

Tumor Cell Identification in Ki-67 Images on Deep Learning

Ruihan Zhang^{1,2}, Junhao Yang¹ and Chunxiao Chen^{1,*}

Abstract: The proportion of cells staining for the nuclear antigen Ki-67 is an important predictive indicator for assessment of tumor cell proliferation and growth in routine pathological investigation. Instead of traditional scoring methods based on the experience of a trained laboratory scientist, deep learning approach can be automatically used to analyze the expression of Ki-67 as well. Deep learning based on convolutional neural networks (CNN) for image classification and single shot multibox detector (SSD) for object detection are used to investigate the expression of Ki-67 for assessment of biopsies from patients with breast cancer in this study. The results focus on estimating the probability heatmap of tumor cells using CNN with accuracy of 98% and detecting the tumor cells using SSD with accuracy of 90%. This deep learning framework will provide an objective basis for the malignant degree of breast tumors and be beneficial to the pathologists for fast and efficiently Ki-67 scoring.

Keywords: Ki-67, breast cancer, convolution neural networks, single shot multibox detector.

1 Introduction

Ki-67, a protein that encoded by the MKI67 gene in human was first identified by Scholzer and Gerdes in the early 1980s [Gerdes, Lemke, Baisch et al. (1984)]. It is a hallmark to determine the growth fraction of a given cell population. Human Ki-67 protein expression is highly associated with tumor cell proliferation and growth, especially for carcinomas of the prostate, the breast and nephroblastoma tumors [Scholzen and Gerdes (2000)]. Currently, a trained laboratory scientist can give a score of tumor cells stained by the antibody by experience in diagnosis [Li, Jiang and Chen (2015)]. However, this method may not only bring subjective errors, but also a tremendous workload in measuring Ki-67.

Convolutional neural network (CNN), one of the essential tools for deep learning, is a feed-forward artificial neural network mostly based on convolutional layers and pooling layers. CNN is commonly used for image analyzing since LeNet by LeCun [LeCun, Bottou and Bengio (1998)]. After works of Krizhevsky [Krizhevsky, Sutskever, Hinton et al. (2012)], VGG [Simonyan and Zisserman (2014)], GoogleLenet [Szegedy, Liu, Jia et

¹ Department of Biomedical Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing, 211106, China.

² The University of Melbourne, Parkville Campus, Melbourne, Victoria, 3010, Australia.

*Corresponding Author: Chunxiao Chen. E-mail: ccxbme@nuaa.edu.cn.

al. (2015)] and Resnet [He, Zhang, Ren et al. (2016)], it exhibits superior performance in image classification, object detection and semantic segmentation [Zhao, Feng, Wu et al. (2017)]. Compared to classic feed-forward neural networks with similarly-sized layers, CNN has much fewer connections and parameters, and do not need feature extraction as it is trained.

Unlike image classification, detection requires localizing object within an image. Deep learning model based on object detection can be divided into two stages [Girshick, Donahue, Darrell et al. (2014)]. One is to generate region proposals, and another is to classify the regions and provide a confidence degree for each object. Some related work contains Fast R-CNN [Girshick (2015)], SPP [He, Zhang, Ren et al. (2014)], Faster R-CNN [Ren, He, Girshick et al. (2015)], R-FCN [Dai, Li, He et al. (2016)] and Mask R-CNN [He, Gkioxari, Dollár et al. (2017)].

Deep learning has achieved unprecedented performance on various tasks, especially in biomedical field. Wang et al. [Wang, Khosla, Gargeya et al. (2016)] segmented the whole breast cancer image into patches and provided tumor probability map based on patch's classification. He obtained an area under the receiver operating curve (AUC) of 0.925 for the task of whole slide image classification and a score of 0.7051 for the tumor localization task. Saha et al. [Saha, Chakraborty, Arun et al. (2017)] built an automatic scoring system for Ki-67 using gamma mixture model (GMM) with expectation-maximization and deep learning approach with results of 93% precision, 0.88% recall and 0.91% *f*-score value. Additionally, end-to-end detecting approaches based on deep learning, such as SSD [Liu, Anguelov, Erhan et al. (2016)], YOLO [Redmon, Divvala, Girshick et al. (2016)] and RON [Kong, Sum, Yao et al. (2017)], can directly predict the size, location and label of an object without any intermediate steps.

