

YOLO Model Testing Report

Model:	yolov8x
Model Size:	130.5 MB
Parameters:	68.2 M
FLOPs (640x640):	258.55 GFLOPs
Run Name:	yolov8x_testing_20251127_215121
Timestamp:	2025-11-27 22:04:05
Dataset:	bdd100k_yolo_limited - test split
Images Processed:	20000
Total Execution Time	146.67s
IoU Threshold:	0.5

Inference Performance

Metric	Value
Average Inference Time	4.06 ms per image
FPS (Frames Per Second)	246.45

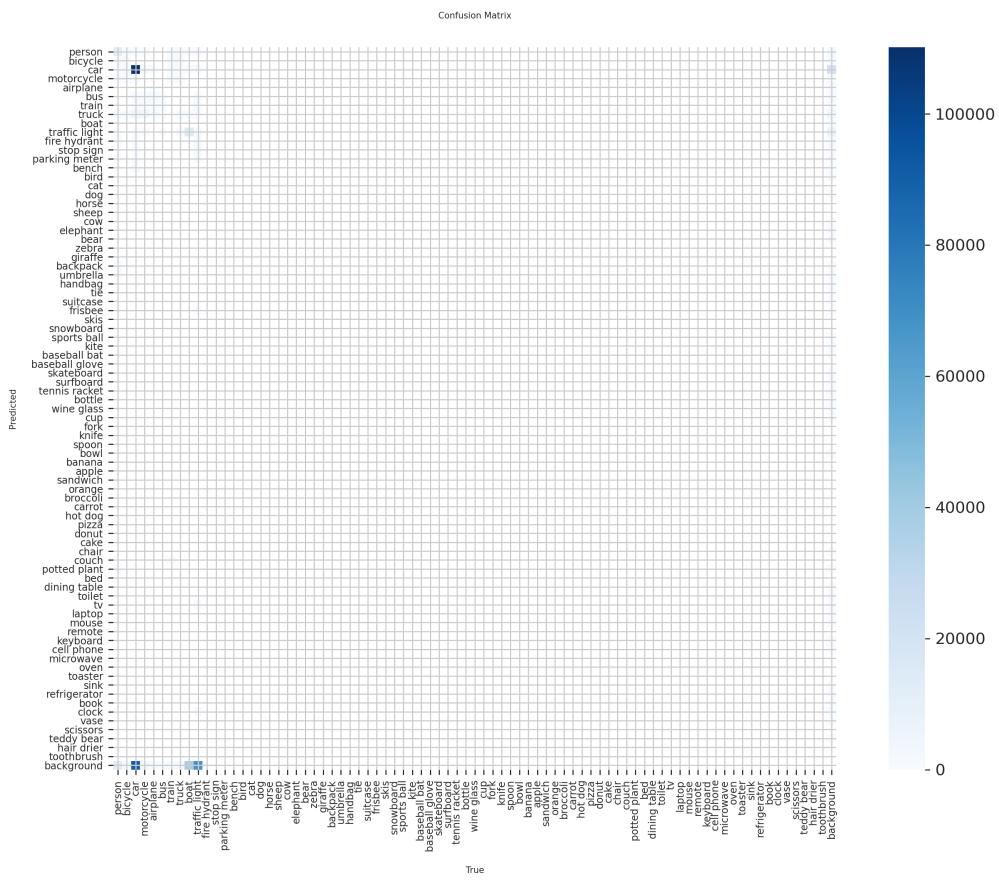
Note: Inference time includes preprocessing, model inference, and postprocessing.

Overall Accuracy Metrics

Metric	Value
Precision	0.3337
Recall	0.6711
F1-Score	0.4457
mAP@0.5	0.1281
mAP@0.5:0.95	0.0710
True Positives	122700
False Positives	245027
False Negatives	60121

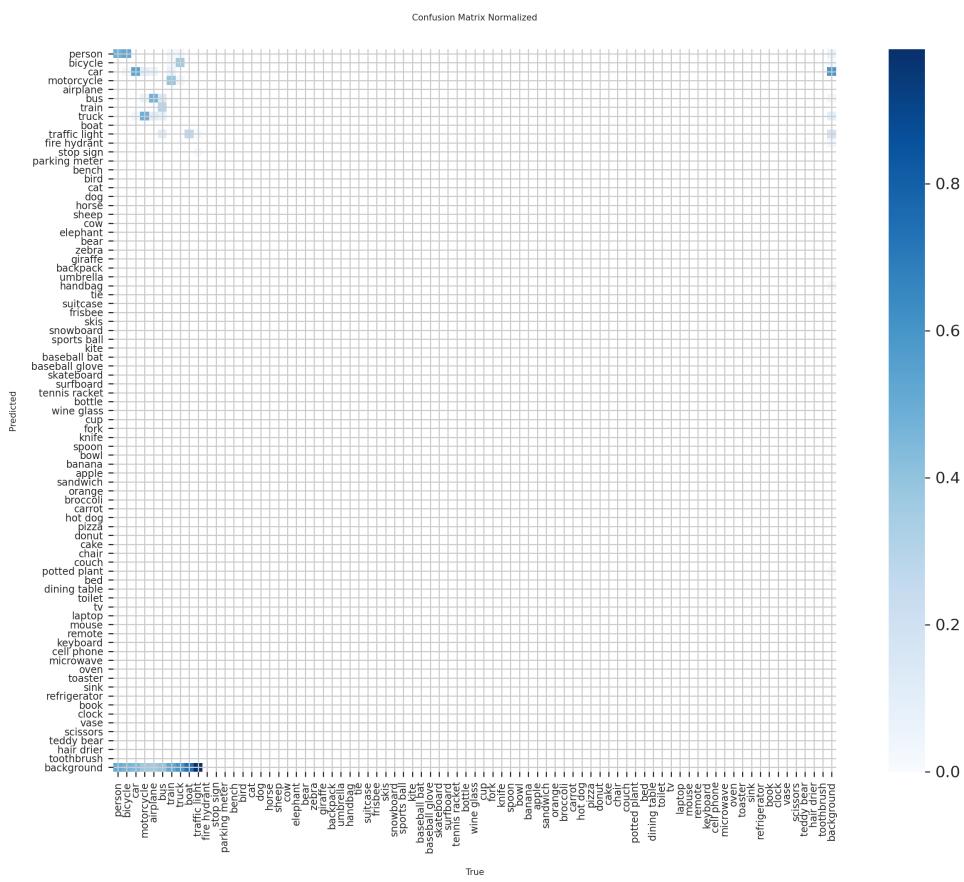
Confusion Matrices

Confusion Matrix (Counts)



Note: Shows raw prediction counts for each class combination.

