity after CN. As previously discussed, our methods guarantee fidelity to the H&E bands separately. Since the CN in (16) modifies the concentration dynamic range, it will reduce the similarity to the original image (e.g. by increasing the contrast between stains) but will not have a negative impact on the stain structure and, hence, the PSNR and SSIM values are not heavily affected.

Table 5: PSNR and SSIM for the normalized CAMELYON17 dataset.

	MAC	VAH	ALS	HID	PER	ZHE	EBKSVD	BKSVD
PSNR	13.80	12.74	11.16	17.77	17.73	**22.20**	19.29	19.54
SSIM	0.7265	0.6490	0.3132	0.8617	0.8644	**0.9603**	0.8594	0.8735

For a visual qualitative analysis, we depict in Figure 5 a sample patch for each center and the corresponding CN by the different methods. The first row shows the reference image and some 224×224 patches extracted from it showing the variance within the reference image at the same scale as the other patches. The remaining rows show, in the first column, the patch to be normalized and the rest of the columns the CN result with different methods. We notice that MAC and VAH normalize the images but do not obtain colors similar to the reference. ALS introduces color artifacts in most of the patches. ZHE, which is trained to reduce NMI, obtains good figures, but tends to over-brighten the images to reduce NMI variation. The Bayesian methods HID and PER also obtain a consistent normalization, but in some cases, they tend to over-estimate the presence of hematoxylin. The proposed BKSVD and EBKSVD obtain the images most similar to the reference image for all centers, producing high quality results and minimizing the inter-center color variations while maintaining clear differences between both stains. The difference between both methods is difficult to appreciate in this figure. Only in the image of the second center (third row), where hematoxylin and eosin are difficult to differentiate, EBKSVD clearly separates them although it introduces some artifacts, while BKSVD and the other methods do not correctly identify the eosin.

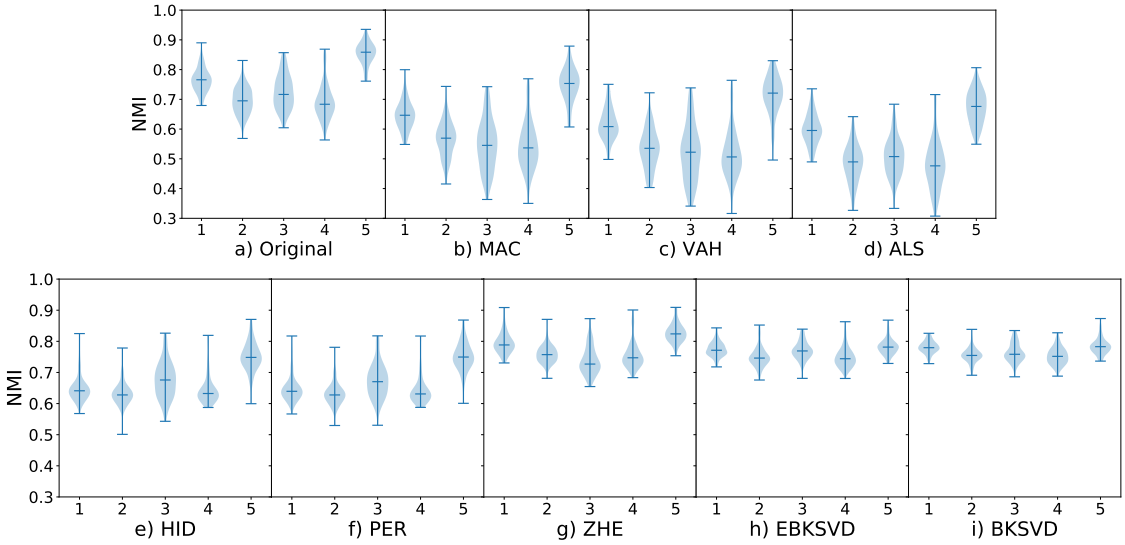

Figure 4: Violin plots of NMI values for each center in CAMELYON17. The blue shadow represents the histogram of NMIs for each plot, the maximum, median and minimum values for each plot are marked with bars. The x-axis indicates the center corresponding to a set of images.

7.4. Data augmentation and cancer classification

The main objective of BCD and CN is to improve the performance of CAD systems, usually based on patch classification systems (Esteban et al., 2019; Tellez et al., 2019). In this section we quantitatively assess the effect of BCD, CN and SCA on a breast cancer detection task (CAMELYON17). For that, we train a VGG19 (Simonyan and Zisserman, 2015) classifier, commonly used in cancer detection (Esteban et al., 2019), on the original, color normalized, and color augmented patches, both from RGB images and OD concentrations.

