BLIND COLOR DECONVOLUTION AND CLASSIFICATION OF HISTOLOGICAL IMAGES USING THE HYPERBOLIC SECANT PRIOR.

Francisco M. Castro-Macías[a,b,*], *Fernando Pérez-Bueno*[a,b], *Miguel Vega*[c], *Javier Mateos*[a], *Rafael Molina*[a], *Aggelos K. Katsaggelos*[d]

[a] Depto. de Ciencias de la Computación e I.A., Universidad de Granada, Spain.
[b] Centro de Investigación en Tecnologías de la Información y las Comunicaciones, Universidad de Granada, Spain.
[c] Depto. de Lenguajes y Sistemas Informáticos, Universidad de Granada, Spain.
[d] Dept. of Electrical Engineering and Computer Science, Northwestern University, USA.

ABSTRACT

In this paper, we present a novel approach to Blind Color Deconvolution, a stain separation technique useful for normalizing, augmenting, and automatically diagnosing histological images. To robustly estimate the stain colors and concentrations, we follow the Bayesian framework and introduce a Gaussian prior distribution on the color vectors and a Hyperbolic Secant prior on the concentrations, which is a seldom explored image model.

We provide a comprehensive mathematical derivation of our inference procedure, outlining the underlying principles and assumptions. To demonstrate its effectiveness, we conduct two experiments. The first experiment assesses the fidelity of the reconstructed images to the underlying tissue characteristics. The second experiment uses it as a preprocessing step for a multicenter breast cancer classification task. Our results reveal the superior performance of the proposed method, underscoring its potential for advancing histological image processing and AI-assisted diagnosis.

Index Terms— Blind Color Deconvolution, Bayesian modeling, histological images

1. INTRODUCTION

Histological images are stained to highlight the tissue structure, enabling pathologists to distinguish the elements in the tissue by their distinctive staining. The main objective of Blind Color Deconvolution (BCD) is to disentangle the stain mixture in the observed RGB image, obtaining a color-vector matrix, and a concentration matrix that represents the amount of stain per pixel. BCD can be used to normalize and augment histological images, but also to feed Computer-Assisted Diagnosis (CAD) systems, which benefit from stain-separated data [1, 2, 3]. Furthermore, BCD has been proven to be especially important in standardizing histological image analysis and mitigating performance variations when dealing with images from different sources. This has been identified as a major issue in CAD system development [4, 5].

BCD methods rely on the Beer-Lambert law [6], which establishes a bilinear relationship between the observed image and the underlying stain color-vector matrix and concentrations in the logarithmically inverted Optical Density (OD) space. Estimating the color-vector matrix and concentrations from the observed image is an ill-posed problem since infinite combinations of color and concentrations can produce the observed image. Different regularization and optimization techniques have been explored to estimate the unknowns (see [4] for a review), such as Non-Negative Matrix Factorization [7, 8], Singular Value Decomposition [9, 10], Independent Component Analysis [11], or clustering [12, 13]. In the Bayesian approach, the regularization constraints are modeled as prior probability distributions. In [14], a similarity prior is proposed for the color-vector matrix, along with a Simultaneous Autoregressive prior on the concentrations to achieve smoothness. In [1], the authors suggested using a Total Variation prior on each stain, while in [2] the use of Super Gaussian (SG) distributions is used to promote sparsity in the concentrations.

In this work, we propose to use the Hyperbolic Secant (HS) distribution as a prior for the concentrations [15]. Since the HS distribution is also a SG distribution, its sparsity-promoting properties make it appropriate for the BCD task. Notably, this distribution admits a Gaussian Scale Mixture (GSM) representation, which allows us to leverage a principled data augmentation approach to approximate the posterior distribution. The proposed approach is evaluated in the stain-separation task, as well as in the breast cancer classification task. In both tasks, the use of the HS prior resulted in improved performance compared to previous BCD methods, especially those based on other SG priors.

The rest of the paper is organized as follows: The proposed Bayesian modeling and inference are presented in

*Corresponding author.

