Deep Bayesian Blind Color Deconvolution using a Gaussian Scale Mixture Prior

Francisco M. Castro-Macías[a,*], Miguel Vega[b], Fernando Pérez-Bueno[c], Rafael Molina[a], Aggelos K. Katsaggelos[d]
[a] Dept. of Computer Science and AI, University of Granada, Granada, Spain.
[b] Dept. Computer Languages and Systems, University of Granada, Granada, Spain.
[c] Koh Young Research Spain.
[d] Dept. of Electrical and Computer Engineering, Northwestern University, Evanston (IL), USA.
[*] Corresponding author. Email: fcastro@ugr.es

***Abstract*—Variability in histological staining protocols across medical centers introduces significant color differences. As a consequence, the performance of Computer-Aided Diagnosis (CAD) systems degrades when tested in images from unseen centers. To mitigate this issue, Blind Color Deconvolution (BCD) enables the separation of stain components. However, traditional analytical BCD methods are computationally expensive, while deep learning (DL)-based approaches offer efficient inference once trained. Despite their efficiency, existing DL-based methods lack effective priors to guide the learning process. In this work, we propose GSMNet, a DL-based model that incorporates a Gaussian Scale Mixture (GSM) prior to promote sparsity in concentration maps. To address the intractability introduced by this prior, we formulate an augmented probabilistic model that enables efficient variational inference. We evaluate GSMNet on two tasks – stain separation and breast cancer classification – demonstrating that it achieves performance superior to or comparable with state-of-the-art BCD methods while being significantly faster.**

***Index Terms*—Blind Color Deconvolution, Deep Learning, Variational Bayes, Gaussian Scale Mixture.**

I. Introduction

Accurate diagnosis in computational pathology relies on histological staining to distinguish tissue components. However, variations in staining protocols across medical centers introduce significant color discrepancies, which can degrade the performance of Computer-Aided Diagnosis (CAD) systems. As a result, CAD models often struggle when tested on images from hospitals not included in the training set [1].

To address this, various approaches have been explored, among which stain separation stands out, as it is a key preliminary step for other tasks [2]. Blind Color Deconvolution (BCD) methods tackle this problem by decomposing the observed RGB image into a color-vector matrix and a concentration matrix, representing stain intensity per pixel [3]. Today, BCD techniques are essential in large-scale applications, such as developing foundation models for pathology and automating image diagnosis [4], [5]. In these applications, the BCD decomposition process must be performed efficiently.

Different approaches have been proposed to address the BCD problem, broadly categorized into analytical and deep learning (DL)-based methods. Analytical methods are typically computationally expensive, as they require an optimization procedure to be performed independently for each image (see [2] for a review). These approaches incorporate various constraints to guide the optimization process, either as regularization terms [6]–[9] or as prior probability distributions [10]–[12]. Notably, enforcing sparsity in the concentration maps has proven effective, often achieved through different forms of Super Gaussian (SG) distributions [7], [11], [12].

In contrast to analytical approaches, DL methods are significantly more efficient, as they can process unseen images in a single forward pass once trained. However, due to the lack of labeled BCD datasets, most of existing DL-based approaches primarily focus on stain normalization or adaptation and cannot perform stain separation. Thus, previous works have proposed learning stain-invariant features within a classifier [13], adapting Generative Adversarial Networks (GANs) for stain normalization [14], and diffusion models for stain adaptation [15].

As mentioned earlier, these methods cannot inherently separate an input image into distinct stains. In contrast, certain DL-based methods do enable stain separation in a blind scenario [16]–[18]. Among them, BCDNet [18] is the best-performing one. It follows a Bayesian approach that does not rely on ground truth data during training yet can still estimate both the color matrix and concentration maps for a given image. However, their modeling employs a simple flat prior on the concentrations which, while flexible, may not effectively guide the learning process.

In this work, inspired by the success of sparsity-promoting priors in analytical approaches, we demonstrate how to adapt these priors for DL-based models. We summarize our main contributions as follows:

- We propose GSMNet, a DL model for BCD that incorporates a sparsity-promoting prior on the concentrations using the Gaussian Scale Mixture (GSM) family.
- To address the intractability introduced by this new prior, we perform inference in an augmented probabilistic model, which remains equivalent to the original model due to the properties of GSM while ensuring a tractable training objective.
- We evaluate GSMNet on two tasks – stain separation and breast cancer classification – demonstrating its superiority

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

over existing DL-based and analytical methods while achieving significantly faster inference.

The remainder of the paper is organized as follows: Sec. II describes the proposed Bayesian model, Sec. III outlines the amortized variational inference procedure, and Sec. IV evaluates the performance of the proposed method. Finally, Sec. V concludes the paper.

