南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Harmonizing Generalization and Personalization in Federated Prompt Learning

Tianyu Cui [1] Hongxia Li [1] Jingya Wang [1] Ye Shi [1 *]
[1] ShanghaiTech University

ICML 2024

Background

• The ability of large pre-trained Vision-Language models (VLM) like CLIP and ALIGN to learn transferable representations across downstream tasks makes them a natural fit for integration with federated learning.
• Due to the millions of parameters in VLM, fine-tuning the entire model in federated learning leads to high communication costs and memory footprint issues.
• Prompt tuning addresses these challenges by adapting pre-trained models to diverse downstream tasks with a reduced parameter count, and its integration of federated learning has been explored in previous research.
• Currently, studies in FPL have not been thoroughly explored i**n terms of personalization and generalization**.

pFedPT

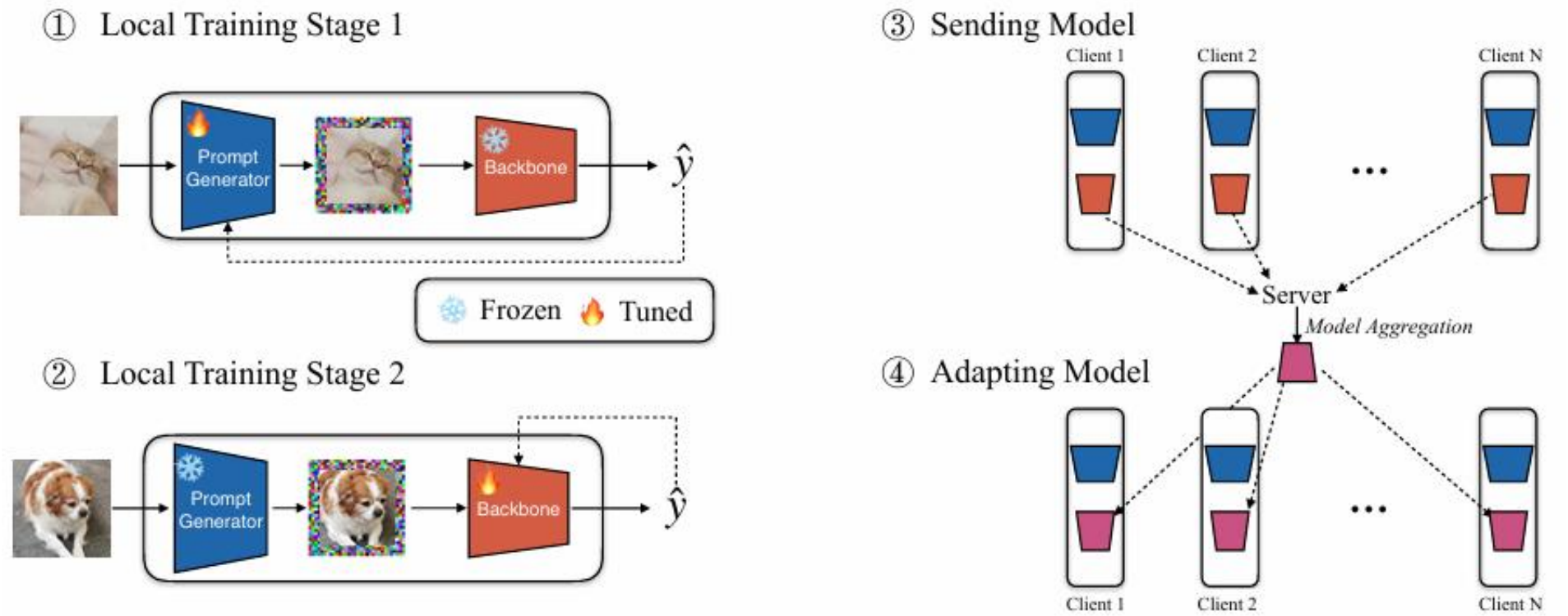


Figure 2: The pipeline of the pFedPT. $\hat{y}$ stands for the predicted logits of all classes. The dashed lines in steps 1 and 2 represent the loss backward for the model update. Each client contains a Prompt Generator, a set of personalized learnable parameters preserved locally, and a Backbone, which the server will aggregate with those of other clients. The raw image input will be added to a visual prompt (colored pixels padded around the image) and then passed into the backbone for prediction.

