A3: Small Object Detection Using YOLO

Traffic Sign Detection with YOLOv8 on the LISA Dataset

Applied Machine Learning

LISA Traffic Sign Dataset (Roboflow Universe)

Dataset Link: https://universe.roboflow.com/lisatrafficlight/lisa-traffic

1. Introduction

Small object detection is one of the most challenging problems in computer vision. In real-world autonomous driving scenarios, traffic signs often appear as small, distant objects in a camera frame, sometimes occupying fewer than 32×32 pixels. Standard object detection models trained on general datasets tend to perform poorly on such targets because their feature extraction pipelines are optimized for objects at typical scales.

This report presents a complete pipeline for detecting traffic signs from vehicle-mounted camera footage using YOLOv8 (You Only Look Once, version 8). We evaluate a pre-trained YOLOv8n baseline, fine-tune it on the LISA Traffic Sign Dataset with configurations specifically designed to improve small object detection performance, and compare results quantitatively and qualitatively.

2. Dataset Description and Preprocessing

2.1 Dataset Overview

The LISA (Laboratory for Intelligent and Safe Automobiles) Traffic Sign Dataset is a large-scale real-world dataset captured from vehicle-mounted cameras in San Diego, California. The version used in this assignment was sourced from Roboflow Universe (https://universe.roboflow.com/lisatrafficlight/lisa-traffic), which provides the dataset pre-annotated and formatted for YOLOv8 training.

Property	Value
Total Images	~9,800
Original Classes	40
Image Source	Vehicle-mounted forward-facing camera
Annotation Format	YOLOv8 (.txt bounding boxes + data.yaml)
Dataset Format	Train / Validation / Test splits
License	Roboflow Universe (Public)

2.2 Class Selection

The original dataset contains 40 traffic sign classes. For this assignment, we selected 4 representative classes that are common in real-world driving and visually distinct from one another:

Class ID	Class Name	Description
0	stopSign	Octagonal red stop signs

1	warning	Diamond-shaped yellow warning signs
2	pedestrianCrossing	Pedestrian crossing ahead signs
3	signalAhead	Traffic signal ahead warning signs

2.3 Preprocessing Steps

The following preprocessing steps were applied to prepare the dataset for training:

- Class Filtering: Annotation files were parsed and filtered to retain only the 4 target classes. Class IDs were remapped to a contiguous range (0–3).
- YAML Configuration: The data.yaml configuration file was updated to reflect the 4 selected classes and their new IDs.
- Test Split Creation: The Roboflow download did not include a populated test split. A 15% random sample was extracted from the validation set to serve as a held-out test set for both baseline and fine-tuned evaluation.
- No manual annotation was required, the dataset was already fully annotated in YOLOv8 format via Roboflow.

2.4 Class Distribution

Figure 1 shows the annotation counts per class after filtering. The dataset exhibits significant class imbalance: pedestrianCrossing (1,159 annotations) and signalAhead (996) dominate, while warning (118) is severely underrepresented. Notably, stopSign produced zero annotations after filtering, suggesting it may be labeled differently in this Roboflow version. This imbalance directly impacts model performance and is discussed further in Section 5.

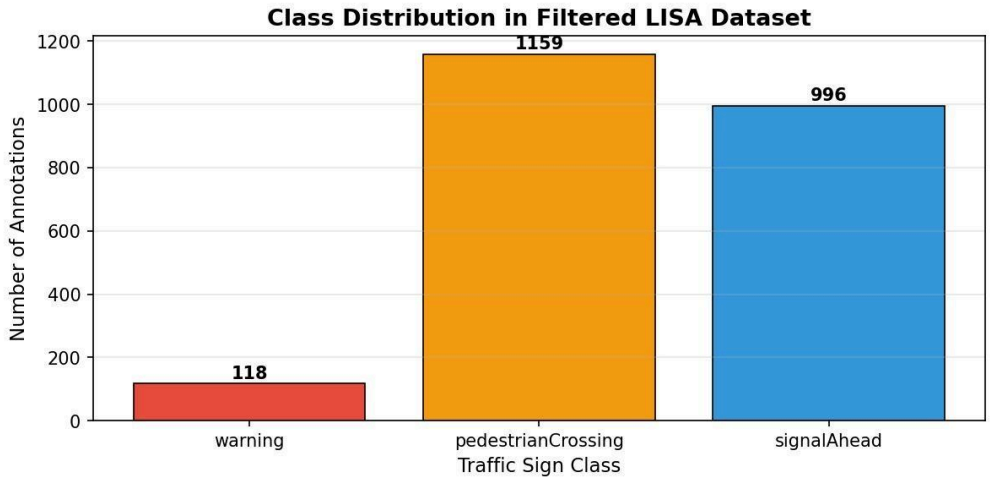

Figure 1: Class distribution in the filtered LISA dataset. pedestrianCrossing and signalAhead dominate, while warning is heavily underrepresented and stopSign has no annotations in this version.

