EXPLORE INNOVATE EXCEL

MULTI-SOURCE AERIAL AND GROUND OBJECT DETECTION SYSTEM USING IR IMAGES

Minor presentation (25261)

Sajal Goel 23106031
Supriya Raj 23106064
Parth Tandon 23106065
Gurnoor Kaur 23106061

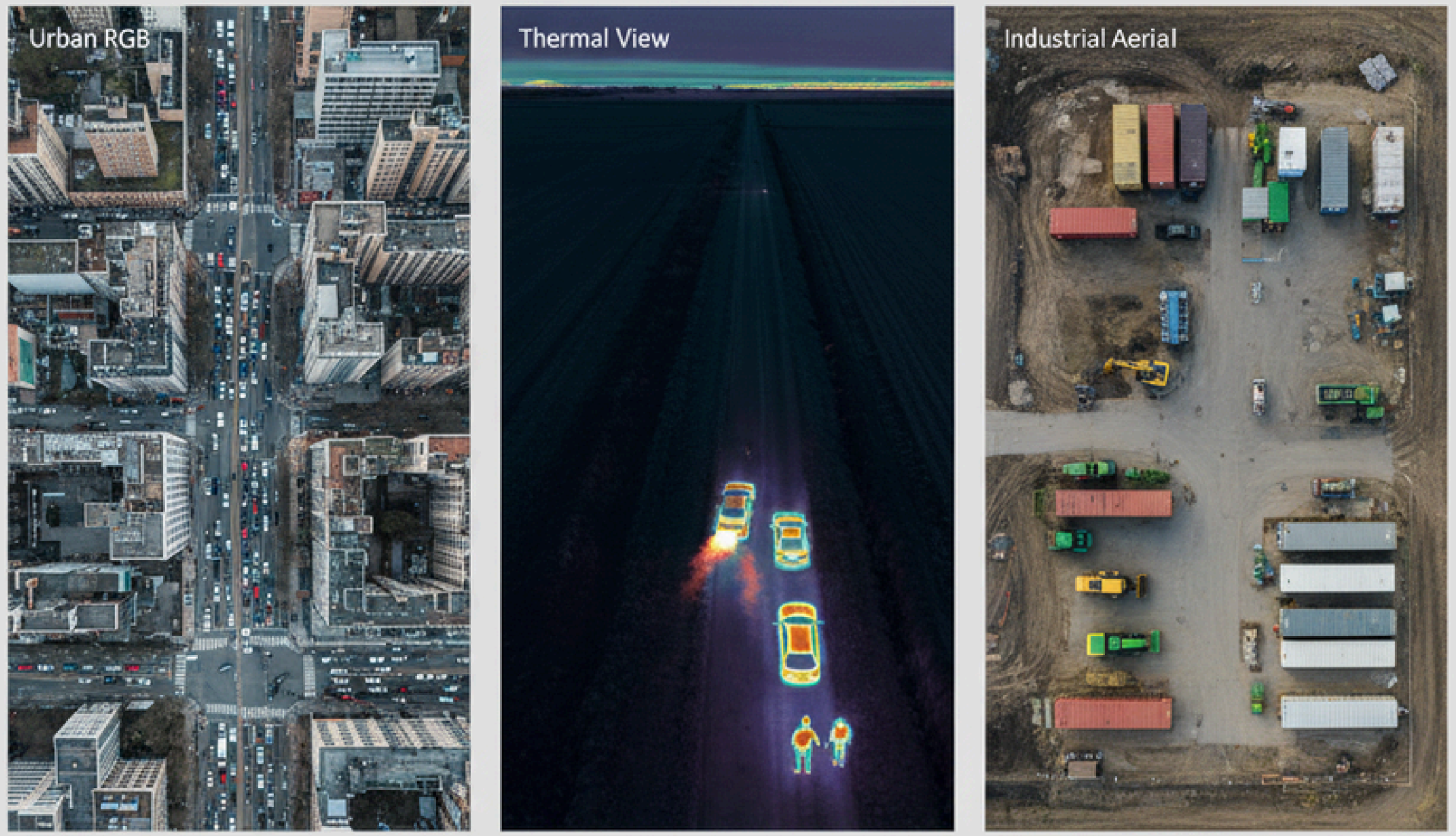
CSE(Data Science)

MENTORS

Dr. Alka Jindal (Assistant Professor)

Dr. Padmavati (Associate Professor)

INDEX



Dataset Diversity for Robust Object Detection: Fusing Diverse Modalities & Environments

PROBLEM STATEMENT

- Infrared (IR) object detection is crucial for UAV-based night surveillance and defense operations, where RGB models fail due to low contrast and texture loss in thermal imagery.

- However, most existing datasets and models are RGB-focused, leaving IR detection underexplored — especially for real-time, lightweight UAV applications.

Our work aims to bridge this gap by developing a dedicated IR-based detection pipeline that ensures:

- Accurate recognition in low-visibility and night conditions.
- Efficient inference on resource-limited UAV systems.
- This research addresses a critical and underdeveloped area in computer vision, contributing toward smarter, defense-grade surveillance intelligence.

