Car Detection on BDD100K using YOLOv8s

1. Experiment Reporta

BDD100K is a large-scale driving dataset containing images captured under diverse environmental conditions, including variations in illumination, weather, scene type, occlusion, and object scale. These factors introduce significant challenges for object detection models and directly influence evaluation metrics such as precision, recall, and mean Average Precision (mAP).

The objective of this study is to analyze the impact of dataset characteristics on detection performance and to evaluate a fine-tuned YOLOv8 model for car detection. The experimental pipeline consists of annotation preprocessing, exploratory data analysis, data downsampling, model fine-tuning, and quantitative evaluation on validation and test splits.

To reduce computational overhead while preserving dataset diversity, the training set was downsampled using a stratified approach. Validation and test sets were retained without modification to ensure reliable performance assessment.

2. Exploratory Data Analysis and Observations

Exploratory data analysis was conducted to understand the distribution of key dataset attributes and their implications for detection performance.The training split contains approximately 70,000 images, while the validation and test splits contain 10,000 and 20,000 images respectively. In total, the dataset includes over 100,000 annotated car instances.

2.1 Object Size Distribution

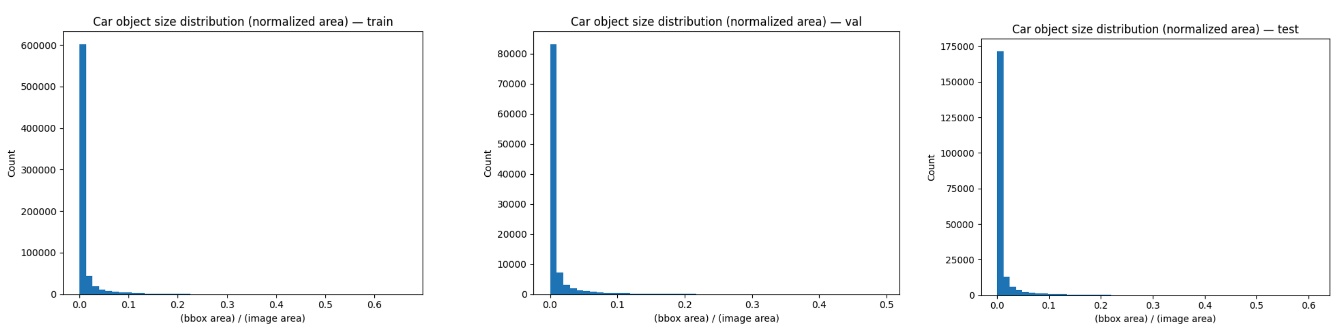

Fig. Car size distribution for train,validation and test datasets

Analysis of normalized bounding box areas revealed a strong skew toward small objects. A substantial proportion of vehicles occupy a limited number of pixels, particularly in highway and urban scenes with long viewing distances. **Small object sizes increase localization difficulty and negatively affect strict localization metrics such as mAP@0.5:0.95.**

2.2 Truncated and Occluded Objects

A non-negligible fraction of annotated vehicles are marked as truncated, indicating partial visibility due to image boundaries. Additionally, occlusion caused by other vehicles and infrastructure is common. **Truncation and occlusion reduce recall and localization accuracy due to incomplete visual information.**