II. Bayesian modeling

Gaussian Scale Mixtures. The family of Super Gaussian (SG) distributions has been extensively used in different image processing problems, including BCD [7], [11], [12]. Intuitively, when a high-pass filter is applied to a natural image, the resulting coefficients are sparse, and SG distributions can be used to enforce this property.

In this work, we are interested in a subclass of the mentioned SG family: distributions whose density allows for a Gaussian Scale Mixture (GSM) representation [19]. Formally, a density $f\colon \mathbb{R} \to \mathbb{R}$ admits a GSM representation if there exists another density $\pi(\omega)\colon \mathbb{R} \to \mathbb{R}$ such that,

Name	$\rho(x)$	τ
sech	$-\log\operatorname{sech}(x)$	2.0
log	$\log(\epsilon + \lvert x\rvert)$	0.001
ℓ_1	$\lvert x\rvert$	0.001
ℓ_2	$0.5\lvert x\rvert^2$	0.01

TABLE I
GSM densities considered in this work. All are obtained as $f(x) \propto \exp(-\rho(x))$.

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

Here, π is called the *mixing* density. The above expression will be fundamental for the inference procedure.

Modeling the problem. A stained histological slide is stored as a RGB image $\mathbf{I} \in \mathbb{R}^{HW\times 3}$ with H rows and W columns. Then, it is transformed into the OD space as $\mathbf{Y} = -\log\left(\mathbf{I}/i_0\right) \in \mathbb{R}^{HW\times 3}$, where i_0 is the maximum incident luminosity, set to 255. The Beer-Lambert law [3] states that

$$\mathbf{Y}^\top = \mathbf{M}\mathbf{C}^\top + \mathbf{N}^\top, \tag{2}$$

where $\mathbf{M} \in \mathbb{R}^{3\times S}$ is the color-vector matrix, S is 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. In the following, we 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 denote $[A] = \{1, \dots, A\}$.

From Eq. 2 we obtain the likelihood

$$\mathrm{p}\left(\mathbf{Y} \mid \mathbf{C}, \mathbf{M}\right) \propto \exp\left(-\lambda \|\mathbf{Y}^\top - \mathbf{M}\mathbf{C}^\top\|_F^2\right), \tag{3}$$

where $\|\cdot\|_F$ represents the Frobenius norm and $\lambda > 0$ is an hyperparameter.

Next, we define a prior for the concentrations using the family of the GSM distributions. Following [11], [12], we consider a set of N high-pass filters $\{\mathbf{F}^1, \dots, \mathbf{F}^N\} \subset \mathbb{R}^{HW\times HW}$. We write $\mathbf{c}_s^n = \mathbf{F}^n \mathbf{c}_s$ for $s \in [S]$ and $n \in [N]$. Denoting the components of $\mathbf{c}_s^n$ by c_{is}^n, we define,

$$\mathrm{p}\left(\mathbf{C}\right) \propto \prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{i=1}^{HW} f\left(c_{is}^n\right), \tag{4}$$

where f is a density that admits a GSM representation. In this work, following [11] we consider four different GSM densities, obtained as $f(x) \propto \exp(-\rho(x))$, see Table I.

For the color-vector matrix we adopt the *Ruifrok prior* [18], which assume that the color vectors are always close to those provided by the Ruifrok reference matrix,

$$\mathrm{p}\left(\mathbf{M}\right) \propto \exp\left(-\gamma \left\|\mathbf{M} - \mathbf{M}^{\text{Rui}}\right\|_F^2\right), \tag{5}$$

where $\mathbf{M}^{\text{Rui}} \in \mathbb{R}^{3\times S}$ is the Ruifrok reference matrix [3], and $\gamma > 0$ is an hyperparameter.

Finally, the joint probability distribution is given by

$$\mathrm{p}\left(\mathbf{Y}, \mathbf{C}, \mathbf{M}\right) = \mathrm{p}\left(\mathbf{Y} \mid \mathbf{C}, \mathbf{M}\right) \mathrm{p}\left(\mathbf{C}\right) \mathrm{p}\left(\mathbf{M}\right). \tag{6}$$

III. Amortized Variational Inference

Following the Bayesian perspective, we aim to utilize the posterior distribution $\mathrm{p}\left(\mathbf{C}, \mathbf{M} \mid \mathbf{Y}\right)$ to estimate $\mathbf{C}$ and $\mathbf{M}$ from $\mathbf{Y}$. Since it can not be obtained in closed form, we approximate it by a variational distribution $\mathrm{q}\left(\mathbf{C}, \mathbf{M} \mid \mathbf{Y}\right)$ using variational Bayesian inference. In contrast to analytical approaches, we parameterize this distribution using a neural network (NN) and optimize its parameters by maximizing the Evidence Lower Bound (ELBO) [18]. Unfortunately, as in [11], [12], the presence of the GSM prior (see Eq. (4)), makes it impossible to obtain a tractable expression for the ELBO.

