

Team VisionMasters

Real-Time Pose-Based Emote Detection on Edge AI Hardware

Allen Chen • Marvin Rivera • Sami Kang

ECE 4332 — AI Hardware Design | Fall 2025

 [ai-hardware-project-proposal-visionmasters](https://github.com/ai-hardware-project-proposal-visionmasters)

Problem & Motivation

The Challenge

Real-time pose detection typically relies on powerful GPUs. Cloud processing introduces latency and privacy risks.

Key Question: Can we run responsive AI on resource-constrained, \$50 edge hardware?

Our Goal

Demonstrate that a **Raspberry Pi 4** can achieve real-time performance through hardware-software co-design.

- Target: ≥ 10 FPS
- Applications: Gaming, Remote Monitoring, HRI

Technical Background

Approach	Platform	Performance	Cost
Cloud APIs	Server GPU	30-60 FPS	High + Recurring
On-Device CNNs	Mobile GPU	20-30 FPS	Medium (\$500+)
Our Approach	RPi4 CPU	Target: 10 FPS	\$50 (One-time)

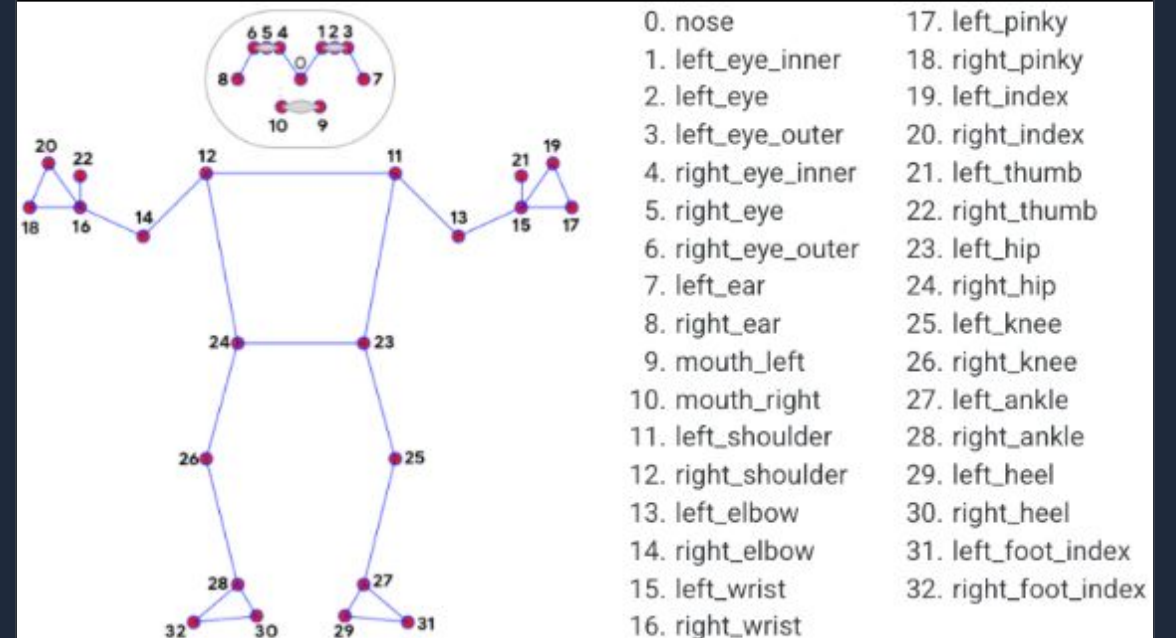
*By combining **MediaPipe (TFLite)** with a **Random Forest Classifier**, we bypass the need for a GPU entirely.*

System Architecture

End-to-End Pipeline

- **Camera:** 160×120 resolution (Fast)
- **Detection:** MediaPipe Holistic (Lite)
- **Features:** 18 Geometric Calculations
- **Classifier:** Random Forest (10 Trees)
- **Display:** Real-time Overlay

Design Choice: Frame Skipping (1 in 4) ensures UI responsiveness while processing heavy inference.