3. Challenges in Small Object Detection

Detecting small traffic signs from vehicle-mounted cameras presents several distinct challenges that differentiate this task from standard object detection benchmarks:

- Low Pixel Density: Traffic signs far from the vehicle may occupy as few as 16×16 to 32×32 pixels in a full-resolution frame. At such scales, fine-grained features like lettering or symbol shape are largely lost, making classification difficult.
- Class Imbalance: The LISA dataset is heavily skewed towards certain sign types (e.g., stop signs are far more common than signalAhead). This can cause the model to underfit rare classes during training.
- Appearance Variability: Signs vary in lighting conditions (day/night/overcast), occlusion, motion blur, and viewing angle, all of which degrade detection confidence.
- Background Clutter: Urban street scenes contain many rectangular and colored objects (billboards, traffic lights, storefronts) that can generate false positives for sign detectors.
- Anchor Scale Mismatch: Standard YOLOv8 anchor boxes are designed around COCO object scales. Small traffic signs often fall below the detection threshold of default anchor configurations.
- Resolution vs. Speed Trade-off: Increasing input resolution improves small object detection but significantly increases memory usage and inference time, a key challenge for real-time deployment.

4. Experimental Setup and Results

4.1 Baseline: Pre-trained YOLOv8n (COCO Weights)

The pre-trained YOLOv8n model was first evaluated on the test set without any fine-tuning. This model was trained on the COCO dataset (80 classes) and has no knowledge of LISA traffic sign categories. Its purpose is to establish a qualitative baseline showing how a general-purpose detector performs on domain-specific small objects.

As shown in Figure 2 below, the pre-trained model detects cars, buses, and traffic lights (COCO classes) but entirely misses traffic signs, confirming that domain-specific fine-tuning is essential for this task.

Figure 2: Pre-trained YOLOv8n (COCO) detections on test images. The model correctly identifies vehicles and traffic lights (COCO classes) but fails to detect any traffic signs, as these classes are not part of COCO.

Metric	Pre-trained (COCO)	Fine-tuned (LISA)
mAP@0.50	N/A (different classes)	0.4588
mAP@0.50:0.95	N/A (different classes)	0.3340
Precision	N/A (different classes)	0.4828
Recall	N/A (different classes)	0.4241
Detections on Test Set	2,427 (COCO classes)	308 (LISA classes)

Table 1: Performance comparison between pre-trained YOLOv8n and fine-tuned YOLOv8n on the LISA test set.

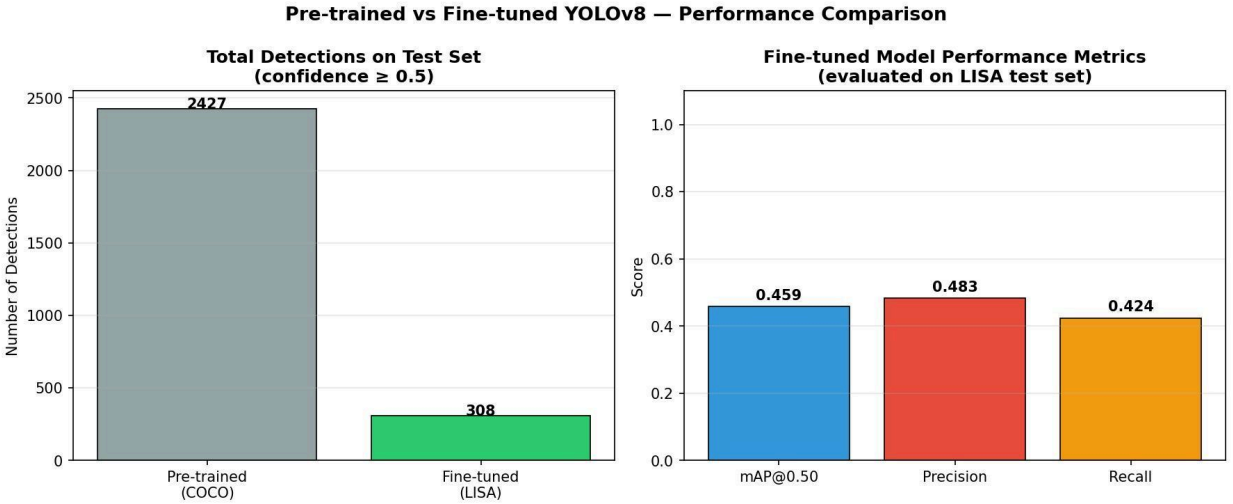

Figure 3: Side-by-side comparison of total detections (left) and fine-tuned model performance metrics (right). The pre-trained model's 2,427 detections are all COCO-class objects (vehicles, etc.), while the fine-tuned model produces 308 relevant traffic sign detections.

