SRI HARSHA MUDUMBA

(515)-916-3011 ◇ srim@iastate.edu ◇ LinkedIn • GitHub • Iowa, IA

EDUCATION

Master's in Computer Engineering, Iowa State University, Iowa, IA Aug 2023 - Aug 2025
GPA: 3.77/4.0
Relevant Coursework: Deep Learning, High-Performance Computing, Computer System Architecture, Hardware Design for ML, Advanced HPC for AI, Distributed Systems, Computational Perception.

TECHNICAL SKILLS

Programming Languages	Python, C++, C, SQL, Shell scripting
AI Frameworks	TensorFlow, PyTorch, ONNX
HPC	OpenMP, MPI, Slurm
Databases	Oracle 11g/12c/19c, Oracle EBS
Other	Benchmarking, Machine Learning, LLMs, RPC, Gem5, Linux, MATLAB, Simulink, Digit

PROJECTS

Lazy Neural Network Optimization (Present) *Python, PyTorch, TensorFlow*
Developing LazyNN with **adaptive early exit**, **multi-modal processing**, and **diffusion-based generative restoration**. Implementing **backtracking-based optimization** to reduce redundant computations, leading to **faster inference**. Applying techniques to optimize **diffusion models** by integrating efficient computation strategies and reducing unnecessary activations.

Benchmarking 1BitLLM with ONNX and IREE *Python, ONNX, IREE, PyTorch*
Designed a framework to analyze **latency and scalability**, optimizing inference using ONNX and IREE compilers for deployment of 1BitLLM.

SnaPEA Neural Network Inference Optimization *Python, TensorFlow, PyTorch*
Developed SnaPEA framework optimizing deep learning by dynamically pruning non-contributive activations using percentile-based thresholds, sparse computation, mixed precision training, and early exit mechanisms, achieving up to **74% faster inference with minimal accuracy trade-offs** across models like GoogLeNet, ResNet, and VGG-Net

SimNet Computer Architecture Simulation *C++, Python, PyTorch*
Built a CNN-based simulator achieving **55x faster throughput** than Gem5, improving fetch latency, branch predictions, and cache behavior with **3.5% reduced simulation errors**.

Vehicle Number Plate Detection System *C++, OpenCV, MPI*
Automated number plate detection using OpenCV and EasyOCR, improving accuracy under varying conditions. Integrated **multi-modal learning** for enhanced feature extraction and applied **low-level vision techniques** such as **spatial and temporal image analysis** to improve detection in challenging environments. MPI parallelization led to a **90% reduction in runtime**.

Music Genre Classification *Python, TensorFlow, Keras*
Developed a CNN-based classifier merging GTZAN and FMA datasets, utilizing **multi-modal foundation models** to enhance feature extraction. Applied **deep learning-based low-level audio processing** techniques to improve classification accuracy. Implemented augmentation techniques and real-time deployment for robustness.

PUBLICATIONS

R. Ippalapally, S. H. Mudumba, M. Adkay, and N. V. H. R., *"Object Detection Using Thermal Imaging,"* IEEE 17th India Council International Conference (INDICON), 2020, doi: 10.1109/INDICON49873.2020.9342179.

PROFESSIONAL EXPERIENCE

Associate Database Administrator Aug 2020 - Jul 2023
Cognizant Technology Solutions, Bengaluru, India
Managed Oracle databases and optimized system performance. Automated database processes reducing patching time by **50%**, improved security with Active Directory integration, and enhanced efficiency with Linux-based database cloning solutions.