

# DEEP BAYESIAN BLIND COLOR DECONVOLUTION OF HISTOLOGICAL IMAGES

Shuowen Yang\*      Fernando Pérez-Bueno<sup>†\*\*</sup>      Francisco M. Castro-Macías<sup>†\*\*</sup>      Rafael Molina<sup>†</sup>  
Aggelos K. Katsaggelos<sup>‡</sup>

\* School of Optoelectronic Engineering, Xidian University, Xi'an, China.

<sup>†</sup>Dpto. Ciencias de la Computación e Inteligencia Artificial of the University of Granada, Spain.

<sup>\*\*</sup> Research Center for Information and Communication Technologies (CITIC-UGR)

<sup>‡</sup>Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA.

## ABSTRACT

Histological images are often tainted with two or more stains to reveal their underlying structures and conditions. Blind Color Deconvolution (BCD) techniques separate colors (stains) and structural information (concentrations), which is useful for the processing, data augmentation, and classification of such images. Classical BCD methods rely on a complicated optimization procedure that has to be carried out on each image independently, i.e., they are not amortized methods. In contrast, once they have been trained, deep neural networks can be used in a fast, amortized manner on unseen inputs. Unfortunately, the lack of large databases of ground truth color and concentrations has limited the development of deep models for BCD. In this work, we propose a deep variational Bayesian BCD neural network (BCD-Net) for stain separation and concentration estimation. BCD-Net is trained by maximizing the evidence lower bound of the observed images, which does not require the use of ground truth examples of stains and concentrations. Results obtained using two multicenter databases (Camelyon-17 and a stain separation benchmark) demonstrate the effectiveness of BCD-Net in the stain separation tasks, while drastically reducing the computation time compared to classical non-amortized methods.

**Index Terms**— Blind Color Deconvolution, Deep Variational Bayes, Stain Separation, Histological Images

## 1. INTRODUCTION

Histological images are stained to highlight the tissue structure. Pathologists, then easily identify the elements in the tissue by their distinctive staining. When using Hematoxylin and Eosin (H&E), the former highlights nuclei and the latter highlights cytoplasm and connective tissue. Blind Color Deconvolution (BCD) techniques provide a framework for identifying the stains in the images and using the separated information (color and structure) of each stain. Stain separation has proven to be useful for the automated diagnosis [1], nuclei segmentation [2], color normalization [3], and data augmentation of histological images [4, 5].

Unfortunately, differences in image acquisition protocol introduce intra and inter-hospital chromatic variability, which complicates the BCD stain separation. Therefore, BCD often requires computationally expensive optimization methods where the parameters

have to be estimated for each image [6, 2, 4]. Thus, the obtained models are not amortized, which makes time-consuming and expensive to separate the stains of large volumes of images. To solve this problem, we propose the use of Deep Learning (DL) approaches to train an amortized model capable of estimating the stain color and structure. Unfortunately, the stain-separated ground truth (color and concentrations) are rarely available, making it challenging to train fully supervised data-driven approaches.

In this work, we propose to address this lack of ground truth by combining analytical and DL approaches [7]. We are inspired by the work in [8], which combines per-image analytical modeling and a DL framework to solve the Blind Image Deblurring (BID) problem. The proposed modeling uses blurred and clean ground truth images to define data-driven priors. Unfortunately, in the case of histological images, the stain-separated ground truth is neither available nor easy to obtain, so we need to find an alternative way to introduce prior information.

