Simultaneous Localization and Classification of Acute Lymphoblastic Leukemic Cells in Peripheral Blood Smears Using a Deep Convolutional Network with Average Pooling Layer

Arna Ghosh, Satyarth Singh and Debdoot Sheet
Department of Electrical Engineering, Indian Institute of Technology Kharagpur
Kharagpur, WB 721302, India
Email: arnatubaikgp@gmail.com, satyarth2305@gmail.com, debdoot@ee.iitkgp.ernet.in

***Abstract*—It is important to analyze and classify the blood cells for the evaluation and diagnosis of many diseases. Acute Lymphoblastic Leukemia (ALL) is a blood cancer mostly found in children below the age of 7-8 years. It can be fatal if left untreated. ALL cells are abnormal lymphocytes that have a condensed appearance to their chromatin. ALL can be detected through the analysis of white blood cells (WBCs) also called as leukocytes. Presently the morphological analysis of blood cells is performed manually by skilled operators, which makes it a time-taking and non-standardized process. This paper presents a novel deep learning approach to automate the process of detecting ALL from whole-slide blood smear images. Previous work in this domain deal with the isolation and classification of WBCs based on certain morphological image features. However, this work uses a deep network for simultaneous localization and classification of WBCs. The network makes use of Average Pooling layers to figure out the hot-spots of WBC locations in whole-slide images. Although this workflow fails to successfully figure out all the ALL lymphocytes in a whole-slide image, but it does performs very well in the task of predicting whether the blood smear image belongs to an ALL patient or not.**

***Index Terms*—Acute lymphoblastic leukemia, deep CNN, digital pathology, average pooling layer, localization.**

I. Introduction

Leukemia is the blood cancer associated with white blood cells [1]. It is a bone marrow disorder that arises when abnormal white blood cells begin to continuously replicate themselves. As they accumulate, they inhibit the production of normal blood cells in the bone marrow, thus leading to anaemia. Acute leukemia is classified according to the French-American-British (FAB) classification into two types: Acute Lymphoblastic Leukemia(ALL) and Acute Myelogenous Leukemia (AML) [2]. ALL is generally common among children and while AML mainly affects adults, it can also occur in children and adolescents.

ALL is a fast-growing cancer and is fatal if left untreated, as it spreads rapidly into the bloodstream as well as other vital organs [3]. But the diagnosis of the disease poses a serious challenge as the symptoms are very similar to flu. The detection of ALL starts with a complete blood count (CBC) [4]. It there are abnormalities in this count, a study of morphological bone marrow and peripheral blood slide analysis is done to confirm the presence of leukemic cells. In order to classify the abnormal cells in their particular types and subtype of leukemia, an expert operator is required to observe some cells under light microscopy looking for the abnormalities presented in the nucleus or cytoplasm of the cells [5]. However, this analysis is a time-taking process and presents a non-standardized accuracy as it is subject to the operators capabilities. This paper presents a fast and effective deep network approach to localize and segment out the blast cells and aid the diagnosis of leukemia.

II. Prior Art

A substantial amount of work has been done for automated detection of ALL from blood smear images. Most of the work has been along the lines of WBC segmentation followed by classification. The use of border identification and segmentation technique [6] and use of morphological operators followed by snake balloon algorithm [7] for segmentation of WBC cells has been tried and tested. [8] presents a segmentation technique on color space images using scale-space filtering of nucleus extraction and watershed algorithm. [9] presents a computationally heavy fuzzy c-means clustering for leukemic image segmentation. Morphological operators coupled with scale-space properties of toggle operator to improve accuracy of segmentation is presented in [10]. A morphological analysis of WBCs are presented in [11]. The most common thresholding technique used in the above works is Otsu's thresholding [12] and Otsu's method with image enhancement [13]. In [14] they consider the use of Zack thresholding algorithm instead of Otsu's method.

