Table 6: Few-shot accuracy (%) on DomainNet. n-shot denotes training with n samples per class and per domain. Our results are marked in blue. The best results are **bolded**.

Method	CLIP-based	full	1-shot	2-shot	4-shot	8-shot	16-shot
Single Domain Tuning	✓	59.1	51.1	51.8	53.2	54.7	56.2
FedAvg (*ResNet-50*)	✗	54.7	-	-	-	-	15.1
FedAvg (*ViT-Base/16*)	✗	55.2	-	-	-	-	19.7
PromptFL	✓	63.2	51.4	51.8	55.2	57.6	61.2
FedCLIP	✓	60.3	50.8	51.2	52.1	53.4	54.6
Fed-DPT (ours)	✓	**68.4**	**55.4**	**57.2**	**60.3**	**62.7**	**64.5**

APPENDIX

Robustness to few-shot learning. One of the primary advantages of prompt learning is the robustness to few-shot scenarios. We investigate if our dual prompt tuning method retains this merit in the context of federated learning. Therefore, we conduct few-shot learning experiments on DomainNet, employing 1, 2, 4, 8, and 16 training samples per category and per domain. We evaluate the other CLIP-based methods with the same setting, yet only test 16-shot performance for FedAvg as it fails to yield reasonable results with fewer training samples. The corresponding results are summarized in Table 6. As is shown, CLIP-based methods exhibit superior robustness against few-shot learning than FedAvg, which again demonstrates the significant benefits of using parameter-efficient approaches. Also, our Fed-DPT consistently outperforms the baselines in few-shot learning.