Santhoshini Gongidi

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

Interested in building scalable, efficient, and reliable ML systems for perception and reasoning

---

Education

Carnegie Mellon University - School of Computer Science
Pittsburgh, PA

Master of Science in Computer Vision *(GPA: 4.20/4.33)* — December 2025

Coursework: Multimodal Machine Learning, ML Systems for Large Language Models (LLM), 3D Vision

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

---

Work Experience

Research Engineer (Intern) | Plus.ai
May 2025 - August 2025

With Perception Team — Santa Clara, CA

- Demonstrated feasibility of extending single-modality **transformer architecture to multimodal architectures**, achieving a 7% relative accuracy improvement and enhanced robustness on complex real-world tasks.
- Developed a modular feature-extraction and fusion framework in PyTorch Lightning with ML unit-test coverage, designed for easy extension to similar data modalities.
- Automated **large-scale data-generation and training pipelines across distributed CPU/GPU clusters**, leveraging Kubeflow orchestration and MLflow logging to support scalable experimentation.

Senior Engineer, Applied ML | Micron Technology
July 2021 - November 2023

With Technology Products Group — Hyderabad, India

- Delivered $5.6M in savings by building **ML-based anomaly-detection** pipelines for large-scale manufacturing data.
- Trained and validated convolutional models on GPU clusters with strong statistical rigor and reproducibility.
- Designed **auto-labeling algorithms and active-learning tools**, expanding edge-case coverage and reducing manual effort.
- Deployed **scalable inference APIs on Google Kubernetes Engine** (GKE) to serve high-throughput predictions in production. Earned two project excellence awards recognizing technical contributions and execution.
- Developed unsupervised embedding methods for high-cardinality tabular data, improving downstream model accuracy.

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]

Research Experience

Foundation Models | Geometry Aware Robot Learning [blog]
January 2025 - Present

With Prof. Jeffrey Ichnowski | Momentum Robotics Lab — CMU

- Trained and evaluated **diffusion based policy models** that learn structured world representations, improving performance and reasoning consistency across tasks in simulated environments.
- Fine-tuning and adapting **large foundation models** (e.g., VGGT, Rayst3R) **as action or reasoning verifiers**, integrating geometric and semantic priors into **vision-language-action models** to strengthen grounding and decision quality.

GPU Optimization | Mixed Precision Attention for Vision Transformers [blog]
October 2025 - Present

With Prof. Zhihao Jia — CMU

- Developing optimized GPU kernels for adaptive mixed-precision inference to accelerate vision models.

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.

Skills

Programming: Python, C++, PyTorch, Triton, Docker, Kubernetes

Frameworks & Tools: MLflow, GCP, AWS, FSDP/DDP, Slurm, Kubeflow, MLFlow

Expertise: Vision & Multimodal ML, 2D and 3D Foundation Model Finetuning, Cloud Deployment

---

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)