YOLOv8 Outdoor Object Detection Finetuning Report

This report summarizes the process and results of finetuning a YOLOv8 object detection model on the BDD100K dataset. The steps include dataset preparation, model training, and analysis of results, with visualizations and metrics included.

---

1. Download and Prepare Dataset

We began by copying the BDD100K dataset (in YOLO format) from the Kaggle input directory to the working directory. This dataset contains images and corresponding labels for 10 object classes relevant to autonomous driving scenarios.

Dataset structure after copying:

- images/train/: Training images
- images/val/: Validation images
- labels/train/: Training labels
- labels/val/: Validation labels

The purpose of this step is to ensure all data is accessible for training and validation.

---

2. Verify Dataset Structure

To confirm the dataset was copied successfully, we listed a few sample images from the training directory.

---

3. Install Required Libraries

We used the Ultralytics YOLO library for model training and evaluation.

---

4. Create Dataset Configuration File

A YAML configuration file was created to specify dataset paths and class names. This file defines 10 object classes and points to the appropriate image directories.

---

5. Train YOLOv8 Model

We initialized and trained a YOLOv8 model (yolov8n.pt variant) on the BDD100K dataset. The following hyperparameters were used:

- **Epochs:** 30
- **Batch size:** 16
- **Image size:** 640
- **Model variant:** YOLOv8n (nano, for faster training and lower resource usage)

---

6. Analyze Training Results

After training, we analyzed the results using the results.csv file, which contains metrics for each epoch. Key metrics include:

- **Best mAP50-95 (B):** Mean Average Precision at IoU 0.5:0.95
- **Best Precision (B):** Model's ability to avoid false positives
- **Best Recall (B):** Model's ability to detect all relevant objects

# epoch	# time	# train/box_loss	# train/cls_loss	# train/dfl_loss
26	21524.9	1.3219	0.7777	
27	22324.4	1.31573	0.77023	
28	23121.2	1.31199	0.76462	
29	23917.6	1.30609	0.7586	
30	24718.4	1.30101	0.75141	

5 rows x 15 cols 10 per page Page 1 of 1

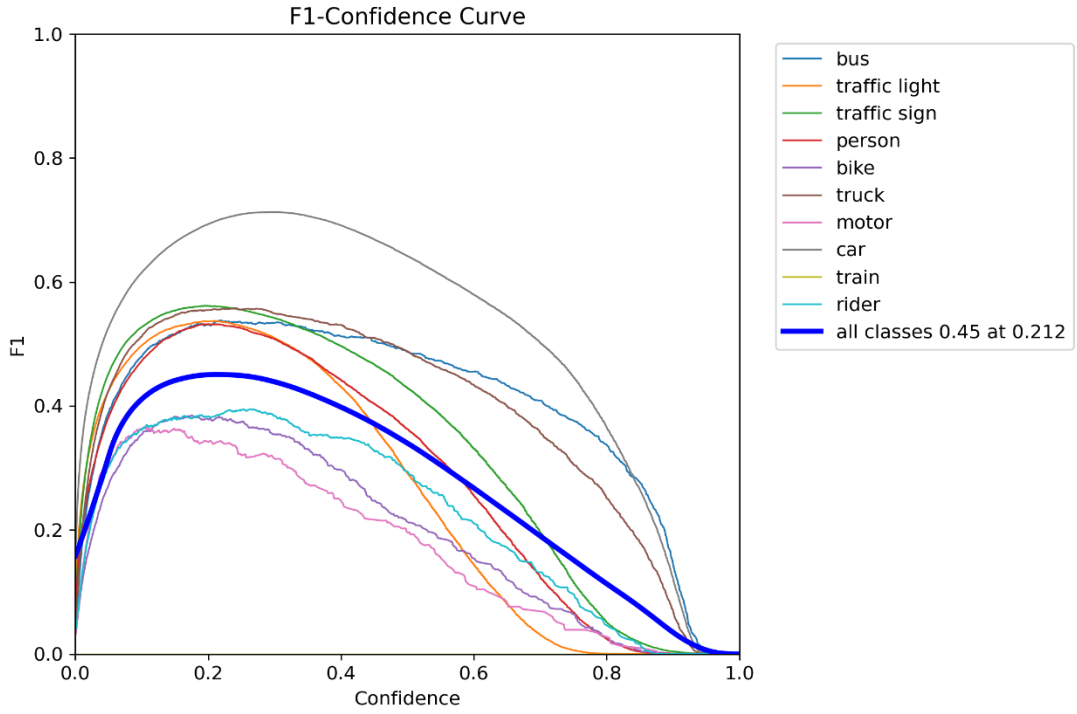
```
Best mAP50-95(B): 0.241
Best Precision(B): 0.656
Best Recall(B): 0.391
```

