Computer Methods and Programs in Biomedicine 211 (2021) 106453

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Blind color deconvolution, normalization, and classification of histological images using general super Gaussian priors and Bayesian inference

Fernando Pérez-Bueno[a,*], Miguel Vega[b], María A. Sales[c], José Aneiros-Fernández[d], Valery Naranjo[e], Rafael Molina[a], Aggelos K. Katsaggelos[f]

[a] Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Spain
[b] Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Granada, Spain
[c] Anatomical Pathology Service, University Clinical Hospital of Valencia, Valencia, Spain
[d] Intercenter Unit of Pathological Anatomy, San Cecilio University Hospital, Granada, Spain
[e] Dpto. de Comunicaciones, Universidad Politécnica de Valencia, Spain
[f] Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA

ARTICLE INFO

Article history:
Received 6 May 2021
Accepted 1 October 2021

Keywords:
Blind color deconvolution
Image normalization
Histopathological images
Variational bayes
Super Gaussian

ABSTRACT

Background and Objective: Color variations in digital histopathology severely impact the performance of computer-aided diagnosis systems. They are due to differences in the staining process and acquisition system, among other reasons. Blind color deconvolution techniques separate multi-stained images into single stained bands which, once normalized, can be used to eliminate these negative color variations and improve the performance of machine learning tasks. *Methods:* In this work, we decompose the observed RGB image in its hematoxylin and eosin components. We apply Bayesian modeling and inference based on the use of Super Gaussian sparse priors for each stain together with prior closeness to a given reference color-vector matrix. The hematoxylin and eosin components are then used for image normalization and classification of histological images. The proposed framework is tested on stain separation, image normalization, and cancer classification problems. The results are measured using the peak signal to noise ratio, normalized median intensity and the area under ROC curve on five different databases. *Results:* The obtained results show the superiority of our approach to current state-of-the-art blind color deconvolution techniques. In particular, the fidelity to the tissue improves 1,27 dB in mean PSNR. The normalized median intensity shows a good normalization quality of the proposed approach on the tested datasets. Finally, in cancer classification experiments the area under the ROC curve improves from 0.9491 to 0.9656 and from 0.9279 to 0.9541 on Camelyon-16 and Camelyon-17, respectively, when the original and processed images are used. Furthermore, these figures of merits are better than those obtained by the methods compared with. *Conclusions:* The proposed framework for blind color deconvolution, normalization and classification of images guarantees fidelity to the tissue structure and can be used both for normalization and classification. In addition, color deconvolution enables the use of the optical density space for classification, which improves the classification performance.

1. Introduction

Histopathological tissues utilized for cancer diagnosis are stained using different dyes, commonly Hematoxylin-Eosin (H&E) [1]. This process facilitates the analysis made by pathologists. The Whole-Slide Images (WSIs) obtained by high-resolution scanners have many advantages: images do not deteriorate over time, they can be easily accessed and shared and, very importantly, enable pathologists to study slides on a screen and the development of

* Corresponding author.
E-mail addresses: fpb@ugr.es (F. Pérez-Bueno), mvega@ugr.es (M. Vega), sales_man@gva.es (M.A. Sales), janeirosf@hotmail.com (J. Aneiros-Fernández), vnaranjo@dcom.upv.es (V. Naranjo), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos).

https://doi.org/10.1016/j.cmpb.2021.106453

Computer-Aided-Diagnosis (CAD) systems. The performance of CAD systems can be significantly affected by color variations of histological images [2]. These variations, which can be inter- and intra-laboratory are introduced in the acquisition procedure. Caused by variables like fixatives, staining manufactures, lab condition and temperatures, and the use of different scanners, among others, see [3] for details. Two main approaches have been proposed to minimize the influence of color variations on the obtained images and their posterior analysis. Blind Color Deconvolution (BCD) and Color Normalization (CN).

BCD techniques separate the stains in an image by estimating its stain color-vectors and the corresponding stain concentrations. The process should lead to structure, nuclei (hematoxylin), cytoplasm and collagen of the stroma (eosin), etc, preservation. BCD can be used for image normalization (by normalizing each stain separately), but this is only one of the possible solutions it offers to deal with color variation. Stain separation also allows CAD systems to use the information provided by each stain separately [4]. Furthermore, concentrations can be directly used for classification [4,5].

CN focuses on transforming histological images to a common color range, usually obtained from a reference WSI. Tosta et al. [3] classifies direct CN methods into histogram matching and color transfer. Histogram matching techniques adjust image colors using histogram information. This is a common solution for general images but it is not appropriate for histological images as it assumes that stains are equally distributed and disregards local information. Stain concentration is closely related to the tissue and cell structures which need to be preserved. Color transfer often includes a segmentation step to identify histological regions or dyes. Then a stain-specific based color correction is applied. However, the selective transformation occasionally causes artifacts on the images. Most Deep Learning (DL) methods are included in this category as they usually perform CN without Color Deconvolution (CD) [6,7].

1.1. Related work

A wide range of solutions have been proposed to find the stain color-vector in the images. They can be experimentally obtained as Ruifrok and Johnston [8] did in one of the pioneer works in the CD field. The empirically obtained color-vectors proposed in [8] do not tackle stain color variation. To take variablity into account, the selection of pixels corresponding to each stain was proposed in [9]. The amount of slides available quickly made this solution obsolete. Formulating the problem as blind source separation, Non-negative Matrix Factorization (NMF) was used in [10]. Using the same principles [11] and [12] further developed this research approach including regularization and sparsity terms which encapsulate the assumption that each stain fixes only to specific tissue structures, forcing most of the pixels to respond to one type of stain only. Singular Value Decomposition (SVD), was applied in [13] for H&E stain separation and then further developed in [14] by considering the interaction between stains. It was recently revisited in [15] where the steps were reorganized to obtain a time-optimized pipeline. The NMF memory and time requirements were reduced in [16] with the use of Non-Negative Least Squares (NNLS). In [17], stain vectors were estimated through clustering in the Maxwellian chromacity plane. In [18], supervised relevant vector machines are used to segment background, hematoxylin and eosin pixels. The color-vector for each stain is then defined as the mean of the pixels in each class. Recently, Salvi et al. [19] have presented a three steps method using Gabor kernels, structure segmentation and a final deconvolution step. Independent Component Analysis (ICA) was utilized in [20] and extended in [21,22], using the wavelet transform that reduces the independence condition between sources. The method in [13] was revised in [23], where the author state that they obtained better result applying it in the linearly inverted RGB-space instead of the (logarithmically inverted) absorbency space. The work by Zheng et al. [24] includes the deconvolution by Ruifrok as a starting point and optimizes the color-vector and concentration values using a prior knowledge-based objective function.

In this work, we develop a Bayesian framework for BCD, CN, and classification of histological images using both normalized and stain separated images. Like the approaches presented in [25] and [5], this work uses Bayesian modeling and inference. In [25], a similarity prior to reference stain color-vectors, together with a smoothness Simultaneous Autoregressive (SAR) prior model on the stain concentrations were used. Since the SAR prior oversmooths edges, in [5], we presented the use of a Total Variation (TV) prior on the stain concentrations. The TV prior preserves sharp edges while reducing noise in the images[26], but unfortunately, in some cases, it tends to flatten areas which, together with the edges, are essential for image classification. For blind natural image deconvolution, we proposed in [26,27] a general framework to model and restore the the image from its blurred and noisy version. We introduced a large class of sparse image priors, the so called Super Gaussians (SGs) which represent well sharp image characteristics. Most sparse image models used in the literature are included in the formulation as special cases. In this work we provide a complete mathematical derivation of how to combine SG prior models with the likelihood associated to blind color deconvolution of histological images. The proposed approach is tested on stain separation, image normalization, and classification problems using five different databases. Preliminary results were presented in [28,29] where a limited theoretical derivation was provided and a reduced set of SG priors and datasets were utilized in the experimental validation. In this work we extend [28,29] by providing a complete and clearer mathematical derivation of the model. We also provide an extensive experimental validation using three additional databases including images from different laboratories. The validation now includes: Application of the SG prior models to stain normalization, a complete evaluation of the stain normalization results, additional classification experiments using normalized images and stain concentrations separately, time comparison of the competing methods, and analysis of the similarity prior on the color-vectors. Furthermore we also evaluate the use of normalized images or stain concentrations for classification tasks, and discuss the use of a third residual stain.

The paper is organized as follows: Section 2 introduces the BCD problem and its mathematical formulation. Section 2.2 presents the modeling and Bayesian inference proposed for the estimation of the color-vector matrix, the stain concentrations, and all the model parameters. In Section 4, we use H&E stained images to evaluate the proposed framework and provide a comparison with classical and state-of-the-art CD methods using four different histolopathology related tasks: BCD stain separation, image normalization, deconvolution based prostate cancer classification, and breast cancer classification using normalized images and stain concentrations. Section 5 includes the discussion and finally, Section 6 concludes the paper.

2. Methods

2.1. Problem formulation

For each WSI, the tissue observed by a brightfield miscroscope is represented as an $MN \times 3$ matrix $\mathbf{I}$. Each color plane is stacked into a $MN \times 1$ column vector $\mathbf{i}_c = (i_{1c}, \ldots, i_{MNc})^{\mathrm{T}}, c \in \{R, G, B\}$. The transmitted light on the color band $c \in \{R, G, B\}$ for the ith pixel in the slide is stored in i_{ic}. Stain deconvolution methods usually ap-

Table 1
Some penalty functions.

Label	$\rho(s)$	$\rho'(s)/\lvert s\rvert$
ℓ_p, $0 < p \leq 1$	$\frac{1}{p}\lvert s\rvert^p$	$\lvert s\rvert^{p-2}$
log	$\log(\epsilon + \lvert s\rvert)$	$(\epsilon + \lvert s\rvert)^{-1}\lvert s\rvert^{-1}$

ply the Beer-Lambert law to transform slide images to the *Optical Density* (OD) space, where the n_s stained slide can be expressed as

$$\mathbf{Y}^{\mathrm{T}} = \mathbf{M}\mathbf{C}^{\mathrm{T}} + \mathbf{N}^{\mathrm{T}}, \tag{1}$$

The observed OD image $\mathbf{Y} \in \mathbb{R}^{MN\times 3}$ contains three channels, i.e., $\mathbf{Y} = [\mathbf{y}_R\ \mathbf{y}_G\ \mathbf{y}_B]$ and each channel $\mathbf{y}_c \in \mathbb{R}^{MN\times 1}$ is defined as $\mathbf{y}_c = -\log_{10}\left(\mathbf{i}_c/\mathbf{i}_c^0\right)$, with $\mathbf{i}_c^0$ the incident light (Typically 255 for RGB images). The values for $\mathbf{y}_c$ are computed element-wise. The matrix $\mathbf{C} \in \mathbb{R}^{MN\times n_s}$ contains the stain concentration, $\mathbf{M} \in \mathbb{R}^{3\times n_s}$ is the color-vector matrix and $\mathbf{N} \in \mathbb{R}^{MN\times 3}$ is a random noise matrix with i.i.d. zero mean Gaussian components with variance β^{-1}.

The BCD aproach aims to estimate both $\mathbf{C}$ and $\mathbf{M}$. In $\mathbf{C}$, the concentration of each stain in the ith $\mathbf{Y}$ pixel value, $\mathbf{y}_{i,:}$, is expressed as the ith row $\mathbf{c}_{i,:}^{\mathrm{T}} = (c_{i1}, \ldots, c_{in_s})$ and the whole contribution of the sth stain to the image is the sth column $\mathbf{c}_s = (c_{1s}, \ldots, c_{MNs})^{\mathrm{T}}$. In the color-vector matrix $\mathbf{M}$, each $\mathbf{m}_s$ column contains the color composition of the sth stain.

2.2. SG Bayesian model

Using the Beer-Lambert model in (1), the observation model is

$$\mathrm{p}(\mathbf{Y}|\mathbf{C}, \mathbf{M}, \beta) = \prod_{i=1}^{MN} \mathcal{N}(\mathbf{y}_{i,:}|\mathbf{M}\mathbf{c}_{i,:}, \beta^{-1}\mathbf{I}_{3\times 3}). \tag{2}$$

The Bayesian approach requires to select a prior distribution on the unknowns. Here we adopt SG distributions as priors for the stain concentrations in the filtered space. SG priors are known to preserve sharp images [26]. They induce sparsity and allow us to find the key values for each stain. We use a set of J high-pass filters noted as $\{\mathbf{D}_v\}_{v=1}^{J}$ to obtain the filtered concentrations $\mathbf{c}_{vs} = \mathbf{D}_v\mathbf{c}_s$. The filtered space remarks the edges in the image that we want to preserve.

$$\begin{aligned}\mathrm{p}(\mathbf{C}|\boldsymbol{\alpha}) &= \prod_{v=1}^{J}\prod_{s=1}^{n_s} \mathrm{p}(\mathbf{c}_{vs}|\alpha_{vs}) \\ &= \prod_{v=1}^{J}\prod_{s=1}^{n_s}\prod_{i=1}^{MN} Z(\alpha_{vs}) \exp\left[-\alpha_{vs}\rho(c_{vs}(i))\right],\end{aligned} \tag{3}$$

with $\alpha_{vs} > 0$ and $Z(\alpha_{vs})$ a partition function. For $\mathrm{p}(\mathbf{c}_{vs}|\alpha_{vs})$ in (3) to be SG, the penalty function $\rho(.)$ has to be symmetric around zero. In addition, $\rho(\sqrt{s})$ has to be increasing and concave for $s \in (0, \infty)$, which is equivalent to $\rho'(s)/s$ being decreasing on $(0, \infty)$. The latter condition allows ρ to be written as follows

$$\rho(c_{vs}(i)) = \inf_{\eta_{vs}(i)>0} L(c_{vs}(i), \eta_{vs}(i)) \tag{4}$$

where $L(c_{vs}(i), \eta_{vs}(i)) = \frac{1}{2}\eta_{vs}(i)\, c_{vs}^2(i) - \rho^*\left(\frac{1}{2}\eta_{vs}(i)\right)$, inf denotes infimum and $\rho^*(\cdot)$ is the concave conjugate of $\rho(\cdot)$ and $\boldsymbol{\eta}_{vs} = \{\eta_{vs}(i)\}_{i=1}^{MN}$ are positive parameters. The relationship dual to (4) is given by Rockafellar [30]

$$\rho^*\left(\frac{1}{2}\eta_{vs}(i)\right) = \inf_{c_{vs}(i)} \frac{1}{2}\eta_{vs}(i)\, c_{vs}^2(i) - \rho(c_{vs}(i)). \tag{5}$$

Table 1 and Fig. 1 show possible choices for the penalty function and their corresponding SG distributions (for additional SG distributions, see [26]).