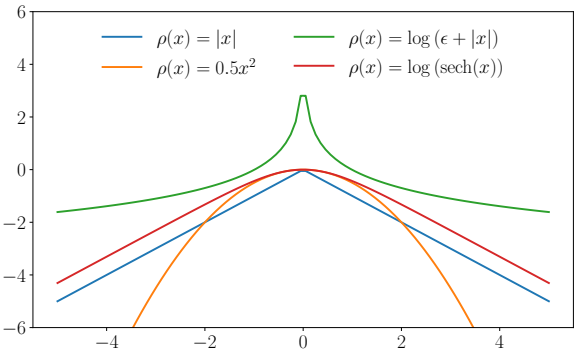

Fig. 1. Penalty functions of different SG distributions [2]: the HS (red), the Gaussian (orange), the $l1$-norm (blue), and the log (green).

sections 2 and 3, respectively. In section 4, the proposed method is experimentally evaluated and compared with other approaches in the literature. Finally, section 5 concludes the work.

2. BAYESIAN MODELING

In this work, we follow a Bayesian approach to solve the BCD problem. Each observed $H \times W$ RGB image $\mathbf{I} \in \mathbb{R}^{HW\times 3}$ is transformed into the the observed OD image $\mathbf{Y} = -\log(\mathbf{I}/i_0) \in \mathbb{R}^{HW\times 3}$, where i_0 is the maximum incident luminosity, set to 255. The Beer-Lambert law states that

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

where $\mathbf{M} \in \mathbb{R}^{3\times S}$ is the color-vector matrix, with S being the number of stains, $\mathbf{C} \in \mathbb{R}^{HW\times S}$ is the stain concentration matrix, and $\mathbf{N} \in \mathbb{R}^{HW\times 3}$ is a random noise matrix with i.i.d. zero mean Gaussian components with precision β. We will refer to the s-th column of $\mathbf{C}$ as $\mathbf{c}_s$, and to the s-th column of $\mathbf{M}$ as $\mathbf{m}_s$. Also, given $A \in \mathbb{N}$, we will denote $[A] = \{1, \ldots, A\}$.

Our goal is to estimate both $\mathbf{C}$ and $\mathbf{M}$ having observed exclusively $\mathbf{Y}$. From Eq. (1) we obtain the observation model,

$$\mathrm{p}(\mathbf{Y} \mid \mathbf{C}, \mathbf{M}, \beta) \propto \exp\left(-\frac{\beta}{2}\left\|\mathbf{Y}^\top - \mathbf{M}\mathbf{C}^\top\right\|_F^2\right), \tag{2}$$

where $\|\cdot\|_F$ represents the Frobenius norm.

Next, we define a prior for the concentrations. We explore the use of the HS distribution, which belongs to the family of the SG distributions [16]. The probability density function is given by

$$f(x \mid \alpha) = \frac{\pi}{\alpha}\,\mathrm{sech}(\alpha x), \quad \forall x \in \mathbb{R}, \tag{3}$$

where $\alpha \in (0, +\infty)$. Its penalty function is represented in Figure 1 together with the penalty functions of other SGs previously used for BCD [2]. The HS distribution has heavier tails and is more peaked than the Gaussian distribution, thus promoting sparsity. Importantly, it belongs to the family of z distributions [16] and therefore, its density can be expressed as a GSM. This representation will be particularly relevant for our inference procedure in the next section.

To take advantage of the sparsity-promoting properties of the HS distribution, we formulate the model on the filtered space [2]. Thus, we apply N high-pass filters $\{\mathbf{F}^n\}_{n=1}^N$ to obtain N filtered concentrations $\mathbf{c}_s^n = \mathbf{F}^n\mathbf{c}_s$. Denoting the components of $\mathbf{c}_s^n$ by c_{is}^n, we define the prior for the concentrations as

$$\mathrm{p}(\mathbf{C} \mid \boldsymbol{\alpha}) \propto \prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{i=1}^{HW} \mathrm{sech}(\alpha_s^n c_{is}^n), \tag{4}$$

where $\boldsymbol{\alpha} = \{\alpha_s^n \colon s \in [S], n \in [N]\}$ are the distribution precision parameters, which define the global partition function.

