SUPPLEMENTARY MATERIAL FOR JOINT ATTENTION FOR MEDICAL IMAGE SEGMENTATION

Mo Zhang[1,2,3], *Bin Dong*[4,1], *Quanzheng Li*[5]

[1]Peking University, Center for Data Science, China;
[2]Peking University, Center for Data Science in Health and Medicine, China;
[3]Beijing Institute of Big Data Research, Laboratory for Biomedical Image Analysis, China;
[4]Peking University, Beijing International Center for Mathematical Research (BICMR), China;
[5]Harvard Medical School, Massachusetts General Hospital, MGH/BWH Center for Clinical Data Science, Center for Advanced Medical Computing and Analysis, Department of Radiology, USA.

1. EXPERIMENTS

Implementation Details. All models are implemented with TensorFlow on a Tesla V100 GPU. We use cross entropy loss, a mini-batch of 4 and Adam optimizer with weight decay 10^{-8} and the initial learning rate 10^{-3}. We train models for 100k iterations on ISIC-2017 dataset and 60k iterations on DRIVE dataset respectively. The additional network in point-wise attention is implemented by two 3×3 convolutional layers (the numbers of channel are 64,1 respectively).

2. ABLATION STUDIES

As shown in Table 1 and Table 1 of the paper, DenseUNet-JA outperforms DenseUNet-SA and DenseUNet-PWA in Dice and Accuracy, which are both more comprehensive indicators. More specifically, DenseUNet-SA and DenseUNet-PWA tend to generate unbalanced results. For instance, on DRIVE dataset, DenseUNet-SA gains a high TNR (0.9881) and a low Recall (0.7691) while DenseUNet-PWA obtains a low TNR (0.9828) and a high Recall (0.8087) conversely. However, the results of our DenseUNet-JA are more balanced, as joint attention leverages the strengths of self-attention and point-wise attention at the same time.

In order to search the optimal design of joint attention, we implement the nine structures in Fig.2 of the paper. The corresponding experimental results on ISIC-2017 dataset are shown in Fig.3. It is obvious that joint attentions in serial with point-wise attention first (JA-7,JA-8,JA-9) are better than the other two settings. Further, JA-9 gets the highest Dice score, suggesting that using two separate shortcuts is more useful and flexible. Note that all DenseUNet-JA in this paper denotes DenseUNet-JA-9 by default.

Bin Dong is supported in part by the NSFC under Grant 12090022, 12090020, 11831002.

3. REFERENCES

[1] Kai Hu, Zhenzhen Zhang, Xiaorui Niu, Yuan Zhang, Chunhong Cao, Fen Xiao, and Xieping Gao, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, 2018.

[2] Md Zahangir Alom, Chris Yakopcic, Mahmudul Hasan, Tarek M Taha, and Vijayan K Asari, "Recurrent residual u-net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, pp. 014006, 2019.

[3] Lei Mou, Li Chen, Jun Cheng, Zaiwang Gu, Yitian Zhao, and Jiang Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE transactions on medical imaging*, 2019.

[4] Yicheng Wu, Yong Xia, Yang Song, Donghao Zhang, Dongnan Liu, Chaoyi Zhang, and Weidong Cai, "Vessel-net: retinal vessel segmentation under multi-path supervision," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 264–272.

[5] Junyan Lyu, Pujin Cheng, and Xiaoying Tang, "Fundus image based retinal vessel segmentation utilizing a fast and accurate fully convolutional network," in *International Workshop on Ophthalmic Medical Image Analysis*. Springer, 2019, pp. 112–120.

[6] Wei Wang, Jiafu Zhong, Huisi Wu, Zhenkun Wen, and Jing Qin, "Rvseg-net: An efficient feature pyramid cascade network for retinal vessel segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 796–805.

Table 1. Quantitative results of different methods on DRIVE dataset.

