

YOLO Model Testing Report

Model:	yolov9s
Model Size:	14.7 MB
Parameters:	7.3 M
FLOPs (640x640):	27.56 GFLOPs
Run Name:	yolov9s_testing_20251127_222004
Timestamp:	2025-11-27 22:35:20
Dataset:	bdd100k_yolo_limited - test split
Images Processed:	20000
Total Execution Time	107.38s
IoU Threshold:	0.5

Inference Performance

Metric	Value
Average Inference Time	2.02 ms per image
FPS (Frames Per Second)	495.74

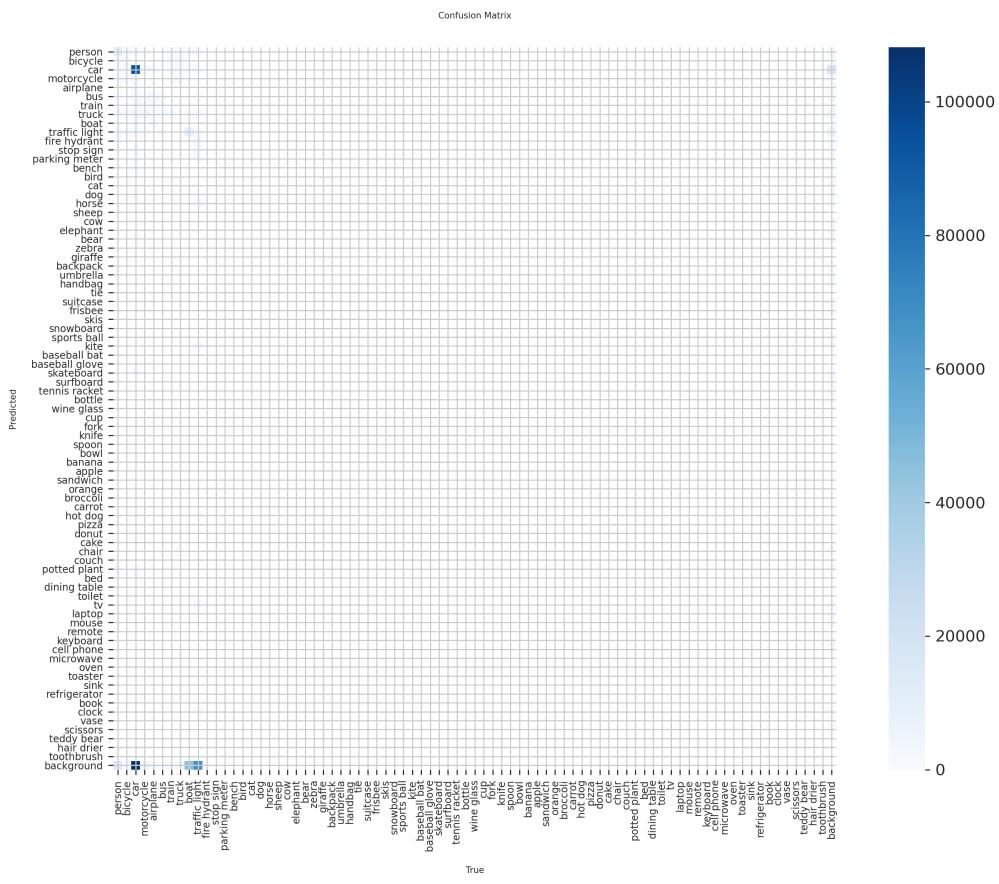
Note: Inference time includes preprocessing, model inference, and postprocessing.

Overall Accuracy Metrics

Metric	Value
Precision	0.2817
Recall	0.6791
F1-Score	0.3982
mAP@0.5	0.1106
mAP@0.5:0.95	0.0596
True Positives	103598
False Positives	264129
False Negatives	48953

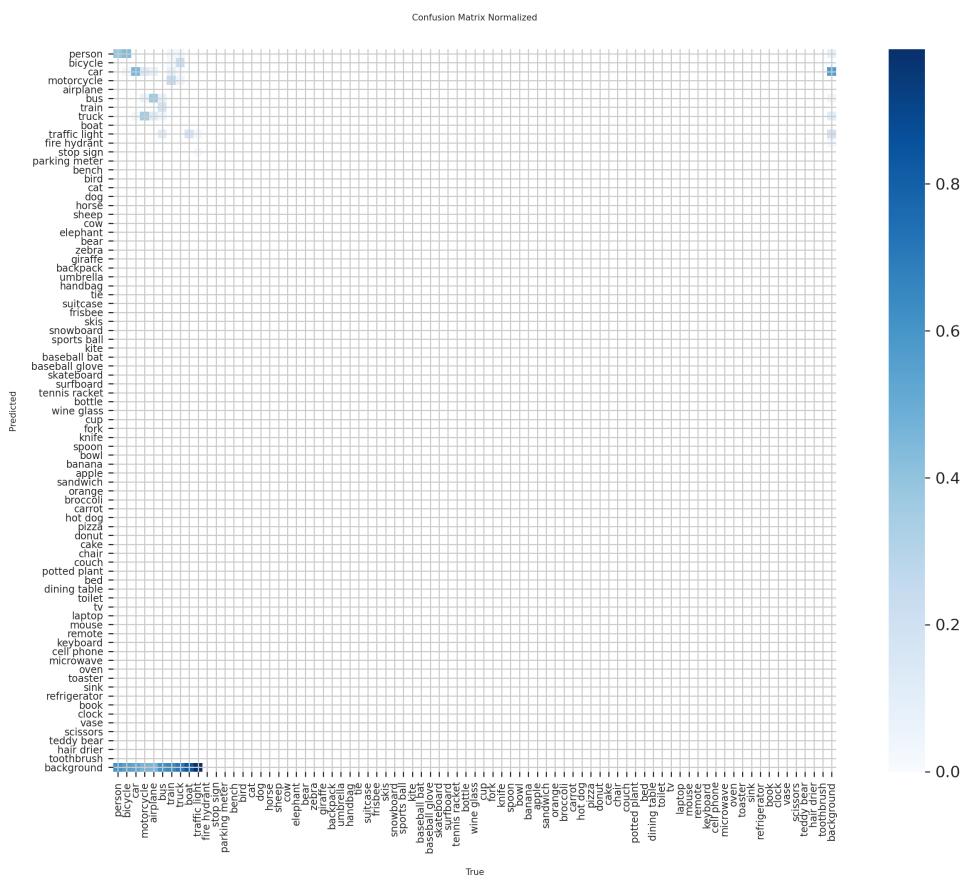
Confusion Matrices

Confusion Matrix (Counts)