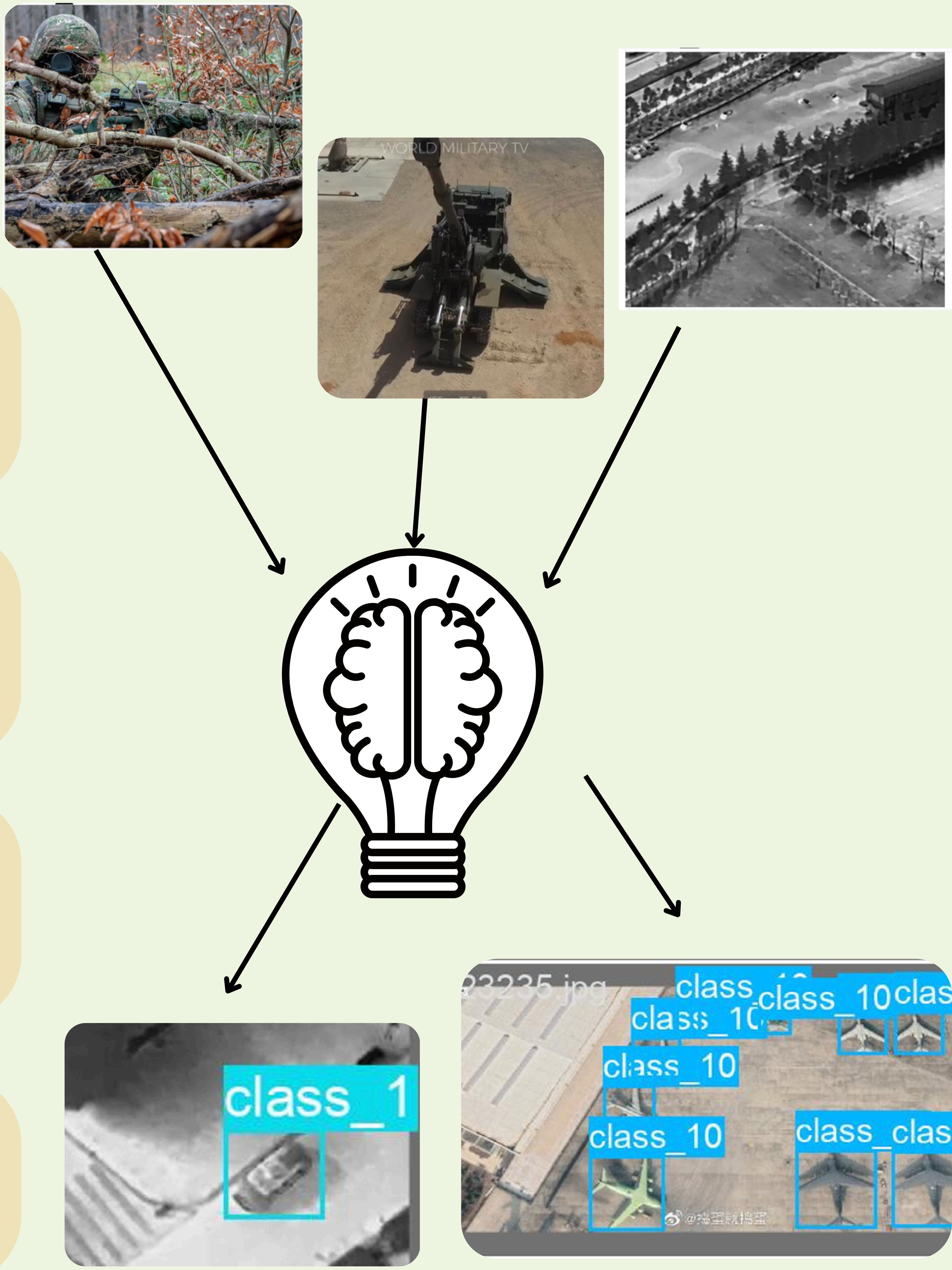
Objectives:

Develop a Robust Object Detector: Create a deep learning model capable of accurately identifying diverse objects in complex aerial and multi-modal imagery

Achieve High Generalization: Ensure effective performance across various environments, scales, and sensor types by integrating multiple heterogeneous datasets

Optimize for Real-Time Performance: Utilize an efficient architecture (YOLOv8s) suitable for deployment on edge devices, enabling rapid inference for dynamic applications

Demonstrate Versatility: Showcase the model's adaptability and utility across different scenarios and data characteristics, anticipating future applications (e.g., IR/Thermal)



LITERATURE OVERVIEW

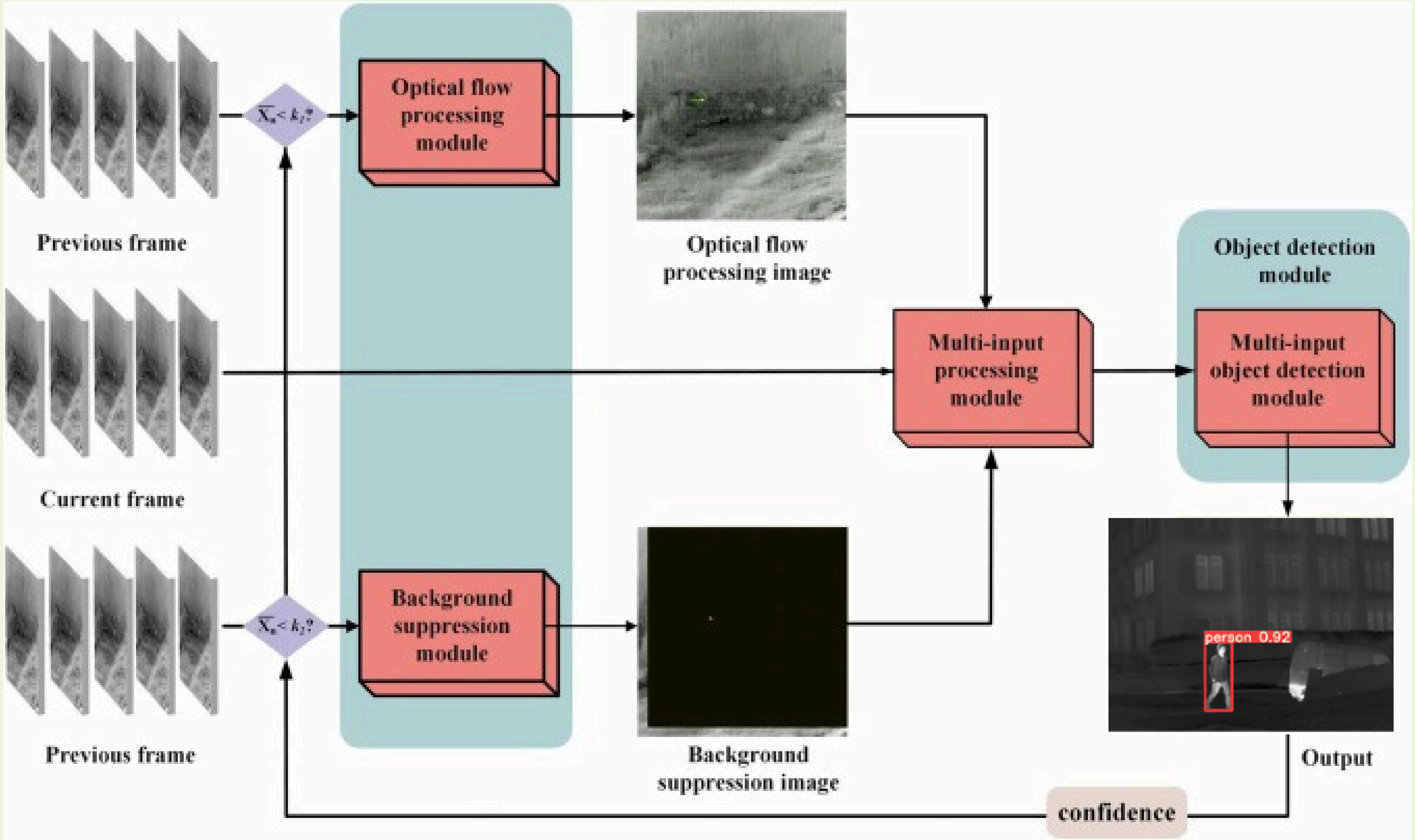
Publication Title	Year	Datasets Used	Size (Total Images)	Model	Preprocessing	Key Results	Classes
YOLO-UIR: A Lightweight and Accurate Infrared Object Detection Network Using UAV Platforms	2025	1. HIT-UAV 2. Drone Vehicle 3. LLVIP	1. **HIT-UAV**: 2,898 infrared images 2. **Drone Vehicle**: 28,439 infrared images 3. **LLVIP**: 15,488 pairs of visible and infrared images (infrared used for experiments)	YOLO-UIR with three variants: YOLO-UIR-n, YOLO-UIR-s, YOLO-UIR-m	**Dataset-specific:** - HIT-UAV: "Other Vehicle" class merged into "Car"; "DontCare" samples removed - Drone Vehicle: Images resized to 640×512 **General Data Augmentation:** - Random blurring, mirroring, flipping	- **Drone Vehicle**: mAP 71.1% (YOLO-UIR-n) - **HIT-UAV**: mAP 90.7% (YOLO-UIR-n) - **LLVIP**: mAP 94.3% (YOLO-UIR-n) - **Efficiency**: 3.0G FLOPs, 47 FPS inference speed	- **HIT-UAV**: Person, Bicycle, Car - **Drone Vehicle**: Car, Truck, Bus, Van, Freight Car - **LLVIP**: Pedestrians
Drones in Defense: Real-Time Vision-Based Military Target Surveillance and Tracking	2025	KIIT-MiTA	1,700 images / 130 MB	VGG16 Faster R-CNN	Ensemble Detection Scheme (WBF): Images aggressively resized to 224×224 pixels. Uses a Weighted Boxes Fusion (WBF) ensemble technique to combine multiple base detectors	$mAP_{0.5}$: 77.02%, $mAP_{0.5:0.95}$: 29.73%. FPS: Low (base detection systems like VGG16 Faster R-CNN achieved 5 FPS on GPU).	Classes(5-8) Person Vehicle Bicycle Motorcycle Building Road Tree
Examining the Impact of Image Augmentation on Object Detection and Classification to Improve Performance of Situational Awareness Systems	2025	Military assets dataset	26,315 labelled images	YOLOv8 (n, s, m variants)	Image augmentations and image compression	- Pre-trained YOLOv8n: 68.2% mAP50 (vs 52.4% retrained) improvement - Friend-foe accuracy dropped from 84.6% → 79.1% (augmentation)	distributed across 12 classes of objects.