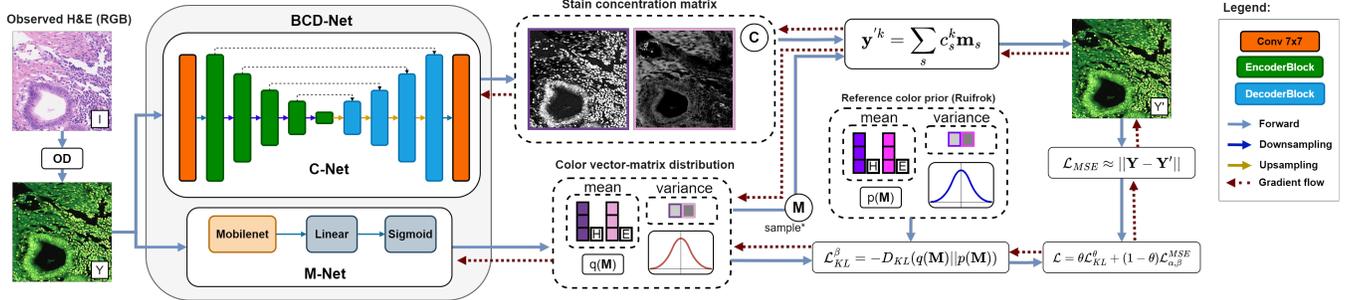
This paper is organized as follows. In section 1.1 we describe related work. In section 2 we detail the proposed deep variational BCD model and the inference process. Section 3 contains the experimental results and analysis. Finally, section 4 reports the conclusions of this work.

### 1.1. Related work

In 2001, Ruifrok *et al.* [9] proposed the use of the Beer-Lambert law and a standard color-vector matrix, which have been widely used since then. However, to deal with color variation a color-vector matrix must be estimated for each image. Since then, several optimization-based methods have been proposed (See [1] for a complete review). They use different techniques for this task, such as Singular Value Decomposition (SVD) [6], Non-Negative Matrix Factorization (NMF) [2], Independent Component Analysis (ICA) [10], or clustering techniques [11, 12]. As mentioned above, these methods are not amortized, i.e., the optimization procedure has to be repeated for each image.

Amortized DL approaches to BCD are scarce. Duggal *et al.* [13] propose a color deconvolution layer to be appended to classification CNNs. The parameters of the layer, which emulate the color-vector matrix, are initialized using [6] and optimized during training. Similarly, in [14], a capsule network is used to generate multiple stain separation candidates that are assembled with a sparse constraint. Both [13] and [14] fix their parameters after training and thus do not take color variation into account. The work by Abousamra *et al.* [15] uses an autoencoder for stain separation of immunohistochemistry images, trained with manually placed dot labels as weak

This work was supported by project B-TIC-324-UGR20 funded by FEDER/Junta de Andalucía and Universidad de Granada and PID2019-105142RB-C22 funded by MCIN/AEI/10.13039/501100011033. The work by Francisco M. Castro-Macías was sponsored by Ministerio de Universidades under FPU contract FPU21/01874.



**Fig. 1:** Overall architecture for BCD-Net including the subnetworks C-Net and M-Net, the overview of the Bayesian framework and loss. For architecture design and details, see section 2. (\*) Indicates that the sampling of  $q(\mathbf{M})$  is done using the reparametrization trick.

supervision. Unfortunately, this work cannot be extended to other staining protocols without a labeled dataset.

Bayesian approaches to BCD have yielded impressive results. In [16] the Beer-Lambert law is used to propose an observation model in the Optical Density (OD) space, together with a reference matrix based prior on the color-vectors, and a Simultaneous Autoregression (SAR) prior model on the stain concentrations. This work is extended in [17] with a Total Variation (TV) prior on the concentrations and in [3] with general Super-Gaussian (SG) priors. The work in [4] uses a dictionary-learning approach and proposes the use of Bayesian K-SVD (BKSVD). These methods have been adapted from Probabilistic Blind Image Deblurring (BID) methods [18], which is closely related to BCD. The goal of BID is to estimate the blur kernel and the underlying clean image from a blurred observed image. In BCD, the goal is to estimate the color-vector matrix and the underlying stain concentrations from a multi-stained image.

DL BID methods usually rely on the easily obtainable ground truth images and blur for training, which has prevented their adaptation to BCD where stain color-vector matrices and concentrations are hardly available. The recently proposed variational network in [19] and its extension in [8] propose a combination of probabilistic analytical (per image) techniques with a probabilistic amortized DL formulation. The authors use a data-driven prior, and a Dirichlet distribution to define the priors on the real underlying image and blur, respectively. They infer two networks that approximate the corresponding posteriors.