So, in case of scoring Ki-67, CNN and SSD can be induced to make the cell's recognition processes more automatic and thus significantly minimize the risk of incorrect classification in the clinic.

However, despite the attractive qualities of CNN, it is necessary to have much larger training sets. The practical analysis of Ki-67 shows that the limited labeled datasets might cause CNN to fall short, overfitting on the training set and producing poor accuracy on the test set. Generative Adversarial Networks (GAN) [Goodfellow, Pouget-Abadie, Mirza et al. (2014)], a novel way to train a generative model, has been successfully induced to generate more artificial samples for data augmentation to improve the efficiency of CNN [Zhang and Liu (2018)]. GAN consists of two adversarial models: a generative model which can be used to captures the tumor cells distribution, and a discriminative model which can be used to estimate the probability that tumor cells come from the training data. Therefore, GAN can be used to generate more labelled samples to improve the performance of classification in Ki-67 investigation.

2 Methodology

2.1 Ki-67 in breast cancer

Ki-67 is one of the essential biomarkers for evaluation of proliferation rate in breast cancer screening and grading. Ki-67 immunohistochemical image is shown in Fig. 1, which contains tumor cells, connected tumor cells, normal cells and background. In the

practical training process, the Ki-67 image is annotated into three patches (tiny image) by experienced researchers corresponding to normal cells, tumor cells and background.

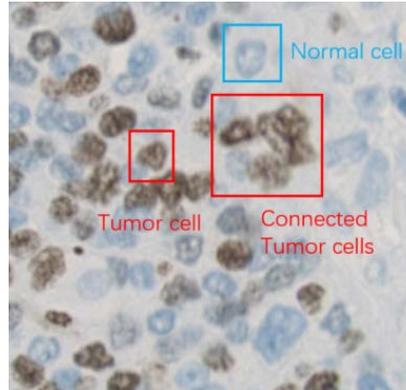


Figure 1: Cells in a Ki-67

2.2 Cell’s heatmap on CNN

CNN is a special kind of multi-layer neural network designed to recognize visual patterns directly from pixel images without feature’s extraction. In Ki-67 image, normal cell, tumor cell and background patches can be classified into probability heatmap according to steps in Fig. 2. The key processes of estimation include generating training dataset and then use CNN image classifier for pixel-wise classification to obtain probability heatmap.