As previously discussed, using the original WSIs implies dealing with inter-center staining variations that produce generalization errors to unseen stain color variations. BCD and CN aim to reduce the generalization error by reducing color-variation in the input data. However, it is also possible to reduce the generalization error by simulating realistic variations of the training data. The SCA approach is a specific technique of data augmentation for histopathological images that produces realistic variations of the stain colors of the available data. As CN, SCA can also be obtained as an additional step after BCD. While Tellez et al. (2018) applies SCA on the concentrations obtained from Ruifrok and Johnston (2001), we propose to use a combination of both CN and SCA as to obtain an augmented OD image $\hat{\mathbf{Y}}^{aug}$ as follows:

$$\hat{\mathbf{Y}}^{aug} = \sum_{s=1}^{n_s} \hat{\mathbf{m}}_s^{ref} \hat{\mathbf{c}}_{s,:}^{aug}, \tag{19}$$

where the augmented concentrations $\hat{\mathbf{c}}_{s,:}^{aug}$ are synthesized as

$$\hat{\mathbf{c}}_{s,:}^{aug} = \alpha_s \hat{\mathbf{c}}_{s,:}^{norm} + \beta_s \cdot \mathbf{1}, \tag{20}$$

being $\hat{\mathbf{c}}_{s,:}^{norm}$ the normalized concentrations obtained using (17) and α_s, β_s random values following uniform distributions $U(1-\sigma, 1+\sigma)$ and $U(-\sigma, \sigma)$, respectively. This procedure leads to augmentation on the objective reference domain, allowing us to combine the advantages

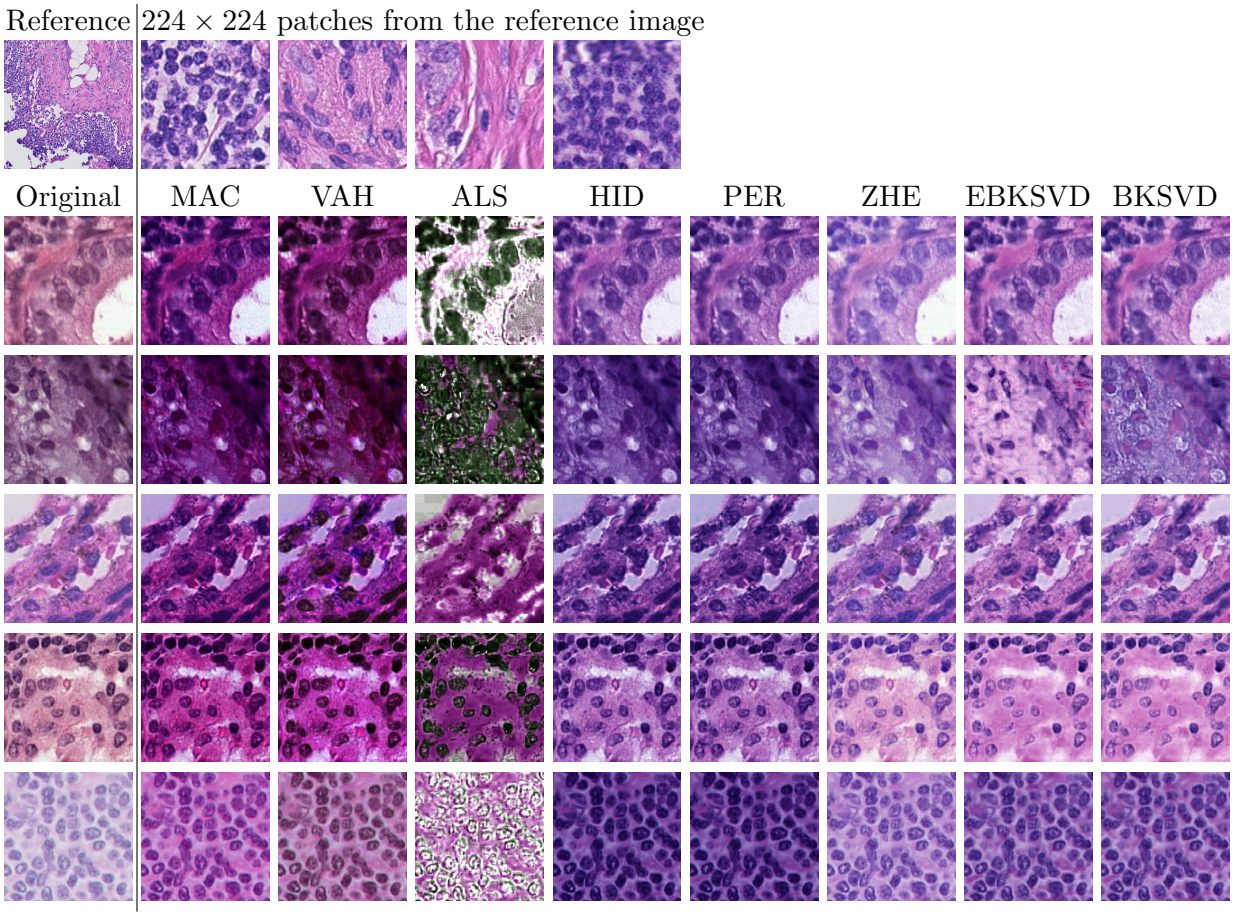

Figure 5: Example 224 × 224 patches from different centers in CAMELYON17. The first row shows the reference image and some 224 × 224 patches extracted from the reference. Rows 2–6 correspond to the different centers in CAMELYON17. The original patch is shown in the first column and the other columns show the CN results with different methods.

of both CN and SCA approaches. CN will reduce the variation between centers and SCA will cover the variations that were not completely captured by the CN.

