

Computational normalization of H&E-stained histological images: Progress, challenges and future potential

Thaína A. Azevedo Tosta^{a,*}, Paulo Rogério de Faria^b, Leandro Alves Neves^c, Marcelo Zanchetta do Nascimento^{a,d}

^a Center of Mathematics, Computing and Cognition, Federal University of ABC, Av. dos Estados, 5001, 09210-580, Santo André, São Paulo, Brazil

^b Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia, Av. Amazonas, S/N, 38405-320, Uberlândia, Minas Gerais, Brazil

^c Department of Computer Science and Statistics, São Paulo State University, R. Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto, São Paulo, Brazil

^d Faculty of Computer Science, Federal University of Uberlândia, Av. João Naves de Ávila, 2121, 38400-902, Uberlândia, Minas Gerais, Brazil

ARTICLE INFO

Keywords:

Histological image analysis
Hematoxylin-eosin
Normalization
Color corrections

ABSTRACT

Different types of cancer can be diagnosed with the analysis of histological samples stained with hematoxylin-eosin (H&E). Through this stain, it is possible to identify the architecture of tissue components and analyze cellular morphological aspects that are essential for cancer diagnosis. However, preparation and digitization of histological samples can lead to color variations that influence the performance of segmentation and classification algorithms in histological image analysis systems. Among the determinant factors of these color variations are different staining time, concentration and pH of the solutions, and the use of different digitization systems. This has motivated the development of normalization algorithms of histological images for their color adjustments. These methods are designed to guarantee that biological samples are not altered and artifacts are not introduced in the images, thus compromising the lesions diagnosis. In this context, normalization techniques are proposed to minimize color variations in histological images, and they are topics covered by important studies in the literature. In this proposal, it is presented a detailed study of the state of art of computational normalization of H&E-stained histological images, highlighting the main contributions and limitations of correlated works. Besides, the evaluation of normalization methods published in the literature are depicted and possible directions for new methods are described.

1. Introduction

Cancer is defined as a tissue mass composed of genetically modified cells that do not respond to regulatory mechanisms of cellular growth. Currently, it is considered one of the main causes of death in the world due to late diagnoses [1,2]. Until 2030, almost 23 million new cases of cancer are estimated all over the world [3]. In Brazil, 600 thousand new cases were estimated between 2016-2017 [4]. The continuous demographic and epidemiological changes indicate the rise of cancer incidence in the next decades, which demands studies for improving diagnosis and treatment for cancer-affected patients [5].

To confirm the diagnosis of cancer, a microscopic analysis of cells and tissue components of a stained cancer sample has to be performed by a pathologist [6]. To do so, such a sample must be prepared by using a specific staining protocol, which allows recognizing histological

structures through the impregnation of dyes that give them colors [7]. The most common stain used in pathological practices is the hematoxylin-eosin (H&E), in which the hematoxylin dye stains acid structures in purple and the eosin dye stains the basic ones in pink [8].

The sample digitization allows to obtain digital histological images [9]. These images advantageously replace physical histological slides for educational purposes, distance consultancy and, mainly, the development of computational algorithms of digital image processing [10]. Histological image processing methods are used for developing computer-aided diagnosis (CAD) systems to help pathologists make diagnoses, which are subjective and time-consuming [11–13].

Among the several challenges imposed to CAD systems development, the color variations of histological images can significantly reduce the performance of segmentation and classification techniques [14]. These variations arise from the preparation and digitization steps

* Corresponding author.

E-mail addresses: tosta.thaina@gmail.com (T.A. Azevedo Tosta), paulo.faria@ufu.br (P.R. de Faria), neves.leandro@gmail.com (L.A. Neves), marcelo.zanchetta@gmail.com (M.Z. do Nascimento).

of histological samples. The use of different fixatives, variations in the digitization systems, and differences in the slides storage are examples of factors that lead to color variations of these images. Besides, the staining step can have a huge influence in the colors presented by the images due to differences in concentration of a staining solution, pH of a staining solution, chemistry oxidation and staining time [15].

In this context, normalization methods are proposed so the images colors can be adjusted [16]. For these corrections be efficient, it is necessary to guarantee that the histological structures are preserved and artifacts are not introduced. Various H&E-stained histological image normalization techniques are proposed in the literature, but each one has its respective limitation and contribution to the state of art. In this study, these techniques are presented with an indication of new perspectives for the development of new normalization methods. To the best of our knowledge, the study of Onder et al. [17] also presents the description, discussion and indication of future researches for histological images normalization based on techniques previously published, being the first review on this topic. In order to contribute to the review of normalization and deconvolution methods of Onder et al. [17], published in 2014, this study includes publications exclusively dedicated to H&E-stained histological image normalization between 2008 and 2017, including 2015 that was the year with the greatest number of publications on this topic and that were not described in [17].

This study is organized into six sections. In Section 2, the steps for obtaining histological images are described, as well as the possibility of color variation associated with them. Section 3 corresponds to the methodological section with the description of the criteria employed to select the normalization papers and the databases to do that. Sections 4 and 5 describe the H&E-stained histological image normalization studies of the literature and their result evaluation methods, respectively. In Section 6, the discussion of limitations of these studies and possible directions for the development of new proposals are presented.

2. Histological sample preparation and digitization: color variations sources

Different protocols of histological samples preparation are employed for obtaining their images [18]. These samples should be prepared to preserve their structures and present an appearance similar to their aspect in the living organism [19]. The steps used for this purpose are: (1) collection, (2) fixation, (3) dehydration and clearing, (4) paraffin embedding, (5) microtomy, (6) staining, and (7) mounting [19]. Such steps are described as follows, as well as their possible color variations:

- **Collection:** This step is responsible for the removal of a tissue sample from an organism still alive or *post mortem* [18];
- **Fixation:** Fixation is used for tissue preservation to prevent its digestion by cellular enzymes or bacteria [20]. In this step, the tissue is immersed in substances called fixatives, which prevent physical and chemical alterations. Due to the variety of available fixatives, it is necessary to choose the most adequate one to the type of tissue to be analyzed [18]. The immediate sample alteration is called autolysis, which is an enzymatic digestion of cells and tissues caused by inadequate fixation, leading to a reduced affinity of dyes to tissue structures [21]. In addition to the selective affinities of the fixatives to the type of tissue, their fixation time can also alter the appearance of the histological samples. Figs. 1 and 2 illustrate the variation of fixation time and the application of different fixatives [22];
- **Dehydration and clearing:** To be cut in thin slices, tissues are submitted to a histological processing that aims to initially remove water from the tissue and then to perfuse it with paraffin, hence giving it enough hardness to facilitate the microtomy. To do so, the sample is dehydrated through a series of graded concentrations of alcohol baths (from 75% to 95%). The clearing is used because the alcohol is not a paraffin solvent, so it has to be removed and

replaced by xylol through successive baths until the tissue becomes enough transparent [19,21]. Whether the dehydration is not adequately performed, water drops can be microscopically observed, which make histological sample regions opaque. As a consequence, microscopic details may be lost and unexpected changes in staining patterns of cells and tissue structures may happen, as illustrated in Fig. 3 [23];

- **Paraffin embedding:** In this step, the tissue is embedded with heated paraffin, which contributes to evaporate the solvent used in the previous step and fills all the spaces within the tissue. At the end of this step, a histological sample-containing paraffin block is obtained [19,21];
- **Microtomy:** After the removal of paraffin excess, the paraffin block is progressively sectioned by the microtome [19,21]. After that, the microtomy cuts are placed on glass slides. This step may have some influence in sample appearance in case the tissue section is not uniformly cut. Samples with 4–5 µm of thickness allow recognizing cellular morphology (nucleus and cytoplasm) and tissue architecture. Thinner samples may give more details of components within the nucleus, such as nucleolus and chromatin distribution [24]. In Fig. 4, the differences of samples with 4 µm (a) and 1 µm (b) are illustrated [24];
- **Staining:** After removing paraffin and hydrating the tissue, the stain is applied for distinguishing cellular and histological components since tissue samples become translucent after submitting to tissue processing [19,21]. Various factors can influence the tissue staining, such as different staining time, and concentration and pH of the solutions [15]. Fig. 5 illustrates an example of color variation due to changes in the staining step with the hematoxylin and eosin dyes [25].
- **Mounting:** Finally, the slides are covered with a mounting media and then with a coverslip to protect the tissue sample and promote a better visualization under the microscope [19]. This step can produce artifacts to the histological tissue like bubbles, dust, or contamination with microorganisms. Fig. 6 shows a histological image obtained with the representation of dust in a dark color, which cannot be altered by the image normalization [23].

2.1. Digitization of histological samples

Sample digitization can also impose color variations to histological images, and such variations can even happen with the use of different scanners. The public dataset in [26] is composed of histological samples digitized by the Aperio XT and Hamamatsu scanners. Images from this dataset are exemplified in Fig. 7, which depicts differences in colors between the images in (a) and (b).

The sample storage condition can also alter the way in which the tissue interacts with the stain, in addition to its natural discoloration [14]. This leads to color variations as exemplified by faded samples illustrated in Fig. 8 [27].