---

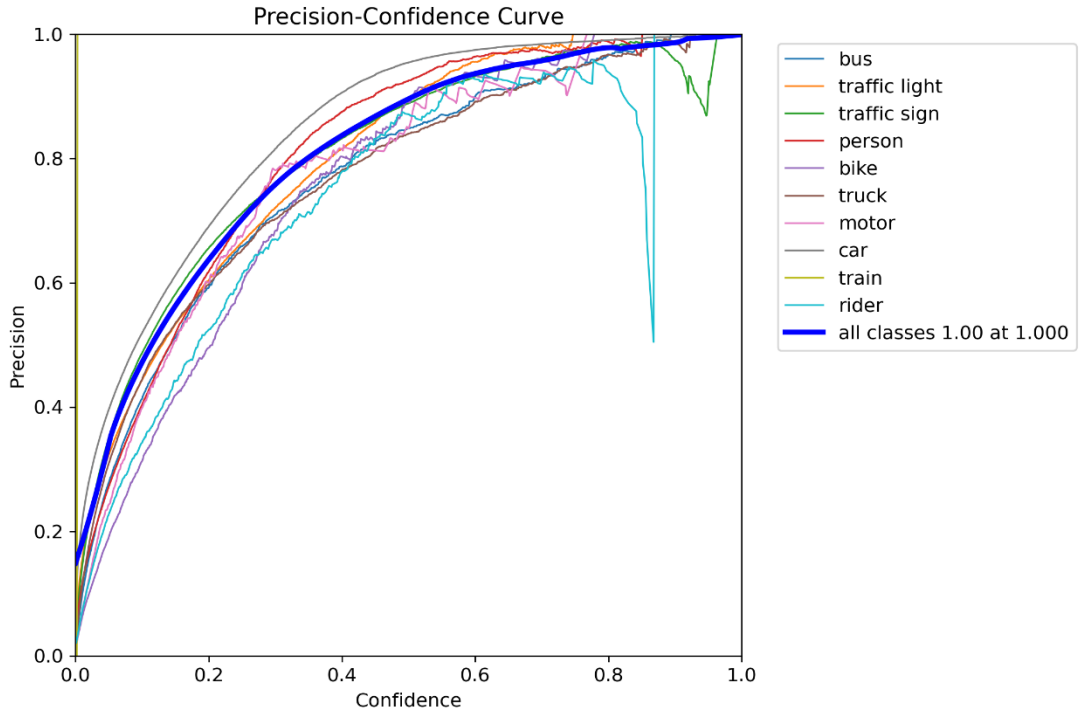
7. Visualize Training Metrics

The following plots illustrate the model's performance during training and validation:

