Santhoshini Gongidi

Website | Google Scholar | LinkedIn | GitHub | sgongidi@andrew.cmu.edu | 412-954-7844

Interested in making autonomy simpler, safer and seamless

Education

Carnegie Mellon University - School of Computer Science — Pittsburgh, PA
Master of Science in Computer Vision *(GPA: 4.20/4.33)* — December 2025
Coursework: Learning for 3D Vision, Multimodal Machine Learning, **ML Systems for Large Language Models** (LLM)

International Institute of Information Technology (IIIT), Hyderabad — Hyderabad, India
Master of Science in Computer Science & Engineering by Research, **Advisor: C.V. Jawahar** — December 2021
Coursework: Computer Vision, Deep Learning, Optimization Methods

Research Experience

3D Foundation Models | Geometry-Driven Robot Learning [blog] — January 2025 - Present
With Prof. Jeffrey Ichnowski | **Momentum Robotics Lab** — CMU

- Trained **diffusion policy** models on 3D scene representations (from VGGT) to match the performance of RGB-D–based 3D DiffusionPolicy in **MetaWorld** and **Adroit** simulation environments.
- Developed a lightweight alignment method built on Mast3r-generated point maps, enabling scalable multi-view spatial alignment and temporally coherent **4D scene reconstruction** for dynamic environments.
- Fine-tuning **3D foundation models** (VGGT, Rayst3R) as action-verifiers to integrate object completion priors into **vision-language-action models** (OpenVLA) to enable 3D scene reasoning and improve action planning.

GPU Optimization | Mixed Precision Attention for Vision Transformers [blog] — October 2025 - Present
With Prof. Zhihao Jia — CMU

- Implementing **Triton kernels for adaptive mixed-precision attention** by dynamically routing image patches to different precision paths within vision transformers, improving inference speed without sacrificing performance.

Image Generation| Text-aware 1D Image Tokenizers [blog] — January 2025 - May 2025
With Prof. Yonatan Bisk — CMU

- Trained a vision-transformer-based, **text-aware tokenizer** that compresses images to 50% of the token count required by conventional VAE models, leading to **compute-efficient generative AI** via a latent 1D diffusion transformer.

Work Experience

Research Engineer (Intern) | **Plus.ai** — May 2025 - August 2025
With Perception Team — Santa Clara, CA

- Proved feasibility of extending a vision-only perception model with **multimodal sensor fusion**, achieving a 7% relative mAP gain on moving-object detection.
- Implemented modular 2.5D feature extractor and multimodal fusion module with comprehensive ML unit tests in PyTorch Lightning. Designed to be extensible to other data modalities.
- **Automated large-scale 3D label generation pipelines** across Oracle GPU and CPU clusters, and **trained transformer-based models** on Oracle GPU clusters with Kubeflow orchestration and MLflow logging.

Senior Engineer, Applied ML | Micron Technology — July 2021 - November 2023
With Technology Products Group — Hyderabad, India

- Delivered **$5.6M in savings** by developing segmentation-based anomaly detection pipelines for wafer imagery.
- Trained and validated large-scale instance segmentation models on GPU clusters, ensuring statistical rigor.
- Expanded edge-case coverage by designing and developing image auto-labeling algorithms and tools.
- Led cross-functional teams to deploy **high-throughput inference APIs on Google Kubernetes Engine (GKE).**
- Designed unsupervised embedding learning methods for high-cardinality tabular data, driving innovation in daily work.
- Earned two **project excellence awards** recognizing technical contributions and execution.

Graduate Research Assistant | IIIT Hyderabad — June 2018 - December 2021
With Prof. C. V. Jawahar | **Computer Vision Lab** — Hyderabad, India

- Strengthened feature encoders in text recognition through joint training with droppable feature auxiliary heads, improving relative accuracy by 15–20% and yielding efficient feature heads. Adaptable to other ML architectures.
- Benchmarked and released a 10M handwritten dataset across 10 Indic scripts via semi-automated annotation pipelines, pioneering scalable labeling and benchmarking strategies applicable to **large-scale AI data curation**.
- Built a **multimodal text-to-image retrieval** engine that achieved 0.86 top-10 precision in **zero-shot retrieval**. [blog]

Skills

Expertise: 3D Reconstruction, 3D Foundation Models, Vision-Language Models (VLMs) Fine-tuning, 2D Perception
Tools: **Python, PyTorch, AWS, Google Cloud, FSDP/DDP, Triton, Slurm**, C++, Docker, Kubernetes, MLflow
Engineering Strengths: Test-Driven ML, Model debugging, Cloud deployment, Scalable data curation, Agile problem-solving

Publications

ICDAR 2021 (Switzerland, Handwritten Dataset & Benchmarking), **CVIP 2021** (India, Image Retrieval),
AAAI 2018 (USA, NLP), **ICDAR 2017** (Japan, Handwriting Recognition), **ACPR 2017** (China, Handwriting Recognition)