PROMPTFL

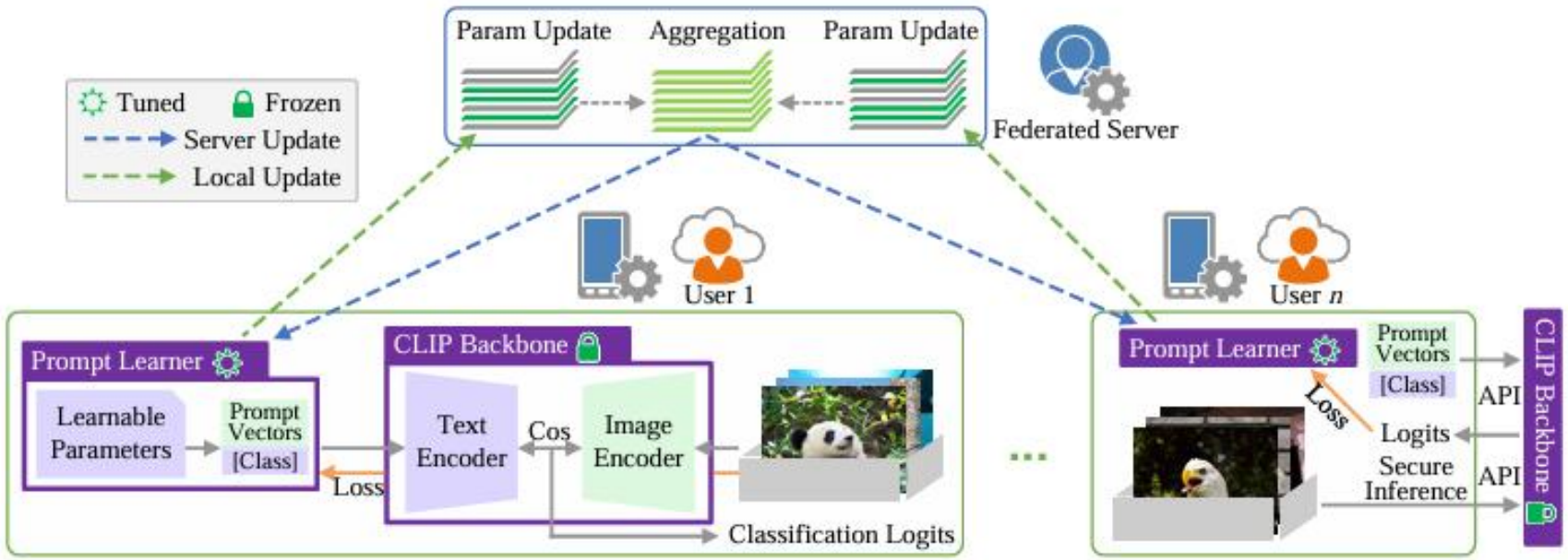


Figure 1: Framework and workflow of PROMPTFL. Each client includes a prompt learner (with only a small amount of trainable parameters) and an out-of-the-box CLIP (with backbone frozen). The federated server aggregates only the parameter updates of prompt learners over multiple users, and transmit the updated parameters back to each user.