Although the color-vector matrix is different for each image, it is expected to be similar to a reference color-vector matrix. For Hematoxylin-eosin (H&E) staining, the color vectors proposed in [6] are typically used. Following [14], we incorporate this knowledge using the following prior,

$$\mathrm{p}(\mathbf{m}_s \mid \gamma) \propto \exp\left(-\frac{\gamma_s}{2}\left\|\mathbf{m}_s - \underline{\mathbf{m}}_s\right\|^2\right), \tag{5}$$

$$\mathrm{p}(\mathbf{M} \mid \boldsymbol{\gamma}) = \prod_{s=1}^{S} \mathrm{p}(\mathbf{m}_s \mid \gamma_s) \tag{6}$$

where $\boldsymbol{\gamma} = \{\gamma_s \colon s \in [S]\}$, and γ_s measures the confidence on the accuracy of the reference $\underline{\mathbf{m}}_s$.

With all these ingredients, the joint probability distribution is given by

$$\mathrm{p}(\mathbf{Y}, \mathbf{C}, \mathbf{M} \mid \boldsymbol{\Lambda}) = \mathrm{p}(\mathbf{Y} \mid \mathbf{C}, \mathbf{M}, \beta)\mathrm{p}(\mathbf{M} \mid \boldsymbol{\gamma})\mathrm{p}(\mathbf{C} \mid \boldsymbol{\alpha}), \tag{7}$$

where $\boldsymbol{\Lambda} = \{\beta, \boldsymbol{\gamma}, \boldsymbol{\alpha}\}$ represents the hyperparameters.

3. BAYESIAN INFERENCE

Following the Bayesian paradigm, the inference procedure will be based on the posterior distribution $\mathrm{p}(\mathbf{C}, \mathbf{M} \mid \mathbf{Y}, \boldsymbol{\Lambda})$. Since this posterior cannot be obtained in a closed form, we approximate it using variational Bayesian inference [17]. The use of the HS prior makes it impossible to obtain a closed-form expression for the approximate variational posterior. Fortunately, as we have already mentioned previously, the density of the HS distribution can be expressed as a GSM,

$$f(x \mid \alpha) = \int_0^{+\infty} \mathcal{N}\left(x \mid 0, \omega^{-1}\right) \widehat{f}(\omega \mid \alpha)\mathrm{d}\omega, \tag{8}$$

where $\widehat{f}(\omega \mid \alpha)$ is a *mixing* density whose expression can be obtained using the Pólya-Gamma distribution [18]. We leverage this representation by considering an augmented model in

which inference is tractable. We define

$$\mathrm{p}(\mathbf{C}, \boldsymbol{\omega} \mid \boldsymbol{\alpha}) = \mathrm{p}(\mathbf{C} \mid \boldsymbol{\omega})\mathrm{p}(\boldsymbol{\omega} \mid \boldsymbol{\alpha}), \tag{9}$$

$$\mathrm{p}(\mathbf{C} \mid \boldsymbol{\omega}) \propto \prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{i=1}^{HW} \mathcal{N}(c_{is}^n \mid 0, (\omega_{is}^n)^{-1}), \tag{10}$$

$$\mathrm{p}(\boldsymbol{\omega} \mid \boldsymbol{\alpha}) = \prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{i=1}^{HW} \widehat{f}(\omega_{is}^n \mid \alpha_s^n), \tag{11}$$

where $\boldsymbol{\omega} = \{\omega_{is}^n \colon i \in [HW], s \in [S], n \in [N]\}$. The original $\mathrm{p}(\mathbf{C} \mid \boldsymbol{\alpha})$ is recovered by integrating out $\boldsymbol{\omega}$ in Eq. (9). The augmented joint probability distribution is given by,

$$\mathrm{p}(\mathbf{Y}, \mathbf{C}, \mathbf{M}, \boldsymbol{\omega} \mid \boldsymbol{\Lambda}) = \mathrm{p}(\mathbf{Y} \mid \mathbf{C}, \mathbf{M}, \beta)\mathrm{p}(\mathbf{M} \mid \boldsymbol{\gamma})\mathrm{p}(\mathbf{C}, \boldsymbol{\omega} \mid \boldsymbol{\alpha}). \tag{12}$$