With the lessons learned from the amortized BID method in [8], and the analytical (per-image) BCD methods in [16, 17, 3], we propose BCD-Net, an amortized deep variational Bayesian model for BCD, with two subnetworks, C-Net and M-Net, that infer the concentrations and stain-color vector. BCD-Net is trained using the Evidence Lower Bound (ELBO) of the log-likelihood of the observations, which does not require the stain-color vectors and concentrations ground truth.

## 2. DEEP VARIATIONAL BAYESIAN BLIND COLOR DECONVOLUTION

Given an observed histological RGB image  $\mathbf{I} = [i^{kc}]$  with  $HW \times 3$  pixels, we transform it into the *Optical Density* (OD) space [3]. For each RGB channel  $c \in \{R, G, B\}$  and pixel  $k \in \{1, \dots, HW\}$ , the corresponding OD value is defined as  $y^{kc} = -\log(i^{kc}/255)$ . From these values, we define the corresponding OD image as  $\mathbf{Y} = [y^{kc}] \in \mathbb{R}^{HW \times 3}$ . According to the Beer-Lambert law,

$$\mathbf{Y}^\top = \mathbf{M}\mathbf{C} + \mathbf{N} \quad (1)$$

where  $\mathbf{M} \in \mathbb{R}^{3 \times N_s}$  is the color-vector matrix,  $\mathbf{C} \in \mathbb{R}^{N_s \times HW}$  is the concentration matrix and  $\mathbf{N} \in \mathbb{R}^{3 \times HW}$  is a noise matrix.

Here, we consider the joint probability  $p(\mathbf{C}, \mathbf{M}, \mathbf{Y}) = p(\mathbf{C})p(\mathbf{M})p(\mathbf{Y} | \mathbf{M}, \mathbf{C})$ , where each factor is defined as follows. Ideally, a data-driven prior could be used for the concentration matrix prior  $p(\mathbf{C})$  and the color matrix prior  $p(\mathbf{M})$  [8]. However, there is not a large enough dataset of stain-separated ground truth examples. To define  $p(\mathbf{C})$ , distributions that provide general information about the concentrations could be explored. Since this increases the complexity of the model, we decide to keep it simple and use an improper flat prior  $p(\mathbf{C}) \propto \text{const}$ . As it will be explained later, this model amounts to using maximum likelihood to estimate the concentrations. To define the color matrix prior we take into account that the staining protocol (e.g. H&E) is known and it is generally accepted that the color vectors are always close to those provided by Ruifrok's reference matrix [9, 20, 16]. We denote this matrix by  $\mathbf{M}^{\text{Rui}}$ , with columns  $\mathbf{m}_s^{\text{Rui}}$ , and use it to define the following prior on  $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{N_s}]^\top$ ,

$$p(\mathbf{M}) = \prod_{s=1}^{N_s} p(\mathbf{m}_s) = \prod_{s=1}^{N_s} \mathcal{N}(\mathbf{m}_s | \mathbf{m}_s^{\text{Rui}}, (\sigma_s^{\text{Rui}})^2 \mathbf{I}). \quad (2)$$

The square root of the variances  $\sigma_1^{\text{Rui}}, \dots, \sigma_{N_s}^{\text{Rui}}$  control the amount of variation allowed in each stain. For this work, we fix these values to  $\sigma^{\text{Rui}} = 0.05$ , assuming a moderate variance from the reference  $\mathbf{M}^{\text{Rui}}$ . Finally, the observation model in equation (1) can be written as

$$p(\mathbf{Y} | \mathbf{M}, \mathbf{C}) \propto \exp \left[ -\lambda^2 \left\| \mathbf{Y}^\top - \mathbf{M}\mathbf{C} \right\|_{\text{F}}^2 \right], \quad (3)$$

where  $\|\cdot\|_{\text{F}}$  denotes the Frobenius norm and  $\lambda > 0$  is the noise variance of the observation model.