OUR PROPOSED MODEL

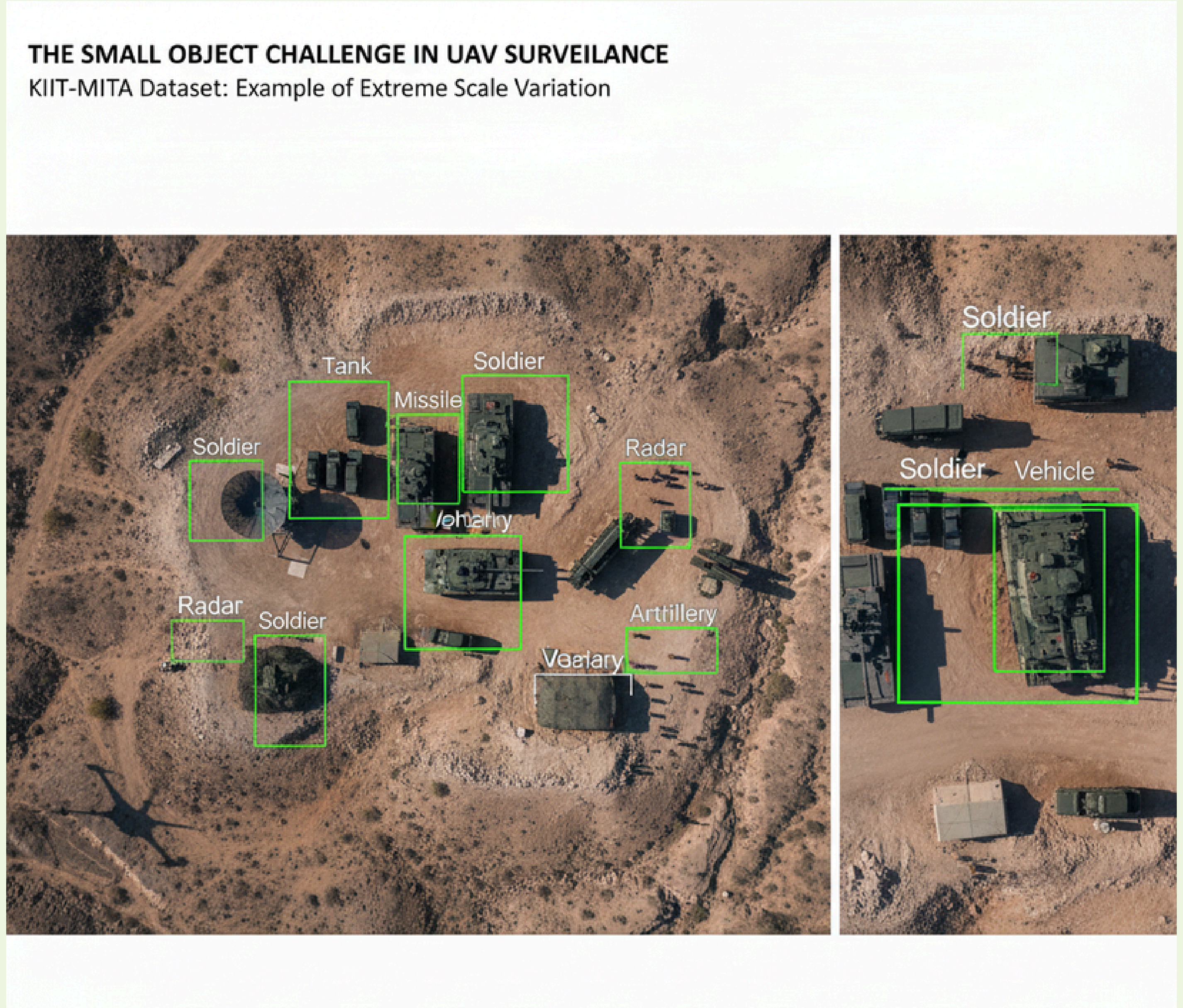
Multi-Source Data Fusion: Combines five diverse aerial datasets across scenes, objects, and sensor types (thermal, large-scale), ensuring robust generalization — a novel approach.

Custom Multi-Phase Sequential Learning: Our innovative 3-phase training plan (Foundation → Specialization → Refinement) systematically builds generalized knowledge and then fine-tunes for specific challenging aspects (e.g., small objects, specific modalities), leading to enhanced robustness

Cross-Domain Generalization: Trained on diverse sources to build a model adaptable to unseen scenarios, including future IR and thermal applications.



