Multi-pixels Classification for nuclei segmentation in digital pathology based on deep machine learning

Yi De XIE
School of traffic and transportation, Beijing Jiaotong University, Beijing, 100044, China
deyxie@163.com

Abstract. The problem this paper solves is identifying the accurate boundary for nuclei in breast cancer digital pathology images. By doing the nuclei segmentation, the output image can be a reference for the doctors to make better diagnosis for breast cancer patients. Using convolutional neural network to solve this problem which is a popular deep learning method for image classification. By generating labelled patches from the original dataset for training and using a sliding window to cut out patches from testing image for classifying pixels in the testing image, the image level classification can be transferred to pixel level classification. A 2x2 receptive field can be classified to 16 classes, and each pixel can be calculated 4 time, which increase the pixel level TPR (true positive rate) from 0.868 to 0.909. And the boundaries are smother than traditional method that the receptive field is 1x1.

1. Introduction
According to World Cancer Report 2014, Breast cancer (BC) has become the leading type of cancer in women, accounting for 25 % of all cases. In 2012, it resulted in 1.68 million new cases and 522,000 deaths [1], which is a great threat to females' health. And in China, the mortality rate of breast cancer (BC) in China has increased by 38.91%. So making improvements in curing BC is worthwhile. As we know, Breast cancers are classified by several clinical grading systems. Different grades and stages of BC will lead to different treatments. Experts using the variability in the number, size, shape, staining of cells and nuclei in ROIs (Region of interests) to make diagnoses. For an ROI, most time the nuclei occurs with a large variance and amount, while the shape is tiny and diversified and sometimes the nuclei are overlapping or clustering.

In this report, the main method to address the nuclei segmentation problem is the convolutional neural network (CNN). CNN is one of the most practical deep machine technologies. It has been proved that CNN has inspiring performance on image classification problems. [2]

2. Pixel Classification Method
2.1. Convolutional Neural Network
The basic neural unit can be seen from figure1. The calculation process as followed. For normal neural networks, with more hidden layers, the network can handle non-linear classification problem. The output can be calculated as \( h_w = f(W^T x) = f(\sum_{i=1}^{n} W_i x_i + b) \)
But when processing an image. If we let each pixel connected to next layer, the parameter is too much which is hard to calculate. So based on convolution operation and human visual principles, scientists put forward CNN to let computer analysis images. And a basic structure of CNN can be seen in Figure1. This is the structure of LeNet, which is the most classic CNN and shows the basic construction of a convolutional neural network.

![Figure 1 Structure of LeNet. [4]](image)

From table 2.1 we can be informed the basic layers and the functions of each layer in CNN. The input image is \( X \), and we use \( H_i \) to represent the output of layer \( i \). So if \( H_i \) is a convolutional layer, the output can be described as formula 2.4 where \( W \) is the weights and \( B \) is the bias.

\[
H_i = f(H_{i-1} \otimes W_i + b_i)
\]

Operator \( \otimes \) means the convolutional kernel implement the convolution calculation to the output from the last layer. The kernel size \( k \times k \) is smaller than the image size \( w \times w \). A convolutional kernel is likened to a local receptor field, where spatially proximal inputs are mapped to a single value through a filter activation. Convolution layer minimize the number of parameters required \( k^2 << w^2 \), but still represent the main features of the original image.

| Main layers                      | Main Functions                          | Type                      | Notes                        |
|----------------------------------|-----------------------------------------|---------------------------|------------------------------|
| Input data                       | Data source                             | Image,                     | The colour images have three channels and grey images have one |
| Convolutional Layer              | Feature extraction                      | Convolution               | Learning Rate, Data Dimension |
| Subsampling Layer                | Feature pooling                         | Max Pooling, Average Pooling | Different pooling method, Data Dimension |
| Local Response Normalization     | lateral inhibition                      | LRN                       |                              |
| Loss Layer                       | comparing an output to a target and assigning cost to minimize | Softmax, Euclidean, Hinge loss, Accuracy, Sigmoid, Cross-Entropy, etc. | Choose proper loss type and norm |
| Activation Function              | Adding Non-linear factors               | ReLU, Sigmoid, Tanh, etc.  | ReLU is faster than other activation function |
| Fully Connected Layer            | Flap the feature map                    | Inner Production          |                              |

