

Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

Guangyu Sun

M.S. University of Rochester, 2022
B.S. University of Missouri-Columbia, 2019
B.E. Shandong University, 2019

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
(COMPUTER SCIENCE)

25 March 2026, 11 A.M.
SPRK (UCF Global) 101

DISSERTATION COMMITTEE

Professor Chen Chen, *chen.chen@crcv.ucf.edu*
Professor Yogesh Singh Rawat, *yogesh@crcv.ucf.edu*
Professor Xueqiang Wang, xueqiang.wang@*ucf.edu*
Professor Xin Li, xin.li@*ucf.edu*

DISSERTATION RESEARCH IMPACT

The emergence of large-scale foundation models has reshaped computer vision and multimodal learning, offering powerful generalization for diverse downstream applications. Progress in this domain has traditionally relied on centralized training with vast datasets, which raises significant privacy and data sovereignty concerns. Federated learning (FL) addresses these issues by enabling collaborative training across decentralized clients without sharing raw data. However, the practical deployment of massive foundation models in FL introduces challenges such as communication bottlenecks, heterogeneous data distributions, limited computational memory, and a scarcity of labeled data.

This dissertation addresses these fundamental challenges in federated foundation model adaptation through several contributions. A parameter-efficient fine-tuning framework (FedPEFT) is introduced to reduce communication overhead by updating only lightweight parameter subsets. To overcome data heterogeneity, a partial model personalization strategy (FedPerfix) is proposed for Vision Transformers to balance global generalization with client-specific adaptation. Additionally, a collaborative framework (FedCola) is developed to tackle multimodal data silos by enabling knowledge exchange across clients. Finally, to accommodate resource-constrained edge clients with unlabeled data, a semi-supervised federated learning paradigm (FedMox) driven by a spatial sparse mixture-of-experts architecture is proposed.

SELECTED PUBLICATIONS

Total Citations: 256, h-index 9

1. **FedPerfix: Towards Partial Model Personalization of Vision Transformers in Federated Learning**, *Guangyu Sun*, Matias Mendieta, Jun Luo, Shandong Wu, Chen Chen, in *IEEE/CVF International Conference on Computer Vision,* 2023.
2. **Towards Multi-modal Transformers in Federated Learning**, *Guangyu Sun*, Matias Mendieta, Aritra Dutta, Xin Li, Chen Chen, in *2024 European Conference on Computer Vision,* 2024.
3. **Exploring Parameter-Efficient Fine-Tuning to Enable Foundation Models in Federated Learning**, *Guangyu Sun*, Umar Khalid, Matias Mendieta, Pu Wang, Chen Chen, in *IEEE International Conference on Big Data,* 2024.
4. **Unlocking Multi-Site Clinical Data: A Federated Learning Approach to Privacy-First Child Autism Behavior Analysis**, *Guangyu Sun*, Wenhan Wu, Zhishuai Guo, Ziteng Wang, Pegah Khosravi, Chen Chen, in *CVPR CV4CHL Workshop*, 2026.
5. **Geo²: Geometry-Guided Cross-view Geo-Localization and Image Synthesis**, Yancheng Zhang, Xiaohan Zhang, *Guangyu Sun*, Zonglin Lyu, Safwan Wshah, Chen Chen, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.
6. **EGGS: Exchangeable 2D/3D Gaussian Splatting for Geometry-Appearance Balanced Novel View Synthesis**, Yancheng Zhang, *Guangyu Sun*, Chen Chen, in *Conference on Neural Information Processing Systems*, 2025, *Spotlight*.
7. **Navigating Heterogeneity and Privacy in One-Shot Federated Learning with Diffusion Models**, Matias Mendieta**,** *Guangyu Sun*, Chen Chen**,** in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2025.
8. **Practical Semi-Supervised Federated Object Detection with Edge Devices**, *Guangyu Sun*, Jingtao Li, Weiming Zhuang, Chen Chen, Chen Chen, Lingjuan Lyu , *in submission*, 2026.

