Srinath Srinivasan

Linkedin: linkedin.com/in/srina1h / Github: github.com/srina1h Website: srina1h.github.io

EDUCATION

North Carolina State University

Raleigh, NC

Master's in Computer Science: GPA: 4.0/4.0

Aug 2023 - Present

Email: srinath.ksrini@gmail.com

Courses: Neural Networks, Compiler Construction, Computer Networks, Software Engineering, Efficient Deep Learning, Design & Analysis Of Algorithms, Deep Learning Beyond Accuracy, Efficient Tensor Processing for AI

Anna University

Chennai, India

Bachelor of Engineering in Computer Science; GPA: 8.8/10 (3.82/4.0)

Aug 2018 - May 2022

Courses: Operating Systems, Data Structures, Analysis Of Algorithms, Artificial Intelligence, Databases, Theory of Computation, Discrete Mathematics, Probability and Statistics, Distributed Systems, Computer Architecture, Cloud Computing, Programming in C

Research Experience

North Carolina State University

Raleigh, USA

Graduate Research Assistant (Advisor: Dr. Jiajia Li) Aug 2023 - Present

- Realized 10% improvement per iteration in compressed LLM inference by identifying bottlenecks in PyTorch components
- o Introduced novel tensor contraction techniques into an existing tensorized transformer model connecting PyTorch to high performance CUDA/C++ libraries. Studied implications/overhead of multi-language library use.
- o Created pipelines to benchmark intermediate layer operations from PyTorch such as Transformer, Dense on NVIDIA GPUs

Robert Bosch Centre for Data Science and Artificial Intelligence

Chennai, India

Research Intern

Apr 2022 - Aug 2022

- Used gradient boosting and graph neural networks to predict ambulance response times based on position in Chennai city
- Project done in collaboration with 108 ambulance service of Chennai (State Govt of Tamil Nadu)

Indian Institute of Technology Madras

Chennai, India

Resesarch Intern

Jun 2020 - Oct 2021

- Trained complex-valued regression models for loudspeaker source localization performing 5 times faster than conventional numerical methods while preserving accuracy. Co-authored a publication in the Journal of the Acoustical Society of America.
- o Co-authored a paper to recognize, localize and reconstruct pressure signals to model it on an adjacent surface

Robert Bosch Centre for Data Science and Artificial Intelligence

Chennai, India

Research Intern

Apr 2021 - Aug 2021

o Worked on predicting student dropouts from NPTEL (MOOC) courses focusing on explainability. Insights gained were used by NPTEL to understand and combat various issues leading to student dropouts

Publications (Google Scholar)

• Extracting Usable Predictions from Quantized Networks through Uncertainty Quantification for OOD Detection (Link)

Rishi Singhal, Srinath Srinivasan

 $ArXiv\ Preprint\ (2024)$

- Machine learning aided near-field acoustic holography based on equivalent source method (Link) S. K. Chaitanya, Siddharth Sriraman, Srinath Srinivasan, K. Srinivasan The Journal of the Acoustical Society of America (2023)
- Data-Driven Neural Networks for Source Localization and Reconstruction Using a Planar Array (Link) Sai Manikanta Kaja, **Srinath Srinivasan**, K. Srinivasan International Journal of Aeroacoustics (2022)
- Exploring Bayesian Uncertainty Modeling for Book Genre Classification (Link) Srinath Srinivasan, SG Shivanirudh, Sujay Sathya, TT Mirnalinee IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (2022)
- Equivalent source method based near-field acoustic holography using machine learning (Link) S.K Chaitanya, Siddharth Sriraman, **Srinath Srinivasan**, K Srinivasan The 51st International Congress and Exposition on Noise Control Engineering (Internoise 2022)
- LRCNs for Stroke Detection in Table Tennis (Link) Siddharth Sriraman, Srinath Srinivasan, Vishnu K Krishnan, J Bhuvana, TT Mirnalinee MediaEval Workshop (2019)