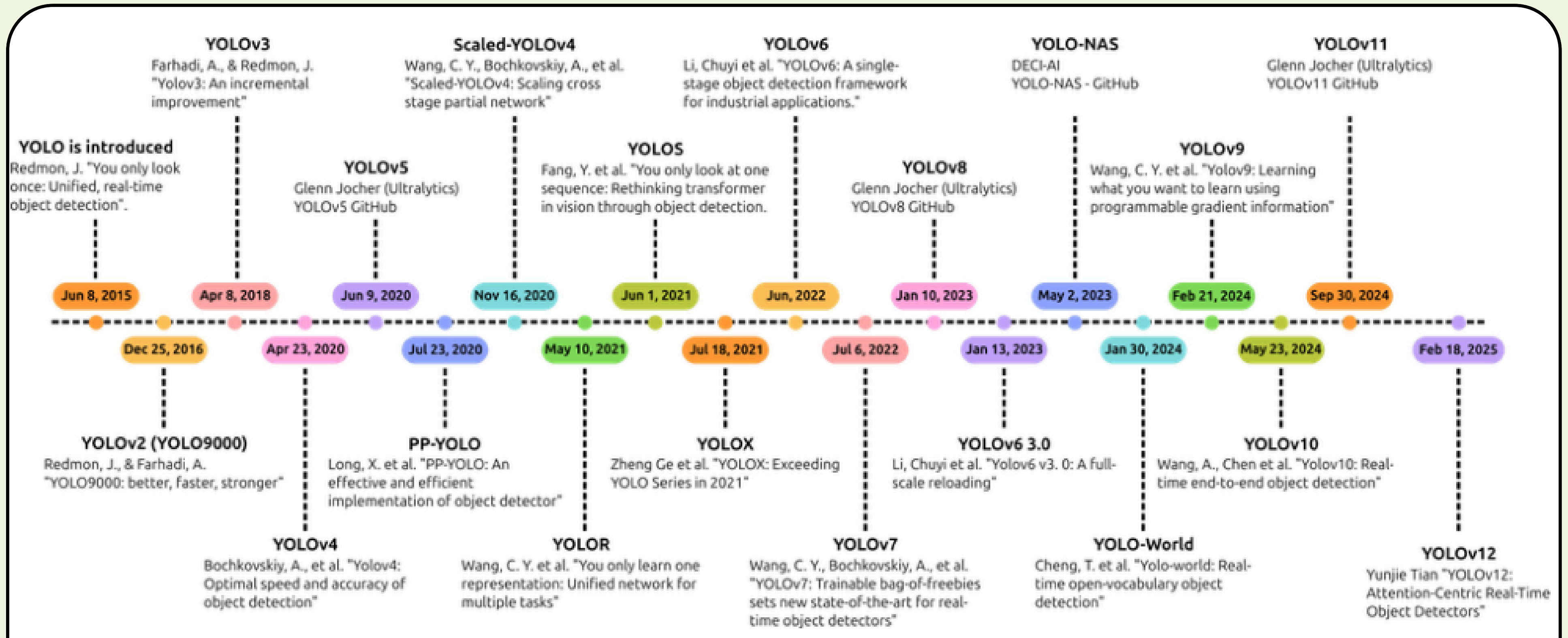
Model Detection Showcase: Generalization
Acrass Acrass Diverse UAV Datasets
Demonrating post-training performance on sampled images across 5 datesets
VisDrone Urban
HIT-UAV Thermal
KIT-MITA Camp
MVRSD Desert
Military Assets
Mixed Rural 1
Mixed Rural 2
Urban Scout
Coastal Patrol



THE SMALL OBJECT CHALLENGE IN UAV SURVEILANCE
KIIT-MITA Dataset: Example of Extreme Scale Variation
Tank
Soldier
Missile
Soldier
Radar
Radar
Soldier
Artillery
Soldier
Soldier
Vehicle

MODEL ARCHITECTURE (YOLOv8s)

- State-of-the-art Detector – YOLOv8: Fast and accurate single-stage object detection.
- Lightweight Variant – YOLOv8s: Efficient for real-time UAV and edge deployment.
- Architectural Upgrades – Anchor-free head & robust backbone for better feature extraction.
- Unified Detection – Predicts all bounding boxes and class probabilities in a single pass.



[YOLO Evolution]

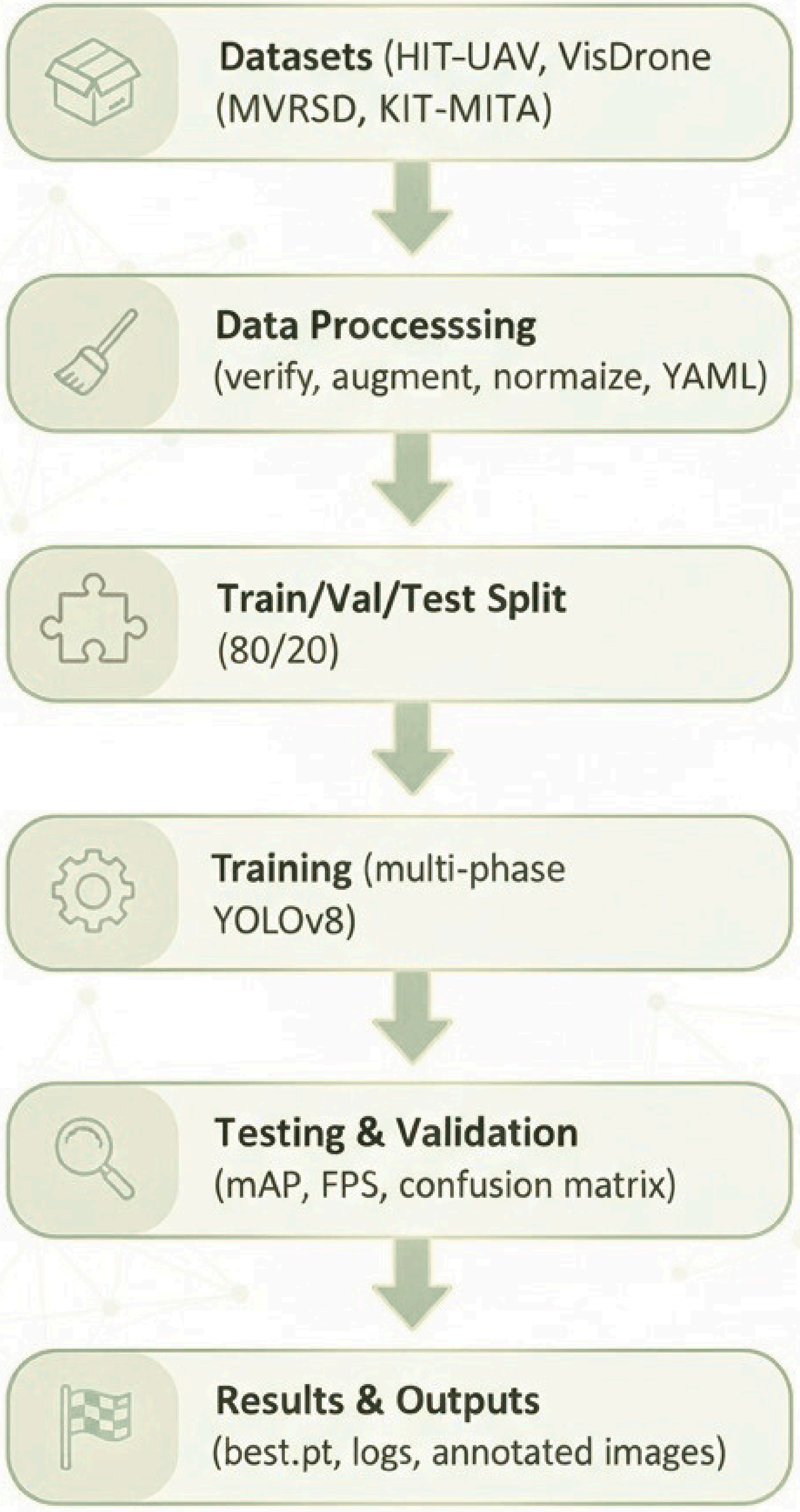
Technical Approach:

🧠Developed entirely in Python libraries — pathlib, yaml, and tqdm for efficient file handling, configuration parsing, and progress tracking.

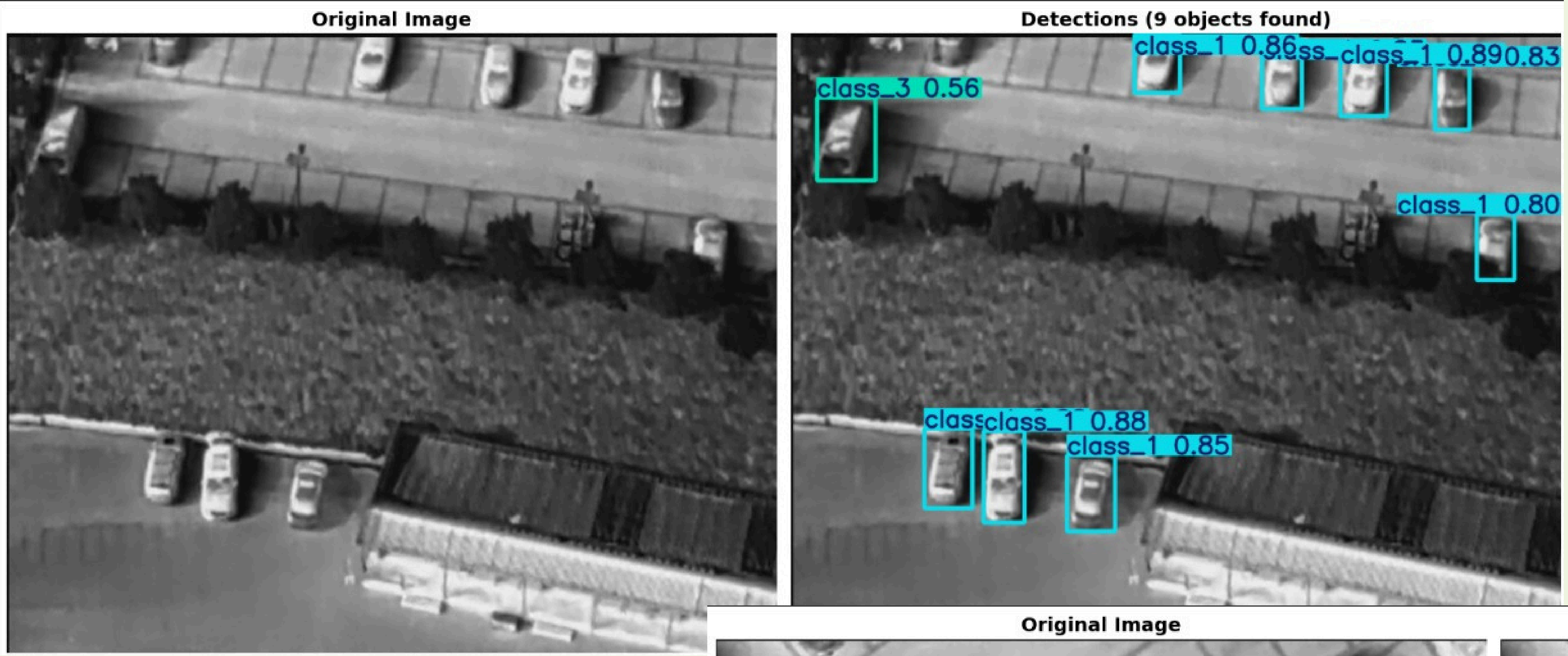
🔍 Automated dataset validation — cross-checks image–label consistency and detects corrupt or missing pairs before training.