The augmented model. To circumvent this intractability, we consider an augmented model in which the ELBO is tractable thanks to the GSM representation presented in Sec. II,

$$\mathrm{p}\left(\mathbf{Y}, \mathbf{C}, \mathbf{M}, \boldsymbol{\omega}\right) = \mathrm{p}\left(\mathbf{Y} \mid \mathbf{C}, \mathbf{M}\right) \mathrm{p}\left(\mathbf{M}\right) \mathrm{p}\left(\mathbf{C} \mid \boldsymbol{\omega}\right) \mathrm{p}\left(\boldsymbol{\omega}\right), \tag{7}$$

$$\mathrm{p}(\boldsymbol{\omega}) = \prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{i=1}^{HW} \pi(\omega_{is}^n), \tag{8}$$

$$\begin{aligned}\mathrm{p}\left(\mathbf{C} \mid \boldsymbol{\omega}\right) &= \prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{i=1}^{HW} \mathcal{N}\left(c_{is}^n \mid 0, \left(\omega_{is}^n\right)^{-1}\right), \\ &\propto \prod_{s=1}^{S} \mathcal{N}\left(\mathbf{c}_s \mid \mathbf{0}, {\mathbf{P}_s}^{-1}\right),\end{aligned} \tag{9}$$

where $\mathbf{P}_s = \sum_{n=1}^{N} \left(\mathbf{F}^n\right)^\top \boldsymbol{\Omega}_s^n \mathbf{F}^n$, and $\boldsymbol{\Omega}_s^n = \mathrm{Diag}(\omega_{1s}^n, \dots, \omega_{HWs}^n) \in \mathbb{R}^{HW\times HW}$. Here, $\boldsymbol{\omega}$ are called *augmentation variables*. Integrating them and using the GSM representation (Eq. 1), we recover the original model in Eq. 6. To obtain an amortized variational posterior we define

$$\mathrm{q}^{(\boldsymbol{\alpha},\boldsymbol{\beta})}\left(\mathbf{C}, \mathbf{M}, \boldsymbol{\omega} \mid \mathbf{Y}\right) = \mathrm{q}^{\boldsymbol{\alpha}}\left(\mathbf{C} \mid \mathbf{Y}\right) \mathrm{q}^{\boldsymbol{\beta}}\left(\mathbf{M} \mid \mathbf{Y}\right) \mathrm{q}\left(\boldsymbol{\omega} \mid \mathbf{Y}\right),$$

$$\mathrm{q}^{\boldsymbol{\alpha}}\left(\mathbf{C} \mid \mathbf{Y}\right) = \sum_{s=1}^{S} \delta\left(\mathbf{c}_s - \boldsymbol{\mu}_{\mathbf{c}_s}^{\boldsymbol{\alpha}}\left(\mathbf{Y}\right)\right), \tag{10}$$

$$\mathbf{q}^{\boldsymbol{\beta}}\left(\mathbf{M} \mid \mathbf{Y}\right) = \prod_{s=1}^{S} \mathcal{N}\left(\mathbf{m}_s \mid \boldsymbol{\mu}_{\mathbf{m}_s}^{\boldsymbol{\beta}}\left(\mathbf{Y}\right), \sigma_{\mathbf{m}_s}^{\boldsymbol{\beta}}\left(\mathbf{Y}\right)^2 \mathbf{I}\right), \tag{11}$$

where $\boldsymbol{\mu}_{\mathbf{c}_s}^{\boldsymbol{\alpha}}\colon \mathbb{R}^{HW\times 3} \to \mathbb{R}^{HW}$, $\boldsymbol{\mu}_{\mathbf{m}_s}^{\boldsymbol{\beta}}\colon \mathbb{R}^{HW\times 3} \to \mathbb{R}^3$, and $\sigma_{\mathbf{m}_s}^{\boldsymbol{\beta}}\colon \mathbb{R}^{HW\times 3} \to \mathbb{R}$ are NNs with parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. In the following, identify $\boldsymbol{\mu}_{\mathbf{C}} = \left[\boldsymbol{\mu}_{\mathbf{c}_1}, \dots, \boldsymbol{\mu}_{\mathbf{c}_S}\right]$ as the concentrations network and $(\boldsymbol{\mu}_{\mathbf{M}}, \boldsymbol{\sigma}_{\mathbf{M}})$ as the color matrix network, with