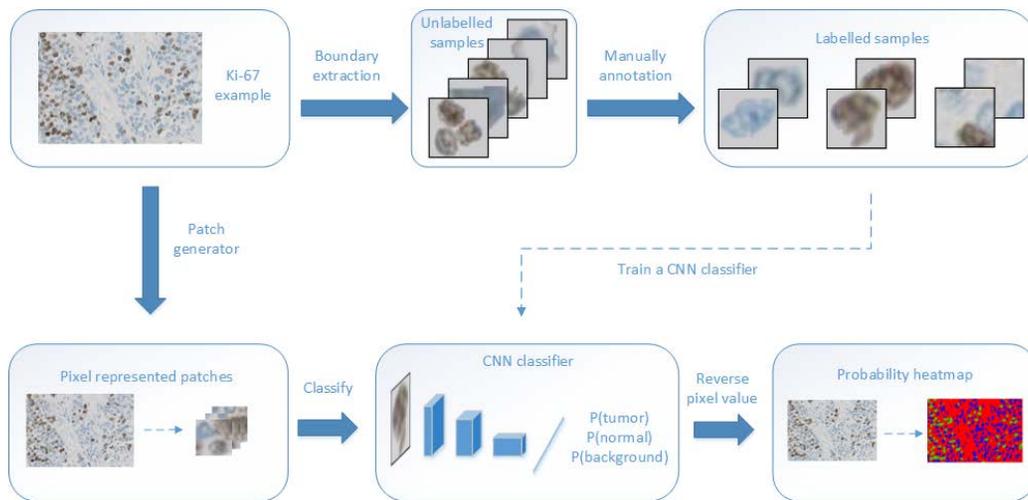


Figure 2: Cell classification overview

First of all, in order to reduce the workload of manual image segmentation, RGB (red, green and blue) color model microscopic image is transformed to HSI (hue, saturation and intensity) model. Multi-channel threshold segmentation method combining component of intensity to coarsely segment tumor cell, normal cell and background

patches in Ki-67 image. Then, manual accurate segmentation and annotation are needed for tumor cell, normal cell and background patches by experts. The size of each patch is set to be 56*56, which is close to the size of cell in the Ki-67 image.

Next, in order to expand the dataset, GAN method is applied to the training data set for higher accuracy. It contains two parts: Generator G which generates new samples and discriminator D which discriminates generated samples. It can be written as followed:

$$\min_G \max_D E_{x \sim P_r} [\log D(x)] + E_{x \sim P_G} [\log(1 - D(x))] \quad (1)$$

While training, generator can generate better samples due to the output of discriminator and the discriminator can be better through generated and real samples. Both of them will converge at Nash Equilibrium Point. After training these models, generator can be used to generate training dataset.

The CNN classifier (Fig. 3) is like VGG but smaller. In this convolutional neural network, every basic structure contains two convolutional layers with 3*3 filters and a maxpooling layer with 2*2 filter. After then, the feature map will halve the size of image and double the number of hidden layers. Repeat this structure for three times, a flatten layer and a dense layer is added to reduce the number of features. At last, a softmax layer converts result into a probability for each class.

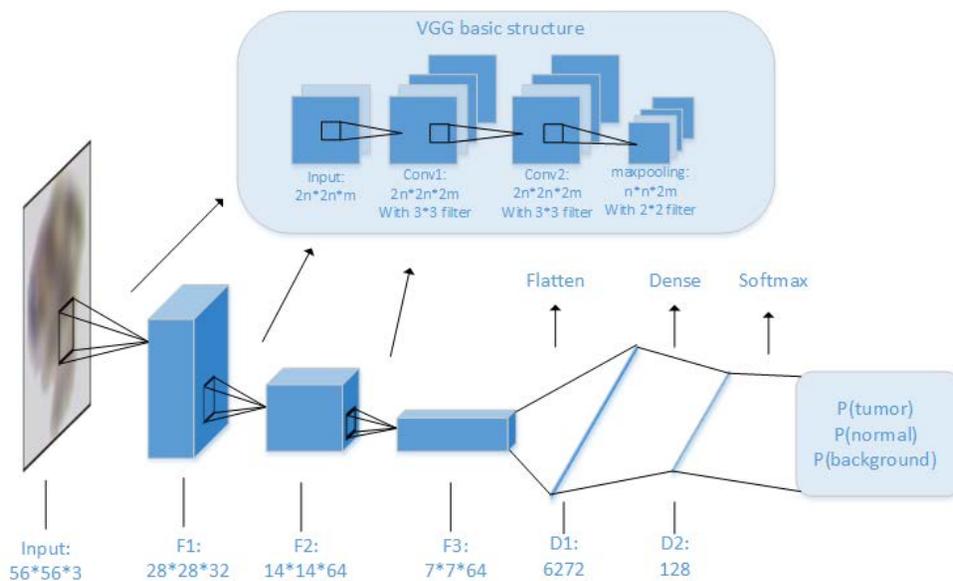


Figure 3: CNN classifier

Gradient descent is an optimization algorithm often used to train the deep learning models. Mini-batch gradient descent (MBGD) is a variation of the gradient descent algorithm that splits the training dataset into small batches that are used to calculate model errors and update model coefficients. MBGD chooses a small batch to train rather than the whole training set and exhibits good performance in speed and accuracy. In CNN classifier, cross entropy is defined as the loss function for training, which is shown

in Eq. (2). Here, n is the number of samples in a batch and m is the number of classes.