### 2.1. Inference

Our goal is to estimate  $\mathbf{C}$  and  $\mathbf{M}$  from each observation  $\mathbf{Y}$ . To do so, we need to compute the posterior  $p(\mathbf{C}, \mathbf{M} | \mathbf{Y})$ . Since it does not admit an analytic expression, we decide to use variational inference and approximate it by  $q(\mathbf{C}, \mathbf{M} | \mathbf{Y}) = q(\mathbf{C} | \mathbf{Y})q(\mathbf{M} | \mathbf{Y})$ . This approach was followed in [16, 17, 3] where different priors were used for  $\mathbf{C}$  (i.e., SAR, TV, and SG), obtaining non-amortized methods. In this work, to build the inference model  $q(\mathbf{C}, \mathbf{M} | \mathbf{Y})$ , we use two DNNs, called C-Net and M-Net. C-Net has network parameters  $\alpha$  and returns the prediction of the concentration matrix  $\mathbf{C}^\alpha(\mathbf{Y}) \in \mathbb{R}^{N_s \times HW}$  for a given image  $\mathbf{Y}$ . The posterior  $q_\alpha(\mathbf{C} | \mathbf{Y})$  is chosen to be the degenerate distribution given by

$$q_\alpha(\mathbf{C} | \mathbf{Y}) = \begin{cases} 1 & \text{if } \mathbf{C} = \mathbf{C}^\alpha(\mathbf{Y}) \\ 0 & \text{otherwise} \end{cases}. \quad (4)$$

M-Net, with parameters  $\beta$ , returns the means  $\mu_1^\beta(\mathbf{Y}), \dots, \mu_{N_s}^\beta(\mathbf{Y})$  and variances  $\sigma_1^\beta(\mathbf{Y})^2, \dots, \sigma_{N_s}^\beta(\mathbf{Y})^2$  for the approximated posterior  $q_\beta(\mathbf{M} | \mathbf{Y})$ , which is chosen to be a product of Gaussian distributions, that is,

$$q_\beta(\mathbf{M} | \mathbf{Y}) = \prod_{s=1}^{N_s} \mathcal{N} \left( \mathbf{m}_s \mid \mu_s^\beta(\mathbf{Y}), \sigma_s^\beta(\mathbf{Y})^2 \mathbf{I}_{3 \times 3} \right). \quad (5)$$

To estimate  $\alpha$  and  $\beta$  we use a very large dataset of histological images  $\mathcal{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}$  and maximize the Evidence Lower Bound (ELBO) of the log-likelihood of the observations,  $\text{ELBO} = \sum_{\mathbf{Y} \in \mathcal{Y}} \text{ELBO}(\mathbf{Y})$ , where

$$\text{ELBO}(\mathbf{Y}) = \mathbb{E}_{q(\mathbf{C}, \mathbf{M} | \mathbf{Y})} \left[ \log \frac{p(\mathbf{C}, \mathbf{M}, \mathbf{Y})}{q(\mathbf{C}, \mathbf{M} | \mathbf{Y})} \right] = \quad (6)$$

$$= -\mathbb{E}_{q_\alpha(\mathbf{C} | \mathbf{Y})} \left[ \log \frac{q_\alpha(\mathbf{C} | \mathbf{Y})}{p(\mathbf{C})} \right] \quad (7)$$

$$- \mathbb{E}_{q_\beta(\mathbf{M} | \mathbf{Y})} \left[ \log \frac{q_\beta(\mathbf{M} | \mathbf{Y})}{p(\mathbf{M})} \right] \quad (8)$$

$$- \lambda^2 \mathbb{E}_{q_\beta(\mathbf{M} | \mathbf{Y})} \left[ \left\| \mathbf{Y}^\top - \mathbf{M} \mathbf{C}^\alpha(\mathbf{Y}) \right\|_{\text{F}}^2 \right] + \text{const.} \quad (9)$$

Since  $p(\mathbf{C})$  is improper and  $q_\alpha(\mathbf{C})$  is degenerate the term in the equation (7) is not properly defined and is not considered. Actually, this term is minus infinity since the divergence between these two distributions is infinite. Note that this is not a problem: if we had used maximum likelihood for the C-Net parameters and the same approximate posterior distribution  $q_\beta(\mathbf{M})$  for the color matrix, this term would have not appeared. Then, the expectation in the equation (8) corresponds to minus the Kullback-Leibler divergence between the two distributions in equations (2) and (5). This divergence, denoted from now on as  $\mathcal{L}_{\text{KL}}^\beta(\mathbf{Y})$ , can be computed in closed form using the well known Kullback-Leibler divergence between two Gaussian distributions,