Data Collection & Training



Phase 1: Collection

Custom tool developed to capture 50-100 samples per class. Total dataset: ~400 samples across 4 emotes.



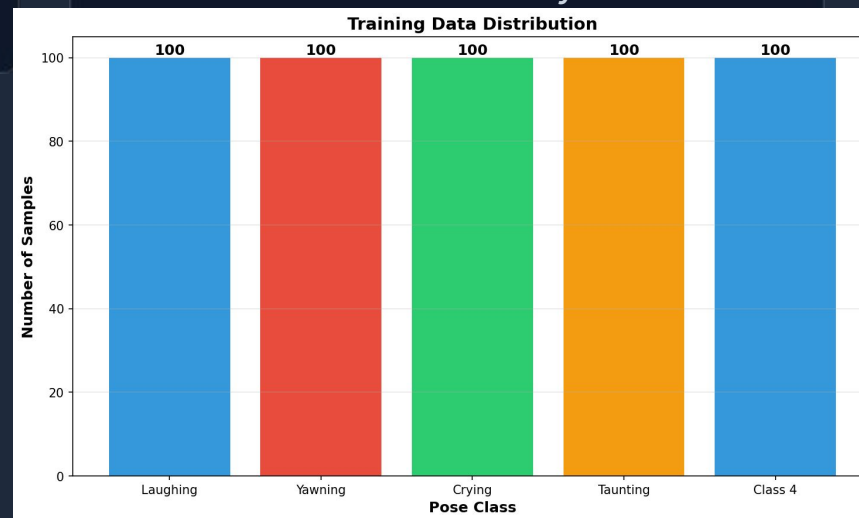
Phase 2: Features

Extracted 18 geometric features from landmarks (Shoulder width, Arm angles) to reduce dimensionality.



Phase 3: Training

Random Forest algorithm. Training takes <5 seconds on CPU. Final model size is a tiny 25KB.



