Stain standardization capsule for application-driven histopathological image normalization

Yushan Zheng, Zhiguo Jiang*, Haopeng Zhang, Fengying Xie, Dingyi Hu, Shujiao Sun, Jun Shi, and Chenghai Xue

Abstract

Color consistency is crucial to developing robust deep learning methods for histopathological image analysis. With the increasing application of digital histopathological slides, the deep learning methods are probably developed based on the data from multiple medical centers. This requirement makes it a challenging task to normalize the color variance of histopathological images from different medical centers. In this paper, we propose a novel color standardization module named stain standardization capsule based on the capsule network and the corresponding dynamic routing algorithm. The proposed module can learn and generate uniform stain separation outputs for histopathological images in various color appearance without the reference to manually selected template images. The proposed module is light and can be jointly trained with the application-driven CNN model. The proposed method was validated on three histopathology datasets and a cytology dataset, and was compared with state-of-the-art methods. The experimental results have demonstrated that the SSC module is effective in improving the performance of histopathological image analysis and has achieved the best performance in the compared methods.

Index Terms

Stain standardization, Histopathological image analysis, Digital pathology, Computational pathology, Capsule network

I. Introduction

Based on the widespread application of digital pathology (DP) in cancer research and clinical diagnosis, an increasing number of methods for histopathological image analysis (HIA) have been developed [1], [2], [3], [4], [5]. In practice, the color appearance of digital whole slide images (WSI) varies due to the diversity in the section fabrication and digitization procedures. The variation makes it a challenging task to establish robust analysis frameworks based on digital histopathological images. Generally, stain standardization (or normalization) is the way to solve the problem of color variances [6], [7], [8], [9].

Recently, the data-driven deep-learning methods, especially the convolutional neural network (CNN) [10], [11], have become the major basis of emerging HIA researches [2], [12], [13], [14], [15]. Correspondingly, the research on data standardization has been focusing on improving the effectiveness and robustness of CNN frameworks. To train a universal model for cancer diagnosis of a certain organ, the histopathological WSIs are now collected from multiple hospitals or medical centers [16]. Meanwhile, the trained model is required to be deployed back to multiple hospitals for sub-region prediction or whole slide image analysis. The color appearance of WSIs is diverse across hospitals or even within the same hospital but in different periods. The dynamic color variations have increased the difficulty of CNN frameworks in learning pathological patterns and thereby raised the demand for good color normalization algorithms.

In this paper, we proposed a novel stain standardization module for CNN-based histopathological image analysis, which is named as stain standardization capsule (SSC). The main structure of the module is modified from the *Capsule Network* [17]. The approach to stain standardization is realized by delivering uniform stain separation results for images in different stain styles. In contrast to existing methods [18], [19], [20] which solve a specific stain separation problem for each image or WSI, we propose learning a group of stain separation parameters based on the color distribution of the entire training data. In the application stage, the most appropriate parameter for stain separation for each image is online selected through dynamic routing (DR) operation, which makes the model more robust than the previous methods. The proposed SSC module was evaluated on three histopathology datasets and a cytology dataset, and compared with six typical methods for stain normalization. The experimental results have demonstrated the effectiveness and scalability of the proposed method.

The remainder of this paper is organized as follows. Section II reviews the related works for histopathological image normalization. Section III presents the methodology of the proposed method. The experiments and corresponding discussions are presented in Section IV. Section V summarizes the conclusion.

This work was supported by the National Natural Science Foundation of China (grant No. 61901018, 61771031, 61906058 and 61471016), the China Postdoctoral Science Foundation (Grant No. 2019M650446) and Tianjin Science and Technology Major Project (Grant No. 18ZXZNSY00260). *Asterisk indicates the corresponding author.*

Y. Zheng, Z. Jiang, H. Zhang, F. Xie, D. Hu, and S. Sun are with Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China, and also with Image Processing Center, School of Astronautics, Beihang University, Beijing 102206, China. E-mail: ({yszheng, jiangzg}@buaa.edu.cn). J. Shi is with School of Software, Hefei University of Technology, Hefei 230601, China. Chenghai Xue is with Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China and also with Wankangyuan Tianjin Gene Technology, Inc, Tianjin, 300220, China.

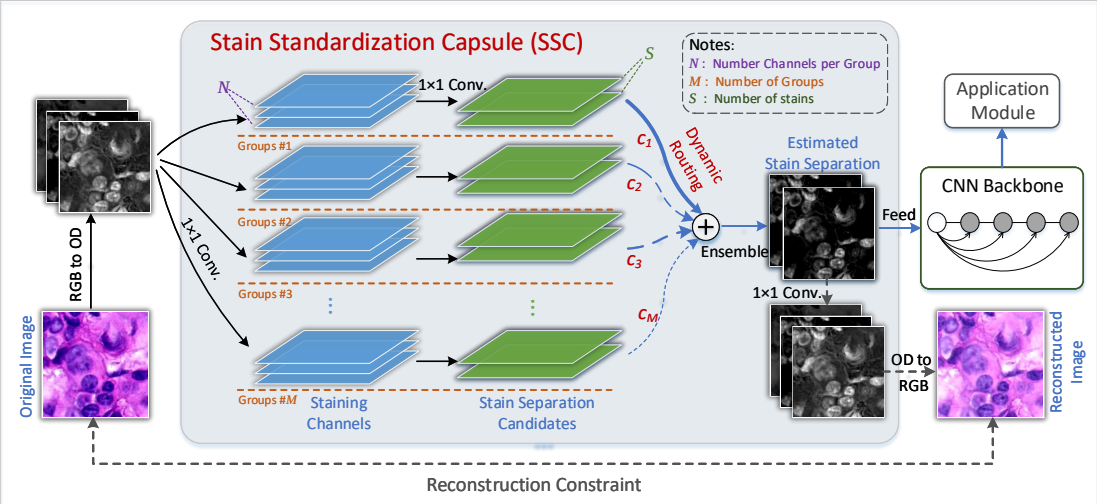


Fig. 1. Structure of the proposed SSC module, where the input RGB-format images are first converted to the optical density space, then projected into M groups of S stain channels via linear transformations, and finally assembled to obtain the stain separation results via the designed sparsity routing algorithm.

II. Related works

The approaches to solving the problem of color variations can be roughly categorized as *Stain transformation*, *Color domain transformation*, and *Color augmentation*. The typical methods of these categorizes are reviewed in this section.

A. Stain transformation

Stain transformation in this paper indicates the normalization approach achieved by converting the color of an image to the template image(s). The early studies referred to the color style transformation methods [21], [22] in natural scene image processing. Then, the prior knowledge of tissue staining, especially the hematoxylin-eosin (HE) staining, is utilized to enhance the stability of normalization [23], [24], [25], [26], [27]. Specifically, the independent stain components are extracted from RGB-format images based on stain separation algorithms and then matched to the stain distribution of the template image(s) that are manually selected by pathologists. The essential of these methods is to precisely estimate the parameters of the stain separation model for both the input image and the template image. Nevertheless, the information in an image patch could not cover the staining phenomena of the entire tissue section, which usually caused misestimation of stain parameters and thus delivered inaccurate normalization results. With the development of whole slide imaging techniques, the objects of computer-aided diagnosis have changed from histopathological image patches to WSIs. Simultaneously, the stain transformation algorithms for WSIs are developed [19], [18], [28], [20]. The stain separation parameters were estimated or optimized with sufficient pixels sampled from the entire WSI. Therefore, the robustness of stain normalization was greatly improved compared to the patch-based transformation methods. However, the dependence of abundant pixels and whole slide images meanwhile narrowed the scope of application.