$$\begin{aligned} \mathcal{L}_{\text{KL}}^\beta(\mathbf{Y}) &= \sum_{s=1}^{N_s} \frac{\| \mu_s^\beta(\mathbf{Y}) - \mathbf{m}_s^{\text{Rui}} \|^2}{2(\sigma_s^{\text{Rui}})^2} + \\ &\frac{3}{2} \sum_{s=1}^{N_s} \left( \frac{\sigma_s^\beta(\mathbf{Y})^2}{(\sigma_s^{\text{Rui}})^2} - \log \frac{\sigma_s^\beta(\mathbf{Y})^2}{(\sigma_s^{\text{Rui}})^2} - 1 \right). \end{aligned} \quad (10)$$

Finally, although the expectation in the equation (9) also admits a closed-form expression, we have found that the training procedure is much more stable if we use the reparameterization trick instead [21]. This trick allows us to obtain an unbiased differentiable estimator of  $\mathbb{E} \left[ \left\| \mathbf{Y}^\top - \mathbf{M} \mathbf{C}^\alpha(\mathbf{Y}) \right\|_{\text{F}}^2 \right]$ , which will be denoted by  $\mathcal{L}_{\text{MSE}}^{\alpha, \beta}(\mathbf{Y})$  in the following.

In summary, the ELBO in equation (6) is approximated as  $\text{ELBO}(\mathbf{Y}) \approx -\mathcal{L}_{\text{KL}}^\beta(\mathbf{Y}) - \lambda^2 \mathcal{L}_{\text{MSE}}^{\alpha, \beta}(\mathbf{Y})$ . Instead of maximizing the approximated ELBO, we equivalently minimize the negative approximated ELBO, which yields the following objective

$$\min_{\alpha, \beta} \sum_{\mathbf{Y} \in \mathcal{Y}} \left[ \mathcal{L}_{\text{KL}}^\beta(\mathbf{Y}) + \lambda^2 \mathcal{L}_{\text{MSE}}^{\alpha, \beta}(\mathbf{Y}) \right]. \quad (11)$$

The two terms in the above equation play very important roles. The first monitors M-Net to provide a color distribution close to the prior and the second combines both C-Net and M-Net to provide a good reconstruction of the observed image according to the observation model. The values of  $\lambda^2$  and  $\sigma_s^{\text{Rui}}$  determine the balance between the two terms. To experimentally determine how this balance affects the results, we redefine the objective in equation (11) to

include a weighting parameter  $0 < \theta = \lambda^2 / (1 + \lambda^2) < 1$  (whose importance will be evaluated in the ablation study in section 3),

$$\min_{\alpha, \beta} \sum_{\mathbf{Y} \in \mathcal{Y}} \left[ \theta \mathcal{L}_{\text{KL}}^\beta(\mathbf{Y}) + (1 - \theta) \mathcal{L}_{\text{MSE}}^{\alpha, \beta}(\mathbf{Y}) \right], \quad (12)$$

In summary, the proposed inference model uses two networks, C-Net and M-Net. Both branches, C-Net and M-Net, jointly define BCD-Net for the Bayesian modeling and inference presented in this section. Each of them has a specific task: to estimate the stain concentrations in the case of C-Net and to estimate the color-matrix posterior in the case of M-Net. However, they are jointly trained to reconstruct the observed image according to the Beer-Lambert model in equation (3). In addition, M-Net is also constrained by the prior defined on the color vectors. Note that using two subnetworks to boost the performance of a joint task, which amounts to an independent assumption in the posterior approximation, is a common approach in blind image deblurring [8, 22] and denoising [19].

## 2.2. Network architecture

To build C-Net we follow [8] and use Unet [23], which was first proposed for biological image segmentation. The output of C-Net has two channels of the same size as the input image, one for the concentration of each stain in the H&E image. We use four scales in both the encoder and decoder, where each encoder and decoder block contains three stacked ResBlocks with LeakyReLU activation and a small kernel size of  $3 \times 3$ . The number of channels per layer is set to 64. Each downsampling block uses a convolution layer with a  $3 \times 3$  filter and a stride of 2. The upsampling uses a transposed convolution with a  $5 \times 5$  filter and a stride of 2. M-Net uses MobileNet V3 Small [24] followed by a linear fully connected layer. The output of M-Net is the estimation of the means and the logarithm of the variances of the variational approximation of the color-vector matrix posterior. As the mean is required to have a unitary norm, we include an L2 constraint [25] for its corresponding M-Net output.

