Supplementary Material for FedRepOpt: Gradient Re-parametrized Optimizers in Federated Learning

Kin Wai Lau[1,2], Yasar Abbas Ur Rehman[1], Pedro Porto Buarque de Gusmão[3], Lai-Man Po[2], Lan Ma[1], and Yuyang Xie[1]

[1] TCL AI LAB, Hong Kong, China
[2] Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China {kinwailau6-c}@my.cityu.edu.hk
[3] University of Surrey, United Kingdom

1 Network architecture used in FedRepOpt

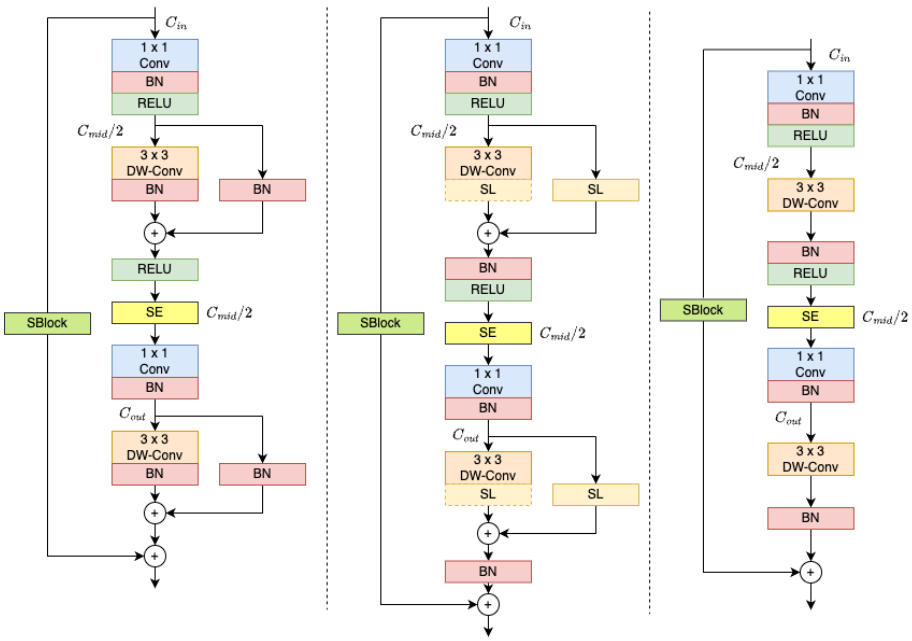

Fig. 1: Network architecture of the Fed-RepGhost-Tr (Left), Fed-CSLA-Ghost (Middle) and Fed-RepGhost-Inf / FedRepOpt-Ghost (Right) [1]. **SE:** Squeeze-and-excitation networks [3]. **SBlock:** Shortcut block [1]. **SL with dotted line:** Constant scaling layer [2]. **SL:** Trainable scaling layer [2]. **DW-Conv:** Depth-wise convolution.

In our proposed FedRepOpt FL framework, we consider two architectures (i.e., RepGhostNet and VGG-style) proposed in the original work on RepOpt [2]. Figure 1 demonstrates the block design of the Fed-RepGhost-Tr, Fed-CSLA-Ghost, and Fed-RepGhost-Inf/FedRepOpt-Ghost in our experiments. More precisely, Fed-RepGhost-Tr contains a parallel fusion layer (i.e., batch normalization (BN) layer), while the Fed-CSLA-Ghost replaces the BN layer with a trainable linear scaling layer. To follow the assumption in [2], where each branch only contains a linear trainable operator, the BN layer followed by the 3×3 depth-wise convolution is replaced by a constant scaling layer. Fed-RepGhost-Inf and FedRepOpt-Ghost follow the same structure mentioned in [2] that removes the parallel fusion BN layer in Fed-RepGhost-Tr. The other architecture used in our experiment is VGG-style, as shown in Figure 2. Similar to Fed-RepGhost-Tr, Fed-RepVGG-Tr contains a 1×1 and a BN branch in parallel. Fed-CSLA-VGG replaces the batch normalization layer with trainable/constant linear scaling layers. Fed-RepVGG-Inf and FedRepOpt-VGG are simplified versions of Fed-RepGhost-Tr without any multi-branch.

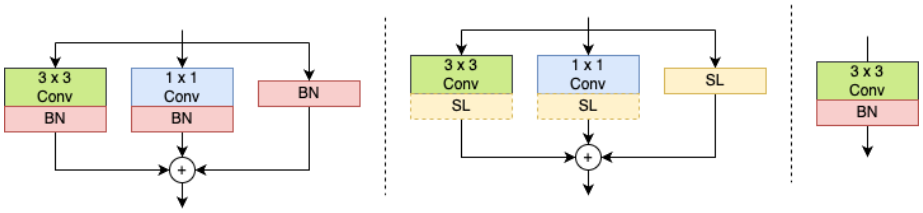

Fig. 2: Network architecture of the Fed-RepVGG-Tr (Left), Fed-CSLA-VGG (Middle) and Fed-RepVGG-Inf / FedRepOpt-VGG (Right) [2]. **SL with dotted line:** Constant scaling layer. **SL:** Trainable scaling layer.

References

1. Chen, C., Guo, Z., Zeng, H., Xiong, P., Dong, J.: Repghost: A hardware-efficient ghost module via re-parameterization. arXiv e-prints pp. arXiv–2211 (2022)
2. Ding, X., Chen, H., Zhang, X., Huang, K., Han, J., Ding, G.: Re-parameterizing your optimizers rather than architectures. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=B92TMCG_7rp
3. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)