B. Color domain transformation

Color domain transformation is another popular approach for histopathological image data normalization. It derives from the study on style transfer [29] in the domain of natural scene images, especially the method based on generative adversarial networks (GANs) [30]. Color domain transformation is designed for the situation where the deep learning models are trained within the data from a single hospital but required to deploy to another hospital [31], [32], [33], [34]. Instead of estimating transform parameters between image pairs or WSI pairs, these methods established a GAN structure to learn the data allocations of the training dataset and the application (testing) dataset. Then, the domain transformation can be completed with the generative networks. The performance of stain transformation has proven very promising. Nevertheless, the present GAN-based methods require to know the full data distribution of the application dataset (or testing dataset) and the transform models are required to be trained in pairs if more than two medical centers provide the data. When the training dataset consists of samples from multiple centers, GAN-based methods are inconvenient in deployment. Recently, Tellez et al. [35] constructed a U-Net model to learn a color correction model where a set of manually selected images are regarded as the ground truth and the same images with randomly performed color biases are used as the input of the network. The trained network is powerful to adjust the various color biases in the images. However, the network contains millions of parameters, which makes it less efficient in computation.

C. Color augmentation

Color augmentation is also an effective approach to handle color variations. Liu et al. [36] performed random color biases in multiple image factors including the brightness, saturation, and contrast to improve the color scalability of CNN-based applications. Considering the characteristic of pathological images, Tellez et al. [37] proposed to produce stain augmentation based on the color deconvolution [38] algorithm to simulate different staining situations. However, the stain information is extracted using fixed model parameters that estimated under the ideal dyeing case. When facing the samples in non-ideal staining styles, the augmented samples would be out of the distribution of real color.

D. The novelty of the proposed method

The novelty of the proposed method to the existing methods can be summarized as follows.

1) We bring the insight of capsule network into histopathological image standardization. Beyond optimizing the normalization parameters for a specific image (or WSI) [24], [19], [18], [20], the stain standardization is achieved by optimizing the forward routes within the trained networks via a designed sparsity routing process. It prevents the standardization results from serious artifacts or even failures.

2) The proposed normalization is a template-free method. Instead of mapping the color of images to manually selected template images [26], [27], [19], [18], [20] or relying on the entire allocation of the testing dataset [32], [33], [34], the proposed SSC module searches for several candidate ways to stain separation only based on the training data that involves various color appearance, and extracts uniform stain components for the following CNN applications in the testing stage. The standardized data is more objective than that from template-based methods.

2) The SSC module is essentially a neural network, which can be seamlessly equipped to the CNN backbones. Therefore, it can be jointly trained with specific HIA tasks and can directly provide uniform input tensors for CNN in the forward propagation. Furthermore, the module contains only tens of parameters, which is much lighter than the CNN-based normalization methods [35], [31], [32]. These properties determine the SSC module is easy-to-use in both the development and deployment of HIA applications.

A preliminary version of this work has been reported [39]. In this paper, we provide more comprehensive reviews of the related works, detail the methodology, provide discussions, conduct extensive experiments to validate the essential structures and hyper-parameters of the SSC module, and further assess the effectiveness of the model in a practical gatric dataset and a cytology dataset.

III. Method

In the proposed SSC module, the stain standardization is achieved by generating uniform stain separation results for images in various color appearances. The structure of the module derives from *Capsule Network* [17], which is first introduced in brief. Then, the methodology of the SSC module is detailed.

A. Capsules network

Capsule Network is a new paradigm of artificial neural networks [17], in which the input and output of the neurons are vectors (as shown in Fig. 2), rather than scalars in traditional neural networks. While the steps to active the capsule neurons are similar with the traditional neurons. Specifically, a set of vectors $\{\mathbf{u}_i \in \mathbb{R}^d | i = 1, 2, ..., M\}$ are first weighted, then assembled by a sum operation and finally activated by an operator named *squash* to obtain the output $\mathbf{v}_j$.

As for the input $\mathbf{u}_i$ is vector, the weighting is defined as a vector operation with a trainable matrix $\mathbf{w}_{ij} \in \mathbb{R}^{d' \times d}$:

$$\hat{\mathbf{u}}_{j|i} = \mathbf{w}_{ij}\mathbf{u}_i. \tag{1}$$

In the algorithm, the weighted vectors $\hat{\mathbf{u}}_{j|i}$ are regarded as multiple candidate routes that could pass the neuron and the goal of DR is to enhance the route that contains discriminating information and weaken the route that is less important to report. Specifically, the routing is realized by a set of scalars $\{c_i | i = 1, 2, ..., M\}$ that are initialized as $c_i = 1/M$, the output of the neuron is defined as

$$\mathbf{v}_j = squash(\mathbf{s}_j), \quad \mathbf{s}_j = \sum_{1}^{M} c_i \hat{\mathbf{u}}_{j|i}, \tag{2}$$

where $squash$ is the activation function that is unfold as

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2} \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}. \tag{3}$$

Next, the authors [17] proposed to assembling the weighted vectors through a *Dynamic Routing* (DR) algorithm. An agreement score $b_{ij} = 0$ is initialized for each route before the routing. Then, in each iteration, the agreement score is updated based on the inner product between $\hat{\mathbf{u}}_{j|i}$ and the output $\mathbf{v}_j$ via equation

$$b_i = b_i + \mathbf{v}_j \cdot \hat{\mathbf{u}}_{j|i}. \tag{4}$$

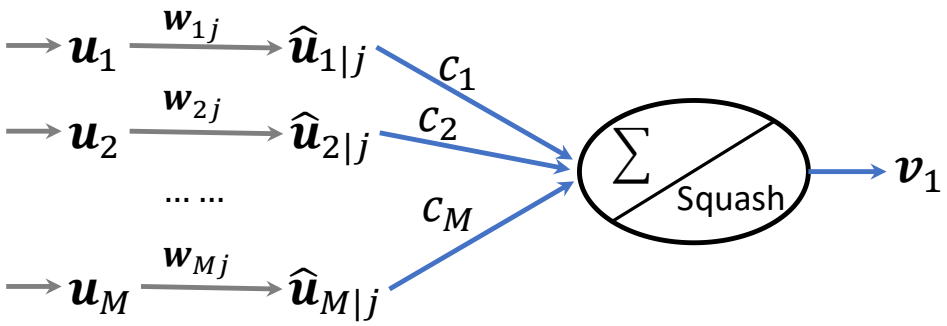


Fig. 2. The structure of a neuron in capsule network, where the $\mathbf{u}_i$ is an input vector of this neuron and also the output of the i-th neuron in the previous layer and $\mathbf{v}_j$ represents the output vector.

Next, the scalars for ensemble are calculated with equation

$$c_i = exp(b_i)/\sum_k exp(b_k), \tag{5}$$

i.e., $\{c_1, c_2, ..., c_M\}$ is the softmax of $\{b_1, b_2, ..., b_M\}$. After several iterations of routing, the final output vector is obtained.

B. Stain separation candidates

Motivated by the insight of capsules network, we proposed to constructing parallel routes based on neural networks to propose multiple candidates of stain separation for an input image, and then assembling the stain candidates $\{\hat{\mathbf{u}}_{i|j}|i = 1, ..., M\}$ through dynamic routing. The routing aims to search for the most appreciate stain separation result for each specific image. The detail of the SSC structure is illustrated in Fig. 1.

