Unrolled projected gradient algorithm for stain separation in digital histopathological images

Aymen Sadraoui, Astrid Laurent-Bellue, Mounir Kaaniche, Amel Benazza-Benyahia, Catherine Guettier, Jean-Christophe Pesquet

UNROLLED PROJECTED GRADIENT ALGORITHM FOR STAIN SEPARATION IN DIGITAL HISTOPATHOLOGICAL IMAGES

Aymen Sadraoui[1] *Astrid Laurent-Bellue*[2,3,4] *Mounir Kaaniche*[1,5]
Amel Benazza-Benyahia[6] *Catherine Guettier*[2,3,4] *Jean-Christophe Pesquet*[1]

[1] Centre de Vision Numérique, Université Paris-Saclay, Inria, CentraleSupélec, Gif-sur-Yvette, France
[2]Department of Pathology, AP-HP. Hôpital Bicêtre, Le Kremlin-Bicêtre, France, [3]Faculté de Médecine, Université Paris-Saclay, Le Kremlin-Bicêtre, France, [4]INSERM U1193, Villejuif, France,
[5]Université Sorbonne Paris Nord, L2TI, UR 3043, Villetaneuse, F-93430, France,
[6]Université de Carthage, SUP'COM, LR11TIC01, COSIM Lab, Ariana, Tunisia.

ABSTRACT

This paper introduces a novel optimization approach for stain separation in digital histopathological images. Our stain separation cost function incorporates a smooth total variation regularization and is minimized by using a projected gradient algorithm. To enhance computational efficiency and enable supervised learning of the hyperparameters, we further unroll our algorithm into a neural network. The unrolled architecture is not only more efficient for solving the stain separation problem, but also allows to design a highly interpretable and flexible method. Experimental results demonstrate the effectiveness of the proposed unrolled projected gradient algorithm in achieving accurate and visually consistent stain separation.

***Index Terms*—** Proximal gradient, unrolling, stain separation, histopathology

1. INTRODUCTION

Digital histopathological image is widely used for the diagnosis of a considerable number of diseases including several types of cancer. To generate such data, tissues are stained using a combination of color dyes to better visualize and analyze their inherent structures. In this respect, one of the most common staining protocol is hematoxylin and eosin (HE) [1]. However, in some countries (e.g., France and Belgium), the staining process also includes the saffron stain, yielding (HES) images. Generally, the staining process may cause inevitable color variations due to differences in color responses of slide scanners, raw materials, manufacturing techniques of stain vendors, and staining protocols across different pathology laboratories [2]. Such variations affect the quantitative analysis of histopathology images and reduce the accuracy of different computer-aided systems. In this context, the standardization of the different color appearances has attracted a lot of attention though the development of stain normalization methods [3, 4, 5]. A key step in these methods is stain separation, which aims to identify and isolate each stain in the original image.

Various approaches have been developed in the literature for unmixing stains, also known as color deconvolution. One of the first developed methods has been proposed by Ruifrok and Johnston [6] using simple pseudo-inversion. Moreover, Non-negative Matrix Factorization (NMF) was used in one of the pioneering works for stain unmixing which is formulated as a blind source separation problem [7]. Based on this framework, Xu *et al.* [8] and Vahadane *et al.* [9] have included sparsity and smoothness constraints, and called it sparse NMF (SNMF). Furthermore, other classical methods use the popular singular value decomposition (SVD) [5, 10] and independent component analysis (ICA) [10, 11]. It was generally reported that NMF based algorithms outperform SVD and ICA methods [12]. However, the developed NMF based algorithms often introduce some hyperparameters, which are set in an empirical manner. Moreover, for efficient stain separation performance, the estimation of these hyperparameters must be performed for each image independently, which is a computationally demanding task. To alleviate this drawback, deep learning based approaches have been recently developed in the literature [13, 14, 15]. It is worth pointing that only few works using neural networks exist in the literature on this topic. This could be explained by the requirement of large size of data samples and the lack of ground truth images for each used stain. It should also be noted that the aforementioned deep learning methods introduce many parameters which need to be estimated.

In this paper, we propose a new approach for stain separation of HES-stained images. The proposed method takes advantage of the projected gradient algorithm and the unrolling paradigm. More precisely, starting from the standard problem formulation used in NMF based methods, we include a total-variation based regularization to impose the smoothness of the separated images. The regularized problem is then solved by a projected gradient algorithm. Finally, the algo-

rithm is unrolled into layers of a neural network.
The remainder of paper is organized as follows. In Section 2, we recall the main background on stain separation. The proposed methodology is then described in Section 3. Finally, experimental results are presented in Section 4, and some conclusions are provided in Section 5.