Classification of the segmented WBCs based on features like shape, perimeter and area of bounding block, form factor etc. using support vector Mmachines (SVM) [15], naïve Bayes classifier [16] and Reinforcement Learning [17]. In [18] they present a brief overview of the techniques used and a detailed description of the feature space.

However, most of the existing methods consider classification and localization as separate problems and deal them with

978-1-5386-1676-5/17/$31.00

varied approach. In this paper, we present a deep learning approach to come up with a localization-cum-classification architecture dedicated to the detection of ALL from whole-slide blood smear images.

III. Methodology

Our approach consists of the following stages:

1) Network training over patch dataset
2) Generation of heatmaps using Average Pooling Layer
3) Extraction of patches from whole-slide images for classification

The methodology is similar to the architecture in [19] and is extended to abnormal WBC detection in blood smear images.

A. Network Training

A pre-trained AlexNet is used and the network is fine-tuned on the patch dataset available from the ALL-IDB dataset [4]. Since the AlexNet is trained for object classification on 227×227 images and the patch dataset available is of dimension 257×257, the linear layer dimensions of the pre-trained AlexNet did not suit the training purpose. Therefore, the convolutional layers of the AlexNet are only used. These essentially store the significant filters required for feature extraction from natural images. We hope to build on this feature space to come up with a better feature space dedicated to ALL detection. Linear layers of suitable dimension are added to come up with a network architecture as shown in 1. This network is then trained on the patch dataset. Hence, the linear part of the network is trained from scratch and the convolutional layers are fine-tuned to learn to classify the patches of WBC as normal and abnormal.

The available dataset contains 130 images of each class. Since we feel a data insufficiency to train our network, we augment data to the dataset using rotation and mirroring of the data. Since the images are rotation and mirroring invariant, applying those operations on the images doesnt change the class the original image belongs to. We perform rotations at an interval of 20° and each of these images are mirrored. Hence the dataset of 260 images becomes a dataset of 9360 images, sufficient for training a deep neural network. We perform a 10-fold cross validation on this data. Training is done for 25 iterations with a learning rate of 0.02 and a learning decay rate of 0.001. This augmentation is done to teach the network the rotation and mirroring invariant property of the WBCs.

The loss function used during training is the negative log-likelihood(NLL) represented as

$$J(\theta) = -\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right)\right] - \left[\sum_{i=1}^{m} \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right] \tag{1}$$

where $y^{(i)}$ represents the probability of the class to occur, $h_\theta\left(x^{(i)}\right)$ represents the probability of the corresponding class predicted by the classifier $h_\theta(\cdot)$ for the given sample $x^{(i)}$.

The entire dataset is divided into 5 bags namely B_1 to B_5 each consisting of $1,872$ images (936 images of each class). The details of training scheme and results are presented in Table I. The training and validation curves for Fold 1 are presented in Fig. 2. Other folds show similar trends.

B. Generation of Heatmap

We follow the idea as presented in [19]. Counting on the network trained on the patch dataset, we use the convolutional layers and append an Average Pooling layer, that averages the excitation of each pixel over all the filters at the end, to form a network that generates the heatmap. We then pass the whole-slide images present in the ALL-IDB dataset [4] through this network to generate filters corresponding to each image. These filters show higher excitation at the position of the WBCs and lower excitation at all other positions. Thus, using a hard threshold gives us reasonable masks to segment out WBCs from the images for further classification.

Using a fixed threshold for all images gives us a huge variance in the output. This is primarily because each image has its own excitation intensity depending on the inherent intensity of image pixels. Hence, the mask obtained is normalized before applying the hard threshold. We also observe certain noise pixels in these images which are primarily bright regions in the whole-slide images. To avoid such anomalies, we convert the RGB image to HED space [20] and threshold on the Eosin channel to segment only the WBC cells from the masked image.