The parallel networks for stain separation is constructed based on the theory of *Color deconvolution* (CD) [38], which is widely utilized in popular stain standardization methods [24], [18], [20]. Referring to [38], independent stain components can be extracted through linear transformations in the optical density (OD) space. Therefore, we proposed to transform the RGB-format image into optical density space and then extract separation candidates using 1 by 1 convolution operations with near linear activation functions. Specifically, the separation candidates are obtained by grouped convolution layers. Letting $\mathbf{I} \in \mathbb{R}^{m\times n\times 3}$ denote a histopathological image in size of $m \times n$ pixels and $\mathbf{o} \in \mathbb{R}^{m\times n\times 3}$ denote the image in optical density space, the projection between $\mathbf{I}$ and $\mathbf{o}$ is formulated as

$$\mathbf{o} = -\log(\mathbf{I} + \epsilon)/I_{max}, \tag{6}$$

where I_{max} is the upper intensity for the digitization and ϵ is a small scalar to protect the log operation. Then, the grouped convolution layers can be represented as

$$\mathbf{u}_i = Conv(\mathbf{o}, \mathbf{W}_i^{(1)}) \in \mathbb{R}^{m\times n\times N}, \tag{7}$$

$$\hat{\mathbf{u}}_i = Conv(\mathbf{u}_i, \mathbf{W}_i^{(2)}) \in \mathbb{R}^{m\times n\times S}, i = 1, 2, ..., M, \tag{8}$$

where $Conv$ represents the convolution followed by an activation function, $\mathbf{W}_i^{(1)}$ and $\mathbf{W}_i^{(2)}$ are the convolutional weights, and M, N and S (see Figure 1) denotes the number of groups, the number of channels in each group and the number of stains involved in the images, respectively. The first layer is designed to extract potential staining channels from the images and the second layer is used to summarize the independent stain components based on these potential channels.

Considering that stain standardization should not change the structure of histopathological images, the spatial sizes of the convolution kernels are constrained to 1×1. It is equivalent to linear projections of feature maps. Correspondingly, the activation function is set as Leakly-Relu to maintain the linear property of the network and meanwhile restrain the negative values.

C. Sparsity routing

A good stain separation should assign the power of a pixel value to one stain channel, i.e. the separated result is desired to be pixel-wise sparse [24], [19], [20]. Therefore, we modified the agreement scoring in the dynamic routing as sparsity scoring and designed a novel *Sparsity Routing* (SR) algorithm. Specifically in each iteration, the weight of the route that proposes

pixel-wise-sparse results will be enhanced in the ensemble. The pseudo-code of SR is given in Algorithm 1 and the score of the pixel-wise sparsity is calculated based on the sparseness measure defined in [40]:

$$\eta_p(\mathbf{x}) = \frac{1}{mn}\sum_{ij}\frac{\sqrt{S} - \sum_k |x_{ijk}+\epsilon| / \sqrt{\sum_k (x_{ijk}+\epsilon)^2}}{\sqrt{S}-1}, \tag{9}$$

where $\mathbf{x} \in \mathbb{R}^{m\times n\times S}$ denotes the tensor to score. To avoid all the data being assigned to a single stain channel, a channel-wise sparseness is additionally defined:

$$eta_c(\mathbf{x}) = \frac{1}{S}\sum_{k}\frac{\sqrt{mn} - \sum_{ij} |x_{ijk}+\epsilon| / \sqrt{\sum_{ij} (x_{ijk}+\epsilon)^2}}{\sqrt{mn}-1}. \tag{10}$$

Finally, the sparsity score is formulated as

$$\eta(\mathbf{x}) = \eta_p(\mathbf{x}) + \eta_c(\mathbf{x}) \tag{11}$$

and referred as $SparseScore(\mathbf{x})$ in Algorithm 1. After SR, the output of SCC is calculated by equation

$$\mathbf{s} = \sum_i^M c_i \cdot \hat{\mathbf{u}}_i. \tag{12}$$

The SR process allows SSC generating refined stain separation results by adjusting the weights $\{c_i\}$ and then allows the following CNNs concentrate on the structural variances of tissue images.

Algorithm 1: Sparsity Routing.

Input:
$\{\hat{\mathbf{u}}_i | i = 1, ..., M\} \leftarrow$ The outputs of the candidate layer.
$R \leftarrow$ The number of iteration.
Output:
$\mathbf{s} \leftarrow$ The stain separation result.

```
for i = 1 to M do
    b_i ← 0, c_i ← 1/M;
end
for r = 1 to R do
    ŝ ← Σ_i c_i · û_i;
    for i = 1 to M do
        b_i ← b_i + SparseScore(û_i + ŝ);
    end
    for i = 1 to M do
        c_i ← exp(b_i) / Σ_i exp(b_i);
    end
end
s = Σ_i c_i · û_i;
return s;
```

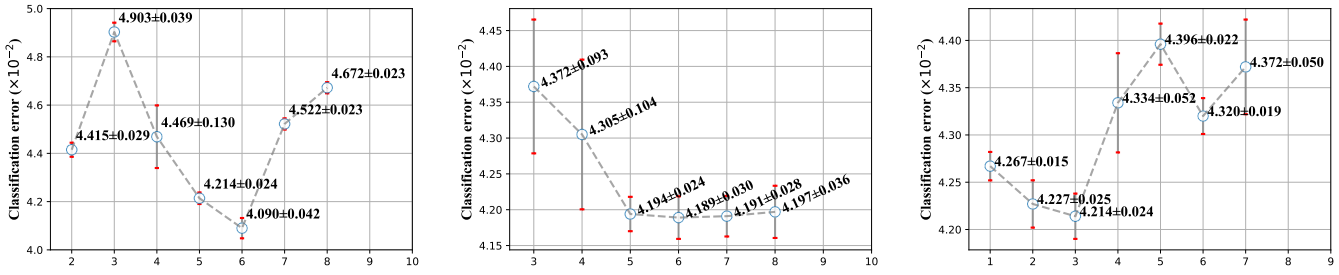
D. Training and Application

The SSC module is essentially a convolutional neural network. The assembled stain separation result $\mathbf{s}$ is the output of the SSC module and meanwhile the input of the following CNN. Therefore, it can be directly equipped to an application-driven CNN and jointly trained along with the objective of the CNN. To improve the generalization of $\mathbf{s}$, a reconstruction layer is appended to the end of SSC and a mean square error (MSE) loss is considered between the original image and the reconstructed results. The MSE loss is added to the loss of the following CNN in the training stage. Note that the SR only processes in the forward stage and the scalar c_i is constant in the backward stage [17].

IV. Experiment

A. Experimental settings

The proposed SSC module was mainly validated via image classification tasks on two public datasets involving histopathology whole slide images. The profiles of the datasets are provided.



(a) The group number M. (b) The width of group N. (c) The iterations of routing R.

Fig. 3. The curves of mean classification errors reported by 5-fold cross-validation as function of the hyper-parameters of SSC, where the standard deviations are marked by red bars.

