



EDUCATION

Program	Institution	%/CGPA	Year of completion
Dual Degree CSE (B Tech + M Tech)	Indian Institute of Technology Madras	7.83	2020
XII (CBSE)	Jawahar Navodaya Vidyalaya, Raipur	93.80	2014

SKILLS

- Languages (Technical): C, C++, CUDA, OpenCL, Java, Python, HDL
- Frameworks and APIs: OpenMP, MPI, SPARK, Tensor-flow, OWL
- Library: cuDNN, cuBLAS, CUB, Thrust, cuML, cuGraph

PROFESSIONAL EXPERIENCE

Associate Engineer, Algorithm R&D at KLA, Chennai (July 2020 - Present)

- Design and implement custom AI and Machine learning algorithms on GPU.

Algorithm and AI Intern at KLA, Chennai (December 2019 – June 2020)

- Developed parallel algorithms for Auto Segmentation, Feature Selection and Random Forest on GPU.
- Optimized Random Forest training on GPU, which is faster than cuML.
- Implemented Object Oriented Inference with dynamic parallelism for efficient load balancing on GPUs.

Algorithm Intern at KLA, Chennai (Summer 2019)

- Implemented parallel Inference and Sampling algorithms on GPU, achieved 2x-8x speedup than OpenMP.
- Scaled up the GPU programs using SPARK 2.1 framework for multi-node/multi-GPU systems.

Software Intern at eClerx, Mumbai (Summer 2018)

- Object detection and localization in Image using CNN, achieved 88% accuracy on the custom dataset.
- Developed an algorithm to verify websites from pdf-based wireframes using OCR.

Android Developer at Machadalo (IITB startup), Mumbai (Summer 2017)

- Developed an Image Auditing mobile application to capture images, verify and upload in the server on Android.
- Used OpenCV for image matching and feature matching to catch fraudulent images in database.

PROJECTS

- **A* algorithm for Dynamic Graphs on GPU** (2019 - 2020)
 - Built a framework for parallel dynamic A* which handles insertions, deletions, and fully-dynamic operations.
 - Proved crucial properties of the dynamic computation, which allowed to implement synchronization effectively.
 - Achieved 24x-54x speedup than static A* for SNAP datasets on 100 batch updates.
 - Applied the framework to different applications of A*: wireless sensor networks, path planning, and game trees.
- **Sparse Tensor Transpose Operation on GPU.** (2018)
 - Parallelized Tensor Transpose operation on GPU, achieved coalesced memory access for both input and output tensor.
 - Performance improvement persists with varying ranks, varying permutations and varying index ranges.
- **ABYSS GENOME Assembler on GPU.** (2018)
 - Parallel de-novo assembling of reads into genome sequence and optimized contig formation.
 - Modified data-structure to reduce 10x space and achieved 6x-8x speedup as compared to OpenMP.
- **Lock Contention aware Scheduler for NUMA architecture** (2017)
 - Implemented *shuffling*, it migrates threads across sockets so that a thread seeking lock can find it on the same socket.
 - Reduces the time spent on acquiring locks and shared data access in the critical section.

POSITION OF RESPONSIBILITY

- Co-Founder and CTO of **edAR labs** (CIN U80302TG2019PTC134107) (2017-2019)
 - edAR is an AR based learning platform for school students that focuses on experiential learning.
 - Responsible for managing the development of the product.