Methods	Recall	TNR	Dice	Accuracy	Auc
Hu[1]	0.7772	0.9793	-	0.9533	0.9759
Alom[2]	0.7792	0.9813	**0.8171**	0.9556	0.9784
Mou[3]	**0.8126**	0.9788	-	0.9594	0.9796
Wu[4]	0.8038	0.9802	-	0.9578	0.9821
Lyu[5]	0.7940	0.9820	-	0.9579	0.9826
Wang[6]	0.8107	0.9845	-	0.9681	0.9817
DenseUNet	0.7635	0.9875	0.8048	0.9677	0.9802
DenseUNet-SA	0.7691	**0.9881**	0.8109	0.9687	0.9838
DenseUNet-PWA	0.8087	0.9828	0.8118	0.9673	0.9826
DenseUNet-JA	0.7907	0.9864	0.8164	**0.9690**	**0.9845**

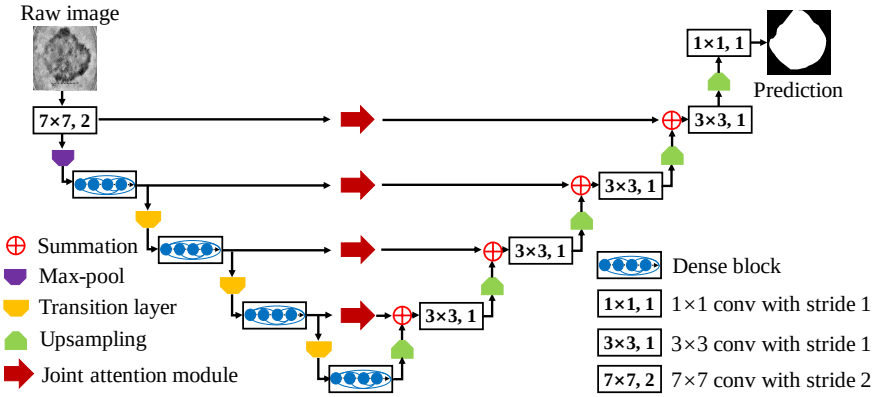

Fig. 1. The architecture of DenseUNet.

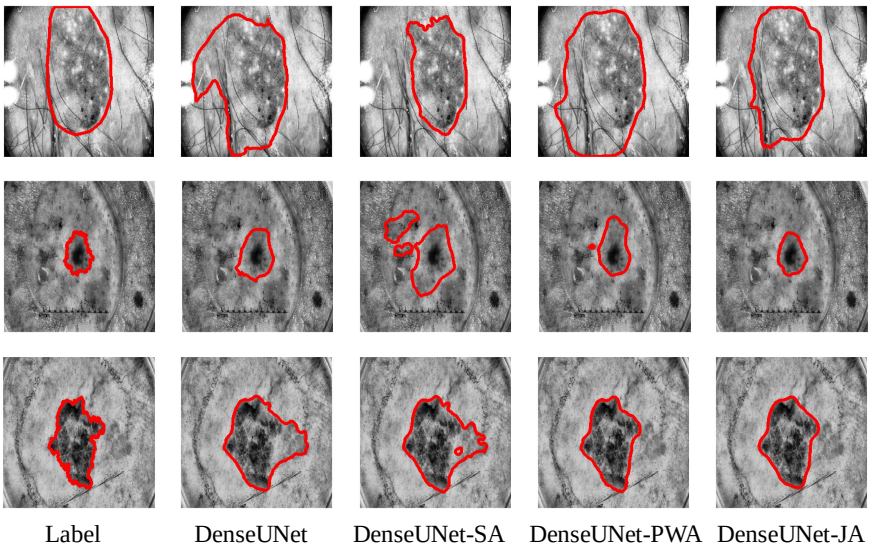

Fig. 2. Comparisons of segmentation maps of different models on ISIC-2017 dataset.

Fig. 3. Comparisons of Dice coefficient among different models on ISIC-2017 dataset. Baseline denotes DenseUNet, and JA-1 corresponds to the first structure of joint attention in Fig.2 of the paper.