3. Materials and methods

This review is composed of studies of the literature, published between 2008 and 2017, that propose computational techniques for color normalization considering its application to H&E-stained histological images. This well-defined application allowed the selection of relevant studies for this paper. To do so, we did not limit this selection by the type of cancer of the images employed for performance evaluation or by the artificial intelligence techniques of their methodologies. Therefore, the described works present a great variety of images and computational methods. In addition to the restriction of application to H&E-stained histological images, the second main criterion employed was the selection of papers exclusively dedicated to the proposal of normalization techniques. Thus, works that used the normalization as a preprocessing for image analysis tasks were not considered in this

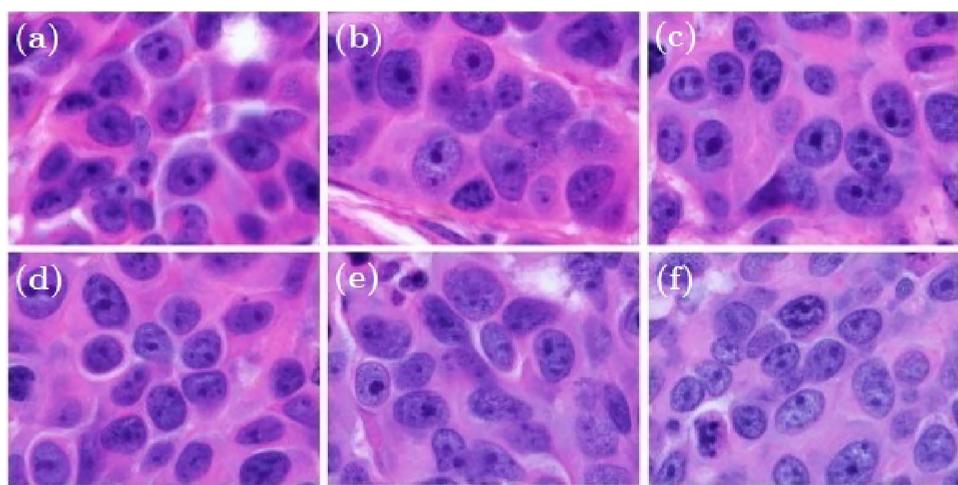


Fig. 1. Histological sample of 60× magnification processed with the alcoholic formalin fixative in intervals of 1 h (a), 3 h (b), 6 h (c), 12 h (d), 24 h (e) and 48 h (f) (figure adapted from [22] © 2010 Elsevier).

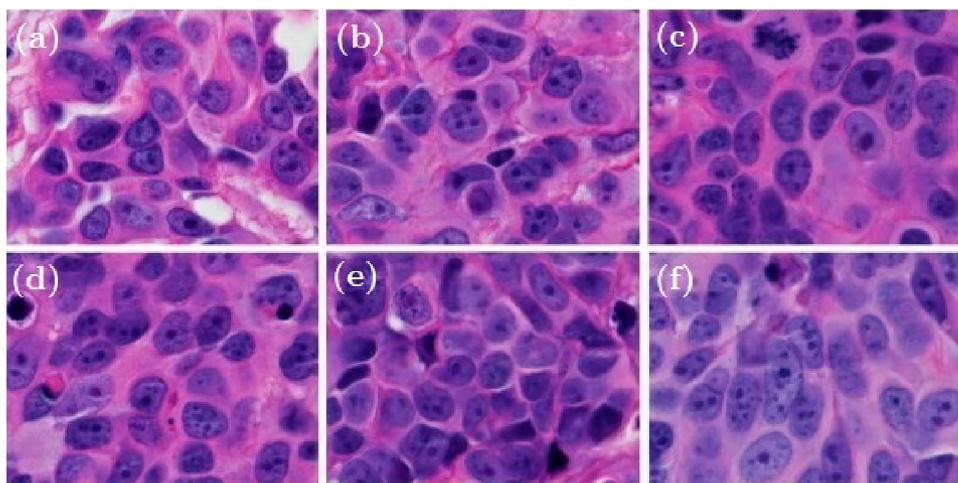


Fig. 2. Histological samples submitted to different intervals of fixation time of 1 h (a), 3 h (b), 6 h (c), 12 h (d), 24 h (e) and 48 h (f) through the zinc formalin fixative (figure adapted from [22] © 2010 Elsevier).

review.

The papers were selected using keywords not limited to “histological images”, “H&E”, “color normalization”, “standardization”, “color correction” and “stain estimation” by the Scopus¹ and Google Scholar² databases. Other papers were also found through the references of the papers found by the aforementioned queries, and by the works that cited them. After that, the papers were classified into the groups of high, medium and low relevance according to the aim of our review. This classification allowed this paper to have only the 23 studies corresponding to the works of the high relevance group.

4. Color normalization methods of H&E-stained histological images

Normalization methods are developed to minimize the influence of color variations in histological image analysis. These variations present high impact on computational analyses of these images, hence making normalization a useful concept for this application [28]. Normalization algorithms of H&E-stained images proposed in the literature are presented in this section.

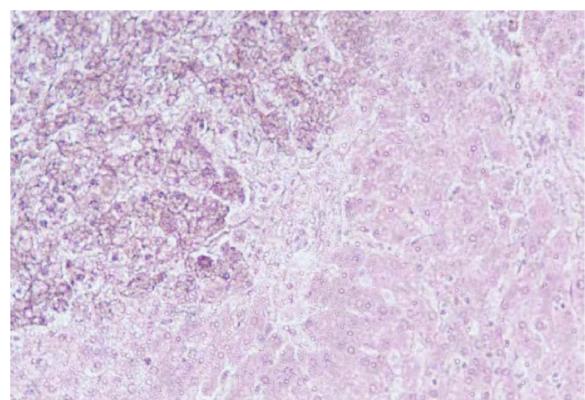


Fig. 3. Digitization of a histological sample submitted to an inefficient step of dehydration (figure extracted from [23] © 2008 Melbourne: Leica Microsystems).

Normalization methods can be classified into histogram matching, color transfer, and spectral matching [14]. Histogram matching techniques adjust images colors in the RGB space through information represented in their histograms. Using this representation, local information are disregarded, which leads to information loss [28].

¹ <https://www.scopus.com/>.

² <https://scholar.google.com.br/>.

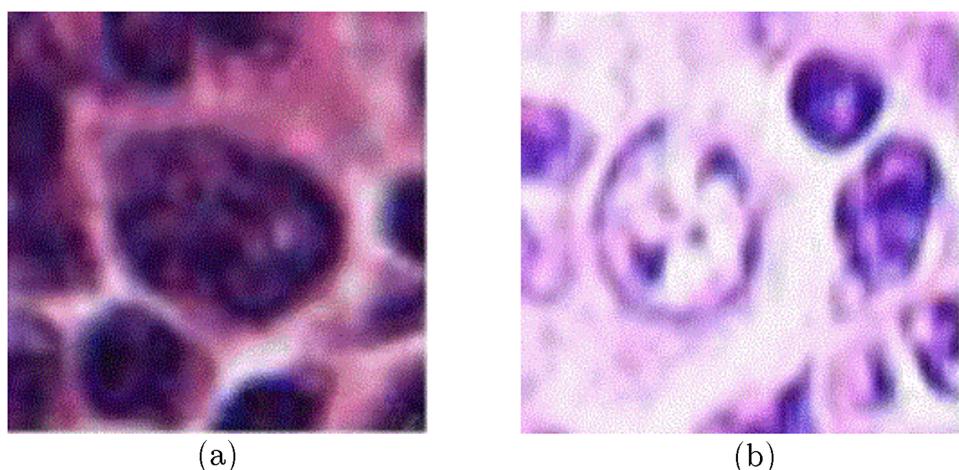


Fig. 4. Histological samples with thicknesses of 4 µm (a) and 1 µm (b) in the step of microtomy (figure extracted from [24] © 2014 IEEE).

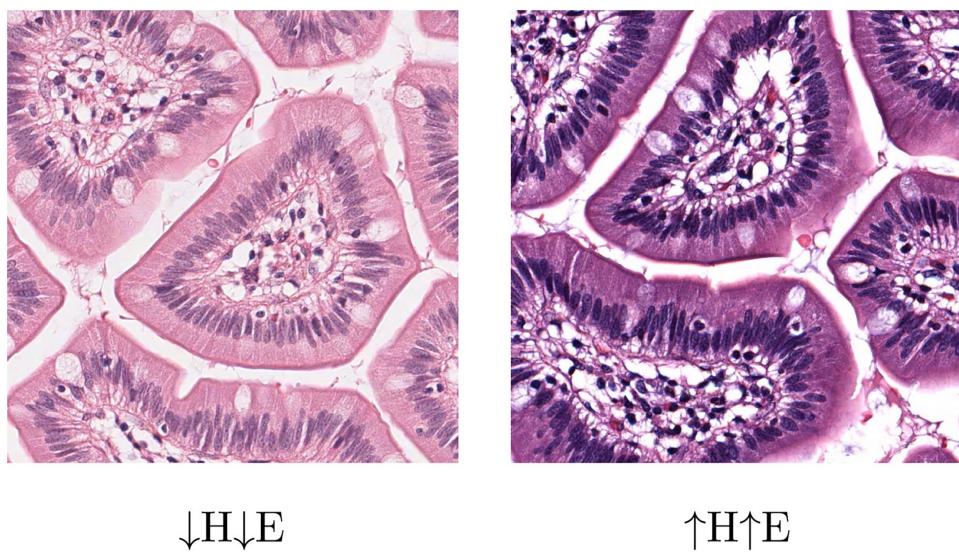


Fig. 5. Histological samples with variation in the staining step with under- (↓) and over-staining (↑) (figure adapted from [25] © 2017 Elsevier).

Therefore, the histogram matching assumes that the histological images have the same proportion of stained tissue components, which is a varying factor in most of these images [29]. Color transfer techniques use statistical correspondences between histological regions identified by a segmentation step or between dyes channels defined by a deconvolution. This class can also degrade histological information by the colors adjustment according to the results of segmentation and deconvolution. The spectral matching performs the normalization by estimating the representation of each dye in the RGB color channels, and the dyes concentrations in each image pixel. This approach stands out due to its potential histological structure preservation [14,28].