We train the network with the RGB normalized images, OD concentrations obtained by the BCD methods, the SCA in Tellez et al. (2018), denoted by TEL, and the SCA using (19), denoted by BKSVD$_{aug}$ and EBKSVD$_{aug}$ depending on whether we use the BKSVD or EBKSVD concentrations. Following Tellez et al. (2018), $\sigma = 0.05$ and $\sigma = 0.2$ were used for light and strong augmentation, respectively.

From CAMELYON17, four centers were used for training and the 5th center, which showed a bigger color difference in the previous section, was used as test set. From the 50 tumor annotated WSIs in CAMELYON17, approximately 55.000 positive patches were sampled for training and 12.500 for testing. Negative patches were sampled from negative WSIs only, obtaining 55.000 for training and 12.500 for testing.

VGG19 was trained from scratch for 100 epochs in each case using 64 sample batches with batch normalization. The learning rate was initially set to 0.01, which is halved every 30 epochs. When using OD concentrations, the architecture was modified to use 2 input channels (H&E) instead of the RGB image. The area under the ROC curve (AUC), shown in

Tables 6 and 7, was calculated on the test set for the best performing epoch during training for each method.

Table 6: AUC performance of the VGG19 classifier for the proposed and competing methods using CN both on the RGB and OD spaces. Bold values indicate the highest performance for each space.

Input	Original	RUI	MAC	ALS	HID	VAH	PER	ZHE	EBKSVD	BKSVD
RBG	0.9491	NA	0.9499	0.9738	0.9479	0.7985	0.9305	0.9755	**0.9817**	0.9711
OD	NA	0.9417	0.9468	0.9725	0.9642	0.6614	0.9508	**0.9864**	0.9834	0.9672

Table 7: AUC performance of the VGG19 classifier for the proposed and competing methods using SCA both on the RGB and OD spaces. Bold values indicate the highest performance for each space.

Input	$\text{TEL}_{aug}^{strong}$	TEL_{aug}^{light}	$\text{EBKSVD}_{aug}^{strong}$	$\text{EBKSVD}_{aug}^{light}$	$\text{BKSVD}_{aug}^{strong}$	$\text{BKSVD}_{aug}^{light}$
RGB	0.9673	0.9601	**0.9716**	0.9647	0.9679	0.9650
OD	0.9654	0.9639	0.9865	**0.9879**	0.9728	0.9790

The results show that, when using RGB images, CN increased the AUC in most cases, increasing it from the original images (0.9491) up to 0.9817 with the proposed EBKSVD. The less sparse BKSVD approach, slightly increases the AUC without reaching the outperforming result of the EBKSVD. CN obtained by ZHE and ALS also increased the classification performance considerably despite the over-brightened images produced by ZHE and the artifacts produced by ALS.

Results using OD concentrations show that most methods increase AUC in comparison to the baseline RUI method. Also, the performance of HID, PER, ZHE and the proposed EBKSVD is better in OD than in normalized RGB space, showing that BCD is able to provide more useful information for the CNN. Separating the structures in the image from the color information, usually produces better results than using the RGB image since the network does not need to extract the structural information from colors. SCA improves the performance with respect to the original images, both using RGB and OD concentrations, obtaining the best performance with the latter. The highest AUC value was obtained using EBKSVD and light SCA in the OD space. Our results show that SCA benefits from the use of EBKSVD instead of the RUI method used by TEL. The difference between light and strong augmentation is minor both in TEL and the proposed augmentation. Our results show that CN and BCD have a bigger impact on classification than SCA when RGB images are used. However, the proposed combination of CN and SCA improves the results on the OD space.

8. Conclusions

In this paper, we have proposed a novel Bayesian approach for blind color deconvolution of histopathological images, based on K-SVD with two possible inference approaches: variational and empirical Bayes. We utilize a hierarchical prior on the concentrations that enforces sparsity in the same way as a Laplacian prior while allowing for a tractable Bayesian inference. The framework presented automatically estimates the stain concentrations, the

color-vector matrix, and all model parameters. The proposed BKSVD and EBKSVD methods guarantee fidelity to the tissue structure on different relevant histopathological tasks such as color normalization, stain color augmentation, and classification of histological images.

The proposed method is designed to work at the highest magnification available. Although the proposed approach has shown a good performance at 20× and 40×, it is unclear how magnification affects the estimation of the color-vector matrix and has never been explored in the literature. This is an interesting topic to be addressed in future research, specially if hierarchical model are to be used.