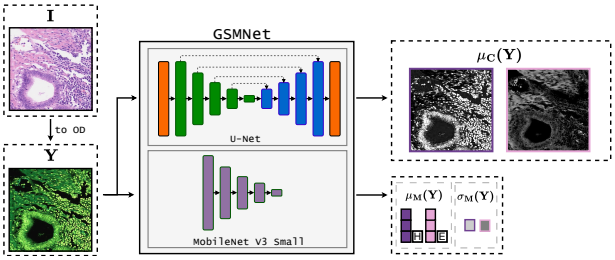

Fig. 1. A forward pass in GSMNet. The input image $\mathbf{I}$ is transformed to the OD space, $\mathbf{Y}$. The concentrations network $\boldsymbol{\mu}_{\mathbf{C}}$ estimates the concentration maps, while the color matrix network $(\boldsymbol{\mu}_{\mathbf{M}}, \boldsymbol{\sigma}_{\mathbf{M}})$ computes the color matrix distribution.

$\boldsymbol{\mu}_{\mathbf{M}} = \left[\boldsymbol{\mu}_{\mathbf{m}_1}, \ldots, \boldsymbol{\mu}_{\mathbf{m}_S}\right]$, and $\boldsymbol{\sigma}_{\mathbf{M}} = \mathrm{Diag}\left(\boldsymbol{\sigma}_{\mathbf{m}_1}, \ldots, \boldsymbol{\sigma}_{\mathbf{m}_S}\right)$, see Figure 1.

Note that $\mathrm{q}^{\alpha}\left(\mathbf{C} \mid \mathbf{Y}\right)$ is defined as a deterministic distribution using the Dirac delta operator. This is conceptually equivalent to choosing a normal distribution with infinitely small and constant variance, and then ignoring the constant terms in the ELBO. Also, as we argue later, the moments of $\mathrm{q}\left(\boldsymbol{\omega} \mid \mathbf{Y}\right)$ can be computed in closed form, and therefore it is not necessary to parameterize it. For ease of notation, in the following we omit the dependence on $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

These NNs will be trained using a large dataset of histological images $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_M\}$ to maximize the ELBO,

$$\sum_{m=1}^{M} \log \mathrm{p}\left(\mathbf{Y}_m\right) \geq \sum_{m=1}^{M} \mathrm{ELBO}\left(\mathbf{Y}_m\right), \tag{12}$$

$$\mathrm{ELBO}\left(\mathbf{Y}\right) = \mathbb{E}_{\mathrm{q}(\mathbf{C},\mathbf{M},\boldsymbol{\omega}|\mathbf{Y})}\left[\log \frac{\mathrm{p}\left(\mathbf{Y},\mathbf{C},\mathbf{M},\boldsymbol{\omega}\right)}{\mathrm{q}\left(\mathbf{C},\mathbf{M},\boldsymbol{\omega} \mid \mathbf{Y}\right)}\right]. \tag{13}$$

Computing the ELBO. The ELBO can be written as

$$\begin{aligned}\mathrm{ELBO}\left(\mathbf{Y}\right) = &- \mathrm{LL}\left(\mathbf{Y}\right) - \mathrm{KL}_{\mathbf{M}}\left(\mathbf{Y}\right) - \mathrm{KL}_{\mathbf{C}}\left(\mathbf{Y}\right) \\ &- \mathrm{KL}_{\boldsymbol{\omega}}\left(\mathbf{Y}\right).\end{aligned} \tag{14}$$

Here, $\mathrm{LL}\left(\mathbf{Y}\right)$ corresponds to the log-likelihood,

$$\mathrm{LL}\left(\mathbf{Y}\right) = \lambda \mathbb{E}_{\mathrm{q}(\mathbf{M}|\mathbf{Y})}\left[\left\|\mathbf{Y}^{\mathrm{T}} - \mathbf{M}\boldsymbol{\mu}_{\mathbf{C}}\left(\mathbf{Y}\right)^{\top}\right\|_F^2\right] + \text{const}, \tag{15}$$

where const is a constant term that does not depend on the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. This term ensures that our NNs correctly reconstructs the original OD image. In our implementation, it is approximated using the reparametrization trick. The rest of the terms act as regularizers, corresponding to the Kullback-Leibler divergence between the prior and variational posterior terms. For the color-vector matrix,

$$\begin{aligned}\mathrm{KL}_{\mathbf{M}}\left(\mathbf{Y}\right) = &\mathrm{KL}\left(\mathrm{q}\left(\mathbf{M} \mid \mathbf{Y}\right), \mathrm{p}\left(\mathbf{M}\right)\right) = \\ = &\gamma\left(\left\|\boldsymbol{\mu}_{\mathbf{M}}\left(\mathbf{Y}\right) - \mathbf{M}^{\mathrm{Rui}}\right\|_F^2 + \mathrm{Trace}\left(\boldsymbol{\sigma}_{\mathbf{M}}\left(\mathbf{Y}\right)\right)\right) + \\ &- \log\left|\boldsymbol{\sigma}_{\mathbf{M}}\left(\mathbf{Y}\right)\right| + \text{const},\end{aligned} \tag{16}$$

