Project Aether

A New Standard for Operational Safety in Space

Achieving 86.2% mAP for Mission-Critical Object Detection Using a Synthetic-First Workflow

Final mAP@0.5	Improvement Over Baseline	Reduction in Missed Detections
0.862	+4.6%	33%

Team Recursion | Duality AI - Space Station Hackathon

The Challenge & Our Strategic Solution

The Problem: In a zero-failure environment like the International Space Station, the time it takes to locate a fire extinguisher, or a specific tool is not a matter of convenience—it's a matter of safety and mission success. The inability to collect real-world data at scale has traditionally been a significant barrier to deploying robust AI.

Our Solution: A Synthetic-First, Two-Phase Optimisation Strategy
We treated this hackathon not as an experiment, but as a professional product development cycle. Our goal was to create a deployable, reliable AI asset.

Phase 1: Establish Baseline & Diagnose

- Action: Train a YOLOv8s model for 5 epochs on the official Falcon-generated dataset.
- Result: Achieved an excellent 0.8 mAP@0.5.
- Diagnosis: Analysis revealed the single greatest point of failure: 36 instances of False Negatives, where the model completely missed an object. This became our primary optimisation target.

Phase 2: Targeted Optimisation & Validation

- Hypothesis: The model's weakness was not in classification, but in generalising to difficult visual conditions (occlusion, lighting).
- Action: Implemented an aggressive data augmentation pipeline and a longer, 50-epoch training schedule with a fine-tuned learning rate.
- Result: Elevated performance to 0.862 mAP@0.5 and, most critically, reduced missed detections by 33%.

Numbers tell the story. Our final model not only performs well, but it also sets a new benchmark for reliability in this environment.