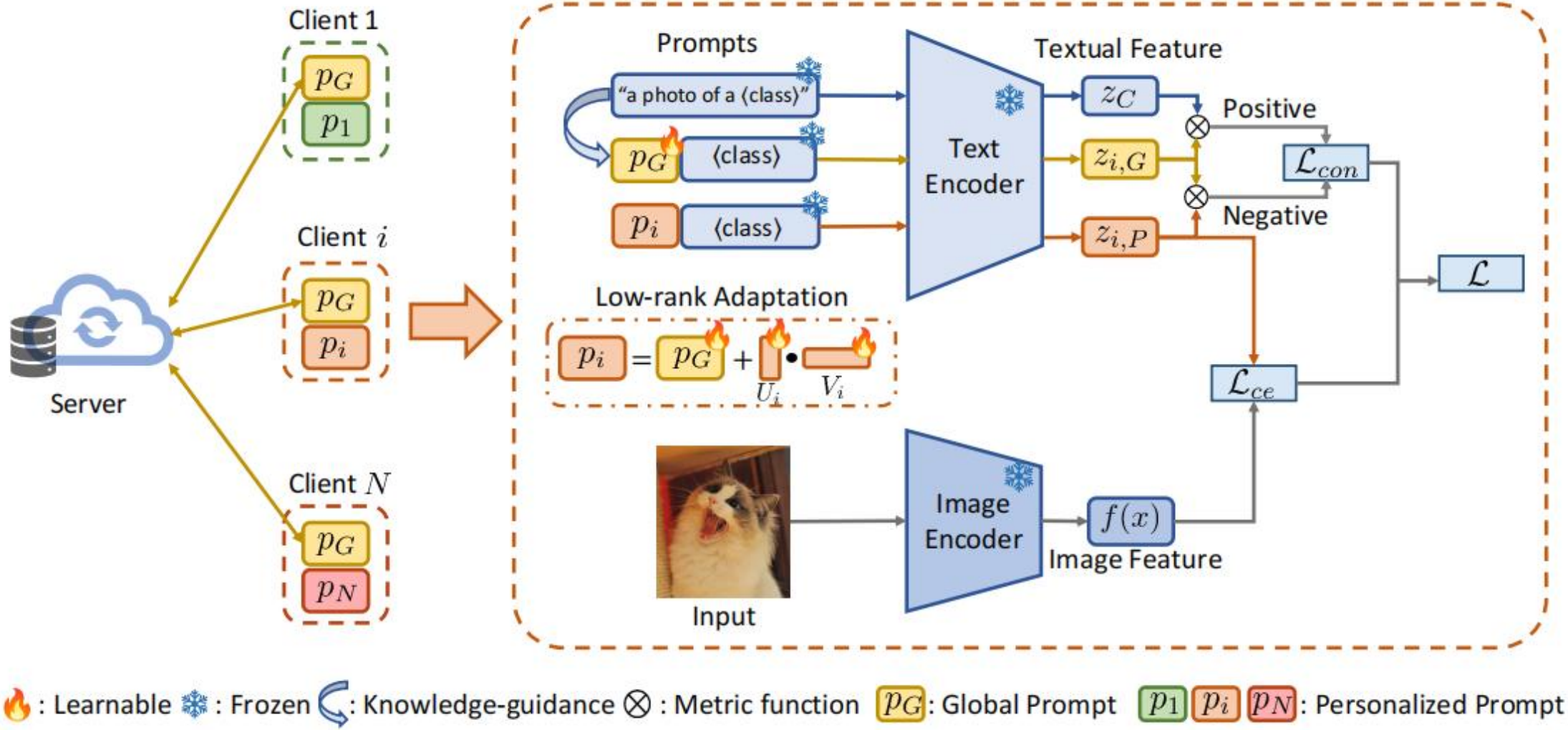
Introduction

• In federated learning, it is essential to account for the generalization capability to **unseen domains or categories**.
• The generalization issues in prompt based VLM have been revealed in recent research,
→ CoOp struggles with generalizing to unseen categories within the same dataset due to **overfitting**, resulting in lower test accuracy on novel categories.

