Neutrophils Identification by Deep Learning and Voronoi Diagram of Clusters *

Jiazhuo Wang[1], John D. MacKenzie[2],
Rageshree Ramachandran[3], and Danny Z. Chen[1]

[1] Department of Computer Science and Engineering, University of Notre Dame, USA
[2] Department of Radiology and Biomedical Imaging, UCSF, USA
[3] Department of Pathology, UCSF, USA

Abstract. Neutrophils are a primary type of immune cells, and their identification is critical in clinical diagnosis of active inflammation. However, in H&E histology tissue slides, the appearances of neutrophils are highly variable due to morphology, staining and locations. Further, the noisy and complex tissue environment causes artifacts resembling neutrophils. Thus, it is challenging to design, in a hand-crafted manner, computerized features that help identify neutrophils effectively. To better characterize neutrophils, we propose to extract their features in a learning manner, by constructing a deep convolutional neural network (CNN). In addition, in clinical practice, neutrophils are identified not only based on their individual appearance, but also on the context formed by multiple related cells. It is not quite straightforward for deep learning to capture precisely the rather complex cell context. Hence, we further propose to combine deep learning with Voronoi diagram of clusters (VDC), to extract needed context. Experiments on clinical data show that (1) the learned hierarchical representation of features by CNN outperforms hand-crafted features on characterizing neutrophils, and (2) the combination of CNN and VDC significantly improves over the state-of-the-art methods for neutrophil identification on H&E histology tissue images.

1 Introduction

Identification of a primary type of immune cells, neutrophils, is of great medical importance. Because the number and the locations are key features for acute inflammation diagnosis [6]; further, quantitative analysis of the distribution patterns of neutrophils may help provide deep insight into acute inflammation.

In H&E histology tissue images (Fig. 1(a)), neutrophils are characterized as having multiple lobes in their nuclei, and almost invisible cytoplasms (Fig. 1(b)). But, in practice, it is highly challenging to identify them for the following reasons. First, there are lots of variations in neutrophil appearances due to, e.g., staining, shape, and size. In fact, large portions of neutrophils do not show common

* This research was supported in part by NSF Grant CCF-1217906, a grant of the National Academies Keck Futures Initiative (NAKFI), and NIH grant K08-AR061412-02 Molecular Imaging for Detection and Treatment Monitoring of Arthritis.

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 226–233, 2015.
DOI: 10.1007/978-3-319-24574-4_27

"textbook" characteristics (Fig. 1(c)). Second, the background is quite noisy and complex with a mixture of different biological structures, such as other types of immune cells (e.g., lymphocytes, eosinophils, and plasma cells, see Fig. 1(d)) and some tissue layers (e.g., glands and villi). These structures greatly complicate the neutrophils identification task. For example, each lobe of a neutrophil could be classified incorrectly as a single lymphocyte. Artifacts resembling neutrophils are often created by other structures in H&E histology tissue images (Fig. 1(e)).

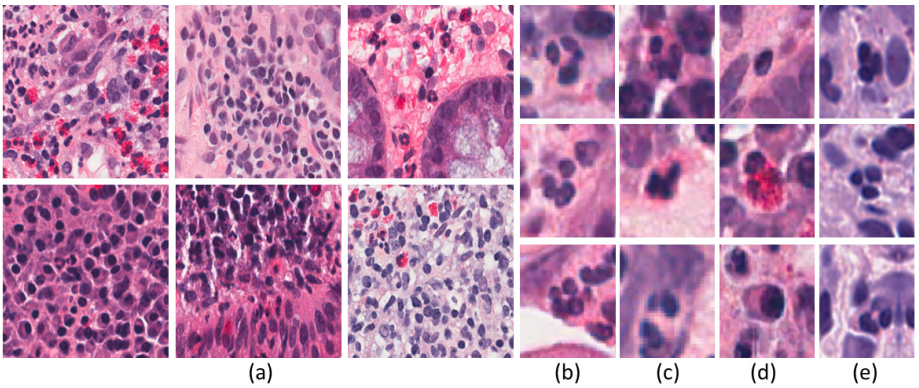

Fig. 1. (a) H&E histology tissue image examples; (b) neutrophils; (c) neutrophils without obvious "textbook" appearances; (d) a lymphocyte (top), an eosinophil (middle), and a plasma cell (bottom); (e) artifacts resembling neutrophils.

