

YOLO Model Testing Report

Model:	yolov8l
Model Size:	83.7 MB
Parameters:	43.7 M
FLOPs (640x640):	165.74 GFLOPs
Run Name:	yolov8l_testing_20251127_213827
Timestamp:	2025-11-27 21:51:02
Dataset:	bdd100k_yolo_limited - test split
Images Processed:	20000
Total Execution Time	126.30s
IoU Threshold:	0.5

Inference Performance

Metric	Value
Average Inference Time	3.08 ms per image
FPS (Frames Per Second)	324.64

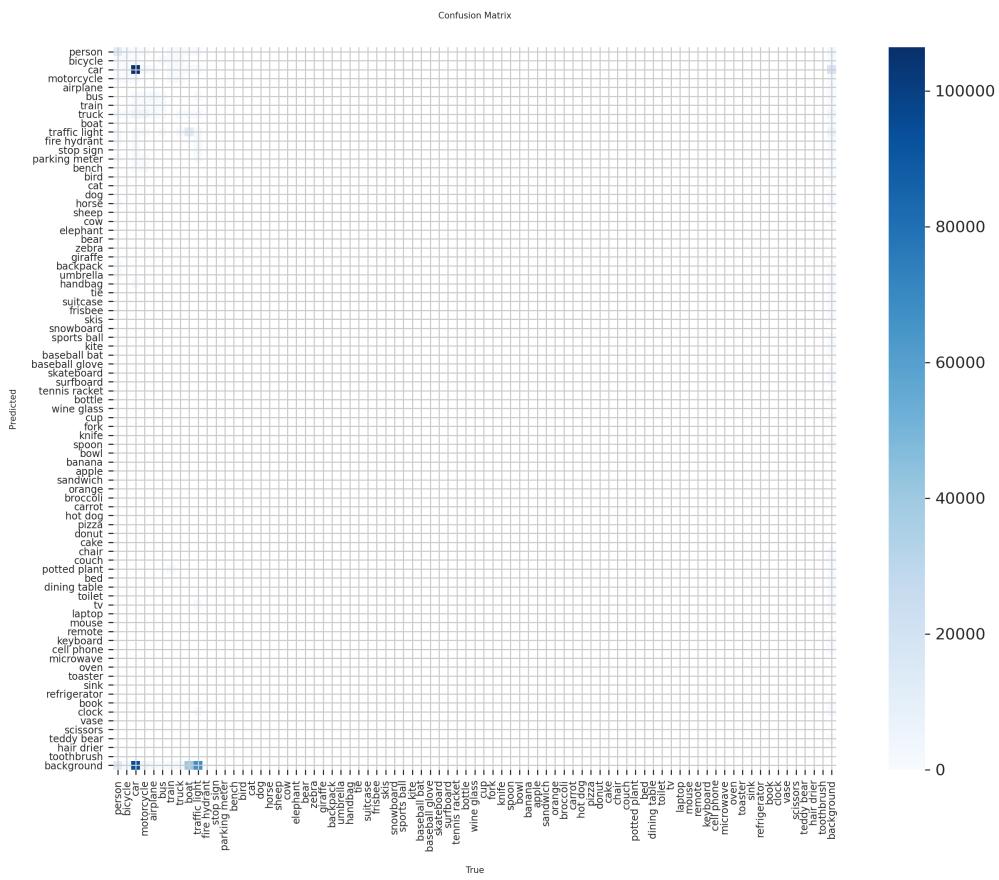
Note: Inference time includes preprocessing, model inference, and postprocessing.

Overall Accuracy Metrics

Metric	Value
Precision	0.3225
Recall	0.6807
F1-Score	0.4376
mAP@0.5	0.1269
mAP@0.5:0.95	0.0700
True Positives	118586
False Positives	249141
False Negatives	55623

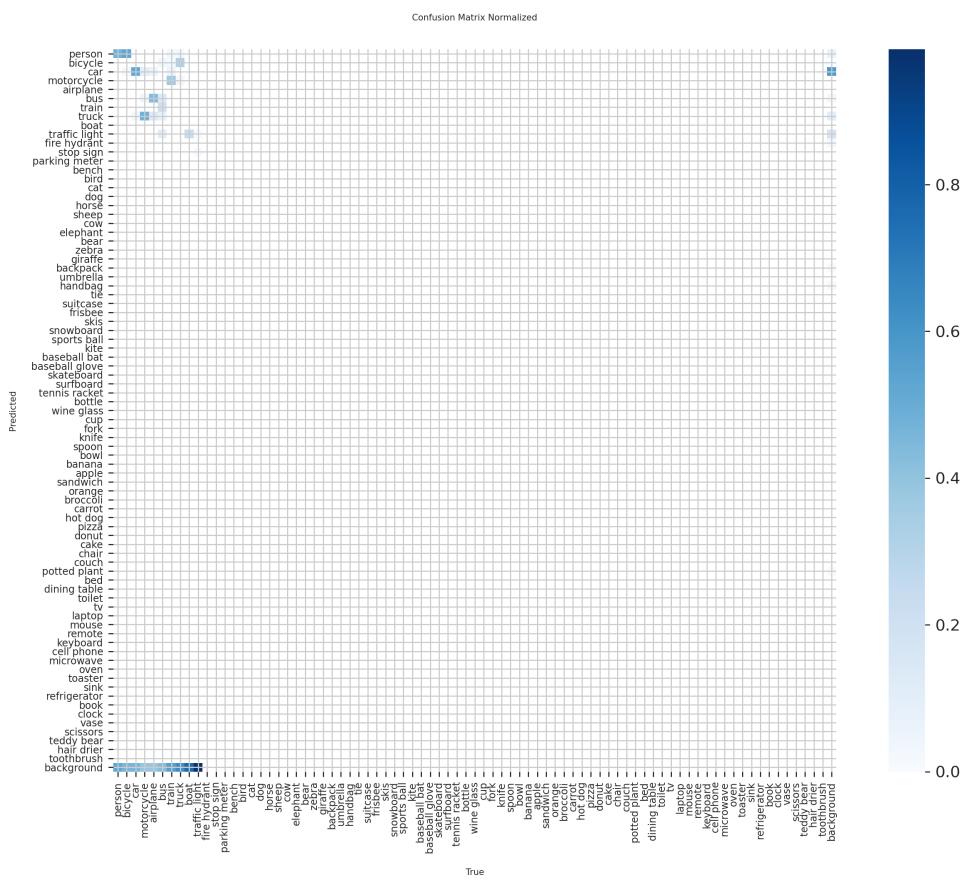
Confusion Matrices

Confusion Matrix (Counts)



Note: Shows raw prediction counts for each class combination.