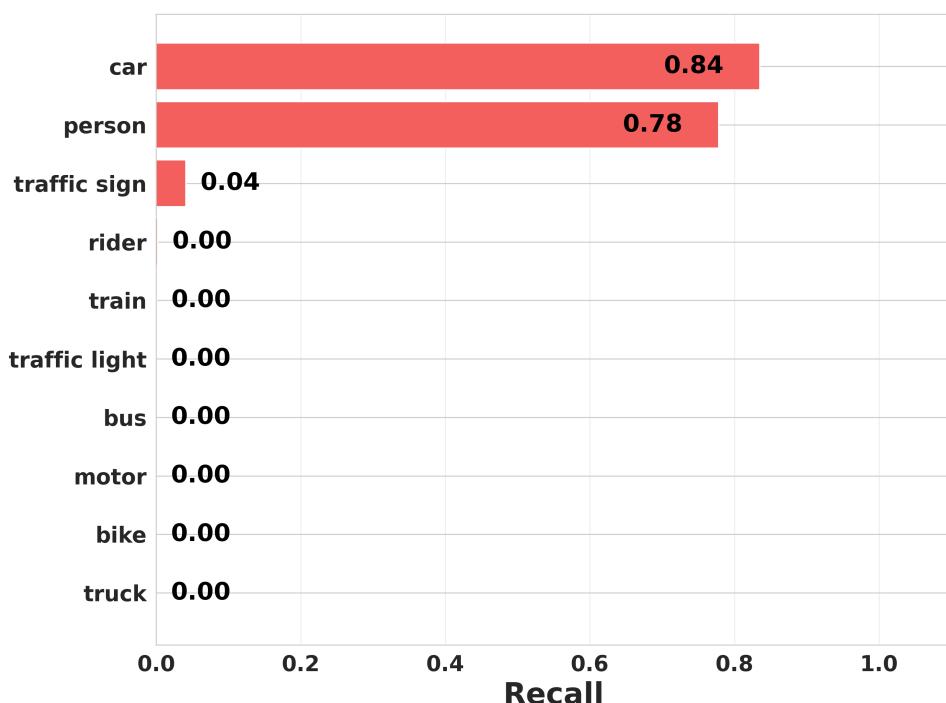


Note: Shows raw prediction counts for each class combination.

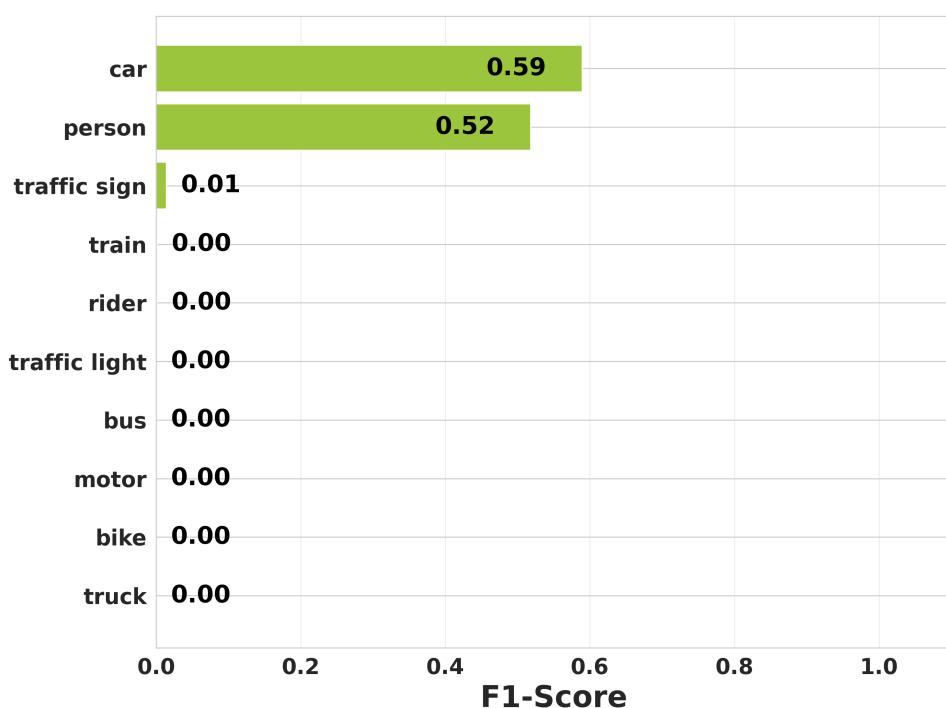
Confusion Matrix (Normalized)