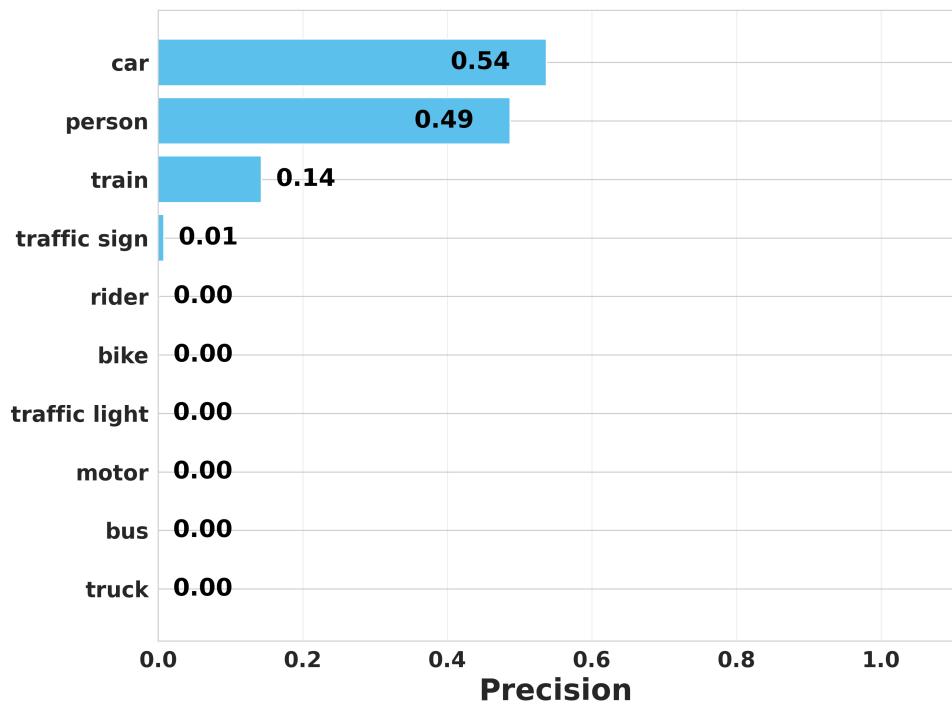
Confusion Matrix (Normalized)



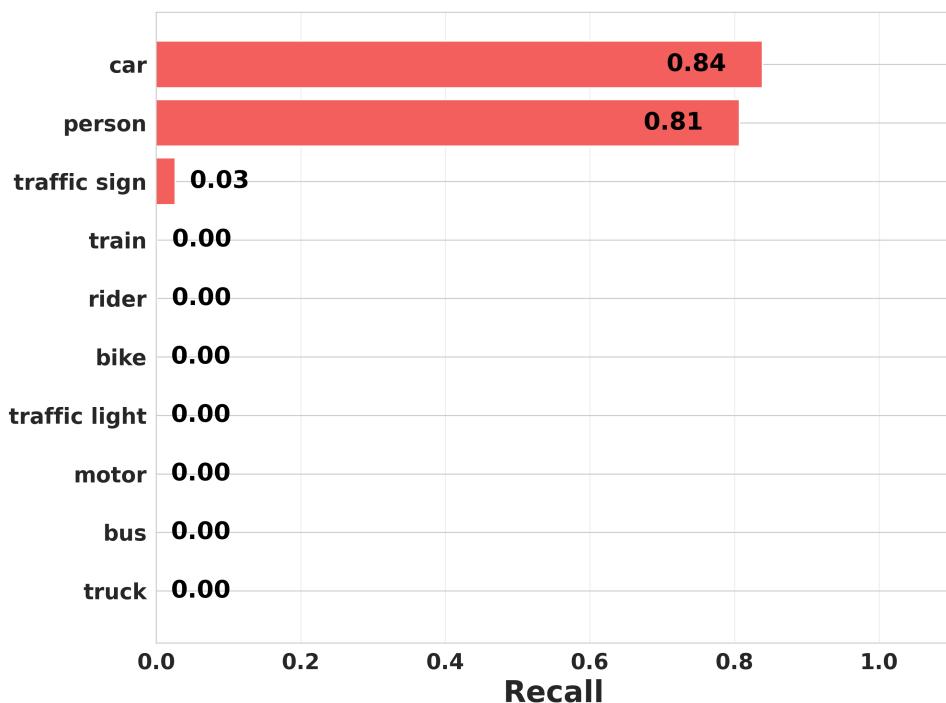
Note: Values are normalized by the number of ground truth instances per class. Shows the proportion of correct vs incorrect classifications.

