

A NOVEL FRAMEWORK TO INTEGRATE CONVOLUTIONAL NEURAL NETWORK WITH COMPRESSED SENSING FOR CELL DETECTION

¹*Yao Xue, ¹Nilanjan Ray, ²Judith Hugh, ²Gilbert Bigras*

1. Department of Computing Science, University of Alberta
2. Cross Cancer Institute, University of Alberta

ABSTRACT

The ability to detect certain types of cells in a microscopy image is important for a wide range of clinical applications. Cells often present huge variations in density and appearance, and often occupy only a small portion of an image. Consequently, general object detection methods in computer vision do not meet accuracy requirements: false / missed detections prevail. In this paper, we apply convolutional neural network (CNN) to regress a fixed length vector from a microscopy image. Then, L_1 minimization / compressed sensing (CS) recovers a variable number of cell locations from this fixed-length predicted vector. Our contribution in this work is combining CS with CNN to solve cell detection problem in a regression framework, which needs to handle a variable number of cells. Our method relies on the observation that the number of pixels indicating cell centroid locations is a tiny fraction of the total image size. Thus, we utilize this sparsity property by CS. The proposed method is evaluated with several state-of-the-art approaches on public cell datasets and it obtains superior or comparable performances.

Index Terms— Cell Detection, Convolutional Neural Network, Compressed Sensing, L_1 Minimization.

1. INTRODUCTION

1.1. Problem Definition

Automatic cell detection is to find whether there are certain types of cells in an input image (like microscopy images) and to localize these cells in the image. It is of great interest to a wide range of medical imaging tasks and clinical applications. An example is the diagnosis of breast cancer, where the proliferating (e.g. Ki67 positive) tumor cells is an important index associated with the severity of the disease. Fig.1 shows three examples of microscopy images with annotated cells.

1.2. Motivation

In recent years, computer vision has recorded significant progress in object detection and localization systems. How-

ever, cell detection and localization task is not simply a sub-task of a general object detection system in computer vision, which typically deals with extended objects, such as human and vehicles, etc. that occupy a significant portion of the field of view in the image. Cell detection and localization rather falls under the category of small object detection and localization (e.g. crowds counting), where individual objects occupy only a fraction of the entire field of view. Add to this challenge the factor that cells belonging to a particular category can appear very sparsely (only in tens), moderately densely (in tens of hundreds) or highly densely (in thousands) in a typical 2000-by-2000 pixel high resolution microscopy image. Additionally, significant variations in the appearance of cells are also present. Furthermore, background tissues may have a say in the classification of cells. These challenges render the cell detection/localization/counting problems far from being solved at the moment, in spite of significant recent progresses in computer vision.

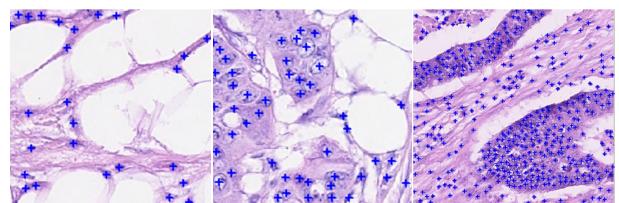


Fig. 1. Three typical example images with cells annotated by a cross on the centroid of each cell.

Prior to the introduction of Deep Learning (DL) methods in computer vision, cell detection and localization depended on segmentation of cells. An effective summary can be found here [1]. Recently, following the success of DL in computer vision tasks, cell detection has been approached by CNN-based prediction followed with ad-hoc post processing in [2]. Expectation maximization has been utilized within the DL framework in an end-to-end fashion for mitosis detection [3]. A cascaded network has been proposed for the same task more recently [4], which is in principle similar to a fully convolutional (FCN) model [5] that tries to predict a probability map indicating the density of cells. Accurately localizing cells from this heatmap requires non-trivial post-processing

This work is done when Yao Xue is with funding from China Scholarship Council.

steps comprising of various image processing techniques (for example: finding local maximum, thresholding) applied to the heatmap. Furthermore, the local maximum step often fails in several scenarios. For example, when cell density is high, cells are spatially close to each other, then it becomes quite difficult for local maximum to distinguish cells based on the heatmap predicted by FCN. Consequently, it misses many true-positives.