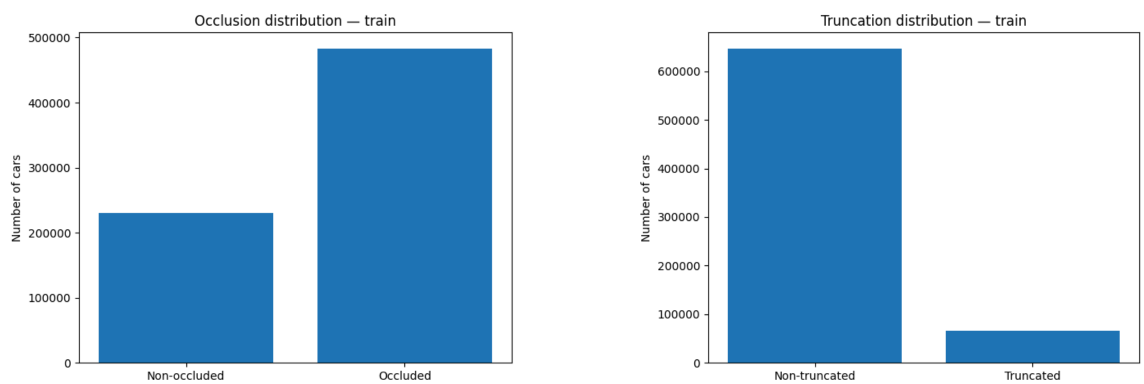

Fig. Occlusion and Truncation in Train dataset

2.3 Time of Day and Weather Conditions

The dataset is dominated by daytime and clear weather conditions, with smaller but significant subsets corresponding to night-time, rainy, foggy, and overcast scenes.**Detection performance is expected to be higher in daytime scenes, while night-time and adverse weather conditions introduce reduced contrast and increased visual noise, leading to lower confidence predictions.**

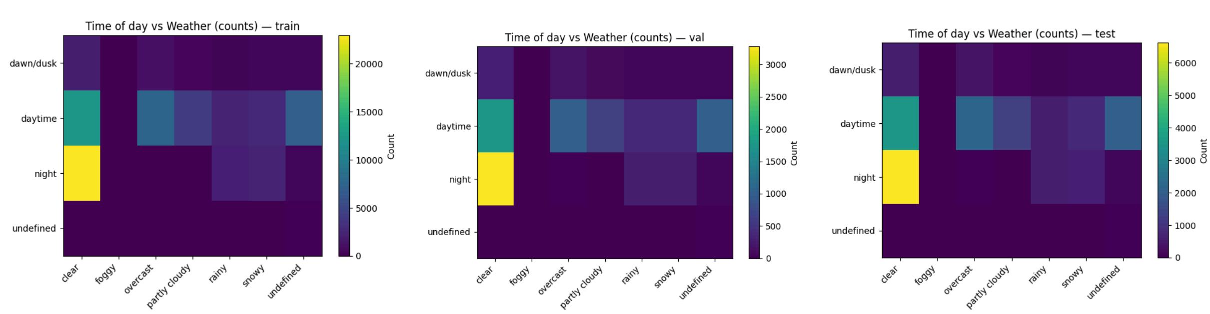

Fig. Heat map of Time of day against weather in train, validation and test data.

3. Data Pre-processing

3.1 Annotation Conversion

BDD100K annotations are originally provided in JSON format. YOLO-based detectors require annotations in a normalized text format specifying class identifiers and bounding box parameters. A custom preprocessing script was implemented to 1. Extract only objects belonging to the **car** category. 2. Clamp bounding boxes within image boundaries. 3. Normalize bounding box coordinates by image width and height. 4. Write YOLO-compatible label files **.txt** format.

3.2 Class Filtering

Only the **car** category was retained. Restricting the task to a single class simplifies the learning objective and allows focused analysis of localization performance.

3.3 Downsampling Strategy

To reduce computational cost during prototyping, the training split was downsampled to 20% of its original size using stratified sampling based on **time-of-day**, ensuring that daytime, night, and dawn/dusk conditions remained proportionally represented. For example, from an original training set of 70,000 images, around 14,000 images were used for training. The validation (10,000 images) and test (20,000 images) splits were not downsampled to preserve the true data distribution and ensure stable, statistically reliable evaluation metrics. This approach significantly reduced training time while maintaining representative training data and unbiased performance assessment.

4. Model Selection and Understanding

4.1 Model Choice

YOLOv8 was selected due to its efficient single-stage detection architecture, strong performance on real-time object detection tasks, and availability of pretrained weights.

4.2 Fine-Tuning Strategy

The model was initialized using COCO-pretrained weights and fine-tuned on the BDD100K car detection task.

- Pretrained weights provide robust low-level feature representations.
- Faster convergence compared to training from scratch.
- Improved generalization on limited training data.

The YOLOv8s variant was chosen as a balance between detection accuracy and computational efficiency.