2. BACKGROUND ON STAIN SEPARATION

The main concept behind the developed stain separation techniques relies on the Beer-Lambert Law [6, 16], which states the existence of a linear relationship between the stain concentration matrix and the stain color-vector matrix in the *Optical Density* (OD) space.
To make this relationship more precise, let $I \in \mathbb{R}^{3\times N}$ be the vectorized HES-stained image in the RGB color space such that N is the number of pixels, and let $I_0 \in \mathbb{R}$ be the incident light (usually set to 255 for 8-bit images). Moreover, let $W \in \mathbb{R}^{3\times r}$ be the stain color-vector matrix whose columns (also called stain vectors) represent the RGB values for each stain, r being the number of stains (i.e., $r = 3$ for an HES stained-image), and let $H \in \mathbb{R}^{r\times N}$ be the stain concentration matrix (also called stain density map), whose rows represent the concentration of each stain. Then, according to the Beer-Lambert law, the image I follows the exponential form:

$$I = I_0 \, \exp(-WH), \tag{1}$$

where the exponential applies componentwise. Let $V \in \mathbb{R}^{3\times N}$ denote the OD of I, given by

$$V = -\log\left(\frac{I}{I_0}\right). \tag{2}$$

It follows from (1) and (2) that

$$V = WH, \tag{3}$$

which establishes the linear relationship existing between the OD of the HES stained-image, the stain color-vector matrix, and the stain concentration matrix.
Thus, the main goal of the stain separation process is to estimate the matrices W and H. However, inspired by [6] and other related works, a good estimation of the stain color-vector matrix W can be experimentally obtained and this matrix can therefore be considered as a known operator. In this case, the stain separation problem reduces to the determination of the stain concentration matrix H, which will be the focus of this work. Based on our experimental observations, it is judicious to consider the hematoxylin stain as two blue shade colors: a darker shade color corresponding mainly to the *nuclei* and a lighter shade color for the *background*, which will be denoted by $\mathrm{h_n}$ and $\mathrm{h_b}$, respectively. As a result, instead of using three stain vectors for the matrix W, the latter will be composed of four column vectors (i.e., $r = 4$), where the first and second columns represent $\mathrm{h_n}$ and $\mathrm{h_b}$, while the third and fourth columns correspond to the eosin (e) and saffron (s) stains.
Therefore, given the observation matrix V and the known stain color-vector matrix W, (3) can be solved by formulating the following optimization problem:

$$\begin{aligned} &\underset{H\in\mathbb{R}^{r\times N}}{\text{minimize}} \;\; \frac{1}{2}\|V - WH\|_{\mathrm{F}}^2 + R(H) \\ &\text{subject to} \quad H \geq 0 \end{aligned} \tag{4}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, and R is a regularization function.
Once the stain density map H is estimated, the reconstruction of each stain image is derived as

$$I_{\mathrm{h}} = I_0 \exp\big(-(W_1H_1 + W_2H_2)\big) \tag{5}$$

$$I_{\mathrm{e}} = I_0 \exp\big(-W_3\, H_3\big) \tag{6}$$

$$I_{\mathrm{s}} = I_0 \exp(-W_4\, H_4). \tag{7}$$

Hereabove I_{h}, I_{e}, and I_{s} are the reconstructed RGB images associated to the hematoxylin, eosin and saffron, respectively. In addition, for every $c \in \{1, \ldots, r\}$, W_c (resp. H_c) is the c-th column (resp. row) of W (resp. H).

3. PROPOSED METHOD

3.1. Notation and definitions

Let $\mathscr{H}$ be a Hilbert space, and let $\Gamma_0(\mathscr{H})$ denote the set of proper, lower semi-continuous convex functions from $\mathscr{H}$ to $\mathbb{R} \cup \{+\infty\}$. The proximity operator of a function $f \in \Gamma_0(\mathscr{H})$ is defined, for every $x \in \mathscr{H}$, by $\mathrm{prox}_f(x) = \mathrm{argmin}_{x\in\mathscr{H}} f(y) + \frac{1}{2}\|x-y\|_2^2$. For a given nonempty closed convex set C, ι_C is its indicator function defined, for every $x \in \mathscr{H}$, by $\iota_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise.

