Efficient Lung Cancer Cell Detection with Deep Convolution Neural Network

Zheng Xu and Junzhou Huang(✉)

Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019, USA
jzhuang@uta.edu

Abstract. Lung cancer cell detection serves as an important step in the automation of cell-based lung cancer diagnosis. In this paper, we propose a robust and efficient lung cancer cell detection method based on the accelerated Deep Convolution Neural Network framework(DCNN). The efficiency of the proposed method is demonstrated in two aspects: (1) We adopt a training strategy, learning the DCNN model parameters from only weakly annotated cell information (one click near the nuclei location). This technique significantly reduces the manual annotation cost and the training time. (2) We introduce a novel DCNN forward acceleration technique into our method, which speeds up the cell detection process several hundred times than the conventional sliding-window based DCNN. In the reported experiments, state-of-the-art accuracy and the impressive efficiency are demonstrated in the lung cancer histopathological image dataset.

1 Introduction

Automatic lung cancer cell detection is the basis of many computer-assisted methods for cell-based experiments and diagnosis. However, at present, very few work has been focused on lung cancer cell detection. The difficulty in lung cancer cell detection problem is basically three-fold. First, the density of lung tumor cells is generally very high in the histopathological images. Second, the cell size might vary and cell clumping is usual. Third, the time cost of cell detection method, especially in high-resolution histopathological images, is very high in cell-based diagnosis. With these challenges mentioned above, it is still in great demand for researchers to develop efficient and robust lung cancer cell detection methods. To alleviate these problems, we propose an efficient and robust lung cancer cell detection method based on the Deep Convolution Neural Network(DCNN) [1]. Other than computationally-intensive frameworks [2,3], or ROI(region of interest)-based detection method [4,5], it exploits the deep architecture to learn the hierarchical discriminative features, which has recently achieved significant success in biomedical image analysis [6,7].

This work was partially supported by U.S. NSF IIS-1423056, CMMI-1434401, CNS-1405985.

G. Wu et al. (Eds.): Patch-MI 2015, LNCS 9467, pp. 79–86, 2015.
DOI: 10.1007/978-3-319-28194-0_10

In the proposed method, the training process is only performed on the local patches centered at the weakly annotated dot in each cell area with the non-cell area patches of the same amount as the cell areas. This means only weak annotation of cell area (a single dot near the center of cell area) are required during labeling process, significantly relieving the manual annotation burden. Another benefit for this technique is to reduce the over-fitting effect and make the proposed method general enough to detect the rough cell shape information in the training image, providing the benefit for further applications, e.g. cell counting, segmentation and tracking.

During testing stage, the conventional sliding window manner for all local pixel patches is inefficient due to the considerable redundant convolution computation. To accelerate the testing process for each testing image, we present a fast forwarding technique in DCNN framework. Instead of preforming DCNN forwarding in each pixel patch, the proposed method performs convolution computation in the entire testing image, with a modified sparse convolution kernel. This technique almost eliminates all redundant convolution computation compared to the conventional pixel-wise classification, which significantly accelerates the DCNN forwarding procedure. Experimental result reports the proposed method only requires around 0.1 s to detect lung cancer cells in a 512×512 image, while the state-of-the-art DCNN requires around 40 s.

To sum up, we propose a novel DCNN based model for lung cancer cell detection in this paper. Our contributions are summarized as three parts: (1) We built up a deep learning-based framework in lung cancer cell detection with modified sliding window manner in both training and testing stage. (2) We modify the training strategy by only acquiring weak annotations in the samples, which decreases both labeling and training cost. (3) We present a novel accelerated DCNN forwarding technology by reducing the redundant convolution computation, accelerating the testing process several hundred times than the traditional DCNN-based sliding window method. To the best of our knowledge, this is the first study to report the application of accelerated DCNN framework for lung cancer cell detection.

2 Methodology

Given an input lung cancer histopathological image I, the problem is to find a set $D = \{d_1, d_2, \ldots, d_N\}$ of detections, each reporting the centroid coordinates for a single cell area. The problem is solved by training a detector on training images with given weakly annotated ground truth information $G = \{g_1, g_2, \ldots, g_M\}$, each representing the manually annotated coordinate near the center of each cell area. In the testing stage, each pixel is assigned one of two possible classes, *cell* or *non-cell*, the former to pixels in cell areas, the latter to all other pixels. Our detector is a DCNN-based pixel-wise classifier. For each given pixel p, the DCNN predicts its class using raw RGB values in its local square image patch centered on p.

2.1 Training the Detector

Using the weakly annotated ground truth data G, we label each patch centered on the given ground truth g_m as positive(*cell*) sample. Moreover, we randomly sample the negative(*non-cell*) samples from the local pixel patches whose center are outside of the boundary of positive patches. The amount of negative sample patches is the same as the positive ones. If a patch window lies partly outside of the image boundary, the missing pixels are fetched in the mirror padded image.