We follow the mean-field variational Bayesian approach to approximate $\mathrm{p}(\mathbf{C}, \mathbf{M}, \boldsymbol{\omega} \mid \mathbf{Y}, \boldsymbol{\Lambda})$ by the variational distribution $\mathrm{q}(\mathbf{C}, \mathbf{M}, \boldsymbol{\omega}) = \mathrm{q}(\mathbf{C})\mathrm{q}(\mathbf{M})\mathrm{q}(\boldsymbol{\omega})$, where $\mathrm{q}(\mathbf{M}) = \prod_s \mathrm{q}(\mathbf{m}_s)$, $\mathrm{q}(\mathbf{C}) = \prod_s \mathrm{q}(\mathbf{c}_s)$, and $\mathrm{q}(\boldsymbol{\omega}) = \prod_{n,s,i} \mathrm{q}(\omega_{is}^n)$. The variational factors, $\mathrm{q}(\mathbf{C})$, $\mathrm{q}(\mathbf{M})$, and $\mathrm{q}(\boldsymbol{\omega})$, are obtained by minimizing the Kullback-Leibler (KL) divergence between the variational approximation and the true posterior [17]. The expression for each factor, given by [17, Eq. (10.9)], is detailed below.

Update of $\mathrm{q}(\mathbf{C})$**.** For $s \in [S]$, from [17, Eq. (10.9)] we obtain $\mathrm{q}(\mathbf{c}_s) = \mathcal{N}\left(\mathbf{c}_s \mid \boldsymbol{\mu}_{\mathbf{c}_s}, \boldsymbol{\Sigma}_{\mathbf{c}_s}\right)$, with

$$\boldsymbol{\mu}_{\mathbf{c}_s} = \beta \boldsymbol{\Sigma}_{\mathbf{c}_s} \mathbf{E}_s \boldsymbol{\mu}_{\mathbf{m}_s}, \tag{13}$$

$$\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1} = \beta v_{\mathbf{m}_s} \mathbf{I} + \sum_{n=1}^{N} \mathbf{F}^{n\top} \boldsymbol{\Theta}_s^n \mathbf{F}^n, \tag{14}$$

where

$$\boldsymbol{\mu}_{\mathbf{m}_s} = \mathbb{E}_{\mathrm{q}(\mathbf{m}_s)}\left[\mathbf{m}_s\right], \quad \boldsymbol{\mu}_{\mathbf{c}_s} = \mathbb{E}_{\mathrm{q}(\mathbf{c}_s)}\left[\mathbf{c}_s\right], \tag{15}$$

$$v_{\mathbf{m}_s} = \mathbb{E}_{\mathrm{q}(\mathbf{m}_s)}\left[\|\mathbf{m}_s\|^2\right] = \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{m}_s}\right) + \left\|\boldsymbol{\mu}_{\mathbf{m}_s}\right\|^2, \tag{16}$$

$$\boldsymbol{\Theta}_s^n = \mathbb{E}_{\mathrm{q}(\boldsymbol{\omega}_s^n)}\left[\mathrm{diag}\left(\boldsymbol{\omega}_s^n\right)\right], \tag{17}$$

$$\mathbf{E}_s = \mathbf{Y} - \sum_{k \neq s} \boldsymbol{\mu}_{\mathbf{c}_k} \boldsymbol{\mu}_{\mathbf{m}_k}^{\top}. \tag{18}$$