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{i,j} \log(p_{ij}) \quad (2)$$

In the process of testing, a set of labelled patches (tiny images) is generated from the Ki-67 image, and classification of pixels can be represented by the classification of their nearby patches. After classifying the patch set, a probability heatmap can be formed by reversing the values of these patches to its original location.

2.3 Cell's detection on SSD

Single shot multibox detector (SSD) is a unified framework for object detection in an image using a single deep neural network. It discretized the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. In order to quantity tumor cells and normal cells, SSD is used for cell detection. The workflow is shown in Fig. 4, which consists of generator and detector.

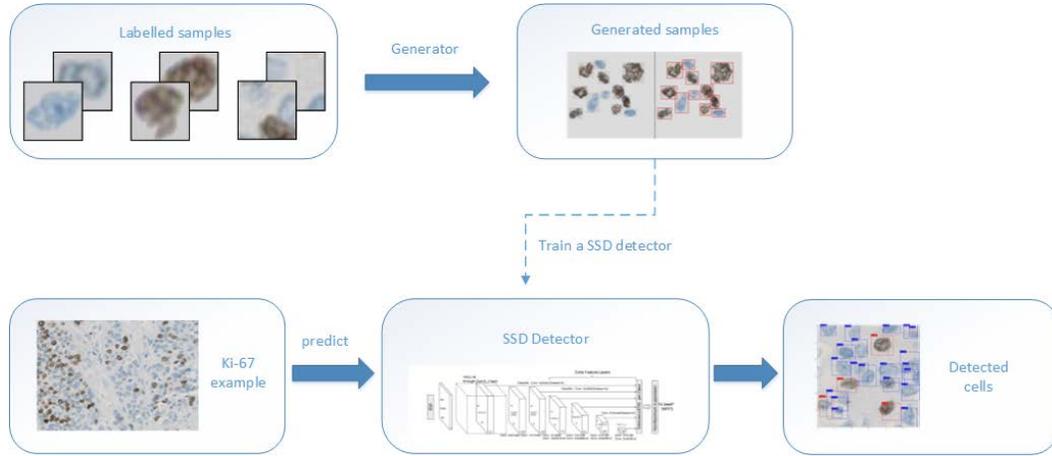


Figure 4: Cell detection on SSD

Generator uses the single cells images to generate artificial Ki-67 image with annotation. In this process, every cell scaled 56×56 should not be covered with others for avoiding distribution mismatch, but proper collision is allowed. Additionally, the ratio for each kind of classes should be balanced for higher recognition accuracy.

During cell's detection on SSD, the image is divided into several parts and predict $(C+4) \times k \times m \times n$ values. $(C+4) \times k$ filters for k anchors and C categories in one cell, $(C+4) \times k \times m \times n$ outputs for an $m \times n$ feature map. Different filters are applied to different scale/ratio anchors. The number of 4 means $[cx; cy; w; h]$, indicating the location and size of an object. Each layer is connected to the output layer to extract features, which is useful for detecting small object. The loss function of detection which requires bounding boxes prediction and cell classification is as follows:

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (3)$$

Here, L_{conf} reflect the loss of confidence for categories, which is shown as Eq. (4):

$$L_{conf}(x, c) = -\sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \text{ where } \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \quad (4)$$

where, x is an indicator tells whether there exist an object and c are predicted classes.

