FedPerfix: Towards Partial Model Personalization of Vision Transformers in Federated Learning

Center for Research in Computer Vision

Guangyu Sun[1], Matias Mendieta[1], Jun Luo[2], Shandong Wu[2], Chen Chen[1]

[1]Center for Research in Computer Vision, University of Central Florida, USA

[2]University of Pittsburgh, USA

{guangyu.sun, matias.mendieta}@ucf.edu; jul117@pitt.edu; wus3@upmc.edu; chen.chen@crcv.ucf.edu

Motivation

Federated Learning:

- A server will broadcast a global model to the clients.
- Each client performs local training and sends the model back to the server
- The server aggregates the received models into a new global model
- Repeat it for several rounds.

All clients share the same global model.

Personalized Federated Learning:

Each client will have its own client model!

- Full model personalization keeps a separate client model for each client, while partial model personalization keeps a subset of the parameters as personalized parameters for each client.

- Partial model personalization shares fewer parameters, providing advantages in terms of computation, communication, and privacy.

- Previous work focuses on CNNs, while ViTs have shown superior performance in many fields.
- Can these methods for CNNs maintain high performance with ViT backbones?
- Where and how to personalize a model with a ViT backbone?

Where to Personalize?

Table 1. **Sensitivity to data distribution of each type of layer in a Vision Transformer.** The *mean* and *standard deviation* of the client's Top-1 Accuracy are reported for each type of layer (**Stand-alone**). Considering the classification head is the most sensitive type of layer, we also report the performance when a type of layer is kept updated locally along with the classification head (**Combined**). The **overall** performance is the mean of the stand-alone accuracy and combined accuracy.

Type	Stand-alone	Combined	Overall
All Local	$34.74_{\pm 9.36}$	-	-
All Global	$23.29_{\pm 11.31}$	-	-
Classification Head	$44.42_{\pm 7.98}$	-	-
Patch Embedding	$27.61_{\pm 10.02}$	$43.45_{\pm 8.69}$	35.53
Position Embedding	$23.80_{\pm 11.42}$	$45.04_{\pm 8.08}$	34.42
LayerNorm	$23.94_{\pm 11.19}$	$44.60_{\pm 7.96}$	34.27
Self-attention	$42.95_{\pm 8.68}$	$44.63_{\pm 8.67}$	**43.79**
MLP	$42.53_{\pm 8.82}$	$42.63_{\pm 8.84}$	42.58

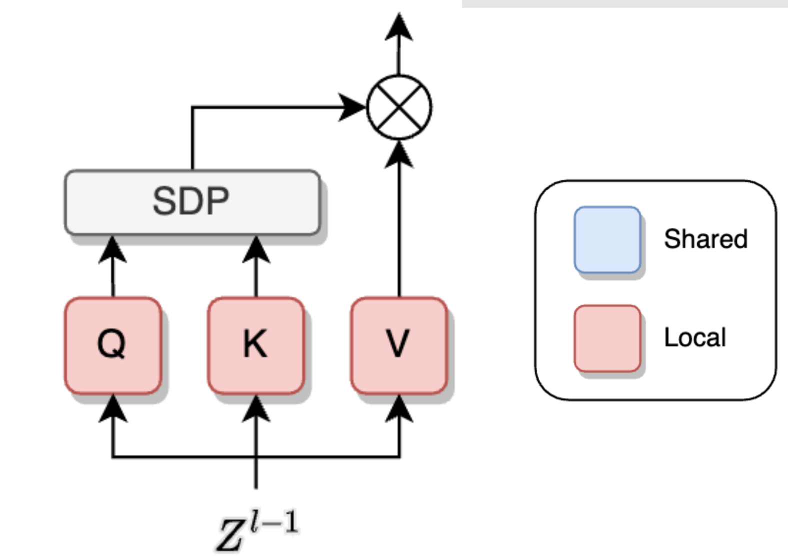
How to Personalize? -- A Baseline

A baseline: Vanilla Attention

Attention layers and the classification head are the most sensitive layers.

keep those layers updated locally

However, Vanilla Attention will also prevent it from learning general information for the global model.



