

SD-Layer: Stain Deconvolutional Layer for CNNs in Medical Microscopic Imaging

Rahul Duggal¹, Anubha Gupta^{1(✉)}, Ritu Gupta², and Pramit Mallick¹

¹ SBILab, Department of ECE, Indraprastha Institute of Information
Technology-Delhi (IIIT-D), Delhi, India
rahulduggal2608@gmail.com, anubha@iiitd.ac.in,
pramitmалlick@gmail.com

² Laboratory Oncology Unit, Dr. BRA.IRCH, AIIMS, Delhi, India
drritugupta@gmail.com,
<http://sбилab.iiitd.edu.in/index.html>

Abstract. Convolutional Neural Networks (CNNs) are typically trained in the RGB color space. However, in medical imaging, we believe that pixel stain quantities offer a fundamental view of the interaction between tissues and stain chemicals. Since the optical density (OD) colorspace allows to compute pixel stain quantities from pixel RGB intensities using the Beer-Lambert's law, we propose a stain deconvolutional layer, hereby named as SD-Layer, affixed at the front of CNN that performs two functions: (1) it transforms the input RGB microscopic images to Optical Density (OD) space and (2) this layer deconvolves OD image with the stain basis learned through backpropagation and provides tissue-specific stain absorption quantities as input to the following CNN layers. With the introduction of only nine additional learnable parameters in the proposed SD-Layer, we obtain a considerably improved performance on two standard CNN architectures: AlexNet and T-CNN. Using the T-CNN architecture prefixed with the proposed SD-Layer, we obtain 5-fold cross-validation accuracy of 93.2% in the problem of differentiating malignant immature White Blood Cells (WBCs) from normal immature WBCs for cancer detection.

Keywords: Deep learning · Classification · Stain deconvolution · Cancer imaging

1 Introduction

In the past few years, Convolutional Neural Networks (CNN's) have attained immense success in medical imaging problems such as detection and classification [2–6, 12, 15]. For example, in [3], a magnification independent framework and CNN model is presented to detect H&E stained breast cancer cells. In [2], a

Authors gratefully acknowledge the research funding support (Grant Number: 1(7)/2014-ME&HI) from the Ministry of Communication and IT, Govt. of India for this research work.

simple 3-layer CNN architecture is presented with data augmentation to classify Immunofluorescence images of HEP-2 cells. In [4, 5], deep neural networks are investigated for mitosis detection. Apart from classification and detection, CNNs have also been used in medical image segmentation [13].

CNNs used in these problems are typically trained in the RGB colorspace. However, the ideal discriminating features in medical microscopic images may not be the pixel intensities in the RGB color space, but the stain quantities that are absorbed and are characteristics of the tissue. Previous works have shown that the stain quantities can be estimated in the Optical Density (OD) space through the application of Beer-Lambert's law [8–10, 14]. This transformation from RGB color space to the stain quantity space is commonly termed as stain deconvolution. Motivated with the above, we propose Stain Deconvolution Layer (hereby named as SD-Layer) that is a biomedically relevant CNN layer and can be prefixed to any CNN model and performs the following functions:

- (i) It transforms the input RGB images to the Optical Density (OD) space.
- (ii) Initialized with the stain basis vector of one of the cell image, this layer learns the optimal stain basis vectors of cell/tissue of interest for class labels through backpropagation.
- (iii) It deconvolves OD image with the learned stain color basis and provides tissue-specific stain absorption quantities that are used as input to the following CNN architecture.

To the best of our knowledge, this is the first work where deep learning based classification of medical images has been employed in the OD space using the Beer-Lambert law based stain deconvolution. We evaluate the performance of the proposed SD-Layer by prefixing it to two standard CNNs (AlexNet and T-CNN) on the challenging problem of differentiating malignant immature White Blood Cells (WBCs) from hematogones (benign immature WBCs) for cancer detection in Acute Lymphoblastic Leukemia (ALL). ALL detection has been carried out in the past using machine learning algorithms on hand-crafted features [11, 15].