## 3. EXPERIMENTS

### 3.1. Experimental settings

**Datasets.** We use two widely used datasets in histological color deconvolution-related tasks. To train and validate BCD-Net, we use data obtained from the Camelyon-17 breast cancer classification challenge [26]. This dataset contains images from five different medical centers in the Netherlands, with intra- and inter-center color variations. We use 500 slides (100 from each center) that were released as the training set for the challenge. We use non-overlapping patches of size  $224 \times 224$  with at least 70% tissue. Our model is trained using images from three different centers and validated using images from the remaining two centers, which allow us to capture a wide range of variations. This dataset does not include a stain separation ground truth. The Warwick Stain Separation Benchmark (WSBB) [10] is used to assess the stain separation performance of BCD-Net. WSSB is of special interest to assess the performance of BCD methods because it includes the ground truth stain separation for 24 images of three different tissue types (breast, lung, and colon). The stain color matrix was obtained by asking pathologists to identify pixels of biological structures that were stained with a single stain. Then, the ground truth concentrations  $\mathbf{C}_{\text{GT}}$  were obtained in [10] as  $\mathbf{C}_{\text{GT}} = \mathbf{M}_{\text{GT}}^+ \mathbf{Y}$ , where  $\mathbf{M}_{\text{GT}}^+$  is the Moore-Penrose pseudo-inverse of the color-matrix ground truth.

**Table 1:** PSNR for BCD-Net with different values of  $\theta$  on the WSSB dataset [10]. The best value is highlighted in bold.

Value of $\theta$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Mean PSNR	22.57	24.03	24.18	<b>24.76</b>	23.78	23.80	22.84	21.81	21.84	21.46	21.11

**Table 2:** PSNR for the best performing methods on WSSB. The best value of each pair (subset, stain) is highlighted in bold.

Subset	Stain	RUI	MAC	VAH	ALS	SAR	TV	SGP	BKSVD	BCD-Net
Lung	H	22.47	19.52	25.87	20.62	32.91	33.10	<b>35.21</b>	32.67	27.32
	E	22.05	18.09	25.53	23.95	30.77	31.02	<b>33.07</b>	30.61	24.96
Breast	H	15.27	26.24	25.46	24.60	28.81	29.14	30.50	<b>32.20</b>	25.58
	E	17.66	23.62	27.68	25.92	26.60	26.76	27.71	<b>29.43</b>	23.11
Colon	H	22.27	23.91	25.83	21.11	28.57	28.62	29.01	<b>34.08</b>	25.03
	E	20.70	21.55	26.29	21.94	27.58	27.60	28.38	<b>33.32</b>	22.55
Mean	H	20.00	23.22	25.72	22.11	30.10	30.29	31.57	<b>32.98</b>	25.98
	E	20.14	21.08	26.50	23.94	28.32	28.46	29.72	<b>31.12</b>	23.54
	Mean	20.07	22.15	26.11	23.03	29.21	29.38	30.65	<b>32.05</b>	24.76

**Implementation.** BCD-Net is built using Pytorch and trained on four NVIDIA GeForce RTX 3090. We use a batch size of 64 (16 per GPU) and the ADAM optimizer with an initial learning rate of  $10^{-4}$ , which is halved every 3 epochs if the loss does not decrease. To facilitate the network to establish the right order in the H&E channels, a pretraining epoch is added in which  $\theta$  is set to 0.99. We train for a maximum of 100 epochs, with an early stopping callback that halts the training procedure if the loss has not decreased for 10 consecutive epochs. The weights corresponding to the best-performing epoch on the validation data are saved and used in the following experiments.