(a) Vanilla Attention

How to Personalize? -- Motivation

Simply keeping the attention layers local prevents learning from other clients

PEFT structures as plugins to capture local information

Aggregate the attention layers but adapt with local personalized modules!

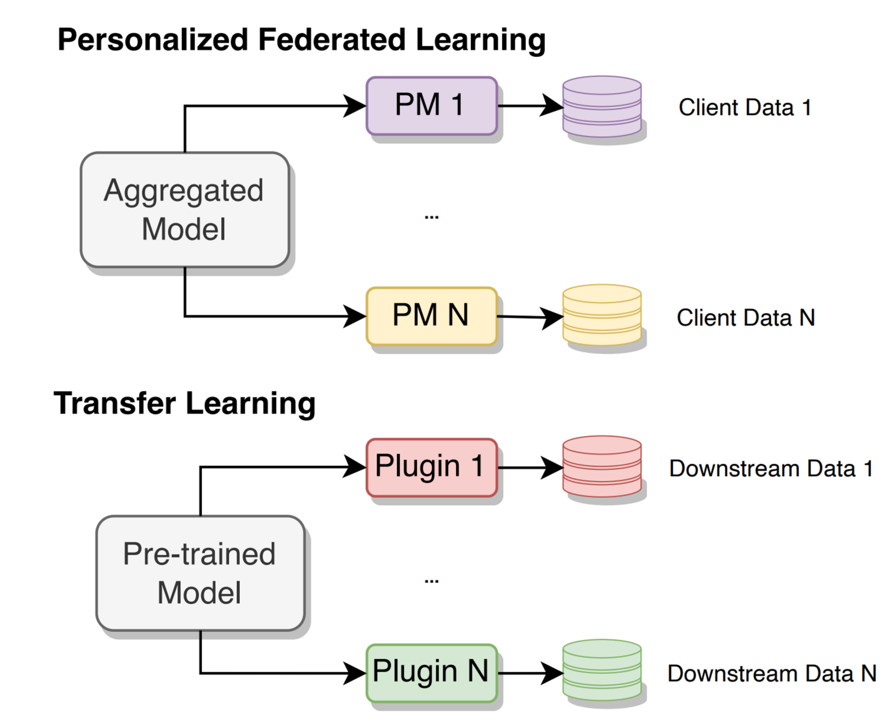


Figure 1. **Analogy between transfer learning and personalized federated learning.** A pre-trained model can be transferred to different downstream data with different plugins. In personalized federated learning, we are seeking different personalization modules (PM) to transfer the aggregated model to different client data.

How to Personalize? -- Plugins

Prompts:

Applied to the input

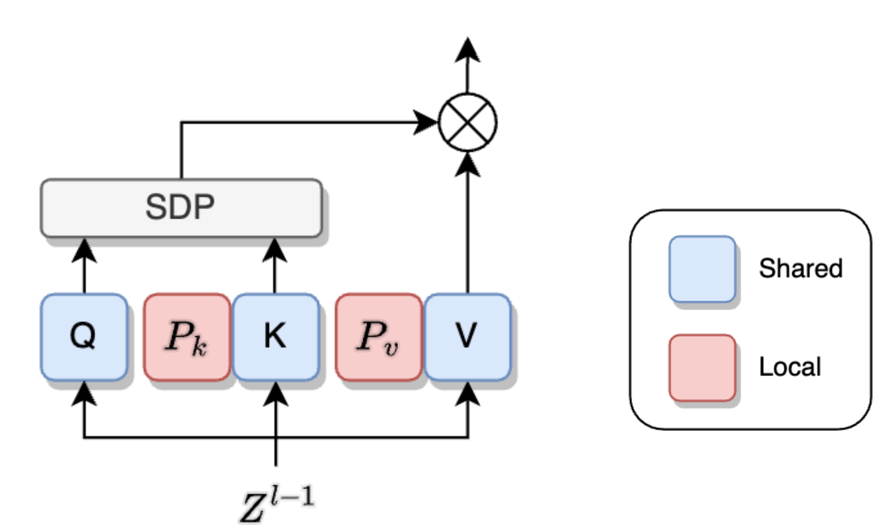
Adapters:

Applied to the MLPs

Prefixes:

Directly applied to attention layers

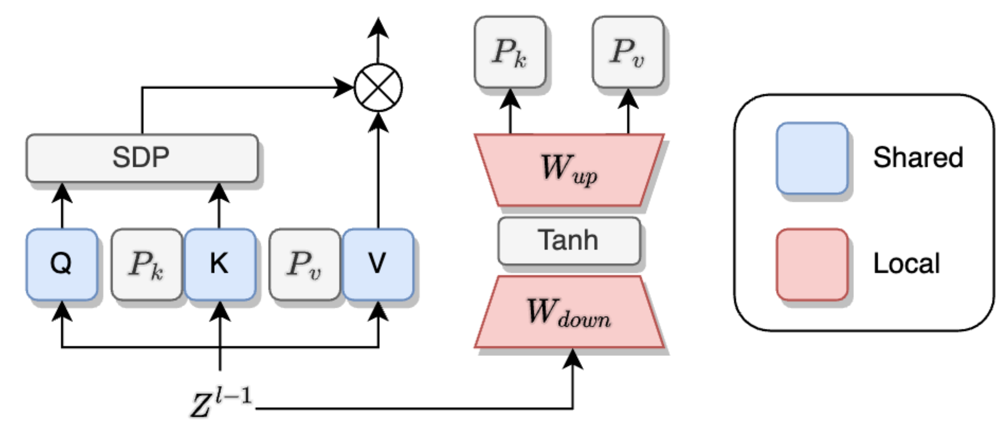
Prefixes are appended to the key and value matrix of the self-attention layer



(b) Vanilla Prefix-tuing

How to Personalize? -- FedPerfix

- In Vanilla Prefix-tuning, the prefixes are randomly initialized, which can result in unstable performance when initialized with different weights.
- To address this issue, we draw inspiration from the parallel attention design, which uses adapters to stabilize the prefixes.



(c) FedPerfix

Generate the Prefixes based on the input

FedPerfix, short for **Fed**erated **Per**sonalized Pre**fix**-tuning

Experiments

Table 2. **Performance and required resources for each method.** The mean and standard deviation of the Top-1 Accuracy among all clients are reported. The bold style indicates the best performance in each dataset.

Method	Performance CIFAR-100	Performance OrganAMNIST	Performance Office-Home	Storage (# Params)	Computation (FLOPs)	Communication (# Params)
FedAvg	$23.29_{\pm 11.31}$	$87.31_{\pm 5.98}$	$21.47_{\pm 6.24}$	21.03M (100%)	65.65M (100%)	21.03M (100%)
Local	$34.74_{\pm 9.36}$	$78.26_{\pm 11.07}$	$20.39_{\pm 7.26}$	21.03M (100%)	65.65M (100%)	0 (0%)
APFL	$44.88_{\pm 10.50}$	$89.74_{\pm 5.83}$	$24.23_{\pm 7.02}$	42.06M (200%)	131.30M (200%)	21.03M (100%)
Per-FedAVG	$33.86_{\pm 8.01}$	$82.81_{\pm 7.13}$	$17.09_{\pm 4.83}$	21.03M (100%)	131.30M (200%)	21.03M (100%)
FedBN	$23.94_{\pm 11.19}$	$87.63_{\pm 5.78}$	$21.25_{\pm 5.89}$	21.03M (100%)	65.65M (100%)	21.01M (100%)
FedBABU	$41.41_{\pm 8.87}$	$88.38_{\pm 7.16}$	$19.50_{\pm 7.71}$	21.03M (100%)	65.65M (100%)	20.66M (98%)
FedRep	$44.42_{\pm 7.80}$	$92.63_{\pm 3.77}$	$23.67_{\pm 5.97}$	21.03M (100%)	65.65M (100%)	20.66M (98%)
Vanilla Attention	$44.63_{\pm 8.67}$	$88.90_{\pm 5.87}$	$22.55_{\pm 6.37}$	21.03M (100%)	65.65M (100%)	13.89M (66%)
FedPerfix (ours)	$\mathbf{48.10}_{\pm 7.76}$	$\mathbf{93.17}_{\pm 3.51}$	$\mathbf{24.38}_{\pm 8.47}$	24.42M (116%)	66.58M (101%)	20.66M (98%)

Project Info

Scan this QR code for more project information!

FedPerfix