- *Camelyon16* [41] contains 400 H&E-stained breast lymph node WSIs, in which 270 WSIs are used for training and the remainder are used for testing. [1]
- *ACDC-LungHP* [42] contains 150 H&E-stained lung cancer WSIs, in which 120 WSIs are randomly sampled for training and the remainder are used for testing. [2]

The regions with cancer in the WSIs are annotated by pathologists. Image patches in size of 224×224 were randomly sampled from the WSIs. The patches containing above 75% cancerous pixels according to the annotation were labeled as positive, the patches containing none cancerous pixels were labeled as negative and the other patches were not used in the experiments. In total, $315,662$ / $83,431$ patches were generated from Camelyon16 dataset for training/testing and $115,899$ / $30,157$ training/testing patches were generated from ACDC-LungHP dataset.

The histopathological image classification application based on the DenseNet-121 [11] backbone was used as the benchmark. The sensitivity, specificity, accuracy and the area under the receiver operating characteristic (ROC) curve for classification were the metrics.

We first conducted 5-fold cross-validation to verify the structure of the SSC module and meanwhile tuned the hyper-parameters with the training set, according to the classification performance. Specifically, the WSIs in the training set were randomly divided into 5 parts and the 5 parts were sequentially regarded as the validation data with the other parts treated as the training data. Then, the mean validation error of the 5-fold evaluations for a certain hyper-parameter value was calculated and recorded as the result for the hyper-parameter value. Finally, the optimized hyper-parameter value was determined according to the lowest mean validation error. Note that the other hyper-parameters were set fixed when one hyper-parameter was tuned. After the determination of hyper-parameters, we trained the CNN model equipped the SSC module based on all the WSIs in the training set and applied the trained model to the testing set to evaluate the normalization performance for unseen histopathological images.

All the experiments were conducted in python with pytorch and run on a computer with double Intel Xeon E5-2670 CPUs, 128 Gb RAM and 4 GPUs of Nvidia GTX 2080Ti.

B. Hyper-parameter setting

The paradigm for the hyper-parameter setting is the same for the two histopathological image datasets and the detail on the Camelyon16 dataset is taken as the instance in this section. The image classification error was applied as the metric in this experiment. The essential parameters involved in our method are the number of groups M, the width of grouped convolution layers N and the number of iteration R in the routing. The results are presented in Fig. 3.

1) The number of groups: The parallel networks implemented by the group convolution are used to learn multiple stain separation functions. The number of the group M is set according to the complexity of color appearance within the dataset. Specifically, M should be large enough to make the module cover the number of color styles and while should not introduce redundant routes which would reduce the efficiency of the module. We tuned the M from 2 to 8 and observed the metric. The result in Fig. 3 (a) indicates that $M = 6$ is adequate to cover the color in Camelyon16. Therefore, 6 parallel networks were used in the following experiment.

2) The width of first convolution layers: As the group number was determined, the width (i.e. the number of channels) of the first convolution layer N in each group was selected. The first layer can be regarded as a buffer that provides stain component proposals to the following output layer and N decides the number of proposals. In this experiment, N was tuned from 3 to 8. The results are shown in Fig. 3 (b). It presents that the performance of the module tends to be stable when N is

[1]https://camelyon16.grand-challenge.org/

[2]https://acdc-lunghp.grand-challenge.org/. Since the annotations of testing part of the data set are not yet published, only the 150 training WSIs of the data were used in this paper.

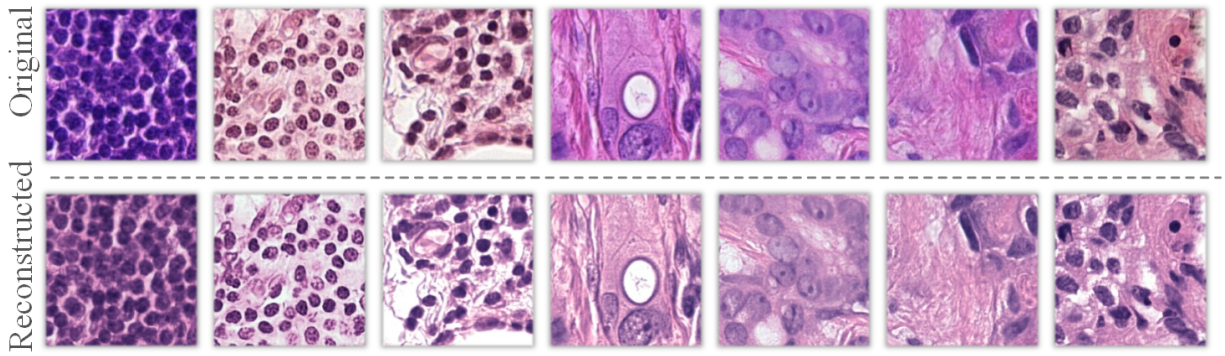


Fig. 4. Joint display of the original images and the reconstructed images.

TABLE I
RESULTS FOR THE ABLATION STUDY ON THE CAMELYON16 DATASET.

Methods	Classification metric			
	Sensitivity	Specificity	Accuracy	AUC
Oringinal	0.851	0.969	0.910	0.957
SSC w/o od	0.860	0.971	0.916	0.967
SSC w/o $\mathbf{u}$	0.861	0.972	0.919	0.966
SSC w/o RC	0.891	0.962	0.927	0.970
SSC w/o n_p	0.848	**0.975**	0.912	0.957
SSC w/o n_c	0.863	0.974	0.918	0.966
SSC	**0.902**	0.970	**0.936**	**0.974**

set above 5. To pursue a low classification error and meanwhile maintain a low computation of the module, we set $N = 5$ in the following experiment.

3) The iteration of routing: The sparsity routing is achieved by iteration operations. As the number of iteration R increases, the scalars $\{c_i\}_1^M$ will become sparser, i.e. the module will tend to report the signal of a specific route and restrain the signals of the other routes in the output. A larger R will highlight the route that delivers the most appreciate stain separation result and will also cause information losses when the correct stain separation for an image relies on the signals from multiple routes. The phenomenon can be quantitatively reflected by the results in Fig. 3 (c). Finally, for better classification performance, we set $R = 3$ in the following experiment.

C. Ablation study

We conducted ablation studies to verify the effectiveness of the model factors. Five degradation models were evaluated in this experiment. These models are introduced as follows.

- SSC w/o od: The input of the module is the RGB-format image rather than the optical density.
- SSC w/o $\mathbf{u}$: The first convolutional layer is removed and the grouped convolution operations are directly performed on the input data.
- SSC w/o RC: The reconstruction layer along with the reconstruction constraint are abandoned.
- SSC w/o η_p: The pixel-wise sparsity $\eta_p(\cdot)$ is not considered in the sparsity routing algorithm.
- SSC w/o η_c: The channel-wise sparsity $\eta_c(\cdot)$ is not considered in the sparsity routing algorithm.

The quantitative comparison for the degradation models and the complete SSC module are shown in Table I. It can be seen that the accuracy and AUC decrease when any of the four components of SSC are removed. The results have demonstrated the necessity of the first convolutional layer, the reconstruction constraint, and the designed sparsity measurement. Particularly, the performance of the model seriously reduced when the primary measurement η_p was removed. It indicates that the sparsity routing is essential in the proposed model. The SSC module will degrade to normal neural networks when the sparsity routing is removed from the structure. The effect of SSC is almost equivalent to deepen the following CNN. Therefore, the performance achieved by SSC w/o n_p could not significantly surpass that without color normalization (Original).

D. Comparison with other methods

In this experiment, four state-of-the-art methods [18], [37], [35], [20] for improving the color scalability of CNNs are compared. Besides, the performance for the universal color augmentations (including random scale in illumination, saturation, hue, and contrast) used in [36] is also compared.