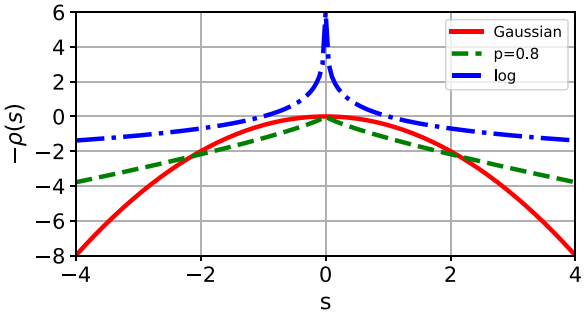


Fig. 1. Penalties corresponding to functions in Table 1. log $|s|$ is bounded for better visualization.

The color-vector matrix $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_{n_s}]$ is also unknow, but it is expected to be similar to a reference color-vector matrix $\underline{\mathbf{M}} = [\underline{\mathbf{m}}_1, \ldots, \underline{\mathbf{m}}_{n_s}]$. Therefore we use a similarty prior as

$$\begin{aligned}\mathrm{p}(\mathbf{M}|\boldsymbol{\gamma}) &= \prod_{s=1}^{n_s} \mathrm{p}(\mathbf{m}_s|\gamma_s) \\ &\propto \prod_{s=1}^{n_s} \gamma_s^{\frac{3}{2}} \exp\left(-\frac{1}{2}\gamma_s\|\mathbf{m}_s - \underline{\mathbf{m}}_s\|^2\right),\end{aligned} \tag{6}$$

where the parameter γ_s, $s = 1, \ldots, n_s$, measures the confidence on the accuracy of the reference $\underline{\mathbf{m}}_s$.

The joint probability distribution is then defined as

$$\begin{aligned}\mathrm{p}(\mathbf{Y}, \mathbf{C}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) &= \mathrm{p}(\mathbf{M}|\boldsymbol{\gamma})\mathrm{p}(\boldsymbol{\gamma})\mathrm{p}(\mathbf{Y}|\mathbf{C}, \mathbf{M}, \beta)\mathrm{p}(\beta) \\ &\quad \times \prod_{v=1}^{J}\prod_{s=1}^{n_s} \mathrm{p}(\mathbf{c}_{vs}|\alpha_{vs})\mathrm{p}(\alpha_{vs}),\end{aligned} \tag{7}$$

where we include the hyperpriors $\mathrm{p}(\boldsymbol{\gamma})$, $\mathrm{p}(\beta)$ and $\mathrm{p}(\alpha_{vs})$ on the model hyperparameters for automatic estimation.

Following the Bayesian paradigm, the estimation of $\mathbf{M}$ and $\mathbf{C}$ is based on our estimation of the posterior distribution $\mathrm{p}(\Theta|\mathbf{Y})$ with $\Theta = \{\mathbf{C}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}\}$ including all the unknowns. Our approach approximates $\mathrm{p}(\Theta|\mathbf{Y})$ using the mean-field variational Bayesian model [31], by the distribution $\mathrm{q}(\Theta)$ of the form $\mathrm{q}(\Theta) = \prod_{s=1}^{n_s}\mathrm{q}(\mathbf{m}_s)\prod_{v=1}^{J}\mathrm{q}(\mathbf{c}_{vs})$ that minimizes the Kullback-Leibler (KL) divergence [32] defined as

$$\mathbf{KL}(\mathrm{q}(\boldsymbol{\Theta})\,||\,\mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y})) = \int \mathrm{q}(\Theta)\log\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, \mathbf{Y})}\,\mathrm{d}\Theta + \log\mathrm{p}(\mathbf{Y}). \tag{8}$$

However, the SG prior for $\mathbf{C}_v$ makes the evaluation of this divergence intractable. To tackle this problem we will make use of the quadratic bound for ρ to bound the prior in (3) with a Gaussian form

$$\mathrm{p}(c_{vs}(i)|\alpha_{vs}) \geq Z(\alpha_{vs}) \exp[-\alpha_{vs}L(c_{vs}(i), \eta_{vs}(i))], \tag{9}$$

$\forall \eta_{vs}(i) > 0$. Then we define

$$\mathcal{M}_v(\mathbf{C}, \boldsymbol{\eta}_v|\boldsymbol{\alpha}_v) = \prod_{s=1}^{n_s}\prod_{i=1}^{MN} Z(\alpha_{vs}) \exp\left[-\alpha_{vs}L(c_{vs}(i), \eta_{vs}(i))\right] \tag{10}$$

and

$$\begin{aligned}\mathrm{F}(\Theta, \mathbf{Y}) &= \mathrm{p}(\mathbf{M}, \boldsymbol{\gamma})\mathrm{p}(\boldsymbol{\gamma})\mathrm{p}(\mathbf{Y}|\mathbf{C}, \mathbf{M}, \beta)\mathrm{p}(\beta) \\ &\quad \times \prod_{v} \mathcal{M}_v(\mathbf{C}, \boldsymbol{\eta}_v|\boldsymbol{\alpha}_v)\mathrm{p}(\boldsymbol{\alpha}_v),\end{aligned} \tag{11}$$

obtaining the bound $\log\mathrm{p}(\Theta, \mathbf{Y}) \geq \log\mathrm{F}(\Theta, \mathbf{Y})$.

Using $\mathrm{F}(\Theta, \mathbf{Y})$ for the posterior distribution in (8) we can now minimize $\mathbf{KL}(\mathrm{q}(\Theta)\,||\,\mathrm{F}(\Theta, \mathbf{Y}))$ instead of $\mathbf{KL}(\mathrm{q}(\Theta)\,||\,\mathrm{p}(\Theta|\mathbf{Y}))$.

As described in [31], $\mathrm{q}(\theta)$, for each unknown $\theta \in \boldsymbol{\Theta}$, can be written as

$$\mathrm{q}(\theta) \propto \exp\langle\log\mathrm{F}(\Theta, \mathbf{Y})\rangle_{\mathrm{q}(\boldsymbol{\Theta}\setminus\theta)}, \tag{12}$$

where $\langle\cdot\rangle$ is the expectation and $q(\Theta\backslash\theta)$ indicates that it is taken with respect to all parameters in $\boldsymbol{\Theta}$ except θ. The mean is used when a point estimation is required.

2.3. Updating the concentrations

We define

$$\mathbf{e}_{i,:}^{-s} = \mathbf{y}_{i,:} - \sum_{k\neq s} \langle c_{ik}\rangle \langle \mathbf{m}_k\rangle$$

$$z_i^{-s} = \langle \mathbf{m}_s\rangle^{\mathrm{T}} \mathbf{e}_{i,:}^{-s}, \quad i = 1,\ldots,MN, \tag{13}$$

from Eq. (12) we can obtain that $q(\mathbf{c}_s) = \mathcal{N}(\mathbf{c}_s|\langle\mathbf{c}_s\rangle, \boldsymbol{\Sigma}_{\mathbf{c}_s})$, where the inverse of the covariance matrix is given by

$$\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1} = \beta\langle \| \mathbf{m}_s \|^2\rangle \mathbf{I}_{MN\times MN} + \sum_{\nu} \alpha_{\nu s} \mathbf{D}_\nu^{\mathrm{T}} \mathrm{diag}(\boldsymbol{\eta}_{\nu s}) \mathbf{D}_\nu \tag{14}$$

and the mean is obtained as

$$\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1} \langle \mathbf{c}_s\rangle = \beta \mathbf{z}^{-s}. \tag{15}$$

2.4. Updating the color-vector matrix

Similarly, from (13), we obtain that $q(\mathbf{m}_s) = \mathcal{N}(\mathbf{m}_s|\langle\mathbf{m}_s\rangle, \boldsymbol{\Sigma}_{\mathbf{m}_s})$, where

$$\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1} = \left(\sum_{\nu=1}^{J} \beta_\nu \sum_{i=1}^{MN} \langle c_{\nu is}^2\rangle + \gamma_s\right) \mathbf{I}_{3\times 3},$$

$$\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1} \langle \mathbf{m}_s\rangle = \left(\sum_{\nu=1}^{J} \beta_\nu \sum_{i=1}^{MN} \langle c_{\nu is}\rangle \mathbf{e}_{\nu i,:}^{-s} + \gamma_s \underline{\mathbf{m}}_s\right). \tag{16}$$

To ensure $\langle\mathbf{m}_s\rangle$ to be a unitary vector, we replace $\langle\mathbf{m}_s\rangle$ by $\langle\mathbf{m}_s\rangle / \| \langle\mathbf{m}_s\rangle \|$ and $\boldsymbol{\Sigma}_{\mathbf{m}_s}$ by $\boldsymbol{\Sigma}_{\mathbf{m}_s} / \| \langle\mathbf{m}_s\rangle \|^2$.

2.5. Updating the variational parameter

The estimation of the $\boldsymbol{\eta}$ matrix, requires to solve, for each $s \in \{1,\ldots,n_s\}$, $\nu \in \{1,\ldots,J\}$ and $i \in \{1,\ldots,MN\}$

$$\begin{aligned}\hat{\eta}_{\nu s}(i) &= \arg\min_{\eta_{\nu s}(i)} \langle L(c_{\nu s}(i), \eta_{\nu s}(i))\rangle_{q(\mathbf{c}_s)} \\ &= \arg\min_{\eta_{\nu s}(i)} \frac{1}{2}\eta_{\nu s}(i)\, u_{\nu s}^2(i) - \rho^*\left(\frac{1}{2}\eta_{\nu s}(i)\right)\end{aligned} \tag{17}$$

where $u_{\nu s}(i) = \sqrt{\langle c_{\nu s}^2(i)\rangle}$. Since

$$\rho^*\left(\frac{\hat{\eta}_{\nu s}(i)}{2}\right) = \min_x \frac{1}{2}\hat{\eta}_{\nu s}(i)x^2 - \rho(x) \tag{18}$$

whose minimum is achieved at $x = u_{\nu s}(i)$. Then, differentiating the right hand side of (18) with respect to x, equating it to zero and substituting the value for x at its minimum, we have,

$$\hat{\eta}_{\nu s}(i) = \rho'(u_{\nu s}(i))/|u_{\nu s}(i)|. \tag{19}$$

2.6. Updating the hyperparameters

The estimates of the parameters controlling the noise and color-vectors confidence are calculated from

$$\hat{\beta}^{-1} = \frac{\mathrm{tr}\langle(\mathbf{Y}^{\mathrm{T}} - \mathbf{M}\mathbf{C}^{\mathrm{T}})(\mathbf{Y}^{\mathrm{T}} - \mathbf{M}\mathbf{C}^{\mathrm{T}})^{\mathrm{T}}\rangle_{q(\boldsymbol{\Theta})}}{3MN}, \tag{20}$$

$$\hat{\gamma}_s^{-1} = \frac{\mathrm{tr}\left(\left\langle(\mathbf{m}_s - \underline{\mathbf{m}}_s)(\mathbf{m}_s - \underline{\mathbf{m}}_s)^{\mathrm{T}}\right\rangle\right)}{3}. \tag{21}$$

Using (12) the distribution for $\alpha_{\nu s}$ is written as follows

$$q(\alpha_{\nu s}) = \mathrm{const} + \sum_{i=1}^{MN} \log Z(\alpha_{\nu s}) \exp\left[-\alpha_{\nu s}\rho(u_{\nu s}(i))\right], \tag{22}$$

where $u_{\nu s}(i)$ was defined in Section 2.5. Estimating $\alpha_{\nu s}$ with the mode of (22), we obtain $\hat{\alpha}_{\nu s}$ from

$$\frac{\partial \log Z(\hat{\alpha}_{\nu s})}{\partial \hat{\alpha}_{\nu s}} = \frac{1}{MN}\sum_{i=1}^{MN} \rho(u_{\nu s}(i)). \tag{23}$$

From the penalty functions shown in Table 1, ℓ_p produces proper priors, where we can evaluate the partition function. However, the log penalty function produces an improper prior. To tackle this problem we examine the behaviour of

$$Z(\alpha_{\nu s}, K)^{-1} = \int_{-K}^{K} \exp\left[-\alpha_{\nu s}\rho(t)\right] dt \tag{24}$$

when $\alpha_{\nu s} \neq 1$, and keeping in $\partial Z(\alpha_{\nu s})/\partial\alpha_{\nu s}$ the term that depends on $\alpha_{\nu s}$. This produces for the log prior

$$\frac{\partial Z(\hat{\alpha}_{\nu s})}{\partial \hat{\alpha}_{\nu s}} = (\hat{\alpha}_{\nu s} - 1)^{-1}. \tag{25}$$

Values for $\hat{\alpha}_{\nu s}$ can be obtained substituting this last expression into (23). Flat hyperpriors have been used for all the hyperparamenters.