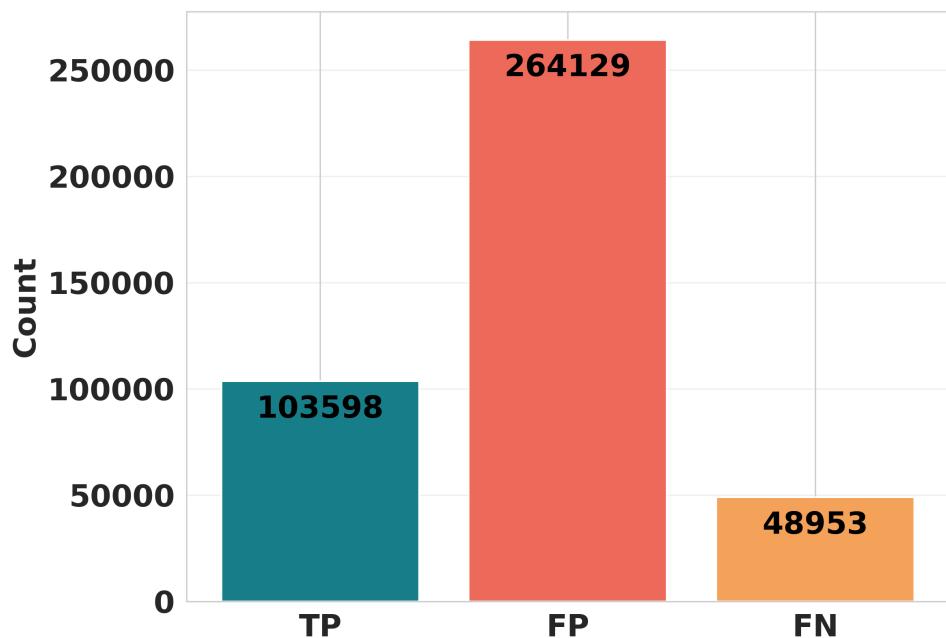
Recall by Class



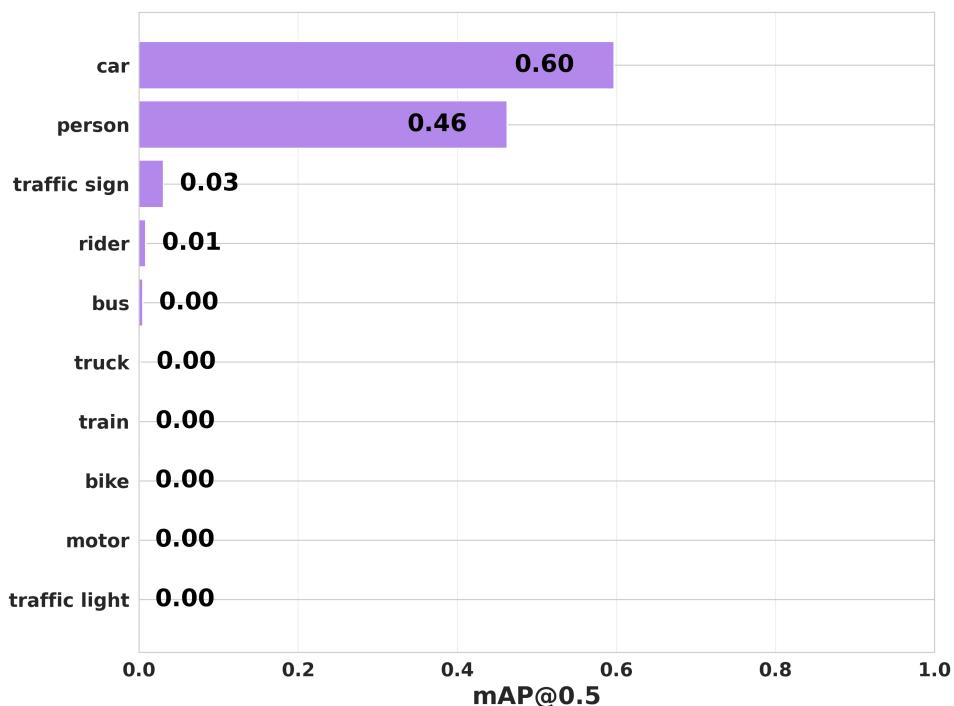
F1-Score by Class



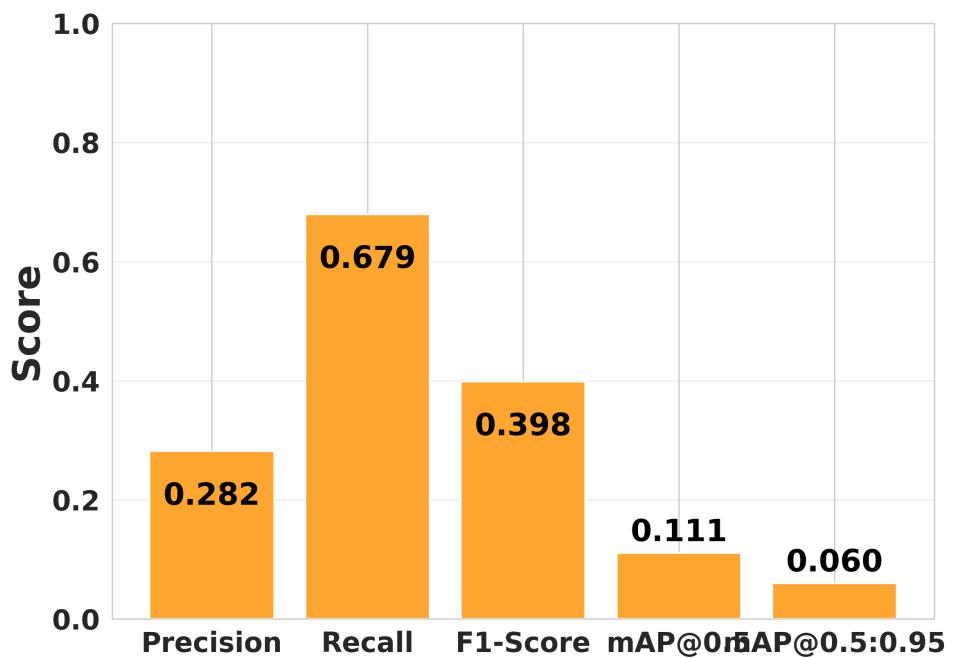
Overall Detection Outcomes



mAP@0.5 by Class



Overall Metrics



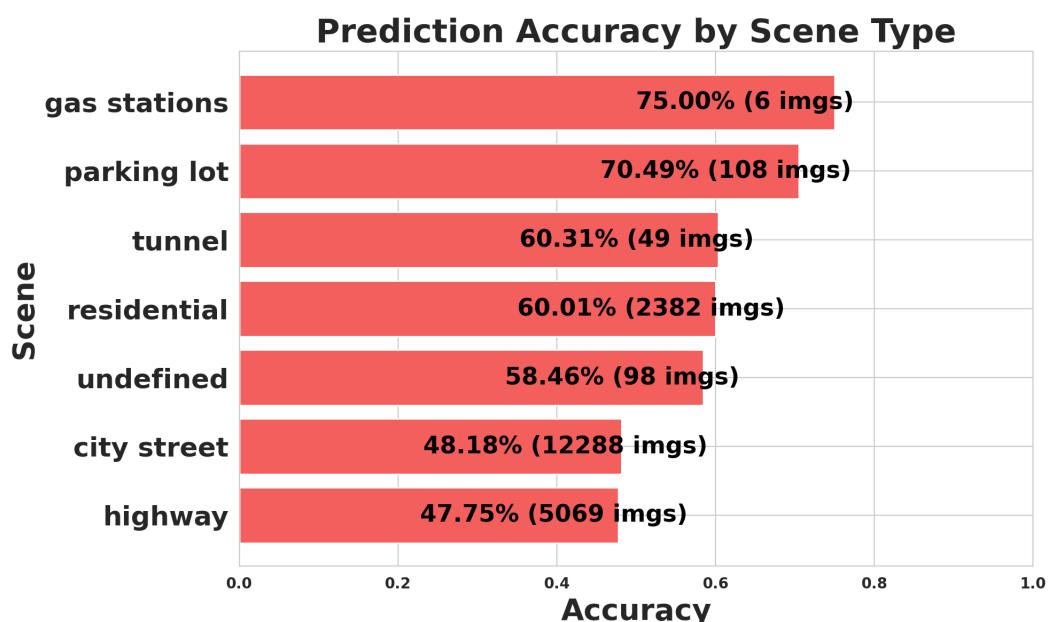
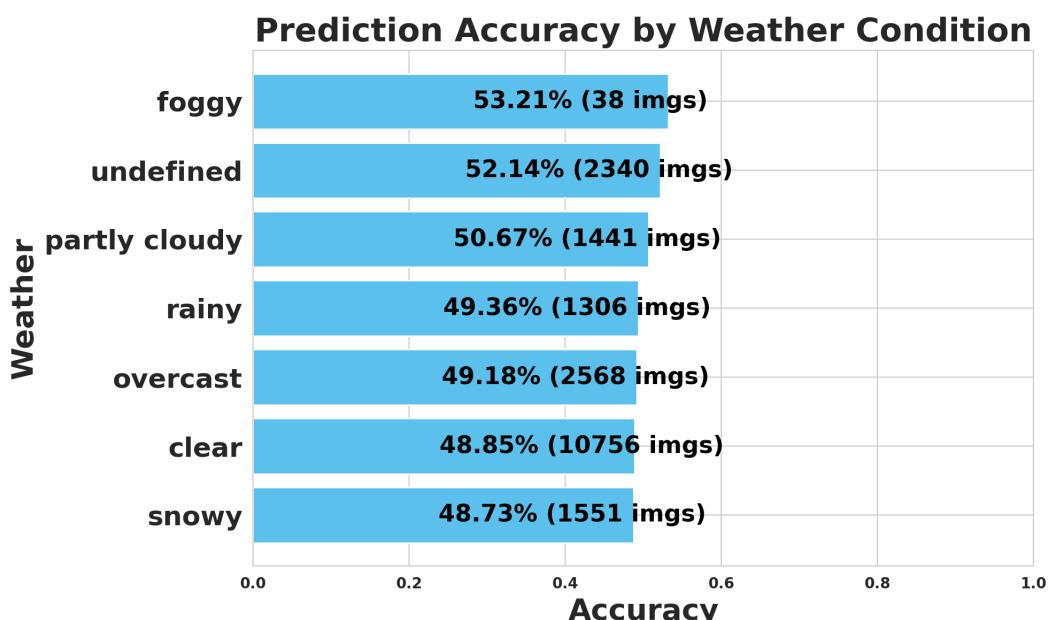
Per-Class Performance

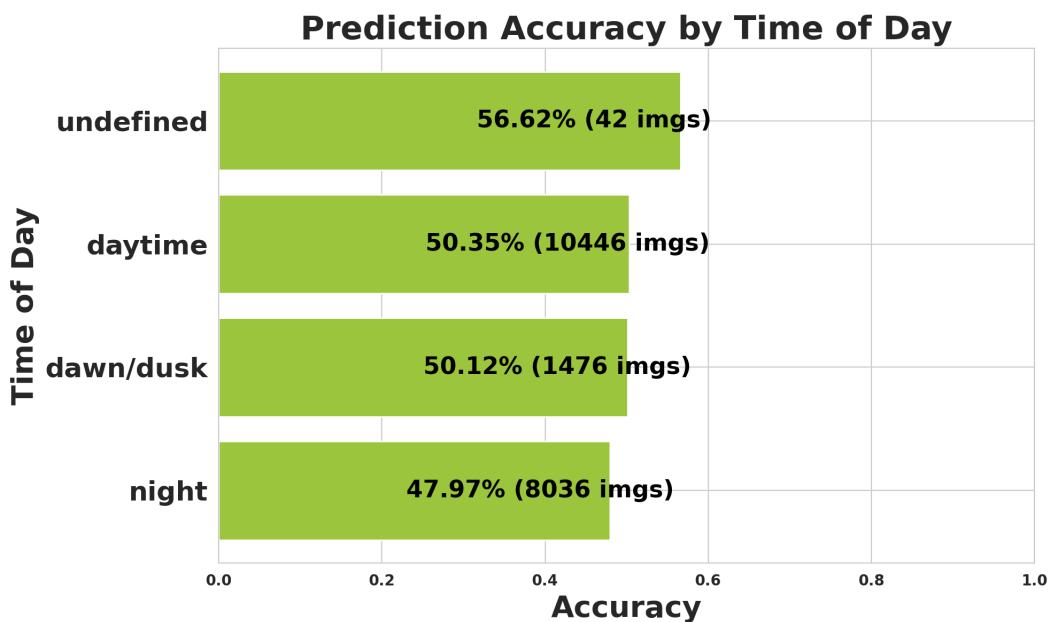
Class	TP	FP	FN	Precision	Recall	F1-Score	mAP@0.5
person	9583	15067	2730	0.3888	0.7783	0.5185	0.4627
rider	1	1293	535	0.0008	0.0019	0.0011	0.0084
car	93425	111724	18413	0.4554	0.8354	0.5895	0.5972
truck	0	8704	433	0.0000	0.0000	0.0000	0.0015
bus	0	3217	20	0.0000	0.0000	0.0000	0.0049
train	2	26	3358	0.0714	0.0006	0.0012	0.0003
motor	0	841	177	0.0000	0.0000	0.0000	0.0001
bike	0	1998	9531	0.0000	0.0000	0.0000	0.0002
traffic light	0	52840	4	0.0000	0.0000	0.0000	0.0000
traffic sign	587	68419	13752	0.0085	0.0409	0.0141	0.0309