TABLE II
COMPARISON OF THE STANDARDIZATION PERFORMANCE FOR HISTOPATHOLOGICAL IMAGE CLASSIFICATION ON THE CAMELYON16 DATASET.

Methods	**Classification metric**			
	Sensitivity	Specificity	Accuracy	AUC
Original	0.851	0.969	0.910	0.957
Color Aug. [36]	0.868	0.950	0.909	0.958
Stain Aug. [37]	0.875	0.946	0.911	0.967
SNMF [18]	0.877	0.962	0.919	0.965
ACD [20]	0.892	0.944	0.918	0.968
U-Net-Norm [35]	0.875	**0.970**	0.922	0.971
SSC (Ours)	**0.902**	**0.970**	**0.936**	**0.974**

TABLE III
COMPARISON OF THE STANDARDIZATION PERFORMANCE FOR HISTOPATHOLOGICAL IMAGE CLASSIFICATION ON THE ACDC-LUNGHP DATASET.

Methods	**Classification metric**			
	Sensitivity	Specificity	Accuracy	AUC
Original	0.822	0.779	0.801	0.882
Color Aug. [36]	0.836	0.760	0.798	0.881
Stain Aug. [37]	0.819	0.778	0.799	0.882
SNMF [18]	0.827	0.780	0.803	0.883
ACD [20]	0.836	0.776	**0.805**	0.886
U-Net-Norm [35]	0.821	**0.788**	0.804	0.886
SSC (Ours)	**0.840**	0.778	**0.805**	**0.887**

1) Results on Camelyon16 dataset: The classification performance in the Camelyon16 testing set is presented in Table II. Overall, the results have shown that our method is the most effective in improving the performance of histopathological image classification. Stain augmentation [37] utilized the prior knowledge of slide staining to augment the color allocation of training images. Therefore, the classification network using [37] performed better than that using the universal color augmentations [36]. However, the method would generate images with unreasonable color styles. Therefore, the classification accuracy cannot surpass ACD, SNMF, and SSC that searched for specific stain separation principles for different images [20].

Figure 4 illustrated original images and the corresponding reconstructed images from the Camelyon16 dataset. Without any template images, SSC appears to have learned a uniform stain style in the reconstruction layer for images in diverse color appearance. It indicates a standard representation in the SSC output layer. The standard representation has allowed the following CNN concentrating on structural discrimination in histopathological images and thus has improved the performance of the HIA application.

2) Results on ACDC-LungHP dataset: The evaluation of the ACDC-LungHP dataset followed the same paradigm with that on the Camelyon16 dataset. The hyperparameters were tuned as $(M, N, R) = (4, 3, 4)$. The results of the compared methods for image patch classification are summarized in Table III. As the color consistency in ACDC-lungHP dataset is much better than that in the Camelyon16 dataset, the performance of data standardization appears to be less effective in this dataset than that in the Camelyon16 dataset. Nevertheless, the experiment on this dataset raised a discussion about the robustness and reliability of the color normalization methods. Generally, a good color normalization method should not decay the performance of the application when the samples are already in high color quality. From the results in Table III, the methods *Color Augmentation* [36] and *Stain Augmentation* [37] did not improve the performance of classification or even decreased the accuracy and AUC compared to the baseline (Original). The probable reason was that the augmentation produced training images in the color that had significantly deviated from the testing images. In this situation, the performance of augmentation is similar to introduce artificial noise into the training set, which has improved the difficulty of the following CNN in the recognition of tumor tissues. In contrast, the other four methods (*SNMF*, *ACD*, *U-Net-Norm*, and *SSC*) are developed to produce standard input images for the CNNs. The classification results have been further improved by these methods. It demonstrates the four methods are more robust in the task of color normalization for histopathological images.

E. Assessment on practical gastric dataset

To comprehensively assess the normalization performance of different methods in the practical applications, we collected a large dataset from a telepathology platform. The dataset contains 673 H&E-stained gastric WSIs from 673 clinical cases submitted by hospitals around China. All the cases were diagnosed by exports on the platform. The collected dataset consists of 5 typical gastric lesions, including *Low-grade intraepithelial neoplasia* (LGIN), *High-grade intraepithelial neoplasia* (HGIN), *Mucinous adenocarcinoma* (MA), *Adenocarcinoma* (A.), *Signet-ring cell carcinoma* (SRCC). All the lesion regions have been annotated by pathologists. Figure 5 displays several WSIs randomly selected from the gastric dataset. It exhibits serious variances in image color.

In this experiment, 202 WSIs (about 30%) from the 673 WSIs were used as the testing set, and the remainder WSIs were used as the training set. Note that there is no intersection in hospitals of the WSIs in the training set and testing set. Therefore,

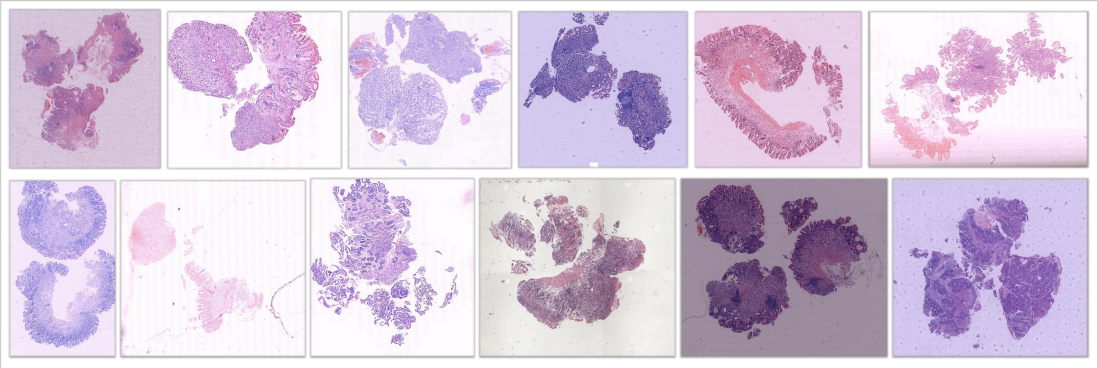


Fig. 5. Display of WSIs from the gastric dataset, where the randomly selected WSIs are presented.

TABLE IV
COMPARISON OF THE STANDARDIZATION PERFORMANCE FOR HISTOPATHOLOGICAL IMAGE CLASSIFICATION ON THE GASTRIC DATASET, WHERE THE BEST VALUE FOR EACH METRIC IS PRINTED AS BOLD AND THE SECOND VALUE IS UNDERLINED.

Methods	**Cancer Identification**				**Lesion Classification**						
	Sensitivity	Specificity	Accuracy	AUC	Normal	LGIN	HGIN	MA	A.	SRCC	Total
Original	0.702	0.883	0.836	0.891	0.810	0.647	**0.237**	0.379	0.684	0.268	0.680
Color Aug. [36]	0.749	0.883	0.848	0.905	0.891	0.556	0.181	0.358	0.607	0.244	0.683
Stain Aug. [37]	0.772	0.873	0.847	0.909	0.886	0.584	0.163	0.382	0.681	0.290	0.701
SNMF [18]	0.763	0.884	0.852	0.910	**0.905**	0.531	0.144	0.297	0.623	0.180	0.682
ACD [20]	0.778	0.873	0.848	0.912	0.861	0.624	0.197	**0.386**	0.664	0.257	0.694
U-Net-Norm [35]	**0.844**	0.814	0.822	0.901	0.882	0.623	0.173	0.376	0.647	0.263	0.698
StainGAN [32]	0.800	0.855	0.841	0.908	0.851	0.629	0.155	0.378	**0.710**	**0.314**	0.697
SSC (Ours)	0.791	**0.898**	**0.867**	**0.927**	0.877	**0.681**	**0.237**	0.374	0.662	0.311	**0.716**

the color profiles in the testing set were unseen in the modeling stage for the normalization methods. In this situation, we additionally assessed a domain transfer model, StainGan [32] in this experiment, besides the methods considered in the previous sections.