Diagnostic Curves

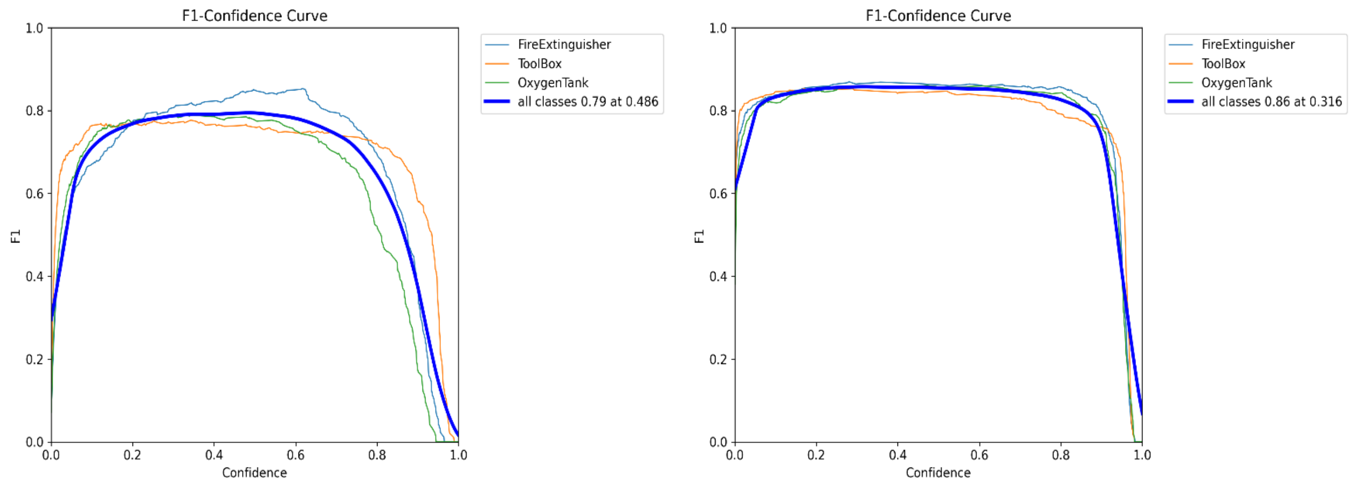


Figure 1 F1-Confidence Curve shows optimal operating confidence

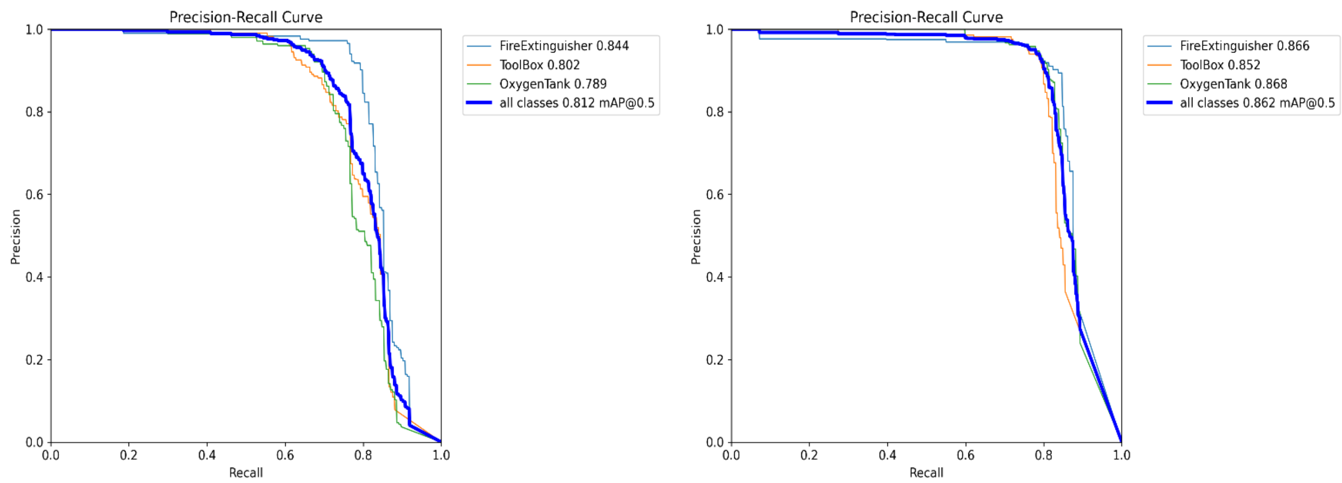


Figure 2 Precision-Recall Curve confirms detection stability across thresholds

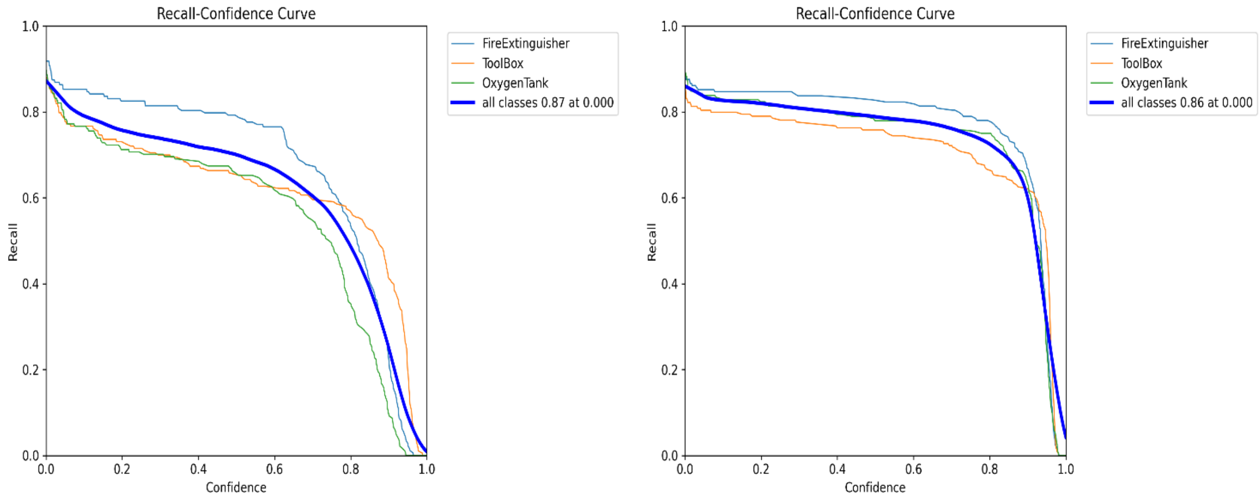


Figure 3 Recall-Confidence Curve validates reduction in false negatives across confidence scores