C. Extraction of Patches

Using the obtained binary image, we extract potential centroids of individual WBCs. This is a relatively simple problem when WBCs are not overlapping or agglomerated. In case of singular WBC, a simple blob detection algorithm should suffice. However, the dataset has multiple instances of agglomerated WBCs and therefore we use a different approach. We follow the following algorithm to determine the centroid of the WBC

1) All black pixels in the binary image (with WBC pixels as white), are labeled as background pixels.
2) For all white pixels, the minimum distance from background in both x-axis and y-axis is calculated.
3) The overall cost of the pixel is calculated as the sum of the distances in the two directions.
4) The local maxima of the cost function are labeled as the centroids. To calculate the local maxima points, we check the 3×3 neighborhood of a pixel and label it as a centroid iff it has the maximum cost among the 9 pixels.

With the obtained centroids, we choose a patch of size 257×257 around each centroid and pass it through the trained network to label it as an abnormal or normal WBC. In the result shown, the red boxes mark the abnormal patches and the green boxes mark the normal ones.

The workflow is illustrated in Fig. 3.

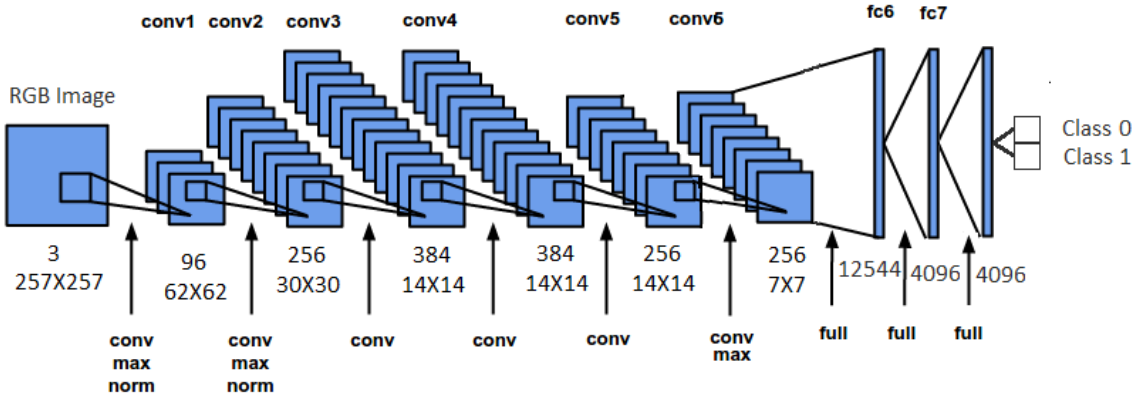


Fig. 1. Architecture of the deep convolutional network used in our approach.

TABLE I
DATASET ARRANGEMENT INTO TRAINING, VALIDATION AND TESTING SETS

Fold	Training	Validation	Testing	True Positives	True Negatives	False Positives	False Negatives	Accuracy	Specificity	Sensitivity	F-score
1	B_1,B_2,B_3	B_4	B_5	936	936	0	0	100%	100%	100%	100%
2	B_1,B_2,B_3	B_5	B_4	936	936	0	0	100%	100%	100%	100%
3	B_2,B_3,B_4	B_5	B_1	936	936	0	0	100%	100%	100%	100%
4	B_2,B_3,B_4	B_1	B_5	936	936	0	0	100%	100%	100%	100%
5	B_3,B_4,B_5	B_1	B_2	936	936	0	0	100%	100%	100%	100%
6	B_3,B_4,B_5	B_2	B_1	936	936	0	0	100%	100%	100%	100%
7	B_4,B_5,B_1	B_2	B_3	936	936	0	0	100%	100%	100%	100%
8	B_4,B_5,B_1	B_3	B_2	936	936	0	0	100%	100%	100%	100%
9	B_5,B_1,B_2	B_3	B_4	936	936	0	0	100%	100%	100%	100%
10	B_5,B_1,B_2	B_4	B_3	936	936	0	0	100%	100%	100%	100%

