Mixture of Experts Made Personalized: Federated Prompt Learning for Vision-Language Models

University of Pittsburgh
UNIVERSITY OF CENTRAL FLORIDA

Jun Luo†, Chen Chen‡, and Shandong Wu †*

† Intelligent Systems Program, *Department of Radiology
University of Pittsburgh, Pittsburgh, PA, USA
(jul117@pitt.edu, wus3@upmc.edu)

‡ Center for Computer Vision Research
University of Central Florida, Orlando, FL, USA
(chen.chen@crcv.ucf.edu)

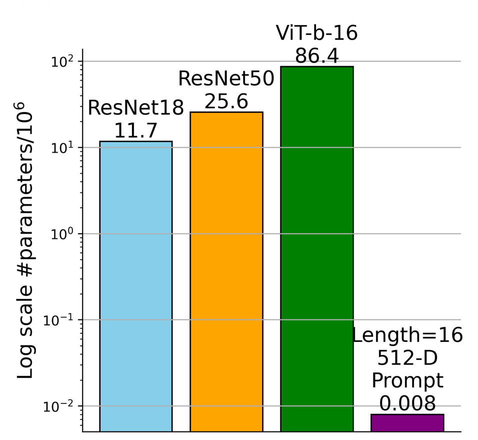
ICLR International Conference On Learning Representations

Project GitHub

Motivation & Objective

- **Federated prompt learning** leverages **CLIP-like Vision-Language Models' (VLMs')** robust representation learning for federated learning (FL), but lacks personalization and restricts clients to a **single aggregated global model** under data heterogeneity, failing to exploit the **lightweight** nature of prompts.

- Issues with the traditional approaches:
  - ✓ Ill-suited restriction for sharing only aggregated prompt
  - ✓ Limited knowledge sharing across clients
  - ✓ Inability to leverage multiple prompt experts locally



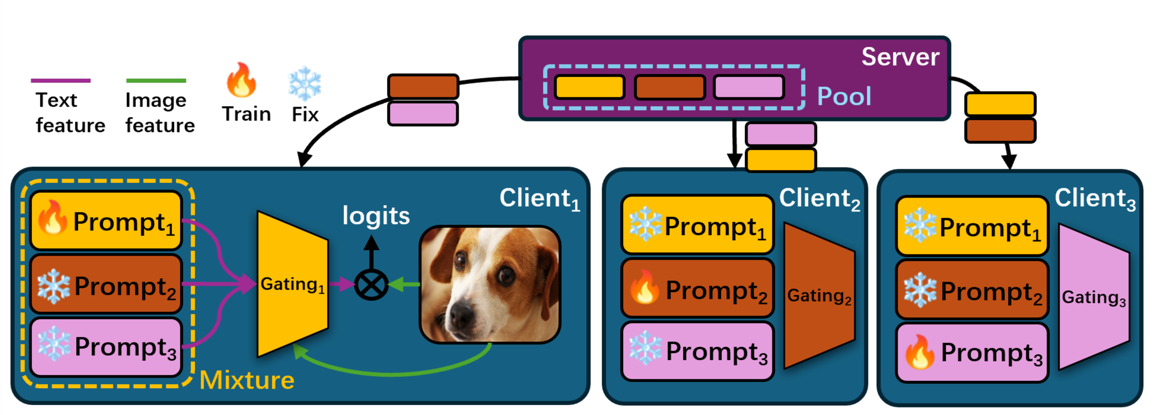
Model parameter counts

- Propose a paradigm shift in personalized federated prompt learning:
  - ✓ Personalized federated prompt learning:

$$\min_{\theta_1,\ldots,\theta_N} F(\theta_1,\ldots,\theta_N) = \min_{\theta_1,\ldots,\theta_N} \sum_{i=1}^{N} p_i F_i(\theta_i)$$

  - ✓ Challenges conventional restriction on sharing a single global model
  - ✓ Encourages sharing of pre-aggregated prompts
  - ✓ Sharing only the prompts reduces communication overhead (ResNet50: ~25.6 M parameters vs. prompt: ~8K parameters)

- Propose **pFedMoAP (Personalized Federated Mixture of Adaptive Prompts)**
  - ✓ Knowledge transfer through pre-aggregated prompts



  - ✓ Attention-based gating network from Mixture of Experts (MoE) perspective
  - ✓ Superior performance across 9 datasets under various federated settings

Method

- **Novel Architecture: pFedMoAP**
  - ✓ Locally updated prompts shared as specialized experts under the scope of Mixture of Experts (MoE).
  - ✓ Client-specific, parameter-efficient attention-based gating
  - ✓ Enables multi-source knowledge utilization