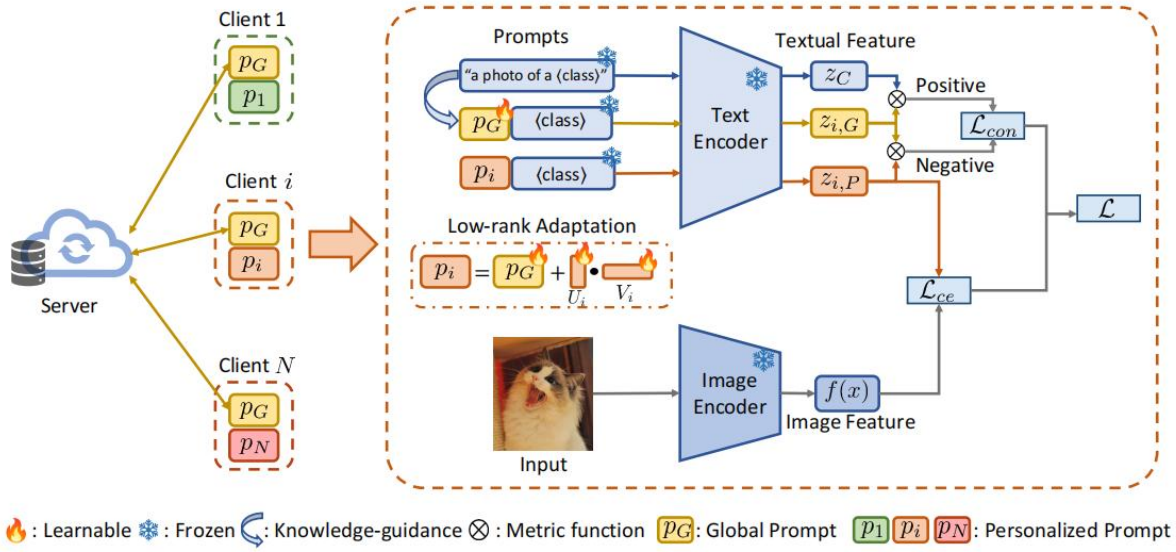
• Data heterogeneity results in another challenge in federated learning.
• If we employ strong personalized techniques to **fully adapt** the prompts to local distributions, it may lead to the **loss of inherent generalization** in VLM.

How can we strike a balance between generalization and personalization in FPL?

Method


Client 1
p_G
p_1
Client i
p_G
p_i
Client N
p_G
p_N
Server
Prompts
"a photo of a ⟨class⟩"
p_G
⟨class⟩
p_i
⟨class⟩
Text Encoder
Textual Feature
z_C
$z_{i,G}$
$z_{i,P}$
Positive
Negative
$\mathcal{L}_{con}$
$\mathcal{L}$
$\mathcal{L}_{ce}$
Low-rank Adaptation
$p_i = p_G + U_i \cdot V_i$
Input
Image Encoder
$f(x)$
Image Feature
: Learnable
: Frozen
: Knowledge-guidance
: Metric function
p_G: Global Prompt
p_1 p_i p_N: Personalized Prompt

Method



$$p_i = p_G + \Delta p_i, \tag{7}$$

$$\min_{p_G, \{\Delta p_i\}_{i=1}^N} \sum_{i=1}^{N} \frac{n_i}{\sum_j n_j} \mathcal{L}_{ce}^{D_i}(p_G + \Delta p_i). \tag{8}$$

Balance generalization and personalization:
Inspired by LoRA:
prompt may also possess a low "intrinsic rank" during the adaptation process.