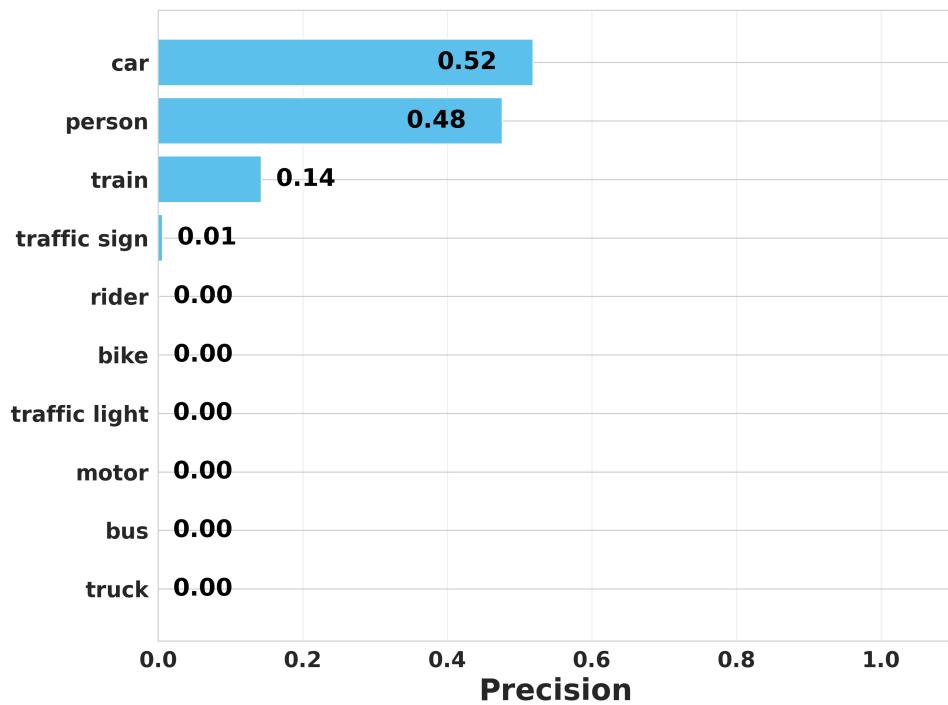
Confusion Matrix (Normalized)



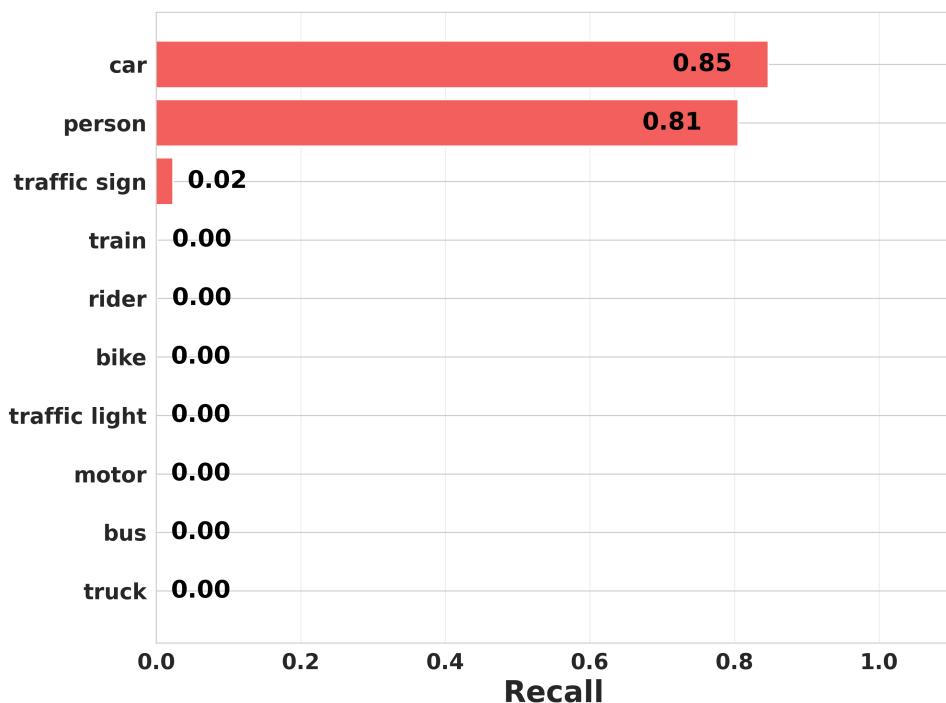
Note: Values are normalized by the number of ground truth instances per class. Shows the proportion of correct vs incorrect classifications.