which keeps the estimated color matrix close to that of Ruifrok. For the concentrations,

$$\begin{aligned}\mathrm{KL}_{\mathbf{C}}\left(\mathbf{Y}\right) = &\mathbb{E}_{\mathrm{q}(\boldsymbol{\omega}|\mathbf{Y})}\left[\mathrm{KL}\left(\mathrm{q}\left(\mathbf{C} \mid \mathbf{Y}\right), \mathrm{p}\left(\mathbf{C} \mid \boldsymbol{\omega}\right)\right)\right] = \\ = &\frac{1}{2}\sum_{s=1}^{S} \boldsymbol{\mu}_{\mathbf{c}_s}\left(\mathbf{Y}\right)^{\top} \mathbb{E}_{\mathrm{q}(\boldsymbol{\omega}|\mathbf{Y})}\left[\mathbf{P}_s\right] \boldsymbol{\mu}_{\mathbf{c}_s}\left(\mathbf{Y}\right) + \text{const},\end{aligned} \tag{17}$$

where $\mathbb{E}_{\mathrm{q}(\boldsymbol{\omega}|\mathbf{Y})}\left[\mathbf{P}_s\right] = \sum_{n=1}^{N} \mathbf{F}^{n\,\mathrm{T}} \boldsymbol{\Theta}_s^n\left(\mathbf{Y}\right) \mathbf{F}^n$, with $\boldsymbol{\Theta}_s^n\left(\mathbf{Y}\right) = \mathrm{Diag}\left(\theta_{1s}^n\left(\mathbf{Y}\right), \ldots, \theta_{HWs}^n\left(\mathbf{Y}\right)\right)$, $\theta_{is}^n\left(\mathbf{Y}\right) = \mathbb{E}_{\mathrm{q}(\omega_{is}^n|\mathbf{Y})}\left[\omega_{is}^n\right]$.

The last term, $\mathrm{KL}_{\boldsymbol{\omega}}\left(\mathbf{Y}\right)$, corresponds to the Kullback-Leibler divergence between $\mathrm{q}\left(\boldsymbol{\omega} \mid \mathbf{Y}\right)$ and $\mathrm{p}\left(\boldsymbol{\omega}\right)$. For most choices of f this term will be intractable since it depends on the exact form of the corresponding mixing density $\pi(\omega)$ and the variational posterior. We address this in the following.

Updating the augmentation variables $\boldsymbol{\omega}$. To compute the rest of the terms we only need the first-order moment of $\mathrm{q}\left(\boldsymbol{\omega} \mid \mathbf{Y}\right)$, see Eq. 17. Fortunately, this moment can be computed in closed form. First, note that the optimal expression of $\mathrm{q}\left(\boldsymbol{\omega} \mid \mathbf{Y}\right)$ in terms of the rest of the variational factors is given by [20, Eq. (10.9)]. Applying these equations, and after some calculations (see [12] for details), we arrive at

$$\theta_{is}^n\left(\mathbf{Y}\right) = \mathbb{E}_{\mathrm{q}(\omega_{is}^n|\mathbf{Y})}\left[\omega_{is}^n\right] = -\frac{f'\left(\xi_{is}^n\left(\mathbf{Y}\right)\right)}{\xi_{is}^n\left(\mathbf{Y}\right) f\left(\xi_{is}^n\left(\mathbf{Y}\right)\right)}. \tag{18}$$

where $\xi_{is}^n\left(\mathbf{Y}\right) = \sqrt{\mathbb{E}_{\mathrm{q}(c_{is}|\mathbf{Y})}\left[\left(c_{is}^n\right)^2\right]}$. This expression is the optimal way to update the parameters θ_{is}^n keeping the rest of the variational distributions fixed.

The training procedure. The procedure to train the proposed model is in algorithm 1. Motivated by the above observation, we take an iterative approach. Having initialized $\theta_{is}^n\left(\mathbf{Y}_m\right)$, we update $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using gradient ascent. Then, we update $\theta_{is}^n\left(\mathbf{Y}_m\right)$ using Eq. 18. We repeat this process until a convergence criterion is met. The training objective is a modified version of the negative of the ELBO,

$$\mathcal{L}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}\right) = \sum_{m=1}^{M} \left(1-\eta\right) \mathrm{LL}\left(\mathbf{Y}_m\right) + \eta\, \mathrm{KL}\left(\mathbf{Y}_m\right), \tag{19}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_s^n\left(\mathbf{Y}_m\right) : n \in [N], s \in [S], m \in [M]\}$, $\mathrm{KL}\left(\mathbf{Y}_m\right) = \mathrm{KL}_{\mathbf{M}}\left(\mathbf{Y}_m\right) + \tau\, \mathrm{KL}_{\mathbf{C}}\left(\mathbf{Y}_m\right)$, and we have introduced two hyperparameters $\eta \in (0,1), \tau \in (0, +\infty)$ to account for the balance between each term.

