Digit. Signal Process. 145 (2024) 104318

Contents lists available at ScienceDirect

Digital Signal Processing

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

BCD-net: Stain separation of histological images using deep variational Bayesian blind color deconvolution

Shuowen Yang [a,1], Fernando Pérez-Bueno [b,c,*,1], Francisco M. Castro-Macías [b,c], Rafael Molina [b], Aggelos K. Katsaggelos [d]

[a] School of Optoelectronic Engineering, Xidian University, Xi'an, China
[b] Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Spain
[c] Research Center for Information and Communication Technologies (CITIC-UGR), Universidad de Granada, Spain
[d] Dept. of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA

ARTICLE INFO

Keywords:
Deep Bayesian modeling
Variational inference
Histological images
Stain separation
Blind color deconvolution

ABSTRACT

Histological images are often tainted with two or more stains to reveal their underlying structures. 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 analytical BCD methods are typically computationally expensive in two distinct ways. First, estimating the colors and concentrations corresponding to a given image is a time-consuming process. Second, the entire estimation procedure must be performed independently for each image.

In contrast, Deep Learning (DL) methods involve high training costs, but once trained, they are able to directly process unseen images. The application of DL to BCD has been limited by the absence of extensive databases containing ground truth color and concentrations. In this work, we propose BCD-Net, a deep variational Bayesian neural network for stain separation and concentration estimation. Under this framework, we address the challenge of lacking ground truth data by leveraging Bayesian modeling and inference techniques.

We propose to use a prior distribution on the stain colors, and a simple flat prior on the concentrations. BCD-Net is trained by maximizing the evidence lower bound of the observed images. The loss function comprises two essential components: fidelity to the observed images and the Kullback-Leibler divergence between the estimated posterior distribution of colors and the selected prior.

The model is trained, validated, and tested on two multicenter databases: Camelyon-17 and Warwick stain separation benchmark. The proposed approach is tested on image reconstruction, stain separation, and cancer classification. It performs well when contrasted with classical non-amortized methods and offers a substantial computational time advantage. This marks a significant step forward in the application of DL techniques to address BCD and paves the way for new approaches.

1. Introduction

Staining is at the core of histological image analysis. Tissues must be tainted using a combination of stains to reveal their underlying structures. Then, they are scanned to obtain histological images that can be analyzed by pathologists and/or Computer-Aided Diagnosis (CAD) systems [14].

Although data-driven CAD systems work well in several areas of diagnosis, their performance degrades significantly when tested on images from hospitals not included in the training set [33]. Intra- and inter-hospital chromatic variability [14,33] caused by differences in the staining of the images is often considered one of the major causes of the loss of performance. Beyond training CAD systems with more multi-hospital data, several approaches have been proposed to tackle chro-

* Corresponding author at: Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Spain.
E-mail addresses: shuowenyang@xidian.edu.cn (S. Yang), fpb@ugr.es (F. Pérez-Bueno), fcastro@ugr.es (F.M. Castro-Macías), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos).
[1] Equal contribution.

https://doi.org/10.1016/j.dsp.2023.104318

Available online 20 November 2023
1051-2004/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

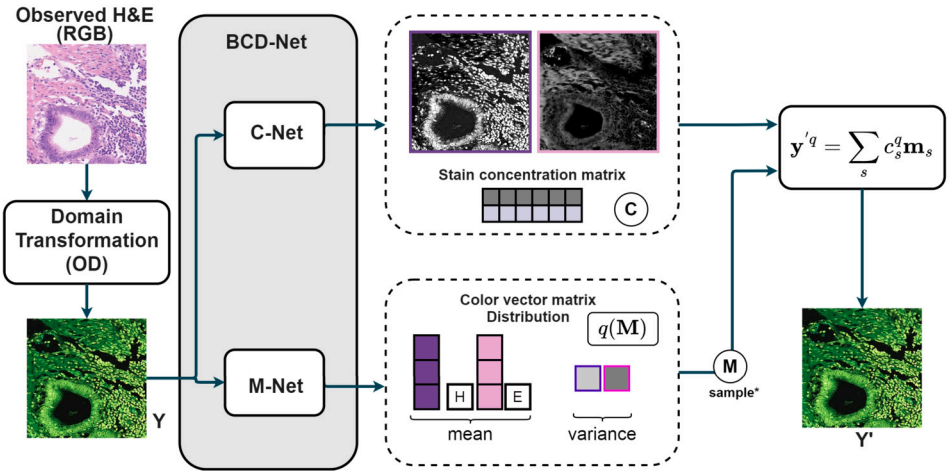


Fig. 1. Graphical abstract. Overall architecture for BCD-Net ($S = 2$) including the subnetworks C-Net and M-Net. First, the image is transformed to the Optical Density (OD) space. Then, it is fed to both subnetworks. C-Net outputs two single-stain concentration images with the same size as the input. M-Net outputs a $3 \times S$ color-vector matrix and S 3×3 diagonal matrices. These correspond, respectively, to the mean vectors and the covariance matrices of the approximated posterior q($\mathbf{M} \mid \mathbf{Y}$). Finally, the outputs from both subnetworks are combined following Eq. (2) to recover the OD image. For architecture design and details, see section 3. (*) Indicates that the sampling of q($\mathbf{M}$) is done using the reparametrization trick.

matic variability. Of all of them, three stand out. The first, called *stain normalization*, aims at transforming the observed images into new ones as if they had been tainted using the same process. The second, called *data augmentation*, aims at hallucinating new images with augmented chromatic variability, with the objective of reducing the generalization error of CAD systems on unseen data. See [14] for more details.

In this work, we focus on the third one, called *stain separation*. Given that stains bind to specific elements in the tissue, chromatic variability jeopardizes the ability of the CAD system to correctly identify the structures of the tissues and their labels. By using stain separation techniques it is possible to identify and separate each stain in the image [14]. Note that the separation is more biologically meaningful than the mixture of stains in the observed RGB image [34]. Furthermore, stain separation has proven to be useful for automated diagnosis [8,9,24], as a preliminary step for stain normalization [19,25,35] and for data augmentation [24,33,41].

To tackle the stain separation problem, Blind Color Deconvolution (BCD) techniques are frequently used. After transforming the RGB image to the Optical Density (OD) space, the stain color vectors and the corresponding stain structures, known as concentrations, are estimated. Most of the BCD methods used so far are *analytical:* they utilize an optimization-based process that has to be repeated for each image. Unfortunately, calculating the color vectors, concentrations, and model parameters, for each image independently, is a daunting and computationally demanding task.

Deep Learning (DL) methods are *amortized*: once they are trained, they provide a fast performance on test images. Unfortunately, they require large datasets for training and the stain separation ground truth is rarely available. Therefore, most DL approaches for histological image processing have avoided the stain separation step and focused on stain normalization [5,39]. Most of these methods assume that the color distribution is laboratory-specific and perform style transfer between centers. However, unless intra-hospital color variations can be identified, they are unable to tackle it and, very importantly, they lose the interpretability that stain separation provides.

Blind Image Deblurring (BID) is a closely related field that shares similar challenges to BCD. From a blurry observed image, the goal of BID is to simultaneously estimate the blur kernel and the underlying clean image. In BCD, the goal is to estimate the color-vector matrix and the underlying stain concentrations from a multi-stained image. Typically, analytical BID techniques have been adapted to the BCD problem. Unfortunately, although ground truth images and blurs are easy to obtain, color-vector matrices and concentrations are hardly available, which has prevented the convenient and very useful adaptation of DL BID techniques to BCD.

In this work, we propose a deep variational Bayesian BCD neural network (BCD-Net) for stain separation and concentration estimation (see Fig. 1), which tackles the lack of ground truth by combining analytical Bayesian modeling and a DL framework [17]. Our work is inspired by Zhao et al. [40], for BID which uses blurred and clean ground truth images to define data-driven priors. Unfortunately, as we have already indicated, in the case of histological images, the stain-separated ground truth is neither available nor easy to obtain, so we must find an alternative way to introduce prior information.

This paper represents, to the best of our knowledge, the first attempt to use variational Bayesian DL techniques to amortizely solve the BCD problem. It is organized as follows. In section 2 we describe related studies, as well as our contributions to this paper. In section 3 we detail the proposed deep variational BCD model and inference process. We then present the network architecture and its two subnetworks in Section 4. Sections 5 and 5.6 contain the experimental results and analysis. Finally, section 6 reports the conclusions of this work. All the acronyms used in this paper have been collected in Table A.7, included at the end of the paper.

2. Related works and contributions

2.1. Blind color deconvolution methods

Ruifrok et al. [29] proposed the use of the logarithmically inverted OD space and a non-blind color deconvolution algorithm to obtain the stain concentrations. They used an experimentally obtained standard color-vector matrix through fixed RGB triplets for hematoxylin, eosin, and diaminobenzidine. However, the color variability between different images was not addressed.

To automatically find the color-vector matrix, the use of Non-negative Matrix Factorization was proposed by Rabinovich et al. [26]. The works by Vahadane et al. [35] and Xu et al. [37] used this technique and included sparsity regularization terms on the stain concentrations. To reduce its computational and memory cost, the use of Non-Negative Least Squares was proposed by Carey et al. [7].

Macenko et al. [19] proposed the use of Singular Value Decomposition (SVD) to find the optimal color-vector matrix and to separate the stains, setting an experimental threshold to remove noisy pixels. This work was later extended by McCann et al. [21] by taking into account

Table 1
Qualitative comparison with BCD related amortized methods.

Method	Addresses Color Variability	Concentration Estimation	Color-vector Matrix Estimation	Ground Truth Used	Applicable to H&E Images
Duggal et al. [8]	NO	YES	NO	Synthetic: Macenko et al. [19]	YES
Zheng et al. [41]	NO	YES	NO	Synthetic: Macenko et al. [19]	YES
Marini et al. [20]	YES	NO	YES	Synthetic: Macenko et al. [19] + tumor classification labels	YES
Abousamra et al. [1]	YES	YES	YES	single-stained dot labels	NO
BCD-Net	YES	YES	YES	Not required	YES

the interaction between dyes. Astola [4] proposed to use this decomposition in the linearly inverted RGB space instead of the OD space, taking into consideration the RGB sensibilities of the camera.

Independent Component Analysis was utilized by Trahearn et al. [34], including a correction step for those cases where the stain vectors were not orthogonal. Later, it was also used by Alsubaie et al. [2,3] to reduce the independence condition among sources.

In [10], stain vectors were estimated by projecting the input color image onto the Maxwellian chromaticity to form clusters. Khan et al. [15] estimated the color-vector matrix by performing Oct-tree quantization and a relevance vector machine to classify RGB pixels of interest. Recently, Salvi et al. [32] have presented a combination of segmentation and clustering strategies to obtain the color-vector matrix and separate the stains. Zheng et al. [42] proposed an adaptive method with an objective function that considers the minimization of residuals, the balance of the stains, and the overall energy using experimentally determined hyperparameters.

