

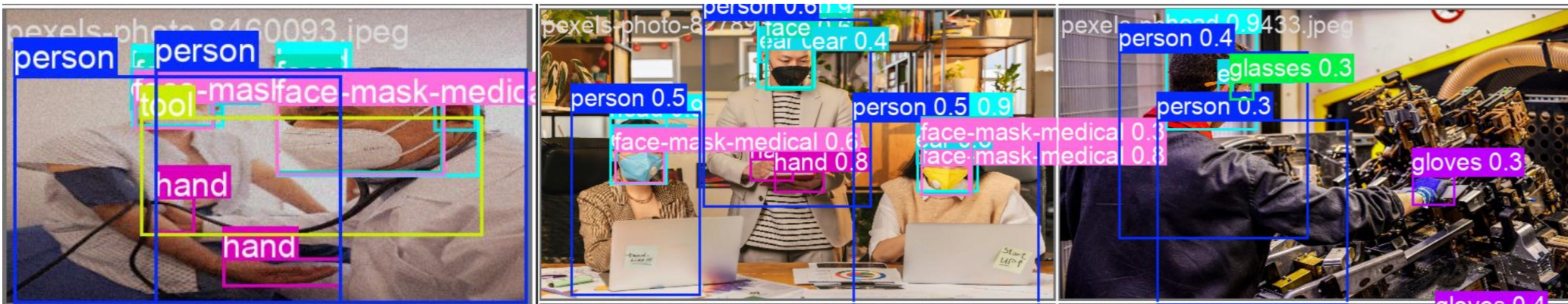
Cost Optimized Deployment of Computer Vision Model for PPE Compliance in Construction

By: Steven Harta Prawira; Academic Supervisor: Asst. Prof. Ian Zhu.

Project Background

Motivation

- Construction industry has highest # of workplace fatalities in the U.S. (1075 deaths in 2023 – U.S. Bureau of Labor Statistics or BLS).
- Construction sector contributes 30% of fatal and major workplace injuries in Singapore (166 cases in 2024 – Singapore MOM).
- 62% workers exposed to hazardous heights, only 31% uses personal protective equipment (PPE) effectively. (U.S. BLS, 2021).
- Increasing research to prove technical feasibility of AI-powered PPE monitoring, but **economic viability remains key**.



Objective

“Beyond feasibility: optimizing PPE detection systems for deployment through cost–performance tradeoff analysis.”

Evaluating practical tradeoffs to move beyond proof-of-concept into real world deployment from lab to construction site.

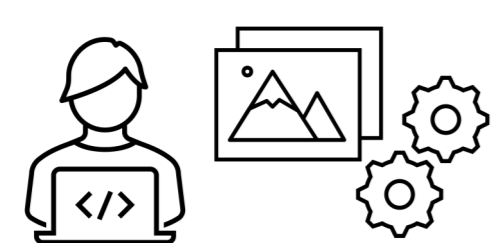
Key Deliverable

A practical framework for selecting the **optimal hardware setup**, balancing **cost, accuracy, and latency** for real-world PPE detection system deployment.



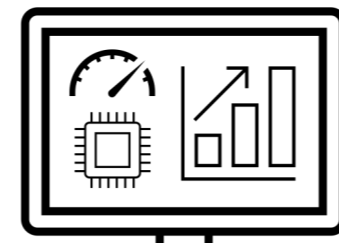
Source: <https://www.avidbeam.com/>

Methodology: A Phased Approach



Model Selection & Image Preprocessing

Object Detection Model: YOLOv8s
 Open-source, YOLO-compatible image dataset
 Noise Augmentation and Brightness adjustments
 Resized (480p, 720p, 1080p, 1440p) with Python PIL library



Model Benchmarking & Data Collection

Key benchmarks: accuracy, inference latency, training duration and cost
 Deployment cost integration
 Edge device simulation: RP4, Jetson Nano, Jetson Xavier NX
 Data collection on 4 image resolutions using 3 GPUs: T4, L4 and A100