The first and state-of-the-art method for neutrophils identification on H&E histology tissue images was proposed very recently in [8]. But further improvements can be made as illustrated below.

One key to effective identification of neutrophils is to well capture the features of individual cells (such features should be invariant to different factors). The method in [8] conducted feature engineering in a hand-crafted manner, which is quite challenging, since the characteristics of neutrophils become highly subtle in the noisy and complex tissue environment. Hence, we advocate to automatically learn features in such settings. In computer vision literature, deep learning [2,1,7] was recently shown to outperform highly-tuned state-of-the-art methods on many problems (e.g., recognition, detection) using the automatically learned hierarchical low, mid, and high level representations of features.

Another key to effective identification of neutrophils is to model cell context, i.e., related cells in a certain neighborhood help determine cell types on one another. In fact, cell context is commonly utilized by pathologists, especially when identifying neutrophils without obvious "textbook" appearances. Therefore, the method in [8] proposed to exploit such cell context, extracted based on Voronoi diagram of clusters (VDC) [3] of "trustable" neutrophils. However, because their "trustable" neutrophils are identified based on hand-crafted features and Random Forest, it is likely that some of the "trustable" neutrophils thus identified

are actually not neutrophils. Clearly, using context built on a not so "trustworthy" version of "trustable" neutrophils means propagating noisy or even wrong information on the appearances of neutrophils to the neighborhood.

Motivated by the above two keys, we propose to combine deep learning and Voronoi diagram of clusters (VDC) into one framework for identifying neutrophils in H&E histology tissue images. Our main idea is to first model the appearances of individual cells by a convolutional neural network (CNN) [5] and identify the "trustable" portion of neutrophils, and then model context by constructing the VDC of the "trustable" neutrophils and utilize the context to resolve the ambiguous cases (evaluated on merely individual cell appearances).

Experimental results on clinical data validate that deep learning and VDC are efficacious complements to each other. Also, the general idea of combing deep learning (for modeling individual object appearances) and VDC (for modeling context) can be applicable to other similar identification problems.

2 Modeling Individual Cell Appearances by CNN

Overview. Assume that the cells (regardless of their types) have been segmented and detected (we actually apply the iterative edge labeling method [8] to segment and detect cells). For each segmented and detected cell, we crop a small image patch containing it. Our goal is to capture the subtle characteristics of individual cells, so that we can subsequently classify the image patches into neutrophils and non-neutrophils.

Our main idea is to apply supervised deep neural networks to learn a hierarchical representation of the low, mid, and high level features of the cells. Such a network maps each input image (patch), via a series of parameterized layers, into probabilities of the target classes (in our case, neutrophils and non-neutrophils). The parameters of the layers are learned during training, and some of them can be viewed as filters. Responses to the filters in lower layers (i.e., closer to the input image) usually contain low (e.g., edges, corners) or mid (e.g., correlations between edges) level features, and responses in higher layers contain high level semantic features related to the target classes.

Deep Learning Model Design. We construct a deep convolutional neural network (CNN) [5]. The overall architecture of our CNN (shown in Fig. 2) basically follows that in [5], but with some important modifications to the hyper parameters, as described below. Our modifications to the hyper parameters are motivated mainly by two considerations. First, the size of our input image patch is quite different from that in [5] (256 pixels). Thus, using the same set of hyper parameters cannot describe well the characteristics of our target objects. For example, the size of the filters in the first convolutional layer of [5] is 11×11, which appears seemingly too big for the first layer filters in our case, which are supposed to capture low level features (e.g., edges and corners). Because the layers are closely related, an inappropriate first layer not functioning well could impose negative influence on the subsequent layers. Second, as stated in [10], by visualizing the network of [5] using the method in [10], some hyper parameters

do not have appropriate values. For example, the large stride 4 of the filters in the first layer causes aliasing artifacts in the second layer. We set the stride to 1, taking into consideration of such aliasing artifacts, our input image size, and the influence to the subsequent layers.