Like the spectral matching methods, color transfer techniques can also use the concepts of stain color appearance matrix and stain density map, both of which are defined here to standardize the terminology employed in this study. Stained histological samples attenuate the incident light on them in a determined color spectrum (RGB color channels), which is dependent on the used dyes and the amount of them absorbed by the histological structures. This relationship between RGB

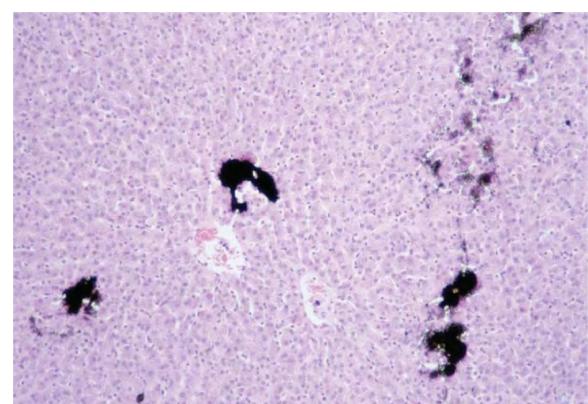


Fig. 6. Histological image with the representation of dust to exemplify the possible introduction of artifacts in the mounting step (figure extracted from [23] © 2008 Melbourne: Leica Microsystems).

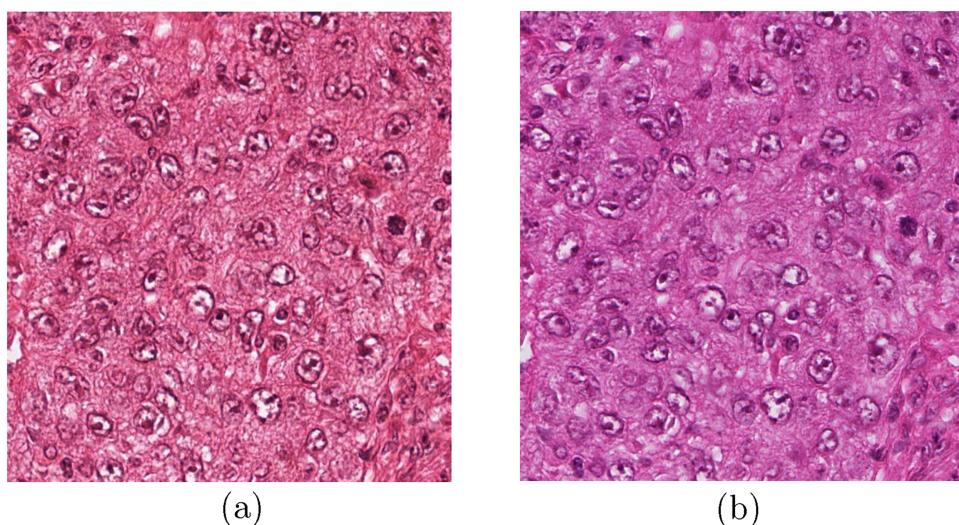


Fig. 7. Samples digitized by the Aperio XT (a) and Hamamatsu (b) scanners for illustration of color variations in the histological samples digitization step [26].

colors and stain density is defined by the Beer-Lambert law. Let I be a RGB image distributed in $m \times n$, where m corresponds to the RGB channels and n to the number of image pixels, I_0 be the incident light on the histological sample, W be the stain color appearance matrix with dimension of $m \times r$, where r represents the H&E stain applied to the sample ($r = 2$), and H corresponding to the stain density map with the dimension of $r \times n$, then the Beer-Lambert law is given by [30]:

$$I = I_0 \cdot \exp(-W \cdot H). \quad (1)$$

The relative optical density (V) is defined by:

$$V = \log \frac{I_0}{I}. \quad (2)$$

Then, the relationship between V and the Beer-Lambert law is established by:

$$V = W \cdot H. \quad (3)$$

By these principles, H&E-stained histological images are resulting from the combination between W and H . The W matrix is composed of coefficients that specify how the dyes are represented by the RGB channels. Then, its (i,j) th coefficient represents the absorption spectrum of the j th dye in the i th RGB color channel [14]. The H map presents the proportion of each dye represented by the image pixels [15]. In that way, the (i,j) th coefficient corresponds to the i th dye quantification in

the j th image pixel.

The description of studies from the literature presented by this work is based on the classes of histogram matching, color transfer and spectral matching. The classes of histogram matching and color transfer are represented in this work by the class called statistic-based color correction. This grouping was used to make the manuscript less complex since only one described study uses a histogram matching technique [16]. This name was chosen because the color transfer methods employ statistical metrics in their color correction and the histogram matching employs the histogram representation, which is a kind of statistical plot [31]. Even that these classes were merged, it is important to highlight that their difference is observed by the global approach used by the histogram matching and the local color correction observed in the color transfer. **Table 1** summarizes the normalization methods published between 2008 and 2017, their image datasets, their distribution among the mentioned classes, and their evaluation methods that allowed to assign them to evaluation classes.

4.1. Statistic-based color correction

Among the studies presented in **Table 1**, the work of Hoffman et al. [16] was the only one that employed histogram matching methods. This study proposed the comparison of five normalization techniques:

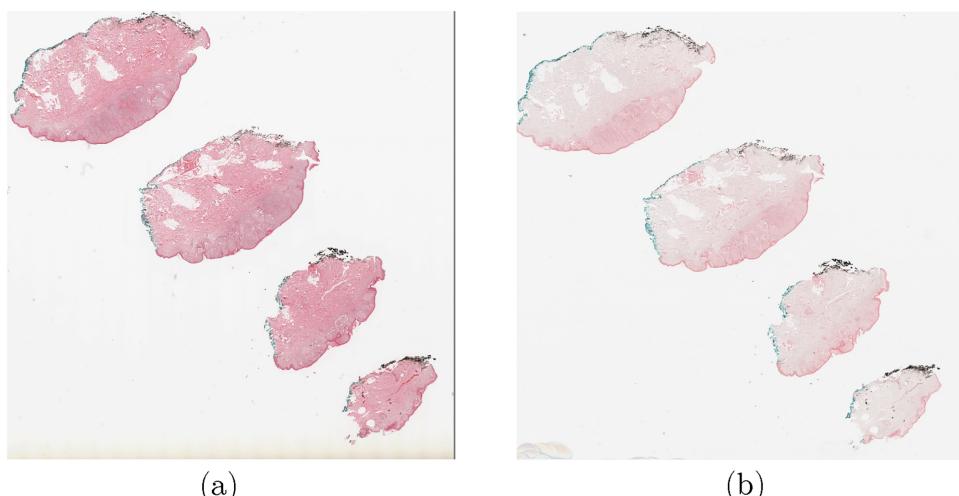


Fig. 8. Digitization of a histological sample that presents effects of fading that lead to color differences: first digitization (a) and second digitization (b) [27].

Table 1

The normalization methods of H&E-stained histological images published between 2008-2017 with their evaluation techniques used for performance analysis through the metrics of normalized mean square error (NMSE), normalized median intensity (NMI), Performance index (PI), receiver operating characteristic (ROC), relative root mean square error (rRMSE), quaternion structural similarity index (QSSIM) and kernel-density estimate (KDE), with indication of studies that used whole slide imaging (WSI) for performance evaluation.

