

DETECTION AND CLASSIFICATION OF BREAST CANCER METASTASES BASED ON U-NET

LIN XU, CHENG XU, YI TONG, YU CHUN SU

GENERATIONSE SOFTWARE SOLUTIONS, INC., CANADA

ABSTRACT

This paper presents U-net based breast cancer metastases detection and classification in lymph nodes, as well as patient-level classification based on metastases detection. The whole pipeline can be divided into five steps: preprocessing and data argumentation, patch-based segmentation, post processing, slide-level classification, and patient-level classification. In order to reduce overfitting and speedup convergence, we applied batch normalization and dropout into U-Net. The final Kappa score reaches 0.91 on training data.

Index Terms— Deep learning, Medical imaging, U-Net, Camelyon17, Breast cancer detection

1. INTRODUCTION

Deep learning has been successfully applied to many tasks including image processing [1], sound/voice processing [2], language translation [2]. Recent progress shows that deep learning can also be applied into medical image processing, such as MRI [3], CT, biopsy, endoscopy. In the field of pathology, most of work still need to be done manually by pathologist using microscopes. Therefore, such tasks are time-consuming and the results vary significantly depends on the skill and condition of individual pathologist. In particular, cancer metastases detection is very challenging and requires extensive microscopic assessment by pathologists as the size of tumor region can be very small. An automated solution would potentially reduce the workload of pathologists as well as reduce the subjectivity in diagnosis [4]. The goal of Camelyon17 is automated detection and classification of breast cancer metastases in whole-slide images of histological lymph node. As the size of tumor regions can be very small (< 200 cells), pathologists are often required using high magnification (20X ~ 40X) for detecting tumor cells. This requirement significantly increases the workload for pathologist and machine learning approaches as well.

Although the goal of Camelyon17 Challenge is to determine a pN-stage for every patient in the test dataset based on sizes/numbers of positive lymph nodes in multiple slides (Macro: metastases > 2.0 mm; Micro: metastases > 0.2 mm / 200 cells, < 2.0 mm), the foundation of such tasks is to accurate segment cancer cells from normal cells. As the

size of biopsy digital image is very large, there is no way to segment a whole image in a single pass, patch-based training/test pipeline becomes standard for handling Gigapixel pathology images. There are two common approaches for addressing this segmentation problem. The first approach solves metastases segmentation problem by classification [5] [6]. The whole slide is divided into multiple overlapping patches. Patch with more than half cancer cells is labelled as a cancer patch (positive), otherwise as a normal patch (negative). Therefore, metastases segmentation can be generated by averaging over classification results for all patches. The second approach is based on patch segmentation [7]. It still generates many patches, then segment cancer cells from normal ones for each patch. In this work, we choose the follow second approach as metastases can be very small and may not be the major component in some patches.

2. U-NET ARCHITECTURE

Our segmentation (pixel-wise classification) is based on U-Net architecture [8] as shown in Figure 1.

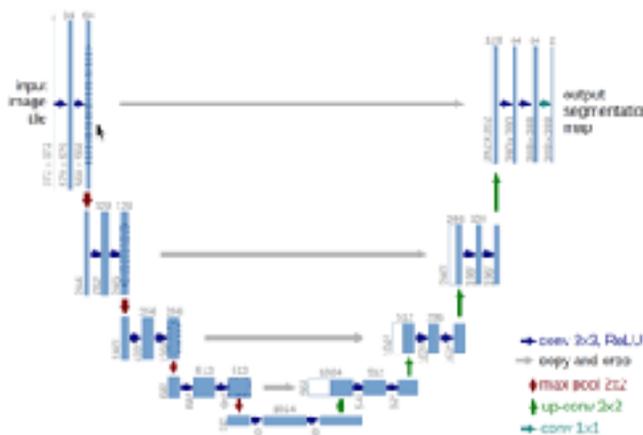


Figure 1: Original U-Net architecture

We made some necessary modifications in order to better fit our purpose. We changed the model input size from $572*572*1$ to $256*256*3$ in order to accept RGB images. Consequently, the following layers are also altered accordingly. Since Camelyon16 and Camelyon17 only provides less than 200 positive slides with masks, we also added batch normalization and dropout to the original implementation of U-Net in order to reduce overfitting and speedup convergence.

3. TRAINING PIPELINE

Our model training pipeline can be divided into 5 steps: preprocessing and data argumentation, patch-based segmentation, post processing, slide-level classification, and patient-level classification.

3.1. Preprocessing and data argumentation

As many deep learning-based approaches, many efforts have been made on data processing. We used all negative slides and positive slides from Camelyon16 and Camelyon17 with masks. Before taking samples from original slides, we generated regions of interest (ROI) by combining results from contours and Otsu's method [9] as shown in Figure 2.

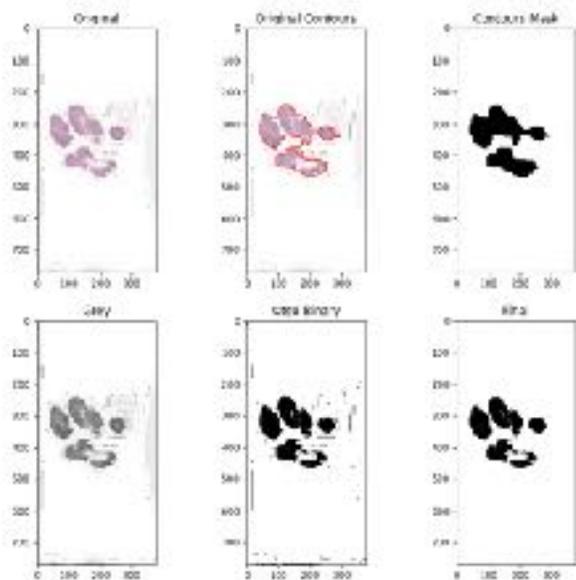


Figure 2: ROI detecting based on contours and Otsu

We selected N patches of size $512*512$ that covered all tumor regions, then randomly selected $4*N$ patches from normal regions from tumor slides as well as normal sides within region of interest (ROI). In the end, we oversampled tumor patches to $2*N$. Therefore, the negative/positive ratio was 2. In order to add variability into training data, heavy data argumentation was used such as flip, HSV color argumentation. Finally, all patches were resized to size $256*256$.

3.2. Patch-based segmentation

We chose to use TensorFlow [10] implementation of U-Net. All positive and negative patches were sampled in advance; data argumentation was performed during training on the fly. Training data samples are shown in Figure 3.

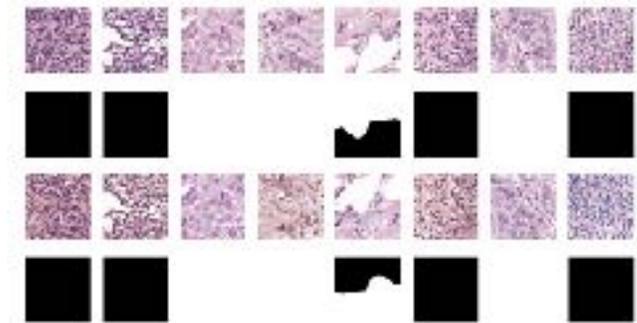


Figure 3: Training patch and corresponding mask without (top) and with argumentation (bottom)

We trained our model on Dell T630 server with 128G memory and 4 Titan X GPUs. Initially, the learning rate was set to 0.001, and reduced to 0.0001 after 20 epochs. The final segmentation performance is evaluated on independent validation slides (see Figure 4).

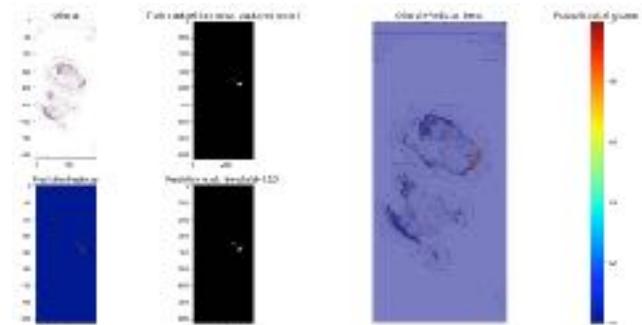


Figure 4: Sample segmentation results

3.3. Post processing

The segmentation produces a heat mask with probability for each pixel being cancer/normal for each slide. We empirically selected a threshold value to separate tumor cells from normal ones. We also conducted experiments with average filter to filter out noise. However, applying such filter may remove small isolated tumor regions. Therefore, we did not use average filter for slide-level classification.

