MUNEEB AHMAD

muneeb0ahmed3@gmail.com

[Website](#) | [GitHub](#)

Davis, California

+1 (408) 497-7947

EDUCATION

University of California, Davis, CA, USA — **Sept '25 - June '27**
Master of Science in **Electrical and Computer Engineering**
    Coursework: Modelling & Optimization in Computer Engineering, High Performance Statistical Computing, Information Theory & Coding

Jamia Millia Islamia, New Delhi, India — **2021 - 2025**
Bachelor of Technology in **Computer Engineering**
    CGPA: **9.05 out of 10**
    Coursework: Computer Architecture, Parallel & Distributed Computing, Operating Systems, Machine Learning, Digital System Design

WORK EXPERIENCE

Undergraduate Researcher at [BeyondDefence Lab](#), University of New Mexico — **Jan '25 - Present**
- Creating a framework to find the optimal deployment strategy for a given deep learning on model on a given edge device. Target use cases are satellites and robotics.
- Benchmarked quantized classification models on the Coral EdgeTPU dev board, achieving a 400% improvement in latency.
- Worked with hardware accelerators such as Coral EdgeTPU and Nvidia Jetson Nano.
- Applied quantization to compress models and transformed formats for compatibility with hardware.

Advanced Application Engineering Intern at Accenture, Gurugram, India — **May '24 - July '24**
- Gained knowledge of data security standards and compliance frameworks.
- Performed cybersecurity audits of 3 organizations as part of third-party risk management.

Intern at Nikah Forever, New Delhi, India — **Jun '23 - Aug '23**
- Developed content-based filtering algorithm for a best match recommendation system with user activity driven results to improve over the existing rule-based system.
- Optimized processing in numpy to reduce the memory footprint by 50% to allow for deployment of the recommendation system on a high traffic platform.

PROJECTS

SpMV on GPUs using DirectX Work Graphs 🔗 — **July '25 - Present**
- Implemented load balanced Sparse Matrix Vector Multiplication (SpMV)— an important operation used to accelerate deep learning models to the GPU —using DirectX12 Work Graphs to achieve better performance despite the irregular nature of the application.
- Currently benchmarking against existing methods such as the cuSparse library.

Vulkan Forward Clustered Renderer 🔗 — **April '23 - July '24**
- Implemented a forward-clustered renderer using Vulkan to handle large numbers of lights.
- Compute shader determines lights relevant to a cluster of space for optimal processing.
- Voxel renderer with compute shader based meshing to reduce draw count by 6 times.
- Focus on easy to modify rendering pipeline.

Engine-2: Real-time 3-D rendering engine 🔗 — **Oct '21 - Jan '23**
- Created a rendering engine for drawing large amounts of dense geometry utilizing visibility buffers in C++/OpenGL.
- Implemented a data-oriented object storage system (ECS) using sparse arrays. 🔗

SKILLS

- **Languages** - C, C++, Python, GLSL, HLSL, Rust, JavaScript, SQL
- **Frameworks** – CUDA, Vulkan, DirectX12, OpenGL, OpenMP, MPI, numpy, PyTorch, Tensorflow
- **Tools** – RenderDoc, NVIDIA Nsight, Visual Studio, perf, CMake