The proposed approach solves the dependency on the reference color-vector matrix of previous Bayesian approaches. However, this also exposes a limitation that affects to many other BCD and CN methods: the common assumption that colors on the image come exclusively from H&E stains might not hold in some scenarios Although the proposed Bayesian approach and the pixel sampling provide a certain robustness to variations, large areas of blood, cauterized tissue (e.g. bladder samples) or other anomalies in the WSIs can affect the BCD results and therefore the CN or SCA performance. This issue, that also affects CNN-based CN methods, has never been explored in the BCD or CN fields and needs to be addressed in future research.

The proposed BKSVD and EBKSVD methods outperform classical and state-of-the-art methods on all the performed experiments obtaining higher fidelity to the tissue structure, a more consistent normalization, and a stain specific color augmentation that improves classification on VGG19. The optimal approach, BKSVD or EBKSVD, varies depending on the task.

We have analyzed the effect of using color normalized images or OD concentrations to feed a CNN classifier. The carried out experiments indicate that using OD concentrations for H&E achieves higher classification performance than feeding the network with RGB images. The dependency on a reference image is a well-known issue for BCD-based CN. The choice of a proper reference image also have an impact on the classification performance. The relevance of this choice needs to be quantified in future research. However, it can be avoided with the use of OD concentrations directly for classification.

Finally, we have shown that stain color augmentation techniques are more beneficial when using high-quality stain concentrations that better represent the real structure of the stains in the image. The use of the OD concentrations as input for the network is also useful when working with augmentation techniques.

Appendix A. Derivation of the Sequential Inference for Sparse Bayesian Models

We detail now the maximization of the marginal likelihood in (12), which we introduce here for the sake of completeness

$$\mathcal{L}(\boldsymbol{\gamma}_q) = \log\left[\mathrm{p}(\boldsymbol{\gamma}_q|\hat{\lambda}_q)\int \mathrm{p}(\mathbf{y}_q|\mathbf{c}_q,\hat{\beta})\mathrm{p}(\mathbf{c}_q|\boldsymbol{\gamma}_q)d\mathbf{c}_q\right], \tag{A.1}$$

where $\mathrm{p}(\mathbf{y}_q|\mathbf{c}_q,\hat{\beta}) \sim \mathcal{N}(\hat{\mathbf{M}}\mathbf{c}_q, \hat{\beta}^{-1}\mathbf{I})$, which is clear from the observation model in (1); $\mathrm{p}(\mathbf{c}_q|\boldsymbol{\gamma}_q) \sim \mathcal{N}(\mathbf{c}_q|\mathbf{0},\boldsymbol{\Gamma}_q)$ as defined in Sec. 3; and, the remaining variables, $\mathbf{M}$, β and λ_q, are fixed to the values estimated with variational inference. The marginal integral in (A.1) is a well-known

result:

$$\mathrm{p}(\mathbf{y}_q|\hat{\beta},\hat{\mathbf{M}},\boldsymbol{\gamma}_q) := \int \mathrm{p}(\mathbf{y}_q|\mathbf{c}_q,\hat{\beta})\mathrm{p}(\mathbf{c}_q|\boldsymbol{\gamma}_q)d\mathbf{c}_q = \mathcal{N}(\mathbf{y}_q|\mathbf{0},\mathbf{X}_q), \tag{A.2}$$

with covariance matrix

$$\mathbf{X}_q = \hat{\beta}^{-1}\mathbf{I}_3 + \hat{\mathbf{M}}\boldsymbol{\Gamma}_q\hat{\mathbf{M}}^{\mathrm{T}}. \tag{A.3}$$

Now we can rewrite the marginal likelihood as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}_q) &= \log \mathrm{p}(\boldsymbol{\gamma}_q|\hat{\lambda}_q)\mathrm{p}(\mathbf{y}_q|\hat{\beta},\hat{\mathbf{M}},\boldsymbol{\gamma}_q) \\ &= -\frac{1}{2}\Big[\log|\mathbf{X}_q| + \mathbf{y}_q^{\mathrm{T}}{\mathbf{X}_q}^{-1}\mathbf{y}_q + \hat{\lambda}_q\sum_s \gamma_{sq}\Big] + \text{const.},\end{aligned} \tag{A.4}$$

where the constant includes all terms not depending on $\boldsymbol{\gamma}_q$.

Notice that we can easily find the posterior distribution of $\mathbf{c}_q$, using (6) and (7), once $\hat{\boldsymbol{\gamma}}_q$ has been calculated. In addition, if $\boldsymbol{\gamma}_{sq} = 0$, then the posterior distribution of c_{sq} will be degenerate at zero.