For these images, we only feed very few patches into the proposed model for training, therefore extremely accelerating the training stage. Besides, this technique also partly eliminates the effect of over-fitting due to the under-sampling usage of sample images (Fig. 1).

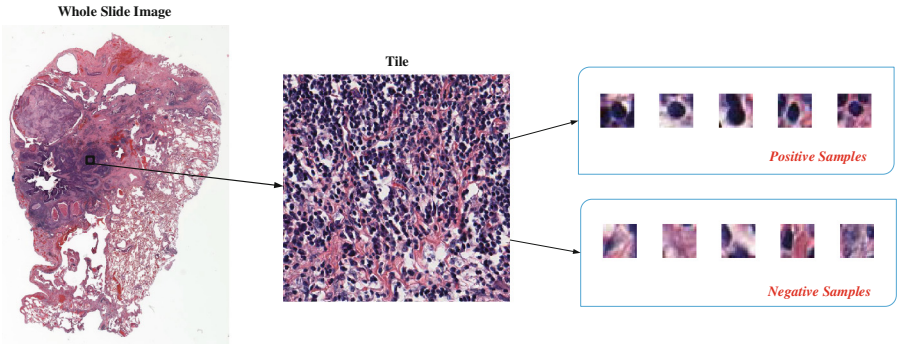

Fig. 1. The illustration of generation of training samples: (1) Tiles are randomly sampled from the whole slide images. (2) The sampled tiles are manually annotated by well-trained pathologists, which construct the weakly annotated information. (3) We only feed the local pixels patches center on the annotated pixels and the randomly sampled non-cell patches of the same amount as the cell ones.

2.2 Deep Convolution Neural Network Architecture

Our DCNN model contains two pairs of convolution and max-pooling layers, followed by a fully connected layer, rectified linear unit layer and another fully connected layer as output. Figure 2 illustrates the network architecture for training stage. Each **convolution layer** performs a 2D-convolution operation with a square filter. If the activation from previous layer contains more than one map, they are summed up first and then convoluted. In the training process, the stride of **max-pooling layer** is set the same as its kernel size to avoid overlap, provide more non-linearity and reduce dimensionality of previous activation map. The **fully connected layer** mixes the output from previous map into the feature vector. A **rectified linear unit layer** is followed because of its superior non-linearity. The output layer is simply another fully connected layer with just two neurons(one for cell class, the other for non-cell class), activated by a softmax function to provide the final possibility map for the two classes. We detail the

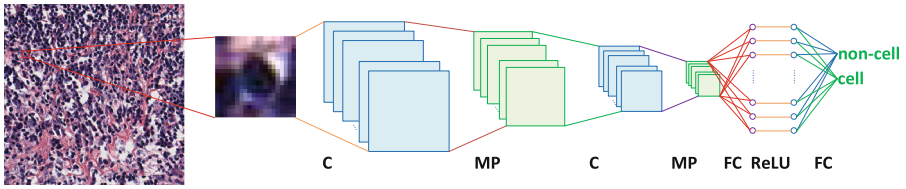

Fig. 2. The DCNN architecture used in the training process of the proposed framework. C, MP, FC, ReLU represents the convolution layer, max pooling layer, fully connected layer and rectified linear unit layer, respectively.

Table 1. Backward (left) and accelerated forward (right) network architecture. M: the number of patch samples, N: the number of testing images. Layer type: I - Input, C - Convolution, MP - Max Pooling, ReLU - Rectified Linear Unit, FC - Fully Connected

Type	Maps and neurons	Filter size	Filter num	Stride	Type	Maps and neurons	Filter size	Filter number	Stride
I	$3 \times 20 \times 20M$	-	-	-	I	$3 \times 531 \times 531N$	-	-	-
C	$20 \times 16 \times 16M$	5	20	1	C	$20 \times 527 \times 527N$	5	20	1
MP	$20 \times 8 \times 8M$	2	-	2	MP	$20 \times 526 \times 526N$	2	-	1
C	$50 \times 4 \times 4M$	5	50	1	C	$50 \times 518 \times 518N$	9	50	1
MP	$50 \times 2 \times 2M$	2	-	2	MP	$50 \times 516 \times 516N$	3	-	1
FC	$500M$	1	-	-	FC(C)	$500 \times 512 \times 512N$	5	-	1
ReLU	$500M$	1	-	-	ReLU	$500 \times 512 \times 512N$	1	-	-
FC	$2M$	1	-	-	FC(C)	$2 \times 512 \times 512N$	1	-	-

layer type, neuron size, filter size and filter number parameters of the proposed DCNN framework in the left of Table 1.