In this paper, we deviate from existing cell detection methods and cast the problem as a regression task from raw pixels. Treating the cell detection/localization as a regression task is tricky, because the length of output cell centroid locations is unknown. Thus, regressing on a fixed length vector is hardly an option. Our key observation for cell detection or localization is that the output space is quite sparse: an automated system needs to label a small fraction of the total pixels as cell centroid locations. Thus, using compressed sensing (CS) techniques, it is possible to turn this variable number of target detection problem into a fixed length vector regression problem. Thus, by introducing CS techniques, a real-valued signal y is built to carry the compressed location information of a variable number of target cells in a fixed vector length. The machine leaner can now regress this fixed length vector y and L_1 minimization can recover the cell centroid locations from y .

2. THE PROPOSED METHOD

2.1. Compressed Sensing into Cell Detection

During the past decade, compressed sensing (CS) [6] has emerged as a new framework for signal acquisition and reconstruction, mainly motivated by the rich theoretical and experimental results shown in [7], [8]. Under the premise of CS, an unknown and sparse signal is observed (sensed) through a limited number of linear observations. Under some conditions, it is possible to recover the unknown and sparse signal from these observations [7], [8].

Suppose, we have a N -dimensional signal f , which carries the location information of target cells. Because f is sparse, the CS theory guarantees that f could be recovered from linear observations y :

$$y = \Phi f, \quad (1)$$

provided the sensing matrix Φ has some properties [7], [8]. Φ is typically a $M \times N$ ($M \ll N$) Random Gaussian Matrix. The observation phase is summarized in Fig.2. If there are at most k non-zero entries in f , we say that f is k -sparse. The number of observations M is much less than N , and obeys: $M \geq kC_M \log(N)$, where C_M is a small constant greater than one.

As shown in Fig.2, for a training image, each cell location is converted to a positive element in the binary signal f . So, each positive element of f samples one column of the sensing

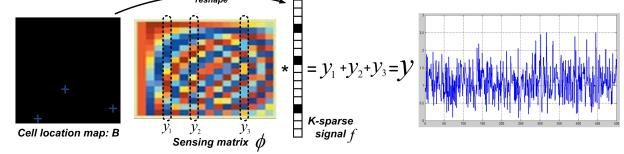


Fig. 2. Converting a binary pixel-wise annotation to N -length real-valued signal by Compressed Sensing.

matrix. This way the annotation mask B is encoded into the observed signal y .

Accurate recovery of f can be obtained from signal y , by solving the following L_1 norm convex optimization problem:

$$\hat{f} = \arg \min_f \|f\|_1 \quad \text{subject to} \quad y = \Phi f \quad (2)$$

After \hat{f} is recovered, it is easy to get the pixel-level annotations B . Since, \hat{f} is equivalent to pixel-level annotations B , by thresholding (T) on \hat{f} and reshaping back to B .

2.2. System Overview

The proposed detection framework consists of three major components: (1) Encoding phase of Compressed Sensing, (2) a CNN based regression model to capture the relationship between cell RGB image and the encoded signal y , (3) Decoding phase of Compressed Sensing. The flow chart of the whole framework is shown in Fig.3.

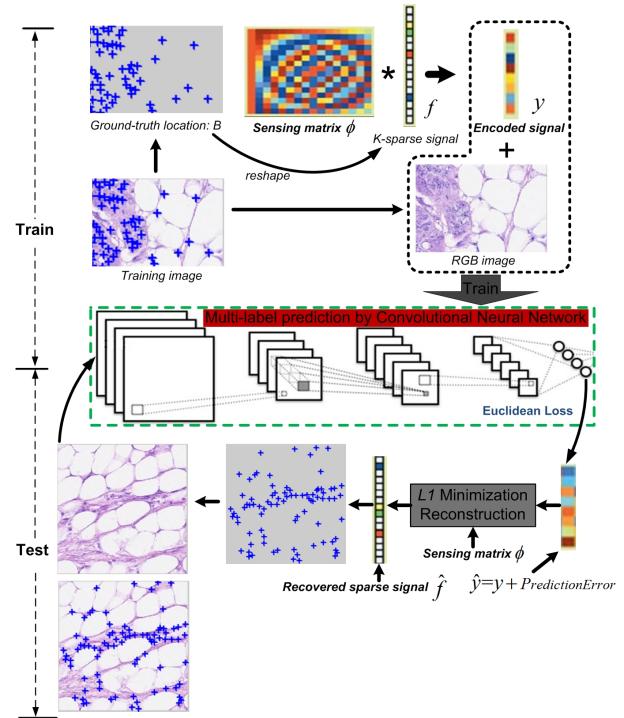


Fig. 3. System overview of the proposed framework.