- **Prompt Learning for CLIP-like**

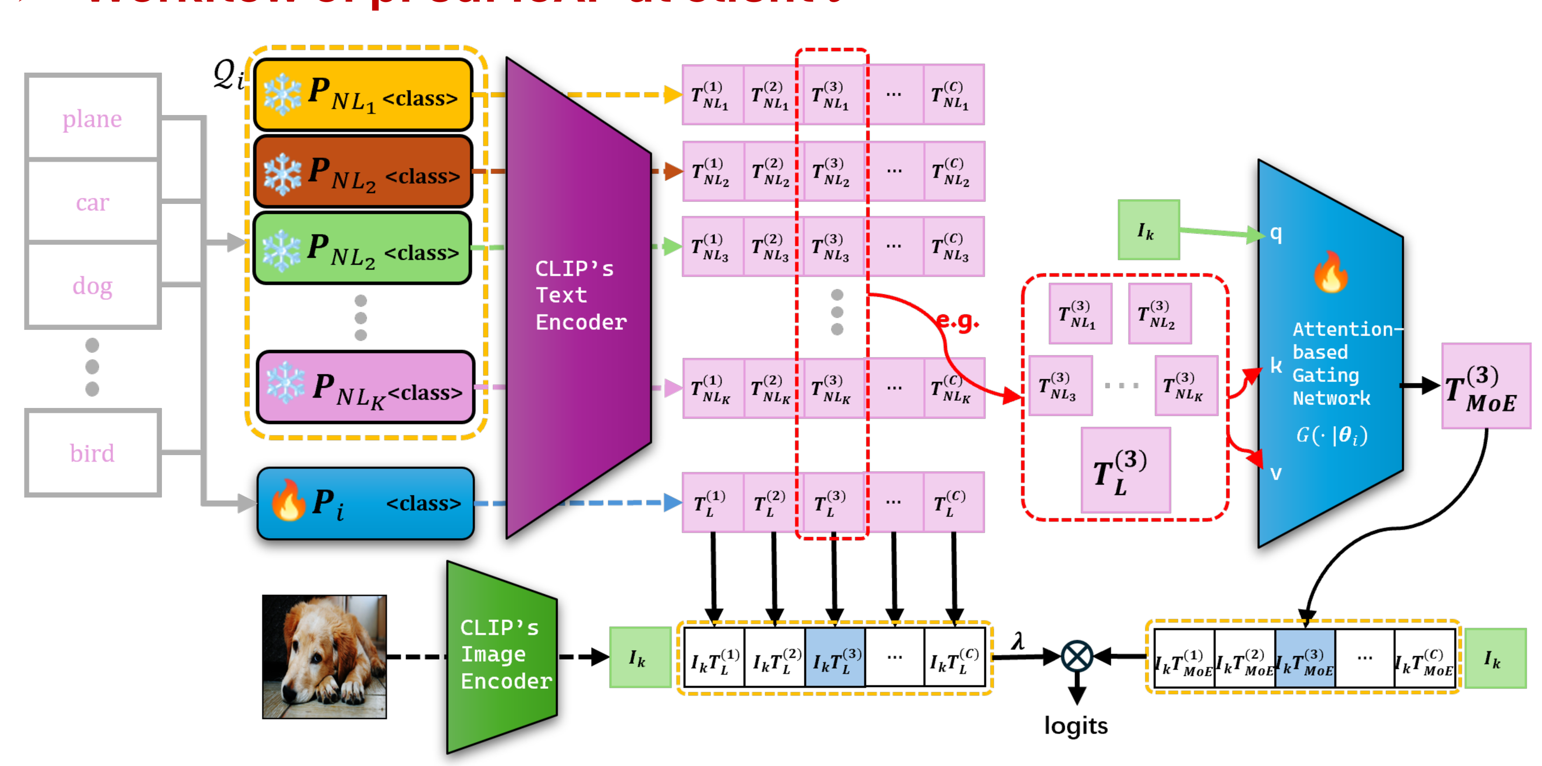
$$\text{logit}^{(c)} = \text{sim}(f(\boldsymbol{x}), g(\boldsymbol{P}^{(c)}))$$

$$p(\hat{y}=c|\boldsymbol{x}) = \frac{\exp(\text{logit}^{(c)})/\tau)}{\sum_{k=1}^{C}\exp(\text{logit}^{(k)}/\tau)}$$

- **Mixture of Experts**

$$MoE(\boldsymbol{x}) = \sum_{i=1}^{N} G(\boldsymbol{x})_i \cdot E_i(\boldsymbol{x})$$

- **Workflow of pFedMoAP at client *i***



- **Server-Maintained Pool of Prompts with KNN-Based Expert Assignment**
  - ✓ Refreshes with new prompts after each round
  - ✓ Maintains diverse expert prompts from clients
  - ✓ Each client assigned K non-local experts based on prompt similarity
  - ✓ Facilitates knowledge sharing from clients with similar data distributions

$$\mathcal{P}_t = \mathcal{P}_{t-1} - \{\boldsymbol{P}_i^{t-1}\}_{i\in\mathcal{P}_{t-1}\cap\mathcal{S}_t} + \{\boldsymbol{P}_j^{t}\}_{j\in\mathcal{S}_t}$$