However, the datasets considered in these studies are typically small and hence, generalization error of these on the real world unseen data may be higher. In this paper, we have applied deep learning based proposed architecture on nearly 9000 immature WBCs (total) for malignant versus normal WBC blast classification with a 5-fold cross validation accuracy of 93.2%. This is to note that the novelty of the paper lies in the proposed deep learning based architecture that can be applied to other classification problems in medical imaging. The remainder of the paper is organized as follows. In Sect. 2, we review some of the relevant theory. In Sect. 3, we propose the SD-Layer formulation. In Sect. 4, we evaluate the performance of the proposed SD-Layer. Section 5 presents a small discussion, followed by some conclusions in Sect. 6.

2 Background

This section presents a brief review of the theory required to understand the proposed work. Assume that a given stained slide is illuminated by light of

intensity $I_{o,c}$ in color channel c (red, green, or blue) and $I(p, c)$ denotes intensity captured by the camera at pixel location p in channel c . Beer Lambert's law is defined as:

$$I(p, c) = I_{o,c} e^{-\sum_{i=1}^N Q(p,i)S(i,c)}, \quad (1)$$

where $Q(p, i)$ is the quantity of the i^{th} stain constituent absorbed at pixel location p , $S(i, c)$ is the characteristic absorbance of the i^{th} stain constituent in the channel c , $I_{o,c}$ can also be viewed as the maximum pixel intensity in channel c where no staining chemical is absorbed, and N is the number of stain constituents. From (1), it is noted that the observed pixel intensity $I(p, c)$ varies non-linearly with the quantity of staining chemical Q in the RGB colorspace. However, the optical density $O(p, c)$ defined as the negative log of (1) varies linearly with Q as below:

$$O(p, c) = -\log_{10} \frac{I(p, c)}{I_{o,c}} = \sum_{i=1}^3 Q(p, i)S(i, c). \quad (2)$$

In the matrix notation, this can be written as

$$\mathbf{O} = \mathbf{Q}\mathbf{S}, \quad (3)$$

where \mathbf{O} and \mathbf{Q} are matrices of dimension $MN \times 3$, \mathbf{S} is the stain color matrix of dimension 3×3 . Each row of \mathbf{S} constitutes one stain basis vector, while each column of \mathbf{Q} refers to the quantity of each of these stain basis vectors present at different pixel positions.

Generally, Beer-Lambert law based deconvolution proceeds as follows. A given input image in RGB space is first converted to OD space via (2) to obtain \mathbf{O} . Next, \mathbf{Q} and \mathbf{S} are estimated through different matrix factorization strategies such as Singular value decomposition (SVD) [9], non negative matrix factorization (NMF) [8], and sparse NMF (S-NMF) [14]. In this paper, we use the widely popular SVD based method to achieve stain deconvolution.

3 Proposed Stain Deconvolution Layer (SD-Layer)

In this section, we present the proposed SD-Layer that is built on the understanding of the staining based imaging of biological tissues.

Firstly, as has been discussed earlier, absorption of stain quantities at different positions correspond to the tissue properties. Since variation in stain absorption by tissues/cells lead to the formation of corresponding medical image, it is more appropriate to design a classifier using stain quantities. Thus, we propose to train the CNN on the stain quantities absorbed \mathbf{Q} obtained via deconvolution of the OD space image \mathbf{O} with the stain basis vectors in \mathbf{S} as below:

$$\mathbf{Q} = \mathbf{O}\mathbf{S}^{-1}. \quad (4)$$

Here, \mathbf{S} can be obtained via SVD of \mathbf{O} [9]. In practice, stain matrix \mathbf{S} determined using (SVD) would vary from image to image due to several factors such

as illumination variation, over/under staining, ageing of the staining chemicals, etc. [8]. Thus, full microscopic images are stain normalized prior to cell segmentation and classification. However, stain normalization carried out on the full slide (containing large number of cells) may still lead to stain variations at the individual cell level. Since classification is required at the cell level, training a CNN on \mathbf{Q} , obtained using (4) via stain matrix \mathbf{S} estimated on the full slide reference image, may not yield desired classification accuracy. Thus, we would like to fine tune the stain matrix \mathbf{S} at the cell level via the proposed SD-Layer.