The marginal likelihood $\mathcal{L}(\boldsymbol{\gamma}_q)$ has interesting properties that result in a sequential maximization strategy which will allow us to add, update or remove a single γ_{sq} in order to increase $\mathcal{L}(\boldsymbol{\gamma}_q)$. Concretely, see how we can isolate the contribution of a single γ_{sq} in the covariance matrix $\mathbf{X}_q$ writing

$$\mathbf{X}_q = \Big[\hat{\beta}^{-1}\mathbf{I}_3 + \sum_{i\neq s}\gamma_{iq}\hat{\mathbf{m}}_i\hat{\mathbf{m}}_i^{\mathrm{T}}\Big] + \gamma_{sq}\hat{\mathbf{m}}_s\hat{\mathbf{m}}_s^{\mathrm{T}} =: \widetilde{\mathbf{X}}_q + \gamma_{sq}\hat{\mathbf{m}}_s\hat{\mathbf{m}}_s^{\mathrm{T}}, \tag{A.5}$$

where, clearly, $\widetilde{\mathbf{X}}_q$ has no dependence on γ_{sq}. Using the determinant identity and the matrix inversion lemma on $\mathbf{X}_q$ we can write

$${\mathbf{X}_q}^{-1} = \widetilde{\mathbf{X}}_q^{-1} - \frac{\widetilde{\mathbf{X}}_q^{-1}\hat{\mathbf{m}}_s\hat{\mathbf{m}}_s^{\mathrm{T}}\widetilde{\mathbf{X}}_q^{-1}}{{\gamma_{sq}}^{-1} + \hat{\mathbf{m}}_s^{\mathrm{T}}\widetilde{\mathbf{X}}_q^{-1}\hat{\mathbf{m}}_s}, \tag{A.6}$$

$$|\mathbf{X}_q| = |\widetilde{\mathbf{X}}_q| \cdot |1 + \gamma_{sq}\hat{\mathbf{m}}_s^{\mathrm{T}}\widetilde{\mathbf{X}}_q^{-1}\hat{\mathbf{m}}_s|\,. \tag{A.7}$$

The previous equations allow us to rewrite (A.4) as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}_q) &= -\frac{1}{2}\Big[\log|\widetilde{\mathbf{X}}_q| + {\mathbf{y}_q}^{\mathrm{T}}\widetilde{\mathbf{X}}_q^{-1}\mathbf{y}_q + \hat{\lambda}_q\sum_{n\neq s}\gamma_{nq}\Big] + \frac{1}{2}\Big[\log\frac{1}{1+\gamma_{sq}g_{sq}} + \frac{h_{sq}^2\gamma_{sq}}{1+\gamma_{sq}g_{sq}} - \hat{\lambda}_q\gamma_{sq}\Big] \\ &=: \mathcal{L}(\{\gamma_{iq}\}_{i\neq s}) + l(\gamma_{sq}),\end{aligned} \tag{A.8}$$

where $g_{sq} = \hat{\mathbf{m}}_s^{\mathrm{T}}\widetilde{\mathbf{X}}_q^{-1}\hat{\mathbf{m}}_s$ and $h_{sq} = \hat{\mathbf{d}}_s^{\mathrm{T}}\widetilde{\mathbf{X}}_q^{-1}\mathbf{y}_q$ and the constant has been omitted as it plays no role in the optimization. Notice that the quantities g_{sq} and h_{sq} do not depend on γ_{sq}. Therefore, the terms related to a single hyperparameter γ_{sq} are now separated from the rest. A closed form solution of the maximization of $\mathcal{L}(\boldsymbol{\gamma}_q)$, when only its sth component is changed, can be found by holding the other hyperparameters fixed, taking its derivative with respect to γ_{sq} and setting it equal to zero, obtaining a unique maximum at

$$\hat{\gamma}_{sq} = \begin{cases}\frac{-(g_{sq}+2\hat{\lambda}_q)+\sqrt{g_{sq}^2-4\hat{\lambda}_q h_{sq}^2}}{2\hat{\lambda}_q g_{sq}}, & h_{sq}^2 - g_{sq} \geq \hat{\lambda}_q \\ 0, & \text{otherwise.}\end{cases} \tag{A.9}$$

In order to effectively reduce the computational burden, this calculation must be performed efficiently. To explain how to carry them out, let us overload slightly the notation. The current (^{c}) covariance matrix of the marginal of the observations is rewritten as

$$\mathbf{X}_q^c = \hat{\beta}^{-1}\mathbf{I}_3 + \sum_{i\in\mathcal{A}} \gamma_{iq}^c \hat{\mathbf{m}}_i \hat{\mathbf{m}}_i^{\mathrm{T}} + \sum_{i\in\overline{\mathcal{A}}} \gamma_{iq}^c \hat{\mathbf{m}}_i \hat{\mathbf{m}}_i^{\mathrm{T}}, \tag{A.10}$$

where $\mathcal{A} = \{i|\gamma_{iq}^c > 0\}$ and $\overline{\mathcal{A}} = \{i|\gamma_{iq}^c = 0\}$. Notice that, the last term on the right hand side of (A.10) is equal to zero and has been included for clarity. Then, applying the Woodbury identity, we obtain