Probabilistic BID methods [30] have been adapted to solve the BCD problem. In BID, the blurred image is written as the convolution of the blur kernel and the latent clean image. Likewise, in BCD, following the Beer-Lambert law, once the observed RGB histological image is transformed to the OD space, it can be expressed as the product of the color-vector matrix and the concentration matrix. In [11,23,25] this observation model was used in combination with different prior models on the concentrations, and the same prior model on the color vectors: normal independent distributions with mean the Ruifrok's reference matrix [11,29,42] and covariances to be estimated. Obviously, these priors do not make use of the original stain separation. For all these probabilistic models, the variational inference was used to obtain, per image, the approximation of the posterior distribution approximation of the color stains and concentrations. The work in [24] formulates the estimation of the color-vector matrix as a dictionary-learning problem using Bayesian K-SVD. Notice that these, as well as all the above-described methods, perform BCD on a per-image basis.

As mentioned earlier, DL approaches to BCD are scarce due to the lack of the stain separation ground truth for training. To be able to train a Deep Neural Network (DNN) with stain-separated images, Duggal et al. [8] introduced a color deconvolution layer that can be appended to the input of the network. The parameters of the deconvolution layer emulate the color-vector matrix and are optimized during training (being initialized using [19]). This approach does not address chromatic variability between images and requires the normalization of the dataset before training. Similarly, in [20] a Convolutional Neural Network (CNN) was proposed to learn stain-invariant features by estimating the color-vector matrix and the classification label at the same time. Zheng et al. [41] proposed a capsule network to produce multiple stain separation candidates that were later assembled using a sparse constraint and utilized for normalization. The deconvolution step was also done with network parameters, therefore it was not able to adapt to unseen color distributions. None of the above-mentioned DL approaches evaluates the quality of the stain separation after their BCD steps. Finally, Abousamra et al. [1] proposed an Autoencoder for stain separation of multiplex immunohistochemistry images stained with six different stains using manually placed dot labels for single-stained pixels as weak supervision.

To conclude, in Table 1 we summarize the key differences between the proposed BCD-Net and the amortized methods mentioned above. Only the method by Abousamra et al. [1] estimates both the color-vector matrix and concentrations, but this approach cannot be extended to Hematoxylin-Eosin (H&E) images without a large annotated dataset. In addition, the other methods rely on synthetic ground truth, which could potentially introduce unexpected biases to the model.

2.2. DL stain normalization methods without BCD

In this section, we include DL-based stain normalization methods that do not use BCD. Although these works do not perform stain separation, they are of interest as they show how DL has been applied to the processing of histological images.

Janowczyk et al. [13] proposed the use of Sparse Autoencoders for stain normalization. Their approach separates the pixels into k clusters and then applies histogram equalization across clusters and RGB channels to obtain a color standardized image. Similarly, Zanjani et al. [39] proposed the use of deep generative models to separate pixels into k tissue classes and then normalize the stains in the images by considering the separation of the source and target image.

Bentaieb and Hamarneh [5] combined classification and stain normalization by using a Generative Adversarial Network where the generator normalizes the images and the discriminator distinguishes between original and normalized images at the same time that discriminates between benign and malign tissues. The training of this network requires images from at least two different centers, where one of them is considered the target appearance for normalized images.

Tellez et al. [33] proposed an U-Net-like architecture for stain normalization fed with heavily color-augmented images and trained to reconstruct their original appearance. When trained with images from a target center, the network should be able to transform new images to the same target color distribution.

Finally, we emphasize that other popular CNN architectures such as Pix2pix [31] or CycleGAN [43], have been adapted to stain normalization. The interested reader can refer to [14] for a complete review.

2.3. DL BID approaches of interest to DL BCD

Analytical probabilistic methods have been used to solve the BID problem [30] while CNNs have been trained to perform BID tasks in an amortized manner. See [36] for a recent survey of deep image deblurring and [17,18] for connections between analytical and DL deblurring methods.

The variational network proposed by Yue et al. [38] and its extension in [40] combine the probabilistic modeling of analytical techniques with an amortized DL formulation also based on probabilistic modeling and inference to solve image processing problems. For a denoising problem, Yue et al. [38] propose the use of a prior based on the underlying real image. Using DL techniques they infer a network whose output is the amortized posterior of the real underlying image given the observed one. In [40] the model is extended to deal with BID problems. The prior

on the image is the same as in [38] while the prior on the blur is defined using a Dirichlet distribution [44]. Using DL techniques once more, the authors infer two networks that approximate the amortized posteriors of the real underlying image and blur, respectively, given the observed image.

It is important to emphasize that, in the DL methods just described, the original blurs and images are used to define the corresponding prior models. Here, for the BCD problem, we have access neither to the original concentrations nor to the color vectors and so we can only make use of the observed histological images.

2.4. Our contributions

We combine ideas from the supervised methods for BID and denoising in [38,40] with analytical BCD methods in [11,23,25] to propose an amortized deep variational Bayesian model for BCD. The method is named BCD-Net and is trained without stain color vectors and concentrations ground truth. Our contributions are detailed next.

- We define an amortized DL probabilistic modeling of the BCD problem with three components: a normal prior on the color-vector matrix, a flat, improper prior on the concentrations, and an observation model given the stain color vectors and the concentrations.
- We design BCD-Net to have two subnetworks, C-Net and M-Net. We configure C-Net to produce the maximum likelihood estimation of the concentrations. Importantly, the use of maximum likelihood principles to estimate the C-Net parameters is simpler than outputting non-degenerate posterior distribution approximations. M-Net is designed to output a Gaussian approximation of the posterior distribution of the stain color vectors.
- For these posterior approximations, the Evidence Lower Bound (ELBO) to be optimized has two terms. The first one measures how close the estimated posterior of the stain color vectors is to the proposed color vector priors, while the second one accounts for the fidelity of the reconstruction to the observed image.
- The proposed BCD-Net is evaluated and compared against state-of-the-art methods in three different tasks: image reconstruction, stain separation, and breast cancer classification. These experiments show that BCD-Net reduces the computation time required by classical non-amortized methods while remaining competitive in terms of performance.

To the best of our knowledge, our model is the first amortized attempt to use variational deep neural networks for BCD, and so provides an interesting framework for future DL BCD methods.

3. Deep variational Bayesian blind color deconvolution

Our model relies on two networks C-Net and M-Net that are used as estimators in a Bayesian framework. Section 3.1 defines the generative model, the priors that introduce the previous knowledge about the unobservable stains and concentrations, and the observation model that relates them to the observed images. Section 3.2 details how the posterior distributions are defined and how the loss function is derived to train the network with the observed data.

3.1. Modeling

Let $\mathbf{I} \in \mathbb{R}^{HW\times 3}$ be a histological RGB image. We write $\mathbf{I} = \left[\mathbf{i}^1, \dots, \mathbf{i}^{HW}\right]^\top$ where $\mathbf{i}^k = \left[i^{Rk}, i^{Gk}, i^{Bk}\right]^\top \in \mathbb{R}^3$. Following [11], we define the transformation to the OD absorbency space as

$$y^{ck} = -\log(i^{ck}/255), \tag{1}$$

where $k \in \{1, \dots, HW\}$ is the pixel index and $c \in [R, G, B]$ is the RGB channel index. We denote $\mathbf{Y} = \left[\mathbf{y}^1, \dots, \mathbf{y}^{HW}\right]^\top \in \mathbb{R}^{HW\times 3}$, where $\mathbf{y}^k = \left[y^{Rk}, y^{Gk}, y^{Bk}\right]^\top \in \mathbb{R}^3$. According to the Beer-Lambert law, each pixel $\mathbf{y}^k$ in the OD image $\mathbf{Y}$ can be modeled as

$$\mathbf{y}^k = \mathbf{M}\mathbf{c}^k + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbf{M} \in \mathbb{R}^{3\times S}$ denotes the color-vector matrix, $\mathbf{c}^k \in \mathbb{R}^S$ is the stain concentration of the $k-th$ pixel and $\boldsymbol{\epsilon} \in \mathbb{R}^3$ is a noise vector. Here, S is the number of stains. We can decompose the contribution from each stain as follows,

$$\mathbf{y}^k = \sum_{s=1}^{S} c_s^k \mathbf{m}_s + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{m}_s \in \mathbb{R}^3$ is the color-vector associated to the $s-th$ stain and c_s^k is the concentration of the $s-th$ stain for pixel k. This way, $\mathbf{M} = \left[\mathbf{m}_1, \dots, \mathbf{m}_S\right]$ and $\mathbf{c}^k = \left[c_1^k, \dots, c_S^k\right]^\top$. In the following, we will use $\mathbf{C} = \left[\mathbf{c}^1, \dots, \mathbf{c}^{HW}\right] \in \mathbb{R}^{S\times HW}$ to denote the matrix with all the concentrations for all pixels.

We consider the following generative model

$$\mathrm{p}(\mathbf{C}, \mathbf{M}, \mathbf{Y}) = \mathrm{p}(\mathbf{C})\mathrm{p}(\mathbf{M})\mathrm{p}(\mathbf{Y} \mid \mathbf{M}, \mathbf{C}), \tag{4}$$

for which we must define the priors $\mathrm{p}(\mathbf{C})$, $\mathrm{p}(\mathbf{M})$ and the observation model $\mathrm{p}(\mathbf{Y} \mid \mathbf{M}, \mathbf{C})$. We start with $\mathrm{p}(\mathbf{C})$, an appealing idea would be to define a data-driven prior as in [40]. Unfortunately, a sufficiently large dataset of stain-separated ground truth concentrations does not exist. Another alternative would be to consider priors that provide general information about the concentrations [11,23,25], but since this would increase the complexity of the model, we decide to keep it simple and use the following improper flat prior

$$\mathrm{p}(\mathbf{C}) \propto \text{const}, \ \forall\ \mathbf{C}. \tag{5}$$

As will be explained later, this choice leads to the use of maximum likelihood to estimate the concentrations.

Now we want to define a prior $\mathrm{p}(\mathbf{M})$ on the color vectors. Again, we do not have enough data to define a data-driven prior. However, since 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 [11,29,42], we choose the following prior on $\mathbf{M}$,

$$\mathrm{p}(\mathbf{M}) = \prod_{s=1}^{S} \mathrm{p}(\mathbf{m}_s) = \prod_{s=1}^{S} \mathcal{N}(\mathbf{m}_s \mid \mathbf{m}_s^{\mathrm{Rui}}, \gamma_s^{\mathrm{Rui}}\mathbf{I}), \tag{6}$$

where $\mathbf{M}^{\mathrm{Rui}} = \left[\mathbf{m}_1^{\mathrm{Rui}}, \dots \mathbf{m}_S^{\mathrm{Rui}}\right]$ is the reference matrix of Ruifrok [29]. The variances $\gamma_1^{\mathrm{Rui}}, \dots, \gamma_S^{\mathrm{Rui}} > 0$ control the amount of variation allowed in each stain. We explore the importance of these values in the ablation study in Section 5.

Finally, assuming that all the components of $\boldsymbol{\epsilon}$ in Eq. (2) are independent and identically distributed according to $\mathcal{N}(0, \lambda_n^2)$, we can write,

$$\mathrm{p}(\mathbf{Y} \mid \mathbf{M}, \mathbf{C}) \propto \exp\left(-\frac{1}{2\lambda_n^2}\left\|\mathbf{Y}^\top - \mathbf{M}\mathbf{C}\right\|_{\mathrm{F}}^2\right), \tag{7}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm and λ_n^2 is the noise variance of the observation model.