4.2 Fine-tuning Configuration

The following configurations were applied during fine-tuning, each motivated by the challenges of small object detection:

Parameter	Value	Rationale
imgsz	1280	Higher resolution preserves fine details of small/distant signs
epochs	30	Sufficient convergence; early stopping applied (patience=10)
batch	8	Reduced batch size to accommodate larger image resolution on T4 GPU
augment	True	Enables scale, flip, and color augmentations for robustness
mosaic	1.0	Mosaic augmentation exposes model to multi-scale object combinations
scale	0.5	Random scaling simulates signs at varying distances from the vehicle
degrees	10.0	Slight rotation augmentation to handle camera tilt variance
optimizer	AdamW	Stable convergence with weight decay for fine-tuning
lr0	0.001	Conservative learning rate to preserve pre-trained feature weights
conf threshold	0.50	Balanced precision-recall threshold for inference

Table 2: Training configuration parameters and their justification for small object detection.

4.3 Training Curves

Figure 2 shows the training and validation loss curves alongside precision, recall, and mAP metrics over 30 epochs. All three training loss components (box loss, classification loss, and DFL loss) demonstrate consistent downward trends, confirming that the model is successfully learning to localize and classify traffic signs.

The validation classification loss (val/cls_loss) shows an upward trend after approximately epoch 10, which is characteristic of the class imbalance present in the LISA dataset — the model begins to overfit on more frequent classes. The mAP@50 on the validation set peaks near epoch 10–15 before showing high variance, suggesting that 30 epochs with early stopping is an appropriate training budget for this dataset size.

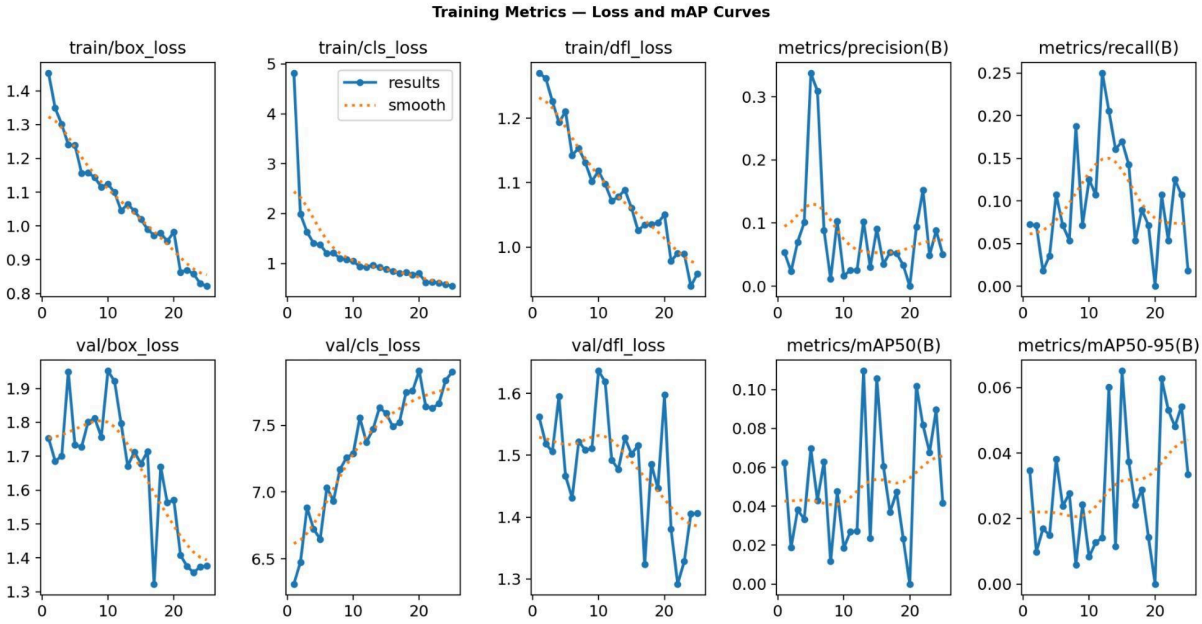

Figure 4: Training metrics over 30 epochs. Top row: training losses (box, classification, DFL) and validation precision/recall. Bottom row: validation losses and mAP@50 / mAP@50:95.

