Yongfan(Steff) Liu

(949)621-1644 yongfal@uci.edu sf-liu.github.io Anticipate Grad. 26.12 CPT/OPT

Education

University of California, Irvine - Irvine, U.S. Sept. 2020 – Present
The Henry Samueli School of Engineering — Ph.D. in Computer Engineering

Southeast University - Nanjing, China Sept. 2016 – June 2020
Chien-Shiung Wu (Honor) College — Bachelor of Computer Science and Technology

RWTH Aachen - Aachen, Germany Jan. 2018 – Mar. 2018
Germany Smart Industrial 4.0 Winter Camps Study Project

Technical Skills

Programming Languages: Python, Java, C++, MATLAB, JavaScript
Frameworks/Tools: PyTorch, TensorFlow, OpenCV, SNPE SDK
Research Areas: Computer Vision, 3D Reconstruction, Edge AI, Light-weight Model, Model Quantization
Development Platforms: NVIDIA Orin Jetson Nano, Qualcomm Snapdragon, Qualcomm Automotive Platform

Experience

ILLIXR (Illinois Extended Reality) Lab Oct. 2024 –Present
Collaborator – Efficient Streaming 3D Reconstruction with XR devices *UIUC, IL & UC Irvine, CA*

- Designed a graph-based memory bank mechanism for VGGT, enabling streaming input and achieving up to 6.3× faster inference while maintaining constant peak memory usage.
- Modeled frame-to-frame relationships using a graph structure, significantly improving ViT efficiency for AR/VR and long-content scenarios; outperforms the SOTA baseline by 20.7%, 5.9%, and 5.3% in camera pose accuracy, depth accuracy, and 3D reconstruction quality, respectively.
- Submitted one paper under review at CVPR 2026.

Intelligent System Architecture Laboratory Mar. 2023 – Nov. 2025
Graduate Researcher – Efficient Depth Estimation with Meta Glasses *UC Irvine, CA*

- Proposed hardware-friendly novel efficient model with MultiHead Cost Volume, decreased inference latency by 7% and decreased error rate by 17% comparing with previous model
- Proposed Rectification Positional Encoding (RPE) to enable fast online stereo rectification, reducing inference latency by 43% and lowering the overall pipeline error rate by 42%.
- Implemented models with cross-compiling and quantization skills on Qualcomm Snapdragon devices, Nvidia Orin Jetson, and next-generation AR glasses, *Aria* .
- Published one paper to **CVPR 2025**

NIO – Electric vehicle manufacturer June 2024 – Sept. 2024
Internship – Digital Cockpit Department *San Jose, CA*

- **Acceleration and deployment of large language models on heterogeneous in-vehicle devices**
- Optimized language model inference with speculative sampling, achieving a 2.5× speedup.
- Deployed large language models on NVIDIA Orin Jetson Nano and Qualcomm SA-8295P platforms.
- Collaborated with other team members on generative models, including diffusion-based image synthesis.

Irvine Vision Laboratory Nov. 2025 – Present
Leading Student – Air quality prediction with hyperspectrum images by SmolVLM *Irvine, CA*

- Integrated pressure, humidity, and hyperspectral imaging data to predict ground-level air quality using Vision-Language Models (VLMs), improving multi-modal environmental sensing performance.

Publications & Patents

[**C.2**] **Yongfan Liu**, Boyuan Tian, Rahul Singh, Sarita Adve, and Hyoukjun Kwon: MBVGGT: Adapting VGGT for Long Video Sequences via Graph Memory Bank. Under review of CVPR 2026

[**C.1**] **Yongfan Liu** and Hyoukjun Kwon: Efficient Depth Estimation for Unstable Stereo Camera Systems on AR Glasses. Accepted by **CVPR 2025**

[**P.1**] Brain tissue segmentation method based on diagonal voxel local binary pattern texture operator, CN110728685B · Issued 2023-04-07

[**J.1**] **Yongfan Liu**, Sen Du and Youyong Kong: Supervoxel Clustering with a Novel 3D Descriptor for Brain Tissue Segmentation. ACMLC 2020 & IJMLC [ISSN: 2010-3700 (Online)]