DISSERTATION

FEDERATED LEARNING IN THE ERA OF FOUNDATION MODELS

The emergence of large-scale foundation models has reshaped computer vision and multimodal learning, but their substantial computational and communication demands pose significant challenges for deployment in federated learning (FL) systems. Federated learning enables collaborative model training across decentralized clients without sharing raw data, offering strong privacy guarantees. However, adapting foundation models in FL remains difficult due to communication bottlenecks, heterogeneous data distributions, limited device resources, and scarce labeled data. This dissertation presents a unified line of research that addresses these challenges through four complementary frameworks, progressively advancing the efficiency, personalization, and practicality of foundation model adaptation in federated settings.

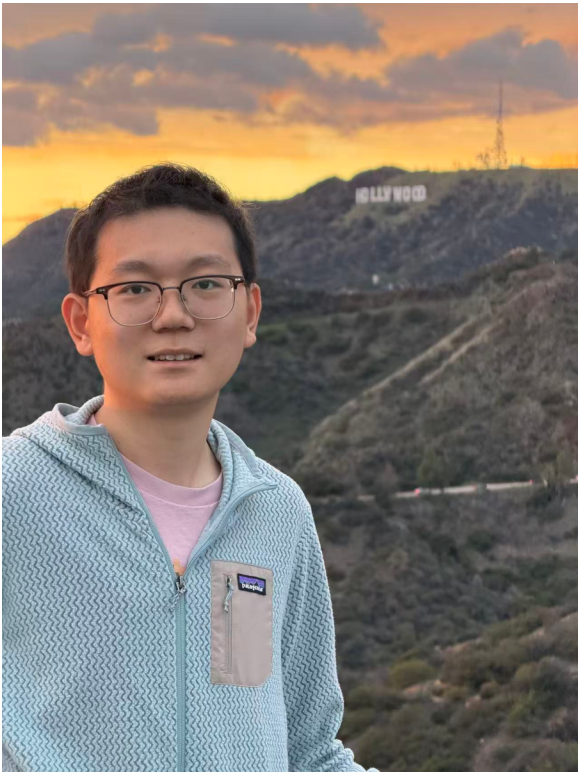
First, FedPEFT investigates parameter-efficient fine-tuning as a means to enable foundation models in FL. By updating and communicating only lightweight parameter subsets, FedPEFT drastically reduces communication overhead while maintaining strong performance and robustness under diverse federated conditions.

Building upon this communication-efficient foundation, FedPerfix focuses on personalized federated learning for Vision Transformers. Through a systematic analysis of transformer components, it identifies heterogeneity-sensitive layers and introduces a selective personalization strategy that balances global generalization with client-specific adaptation.

The third contribution, FedCola, extends federated learning to multi-modal transformers. It introduces a collaborative framework that enables clients with unpaired vision and language data to jointly train a unified multi-modal model. By addressing both in-modality and cross-modality gaps using parameter-based collaboration, FedCola establishes a systematic approach to multi-modal federated learning.

Finally, FedMox advances toward real-world deployment by introducing Practical Semi-Supervised Federated Learning, which models settings where edge clients hold unlabeled, low-resolution data and limited resources. Through a sparse mixture-of-experts architecture with spatial routing and soft aggregation, FedMox enables efficient foundation model adaptation for complex tasks such as object detection.

Together, these contributions demonstrate that foundation models can be efficiently trained, personalized, and deployed in federated environments, advancing FL toward scalable and practical real-world applications.



Guangyu Sun

1997	Born in Shandong, China
2015-2019	B.E., Shandong University, China
2017-2019	B.S., University of Missouri, Columbia, MO
2020-2022	M.S., University of Rochester, Rochester, NY
2024	Research Intern, Sony AI America, Remote
2025	Applied Scientist Intern, Amazon, Seattle, WA
2022-2026	Ph.D., University of Central Florida
2026-	Research Engineer, Meta, Menlo Park, CA