3.2. Problem formulation

Let us rewrite the main optimization problem (4) as the minimization of the sum of a data fidelity term $\tilde{g}$ involving W and H, a regularization term R, and an indicator function encoding the nonnegativity constraint on the pixel intensities:

$$\underset{H\in\mathbb{R}^{r\times N}}{\text{minimize}} \;\; \tilde{g}(H) + R(H; \lambda_1, \lambda_2, \varepsilon) + \iota_{[0,+\infty[^{r\times N}}(H). \tag{8}$$

The data fidelity term is thus defined as $\tilde{g}(H) = \frac{1}{2}\|V - WH\|_{\mathrm{F}}^2$, and the regularization function R is given by

$$R(H; \lambda_1, \lambda_2, \varepsilon) = \frac{\lambda_1}{2}\|H\|_{\mathrm{F}}^2 + \lambda_2\, \mathrm{STV}_\varepsilon(H), \tag{9}$$

with λ_1 and λ_2 are positive regularization parameters, and $\mathrm{STV}_\varepsilon(H)$ is the smoothed total variation term (with parameter $\varepsilon > 0$), which aims to ensure smoothness in homogeneous

color regions while preserving sharp edges. The smoothed total variation is expressed as

$$\begin{aligned}\mathrm{STV}_\varepsilon(H) &= \sum_{c=1}^{r} \mathrm{STV}_\varepsilon(H_c) \\ &= \sum_{c=1}^{r}\sum_{i=1}^{N} \sqrt{(D_v H_c^\top)_i^2 + (D_h H_c^\top)_i^2 + \varepsilon^2},\end{aligned}$$

where $D_v \in \mathbb{R}^{N\times N}$ and $D_h \in \mathbb{R}^{N\times N}$ are the vertical and horizontal discrete gradient operators, respectively. To the best of our knowledge, the existing NMF-based stain separation methods have mainly considered quadratic and/or sparse regularization terms. However, TV regularization has not yet been investigated in this specific problem, except in [17], where a TV prior model is adopted within a variational Bayesian framework.

3.3. Optimization algorithm

According to our choices for the regularization functions, Problem (8) can be seen as the minimization of the sum of two functions g and f expressed as

$$g = \tilde{g} + R \tag{10}$$

$$f = \iota_{[0,+\infty[^{r\times N}}. \tag{11}$$

It can be observed that function $g \in \Gamma_0(\mathbb{R}^{r\times N})$ is differentiable with a Lipschitz continuous gradient, whose Lipschitz constant will be denoted by $L > 0$. Moreover, $f \in \Gamma_0(\mathbb{R}^{r\times N})$ is a function whose proximity operator reduces to the projection $\mathrm{proj}_{[0,+\infty[^{r\times N}}$ onto the nonnegative orthant $[0,+\infty[^{r\times N}$. Based on these observations, our optimization problem can be solved using Algorithm 1. This algorithm is a projected gradient algorithm (PGA), which is a special case of the proximal gradient algorithms, one of the most popular proximal methods [18].

Algorithm 1 Projected Gradient Algorithm (PGA)

Input: Initial point $H_0 \in \mathbb{R}^{r\times N}$, fixed stepsize $\gamma \in]0, \frac{2}{L}[$ and number of iterations $K \in \mathbb{N}^*$.
for $k = 0, 1, \ldots, K-1$ **do**
$H_{k+1} = \mathrm{proj}_{[0,+\infty[^{r\times N}}\left(H_k - \gamma \nabla g(H_k; \lambda_1, \lambda_2, \varepsilon)\right)$
end for

It is worthy to note that the Lipschitz constant L of the gradient ∇g is given by

$$L = \|W\|_{\mathrm{S}}^2 + \lambda_1 + 8\frac{\lambda_2}{\varepsilon}, \tag{12}$$

where $\|W\|_{\mathrm{S}}$ denotes the spectral norm of W.

3.4. Unrolled architecture

The main difficulty in Algorithm 1 as well as most iterative optimization algorithms is the optimal setting of their hyperparameters. For this reason, we adopt in this paper a supervised learning strategy to set these hyperparameters using a training set of images. This is achieved by resorting to the unrolling paradigm. The main idea behind this concept is to map each iteration $k \in \{0, \ldots, K-1\}$ of Algorithm 1 to a network layer and stack the layers together [19, 20, 21]. By unrolling the algorithm in a neural network, the regularization parameters $\{\lambda_1, \lambda_2, \varepsilon\}$ and the stepsize γ are untied across the network, which will offer more flexibility to our approach and expand the exploration space. Such strategy leads to a vector of weight parameters $\Theta_k = [\lambda_{1,k}, \lambda_{2,k}, \varepsilon_k, \gamma_k]^\top$ to be learned for each layer index $k \in \{0, \ldots, K-1\}$. The update rule at a given iteration k reads

$$\begin{aligned}H_{k+1} &= \mathcal{A}(H_k; \Theta_k) \\ &= \mathrm{proj}_{[0,+\infty[^{r\times N}}(H_k - \gamma_k \nabla g(H_k; \lambda_{1,k}, \lambda_{2,k}, \varepsilon_k)).\end{aligned} \tag{13}$$