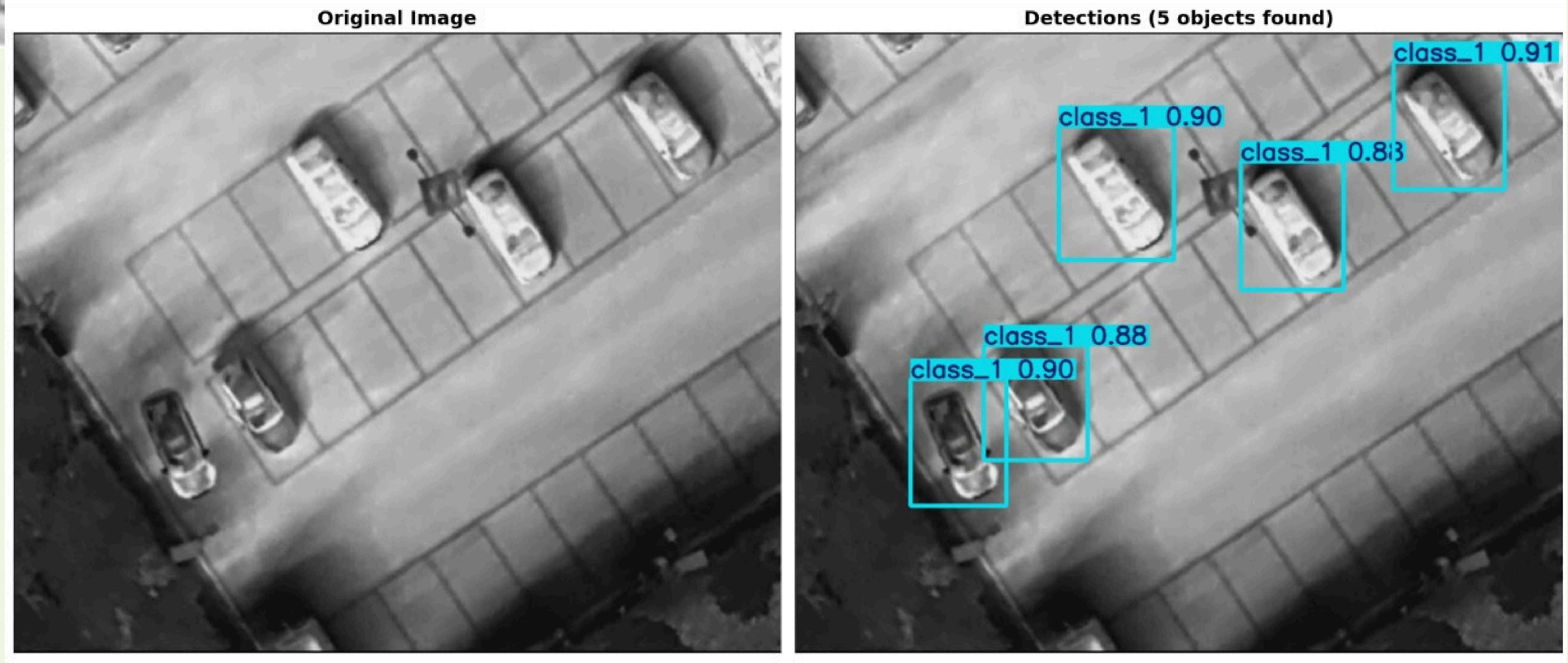
🧾 Dynamic YAML generation — creates standardized YOLOv8/YOLOv11 configuration files directly from dataset metadata.

🔄 Scalable modular architecture — easily extendable for new datasets or additional preprocessing pipelines.