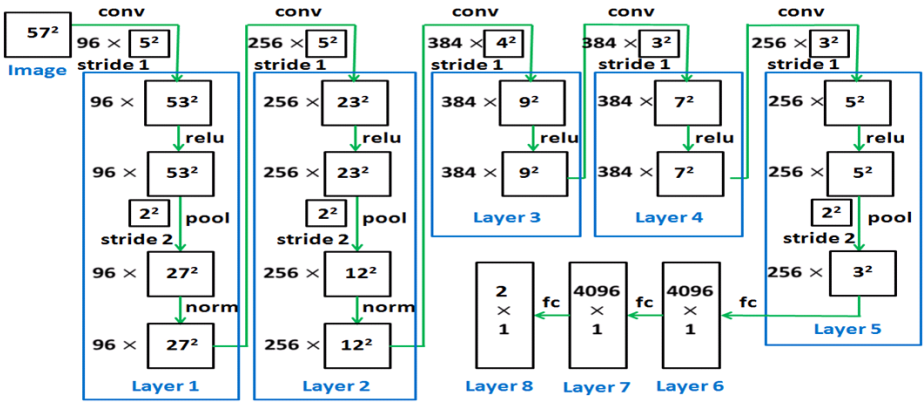

Fig. 2. Our CNN structure. Our CNN has 8 layers. The first 5 layers are convolutional. For example, the input image patch of size 57×57 to the first convolutional layer is first convolved (conv) by a set of 96 learned filters of size 5×5; the resulting 96 feature maps of size 53×53 then pass through a rectified linear function (relu), go through max pooling with stride 2 on each direction, and finally are locally normalized before output to the next layer. The rest 3 layers are fully connected (fc). We minimize the softmax cost function at the last layer. Please see [5] for the roles played by layers.

Training. *Learning Algorithm.* We apply stochastic gradient descent with a batch size of 512 to learn parameters in CNN. We initialize learning rate as 0.01, momentum as 0.9, and weight decay as 0.0005. We also decrease learning rate three times before termination. We iterate 30000 times over the batches.

Overfitting. To combat overfitting, we apply data augmentation [5] and drop out [4]. For data augmentation, we sub-crop 10 images (corners and centers with or without horizontal flips) of size 57×57 from the original image patch of size 64×64. We apply drop out at the fully connected layers (6 and 7) and set the drop out rate as 0.5. Please see [5,4] for more details of these techniques.

Conservativeness. In our clinical setting of studying inflammatory bowl diseases, neutrophils are usually rare (but important) comparing to other types of immune cells (such as lymphocytes and plasma cells). But, the noisy and complex environment of histology tissue images may cause many artifacts that look similar to neutrophils. Hence, we adopt a conservative strategy to identify neutrophils, by penalizing more when a non-neutrophil is misclassified as a neutrophil than when a neutrophil is misclassified as a non-neutrophil, during the training of CNN. We did so via multiplying by two the computed softmax loss when a non-neutrophil is misclassified as a neutrophil.

3 Modeling Cell Context by VDC

Overview. In clinical practice, when pathologists identify neutrophils, clusters of neutrophils usually attract their attention (bigger clusters may attract more attention), because such clusters can provide some cues about the local appearances difference between neutrophils and non-neutrophils in the neighborhood of the clusters, and help identify near by neutrophils that may appear ambiguous individually. To our best knowledge, no straightforward way is known for deep learning to model precisely such context information. Thus, to complement deep learning and capture cell context, our main idea is to apply the model of Voronoi diagram of clusters (VDC) [8,3], based on the "trustable" portion of neutrophils. Note that, different from [8], we incorporate and utilize the high quality CNN features and CNN probabilities in VDC (as described later in detail below), which make the extracted context information more robust.

Voronoi Diagram of Clusters. We first discuss general concepts for VDC. Different from the common Voronoi diagram, in which each site is a single point, every site of VDC can also be a cluster of points. Due to this feature, VDC can capture the influence of each cluster (i.e., the collective influence of all site points belonging to the cluster) to the neighborhood regions. The tessellation of VDC is built based on the influence function $D(\cdot,\cdot)$, of each cluster C (with site points $\{p_1, p_2, \ldots, p_m\}$) on any point p, which is defined as

$$D(C,p) = \sum_{i=1}^{m} F(p,p_i), \qquad F(p,p_i) = \begin{cases} \frac{w(p_i)}{d(p,p_i)^2}, & \text{if } d(p,p_i) \le dist_T \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $F(p,p_i)$ measures the influence of a single point $p_i \in C$ to the point p, which depends on both the weight $w(p_i)$ (for the "importance" of p_i) and the Euclidean distance $d(p,p_i)$ between p and p_i. Note that we truncate $F(p,p_i)$ using a distance threshold $dist_T$ to reflect that a point p which is too far away from a cluster C is not influenced by C. In this way, a point p would be influenced more by a cluster C whose size is big, and/or its site points have larger weights, and are closer to p. Thus p belongs to the geometric VDC cell of C (i.e., this is not a biological cell) in the tessellation of VDC.