The basic experimental settings are the same as those on the Camelyon16 and ACDC-LungHP datasets. Consequently, 270,892 training patches and 118,734 testing patches were extracted for classification experiments. All the compared methods were evaluated on two classification tasks, which were designed as follows.

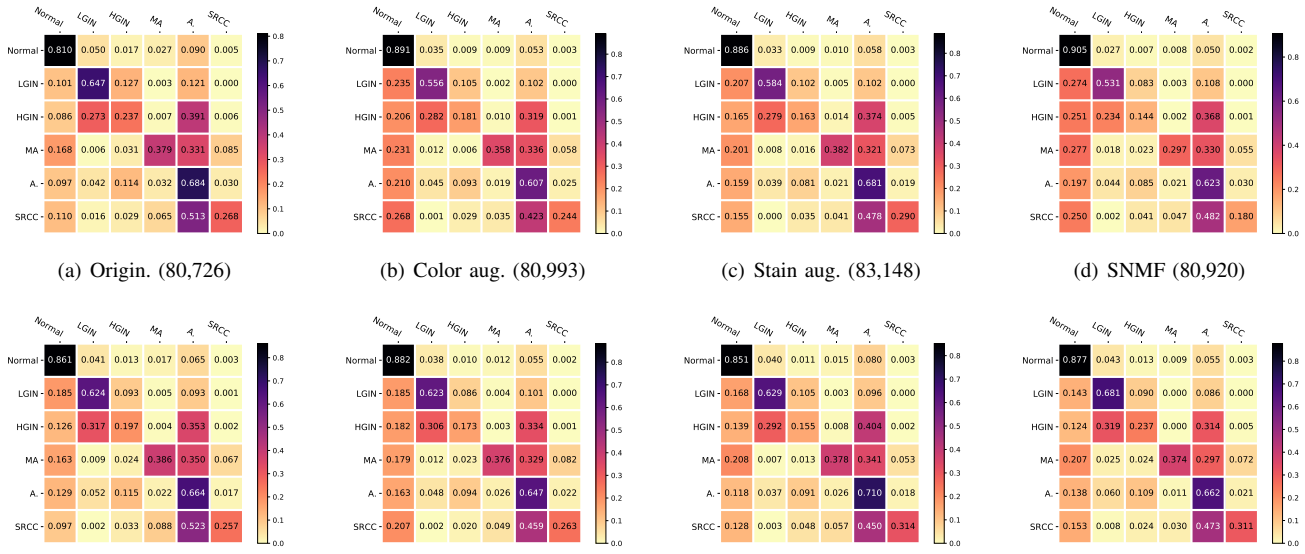
- Cancer Identification: The patches sampled within the annotations of benign lesions (including the LGIN and HGIN) and normal regions are labeled as negative samples and the patches from cancerous annotations (including A., MA, and SRCC) are labeled as positive samples. The CNN with a binary classification output is trained to identify the cancerous patches.
- Lesion Classification: The patches are labeled as 5 classes according to the lesion types of the regions the patches sampled. And the patches from the normal regions are labeled as an additional class. Then, we trained CNNs with 6 output neurons for sub-type classification of gastric lesions.

The hyper-parameters for each compared method are tuned within the training set and the testing set was normalized with the optimized model. For the SSC module, the hyper-parameters were tuned as $(M, N, R) = (6, 3, 4)$. The classification CNN models were trained with different color normalization methods within the training set and applied to the testing images for evaluation. Specifically for the method StainGAN [32], we trained the model to transfer the color profile of the testing set to that of the training set. Then the CNN model trained with color augmentation [36] method was employed to classify the transferred testing images.

The results for the two classification tasks are provided in Table IV and the confusion matrices for the lesion classification task are presented in Figure 6. Overall, the proposed SSC model achieved the best normalization performance, which improved the accuracy of gastric cancer identification by 3.1% and the accuracy of gastric lesion classification by 3.6%. The result has demonstrated the proposed method is effective in practical histopathological images.

Notice that the accuracy improvement achieved by the compared methods on the gastric dataset is more significant than that on the Camelyon16 or ACDC-LungHP dataset. The main reason is that the color variation in the gastric dataset is much more complex than those on the other two datasets, which has increased the difficulty of the CNN model in extracting the pathology patterns from the images. Obviously, the color normalization has successfully relieved the confusion caused by color variations. The result has indicated the color normalization is even necessary for current deep learning models developed for practical histopathological image analysis systems.

The GAN-based methods, e.g. StainGan [32] are designed for the situation where the model is trained by data from one



(a) Origin. (80,726) (b) Color aug. (80,993) (c) Stain aug. (83,148) (d) SNMF (80,920)

(e) ACD (82,318) (f) CNN (82,872) (g) StainGAN (82,736) (h) SSC (84,895)

Fig. 6. Comparison of confusion matrices for lesion classification on the gastric dataset with different normalization methods, where the number of correctly classified images in the testing set is given in the brackets.

TABLE V

COMPARISON OF MODEL PROPERTIES OF DIFFERENT STAIN NORMALIZATION METHODS, INCLUDING THE NUMBER OF MODEL PARAMETERS (n_{param}), WHETHER TO RELY ON MANUALLY SELECTED TEMPLATE IMAGES, THE REQUIREMENT OF PARAMETER ESTIMATION IN THE TESTING STAGE AND THE FRAMES PER SECOND (FPS) IN THE INFERENCE STAGE.

Methods	# Parameter	Template image(s)	Re-estimation	Frames per Second
Original	–	–	–	615
Color Aug. [36]	None	No	No	615
Stain Aug. [37]	None	No	No	612
SNMF [18]	$< 10^1$	Yes	Yes	50.5
ACD [20]	$< 10^1$	Yes	Yes	109
U-Net-Norm [35]	$> 10^7$	Yes	No	85.9
StainGAN [32]	$> 10^6$	Yes	No	50.0
SSC (Ours)	$< 10^2$	No	No	570

medical center or a few centers but needed to be deployed to one other center. Generally, the color profile within the target center is consistent and therefore can be transferred to the training profile. However, in the situation of telepathology we discussed in this experiment, the images in the practical application are also from different medical centers and in various color profiles. The situation has deviated from the hypothesis of the GAN-based method. This is the main reason the classification performance of StainGAN [32] on the gastric dataset (Table IV) is inferior to Stain Augmentation [37] and SSC.

The efficiency and convenience of a stain normalization method are also critical in practical application. In this experiment, we have assessed the speed of CNN applications using different stain normalization methods. The frames per second (FPS) of the CNN inference are presented in Table V, where the properties of these normalization models are also checked. The methods ACD [20] and SNMF [18] contain several parameters. Nevertheless, they require individual estimation process of standardization parameters for each testing WSI. The estimations have seriously reduced the FPS of CNN inference from about 600 to less than 100. The normalization methods based on U-Net [35] and StainGAN [32] involve millions of model parameters. The computation amount for stain normalization is comparable or even superior to the following HIA application. It has obviously decreased FPS in practical applications. In comparison, our SSC module involves only tens of model parameters and has no additional parameter estimation process in the prediction stage, which maintains the FPS of the CNN-based image classification application up to 570. These properties determine the SSC module is more efficient and convenient than the present methods in the deployment for HIA applications.