2.7. Covariance matrices for the concentration

We have to find the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{c}_s}$ in order to calculate its trace as well as $\hat{\eta}_{\nu s}(i)$. Unfortunately, this is computationally intensive. To reduce the impact of the calculation, we propose to approximate $\boldsymbol{\Sigma}_{\mathbf{c}_s}$ as follows. First, we approximate $\mathrm{diag}(\boldsymbol{\eta}_{\nu s})$ by

$$\mathrm{diag}(\boldsymbol{\eta}_{\nu s}) \approx z(\boldsymbol{\eta}_{\nu s})\mathbf{I}, \tag{26}$$

where we use the mean of the diagonal values to calculate $z(\boldsymbol{\eta}_{\nu s})$. Then we approximate

$$\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1} \approx \beta\langle \| \mathbf{m}_s \|^2\rangle \mathbf{I}_{MN\times MN} + \sum_{\nu} \alpha_{\nu s} z(\boldsymbol{\eta}_{\nu s}) \mathbf{D}_\nu^{\mathrm{T}} \mathbf{D}_\nu = \mathbf{B}.$$

Finally we have $\langle c_{\nu s}^2(i)\rangle \approx (\langle c_{\nu s}(i)\rangle)^2 + \frac{1}{MN}\mathrm{tr}\left[\mathbf{B}^{-1}\mathbf{D}_\nu^{\mathrm{T}}\mathbf{D}_\nu\right]$.

2.8. Proposed algorithm

Considering the previous inference, we propose the Fully Variational Bayesian SG BCD in Algorithm 1. Fig. 2 depicts the pipeline

Algorithm 1 Fully Variational Bayesian SG BCD.

Require: Observed RGB image $\mathbf{I}$ and reference (prior) color-vector matrix $\underline{\mathbf{M}}$.

Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$ and set $\langle\mathbf{m}_s\rangle^{(0)} = \underline{\mathbf{m}}_s$, $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{(0)} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(0)} = \mathbf{0}$, $\langle\mathbf{c}_s\rangle^{(0)}$, $\forall s = 1,\ldots,n_s$, from the matrix $\mathbf{C}$ obtained as $\mathbf{C}^{\mathrm{T}} = \underline{\mathbf{M}}^{+}\mathbf{Y}^{\mathrm{T}}$, with $\underline{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$, and $n = 0$.

while convergence criterion is not met **do**

1. Set $n = n + 1$.
2. Obtain $\beta^{(n)}$, $\gamma_s^{(n)}$ and $\alpha_{\nu s}^{(n)}$ from (20), (21) and (23).
3. Using $\langle\mathbf{c}_s\rangle^{(n-1)}$ and $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(n-1)}$ $\forall s$, update variational parameters $\hat{\boldsymbol{\eta}}_{\nu s}^{(n)}$ from (19) $\forall\nu$.
4. Using $\langle\mathbf{c}_s\rangle^{(n-1)}$, $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(n-1)}$ and $\langle\mathbf{m}_s\rangle^{(n-1)}$ update $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{(n)-1}$ and solve (16) for the color-vectors $\langle\mathbf{m}_s\rangle^{(n)}$, $\forall s$.
5. Using $\langle\mathbf{m}_s\rangle^{(n)}$, $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{(n)}$ and $\hat{\boldsymbol{\eta}}_{\nu s}^{(n)}$ $\forall\nu$ update $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(n)-1}$ from (14) and solve (15) for the concentrations $\langle\mathbf{c}_s\rangle^{(n)}$, $\forall s$.

end while

Output color-vector $\hat{\mathbf{m}}_s = \langle\mathbf{m}_s\rangle^{(n)}$ and $\hat{\mathbf{c}}_s = \langle\mathbf{c}_s\rangle^{(n)}$.

followed by the proposed framework. We use the Conjugate Gradient approach to solve the linear equation problem in step 4 of

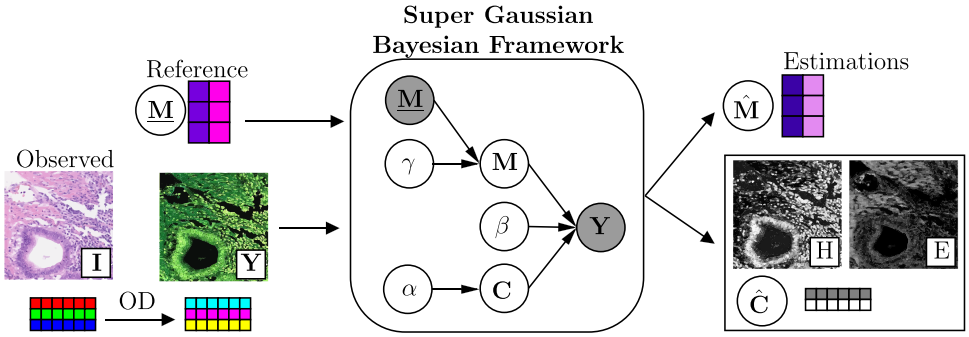


Fig. 2. Pipeline of the proposed framework. First, the H&E image is converted to the OD space. The OD image $\mathbf{Y}$ and the reference matrix $\underline{\mathbf{M}}$ are given to the SG Bayesian framework. The values of all parameters are automatically estimated during the inference procedure using the KL divergence. Finally, the estimated color-vector matrix $\hat{\mathbf{M}}$ and concentrations $\hat{\mathbf{C}}$ are obtained.

Alg. 1. The inference procedure iterates between concentration update, color-vector update, variational parameter update, and parameter update. When necessary, a single-stain RGB image $\hat{\mathbf{I}}_s^{\text{sep}}$, can be obtained from the outputs in Algorithm 1 as follows

$$(\hat{\mathbf{I}}_s^{\text{sep}})^{\mathrm{T}} = \exp_{10}\left(-\hat{\mathbf{m}}_s\hat{\mathbf{c}}_s^{\mathrm{T}}\right) \tag{27}$$

2.9. Use of the algorithm on WSIs

The size of WSI images is usually on the order of Gigapixels, making their processing challenging. The proposed method could, in principle, be used directly on WSIs but Bayesian methods are computationally expensive and the computational burden would be considerable, notice that M and N would be huge. However, WSIs are not usually processed at once and most classification or analysis tasks require patching [4,33] or focusing only on Regions-of-Interest (RoI) [22].

For classification purposes it is possible to deconvolve patches separately. This approach can tackle local variations but will create variations in the estimated color-vector matrix for each patch. Another possible solution is to select a RoI in order to obtain the color matrix. This is the approach we follow in this paper. First, we select the biggest connected RoI within the patches of interest and estimate the color-vector matrix $\hat{\mathbf{M}}$ for the complete WSI. Then, the concentrations of the remaining patches are obtained using $\mathbf{C}^{\mathrm{T}} = \hat{\mathbf{M}}^{+}\mathbf{Y}^{\mathrm{T}}$, with $\hat{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\hat{\mathbf{M}}$. Notice that a single color-vector matrix is obtained for all patches belonging to the same WSI and that they can be stitched together without artifacts if necessary.

An alternative approach is to use the prior on the concentrations in Eq. (3) and the observation model in Eq. (2) for all the patches in the WSI we want to use. In other words, Eq. (2) becomes a product over patches of interest. Notice that the new variational distributions are similar to those derived in the paper but now have to consider all the utilized patches.

3. Data material

Five databases, were used in the experiments *Warwick Stain Separation Benchmark* (WSSB) [22], SICAPv1 [4], SICAP-GR, Camelyon-16 [34] and Camelyon-17 [35]. Details for each database are provided below:

3.1. WSSB

WSSB is a multi-tissue dataset (breast, colon, and lung) that contains 24 H&E stained images from different laboratories and captured with different microscopes. Colon images were captured at 20× magnification and Breast and Lung at 40×. Hematoxylin- and Eosin-only pixels manually selected by expert pathologists were used to obtain the ground truth stain color-vector matrix for each image. Then, the ground truth concentration is calculated in [22] as

$$\mathbf{C}_{GT}^{\mathrm{T}} = \mathbf{M}_{GT}^{+}\mathbf{Y}^{\mathrm{T}}. \tag{28}$$

Then using (27), a single-stain RGB image was calculated for both hematoxylin and eosin. This database will be used for BCD evaluation.

3.2. SICAPv1

This database comes from *Hospital Clínico Universitario de Valencia*, Spain, it contains 79 H&E WSI from 48 patients, 19 benign prostate tissue biopsies (negative class) and 60 pathological prostate tissue biopsies (positive class). The images were digitized using a Ventana iScan Coreo scanner at 40× magnification. Malignant regions of each pathological WSI were annotated by expert pathologists. 60 WSI (17 benign and 43 pathological) were used as training set and the remaining 19 WSI (2 benign and 17 pathological) were utilized for testing. This database will be used for classification purposes and some of its slides will also be used for CN as we describe next.

3.3. SICAP-HUVNGR

This dataset contains 26 prostate H&E WSI: 13 slides at 40× magnification from *Hospital Universitario Virgen de las Nieves de Granada (HUVNGR)* and 13 slides from *Hospital Clínico Universitario de Valencia* (randomly extracted from SICAPv1 dataset). These WSIs will be used for CN evaluation.

3.4. Camelyon-16 and 17

These two databases are part of the Camelyon challenge[1] for cancer metastasis detection in the lymph node. We will use them in CN and classification experiments. Both Camelyon databases were scanned at 40×. They are described below.

- Camelyon-16 contains 400 H&E-stained lymph node multiresolution WSIs from 2 different laboratories. 270 are used from training (159 referred as normal and 111 as tumor) and 130 for testing. Cancer regions were annotated by expert pathologists in tumor and test images. All the annotations are available.
- Camelyon-17 contains 1000 WSIs from 5 medical centers. Only the training set, which contains 500 WSIs, was used since the annotations for the testing WSIs are not yet available. The

[1] https://camelyon17.grand-challenge.org/.

Table 2
Camelyon 17 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
no annotated	450	318	20	42	70

Table 3
Experiments performed for each database.

database	Stain separation	Color normalization	Classification
WSSB	✓		
SICAPv1		✓	✓
HUVNGR			
Camelyon-16		✓	✓
Camelyon-17		✓	✓

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 2 for details.

For a clearer perspective, we include Table 3 that shows the experiments performed for each database.

4. Experiments and results

As mentioned previously, BCD techniques are used to facilitate the visual analysis and to improve the automatic classification of WSIs. These are frequently conflicting goals due to the differences between the human eye and computer vision. Usually, the highest classification performance is not obtained with the most accurate color deconvolved images, where each stain is accurately separated.

We have designed a set of experiments to test the performance of the algorithms on the most common histological color deconvolution related tasks: stain separation, image normalization, and classification. Our first experiment is devoted to assess the quality of the stain separated images, that is, of the concentration matrices and color vector. Then, in the second one, we analyze the quality of these matrices in CN. In the CN step a reference WSI is selected and the color-vectors of the image to normalize are substituted by those of the reference image, keeping the concentrations. Finally, the obtained deconvolved and normalized images are evaluated on histological classification problems.

The proposed SG framework was compared with the following (B)CD methods frequently used in the literature: the classical non-blind CD method by Ruifrok and Johnston [8] and the BCD methods by Macenko et al. [13], Vahadane et al. [11], Alsubaie et al. [22], Hidalgo-Gavira et al. [25], Pérez-Bueno et al. [5] and Zheng et al. [24]. They will be denoted by RUI, MAC, VAH, ALS, HID, PER, and ZHE, respectively. Since SG represents a family of prior distributions, we have selected two of its representative members, the corresponding to ℓ_p and log energy functions. They will be denoted by L1 and LOG, respectively, in the experiments.[2] For the ℓ_p function we experimentally compared values for p in the interval $[0.6, 1]$ and found no significant differences. For simplicity, we choose ℓ_1. The proposed L1 and LOG methods were run until the criterion $\|\langle \mathbf{c}_s \rangle^{(n)} - \langle \mathbf{c}_s \rangle^{(n-1)}\|^2 / \|\langle \mathbf{c}_s \rangle^{(n)}\|^2 < 10^{-3}$ was met by all the stains. Vertical, horizontal and diagonal differences were used as high-pass filters in the concentration prior (Eq. (3)). All the model parameters are automatically estimated.

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

4.1. BCD stain separation experiments

We begin the experimental assessment by comparing the fidelity to the H&E separation obtained by the different BCD methods on the WSSB database, see Section 3. From this dataset, we show an observed RGB image (Fig. 3(a)) and the corresponding ground truth H&E-only RGB image (Fig. 3(b)).

To set an adequate prior for our method, we consider that the stain color properties may change for the different tissues types in WSSB (Colon, breast, lung). For each tissue, an H&E reference color-vector matrix $\underline{\mathbf{M}}$ was selected by a non-medical expert using a single pixel for each stain. Following the widely used implementation [36] of Ruifrok's method, when a third residual component is used, the reference color-vector is calculated using the vector product of the H&E components in the color matrix.

