SRECHARAN SELVAM

📞 +1 917-495-3410 ✉ sselvam@andrew.cmu.edu in linkedin.com/srecharan 🌐 Srecharan.com github.com/Srecharan

Education

Carnegie Mellon University   **May 2025**
Master of Science, Research: Machine Learning   *Pittsburgh, PA*
Relevant Coursework: Machine Learning, Deep Learning, Advanced Computer Vision, Visual Recognition   GPA - 4/4

Experience

Kantor Lab, Carnegie Mellon University [GitHub ↗]   **Aug 2023 – Present**
Graduate Research Assistant
- Developed self-supervised learning pipeline combining RAFT-Stereo depth estimation, YOLOv8 segmentation & geometric algorithms, eliminating **100%** manual annotation requirements to train GraspPointCNN for robotic leaf manipulation.
- Trained attention-based GraspPointCNN using MLflow to track **60+** model experiments for grasp point optimization.
- Parallelized 1.5M-pixel depth-to-3D projection via CUDA kernels for efficient real-time 3D reconstruction **(150→30ms)**.
- Compiled vision models into TensorRT engines with FP16 precision to boost inference throughput from **20 to 27 FPS**.
- Deployed Docker-containerized grasping system to a 6-DOF robot, achieving **82%** leaf grasp success rate in field tests.

Hanon Systems [GitHub ↗]   **Jan 2023 – Jun 2023**
Machine Learning Engineer Intern
- Built a GRU-based model trained on custom dataset with EKF for real-time 3D hand gesture (dynamic/static) recognition in AR system (Unity Engine), allowing **100+** automotive technicians to safely simulate HVAC assembly procedures.
- Optimized inference using CUDA kernels and ONNX quantization, cutting latency by **33%** & memory footprint by **50%**.
- Exposed gesture and depth modules via Flask REST APIs for modular model serving to downstream applications.
- Containerized complete pipeline by utilizing Docker for consistent, scalable deployment across **3** training centers.

Vee Ess Engineering [GitHub ↗]   **Jul 2022 – Dec 2022**
Computer Vision Engineer Intern
- Engineered distributed Apache Spark pipeline to boost recyclable material recovery on high-speed conveyors by handling terabytes of multi-camera footage and auto-annotating frames with Mask R-CNN to create **43,000+** segmented images.
- Streamlined dataset storage & versioning via AWS S3 to reduce I/O overhead by **24%** and accelerate training iterations.
- Integrated custom-trained YOLOv5 model across real-time camera streams yielding **96% mAP@[0.5:0.95]** with **<15ms** latency, enabling conveyor speed modulation based on detection density to cut manual sorting labor by **6+ hrs/week.**

Projects

VLM-Based Tool Recognition System for Industrial Safety Applications [GitHub ↗]   **Jan 2025 – May 2025**
- Fine-tuned Qwen-2.5-VL-7B & LLaMA-3.2-11B-V using LoRA on custom dataset via LLM-guided prompt engineering (**8K** images, **29K** annotations) for real-time multi-modal industrial tool recognition and safety guidance generation.
- Built a RAG pipeline using LangChain and Pinecone to ground tool-specific information, reducing hallucinations by **55%**.
- Implemented RLHF (GRPO) on AWS SageMaker to optimize preference learning on paired responses for VLM alignment.
- Orchestrated LLM-based evaluation pipeline (OpenAI API) with Kubernetes, scoring **4K+** outputs for **8** model variants.

Multi-Model Stock Prediction with NLP and Automated Trading [GitHub ↗]   **Oct 2024 – Feb 2025**
- Spearheaded distributed training infrastructure using data parallelism across 4x V100 GPUs to train ensemble ML models (bidirectional LSTM + XGBoost with **35+** features) for algorithmic trading system across multiple timeframes.
- Created Kafka streaming pipeline to ingest **9K** financial events/day from multiple APIs for low-latency trading decisions.
- Automated Apache Airflow workflows managing FinBERT sentiment analysis, boosting prediction accuracy by ~**5%**.
- Designed automated trading system with CI/CD model retraining and Tradier API execution, delivering **58.5%** win rate.

Bird Image Generation using GANs, VAEs, and Diffusion Models [GitHub ↗]   **Apr 2024 – Jul 2024**
- Trained WGAN-GP with gradient penalty, β-VAEs, and diffusion models on CUB-200-2011 dataset for image synthesis.
- Developed custom loss functions to optimize image quality, achieving best performance with WGAN-GP **(33.07 FID)**.
- Accelerated diffusion inference **10x** (1000 to 100 time-steps) via DDIM sampling while maintaining same image quality.

Skills

Languages & Frameworks: Python, C++, SQL, PyTorch, TensorFlow, OpenCV, scikit-learn, Transformers, ONNX, Git
ML Training: RAG, RLHF (GRPO, DPO), SFT, PEFT, LoRA, VLM/LLM Fine-tuning, Distributed Training, LangChain
Infrastructure: AWS (SageMaker, EC2, S3), GCP, Docker, Kubernetes, TensorRT, CUDA, MLflow, Kafka, Spark, Flask

Publications

Srecharan Selvam, Abhisesh Silwal, George Kantor **"Self-Supervised Learning for Robotic Leaf Manipulation: A Hybrid Geometric-Neural Approach"**, arXiv:2505.03702, Under review at **ICCV 2025**. [**Project Page** ↗] [**PDF** ↗]