Performance Visualizations

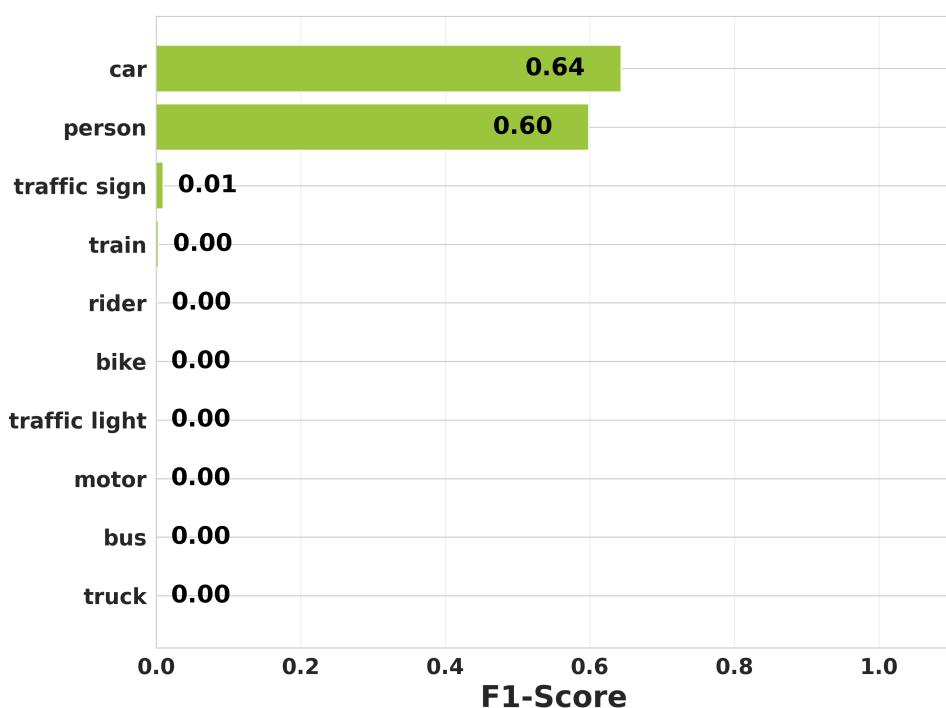
Precision by Class



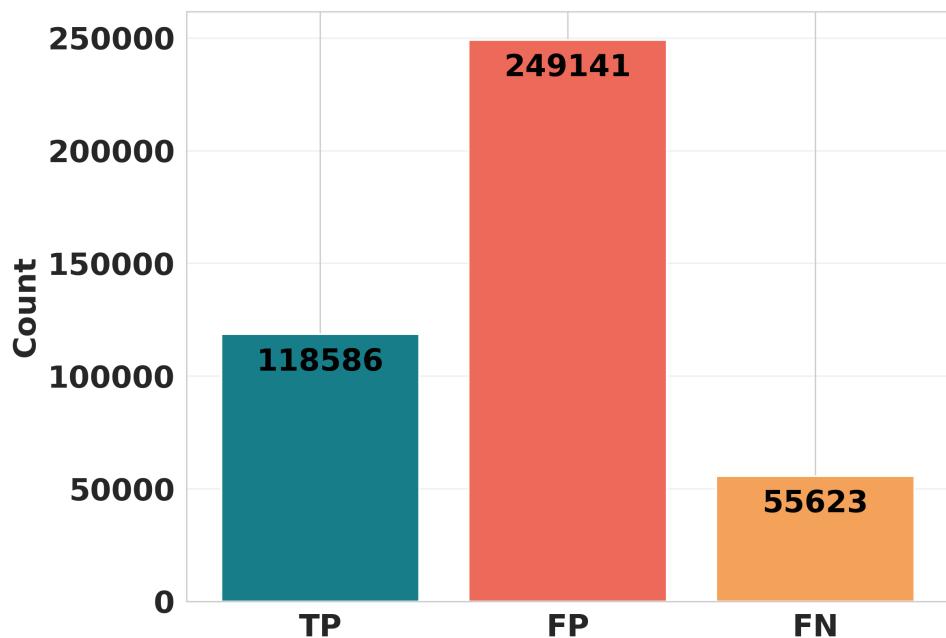
Recall by Class



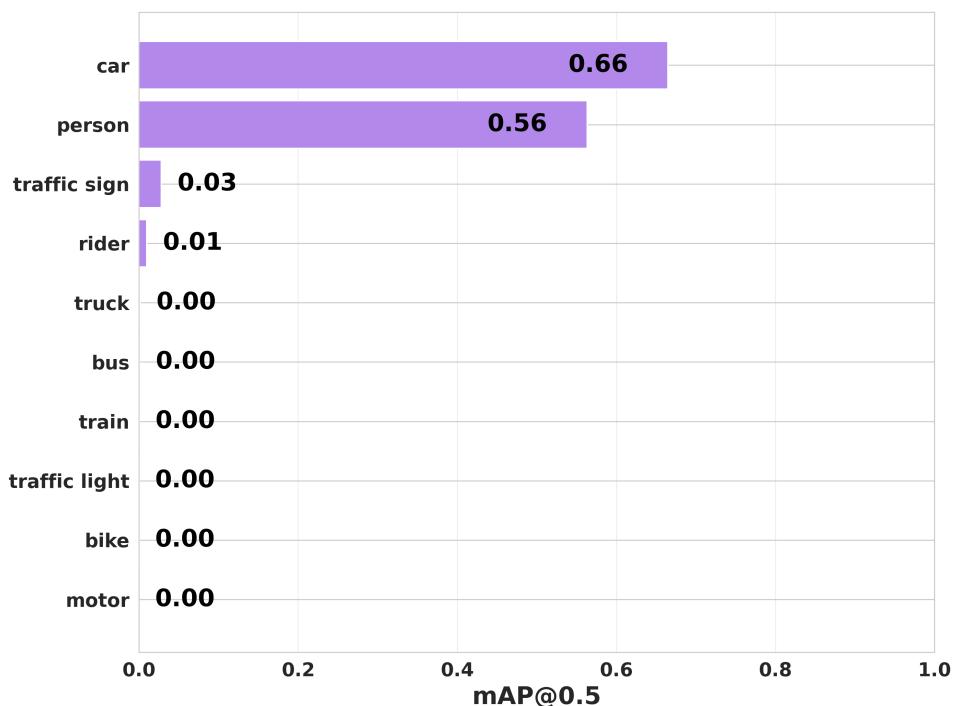
F1-Score by Class