In order to realize this, we interpret the matrix multiplication between \mathbf{O} and \mathbf{S}^{-1} in (4) as convolution between the rows of \mathbf{O} and the columns of \mathbf{S}^{-1} as shown in Fig. 1. Thus, each column of \mathbf{S}^{-1} is equivalent to a convolution filter of dimension $1 \times 1 \times 3$ and stride 1. This interpretation allows to learn \mathbf{S}^{-1} optimally at the cell level through backpropagation. It is important to note that accurate learning of \mathbf{S}^{-1} is heavily dependent on its initialization. We found that initializing the convolution filters using the columns of \mathbf{S}^{-1} , determined through SVD on the reference image, led to good results. We experiment with other initializations in the experiments section.

To sum up the discussion, the SD-Layer (shown in Fig. 2) performs two functions. Firstly, it transforms the input image from RGB to OD space using (2). Secondly, it determines the stain quantities present at each pixel using (4). The stain matrix, initialized through stain deconvolution of the reference image, is

$$\begin{array}{c}
 \mathbf{O} \qquad \qquad \qquad \mathbf{S}^{-1} \qquad \qquad \qquad \mathbf{Q} \\
 \begin{bmatrix} 0.5614 & 0.9294 & 0.4334 \end{bmatrix} \times \begin{bmatrix} -0.4641 & -0.7960 & -0.3886 \\ -0.8050 & 0.1960 & 0.5600 \\ -0.3697 & 0.5727 & -0.7316 \end{bmatrix} = \begin{bmatrix} -1.1690 & -0.0165 & -0.0148 \end{bmatrix} \\
 \qquad \qquad \qquad \phi_1 \qquad \qquad \phi_2 \qquad \qquad \phi_3 \\
 \updownarrow \\
 \begin{bmatrix} [0.5614 & 0.9294 & 0.4334] & [0.5614 & 0.9294 & 0.4334] & [0.5614 & 0.9294 & 0.4334] \\ \qquad \qquad \qquad * & \qquad \qquad \qquad * & \qquad \qquad \qquad * \\ [-0.4641 & -0.8050 & -0.3697] & [-0.7960 & 0.1960 & 0.5727] & [-0.3886 & 0.5600 & -0.7316] \\ \qquad \qquad \qquad \phi_1 & \qquad \qquad \qquad \phi_2 & \qquad \qquad \qquad \phi_3 \end{bmatrix}
 \end{array}$$

Fig. 1. Interpreting matrix multiplication in (4) as convolution. Each element of \mathbf{Q} can be viewed as convolution of rows of \mathbf{O} with columns of \mathbf{S}^{-1} .

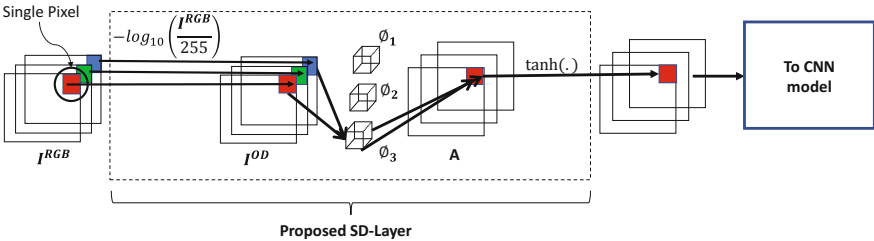


Fig. 2. Illustration of SD-Layer. ϕ_i 's are learnable $1 \times 1 \times 3$ sized convolutional filters.

optimally learned at the cell level through backpropagation. This introduces only 9 additional learnable parameters that is insignificant compared to the total number of weights in the model. Thus, the gain in classification accuracy as presented in the next section is due to the more biologically relevant input image representation rather than the enhanced model capacity.