Performance Visualizations

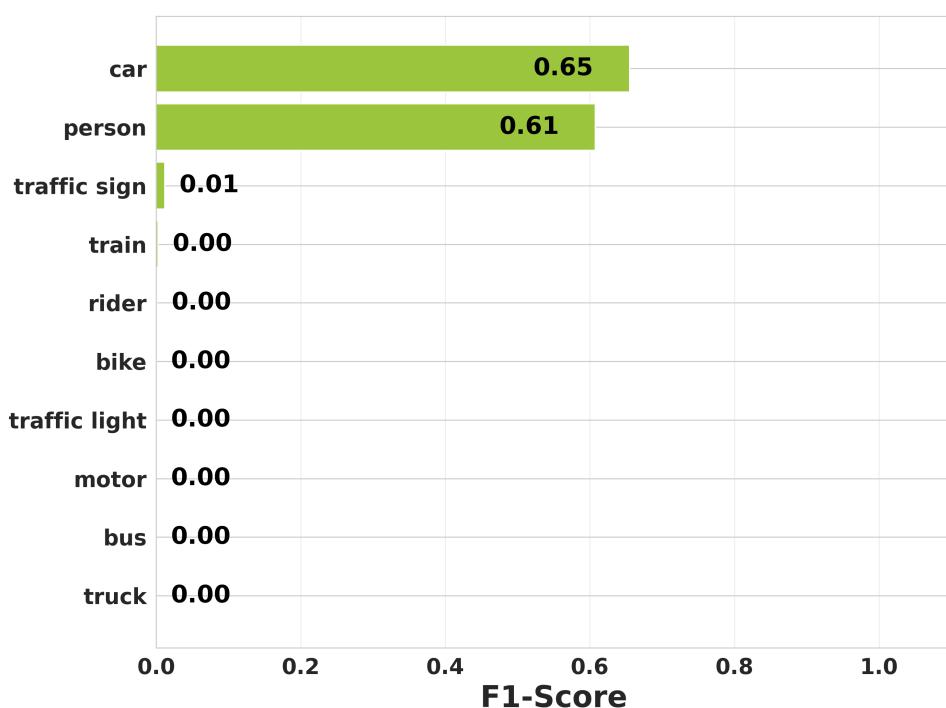
Precision by Class



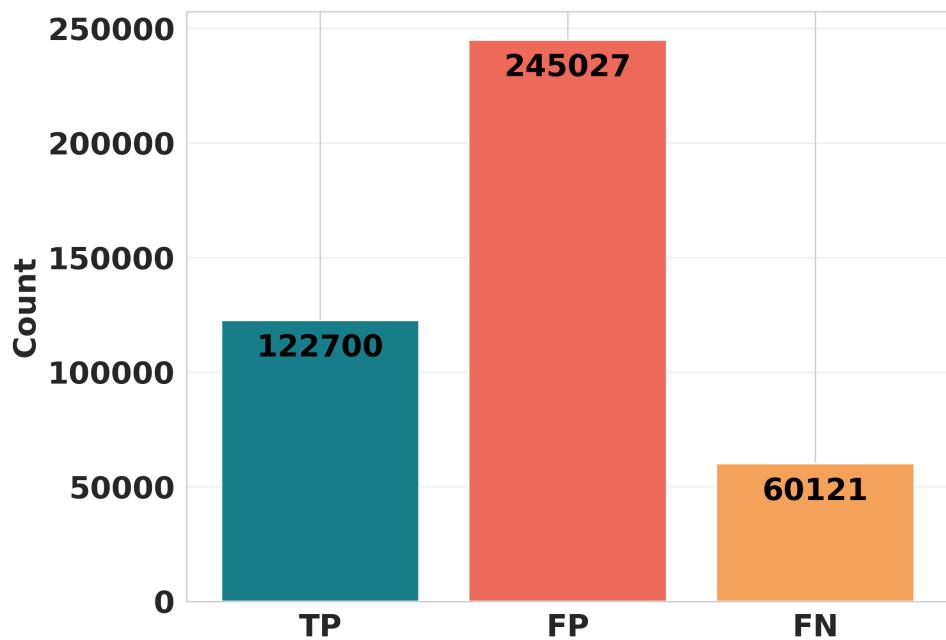
Recall by Class



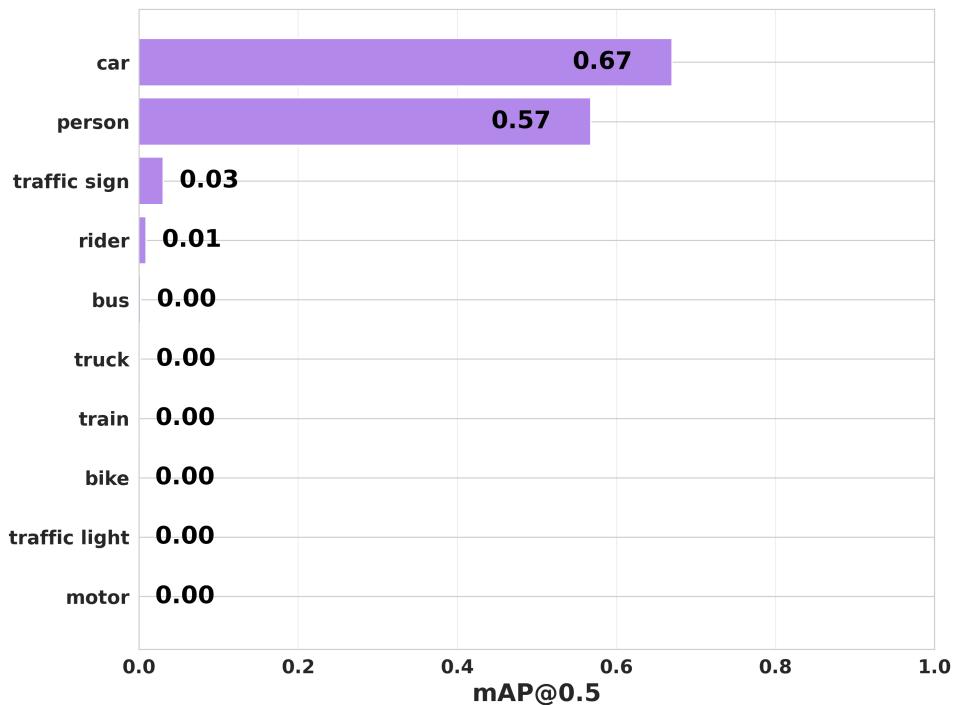
F1-Score by Class



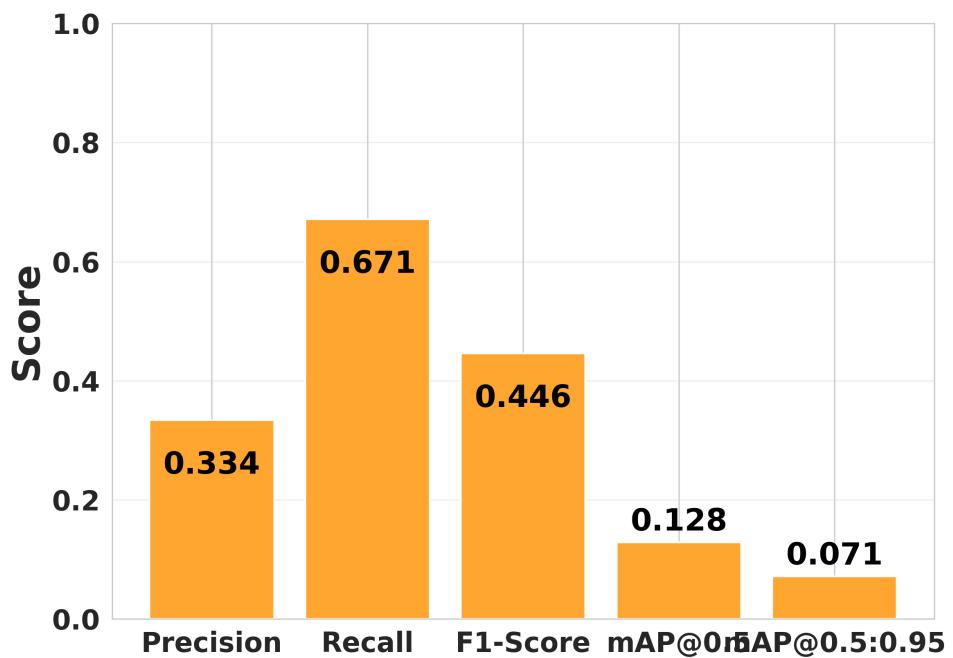
Overall Detection Outcomes



mAP@0.5 by Class



Overall Metrics



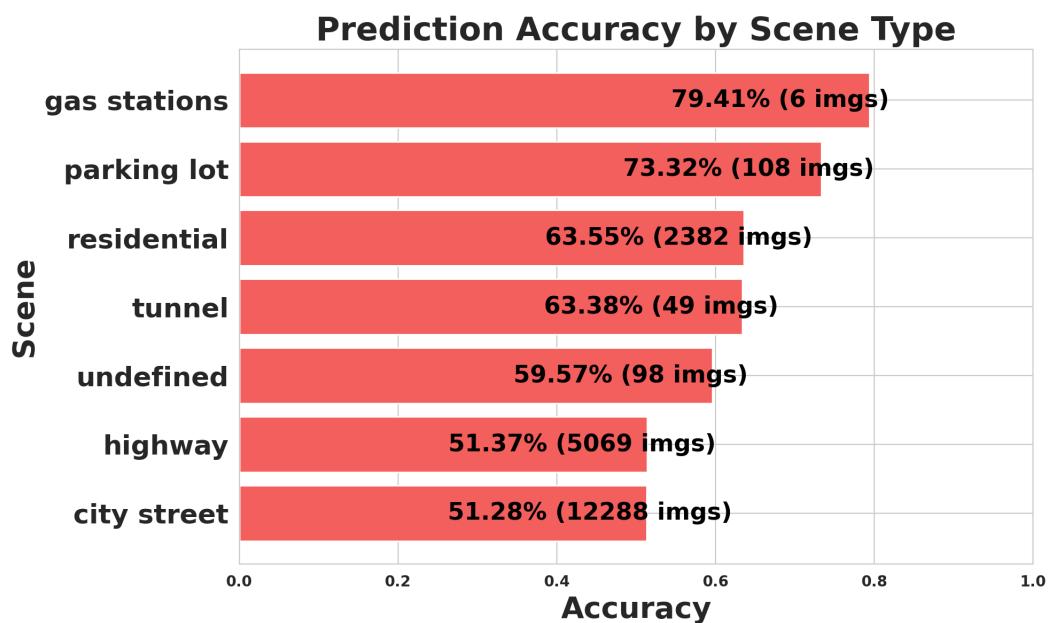
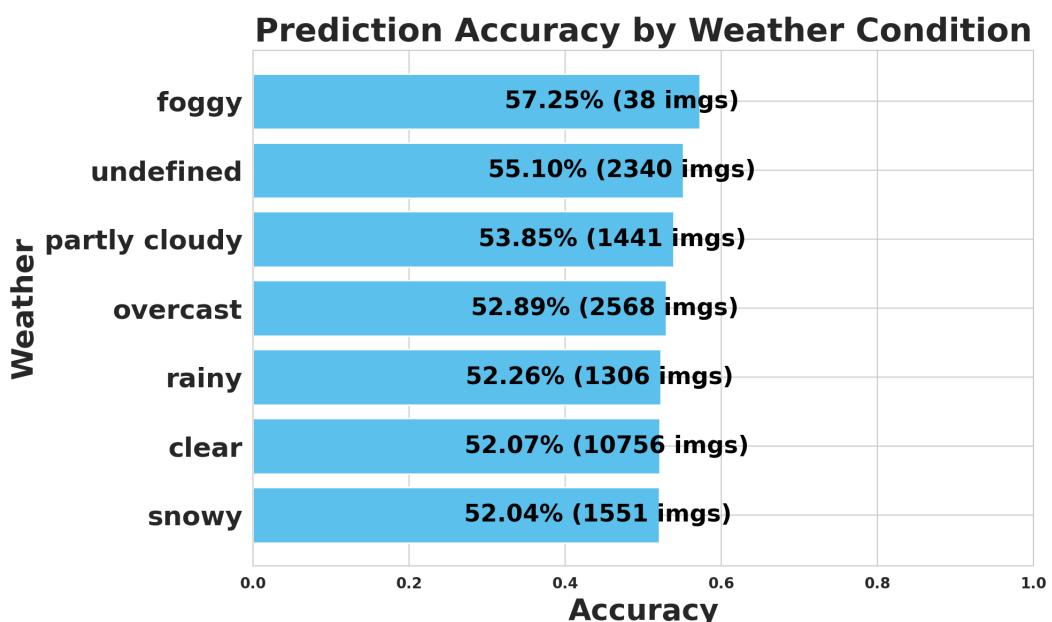
Per-Class Performance

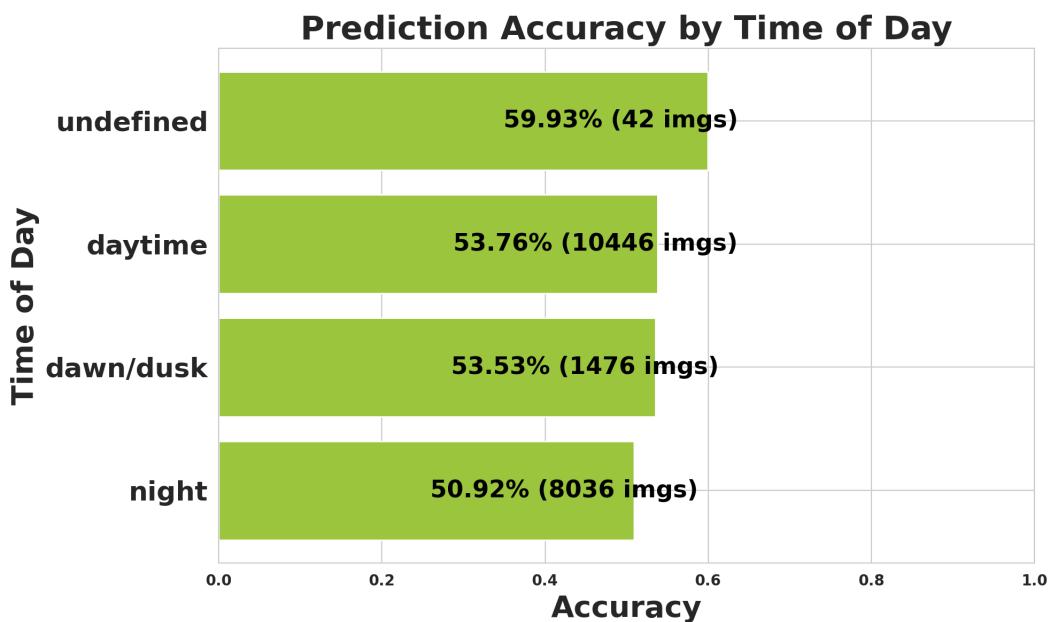
Class	TP	FP	FN	Precision	Recall	F1-Score	mAP@0.5
person	12004	12646	2872	0.4870	0.8069	0.6074	0.5680
rider	1	1293	907	0.0008	0.0011	0.0009	0.0087
car	110161	94988	21176	0.5370	0.8388	0.6548	0.6702
truck	0	8704	512	0.0000	0.0000	0.0000	0.0011
bus	0	3217	21	0.0000	0.0000	0.0000	0.0021
train	4	24	3129	0.1429	0.0013	0.0025	0.0005
motor	0	841	155	0.0000	0.0000	0.0000	0.0000
bike	1	1997	11713	0.0005	0.0001	0.0001	0.0001
traffic light	0	52840	17	0.0000	0.0000	0.0000	0.0000
traffic sign	529	68477	19619	0.0077	0.0263	0.0119	0.0304