$$\Delta p_i = U_i V_i. \tag{9}$$

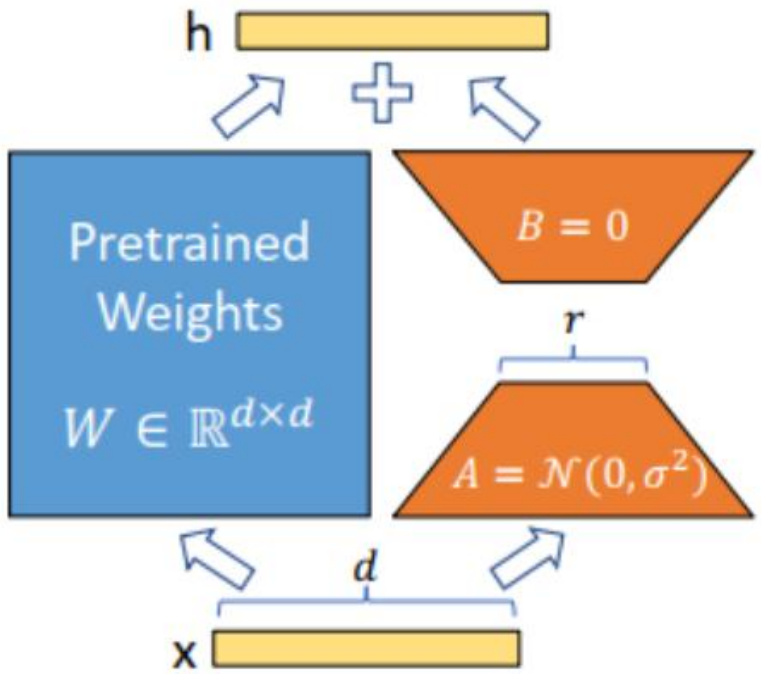
$$\min_{p_G, \{U_i, V_i\}_{i=1}^N} \sum_{i=1}^{N} \frac{n_i}{\sum_j n_j} \mathcal{L}_{ce}^{D_i}(p_G + U_i V_i). \tag{11}$$

For better personalization:
increase dissimilarity between representations of global and personalized prompts.

$$\mathcal{L}_{con} = -\log \frac{\exp(\mathrm{sim}(z_G, z_C)/\tau)}{\exp(\mathrm{sim}(z_G, z_C)/\tau) + \exp(\mathrm{sim}(z_G, z_i)/\tau)}, \tag{12}$$

$$\mathcal{L} = \mathcal{L}_{ce} + \mu \mathcal{L}_{con}, \tag{13}$$

LoRA



	Weight Type	$r = 1$	$r = 2$	$r = 4$	$r = 8$	$r = 64$
WikiSQL(±0.5%)	W_q	68.8	69.6	70.5	70.4	70.0
	W_q, W_v	73.4	73.3	73.7	73.8	73.5
	W_q, W_k, W_v, W_o	74.1	73.7	74.0	74.0	73.9
MultiNLI (±0.1%)	W_q	90.7	90.9	91.1	90.7	90.7
	W_q, W_v	91.3	91.4	91.3	91.6	91.4
	W_q, W_k, W_v, W_o	91.2	91.7	91.7	91.5	91.4

	$r = 4$			$r = 64$		
	ΔW_q	W_q	Random	ΔW_q	W_q	Random
$\|\|U^\top W_q V^\top\|\|_F =$	0.32	21.67	0.02	1.90	37.71	0.33
$\|\|W_q\|\|_F = 61.95$	$\|\|\Delta W_q\|\|_F = 6.91$			$\|\|\Delta W_q\|\|_F = 3.57$		

$$W_0 + \Delta W = W_0 + BA \qquad B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times k} \quad and \quad r \ll min(d, k) \quad (3)$$

$$h = W_0 x + \Delta W x = W_0 x + BAx \qquad (4)$$

Experiments

Table 1: Accuracy comparison (%) on clients' local classes and Base-to-novel generalization.

(a) Average over 5 datasets.

Methods	Local	Base	Novel	HM
CLIP (Radford et al., 2021)	79.18	79.83	**83.25**	80.72
CoOp (Zhou et al., 2022b)	94.28	69.40	73.16	77.55
PromptFL (Guo et al., 2023b)	90.00	85.65	78.53	84.46
Prompt+Prox (Li et al., 2020)	89.84	85.04	77.40	83.78
FedMaPLe	90.81	84.90	81.49	85.56
FedCoCoOp	90.01	85.08	81.4	85.35
FedOTP (Li et al., 2024)	**98.16**	59.73	71.08	73.17
FedPGP	95.67	**85.69**	81.75	**87.33**

(b) OxfordPets.

Methods	Local	Base	Novel	HM
CLIP (Radford et al., 2021)	89.34	89.31	96.86	91.70
CoOp (Zhou et al., 2022b)	95.33	82.51	92.92	89.90
PromptFL (Guo et al., 2023b)	95.12	95.16	91.89	94.03
Prompt+Prox (Li et al., 2020)	95.95	95.24	91.25	94.10
FedMaPLe	93.75	95.53	**97.45**	95.55
FedCoCoOp	96.02	96.01	97.25	96.42
FedOTP (Li et al., 2024)	**99.93**	63.92	80.56	78.81
FedPGP	96.65	**95.87**	97.33	**96.61**

(c) Flowers102.