Algorithm 1: Training procedure of GSMNet

Require: Dataset $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_M\}$, number of iterations T.
Initialize $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}$.
for $t = 1, \ldots, T$ **do**
  Update $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using $\nabla_{(\boldsymbol{\alpha},\boldsymbol{\beta})}\mathcal{L}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}\right)$.
  Update $\theta_{is}^n\left(\mathbf{Y}_m\right)$ using Eq. 18 and stopping gradients.
end for

Making predictions. Once the training has finished, we estimate the concentrations and the color matrix for a given OD image $\mathbf{Y}$ using the learned variational distributions. For simplicity, we consider the mean of each distribution to estimate these variables, although other approaches are also possible.

IV. Experiments

We compare the performance of the proposed GSMNet against state-of-the-art (SOTA) methods in two different tasks: stain separation and breast cancer classification.

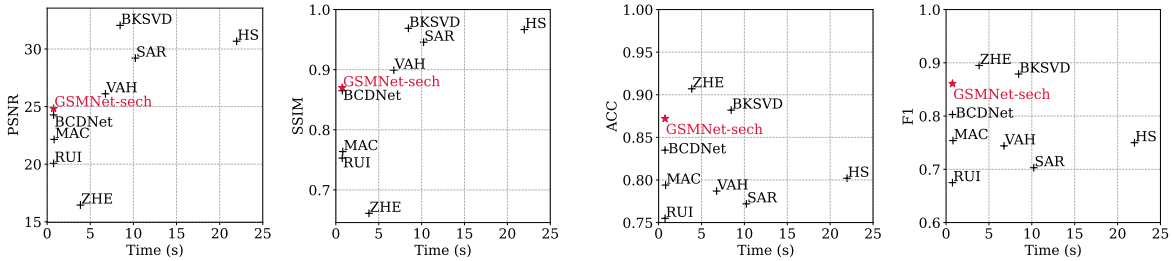

Fig. 2. Inference time vs. performance across two tasks: stain separation (PSNR, SSIM) and breast cancer classification (ACC, F1). The proposed GSMNet-sech achieves the lowest inference time while maintaining high-quality estimations. Inference time is measured as the time required to process a 2000×2000 image from the WSSB dataset.

A. Experimental setup

Datasets. We use two well-known datasets: Camelyon17 [21] and the Warwick Stain Separation Benchmark (WSBB) [8]. Camelyon17 contains Whole Slide Images (WSIs) from five different medical centers for which the ground truth color matrix is not available. We use 100 WSIs from each center. From each WSI we extract 500 non-overlapping patches of size 224×224. WSSB provides the ground truth color matrix for 24 images of three different tissue types. Using this matrix, ground truth concentrations and stains are obtained following [8].

Methods considered. We compare the proposed model against SOTA methods in BCD, considering two types of baselines: amortized and non-amortized methods. As an amortized baseline, we use BCDNet [18], upon which our model is built. For non-amortized methods, we consider several popular analytical approaches, including the methods by Ruifrok et al. [3] (RUI), Macenko et al. [6] (MAC), Vahadane et al. [7] (VAH), and Zheng et al. [22] (ZHE). Additionally, we consider popular Bayesian methods such as the Simultaneous Autoregressive (SAR) method [10], the K-SVD-based method (BKSVD) [23], and the Hyperbolic Secant prior method (HS) [12]. Non-amortized methods generally yield more accurate estimations but at the cost of increased computational time; we analyze this trade-off in the following discussion. For the proposed GSMNet, we analyze four variants, corresponding to four different choices for the GSM density f, see Eq. 4 and Table I.