Comprehensive Failure Analysis

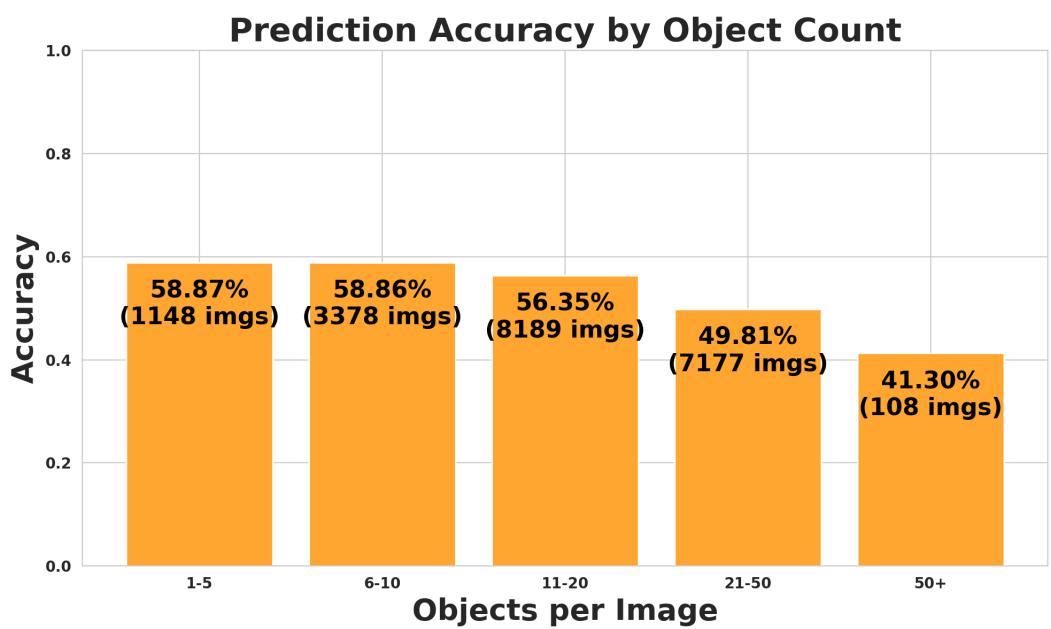
Metric	Value
Total Images Analyzed	20000
Total Expected Objects	367728
Total Matched Objects	193910
Overall Accuracy	52.73%