- **Attention-Based Gating Network**
  - ✓ Provides robustness to adaptive experts
  - ✓ Serves as enhanced linear probing with higher capacity
  - ✓ Leverages CLIP's feature alignment through attention mechanism
  - ✓ Remains agnostic to experts' ordering.
  - ✓ Multi-head attention (MHA) with image features as query, text features as key & value

$$\boldsymbol{T}_{MoE}^{(c)} = G(\boldsymbol{I}_k, \boldsymbol{T}_L^{(c)}, \boldsymbol{T}_{NL}^{(c)}|\theta_i) = \text{MHA}(Q,K,V) = \text{Concat}(\text{head}_1,\ldots,\text{head}_h)W^O$$

- **Combined Logits Computation**
  - ✓ Balances mixed experts and local prompts contributions
  - ✓ λ controls personalization vs. knowledge sharing

$$\forall c\in[C],\ \text{logit}^{(c)} = \text{sim}(\boldsymbol{I}_k, \boldsymbol{T}_{MoE}^{(c)}) + \lambda\cdot\text{sim}(\boldsymbol{I}_k, \boldsymbol{T}_L^{(c)})$$

Experiments

- Few-shot performance on CLIP under pathological non-IID setting
  - ✓ **pFedMoAP outperforms other baselines across 5 datasets**

	Flowers102	OxfordPets	Food101	Caltech101	DTD
ZS CLIP	62.17±0.12	84.47±0.01	75.27±0.05	85.14±0.24	40.21±0.12
CoOp	70.14±0.76	83.21±1.30	70.43±2.42	87.37±0.44	44.23±0.63
PromptFL	72.80±1.14	90.79±0.61	77.31±1.64	89.70±1.99	54.11±0.22
PromptFL+FT	72.31±0.91	91.23±0.50	77.16±1.56	89.70±0.25	53.74±1.36
Prompt+FedPer	72.11±1.35	89.50±1.62	71.29±1.87	86.72±1.45	50.23±0.82
Prompt+FedProx	66.40±0.29	89.24±0.41	76.24±1.94	89.41±0.55	44.26±1.11
Prompt+FedAMP	69.10±0.13	80.21±0.44	74.48±1.71	87.31±1.60	47.16±0.92
pFedPrompt	86.46±0.15	91.84±0.41	92.26±1.34	96.54±1.31	77.14±0.09
FedOTP	96.23±0.44	98.82±0.11	92.73±0.15	97.02±0.36	87.64±0.70
pFedMoAP (λ=0.0)	97.61±0.11	94.83±0.65	86.71±0.15	95.71±0.37	85.64±0.34
pFedMoAP (λ=0.5)	98.41±0.04	99.06±0.09	93.39±0.09	97.95±0.07	89.13±0.54

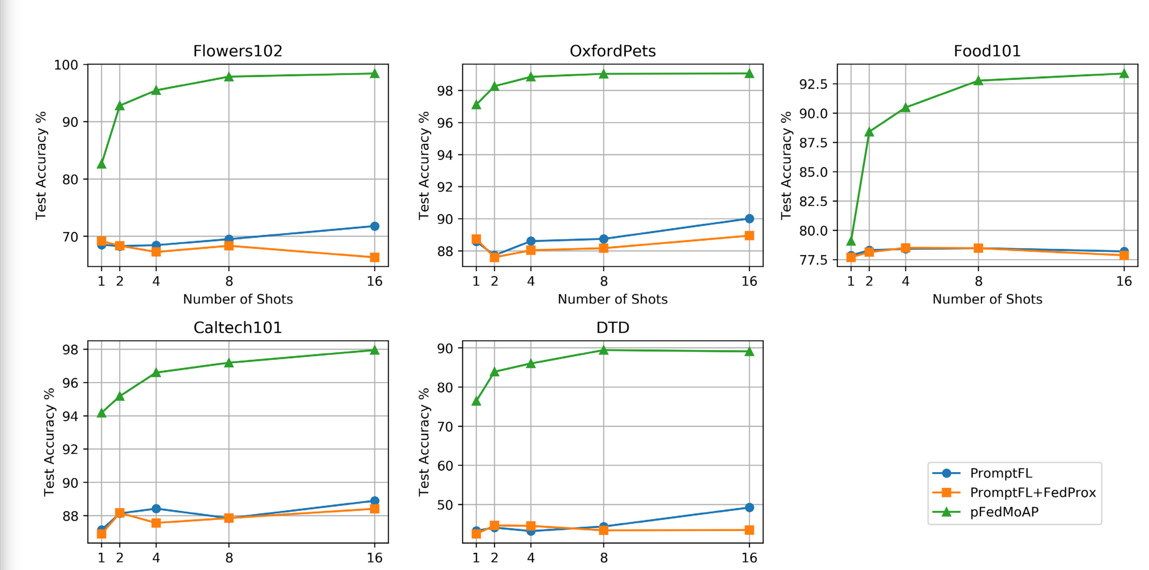
- CIFAR10 and CIFAR100 with label shift with (Dir(α = 0.5)) partition into 100 clients
  - ✓ **pFedMoAP outperforms baselines by introducing non-local experts to clients, effectively addressing label shifts**

