Unrolled Projected Gradient Algorithm for Stain Separation in Digital Histopathological Images

A. Sadraoui[1] A. Laurent-Bellue[2] M. Kaaniche[1,3]
A. Benazza-Benyahia[4] C. Guettier[2] J.-C. Pesquet[1]

[1]Centre for Visual Computing, CentraleSupélec, Université Paris-Saclay, Inria, OPIS, Gif-Sur-Yvette, France
[2]Department of Pathology, AP-HP. Hôpital Bicêtre, INSERM U1193, Le Kremlin-Bicêtre, France,
[3]Université Sorbonne Paris Nord, L2TI, UR 3043, Villetaneuse, F-93430, France,
[4]Université de Carthage, SUP'COM, LR11TIC01, COSIM Lab, Ariana, Tunisia.

October 30th, 2024

Outline

Outline

Introduction

Histopathological images

Staining the tissue of a given organ using a combination of color dyes

- Hematoxylin (H): bluish-purple stain strongly related to the nuclei.
- Eosin (E): red-pink stain that highlights the cytoplasm of the nucleus.
- Saffron (S): yellow stain used to detect connective tissues.

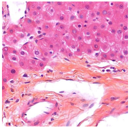

Figure 1: Example of HES-stained image.

Introduction

Histopathological images

Staining the tissue of a given organ using a combination of color dyes

- Hematoxylin (H): bluish-purple stain strongly related to the nuclei.
- Eosin (E): red-pink stain that highlights the cytoplasm of the nucleus.
- Saffron (S): yellow stain used to detect connective tissues.

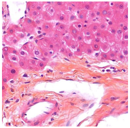

Figure 1: Example of HES-stained image.

Challenge: Variation in staining protocols

- Digital histopathological images, particularly HES-stained images, suffer from color variations due to differences in staining protocols, and materials.
- These color variations affect the accuracy of computer-aided systems used for disease diagnosis, especially in cancer detection.

Introduction

Histopathological images

Staining the tissue of a given organ using a combination of color dyes

- Hematoxylin (H): bluish-purple stain strongly related to the nuclei.
- Eosin (E): red-pink stain that highlights the cytoplasm of the nucleus.
- Saffron (S): yellow stain used to detect connective tissues.

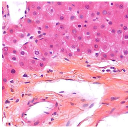

Figure 1: Example of HES-stained image.

Challenge: Variation in staining protocols

- Digital histopathological images, particularly HES-stained images, suffer from color variations due to differences in staining protocols, and materials.
- These color variations affect the accuracy of computer-aided systems used for disease diagnosis, especially in cancer detection.

☛ Need for the standardization/normalization of the different stain appearances to ensure consistent results.

Solution: Stain separation

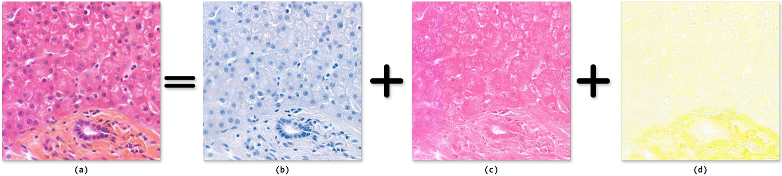

Figure 2: The principle of stain separation: (a) HES-stained image. (b) H-stained image. (c) E-stained image. (d) S-stained image.

Solution: Stain separation

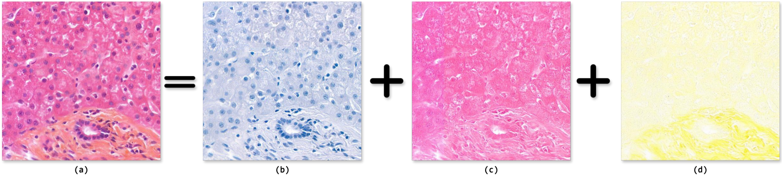

Figure 2: The principle of stain separation: (a) HES-stained image. (b) H-stained image. (c) E-stained image. (d) S-stained image.

