UNFL: Lightweight U-Net-Based Framework for Heterogeneous Federated Learning

Beomseok Seo, Minjung Kim, JaeYeon Park
Department of Mobile Systems Engineering, Dankook University
Yong-in, Republic of Korea
beomseok.seo@dankook.ac.kr, minjung.kim@dankook.ac.kr, jaeyeon.park@dankook.ac.kr

***Abstract*—We present UNFL, a federated learning framework for heterogeneous clients that share a lightweight U-Net module across all participants. Unlike prior approaches that require architectural alignment or distillation, UNFL enables efficient knowledge transfer through a shared U-Net module while allowing each client to retain its own backbone. Experiments with VGG variants demonstrate strong generalization and low communication overhead.**

***Index Terms*—Federated Learning, U-Net, Heterogeneous Models**

I. Introduction

Federated learning (FL) allows distributed clients to train models collaboratively without sharing raw data, making it well-suited for privacy-sensitive domains such as mobile computing and healthcare. Traditional FL algorithms like FedAvg [1] assume that all clients use the same model architecture, which simplifies aggregation but limits deployment flexibility. In reality, clients often have diverse hardware capabilities and model preferences, making homogeneous FL impractical. Although recent works propose support for heterogeneous FL through knowledge distillation or feature alignment, they typically require shared datasets, complex synchronization, or costly computation, reducing the practical advantages of FL [2].

To overcome these limitations, we propose UNFL, a lightweight and architecture-agnostic framework that supports heterogeneous backbones by sharing a compact U-Net [3] module across all clients. This shared module learns client-invariant features and promotes knowledge transfer without needing structural uniformity or additional datasets.

II. Motivation

Federated learning (FL) has emerged as a promising paradigm for privacy-preserving machine learning by enabling decentralized model training without direct access to local data. However, a major limitation of most existing FL methods is the assumption that all clients share the same model architecture. This constraint is impractical in real-world deployments, where clients often differ significantly in computational capabilities, resource constraints, and application-specific requirements.

To address this, prior work has explored heterogeneous FL through knowledge distillation, prototype aggregation, or shared classifiers [4]–[6]. Despite some success, these methods commonly rely on public datasets, complex model alignment, or significant computation and communication overhead [2]. Moreover, they often fail to generalize well across structurally dissimilar models. These limitations highlight the need for a lightweight, architecture-agnostic mechanism that enables efficient knowledge sharing among diverse client models without compromising privacy or scalability. To this end, we adopt a compact U-Net as the shared module, where the multi-scale encoder aggregates global shape and layout and the decoder with lateral skips restores high-frequency details on the input grid, yielding pixel-aligned, client-invariant descriptors that transfer across heterogeneous backbones at low overhead [7]–[9]. Without this decoding and skip pathway, sharing only shallow heads produces backbone-specific latents that misalign across clients, leading to feature drift and higher per-client variance under non-IID splits.

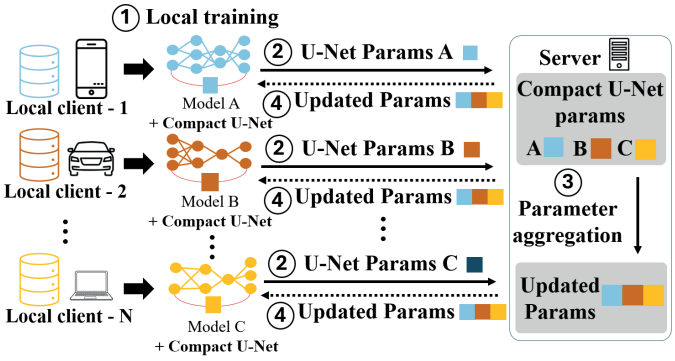

Fig. 1: Overall architecture

III. Framework

We propose UNFL, a novel federated learning (FL) framework designed to support heterogeneous client architectures by attaching a shared, lightweight U-Net module to each client. Unlike existing approaches that require uniform model structures or share only parts of the model, UNFL synchronizes the entire compact U-Net module across all clients, while allowing each client to retain its own backbone model (e.g., VGG11, VGG16, or VGG19).

To reduce the communication and computational cost of sharing the U-Net module, we significantly compress its structure. Specifically, we reduce the encoder and decoder depth from 4 to 2 and decrease the number of channels in all convolutional layers. This results in a 4× reduction in the

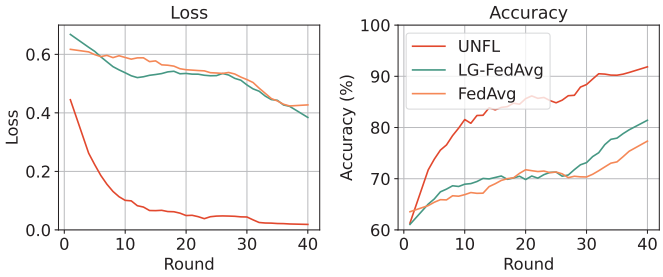

Fig. 2: Evaluation of heterogeneous model support based on the VGG11, VGG16, and VGG19 using CIFAR10 dataset.