The single stain images obtained from the observed image in Fig. 3(a) are shown in Fig. 3(c-k). The standard color vector used by RUI obtains a separation that do not represent the ground truth. The proposed methods, L1 and LOG, and MAC, HID, and PER are able to find colors that are close to the ground truth separation in Fig. 3(b). The Bayesian methods HID, PER and the proposed ones share the same prior for the color-vectors, but their differences lay on the concentration prior. HID uses a SAR model, that tends to oversmooth images. The TV based method PER keeps edges sharp, but flattens the inner area of the tissues. The proposed SG methods does not suffer from the Gaussian oversmoothing, obtaining sharper edges depending on the prior chosen and richer details than MAC and the just described methods.

The quantitative comparison on the stain separated images was performed using the Quaternion Structural Similarity (QSSIM) [37] and the Peak Signal to Noise Ratio (PSNR) metrics. The mean value for each tissue in the dataset is presented in Table 4. The results show that the proposed L1 with $n_s = 2$ achieves outperform the competitors. The proposed LOG slightly improves the results of the TV based method PER. This table also includes the performance of our proposed methods when three color vectors are used. As we will later show, the use of a residual component facilitates the classification task, see also [5]. Although the use of three components deteriorates the quality of the stained separated images, our methods perform similarly to some other methods (not the worst ones) in terms of PSNR and QSSIM values.

As it can be observed in Figs 3(j-m) the differences when a third component is used are difficult to distinguish. For a better visual comparison, Fig. 4 shows zoomed in details from Fig. 3(k&m). Notice that we report L1 results since this method obtains the best PSNR and QSSIM values with $n_s = 2$ and the difference with the $n_s = 3$ results is wider. The difference between hematoxylin (Figs 4(a&d) and eosin (Fig. 44(b&e) colors is small. The third component captures only residual information extracted from the H&E bands. The third band is discarded, which implies less fidelity to the original image. Then, the experimental design in [22] implies that removing information will lead to lower PSNR and QSSIM values. In spite of the lower figures of merit, we will see in following sections that the use of a third component leads to better classification performances.

4.1.1. Dependency on the reference color-vector $\underline{\mathbf{M}}$

The similarity prior in (6) requires the use of a reference color-vector matrix $\underline{\mathbf{M}}$. On one hand, the prior on $\mathbf{M}$ ensures that the obtained result agrees with our previous knowledge on the H&E channels. On the other hand, it reduces the search space of feasible solutions. The prior for our model should be as accurate as

Table 4
Mean PSNR and QSSIM values for all the methods on the WSSB dataset [22].

		RUI	MAC	VAH	ALS	HID	PER	ZHE	LOG	LOG	L1	L1
		$n_s = 2$	$n_s = 3$	$n_s = 2$	$n_s = 3$							
Image	Stain											
Colon	H	22.27	23.91	25.83	21.11	28.57	28.62	17.89	28.66	24.12	**29.01**	24.12
	E	20.70	21.55	26.29	21.94	27.58	27.60	14.76	27.74	25.31	**28.38**	25.31
Breast	H	15.27	26.24	25.46	24.60	28.81	29.14	15.31	29.23	27.56	**30.50**	27.56
	E	17.66	23.62	27.68	25.92	26.60	26.76	14.99	26.74	27.19	**27.71**	27.19
Lung	H	22.47	19.52	25.87	20.62	32.91	33.10	19.51	31.21	24.69	**35.21**	24.69
	E	22.05	18.09	25.53	23.95	30.77	31.02	16.23	29.99	25.50	**33.07**	25.50
Mean	H	20.00	23.22	25.72	22.11	30.10	30.29	17.57	29.70	25.46	**31.57**	25.46
	E	20.14	21.08	26.50	23.94	28.32	28.46	15.33	28.16	26.00	**29.72**	26.00
QSSIM												
Image	Stain											
Colon	H	0.8841	0.8581	0.9536	0.5369	0.9635	0.9163	0.7490	0.9556	0.9168	**0.9696**	0.9168
	E	0.5670	0.6133	0.8656	0.7642	0.8713	0.6111	0.4407	0.8455	0.8404	**0.9011**	0.8404
Breast	H	0.7721	0.9859	0.9881	0.7347	**0.9919**	0.6813	0.5231	0.9903	0.9852	0.9918	0.9852
	E	0.7721	0.8907	**0.9695**	0.8068	0.9598	0.5527	0.3108	0.9567	0.9594	0.9605	0.9594
Lung	H	0.9206	0.6973	0.9489	0.4603	**0.9959**	0.9519	0.7747	0.9894	0.9442	0.9957	0.9442
	E	0.5368	0.3500	0.8064	0.7983	0.9401	0.6226	0.3359	0.8807	0.8405	**0.9433**	0.8405
Mean	H	0.8589	0.8471	0.9635	0.5773	0.9838	0.8499	0.6823	0.9784	0.9488	**0.9857**	0.9488
	E	0.6253	0.6180	0.8805	0.7898	0.9237	0.5955	0.3624	0.8943	0.8801	**0.9349**	0.8801

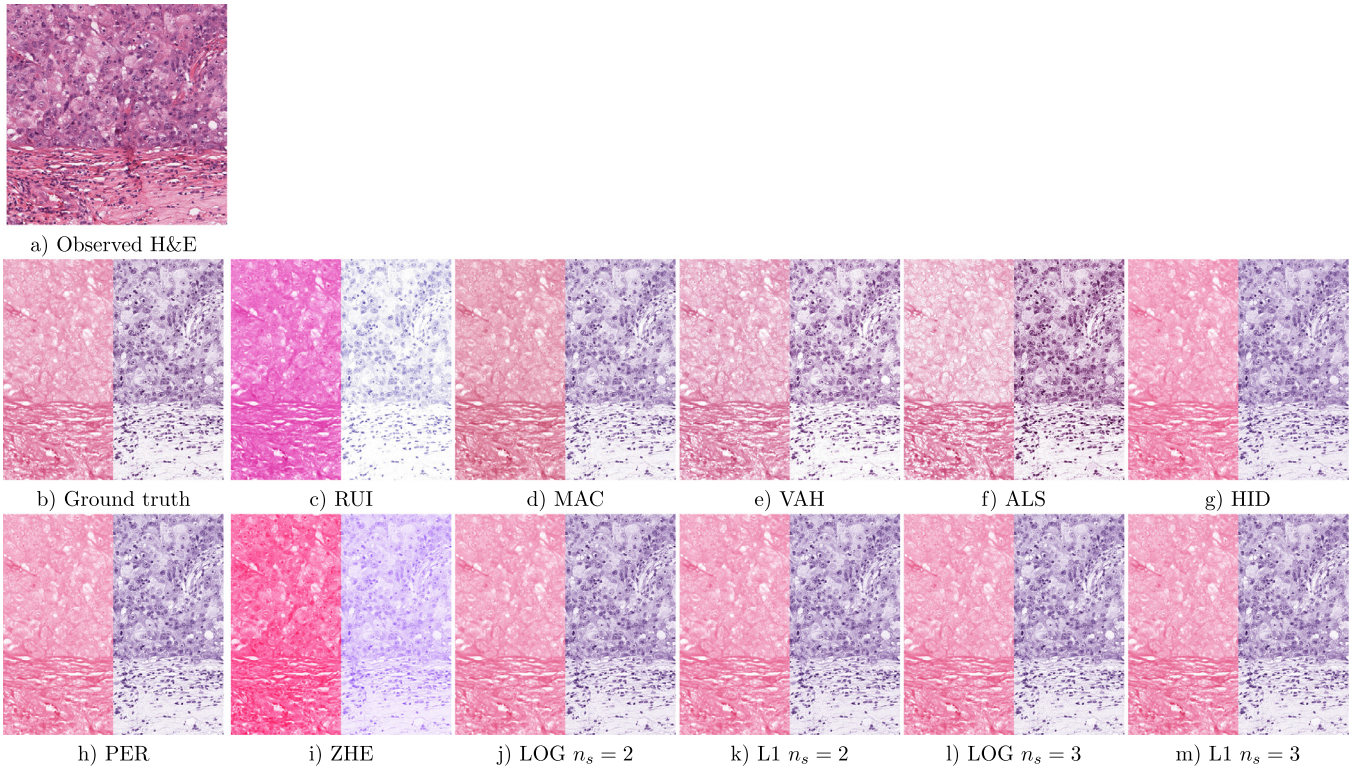


Fig. 3. Single breast observed H&E RGB image from WSSB [22], corresponding ground truth single stain E-only and H-only images and separation obtained by the BCD methods. Eosin and hematoxylin separations are presented on the left and right hand sides of each image, respectively.

possible. However, the color variability in the WSIs might hamper the accuracy of our prior. To assess the impact of the reference matrix $\mathbf{M}$ we have evaluated a breast image on the WSSB dataset using different values of $\mathbf{M}$. Variations of $\mathbf{M}$ were obtained by adding random values sampled from an uniform distribution $U(-\sigma, \sigma)$, with $\sigma \in [0.05, 0.3]$. Then, each row is normalized to achieve $\| \mathbf{m}_s \| = 1$. Twenty different color-vectors were used as prior for the L1 method. Fig. 5 depicts some values for $\mathbf{M}$, PSNR, and QSSIM as σ increases. Values of $\sigma \geqslant 0.2$ produce low quality values for the prior, as they do not represent the stains in the image and even reach unreal values for the H&E channels. The variations on the prior have a considerable impact on the obtained separation. The proposed method is able to deal with variations up to $\sigma = 0.1$ while obtaining values comparable to the competing methods.

4.1.2. Time comparison

Using a single WSSB image, we measured the time needed for the competing methods to deconvolve the image. The comparison is shown in Fig. 6 as a joint plot with the PSNR values obtained. RUI, which does not require color-vector estimation, obtains the lowest time. More complex blind methods require more compu-

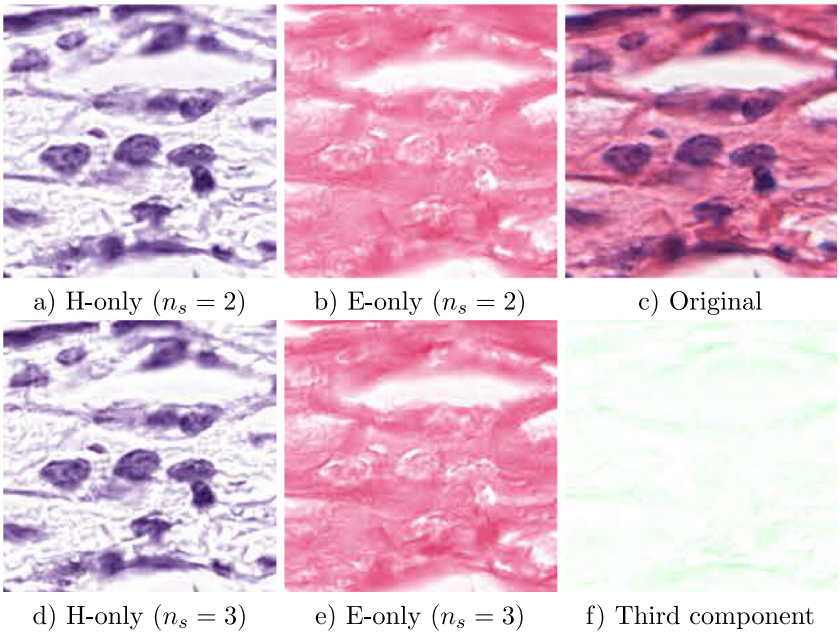


Fig. 4. Detail of the of the bottom left corner of Fig. 3(a) and its H-only, E-only and third component separations. Separations in the top and bottom rows were obtained with the proposed method L1 with two components Fig. 3(k) and three components Fig. 3(m), respectively.

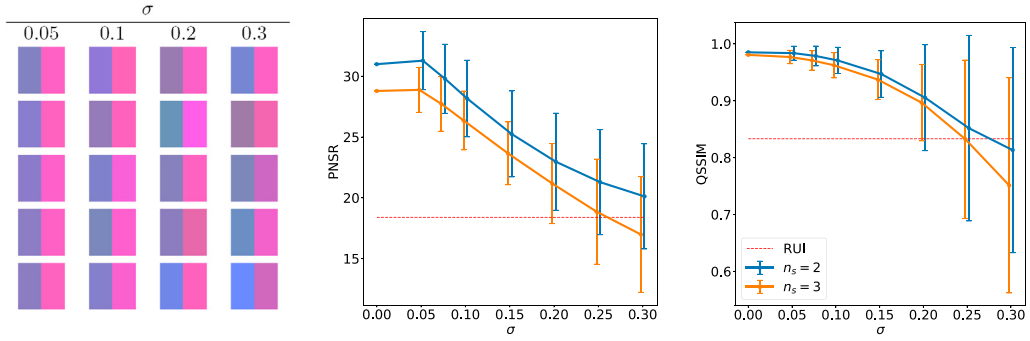


Fig. 5. Evolution of the results for different values of **M**. a) Different combinations of H&E color-vectors used as **M**. Each column shows different values obtained with a fixed variance. b&c) Evolution of PSNR and QSSIM as the variance in **M** increases, respectively. The red dashed line indicates the performance of the separation by RUI.