Year	Ref.	Histological samples	Classes	Evaluation classes	Evaluation methods
2008	[32]	Four liver samples with 2000×2000 pixels.	Statistic-based color correction	Estimates of W and H , and colors intensities	Minimization of the NMSE metric between V and $W \cdot H$ by 200 points of nucleus, cytoplasm and red blood cells, and minimization of average difference of Lab colors model between red blood cells of the normalized and reference images.
2009	[33]	Five histological samples of melanoma and nine of benign nevi.	Spectral matching	Applications in histological image processing	Statistical analysis of mean and covariance matrix metrics of hematoxylin, eosin and background regions.
2009	[34]	48 samples of cirrhotic liver digitized with dimensions of 1000×1000 pixels.	Statistic-based color correction	Colors intensities	Minimization of deviation, in degrees, of estimated plane normal vectors and the reference ones.
2010	[35]	Synthetic images and 13 samples of melanocytic lesions with dimensions of 1000×1000 pixels.	Spectral matching	Estimates of W and H	Presentation and description of colors of H&E-stained histological regions.
2012	[15]	Five histological samples of prostate, cervix and colon polyp with high-grade dysplasia with dimensions of 960×960 pixels.	Spectral matching	Visual analysis	Minimization of NMI standard deviation and coefficients of variation for evaluation of nuclear segmentation, minimization of difference of nine percentiles of G channel histogram of segmented nuclei from the normalized and reference images, and visual analysis of nuclear segmentation quality.
2013	[36]	19 prostate images and 26 oropharyngeal images with dimensions of 500×500 pixels.	Statistic-based color correction	Applications in histological image processing, colors intensities and visual analysis	Visual analysis of normalized quality.
2013	[37]	No definition of the used histological samples.	Statistic-based color correction	Visual analysis	Minimization of Euclidean distance between V and $W \cdot H$.
2013	[38]	20 inflamed mice skin images with dimensions of 1300×1300 pixels.	Spectral matching	Estimates of W and H , and computational performance	minimization of the maximum number of iterations, minimization of the execution time of the evaluated algorithms, and maximization of the PI metric.
2014	[16]	50 ovarian serious adenocarcinoma images, 52 glioblastoma multiform images, with dimensions of 1024×1024 pixels, and 102 renal cell carcinoma images with dimensions of 1600×1200 pixels.	Statistic-based color correction	Applications in histological image processing and computational performance	Maximization of pixels classification accuracy into hematoxylin, eosin, erythrocytes and stained-free regions, and minimization of execution time.
2014	[39]	320 images with dimensions of 400×400 pixels extracted from eight histological samples, among liver, breast, lymph, intestine and brain samples.	Statistic-based color correction	Colors intensities and visual analysis	Minimization of perceptual colors difference for normalization evaluation on images obtained from different scanners, and visual analysis of colors distributions of nucleus, cytoplasm and red blood cells.
2014	[40]	135 lymph nodes images with dimensions of 2300×3300 pixels.	Statistic-based color correction	Applications in histological image processing	Minimization of NMI standard deviation and coefficient of variation.
2014	[41]	48 liver images, 12 esophageal images, and 100 breast cancer images.	Spectral matching	Applications in histological image processing	Maximization of pixels classification accuracy into hematoxylin, eosin and background classes, maximization of the Dice coefficient for tumor segmentation, and minimization of execution time.
2015	[8]	Five breast cancer samples.	Spectral matching	Colors intensities	Maximization of Pearson correlation for normalization evaluation on images from different scanners.
2015	[14]	375 lymphoma images with dimensions of 1388×1040 and 58 normal and cancerous breast cases, with dimensions of 896×768 pixels.	Spectral matching	Estimates of W and H	Minimization of standard deviation of W matrix coefficients among images from the same histological sample.
2015	[27]	Synthetic experiments, 13 real images of melanocytic lesions, 23 pairs of cutaneous melanoma images, 31 images of benign nevi and 21 of melanoma.	Spectral matching	Estimates of W and H , applications in histological image processing and colors intensities	Minimization of deviation, in degrees, of estimated plane normal vectors and the reference ones, maximization of accuracy and area under the ROC curve by nuclear features classification, and minimization of earth mover's distance through RGB channels histograms of original and faded samples.
2015	[42]	20 liver images with dimensions of 1280×960 pixels.	Statistic-based color correction	Applications in histological image processing and colors intensities	Maximization of pixels classification accuracy into nucleus, cytoplasm, red blood cells and stained-free regions, and minimization of perceptual colors difference of nucleus, cytoplasm and red blood cells.

(continued on next page)

Table 1 (continued)

Year	Ref.	Histological samples	Classes	Evaluation classes	Evaluation methods
2015	[43]	Six images of human placenta.	Statistic-based color correction	Colors intensities and pathologists evaluation	Minimization of average differences of RGB histograms for images evaluation with illumination variations, and maximization of indices defined by images sorting according to visual analysis of pathologists. Maximization of Pearson correlation and multi-channel normalized mutual information for evaluation of images from different scanners, and maximization of scores from 0 to 5 assigned to H&E stain deconvolution by visual analysis of pathologists.
2015	[44]	650 images of breast cancer.	Spectral matching	Colors intensities and pathologists evaluation	Minimization of rRMSE for evaluation of W estimation by manual marking of regions representative of the H&E stain, maximization of Pearson correlation for evaluation of H estimation in comparison to a H reference map, presentation of 3D histograms for structures preservation evaluation, minimization of execution time, maximization of the Pearson correlation and QSSIM metrics for evaluation of images from different scanners, maximization of scores from 0 to 5 assigned by pathologists visual evaluation for stain deconvolution.
2016	[30]	71 stomach images, 70 of prostate, 70 of colon and 84 of bladder, 650 breast cancer images, and 21 colon WSIs.	Spectral matching	Estimates of W and H , computational performance, colors intensities and pathologists evaluation	Minimization of Euclidean distance between W matrices of reference, obtained from histological regions representative of the H&E stain, and the estimated one by this proposal, minimization of NMI standard deviation and coefficient of variation from segmented nuclei and eosin regions respectively identified in unsupervised and manually approaches, maximization of accuracy and presentation of ROC curve for necrosis detection, and minimization of execution time.
2016	[45]	125 WSIs of lymph node and 10 samples of rat liver.	Statistic-based color correction	Estimates of W and H , applications in histological image processing and computational performance	Maximization of classification accuracy of two subtypes of lung cancer by shape, color and texture features, maximization of Pearson and canonical correlations of these features from histological regions of the original and normalized images.
2016	[46]	260 adenocarcinoma images, 32 non-small cell lung cancer images, 68 squamous cell carcinoma images and 74 other lung cancer tumors.	Statistic-based color correction	Applications in histological image processing	Maximization of the Dice coefficient for segmentation of cell nuclei, minimization of the sum of squared difference between 128 bins histograms of RGB channels for evaluation of samples digitized by different scanners and staining protocols.
2017	[25]	200 breast cancer images with dimensions of 1000×1000 pixels, and 25 images of gastrointestinal cancer with dimensions of 1000×1000 pixels.	Statistic-based color correction	Applications in histological image processing and colors intensities	Minimization of mean and standard deviation of the metrics of KDE and sum of squared differences on 256 bins histograms of RGB channels, minimization of the structural dissimilarity index, and minimization of color volume.
2017	[47]	Five lung cancer blocks sectioned into nine samples divided into five regions of 1000×1000 pixels, totaling 225 images.	Statistic-based color correction	Colors intensities	

two histogram matches and three color transfers, as described below. In this study, Hoffman et al. [16] used the all-pixel quantile method from the histogram matching class that employs a mapping function from the sorted intensity levels of each RGB color channel of the original image to the sorted intensity levels of the corresponding channel of the reference image. Also from this class, the color map normalization technique, presented by Hoffman et al. [16], employed a mapping function by the unique RGB tuples of the original and reference images. Even with the histological information loss by their methodologies, it is worth highlighting that the all-pixel quantile method allows to have a more structural preservation as a result of intensity distributions of image morphological variations. The color map normalization leads to more losses by presenting similar color distributions of images with different contents [48].

Among the color transfer proposals, the studies of Abe et al. [32], Magee et al. [34], Hoffman et al. [16], Bautista and Yagi [42], and Tam et al. [46] used the concepts of the Beer-Lambert law. The color transfer techniques by Hoffman et al. [16] employed a deconvolution step. In order to do so, an empirically defined W matrix was used to obtain the H map by the Beer-Lambert law. After obtaining a color channel corresponding to each dye, the methods of (1) K-means, (2) Gaussian mixture model with expectation–maximization, and (3) Gaussian mixture model with variational Bayesian inference were evaluated for pixels separation between foreground (strong staining) and background (weak staining). With the identification of these regions, one of the color transfers of Magee et al. [34] was applied with the best results achieved by the K-means.

In the work of Bautista and Yagi [42], the H map was calculated by the multiplication between V densities and the inverse of the W matrix, which was defined by the individual staining of histological samples with hematoxylin and eosin and unstained tissue samples. After calculating the H maps from the original and reference images, the minimum and maximum values of each component (hematoxylin, eosin, and red blood cells) were employed to define 21 color intervals equally spaced, which form their look-up tables (LUTs). The color transfer was performed by the mapping between the LUTs color intervals from the original and reference images through a linear interpolation function. The RGB image reconstruction was given by the inverse application of the Beer-Lambert law.

The study of Abe et al. [32] is similar to [42] because it also employed unstained tissue samples and samples separately stained with hematoxylin and eosin to define the W matrix. Abe et al. [32] proposed the color correction of red blood cells as a way to improve previous studies. For this purpose, manual segmentations of the nucleus, cytoplasm and red blood cells in the original and reference images were performed. The relationship between these histological components colors and the H&E stain was established by the W matrix and the H map defined by the least square method. The Beer-Lambert law was modified by weighting factors of hematoxylin, eosin, and red blood cells, and they were defined by the ratio between the average of the H coefficients from the original and reference images. Based on these factors, the original image colors were modified according to the reference image H map.

Like Abe et al. [32], Magee et al. [34] and Basavanhally and Madabhushi [36] employed a histological component segmentation step, but in an unsupervised way. Basavanhally and Madabhushi [36] proposed modeling the original and reference images pixels as Gaussian mixtures of the nucleus, epithelium, stroma, and white regions (lumen and background). The parameters of this model were initially estimated by the K-means on the RGB values and subsequently optimized by the expectation-maximization method. To perform the normalization, the correspondences between the histological components from the original and reference images were obtained by minimizing the Euclidean distances of the average RGB values from each component. The original image colors were altered by evenly-spaced percentiles interpolation from RGB channels distributions from each component in relation to

the reference image.

Magee et al. [34] presented four methods with the application of the normalization proposed by Reinhard et al. [49] to deal with its insufficient preservation of structures. Employing the deconvolution described by Ruifrok and Johnston [50], Magee et al. [34] applied the mean and standard deviation of hematoxylin and eosin channels from the original and reference images for application of the technique of Reinhard et al. [49]. In the second method, Magee et al. [34] identified the two 1D Gaussian mixture models, firstly, performing the deconvolution and, secondly, detecting the foreground and background regions in the hematoxylin and eosin channels, so the colors of such regions could be separately corrected in each dye.

In addition to the normalization by deconvolution, two of Magee et al.'s [34] proposals employed a segmentation step by assigning the pixels to the classes of hematoxylin, eosin, and background. The variations of the method of Reinhard et al. [49] assigned the pixels classes according to either (1) the highest probability obtained by the variational Bayesian Gaussian mixture model (VB-Reinhard-Hard) or (2) by the use of these probabilities as the weighting of mean and standard deviation employed by Reinhard et al. [49] (VB-Reinhard-Weighted).