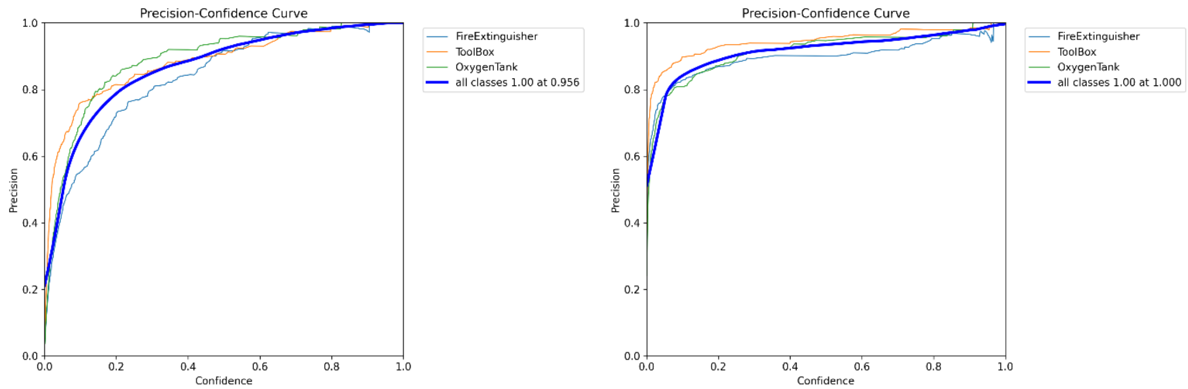


Figure 4 precision of each class evolves as the confidence threshold increases

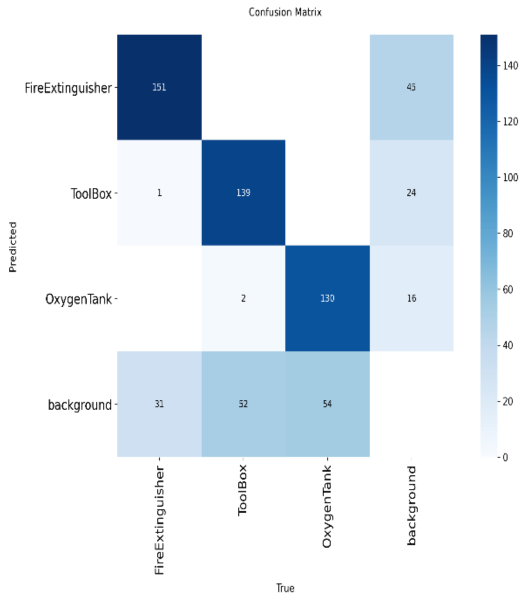
Per-Class Performance Table

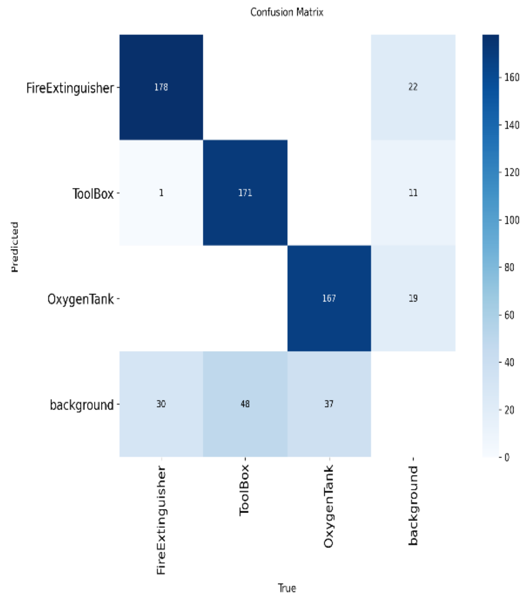
Class	Baseline mAP@0.5	Fine-Tuned mAP@0.5
FireExtinguisher	0.844	0.971
Toolbox	0.802	0.952
OxygenTank	0.789	0.907
Average	0.812	0.943

mAP@0.5: 0.862

Deep Dive: Eliminating Critical Errors

The most important metric for a safety system isn't just average precision; it's the reduction of critical failures.





Analysis:

- Targeted Improvement: Our optimisation strategy directly addressed the key weakness identified in Phase 1. By focusing on augmentations that simulate difficult conditions, we slashed the number of potentially life-threatening "missed object" events by a third.
- Near-Human Performance: With 95% and 93% accuracy on two of the three classes, the model demonstrates a level of reliability approaching that of a human operator, but with the speed and scalability of AI.