Update of $\mathrm{q}(\mathbf{M})$**.** For $s \in [S]$, from [17, Eq. (10.9)] we deduce $\mathrm{q}(\mathbf{m}_s) = \mathcal{N}\left(\mathbf{m}_s \mid \boldsymbol{\mu}_{\mathbf{m}_s}, \boldsymbol{\Sigma}_{\mathbf{m}_s}\right)$, with

$$\boldsymbol{\mu}_{\mathbf{m}_s} = \boldsymbol{\Sigma}_{\mathbf{m}_s}\left(\gamma_s \underline{\mathbf{m}}_s + \beta \mathbf{E}_s^{\top} \boldsymbol{\mu}_{\mathbf{c}_s}\right), \tag{19}$$

$$\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1} = \left(\gamma_s + \beta \sum_{n=1}^{N} v_{\mathbf{c}_s}^n\right) \mathbf{I}, \tag{20}$$

where $v_{\mathbf{c}_s}^n = \mathbb{E}_{\mathrm{q}(\mathbf{c}_s)}\left[\|\mathbf{F}^n \mathbf{c}_s\|^2\right] = \mathrm{Tr}\left(\mathbf{F}^{n\top}\mathbf{F}^n \boldsymbol{\Sigma}_{\mathbf{c}_s}\right) + \left\|\mathbf{F}^n \boldsymbol{\mu}_{\mathbf{c}_s}\right\|^2$. Following [2], we force the color-vectors to be unitary by replacing $\boldsymbol{\mu}_{\mathbf{m}_s}$ by $\boldsymbol{\mu}_{\mathbf{m}_s} / \left\|\boldsymbol{\mu}_{\mathbf{m}_s}\right\|$ and $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1}$ by $\left\|\boldsymbol{\mu}_{\mathbf{m}_s}\right\|^2 \boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1}$.

Update of $\mathrm{q}(\boldsymbol{\omega})$**.** For $n \in [N], i \in [HW]$, applying [17, Eq. (10.9)] we obtain

$$\log \mathrm{q}(\omega_{is}^n) = \mathbb{E}_{\mathrm{q}(c_{is}^n)}\left[\log \mathcal{N}(c_{is}^n \mid 0, (\omega_{is}^n)^{-1})\right] + \tag{21}$$

$$+ \log \widehat{f}(\omega_{is}^n \mid \alpha_n) + \text{const.} \tag{22}$$

Next, we observe that $\mathbb{E}_{\mathrm{q}(c_{is}^n)}\left[\log \mathcal{N}(c_{is}^n \mid 0, (\omega_{is}^n)^{-1})\right] = \log \mathcal{N}(\xi_{is}^n \mid 0, (\omega_{is}^n)^{-1})$, with $\xi_{is}^n = \sqrt{\mathbb{E}_{\mathrm{q}(c_{is}^n)}\left[(c_{is}^n)^2\right]}$. Therefore, we obtain $\mathrm{q}(\omega_{is}^n) \propto \mathcal{N}(\xi_{is}^n \mid 0, (\omega_{is}^n)^{-1}) \widehat{f}(\omega_{is}^n \mid \alpha_n)$, that, after normalization, becomes

$$\mathrm{q}(\omega_{is}^n) f(\xi_{is}^n \mid \alpha_s^n) = \mathcal{N}(\xi_{is}^n \mid 0, (\omega_{is}^n)^{-1}) \widehat{f}(\omega_{is}^n \mid \alpha_s^n). \tag{23}$$

Finally, we note that we only require $\mathbb{E}_{\mathrm{q}(\omega_{is}^n)}\left[\omega_{is}^n\right]$ to compute $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1}$. We calculate it in closed form by adapting the procedure in [19]. We differentiate under the integral sign in Eq. (8), and use (23) to obtain

$$\mathbb{E}_{\mathrm{q}(\omega_{is}^n)}\left[\omega_{is}^n\right] = \frac{\alpha_n \tanh(\alpha_n \xi_{is}^n)}{\xi_{is}^n}. \tag{24}$$

Approximating the covariance matrix of $\mathrm{q}(\mathbf{C})$. Note that to compute $v_{\mathbf{c}_s}$ we need the inverse of $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1}$, which is a $HW \times HW$ matrix. Since this is computationally prohibitive, we follow [2] and approximate it as $\boldsymbol{\Sigma}_{\mathbf{c}_s} \approx \left(\beta v_{\mathbf{m}_s} \mathbf{I} + \sum_{n=1}^{N} \theta_s^n \mathbf{F}^{n\top}\mathbf{F}^n\right)^{-1}$, where θ_s^n is the mean of the diagonal values of $\boldsymbol{\Theta}_s^n$.