Methods	Local	Base	Novel	HM
CLIP (Radford et al., 2021)	67.69	68.85	**77.23**	71.01
CoOp (Zhou et al., 2022b)	96.39	55.91	64.47	68.54
PromptFL (Guo et al., 2023b)	94.32	76.19	70.1	78.96
Prompt+Prox (Li et al., 2020)	92.73	73.06	66.09	75.75
FedMaPLe	94.89	77.49	70.46	79.71
FedCoCoOp	94.57	77.88	74.39	81.4
FedOTP (Li et al., 2024)	99.63	44.16	56.57	59.57
FedPGP	**99.68**	**78.48**	75.11	**83.13**

(d) DTD.

Methods	Local	Base	Novel	HM
CLIP (Radford et al., 2021)	53.79	54.62	**58.20**	55.47
CoOp (Zhou et al., 2022b)	86.38	39.2	37.65	47.13
PromptFL (Guo et al., 2023b)	72.71	71.41	49.28	62.44
Prompt+Prox (Li et al., 2020)	74.07	**71.84**	50.20	63.37
FedMaPLe	78.37	65.35	55.85	65.26
FedCoCoOp	72.61	68.20	54.4	64.08
FedOTP (Li et al., 2024)	**94.91**	43.87	43.68	53.36
FedPGP	89.07	69.65	55.25	**68.15**

(e) Caltech101.

Methods	Local	Base	Novel	HM
CLIP (Radford et al., 2021)	95.72	96.96	93.99	95.54
CoOp (Zhou et al., 2022b)	99.39	86.37	86.12	90.22
PromptFL (Guo et al., 2023b)	97.04	97.27	92.79	95.65
Prompt+Prox (Li et al., 2020)	96.76	**97.34**	91.99	95.30
FedMaPLe	96.47	96.7	**94.32**	95.82
FedCoCoOp	96.65	95.45	92.46	94.82
FedOTP (Li et al., 2024)	**99.68**	87.49	89.33	91.86
FedPGP	99.46	96.09	93.62	**96.33**

(f) Food101.

Methods	Local	Base	Novel	HM
CLIP (Radford et al., 2021)	89.38	89.39	**89.98**	89.58
CoOp (Zhou et al., 2022b)	93.92	82.99	84.62	86.92
PromptFL (Guo et al., 2023b)	90.79	88.22	88.6	89.19
Prompt+Prox (Li et al., 2020)	89.68	87.72	87.49	88.29
FedMaPLe	90.59	**89.43**	89.38	89.80
FedCoCoOp	90.18	87.86	88.51	88.84
FedOTP (Li et al., 2024)	**96.65**	59.19	85.28	76.98
FedPGP	93.51	88.37	88.44	**90.04**

Experiments

Table 2: The average classification accuracy using leave-one-domain-out validation on Offica-Caltech10 and DomainNet.

Datasets	Office-Caltech10					DomainNet						
Domains	A	C	D	W	Avg	C	I	P	Q	R	S	Avg.
CLIP (Radford et al., 2021)	19.40	18.32	21.87	18.59	19.55	49.89	47.23	53.61	32.10	48.19	50.79	46.96
CoOp (Zhou et al., 2022b)	41.54	15.55	56.04	43.60	39.18	83.42	53.28	80.80	49.41	75.18	82.88	70.83
PromptFL (Guo et al., 2023b)	96.34	91.57	97.96	98.30	96.04	95.28	73.72	94.50	61.60	95.72	95.43	86.04
Prompt+Prox (Li et al., 2020)	96.13	**92.52**	97.57	97.96	96.05	95.47	69.44	94.95	61.24	75.18	95.41	81.95
FedOTP (Li et al., 2024)	95.88	92.13	**99.15**	97.15	96.07	94.10	70.57	89.88	55.80	94.93	92.73	83.00
FedPGP	**96.55**	91.92	98.93	**98.75**	**96.54**	**96.45**	**74.46**	**95.43**	**62.12**	**96.06**	**96.05**	**86.76**

Table 4: The detailed classification accuracy using leave-one-domain-out validation on Offica-Caltech10 dataset.