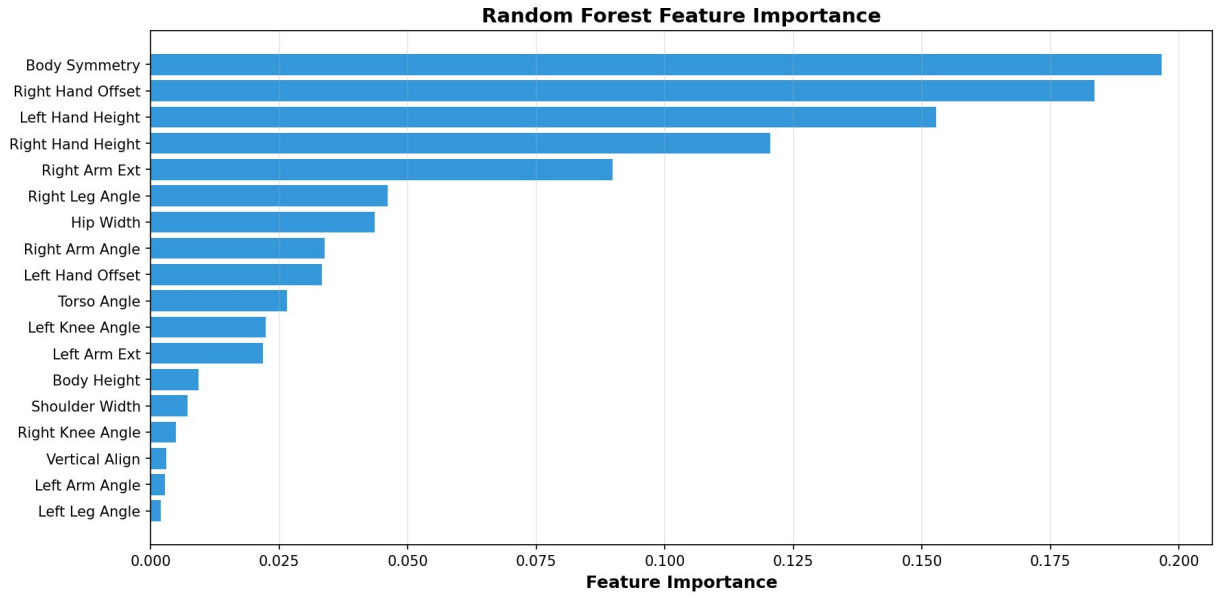
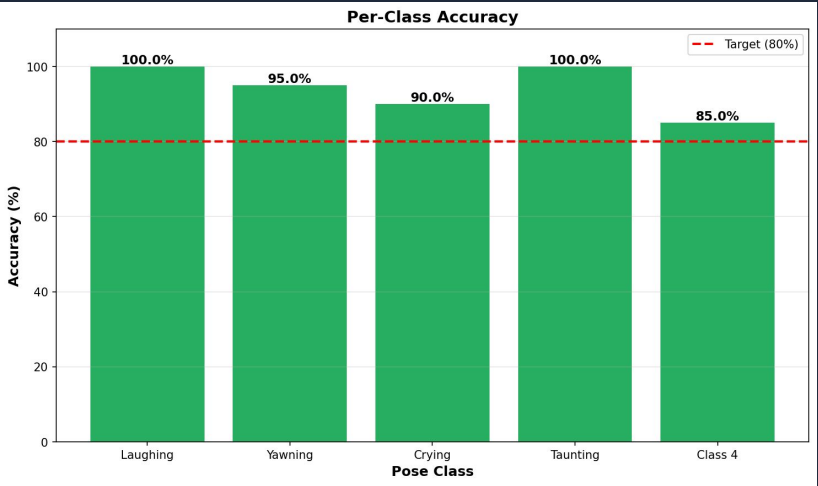
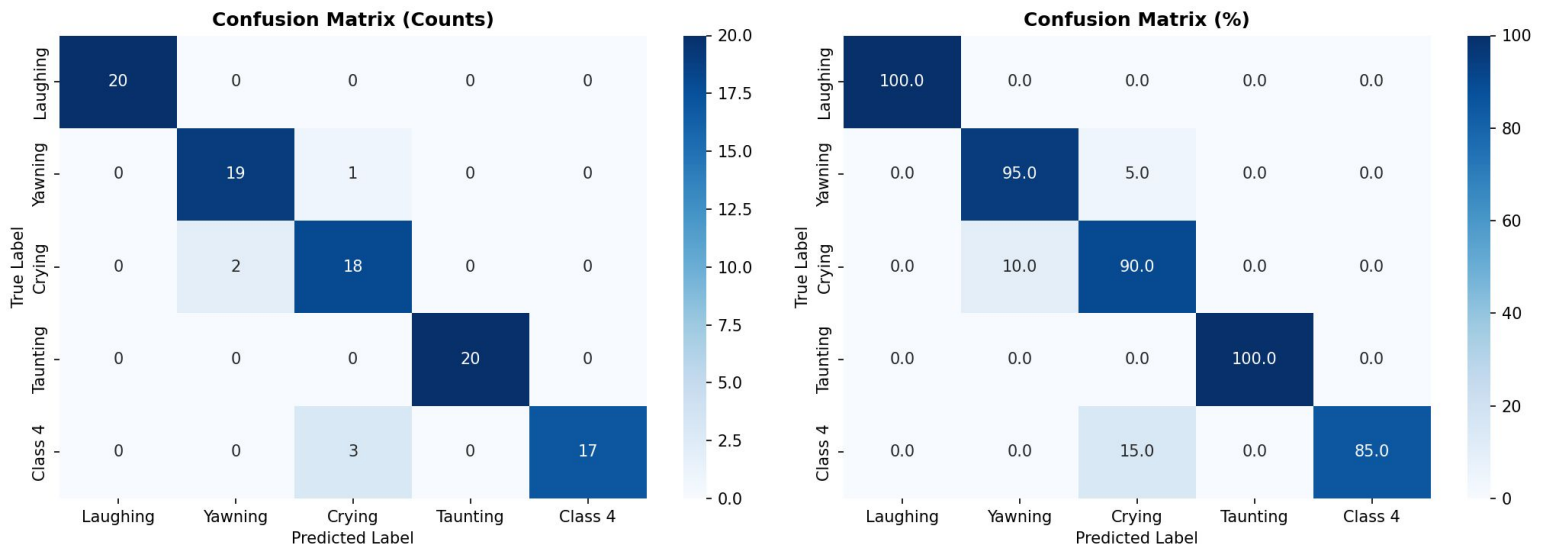
Model Evaluation Results