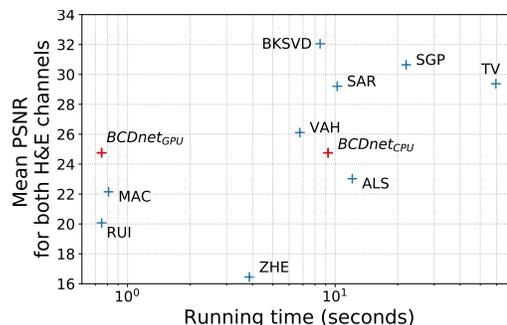
### 3.2. Stain separation performance

To quantitatively assess the model performance on the stain separation task we analyze the fidelity of the stain-separated RGB images to those obtained using the ground truth. We compare them using the Peak Signal-to-Noise ratio (PSNR). For each image, we obtain a PSNR value for each stain image (Hematoxylin and Eosin) and a PSNR value corresponding to the mean of both. The mean PSNR obtained for each value of  $\theta$  is presented in table 1. Recall that small values of  $\theta$  increase the importance of the reconstruction term  $\mathcal{L}_{MSE}^{\alpha,\beta}(\mathbf{Y})$ , while large values of  $\theta$  balance the loss towards similarity to the prior. These results show that the regularization term imposed by  $\mathcal{L}_{KL}^{\beta}(\mathbf{Y})$  is relevant for a better stain separation. However, excessive fidelity to the prior hampers a faithful estimation of the stains in the image.

In table 2 we present a comparison between BCD-Net and other non-amortized state-of-the-art methods. These include the classical non-blind CD method by Ruifrok *et al.* (RUI) [9], the non-amortized methods by Macenko *et al.* (MAC) [6], Vahadane *et al.* (VAH) [2], Alsubaie *et al.* (ALS) [10], and Zheng *et al.* (ZHE) [12]. We also include the Bayesian methods Simultaneous Autoregressive (SAR) [16], Total Variation (TV) [17], Super Gaussian Priors (SGP) [3], and Bayesian K-SVD (BKSVD) [4]. The comparison shows that the proposed method is still far from the latest non-amortized methods, but the results are promising. BCD-Net outperforms RUI, ZHE, MAC, and ALS, and it is competitive with the commonly used VAH. The figures of merit show that the amortized BCD-Net model produces an appropriate estimation of the concentration and color-vector matrices, even though it is trained without stain-separated ground truth samples.

### 3.3. Computational Efficiency

We analyze the time taken by each of the methods in the table 2 to perform the stain separation of a  $2000 \times 2000$  image from the WSSB dataset. The proposed BCD-Net is run on an NVIDIA GeForce GTX3090 X GPU and in the same CPU as the rest of the methods. The results are presented in a biplot (see figure 2) with the average required time versus the mean PSNR from table 2. These figures allow us to visually compare the time efficiency and the stain separation quality of each model. The best methods are those near the upper left corner. When using the GPU, the proposed method is the fastest, requiring an average of 0.74 seconds. Note that this makes our method even faster than the non-blind method RUI, which requires 0.81 seconds. If the GPU is not available, the proposed BCD-Net can also run on the CPU, where it takes 9.24 seconds, making it competitive in time with the state-of-the-art non-amortized methods, and still faster than SAR, ALS, SGP, and TV.



**Fig. 2:** Mean PSNR vs running time for deconvolving a  $2000 \times 2000$  image. The proposed BCD-Net is marked in red.

## 4. CONCLUSIONS

In this paper, we have proposed a novel Deep Variational Bayesian Blind Color Deconvolution Neural Network (BCD-Net) for stain separation of histological images. BCD-Net combines DL with analytical Bayesian modeling to train a DNN by maximizing the ELBO of the observed optical density images without any stain separation ground truth. The proposed model includes two subnetworks, C-Net and M-Net, which jointly estimate a posterior distribution for the color-vector matrix and the stain concentrations using maximum likelihood in an amortized inference manner. Both outputs are used to reconstruct the observed image according to the Beer-Lambert law. This innovative approach has provided promising results in the stain separation task while significantly reducing the computational demands of non-amortized models used in the literature. Moreover, since our model relies on maximum likelihood to estimate the C-Net parameters, we believe there is potential for improvement by considering different prior distributions for the stain concentrations. Finally, we believe that this work paves the way for further exploration of the BCD problem from a DL perspective and can be used to further explore the field of histological image analysis.