Accuracy by Environmental Attributes

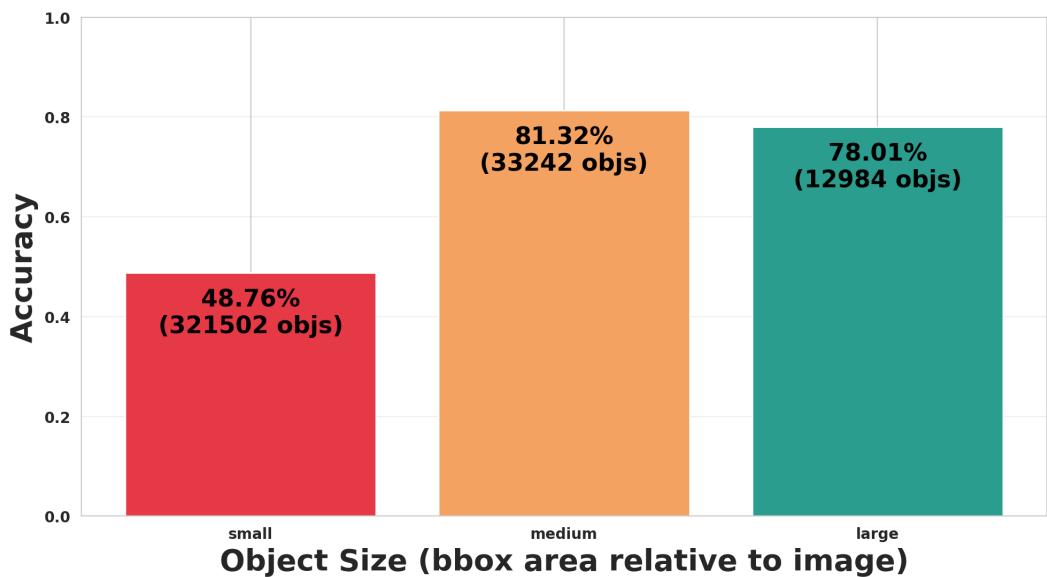




Accuracy by Object Characteristics

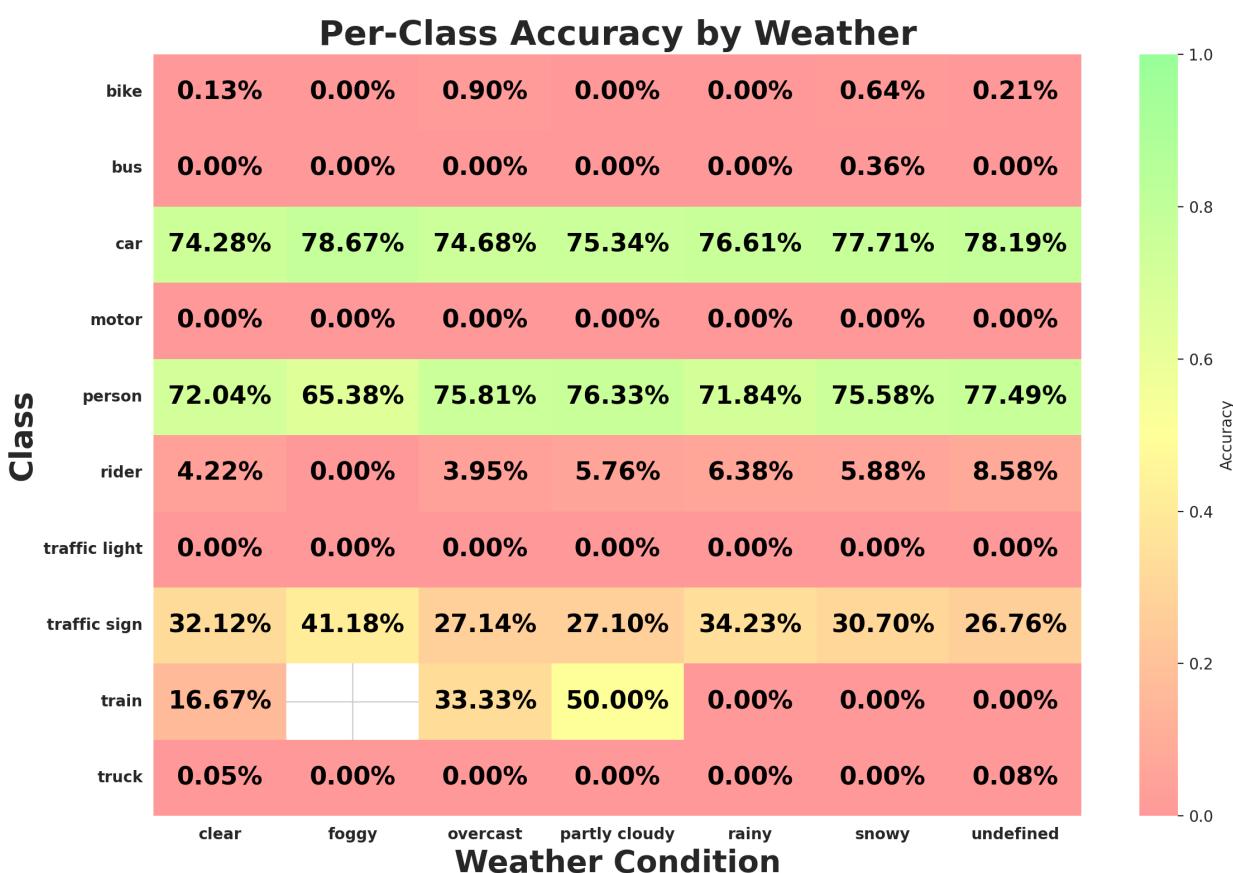
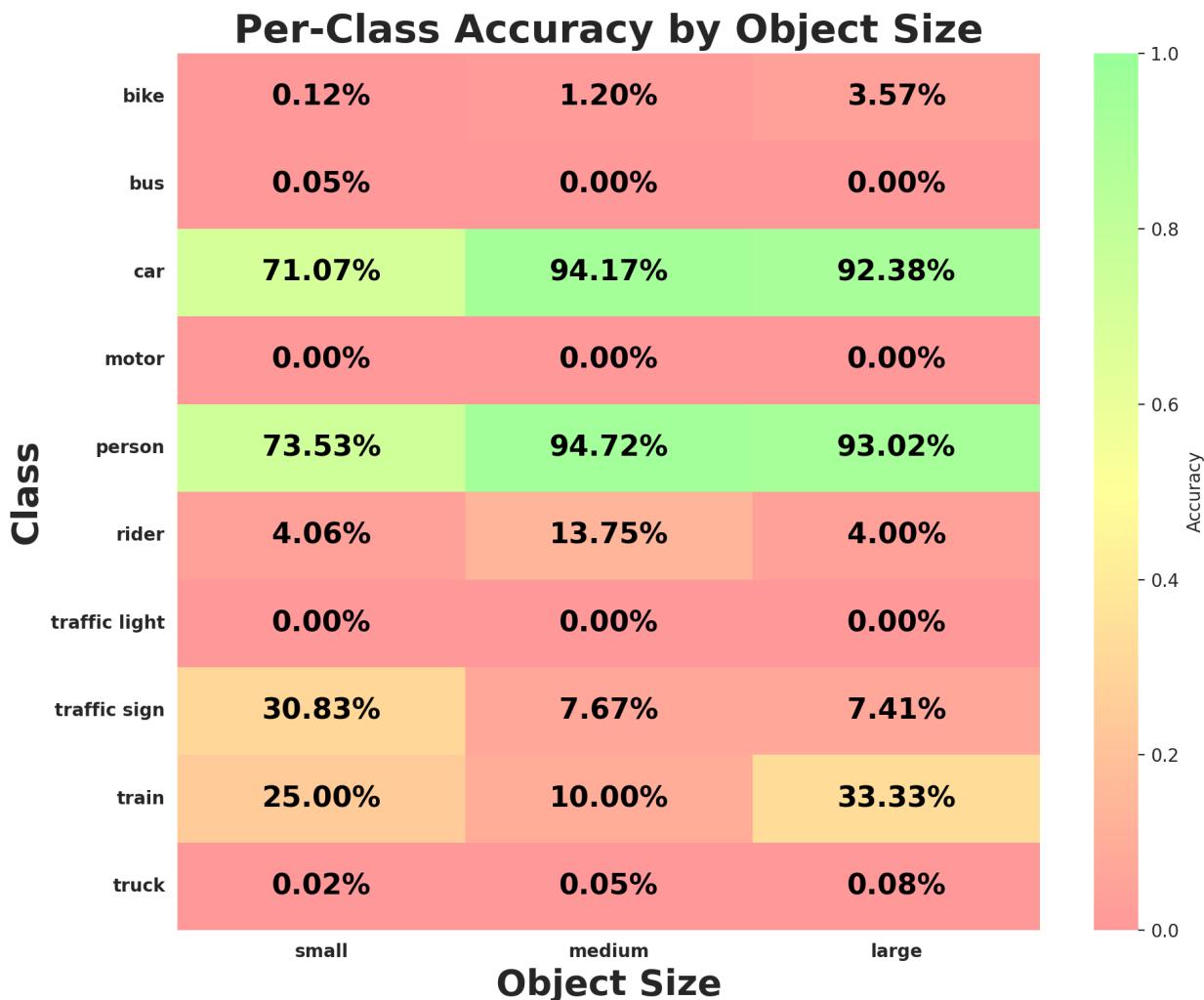