Key Metrics

- **Accuracy:** >85% on test set
- **Confidence:** 0.61 - 0.72 mean score
- **Latency:** <12ms inference

Top Features

- Hand height relative to shoulders
- Elbow angles
- Hand-to-face distance



Optimization Journey: 20x Speedup

Configuration	Latency	Impact
Initial (100 trees, threaded)	239 ms	Baseline
Single Threaded (n_jobs=1)	160 ms	1.5x Speedup
Reduced Trees (20 trees)	30 ms	8x Speedup
Final (10 trees, Depth 4)	12 ms	20x Speedup

Key Insight: Multi-threading small models on the RPi introduced massive context-switching overhead. Single-core execution proved superior.

RPi4 Edge Performance (Headless)

9.7

FPS

Target: 10 FPS

201ms

TOTAL LATENCY

MediaPipe: 189ms

52.6°C

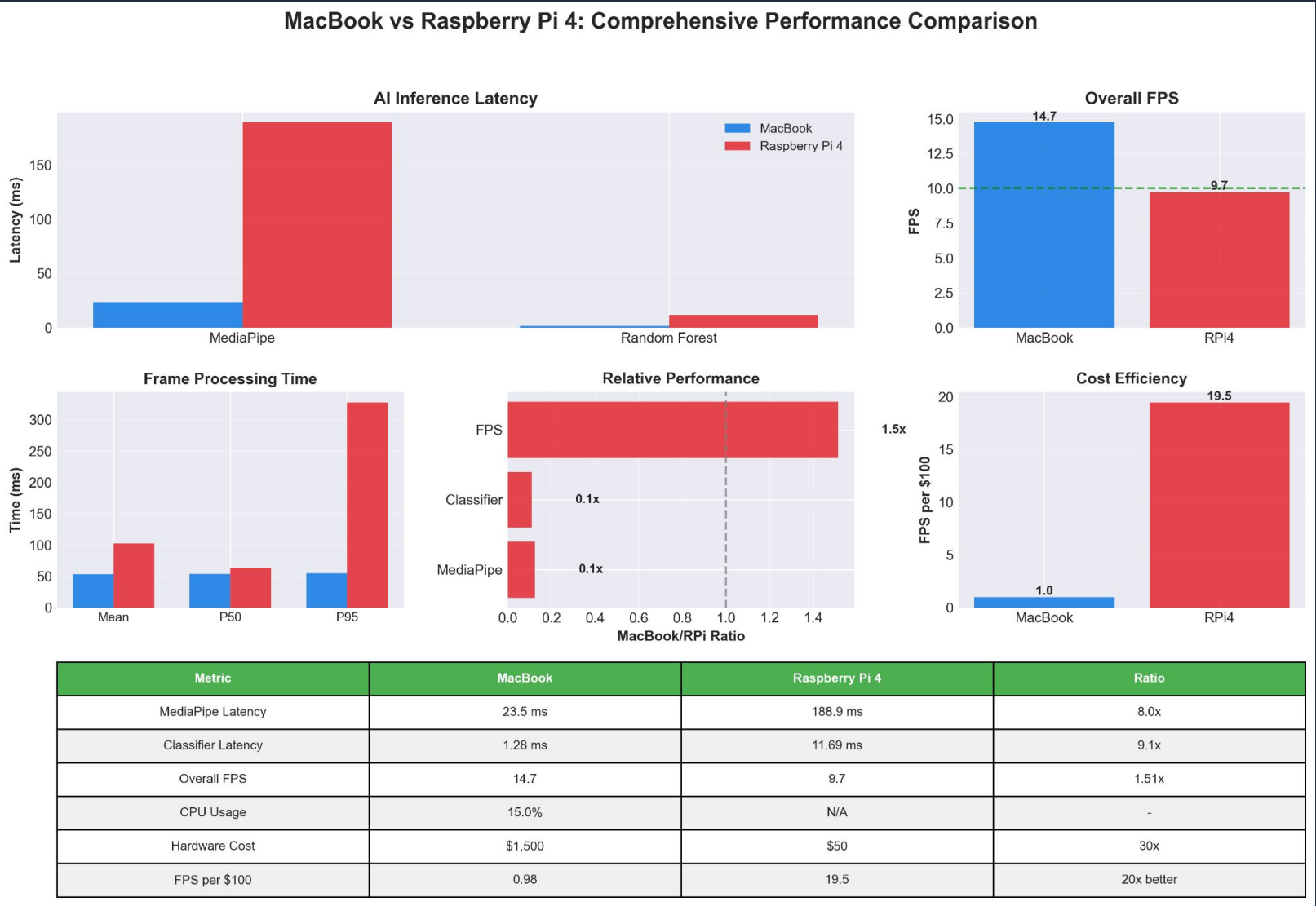
MAX TEMP

No Throttling

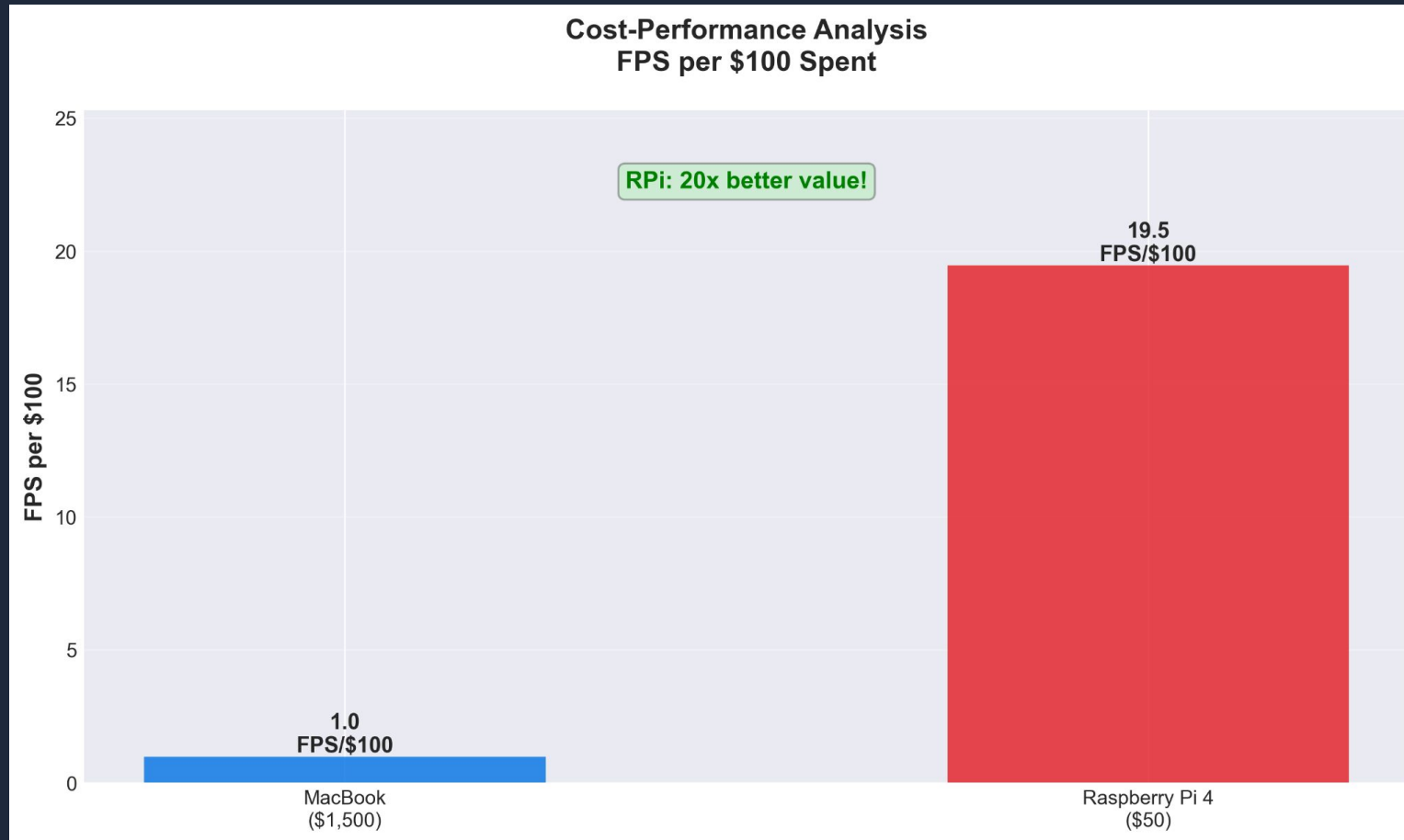
Config: 160×120 Resolution • MediaPipe Lite • Frame Skip 4