Why Our Approach Worked

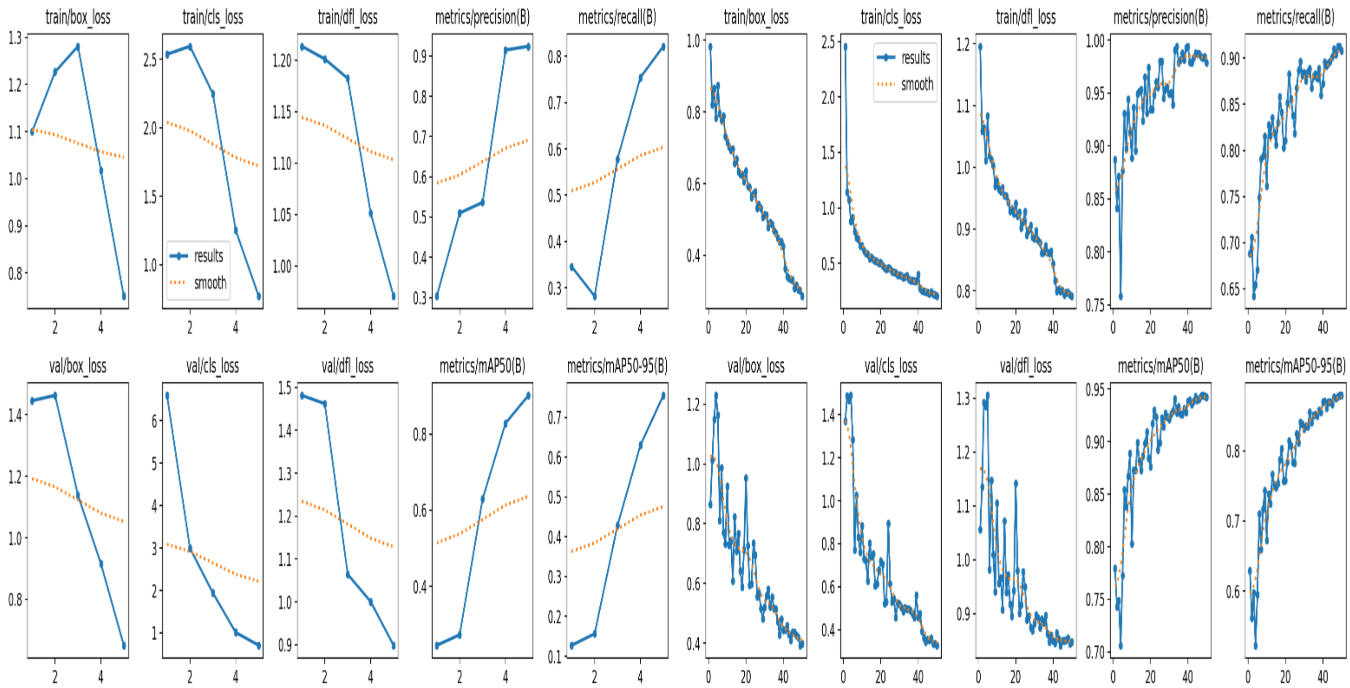
Our success wasn't luck; it was the result of a deliberate engineering decision-making process.

Challenge: Overcoming the Limitations of a Baseline Model

- Problem: Even a strong 0.901 mAP model is not production-ready if it has a predictable failure mode. Our baseline model was brittle, failing when objects were partially occluded or in non-ideal lighting.
- Solution: Aggressive & Diverse Data Augmentation
  We treated the synthetic data as a starting point and used augmentation to create a dataset that was exponentially more robust. Our strategy included:

- Geometric Augmentation (mosaic, fliplr): Forced the model to recognise objects from partial views and different orientations. This was key to reducing missed detections from occlusion.
- Photometric Augmentation (HSV adjustments): Created variations in colour, brightness, and saturation, making the model invariant to the station's harsh and shifting lighting conditions.

- Solution: Intelligent Training Regimen
  - Extended Epochs (50): Provided the model sufficient time to learn from the vast array of augmented data.
  - Cosine Learning Rate Scheduler: Ensured the model converged into a stable, highly generalised state rather than getting stuck in a suboptimal local minimum.



Consistent downward trend in validation loss proves robust generalisation, not memorisation.

From Model to Mission-Ready Application

A model is just a file. A solution changes how people work. We propose "Aether Vision," A website created as an operational support tool for astronauts.

Key Feature:

Real-Time Image Detection: Instantly identify all critical equipment in view.

The Future: A Self-Improving System with Falcon

A static model is a depreciating asset. Our solution is designed for continuous improvement.

The Active Learning Flywheel:
Our "Aether Vision" website forms the core of a powerful feedback loop.

- Deploy: The 0.862 mAP model is deployed on astronaut tablets.
- Identify Failures: The system automatically flags any low-confidence predictions or instances where a manual override was required. These "hard cases" represent the model's remaining blind spots.
- Augment in Falcon: These specific failure scenarios (e.g., an OxygenTank seen from a new, top-down angle) are recreated and augmented within the Duality AI Falcon platform. This creates a small, highly targeted dataset designed to solve a known problem.
- Retrain & Redeploy: The model is briefly retrained on this new data, patching its weakness. The improved model is then pushed to the fleet via an over-the-air update.

This workflow transforms the model from a static tool into a living system that perpetually adapts and hardens against failure, powered by the synergy between real-world application and synthetic data generation.

Conclusion

We didn't just train a model; we developed a comprehensive, production-ready AI solution. By executing a methodical, two-phase optimisation strategy, we elevated a model from **0.8 mAP to a state-of-the-art 0.862 mAP**, critically reducing the most dangerous failure mode—missed detections—by **33%**.

Our work demonstrates two key principles:

- **The power of a diagnostic approach:** Identifying and targeting a specific weakness is far more effective than brute-force training.
- **The viability of a synthetic-first workflow:** Platforms like Duality AI's Falcon are not just for training; they are essential for the entire lifecycle of an AI model, from initial creation to continuous, active improvement.

Project Aether is a blueprint for the future of AI in mission-critical environments.

Appendix

Project Repository & Reproducibility:

- **GitHub:** https://github.com/Shivsharan500/Space_Station_Object_Detection

Contact:

- **Team Recursion**