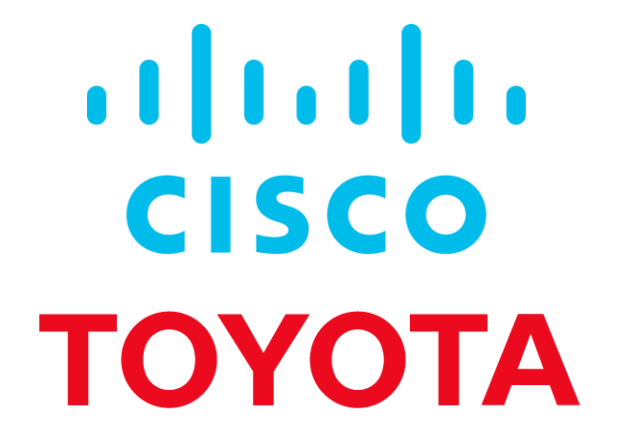


VLM-powered Video Filtering at the Edge: Scalable Data Pipelines for Data First Initiative

Cisco, Toyota



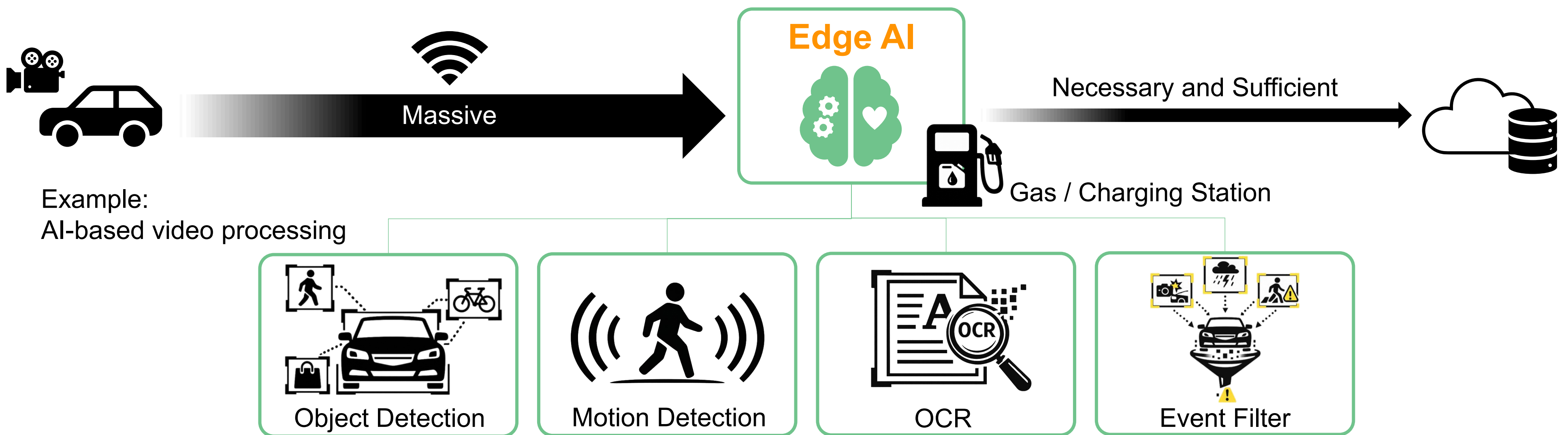
Problem & Goal

Goal: Enable scalable **AI-based video filtering** at the Edge Data Center

Connected cars generate massive data. However, only a small fraction contains useful content. Uploading raw data to the cloud causes bandwidth and storage challenges.

Concept

Edge preprocessing filters vehicular big data, sending only what is needed to cloud



PoC: VLM-powered Video Filtering

PoC Scenario:

For AI model training, a software development team wants to collect dashcam footage of specific scenes



Example prompts for Edge AI:

- Is a pedestrian crossing?
- Is an emergency vehicle visible?

AI extracts relevant frames from video streams.



Evaluation

Dataset:

- 22 scenarios
- 8 positive cases
- 14 negative cases

Metrics:

- Accuracy
- Response time
- Throughput

Platform

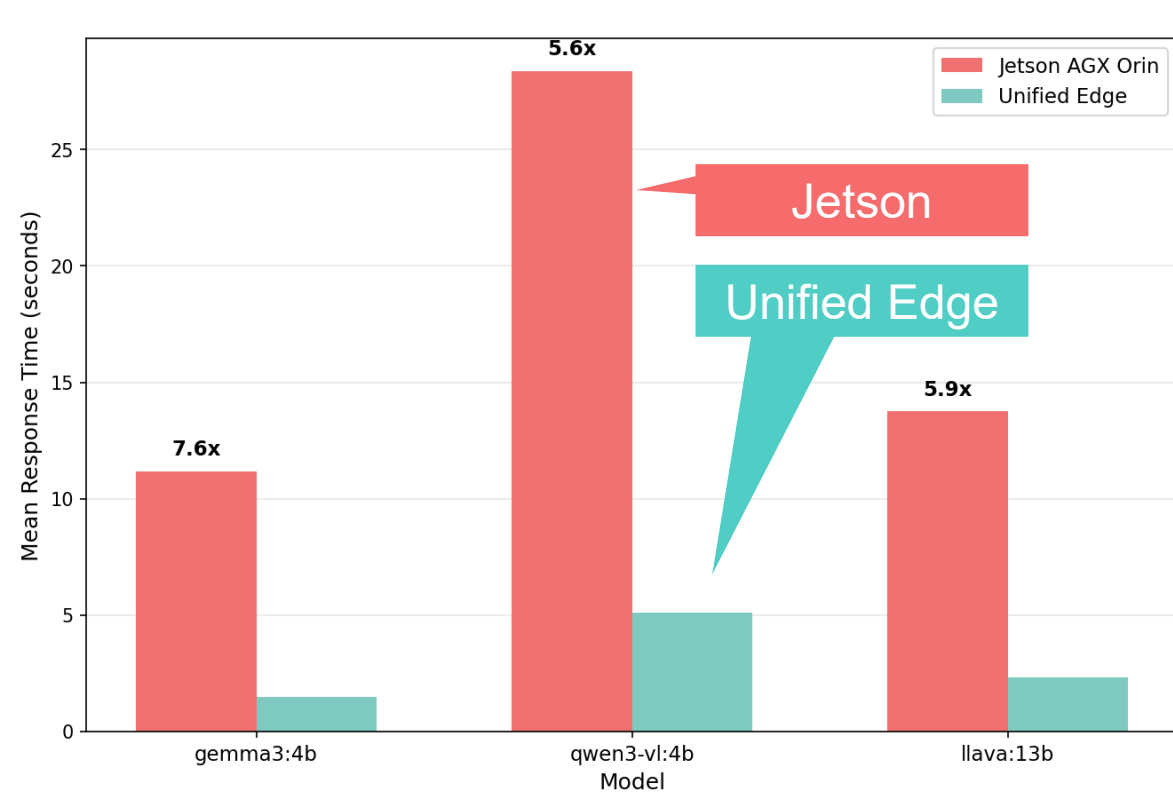
Cisco Unified Edge:

- Intel Xeon SoC
- NVIDIA L4 GPU
- 128GB Memory
- Managed by Cisco Intersight

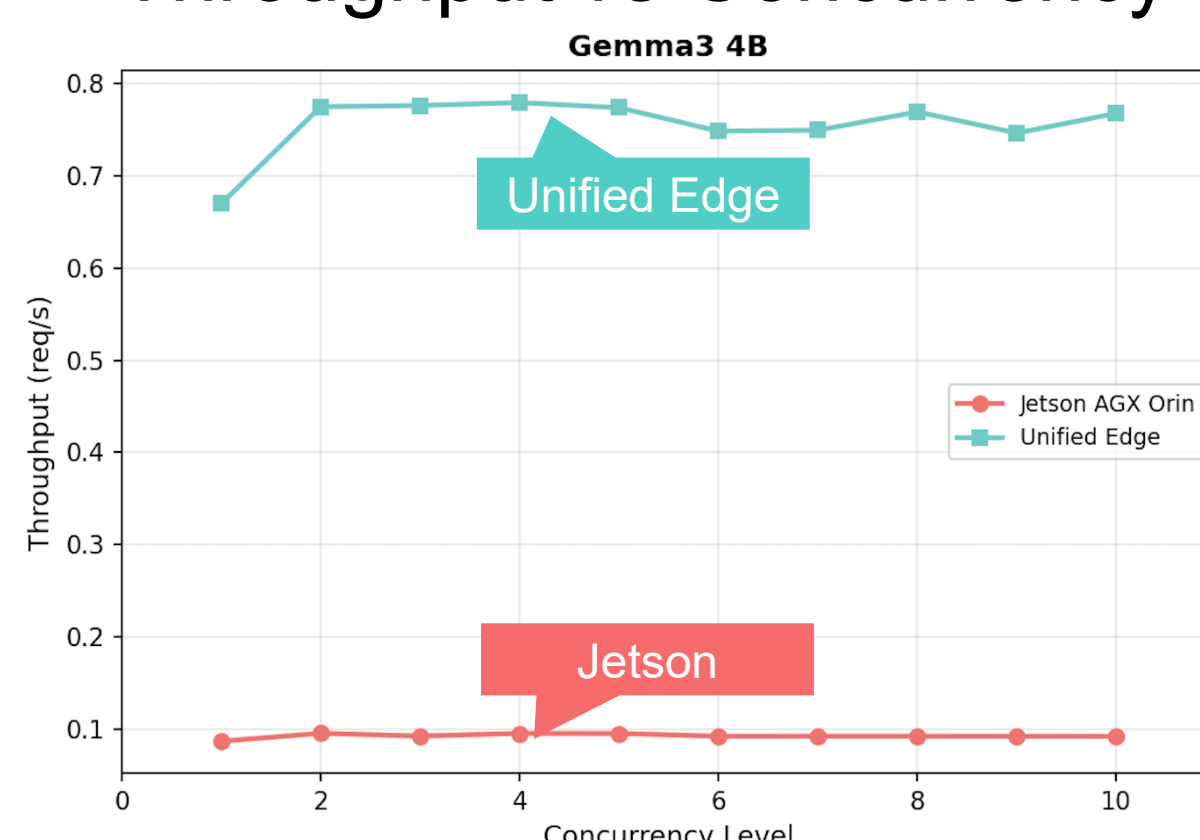


Key Results

Response time



Throughput vs Concurrency



Key findings:

- **6~8x faster** than Jetson AGX Orin
- Real-time inference latency
- Multi-stream processing capability
- Edge DC can process multiple vehicle workloads simultaneously.