Therefore, for every $k \in \{0, \ldots, K-1\}$, the k-th layer can be seen as the combination of a hidden structure, $\mathcal{L}_k$, followed by the update $\mathcal{A}$. The hidden structure $\mathcal{L}_k$ aims to generate the vector of parameters Θ_k. Since these output parameters must be positive, we propose to consider the Softplus function [22], which can be seen as a smooth approximation of the ReLU activation function. Thus, the vector Θ_k is obtained as follows:

$$\Theta_k = \mathrm{Softplus}(\Psi_k) = \log(1 + \exp(\Psi_k)), \tag{14}$$

where $\Psi_k = [a_k, b_k, c_k, d_k]^\top$ represents a vector of parameters learned during the training.

While considering different set of parameters across the different layers, the resulting neural network is trained by minimizing the following loss function:

$$\mathcal{L}(\Theta) = \frac{1}{3}\sum_{\mathrm{c}\in\{\mathrm{h,e,s}\}} \ell(I_{\mathrm{c}}^{(\mathrm{GT})}, I_{\mathrm{c}}(\Theta)), \tag{15}$$

where $\Theta = (\Theta_k)_{0\leq k\leq K-1}$ represents the global set of parameters used in the unfolded network and ℓ is a given criterion used to compare the reconstructed image I_{c} associated to the stain c with its corresponding ground truth $I_{\mathrm{c}}^{(\mathrm{GT})}$.

4. EXPERIMENTAL RESULTS

4.1. Dataset and experimental settings

Our experiments are carried out on data acquired at the Kremlin-Bicêtre hospital. First, from a liver tissue volume, four closely adjacent whole slide images (WSIs) have been extracted. Then, three were exclusively stained with a single

pure stain, which will serve as ground truth, while the fourth underwent staining under the conventional HES protocol. Finally, from the acquired WSIs, we created a dataset of 325 HES-stained images, of size 512×512 pixels, with their corresponding hematoxylin, eosin, and saffron single stained images. The dataset is divided into a training set (of 225 images) and a test set (of 100 images).

Our unrolled architecture composed of 20 layers, results in limited number (80) of parameters to be learned. To train our architecture, the Perceptual Image-Error Assessment through Pairwise Preference (PieAPP) criterion [23] has been used as a loss function[1] in (15). Such criterion was found to be better correlated with human perception and more efficient in capturing image features compared to pixel-based metrics [24]. Our model is implemented in Pytorch, using the ADAM optimizer with an initial learning rate of 0.01, while applying a decay of 0.1 when there is no improvement every 5 epochs. The batch size and number of epochs were set to 5 and 150, respectively.

4.2. Results and discussion

Our proposed methods have been compared to different state-of-the-art (SOA) methods, including Ruifrok and Johnston [6], Xu *et al.* [8], Vahadane *et al.* [9], and Yang *et al.* [15]. For a fair comparison, in our proposed PGA-based method as well as the SNMF-based methods [8, 9], the hyperparameters have been optimized on the training dataset using the Nelder-Mead method [25]. The different methods have been firstly evaluated in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and PieAPP, as depicted in Table 1. It can be noticed that the proposed PGA-based method outperforms the SOA methods. Significant improvements are achieved through the unrolled architecture. A subjective comparison is performed by illustrating the separated stain images for a given test sample. Fig. 1 displays the ground truth stain images as well as the obtained ones using the proposed approaches and the best three SOA methods. Again, it can be observed that our proposed method yields the best quality for the different separated stain images, especially for the hematoxylin and saffron stain images. For illustrating the hematoxylin stain estimation process, Fig. 2 shows the two reconstructed images associated to $\mathrm{h_n}$ and $\mathrm{h_b}$ colors, which are combined to generate the final image I_{h}.