Teaching & Service

- Teaching Assistant for CSC591: Parallel Algorithms Fall 2024 at North Carolina State University
 - Sole TA for Parallel Algorithms, a research oriented course by Dr. Jiajia Li. Course objectives were to review several publications in the HPC domain on EasyChair (a conference portal) and a course project
 - Prepared key points from student's reviews for in-class discussion validating them against OpenReview comments
 - o Graded assignments and organized a project Symposium to showcase student's project posters

Work Experience

Data Aces

Sugar Land, TX, USA Jun 2024 - Aug 2024

ML Engineering Intern

- o Developed a Gen AI aided solution to explain flagging of fraudulent transactions to banking customers
- Designed a RAG pipeline to feed the LLM learned knowledge from XGBoost model along with an individual user's recent transaction data compounding information from both vector and relational databases. Improved explainability of fraud.

Poshmark

Chennai, India

Software Engineer - 1

Aug 2022 - Jul 2023

- Developed key infrastructure for the Poshmark iOS app such as screen pagination algorithms from scratch in the SwiftUI Combine based publisher-subscriber design paradigm
- \circ Migrated multiple legacy objective C components to the SwiftUI MVVM design pattern, resulting in a 60% reduction in code complexity and 10% improvement in screen load times. Oversaw slow-roll out of new components through A/B testing

Kaleris (formerly Navis)

Chennai, India

 $Associate\ Software\ Engineer\ -\ Intern$

Jan 2022 - July 2022

- Resolved mission critical bugs in the Navis N4 Terminal Operating System, increasing system stability and code coverage
- \circ Worked on an adjacent AI-aided solution to enable smart docking of ships saving $\sim \! 15$ minutes of manual calculation

Renault Nissan

Chennai, India

Software Development Intern

Aug 2021 - Oct 2021

- Designed and developed a scalable cloud-based pipeline to enable car dealerships to remotely communicate key data points to analysts. Pushed the MVP into production with minimal errors
- \circ Application reduces delay in creating analytical reports & minimizes manual entry errors through multiple validation checks bringing down weekly engineering efforts of data transfer and validation from \sim 3 hours to a few minutes

PROJECTS

- Efficient Monte Carlo dropout for uncertainty estimation Designed a model splitting technique speeding up model uncertainty estimation by upto 33x. Devised technique can be applied to pre-existing models as well. (Link). Project guided by Prof. Dongkuan Xu at NC State
- Synthetic cough generation using diffusion models Used a diffusion model to generate synthetic data for the small COUGHVID dataset for better Covid-19 cough detection. Improved efficiency using latent diffusion to achieve similar performance with lower training time. (Link)
- LLVM tool for C code optimization (in progress) Creating a 2 part tool to parse LLVM IR and based on variable reuse, identify the most influential variables on runtime performance and judge libraries adding most overhead
- Retinal OCT Disease Classification Retinal disease identification using Optical Coherence Tomography images. Comparison of VGG-19 and lightweight MobileNet v2 based transfer learning models. Highlights potential of lightweight models in complex tasks.(Link)

Test Scores

- GRE: 332 170 Quant, 162 Verbal, 4.0 AWA (Link to Score Report)
- TOEFL: 116 30 Reading, 27 Listening, 30 Speaking, 29 Writing (Link to Score Report)

Honors and Awards

- 1st place, Shaastra AI Games Challenge hosted by IIT Madras 2021 (Link)
- Top 5%, NPTEL Course Deep Learning 2021 (Link)
- National Rank 1, NPTEL Course Introduction to internet of things 2021 (Link)

SKILLS

- Programming Languages: Python, C/C++, Swift Java, JavaScript, SQL, HTML, Linux Scripting
- Tools/Technologies: Git, PyTorch, TensorFlow, CUDA NumPy, Pandas, CuPy

Extracurriculars

- Karate Black belt level 1 at Hayashi Ha Karate Do (Japanese Karate organization)
- Trained Carnatic (Indian classical) music vocalist Performed at many community events