L_{loc} means the loss of locations and boundary box sizes, which is shown as Eqs. (5-7):

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}(l_i^m - \hat{g}_j^m) \quad (5)$$

$$\hat{g}_j^{cx} = \frac{g_j^{cx} - d_i^{cx}}{d_i^w} \quad \hat{g}_j^{cy} = \frac{g_j^{cy} - d_i^{cy}}{d_i^h} \quad (6)$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right) \quad \hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right) \quad (7)$$

Where, g is the ground truth box, d is the default box and l is the predicted box.

3 Results and discussion

The Ki-67 images come from <http://pan.baidu.com/s/1mhGmSOS>, a database of breast cancer Ki-67 immunohistochemistry. The 3,000 samples of each kind of classes (tumor cell, normal cell and background) identified by experts are augmented to 30,000 with generative adversarial networks (GAN) for improving the accuracy of classifier. In the experiment, 90% training samples and 10% validating samples randomly selected from 30000 samples for each kind of cells. The network was trained on GPU (GTX1050) while the algorithms of tumor classification and detection were done using Python and Tensorflow.

3.1 Classification and heatmap of tumor cells

We use the batch size 64 and learning rate 1e-3. Fig. 5 shows the accuracy and loss for each step when training. The results converge reasonably well while step is approximate 850 and epoch is 2. CNN classifier achieves an accuracy of around 98% and a loss under 0.1. Back to original 3,000 samples, the classifier can still achieve 98.2% accuracy.

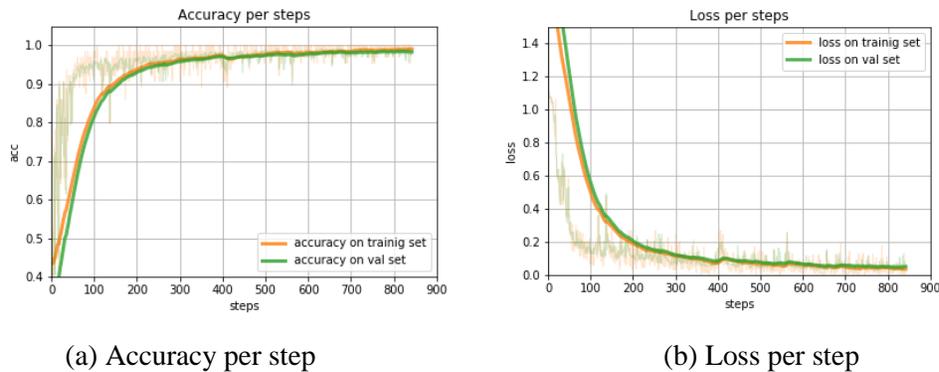


Figure 5: Paragraph selection model

Then, moving every 10 pixels to get a 56*56 patch from a Ki-67 image, an unclassified Ki-67 image with a size of 1600*900 will result in 14,400 patches. After classifying by CNN, reverse these patches with annotation back to pixels to form a probability heatmap.

Fig.6 shows some examples of the probability map which directly reflects the distributions of cell. In the probability heatmap, tumor cell, normal cell and background are marked by green, blue and red respectively. The ratio of tumor cell can be an indicator of the Ki-67 expression.