$$\hat{\mathbf{m}}_s^{\mathrm{T}} \mathbf{X}_q^{c\,-1} \hat{\mathbf{m}}_s = \hat{\beta}\hat{\mathbf{m}}_s^{\mathrm{T}}\hat{\mathbf{m}}_s - \hat{\beta}^2 \hat{\mathbf{m}}_s^{\mathrm{T}} \hat{\mathbf{M}}^c \boldsymbol{\Sigma}_{\mathbf{c}_q}^c (\hat{\mathbf{M}}^c)^{\mathrm{T}} \hat{\mathbf{m}}_s =: G_{sq} \tag{A.11}$$

where $\boldsymbol{\Sigma}_{\mathbf{c}_q}^c$ is obtained from $\boldsymbol{\Sigma}_{\mathbf{c}_q}$ by keeping only the columns and rows associated to the indices in $\mathcal{A}$. We apply the same restriction to the columns of $\hat{\mathbf{M}}^c$, that is, we keep in $\hat{\mathbf{M}}^c$ the columns associated to $\gamma_{iq}^c > 0$. From (A.6), for $s \in \mathcal{A} \cup \overline{\mathcal{A}}$, we have

$$g_{sq} = \frac{G_{sq}}{1 - \gamma_{sq}^c G_{sq}}. \tag{A.12}$$

Furthermore

$$\hat{\mathbf{m}}_s^{\mathrm{T}} \mathbf{X}_q^{c\,-1} \mathbf{y}_q = \hat{\beta}\hat{\mathbf{m}}_s^{\mathrm{T}}\mathbf{y}_q - \hat{\beta}^2 \hat{\mathbf{m}}_s^{\mathrm{T}} \hat{\mathbf{M}}^c \boldsymbol{\Sigma}_{\mathbf{c}_q}^c (\hat{\mathbf{M}}^c)^{\mathrm{T}} \mathbf{y}_q =: H_{sq} \tag{A.13}$$

Using an analogous procedure we can write

$$h_{sq} = \frac{H_{sq}}{1 - \gamma_{sq}^c G_{sq}}. \tag{A.14}$$

Given $\boldsymbol{\Sigma}_{\mathbf{c}_q}^c$ we can now efficiently check whether we should add $\gamma_{sq}, s \in \overline{\mathcal{A}}$, or update, or remove $\gamma_{sq}, s \in \mathcal{A}$. Moreover, the amount the marginal log likelihood is improved by each single addition, update, or removal is easily calculated from (A.8). Finally,we notice that $\boldsymbol{\Sigma}_q^c$ and $\hat{\mathbf{c}}_q^c$ can be updated very efficiently considering only a single coefficient γ_{sq}, see Tipping and Faul (2003).

Acknowledgements

This work was supported by project PID2019-105142RB-C22 funded by MCIN / AEI / 10.13039 / 501100011033 and project P20_00286 funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades. The work by Fernando Pérez-Bueno was sponsored by Ministerio de Economía, Industria y Competitividad under FPI contract BES-2017-081584.

References

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54, 4311–4322.

Alsubaie, N., Raza, S.E.A., Rajpoot, N., 2016. Stain deconvolution of histology images via independent component analysis in the wavelet domain, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 803–806.

Alsubaie, N., Trahearn, N., Raza, S.E.A., Snead, D., Rajpoot, N., 2017. Stain deconvolution using statistical analysis of multi-resolution stain colour representation. PLOS ONE 12, e0169875.

Astola, L., 2016. Stain separation in digital bright field histopathology, in: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–6.

Babacan, S.D., Luessi, M., Molina, R., Katsaggelos, A.K., 2012. Sparse Bayesian methods for low-rank matrix estimation. IEEE Transaction on Signal Processing 60, 3964–3977.

Babacan, S.D., Molina, R., Katsaggelos, A.K., 2010. Bayesian compressive sensing using Laplace priors. IEEE Transactions on Image Processing 19, 53–63.

Basavanhally, A., Madabhushi, A., 2013. EM-based segmentation-driven color standardization of digitized histopathology, in: Progress in Biomedical Optics and Imaging - Proceedings of SPIE, p. 86760G.

Bejnordi, B.E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., Laak, J.A.v.d., 2016. Stain specific standardization of whole-slide histopathological images. IEEE Transactions on Medical Imaging 35, 404–415.

Bentaieb, A., Hamarneh, G., 2018. Adversarial stain transfer for histopathology image analysis. IEEE Transactions on Medical Imaging 37, 792–802.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer. pp. 454–455.

Bándi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermsen, M., Ehteshami Bejnordi, B., Lee, B., Paeng, K., Zhong, A., Li, Q., Zanjani, F.G., Zinger, S., Fukuta, K., Komura, D., Ovtcharov, V., Cheng, S., Zeng, S., Thagaard, J., Dahl, A.B., Lin, H., Chen, H., Jacobsson, L., Hedlund, M., Çetin, M., Halıcı, E., Jackson, H., Chen, R., Both, F., Franke, J., Küsters-Vandevelde, H., Vreuls, W., Bult, P., van Ginneken, B., van der Laak, J., Litjens, G., 2019. From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge. IEEE Transactions on Medical Imaging 38, 550–560.