3.2. Inference

To estimate $\mathbf{C}$ and $\mathbf{M}$ for each observation $\mathbf{Y}$ we need to compute the posterior $\mathrm{p}(\mathbf{C}, \mathbf{M} \mid \mathbf{Y})$. Since it does not admit an analytical expression, we use variational inference and approximate it by $\mathrm{q}(\mathbf{C}, \mathbf{M} \mid \mathbf{Y})$. In contrast to previous works [11,23,25], where the variational approach was also used, here we proceed in a different -*amortized*- manner. To build the inference model $\mathrm{q}(\mathbf{C}, \mathbf{M} \mid \mathbf{Y}) = \mathrm{q}(\mathbf{C} \mid \mathbf{Y})\mathrm{q}(\mathbf{M} \mid \mathbf{Y})$ we consider two DNNs. The first DNN, C-Net, is used to define $\mathrm{q}(\mathbf{C} \mid \mathbf{Y})$. It has parameters $\boldsymbol{\alpha}$, takes as input the OD image $\mathbf{Y}$, and outputs the concentration

estimates $\mathbf{C}^{\boldsymbol{\alpha}}(\mathbf{Y}) \in \mathbb{R}^{S\times HW}$. We decide to estimate $\mathbf{C}$ deterministically using a degenerate distribution,

$$\mathrm{q}^{\boldsymbol{\alpha}}(\mathbf{C} \mid \mathbf{Y}) = \begin{cases} 1 & \text{if } \mathbf{C} = \mathbf{C}^{\boldsymbol{\alpha}}(\mathbf{Y}), \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Under this distribution, $\mathbf{C}$ takes the value $\mathbf{C}^{\boldsymbol{\alpha}}(\mathbf{Y})$ with probability 1 and, therefore, its variance is zero.

The second network, M-Net, defines $\mathrm{q}(\mathbf{M} \mid \mathbf{Y})$. M-Net has parameters $\boldsymbol{\beta}$, takes as input the OD image $\mathbf{Y}$ and outputs the means $\boldsymbol{\mu}_1^{\boldsymbol{\beta}}(\mathbf{Y}), \ldots, \boldsymbol{\mu}_S^{\boldsymbol{\beta}}(\mathbf{Y}) \in \mathbb{R}^3$ and variances $\sigma_1^{\boldsymbol{\beta}}(\mathbf{Y})^2, \ldots, \sigma_S^{\boldsymbol{\beta}}(\mathbf{Y})^2 \in \mathbb{R}$ for the approximated posterior,

$$\mathrm{q}^{\boldsymbol{\beta}}(\mathbf{M} \mid \mathbf{Y}) = \prod_{s=1}^{S} \mathrm{q}^{\boldsymbol{\beta}}\left(\mathbf{m}_s\right) = \prod_{s=1}^{S} \mathcal{N}\left(\mathbf{m}_s \mid \boldsymbol{\mu}_s^{\boldsymbol{\beta}}(\mathbf{Y}), \sigma_s^{\boldsymbol{\beta}}(\mathbf{Y})^2 \mathbf{I}_{3\times 3}\right). \tag{9}$$

To estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ we will use a large dataset of OD histological images, denoted by $\mathcal{Y}$. Both networks, C-Net and M-Net, will be trained to maximize the following Evidence Lower Bound (ELBO) of the log-likelihood of the observations,

$$\mathrm{ELBO}(\mathcal{Y}) = \sum_{\mathbf{Y}\in\mathcal{Y}} \mathrm{ELBO}(\mathbf{Y}) \tag{10}$$

where

$$\begin{aligned} \mathrm{ELBO}(\mathbf{Y}) =& \mathbb{E}_{\mathrm{q}(\mathbf{C},\mathbf{M}\mid\mathbf{Y})}\left[\log \frac{\mathrm{p}(\mathbf{C},\mathbf{M},\mathbf{Y})}{\mathrm{q}(\mathbf{C},\mathbf{M}\mid\mathbf{Y})}\right] = \underbrace{-\mathbb{E}_{\mathrm{q}^{\boldsymbol{\alpha}}(\mathbf{C}\mid\mathbf{Y})}\left[\log \frac{\mathrm{q}^{\boldsymbol{\alpha}}(\mathbf{C}\mid\mathbf{Y})}{\mathrm{p}(\mathbf{C})}\right]}_{A_1} + \\ & \underbrace{-\mathbb{E}_{\mathrm{q}^{\boldsymbol{\beta}}(\mathbf{M}\mid\mathbf{Y})}\left[\log \frac{\mathrm{q}^{\boldsymbol{\beta}}(\mathbf{M}\mid\mathbf{Y})}{\mathrm{p}(\mathbf{M})}\right]}_{A_2} \\ & \underbrace{-\frac{1}{2\lambda_n^2}\mathbb{E}_{\mathrm{q}^{\boldsymbol{\beta}}(\mathbf{M}\mid\mathbf{Y})}\left[\left\|\mathbf{Y}^\top - \mathbf{M}\mathbf{C}^{\boldsymbol{\alpha}}(\mathbf{Y})\right\|_{\mathrm{F}}^2\right]}_{A_3} + \mathrm{const.} \end{aligned} \tag{11}$$

Since $\mathrm{p}(\mathbf{C})$ is improper and $\mathrm{q}^{\boldsymbol{\alpha}}(\mathbf{C} \mid \mathbf{Y})$ is degenerate the term A_1 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 $\mathrm{q}_{\boldsymbol{\beta}}(\mathbf{M})$ for the color matrix, this term would have not appeared.

The term A_2 corresponds to the negative of the Kullback-Leibler divergence between the two Gaussian distributions in Eq. (6) and Eq. (9),

$$\begin{aligned} A_2 =& \mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y}) = \frac{1}{2}\sum_{s=1}^{S} \frac{\left\|\boldsymbol{\mu}_s^{\boldsymbol{\beta}}(\mathbf{Y}) - \mathbf{m}_s^{\mathrm{Rui}}\right\|^2}{\gamma_s^{\mathrm{Rui}}} \\ & + \frac{3}{2}\sum_{s=1}^{S}\left(\frac{\sigma_s^{\boldsymbol{\beta}}(\mathbf{Y})^2}{\gamma_s^{\mathrm{Rui}}} - \log\frac{\sigma_s^{\boldsymbol{\beta}}(\mathbf{Y})^2}{\gamma_s^{\mathrm{Rui}}} - 1\right). \end{aligned} \tag{12}$$

The term A_3 admits a closed-form expression as follows,

$$A_3 = -\frac{1}{2\lambda_n^2}\left(\left\|\mathbf{Y}^\top - \boldsymbol{\mu}^{\boldsymbol{\beta}}(\mathbf{Y})\mathbf{C}^{\boldsymbol{\alpha}}(\mathbf{Y})\right\|_{\mathrm{F}}^2 + 3\sum_{k=1}^{HW}\sum_{s=1}^{S} c_s^{k\,\boldsymbol{\alpha}}(\mathbf{Y})^2 \sigma_s^{\boldsymbol{\beta}}(\mathbf{Y})^2\right), \tag{13}$$

where $\boldsymbol{\mu}^{\boldsymbol{\beta}}(\mathbf{Y}) = \left[\boldsymbol{\mu}_1^{\boldsymbol{\beta}}(\mathbf{Y}), \ldots, \boldsymbol{\mu}_S^{\boldsymbol{\beta}}(\mathbf{Y})\right]$. However, we have found the training procedure to be much more stable if we use the reparameterization trick instead [16]. Thus, we approximate $A_3 \approx -0.5\lambda_n^{-2}\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ where

$$\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y}) = \frac{1}{N_{\mathbf{M}}}\sum_{i=1}^{N_{\mathbf{M}}}\left[\left\|\mathbf{Y}^\top - \mathbf{M}_i^{\boldsymbol{\beta}}(\mathbf{Y})\mathbf{C}^{\boldsymbol{\alpha}}(\mathbf{Y})\right\|_{\mathrm{F}}^2\right], \tag{14}$$

and $\left\{\mathbf{M}_1^{\boldsymbol{\beta}}, \ldots, \mathbf{M}_{N_{\mathbf{M}}}^{\boldsymbol{\beta}}\right\}$ are samples drawn from $\mathrm{q}^{\boldsymbol{\beta}}(\mathbf{M} \mid \mathbf{Y})$. Therefore, the ELBO in Eq. (11) is approximated as $\mathrm{ELBO}(\mathbf{Y}) \approx -\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y}) - 0.5\lambda_n^{-2}\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y}) + \mathrm{const}$. Instead of maximizing the approximated lower bound, we equivalently minimize the negative approximated ELBO, which yields the following objective

$$\mathcal{L}(\mathcal{Y}) = \sum_{\mathbf{Y}\in\mathcal{Y}}\left[\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y}) + \frac{1}{2\lambda_n^2}\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})\right]. \tag{15}$$

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. As for the values of λ_n^2 and $\gamma_1^{\mathrm{Rui}}, \ldots, \gamma_S^{\mathrm{Rui}}$, in this work we choose not to automatically estimate them and study their effect in the loss function instead (see the ablation study in Section 5). For the noise variance, we assume it to be independent of the observed image, and the same for every image. For the color-vector variance, we assume it to be the same for every stain, i.e., $\gamma^{\mathrm{Rui}} = \gamma_1^{\mathrm{Rui}} = \cdots = \gamma_S^{\mathrm{Rui}}$. Thus being defined, λ_n^2 and γ^{Rui} determine the balance between the terms involved in the loss function. To experimentally determine how this balance affects the results once γ^{Rui} is fixed, we redefine the objective in Eq. (15) to include a weighting parameter $0 < \theta = 1/(1 + 0.5\lambda_n^{-2}) < 1$,

$$\mathcal{L}(\mathcal{Y}) = \sum_{\mathbf{Y}\in\mathcal{Y}}\left[\theta\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y}) + (1-\theta)\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})\right], \tag{16}$$

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 (depicted in Fig. 2) 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 Eq. (7). 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 independence assumption in the posterior approximation, is a common approach in blind image deblurring [22,40] and denoising [38].

4. Network architecture

The proposed network is depicted in detail in Fig. 2. For the design of C-Net, we follow [40] and rely on the commonly used Unet [28] (proposed for biological image segmentation) to estimate the stain concentrations. The output of C-Net has two layers, one for each stain in the H&E image, and the same size in pixels as the input image. We use four scales in both encoder and decoder, where each encoder and decoder block includes three stacked ResBlocks with LeakyReLU activation and a small kernel size of 3×3. The number of channels per layer is set to 64. Each downsampling block uses a convolution layer with a 3×3 filter and a stride of 2. The upsampling uses a transposed convolution with a 5×5 filter and a stride of 2. This architecture is expected to capture multi-scale features and to accurately estimate the concentration of the stains in the image. The detailed structures of each block are depicted in Fig. 3.