2.2. Image level to Pixel level

As we know, CNN works well for image classification which means CNN can label the whole input images to right classes. But for digital pathology images, it is worthless to label the whole image. And in most cases, there is no label for image. What people interested in is which class does the pixel in the images belongs to. In nuclei segmentation jobs, it is more obvious that the main challenge is
classifying the pixels accurately. [10]The tutorial [3] provide a method which Enables the CNN network do the pixel level classification. It is a process called patches generation. (see Figure 2)

From the mask detect the nuclei pixels and non-nuclei pixels. Making the detected pixel which size is 1x1 from step1 ant at the center of a patch. The patch size is 32x32 [6][7], and if the index of the pixel start from 1, then the position of the central pixel is (17,17). Crop the patch from the mask, if the central pixel is nuclei, label the patch as nuclei (label=1), if the central pixels non-nuclei, label the patch as non-nuclei(label=0). Go back to corresponding original colour pathology, find the same position in the image. Finally, cropping the patch out and labeling them.

Using this method we can generate patches for each image and each patch has a label tells the central pixel is nuclei or not. And this is the Ground truth. For every image we generate 2500 nuclei patches at most and 3200 non-nuclei patches at least. Because the annotated nuclei are incomplete, so we do the data augmentation which means rotate each patch 90 degrees. So for each image we have approximately 11420 patches. And about 1600000 patches totally. Figure 3 shows some examples of generated patches.

Now the patches can be used for training. After training, the CNN model will be able to output the label of each patch. Suppose we have a trained model, and the steps for generate the segmentation result are as follow:

Padding the original image from 2000x2000 to 2032x2032. So that every pixel on the original image will be calculated. From left top pixel, using a window which size is 32x32 to slide the image pixel by pixel. And crop the 32x32 sub-image out as a patch. So for a 2000x2000 image, there will be 4,000,000 patches. Feed these patches into the trained model, and the model will classify them into 0(non-nuclei) or 1(nuclei).The output label represent the class of center pixel in the patch. So that every pixel in the original image will have a class. Using these binary labels mark the result image. If all the patches have been calculated, we will get the result image show which pixels are nuclei and which are not.
And by this method we can do the classification for each pixel in a full-size image, and identify the nuclei.

3. Experiment Environment

3.1. Dataset

The dataset was provided by Janowczyk A. and Madabhushi A. [3]. There are 141 original digital images and 141 corresponding binary mask images. The images are 40x ROIs (region of interests) of estrogen receptor positive (ER+) breast cancer with nuclei in various shape and amount. The masks contain 12000 manually annotated nuclei which are done by the expert. We can notice that not all nuclei are segmented in the masks and that means in this dataset the ground truth is incomplete.

The size of original image and mask is both 2000x2000. The White pixels in masks is the nuclei. The first number of the image name represent the patient ID. Figure 5 shows the dataset.

3.2. Experiments Environment

Because training a CNN model is a calculation intensive job, so we can’t do all the experiments on my personal computer. So that we need to apply for Bluebear access. Bluebear is the HPC (High-Performance Computing) environment in University of Birmingham which can provide powerful computing resource.

4. Experiments and Evaluation

4.1. Implementation of Multi-pixels Classification Network
The network structure is based on AlexNet which is the champion model of Imagenet competition in 2012. AlexNet demonstrates the effectiveness of CNN in complex models, and then the GPU implementation makes the training in the acceptable time range to get the results, really promote the development of a supervised deep learning. Compare with the network in tutorial, AlexNet has a special layer called Norm layer. The function of this layer is smoothing. The AlexNet is deeper.[11]

The structure can be seen from Figure 6. There are 5 convolution layers, 3 pooling layers, 2 norm layers and 3 fully connected layer. And the hyper parameters of Multi-pixels Classification Network can be seen from Table 2.

![Figure 6 The structure of AlexNet.](image)

| Learning rate | Batch size | Input size | Weight decay | Max steps | Learning rate schedule |
|---------------|------------|------------|--------------|-----------|------------------------|
| 0.001         | 256        | 32 x 32 x 32 | 0.004        | 1,000,000 | Adagrad                |

4.2. Multi-pixels Classification

As we know AlexNet and VGGNet are able to discern 1000 classes images. So we can try another method. In this method, the judgment area is enlarged from 1 pixel to 4 pixels, which can be seen from Figure 7. So the class increase to 16.