Extracting Cell Context. We extract the cell context to classify individually ambiguous cases by taking the following steps.

(1) Obtain a set of "trustable" neutrophils by thresholding the neutrophil probability of each cell produced by CNN. (Actually, we also obtain a set of "trustable" non-neutrophils for the non-neutrophil probability, based on the same threshold value.) All the remaining cells (i.e., not "trustable" neutrophils nor "trustable" non-neutrophils) are taken as individually ambiguous cells (thus to be classified later on by context). Note that since the learned hierarchical feature representation can capture well the characteristics of individual cells, the "trustable" ones produced by CNN are actually more "trustable" than those produced using hand-crafting features and simple classifier as in [8], even with

similar thresholding. This reduces the chance of propagating noisy or wrong information when we resolve ambiguous cases based on the influence of clusters of "trustable" neutrophils later on.

(2) Apply density-based clustering [9] to the "trustable" neutrophils to capture their potential clustering behaviors.

(3) Compute VDC: Each site of VDC is either a cluster of "trustable" neutrophils found in Step (2), or an individual "trustable" neutrophil that does not belong to any such cluster. Assign the weight of each "trustable" neutrophil as the neutrophil probability computed by CNN (as opposed to 1 in [8]), in the influence function $F(\cdot, \cdot)$ defined above. By doing so, clusters of neutrophils that are more likely to contain correct information would have larger influence to the neighborhood. Set $dist_T = 160$ pixels. Some examples are given in Fig. 3.

(4) For each geometric VDC cell, build a Random Forest classifier, using the "trustable" neutrophils and non-neutrophils within this VDC cell as respectively the "+" and "-" training examples. Here, we simply use the filters' responses in layer 7 of our CNN as the features, because they contain high level semantic information on the cell types. Such a classifier captures locally what neutrophils and non-neutrophils look like in the neighborhood, and thus is referred to as the *LocalExpert* (LE) for the corresponding geometric VDC cell.

(5) For each ambiguous cell c, define its *context* as all the geometric VDC cells overlapping with the circle of radius R centered at c. (Note that if the context is empty, then it essentially means that likely no neutrophil cluster or single neutrophil is around c, and thus we simply take c as a non-neutrophil.) We use the weighted average probabilities of the LEs of these geometric VDC cells to determine whether c is a neutrophil. We set the weight of an LE as the collective influence from the corresponding cluster to c, to reflect that the decision on c's cell type favors more the "opinion" from a cluster with a larger influence to c.

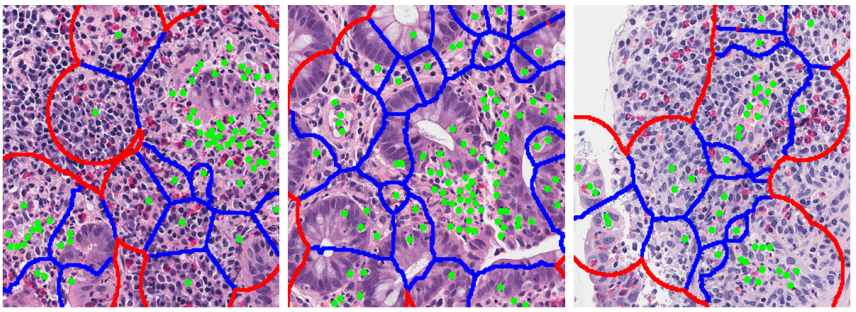

Fig. 3. Some examples of VDC tessellation: The green points are "trustable" neutrophils; the blue curves are boundaries between adjacent geometric VDC cells; the red curves are boundaries between adjacent geometric VDC cells and regions that exceed the distance threshold $dist_T$ based on the influence function $F(\cdot, \cdot)$.

