A Simple Data Augmentation for Feature Distribution Skewed Federated Learning

Supplementary Material

A. More Details

A.1. Details of Dataset

There are 4 subsets for Office-Caltech-10 [4] and 6 subsets for DomainNet [28] and ProstateMRI [24]. These subsets from different domains, incurring feature distribution skew. Following previous work [22, 48], we employ the subsets as clients when conducting experiments on each dataset. Therefore, the number of clients for Office-Caltech-10, DomainNet, and ProstateMRI is 4, 4 and 6, respectively. Both Office-Caltech-10 and DomainNet consist of 10 classes each. ProstateMRI, on the other hand, is a binary segmentation task involving lesions and background. The number of samples in `train`, `val` and `test` for each client can be seen in Table 6.

A.2. Details of Baselines

There are some important hyper-parameters for some FL methods. For instance, the FedProx and FedProto have μ to control the contribution of an additional loss function. We empirically set μ to 0.001 for FedProx and 1 for FedProto on all datasets. Besides, FedAvgM has a momentum hyper-parameter to control the momentum update of the global model parameters, which is set to 0.9 for two image classification datasets and 0.01 for ProstateMRI. FedMix also has two hyper-parameters, *i.e.*, the batch size of mean images and λ to control images fusing. We adopted the reported configuration from the original paper. The batch size is set to 5 and λ is set to 0.1.

B. More Experiments

B.1. Stability of Our method

In this section, we further evaluate the stability of our method. To this end, we conducted two additional independent trials with different random seeds and reported the mean and standard derivation (std) of the average result of all clients across the three trials. Besides, we also performed statistical significance testing by conducting paired t-test between each method and our method, and we reported the corresponding p-value.

Tables 7 and 8 shows the results on three datasets, indicating FedRDN achieves a higher mean while maintaining a low std on three datasets compared to other methods. This demonstrates that the randomness introduced by Eq. (7) in FedRDN does not lead to performance instability. By statistical significance testing, we observed that the p-values across three datasets are less than 0.05, indicating that the performance improvement of FedRDN over other methods is significant.

Table 7. **Results of Stability** on Office-Clatch-10 [4] and DomainNet [28]. We report the mean and standard derivation of the average result of all clients across the three trials: (mean±std). Besides, we perform paired t-test between each method and our method, and reported the corresponding p-value.

Method	**Office-Caltech-10**		**DomainNet**	
	Accuracy	p-value	Accuracy	p-value
FedAvg	61.74±0.76	0.0001	42.66±1.14	0.0128
+ *norm*	61.55±0.40	0.0031	42.64±1.55	0.0160
+ *FedMix*	63.22±1.08	0.0028	43.05±1.43	0.0047
+ *FedRDN*	**69.18±0.63**	-	**43.94±1.35**	-

Table 8. **Results of Stability** on ProstateMRI [24]. We report the mean and standard derivation of the average result of all clients across the three trials: (mean±std). Besides, we perform paired t-test between each method and our method, and reported the corresponding p-value.

Method	**ProstateMRI**	
	Accuracy	p-value
FedAvg	90.44±0.99	0.0237
+ *norm*	90.60±0.94	0.0035
+ *FedRDN*	**92.19±0.86**	-

Table 6. **The number of samples in** `train`, `val` and `test` for each client on three datasets.

Samples	**Office-Caltech-10** [4]				**DomainNet** [28]						**ProstateMRI** [24]					
	Amazon	Catech	DSLR	WebCam	Clipart	Infograph	Painting	Quickdraw	Real	Sketch	BIDMC	HK	I2CVB	BMC	RUNMC	UCL
`train`	459	538	75	141	672	840	791	1280	1556	708	156	94	280	230	246	105
`val`	307	360	50	95	420	525	494	800	972	442	52	31	93	76	82	35
`test`	192	225	32	59	526	657	619	1000	1217	554	52	31	93	76	82	35