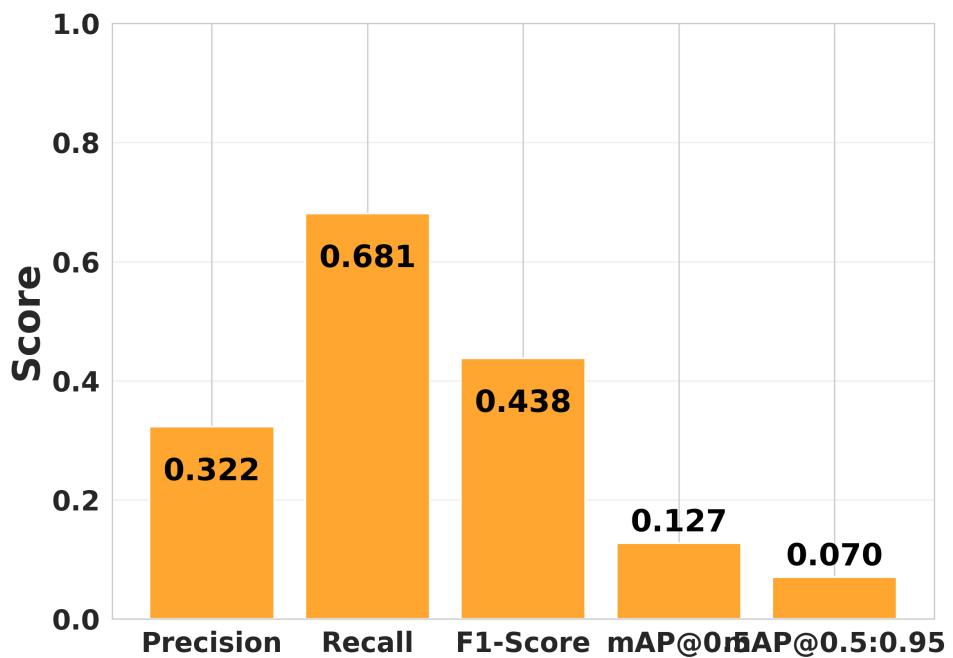
Overall Detection Outcomes



mAP@0.5 by Class



Overall Metrics



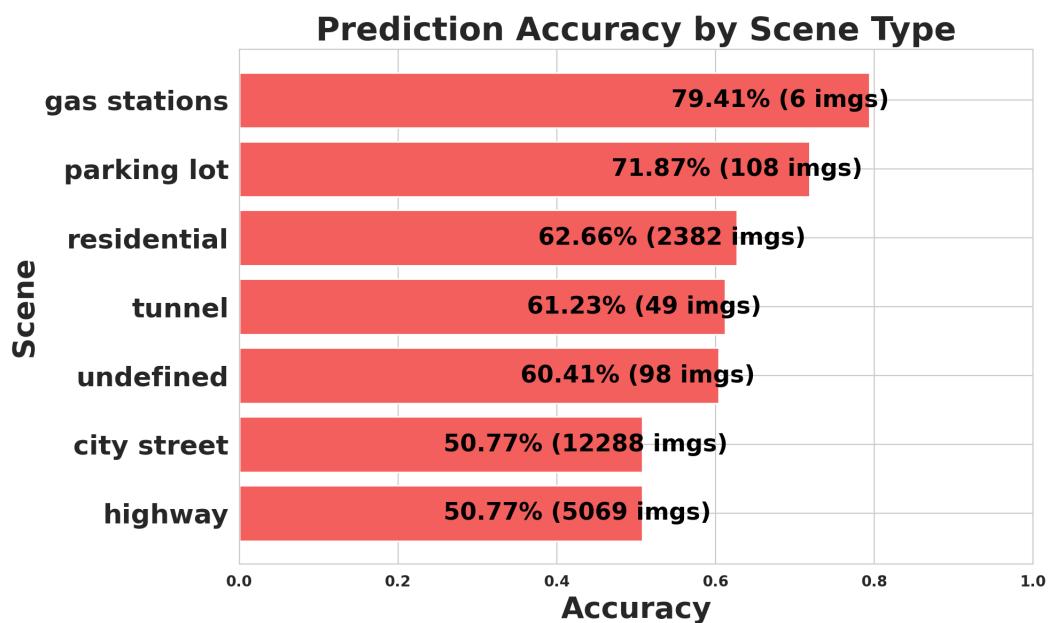
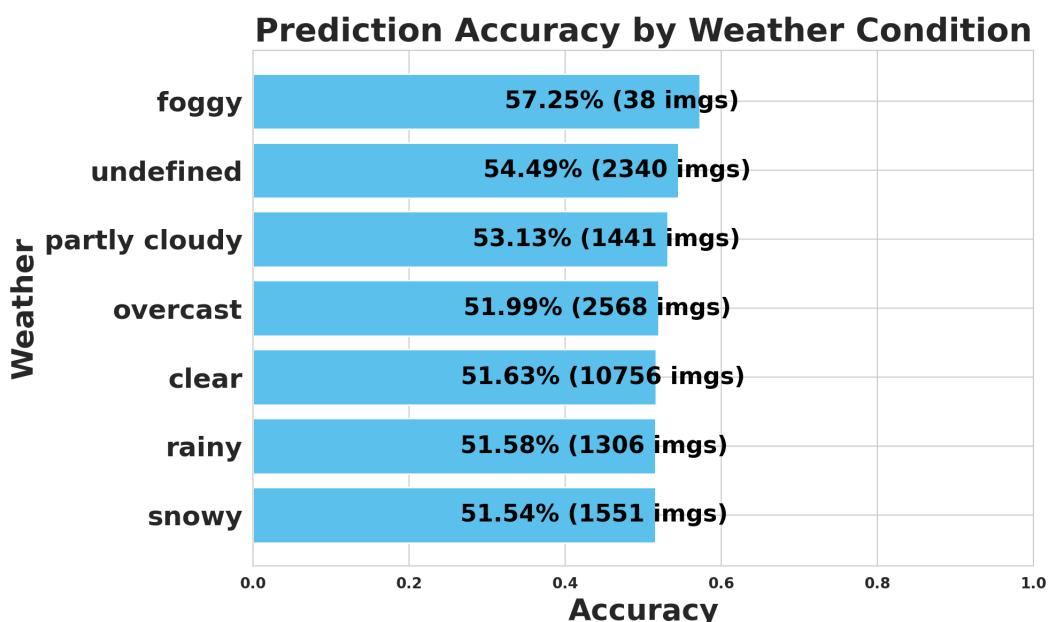
Per-Class Performance