4 Experiments

In this section, we evaluate the performance of SD-Layer appended to the front end of two CNN architectures: AlexNet [7] and Texture-CNN [1]. AlexNet is a widely studied standard CNN model. It consists of 5 convolution layers followed by 2 fully connected layers, followed by a softmax layer. For input image dimension of our dataset, AlexNet contains ≈ 146 million weights.

Texture-CNN (T-CNN) was recently proposed in [1] and was shown to achieve superior results on texture datasets. It modifies AlexNet, by computing features from the 3-D activation map of the last convolutional layer instead of simply flattening it. These features act as order-less texture descriptors. So, for a 3-D map of dimension $H \times W \times D$, computing channel-wise mean results in D number of features whereas flattening would give HWD features. With the reduced number features that are fed to the subsequent fully connected layers, T-CNN contains ≈ 20 million learnable parameters.

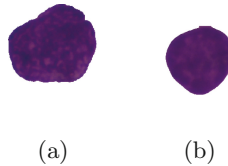


Fig. 3. Example images from our dataset. (a) Nucleus of normal WBC immature cell (b) Nucleus of malignant WBC blast.

4.1 Dataset

Our Data consists of microscopic image slides prepared from the bone marrow aspirate of normal and ALL subjects. These images are stained with Jenner-Giemsa stain. A trained oncologist hand-labeled the normal and malignant WBC immature cells. All the images were normalized using [9] for stain variation. The nuclei of the labeled cells were then segmented. In total, our dataset consists of 8938 cell nuclei, 4469 nuclei of each class. We used random rotations through 180 degrees and vertical flipping in each epoch as two data augmentation strategies during the training phase. To account for the varying sizes of the segmented nuclei, we embed the nuclei in a 400×400 black colored patch. Re-sizing of cell images provide poor results since texture is an important feature that gets altered with re-sizing. Example images from our dataset are shown in Fig. 3.

4.2 Experiment 1: AlexNet vs T-CNN with and Without SD-Layer

To establish a baseline performance, we evaluate performance of two models: AlexNet [7] and Texture-CNN [1] on our dataset. All models were trained using stochastic gradient descent (SGD) for 400 epochs. The initial learning rate for AlexNet was set to 0.01 and for AlexNet with SDLayer to 0.001. For T-CNN, with and without SD-Layer, the initial learning rate was set to 0.01, which was reduced by a factor of 10 on epochs 300 and 350. The momentum and decay were set to 0.9 and 10^{-6} for all models. The 5-fold cross validation accuracy and f-score are shown in the first two rows of Table 1. Since texture is an important discriminative feature that AlexNet is unable to tap despite its larger model capacity, we note that T-CNN outperforms Alexnet by a large margin.

Next, we prefix SD-Layer at the front end of both models. Two settings are considered for the SD-Layer: (1) frozen - convolution filters are not allowed to train post initialization and (2) trainable - filters are allowed to train to the best possible representation. In the first case, CNN performs poorly as is evident from Fig. 4a. This is because the stain vector initialized using SVD on the full slide reference image cannot fully overcome cell-level stain variations. On the other hand, significantly higher test accuracy is obtained on the second

Table 1. 5-fold cross-validation accuracy of Alexnet and T-CNN, with and without SD-Layer

Model architecture	Accuracy (in %)	F-Score (in %)
AlexNet	87.9	88.12
T-CNN	92.48	92.7
AlexNet+SD-Layer	88.5	88.32
T-CNN+SD-Layer	93.2	93.08