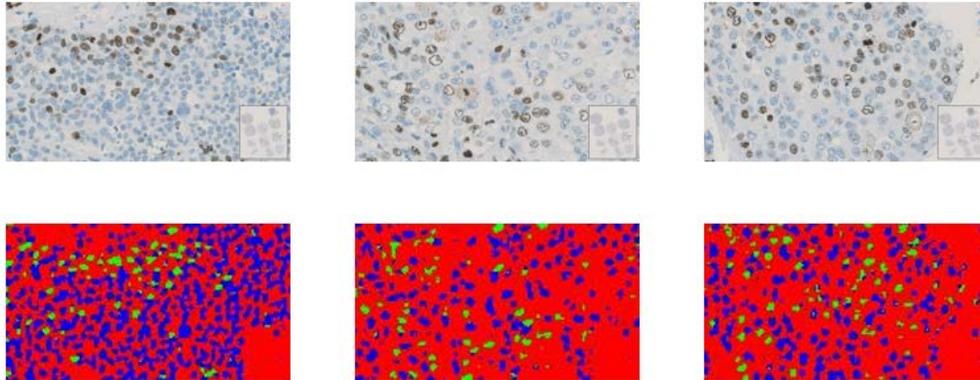
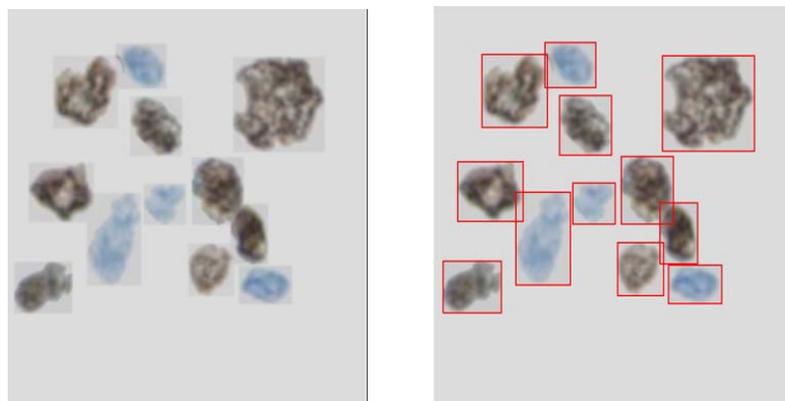


Figure 6: Examples of original images (above) and corresponding probability heatmaps (below)

3.2 Detection of tumor cells

In the process of generating artificial Ki-67 image for avoiding more samples required to be labelled by experts, the ratio for two cell classes is 0.5, scaling factor is 0.2 (scale single cell image to 1 ± 0.2 times) and less than 20% image collision is allowed. Fig. 7 gives an example of a generated Ki-67 image in cell detection training set (Fig. 7a). It contains 12 cells and overlap of cells is prohibited. Fig. 7b shows the bounding boxes of generated cells.



(a) Generated Ki-67 image (b) Generated image with bounding boxes

Figure 7: Example of generated Ki-67 image

Training for 10 epochs, the SSD approach can provide a detector with 90% accuracy for generated Ki-67 images. In Fig. 8, green boxes are true object bounding boxes and blue boxes are predicted object bounding boxes.