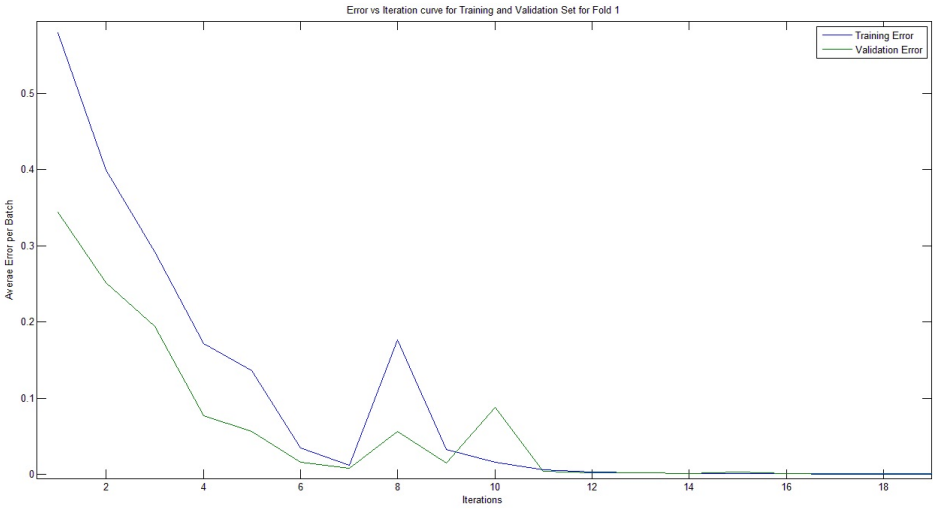


Fig. 2. Error Curve for training and validation sets for Fold 1

IV. RESULTS AND DISCUSSION

A. Accuracy of workflow

Although we are unable to attain a satisfactory cell-level accuracy, our method attains very good slide-level accuracy. Out of available 108 patches, 105 are labeled correctly. The detailed analysis and accuracy metrics are provided in Table II. A step-by-step workflow results are shown for a successful detection case in Fig. 4 and a failure case in Fig. 5

The poor cell-level accuracy can be attributed to the absence of agglomerated WBCs in the patch dataset. We see that the network performs fairly well on patches with a single isolated WBC. But it fails when presented with patches having 2 or