5. CONCLUSION AND PERSPECTIVES

This paper proposes an iterative projected gradient algorithm for the stain separation problem, incorporating smooth a total variation regularization. The algorithm is then unrolled in a neural network architecture. Our experiments demonstrate the efficiency of the proposed method, yielding significant objective and subjective gains compared to state-of-the-art methods.

Table 1. Stain separation performance.

	PSNR	SSIM	PieAPP
Ruifrok and Johnston [6]	44.841	0.355	2.134
Xu *et al.* [8]	44.315	0.369	1.979
Vahadane *et al.* [9]	42.220	0.368	1.851
Yang *et al.* [14]	42.755	0.355	1.841
PGA	45.254	0.394	1.477
Unrolled PGA	**46.525**	**0.427**	**1.161**

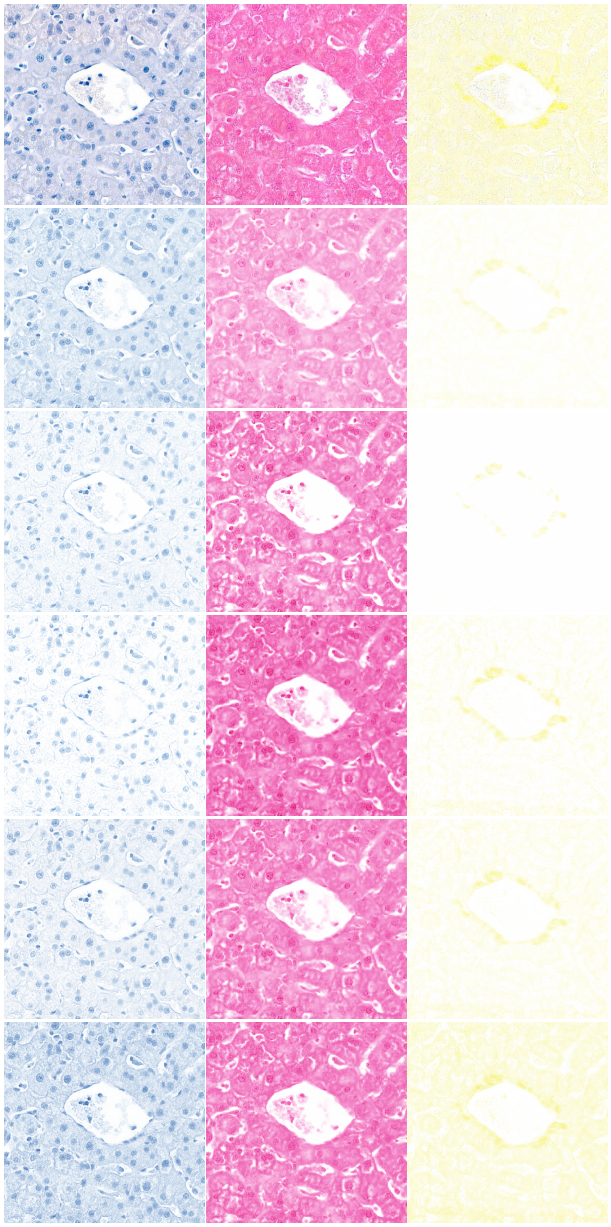

Fig. 1. Illustration of the separated stain images I_{h}, I_{e} and I_{s}. First row: Ground truth. Second row: Xu *et al.* [8]. Third row: Vahadane *et al.* [9]. Fourth row: Yang *et al.* [14]. Fifth row: PGA. Last row: unrolled PGA.

[1] https://github.com/photosynthesis-team/piq

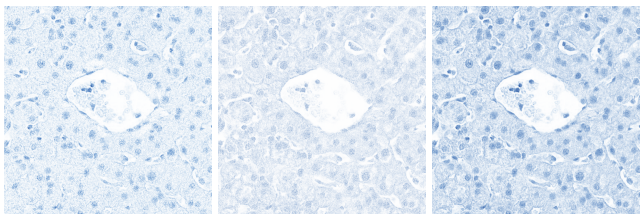

Fig. 2. From left to right, $I_{\mathrm{h_n}}$, $I_{\mathrm{h_b}}$, and I_{h} images.

Acknowledgment

We would like to thank the surgical team of the Centre Hépato-Biliaire of the Hôpital Paul Brousse, and the technicians of the Pathology department of the Hôpital Bicêtre.

Ethics approval

The study conformed to the General Data Protection Regulation (GDPR) and was approved by the Institutional Review Board of Mondor Hospital (IRB#00011558) (notification number: 2022-135).