Contribution

- State-of-the-art methods: SVD, ICA and NMF
- Traditional stain separation methods often require image-specific parameter tuning, which is set in an empirical manner and computationally expensive.

Solution: Stain separation

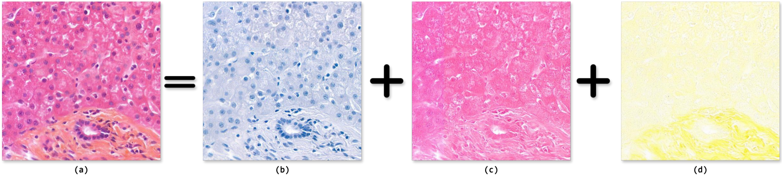

Figure 2: The principle of stain separation: (a) HES-stained image. (b) H-stained image. (c) E-stained image. (d) S-stained image.

Contribution

- State-of-the-art methods: SVD, ICA and NMF
- Traditional stain separation methods often require image-specific parameter tuning, which is set in an empirical manner and computationally expensive.

☛ Goal: design an efficient and robust stain separation method, and enable supervised learning of the hyperparameters

Solution: Stain separation

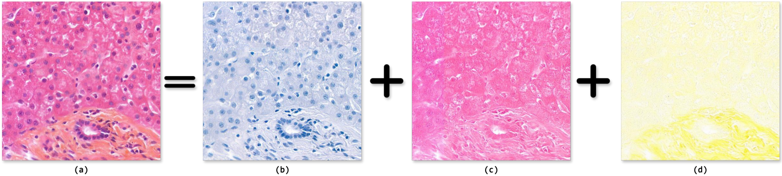

Figure 2: The principle of stain separation: (a) HES-stained image. (b) H-stained image. (c) E-stained image. (d) S-stained image.

Contribution

- State-of-the-art methods: SVD, ICA and NMF
- Traditional stain separation methods often require image-specific parameter tuning, which is set in an empirical manner and computationally expensive.

☛ Goal: design an efficient and robust stain separation method, and enable supervised learning of the hyperparameters
→ Main ideas: Projected Gradient algorithm and unrolling paradigm

Outline

Problem formulation

Beer-Lambert law [1]

$$I = I_0 \cdot \exp(-WH) \quad \underset{\text{Optical Density}}{\Longrightarrow} \quad V = -\log\left(\frac{I}{I_0}\right)$$

$$V = WH \tag{1}$$

- $I \in \mathbb{R}^{3\times N}$: vectorized HES-stained image
- I_0: incident light intensity
- $V \in \mathbb{R}^{3\times N}$: Optical Density (OD) version of I
- $W \in \mathbb{R}^{3\times r}$: stain-color vector matrix (can be experimentally estimated)
- $H \in \mathbb{R}^{r\times N}$: stain concentration matrix
- N is the image size
- r is the number of stains

Problem formulation

Beer-Lambert law [1]

$$I = I_0 \cdot \exp(-WH) \underset{\text{Optical Density}}{\Longrightarrow} V = -\log\left(\frac{I}{I_0}\right)$$

$$V = WH \tag{1}$$

- $I \in \mathbb{R}^{3\times N}$: vectorized HES-stained image
- I_0: incident light intensity
- $V \in \mathbb{R}^{3\times N}$: Optical Density (OD) version of I
- $W \in \mathbb{R}^{3\times r}$: stain-color vector matrix (can be experimentally estimated)
- $H \in \mathbb{R}^{r\times N}$: stain concentration matrix
- N is the image size
- r is the number of stains

☛ **<u>Goal:</u> Estimate H given an observed V and a known W.**

Problem formulation

Equation (1) can be solved by formulating the following optimization problem:

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

Problem formulation

Equation (1) can be solved by formulating the following optimization problem:

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

Problem (2) can be rewritten as follows:

$$\underset{H\in\mathbb{R}^{r\times N}}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|V - WH\|_{\mathrm{F}}^2}_{\text{Data fidelity term}} + \underbrace{R(H;\lambda_1,\lambda_2,\varepsilon)}_{\text{Regularization term}} + \underbrace{\iota_{[0,+\infty[^{r\times N}}(H)}_{\text{nonnegativity constraint}} \tag{3}$$

Problem formulation

- The Regularization term R is given by

$$R(H;\lambda_1,\lambda_2,\varepsilon) = \underbrace{\frac{\lambda_1}{2}\|H\|_{\mathrm{F}}^2}_{\text{quadratic term}} + \underbrace{\lambda_2 \sum_{c=1}^{r}\sum_{i=1}^{N}\sqrt{(D_v H_c^\top)_i^2 + (D_h H_c^\top)_i^2 + \varepsilon^2}}_{\text{smoothed total variation (STV)}} \quad (4)$$

where λ_1 and λ_2 are positive regularization parameters, ε is the STV parameter and, $D_v \in \mathbb{R}^{N\times N}$ and $D_h \in \mathbb{R}^{N\times N}$ are the vertical and horizontal discrete gradient operators, respectively.

Problem formulation

Our minimization problem can be seen as the minimization of two functions g and f

$$\underset{H \in \mathbb{R}^{r \times N}}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|V - WH\|_F^2 + R(H; \lambda_1, \lambda_2, \varepsilon)}_{g(H;\lambda_1,\lambda_2,\varepsilon)} + \underbrace{\iota_{[0,+\infty[^{r\times N}}(H)}_{f(H)} \tag{5}$$

- f and g are proper lower-semicontinuous convex functions on $\mathbb{R}^{r\times N}$
- g is differentiable with an L-Lipschitzian gradient with respect to H.
- f is a function whose proximity operator reduces to the projection $\mathrm{proj}_{[0,+\infty[^{r\times N}}$ onto the nonnegative orthant $[0, +\infty[^{r\times N}$.

Problem formulation

Our minimization problem can be seen as the minimization of two functions g and f

$$\underset{H\in\mathbb{R}^{r\times N}}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|V - WH\|_{\mathrm{F}}^2 + R(H;\lambda_1,\lambda_2,\varepsilon)}_{g(H;\lambda_1,\lambda_2,\varepsilon)} + \underbrace{\iota_{[0,+\infty[^{r\times N}}(H)}_{f(H)} \tag{5}$$

- f and g are proper lower-semicontinuous convex functions on $\mathbb{R}^{r\times N}$
- g is differentiable with an L-Lipschitzian gradient with respect to H.
- f is a function whose proximity operator reduces to the projection $\mathrm{proj}_{[0,+\infty[^{r\times N}}$ onto the nonnegative orthant $[0,+\infty[^{r\times N}$.

☛ Problem (5) can be solved using **Projected Gradient Algorithm (PGA)**, which is a special case of the proximal gradient algorithms [2].

Optimization algorithm

Algorithm 1 Projected Gradient Algorithm (PGA)

Input: Initial point $H_0 \in \mathbb{R}^{r \times N}$, fixed stepsize $\gamma \in \left]0, \frac{2}{L}\right[$ and number of iterations $K \in \mathbb{N}^*$.

for $k = 0, 1, \ldots, K-1$ **do**

$H_{k+1} = \mathrm{proj}_{[0,+\infty[^{r \times N}} \big(H_k - \gamma \nabla g(H_k; \lambda_1, \lambda_2, \varepsilon) \big)$

end for

where L is the Lipschitz constant of the gradient ∇g given by

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

Optimization algorithm

Algorithm 2 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**
$\quad H_{k+1} = \operatorname{proj}_{[0,+\infty[^{r \times N}} \big(H_k - \gamma \nabla g(H_k; \lambda_1, \lambda_2, \varepsilon)\big)$
end for

where L is the Lipschitz constant of the gradient ∇g given by

$$L = \|W\|_S^2 + \lambda_1 + 8\frac{\lambda_2}{\varepsilon}.$$

Difficulty

Hyperparameters setting (γ, λ_1, λ_2, ϵ)

Outline

Unrolled PGA

Advantages

- Deployment of a **neural network** architecture.
- **Learning the hyperparameters** from a training dataset.
- **Interpretable and flexible** algorithm.
- Reducing the required number of iterations (**faster convergence**).