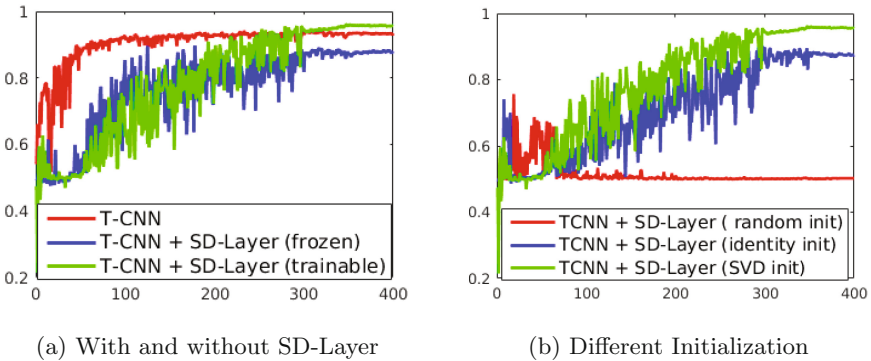


Fig. 4. Plots of test accuracy v/s epochs for T-CNN model without and with SD-Layer, evaluated on a single fold.

setting of ‘trainable’, wherein filters are allowed to fine tune to cell level stain normalization. We report 5-fold cross validation results for AlexNet and TCNN prefixed with the trainable SD-Layer in the bottom two rows of Table 1. We note a significant jump in accuracy, from 87.9% to 88.5% for AlexNet and from 92.48% to 93.2% for T-CNN, with the proposed SD-Layer.

4.3 Experiment 2: Results with Different Initializations of SD-Layer

As stated earlier, the initialization of filters in the SD-Layer plays a significant role. In this sub-section, we evaluate the performance of T-CNN with SD-Layer initialized using three different strategies: (1) Identity matrix, (2) uniform random distribution from $[-0.05, 0.05]$, and (3) with the columns of \mathbf{S}^{-1} determined using SVD on the reference image. Table 2 summarizes maximum test accuracy achieved using each initialization on a single fold. The corresponding, test accuracy v/s epochs plots are shown in Fig. 4b. From this table and figure, we note that the randomly initialized model fails to train. This is expected since, in this case, the input to the CNN is a random image. For the identity initialization though, the model trains towards some intermediate representation starting from the original RGB image. However this representation neither improves accuracy, nor allows us to draw some understandable interpretation. The best test accuracy is achieved through SVD based initialization.

Table 2. Classification accuracy of T-CNN+SD-Layer with different \mathbf{S} initialization with single fold.

Model architecture	Accuracy (in %)	F-Score (in %)
T-CNN+SD-Layer (identity initialization)	87.5	85.8
T-CNN+SD-Layer (SVD initialization)	95.5	95.4

5 Discussion

We claim that SD-Layer trains the stain colour matrix to a representation better suited to classification. This can be verified by generating RGB images using (3), by preserving only a single column of \mathbf{S} at a time and setting the other two to zeroes. This scheme, is equivalent to generating images containing only a single stain. This visualization, for the case of \mathbf{S} obtained through (1) SVD and (2) after training SD-Layer using T-CNN, are shown in Fig. 5. It is observed that initial images of (b)-(d) of malignant blast are modified to (e)-(g), wherein (e) seems to capture shape, (f) seems to capture texture, while (g) is having no information. Similar observation is observed for the normal cell shown in the bottom row of Fig. 5.

