

FedPromo: Federated Lightweight Proxy Models At The Edge Bring New Domains To Foundation Models

Matteo Caligiuri, Francesco Barbato, Donald Shenaj, Umberto Michieli, Pietro Zanuttigh
University of Padova, Padova, Italy

Abstract

Federated Learning (FL) is an established paradigm for training deep learning models on decentralized data. However, as the size of the models grows, conventional FL approaches often require significant computational resources on client devices, which may not be feasible. We introduce FedPromo, a novel framework that enables efficient adaptation of large-scale foundation models stored on a central server to new domains encountered only by remote clients. Instead of directly training the large model on client devices, FedPromo optimizes lightweight proxy models via FL, significantly reducing computational overhead while maintaining privacy. Our method follows a two-stage process: first, server-side knowledge distillation aligns the representations of a large-scale foundation model (*e.g.*, a transformer) with those of a compact counterpart (*e.g.*, a CNN). Then, the compact model encoder is deployed to client devices, where trainable classifiers are learned locally. These classifiers are subsequently aggregated and seamlessly transferred back to the foundation model, facilitating personalized adaptation without requiring direct access to user data. Through novel regularization strategies, our framework enables decentralized multi-domain learning, balancing performance, privacy, and resource efficiency. Extensive experiments on five image classification benchmarks demonstrate that FedPromo outperforms existing methods while assuming limited-resource clients.

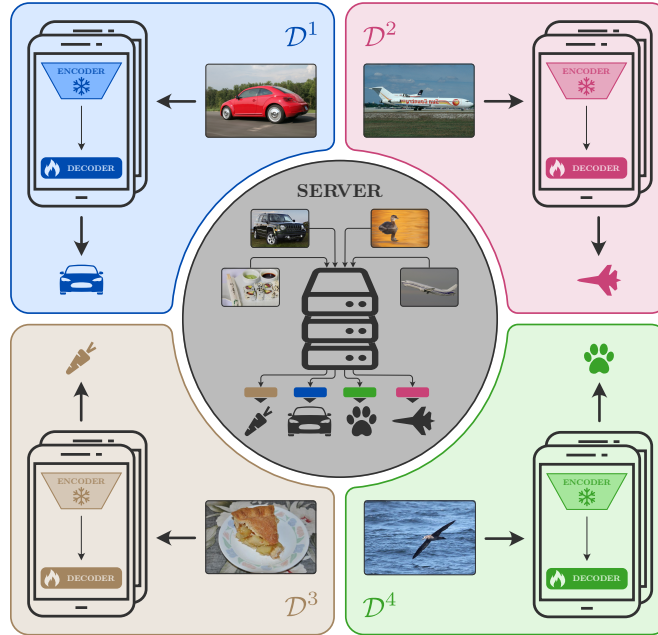


Figure 1. FedPromo enables large-scale model adaptation via Federated Learning through compact proxy models. By combining cross-architectural distillation with specialized training strategies, it achieves private, decentralized, multi-domain adaptation to data on resource-constrained clients.