The assigning of dyes to the pixels was also performed by Bejnordi et al. [40] and Bejnordi et al. [45] through the hue-saturation-density (HSD) color space. The HSD transform modeled the RGB data into chromatic components (c_x and c_y) independent of the stain amount and a density component that presents a linear relation with the amount of each dye. In the method of Bejnordi et al. [45], before the pixels were assigned to the classes of hematoxylin, eosin, and background by the K-nearest neighbors (K-nn) classifier, with $K = 7$, its training set was defined in an unsupervised way. After the K-nn training, the original and reference images pixels were classified by their HSD values. The distribution of chromatic data of each class from the original image was altered by distribution parameters of these data from the reference image. Considering that each pixel is a stain mixture, the new distribution of each class was multiplied by the weight defined by the posterior probability of the Bayesian classifier. These weights were also employed for altering the density component distribution of the original images.

Bejnordi et al. [40] also used the HSD model but with a different assignment of hematoxylin and eosin classes to the pixels. The conversion from RGB images to HSD allowed the application of the singular value decomposition (SVD) on the pixels not identified as background in the $c_x c_y$ plane. Two vectors with the largest singular values were computed and the angles between each data point and the second singular vector allowed to divide these points into the classes of hematoxylin (points with angles below 90°) and eosin (the remaining). The color transfer between the corresponding dyes in the original and reference images was given by the interpolation of percentiles of their $c_x c_y$ planes. To represent each pixel as a stain mixture, the transformed values of c_x and c_y of each dye were multiplied by the posterior probabilities of a Gaussian mixture model with two components that responded to the H&E stain.

The segmentation of regions with no assignment to histological components or dyes was employed by Mosquera-Lopez and Agaian [37], Tam et al. [46] and Janowczyk et al. [25]. In the study of Janowczyk et al. [25], it was presented a normalization method with the use of a deep learning auto-encoder for obtaining a feature space of high level of abstraction. In this work, the auto-encoder parameters were estimated by the stochastic gradient descent method. By these features, the K-means method was applied to cluster the pixels into 50 regions. Finally, the normalization was performed by a histogram matching with 128 bins between each one of the 50 identified classes from the original and reference images.

The method proposed by Mosquera-Lopez and Agaian [37] aimed to correct color variation from the digitization step in natural and histological images. In their proposal, the original and reference images were initially converted to the Lab color model. Afterward, their pixels

were clustered into different regions by the fuzzy c -means membership degrees. After the clusters definition, it was necessary to find their correspondences in the original and reference images by minimizing the Euclidean distances between their centroids. The color transfer was performed by a linear function using the mean and standard deviation of the Lab values of the original and reference images. Besides, the fuzzy c -means membership degrees were employed as weighting coefficients between pixels and clusters, which makes it possible to check the influence of the colors of each cluster on the image pixels.

Tam et al. [46] presented the only color transfer that did not use a reference image. In this regard, the original images were segmented by the superpixels method allied to a deconvolution step by the application of the method proposed by Ruifrok and Johnston [50] who used a W matrix empirically defined. The proposed normalization employed the definition of centroids of each superpixel from the hematoxylin and eosin channels by the average intensity of their histograms. The intensities beyond the centroids were scaled to the interval [128, 255] and the intensities below the centroids were scaled to the interval [0, 128]. To avoid artifacts introduction in the superpixels borders, the contrast-limited adaptive histogram equalization method was applied on regions with 32×32 pixels with at least 75% of its pixels representing tissue, identified by an unsupervised segmentation step.

Like Janowczyk et al. [25], the color transfer of Bug et al. [47] also employed a deep learning method with optimization of their parameters by the stochastic gradient descent. For this purpose, the original and reference images were submitted to convolutions for learning representations that compose a feature extractor of contextual information. The reference image features were then used for color correction by their mean and variance metrics and by shifting and scaling parameters of the network. With the use of this methodology, a training set was necessary for application of the proposed normalization.

Among the color transfer techniques, Bautista et al. [39], and Ţerbănescu and Pleșea [43] used synthetic samples for color adjustment. The normalization of Bautista et al. [39] corrected only the color variation from the samples digitization. For this purpose, the synthetic calibration sample with nine uniform regions of basic and H&E colors was digitized by the same system of the histological sample. Initially, the histological and calibration images were converted to the linear RGB color space by the inverse gamma correction application to correct the exporting file system of the used scanner. The color normalization was performed by multiplying the linear RGB values of the original images and a color correction matrix. This matrix was defined by the least square method between the known colors from the calibration sample and its colors represented by the scanner.

Şerbănescu and Pleşea [43] also used a synthetic sample called standard protein stain marker (SPSM). The synthetic and histological samples were prepared, sectioned, stained, and digitized together, which exposed the SPSM to the same color variations sources of the histological samples. The color correction was performed by the method of Reinhard et al. [49], however, the mean metric was defined by the difference of intensity between the SPSM images with color variations and the one chosen as reference.

4.2. Spectral matching

Among the spectral matching methods, the non-negative matrix factorization (NMF) technique was the most frequent in the literature, and it was employed by Bilgin et al. [15], Saraswat and Arya [38], Li and Plataniotis [14], Vahadane et al. [44,30]. This technique is adequate to estimate W and H since it has a non-negativity restriction that represents that histological densities can only be zero or positive. When represented by negative values, these densities indicate light emission, an invalid biological condition [30].

The work of Vahadane et al. [30] is an extension of the method in [44], with new quantitative evaluations, more details about the proposed normalization, and its application on WSI. In these studies, the

estimates of W and H were given by the NMF. In addition to the non-negativity restriction, these proposals employed the sparsity concept, which considers that histological image pixels colors are a sparse mixture of the H&E stain. This condition was represented by a sparsity parameter (λ) multiplied by the H map. The H estimation was called sparse coding, which was performed by the least angle regression (LARS) method with an implementation based on Cholesky; and the W estimation was called dictionary learning, which was made by the parameter free block-coordinate descent with warm restart technique. In both estimates, the sparse modeling software (SPAMS) toolbox [51] was employed in order to implement the proposed normalization. After these estimates for the original and reference images, the original image H map was combined with the reference image W matrix. The application of this method on the WSIs was performed by the division of these images into grids. Then, the W matrices were estimated for these regions and combined by the median of their coefficients for obtaining the W matrix of the WSI. In this application, the H map was defined as proposed by Ruifrok and Johnston [50]. After these estimates for the original and reference images, the normalization was performed in the same way as images with lower dimensions.

Besides the NMF application, Bilgin et al. [15], and Saraswat and Arya [38] employed initialization techniques of the H map. Bilgin et al. [15] presented an initialization of this map by operations on the pixels RGB values so the eosin pink color, represented by small coefficients in the R channel, could be expressed in the H initialization. This work also estimated the camera dark current term (D), but no detail about the criteria for this application was shown. The estimates of the W , H and D matrices were alternately performed on 15 iterations. For structures like red blood cells could be represented in the color adjustment results, the combination of the W , H and D matrices was subtracted from the original image to obtain the residual components in the R , G and B channels. The color adjustment was performed by the gamma correction applied to the H map. For this application, the adjusted image was obtained by the combination between the new H map and the W and D matrices. Finally, the proposed method result was given by the addition of the residual components to the adjusted image.

Saraswat and Arya [38] proposed the H initialization by the a and b channels from the Lab model. Besides the H initialization, this study also presented an empirical initialization of the W matrix. After the NMF application, the hematoxylin and eosin dyes from the original and reference images were obtained in the RGB color model. The color transfer between the corresponding dyes of these images was performed by the conversion of the RGB model to the uncorrelated colour space [52] with posterior application of the method of Reinhard et al. [49] on each dye channel. After the color adjustment, the original image dyes channels were again converted to the RGB model and combined by its addition.

Like Saraswat and Arya [38], Li and Plataniotis [14] also presented an initialization of W and the application of NMF. The initialization of W by Li and Plataniotis [14] was employed to guarantee the convergence to the global optimal solution of NMF and limit the influence of noise and poorly stained pixels. For this purpose, a saturation-weighted hue histogram generated by the HSV channels was employed where null weighting coefficients were assigned to the poorly stained pixels. After obtaining this histogram, the K-means method was applied for identification of H&E representative hues. By these hues, the S and V values from the HSV model were defined and converted to the optical density space, therefore composing the W matrix. In addition to the spectral normalization by replacing the original W by the reference W , Li and Plataniotis [14] presented an illumination normalization step. The resulting image was given by multiplying each original image channel by the ratio between the background intensities from the corresponding channels of the original and reference images. This correction was conditioned to background intensity levels higher than 200, so only images with background could be corrected in this step.

Different from the NMF, Macenko et al. [33], Niethammer et al.

[35] and Vicory et al. [27] presented the estimation of W by a plane defined by stain vectors, corresponding to the W matrix columns. Vicory et al. [27] presented an extension of the method of Niethammer et al. [35] with new applications and validations. The proposal of Niethammer et al. [35] represented an extension of the normalization of Macenko et al. [33], but with its limitation improved. Macenko et al. [33] aimed to define the limits of pixels color distributions, which corresponded to the W matrix columns, by the SVD calculation with the V densities. To avoid the poorly stained pixels compromising the W estimation, pixels with densities below 0.15 were removed. A plane was created by the two largest singular values obtained by the SVD. This plane was employed for projecting V , with the posterior calculation of the angles in relation to the first SVD direction. The 1st and 99th percentiles of the angles set were calculated and converted to the V space to obtain the W matrix. To establish the equivalence in Eq. (3), it was possible to determine H after defining W .