Comprehensive Failure Analysis

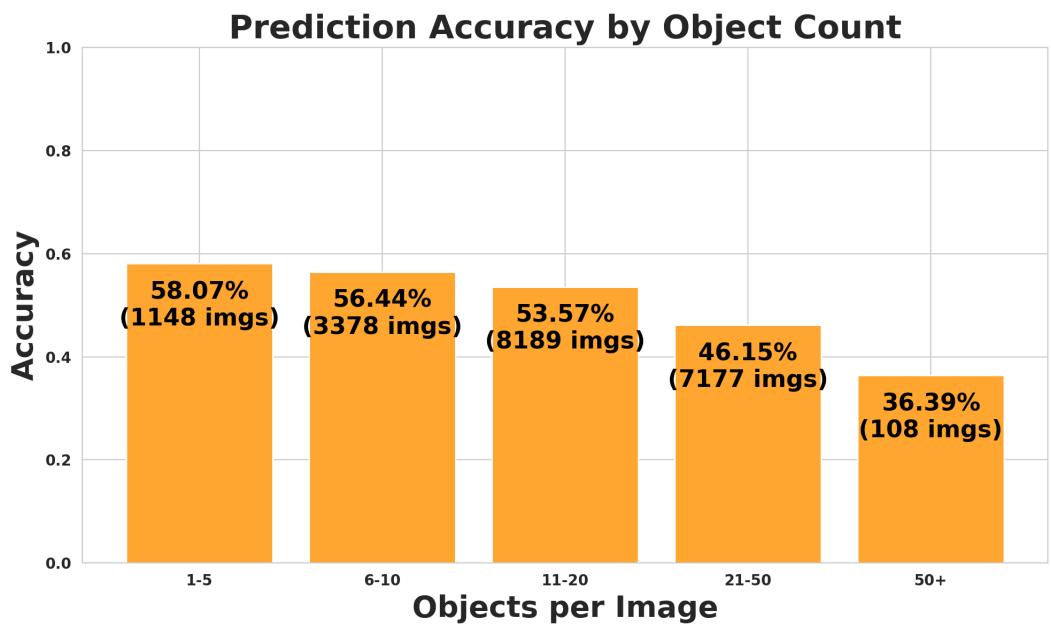
Metric	Value
Total Images Analyzed	20000
Total Expected Objects	367728
Total Matched Objects	181971
Overall Accuracy	49.49%