GSMNet training configuration. To ensure a fair comparison, we adopt the same configuration as in [18]. We adopt the same network architecture, which consists of a U-Net for the concentration network $\boldsymbol{\mu}_{\mathbf{C}}$ and a MobileNetV3 Small for the color-matrix network $(\boldsymbol{\mu}_{\mathbf{M}}, \boldsymbol{\sigma}_{\mathbf{M}})$ (see Fig. 1). Our model is trained for 100 epochs using the patches from centers 0, 1, and 2 of Camelyon17. We use the ADAM optimizer with an initial learning rate of 10^{-4}. Following [18], we set $\lambda = 1, \gamma = 0.05$ and $\eta = 0.3$. During the first epoch, we set $\eta = 0.99$ to help the network determine the correct stain order. We also found it beneficial to set $\tau = 0$ for the first 50 epochs, as this allows the network to first learn the reference matrix and then refine the reconstructions. After that, the value of the τ hyperparameter depends on the selected GSM density, see Table I. Due to space constraints, we omit the corresponding ablation study.

Method	Stain separation (WSSB) PSNR	SSIM	Classification (Camelyon17) ACC	F1
BCDNet	$24.277_{0.525}$	$0.865_{0.010}$	$0.835_{0.031}$	$0.803_{0.044}$
GSMNet-sech	$\mathbf{24.836}_{0.047}$	$\mathbf{0.871}_{0.001}$	$\mathbf{0.872}_{0.017}$	$\mathbf{0.861}_{0.025}$
GSMNet-log	$24.748_{0.123}$	$0.865_{0.006}$	$0.851_{0.034}$	$0.815_{0.062}$
GSMNet-ℓ_1	$24.728_{0.206}$	$0.870_{0.004}$	$0.851_{0.081}$	$0.819_{0.121}$
GSMNet-ℓ_2	$24.799_{0.227}$	$0.863_{0.001}$	$0.852_{0.021}$	$0.827_{0.029}$

TABLE II
Stain separation and breast cancer classification results (mean and standard deviation from three independent runs). All variants of the proposed model exhibit superior performance to that of BCDNet. GSMNet-sech stands out as the best-performing one.

B. Stain separation

The stain separation task evaluates whether a method can accurately separate the contribution of each stain. To assess this, we use the WSSB dataset, which provides ground-truth stain separation. We use the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) as evaluation metrics. The results are presented in Table II and Figure 2.

Improvement upon BCDNet. The results in Table II demonstrate that incorporating a sparsity-promoting GSM prior enhances stain separation performance. Two variants of the proposed model (sech and ℓ_1) outperform BCDNet in terms of PSNR and SSIM, while the remaining two variants achieve higher PSNR and comparable SSIM. Notably, BCDNet currently surpasses other DL-based methods [18]. As we will see in Sec. IV-C, these improvements in stain separation translate into even greater gains in the breast cancer classification task.

Closing the gap with analytical methods. Following [18], we compare the proposed approach against non-amortized methods, considering inference time, see Figure 2. The proposed GSMNet-sech is faster and separates better the stains than the widely used MAC, RUI and ZHE methods. Although BKSVD, HS, SAR, VAH outperform our method, their inference time may become prohibitive in certain settings. For instance, VAH requires over 5 seconds, while BKSVD, HS, and SAR require over 7.5 seconds. In contrast, the proposed GSMNet produces comparable estimations in just 0.75 seconds.

C. Breast cancer classification

We illustrate how concentration maps mitigate performance degradation due to color variations. Following [18], we train a VGG19 classifier to predict whether an input patch is cancerous. The training dataset consists of patches from

centers $\{0, 1, 2, 3\}$ of Camelyon17. Center 4, which exhibits larger color differences, is reserved for testing. Class-balanced patches were sampled from the annotated WSIs, yielding approximately $107 \cdot 10^3$ training patches and $26 \cdot 10^3$ testing patches. For each method, we extract concentration maps and train the same classifier with identical initial weights for 50 epochs with the ADAM optimizer at an initial learning rate of 10^{-3}. Results are presented in Table II and Figure 2.

Improvement upon BCDNet. Table II also presents the results of the proposed GSMNet and the baseline BCDNet, evaluated based on the accuracy (ACC) and F1 scores. Notably, all four variants of the proposed GSMNet outperform BCDNet by a considerable margin, with the sech variant achieving the best performance. These improvements come from the introduction of a novel prior on concentrations, which are being used in this experiment for classification.

Reaching the performance of analytical methods. We compare the proposed approach against non-amortized methods, considering inference time (see Figure 2). The proposed GSMNet-sech sits in the top 3 in this task, achieving comparable performance to BKSVD and ZHE, while requiring only a fraction of the computational cost of other methods. Note that the performance gap with BKSVD and ZHE is minimal (less than 0.03 for both metrics). Notably, ZHE, which performed the poorest in stain separation, ranks first for this task.

V. Conclusions

In this work, we introduced GSMNet, a DL-based model for the BCD problem grounded in a Bayesian framework. GSMNet integrates insights from classical analytical methods through a novel GSM prior on the concentrations, promoting sparsity in the concentration maps. To address the intractability introduced by this prior, we developed an augmented model that leverages the properties of GSM densities, resulting in an equivalent formulation with a tractable training objective.