The proposed BCD method. In light of the derived variational scheme, we propose the Variational Bayesian HS BCD method outlined in Algorithm 1. Given a OD image $\mathbf{Y}$, the variational updates are applied iteratively until a convergence criteria is met. Note that to avoid inverting $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1}$ and $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1}$ when computing $\boldsymbol{\mu}_{\mathbf{c}_s}$ and $\boldsymbol{\mu}_{\mathbf{m}_s}$, we solve their associated linear systems using the Conjugate Gradient method. Finally, the color-vector and concentration for each stain s are estimated as $\boldsymbol{\mu}_{\mathbf{m}_s}$ and $\boldsymbol{\mu}_{\mathbf{c}_s}$, respectively.

4. EXPERIMENTS AND RESULTS

To assess the performance of the proposed method, we conducted two experiments. The first one evaluates the fidelity to the underlying tissue characteristics. The second one employs the processed images for a multicenter breast cancer classification task.

4.1. Stain separation quality

To assess the quality of the proposed method in stain separation, we use the Warwick Stain Separation Benchmark (WSSB) in [11], which includes 24 H&E stained images of breast, colon, and lung tissues whose ground truth stain color-vector matrices, $\mathbf{M}_{GT}$, were manually selected by a pathologist. The ground truth concentrations were obtained

Table 1. PSNR and SSIM values for the different methods on the WSSB dataset [11]. Best values are marked in bold.

Image	S	RUI [6]		MAC [9]		VAH [7]		ALS [11]		SAR [14]		TV [1]		SG log [2]		SG ℓ_1 [2]		HS (proposed)	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Colon	H	22.27	0.8141	23.91	0.8095	25.83	0.8851	21.11	0.7241	28.57	0.9542	28.62	0.9544	28.66	0.9531	**29.01**	**0.9638**	29.00	0.9636
	E	20.70	0.7456	21.55	0.6365	26.29	0.8904	21.94	0.8540	27.58	0.9139	27.60	0.9161	27.74	0.9212	28.38	0.9414	**28.43**	**0.9421**
Breast	H	15.27	0.6215	26.24	0.9552	25.46	0.9239	24.60	0.8068	28.81	0.9528	29.14	0.9560	29.23	0.9464	**30.50**	**0.9751**	30.49	0.9749
	E	17.66	0.7644	23.62	0.9336	27.68	0.9550	25.92	0.9380	26.60	0.9464	26.76	0.9492	26.74	0.9444	27.71	0.9645	**27.82**	**0.9650**
Lung	H	22.47	0.7987	19.52	0.7389	25.87	0.8912	20.62	0.5551	32.91	0.9763	33.10	0.9757	31.21	0.9415	**35.21**	**0.9898**	**35.21**	0.9897
	E	22.05	0.7734	18.09	0.5088	25.53	0.8195	23.95	0.8939	30.77	0.9306	31.02	0.9353	29.99	0.9338	33.07	0.9654	**33.14**	**0.9659**
Mean	H	20.00	0.7448	23.22	0.8345	25.72	0.9100	22.11	0.6953	30.10	0.9611	30.29	0.9621	29.70	0.9470	**31.57**	**0.9762**	31.56	0.9760
	E	20.14	0.7611	21.08	0.6930	26.50	0.8883	23.94	0.8953	28.32	0.9303	28.46	0.9336	28.16	0.9331	29.72	0.9571	**29.79**	**0.9577**

Table 2. Area Under the Curve (AUC) values of the VGG19 classifier using the original and preprocessed images. Bold values indicate the highest performance.