Accuracy by Environmental Attributes

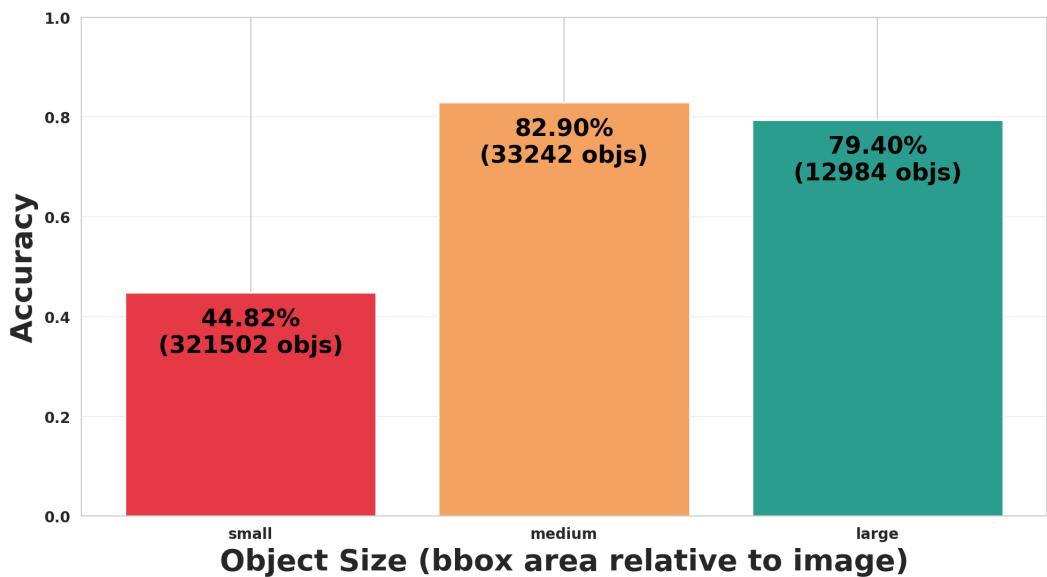




Accuracy by Object Characteristics

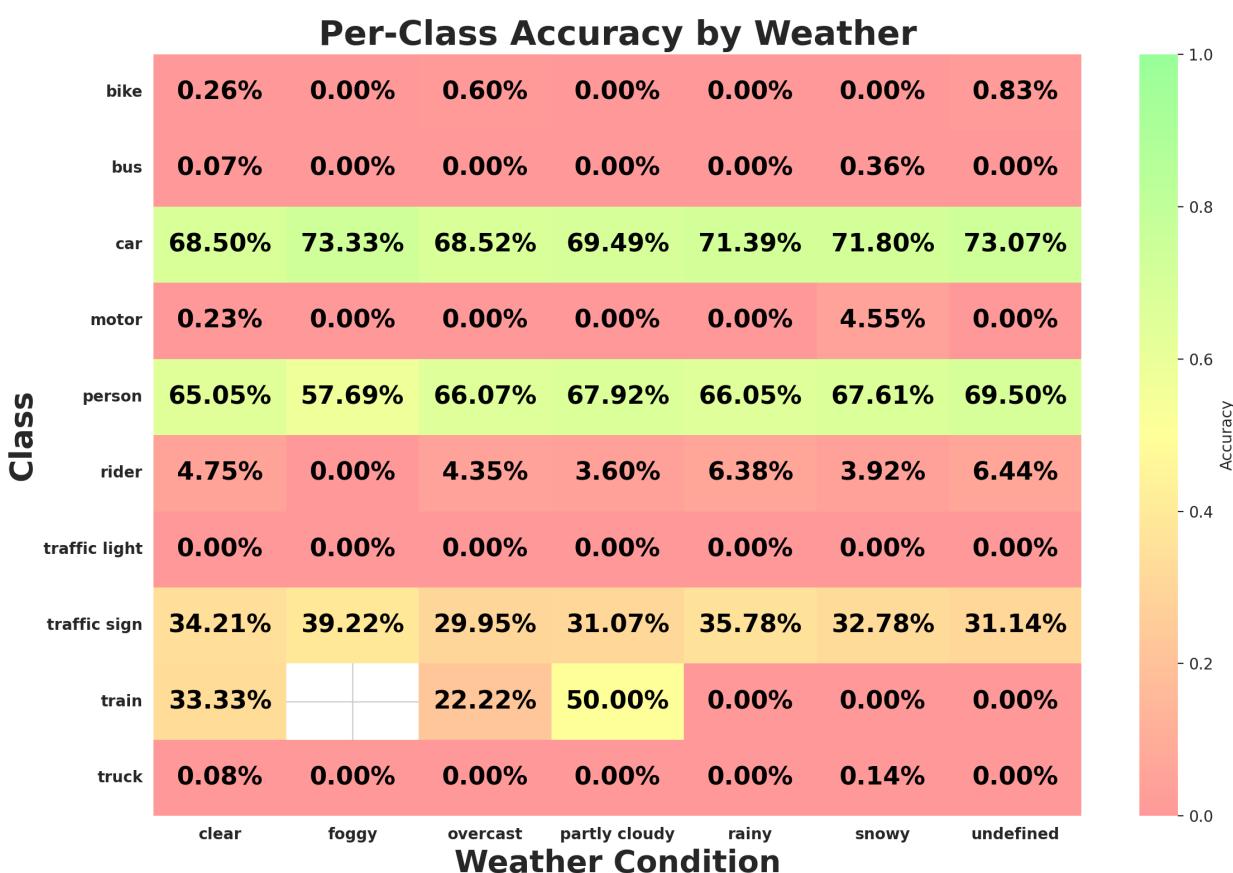
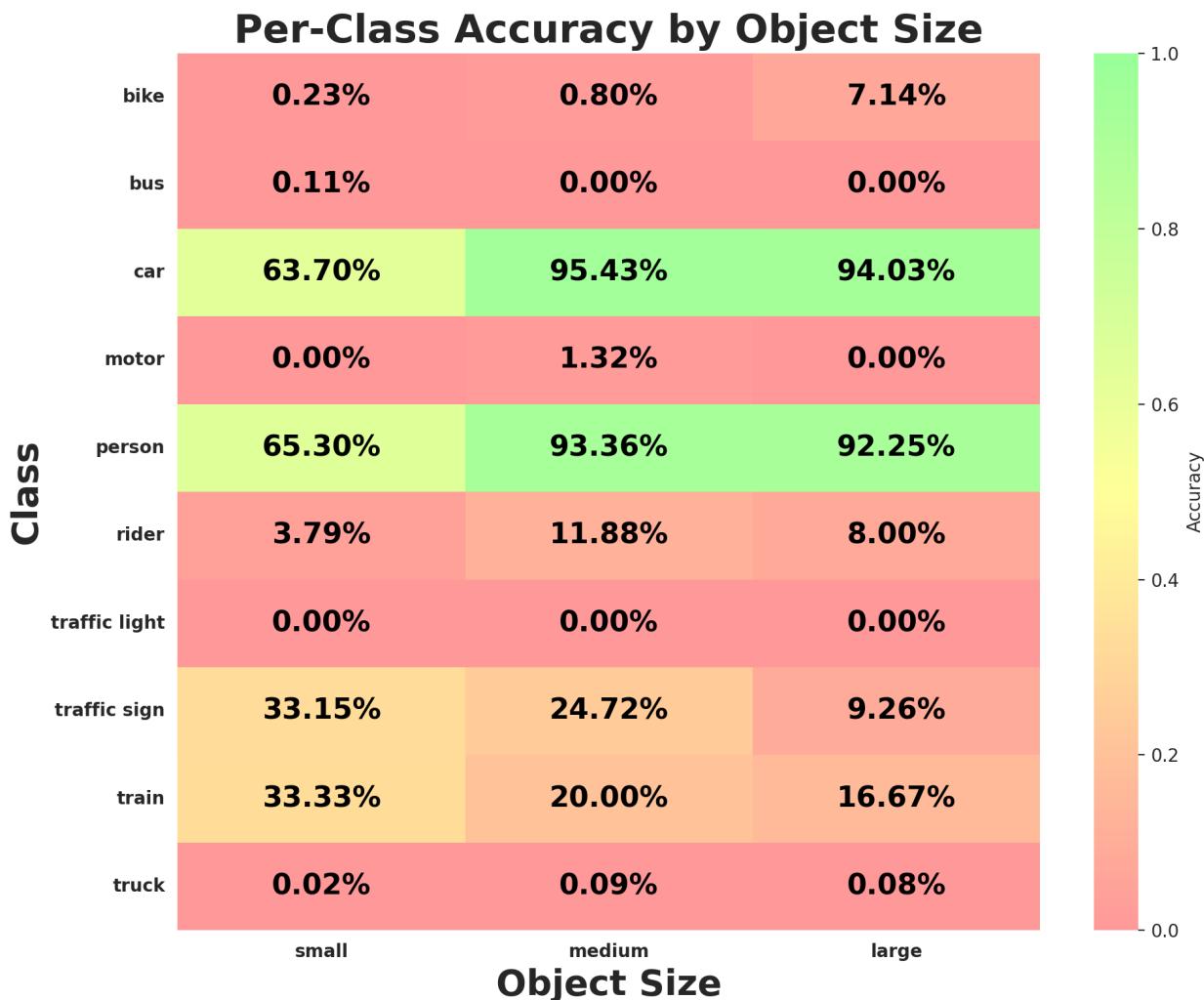