We evaluated the proposed method across two tasks – stain separation and breast cancer classification – showing that GSMNet achieves comparable performance to SOTA methods while being significantly faster, making it suitable for large-scale computational pathology applications.

Future work will explore the integration of additional priors, further refinement of the model architecture, and the expansion of the evaluation dataset.

References

[1] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J.-M. Bokhorst, F. Ciompi, and J. Van Der Laak, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *Med. Imag. Anal.*, vol. 58, p. 101544, 2019.

[2] N. Kanwal, F. Pérez-Bueno, A. Schmidt, K. Engan, and R. Molina, "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. 58 821–58 844, 2022.

[3] A. C. Ruifrok, D. A. Johnston *et al.*, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, no. 4, pp. 291–299, 2001.

[4] R. J. Chen, T. Ding, M. Y. Lu, D. F. Williamson, G. Jaume, A. H. Song, B. Chen, A. Zhang *et al.*, "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, vol. 30, no. 3, 2024.

[5] M. Y. Lu, B. Chen, D. F. Williamson, R. J. Chen, I. Liang, T. Ding, G. Jaume, I. Odintsov *et al.*, "A visual-language foundation model for computational pathology," *Nature Medicine*, vol. 30, no. 3, 2024.

[6] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2009.

[7] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. on Medical Imaging*, vol. 35, no. 8, 2016.

[8] N. Alsubaie, N. Trahearn, S. E. A. Raza, D. Snead, and N. M. Rajpoot, "Stain deconvolution using statistical analysis of multi-resolution stain colour representation," *PloS one*, vol. 12, no. 1, p. e0169875, 2017.

[9] M. Salvi, N. Michielli, and F. Molinari, "Stain color adaptive normalization (scan) algorithm: Separation and standardization of histological stains in digital pathology," *Computer methods and programs in biomedicine*, vol. 193, p. 105506, 2020.

[10] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind color deconvolution of histopathological images," *IEEE Trans. on Image Proc.*, vol. 29, pp. 2026–2036, 2019.

[11] F. Pérez-Bueno, M. Vega, M. 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," *Computer Methods and Programs in Biomedicine*, vol. 211, p. 106453, 2021.

[12] F. M. Castro-Macías, F. Pérez-Bueno, M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Blind color deconvolution and classification of histological images using the hyperbolic secant prior," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024.

[13] N. Marini, M. Atzori, S. Otálora, S. Marchand-Maillet, and H. Müller, "H&e-adversarial network: a convolutional neural network to learn stain-invariant features through hematoxylin & eosin regression," in *IEEE/CVF International Conference on Computer Vision*, 2021.

[14] C. Cong, S. Liu, A. Di Ieva, M. Pagnucco, S. Berkovsky, and Y. Song, "Colour adaptive generative networks for stain normalisation of histopathology images," *Med. Imag. Anal.*, vol. 82, p. 102580, 2022.

[15] C.-C. Tsai, Y.-C. Chen, and C.-S. Lu, "Test-time stain adaptation with diffusion models for histopathology image classification," in *European Conference on Computer Vision*. Springer, 2024, pp. 257–275.

[16] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, D. Hu, S. Sun, J. Shi, and C. Xue, "Stain standardization capsule for application-driven histopathological image normalization," *IEEE journal of biomedical and health informatics*, vol. 25, no. 2, pp. 337–347, 2020.

[17] S. Abousamra, D. Fassler, L. Hou, Y. Zhang, R. Gupta, T. Kurc, L. F. Escobar-Hoyos, D. Samaras *et al.*, "Weakly-supervised deep stain decomposition for multiplex ihc images," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020.

[18] S. Yang, F. Pérez-Bueno, F. M. Castro-Macías, R. Molina, and A. K. Katsaggelos, "Bcd-net: Stain separation of histological images using deep variational bayesian blind color deconvolution," *Digital Signal Processing*, vol. 145, p. 104318, 2024.

[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*. University of California, San Diego, 2006.

[20] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[21] P. Bandi, O. Geessink, Q. Manson, M. Van Dijk, M. Balkenhol, M. Hermsen *et al.*, "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge," *IEEE Trans. Medical Imaging*, vol. 38, no. 2, 2018.

[22] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, "Adaptive color deconvolution for histological wsi normalization," *Computer methods and programs in biomedicine*, vol. 170, pp. 107–120, 2019.

[23] F. Pérez-Bueno, J. G. Serra, M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Bayesian k-svd for h and e blind color deconvolution. applications to stain normalization, data augmentation and cancer classification," *Comp. Med. Imag. and Graphics*, vol. 97, p. 102048, 2022.