The W estimation presented by Niethammer et al. [35] and Vicory et al. [27] proposed an enhancement of the study of Macenko et al. [33], who supposed proportional amounts of the H&E stain in the images. In the W estimation, the Beer-Lambert law was used for the image colors to be interpreted by its distribution on a subset of optical density plane defined by the plane fitting algorithm. However, to guarantee the good performance of this proposal on insufficient amount of one or both dyes, an energy function was minimized to define the plane by weighting coefficients represented by the number of points of each dye defined by the Otsu method. This optimization was performed by alternating between the optimal plane definition and the classification points into hematoxylin and eosin. After the plane identification, the W matrix was defined by the percentiles of the aforesaid classes.

In the W estimation proposed by Celis and Romero [8], the images were initially converted from the RGB space to the $R - B, G$ plane. After defining sub-planes along the G axis, the $R - B$ mean values of each sub-plane were calculated, defining the so-called color characterization curve. In addition to defining the reference image, the color characterization curve allowed to identify the hematoxylin and eosin colors in its distribution by the minimum and maximum values of the $R - B$ means, respectively. Identifying these points in the curve, it was possible to determine the RGB values of each dye, which defined the W coefficients.

Among the spectral matching techniques, the work of Khan et al. [41] employed a manual definition of a training set for the W matrix estimation. The training and test images were quantized by the Oct-tree method and resulted in histograms with 255 colors. Subsequently, these histograms were reduced to a r dimension by the mean values and the covariance eigenvectors of the training set histograms. This low dimension color representation was called principal color histograms (PCH) and the r variable represented the stain color description (SCD). To estimate the W matrix, the classes of hematoxylin, eosin, and background were assigned to the test images pixels by the relevance vector machine (RVM) trained by the RGB pixels values and the SCD of the manually defined training set. After the images division into hematoxylin, eosin, and background, the mean RGB values of each one of these regions allowed to obtain the W matrix. The hematoxylin, eosin, and background channels from the original and reference images were divided into the classes of stained regions, background, and others by the probabilistic metrics of the RVM model. By the optical densities of these regions, the metrics of mean, and 5th and 95th percentiles, which were intuitively chosen, were calculated and employed for mapping the original and reference images colors by the B-spline method. The normalized image was reconstructed through the combination of the hematoxylin, eosin, and background channels with the W matrix.

5. Evaluation of H&E-stained histological images normalization techniques

The evaluation methods of normalization can be divided into (1) estimates of W and H , (2) applications in histological image processing, (3) computational performance, (4) colors intensities, (5) pathologists evaluation, and (6) visual analysis. The evaluation methods of normalization published in the literature are summarized in Table 1. In this table, normalization performance evaluations are categorized according to the division defined by this paper, with a brief description of the employed methods.

The evaluation of the W and H estimates was performed by Niethammer et al. [35] and Vicory et al. [27] through deviations, in degrees, between the estimated plane normal vector and the reference one. For evaluation of these estimates, Saraswat and Arya [38] calculated the Euclidean distance between the values of the V densities and $W \cdot H$. Thus, the lower this distance, the better the estimates.

For the calculation of NMSE between the V densities and $W \cdot H$, Abe et al. [32] employed 200 points of nuclei, cytoplasm, and red blood cells. Li and Plataniotis [14] evaluated the W estimation among images extracted from the same histological sample by the standard deviation of their coefficients. By this metric, the lower the standard deviation, the more robust the W estimation is since it is expected that the considered images matrices are similar in their H&E-representing RGB.

The studies of Bejnordi et al. [45] and Vahadane et al. [30] were similar as they evaluated the W estimation using manual marks in the H &E-stained histological regions. With that, the W reference matrix was defined by the median of the R, G and B channels of the hematoxylin and eosin regions. Bejnordi et al. [45] employed the Euclidean distance between the W reference and its estimation, and Vahadane et al. [30] employed the metric of rRMSE. Vahadane et al. [30] also evaluated the H estimation by the manual marking, which allowed to obtain the H reference map by the method proposed by Ruifrok and Johnston [50]. In this evaluation, the Pearson correlation index was employed to quantify the similarity between the H reference map and its corresponding estimation by the proposed method.

The evaluation of normalization performance in histological image processing was performed by the segmentation and classification steps. Macenko et al. [33], Vicory et al. [27] and Tam et al. [46] evaluated their normalization performances by classifying images of different types of cancer. Macenko et al. [33] did not present quantitative results, but only the distributions of shape and color features from nuclear regions. Vicory et al. [27] presented the results of images classification by the distance-weighted discrimination method through the metrics of accuracy and area under the ROC curve using nuclear features extracted from the normalized images. Also with the accuracy metric, Tam et al. [46] presented their results by the least absolute shrinkage and selection operator regression model using features of shape, color, and texture from normalized images regions.

The pixels classification into dyes was evaluated by Hoffman et al. [16] and Khan et al. [41] through the accuracy metric. In the study of Hoffman et al. [16], they took into consideration the hematoxylin, eosin, erythrocyte and stain-free regions. Khan et al. [41] used only the assignments of hematoxylin, eosin, and background. In both evaluations, the results were obtained by comparing to reference marks. The accuracy metric was also applied by Bautista and Yagi [42], who evaluated the pixels' classification into nucleus, cytoplasm, red blood cells, and free-stain regions.

The cell nuclei segmentation was a common evaluation among the works of Basavanhally and Madabhushi [36], Bejnordi et al. [40], Khan et al. [41], Bejnordi et al. [45] and Janowczyk et al. [25]. Among them, the studies of Basavanhally and Madabhushi [36] and Bejnordi et al.

[40,45] employed the metric of NMI to evaluate the color constancy of the segmented nuclei. For this segmentation, Basavanhally and Madabhushi [36], and Bejnordi et al. [40] employed empirical thresholding values on the HSI color model and on grayscale images, respectively. Bejnordi et al. [45] employed the combination between fast radial symmetry transform and marker-controlled watershed with the removal of false positive nuclei by area and shape criteria. In addition to the evaluation by the nuclei NMI metric, Bejnordi et al. [45] also investigated the color constancy of the eosin regions identified by manual marks. The difference between the NMI metrics employed by these studies was that Bejnordi et al. [45] proposed altering its calculation, so noise pixels would not be considered. With this modification, the NMI metric of Basavanhally and Madabhushi [36], and Bejnordi et al. [40] (Eq. (4)) was updated to Eq. (5) [45]:

$$\text{NMI} = \frac{\text{median}(I\{D\})}{\max(I\{D\})}, \quad (4)$$

$$\text{NMI} = \frac{\text{median}(I\{D\})}{P_{95}(I\{D\})}, \quad (5)$$

where, $I\{D\}$ corresponds to the intensity set of segmented regions and P_{95} represents the 95th percentile. The minimization of standard deviation and coefficient of variation (standard deviation divided by mean) of the NMI in image sets indicates that the color distributions of nuclear and eosin regions are similar, which makes the segmentation consistent.

Khan et al. [41] and Janowczyk et al. [25] evaluated, respectively, the segmentation of tumors and cell nuclei by the Dice coefficient in comparison with manual marks of pathologists. Bejnordi et al. [45] evaluated the normalization application in the CAD systems to detect necrotic tissues in the WSIs. For this purpose, the metrics of accuracy and area under the ROC curve were used with evaluation of training and test sets variations.

In addition to the images classification, Tam et al. [46] also evaluated their method with the Pearson and canonical correlations of shape, color, and texture features from histological regions of the original and normalized images. These metrics allow to quantitatively determine the consistency of these features. The higher the Pearson correlation, the more similar the features are between the images pairs. The canonical correlations quantify how much the features set of the original images is linearly related to the set of the normalized images [46].

The computational performance of normalization was evaluated by Saraswat and Arya [38], Hoffman et al. [16], Khan et al. [41], Bejnordi et al. [45] and Vahadane et al. [30]. To compare different NMF techniques, Saraswat and Arya [38] used the number of iterations of these algorithms and time elapsed, in seconds, for their executions. Furthermore, Saraswat and Arya [38] employed the so-called PI defined by the combination of the mean error, the number of iterations, and time elapsed. The time elapsed for execution was also employed by Bejnordi et al. [45] and Vahadane et al. [30] to compare the performance of methods already published in the literature and different approaches of WSIs normalization, respectively. Hoffman et al. [16] and Khan et al. [41] also used the time elapsed metric in their evaluation.

The evaluation of normalization by color intensities of the RGB and Lab models was the most frequent method of analysis used among the studies depicted in Table 1. By the Lab color model, Abe et al. [32], Bautista et al. [39], Bautista and Yagi [42] and Bug et al. [47] quantified the normalization performance. Abe et al. [32] employed the color mean difference metric, in the Lab model, of red blood cells identified in the reference and normalized images. The smaller this value, the higher the similarity among the intensities of the regions of the considered images, which indicates a better performance of the proposed normalization.

The Euclidean distance between the Lab color model channels intensities, called perceptual color difference, was employed by Bautista

et al. [39] and Bautista and Yagi [42]. Bautista et al. [39] used this metric for evaluation of histological samples digitized by different scanners, and Bautista and Yagi [42] applied it to evaluate the color differences of nucleus, cytoplasm, and red blood cells. Images pairs with perceptual color differences lower than 2.2 are considered perceptually similar [39]. By regions evaluation, the perceptual color difference between objects lower than 3 indicates that their colorimetric features are indistinguishable to the human eye [42].