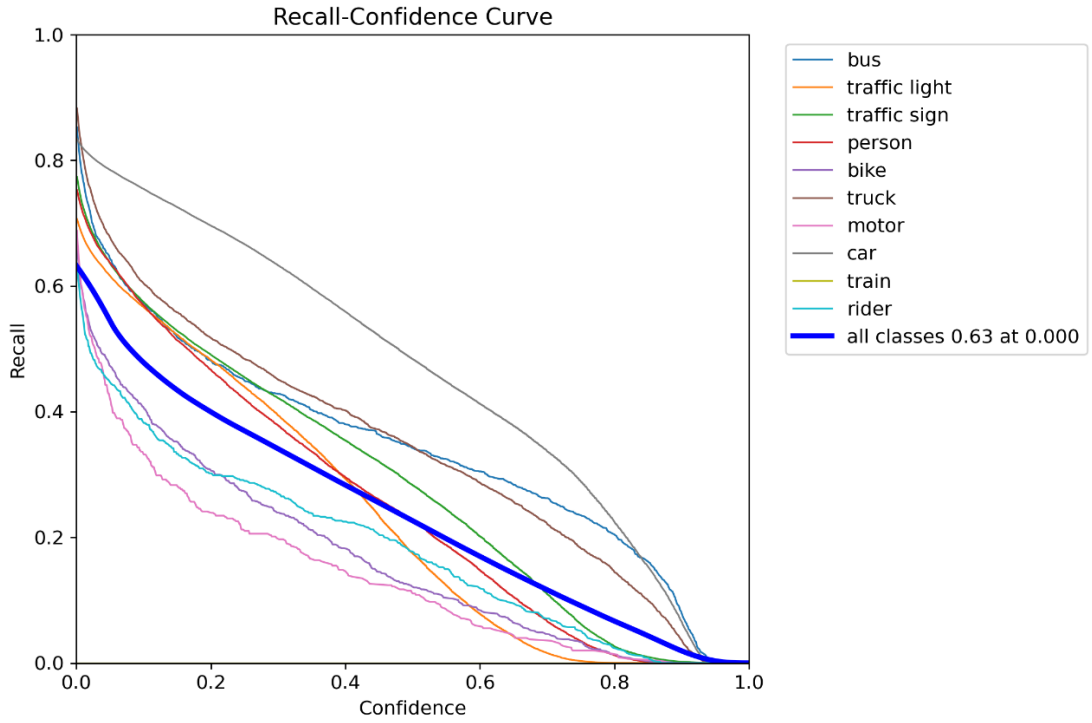
- **BoxF1_curve.png:** F1 score for bounding box predictions