Unrolled PGA

One layer $\mathcal{L}_k$ of the Unrolled PGA mirrors one iteration for PGA by

$$\begin{aligned} H_{k+1} &= \mathrm{proj}_{[0,+\infty[^{r\times N}} \left(H_k - \gamma_k \nabla g(H_k; \lambda_{1,k}, \lambda_{2,k}, \varepsilon_k)\right) \\ &= \mathrm{proj}_{[0,+\infty[^{r\times N}} \left(H_k - \gamma_k (A(H) + B(H; \lambda_{1,k}) + C(H; \lambda_{2,k}, \varepsilon_k))\right) \end{aligned} \tag{6}$$

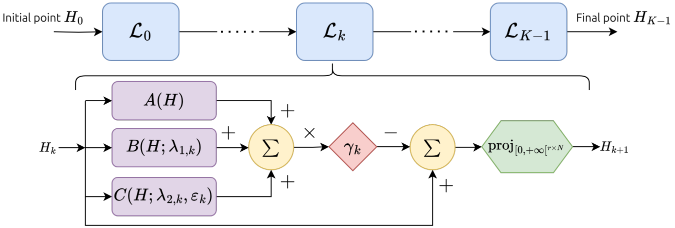

Figure 3: Unrolled PGA architecture.

where,

- $A(H) = W^\top (WH - V)$
- $B(H; \lambda_{1,k}) = \lambda_{1,k} H$
- $C(H; \lambda_{2,k}, \varepsilon_k) = \nabla STV(H; \lambda_{2,k}, \varepsilon_k)$

Unrolled PGA

Hyperparameters learning

To obtain the vector of parameters $\Theta_k = [\lambda_{1,k}, \lambda_{2,k}, \varepsilon_k, \gamma_k]^\top$, and ensure its positivity, we consider:

$$\forall k \in \{0, .., K-1\} \quad \Theta_k = \text{Softplus}(\Psi_k), \tag{7}$$

where Ψ_k is a vector of parameters learned during the training.

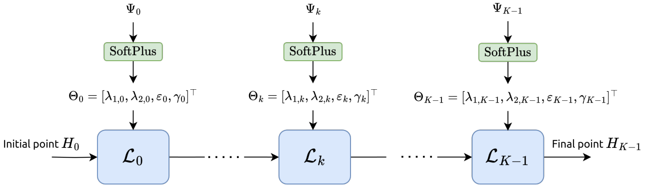

Figure 4: Unrolled PGA architecture with parameters learning.

Unrolled PGA

Loss function

The resulting neural network architecture is trained by minimizing:

$$\mathcal{L}(\Theta) = \frac{1}{3} \sum_{\mathrm{c} \in \{\mathrm{h},\mathrm{e},\mathrm{s}\}} \ell(I_{\mathrm{c}}^{(\mathrm{GT})}, I_{\mathrm{c}}(\Theta)), \tag{8}$$

where

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

Outline

Experimental settings

- Our data was acquired at the Kremlin-Bicêtre hospital, France.
- The unrolled PGA was composed of 20 layers $\Rightarrow$ **80 trainable parameters.**
- Our model is implemented in Pytorch, using the ADAM optimizer with an initial learning rate set to 0.01.
- The batch size and number of epochs were set to 5 and 150, respectively.

Numerical results

Table 1: Stain separation performance.