2.3 Acceleration of Forward Detection

The traditional sliding window manner requires the patch-by-patch scanning for all the pixels in the same image. It sequentially and independently feeds patches to DCNN and the forward propagation is repeated for all the local pixel patches. However, this strategy is time consuming due to the fact that there exists a lot of redundant convolution operations among adjacent patches when computing the sliding-windows.

To reduce the redundant convolution operations, we utilize the relations between adjacent local image patches. In the proposed acceleration model, at the testing stage, the proposed model takes the whole input image as input and can predict the whole label map with just one pass of the accelerated forward propagation. If a DCNN takes $n \times n$ image patches as inputs, a testing image of size $h \times w$ should be padded to size $(h + n - 1) \times (w + n - 1)$ to keep the size consistency of the patches centered at the boundary of images. The proposed method, in the testing stage, uses the exact weights solved in the training stage to generate the exactly same result as the traditional sliding window method does. To achieve this goal, we involve the k-sparse kernel technique [8] for convolution

and max-pooling layers into our approach. The k-sparse kernels are created by inserting all-zero rows and columns into the original kernels to make every two original neighboring entries k-pixel away. To accelerate the forward process of fully connect layer, we treat fully connected layer as a special convolution layer. Then the fully connect layer could be accelerated by the modified convolution layer. The proposed fast forwarding network is detailed in Table 1(right). Experimental results show that around 400 times speedup is achieved on 512×512 testing images for forward propagation (Fig. 3).

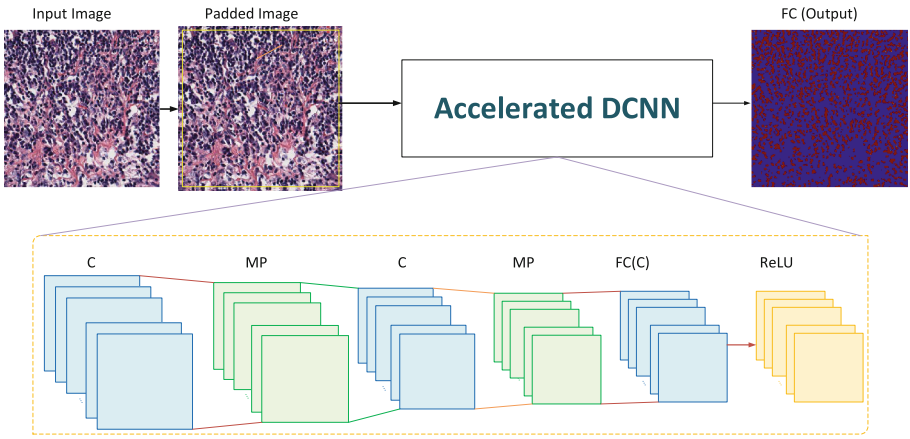

Fig. 3. The illustration of acceleration forward net: (1) The proposed method takes the whole image as input in testing stage. (2) The input image is mirror padded as the sampling process in the training stage. (3) The padded image is then put into the accelerated forward network which generates the whole label map in the rightmost. Note that the fully connected layer is implemented via a modified convolution layer to achieve acceleration.

3 Materials, Experiments and Results

3.1 Materials and Experiment Setup

Data Set. The proposed method is evaluated on part of the National Lung Screening Trial (NLST) data set [9]. Totally 215 tile images of size 512×512 are selected from the original high-resolution histopathological images. The nuclei in these tiles are manually annotated by the well-trained pathologist. The selected dataset contains a total of 83245 nuclei objects.

Experiments Setup. We partition the 215 images into three subsets: training set (143 images), validation set (62 images) and evaluation set (10 images). The evaluation result is reported on evaluation subset containing 10 images. We compare the proposed method with the state-of-the-art method in cell detection [4] and the traditional DCNN-based sliding window method [1].

Table 2. F_1 scores on the evaluation set

	1	2	3	4	5	6	7	8	9	10	Mean
MSER [4]	0.714	0.633	0.566	0.676	**0.751**	0.564	0.019	0.453	0.694	0.518	0.559
Proposed	**0.790**	**0.852**	**0.727**	**0.807**	0.732	**0.804**	**0.860**	**0.810**	**0.770**	**0.712**	**0.786**

Table 3. Mean time cost comparison on the evaluation set

	1	2	3	4	5	6	7	8	9	10	Mean
MSER [4]	37.897	29.000	37.172	43.332	42.806	37.843	28.548	41.570	38.346	37.012	37.353
Pixel-wise [10]	38.936	38.923	38.306	38.080	37.126	38.038	37.030	37.398	37.407	38.470	37.972
Proposed	**0.128**	**0.124**	**0.116**	**0.115**	**0.114**	**0.125**	**0.115**	**0.127**	**0.116**	**0.126**	**0.121**