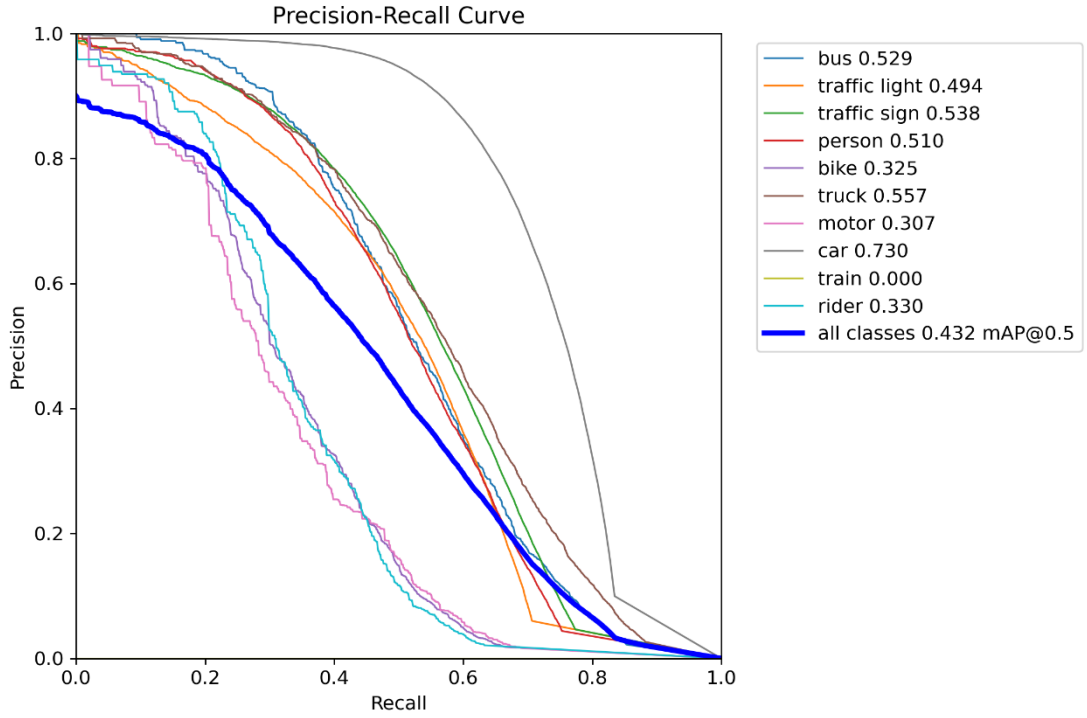
- **BoxP_curve.png:** Precision curve



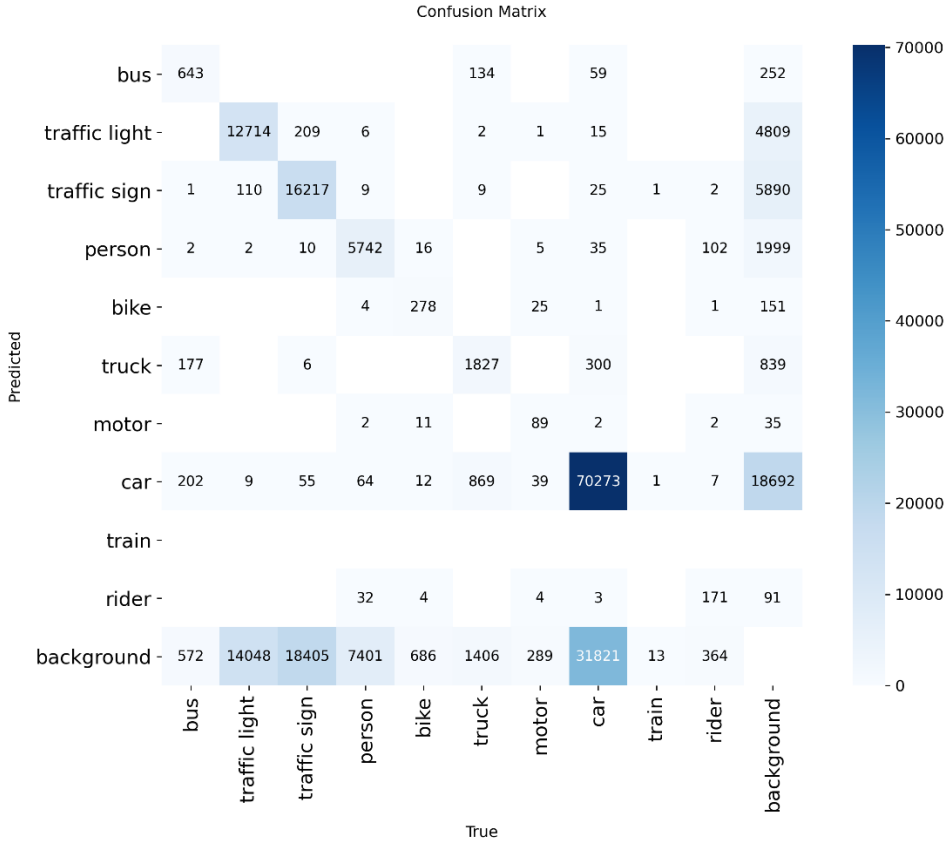