During the training phase, a k -sparse signal f is generated by reshaping B into a vector. Then, encoded vector y is obtained by $y = \Phi f$. In our framework, Φ is a random Gaussian matrix. Thus, a training pair consists of a RGB image and the compact code y , on which a Convolutional Neural Network (CNN) is trained to work as a multi-label regression model. We employ the Euclidean loss function during training, because it is suitable for regression. Image rotations are performed on the training set for the purpose of data augmentation as well as making the system more robust to rotations.

In the test phase, the trained network is responsible for providing an estimated compact code \hat{y} for each test image. After that, we perform L_1 recovery, which predicts a estimated measurement \hat{f} with the given sensing matrix Φ and estimated compact code \hat{y} . For L_1 minimization reconstruction, we choose the dual augmented Lagrangian algorithm [9]. Finally, the final cell location map is obtained by reshaping from the estimated measurement \hat{f} .

3. EXPERIMENT AND PERFORMANCE

3.1. Datasets

First, we describe three cell datasets, on which the proposed method and other comparison methods are evaluated. The first dataset [10] involves 100 H&E stained histology images of colorectal adenocarcinomas. The second dataset [11] consists of 200 highly realistic synthetic emulations of fluorescence microscopic images of bacterial cells. The third dataset [12] comprise of 55 high resolution microscopic images of proliferative tumor cells area. For each dataset, the dotted annotation that represents the location of cells is shown in Fig.4. Details of datasets are summarized in Table.1.

Table 1. *Size* is the image size; *Ntr/Nte* is the number of images selected for training and testing; *AC* indicates the average number of cells; *MinC-MaxC* is the minimum and maximum numbers of cells.

Cell Dataset	Size	Ntr/Nte	AC	MinC-MaxC
Nuclei [10]	500×500	50/50	310.22	1-1189
Bacterial [11]	256×256	100/100	171.47	74-317
Ki67 Cell [12]	1920×2560	45/10	2045.85	70-4808

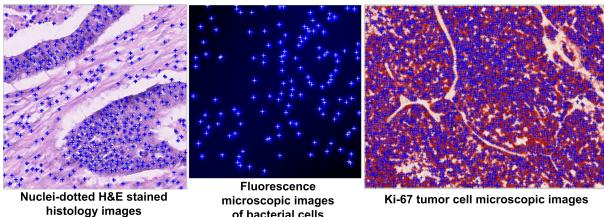


Fig. 4. Dataset examples and their dotted annotation.

3.2. Evaluation and Results

We adopt the criteria of the MITOSIS challenges [13] for method evaluation, a detection would be counted as true positive (TP) if its centroid Euclidian distance to a ground truth cell centroid is less than ρ . Otherwise, a detection is considered as false positives (FP). The missed ground truth cells are counted as false negatives (FN). In our experiments, ρ is set to the radius of the smallest cell in the dataset. Thus, only centroids that are detected to lie inside cells are considered correct. The results are reported in terms of Precision=TP/(TP+FP) and Recall=TP/(TP+FN) and F₁-score in the following sections. To evaluate, we carry out experimental performance comparison between the proposed method and two state-of-the-art cell detection approaches (presented in [14], [15]).

As mentioned in section 2.1, our method has a threshold T to apply on the recovered sparse signal \hat{f} before re-shaping it to binary image B . This threshold is used to do cell or non-cell binary classification and can be treated as a hyper-parameter during training. In "FCN-based" [15], there is also a threshold applied to the local probability-maximum candidate points to make final decision about cell or non-cell. Similarly, in the first step of "Le.detect" [14], researchers use a MSER-detector (a stability threshold involved here) to produce a number of candidate regions, on which their learning procedure determines that which of these candidates regions correspond to cells. In the first experiment, we analyze the three methods under Precision-Recall curve, by varying their own thresholds respectively.

Fig.5 presents Precision-Recall curves, and green lines show the F₁-score ranging from 0.3 (SouthWest) to 0.9 (NorthEast). All the three methods give reliable detection performances in the range of recall=[0.1-0.4]. After about recall=0.6, the precision of "FCN-based" [15] drops much faster. This can be attributed to the fact that "FCN-based" [15] works by finding local maximum points on a cell density map. However, the local maximum operation fails in several scenarios, for example when two cell density peaks are close to each other, or large peak may covers neighbouring small peaks. Consequently, it means that to obtain the same level of recall, "FCN-based" [15] actually provides many false detections.