3.4. Slide-level classification

After post processing, we obtained tumor regions. We adopted random forest approach with 12 hand crafted features mainly focused on the size of the tumor region. In order to reduce the bias introduced by manually selecting heat map threshold, our features are based on three different thresholds, namely 0.5, 0.7, 0.9.

Figure 5 shows confusion matrix on validation data. It is worth to note that classification accuracy for ITC (tumor regions with size < 0.2 mm or < 200 cells) is very low, and ITC almost always be misclassified as negative.

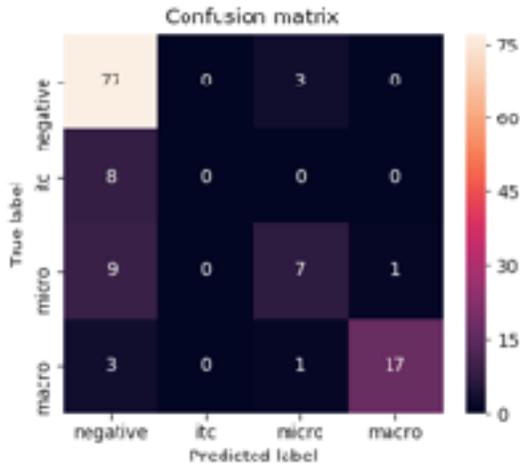


Figure 5: confusion matrix for slide-level classification on validation data (100 slides)

3.5. Patient-level classification

We did not perform additional training for patient-level classification. Rather, we followed simple Pathologic lymph node classification (pN-stage) standard defined in CAMELYON17 [4].

- pN0: No micro-metastases or macro-metastases or ITCs found.
- pN0(i+): Only ITCs found.
- pN1mi: Micro-metastases found, but no macro-metastases found.
- pN1: Metastases found in 1–3 lymph nodes, of which at least one is a macro-metastasis.
- pN2: Metastases found in 4–9 lymph nodes, of which at least one is a macro-metastasis.

4. SUMMARY OF UPDATE

We have done the data preprocessing again due to the changes in both CAMELYON16 and CAMELYON17 datasets recently.

5. RESULT AND CONCLUSION

We tested our pipeline on validation data and obtained Kappa score of 0.8734. The final Kappa score is computed using scripts provide by Camelyon17.

Obviously, there are many room for improvement. We noticed that ITC is very hard to predict. This may due to the patterns of isolated tumor cells are somewhat unique, and there are not enough training data for the model to capture such pattern. By random sampling, we did not obtain enough negative samples on such special regions. Therefore, we should perform important sampling or hard mining to introduce more, harder patches into training data and make the model more focus on mistakes. Another way to improve model prediction accuracy can be applying model assembly. We are working on training models with different architectures, even combining classification approach with segmentation.

Overall, computer-aided diagnosis system will be one of the most useful applications of deep learning. A technology that improves the detection of tumors for breast cancer can be easily adopted for detecting many other types of diseases.

6. REFERENCES

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 2009.
- [2] I. Sutskever, O. Vinyals and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," arXiv: 1409.3215v3, 2014.
- [3] S. Sarraf and G. Tofghi, "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Future Technologies Conference*, San Francisco, 2016.
- [4] CAMELYON17, "CAMELYON17 challenge," 2017. [Online]. Available: <https://camelyon17.grand-challenge.org/>. [Accessed 12 2018].
- [5] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng and M. C. Stumpe, "Detecting Cancer Metastases on Gigapixel Pathology Images," arXiv:1703.02442v2, 2017.
- [6] D. Wang, A. Khosla, R. Gargeya, H. Irshad and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," arXiv:1606.05718v1.
- [7] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," arXiv: 1606.00915v2, 2016.
- [8] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234-241, 2015.
- [9] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62 -- 66, 1979.
- [10] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz and L. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," arXiv: 1603.04467v2, 2016.