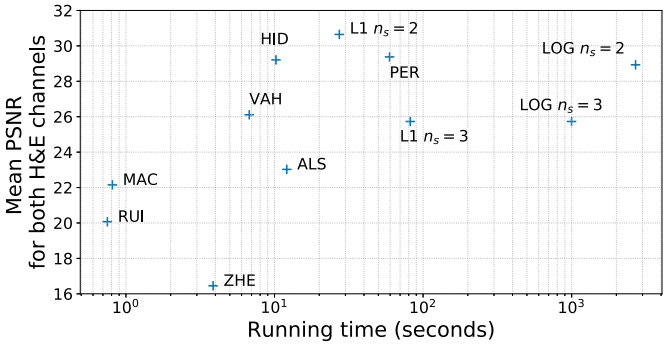


Fig. 6. Joint plot of mean PSNR and running time for deconvolving a 2000×2000 image. The time is counted in seconds and the x axis is presented in logarithmic scale. The time was measured in a shared server running CentOS 7 with 32 CPU Intel(R) Xeon(R) (2.4 GHz).

tational time to estimate the color-vector matrix. ZHE implements a similar deconvolution step to RUI using a similar time. ALS requires as much time as HID but its PSNR and QSSIM values are lower. The proposed approach is severely impacted by the chosen prior. Using LOG the proposed method is expensive in time cost. However L1 reduces by half the time spent by the TV-based method, PER. L1 requires a longer time than some of the competing methods but also obtains the best figures-of-merit as already reported in Table 4. Considering a third stain component increases the time required by L1 but reduces it for LOG. This is due to a higher number of parameters to estimate but less iterations required to converge, specially for LOG. L1 required 6 iterations to deconvolve the image in both cases while LOG used 10 and 6 with $n_s = 2$ and $n_s = 3$, respectively. Notice, also, that the proposed fully Bayesian approach includes estimation of all model parameters together with the stain concentrations and color-vector matrix All these estimations increase the running time but make our methods parameter free.

4.2. CN experiments

Deep learning based CAD systems usually make use of the observed H&E images instead of the separated bands [38]. Therefore, they are highly affected by stain color variations. CN aims to provide an improved input to CAD system. The images are prepro-

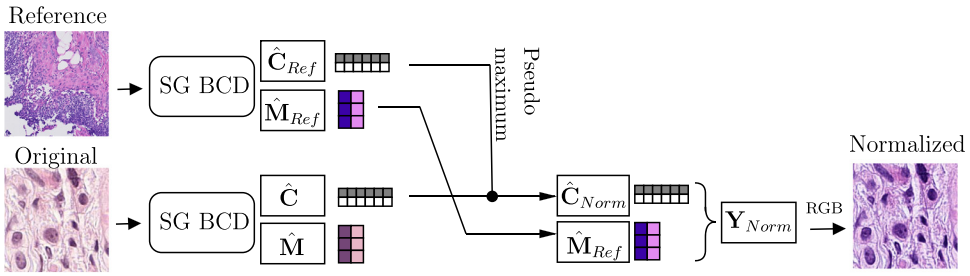


Fig. 7. Pipeline of the normalization procedure. Both, reference and original image are color deconvolved. To obtain the normalized image, the dynamic range of the concentration $\hat{\mathbf{C}}$ is adjusted to be the same as that of $\hat{\mathbf{C}}_{reference}$ and the color matrix $\hat{\mathbf{M}}$ is substituted by $\hat{\mathbf{M}}_{reference}$. Then, the normalized image $\hat{\mathbf{Y}}_{Normalized}$ is transformed back to RGB space.

cessed to reduce the staining variations without modifying their structure. CN can easily be achieved as an additional step after BCD, as stain color information is separated from the structure of stain concentration. This section performs a comparison on the color variations between the original data and the CN obtained by all the competing methods.

To normalize the images a reference image, $\mathbf{I}_{reference}$ is used. Let $\hat{\mathbf{M}}_{reference}$ and $\hat{\mathbf{C}}_{reference}$ be the estimated color and concentration matrices in the OD space obtained using one of our proposed methods on the image $\mathbf{Y}_{reference}$ (obtained from $\mathbf{I}_{reference}$). Following [11], given a new image $\mathbf{I}$, the dynamic range of its corresponding $\hat{\mathbf{C}}$ is adjusted to be the same as that of $\hat{\mathbf{C}}_{reference}$ and the color matrix $\hat{\mathbf{M}}$ is substituted by $\hat{\mathbf{M}}_{reference}$ to obtain the normalized image as follows:

$$(\hat{Y}_{normalized})^T = \sum_{s=1}^{n_s} -(\hat{\mathbf{m}}_s)_{reference}\hat{\mathbf{c}}_s^T \frac{P_{99}((\hat{\mathbf{c}}_s)_{reference})}{P_{99}(\hat{\mathbf{c}}_s)} \tag{29}$$

where $P_{99}(\mathbf{v})$ represents the pseudo maximum (99%) of vector $\mathbf{v}$. The normalized RGB image $\hat{\mathbf{I}}_{normalized}$ is then

$$\hat{\mathbf{I}}_{normalized} = \exp_{10} \hat{\mathbf{Y}}_{normalized} \tag{30}$$

Fig. 7 depicts the pipeline followed to obtain the normalized image.

To measure the quality of a CN procedure, we use the normalized median intensity (NMI) measure [39] defined as

$$NMI(\mathbf{I}) = Median(\mathbf{u})/P_{95}(\mathbf{u}) \tag{31}$$

where $\mathbf{I}$ denotes a WSI and $\mathbf{u}$ is a vector where each u_i component is the mean value of the R, G, and B channels at the ith pixel, [40].

The NMI value is calculated for each WSI in a given dataset. However, we require information about the distribution of the NMI values in the dataset. Then, the standard deviation of the NMI values in the dataset (NMI SD) and the coefficient of the variation (NMI CV), i.e., NMI SD divided by mean of the dataset, were used as metrics. Lower values of NMI SD and NMI CV indicate a more consistent normalization.

Three datasets containing images from different centers were used in this section. SICAP-HUVNGR, Camelyon-16 and Camelyon-17, see Section 3.

In the SICAP-HUVNGR dataset, to avoid the influence of large background regions, 512 × 512 pixel patches at 40× magnification, with at least 70% tissue, were sampled from each WSI. This patch size is motivated by the prostate slide appearance. They are narrow tissue segments surrounded by background which is also visible inside glands. The use of a larger patch size, while maintaining the above tissue percentage, discard most patches containing glands and keep only stroma patches mainly stained with eosin, because they have low nuclear density. The NMI for each WSI is calculated over all the pixels in the patches. The number of patches used from each WSI was evaluated from 20 to 120, observing that beyond 60 the NMI value did not change.

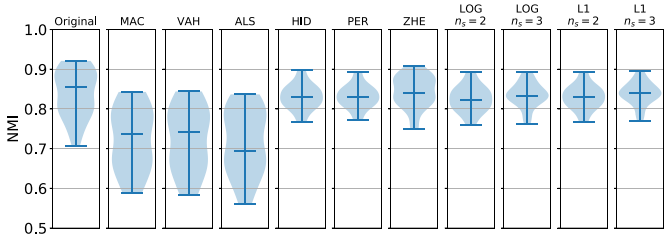


Fig. 8. Violin plots of NMI values for the original and normalized images by the compared methods on the SICAP-HUVNGR dataset. The bars mark the maximum, median and minimum values for each plot.

For Camelyon-16 and Camelyon-17 datasets, 224 × 224 pixel patches, with at least 70% tissue, were sampled from each WSI. This will also be the patch size used for classification, see Section 4.4. Following [24], 500 patches were sampled from each WSI in the datasets for CN and classification purposes.

Let us now describe the obtained results. First we notice that RUI does not estimate the color-vectors in the images, therefore it is not possible to use it for CN. Furthermore, the prior color-vector matrix $\underline{\mathbf{M}}$ used by our method is fixed to the standard proposed by Ruifrok and Johnston [8].

NMI values for the SICAP-HUVNGR dataset are shown in Table 5. The proposed methods, LOG and L1, reduce by half the NMI SD and NMI CV values of the original data. L1 obtains the best value with $n_s = 3$. ZHE significantly reduces both values, but the results are not as clustered as the obtained by HID and PER. MAC, VAH, and ALS do not improve the initial NMI values. Fig. 8 depicts the distribution of NMI values using violin plots. In the first column of the figure, two different NMI distributions can be appreciated on the original data. They correspond to the two centers the images come from. The two centers are still visible when MAC, VAH, ALS and ZHE are used, but disappear when HID, PER, and the proposed L1 and LOG are utilized. The proposed L1 and LOG

Table 5
NMI SD and NMI CV comparison for diferent normalization methods on SICAP-HUVNGR dataset.

Method	NMI SD	NMI CV
Original Data	0.0591	0.0705
MAC	0.0782	0.1079
VAH	0.0796	0.1099
ALS	0.0799	0.1114
HID	0.0313	0.0378
PER	0.0296	<u>0.0356</u>
ZHE	0.0398	0.0472
LOG $n_s = 2$	0.0330	0.0400
LOG $n_s = 3$	0.0307	0.0368
L1 $n_s = 2$	<u>0.0306</u>	0.0368
L1 $n_s = 3$	**0.0287**	**0.0342**

Table 6
NMI SD and NMI CV comparison for diferent normalization methods on Camelyon-16.

database	Camelyon-16							
subset	All images		Tumor		Normal		Test	
Method	SD	CV	SD	CV	SD	CV	SD	CV
Original Data	0.0629	0.0860	0.0497	0.0693	0.0528	0.0684	0.0538	0.0778
Macenko	0.0799	0.1359	0.0553	0.0826	0.0629	0.1122	0.0678	0.1221
Vahadane	0.1127	0.2112	0.0877	0.1404	0.0741	0.1471	0.1274	0.2573
Alsubaie	0.0698	0.1262	0.1186	0.2015	0.1048	0.1540	0.1923	0.3271
Hidalgo-Gavira	0.0645	0.0915	**0.0373**	**0.0480**	0.0552	0.0795	0.0378	0.0572
Pérez-Bueno	0.0624	0.0900	0.0375	0.0492	0.0506	0.0740	**0.0351**	**0.0539**
Zheng	**0.0477**	**0.0616**	0.0394	0.0519	0.0396	**0.0516**	0.0551	0.0693
ℓ_1 prior $n_s = 2$	0.0532	0.0775	0.0376	0.0491	**0.0357**	0.0549	0.0532	0.0785
ℓ_1 prior $n_s = 3$	0.0793	0.1136	0.0493	0.0622	0.0617	0.0910	0.0457	0.0708

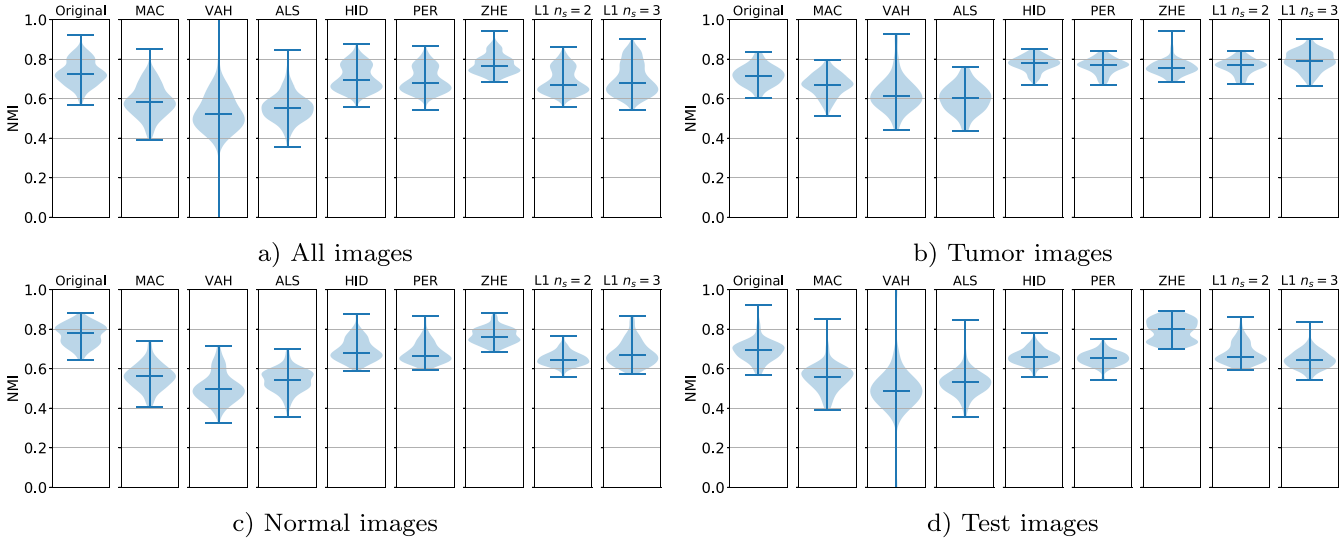


Fig. 9. Violin plots of NMI values for the normalized patches from Camelyon-16 dataset by the compared methods. The bars mark the maximum, median and minimum values for each plot.

correctly identify the H&E distribution on the WSIs. When CN is applied, the color distribution is equalized for all the WSIs and the color properties of each stain are fixed to those in the reference image, reducing the NMI SD and CV values.

The CN analysis on Camelyon databases is provided below. Due to the computational cost of CD and parameter estimation (See Fig. 6) on the large volume of WSIs in those databases and also to the superior performance in previous experiments (See Tables 4 & 5) only the proposed L1, and not LOG, was used in the comparison.

In addition to undesired color variance due to the staining procedure and also to the acquisition system used, pathology related color variations also appear in the WSIs (e.g: tumor images usually have a higher percentage of hematoxylin pixels). The fully labeled Camelyon-16 allows us to study the pathological color variance. For that matter, NMI SD and NMI CV were calculated for the whole dataset and for the tumor, normal and test WSIs as separated subsets. NMI SD and NMI CV values obtained for the Camelyon-16 dataset are shown in Table 6 and Fig. 9. The best result for the complete dataset is obtained by ZHE, closely followed by our proposed L1. However, in the separated normal and tumor subsets, the proposed method obtained the best values. Images normalized by our method are more similar to those in the same subset, but the difference between classes is preserved. The proposed L1 method with $n_s = 3$ obtains higher NMI values than the original dataset when all images are considered, however it is reduced in the normal and tumor subsets. This is caused by a wide separation on the colors for the hematoxylin and eosin channels, that will be useful for classification as we will see in the following sections.

