KELING YAO

+1 (878)-256-9379 | kennyy@andrew.cmu.edu | Home Page | Linkedin | github | Pittsburgh, PA

EDUCATION

Carnegie Mellon University **Dec 2025**
Master of Science in Computer Vision *Pittsburgh, PA*

The Chinese University of Hong Kong **Jul 2024**
Bachelor of Science in Data Science and Big Data Technology (CS track) *Shenzhen, China*
- Major GPA: 3.87/4.00 (top 6%), Cumulative GPA: 3.72/4.00, Graduated with First Class Honors
- Academic Performance Scholarship (Top 1%), Dean's List Award, Harmonia College Scholarship (Top 0.1%)

University of California, Berkeley **May 2023**
Exchange Student in Computer Science *Berkeley, CA*
- Major GPA: 4.00/4.00 (top 1%), Cumulative GPA: 4.00/4.00

SKILLS

Programming Languages: Python, C, C++, C#, Java, JavaScript, SQL, MATLAB, PowerShell, RISC-V, LaTeX, R
Developer Tools: Docker, AWS(EC2), Unity, Blender, Git, Azure, OpenGL
Frameworks: Numpy, PyTorch, TensorFlow, OpenCV, ROS, scikit learn, pandas
Other: Machine Learning, Computer Vision, Image Segmentation, VR/AR/XR, Generative AI, Large Language Models, Multimodal Models, Diffusion Models, Data-driven, Data Analytics, Robot Learning, Reinforcement Learning, Graphics, DNN, R&D

WORK EXPERIENCE

DeepMirror **May 2024 – Aug 2024**
Machine Learning Engineer Intern *Guangzhou, China*
- Implemented a real-time 3D bounding box generator by training a high-resolution saliency detection model, significantly improving the surface detail of meshes and resolving floater issues in NerF 3D reconstruction.
- Implemented and experimented on distortion loss in NeRF 3D reconstruction, leading to a significant reduction in air floaters and mesh surface discontinuities, resulting in a 0.3 PSNR improvement.
- Developed a construction-specific blurry image detection model with an F1 score of 0.99, deployed into mobile devices at 130 FPS, enhancing data robustness for Structure from Motion (SfM) 3D reconstruction pipelines.

RESEARCH PROJECTS

University of California, Berkeley **Dec 2022 – May 2024**
Research Assistant, FHL Vive Center for Enhanced Reality *Berkeley, CA*
- **Published** "Robust 6DoF Pose Estimation Against Depth Noise and a Comprehensive Evaluation on a Mobile Dataset" tailed for **AR/VR mobile applications at DMLR @ ICML as co-first author.**
- Contributed to a Transformer-based 6DoF pose estimation model designed to withstand the low-quality depth data from mobile LiDAR sensors, surpassing all existing baselines with AUC results of ADD and ADD-S as 73.99 and 88.10.
- Led and established the **first** mobile RGBD 6DoF pose estimation dataset (DTTDv2), by developing an efficient data collection paradigm **(open-source)**, utilizing the OptiTrack system to automatic labelling and improve 80% time.
- Conducted LiDAR depth analysis on 47668 frames and introduced a novel depth-ADD metric for quantitative measurement.

Microsoft Research Asia **Jun 2023 – Sep 2023**
Research Assistant, Multimedia Search and Mining *Beijing, China*
- **Published** "Transferring Foundation Models for Generalizable Robotic Manipulation" **at WACV as co-author.**
- Collaborated in designing an end-to-end policy model by leveraging Vision Foundation Models and multi-modal multi-view fusion of semantic, geometric, and temporal observations using self-attention mechanism, achieving sample-efficient generalization learning with 81.25% accuracy on real world robot pick and place problem.
- Engineered real-time, precise, and rich human-like data collection pipeline by developing a 30Hz robotic manipulation framework that uses a VR headset to teleport a Franka robot and dexterous hands via a Unity app and UDP connections.

PROJECTS

CaLoRAify: Fine-tuning Visual Language Models for Calorie Estimation | *CMU* **Aug 2024 – Present**
- Use LoRA method to finetune Visual Language Model as multimodal ingredient prediction and achieve 68% accuracy on Recipe1M+.
- Use Retrieval-Augmented Generation (RAG) on USDA Food Calorie Database to improve Calorie Estimation accuracy to 82%.

InfraGen: multi-modal real-word dataset Generation | *Microsoft Fabric AI hackathon* **Aug 2024 – Present**
- Constructed an image retrieval pipeline on Microsoft Fabric Data Factory and organized 1000+ images in 1 minutes.
- Finetune and integrate Visual Language Models to generate multimodal dataset based on the user's text-prompt.
- Use Caption-Filtering Pipeline to increase the quality of multi-modal dataset with average CILP score 10%.

Zi2Calli: a GAN-Based Model for Chinese Fonts Transfer | *The Chinese University of Hong Kong* **Sep 2023 – May 2024**
- Implemented and trained Zi2Calli (open-source), a GAN-based model with category embeddings and category loss for enhanced one-to-many style transfer, achieving diverse and high-quality calligraphy/font generation in a single framework.
- Curated the first comprehensive dataset of President Xu's calligraphy, consisting of 484-character images from 95 works using Tesseract OCR.