For the estimation of the color-vector matrix posterior, M-Net uses the bottleneck MobileNet V3 Small [12] followed by a linear fully connected layer. MobileNet V3 reduces the number of parameters with techniques such as squeeze-and-excite, reducing the number of initial filters, or by using bottleneck layers, which considerably reduces the computational cost of the Resnet used in [40]. This choice of network is motivated by the reduced size of the color-vector matrix, which is often smaller than a blur kernel. 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 re-

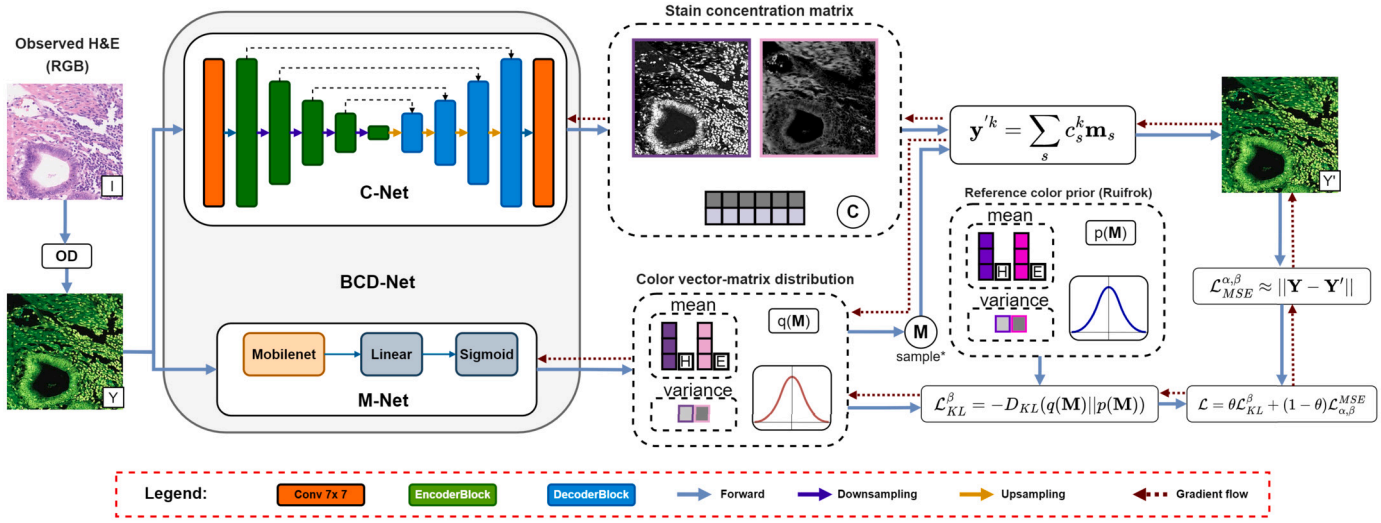


Fig. 2. Detailed architecture for BCD-Net including the subnetworks C-Net and M-Net, the overview of the Bayesian framework and loss. C-Net implements a Unet architecture and M-Net the bottleneck MobileNet V3 Small. The loss function includes two terms. $\mathcal{L}_{\text{MSE}}^{\alpha,\beta}(\mathbf{Y})$ combines both C-Net and M-Net to achieve a reconstruction of the input image. $\mathcal{L}_{\text{KL}}^{\beta}(\mathbf{Y})$ monitors M-Net to provide a color distribution close to the prior. See section 3 for inference details.

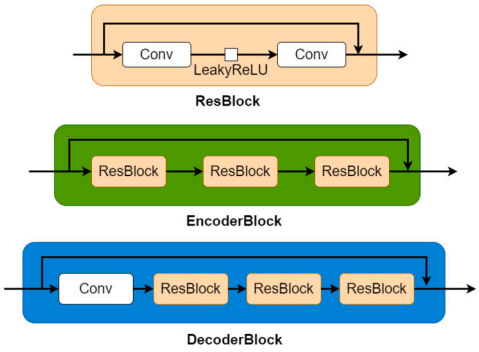


Fig. 3. Detailed structures for each block in C-Net.

quired to have a unitary norm, we include an L2 constraint [27] for its corresponding output.

5. Experiments and training details

In this section, we carry out an empirical validation of our proposed BCD-Net. To this end, we first describe the public datasets used in this work and provide detailed information about the implementation of BCD-Net. Then, we perform an empirical study to select the hyperparameters of our model: the standard deviation γ_s of the color-vector matrix prior in Eq. (6) and the θ parameter of the loss function of Eq. (16). The effect of the parameters is investigated through an ablation study that takes into account the reconstructed images and the stain separation quality. The results are compared with the current state-of-the-art methods for stain separation in terms of performance and computational efficiency. Finally, we consider a cancer classification problem to show how deconvolved images can be used to improve the performance of a CNN-based CAD system.

5.1. Datasets

5.1.1. Camelyon-17

This database was created for the Camelyon-17 breast cancer classification challenge [6] in collaboration with 5 different medical centers in the Netherlands. The database includes intra- and inter-center color variations. An example is depicted in Fig. 4. Following [24,42], we use the 500 slides (100 from each center) that were released as the training set for the challenge, from which we extracted 500 non-overlapping patches of size 224×224 from each slide, only patches with at least 70% tissue were considered. This database has been labeled for cancer classification but does not contain any stain separation ground truth. Our model is trained using the images from the centers identified as 0, 2, and 4 in [6] and validated using the images from the remaining two centers, allowing us to capture a wide range of variations and study whether BCD-Net is able to generalize to multiple centers.

5.1.2. Warwick Stain Separation Benchmark (WSSB)

This dataset was collected by Alsubaie et al. [3] and considers the color variation between tissue types and laboratories. It is of particular interest for evaluating the performance of BCD methods because it contains the ground truth stain matrix for the 24 images of three different tissue types (breast, lung, and colon) that comprise the dataset. The stain color matrix was obtained by asking pathologists to identify pixels of biological structures that were stained with a single stain. That is, nuclei pixels for hematoxylin and cytoplasm for eosin. The manually selected pixels were used to calculate the median color-vector for each stain. Then the ground truth concentrations $\mathbf{C}_{GT}$ were obtained as

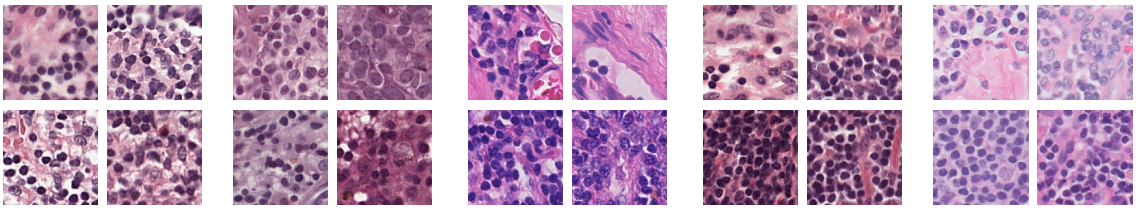
$$\mathbf{C}_{GT} = \mathbf{M}_{GT}^{+}\mathbf{Y}, \tag{17}$$

where $\mathbf{M}_{GT}^{+}$ is the Moore-Penrose pseudo-inverse of the ground truth color-matrix. From these ground-truth concentrations and color vectors, a separate RGB image is obtained for each stain. WSSB is used only for testing the proposed BCD-Net in the stain separation task.

5.2. Implementation details

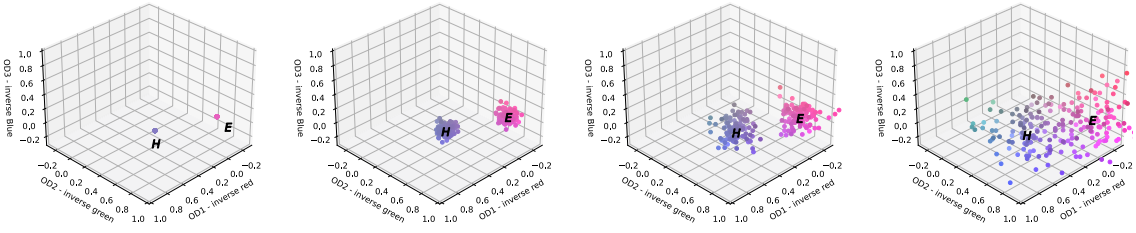
BCD-Net is built using Pytorch.[2] The experiments are performed on four NVIDIA GeForce RTX 3090. We used patches of size 224×224 and the batch size was set to 64 (16 per GPU). We utilized the ADAM optimizer with an initial learning rate of 10^{-4}, which is halved every 3 epochs if the loss does not decrease. We train BCD-Net for a total of 100 epochs, but we include an early stopping callback to halt the training procedure if the loss has not decreased for 10 consecutive epochs. The results reported in the following experiments correspond to the best-performing epoch on the validation data.

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



(a) Center 0. (b) Center 1. (c) Center 2. (d) Center 3. (e) Center 4.

Fig. 4. 224×224 patches extracted from different images within the five different centers of Camelyon-17 to illustrate color variations between the centers.



(a) $\gamma_{\mathrm{H}}^{\mathrm{Rui}} = \gamma_{\mathrm{E}}^{\mathrm{Rui}} = 0.001$ (b) $\gamma_{\mathrm{H}}^{\mathrm{Rui}} = \gamma_{\mathrm{E}}^{\mathrm{Rui}} = 0.05$ (c) $\gamma_{\mathrm{H}}^{\mathrm{Rui}} = \gamma_{\mathrm{E}}^{\mathrm{Rui}} = 0.1$ (d) $\gamma_{\mathrm{H}}^{\mathrm{Rui}} = \gamma_{\mathrm{E}}^{\mathrm{Rui}} = 0.2$

Fig. 5. Graphical representation with 100 samples of the prior p(**M**) in Eq. (6) for different values of $\gamma_{\mathrm{H}}^{\mathrm{Rui}}$ and $\gamma_{\mathrm{E}}^{\mathrm{Rui}}$. Each sample is represented in the OD space according to the coordinates of each channel. Each sample is colored with the corresponding RGB value.

While each $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$ term in Eq. (16) is sensitive to the channel order due to the prior means for both stains, where hematoxylin and eosin should correspond to the first and second channels, respectively, the $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ term in the same equation is not channel-sensitive as long as the order of the channels of the C-Net and M-Net is the same. Obviously, the order of the stains is not important for the reconstruction of the image. A pre-training epoch is added to help the network determine the correct order by setting the weighting hyperparameter θ to 0.99. This forces the network to initially focus on optimizing $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$ instead of $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$. After the pretraining epoch, θ is fixed to the values specified in the ablation study in Section 5.3.

5.3. Ablation study

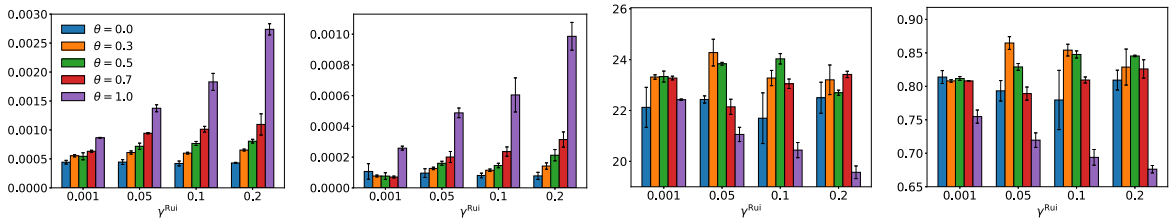
The behavior of the proposed BCD-Net depends on the hyperparameters γ^{Rui} and θ, which controls our confidence in the reference color-vector matrix in Eq. (6) and the balance between the prior and reconstruction terms in the loss function of Eq. (16), respectively. Notice that this is equivalent to balancing the weight that both subnetworks have on the training of BCD-Net, completely removing the terms $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$ or $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ when $\theta = 0$ or $\theta = 1$, respectively. In this section, we study the joint effect of these hyperparameters on two tasks: (i) image reconstruction and (ii) stain separation.