Bug et al. [47] employed the metrics of mean and standard deviation of KDE and the sum of squared differences between the RGB color channels histograms with 256 bins. The structural dissimilarity index was also used to quantify the perceptual dissimilarity by comparing statistics of the first and the second orders in the original and normalized images. The Lab color model was also employed to define the metric called color volume, corresponding to the product among channels' standard deviations of this model. This metric was proposed to quantify the perceptible colors in the normalized images, a condition reached by its minimization.

Magee et al. [34], and Basavanhally and Madabhushi [36] employed the identification of specific regions comparing their colors in the RGB model. Magee et al. [34] statistically evaluated their results with the mean and covariance metrics of hematoxylin, eosin and background representations. Basavanhally and Madabhushi [36] presented quantitative results by the difference of nine percentiles of the G channel histograms of segmented nuclei. The lower this difference, the higher the similarity of the intensity levels of nuclear regions of the considered images.

The normalization evaluation of histological images from two digitizations of each sample by different scanners was performed by Celis and Romero [8], Vahadane et al. [44,30], and Janowczyk et al. [25]. With the same content, these images were evaluated with the Pearson correlation and the multi-channel normalized mutual information by Vahadane et al. [44], and with the Pearson correlation and the QSSIM in the study of Vahadane et al. [30]. Celis and Romero [8] employed the Pearson correlation for this evaluation as well. The color differences from different digitization were also quantified by Janowczyk et al. [25] through the sum of the squared difference of RGB channels histograms with 128 bins. After normalization, it was expected that this result had a small value, as indicated by high color similarity between the images. This metric was also used by Janowczyk et al. [25] to evaluate samples with color variation from different staining protocols.

In order to evaluate digitization with illumination differences, Serbănescu and Pleșea [43] employed the mean differences of RGB histograms of images with these variations and the corresponding reference illumination. Vicory et al. [27] also used the RGB color channels histograms to get the measures of earth mover's distance. By this metric, Vicory et al. [27] evaluated the normalization of histological images characterized by fading of the tissue samples.

The normalization can also be evaluated by pathologists, as performed by Serbănescu and Pleșea [43] and Vahadane et al. [44,30]. Serbănescu and Pleșea [43] proposed this evaluation with six pathologists by an ascending ranking of the original and normalized images according to vague criteria of appearance and details. Applying such an approach, scores were obtained and allowed to quantify this evaluation. The evaluation of Vahadane et al. [44,30] by pathologists qualitatively analyzed the separation of dyes obtained by the normalization technique. In these studies, two pathologists assigned scores from 0 to 5 for the stain separation, but details about the criteria employed were not described.

Among the normalization evaluation, visual analysis was employed by Bilgin et al. [15], Basavanhally and Madabhushi [36], Mosquera-Lopez and Agaian [37], and Bautista et al. [39]. In these works, no quantitative evaluation was used, which makes their validations not objective. Bautista et al. [39] presented the color distributions of nucleus, cytoplasm, and red blood cells by the Lab color model. Bilgin et al. [15] presented and described the colors of histological regions

that had chemical affinities to the hematoxylin and eosin dyes. In this study, the color preservation of white regions, red blood cells, brown pigments, and the colors of folded tissue samples was also evaluated by visual analysis. Mosquera-Lopez and Agaian [37] limited their evaluation by the visual analysis of color similarities of the normalized images. Also by visual criteria, Basavanally and Madabhushi [36] qualitatively compared the nuclear segmentation step between the original and normalized images. This evaluation demonstrated that the normalization allowed more consistent results in the segmentation than the original images.

5.1. Complexity analysis of H&E-stained histological images normalization techniques

A possible way to compare the described normalization methods is by their computational complexity. In this study, we performed this analysis through the information provided by the very own authors in the considered papers on this topic, so that any misleading conclusion could not be obtained. A deeper investigation of the normalization algorithms complexity is still an open issue, and since it is an extremely difficult task it can be performed in future researches.

According to Niethammer et al. [35], the Otsu thresholding method employed in their methodology is efficient since its computations are performed only in the image histogram. Based on this, despite their information loss, histogram matching techniques can be considered simple because they also employ only the histograms representation for the normalization.

In the study of Hoffman et al. [16], the clustering algorithms of *K*-means, Gaussian mixture model with expectation–maximization, and Gaussian mixture model with variational Bayesian inference were compared to segment the images pixels into foreground and background. According to the authors, their computational complexities were significantly different with a more complex algorithm identified in the variational Bayesian method.

Through the LUTs structures, Bautista and Yagi [42] identify their color transfer proposal with a low computational complexity, since it did not employ any complex pattern recognition algorithm. In contrast, Janowczyk et al. [25] indicate their normalization with a deep learning auto-encoder as a computationally burdensome method. By also using a deep learning method, Bug et al. [47] indicate that a more efficient network structure with less parameters can replace the employed network architecture. Finally, among the color transfer techniques with any reference to its computational complexity, Bejnordi et al. [45] explain the use of the naive Bayesian classifier because it is a simple method with no parameters.

Among the spectral matching methods, the NMF was a popular technique employed by many normalization proposals. According to Bilgin et al. [15], the NMF is a NP-hard problem. However, the authors described that different NMF algorithms have different accuracy and complexity, related to their number of iterations. For this reason, it is very difficult to compare NMF normalization methods if they are not clearly defined and described, as in [14]. Then again, we present the complexity analysis of spectral matching techniques based on only the indications of the very own authors of the papers.

One of the NMF techniques that had its performance analyzed was the WSI normalization technique of Vahadane et al. [30]. This analysis indicated that the acceleration scheme proposed in that paper was successfully by using a parallel processing on different CPU cores. The last NMF proposal that was analyzed by its authors was the study of Saraswat and Arya [38]. In that paper, Saraswat and Arya [38] indicated that their NMF initialization method was able to reduce the computational complexity in comparison to the use of random initializations of the *W* matrix and the *H* map. Besides, Saraswat and Arya [38] compared different NMF algorithms, and it is possible to observe a relevant performance difference between them, which corroborates the NMF analysis of Bilgin et al. [15].

Still considering the spectral matching class, Khan et al. [41] described the RVM classifier as a computationally efficient method with a better computational performance than the support vector machine classifier. Among the spectral matching techniques that did not estimate the *W* matrix by the NMF, Macenko et al. [33] and Niethammer et al. [35] were the only studies that briefly analyzed their algorithms. Macenko et al. [33] indicated that their method has a better computation time than NMF methods. Niethammer et al. [35] indicated the good performance of the employed Otsu method since it is based on the histogram representation, which makes it efficient.

6. Discussion and future directions

Taking into consideration the publications summarized in Table 1, it is possible to observe that spectral matching methods were more popular than the color transfer techniques only in 2015. The representation of biological and H&E stain properties by spectral matching techniques can be considered an important factor for the investigation of this class. Besides, the great challenges imposed by the segmentation and deconvolution in color transfer methods make their development complex, which also promotes the investigation of spectral matching as a new alternative for the normalization. However, with both methodologies, challenges are still observed, and new studies can be presented for H&E-stained histological image normalization. In color transfer techniques, it is necessary to guarantee that the steps of segmentation and deconvolution reach robust results, even when applied to images with great color variations. The methods of spectral matching pose challenges by the search for robust *W* matrix and *H* map estimates.

One of the limitations of the spectral matching works is the convergence of the *W* and *H* estimates to local minima, as indicated by Khan et al. [41] in the study of Niethammer et al. [35]. Among the proposed works, possible solutions for this limitation are only presented for the NMF methods. A proposal to avoid the local convergence of the *H* estimation is presented by Bilgin et al. [15], who applied the Gaussian filter in each iteration to remove noises that could lead to this undesirable convergence. For the global convergence of the *W* estimation, Li and Plataniotis [14] presented an initialization of this matrix specific to each image, and with the removal of achromatic pixels that could represent noise and regions poorly stained.

The NMF convergence employed by Bilgin et al. [15] was limited by a maximum number of iterations, which can be insufficient for application on large images datasets. As the NMF convergence would be faster, Saraswat and Arya [38] presented the initialization of the *W* matrix, but with empirical coefficients, without the singularity of the method of Li and Plataniotis [14]. Furthermore, to enhance the NMF convergence, Saraswat and Arya [38] showed an initialization method of the *H* map. With the statistical analyses of the elapsed time, the number of iterations necessary for convergence and the mean error performed by Saraswat and Arya [38] demonstrated a significant difference between the application of random initializations and the proposed ones, thus encouraging the use of automatic initializations for the NMF methods.

The use of the NMF techniques for spectral matching allows to represent biological properties of stain densities by the non-negativity restriction. With this concepts in mind, normalization techniques that do not use NMF methods can present negative values in their estimates, which makes their biological interpretation complex [30]. Vahadane et al. [30] indicated that the *W* estimation by Macenko et al. [33] could present negative coefficients, and the work of Bautista and Yagi [42] was also susceptible to negative values in the *H* map. Thus, the use of the NMF methods has important biological properties, but for being classified as a NP-hard problem [15], its high computational complexity must be considered in the development of new methods.

Besides the application of the Beer–Lambert law for establishing the relationship between the dyes absorbed by the histological structures and their colors, the HSD color model [40,45] can also be employed in

new proposals. The representation of microscopic images by this model is better-suited than the RGB and the HSI color models [45], leading to its use in normalization methods.