In Fig. 9(a–c) we observe that the NMI variation on Camelyon-16 dataset comes not only from different centers but also from different pathologies. Images in tumor, and normal image subsets show different distributions on the original data. The normalized images by HID, PER, and the proposed L1 preserve those differences, keeping a separation on the median NMI value of both subsets. ZHE, designed to optimize NMI values, tends to overnormalize the images, eliminating most of the NMI difference between tumor and normal subsets.

Table 7
NMI SD and NMI CV comparison for diferent normalization methods on Camelyon-17 dataset.

	All images		Non-Negative		Negative	
Method	SD	CV	SD	CV	SD	CV
Original Data	0.0773	0.1035	0.0812	0.1087	0.0750	0.1004
Macenko	0.1031	0.1689	0.0993	0.1581	0.1040	0.1731
Vahadane	0.1058	0.1823	0.1010	0.1685	0.1069	0.1878
Alsubaie	0.0992	0.1806	0.0989	0.1753	0.0984	0.1819
Hidalgo-Gavira	0.0635	0.0948	0.0671	0.0987	0.0606	0.0913
Pérez-Bueno	0.0629	0.0941	0.0668	0.0984	0.0598	0.0902
Zheng	**0.0489**	**0.0631**	**0.0488**	**0.0628**	**0.0489**	**0.0632**
ℓ_1 prior $n_s = 2$	0.0624	0.0935	0.0720	0.1051	0.0534	0.0813
ℓ_1 prior $n_s = 3$	0.0793	0.1136	0.0684	0.1037	0.0638	0.0994

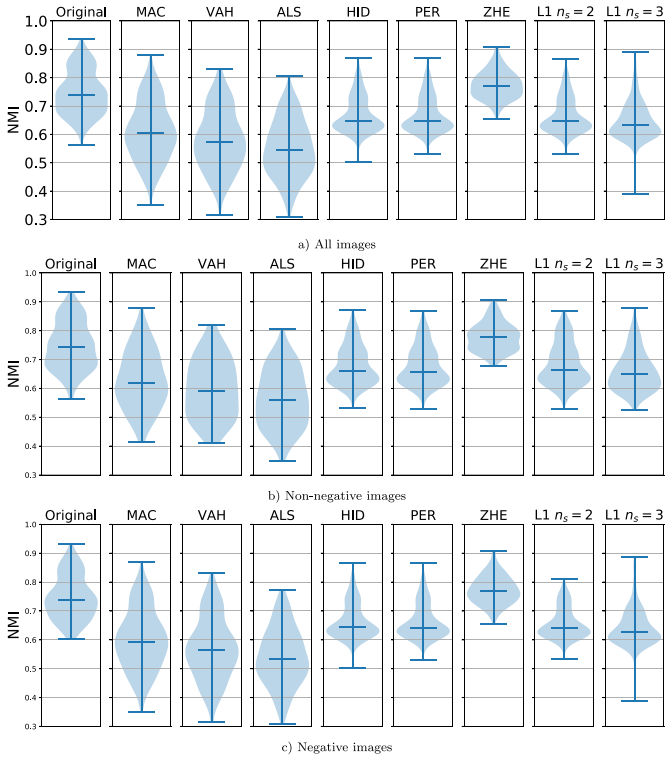


Fig. 10. Violin plots of NMIs for the normalized patches from Camelyon-17 dataset by the compared methods. The bars mark the maximum, median and minimum values for each plot.

The NMI values obtained for Camelyon-17 are shown in Table 7 and Fig. 10. The labeling is more complex on this dataset (See Table 2), so the NMI is calculated for the negative (normal WSIs) and non negative (itc, micro and macro) subsets, along with the full dataset. The original WSIs from Camelyon-17 have larger color variations than previous datasets, although they are not as balanced in terms of normal and tumoral WSIs. From Fig. 10(a–c) it can be appreciated that the subset distributions are similar to the whole dataset distribution, meaning that the NMI differences caused by pathologies are overwhelmed by the differences between centers. In the non-negative subset there is also variation due to the significant differences between itc, micro and macro. The lower NMI SD and CV values are obtained by ZHE. The proposed L1 with $n_s = 2$ obtains the second lowest value in most cases. L1 obtained its lowest values on the negative subset while mantaining a wide distribution on the non negative, probably due to the inter-subset differences mentioned. The Bayesian methods HID and PER show similar results to the proposed one.

To conclude this section we include Fig. 11 to qualitatively compare the CNs obtained by the competing methods. The reference image and some of their 224×224 extracted patches are shown in the first row. The remaining rows contain patches from different WSIs in the Camelyon-16 dataset normalized using the competing methods. MAC and VAH tend to saturate the color in the images. ALS introduces artifacts in some of the patches. ZHE overbrighten the images. HID and PER effectively transformed the color to that of the reference image. The proposed L1 keeps the structure and tissue differences but set the stain properties to those observed in the reference image. The normalization with the proposed L1 and $n_s = 2$ is the most similar to the reference image. When using a third component, the eosin is clearer. and more distinguishable from the hematoxylin. The difference between patches is higher, but the stains keep the common color properties. The effect of the residual component is clearly appreciated in the first and third rows of the last column. Although the removal of the residual produces artifacts, small hematoxylin structures are eliminated and nuclei appear more clearly separated. As discussed in previous sections, discarding the residual reduces the fidelity to the original image. In the following sections, we will demonstrate the beneficial effect of the third component on classification tasks.

4.3. Deconvolution based classification

BCD allows CAD systems to use the single stained bands separately, which can improve the classifier performance [4]. The separated H&E concentrations are used to extract features and train four different classifiers. The prostate cancer histopathological SICAPv1 database [4] was used for this purpose. In the $10\times$ scale, we use patches of size 1024×1024 pixels with the purpose of capturing complete glands within the patches. Training patches have 50% overlap and we discarded those containing mostly background (75%). From the WSI annotated as benign we obtained 1909 negative patches. A minimum of 25% of malignant tissue was required for malignant patches, obtaining 344 pathological ones.

The proposed and competing methods were used to color deconvolve the dataset. Following [4], the hematoxylin and eosin OD concentration images were used to extract the concatenation of Geodesic granulometries (GeoGran) [4] and Local Binary Patterns Variance (LBPV) [41] features. The H&E GeoGran descriptor was proposed in [4] for prostate cancer classification, obtaining stain-specific information. From the hematoxylin, it recovers the gland frontiers formed by the nuclei structure (those that enclosed their lumen and cytoplasm). From the eosin, it encodes how the stroma is affected by the lumen and nuclei structure. LBPV features are extracted from the hematoxylin band to capture texture and contrast information. The use of both Geogran and LBPV features, recovers texture and structural information in the stain separated bands, and has been proven to be an accurate descriptor for histopathological image classification [4].

With the described descriptors, the following set of state-of-art classifiers were trained: Random Forest (RF) [42,43], Extreme Gradient Boosting (XgBoost) [44], Gaussian Processes (GP) [45] and Deep Gaussian Processes (DGP)[46]. The classifiers were configured following [4] to achieve an unbiased classification benchmark. For RF and XgBoost we use 1000 estimators and a maximum depth of 20 and 30, respectively. The learning rate for XgBoost is fixed to 0.01. GP and DGP classifiers were configured following the same approach as in [4]. A GP classifier with Radial Basis Function (RBF) kernel [47] using variational inference and a three-layer DGP classifier with RBF kernel and 100 inducing points per layer, following the doubly stochastic inference proposed in [48]. DGP uses a minibatch size of 1000 and the inducing points were initialized using kmeans. Both models GP and DGP were optimized using Adam with a learning rate of 0.01.

To tackle the unbalance of positive and negative patches (common in cancer classification), we use a five-fold cross-validation. Each patient is assigned to a single fold to avoid correlation between training and testing sets. With this configuration, each classifier was built using all positive patches and a subset of the negative ones. The classifiers were trained from scratch using each deconvolution method

AUCs obtained by all the compared methods are shown in Table 8 and Fig. 12. Since HID oversmooths the images, it performs worse as it happens to methods like MAC which obtain less detailed images. ZHE scores poorly even when its deconvolution step is based on RUI. Although L1 and LOG using $n_s = 2$ do not obtain the best results, the use of $n_s = 3$ leads to a performance comparable to RUI and PER. With XgBoost, LOG obtains the best result. Notice that the best AUC (0.9856) is obtained using PER and GP, closely followed by RUI, L1, and LOG. Notice also that L1 and

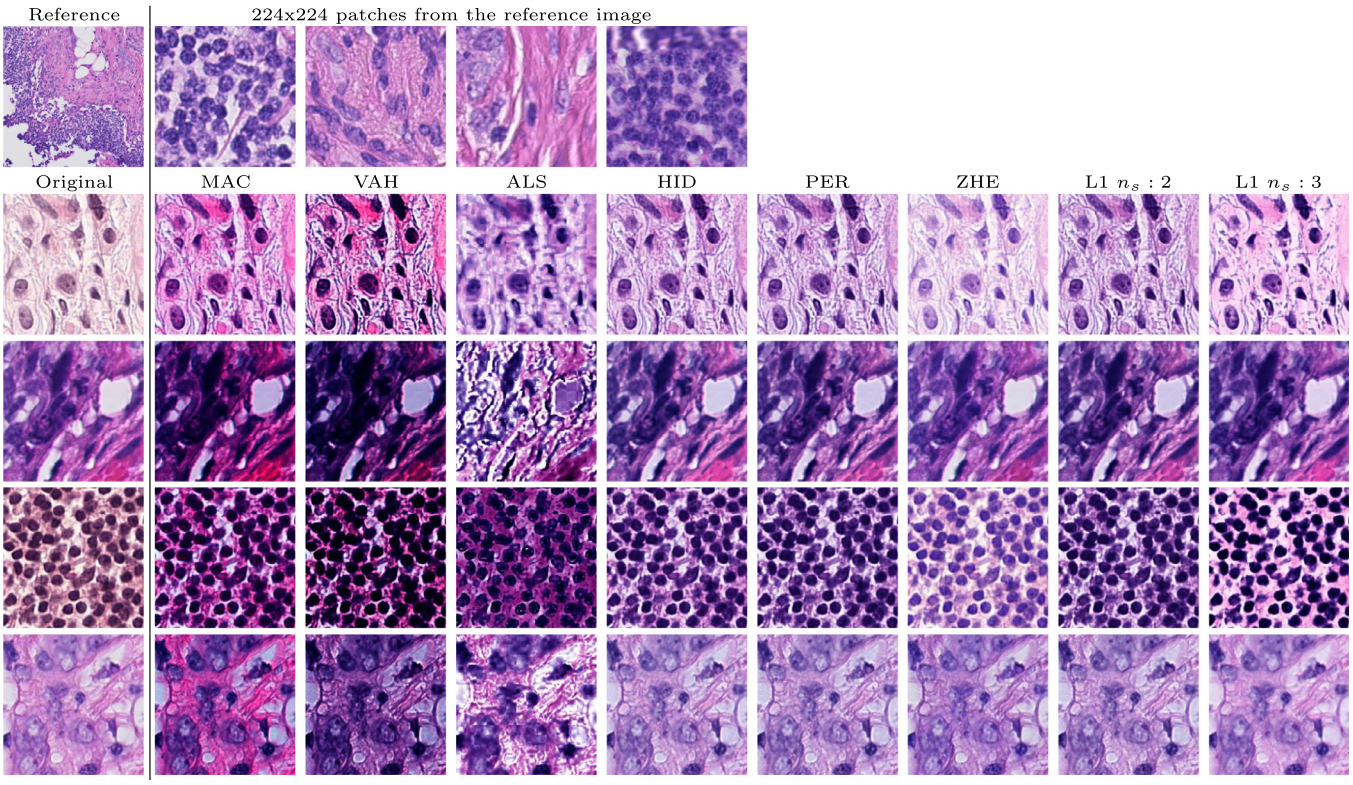


Fig. 11. Example patches from different WSIs in Camelyon-16. The first row shows the reference image used an a sample of 224×224 patches extracted from the reference. The original patch is shown in the first column.