In the first task, we analyze the fidelity to the observed optical density image, which indicates whether BCD-Net preserves the information in the image globally. For this purpose, we will use the Mean Square Error (MSE) between the observed and the reconstructed optical density image. The reconstruction MSE can be computed for Camelyon17 and WSSB, described in Subsections 5.1.1 and 5.1.2 respectively. The second task, stain separation, will assess whether the outputs of the C-Net and M-Net subnetworks correctly represent the structure and appearance of the tissue for each stain. We will use two well-known metrics: Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [3,24]. Note that stain-separated PSNR and SSIM values can only be calculated for the WSSB dataset. Before we discuss how γ_s^{Rui} and θ affect these tasks, let us justify what values we are going to assign to these hyperparameters.

Choosing values for γ^{Rui}. As previously mentioned, the Ruifrok matrix [24,29,33,42] is set as the mean of the color-vector prior that represents the H&E staining. For the prior variance, we consider the values $\gamma^{\mathrm{Rui}} = \gamma_{\mathrm{H}}^{\mathrm{Rui}} = \gamma_{\mathrm{E}}^{\mathrm{Rui}} \in \{0.001, 0.05, 0.1, 0.2\}$ (the same variance for each stain). This amounts to assuming the same diagonal covariance matrix $\gamma^{\mathrm{Rui}}\mathbf{I}$ for all stains. To illustrate the effect of these choices for γ^{Rui}, Fig. 5 depicts random samples from the corresponding prior distributions. Small values of γ^{Rui} result in a priori low uncertainty on the prior mean values prior, and therefore all samples are close to the Ruifrok matrix. Large values increase the uncertainty and produce more variable stain pairs, including unrealistic stain colors or the possibility of confusing the hematoxylin and eosin channels. An optimal value of γ^{Rui} should produce enough variation while keeping the H&E stains well separated. Also, note that small values of γ^{Rui} give more importance to those terms in $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$ (see Eq. (12)). Therefore, a lower value of θ might be necessary to balance the loss function.

Choosing values for θ. If we consider γ^{Rui} fixed, this hyperparameter determines the importance of $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$ and $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ in the loss function. Note that choosing a value for θ amounts to choosing a value for λ_n^2, but it also has the effect of giving more weight to one term or another in the loss function. On the one hand, small values of θ increase the relevance of the fidelity term $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$. Setting $\theta = 0.0$ removes the term $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$ and trains the whole network only with $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$. On the other hand, large values of θ balance the loss towards similarity to the prior p(**M**). Setting $\theta = 1.0$ removes the term $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ and, therefore, C-Net does not participate in the training procedure. Thus, the ultimate goal of BCD-Net becomes to output p(**M**). To study how this balance affects the results, we consider the values $\theta \in \{0.0, 0.3, 0.5, 0.7, 1.0\}$, which roughly corresponds to $\lambda_n^2 \in \{0.0, 0.0555, 0.2142, 0.5, 1.1666, \infty\}$. As we approach $\theta = 1.0$, the noise variance increases towards the infinite and the information from the prior becomes more relevant.

The results for each combination of hyperparameters can be found in Table 2 and are summarized in Fig. 6. Figs. 6a and 6b show that large values of θ tend to produce worse reconstructions with higher variability, especially when γ^{Rui} is large. This is what we expected, since as θ increases the importance shifts from $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ to $\mathcal{L}_{\mathrm{KL}}^{\boldsymbol{\beta}}(\mathbf{Y})$. Interestingly, small values of γ^{Rui} produce high-quality reconstructions even when $\theta = 1.0$ (in this case the term $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ is only used in the pretraining epoch). Small values of γ^{Rui} translate into the posterior distribution q(**M**) having lower variance (see Eq. (12)). Since $\mathcal{L}_{\mathrm{MSE}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{Y})$ is computed by sampling from q(**M**), the generated samples are close and produce



(a) Rec. MSE (Camelyon-17 Validation). (b) Rec. MSE (WSSB). (c) PSNR (WSSB). (d) SSIM (WSSB).

Fig. 6. Reconstruction MSE, PSNR, and SSIM for each of the considered values of γ^{Rui} and θ.

Table 2
Reconstruction MSE, PSNR, and SSIM for each combination of the considered values of γ^{Rui} and θ. MSE has been calculated for Camelyon-17 (Train and Validation sets) and WSSB observed images. PSNR and SSIM are calculated on WSSB using the stain-separated ground truth.

γ^{Rui}	θ	λ_n^2	Rec. MSE $\times 10^2$ (Camelyon-17 Train)	Rec. MSE $\times 10^2$ (Camelyon-17 Valid.)	Rec. MSE $\times 10^2$ (WSSB)	PSNR (WSSB)	SSIM (WSSB)
	0.0	0.0	0.0127 ± 0.0002	0.0443 ± 0.0032	0.0107 ± 0.0051	22.12 ± 0.79	0.8139 ± 0.0096
	0.3	0.2142	0.0145 ± 0.0006	0.0554 ± 0.0019	0.0077 ± 0.0006	23.32 ± 0.09	0.8081 ± 0.0019
0.001	0.5	0.5	0.0155 ± 0.0002	0.0545 ± 0.0061	0.0077 ± 0.0021	23.34 ± 0.21	0.8117 ± 0.0029
	0.7	1.1666	0.0166 ± 0.0004	0.0632 ± 0.0017	**0.0070** ± 0.0006	23.28 ± 0.07	0.8081 ± 0.0001
	1.0	∞	0.0221 ± 0.0002	0.0864 ± 0.0007	0.0258 ± 0.0013	22.42 ± 0.03	0.7547 ± 0.0098
	0.0	0.0	0.0120 ± 0.0008	0.0445 ± 0.0042	0.0096 ± 0.0028	22.43 ± 0.14	0.7933 ± 0.0153
	0.3	0.2142	0.0169 ± 0.0005	0.0610 ± 0.0027	0.0126 ± 0.0008	**24.28** ± 0.53	**0.8647** ± 0.0095
0.05	0.5	0.5	0.0218 ± 0.0019	0.0718 ± 0.0054	0.0160 ± 0.0013	23.84 ± 0.06	0.8290 ± 0.0051
	0.7	1.1666	0.0304 ± 0.0026	0.0943 ± 0.0010	0.0201 ± 0.0036	22.15 ± 0.30	0.7891 ± 0.0098
	1.0	∞	0.0536 ± 0.0047	0.1375 ± 0.0060	0.0488 ± 0.0031	21.06 ± 0.28	0.7196 ± 0.0109
	0.0	0.0	0.0119 ± 0.0007	**0.0421** ± 0.0042	0.0081 ± 0.0015	21.70 ± 1.00	0.7796 ± 0.0443
	0.3	0.2142	0.0171 ± 0.0008	0.0599 ± 0.0015	0.0115 ± 0.0008	23.28 ± 0.30	0.8541 ± 0.0088
0.1	0.5	0.5	0.0217 ± 0.0013	0.0766 ± 0.0035	0.0145 ± 0.0014	24.03 ± 0.20	0.8476 ± 0.0053
	0.7	1.1666	0.0318 ± 0.0007	0.1012 ± 0.0050	0.0236 ± 0.0030	23.05 ± 0.19	0.8095 ± 0.0045
	1.0	∞	0.0810 ± 0.0107	0.1830 ± 0.0147	0.0605 ± 0.0111	20.44 ± 0.30	0.6939 ± 0.0118
	0.0	0.0	**0.0117** ± 0.0011	0.0430 ± 0.0010	0.0078 ± 0.0023	22.50 ± 0.61	0.8093 ± 0.0148
	0.3	0.2142	0.0190 ± 0.0004	0.0653 ± 0.0016	0.0142 ± 0.0021	23.21 ± 0.58	0.8287 ± 0.0269
0.2	0.5	0.5	0.0241 ± 0.0015	0.0805 ± 0.0033	0.0212 ± 0.0037	22.70 ± 0.10	0.8453 ± 0.0011
	0.7	1.1666	0.0367 ± 0.0081	0.1095 ± 0.0183	0.0314 ± 0.0050	23.42 ± 0.12	0.8260 ± 0.0137
	1.0	∞	0.1346 ± 0.0057	0.2738 ± 0.0097	0.0986 ± 0.0090	19.57 ± 0.25	0.6762 ± 0.0055

similar reconstructions, making it easier for C-Net to estimate the right concentrations.

The best reconstruction performance in Camelyon-17 (in both train and validation subsets) is obtained with $\theta = 0.0$, which removes $\mathcal{L}^{\beta}_{\text{KL}}(\mathbf{Y})$ and forces BCD-Net to focus on the reconstruction term. However, better reconstructions in the validation subset are obtained with a lower value of γ^{Rui}, which gives more importance to the term $\mathcal{L}^{\beta}_{\text{KL}}(\mathbf{Y})$ in the pretraining epoch (where $\theta = 0.99$), see Table 2. Furthermore, in WSSB, which is used exclusively as a test set, the best reconstructions are obtained when $\theta = 0.7$ and $\gamma^{\text{Rui}} = 0.001$, which shifts the importance to $\mathcal{L}^{\beta}_{\text{KL}}(\mathbf{Y})$. As expected, this term acts as a regularizer that prevents BCD-Net from overfitting to the training set, while θ and γ^{Rui} jointly balance the importance of the terms in the loss function.

We now turn our attention to the stain separation task (Figs. 6c and 6d). Not surprisingly, setting $\theta = 1.0$ produces the worst results since BCD-Net is trained using only $\mathcal{L}^{\beta}_{\text{KL}}(\mathbf{Y})$. Although training the model using only $\mathcal{L}^{\alpha,\beta}_{\text{MSE}}(\mathbf{Y})$ ($\theta = 0.0$) gives acceptable results, the highest PSNR and SSIM values are obtained with intermediate values of θ (this is true no matter what value we choose for γ^{Rui}). This shows that both terms are necessary to obtain a proper stain separation. The best results are obtained when $\gamma^{\text{Rui}} \in \{0.05, 0.1\}$, which produces enough variation while keeping the H&E samples well separated (see Fig. 5). The best performance is obtained with $\gamma^{\text{Rui}} = 0.05$ and $\theta = 0.3$. These values of the hyperparameters also show an acceptable reconstruction in terms of MSE, and the same color variance value ($\gamma^{\text{Rui}} = 0.05$) has been shown to provide realistic H&E variations in [25].

In conclusion, our results indicate that the use of the color-vector matrix prior improves the generalization of our model to unseen images as well as the quality of the stain separation. The right balance between the reconstruction term and the proximity to the color-vector matrix prior is of paramount importance. Giving too much weight to the former leads to a model where the separation may not be related to H&E, while not giving enough weight to it leads to overfitting to the color-vector matrix prior and poor structural reconstruction. Since no stain separation ground truth is used during the training of BCD-Net, both terms are necessary to obtain a model that leads to realistic stain separations and is able to handle color variation correctly. Finally, we would like to emphasize that, although the best hyperparameters might change when using other datasets, we believe that the need to promote the interaction between C-Net and M-Net through intermediate values of θ will remain regardless of the dataset chosen to train and validate the network.