The segmentation of the color transfer methods presents limitation in supposing whether the original and reference images have all the defined histological regions and small color variations [45]. These limitations can be improved by the methodology employed by Janowczyk et al. [25], who defined a large number of regions to make this step robust to images with great color variations, different from the Gaussian mixture employed by some studies [45]. The pixels classification into hematoxylin and eosin dyes must also consider the great color variations in its methodology, so its performance would not be affected by the disproportion of the dyes, as indicated by Bejnordi et al. [45].

For reproducibility of methods and results, it is necessary that satisfactory amount of H&E-stained histological samples should be evaluated. Macenko et al. [33], Bejnordi et al. [40], and Tam et al. [46] suggested that their methods could be applied to different types of tissues. In contrast, Bautista and Yagi [42] indicated that it is necessary to check the reproducibility of their results on images of different tissues from the ones employed for evaluation.

Although Bejnordi et al. [45] proposed the WSIs normalization, a great portion of their evaluation was performed on regions selected in these images. According to the authors, it was necessary to use this selection to compare their normalization with other methods of the literature. However, Vahadane et al. [30] presented an algorithm for application of the normalization on the WSIs that allowed different normalization techniques to be employed and compared. Therefore, optimization methods for the WSIs normalization should still be investigated.

The evaluation of normalization should be extended to color variations from the sample preparation. Khan et al. [41], and Bautista and Yagi [42] evaluated their normalization on samples stained with different stain batches. Abe et al. [32], Tam et al. [46], Bug et al. [47] and Janowczyk et al. [25] used images from adjacent sections of histological samples with different intensities of the hematoxylin and eosin dyes. Abe et al. [32] defined the H&E intensity variation using different staining time of samples, and Bug et al. [47] used different staining solution concentrations, as well as different tissue thickness.³ Tam et al. [46] and Janowczyk et al. [25] did not present such details.

The color variations from the digitization step should also be evaluated in the normalization, as performed by Bautista et al. [39], Khan et al. [41], Celis and Romero [8], Vahadane et al. [44,30], and Janowczyk et al. [25], who used different digitization of each histological sample in such an evaluation. Janowczyk et al. [25] used three digitization from one scanner and one from a second evaluated scanner, which allowed them to analyze intra and inter-scanners color variations. The color variations from the digitization presented by Ţerbănescu and Pleșea [43] were given by different exposure times to the light, which led to illumination differences.

In the study of Bejnordi et al. [40], a statistically significance of color differences was observed by variations of the laboratory, which means differences of sample preparation protocols, and the day when this preparation is performed. However, no statistical difference was observed in images from different patients' tissues. Thus, it suggests that the normalization methods should be evaluated on samples from different laboratories and days, as used by Bejnordi et al. [40,45], Khan et al. [41], Bautista and Yagi [42], and Li and Plataniotis [14].

The normalization evaluation on faded samples was performed only by Vicory et al. [27]. In this work, two digitization of melanoma samples were performed with differences between two and seven years. By the use of this dataset,⁴ we can conclude that the development of

normalization methods for color correction of faded samples is still an open issue.

Some studies presented an insufficient evaluation of their results. Bilgin et al. [15] and Mosquera-Lopez and Agaian [37] did not use quantitative metrics, making their evaluation subjective. In contrast, the study of Vahadane et al. [30], who use the 3D histogram, allowed to verify the structure preservation obtained by its normalization. In spectral matching methods, the consistency of color representation of the H&E stain in histological regions can be evaluated by the deconvolution defined by the *H* estimation.

Despite its rare use, the evaluation performed by pathologists represents a significant normalization performance verification. Among the described studies, only Ţerbănescu and Pleșea [43] and Vahadane et al. [44,30] employed this type of evaluation for visual quality analysis. To evaluate red blood cells normalization [32], estimation of the *W* matrix [30,45], and the *H* map [30], a manual marking of histological regions was also performed, so their colors could be evaluated by quantitative metrics.

Bautista et al. [39] presented limitations for colors correction of red blood cells. The work of Abe et al. [32] is dedicated to solve this correction, and Bilgin et al. [15] also presented a consistent color representation method of these cells, macrophages, and brown pigments. In the study of Bilgin et al. [15], the red blood cells colors were preserved. Thus, it is possible to investigate the colors correction of red blood cells from the original images according to a reference image by the residual components proposed by Bilgin et al. [15].

Different from the H&E stain concentration preservation by the techniques of Vahadane et al. [44,30], Bejnordi et al. [45] proposed the HSD density channel alteration, which is linearly related to the stain concentration. Bilgin et al. [15] and Bautista and Yagi [42] also presented stain concentration alterations. These alterations could modify the histological information representation, which harms the preservation of structures in the normalization results.

Among the limitations of the proposed normalization methods are the manual definitions of parameters and the *W* matrix coefficients, preparation of synthetic samples, and the use of training sets. Vahadane et al. [44,30] employed a manual assignment to the λ parameter for representation of the stain sparsity property. By the sensitivity analysis of this parameter [30], its automatization should be investigated. Synthetic samples used by Bautista et al. [39], and Ţerbănescu and Pleșea [43] also represented a great limitation for practical applications of these normalization techniques.

In the color transfer methods of Hoffman et al. [16], and Tam et al. [46], empirical coefficients defined by Ruifrok and Johnston [50] were used for the *W* matrix definition, which obtained the hematoxylin and eosin channels in the deconvolution step. Performing such a methodology, the H&E stain representation in the RGB color model can be inadequate in datasets composed of images with great color variations, which makes the empirical *W* matrix definition poorly robust. Moreover, the *W* coefficients of the methods defined by Abe et al. [32] and Bautista and Yagi [42] were obtained by samples separately stained by the hematoxylin and eosin dyes. Besides, Abe et al. [32] also required the identification of nucleus, cytoplasm, and red blood cells for applying their normalization.

Regarding the *W* matrix estimation, the studies of Magee et al. [34] and Khan et al. [41] demanded the use of manual markings of the H&E-stained regions and training images, respectively. The manual interaction employed by Magee et al. [34] to estimate the stain vectors limits its application in large images datasets [27,45]. By the use of training sets, Khan et al. [41] presented limitations to color variations of the images that composed these sets [8,30,45], as well as the color transfer of Bug et al. [47]. Due to the great difficulty in composing H&E image datasets with all possible color variations [30], normalization methods that do not use a training step can be more robust.

Some studies indicated limitations considering structures preservation and artifacts introduction in the evaluation of different

³ Available at: <https://github.com/stes/fan/releases>.

⁴ Available at: <http://midas3.kitware.com/midas/folder/11138>.

- [37] Mosquera-Lopez C, Agaian S. Iterative local color normalization using fuzzy image clustering. *SPIE defense, security, and sensing* 2013;875518.
- [38] Saraswat M, Arya KV. Colour normalisation of histopathological images. *Comput Methods Biomech Biomed Eng Imag Visual* 2013;1:185–97.
- [39] Bautista PA, Hashimoto N, Yagi Y. Color standardization in whole slide imaging using a color calibration slide. *J Pathol Inform* 2014;5.
- [40] Bejnordi BE, Timofeeva N, Otte-Höller I, Karssemeijer N, van der Laak J. Quantitative analysis of stain variability in histology slides and an algorithm for standardization. *Proc Med Imag Digital Pathol* 2014;9041.
- [41] Khan AM, Rajpoot N, Treanor D, Magee D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans Biomed Eng* 2014;61:1729–38.
- [42] Bautista PA, Yagi Y. Staining correction in digital pathology by utilizing a dye amount table. *J Digital Imag* 2015;28:283–94.
- [43] erb nescu MS, Plesea IE. A hardware approach for histological and histopathological digital image stain normalization. *Rom J Morphol Embryol* 2015;56:735–41.
- [44] Vahadane A, Peng T, Albarqouni S, Baust M, Steiger K, Schlitter AM, et al. Structure-preserved color normalization for histological images. 12th international symposium on biomedical imaging, 2015, p. 1012–5.
- [45] Bejnordi BE, Litjens G, Timofeeva N, Otte-Höller I, Homeyer A, Karssemeijer N, et al. Stain specific standardization of whole-slide histopathological images. *IEEE Trans Med Imag* 2016;35:404–15.
- [46] Tam A, Barker J, Rubin D. A method for normalizing pathology images to improve feature extraction for quantitative pathology. *Med Phys* 2016;43:528–37.
- [47] Bug D, Schneider S, Grote A, Oswald E, Feuerhake F, Schüler J, et al. Context-based normalization of histological stains using deep convolutional features. Deep learning in medical image analysis and multimodal learning for clinical decision support 2017:135–42. Springer.
- [48] Kothari S, Phan JH, Moffitt RA, Stokes TH, Hassberger SE, Chaudry Q, et al. Automatic batch-invariant color segmentation of histological cancer images. *IEEE international symposium on biomedical imaging: from nano to macro*. 2011. p. 657–60.
- [49] Reinhard E, Adhikmin M, Gooch B, Shirley P. Color transfer between images. *IEEE Comput Graph Appl* 2001;21:34–41.
- [50] Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quantit Cytol Histol* 2001;23:291–9.
- [51] Mairal J, Bach F, Ponce J, Sapiro G, Jenatton R, Obozinski G. Sparse modeling software. 2017 (accessed 26.10.2017). <http://spams-devel.gforge.inria.fr/index.html>.
- [52] Liu C. Learning the uncorrelated, independent, and discriminating color spaces for face recognition. *IEEE Trans Inform Foren Secur* 2008;3:213–22.
- [53] Sethi A, Sha L, Vahadane AR, Deaton RJ, Kumar N, Macias V, et al. Empirical comparison of color normalization methods for epithelialstromal classification in H and E images. *J Pathol Inform* 2016;7.