Muhammed Jaseem

Banglore, India | jaseemseems@gmail.com | +91 7511135725 | jasseeeem | jasseeeem

About Me

I am a Machine Learning Engineer with 2.5+ years of experience building scalable real-time computer vision systems. Skilled in GPU-accelerated streaming pipelines, Python/C++, API and model latency optimisation, and deploying ML systems at scale. Focused on customer-driven solutions and passionate about contributing to open-source projects.

Skills

Languages: Python, C++, Go, Java, JavaScript/TypeScript, SQL, Terraform

Technologies: Docker, Kubernetes, CUDA, Knative, OCI, Prometheus, SWIG, Pybind

Frameworks & Tools: DeepStream, Gstreamer, FastAPI, Drogon, NVIDIA Triton, React, Node.js, PyTorch, ONNX

Experience

Member of Technical Staff — *Banglore*
Oracle — *June 2023 – Present*

- Designed and deployed a high performance asynchronous **Go server integrating a C++ DeepStream/Gstreamer pipeline** for **real-time tracking of faces, weapons, and objects** across 15 concurrent RTSP/WebRTC/LiveKit streams on a single GPU. Migrated from Python to Go and C++ using SWIG for Go–C++ integration.
- Developed a **Python service for real-time detection and tracking** of persons and weapons using the OCSort tracker for public safety applications.
- Migrated Object Detection, Weapon and Face Tracker code to **C++ from Python**, cutting **CPU usage by 50%**, improving **latency by 25%**, and **reducing failures** in production.
- Accelerated multiple vision models by **50–100×** by migrating from CPU to GPU using **MNN and CUDA**.
- Reduced infrastructure costs by over **$60,000/month** by benchmarking and minimising Kubernetes resource consumption and optimising NoSQL database usage.
- **Tritonised** PaddleOCR model to **double the throughput** and **reduce memory usage by more than 20%**.
- Significantly reduced model serving container restarts in production by identifying and fixing critical memory leaks.
- Migrated pretrained vision models from **K8s HPA to Knative** for scale-to-zero capability and faster autoscaling.

Server Technology Intern — *Banglore*
Oracle — *May 2022 – July 2022*

- Designed a generic asynchronous notification infrastructure and implemented a POC web application to notify users about the completion of various actions in OIC web application.

Education

National Institute of Technology, Calicut — *Aug 2019 – May 2023*
B. Tech in Computer Science and Engineering

- CGPA: 8.94 / 10

Positions of Responsibility

Executive, FOSSCell — Core organizer of FOSSMeet 2023 - NIT Calicut's annual open-source conference.

Executive, Computer Science and Engineering Association — Led design for Threads, the official magazine of CSEA, and contributed to a wide range of creative initiatives including posters, promotional videos, and event branding.

Member, Ragam and Tathva Tech Team — Built official websites for South India's largest tech fest, including an auto-generated certificate system.

Projects

Rain App — *2022*

- Built a full-stack web application in collaboration with Wayanad district officials and alumni to visualize rain gauge data and monitor rain gauges across the district, supporting efforts to mitigate the impact of landslides.

AR Sudoku Solver — *https://git.io/JtBeY*

- Developed a script to extract Sudoku grids from newspaper images, solve them programmatically, and render the solution back onto the source image.