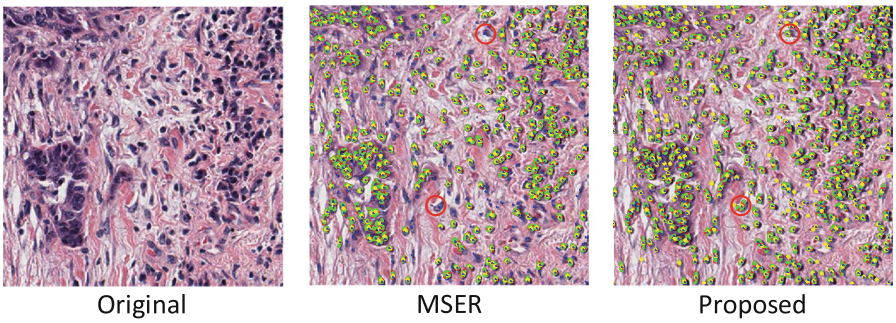

Fig. 4. Visual Comparison between the proposed method and MSER-based method [4]. The green area denotes the detected cell area by the corresponding method. Blue dots denote the ground-truth annotation. The proposed method is able to detect the cell area missed by the MSER-based method as denoted in red circle. Better viewed in $\times 4$ pdf (Color figure online).

3.2 Results

Training Time Cost. The mean training time for the proposed method is 229 s for the training set described below. The unaccelerated version with the same training strategy costs the same time as the proposed method. Besides, the state-of-the-art MSER-based method [4] costs more than 400000 s, roughly 5 days for training 143 images of size 512×512. The proposed method is able to impressively reduce several thousand times time cost of training stage than the state-of-the-art MSER-based method due to the proposed training strategy.

Accuracy of Testing. Table 2 reports the F_1 score metric comparison between the proposed method and MSER-based method. The proposed method outperforms the state-of-the-art method in almost all of the evaluation images in terms of F_1 scores. We also visually compares our results with the MSER-based method in Fig. 4. The proposed method detects almost all of the cell regions even in images with intensive cells.

Testing Time Cost. As shown in Table 3, the proposed method only costs around 0.1 s for a single 512×512 tile image, which is the fastest among the three methods. The proposed method accelerates the forwarding procedure around 400 times compared with the traditional pixel-wise sliding-window method, which is due to the accelerated forwarding technique.

4 Conclusion

In this paper, we propose an efficient and robust lung cancer cell detection method. The proposed method is designed based on the Deep Convolution Neural Network framework [10], which is able to provide state-of-the-art accuracy with only weakly annotated ground truth. For each cell area, only one local patch containing the cell area is fed into the detector for training. The training strategy significantly reduces the time cost of training procedure due to the fact that only around one percent of all pixel labels are used. In the testing stage, by utilizing the relation of adjacent patches, the proposed method provides the exact same results within a few hundredths time. Experimental results clearly demonstrate the efficiency and effectiveness of the proposed method for large-scale lung cancer cell detection. In the future, we shall attempt to combine the structured techniques [11–13] to further improve the accuracy.

Acknowledgments. The authors would like to thank NVIDIA for GPU donation and the National Cancer Institute for access to NCI's data collected by the National Lung Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

References

1. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
2. Bernardis, E., Stella, X.Y.: Pop out many small structures from a very large microscopic image. Med. Image Anal. **15**(5), 690–707 (2011)
3. Nath, S.K., Palaniappan, K., Bunyak, F.: Cell segmentation using coupled level sets and graph-vertex coloring. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 101–108. Springer, Heidelberg (2006)
4. Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A.: Learning to detect cells using non-overlapping extremal regions. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 348–356. Springer, Heidelberg (2012)
5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Computer Vision and Pattern Recognition (2014)
6. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)

7. Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S.: Deep learning based imaging data completion for improved brain disease diagnosis. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part III. LNCS, vol. 8675, pp. 305–312. Springer, Heidelberg (2014)
8. Li, H., Zhao, R., Wang, X.: Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification (2014). arXiv preprint arXiv:1412.4526
9. Team, N.L.S.T.R., et al.: The national lung screening trial: overview and study design. Radiology **258**, 243–253 (2011)
10. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding (2014). arXiv preprint arXiv:1408.5093
11. Huang, J., Huang, X., Metaxas, D.: Simultaneous image transformation and sparse representation recovery. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
12. Huang, J., Huang, X., Metaxas, D.: Learning with dynamic group sparsity. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 64–71. IEEE (2009)
13. Huang, J., Zhang, S., Li, H., Metaxas, D.: Composite splitting algorithms for convex optimization. Comput. Vision Image Underst. **115**(12), 1610–1622 (2011)