4 Experiments and Evaluation

We collected histology tissue slides (scanned at 40X magnification, which pathologists commonly use to identify cell types) from patients suspected of having inflammatory bowl diseases (the study was performed under an IRB and in compliance with the privacy provisions of HIPPA of 1996). We manually marked 810 neutrophils (we automatically detected the cells regardless of their types; the detected cells that are not marked as neutrophils are taken as non-neutrophil ground truth), from 53 images as ground truth. This a larger and more challenging dataset than that in [8]. Each time, we randomly selected 21 images for training, and the rest images for testing (we repeat 5 times to obtain the average performance). To measure the performance quantitatively, we compute two metrics, $precision = \frac{TP}{TP+FP}$ and $recall = \frac{TP}{TP+FN}$, for neutrophils.

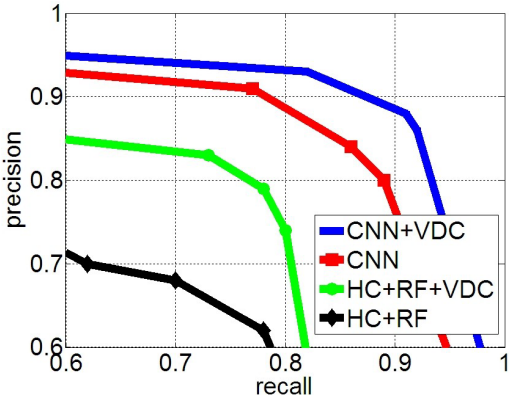

Fig. 4. The precision and recall curves.

Evaluation on Modeling Individual Cell Appearances. To demonstrate the effectiveness of deep learning on capturing the characteristics of cells, we compare the performance based on merely thresholding the probabilities output by CNN (i.e., without VDC), and two versions of the state-of-the-art method [8]: Hand-crafted features (color, texture, etc) + Random Forest (HC + RF), and HC + RF + VDC. The precision and recall curves (obtained by varying the probability threshold for both CNN and HC + RF, and varying the threshold for determining "trustable" neutrophils for HC + RF + VDC) are shown in Fig. 4.

One can see that the combination of hand-crafted features and simple classifier, even after using the additional context information extracted based on VDC, is outperformed by CNN. This validates that the learned hierarchical representation of features by CNN captures well the subtle characteristics of individual cells, which are often further complicated by the noisy and complex environment of H&E histology tissue images.

Evaluation on Modeling Cell Context. The precision and recall curve (obtained by varying the threshold for determining "trustable" neutrophils) of combining CNN and VDC is also shown in Fig. 4. One can see that although CNN alone already performs well, using additional context information extracted from VDC further improves the performance. This validates that VDC is a good complement to CNN, and modeling cell context can considerably help neutrophils identification.

5 Conclusions

In this paper, we present a robust new method that significantly improves the state-of-the-art for neutrophil identification in clinical tissue samples. Our method constructs a CNN to characterize the features of neutrophils, which are subtle and complicated given the cells are embedded in the noisy and complex tissue environment. It also combines CNN with VDC to extract needed cell context that helps identify neutrophils more effectively.

References

1. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning 2(1), 1–127 (2009)
2. Bengio, Y., Courville, A.C., Vincent, P.: Unsupervised feature learning and deep learning: A review and new perspectives. IEEE Trans. PAMI, special issue Learning Deep Architectures (2012)
3. Chen, D.Z., Huang, Z., Liu, Y., Xu, J.: On clustering induced Voronoi diagrams. In: FOCS, pp. 390–399 (2013)
4. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors. CoRR abs/1207.0580 (2012)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1106–1114 (2012)
6. Naini, B.V., Cortina, G.: A histopathologic scoring system as a tool for standardized reporting of chronic (ileo) colitis and independent risk assessment for inflammatory bowel disease. Human Pathology 43, 2187–2196 (2012)
7. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Networks 61, 85–117 (2014)
8. Wang, J., MacKenzie, J.D., Ramachandran, R., Chen, D.Z.: Identifying neutrophils in H&E staining histology tissue images. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 73–80. Springer, Heidelberg (2014)
9. Xu, B., Chen, D.Z.: Density-based data clustering algorithms for lower dimensions using space-filling curves. In: Zhou, Z.-H., Li, H., Yang, Q. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4426, pp. 997–1005. Springer, Heidelberg (2007)
10. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part I. LNCS, vol. 8689, pp. 818–833. Springer, Heidelberg (2014)