	Original	RUI [6]	MAC [9]	VAH [7]	ALS [11]	SAR [14]	TV [1]	SG ℓ_1 [2]	HS (proposed)
AUC	0.9491	0.9417	0.9499	0.7985	0.9738	0.9479	0.9305	0.9617	**0.9788**

Algorithm 1 Variational Bayesian HS BCD

Require: Observed RGB image $\mathbf{I}$, reference color-vector matrix $\underline{\mathbf{M}}$, and hyperparameters $\mathbf{\Lambda}$.

Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$ and set, for each s, $\boldsymbol{\mu}_{\mathbf{m}_s}^{(0)} = \underline{\mathbf{m}}_s$, $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1\,(0)} = \mathbf{0}$, and $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1\,(0)} = \mathbf{0}$. Set $\boldsymbol{\mu}_{\mathbf{c}_s}^{(0)}$ as the s-th row of $\underline{\mathbf{M}}^{+}\mathbf{Y}^{\mathrm{T}}$, with $\underline{\mathbf{M}}^{+} \in \mathbb{R}^{S\times 3}$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$. Set $t = 0$.

while convergence criterion is not met **do**

1. Set $t = t + 1$.
2. For each s, update $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1\,(t)}$ using Eq. (20) and solve the linear system associated to Eq. (19) to obtain $\boldsymbol{\mu}_{\mathbf{m}_s}^{(t)}$.
3. Update $(\boldsymbol{\Theta}_s^n)^{(t-1)}$ using Eq. (24).
4. For each s, update $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1\,(t)}$ using Eq. (14) and solve the linear system associated to Eq. (13) to obtain $\boldsymbol{\mu}_{\mathbf{c}_s}^{(t)}$.

end while

Output $\hat{\mathbf{m}}_s = \boldsymbol{\mu}_{\mathbf{m}_s}^{(t)}$ and $\hat{\mathbf{c}}_s = \boldsymbol{\mu}_{\mathbf{c}_s}^{(t)}$.

using $\mathbf{C}_{GT}^{\top} = \mathbf{M}_{GT}^{+}\mathbf{Y}^{\top}$, where $\mathbf{M}_{GT}^{+} \in \mathbb{R}^{S\times 3}$ is the Moore-Penrose inverse of $\mathbf{M}_{GT}$.

The stain-separated H-only and E-only images were numerically compared using the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) metrics. In Table 1 we report the results obtained by our method and the methods by Ruifrok *et al.* [6] (RUI), Macenko *et al.* [9], Vahadane *et al.* [7] (VAH), Alsubaie *et al.* [11] (ALS), and the Bayesian methods using the SAR prior [14], TV [1], and the SG log and SG ℓ_1 priors [2]. The SG ℓ_1 prior and the proposed approach achieve the best tissue fidelity, showing the advantage of using moderate kurtosis SG priors for the BCD problem. For the E-only image, the proposed HS method achieves a better reconstruction than the SG ℓ_1 method while it is competitive for the H-only image.

4.2. Breast Cancer classification

In this section, we study how different BCD methods affects the performance of a VGG19 classifier on the Camelyon-17 database [20]. We use 500 breast tissue labeled WSIs from 5 medical centers in the Netherlands. Four centers are used for training the VGG19, while the fifth is reserved for testing. This configuration serves as a benchmark to evaluate the presence of challenging color differences. Training and testing sets contain approximately 100.000 and 25.000 patches of size 224×224, respectively. Both sets are balanced in positive and negative samples. The classifier is trained from scratch using both the original and the normalized images [2], given by the outputs of the different methods. The SG log prior is not included in this experiment due to its excessive computational cost [2].

The results, presented in Table 2, indicate that the image preprocessing affects the classification results. Among the evaluated methods, four lead to an improvement in performance on unseen data, with the proposed HS method being the most effective. Note that, although the proposed method performs similarly to the SG ℓ_1 in the stain separation task, the classification performance demonstrate its superiority.