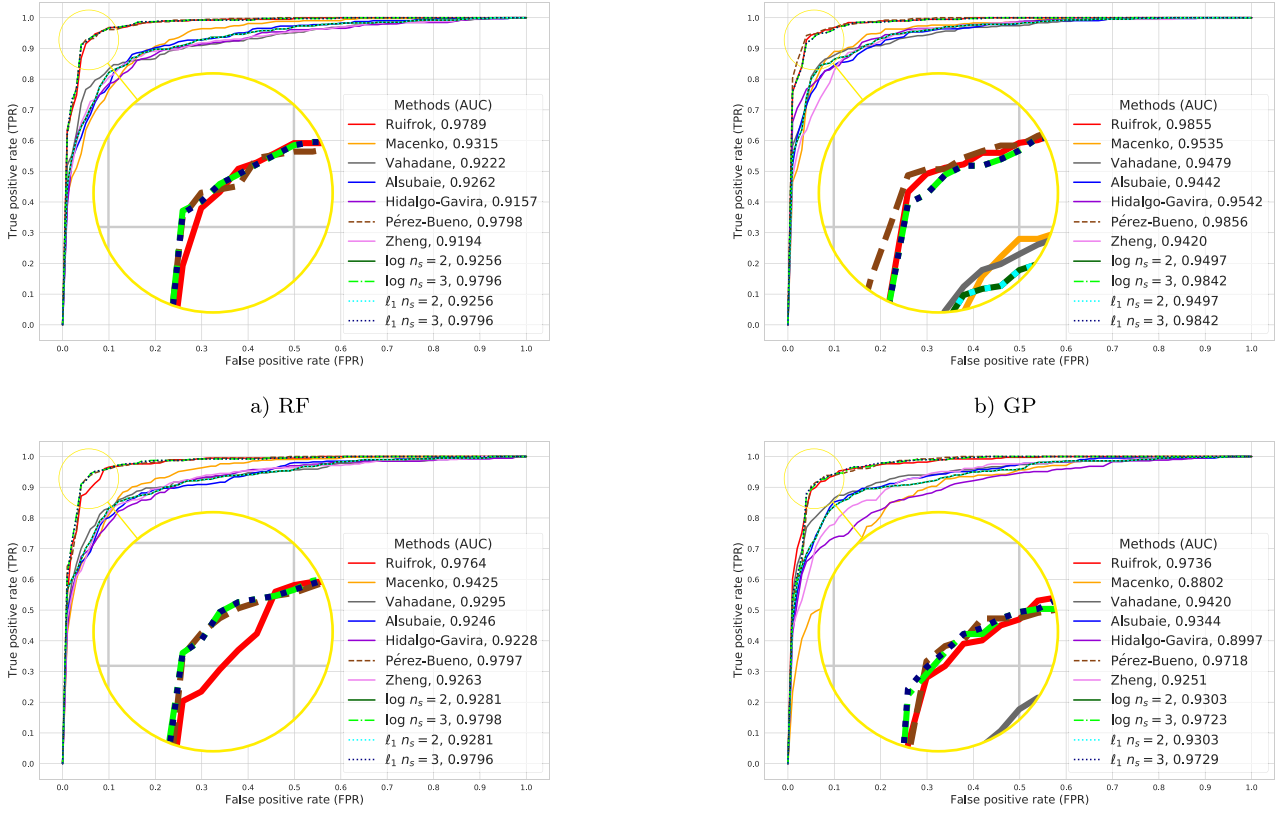


Fig. 12. ROC curves and AUC for the competing methods and classifiers on the SICAPv1 dataset. Each sub-image contains a single classifier trained with all deconvolution methods.

Table 8
AUC obtained by different classifiers when trained with different deconvolution methods on the SICAPv1 cross validation.

Method	RF	GP	XgBoost	DGP
RUI	0.9789	0.9855	0.9764	**0.9737**
MAC	0.9315	0.9535	0.9425	0.8802
VAH	0.9222	0.9479	0.9295	0.9420
ALS	0.9262	0.9442	0.9246	0.9344
HID	0.9157	0.9542	0.9228	0.8997
PER	**0.9798**	**0.9856**	0.9797	0.9718
ZHE	0.9194	0.9420	0.9263	0.9251
LOG $n_s = 2$	0.9256	0.9497	0.9281	0.9303
LOG $n_s = 3$	0.9796	0.9842	**0.9798**	0.9723
L1 $n_s = 2$	0.9256	0.9497	0.9281	0.9303
L1 $n_s = 3$	0.9796	0.9842	0.9796	0.9729

LOG perform very similarly. This is due to the very close estimated color vector matrix which leads to very similar extracted features.

The results obtained by L1 and LOG are in agreement with those obtained in our previous work [5]. Including a third residual component ($n_s = 3$) in the deconvolution step leads to better classification performance although the obtained stain separation is not as close to the ground-truth separation as that obtained using $n_s = 2$. Despite of a lower fidelity, the information captured by the residual channel makes the nuclei in the hematoxylin channel to appear more clearly separated and with less noise. The distribution of nuclei is usually considered be the most determinant feature for classification [4]. We believe this is the most plausible reason for the discriminative power of the residual band.

4.4. Normalization based classification

As we have already indicated, CN can be considered as a preprocessing step whose goal is to increase the performance of CAD systems[49], specially those using as input the original RGB images. To conclude the experimental section, in our last experiment we compare the performance of VGG19 [50], a common CNN used in cancer classification [4,38], when it is fed with the original and color normalized patches. We also analyze the VGG19 performance when trained and tested using the OD concentrations obtained by the different methods, as they can be seen as a two channel image. Fig. 13 shows an example of Camelyon-16 patches and their OD concentration channels.

From the patches extracted in Section 4.2, 55,000 tumor annotated (positive class) patches and 55,000 normal (negative class) patches from negative WSIs were randomly sampled from each Camelyon dataset training set, see Section 3. Since Camelyon-17 contains only 50 tumor annotated WSIs, to complete its 55,000 tumor annotated patches, additional tumor patches were extracted following the procedure described in Section 4.2. Using the above protocol, Camelyon-16 testing set contains approximately 19000 tumor patches, and from this testing dataset 19000 normal patches were sampled. VGG19 was trained and tested in two scenarios. In the first case, we explore how normalization affects performance within a single database (using Camelyon-16 training and testing set). In the second scenario we use Camelyon-17 for training and since Camelyon-17 test set labels are not available, Camelyon-16 is used for test. This experiment provides information on the inter-generalization capabilities of the model.

VGG19 with batch normalization was trained during 100 epochs in each case, which was enough for the network to converge. A batch size of 64 samples was used, constrained by the available memory of the Nvidia Titan X GPU utilized in this work. The learning rate was initially set to 0.01 and was reduced by factor 0.5 each 30 epochs. AUCs were calculated on the test set using the training best performing epoch for each method.

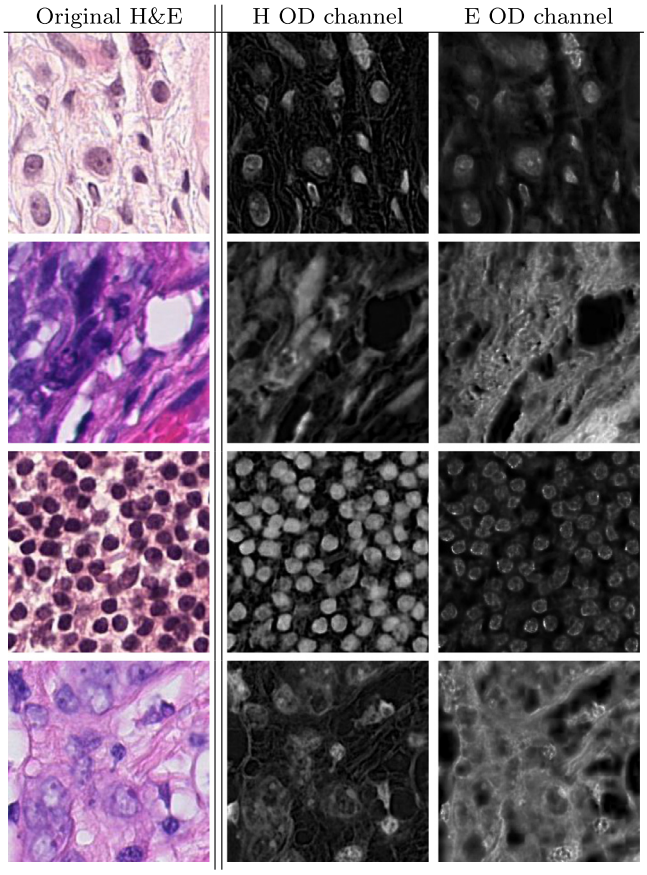


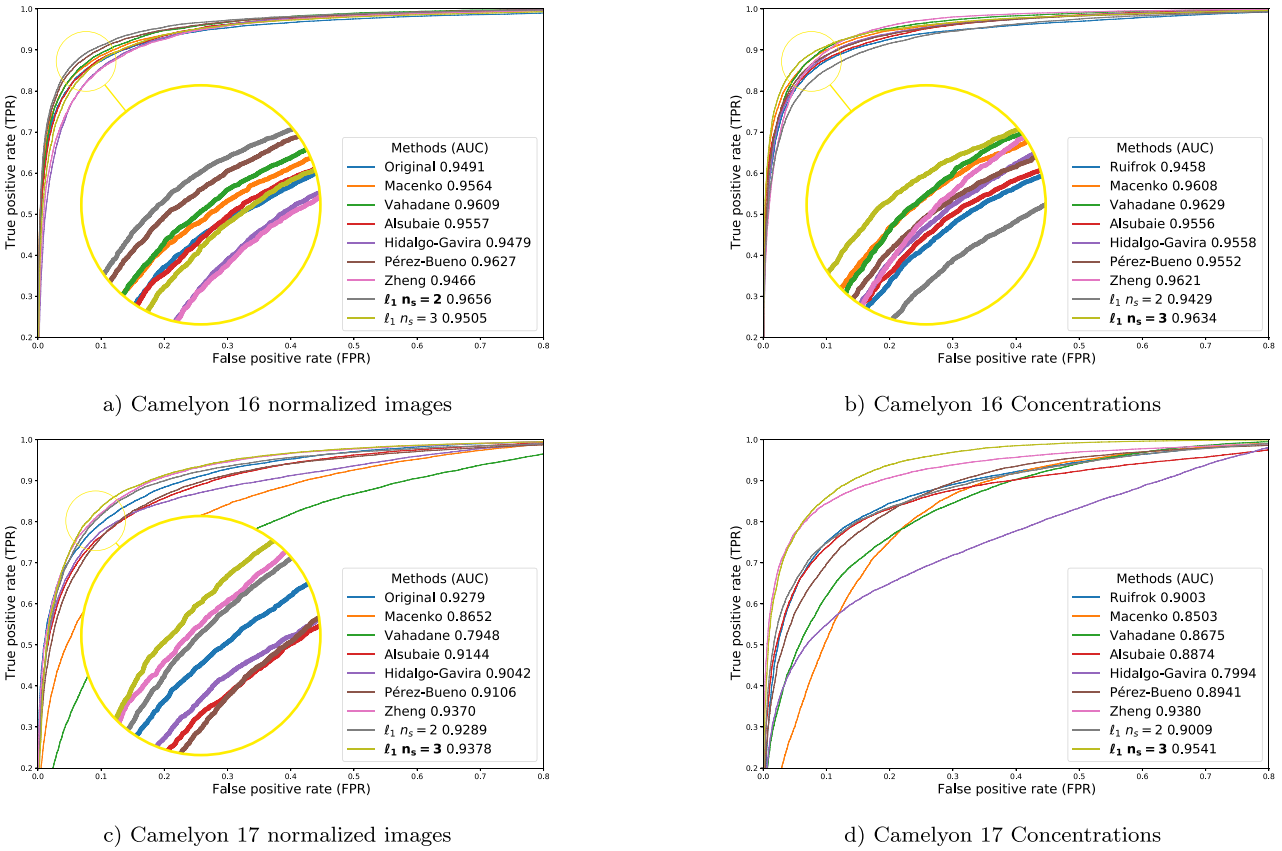
Fig. 13. Examples of the OD concentrations channels obtained by the proposed method L1 $n_s = 2$ for different patches and used to train VGG19 using a 2-band image as input.

The obtained AUCs are shown in Table 9 and the ROC curves in Fig. 14. Notice that for Camelyon-16, VGG19 performs well on the original images (better than some of the methods). The proposed L1 with $n_s = 2$ is however the best feed to VGG19 since its AUC increases from 0.9479 (original data) to 0.9656. The oversmoothing of the edges by HID and the overnormalization of ZHE obtained a slightly lower value than the non-normalized original data. The use as input to VGG19 of Camelyon-16 OD concentrations, was a boost for the methods ZHE and the proposed L1 using $n_s = 3$, and had a slightly beneficial effect for most methods.

Camelyon-17 training set contains more WSIs than Camelyon-16, furthermore its color variance is considerable as images come from 5 different centers. An adequate preprocessing has a higher impact on the generalization capability of the CNN. In this case,

Table 9
AUC Performance of the VGG19 over Cameylon-16 testing set using CN images and OD concentrations obtained by the proposed and competing deconvolution methods.

Method	Training set			
	Camelyon-16		Camelyon-17	
	CN	OD	CN	OD
Original images	0.9491	NA	0.9279	NA
RUI	NA	0.9458	NA	0.9003
MAC	0.9564	0.9608	0.8652	0.8503
ALS	0.9557	0.9556	0.9144	0.8874
HID	0.9479	0.9558	0.9042	0.7994
PER	0.9627	0.9552	0.9106	0.8941
ZHE	0.9466	0.9621	0.9370	0.9380
L1 $n_s = 2$	**0.9656**	0.9429	0.9289	0.9009
L1 $n_s = 3$	0.9505	**0.9634**	**0.9378**	**0.9541**



a) Camelyon 16 normalized images

b) Camelyon 16 Concentrations

c) Camelyon 17 normalized images

d) Camelyon 17 Concentrations

Fig. 14. ROC curve and AUC obtained by the VGG19 when trained with normalized images and OD concentrations by competing methods. Testing set is always the one from Camelyon 16. a) Training set from Camelyon 16 normalized. b) Training set from Camelyon 16 concentrations. c) Training set from Camelyon 17 normalized. d) Training set from Camelyon 17 concentrations.

the original data reached an AUC = 0.9279. Using CN, the proposed L1 with $n_s = 3$ obtained the best result with 0.9377. ZHE performs better in this experiment than in the previous one. In this inter database case, only ZHE and L1 with $n_s = 3$ were boosted by the use of OD concentrations to train the network. The L1 AUC raised to 0.9541 when using $n_s = 3$.