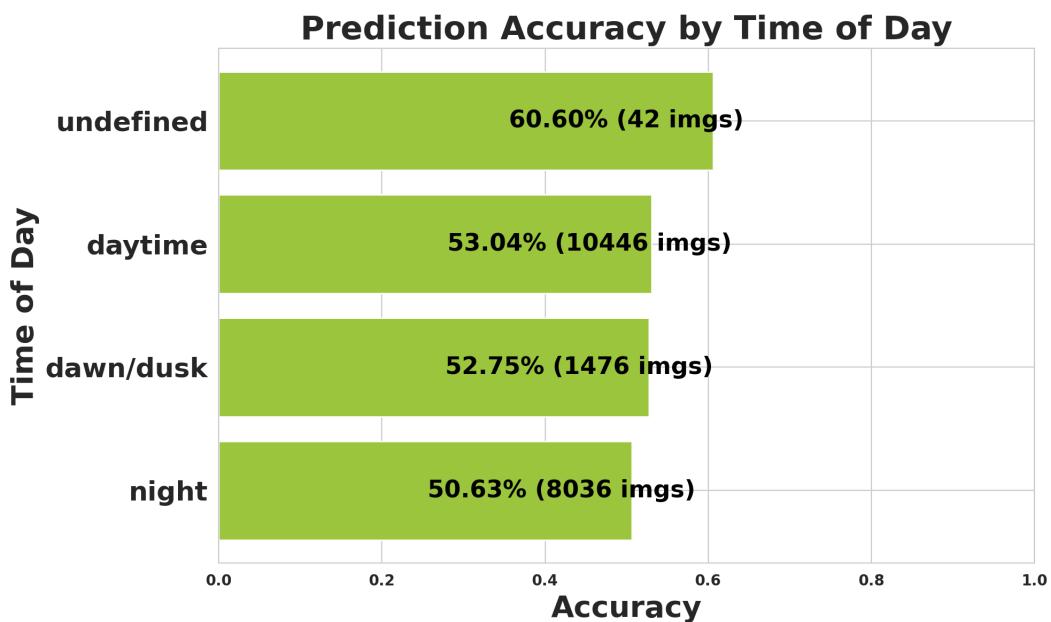
Class	TP	FP	FN	Precision	Recall	F1-Score	mAP@0.5
person	11740	12910	2838	0.4763	0.8053	0.5986	0.5633
rider	1	1293	783	0.0008	0.0013	0.0010	0.0098
car	106418	98731	19269	0.5187	0.8467	0.6433	0.6647
truck	0	8704	499	0.0000	0.0000	0.0000	0.0016
bus	0	3217	17	0.0000	0.0000	0.0000	0.0007
train	4	24	2751	0.1429	0.0015	0.0029	0.0005
motor	0	841	190	0.0000	0.0000	0.0000	0.0000
bike	1	1997	11375	0.0005	0.0001	0.0001	0.0001
traffic light	0	52840	9	0.0000	0.0000	0.0000	0.0002
traffic sign	422	68584	17892	0.0061	0.0230	0.0097	0.0281

Comprehensive Failure Analysis

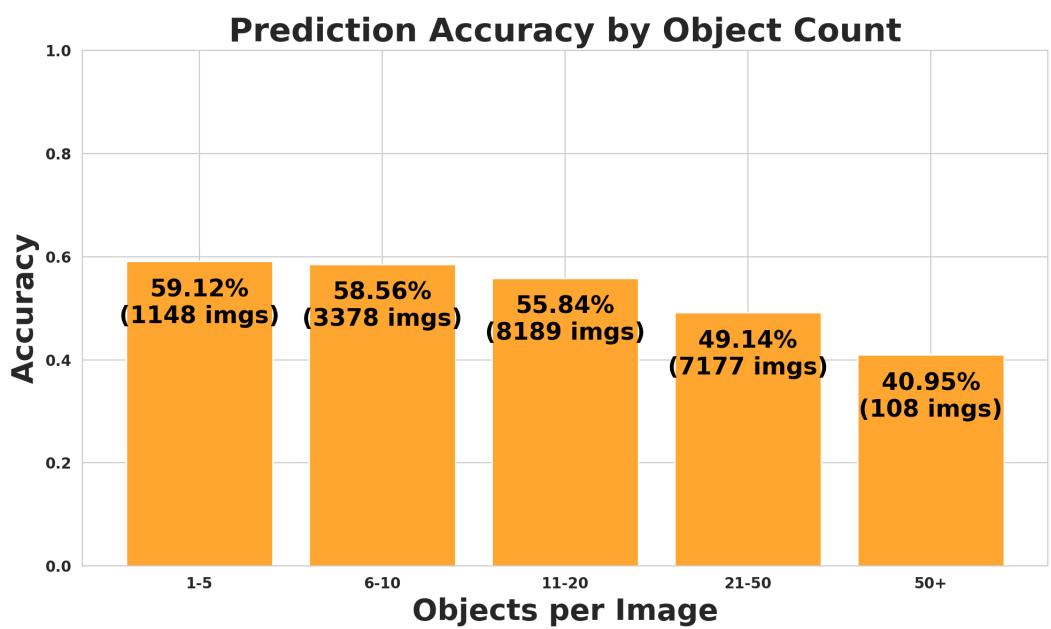
Metric	Value
Total Images Analyzed	20000
Total Expected Objects	367728
Total Matched Objects	191809
Overall Accuracy	52.16%