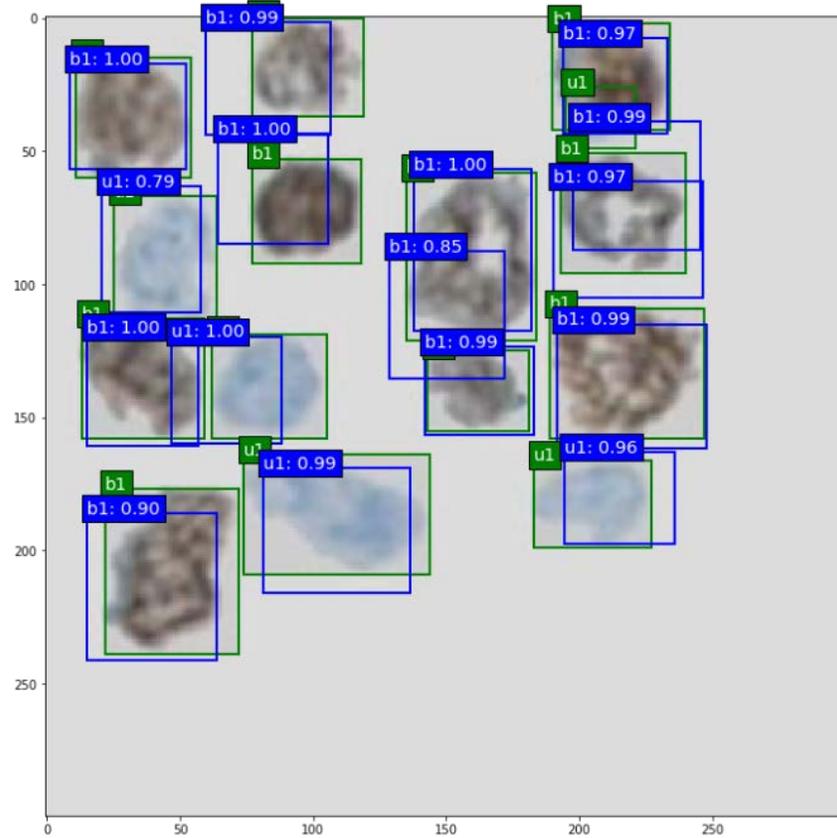


Figure 8: Example of cell detection

Applying SSD approach on real Ki-67 image, the detector can orient most of the cells and classify them correctly as shown in Fig. 9. The tumor cells are in the red boxes and the normal cells are in the blue boxes.

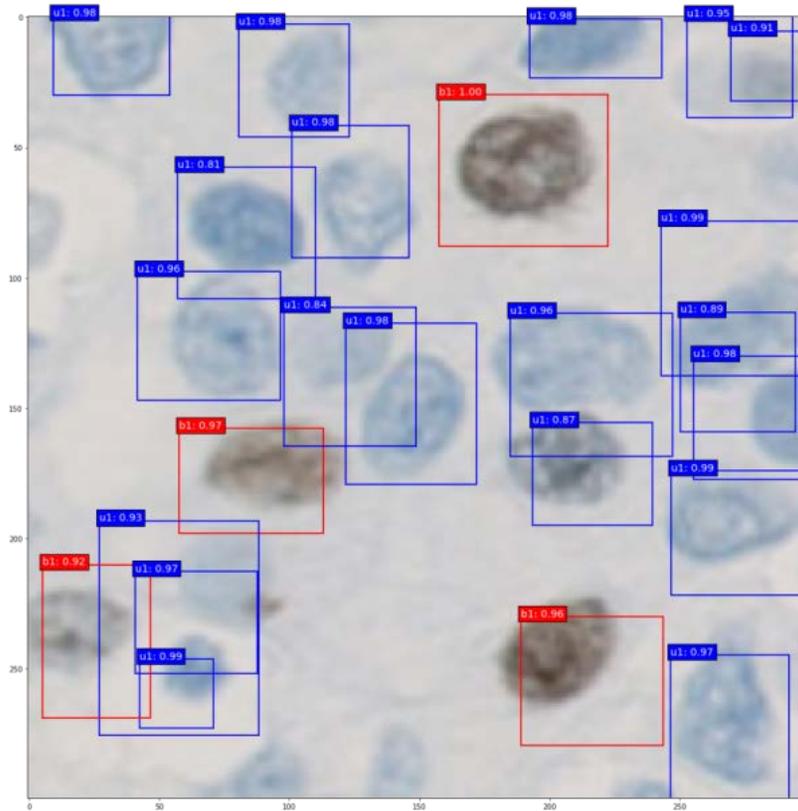


Figure 9: Orientation of cells in Ki-67

4 Conclusion and future work

Two automatic analysis approaches of the Ki-67 expression are discussed in this study. The first one is to build a probability heatmap to indicate the expression of Ki-67 with an accuracy of 98% using CNN image classifier. The other approach is to detect tumor cells and normal cells using SSD with an accuracy of 90%.