Optimization & Sensitivity Analysis

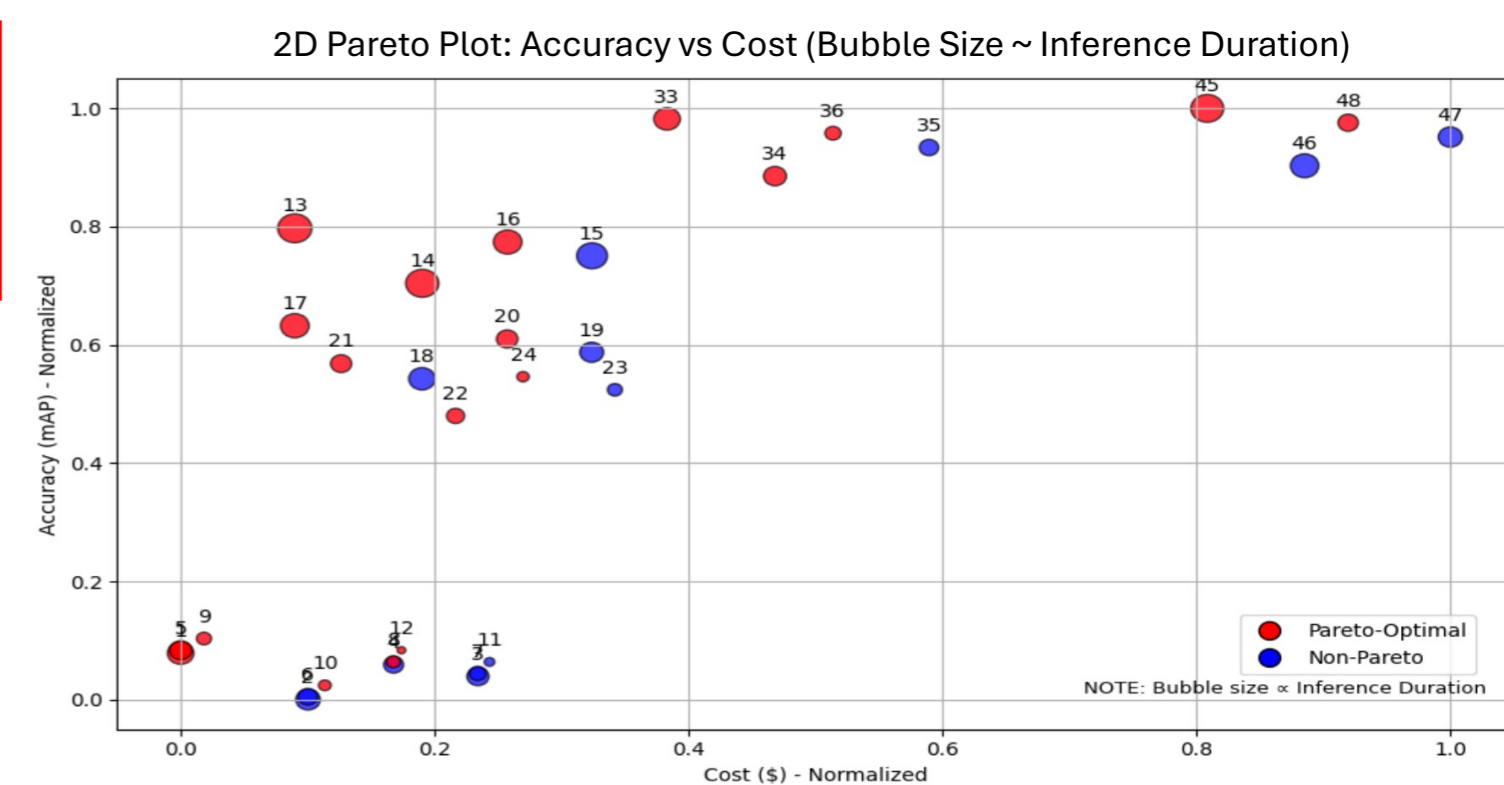
Utility Scores for: Accuracy-Driven, Cost-Constrained, Real-time Priority, Training Conscious and Balanced
 Robust configuration combining Utility Score and Pareto Efficiency
 Pareto Frontier Analysis