	PSNR	SSIM	PieAPP
Ruifrok and Johnston [3]	44.841	0.355	2.134
Xu *et al.* [4]	44.315	0.369	1.979
Vahadane *et al.* [5]	42.220	0.368	1.851
Yang *et al.* [6]	42.755	0.355	1.841
PGA	45.254	0.394	1.477
Unrolled PGA	**46.525**	**0.427**	**1.161**

Key observations: PGA and Unrolled PGA vs SOA Methods

- Both PGA and Unrolled PGA demonstrated superior performance compared to the SOA methods.
- The Unrolled PGA outperformed all methods across all metrics.

Visual results

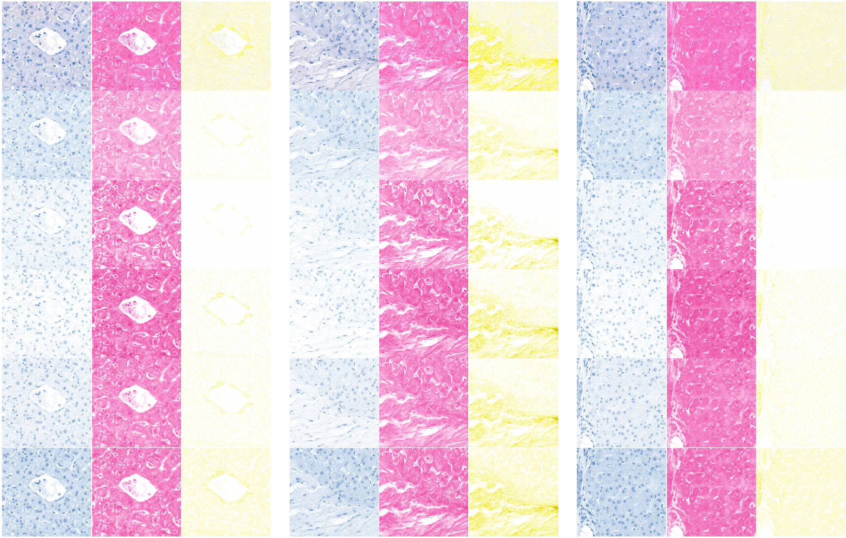

Figure 5: Illustration of the separated stain images (3 examples). 1st row: Ground truth. 2nd row: Xu *et al.* [4]. 3rd row: Vahadane *et al.* [5]. 4th row: Yang *et al.* [6]. 5th row: PGA. Last row: unrolled PGA.

Outline

Conclusion and Perspectives

Conclusion

- ✔ Designing an iterative projected gradient algorithm for the stain separation problem.
- ✔ Incorporating smooth total variation regularization to improve the quality of separation.
- ✔ Unrolling the algorithm into a neural network architecture
- ✔ The proposed method demonstrated significant objective improvements, in terms of PSNR, SSIM, and PieAPP, over state-of-the-art methods.
- ✔ Subjective evaluations showed that our approach produced visually superior results in stain separation.

Conclusion and Perspectives

Conclusion

- ✔ Designing an iterative projected gradient algorithm for the stain separation problem.
- ✔ Incorporating smooth total variation regularization to improve the quality of separation.
- ✔ Unrolling the algorithm into a neural network architecture
- ✔ The proposed method demonstrated significant objective improvements, in terms of PSNR, SSIM, and PieAPP, over state-of-the-art methods.
- ✔ Subjective evaluations showed that our approach produced visually superior results in stain separation.

Perspectives

- Extending the optimization problem to estimate both the matrices $\mathbf{W}$ and $\mathbf{H}$ simultaneously
- Going beyond stain separation to tackle stain normalization as well as other potential applications

References

[1] D. F. Swinehart, "The Beer-Lambert Law," *Journal of Chemical Education*, vol. 39, p. 333, July 1962.

[2] 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.

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

[4] J. Xu, L. Xiang, G. Wang, S. Ganesan, M. Feldman, N. N. 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.

[5] 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, pp. 1962–1971, August 2016.

[6] 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), pp. 710–714, October 2023.