Supplementary material for FedDSPG

To further validate the effectiveness of the proposed two-stage training strategy and conduct an in-depth analysis of the applicability of different generative modeling approaches in the context of Federated Domain Generalization (FDG), we supplement the following experiments.

.1. Two-Stage vs. End-to-End Training

Although the proposed two-stage pipeline introduces a marginal increase in engineering complexity compared with end-to-end training, it significantly improves both performance and interpretability by decoupling optimization objectives. To quantify this benefit, we design an end-to-end baseline (**Ours-end2end**) that replaces the first-stage fine-tuned domain-specific prompts (DSPs) with randomly-initialized soft prompts. Results in Table 1 show consistent accuracy drops across the three datasets and roughly 43% longer convergence iterations (286 vs. 200), validating the necessity of stage-wise optimization.

.2. Generative Backbone Choice: GAN vs. Diffusion

Federated learning imposes strict constraints on model size and communication. We therefore adopt a lightweight GAN (FedDSPG). To quantify the trade-off, we implement a diffusion variant (**Ours-diffusion**) based on DDPM with a compressed 10-step schedule. Table 1 reveals that while diffusion marginally improves accuracy (*e.g.*, +0.06% on `Office`), it demands $3.5\times$ more communication rounds and $5\times$ higher compute per forward pass, rendering it impractical for FL deployment.

Dataset	Office	DomainNet	PACS	Iterations
Ours-end2end	80.88	67.21	96.62	286
Ours-diffusion	85.42	75.72	97.55	**700**
FedDSPG	85.36	75.64	97.43	200

Table 1. Accuracies (%) of the Ours-end2end method and Ours-diffusion and FedDSPG on three datasets (the data above are the average values of all domains) and the training iterations required for convergence.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.