Furthermore, it also can be observed that the proposed method has an improvement over "Le.detect" [14]. (red line clearly outperforms black line under varying recall values). This can be largely explained by the fact that traditional methods (no matter [14] or [15]) always try to predict the coordinates of cells directly on a 2-D image. But these coordinates are very sensitive to system prediction bias or error, considering the nature of cell detection that cells are small and quite dense in most cases. It is not surprising that "Le.detect" [14] will miss some cells and/or detect other cells in a wrong locations. In comparison, the proposed method integrates Com-

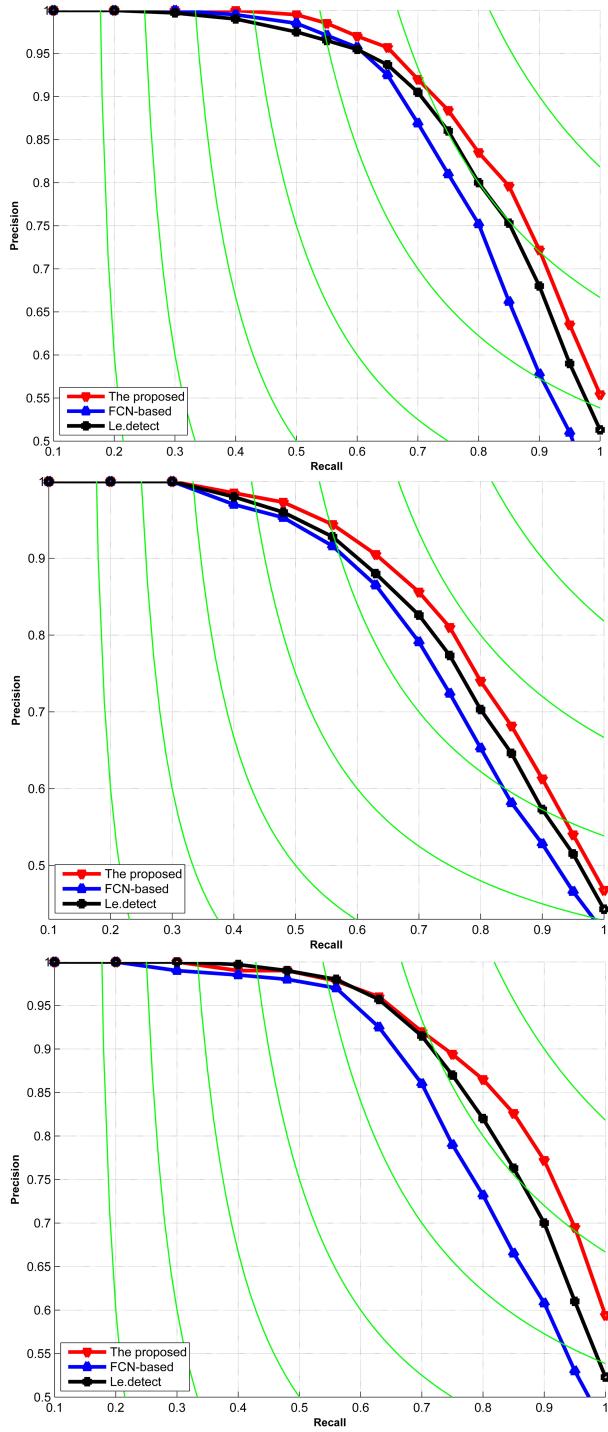


Fig. 5. Precision and Recall curve of three methods on Nuclei (top), Bacterial (middle) and Ki67-cell (bottom) datasets. Green lines show the F_1 -score ranging from 0.3 (SouthWest) to 0.9 (NorthEast).

pressed Sensing (CS) into a Convolutional Neural Network (CNN)-based regression model. We cast the problem of detecting variable number of small cells as a problem of predicting a fixed length signal. All the training and testing phase of CNN is based on this signal. And cell detection is solved by doing recovery from the signal. As a result, the proposed method gives more reliable detection performance as shown in Fig.5.

To get a better idea of the proposed method, we visualize a set of cell images with their detected cells and ground-truth cells in Fig.6. It can be observed that the proposed method is able to accurately detect most cells under a variety of conditions.