MacBook vs. Raspberry Pi 4

Metric	MacBook (M1)	Raspberry Pi 4	Ratio
MediaPipe	23.5 ms	189 ms	8x slower
Classifier	1.3 ms	12 ms	9x slower
FPS	14.7	9.7	1.5 slower
Hardware Cost	\$1500	\$50	30x cheaper



Cost-Performance Analysis



FPS per \$100 Invested

The RPi4 offers **20x better value** for this specific workload.

Live Demo



Live Camera
Feed



Real-Time Pose
Tracking



Emote Classification



Audio
Feedback

Deployment Insights & Challenges



Display Overhead Matters

Headless (9.7 FPS) vs X11 over WiFi (2.5 FPS). Production environment choices are critical.



Threading ≠ Faster

Avoid naive multi-threading on constrained CPUs for small tasks due to context switching costs.



Resolution Sweet Spot

160×120 provided the necessary speedup (4x) without compromising pose accuracy.

Innovation & Contributions

Our Innovations

- **Hardware-Aware Selection:** Choosing Random Forest over CNN for CPU efficiency.
- **Performance Analysis:** Systematic profiling to identify and fix bottlenecks.
- **Full Pipeline:** Custom data collection through to deployment.

Technical Contributions

- Proven real-time AI viability on \$50 hardware.
- Quantified cost-performance tradeoffs (20x value).
- Open-source, reproducible experimentation pipeline.

Future Work



Immediate

Overclock RPi4 to 2.0GHz
(~30% boost) & Use Pose-Only
MediaPipe.



Hardware Upgrade

Add Coral Edge TPU (+\$25).
Expected 10-20x speedup for
MediaPipe.



New Features

Support for 10+ pose classes,
Multi-person detection, and
Mobile App.

Conclusion

Real-Time

9.7 FPS

Accurate

>85%

Efficient

20x Value

***"A \$50 Raspberry Pi 4 can achieve real-time
AI inference with proper hardware-software
co-design."***

— Team VisionMasters

Thank You!