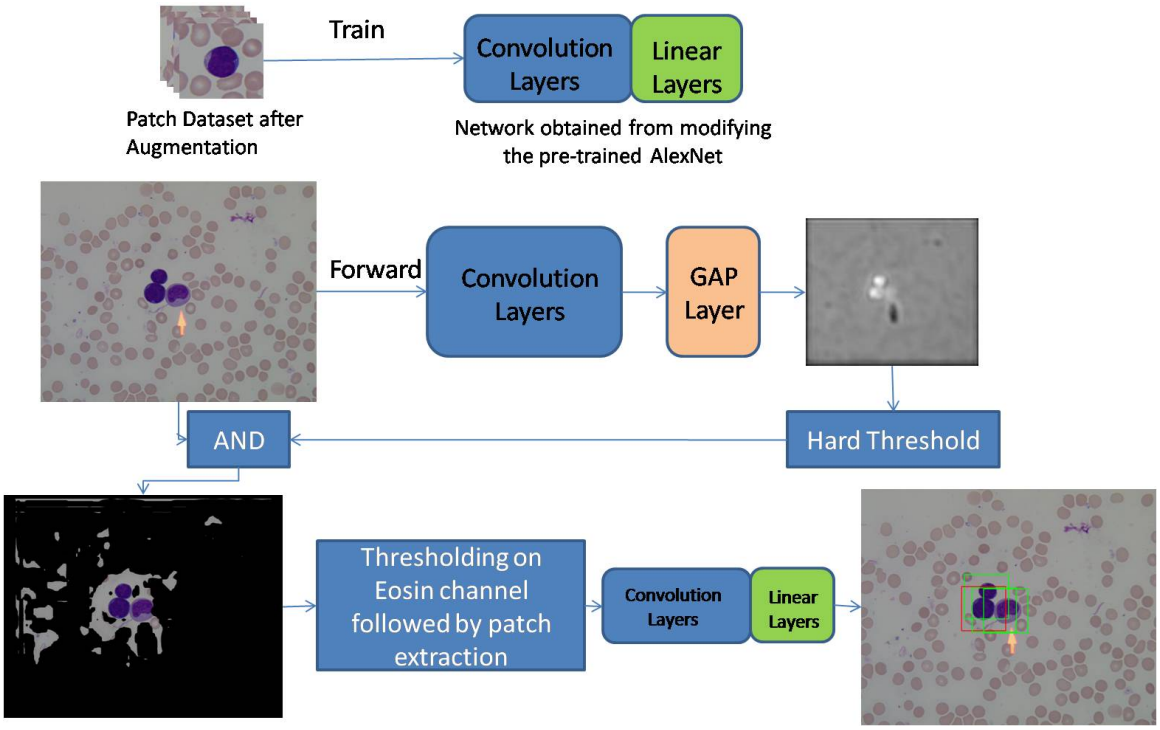


Fig. 3. Workflow illustrated

TABLE II
RESULTS ON WHOLE-SLIDE IMAGES PRESENT IN ALL_IDB1 [4]

Metric	Value
True Positives	49
True Negatives	56
False Positives	3
False Negatives	0
Accuracy	97.22%
Specificity	94.9153%
Sensitivity	100%
F-score	97.03%

more WBC cells sticking together. Therefore, adding such patches to the patch dataset on which the network is trained could improve the patch-level accuracy.

B. Weights associated with learned filters

We try to figure out the individual weights of each filter learned in the network without using the Average Pooling Layer. To figure out individual weights, we pass the patch dataset through the network and blind out the output of the filtermap we are considering. The classification error over the entire dataset is taken as a metric of the weight of that filter. Closely observing the weight vector obtained, we see that all the filters have almost the same accuracy and it is close to the overall error of the network. Hence we come to the conclusion that although we block the features represented by a particular filter, a linear combination of other filters is sufficient to make-up for the absence of that filter. This leads us to believe the existence of redundancy in the filter space.

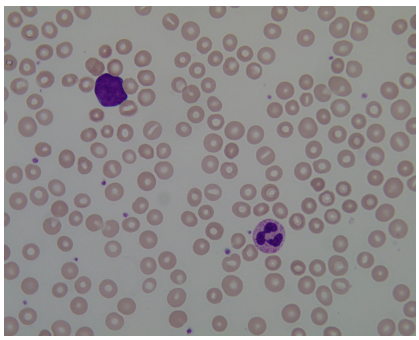
C. Analysis of filters learned

We first pass a white image through the network and obtain filter maps from the last convolutional layer of the network. We find out the Jensen-Shannon(JS) distance between these filters and represent the JS distance matrix as an image as shown in Fig. 6. The intensity of the image pixels are representative of the JS distance values. Arranging the filters to have the most similar ones closer, we get Fig. 7. The existence of dark blocks along the diagonal imply the presence of equivalent filters. This confirms our hypothesis from the previous section.

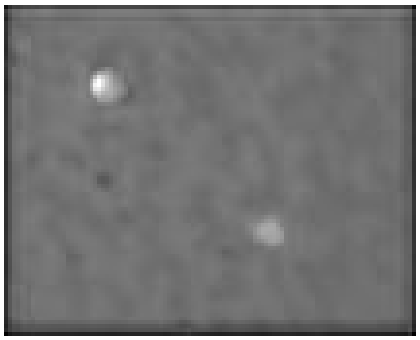
This presents an opportunity to reduce the size of the network while deploying to improve space utility and processing time without compromising on accuracy. While calculating the individual weights of filters, we can hard threshold the similarity matrix image with a small value and thus form a list of similar filters. While calculating the weight of a filter, we blind the output of all filters similar to that filter and assign all similar filters same weightage.