Datasets	Office-Caltech10					
Source Domains		Amazon	Caltech	DSLR	Webcam	Avg.
CoOp (Zhou et al., 2022b)	Amazon	——	89.03	16.49	19.1	41.54±41.15
	Caltech	26.89	——	5.87	13.89	15.55±10.61
	DSLR	64.96	86.62	——	16.56	56.04±35.87
	Webcam	50.16	76.94	3.72	——	43.6±37.05
FedPGP	Amazon	——	96.45	96.03	97.18	**96.55±0.58**
	Caltech	94.66	——	86.92	93.59	**91.92±4.19**
	DSLR	98.08	99.36	——	99.36	**98.93±0.74**
	Webcam	98.98	98.98	98.3	——	**98.75±0.39**

Experiments

Table 5: Accuracy comparison (%) on the Dirichlet Non-IID setting in CIFAR-10 and CIFAR-100 over 100 clients.

Methods	CIFAR-10	CIFAR-100
CLIP (Radford et al., 2021)	87.52±0.56	64.83±0.49
CoOp (Zhou et al., 2022b)	93.13±0.34	74.78±0.41
PromptFL (Zhou et al., 2022b)	92.32±0.79	73.72±0.61
Prompt+Prox (Li et al., 2020)	91.79±0.46	71.08±0.89
FedPGP	**94.82±0.37**	**77.44±0.15**

Table 6: Accuracy (%) of ablation study on adaption and additional loss for clients' local classes and Base-to-novel generalization.

Methods	Local	Base	Novel	HM
FedPGP w/o Positive	94.63	84.68	77.75	85.13
FedPGP w/ Full-rank Adaption	**98.57**	48.00	63.40	64.17
FedPGP	95.67	**85.69**	**81.75**	**87.33**

Table 7: Accuary (%) of ablation study on additional loss for personalization.

Methods	OxfordPets	Flowers102	DTD	Caltech101	Food101
FedPGP w/o Negative	97.65±0.20	98.63±0.11	90.78±0.31	98.48±0.17	94.72±0.18
FedPGP	**98.96±0.42**	**99.29±0.03**	**91.52±0.41**	**98.90±0.19**	**95.52±0.15**

$$\mathcal{L}_{neg} = 1 - \mathrm{sim}(z_G, z_i) \quad \mathcal{L}_{pos} = \mathrm{sim}(z_G, z_C)$$

Experiments

Table 11: Quantitative comparisons on 4 datasets across varying number of shots with different number of bottleneck in FedPGP over 10 clients.

Dataset	Bottleneck	1 shot	2 shots	4 shots	8 shots	16 shots
Oxford Pets	1	92.4	92.89	93.96	94.28	95.12
	2	92.39	93.04	94.93	95.91	96.39
	4	92.51	**93.62**	94.66	96.72	97.32
	8	**93.16**	93.12	**96.31**	**97.93**	**97.81**
Flowers102	1	86.89	91.92	96.26	98.56	98.75
	2	87.79	93.95	96.28	97.60	98.71
	4	87.77	94.86	97.61	**98.92**	**99.37**
	8	**89.74**	**96.55**	**97.64**	98.88	99.05
DTD	1	53.13	60.52	70.41	85.61	83.00
	2	52.63	58.77	73.97	87.75	91.05
	4	55.02	66.05	76.80	**89.27**	90.08
	8	**55.47**	**69.91**	**85.27**	89.16	**92.00**
Caltech101	1	93.46	93.93	96.06	97.62	98.40
	2	93.27	94.36	96.69	97.89	98.33
	4	94.44	**96.32**	97.02	98.20	98.30
	8	**95.74**	95.10	**98.09**	**98.28**	**99.00**

Thanks