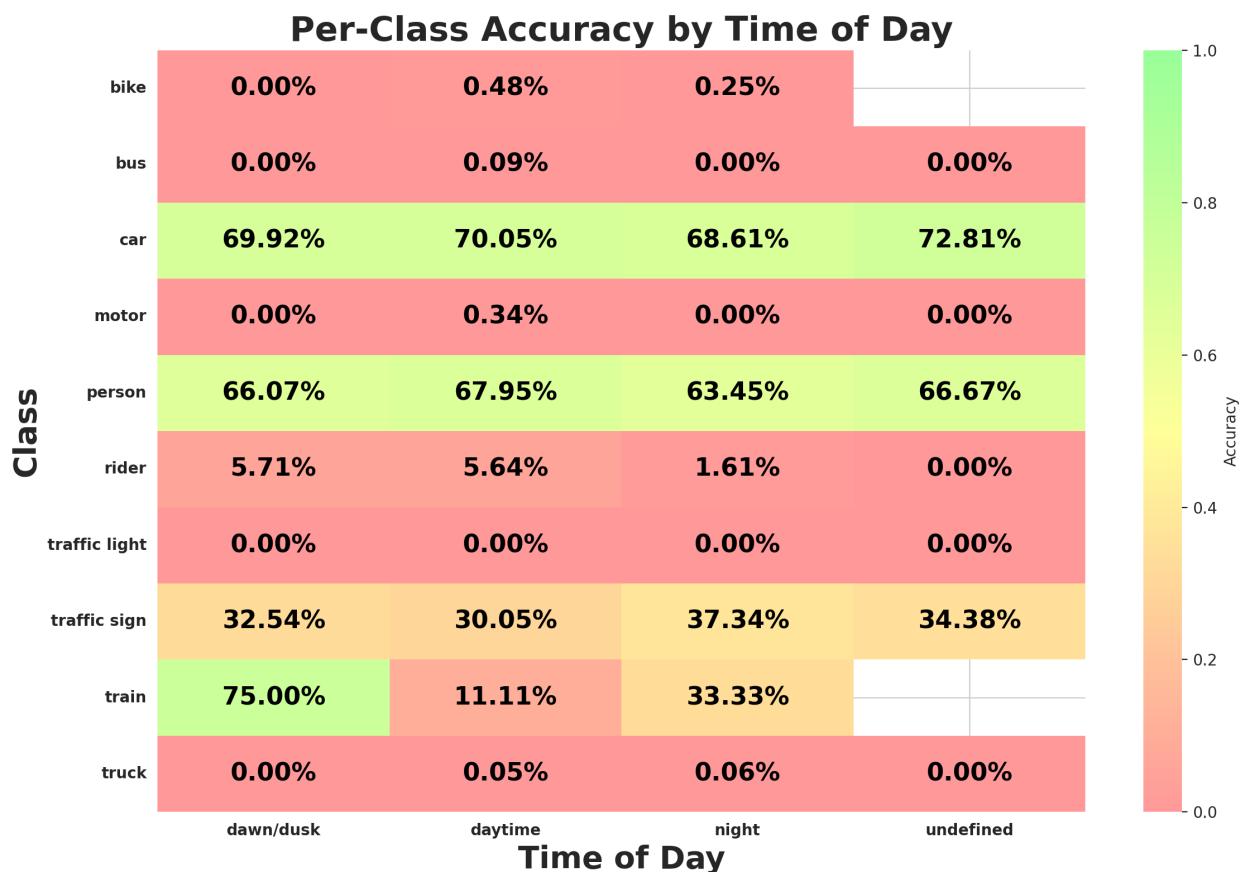
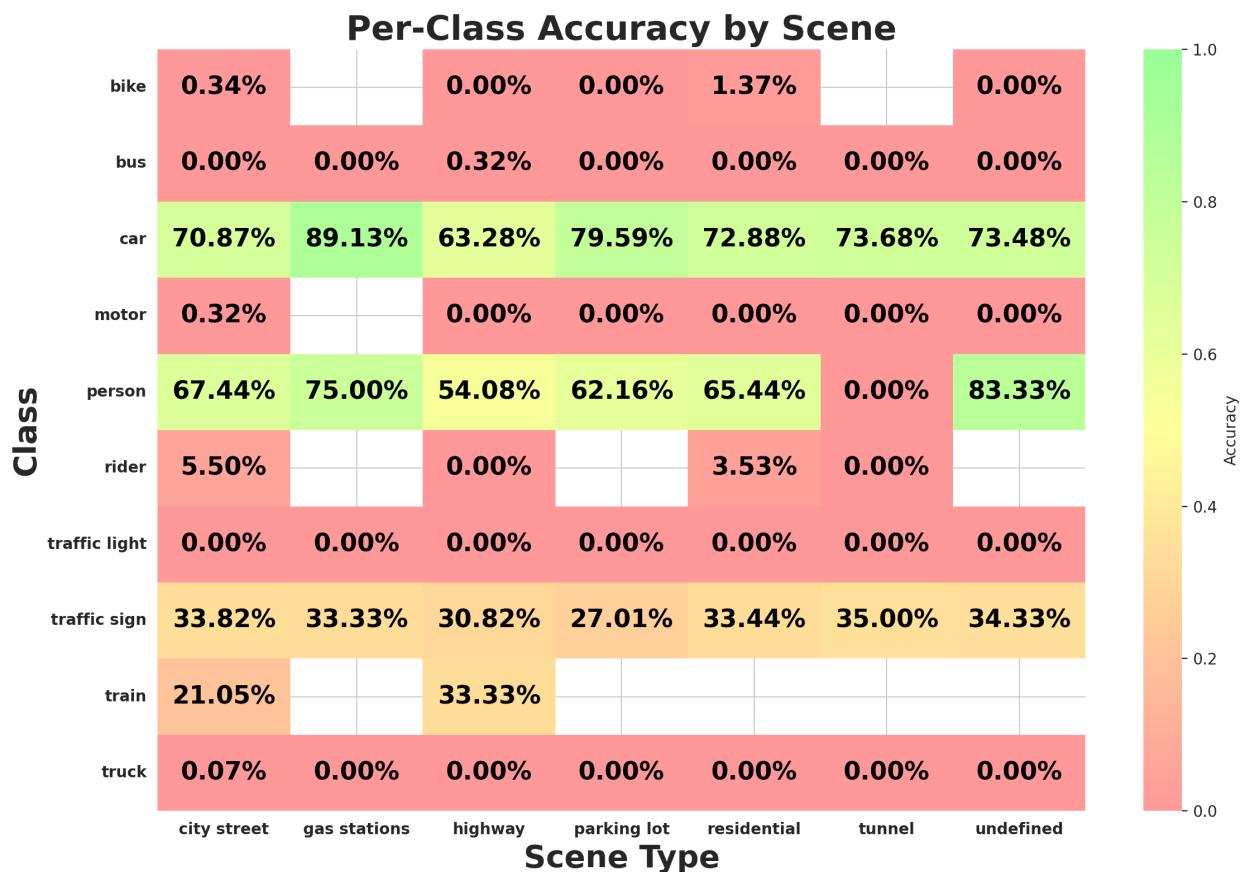
Prediction Accuracy by Object Size



