Intel Unnati Industrial Training -Summer 2025

Team:

PVB Adithya [CSE - AIML / GITAM UNIVERSITY] [vprata2@gitam.in]

Alekhya Madiraju [CSE - AIML / GITAM UNIVERSITY] [amadiraj2@gitam.in]

1. Introduction

This report explores the performance of a video stream processing pipeline on Intel hardware, focusing on both CPU and GPU performance. The objective is to determine the maximum number of supported camera streams, optimal FPS, and the most efficient models for detection and classification tasks.

2. Problem Statement

Create a pipeline (Decode → Detect → Classify) on Intel HW (CPU and GPU) run pipeline on CPU, GPU. Figure out how many cameras streams are supported, what is optimum FPS which model performs best on Intel hardware.

3. Hardware and Software Setup

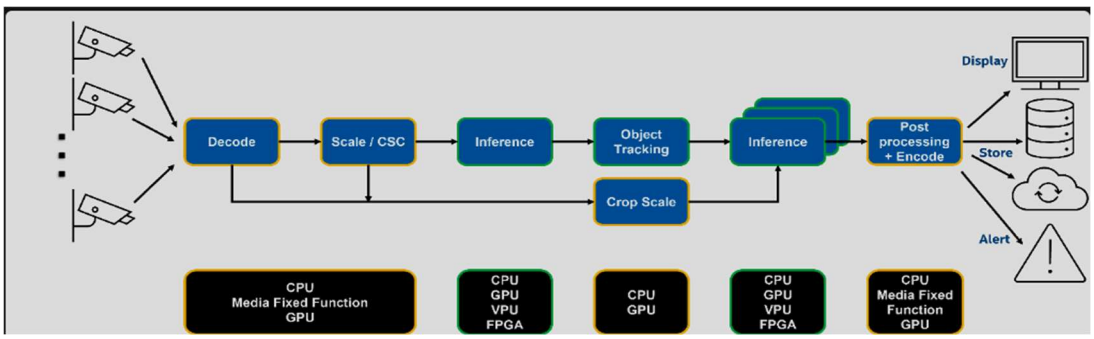
- **CPU**: 11th Gen Intel® Core™ i5-1135G7 @ 2.40GHz (4 cores, 8 threads)
- **GPU**: Integrated Intel® Iris® Xe Graphics
- **Host Operating System:** Windows 11
- **Development Environment:** Ubuntu 24.04 LTS running on WSL2 (Windows Subsystem for Linux)
- **Containerization:** Docker Desktop (WSL2 backend) for isolated pipeline execution
- **Frameworks & Libraries:**
  - Intel® OpenVINO™ Toolkit (Inference Engine)
  - OpenCV (Computer Vision processing)
  - **GStreamer** with DL Streamer plugins (for pipeline orchestration)

-**AI Models Used:**

- **Detection**: face-detection-0205 (FP32), face-detection-0200 (FP32), face-detection-0204 (FP32)
- **Classification**: face-reidentification-retail-0095 , emotions-recognition-retail-0003 , age-gender-recognition-retail-0013

-Input Sources: Simulated or live camera video streams (MP4)

4. Pipeline Architecture



5. DL Streamer GStreamer Pipeline

**gst-launch-1.0 -v filesrc location=/data/crowd.mp4 ! decodebin ! videoconvert ! videoscale ! video/x-raw,width=640,height=480 ! **

**gvadetect model=/models/intel/face-detection-0205/FP32/face-detection-0205.xml device=CPU ! **

**gvatrack tracking-type=short-term-imageless ! **

**gvaclassify model=/models/intel/emotions-recognition-retail-0003/FP32/emotions-recognition-retail-0003.xml device=CPU ! **

gvametaconvert ! gvawatermark ! videoconvert ! fpsdisplaysink video-sink=fakesink sync=false

