Personalizing Federated Medical Image Segmentation via Local Calibration: Supplementary Material

Jiacheng Wang[1], Yueming Jin[2], and Liansheng Wang[1](✉)

[1] Department of Computer Science at School of Informatics,
Xiamen University, Xiamen, China
jiachengw@stu.xmu.edu.cn, lswang@xmu.edu.cn
[2] Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) and
Department of Computer Science, University College London
yueming.jin@ucl.ac.uk

1 Communication Cost

Our LC-Fed involves very little parameter communication costs and affects the efficiency slightly. The extra costs mainly comes from the HC module which gathers prediction heads from other sites, while the head with a single full-connection layer brings negligible model parameters compared with the entire framework.

We further intuitively analyze efficiency, by ***quantitatively*** counting **i)** the amount of model parameters during the communication, and **ii)** the time cost for communication under a general speed (100 Mbps) in the PMR dataset. Results are shown in Table 1. We see that LC-Fed only brings 13 KB extra parameters in the transmission, which are almost negligible compared to the entire framework of 12,000+ KB. As for the time cost, the additional communication cost of 0.001s is also slight compared to the entire training time (generally larger than one minute in each communication round). In contrast, LC-Fed excitingly brings a performance boost (3.09% IoU on PMR), and consistently improvements on other datasets. Therefore, our LC-Fed establishes a ***worthwhile trade-off*** with promising result enhancement yet few extra costs.

Table 1: Quantitative analysis of the communication costs on the PMR dataset.

Method	Para. (KB)	Time (s)	IoU (%)
FedAVG [3]	12440 (+12)	0.9718 (+0.001)	80.85 (+3.37)
FedRep [2]	12439 (+13)	0.9718 (+0.001)	81.13 (+3.09)
LC-Fed	12452	0.9728	84.22

2 Comparison Results

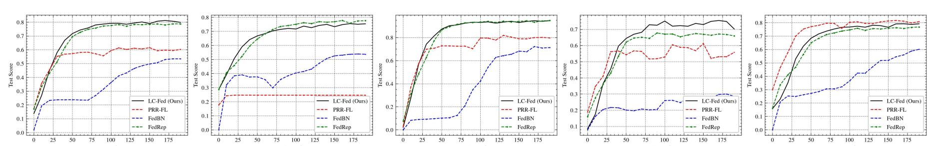
We split the datasets in patient-wise and calculate the metrics in patient-aware including their mean scores and the standard deviations. Table 2 shows calculated statistics on ours and advanced comparative methods. It includes the mean IoU scores and their standard deviations. The Wilcoxon test results have also been shown in the table.

Table 2: Detailed statistics on the EndoPolyp dataset.

Method	A	B	C	D
FedAVG[3]	64.56 (±24.42)	86.76 (±4.50)	61.28 (±31.48)	65.93 (±29.13)
FT[5]	65.95 (±23.36)	87.45 (±4.82)	60.63 (±32.21)	69.04 (±28.23)
FedRep[2]	67.23 (±23.82)	89.94 (±4.43)	61.17 (±32.94)	69.56 (±29.81)
LC-Fed	69.21 (±22.58)	88.51 (±4.82)	68.10 (±25.16)	76.68 (±26.77)

3 Learning Curves

The test score curves versus communication rounds have been shown in Fig. 1. The result shows that our method has the best average segmentation performance and the fastest learning speed. The improvement is particularly large on the Site C. It is also noticed that our model is slightly poorer than PRR-FL [1] on the Site D, while PRR-FL performs much worse on other sites and our model consistently performs well.



(a) Average (b) Site A (c) Site B (d) Site C (e) Site D

Fig. 1: Test score curves on the EndoPolyp dataset, including the (a) average score curve and (b-e) each site's score curve.

4 Future Direction

As this work focuses more on learning scheme instead of network architecture design, we consistently choose the most commonly-used framework in medical segmentation (U-Net [4]) for a direct and clear comparison. To comprehensively verify our method, we include various FL schemes across three different modalities for comparison. Moreover, our FL scheme based on channel selection and prediction calibration can be readily integrated to diverse deep model architectures, including graph-based networks, and even Transformer models. We believe that leveraging both has promising potential, which we shall explore in future work.

References

1. Chen, Z., Zhu, M., Yang, C., Yuan, Y.: Personalized retrogress-resilient framework for real-world medical federated learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 347–356. Springer (2021)
2. Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S.: Exploiting shared representations for personalized federated learning. arXiv preprint arXiv:2102.07078 (2021)
3. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
4. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
5. Wang, K., Mathews, R., Kiddon, C., Eichner, H., Beaufays, F., Ramage, D.: Federated evaluation of on-device personalization. arXiv preprint arXiv:1910.10252 (2019)