The effect of using a third component was limited using normalized images in the Camelyon-16 dataset. However, this configuration obtained the best performance in OD and in both cases when using Camelyon-17. As discussed in previous sections, the most plausible reason is that the third residual component makes nuclei to appear more different from other structures. This effect can be also appreciated on normalized images in Fig. 11. Where the patches in the last column show a bigger difference between hematoxylin and eosin colors.

5. Discussion

BCD is a critical step towards normalization and classification of histological images. The stain separation allows to measure the fidelity to the tissue and facilitate feature extraction. The obtained results clearly show that SG priors are a good choice for color deconvolution of histopathological images. As previously indicated, each stain should fix only and completely to specific proteins on the tissue, leading to sparse stain concentration differences at neighbouring pixels [11]. However, the experimental results show that the sparsity on the differences is moderated. The ℓ_1 prior, with a lower kurtosis than the log prior, allows to keep more non-zero values. This makes ℓ_1 a good prior for this problem, as its induced sparsity is softer than that of the log prior.

We have analyzed the effect of using two or three stain components in our proposed approach to deconvolution. The carried out experiments indicate that using two components produces stains closer to the original ones and also provides good CN. The use of a third component to capture residual information from the H&E images, makes it possible to obtain a clearer stain separation. In the hematoxylin band, the nuclei appear more clearly enhancing nucleus information and the noisy background is reduced. The effect of the third component in the eosin band is reduced but the contrast is increased. Then, we should choose whether to use the third component for BCD depending on our goal. Its use may reduce the fidelity to the tissue in terms of PSNR and SSIM values, but it improves the performance of feature based and CNN classification methods, improving class separation and helping the descriptors or CNN layers to capture the relevant information.

Finally, the use of BCD allows to extract stain-specific information from H&E channels. Our comparison between classification using the normalized images and OD concentrations have shown that CN of histopathological images improves the performance of CNN methods, however the use of CD to obtain the separated H&E concentrations leads to better performance. The H&E separation is directly provided to the CNN by the OD concentration and directly related to a better class separation.

6. Conclusions

In this work we have proposed the use of SG priors for blind color deconvolution of histological images. The framework presented includes a novel variational Bayesian blind color deconvolution algorithm which automatically estimates the color-vector ma-

trix, the concentration of stains, and all the model parameters. SG priors are used to model neighbouring pixel differences. The use of the SG family is a powerful tool to fine tuning the sparsity of concentration differences, reducing the noise in the images while preserving the tissue structure without oversmoothing the edges. Two penalty functions, named L1 and LOG, corresponding to SG distribution have been used. The information obtained through the proposed deconvolution guarantees fidelity to the tissue structure and can be used both for normalization and classification of histological images.

The proposed LOG and L1 methods have been experimentally compared to classical and state-of-art methods on a set of experiments covering the most common histological color deconvolution related tasks: stain separation, image normalization and cancer classification. They obtained very good results on all the performed experiments.

We have analyzed the effect of using a third residual stain component during deconvolution, showing that an affordable reduction of the fidelity to the tissue improves classification performance using descriptors or CNN classifiers

Finally, our study includes a comparison between classification using the normalized images and OD concentrations showing that although CN improves the performance of classifiers over the raw data, stain separated OD concentrations lead to better classification performance.

Statements of ethical approval

The data used in this publication has been either publicly available or have been ethically approved for scientific use by the Ethics committee of the San Cecilio University Hospital (Granada, Spain).

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Acknowledgments

This work was sponsored in part by the Agencia Estatal de Investigación under project PID2019-105142RB-C22 / AEI / 10.13039/501100011033, Junta de Andalucía under project PY20_00286,and the work by Fernando Pérez-Bueno was sponsored by Ministerio de Economía, Industria y Competitividad under FPI contract BES-2017-081584. Funding for open access charge: Universidad de Granada / CBUA.

References

[1] A.H. Fischer, K.A. Jacobson, J. Rose, R. Zeller, Hematoxylin and eosin staining of tissue and cell sections, Cold Spring Harbor Protocols (2008).
[2] M. Salvi, U.R. Acharya, F. Molinari, K.M. Meiburger, The impact of pre- and post-image processing techniques on deep learning frameworks: a comprehensive review for digital pathology image analysis, Comput. Biol. Med. 128 (2021) 104129, doi:10.1016/j.compbiomed.2020.104129.
[3] T.A.A. Tosta, P.R. de Faria, L.A. Neves, M.Z. do Nascimento, Computational normalization of H&E-stained histological images: progress, challenges and future potential, Artif. Intell. Med. 95 (2019) 118–132.
[4] A.E. Esteban, M. Lopez-Perez, A. Colomer, M.A. Sales, R. Molina, V. Naranjo, A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes, Comput. Methods Programs Biomed. 178 (2019) 303–317.
[5] F. Prez-Bueno, M. Lpez-Prez, M. Vega, J. Mateos, V. Naranjo, R. Molina, A.K. Katsaggelos, A TV-based image processing framework for blind color deconvolution and classification of histological images, Digit. Signal Process. 101 (2020) 102727.
[6] F.G. Zanjani, S. Zinger, B.E. Bejnordi, J.A.W.M. van der Laak, P.H.N. de With, Stain normalization of histopathology images using generative adversarial networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI), 2018, pp. 573–577.
[7] M. Shaban, C. Baur, N. Navab, S. Albarqouni, StainGAN: stain style transfer for digital histological images, vol. 2019-April, 2019, pp. 953–956, doi:10.1109/ISBI.2019.8759152. Cited By 41
[8] A.C. Ruifrok, D.A. Johnston, Quantification of histochemical staining by color deconvolution, Anal. Quant. Cytol. Histol. 23 (2001) 291–299.
[9] E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images, IEEE Comput. Graph. Appl. 21 (5) (2001) 34–41.
[10] A. Rabinovich, S. Agarwal, C. Laris, J.H. Price, S.J. Belongie, Unsupervised color decomposition of histologically stained tissue samples, in: Advances in Neural Information Processing Systems, 2004, pp. 667–674.
[11] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A.M. Schlitter, I. Esposito, N. Navab, Structure-preserving color normalization and sparse stain separation for histological images, IEEE Trans. Med. Imaging 35 (2016) 1962–1971.
[12] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N.N. Shih, H. Gilmore, A. Madabhushi, Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis, Comput. Med. Imaging Graph. 46 (2015) 20–29.
[13] M. Macenko, M. Niethammer, et al., A method for normalizing histology slides for quantitative analysis, in: International Symposium on Biomedical Imaging (ISBI), 2009, pp. 1107–1110.
[14] M.T. McCann, J. Majumdar, et al., Algorithm and benchmark dataset for stain separation in histology images, in: International Conference on Image Processing (ICIP), 2014, pp. 3953–3957.
[15] A. Anghel, M. Stanisavljevic, S. Andani, N. Papandreou, J.H. Rschoff, P. Wild, M. Gabrani, H. Pozidis, A high-performance system for robust stain normalization of whole-slide images in histopathology, Front. Med. 6 (2019) 193, doi:10.3389/fmed.2019.00193.
[16] D. Carey, V. Wijayathunga, A. Bulpitt, D. Treanor, A novel approach for the colour deconvolution of multiple histological stains, in: Proceedings of the 19th Conference of Medical Image Understanding and Analysis, 2015, pp. 156–162.
[17] M. Gavrilovic, J.C. Azar, et al., Blind color decomposition of histological images, IEEE Trans. Med. Imaging 32 (2013) 983–994.
[18] A.M. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, IEEE Trans. Biomed. Eng. 61 (6) (2014) 1729–1738.
[19] M. Salvi, N. Michielli, F. Molinari, Stain color adaptive normalization (scan) algorithm: separation and standardization of histological stains in digital pathology, Comput. Methods Programs Biomed. 193 (2020) 105506.
[20] N. Trahearn, D. Snead, I. Cree, N. Rajpoot, Multi-class stain separation using independent component analysis, in: Medical Imaging 2015: Digital Pathology, 2015, p. 94200J.
[21] N. Alsubaie, S.E.A. Raza, N. Rajpoot, Stain deconvolution of histology images via independent component analysis in the wavelet domain, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016, pp. 803–806.
[22] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, N. Rajpoot, Stain deconvolution using statistical analysis of multi-resolution stain colour representation, PLOS ONE 12 (2017) e0169875.
[23] L. Astola, Stain separation in digital bright field histopathology, in: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2016, pp. 1–6.
[24] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, C. Xue, Adaptive color deconvolution for histological WSI normalization, Comput. Methods Programs Biomed. 170 (2019) 107–120.
[25] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, A.K. Katsaggelos, Variational Bayesian blind color deconvolution of histopathological images, IEEE Trans. Image Process. 29 (1) (2020) 2026–2036.
[26] S.D. Babacan, R. Molina, M.N. Do, A.K. Katsaggelos, Blind deconvolution with general sparse image priors, ECCV, 2012.
[27] X. Zhou, M. Vega, F. Zhou, R. Molina, A.K. Katsaggelos, Fast Bayesian blind deconvolution with Huber super Gaussian priors, Digit. Signal Process. 60 (2017) 122–133.
[28] F. Pérez-Bueno, M. Vega, V. Naranjo, R. Molina, A.K. Katsaggelos, Super Gaussian priors for blind color deconvolution of histological images, in: International Conference on Image Processing (ICIP), 2020.
[29] F. Pérez-Bueno, M. Vega, V. Naranjo, R. Molina, A.K. Katsaggelos, Fully automatic blind color deconvolution of histological images using super Gaussians, in: 27th European Signal Processing Conference, EUSIPCO 2020, 2020.
[30] R. Rockafellar, Convex Analysis, Princeton University Press, 1996.
[31] C. Bishop, Pattern Recognition and Machine Learning, Springer, pp. 454–455.
[32] S. Kullback, Information Theory and Statistics, Dover Pub., 1959.
[33] C.L. Srinidhi, O. Ciga, A.L. Martel, Deep neural network models for computational histopathology: a survey, 2019.
[34] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q.F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, J. van der Laak, 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset, GigaScience 7 (6) (2018). Giy065
[35] P. Bndi, O. Geessink, Q. Manson, M. Van Dijk, M. Balkenhol, M. Hermsen, B. Ehteshami Bejnordi, B. Lee, K. Paeng, A. Zhong, Q. Li, F.G. Zanjani, S. Zinger, K. Fukuta, D. Komura, V. Ovtcharov, S. Cheng, S. Zeng, J. Thagaard, A.B. Dahl, H. Lin, H. Chen, L. Jacobsson, M. Hedlund, M. etin, E. Halc, H. Jackson, R. Chen, F. Both, J. Franke, H. Ksters-Vandevelde, W. Vreuls, P. Bult, B. van Ginneken, J. van der Laak, G. Litjens, From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge, IEEE Trans. Med. Imaging 38 (2) (2019) 550–560.
[36] G. Landini, Colour deconvolution, https://blog.bham.ac.uk/intellimic/g-landini-software/colour-deconvolution/, Accessed: 2019-10-30.

[37] A. Kolaman, O. Yadid-Pecht, Quaternion structural similarity: a new quality index for color images, IEEE Trans. Image Process. 21 (2011) 1526–1536, doi:10.1109/TIP.2011.2181522.
[38] C.L. Srinidhi, O. Ciga, A.L. Martel, Deep neural network models for computational histopathology: a survey, Med. Image Anal. 67 (2021) 101813, doi:10.1016/j.media.2020.101813.
[39] A. Basavanhally, A. Madabhushi, Em-based segmentation-driven color standardization of digitized histopathology, vol. 8676, 2013, p. 86760G.
[40] B.E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Hller, A. Homeyer, N. Karssemeijer, J.A.v.d. Laak, Stain specific standardization of whole-slide histopathological images, IEEE Trans. Med. Imaging 35 (2) (2016) 404–415.
[41] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recognit. 43 (3) (2010) 706–719.
[42] M. Valkonen, K. Kartasalo, K. Liimatainen, M. Nykter, L. Latonen, P. Ruusuvuori, Metastasis detection from whole slide images using local features and random forests, Cytometry Part A 91 (6) (2017) 555–565.
[43] A. Paul, D.P. Mukherjee, P. Das, A. Gangopadhyay, A.R. Chintha, S. Kundu, Improved random forest for classification, IEEE Trans. Image Process. 27 (8) (2018) 4012–4024.
[44] A. Pimkin, G. Makarchuk, V. Kondratenko, M. Pisov, E. Krivov, M. Belyaev, Ensembling neural networks for digital pathology images classification and segmentation, Lect. Notes Comput. Sci. 10882 LNCS (2018) 877–886.
[45] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2006.
[46] A. Damianou, N. Lawrence, Deep Gaussian processes, J. Mach. Learn. Res. 31 (2013) 207–215.
[47] M. Opper, C. Archambeau, The variational Gaussian approximation revisited, Neural Comput. 21 (3) (2009) 786–792.
[48] H. Salimbeni, M. Deisenroth, Doubly stochastic variational inference for deep Gaussian processes, in: Advances in Neural Information Processing Systems, 2017, pp. 4591–4602.
[49] D. Tellez, G. Litjens, P. Bndi, W. Bulten, J.-M. Bokhorst, F. Ciompi, J. van der Laak, Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology, Med. Image Anal. 58 (2019) 101544, doi:10.1016/j.media.2019.101544.
[50] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR (2015) arXiv:1409.1556.