D. Comparison with existing methods

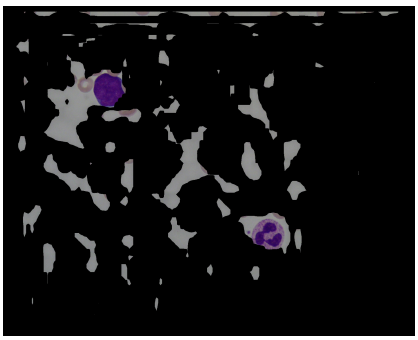
Most the prior work use their own dataset for experiments so it is difficult to compare result. [6] [7] [8] [9] [10] have cited accuracy results for segmentation only and there is no quantitative results on ALL classification or detection. Only [2] reports classification accuracy of 92.5% that too on self-collected dataset for ALL detection. Hence there is a dearth of work on a standard dataset which makes it difficult to compare the results of previous studies. The absence of a common reference framework restricts the authors from drawing a quantitative comparison to previous studies.



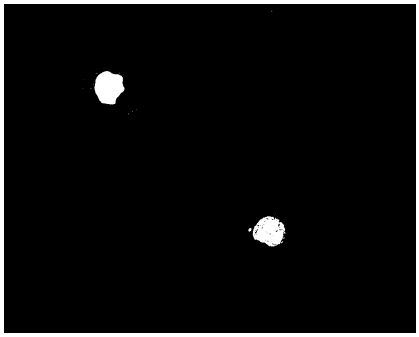
(a) Original Image



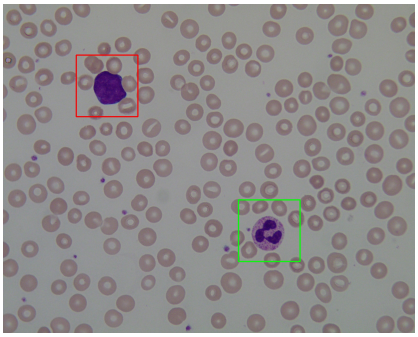
(b) Filter after Average Pooling layer



(c) Image masked with filter result obtained from average pooling

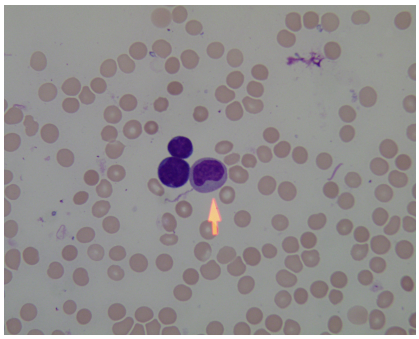


(d) Masked Image thresholded on Eosin channel

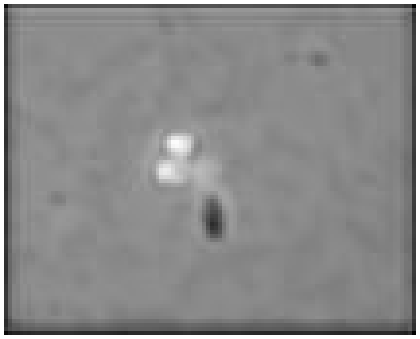


(e) Final Result after patch classification

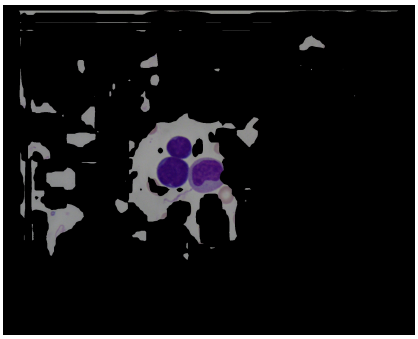
Fig. 4. Results of the different stages of workflow on an image with isolated WBCs