	CIFAR10	CIFAR100
ZS CLIP	53.46±0.21	32.68±0.00
CoOp	80.84±0.39	48.74±0.17
PromptFL	73.29±0.37	45.00±0.62
Prompt+FedProx	73.32±0.34	45.63±0.75
pFedMoAP	83.46±0.53	53.42±0.22

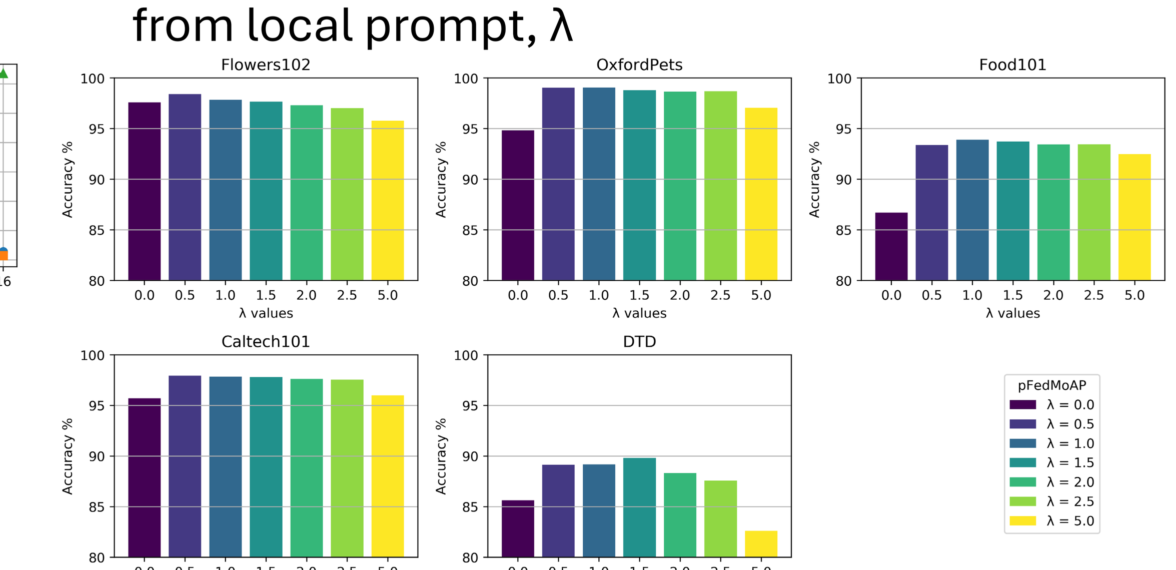
- DomainNet with feature shift and label shift with Dir(α = 0.3) partition into 5 clients/domain
  - ✓ **pFedMoAP remains better than local training and FL with up to 5.94% accuracy boost**

	Clipart	Infograph	Painting	Quickdraw	Real	Sketch	Average
ZS CLIP	9.18±0.62	10.03±0.16	9.93±0.51	10.25±0.40	9.90±1.30	9.54±1.13	9.81±0.30
CoOp	43.84±3.51	45.72±0.85	29.94±0.46	36.83±1.17	31.64±0.49	33.97±0.78	36.99±0.79
PromptFL	27.63±16.41	27.69±18.07	21.62±8.34	23.45±13.49	20.62±11.03	25.90±8.10	24.48±12.52
Prompt+FedProx	22.23±15.42	21.75±17.00	18.58±8.15	19.40±12.59	17.17±10.25	22.49±8.44	20.27±11.83
pFedMoAP	47.49±0.64	46.73±0.71	32.74±0.84	37.16±0.34	31.02±0.59	37.67±0.72	38.80±0.11

- Ablation study on the number of shots



- Ablation study on the coefficient for the logits from local prompt, λ



Conclusions

- **pFedMoAP** pioneers sharing of pre-aggregated personalized prompts for federated VLM under the scope of MoE, achieving superior performance with extreme data heterogeneity.

Acknowledgements

- NIH/NIC 1R01CA218405
- NSF CICI: SIVD: 2115082
- NSF/NIH 1R01EB032896
- Bridges-2 by ACCESS program, NSF ACI- 2138259, 2138286, 2138307, 2137603, 2138296