Supplementary Material: Multi-Stage Pathological Image Classification using Semantic Segmentation

Shusuke Takahama[1] Yusuke Kurose[1,2] Yusuke Mukuta[1,2] Hiroyuki Abe[1,3]
Masashi Fukayama[1,3] Akihiko Yoshizawa[3,4] Masanobu Kitagawa[3,5] Tatsuya Harada[1,2,6]

[1] The University of Tokyo [2] RIKEN [3] The Japanese Society of Pathology
[4] Kyoto University [5] Tokyo Medical and Dental University
[6] Research Center for Medical Bigdata, National Institute of Informatics
{takahama, kurose, mukuta, harada}@mi.t.u-tokyo.ac.jp

In this supplementary material, we discuss additional analyzes that could not be included in the main paper.

A. Model Structure Analysis

In this section, we performed detailed experiments of the evaluation with various model structures. In the main paper, we used GoogLeNet [7] as feature extractor and U-Net [5] as segmentation model. And we set the size of feature vectors of patches as 16. We evaluated the performance when changing the structure of the feature extractor model, the structure of the segmentation model, and the dimension of the feature vector. We evaluated the performance in terms of accuracy and PR-AUC on Stomach biopsy dataset.

We used optimization method1: Separate Learning. The learning rate of the feature extractor model is 1e-4, and the model is trained 30 epochs with a batch size of 128. The learning rate of the segmentation model is 1e-4, and the model is trained 50 epochs with a batch size of 32. We performed each method three times with different initial weights and evaluated the mean value and standard deviation.

A.1. Feature extractor model evaluation

First, we changed the structure of the feature extractor model. We evaluated the performance of VGG16 [6], GoogLeNet [7] and ResNet101 [3], which are commonly used in image recognition tasks. We set the dimension of feature vector 16 and used U-Net as the segmentation model. The result is shown in Table 1. It is revealed that GoogLeNet achieves the best performance compared with other models.

A.2. Segmentation model evaluation

Next, we changed the structure of the segmentation model. We evaluated the performance of FCN [4], SegNet [1], U-Net [5], PSPNet [8] and DeepLabv3+ [2], which are commonly used in semantic segmentation tasks. We set the dimension of feature vector 16 and used GoogLeNet as the feature extractor model. The result is shown in Table 2. U-net achieved the best classification performance. Since this classification task is relatively simple with two classes (tumor or normal), simple structure U-Net appears to be trained more stably than other complex models such as PSPNet or DeepLabv3+.

Table 1. Result of the feature extractor model evaluation.

Model	Accuracy (%)	PR-AUC (%)
VGG16	98.21±0.09	94.90±0.46
GoogLeNet	**98.53±0.03**	**99.30±0.02**
ResNet101	98.38±0.06	98.69±0.08

Table 2. Result of the segmentation model evaluation.

Model	Accuracy (%)	PR-AUC (%)
FCN	98.10±0.15	98.95±0.16
SegNet	98.10±0.12	98.19±0.09
U-Net	98.53±0.03	**99.30±0.02**
PSPNet	98.11±0.11	97.33±0.36
DeepLabv3+	**98.56±0.04**	99.18±0.03

Table 3. Result of the dimension of the feature vector evaluation.

Dimension	Accuracy (%)	PR-AUC (%)
4	98.29±0.03	99.17±0.02
16	**98.53±0.03**	**99.30±0.02**
64	98.36±0.03	99.15±0.01
256	98.37±0.03	99.12±0.03

Table 4. The result of the classification performance with 100 train slide data of Stomach biopsy dataset and with model using Segnet.

Method	Accuracy(%)	PR-AUC(%)
Classifier Only	93.81±0.07	92.28±0.10
Ours (Separate)	96.67±0.06	96.71±0.19
Ours (End-to-End)	**96.99±0.02**	**97.16±0.03**

A.3. Feature vector dimension evaluation

Finally, we changed the dimension of feature vector extracted from the feature extractor model. We used GoogLeNet as the feature extractor model and U-Net as the segmentation model. We evaluated the performance of four kind of the feature vector dimension, 4, 16, 64 and 256. They are smaller than 1024, which is the output dimension of the layer before the final layer of GoogLeNet, in order to reduce memory consumption. The result is shown in Table 3. Although the performance is the best at the dimension of 16, there was only a slight difference in performance depending on the dimension of the feature vector.

Considering this result, we used GoogLeNet, U-Net and feature vector dimension 16 in the experiments of the main paper.

B. Additional evaluation of two proposed methods

In this section, we conducted the experiment of two proposed optimization methods, Separate learning and End-to-end learning, with different settings from the main paper. When the two methods were compared in Sec.4.2.1 in the main paper, the performance of End-to-end learning was better but there was no significant difference between them. We thought the main reason for the slight difference was the high classification score of Separate learning, which left a limited room for improvement. We changed the experimental setting to intentionally lower the performance, and examined whether there is a difference between the two methods. Specifically, we conducted experiments with fewer training data (100 slides) and using Segnet as a segmentation model which showed relatively lower score in the experiment of Sec.A.2. Other settings was the same as the main paper.

Table 4 shows the results. In this situation the difference becomes noticeable. This result reveals that End-to-end learning yields better performance than previous "classifier only" method and Separate learning.

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:2481–2495, 2017.

[2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv preprint arXiv:1802.02611*, 2018.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[4] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2016.

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2015.

[7] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, 2014.

[8] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *CVPR*, 2017.