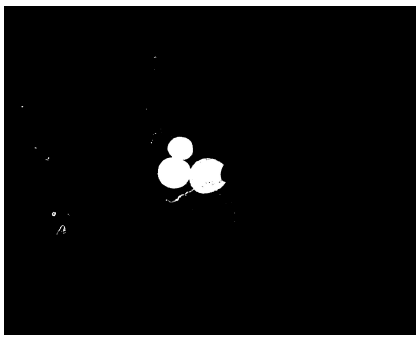
(a) Original Image



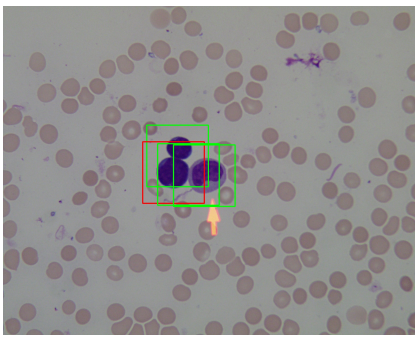
(b) Filter after Average Pooling layer



(c) Image masked with filter result obtained from average pooling



(d) Masked Image thresholded on Eosin channel



(e) Final Result after patch classification

Fig. 5. Results of the different stages of workflow on an image with agglomerated WBCs

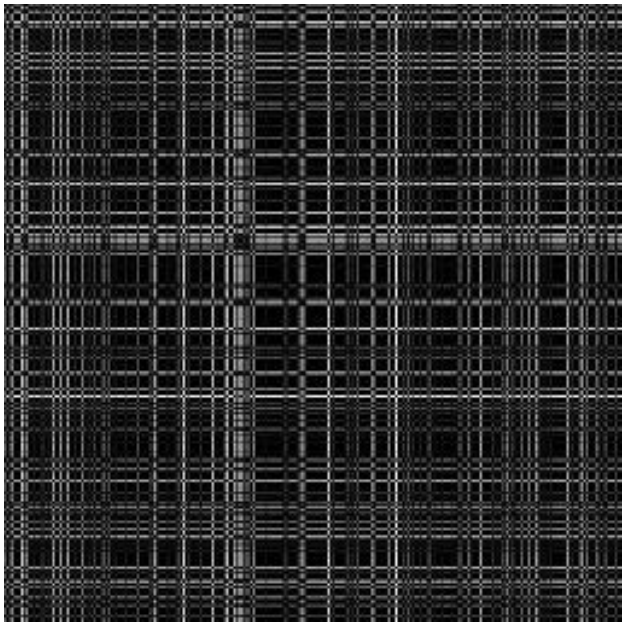


Fig. 6. Similarity matrix of size 256 × 256 represented as image

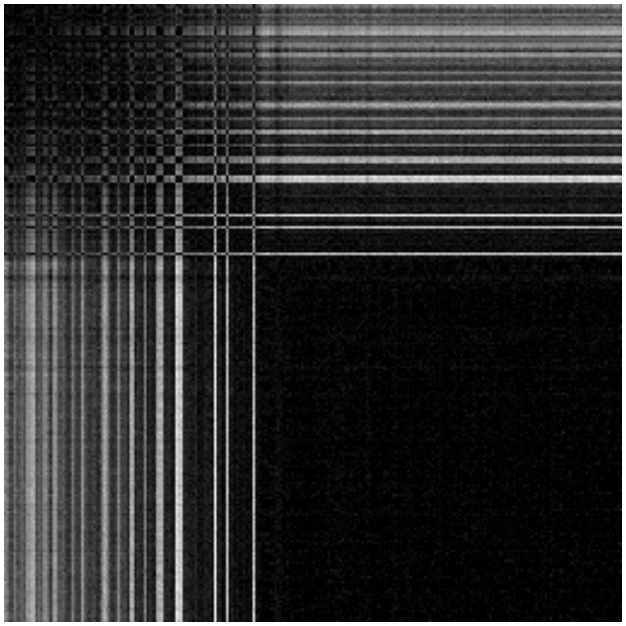


Fig. 7. Similarity matrix sorted

V. Conclusion

The main aim of the presented work is to prepare a end-to-end framework for ALL detection from whole-slide images. The present framework shows promising results and with certain modifications could help solve the problem of ALL detection. This also provides sufficient confidence that using deep networks for localization-cum-classification problems can beat traditional feature extraction based methods in other domains as well. The results obtained encourage future work, as discussed in previous sections to improve the system.