- **BoxR_curve.png:** Recall curve



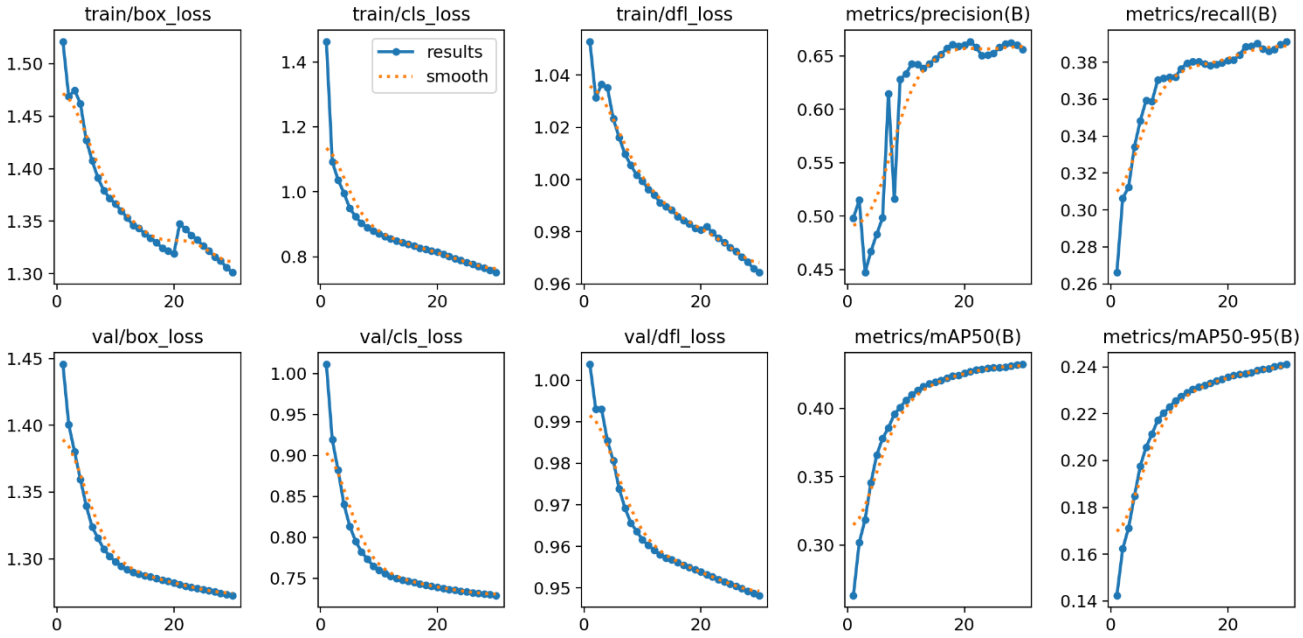
- **BoxPR_curve.png:** Precision-Recall curve