6. REFERENCES

[1] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold Spring Harbor Protocols*, vol. 2008, pp. pdb.prot4986, May 2008.

[2] Y. Yagi, "Color standardization and optimization in Whole Slide Imaging," *Diagnostic Pathology*, vol. 6, no. S15, pp. 1–12, December 2011.

[3] N. Gurcan, L.E Boucheron, A. Can, A. Madabhushi, N.M Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[4] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1729–1738, June 2014.

[5] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, G. Xiaojun, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA, USA, June 2009, pp. 1107–1110.

[6] A.C. Ruifrok and D.A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and Quantitative Cytology and Histology*, vol. 23, no. 4, pp. 291–299, 2001.

[7] A. Rabinovich, S. Agarwal, C. Laris, J. Price, and S. Belongie, "Unsupervised color decomposition of histologically stained tissue samples," in *Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, April 2003, p. 667–674.

[8] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N. Nc Shih, H. Gilmore, and A. Madabhushi, "Sparse Non-negative Matrix Factorization (SNMF) based color unmixing for breast histopathological image analysis," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 20–29, December 2015.

[9] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 8, pp. 1962–1971, August 2016.

[10] M. T. McCann, J. Majumdar, C. Peng, C. A. Castro, and J. Kovacevic, "Algorithm and benchmark dataset for stain separation in histology images," in *IEEE International Conference on Image Processing (ICIP)*, Paris, France, October 2014, p. 3953–3957.

[11] N. Alsubaie, S. E. A. Raza, and N. Rajpoot, "Stain deconvolution of histology images via independent component analysis in the wavelet domain," in *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, Prague, Czech Republic, April 2016, p. 803–806.

[12] M. Salvi, N. Michielli, and F. Molinari, "Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology," *Computer Methods and Programs in Biomedicine*, vol. 193, pp. 105506, September 2020.

[13] A. Moyes, K. Zhang, M. Ji, H. Zhou, and D. Crookes, "Unsupervised deep learning for stain separation and artifact detection in histopathology images," in *Medical Image Understanding and Analysis*, April 2020, pp. 221–234.

[14] S. Yang, F. Pérez-Bueno, F. M. Castro-Macias, R. Molina, and A. K. Katsaggelos, "Deep Bayesian blind color deconvolution of histological images," in *IEEE International Conference on Image Processing (ICIP)*, Kuala Lumpur, Malaysia, October 2023, pp. 710–714.

[15] S. Yang, F. Pérez-Bueno, F. M. Castro-Macias, R. Molina, and A. K. Katsaggelos, "BCD-Net: Stain separation of histological images using deep variational

Bayesian blind color deconvolution," *Digital Signal Processing*, vol. 135, October 2024.

[16] D. F. Swinehart, "The Beer-Lambert Law," *Journal of Chemical Education*, vol. 39, no. 7, pp. 333, July 1962.

[17] F. Pérez-Bueno, M. Lopez-Pérez, M. Vega, J. Mateos, V. Naranjo, R. Molina, and A. K. Katsaggelos, "A TV-based image processing framework for blind color deconvolution and classification of histological images," *Digital Signal Processing*, vol. 101, October 2020.

[18] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer-Verlag*, vol. 49, pp. 185–212, December 2009.

[19] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, March 2021.

[20] C. Bertocchi, E. Chouzenoux, M.-C. Corbineau, J.-C. Pesquet, and M. Prato, "Deep unfolding of a proximal interior point method for image restoration," *Inverse Problems, Special Issue on Variational Methods and Effective Algorithms for Imaging and Vision*, vol. 36, no. 3, pp. 1–27, 2020.

[21] M. Savanier, E. Chouzenoux, J.-C. Pesquet, and C. Riddell, "Deep unfolding of the DBFB algorithm with application to ROI CT imaging with limited angular density," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 502–516, May 2023.

[22] C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia, "Incorporating second-order functional knowledge for better option pricing," in *Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, Dec 2001, p. 472–478.

[23] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "PieAPP: Perceptual Image-Error Assessment Through Pairwise Preference," in *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, June 2018, pp. 1808–1817.

[24] T. Dardouri, M. Kaaniche, A. Benazza-Benyahia, and J.-C. Pesquet, "Dynamic neural network for lossy-to-lossless image coding," *IEEE Transactions on Image Processing*, vol. 31, pp. 569–584, December 2021.

[25] J. Nelder and R. Mead, "A simplex method for function minimization," *IEEE Transactions on Computational Imaging*, vol. 7, no. 4, pp. 308–313, 1965.