However, both approaches still have some shortages. The workflow of obtaining heatmap based on probability involves generating patch set, classifying these patches and reversing them with annotation back to pixels to form probability map and thus is very complex. Some of the newest methods, such as FCN [Long, Shelhamer and Darrell (2015)], Segnet [Badrinarayanan, Kendall and Cipolla (2017)] or Deeplab [Chen, Papandreou, Kokkinos et al. (2018)], may be good choices to generate heatmap with more effectiveness and efficiency. Object detection methods, such as Faster-RNN and YOLO v3 should also be considered to orient the cells in the future work to improve the accuracy. Additionally, in the augmentation of data set, the newly conditional adversarial net (cGAN) with some extra information, an extending model of GAN, outperform than

non-conditional adversarial nets [Mirza and Osindero (2014)]. Therefore, some aspects of the somatic of deep learning still needs to be further optimized.

References

- Badrinarayanan, V.; Kendall, A.; Cipolla, R.** (2017): Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495.
- Chen, L. C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A. L.** (2018): Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848.
- Dai, J.; Li, Y.; He, K.; Sun, J.** (2016): R-FCN: Object detection via region-based fully convolutional networks. *Advances in Neural Information Processing Systems*, pp. 379-387.
- Gerdes, J.; Lemke, H.; Baisch, H. E. I. N. Z.; Wacker, H. H.; Schwab, U. et al.** (1984): Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *Journal of Immunology*, vol. 133, no. 4, pp. 1710-1715.
- Girshick, R.** (2015): Fast R-CNN. *IEEE International Conference on Computer Vision*, pp. 1440-1448.
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J.** (2014): Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D. et al.** (2014): Generative adversarial nets. *Advances in Neural Information Processing Systems*, pp. 2672-2680.
- He, K.; Gkioxari, G.; Dollár, P.; Girshick, R.** (2017): Mask R-CNN. <https://arxiv.org/abs/1703.06870>.
- He, K.; Zhang, X.; Ren, S.; Sun, J.** (2014): Spatial pyramid pooling in deep convolutional networks for visual recognition. <https://arxiv.org/abs/1406.4729>.
- He, K.; Zhang, X.; Ren, S.; Sun, J.** (2016): Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.
- Kong, T.; Sun, F.; Yao, A.; Liu, H.; Lu, M. et al.** (2017): Ron: Reverse connection with objectness prior networks for object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5244-5252.
- Krizhevsky, A.; Sutskever, I.; Hinton, G. E.** (2012): Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 1097-1105.
- LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P.** (1998): Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324.

Li, L. T.; Jiang, G.; Chen, Q.; Zheng, J. N. (2015): Ki67 is a promising molecular target in the diagnosis of cancer. *Molecular Medicine Reports*, vol. 11, no. 3, pp. 1566-1572.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. et al. (2016): SSD: Single shot multibox detector. *European Conference on Computer Vision*, pp. 21-37.

Long, J.; Shelhamer, E.; Darrell, T. (2015): Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.

Mirza, M.; Osindero, S. (2014): Conditional generative adversarial nets. <https://arxiv.org/abs/1411.1784>.

Ren, S.; He, K.; Girshick, R.; Sun, J. (2015): Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, pp. 91-99.

Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. (2016): You only look once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788.

Saha, M.; Chakraborty, C.; Arun, I.; Ahmed, R.; Chatterjee, S. (2017): An advanced deep learning approach for Ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer. *Scientific Reports*, vol. 7, no. 1, pp. 3213.

Scholzen, T.; Gerdes, J. (2000): The Ki-67 protein: from the known and the unknown. *Journal of Cellular Physiology*, vol. 182, no. 3, pp. 311-322.

Simonyan, K.; Zisserman, A. (2014): Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S. et al. (2015): Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9.

Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A. H. (2016): Deep learning for identifying metastatic breast cancer. <https://arxiv.org/abs/1606.05718>.

Zhang, Q.; Liu, Y. (2018): Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks. <https://arxiv.org/abs/1806.07108>.

Zhao, B.; Feng, J.; Wu, X.; Yan, S. (2017): A survey on deep learning-based fine-grained object classification and semantic segmentation. *International Journal of Automation and Computing*, vol. 14, no. 2, pp. 119-135.