According to the results obtained, we will set $\gamma^{\text{Rui}} = 0.05$ and $\theta = 0.3$ in the following sections. These values involve both $\mathcal{L}^{\beta}_{\text{KL}}(\mathbf{Y})$ and $\mathcal{L}^{\alpha,\beta}_{\text{MSE}}(\mathbf{Y})$ in the loss function with more weight given to the latter. They have been shown to properly reconstruct the images and achieve the best stain separation performance.

5.4. Comparison with state-of-the-art methods

In the previous section, we have studied the role γ^{Rui} and θ play and how they affect the reconstruction and stain separation tasks. In

Table 3
MSE·10^1 between the observed optical density and reconstructed images for the different methods on the WSSB dataset [3]. The best value of each subset is highlighted in bold.

Subset	Non-blind	Non-amortized								Amortized
	RUI	MAC	VAH	ALS	SAR	TV	SGP	BKSVD	ZHE	BCD-Net
Lung	0.2033	0.1746	0.1691	0.2427	0.1676	0.1719	0.1703	0.0036	0.1361	**0.0003** ± 0.0000
Breast	0.5816	0.3539	0.3594	0.2667	0.3586	0.3630	0.3607	0.0025	0.6833	**0.0014** ± 0.0006
Colon	0.2559	0.2311	0.2287	0.4778	0.2277	0.2252	0.2235	0.0029	2.3460	**0.0010** ± 0.0003
Mean	0.3469	0.2532	0.2524	0.3291	0.2513	0.2534	0.2515	0.0030	1.0551	**0.0009** ± 0.0002

Table 4
PSNR for the different methods on the WSSB dataset [3]. The best value of each pair (subset, stain) is highlighted in bold.

Subset	Stain	Non-blind	Non-amortized							Amortized
		RUI	MAC	VAH	ALS	SAR	TV	SGP	BKSVD	ZHE	BCD-Net
Lung	H	22.47	19.52	25.87	20.62	32.91	33.10	**35.21**	32.67	19.51	27.22 ± 0.50
	E	22.05	18.09	25.53	23.95	30.77	31.02	**33.07**	30.61	16.23	24.76 ± 0.19
Breast	H	15.27	26.24	25.46	24.60	28.81	29.14	30.50	**32.20**	15.31	24.35 ± 1.20
	E	17.66	23.62	27.68	25.92	26.60	26.76	27.71	**29.43**	14.99	22.18 ± 1.12
Colon	H	22.27	23.91	25.83	21.11	28.57	28.62	29.01	**34.08**	17.89	24.71 ± 0.28
	E	20.70	21.55	26.29	21.94	27.58	27.60	28.38	**33.32**	14.76	22.44 ± 0.10
Mean	H	20.00	23.22	25.72	22.11	30.10	30.29	31.57	**32.98**	17.57	25.43 ± 0.59
	E	20.14	21.08	26.50	23.94	28.32	28.46	29.72	**31.12**	15.33	23.13 ± 0.46
	Mean	20.07	22.15	26.11	23.03	29.21	29.38	30.65	**32.05**	15.45	24.28 ± 0.53

Table 5
SSIM for the different methods on the WSSB dataset [3]. The best value of each pair (subset, stain) is highlighted in bold.

Subset	Stain	Non-blind	Non-amortized								Amortized
		RUI	MAC	VAH	ALS	HID	TV	SGP	BKSVD	ZHE	BCD-Net
Lung	H	0.7987	0.7389	0.8912	0.5551	0.9763	0.9757	**0.9898**	0.9764	0.8116	0.9517 ± 0.0028
	E	0.7734	0.5088	0.8195	0.8939	0.9306	0.9353	**0.9654**	0.9461	0.5390	0.8751 ± 0.0156
Colon	H	0.8141	0.8095	0.8851	0.7241	0.9542	0.9544	0.9638	**0.9826**	0.7894	0.9251 ± 0.0131
	E	0.7456	0.6365	0.8904	0.8540	0.9139	0.9161	0.9414	**0.9646**	0.4625	0.7924 ± 0.0255
Breast	H	0.6215	0.9552	0.9239	0.8068	0.9528	0.9560	0.9751	**0.9801**	0.6488	0.9359 ± 0.0024
	E	0.7644	0.9336	0.9550	0.9380	0.9464	0.9492	0.9645	**0.9632**	0.7150	0.7080 ± 0.0147
Mean	H	0.7448	0.8345	0.9100	0.6953	0.9611	0.9621	0.9762	**0.9797**	0.7500	0.9376 ± 0.0058
	E	0.7611	0.6930	0.8883	0.8953	0.9303	0.9336	0.9571	**0.9580**	0.5722	0.7919 ± 0.0148
	Mean	0.7530	0.7638	0.8992	0.7953	0.9457	0.9479	0.9667	**0.9684**	0.6611	0.8647 ± 0.0095

this section, we fix them to the values that obtain the best performance and compare BCD-Net to the following approaches: the classical non-blind color deconvolution method by Ruifrok et al. [29], the non-amortized methods by Macenko et al. [19], Vahadane et al. [35], Alsubaie et al. [3], and Zheng et al. [42]. They will be denoted as RUI, MAC, VAH, ALS, and ZHE, respectively. We also include the Bayesian methods using Simultaneous Auto-regressive (SAR) [11], Total Variation (TV) [23], Super Gaussian Priors (SGP) [25], and Bayesian K-SVD (BKSVD) [24]. Notice that all these non-amortized methods require a function to be optimized for each image independently. The comparison is carried out on the WSSB dataset using MSE, PSNR, and SSIM as in the previous section.

First, we compare how accurately the considered methods reconstruct the observed image. The results are shown in Table 3. The most accurate reconstruction is obtained by the newly proposed BCD-Net, although the other methods are optimized for each image. In particular, the proposed method outperforms RUI, from which we borrow the mean for the color-vector matrix prior. Recall that BCD-Net is trained to obtain a good reconstruction via the term $\mathcal{L}^{\alpha,\beta}_{\text{MSE}}(\mathbf{Y})$ in Eq. (16).

The comparison of the stain separation quality is presented in Tables 4 and 5. The proposed method is still far from the most recent non-amortized methods in terms of PSNR and SSIM. This is unsurprising, as non-amortized methods recalculate all model parameters per image, usually achieving superior performance. In addition, SAR, TV, SGP, and BKSVD methods also implement a more complex prior on the concentrations p($\mathbf{C}$), which has been proven to have a positive impact on the results [24]. However, BCD-Net results are promising: despite being trained in an amortized way and without ground truth data, BCD-Net outperforms RUI, ZHE, MAC, and ALS, and it is competitive with the commonly used method by VAH. The figures of merits show that the amortized BCD-Net model produces an appropriate estimation of the concentration and color-vector matrices. We also notice that non-amortized methods also achieve higher fidelity to the hematoxylin than to the eosin channel, which was already mentioned in Section 5.3 for BCD-Net. Only the methods by VAH (PSNR) and ALS (both metrics) obtain higher figures of merits in the eosin.

A qualitative comparison is shown in Fig. 7 using a lung image from the WSSB dataset. Most of the non-amortized methods are able to correctly identify the stain appearance (color-vector matrix) in the image. The methods by MAC and ALS do not estimate correctly the hematoxylin color. The proposed BCD-Net obtains a mean color-vector matrix that is not exactly the ground truth but clearly differentiates the stains in the image. The differences in the structure (concentrations) of each stain are more difficult to notice, Fig. 8 presents a zoomed-in version that makes it easier to appreciate. At the ground truth, the eosin shows gaps in the tissue corresponding to pixels completely stained with hema-

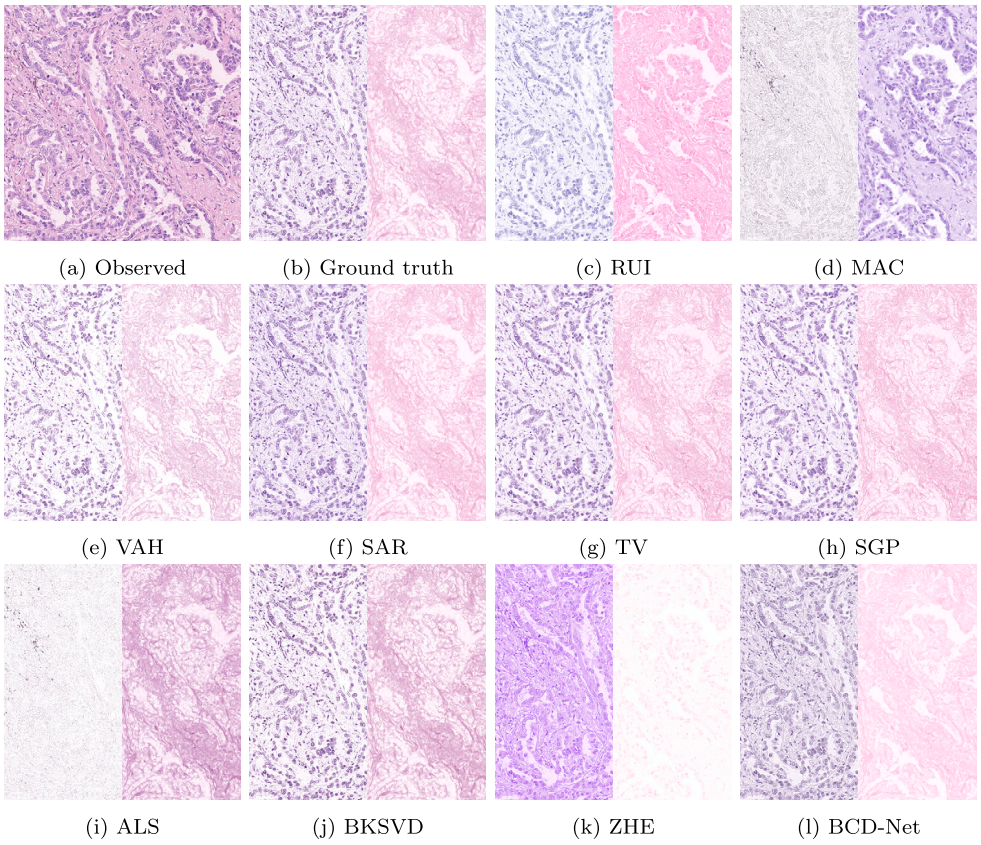


Fig. 7. (a) Depicts a single colon observed H&E image from WSSB, (b) shows the corresponding ground truth single stain H-only and E-only images, and (c)-(l) show the separation obtained by the competing methods. Hematoxylin and eosin separations are presented on the left and right-hand sides of each image, respectively.

toxylin. RUI and MAC do not capture this separation. ZHE and BKSVD move too much information from the eosin to the hematoxylin channels, creating bigger gaps. VAH discards some information from both channels. BCD-Net improves the estimation by RUI, especially on the hematoxylin channel, which is richer than the competitors. BCD-Net's eosin channel captures the connective tissue better than MAC RUI, and ZHE but lacks some details and shows a blurring effect similar to SAR, which explains the low PSNR and SSIM values obtained for this channel.