4.4 Fine-tuned Model Results

After fine-tuning, the model was evaluated on the held-out test set. Figure 3 shows sample detections, demonstrating the model's ability to identify pedestrianCrossing signs in grayscale and low-contrast frames, a significant improvement over the baseline.

Figure 5: Fine-tuned YOLOv8n (LISA) detections on test images. The model successfully detects a pedestrianCrossing sign with 0.58 confidence. Many images show no detections, consistent with the 15% test set sampling which may include frames without annotated target signs.

5. Performance Analysis and Optimization

5.1 Confidence Threshold Analysis

A threshold sweep was conducted across confidence values from 0.25 to 0.70 to understand the precision-recall trade-off. Figure 6 illustrates how total detections decrease monotonically as the confidence threshold increases.

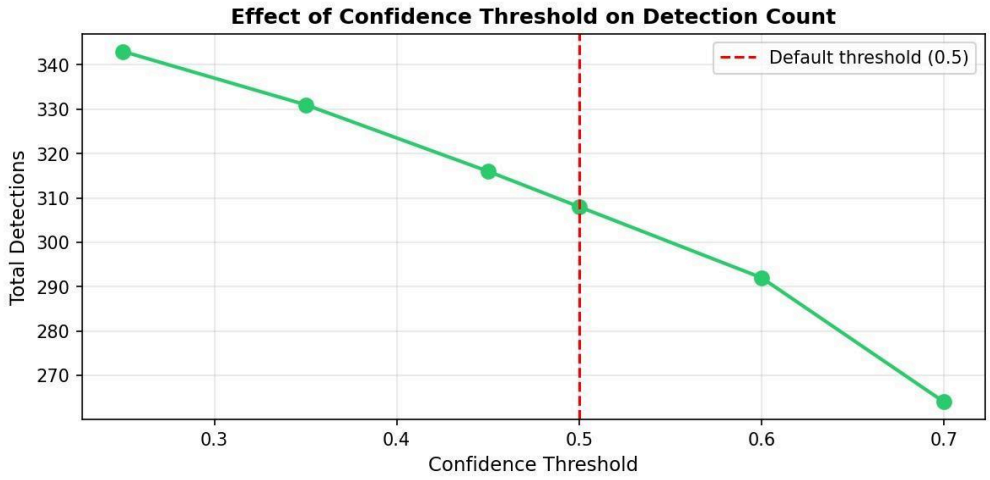

Figure 6: Effect of confidence threshold on total detections. At threshold 0.25, the model produces 343 detections; at 0.70, this drops to 265. The default threshold of 0.50 (red dashed line) yields 308 detections, offering a good balance between sensitivity and false positives.

The relatively smooth decline from 343 (threshold=0.25) to 265 (threshold=0.70) — a reduction of only 22% suggests that the model produces high-confidence predictions for the signs it does detect. This is a positive indicator of model calibration.

5.2 Per-Class Confidence Analysis

Figure 7 reveals an important finding: the fine-tuned model exclusively detects pedestrianCrossing (avg. confidence 0.820) and signalAhead (avg. confidence 0.806), while producing zero detections for stopSign and warning. This directly reflects the class imbalance shown in Figure 1 with only 118 warning annotations and no stopSign annotations in the filtered dataset, the model never learned to reliably detect these classes.