Key Results & Business Implication

Key Business Implication

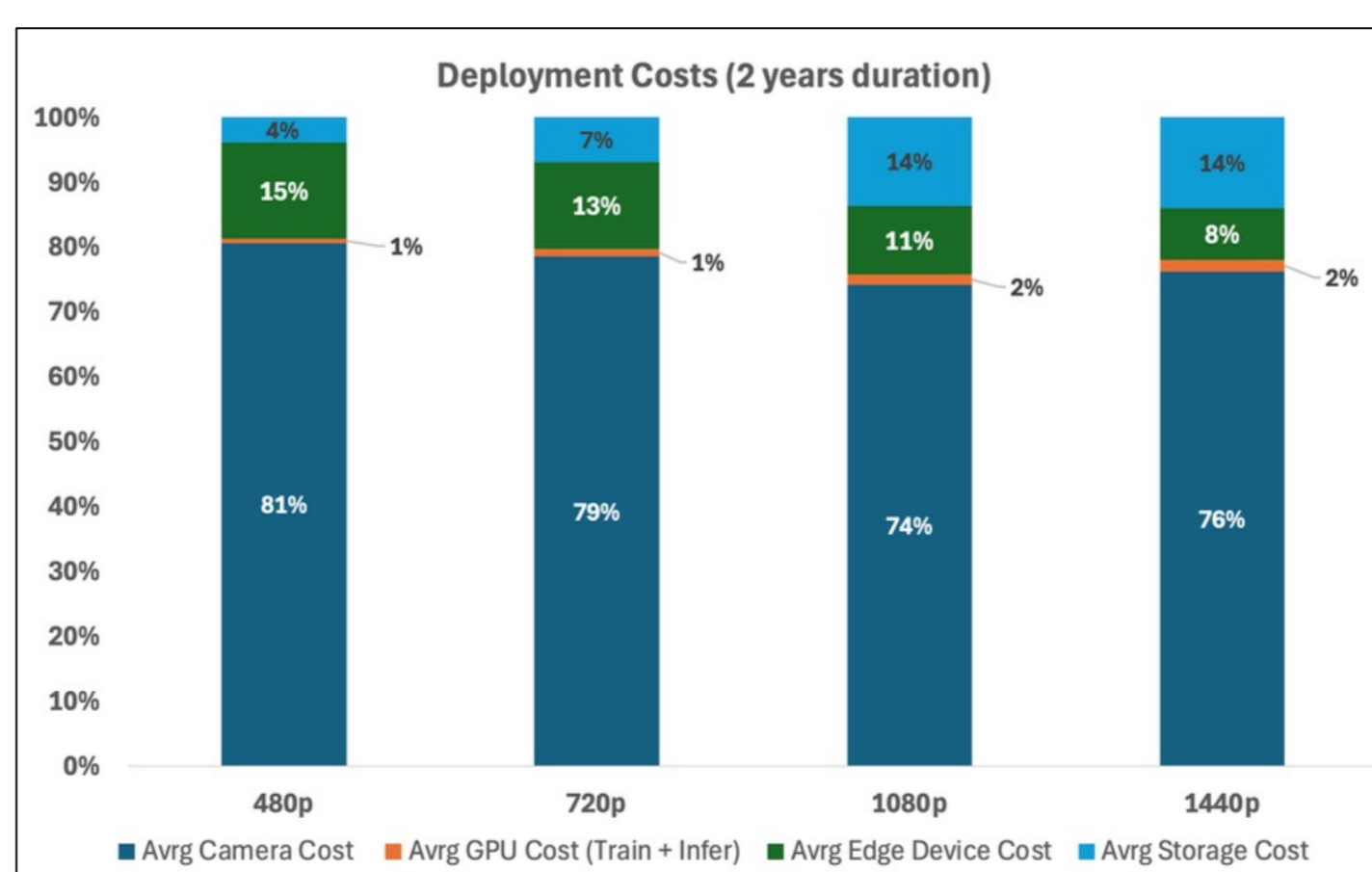
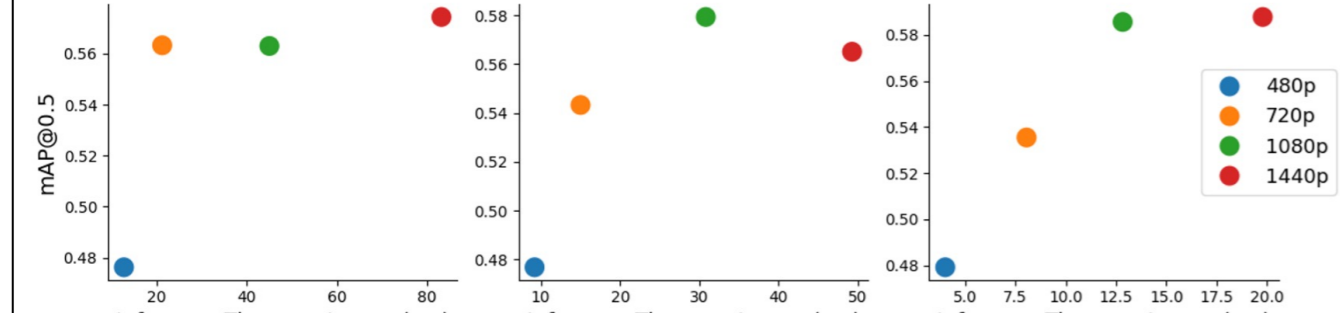
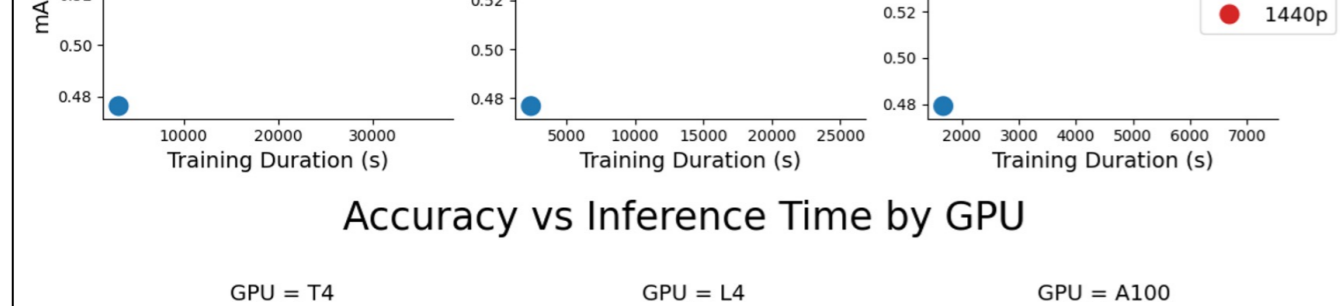
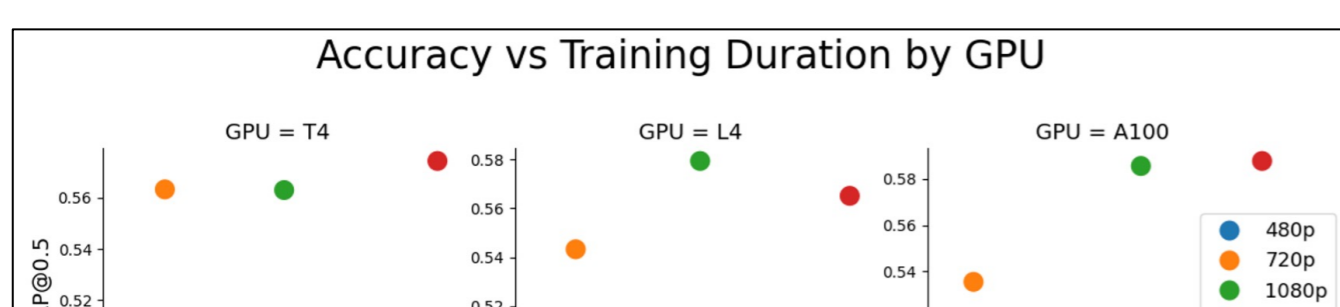
- Start with end in mind - beyond technical feasibility study, to identifying economical deployment.
- In key business stakeholders' mind: *“Are deploying a cost-effective system to achieve the intended outcome? Is this the right hardware choice for the n years deployment?”*

Tradeoff Analysis Outcome



SetupID	Resolution	GPU	Edge	Avg. Utility Rank	Pareto Optimal
24	720p	A100	Jetson Xavier NX	3.2	YES
21	720p	A100	NIL	3.7	YES
9	480p	A100	NIL	5.7	YES
22	720p	A100	RaspberryPi4	6.3	YES
12	480p	A100	Jetson Xavier NX	7.5	YES

Benchmark Performance



- High-res images** improve accuracy but up training time and slow down inference.
- Camera cost** the biggest component of total cost.
- Cloud Storage cost** increases with image resolution and project duration.
- Edge Device:** 1) improves inference time but penalizes accuracy; 2) ~10-15% of total cost.
- GPU** key differentiator to improve inference latency but only small component of total costs.

- Pareto Plot** suggests diminishing returns on Accuracy vs Cost at high resolution.
- Accuracy improvement plateaus at 1080p, while costs, training and inference time up significantly.
- Table lines up Top 5** robust configurations with high utility rank and pareto efficiency.
- Mid-range resolution (720p) with high performance GPU** shown as the 'sweet spot' and costs ~300k\$.
- Low-range resolution (480p) with high spec GPU as possibility for lower risk deployment option.