F. Performance for cervical cell classification

The normalization module is designed based on the properties of histopathology images. While it is potential to be applied to other pathology images. In this section, we evaluated the performance of the SSC module for cell classification on a dataset of cervix cytology images. The dataset contains 25,378 cell patches and involves 8 categories of cervical cells (including

Superficial, *Intermediate*, *Granulocyte*, *Glandular*, *Atypical*, *Koilocytotic*, and *High-N/C-Ratio* cells) annotated by pathologists. The detail of the dataset can refer to [43].

The backbone of the classification network was the same DenseNet-121 structure used in the histopathological image classification tasks. The methodology of the experiment follows that on the Camelyon16 dataset. Specifically, 70% cell patches were randomly sampled to train the network and the remainder was used for the test. The hyperparameters were determined as $(M, N, R) = (3, 6, 3)$. The overall classification accuracy and the confusion matrix for multi-category-classification were used as the metrics. The experimental results are shown in Fig. 7. It shows that the precisions of several types have been apparently improved. Especially for the abnormal cell types including *Atypical*, *Koilocytotic*, and *High-N/C-Ratio*, the precision has been improved by 3.1% to 4.6%. It is significant for applications of cervical cancer screening. Benefited from the improvement in the sub-categories, the overall accuracy of the classification increased from 0.934 to 0.941. The experimental results have indicated the proposed SSC module is also available for cytology image analysis.

G. Discussion

Our method is mainly developed to enhance the accuracy and robustness of the application-driven neural networks for histopathological image analysis, rather than to improve the quality of the visual assessment for pathologists. The reconstruction in our module is designed as a type of holistic regularization, rather than a sample-wise constraint. The purpose of the regularization is to force the normalized color profile being located around the center of color distribution of the training set, instead of overfitting to a certain profile. The ablation study in Table I has verified the effect of the reconstruction constraint. The reconstruction routine is inactive in the application stage of the networks. While it is also available to be activated if there is a requirement for output normalized images. In this situation, the image normalization and analysis can be simultaneously completed, even for whole slide images.

Since color normalization or stain normalization is generally used as the pre-processing for histopathology image analysis, the efficiency and the conveniences of the pre-processing model are as important as the normalization performance. One of the motivations that we studied the SSC module is to pursue a more efficient and convenient stain normalization method for CNN-based applications. After training, the SSC module is directly stored as a part of the CNN model files. Hence, no additional operation is required in the practical application beyond reloading the model files of the network and performing the forward propagation. Also, there is no restriction on CPU or GPU devices.

We have considered using the contextual information into the stain normalization. We tried to enlarge the size of the convolution kernel in Eq. 7 and 8 to 3×3 and 5×5 but found the accuracy of classification changed little. Finally, to minimize the number of parameters in our model, we set the kernel size of the two convolution layers to 1×1.

The researches [9], [44] have introduced the Capsule Network into breast histology image classification. In the two works, the structure of the Capsule Network is directly applied to classify the images. In contrast, we designed a novel stain normalization module following the insight of Capsule Network. The module can be equipped ahead any convolutional networks to enhance the robustness of the network in the aspect of stain color variance.

This work is a preliminary study to apply the routing approach to online determining the stain standardization results in the inference of the application-driven neural networks. The current routing algorithm relies on all the pixels of the stain separation candidates. One future work will try to establish a side network to estimate the routing scalars with only part of the pixels and then apply the scalars to merge the candidates. Hence, the computational efficiency can be further improved. Another work will focus on optimizing the routing principle.

The establishment of the SSC module is based on a good property of pathological images that the independent stain components can be extracted by sample transformation (Linear transformation) and an efficient observation (the sparsity) for optimizing the normalization. Intuitively, this prototype based on capsule networks can be further used to the normalization of other medical images (e.g. X-ray) that share the similar imaging characteristic to that of the pathological images.

V. Conclusion

In this paper, we propose a novel stain standardization module named stain standardization capsule for histopathological image analysis based on the optical properties of digital slides. The proposed module is developed in the domain convolutional neural network and can be directly equipped to CNN-based histopathological image analysis applications. The proposed method was evaluated with the application of histopathological image classification of breast, lung, and gastric tumors and cell classification on cervical cytology images. The results have demonstrated the proposed method is effective and robust for color standardization of pathological images. One future work will focus on refining the network structure to improve the efficiency of the module. Another work will focus on more effective routing algorithms.

References

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

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[3] S. Zhang and D. Metaxas, "Large-Scale medical image analytics: Recent methodologies, applications and Future directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.

[4] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Medical Image Analysis*, vol. 43, pp. 66–84, 2018.

[5] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwiggelaar, "Deep learning in mammography and breast histology, an overview and future trends," *Medical image analysis*, vol. 47, pp. 45–67, 2018.

[6] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. De Souza, A. Baidoshvili, G. J. S. Litjens, B. Van Ginneken, I. D. Nagtegaal, and J. A. W. M. V. Der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *IEEE International Symposium on Biomedical Imaging*, 2017, pp. 160–163.

[7] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PloS one*, vol. 12, no. 6, p. e0177544, 2017.

[8] A. Hamidinekoo and R. Zwiggelaar, "Stain colour normalisation to improve mitosis detection on breast histology images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 213–221.

[9] M. Anupama, V. Sowmya, and K. Soman, "Breast cancer classification using capsule network with preprocessed histology images," in *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2019, pp. 0143–0147.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[12] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[13] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, and Y. Zhao, "Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification," *Pattern Recognition*, vol. 71, pp. 14–25, 2017.

[14] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 737–744.

[15] A. Gertych, Z. Swiderska-Chadaj, Z. Ma, N. Ing, T. Markiewicz, S. Cierniak, H. Salemi, S. Guzman, A. E. Walts, and B. S. Knudsen, "Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides," *Scientific reports*, vol. 9, no. 1, p. 1483, 2019.

[16] P. Bandi, O. Geessink, Q. Manson, M. Van Dijk, M. Balkenhol, M. Hermsen, B. E. Bejnordi, B. Lee, K. Paeng, A. Zhong *et al.*, "From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 550–560, 2018.

[17] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, 2017, pp. 3856–3866.

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

[19] B. E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Höller, A. Homeyer, N. Karssemeijer, and J. A. V. D. Laak, "Stain specific standardization of whole-slide histopathological images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 404–415, 2016.

[20] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, "Adaptive color deconvolution for histological wsi normalization," *Computer methods and programs in biomedicine*, vol. 170, pp. 107–120, 2019.

[21] Y. Y. Wang, S. C. Chang, L. W. Wu, S. T. Tsai, and Y. N. Sun, "A color-based approach for automated segmentation in tumor tissue classification." in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, p. 6577.

[22] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[23] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "Colour normalisation in digital histopathology images," in *IEEE International Symposium on Bio Imaging*, 2009, pp. 1107–1110.

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

[25] X. Li and K. N. Plataniotis, "A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1862–1873, July 2015.

[26] J. Vicory, H. D. Couture, N. E. Thomas, D. Borland, J. S. Marron, J. T. Woosley, and M. Niethammer, "Appearance normalization of histology slides," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 89–98, 2015.

[27] N. Hidalgo-Gavira, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "Fully automated blind color deconvolution of histopathological images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2018, pp. 183–191.