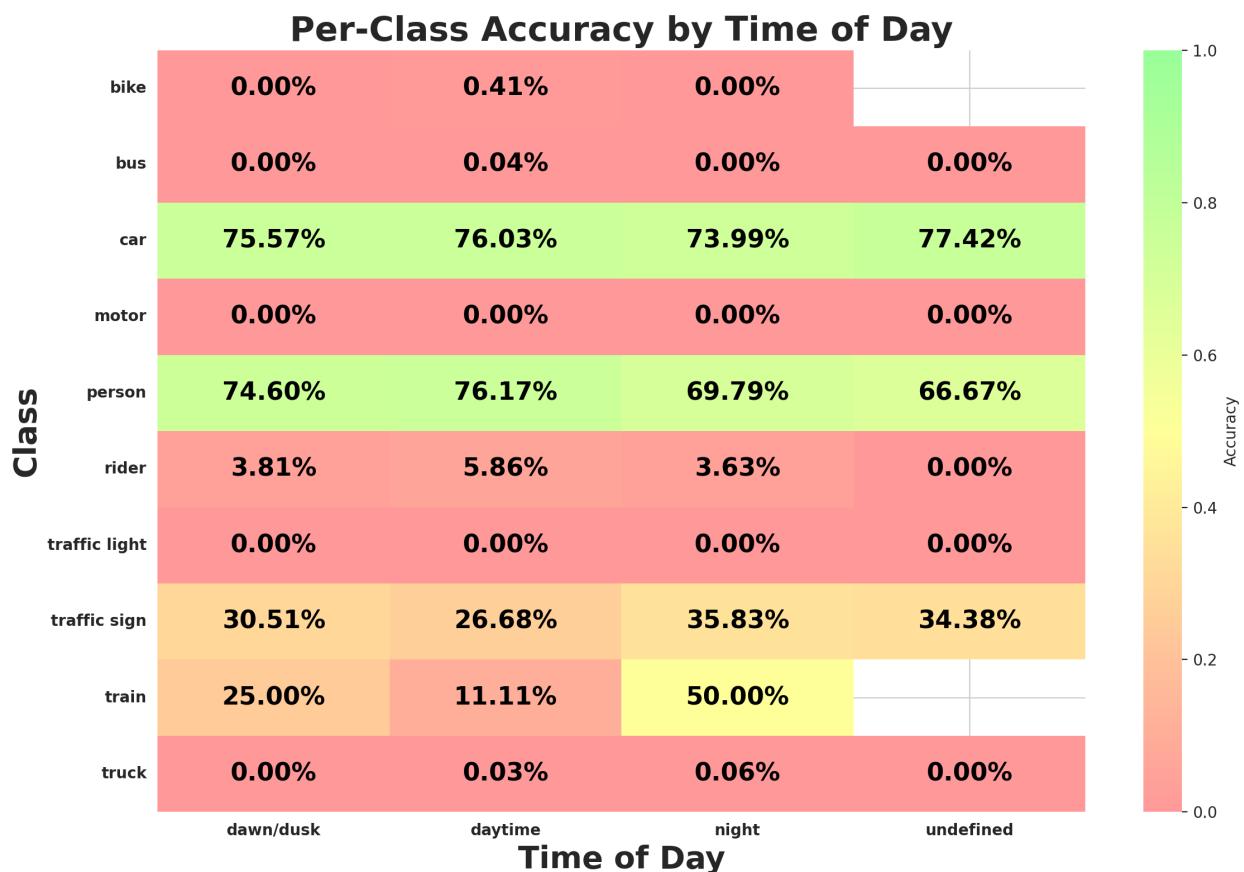
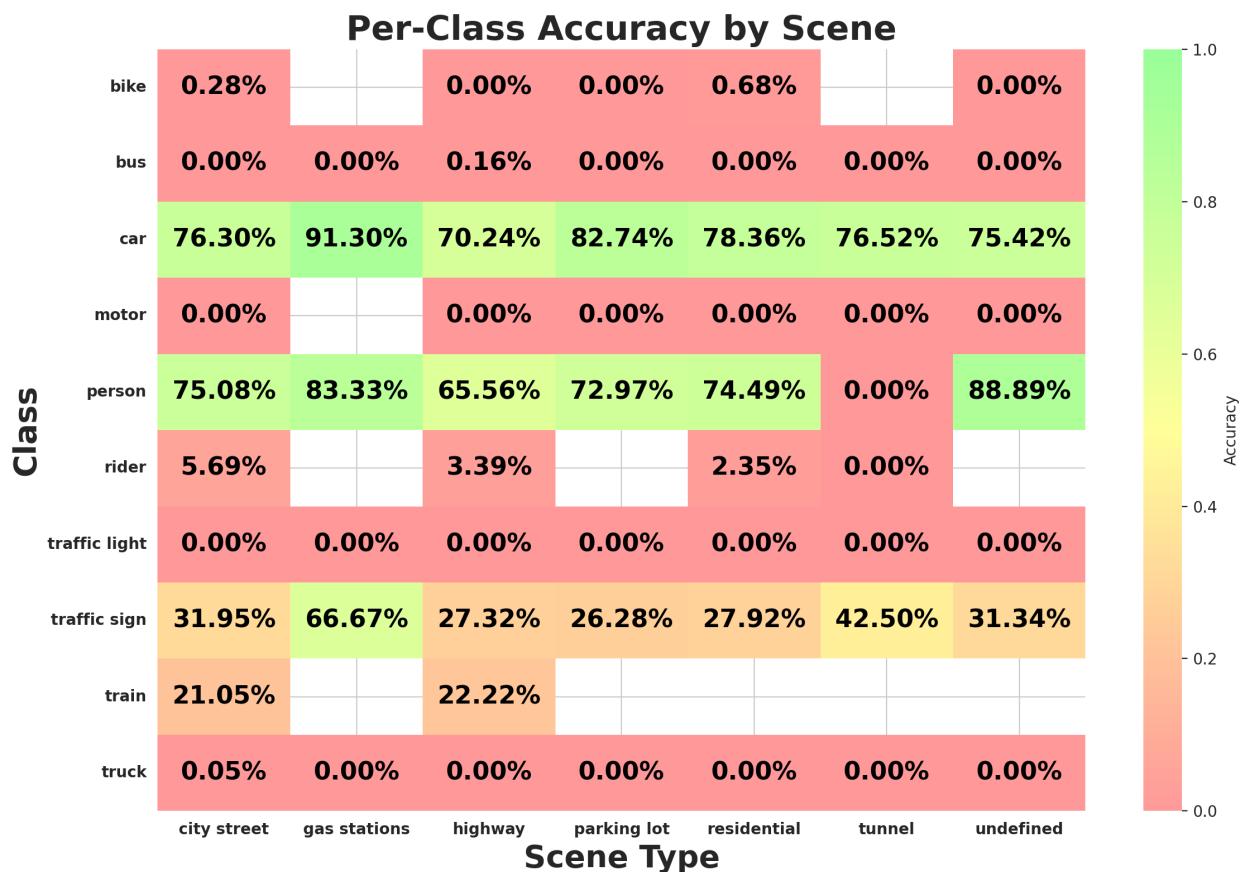
Prediction Accuracy by Object Size