## 5. REFERENCES

- [1] Neel Kanwal, Fernando Pérez-Bueno, Arne Schmidt, Rafael Molina, and Kjersti 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*, pp. 1–1, 2022.
- [2] 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.
- [3] Fernando Pérez-Bueno, Miguel Vega, María A. Sales, José Aneiros-Fernández, Valery Naranjo, Rafael Molina, and Aggelos 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.
- [4] F. Pérez-Bueno, J.G. Serra, M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, “Bayesian k-svd for h&e blind color deconvolution. applications to stain normalization, data augmentation, and cancer classification,” *Comput. Med. Imaging Graph.*, 2022.
- [5] David Tellez et al., “Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology,” *Med. Image Anal.*, vol. 58, pp. 101544, 2019.
- [6] M. Macenko et al., “A method for normalizing histology slides for quantitative analysis,” in *Int. Symp on biomed Imaging (ISBI)*, 2009, pp. 1107–1110.
- [7] A. Lucas, M. Iliadis, R. Molina, and A.K. Katsaggelos, “Using deep neural networks for inverse problems in imaging,” *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 20–36, 2018.
- [8] Qian Zhao et al., “A deep variational bayesian framework for blind image deblurring,” *Knowl. Based Syst.*, p. 109008, 2022.
- [9] A. C. Ruifrok and D. A. Johnston, “Quantification of histochemical staining by color deconvolution,” *Anal. Quant. Cytol. Histol.*, vol. 23, pp. 291–299, 2001.
- [10] N. Alsubaie et al., “Stain deconvolution using statistical analysis of multi-resolution stain colour representation,” *PLOS ONE*, vol. 12, pp. e0169875, 2017.
- [11] Massimo 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.
- [12] Yushan Zheng et al., “Adaptive color deconvolution for histological WSI normalization,” *Comput. Meth. Prog. Bio.*, vol. 170, pp. 107–120, Mar. 2019.
- [13] Rahul Duggal et al., “SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging,” in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, 2017, vol. 10435, pp. 435–443.
- [14] Yushan Zheng et al., “Stain standardization capsule for application-driven histopathological image normalization,” *IEEE J. biomed and Health Inform.*, vol. 25, no. 2, pp. 337–347, 2021.
- [15] Shahira Abousamra et al., “Weakly-supervised deep stain decomposition for multiplex ihc images,” *IEEE Int. Symp. Biomed. Imaging (ISBI)*, pp. 481–485, 2020.
- [16] 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.
- [17] Fernando Pérez-Bueno, Miguel López-Pérez, Miguel Vega, Javier Mateos, Valery Naranjo, Rafael Molina, and Aggelos 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.
- [18] Pablo Ruiz, Xu Zhou, Javier Mateos, Rafael Molina, and Aggelos K. Katsaggelos, “Variational bayesian blind image deconvolution: A review,” *Digit. Signal Process.*, vol. 47, pp. 116–127, 2015.
- [19] Zongsheng Yue et al., “Variational denoising network: Toward blind noise modeling and removal,” in *Adv. Neural Inf. Process. Syst.* 32, pp. 1690–1701, 2019.
- [20] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Jun Shi, and Chenghai Xue, “Adaptive color deconvolution for histological wsi normalization,” *Comput. Meth. Prog. Bio.*, vol. 170, pp. 107–120, 2019.
- [21] Diederik P Kingma and Max Welling, “Auto-encoding variational bayes,” 2014.
- [22] Jinshan Pan et al., “Physics-based generative adversarial models for image restoration and beyond,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2449–2462, 2021.
- [23] Olaf Ronneberger et al., “U-net: Convolutional networks for biomedical image segmentation,” in *Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
- [24] Andrew Howard et al., “Searching for mobilenetv3,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.
- [25] Rajeev Ranjan et al., “L2-constrained softmax loss for discriminative face verification,” *arXiv:1703.09507 [cs.CV]*, 2017.
- [26] Péter Bándi et al., “From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge,” *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 550–560, 2019.