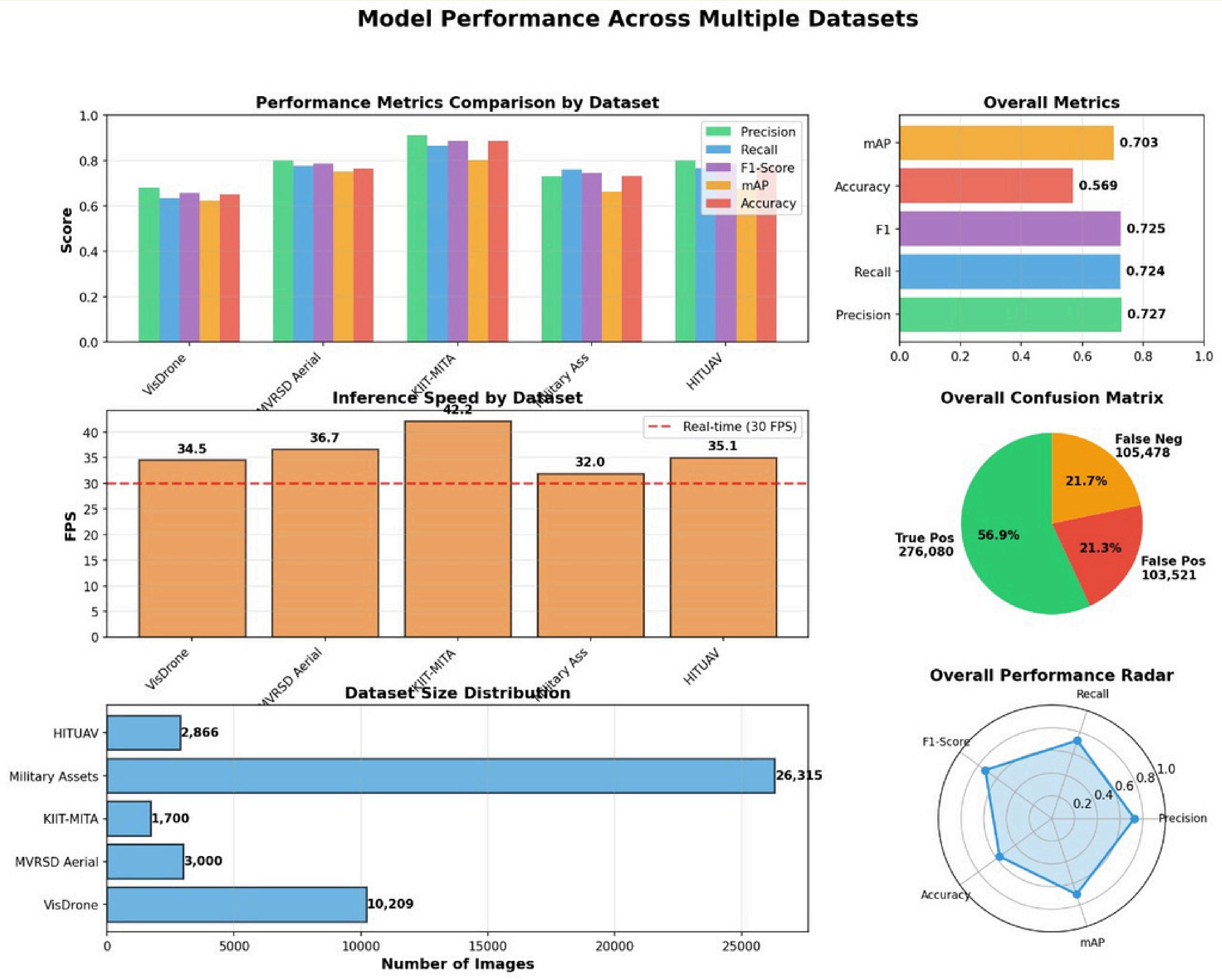


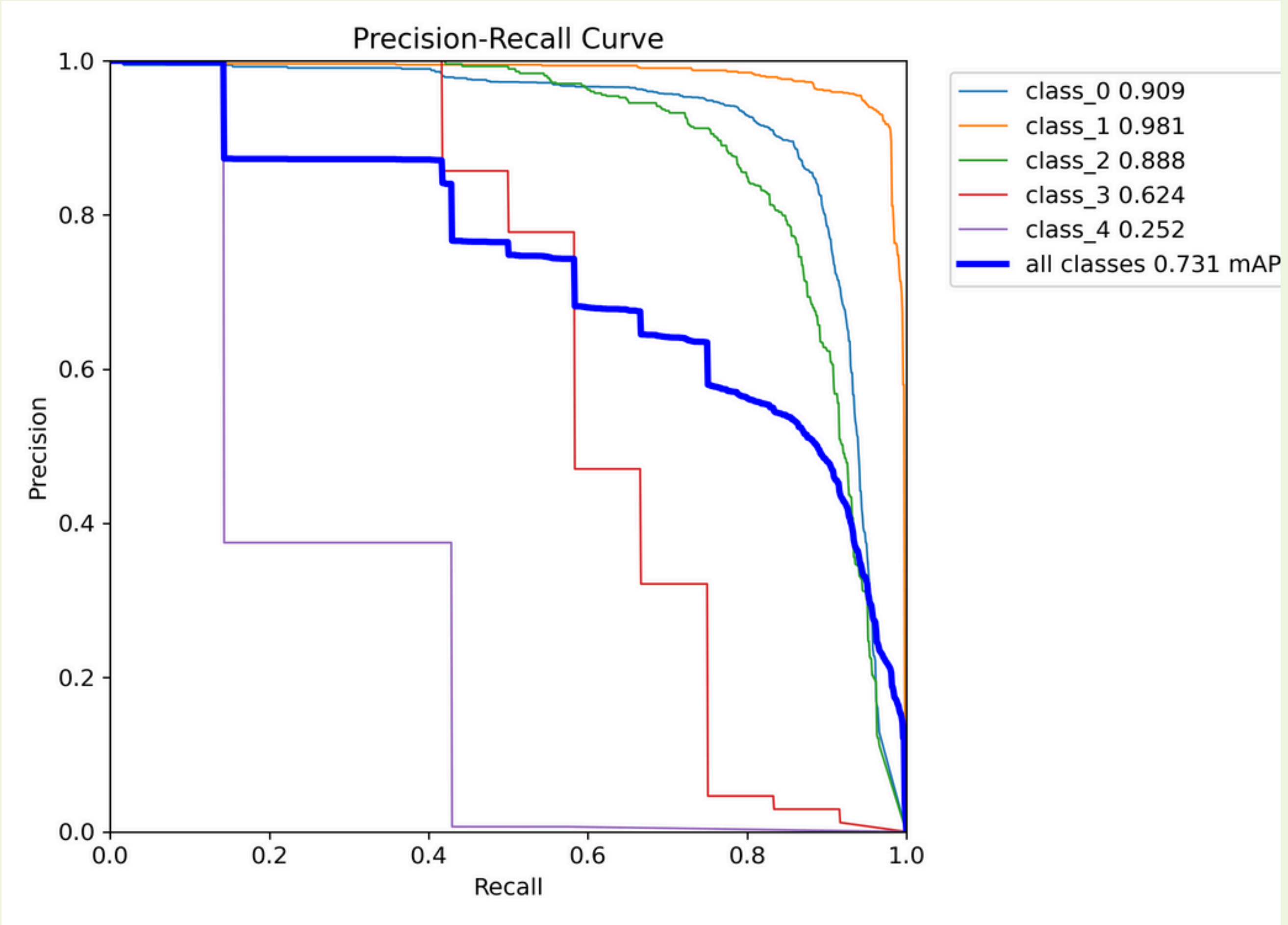
Workflow

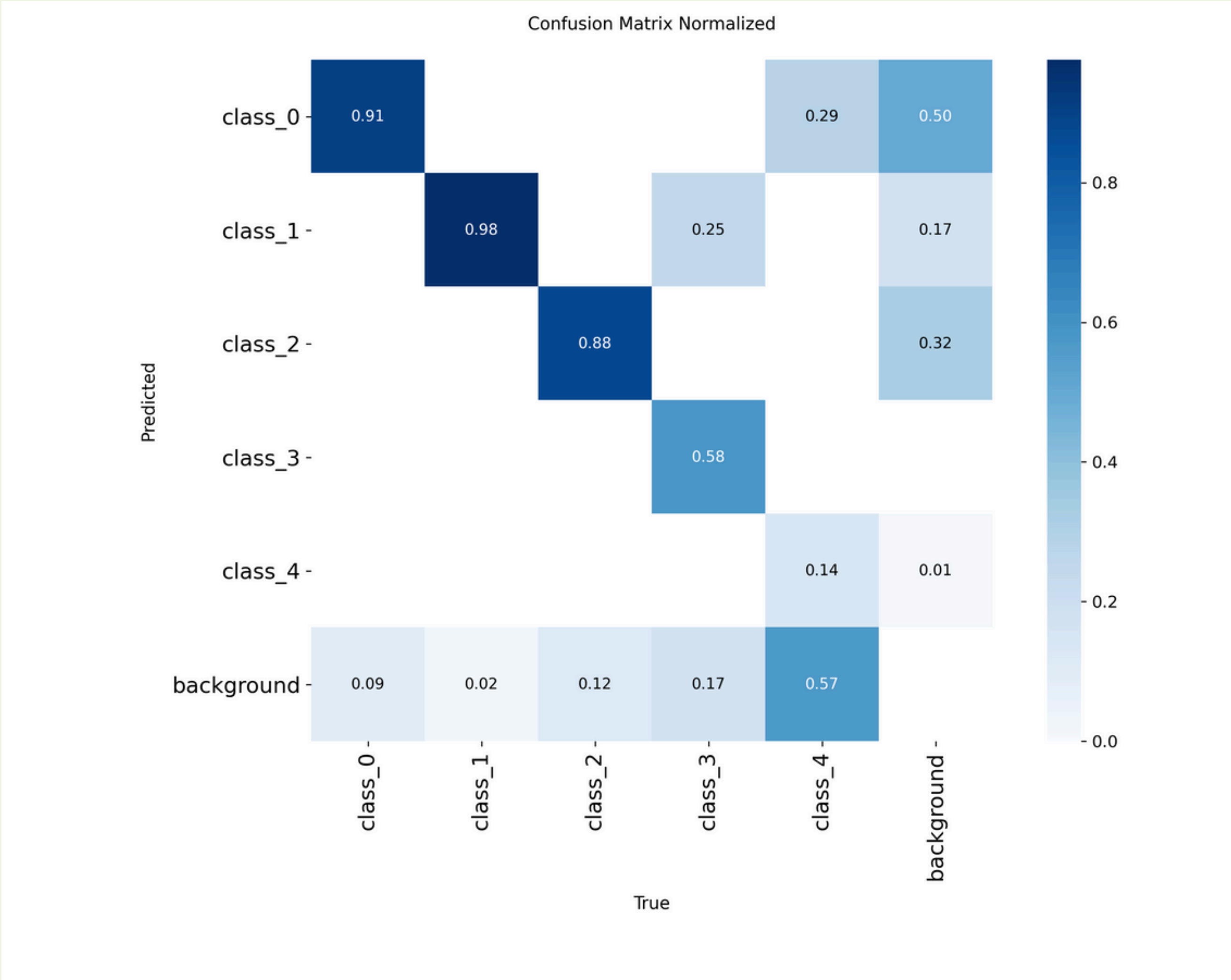

Original Image
Detections (9 objects found)
class_3 0.56
class_1 0.86
class_1 0.89
0.83
class_1 0.80
class_1 0.88
class_1 0.85

