Sai Madhusudan Gunda

Hyderabad, Telangana
gundasaimadhusudan@gmail.com | Sai Madhusudan Gunda | SaiMadhusudan

Education

International Institute of Information Technology, Hyderabad (IIIT-H) **07-2025 – 07-2026**
Master of Science in Computer Science and Engineering by Research (Dual Degree Programme) *CGPA: 9.0/10*

International Institute of Information Technology, Hyderabad (IIIT-H) **08-2021 – 07-2025**
B.Tech (Honours) in Computer Science and Engineering (Dual Degree Programme) *CGPA: 8.35/10*

Experience

BharatGen – National AI Mission **Jan 2025 – Present**
Generative AI Engineer *IIIT Hyderabad*

- Developed **Patram-7B-Instruct**, a multilingual vision-language model for document understanding, OCR, layout parsing, question answering, classification, and summarization across **Indian languages**.
- Built scalable data pipelines for synthetic QA generation, key-value extraction, and summarization, producing large-scale multimodal training samples for instruction tuning and supervised fine-tuning.
- Implemented efficient multi-node distributed training framework using **DeepSpeed ZeRO-3** on **256 H100 GPUs**.

Center for Visual Information Technology (CVIT), IIIT Hyderabad **May 2023 – Present**
Research Student – Computer Vision *Advisor: Prof. Ravi Kiran Sarvadevabhatla*

- Built ML systems for grounded document question answering, historical manuscript OCR, and multimodal document understanding, with a focus on scalable training, evaluation, and inference.
- Published **2 research papers** in top computer vision venues including CVPR and WACV, spanning document grounding, OCR, and multilingual tasks.

Language Technologies Research Centre (LTRC), IIIT Hyderabad **Jan 2024 – Apr 2024**
Undergraduate Researcher – Natural Language Processing *Advisor: Prof. Radhika Mamidi*

- Developed classification models for Indian languages, achieving **92%** accuracy on multilingual datasets by fine-tuning.
- Conducted experiments on cross-lingual transfer and low-resource language classification, contributing to improved understanding of multilingual NLP for Indic languages.

Publications

CURIO: Curvature-Aligned and Efficient OCR for Low-Resource Historical Manuscripts **WACV 2026**
Project Page
Proposes a curvature-aware OCR model that handles warped and degraded historical Indic manuscripts, improving recognition accuracy for low-resource scripts.

M3Grounder: Mask-Based Multi-Span and Multi-Granular Grounding for Document QA **CVPR 2026**
Project Page
Introduces a mask-based grounding framework that localizes multi-span evidence at phrase, line, and block levels for document question answering.

UniLipi: A Unified Multi-Script OCR for Historical Indic Manuscripts **ICDAR 2026**
Project Page
Unified OCR model across 13 Indic scripts with script-aware synthetic data, achieving 6.9% CER and strong cross-script generalization.

Multi Script OCR System for Handwritten Indic Manuscripts **DAS 2026**
Project Page
Demo of Unified OCR model across Indic scripts.

DoCoG: Mask-based Multi-Type Grounded Chain-of-Thought for Document QA (Oral) **ECCV 2026**
Project Page
Grounded document question answering with step-wise reasoning .

DocLayout-VL: Hierarchical, Open-set, and Promptable Document Layout Segmentation **ECCV 2026**
Project Page
Promptable foundational document layout model achieving SOTA results across document layout benchmarks.

Projects

Scalable Vision-Language Document Inference with vLLM | *Python, PyTorch, vLLM*
- Built high-throughput inference pipelines for document models using **vLLM**, enabling efficient processing of **2M+ documents**.
- Reduced inference latency by **2.5**× through prompt batching, cached image embeddings, and dynamic request scheduling.
- Benchmarked throughput, memory footprint, and response quality across different configurations for document QA workloads.

Multi-Page Document Grounding with Agentic RAG | *Python, PyTorch, RAG, Vision-Language Models*
- Extended **M3Grounder** to a multi-page retrieval-augmented generation pipeline, enabling grounded QA over long documents with evidence spanning multiple pages.
- Built a page-level retriever and grounding-aware answerer that uses fine-grained evidence localization to generate answers tied to supporting regions.

Document QA and Key-Value Extraction Pipeline | *Python, PyTorch, OCR, Transformers*
- Designed an end-to-end pipeline for document question answering and key-value extraction using OCR, layout-aware features, and multimodal transformers.
- Integrated synthetic data generation, grounding, and evaluation workflows to support scalable experimentation on document intelligence tasks.

Mixture-of-Experts Routing for Document Models | *PyTorch, Distributed Inference*
- Explored routing strategies to direct document inputs to specialized experts, achieving **15%** compute savings with less than **1%** accuracy drop.
- Evaluated accuracy-latency-compute trade-offs to identify efficient inference strategies for multimodal models.

Recommendation Engine with Collaborative Filtering | *Python, Scikit-learn, SQL, Flask*
- Developed a hybrid recommendation system combining collaborative filtering and content-based methods on the **MovieLens 25M** dataset and served predictions via a REST API.

Real-Time Object Detection & Tracking Pipeline | *Python, PyTorch, OpenCV*
- Built an end-to-end detection and tracking system using **YOLOv8** for multi-object tracking on surveillance videos, with ONNX export for optimized deployment.

Technical Skills

Programming: Python, C/C++, JavaScript, SQL, HTML/CSS
Deep Learning: PyTorch, Hugging Face Transformers, DeepSpeed, vLLM, Ms Swift, Unsloth
Tools: Docker, Git, Linux, AWS (S3, EC2), MongoDB, MERN Stack
Core Areas: Computer Vision, Document AI, NLP, Distributed Training, Model Optimization

Achievements

- Awarded **Dean's List** at IIIT Hyderabad for two consecutive years (2021, 2022).
- Secured **1st Place** in **Megathon** Hackathon conducted by IIIT Hyderabad and Qualcomm.
- Achieved **7th position** in **Amazon ML Challenge** (2024).
- National-level exam ranks: UGEE (19), JEE Mains (5055), JEE Advanced (3240 AIR).