Duggal, R., Gupta, A., Gupta, R., Mallick, P., 2017. SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging, in: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. Lecture Notes in Computer Science. Springer, Cham, pp. 435–443.

Esteban, A.E., Lopez-Perez, M., Colomer, A., Sales, M.A., Molina, R., Naranjo, V., 2019. A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes. Computer Methods and Programs in Biomedicine 178, 303–317.

Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R., 2008. Hematoxylin and Eosin Staining of Tissue and Cell Sections. Cold Spring Harbor Protocols.

Gavrilovic, M., Azar, J.C., Lindblad, J., Wählby, C., Bengtsson, E., Busch, C., Carlbom, I.B., 2013. Blind color decomposition of histological images. IEEE Transactions on Medical Imaging 32, 983–994.

Hidalgo-Gavira, N., Mateos, J., Vega, M., Molina, R., Katsaggelos, A.K., 2018. Blind color deconvolution of histopathological images using a variational Bayesian approach, in: International Conference on Image Processing (ICIP), pp. 983–987.

Hidalgo-Gavira, N., Mateos, J., Vega, M., Molina, R., Katsaggelos, A.K., 2020. Variational Bayesian blind color deconvolution of histopathological images. IEEE Transactions on Image Processing 29, 2026–2036.

Hoque, M.Z., Keskinarkaus, A., Nyberg, P., Seppänen, T., 2021. Retinex model based stain normalization technique for whole slide image analysis. Computerized Medical Imaging and Graphics 90, 101901. URL: https://www.sciencedirect.com/science/article/pii/S0895611121000495, doi:https://doi.org/10.1016/j.compmedimag.2021.101901.

Janowczyk, A., Basavanhally, A., Madabhushi, A., 2017. Stain normalization using sparse autoencoders (StaNoSA): Application to digital pathology. Computerized Medical Imaging and Graphics 57, 50–61.

Khan, A.M., Rajpoot, N., Treanor, D., Magee, D., 2014. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE Transactions on Biomedical Eng. 61, 1729–1738.

Lan, J., Cai, S., Xue, Y., Gao, Q., Du, M., Zhang, H., Wu, Z., Deng, Y., Huang, Y., Tong, T., Chen, G., 2021. Unpaired stain style transfer using invertible neural networks based on channel attention and long-range residual. IEEE Access 9, 11282–11295. doi:10.1109/ACCESS.2021.3051188.

Liu, Y., Gadepalli, K.K., Norouzi, M., Dahl, G., Kohlberger, T., Venugopalan, S., Boyko, A.S., Timofeev, A., Nelson, P.Q., Corrado, G., Hipp, J., Peng, L., Stumpe, M., 2017. Detecting cancer metastases on gigapixel pathology images. arXiv Also presented at the 2017 MICCAI tutorial, Deep Learning for Medical Imaging: https://sites.google.com/view/miccai2017-deeplearning.

Macenko, M., Niethammer, M., et al., 2009. A method for normalizing histology slides for quantitative analysis, in: International Symposium on Biomedical Imaging (ISBI), pp. 1107–1110.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2009. Online dictionary learning for sparse coding, in: Proceedings of the 26th Annual International Conference on Machine Learning, Association for Computing Machinery, New York, NY, USA. p. 689–696.

McCann, M.T., Majumdar, J., et al., 2014. Algorithm and benchmark dataset for stain separation in histology images, in: International Conference on Image Processing (ICIP), pp. 3953–3957.

Mpinda Ataky, S.T., de Matos, J., Britto, A.d.S., Oliveira, L.E.S., Koerich, A.L., 2020. Data augmentation for histopathological images based on gaussian-laplacian pyramid blending, in: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. doi:10.1109/IJCNN48605.2020.9206855.

Pérez-Bueno, F., López-Pérez, M., Vega, M., Mateos, J., Naranjo, V., Molina, R., Katsaggelos, A.K., 2020. A TV-based image processing framework for blind color deconvolution and classification of histological images. Digital Signal Processing 101, 102727.

Pérez-Bueno, F., Vega, M., Sales, M.A., Aneiros-Fernández, J., Naranjo, V., Molina, R., Katsaggelos, A.K., 2021. Blind color deconvolution, normalization, and classification of histological images using general super gaussian priors and bayesian inference. Computer Methods and Programs in Biomedicine 211, 106453. URL: https://www.sciencedirect.com/science/article/pii/S0169260721005277, doi:https://doi.org/10.1016/j.cmpb.2021.106453.

Rabinovich, A., Agarwal, S., Laris, C., Price, J.H., Belongie, S.J., 2004. Unsupervised color decomposition of histologically stained tissue samples, in: Advances in Neural Information Processing Systems, pp. 667–674.

Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P., 2001. Color transfer between images. IEEE Computer Graphics and Applications 21, 34–41.

Ruifrok, A.C., Johnston, D.A., 2001. Quantification of histochemical staining by color deconvolution. Analytical and quantitative cytology and histology 23, 291–299.

Runz, M., Rusche, D., Schmidt, S., Weihrauch, M.R., Hesser, J., Weis, C.A., 2021. Normalization of HE-stained histological images using cycle consistent generative adversarial networks. Diagnostic Pathology 16, 71. URL: https://doi.org/10.1186/s13000-021-01126-y, doi:10.1186/s13000-021-01126-y.

Salehi, P., Chalechale, A., 2020. Pix2pix-based stain-to-stain translation: A solution for robust stain normalization in histopathology images analysis, in: 2020 International Conference on Machine Vision and Image Processing (MVIP), pp. 1–7. doi:10.1109/MVIP49855.2020.9116895.

Salvi, M., Michielli, N., Molinari, F., 2020. Stain color adaptive normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology. Computer Methods and Programs in Biomedicine 193, 105506.

Serra, J., Testa, M., Molina, R., Katsaggelos, A.K., 2017. Bayesian K-SVD using fast variational inference. IEEE Transactions on Image Processing 26, 3344–3348.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Tellez, D., Balkenhol, M., Otte-Höller, I., van de Loo, R., Vogels, R., Bult, P., Wauters, C., Vreuls, W., Mol, S., Karssemeijer, N., Litjens, G., van der Laak, J., Ciompi, F., 2018. Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. IEEE Transactions on Medical Imaging 37, 2126–2136.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.M., Ciompi, F., van der Laak, J., 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical Image Analysis 58, 101544.

Tipping, M., Faul, A., 2003. Fast marginal likelihood maximisation for sparse Bayesian models, in: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistic, pp. 3–6.

Tosta, T.A.A., de Faria, P.R., Servato, J.P.S., Neves, L.A., Roberto, G.F., Martins, A.S., do Nascimento, M.Z., 2019a. Unsupervised method for normalization of hematoxylin-eosin stain in histological images. Computerized Medical Imaging and Graphics 77, 101646.

Tosta, T.A.A., de Faria, P.R., Neves, L.A., do Nascimento, M.Z., 2019b. Computational normalization of H&E-stained histological images: Progress, challenges and future potential. Artificial Intelligence in Medicine 95, 118 – 132.

Trahearn, N., Snead, D., Cree, I., Rajpoot, N., 2015. Multi-class stain separation using independent component analysis, in: Medical Imaging 2015: Digital Pathology, p. 94200J.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N., 2016. Structure-preserving color normalization and sparse stain separation for histological images. IEEE Transactions on Medical Imaging 35, 1962–1971.

Vicory, J., Couture, H.D., Thomas, N.E., Borland, D., Marron, J., Woosley, J., Niethammer, M., 2015. Appearance normalization of histology slides. Computerized Medical Imaging and Graphics 43, 89–98.

Vijh, S., Saraswat, M., Kumar, S., 2021. A new complete color normalization method for H&E stained histopatholgical images. Applied Intelligence 51, 7735–7748. URL: https://doi.org/10.1007/s10489-021-02231-7, doi:10.1007/s10489-021-02231-7.

Wei, J., Suriawinata, A., Vaickus, L., Ren, B., Liu, X., Wei, J., Hassanpour, S., 2020. Generative image translation for data augmentation in colorectal histopathology images, in: Proceedings of the Machine Learning for Health NeurIPS Workshop, PMLR 116:10-24, p. 16.

Xiang, Y., Chen, J., Liu, Q., Liang, Y., 2020. Disentangled representation learning based multidomain stain normalization for histological images, in: 2020 IEEE International Conference on Image Processing (ICIP), pp. 360–364. doi:10.1109/ICIP40778.2020.9190757.

Xu, J., Xiang, L., Wang, G., Ganesan, S., Feldman, M., Shih, N.N., Gilmore, H., Madabhushi, A., 2015. Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis. Computerized Medical Imaging and Graphics 46, 20–29.

Zanjani, F.G., Zinger, S., Bejnordi, B.E., van der Laak, J.A.W.M., de With, P.H.N., 2018. Stain normalization of histopathology images using generative adversarial networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI), pp. 573–577.

Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Hu, D., Sun, S., Shi, J., Xue, C., 2021. Stain standardization capsule for application-driven histopathological image normalization. IEEE Journal of Biomedical and Health Informatics 25, 337–347. doi:10.1109/JBHI.2020.2983206.

Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Shi, J., Xue, C., 2019. Adaptive color deconvolution for histological WSI normalization. Computer Methods and Programs in Biomedicine 170, 107–120.

Zhou, M., Chen, H., Ren, L., Sapiro, G., Carin, L., Paisley, J., 2009. Non-parametric Bayesian dictionary learning for sparse image representations, in: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 2295–2303.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251. doi:10.1109/ICCV.2017.244.