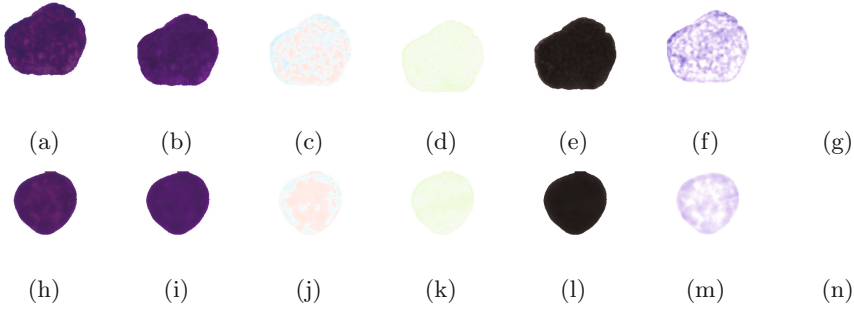


Fig. 5. Top row is an example image of a lymphoblast nucleus (malignant). (a) Original image, (b)-(d) Stain deconvolved images using stain vectors obtained via SVD, (e)-(g) Stain deconvolved images using stain vectors learned via T-CNN+SD-Layer over 400 epochs. Bottom row shows corresponding images for the normal WBC nucleus.

6 Conclusion

In this paper, we have proposed a biomedical microscopic imaging relevant deep CNN network architecture where the staining of tissues/cells are involved. We have proposed stain deconvolution layer (SD-Layer) that operates in the Optical Density space and offers a more fundamental view of the tissue and stain interactions to the following CNN architecture. The concept of initializing and tuning the stain matrix has been incorporated into the SD-Layer that will deal with stain variations present at the cell level. With only an 9 additional learnable parameters, we are able to achieve significant gain in the classification accuracy on two standard models AlexNet and T-CNN fitted with SD-Layer. This suggests that SD-Layer leads to a better representation of the input image.

References

1. Andrearczyk, V., Whelan, P.F.: Using filter banks in convolutional neural networks for texture classification. *Pattern Recogn. Lett.* **84**, 63–69 (2016)
2. Bayramoglu, N., Kannala, J., Heikkilä, J.: Human epithelial type 2 cell classification with convolutional neural networks. In: 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 1–6. IEEE (2015)
3. Bayramoglu, N., Kannala, J., Heikkilä, J.: Deep learning for magnification independent breast cancer histopathology image classification. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2440–2445. IEEE (2016)
4. Chen, H., Dou, Q., Wang, X., Qin, J., Heng, P.A.: Mitosis detection in breast cancer histology images via deep cascaded networks. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pp. 1160–1166. AAAI Press (2016)
5. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 411–418. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40763-5_51](https://doi.org/10.1007/978-3-642-40763-5_51)

6. Cruz-Roa, A.A., Arevalo Ovalle, J.E., Madabhushi, A., González Osorio, F.A.: A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 403–410. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40763-5_50](https://doi.org/10.1007/978-3-642-40763-5_50)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
8. Li, X., Plataniotis, K.N.: A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics. *IEEE Trans. Biomed. Eng.* **62**(7), 1862–1873 (2015)
9. Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: ISBI 2009, pp. 1107–1110. IEEE (2009)
10. Ruifrok, A.C., Johnston, D.A., et al.: Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**(4), 291–299 (2001)
11. Singhal, V., Singh, P.: Local binary pattern for automatic detection of acute lymphoblastic leukemia. In: 2014 Twentieth National Conference on Communications (NCC), pp. 1–5. IEEE (2014)
12. Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.W., Snead, D.R., Cree, I.A., Rajpoot, N.M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans. Med. Imaging* **35**(5), 1196–1206 (2016)
13. Wang, J., MacKenzie, J.D., Ramachandran, R., Chen, D.Z.: A deep learning approach for semantic segmentation in histology tissue images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 176–184. Springer, Cham (2016). doi:[10.1007/978-3-319-46723-8_21](https://doi.org/10.1007/978-3-319-46723-8_21)
14. Xu, J., Xiang, L., Wang, G., Ganesan, S., Feldman, M., Shih, N.N., Gilmore, H., Madabhushi, A.: Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis. *Comput. Med. Imaging Graph.* **46**, 20–29 (2015)
15. Zhao, J., Zhang, M., Zhou, Z., Chu, J., Cao, F.: Automatic detection and classification of leukocytes using convolutional neural networks. *Med. Biol. Eng. Comput.*, 1–15 (2016)