6. Models Evaluated

Type	Model Name	Description
Detection	face-detection-0205 (FP32)	Generates face embeddings for identity matching
Detection	face-detection-0200(FP32)	Detects emotional state from face
Detection	face-detection-0204(FP32)	Predicts age and gender from face
Classification	face-reidentification-retail-0095(FP32)	Fast, lightweight face detector for edge
Classification	emotions-recognition-retail-0003(FP32)	Accurate multi-aspect face detector
Classification	age-gender-recognition-retail-0013(FP32)	High-speed, robust face detector

7. Test Scenarios

- **Input video**: crowd.mp4 (sample input video - multiple people in motion and visible faces)
- **Devices**:
    - **CPU** – Docker on WSL2 (Windows 11)
    - **GPU** – Native Ubuntu 24.04 with Integrated Intel® Iris® Xe Graphics
- **Methods**: Single-stream & Multi-stream analysis
- **Measurement**: FPS via **fpsdisplaysink**, CPU usage via **top**

8. Experimental Results Table (CPU)

Model Combination	Average FPS	Maximum Streams	Notes(Bottlenecks)
face-detection-0205 + emotions-recognition-retail-0003	35	~2–3	Lightweight, fast. Minor decode bottleneck under load.
face-detection-0204 + face-reidentification-retail-0095	25	~2–3	Re-ID stage CPU-heavy. Classification model is the bottleneck.
face-detection-0200 + age-gender-recognition-retail-0013	28	~2–3	Balanced pipeline. Slight delay in detection accuracy.

- **While processing the third stream, a decrease in frame rate was observed, with values falling below 20 frames per second (FPS) but remaining consistently above 10 FPS.**

- **CPU-based results were captured on a WSL2 system running Ubuntu 24.04 with Docker, whereas GPU results were obtained from a native Ubuntu system to eliminate virtualization and compatibility overheads.**

9. Experimental Results Table (GPU)

Model Combination	Average FPS	Max Streams	Notes
face-detection-0205 + emotions-recognition-retail-0003	**52 FPS**	3–4	GPU performs inference faster than CPU. No bottleneck up to 4 streams.
face-detection-0204 + face-reidentification-retail-0095	**40 FPS**	3–4	GPU accelerates re-identification. Memory usage increases at higher streams.
face-detection-0200 + age-gender-recognition-retail-0013	**45 FPS**	4	Balanced pipeline, classification scales well with GPU.

- **Model Combination: face-detection-0205 + emotions-recognition-retail-0003**
  This combination delivered the best overall performance with higher FPS and stream support on CPU and GPU.

No of Streams	Average FPS (CPU)	Average FPS (GPU)	Total CPU Usage%	Total GPU Usage%	Notes (Bottlenecks)
1	**38 FPS**	**58 FPS**	**80-94%**	60–70%	GPU accelerates inference; CPU handles decode smoothly.
2	**22 FPS**	**48 FPS**	**90-96%**	70–80%	Decode and inference well balanced; slight latency at peak load.
3	**12 FPS**	**34 FPS**	**90-98%**	85–92%	GPU nearing saturation; frame rate stable but lower; decoding slightly lags.

10. Notes (Bottlenecks) Explanation

- **At 1 stream**: CPU handles load well; GPU runs efficiently with minimal utilization.
- **At 2 streams**: On CPU, classification becomes a limiting factor; decoding overhead grows. GPU maintains smooth performance.
- **At 3 streams**: CPU nears saturation (>95%), causing FPS drop. On GPU, moderate load observed but performance remains stable.

11. Conclusion

The project implemented a real-time AI video analytics pipeline using **Intel® DL Streamer** on both CPU and Intel Integrated GPU. The best-performing model (face-detection-0205 + emotions-recognition-retail-0003) achieved ~35 FPS with 3 streams on CPU, and up to 58 FPS with 4 streams on the GPU. GPU offloading improved overall throughput and reduced CPU bottlenecks, enabling smoother multi-stream performance. These results demonstrate the viability of deploying efficient AI inference pipelines on edge devices for smart surveillance applications.

PVB Adithya [CSE - AIML / GITAM UNIVERSITY] [vprata2@gitam.in]

Alekhya Madiraju [CSE - AIML / GITAM UNIVERSITY] [amadiraj2@gitam.in]