[28] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, and Y. Zhao, "Histopathological whole slide image analysis using context-based cbir," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1641–1652, 2018.

[29] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2414–2423.

[30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[31] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, and P. H. de With, "Stain normalization of histopathology images using generative adversarial networks," in *IEEE International Symposium on Biomedical Imaging*. IEEE, 2018, pp. 573–577.

[32] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "Staingan: Stain style transfer for digital histological images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 953–956.

[33] A. Lahiani, N. Navab, S. Albarqouni, and E. Klaiman, "Perceptual embedding consistency for seamless reconstruction of tilewise style transfer," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing, 2019, pp. 568–576.

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

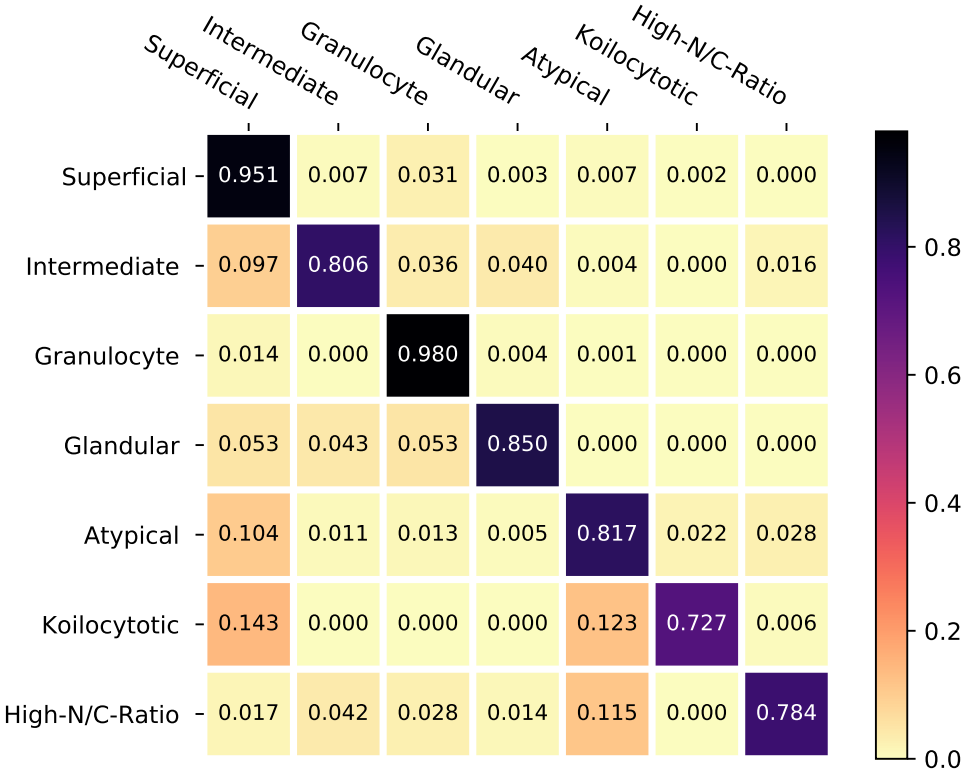
[35] D. Tellez, G. Litjens, P. Bandi, W. Bulten, J.-M. Bokhorst, F. Ciompi, and J. van der Laak, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *arXiv preprint arXiv:1902.06543*, 2019.

[36] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado *et al.*, "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017.

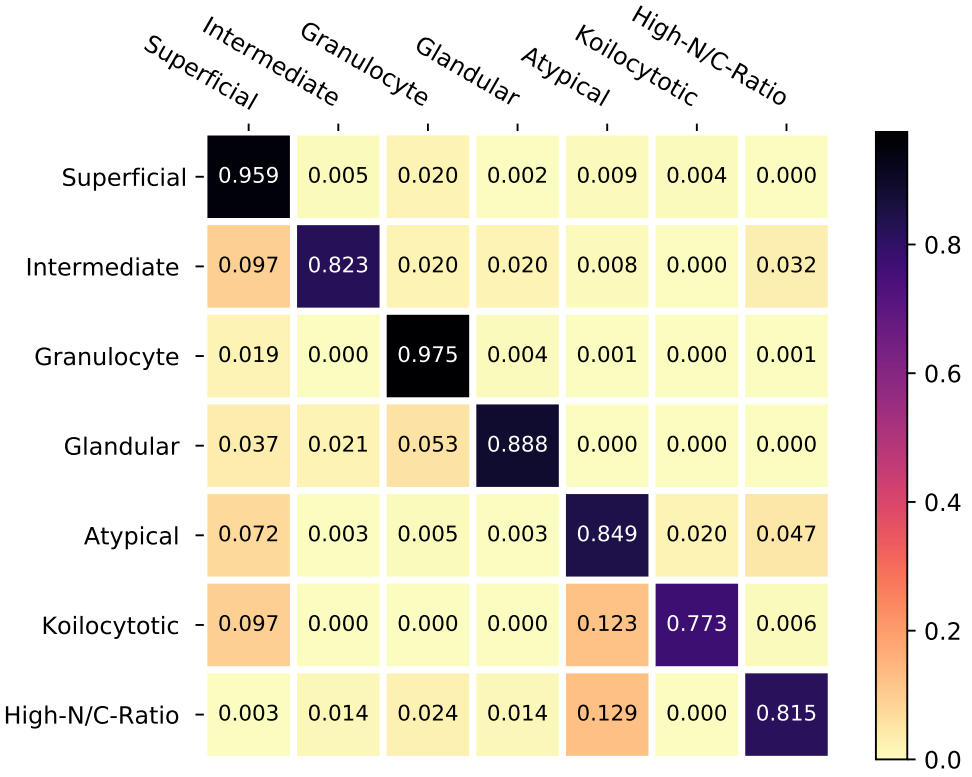
[37] D. Tellez, M. Balkenhol, I. Otte-Höller, R. van de Loo, R. Vogels, P. Bult, C. Wauters, W. Vreuls, S. Mol, N. Karssemeijer *et al.*, "Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2126–2136, 2018.

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

[39] Y. Zheng, Z. Jiang, H. Zhang, J. Shi, and F. Xie, "Stain standardization capsule: A pre-processing module for histopathological image analysis," in *MICCAI 2019 Workshop – COMPAY19*, 2019.
[40] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, no. 11, pp. 1457–1469, 2004.
[41] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. J. S. Litjens, J. A. W. M. V. Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.
[42] Z. Li, Z. Hu, J. Xu, T. Tan, H. Chen, Z. Duan, P. Liu, J. Tang, G. Cai, Q. Ouyang *et al.*, "Computer-aided diagnosis of lung carcinoma using deep learning-a pilot study," *arXiv preprint arXiv:1803.05471*, 2018.
[43] J. Shi, R. Wang, Y. Zheng, Z. Jiang, and L. Yu, "Graph convolutional networks for cervical cell classification," in *MICCAI 2019 Workshop – COMPAY19*, 2019.
[44] T. Iesmantas and R. Alzbutas, "Convolutional capsule network for classification of breast cancer histology images," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 853–860.



(a) Original (7110)



(b) by SSC (7164)

Fig. 7. Comparison of confusion matrices for cell classification on the Motic-Cervix dataset before and after applying the SSC module, where (a) is the original result of the classification network and (b) is the result of the network equipped with the SSC module, and the number of correctly classified images in the testing set is given in the brackets.