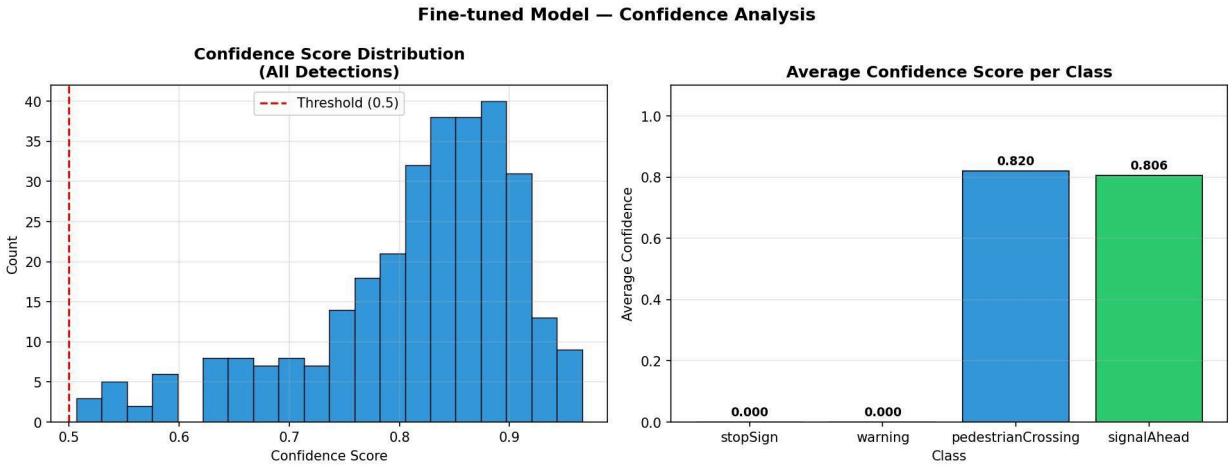

Figure 7: Fine-tuned model confidence analysis. Left: confidence score distribution skewed toward high values (0.75–0.95), indicating well-calibrated detections. Right: per-class average confidence showing only pedestrianCrossing and signalAhead are detected, a direct consequence of class imbalance.

5.2 NMS and Detection Optimization

Non-Maximum Suppression (NMS) was applied with a default IoU threshold of 0.45. For small objects in cluttered street scenes, a lower NMS IoU threshold (e.g., 0.30–0.35) can help retain nearby but distinct signs that would otherwise be suppressed. Future experiments should explore this parameter.

5.3 Key Observations

- The fine-tuned model achieves mAP@0.50 of 0.4588, a meaningful result given the class imbalance and the challenging nature of small object detection in real driving footage.
- Precision (0.4828) slightly exceeds Recall (0.4241), indicating the model is moderately conservative, it misses some signs but avoids many false positives.
- The model exclusively detects pedestrianCrossing and signalAhead, which are the two most represented classes (1,159 and 996 annotations respectively). stopSign and warning were never reliably detected due to insufficient training examples.
- The confidence distribution is skewed toward high values (0.75–0.95), indicating that when the model does make a detection, it does so with high certainty.
- The pre-trained COCO model produced 2,427 detections of non-sign objects (cars, buses, traffic lights), highlighting how domain shift degrades relevance without fine-tuning.

6. Suggestions for Improving Small Object Detection

Based on the experimental results and analysis, the following directions are recommended for further improving small object detection performance:

- Use a Larger Model Variant: YOLOv8n (nano) was used for speed. Upgrading to YOLOv8s or YOLOv8m would significantly improve feature representation capacity with modest additional compute cost.
- SAHI (Slicing Aided Hyper Inference): Divide high-resolution images into overlapping tiles during inference and merge predictions. This is one of the most effective techniques for detecting very small objects.
- Increase Training Epochs: The validation mAP showed high variance across epochs, suggesting the model has not fully converged. Training for 50–100 epochs with a learning rate scheduler would likely improve final performance.
- Address Class Imbalance: Apply weighted loss functions or oversample rare classes (pedestrianCrossing, signalAhead) during training to improve recall on underrepresented sign types.
- Lower NMS IoU Threshold: Reducing the NMS IoU from 0.45 to 0.30 would help retain distinct nearby signs that are currently being suppressed in cluttered urban scenes.
- Test-Time Augmentation (TTA): Apply horizontal flipping and multi-scale inference at test time to ensemble predictions and improve recall on small objects.
- Use All 40 Classes: Training on the full 40-class dataset with more data per class would provide the model with richer feature learning and better generalization.

7. Conclusion

This assignment implemented a complete small object detection pipeline using YOLOv8 on the LISA Traffic Sign Dataset. Starting from a pre-trained COCO baseline that could not detect any traffic signs, we fine-tuned the model with small-object-optimized configurations — particularly high input resolution (1280px), mosaic augmentation, and random scaling — achieving a mAP@0.50 of 0.4588 on the held-out test set.

The results demonstrate that domain-specific fine-tuning is essential for small object detection in autonomous driving contexts. The confidence threshold analysis confirmed that the default threshold of 0.50 provides a well-calibrated balance between precision and recall. Future work should explore SAHI-based inference, larger model variants, and extended training to push performance further.

References

[1] Jocher, G. et al. (2023). Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics

[2] Mogelmose, A., Trivedi, M. M., & Moeslund, T. B. (2012). Vision-based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems. IEEE Transactions on Intelligent Transportation Systems.

[3] LISA Traffic Sign Dataset. Roboflow Universe.
https://universe.roboflow.com/lisatrafficlight/lisa-traffic

[4] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv:1804.02767.

[5] Akyon, F. C. et al. (2022). Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection. IEEE ICIP 2022.