Note: Object size categories are defined based on bounding box area as a percentage of total image area: **Small** objects occupy less than 1% of the image (typically distant or small objects), **Medium** objects occupy between 1% and 5% of the image (moderate-sized objects at medium distances), and **Large** objects occupy more than 5% of the image (close or large objects). These thresholds help identify how object scale affects detection accuracy.

Per-Class Performance Analysis





Sample Predictions: Ground Truth vs Model

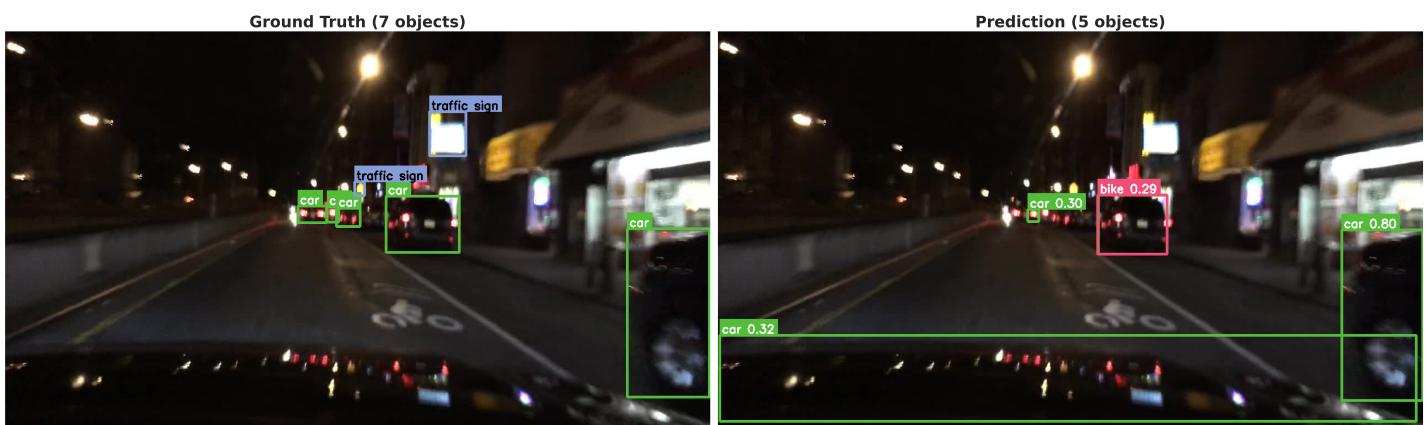
Sample #1: ce61357c-8e1b9a63.jpg

Ground Truth Objects: 28 | Predicted Objects: 14
Weather: clear | Scene: residential | Time of Day: daytime



Sample #2: dbc55d10-c6c15992.jpg

Ground Truth Objects: 7 | Predicted Objects: 5
Weather: clear | Scene: highway | Time of Day: night



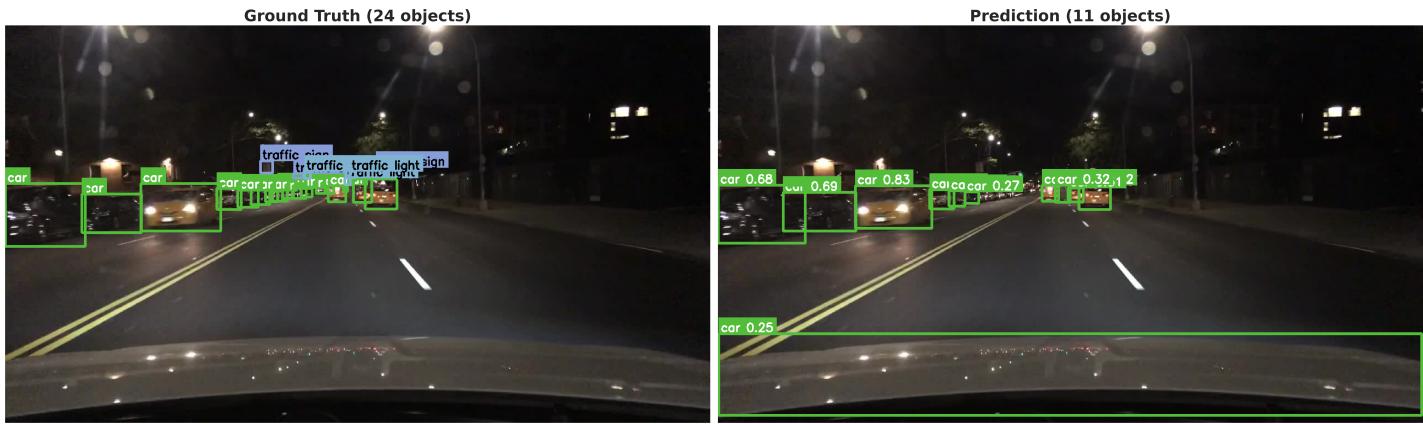
Sample #3: dc37c90c-91e4513a.jpg

Ground Truth Objects: 42 | Predicted Objects: 22
Weather: clear | Scene: city street | Time of Day: daytime



Sample #4: e5702ae4-754312e0.jpg

Ground Truth Objects: 24 | Predicted Objects: 11
Weather: clear | Scene: residential | Time of Day: daytime



Sample #5: f25bceab-ad2ea91b.jpg

Ground Truth Objects: 25 | Predicted Objects: 14

Weather: partly cloudy | Scene: city street | Time of Day: daytime



Sample #6: fcd3dc6b-832307c4.jpg

Ground Truth Objects: 25 | Predicted Objects: 13

Weather: undefined | Scene: city street | Time of Day: daytime



Generated by YOLO Testing Pipeline