	FedAvg	LG-FedAvg	UNFL
Communication Cost (MB)	**27.48**	**1.32**	**0.06**

TABLE I: Communication cost compared to the FL baselines.

number of channels and greatly shrinks the overall size of the module, making it suitable for resource-constrained environments. During local training, the U-Net module processes the input image in parallel with the client's backbone network. The output of the U-Net, which captures client-invariant and spatially rich features, is combined with the backbone's feature map before classification. Since the U-Net is shared globally, it acts as a universal feature extractor, aligning diverse representations into a common latent space. Meanwhile, the backbone continues to specialize on client-specific data. This design enables collaborative feature learning without requiring architectural alignment or shared datasets. It also maintains communication efficiency by synchronizing only a compact and optimized module, rather than the full model. As a result, UNFL effectively bridges the gap between flexibility and efficiency in FL across heterogeneous models.

IV. Results

To evaluate the effectiveness of UNFL in heterogeneous federated learning scenarios, we conduct experiments using the CIFAR-10 dataset under a non-IID setting. Specifically, we simulate a federation of 50 clients, where each client is assigned one of three different VGG variants (i.e., VGG11, VGG16, or VGG19) as its backbone model. In each communication round, 5 clients are randomly selected to participate, and training proceeds for a total of 40 rounds with a learning rate of 10^{-4} and batch size of 32. Data is partitioned such that each client holds an average of 256.21 samples.

We compare UNFL against two widely used FL baselines. (1) FedAvg [1], which assumes homogeneous models and performs full model averaging across all layers that are structurally compatible among the VGG variants. (2) LG-FedAvg [10], which supports limited heterogeneity by separating local and global layers and setting only the fully connected layers as global. Unlike these baselines, UNFL allows each client to maintain its own full backbone architecture while sharing a globally synchronized lightweight U-Net module. Figure 2 shows that UNFL achieves significantly better accuracy while maintaining low communication cost. In the VGG11, VGG16, VGG19 mixed setting, UNFL reaches an accuracy of 92.16%, far surpassing FedAvg and LG-FedAvg, which achieve 78.47% and 81.43% respectively under the same non-IID conditions. Despite the architectural heterogeneity, UNFL enables effective cross-client knowledge transfer through the shared U-Net module, leading to robust generalization and stable convergence across diverse backbones by sharing only **0.06** MB as shown in Table I. These findings demonstrate that UNFL can effectively bridge structural differences between clients without requiring architectural alignment or full model sharing, making it a practical solution for real-world federated systems with device and model diversity.

V. Conclusion

We proposed UNFL, a lightweight federated learning framework that supports heterogeneous client models by sharing a compact U-Net module. Unlike existing methods, UNFL enables effective collaboration without architectural alignment or shared datasets. Experiments with VGG variants show that UNFL achieves high accuracy and low communication cost, making it a practical solution for real-world FL deployments with diverse devices.

Acknowledgment

This research was supported by the MSIT, Korea, under the National Program for Excellence in SW, supervised by the IITP in 2024 (2024-0-00035).

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] D. Li and J. Wang, "Fedmd: Heterogenous federated learning via model distillation," *arXiv preprint arXiv:1910.03581*, 2019.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[4] J. Park, K. Lee, S. Lee, M. Zhang, and J. Ko, "Attfl: A personalized federated learning framework for time-series mobile and embedded sensor data processing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 3, pp. 1–31, 2023.

[5] J. Park and J. Ko, "Fedhm: Practical federated learning for heterogeneous model deployments," *ICT Express*, vol. 10, no. 2, pp. 387–392, 2024.

[6] Y. Shin, K. Lee, S. Lee, Y. R. Choi, H.-S. Kim, and J. Ko, "Effective heterogeneous federated learning via efficient hypernetwork-based weight generation," in *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*, 2024, pp. 112–125.

[7] J. Park, K. Lee, N. Park, S. C. You, and J. Ko, "Self-attention lstm-fcn model for arrhythmia classification and uncertainty assessment," *Artificial Intelligence in Medicine*, vol. 142, p. 102570, 2023.

[8] J. Park, H. Cho, R. K. Balan, and J. Ko, "Heartquake: Accurate low-cost non-invasive ecg monitoring using bed-mounted geophones," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–28, 2020.

[9] J. Park, W. Nam, J. Choi, T. Kim, D. Yoon, S. Lee, J. Paek, and J. Ko, "Glasses for the third eye: Improving the quality of clinical data analysis with motion sensor-based data filtering," in *Proceedings of the 15th ACM conference on embedded network sensor systems*, 2017, pp. 1–14.

[10] P. P. Liang, T. Liu, L. Ziyin, N. B. Allen, R. P. Auerbach, D. Brent, R. Salakhutdinov, and L.-P. Morency, "Think locally, act globally: Federated learning with local and global representations," *arXiv preprint arXiv:2001.01523*, 2020.