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

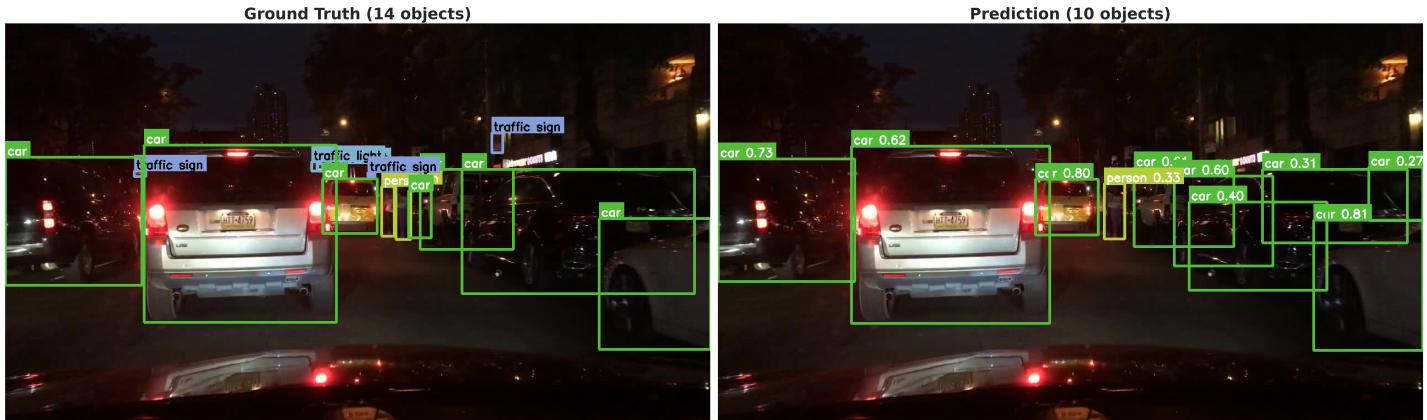




Sample Predictions: Ground Truth vs Model

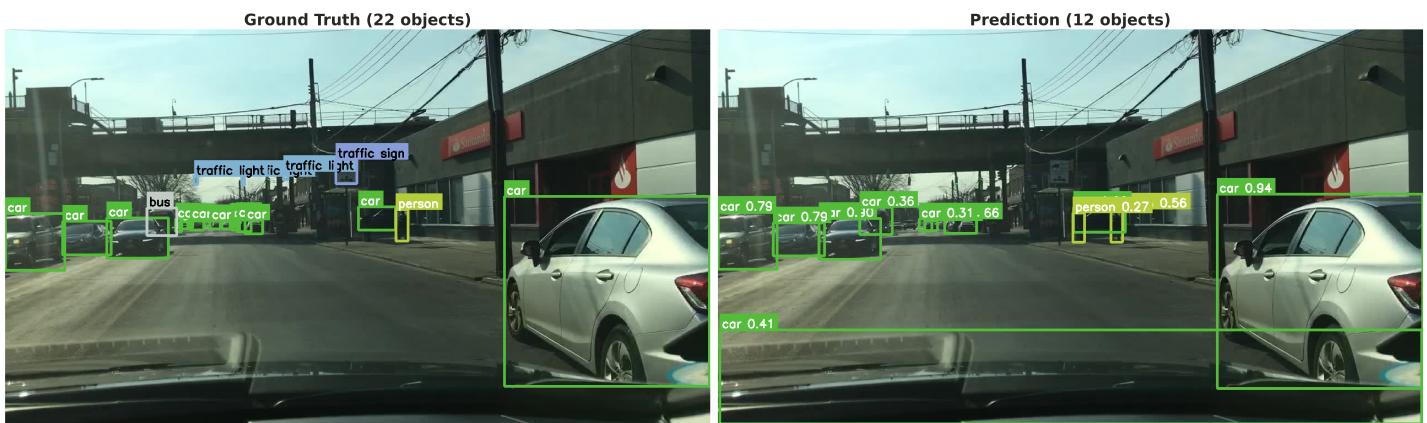
Sample #1: f3dc8936-15a9c200.jpg

Ground Truth Objects: 14 | Predicted Objects: 10
Weather: clear | Scene: city street | Time of Day: night



Sample #2: fbcb233b-4810d2aa.jpg

Ground Truth Objects: 22 | Predicted Objects: 12
Weather: overcast | Scene: city street | Time of Day: daytime



Sample #3: decd56ae-780814b9.jpg

Ground Truth Objects: 29 | Predicted Objects: 11
Weather: overcast | Scene: city street | Time of Day: daytime



Sample #4: f1205e35-f2682974.jpg

Ground Truth Objects: 24 | Predicted Objects: 10
Weather: snowy | Scene: city street | Time of Day: daytime



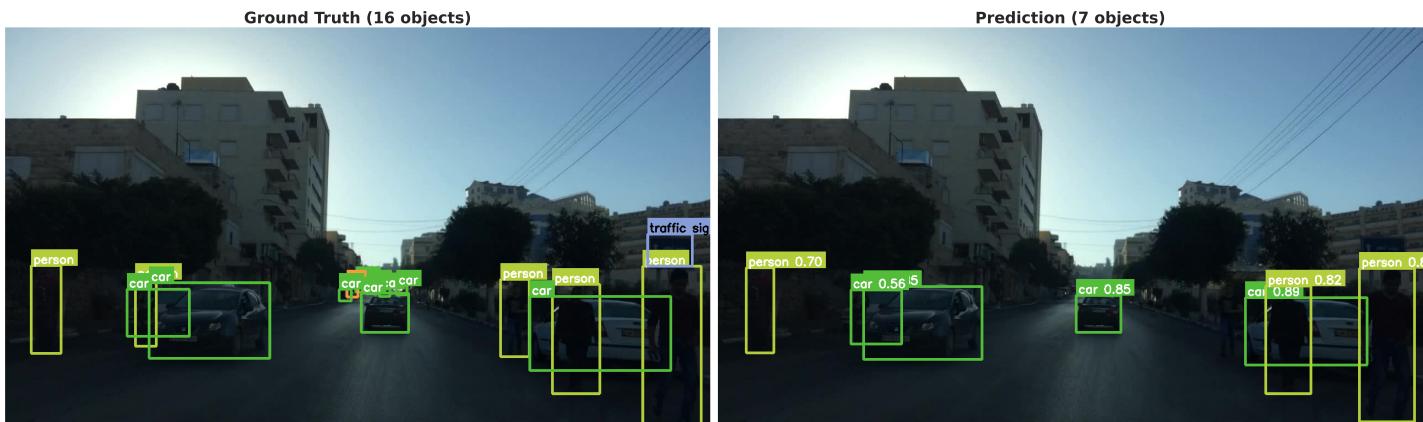
Sample #5: d12a20bb-791f070b.jpg

Ground Truth Objects: 9 | Predicted Objects: 3
Weather: clear | Scene: highway | Time of Day: night



Sample #6: e8a5392b-90fcc4cf.jpg

Ground Truth Objects: 16 | Predicted Objects: 7
Weather: clear | Scene: city street | Time of Day: dawn/dusk



Generated by YOLO Testing Pipeline