Nishanth Marer Prabhu

Google Scholar | nishanth.marerprabhu@gmail.com | +1(857)-390-4372 | LinkedIn | GitHub

EDUCATION

Northeastern University, Boston, MA **December 2024**

Master of Science in Electrical and Computer Engineering (Specialization in Machine Learning) **GPA - 4/4**

Coursework: Introduction of Machine Learning, Advances in Deep Learning, Parallel Processing for Data Analytics, Computer Vision, Fundamentals of Computer Engineering, High-Level Design of Hardware/Software Systems

Audit: Reinforcement Learning, Advanced Algorithms, Computer Architecture

BNM Institute of Technology, Bengaluru, India **May 2019**

Bachelor of Engineering: Electronics and Communication Engineering

Coursework: Digital Signal Processing, Microprocessor and Microcontrollers, Wireless Cellular and LTE 4G, Image Processing

TECHNICAL SKILLS

Programming Languages: C\C++, Python, SystemC, PowerShell, Perl, C++ with Qt, SPARQL, C#, MySQL, Java, Matlab, HTML

Frameworks: Data Structures and Algorithms, Design Patterns, PyTorch, Tensorflow, PySpark, Joblib, ONNX, TensorRT

Technologies / Tools: NVIDIA Jetson Orin, PySpark, Dockers, Linux**,** Jenkins, AWS EC2, S3, Git, CUDA, NVIDIA Nsight Compute

EXPERIENCE

Embedded System Laboratory, Northeastern University, Boston, USA **January 2024 - Present**

***Research Assistant* | Advisor: Prof. Gunar Schirner**

PyTorch to TensorRT model conversion for improved inference speed on NVIDIA Jetson Orin Edge Device

- Achieved up to **20x** speedup in **CNN, Vision Transformer** models using mixed precision, optimizing inference across batch sizes.
- Utilized **NVIDIA Nsight Compute** and **Netron** to validate **ONNX** conversions and analyse computational graphs, identifying **CNN-LSTM** hybrid model bottlenecks from sequential unrolling that limited parallel processing.
- Conducted **layer-wise precision control**, manually controlled layer-wise precision **(FP32/FP16)** for optimal performance and accuracy.
- Employed version-specific **Docker containers** to encapsulate dependencies, ensuring consistent and reliable inference environments.

Terrain Augmented Channel Model (TACM) for Automatic Modulation Recognition (AMR)

- Developed a **Convolutional Vision Transformer (CvT)** that reduced the number of training parameters compared to the standard **Vision Transformer**, improving model efficiency and performance on the TACM2024 dataset.
- Designed a **novel receiver placement algorithm** to balance load proportionally by receiver capacity, simplifying calculations, reducing compute time, and enhancing dataset generation efficiency.
- Utilized **Pytorch Lightning** to modularize training, control logging, and enable batch randomization, incorporating **TensorBoard** for visualizing training loss curves and model's performance metrics.
- Employed **Linear Sum Assignment**, **Mixed-Integer Programming (MIP),** and **Quadratic Least Squares Solvers** to improve receiver placement accuracy, enhancing spatial signal recognition.
- Conducted detailed **CUDA timing analysis** to identify and resolve inefficiencies due to multiple kernels launches and identifying bottlenecks reducing execution time from **4ms to 950µs**.

Siemens, Bengaluru, India **July 2019 - December 2022**

***Senior Software Engineer* | R&D at Smart Grid Infrastructure (C++/Machine Learning)**

- Filed a **Patent** and **Invention Disclosure** for a **Proximity Search-Based Algorithm** to optimize numerical solvers' performance.
- Implemented **ML techniques** like **KNN** and **Graph NN** to enhance the **convergence rate of compute-intensive applications**.
- Designed and developed applications using **C++**, **C#,** and **MySQL,** utilizing **Design Pattern** methodology for better logic flow.
- Built an OS deployment tool on **Jenkins**, cutting setup time from 1 hour per machine to 20 minutes, boosting system test efficiency.
- **Won STAR Performer Award five times** and **2nd place in a Siemens global AI/ML Hackathon** organized by **Architects.**

PUBLICATIONS

- Utilizing terrain-generation to derive realistic channel models for automatic modulation recognition, Proc. SPIE 13035, Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications II, 130351B (7 June 2024)
- Enhancing Automatic Modulation Recognition for IoT Applications Using Transformers, IoT, vol. 5, no. 2, Art. no. 2, Jun. 2024

PROJECTS

Shallow Convolutional Neural Network for Image Classification **February - April 2024**

- Designed **parallel CNN streams** using **VGG** and **ResNet** blocks for **feature map retention** and **multi-scale feature extraction**.
- Achieved **89.5%** accuracy on CIFAR10 with **1.3M** trainable parameters versus 6.5M in the benchmark.

Parallelization of Genetic Algorithms for Optimal Feature Selection **October - December 2023**

- Leveraged **PySpark, Joblib,** and **Genetic Algorithms** for feature selection, **boosting task distribution speed** by **2x to 25x**.
- Achieved speed enhancements for **scikit-learn models** (MLP, Logistic Regression, XGBoost) while maintaining high accuracy.

Medically Informed Stable Diffusion **November - December 2023**

- Fine-tuned **Stable Diffusion** model from **Hugging Face** using detailed prompts generated using **LLMs** on custom brain scan datasets.
- Enhanced image generation with a **control net**, producing anonymized, **medically relevant brain scans for research**.

Reinforcement Learning Solution for Multi-hour Unit Commitment and Economic Dispatch **April - July 2020**

- Formulated a **Markov Decision Process** using **Python** for **multi-horizon security constraints** for UI and ED in power systems.
- Increased **efficiency by 98%** by eliminating intangible states, demonstrating the feasibility of RL for these tasks.