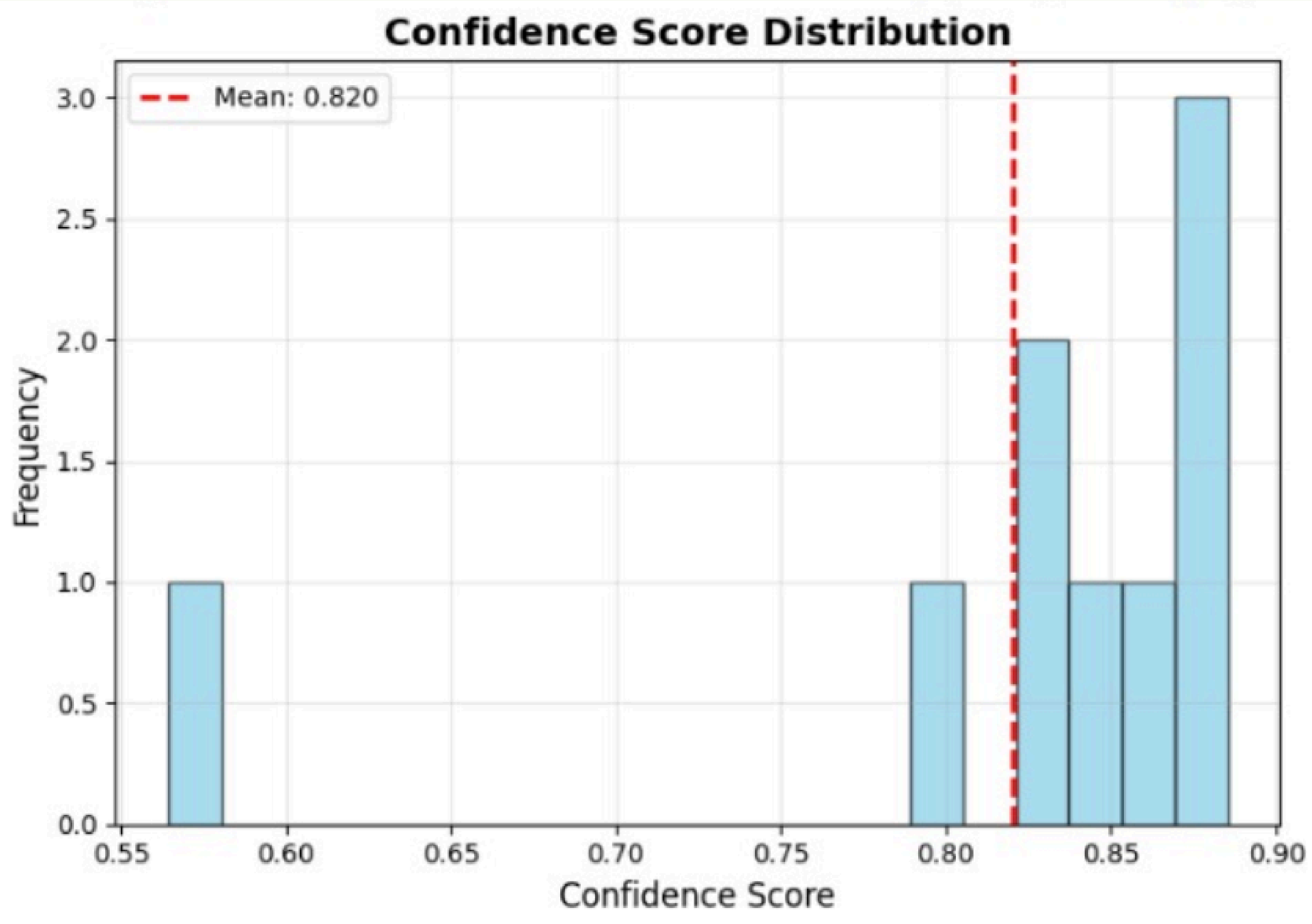

Original Image
Detections (5 objects found)
class_1 0.91
class_1 0.90
class_1 0.88
class_1 0.88
class_1 0.90

Results and Insights:










```
==========================================
📊 OVERALL METRICS (ALL DATASETS COMBINED)
==========================================

🎯 Overall Performance:
   Total Images: 44090
   Precision: 0.7273 (72.73%)
   Recall: 0.7236 (72.36%)
   F1-Score: 0.7254
   Accuracy: 0.5691 (56.91%)
   Average mAP: 0.7031 (70.31%)
   Average FPS: 36.09

📊 Confusion Matrix (Overall):
   True Positives: 276080
   False Positives: 103521
   False Negatives: 105478
```

Dataset	Images	Precision	Recall	F1	mAP	FPS
VisDrone	10209	68.12%	63.34%	0.656	62.43%	34.5
MVRSD Aerial	3000	80.10%	77.60%	0.788	74.97%	36.7
KIIT-MITA	1700	91.12%	86.39%	0.887	80.27%	42.2
Military Assets	26315	73.08%	76.03%	0.745	66.29%	32.0
HITUAV	2866	79.99%	76.75%	0.783	67.59%	35.1
OVERALL	44090	72.73%	72.36%	0.725	70.31%	36.1

Key Achievements:

- **Validated Robustness:** Successfully developed a highly generalized YOLOv8s model capable of accurate object detection across diverse aerial imagery, including varying environments and characteristics.
- **Effective Multi-Phase Learning:** Demonstrated the efficacy of our novel multi-phase sequential training strategy, proving its ability to foster superior model robustness and adaptability from heterogeneous datasets.
- **Real-world Applicability:** Achieved strong performance for challenging detection tasks, such as small object detection and interpretation of multi-modal data characteristics, paving the way for real-time applications.
- **Foundation for Expansion:** Our approach establishes a flexible and performant framework readily extensible to new modalities like dedicated Infrared (IR) and Thermal imagery, addressing critical needs in dynamic visual intelligence.

To Do:

1

Video-Based Object Detection: Develop and implement a pipeline for real-time object detection on live video streams, which will demonstrate the model's performance in dynamic scenarios

2

Model Accuracy Enhancement: Refine the deep learning model to further improve its accuracy, particularly for challenging objects and conditions, by exploring advanced training techniques or architectural modifications.

3

User Interface (UI) Development: Create a user-friendly interface to visualize the model's output, enabling easy interaction and a practical demonstration of the system's capabilities.

Thank You