References

[1] C. Haworth, A. Heppleston, P. M. Jones, R. Campbell, D. Evans, and M. K. Palmer, "Routine bone marrow examination in the management of acute lymphoblastic leukaemia of childhood.," *Journal of clinical pathology*, vol. 34, no. 5, pp. 483–485, 1981.

[2] H. T. Madhloom, S. A. Kareem, and H. Ariffin, "A robust feature extraction and selection method for the recognition of lymphocytes versus acute lymphoblastic leukemia," in *Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on*, pp. 330–335, IEEE, 2012.

[3] T. Breden, K.and Schorr and J. B. Schorr, "Blood," *Colliers Encyclopaedia*, vol. 4, 1978.

[4] R. D. Labati, V. Piuri, and F. Scotti, "All-idb: The acute lymphoblastic leukemia image database for image processing," in *Image processing (ICIP), 2011 18th IEEE international conference on*, pp. 2045–2048, IEEE, 2011.

[5] B. J. Bain, *A beginner's guide to blood cells*. Blackwell Publishing, 2004.

[6] N. Ritter and J. Cooper, "Segmentation and border identification of cells in images of peripheral blood smear slides," in *Proceedings of the thirtieth Australasian conference on Computer science-Volume 62*, pp. 161–169, Australian Computer Society, Inc., 2007.

[7] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac, "An automated differential blood count system," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 3, pp. 2583–2586, IEEE, 2001.

[8] K. Jiang, Q.-M. Liao, and S.-Y. Dai, "A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering," in *Machine Learning and Cybernetics, 2003 International Conference on*, vol. 5, pp. 2820–2825, IEEE, 2003.

[9] S. Rubhala and J. Ramya, "Image segmentation of leukemia using kernel based fuzzy c means clustering," *International Journal of Communications and Engineering*, vol. 3, no. 3, pp. 291–299, 03 March2012.

[10] L. B. Dorini, R. Minetto, and N. J. Leite, "White blood cell segmentation using morphological operators and scale-space analysis," in *Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on*, pp. 294–304, IEEE, 2007.

[11] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in *IEEE international conference on computational intelligence for measurement systems and applications*, vol. 96101, 2005.

[12] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[13] T. Kulkarni and D. Bhosale, "A fast segmentation scheme for acute lymphoblastic leukemia detection," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 3, no. 2, pp. 7253–7258, 2014.

[14] S. Sanal, "Automated detection of acute lymphocytic leukemia-a survey," *International Journal of Engineering Research and General Science*, vol. 3, pp. 168–171, 2015.

[15] T. Markiewicz and S. Osowski, "Data mining techniques for feature selection in blood cell recognition.," in *ESANN*, pp. 407–412, 2006.

[16] M. Ghosh, D. Das, S. Mandal, C. Chakraborty, M. Pala, A. K. Maity, S. K. Pal, and A. K. Ray, "Statistical pattern analysis of white blood cell nuclei morphometry," in *Students' Technology Symposium (TechSym), 2010 IEEE*, pp. 59–66, IEEE, 2010.

[17] F. Kasmin, A. S. Prabuwono, and A. Abdullah, "Detection of leukemia in human blood sample based on microscopic images: A study.," *Journal of Theoretical & Applied Information Technology*, vol. 46, no. 2, 2012.

[18] B. S. Ansha and R. Remya, "An overview on acute lymphocytic leukemia detection using cell image segmentation," in *National Conference in Emerging Technologies*, pp. 22–29, 2014.

[19] M. Bolanos and P. Radeva, "Simultaneous food localization and recognition," *arXiv preprint arXiv:1604.07953*, 2016.

[20] A. C. Ruifrok, D. A. Johnston, *et al.*, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, no. 4, pp. 291–299, 2001.