![Figure 7 The different between single pixel classification Network and Multi-pixels classification Network in patches generating process.](image)

And when we output a testing image from this model. The patch will slide over the whole testing image, so each pixel will be calculated 4 times. The process can be seen from Figure 8 (a) is the recognized filed, and the red blocks mean the pixels we focus on. And Figure 8 (b) show the positional relation between sliding window and testing image. After sliding, from Figure 8 (c) we can know that the central pixel has been calculated four times. So it will have four labels. Then I sum the labels, so the max value of one pixel is 4 and the min value is 0. After all the pixels have been calculated, rescale them and draw the grey image.
4.3. Evaluation

F-score is an evaluation method that considered both precision value and recall value. TPR shows TPR shows the classify performance in positive data points and PPV shows how many predicted positive data points are misclassified. We can only calculate the TPR for pixel level evaluation. And for image level evaluation, we can calculate all these indexes.

Table 3 Comparison Between Multi-pixels Classification Net and popular DP net

|                | AlexNet | VGGNet | Multi-pixels Classification Net |
|----------------|---------|--------|---------------------------------|
| True Positive  | 37.2%   | 40%    | 35%                             |
| True Negative  | 50.8%   | 56%    | 54%                             |
| False Positive | 5.8%    | 1.5%   | 8%                              |
| False Negative | 6.5%    | 2.5%   | 3%                              |
| F-score        | 0.90395 | 0.9456 | 0.864198                        |
| TPR            | 0.75366 | 0.95176| 0.921053                        |
| PPV            | 0.86555 | 0.963855| 0.813953                       |

But as we can see that the performance of VGGNet and AlexNet is not as good as Multi-pixels Classification Net. But we should know that the VGG net has the highest accuracy in training patch. So we using MATLAB to draw which can been seen in Figure 9.

Figure 9 (a), (b), (c), (d) is the magnify drawn boundaries of Tutorial Net, AlecNet, VggNet, Multi-pixels Classification Net.
Although the accuracy of Multi-pixels Classification Net didn’t raise much in training process, it can provide a smoother boundary. Focus on area A, the multi-pixels Classification Net detects one nuclei that the original AlexNet didn’t. So at pixel level, it may make improvement at:

1. Boundary. More smoother boundaries will result in better segmentation performance.
2. Detecting. It may detect the nuclei which are hard to be detected by AlexNet or VGGNet.

Figure 10 magnified the result and made a comparison across (c) Multi-pixels Classification Net, (d) VGGNet, (e) AlexNet and Turorial Net where (a) is testing image and (b) is ground truth.

5. Conclusion
This paper using new training data which has large receptive field (2x2) and 16 classes. The accuracy is similar as normal AlexNet but the output is smoother. Now the output result can provide a support for clinical grading scheme of breast cancer diagnose. It can help the experts identify the nuclei faster in ROIs of digital pathology images. Compared with the traditional methods, the deep learning approach is more flexible, because there is no need to analyze the image details and build a mathematical model. CNN can abstract and learn the features from images. From partial ground truth, CNN can infer the class of every pixel in the input image.

References
[1] McGuire, S. (2016). World cancer report 2014. Geneva, Switzerland: World health organization, international agency for research on cancer, who press, 2015. Advances in Nutrition: An international Review Journal, 7(2):418–419.
[2] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324.
[3] Janowczyk, A. and Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of pathology informatics, 7.
[4] Wikipedia (2017a). Cross-validation (statistics) — wikipedia, the free encyclopedia. [Online; accessed 11-September-2017.
[5] Irshad, H., Veillard, A., Roux, L., and Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE reviews in biomedical engineering, 7:97–114.
[6] Lu, H., Zhang, Q., et al. (2016). Applications of deep convolutional neural network in computer vision. Data collection and analysis, (2016-01):1–17.
[7] Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.
[8] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[9] Xiandao, L. (2017). From lenet to alexnet. [Online; accessed 8-September-2017.
[10] Deng, L., Yu, D., et al. (2014). Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3–4):197–387.
[11] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105.

[12] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324.

[13] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10), pages 807–814.

[14] Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323.