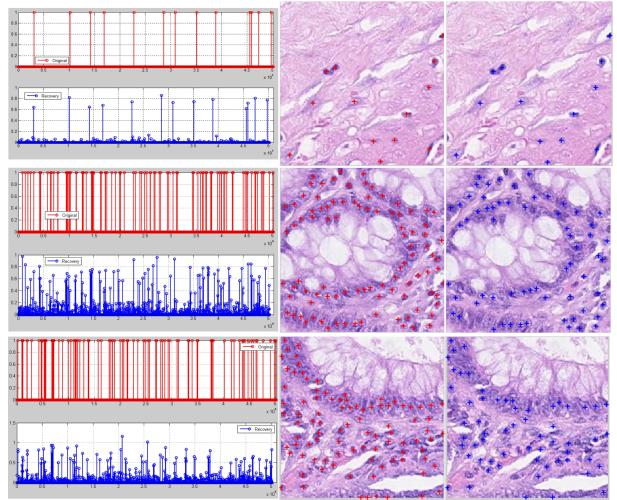


Fig. 6. Some detection results. Ground-truth: red, Prediction: blue. Left part shows the original sparse signal f and recovered \hat{f} that carry the location information of cells; right part shows the ground-truth and detected cells.

4. CONCLUSIONS

Considering the nature of cell detection and localization tasks (for example: variable number of target cells, small size, sparse/high density, etc), it is challenging to adapt classic object detection algorithms to cell detection. In this paper, we propose to integrate Compressed Sensing (CS) with Convolutional Neural Network (CNN) into a detection framework to solve the cell detection and localization problem. The proposed method successfully overcomes the challenges of cell detection problem, and is able to accurately detect cells by casting the problem of detecting a variable number of small cells as a problem of sensing and recovering a fixed length signal. Experiments have demonstrated that the proposed approach achieved superior performance compared with two state-of-the-art cell detection methods.

5. REFERENCES

- [1] Meijering Erik, “Cell segmentation: 50 years down the road [life sciences],” *IEEE Signal Process. Mag.*, 2012.
- [2] Dan C. Cirean, Alessandro Giusti, Luca M. Gambardella, and Jrgen Schmidhuber, “Mitosis detection in breast cancer histology images with deep neural networks,” *MICCAI*, 2013.
- [3] Albarqouni Shadi, Baur Christoph, Achilles Felix, Bela-giannis Vasileios, Demirci Stefanie, and Navab Nassir, “Aggnets: Deep learning from crowds for mitosis detection in breast cancer histology images,” *IEEE Trans. on Medical Imaging*, 2016.
- [4] Hao Chen, Qi Dou, Xi Wang, Jing Qin, and Pheng-Ann Heng, “Mitosis detection in breast cancer histology images via deep cascaded networks,” *AAAI*, 2016.
- [5] Jonathan Long, Evan Shelhamer, and Trevor Darrell, “Fully convolutional models for semantic segmentation,” *CVPR*, 2015.
- [6] David L. Donoho, “Compressed sensing,” *IEEE Trans. on Information Theory*, 2006.
- [7] Emmanuel Candes and Justin Romberg, “Practical signal recovery from random projections,” *IEEE Trans. on Signal Process*, 2005.
- [8] Emmanuel Candes, Justin Romberg, and Terence Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. on Information Theory*, 2006.
- [9] Ryota Tomioka, Taiji Suzuki, and Masashi Sugiyama, “Super-linear convergence of dual augmented lagrangian algorithm for sparsity regularized estimation,” *The Journal of Machine Learning Research*, 2011.
- [10] K. Sirinukunwattana, S.E.A. Raza, Y.W Tsang, I.A. Creeand D.R.J. Snead, and N.M. Rajpoot, “Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images,” *IEEE Transactions on Medical Imaging*, 2016.
- [11] Carlos Arteta, Victor Lempitsky, Alison Noble, and Andrew Zisserman, “Learning to count objects in images,” *NIPS*, 2010.
- [12] Yao Xue, Nilanjan Ray, Judith Hugh, and Gilbert Bi-gras, “Cell counting by regression using convolutional neural network,” *ECCV workshop on BioImage Computing*, 2016.
- [13] ICPR-2014, “Mitos-challenge,” <https://mitos-atypia-14.grand-challenge.org/home/>.
- [14] Carlos Arteta, Victor Lempitsky, J. Alison Noble, and Andrew Zisserman, “Learning to detect cells using non-overlapping extremal regions,” *MICCAI*, 2012.
- [15] Weidi Xie, J. Alison Noble, and Andrew Zisserman, “Microscopy cell counting with fully convolutional regression networks,” *MICCAI workshop on Deep Learning*, 2015.