Accuracy by Environmental Attributes

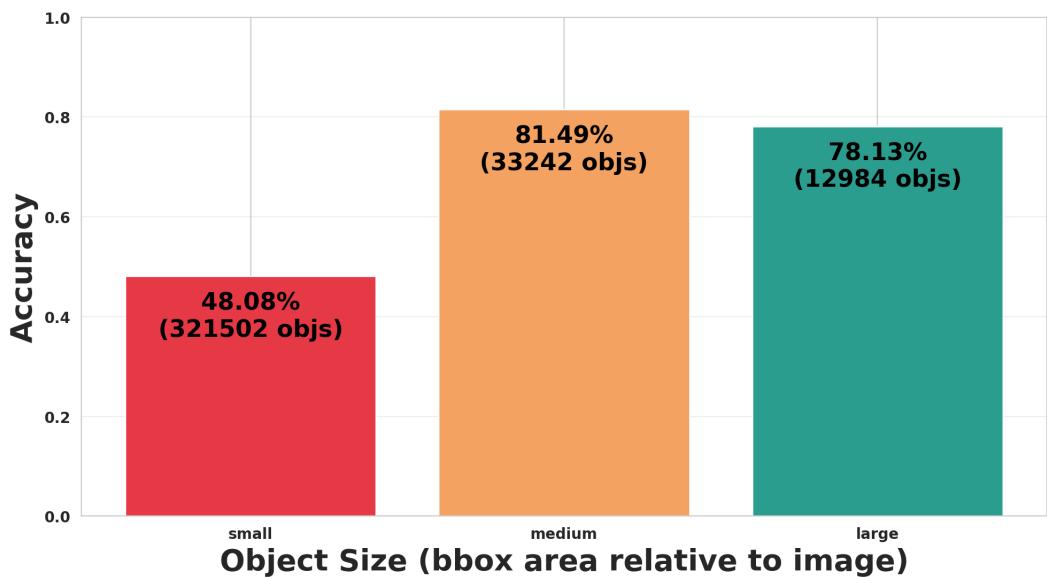




Accuracy by Object Characteristics

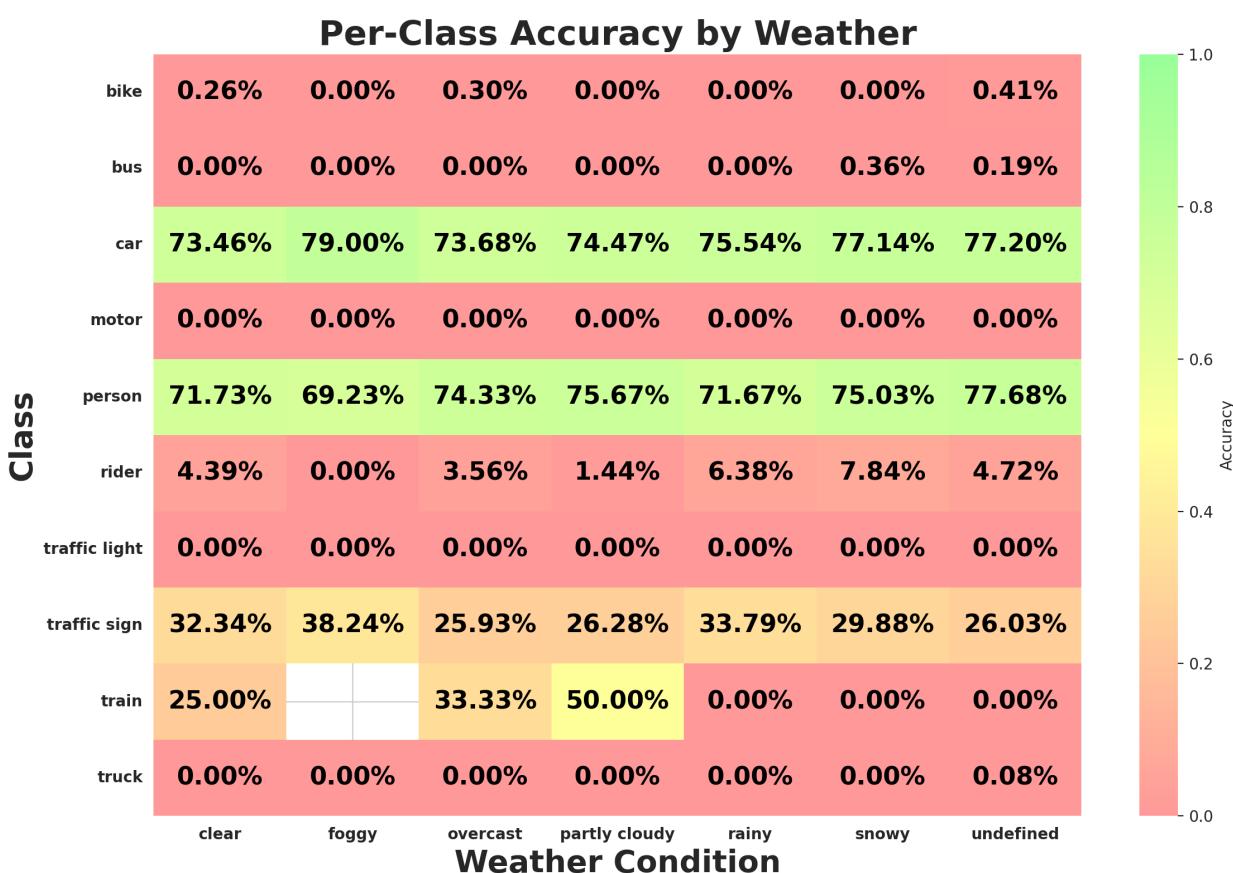
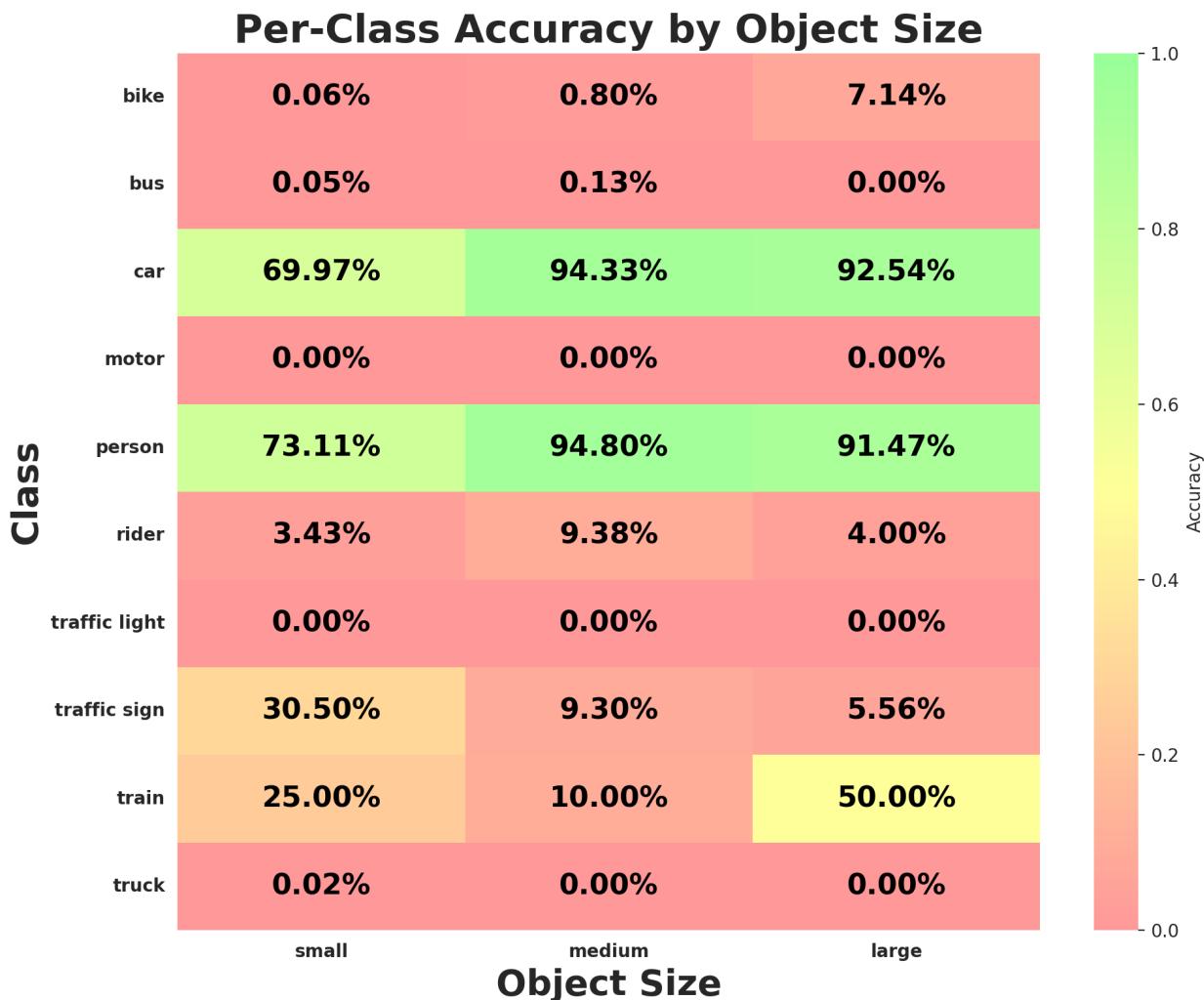


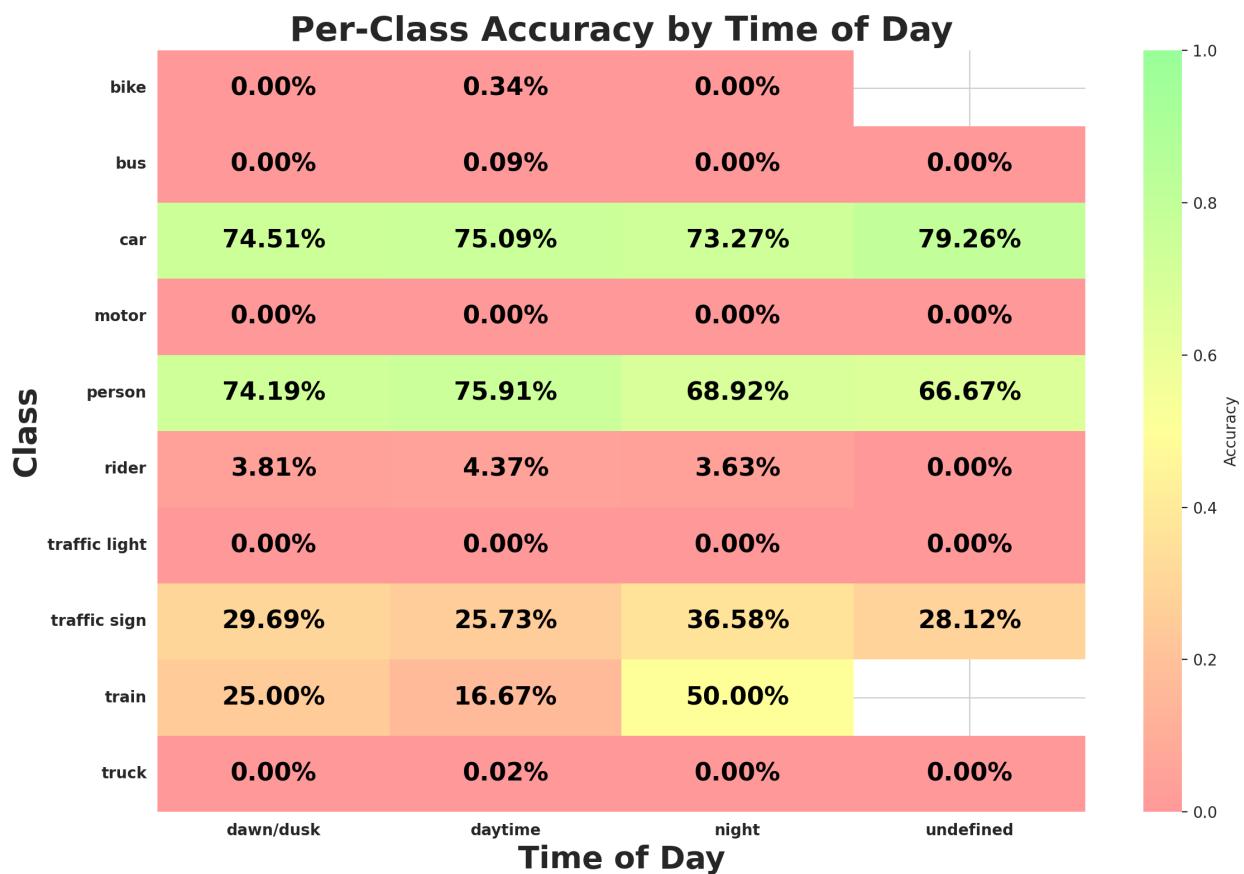
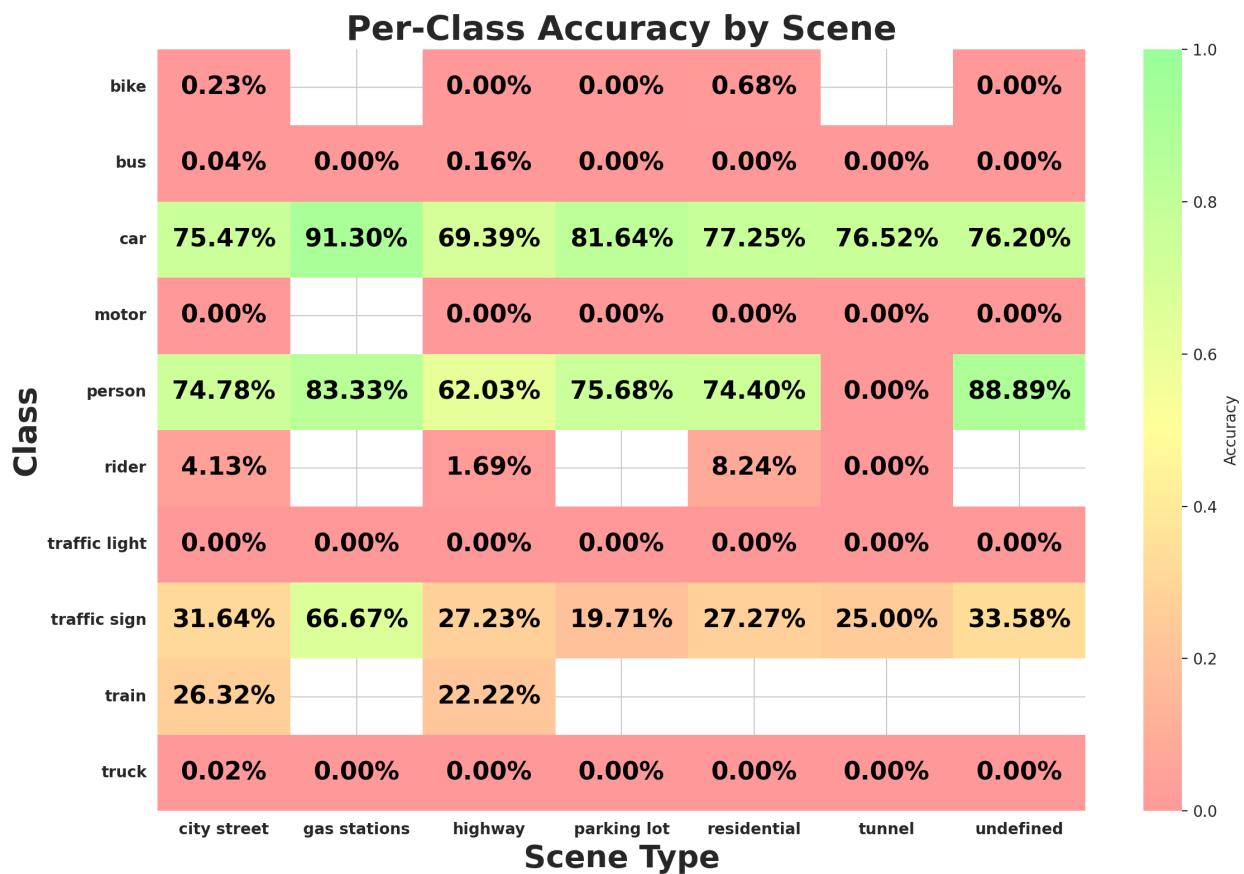
Prediction Accuracy by Object Size



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

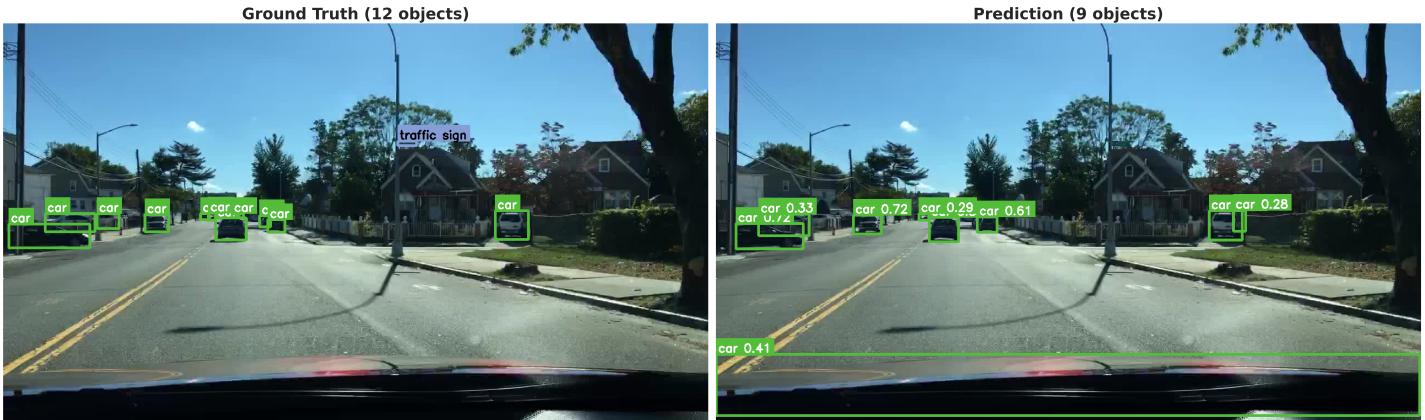




Sample Predictions: Ground Truth vs Model

Sample #1: d58ba9ee-2a2e2d0f.jpg

Ground Truth Objects: 12 | Predicted Objects: 9
Weather: clear | Scene: residential | Time of Day: daytime



Sample #2: fceead05-b73f9eff.jpg

Ground Truth Objects: 18 | Predicted Objects: 10
Weather: clear | Scene: city street | Time of Day: daytime



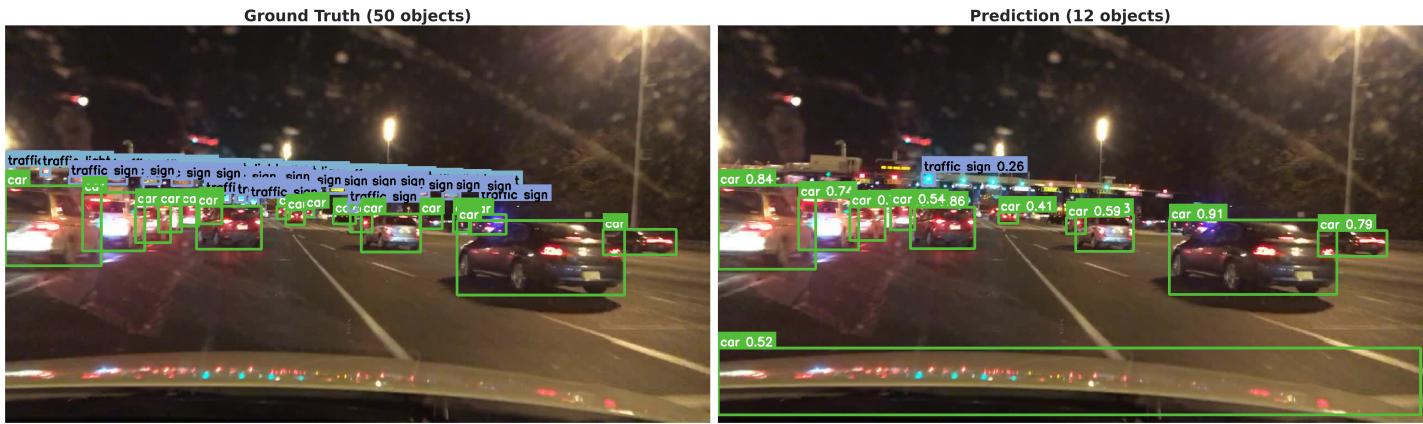
Sample #3: ec17708d-30eea956.jpg

Ground Truth Objects: 32 | Predicted Objects: 21
Weather: overcast | Scene: city street | Time of Day: daytime



Sample #4: e3b59340-e82859a7.jpg

Ground Truth Objects: 50 | Predicted Objects: 12
Weather: snowy | Scene: highway | Time of Day: night



Sample #5: ed9d9554-4a20f791.jpg

Ground Truth Objects: 15 | Predicted Objects: 5
Weather: clear | Scene: residential | Time of Day: night



Sample #6: cc3d9a74-332c6b55.jpg

Ground Truth Objects: 6 | Predicted Objects: 7
Weather: clear | Scene: highway | Time of Day: daytime



Generated by YOLO Testing Pipeline