5.5. Computational efficiency

We analyze the time required by each method to perform the stain separation of a 2000×2000 image in the WSSB dataset. The methods in the comparison were evaluated using CPU, as they are not implemented to run in GPU. The proposed BCD-Net was evaluated using an NVIDIA TITAN X GPU and the same CPU as the rest of the methods. The results are presented in Fig. 9. The biplots present the average required time vs. the mean PSNR and SSIM metrics from Tables 4 and 5. These figures allow us to visually compare the time efficiency and the stain separation quality of each model. The best methods are those close to the left-top corner in both cases. When using the GPU, the proposed method is the fastest, requiring an average of 0.74 seconds. Notice that this makes our method even faster than the non-blind method RUI, which requires 0.81 seconds. BCD-Net is also 9.01 times faster than VAH (6.76 s) while being competitive in terms of PSNR and SSIM. If the GPU is not available, the proposed BCD-Net can also run on CPU, where it requires 9.24 seconds, being competitive in time against the state-of-the-art non-amortized methods, and being still faster than SAR, ALS, SGP, and TV.

The superior performance in terms of computational cost is a noticeable advantage of BCD-Net as an amortized method. The excessive computational cost of other methods renders them impractical for real-world applications. BCD-Net's efficiency during test time makes it a compelling tool for seamless integration with other DL approaches for classification, eliminating the need for database preprocessing beforehand.

5.6. Classification performance

Finally, we evaluate the performance of BCD-Net on a breast cancer classification task. We use the best-performing model from the previous sections. Following the approach described in [24], we use a VGG19 classifier as a benchmark, which is trained using four centers from the Camelyon-17 database. The fifth center, which was found to have larger color differences [24], is utilized for testing. VGG19 is trained and tested with the original images and the OD concentrations obtained by the compared methods. We also include the BCD-based augmentation method proposed by Tellez et al. [33], which is denoted by TEL.

Class-balanced patches were sampled from the annotated and negative WSIs in Camelyon-17, yielding approximately 110,000 patches for training and 55,000 for testing. VGG19 is trained for 100 epochs using 64 sample batches with batch normalization and a learning rate of 0.001, which is halved every 30 epochs. The area under the ROC curve (AUC), shown in Table 6, is calculated on the test set for the best-performing epoch during training for each method.

The proposed BCD-Net achieves an AUC of 0.9652, which outperforms the use of the original images and RUI, MAC, VAH, SAR, TV, and SGP. BCD-Net results are also competitive with the augmentation method in TEL. Although the best classification performance is obtained by the recent non-amortized method BKSVD, these results show that the proposed method is suitable for its use in classification tasks.

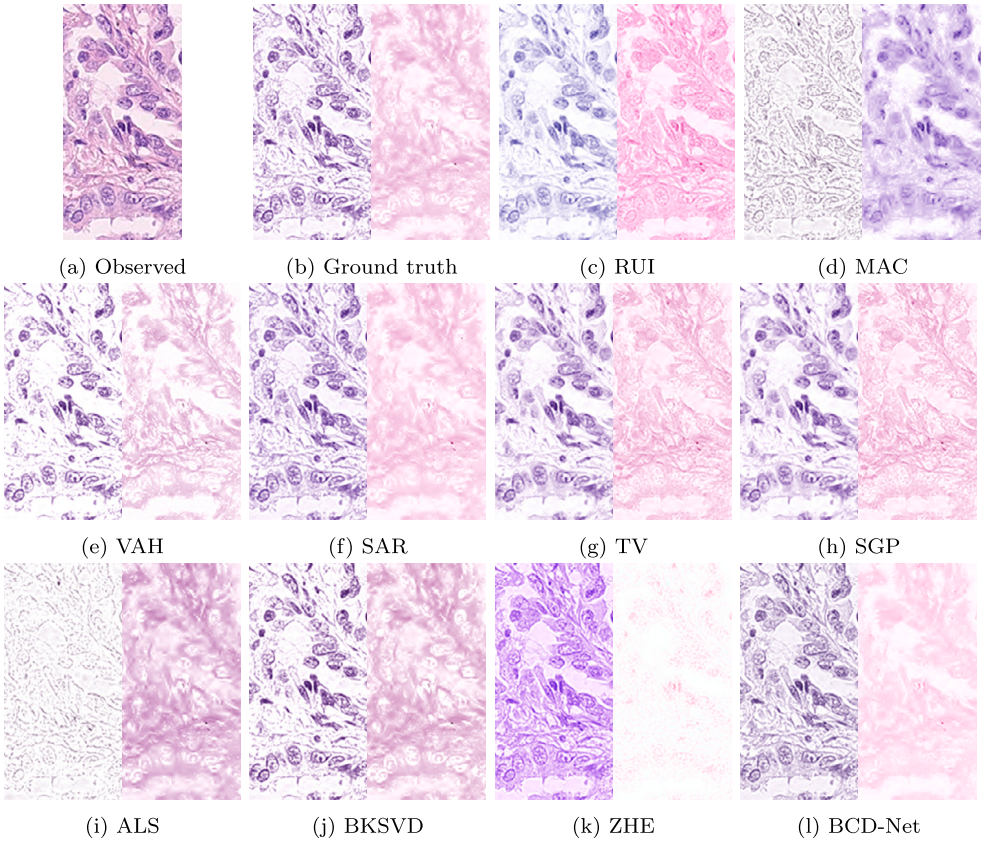
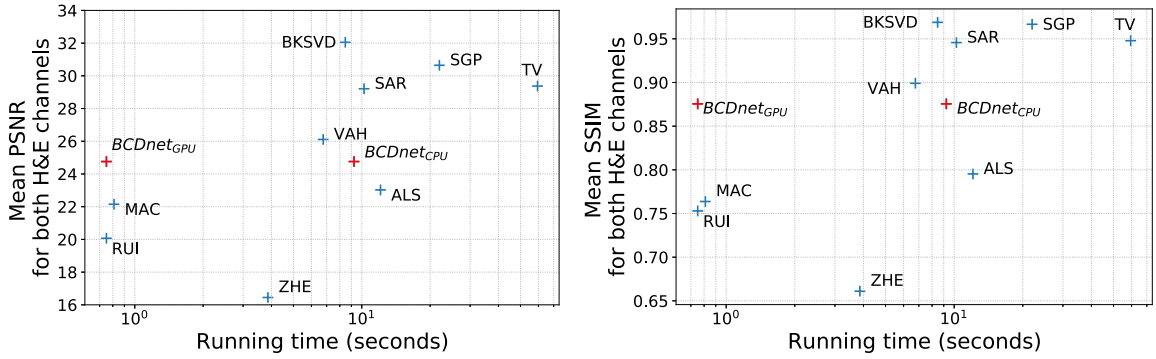


Fig. 8. Zoomed-in sample from the left-bottom corner of the images in Fig. 7.(a) Depicts the observed H&E image from WSSB, (b) shows the corresponding ground truth single stain H-only and E-only images, and (c)-(l) show the separation obtained by the BCD methods. Hematoxylin and eosin separations are presented on the left and right-hand sides of each image, respectively.



(a) PSNR. (b) SSIM.

Fig. 9. Mean quality metrics vs running time for deconvolving a 2000×2000 image. The proposed BCD-Net is marked in red.

Table 6
AUC performance of the VGG19 classifier for the proposed and competing methods using the OD concentrations of the H&E channels. Bold values indicate the highest performance.

Original (RGB)	RUI	MAC	VAH	ALS	SAR	TV	SGP	ZHE	BKSVD	TEL	BCD-Net
0.9491	0.9417	0.9468	0.6614	0.9725	0.9642	0.9508	0.9650	0.9864	**0.9879**	0.9654	0.9652

6. Conclusions and future work

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 CNN by maximizing the Evidence Lower Bound of the observed optical density images and without a stain separation ground truth. The proposed model includes two subnetworks, C-Net and M-Net, that jointly estimate a posterior distribution for the color-vector matrix and the stain concentrations using maximum likelihood in an amortized inference manner.

As other amortized methods for BCD have not been developed yet, the proposed method has been compared against nine non-amortized methods in the literature. While non-amortized methods typically achieve superior performance, BCD-Net shows promising results. It is competitive with the most used method in the literature for stain separation of different tissue types and breast cancer classification. Moreover, the amortized training of BCD-Net significantly improves computational efficiency compared to non-amortized models. Being trained without stain separation ground truth, this work mitigates the need for large labeled databases for DL BCD approaches and opens new research lines to interpretable processing of histological images using DL. Contrary to analytical non-amortized methods, the proposed BCD-Net can be combined with DL classification approaches in end-to-end training [5,8], which will likely improve both stain separation and classification performance.

The proposed model uses the maximum likelihood principles to estimate the C-Net parameters, which keeps the model simpler but provides results that are still outperformed by non-amortized models. Adapting to our framework prior concentration models that correspond to smoothness constraints, as were utilized in the analytical methods in [11,23,25], as well as the fully automatic estimation of the model parameters, will be investigated in the future. This will, very likely, improve the performance of the proposed method in comparison with these amortized methods. Finally, the end-to-end training of amortized BCD and classification techniques will also be investigated.

CRediT authorship contribution statement

Shuowen Yang: Investigation, Software. **Fernando Pérez-Bueno:** Formal analysis, Supervision, Writing – original draft. **Francisco M. Castro-Macías:** Methodology, Software, Writing – review & editing. **Rafael Molina:** Funding acquisition, Project administration, Supervision. **Aggelos K. Katsaggelos:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data belong to publicly available datasets. Details included in the manuscript.

Acknowledgments

This work was supported by projects PID2022-140189OB-C22 funded by MCIN / AEI / 10.13039 / 501100011033 and B-TIC-324-UGR20 funded by FEDER/Junta de Andalucía and Universidad de Granada. The work by Francisco M. Castro-Macías was sponsored by Ministerio de Universidades under FPU contract FPU21/01874. The authors would like to thank CITIC-UGR for their support with the proofreading of this paper.

Appendix A. Acronyms

Table A.7
Abbreviations and definitions used in this paper.

Abbreviation	Definition
ALS	Method by Alsubaie et al. [3]
BCD	Blind Color Deconvolution
BKSVD	Bayesian K-Singular Value Decomposition [24]
BID	Blind Image Deblurring
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
ELBO	Evidence Lower Bound
H&E	Hematoxylin and Eosin
MAC	Method by Macenko et al. [19]
MSE	Mean Square Error
OD	Optical Density
PSNR	Peak Signal to Noise Ratio
RUI	Method by Ruifrok et al. [29]
SAR	Simultaneous Auto-regressive [11]
SGP	Super Gaussian Priors [25]
SSIM	Structural Similarity
SVD	Singular Value Decomposition
TV	Total Variation [23]
VAH	Method by Vahadane et al. [35]
WSSB	Warwick Stain Separation Benchmark [3]
ZHE	Method by Zheng et al. [42]

References

[1] S. Abousamra, D.J. Fassler, L. Hou, Y. Zhang, R.R. Gupta, T.M. Kurç, L.F. Escobar-Hoyos, D. Samaras, B. Knudson, K.R. Shroyer, J. Saltz, C. Chen, Weakly-supervised deep stain decomposition for multiplex ihc images, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 481–485.