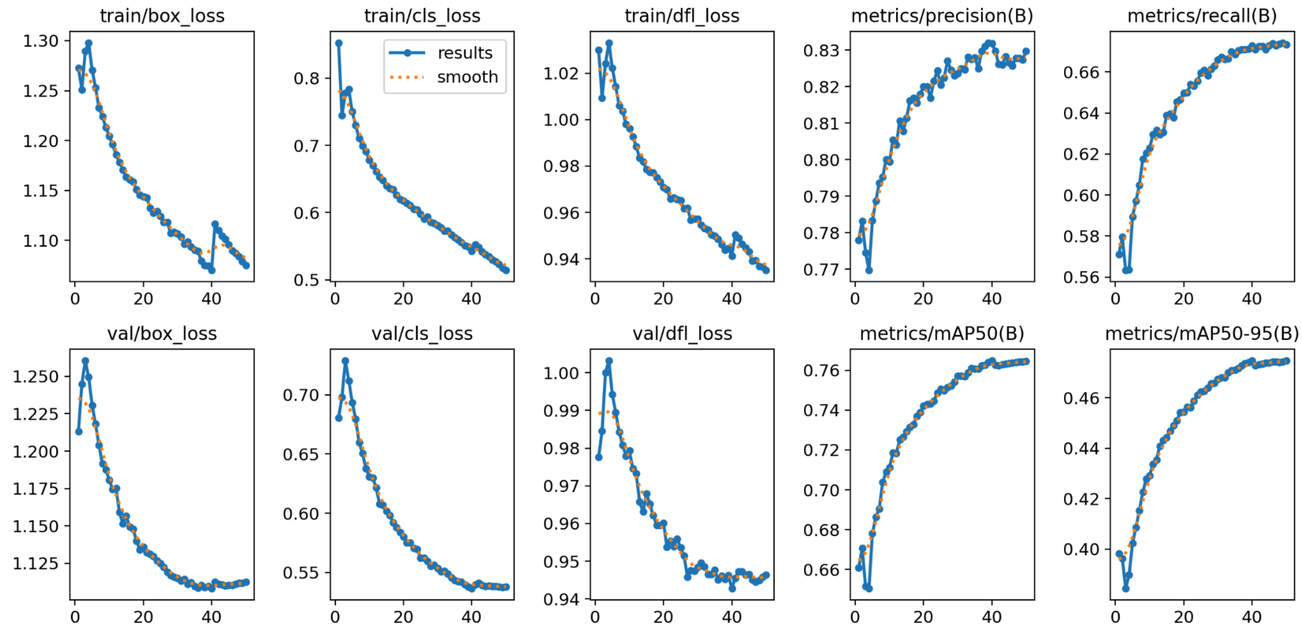

Fig. Metrics for fine tuning.

5. Model Evaluation and Metrics

5.1 Evaluation Metrics

The following standard object detection metrics were used:

- **Precision**: Proportion of correct positive detections.
- **Recall**: Proportion of ground-truth objects detected.
- **mAP@0.5**: Mean Average Precision at IoU ≥ 0.5.
- **mAP@0.5:0.95**: Mean Average Precision averaged over stricter IoU thresholds

5.2 Quantitative Results

Split	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Validation	0.832	0.673	0.765	0.475
Test	0.833	0.671	0.766	0.475

Fig. Qualitative visualization of car detections on test data

5.3 Interpretation

The model achieves high precision, indicating a low false-positive rate. Recall is moderate, reflecting missed detections primarily associated with small, distant, occluded, and low-light vehicles. The gap between mAP@0.5 and mAP@0.5:0.95 highlights challenges in precise localization under real-world driving conditions.Validation and test metrics are closely aligned, indicating good generalization and absence of overfitting.

6. Conclusion

This study presents a complete and reproducible car detection pipeline using YOLOv8 on the BDD100K dataset. The results demonstrate that dataset characteristics such as object scale, truncation, and illumination significantly influence detection performance. Fine-tuning a pretrained YOLOv8 model yields strong generalization while maintaining real-time inference capability.Future work may include higher input resolutions, targeted data augmentation for night-time scenes, and multi-scale training strategies to improve small object detection.