- **confusion_matrix.png:** Confusion matrix for class predictions



- **results.png:** Summary of training and validation losses and metrics

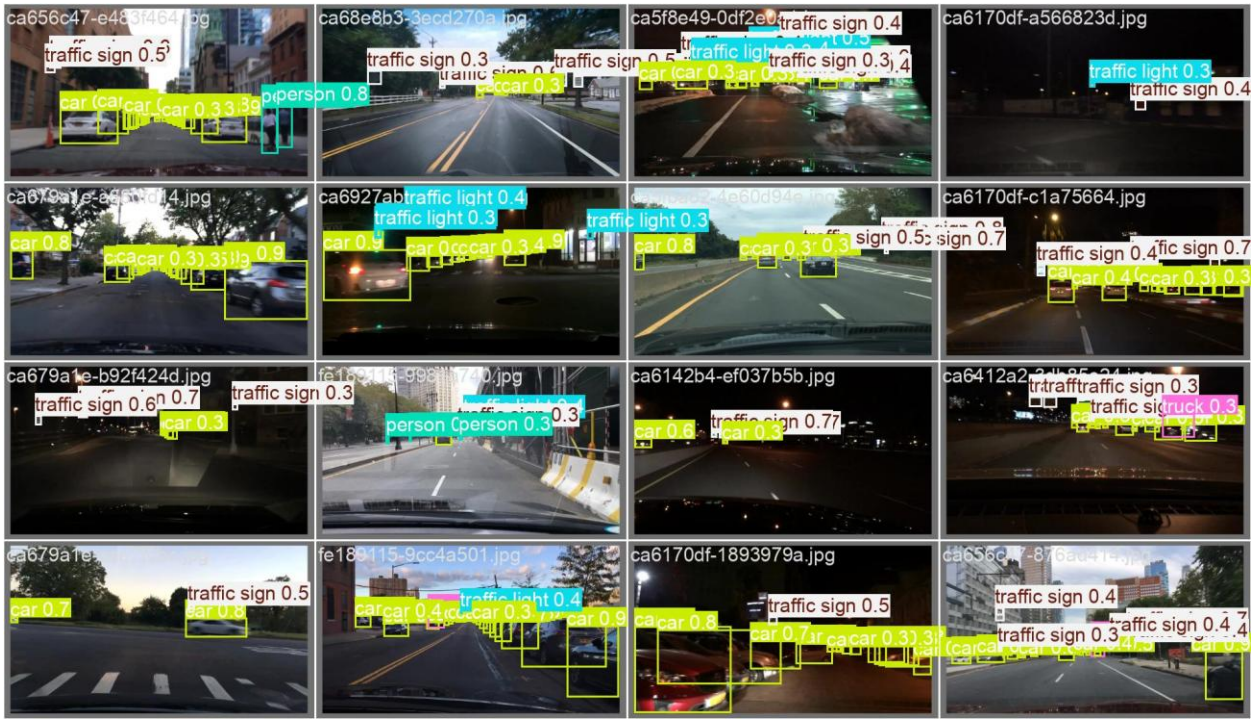


8. Display Example Predictions

Below are example images showing the model's predictions and the ground truth labels on validation data. These qualitative results help assess the model's real-world performance.

- **val_batch0_labels.jpg:** Ground truth labels for a batch of validation images