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

[3] N. Alsubaie, N. Trahearn, S.E.A. Raza, D. Snead, N.M. Rajpoot, Stain deconvolution using statistical analysis of multi-resolution stain colour representation, PLoS ONE 12 (2017) e0169875.

[4] 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.

[5] A. Bentaieb, G. Hamarneh, Adversarial stain transfer for histopathology image analysis, IEEE Trans. Med. Imaging 37 (2018) 792–802.

[6] P. Bándi, O. Geessink, et al., From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge, IEEE Trans. Med. Imaging 38 (2019) 550–560.

[7] D. Carey, V. Wijayathunga, A.J. Bulpitt, D. Treanor, A novel approach for the colour deconvolution of multiple histological stains, in: Proceedings of the 19th Conference of Medical Image Understanding and Analysis, BMVA, 2015, pp. 156–162.

[8] R. Duggal, A. Gupta, R. Gupta, P. Mallick, Sd-layer: stain deconvolutional layer for cnns in medical microscopic imaging, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2017, pp. 435–443.

[9] A.E. Esteban, M. Lopez-Pérez, A. Colomer, M.A. Sales, R. Molina, V. Naranjo, A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes, Comput. Methods Programs Biomed. 178 (2019) 303–317.

[10] M. Gavrilovic, J.C. Azar, J. Lindblad, C. Wählby, E. Bengtsson, C. Busch, I.B. Carlbom, Blind color decomposition of histological images, IEEE Trans. Med. Imaging 32 (2013) 983–994.

[11] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, A.K. Katsaggelos, Variational Bayesian blind color deconvolution of histopathological images, IEEE Trans. Image Process. 29 (2019) 2026–2036.

[12] A. Howard, M. Sandler, G. Chu, L.C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, et al., Searching for mobilenetv3, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1314–1324.

[13] A. Janowczyk, A. Basavanhally, A. Madabhushi, Stain normalization using sparse autoencoders (stanosa): application to digital pathology, Comput. Med. Imaging Graph. 57 (2017) 50–61.

[14] N. Kanwal, F. Pérez-Bueno, A. Schmidt, K. Engan, 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 10 (2022) 58821–58844.

[15] A.M. Khan, N. Rajpoot, D. Treanor, D. Magee, A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution, IEEE Trans. Biomed. Eng. 61 (2014) 1729–1738.

[16] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, in: 2nd International Conference on Learning Representations (ICLR), 2014.

[17] A. Lucas, M. Iliadis, R. Molina, A. Katsaggelos, Using deep neural networks for inverse problems in imaging, IEEE Signal Process. Mag. 35 (2018) 20–36.

[18] S. López-Tapia, R. Molina, A.K. Katsaggelos, Deep learning approaches to inverse problems in imaging: past, present and future, Digit. Signal Process. 119 (2021) 103285.

[19] M. Macenko, M. Niethammer, J.S. Marron, D. Borland, J.T. Woosley, X. Guan, C. Schmitt, 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, pp. 1107–1110.

[20] N. Marini, M. Atzori, S. Otálora, S. Marchand-Maillet, H. Müller, H&e-adversarial network: a convolutional neural network to learn stain-invariant features through hematoxylin & eosin regression, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 601–610.

[21] M.T. McCann, J. Majumdar, C. Peng, C.A. Castro, J. Kovačević, Algorithm and benchmark dataset for stain separation in histology images, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE, 2014, pp. 3953–3957.

[22] J. Pan, J. Dong, Y. Liu, J. Zhang, J. Ren, J. Tang, Y.W. Tai, M.H. Yang, Physics-based generative adversarial models for image restoration and beyond, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 2449–2462.

[23] F. Pérez-Bueno, M. López-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, A.K. Katsaggelos, A tv-based image processing framework for blind color deconvolution and classification of histological images, Digit. Signal Process. 101 (2020) 102727.

[24] F. Pérez-Bueno, J.G. Serra, M. Vega, J. Mateos, R. Molina, 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).

[25] F. Pérez-Bueno, M. Vega, M.A. Sales, J. Aneiros-Fernández, V. Naranjo, R. Molina, A.K. Katsaggelos, Blind color deconvolution, normalization, and classification of histological images using general super Gaussian priors and Bayesian inference, Comput. Methods Programs Biomed. 211 (2021) 106453.

[26] A. Rabinovich, S. Agarwal, C. Laris, J. Price, S. Belongie, Unsupervised color decomposition of histologically stained tissue samples, Adv. Neural Inf. Process. Syst. 16 (2003) 667–674.

[27] R. Ranjan, C.D. Castillo, R. Chellappa, L2-Constrained Softmax Loss for Discriminative Face Verification, 2017.

[28] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer International Publishing, Cham, 2015, pp. 234–241.

[29] A.C. Ruifrok, D.A. Johnston, et al., Quantification of histochemical staining by color deconvolution, Anal. Quant. Cytol. Histol. 23 (2001) 291–299.

[30] P. Ruiz, X. Zhou, J. Mateos, R. Molina, A.K. Katsaggelos, Variational Bayesian blind image deconvolution: a review, Digit. Signal Process. 47 (2015) 116–127, Special Issue in Honour of William J. (Bill) Fitzgerald.

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

[32] M. Salvi, N. Michielli, F. Molinari, Stain color adaptive normalization (scan) algorithm: separation and standardization of histological stains in digital pathology, Comput. Methods Programs Biomed. 193 (2020) 105506.

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

[34] N. Trahearn, D. Snead, I. Cree, N. Rajpoot, Multi-class stain separation using independent component analysis, in: Medical Imaging 2015: Digital Pathology, International Society for Optics and Photonics, 2015, p. 94200J.

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

[36] L. Wang, K.J. Yoon, Deep learning for hdr imaging: state-of-the-art and future trends, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 8874–8895.

[37] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N.N. Shih, H. Gilmore, A. Madabhushi, Sparse non-negative matrix factorization (snmf) based color unmixing for breast histopathological image analysis, Comput. Med. Imaging Graph. 46 (2015) 20–29.

[38] Z. Yue, H. Yong, Q. Zhao, D. Meng, L. Zhang, Variational denoising network: toward blind noise modeling and removal, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc., 2019, pp. 1690–1701.

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

[40] Q. Zhao, H. Wang, Z. Yue, D. Meng, A deep variational Bayesian framework for blind image deblurring, Knowl.-Based Syst. 109008 (2022).

[41] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, D. Hu, S. Sun, J. Shi, C. Xue, Stain standardization capsule for application-driven histopathological image normalization, IEEE J. Biomed. Health Inform. 25 (2020) 337–347.

[42] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, C. Xue, Adaptive color deconvolution for histological wsi normalization, Comput. Methods Programs Biomed. 170 (2019) 107–120.

[43] N. Zhou, D. Cai, X. Han, J. Yao, Enhanced cycle-consistent generative adversarial network for color normalization of h&e stained images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2019, pp. 694–702.

[44] X. Zhou, J. Mateos, F. Zhou, R. Molina, A.K. Katsaggelos, Variational Dirichlet blur kernel estimation, IEEE Trans. Image Process. 24 (2015) 5127–5139.

Shuowen Yang received the degree in electronic science and technology from the Xidian University in 2016. He started the Ph.D. degree in 2018 at the Xidian University under the supervision of Prof. Hanlin Qin. His research interests focus on the infrared spectral imaging, computational imaging and image processing.

Fernando Pérez Bueno received the MSc in Data Science and Ph.D. degree in Information and Telecommunication Technologies from the Universidad de Granada in 2019 and 2022, respectively. He is currently a post-doctoral researcher at the University of Granada, member of the Visual Information Processing Group in the Department of Computer Science and Artificial Intelligence. His research interests focus on the use of Bayesian modeling and inference to solve different problems related to image restoration and machine learning. Centered in cancer histopathological images enhancement and classification. During his research, he has addressed problems, such as blind image deconvolution, pansharpening or anomaly detection.

Francisco Miguel Castro Macías obtained the Degree in Computer Science and the Degree in Mathematics from the University of Granada in 2021. He obtained the MSc degree in Data Science and Computer Engineering from the University of Granada in 2022. Since July 2021 he is member of the Visual Information Processing Group in the Department of Computer Science and Artificial Intelligence of the University of Granada. Since January 2022 he is a Ph.D. student under the supervision of Prof. Rafael Molina Soriano and Pablo Morales Álvarez. He is interested in the development of new deep probabilistic models for image restoration, deconvolution and classification under weak labeling paradigms, with application to medical imaging.

Rafael Molina received the degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He became Professor of Computer Science and Artificial Intelligence at the University of Granada, Granada, Spain, in 2000. He is the former Dean of the Computer Engineering School at the University of Granada (1992–2002) and Head of the Computer Science and Artificial Intelligence department of the University of Granada (2005–2007). His research interest focuses mainly on using Bayesian modeling and inference in problems like image restoration (applications to astronomy and medicine), superresolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low-rank matrix decomposition, active learning, fusion, and classification. Prof. Molina have served as an Associate Editor of Applied Signal Processing; the IEEE Transactions on Image Processing; and Progress in Artificial Intelligence; and an Area Editor of Digital Signal Processing. He is the recipient of an IEEE International Conference on Image Processing Paper Award (2007) and an ISPA Best Paper Award (2009). He is a coauthor of a paper awarded the runner-up prize at Reception for early-stage researchers at the House of Commons.

Aggelos K. Katsaggelos (F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. In 1985, he joined the Department of Electrical Engineering and Computer Science at Northwestern University, where he is currently a professor, holder of the Joseph Cummings Chair. He is a member of the Academic Staff, NorthShore University Health System, an affiliated faculty at the Department of Linguistics and he has an appointment with the Argonne National Laboratory. He was previously the holder of the Ameritech Chair of Information Technology and the AT&T Chair and the co-Founder and Director of the Motorola Center for Seamless Communications. He has published extensively in the areas of multimedia processing and communications (over 250 journal papers, 500 conference papers and 40 book chapters) and he is the

holder of 26 international patents. He is the co-author of Rate-Distortion Based Video Compression (Kluwer, 1997), Super-Resolution for Images and Video (Claypool, 2007), Joint Source-Channel Video Transmission (Claypool, 2007), Machine Learning Refined (Cambridge University Press, 2016) and The Essentials of Sparse Modeling and Optimization (Springer, 2017, forthcoming). He has supervised 56 PhD dissertations so far. Among his many professional activities Prof. Katsaggelos was Editor-in-Chief of the IEEE Signal Processing Magazine (1997-2002), a BOG Member of the IEEE Signal Processing Society (1999-2001), a member of the Publication Board of the IEEE Proceedings (2003-2007), and he is currently a member of the Awards Board of the Signal Processing Society. He is a Fellow of the IEEE (1998) and SPIE (2009) and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), the IEEE Signal Processing Society Technical Achievement Award (2010), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE ICME Paper Award (2006), an IEEE ICIP Paper Award (2007), an ISPA Paper Award (2009), and a EUSIPCO paper award (2013). He was a Distinguished Lecturer of the IEEE Signal Processing Society (2007-2008).