5. CONCLUSIONS

This work presents, for the first time in literature, the use of the Hyperbolic Secant (HS) prior for BCD and examines its use in stain separation and breast cancer classification tasks. The HS prior belongs to the SG family and admits a GSM representation that makes the inference tractable without requiring a quadratic bound. It induces sparsity on the stain concentration differences at neighboring pixels, which is a theoretical feature of the stained tissue. In the stain separation task, the proposed method demonstrates a proper stain separation quality, improving the performance on the E-channel compared to other methods. In terms of classification performance, this enhancement resulted in the best classification performance on unseen data from a new center.

A. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access [11, 20]. Ethical approval was not required as confirmed by the license attached with the open access data.

B. ACKNOWLEDGMENTS

This work was supported by project PID2022-140189OB-C22 funded by MCIN / AEI / 10.13039 / 501100011033, Spain. The work by Francisco M. Castro-Macías was sponsored by Ministerio de Universidades under FPU contract FPU21/01874.

C. REFERENCES

[1] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, and A. K. Katsaggelos, "A TV-based image processing framework for blind color deconvolution and classification of histological images," *Digit. Signal Process.*, vol. 101, pp. 102727, 2020.

[2] F. Pérez-Bueno, M. Vega, María A. Sales, J. Aneiros-Fernández, V. Naranjo, R. Molina, and A. K. Katsaggelos, "Blind color deconvolution, normalization, and classification of histological images using general Super Gaussian priors and Bayesian inference," *Comput. Meth. Prog. Bio.*, vol. 211, pp. 106453, 2021.

[3] R. Duggal et al., "SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. Lecture Notes in Computer Science. Springer, Cham*, 2017, vol. 10435, pp. 435–443.

[4] N. Kanwal, F. Pérez-Bueno, A. Schmidt, R. Molina, and K. Engan, "The devil is in the details: Whole slide image acquisition and processing for artifacts detection, color variation, and data augmentation. a review.," *IEEE Access*, vol. 10, pp. 58821–58844, 2022.

[5] Tekla S. Perry, "Andrew Ng X-Rays the AI Hype," IEEE Spectrum, (2021).

[6] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, pp. 291–299, 2001.

[7] A. Vahadane et al., "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, pp. 1962–1971, 2016.

[8] T. A. Azevedo Tosta et al., "A stain color normalization with robust dictionary learning for breast cancer histological images processing," *Biomedical Signal Processing and Control*, vol. 85, pp. 104978, Aug. 2023.

[9] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in *Int. Symp. on Biomedical Imaging (ISBI)*, 2009, pp. 1107–1110.

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

[11] N. Alsubaie et al., "Stain deconvolution using statistical analysis of multi-resolution stain colour representation," *PLOS ONE*, vol. 12, pp. e0169875, 2017.

[12] M. Gavrilovic et al., "Blind color decomposition of histological images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 983–994, 2013.

[13] M. Salvi et al., "Stain color adaptive normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology," *Comput. Meth. Prog. Bio.*, vol. 193, pp. 105506, 2020.

[14] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "Variational Bayesian blind color deconvolution of histopathological images," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 2026–2036, 2020.

[15] W. L. Harkness and M. L. Harkness, "Generalized Hyperbolic Secant distributions," *Journal of the American Statistical Association*, vol. 63, no. 321, pp. 329–337, 1968, Publisher: Taylor & Francis.

[16] O. Barndorff-Nielsen et al., "Normal variance-mean mixtures and z distributions," *International Statistical Review/Revue Internationale de Statistique*, pp. 145–159, 1982.

[17] C. M Bishop and N. M Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4, Springer, 2006.

[18] Nicholas G. Polson et al., "Bayesian inference for logistic models using pólya–gamma latent variables," *Journal of the American statistical Association*, vol. 108, no. 504, pp. 1339–1349, 2013.

[19] J. A. Palmer, *Variational and scale mixture representations of non-Gaussian densities for estimation in the Bayesian linear model: Sparse coding, independent component analysis, and minimum entropy segmentation*, Ph.D. thesis, University of California, San Diego, 2006.

[20] P. Bándi et al., "From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 550–560, 2019.