- **val_batch0_pred.jpg:** Model predictions for the same batch

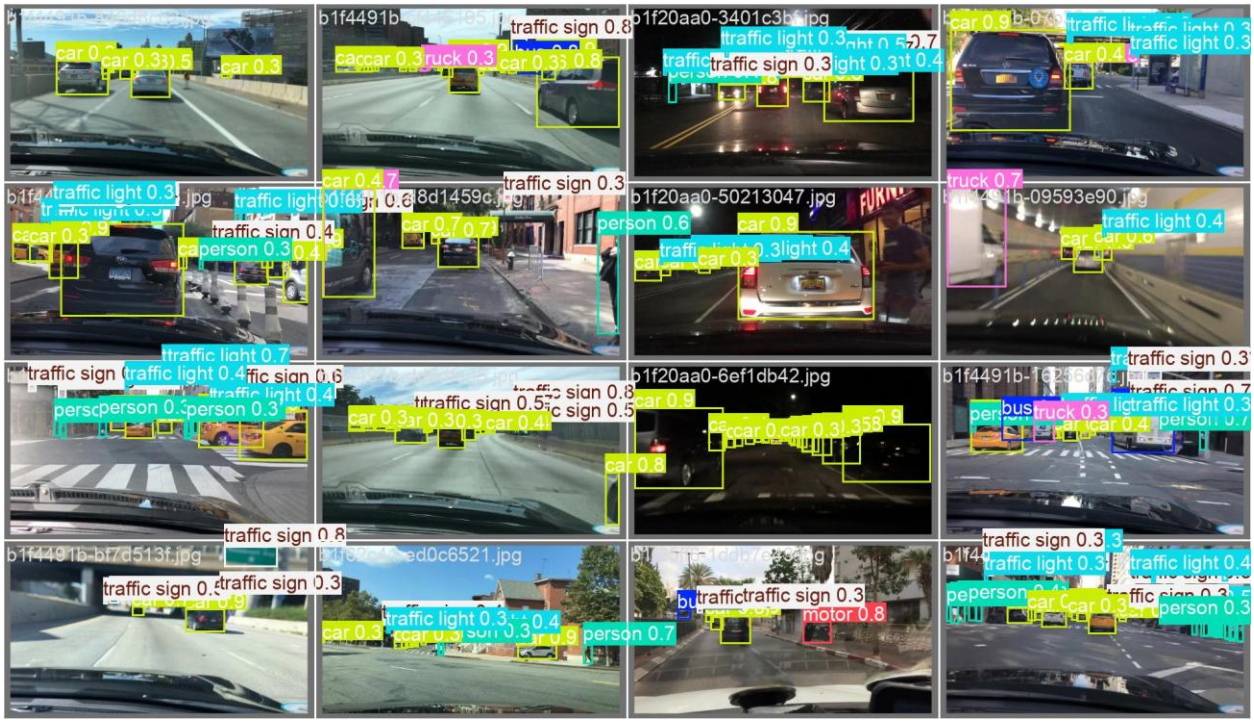


- **val_batch1_labels.jpg:** Ground truth labels for batch 1

val_batch1_pred.jpg: Model predictions for batch 1

- **val_batch2_labels.jpg:** Ground truth labels for batch 2

- **val_batch2_pred.jpg:** Model predictions for batch 2



---

9. Conclusion

Below are examples showing the results of tests done on the best_model.pt our model trained on BDD100K before finetuning again on our custom made dataset. The images used are directly brought from google.

- Image 1: **0: 480x640 1 traffic light, 1 traffic sign, 12 persons, 1 car, 352.3ms Speed: 9.7ms preprocess, 352.3ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)**

Image **Results**

- Image 2: **0: 480x640 2 traffic lights, 13 persons, 1 bike, 3 cars, 1 rider, 257.4ms**
  Speed: 5.7ms preprocess, 257.4ms inference, 13.7ms postprocess per image at shape (1, 3, 480, 640)

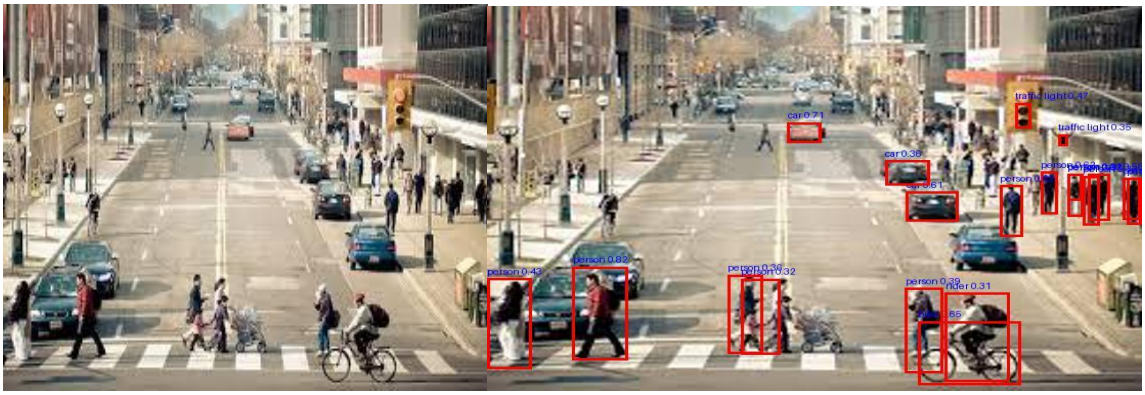


Image **Results**

- Image 3: **0: 480x640 13 persons, 1 bike, 1 car, 1 rider, 326.0ms**
  Speed: 9.5ms preprocess, 326.0ms inference, 4.4ms postprocess per image at shape (1, 3, 480, 640)

Image **Results**

---

10. Conclusion: (1)

This report demonstrates the successful finetuning of a YOLOv8 model on the BDD100K dataset, with both quantitative metrics and qualitative results showing the model's effectiveness in detecting multiple object classes relevant to autonomous driving.

---

11. Fine Tuning on Custom Data set: