Course Name: Computer Vision

Weekly Report: 5

Group Name: XYZ

Submitted to faculty:

Mehul Raval

Date of Submission:

29 Mar 2025

Student Details

Roll No.	Name of the student	Name of the program
AU2240106	Meet Rathi	B.Tech in CSE
AU2240160	Harsh Panchal	B.Tech in CSE
AU2240153	Aditya Agarwal	B.Tech in CSE

Table of Contents.

WORK DONE THIS WEEK

A performance evaluation of code to achieve Jetson compatibility.

A basic performance assessment is mandatory for the existing code before implementing deep optimization for Jetson Orin AGX. The evaluation targets locating weaknesses combined with opportunities for enhancement.

Key Areas of Review

A memory usage optimization process should reduce duplicate tensor storage combined with eliminating unneeded data loading activities.

The computational delivery process needs optimization to shift intensive CPU operations onto GPU acceleration for better speed.

The system needs to handle data through efficient batch processing together with minimized CPU-GPU data transmission operations.

Parallelism & Hardware Utilization – Analyze all possible opportunities to use multi-threading while integrating CUDA acceleration and TensorRT technology.

The code requires refinement through elimination of redundant sections along with architectural improvement while reducing logging processes.

WORK TO BE DONE NEXT WEEK

TensorRT Conversion for Inference Optimization

- The conversion of the model to TensorRT will optimize its speed when running inference operations on Jetson Orin AGX.
- Benchmark performance improvements in terms of latency and power efficiency.

Optimize Resource Scheduling

- During inference operations it is important to apply methods which optimize CPU-GPU workload interaction to prevent system performance decreases from bottlenecks.

Fine-Tune Model Hyperparameters

- A series of efficiency improvements can be achieved through tuning learning rate and batch size and epoch number settings.

The deployment process of the model takes place on Jetson Orin AGX.

- Apply and test the person retrieval system on the edge hardware platform.
- Run proper tests to determine the actual operational performance while using the restricted hardware resources.

Evaluate Performance Metrics on Jetson

- Measure speed, power consumption, and accuracy post-deployment.