Weather-Robust Vehicle Detection on BDD100K using YOLOv8

1. Experiment Report: Data and Its Effect on Metrics

- This work uses a **10% subset of the BDD100K dataset**, created using uniform random sampling with a fixed seed. The subset contains **6,898 training images**, **1,024 validation images**, and **2,078 test images**, all with a resolution of **1280×720**.
- Every image includes at least one labelled object. The scenes are **highly crowded**, with an average of about **18 objects per image**, and some images containing **nearly 90 objects**, which reflects real-world urban traffic conditions.
- The dataset is **not evenly balanced across categories**. Cars appear most frequently, followed by traffic signs and traffic lights, while trucks, buses, bicycles, and riders occur much less often. This imbalance affects how well the model performs on different classes.
- Object sizes vary significantly. Most objects are **small and far from the camera**, with a median bounding box size of around **29 pixels**, while a smaller number of **large, nearby vehicles** increase the average size, creating a long-tailed distribution.
- Most images are captured in **clear weather**, with fewer examples of overcast, rainy, or snowy conditions, and very few foggy scenes. The dataset mainly includes **daytime and nighttime images**, with limited samples from dawn and dusk, which influences model robustness across different environments.

---

2. EDA and Take-Aways (Important Segments)

BDD100K Distributions

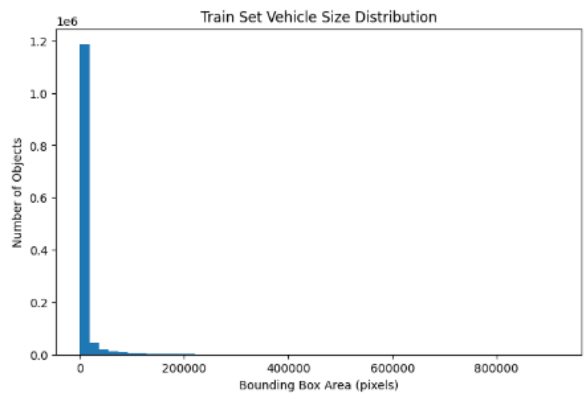
Table 1: Class Distribution and Dataset Overview

Metric	Train	Val	Test
Total Images	6,898	1,024	2,078
Total Boxes	127,228	19,181	38,152
Avg. Boxes / Image	18.4	18.7	18.4

Dominant class counts (overall): - car: ~69,000 - traffic sign: ~24,000 - traffic light: ~19,000 - truck: ~3,000 - bus: ~1,200

A total of **17 unique categories** are present, and every image contains at least one bounding box (no background-only images).

Trian Set Vehicle size Distribution



Weather and Time-of-Day Segments

Table 4: Weather Distribution (Train Split)

Weather	Count	%
clear	3,695	54%
overcast	839	12%
undefined	812	12%
snowy	575	8%
rainy	500	7%
partly cloudy	467	7%
foggy	10	0.1%

Table 5: Top Weather–Time Combinations (Train)

Combination	Count	%
clear / night	2,291	33%
clear / daytime	1,219	18%
overcast / daytime	720	10%

Overall time-of-day split: **daytime (52%)**, **night (41%)**, **dawn/dusk (7%)**. The presence of many “undefined” labels indicates noisy metadata in some samples.

Key EDA Take-Aways

- The dataset is strongly vehicle-centric, making it suitable for vehicle detection, but **class imbalance must be considered** when interpreting results.
- **Small objects and scale variation** are the main factors limiting localization quality and mAP at stricter IoU thresholds.
- The dominance of clear-weather data motivates training a clear-weather baseline while evaluating robustness on mixed-weather scenes.

3. Data Pre-Processing (WHAT, WHY, HOW)

WHAT

- **Training input:** A **clear-weather subset of BDD100K** containing **6,000 images**, used for model training.
- **Validation input:** A **mixed-weather validation set** consisting of **1,200 images**, used for performance monitoring during training.
- **Testing input:** An **adverse-weather test set** with **1,200 images**, reserved for final evaluation.

WHY

- **Clear-weather training:** Clear scenes are the most frequent in BDD100K. Training on clear weather provides a **stable and noise-free baseline** for vehicle detection.
- **Mixed-weather validation:** Including multiple weather conditions during validation helps assess **generalization during training**.
- **Adverse-weather testing:** Evaluating on adverse conditions (e.g., rain, fog, snow) allows **explicit measurement of robustness and domain shift**.
- **Controlled data sizes:** Fixing the number of images per split ensures **balanced and comparable evaluation**.

.

HOW

- Group images by **weather category** using BDD100K metadata.
- Fix the **number of images per split** to prevent over-representation.
- Sample images **uniformly at random** with a **fixed seed**.
- Remove images without **valid vehicle bounding boxes**.
- Apply the **same preprocessing rules** across all splits.

4. Model Understanding (Reasoning for Model Selection)

Choice of YOLOv8

- Designed for **real-time object detection**, matching the requirements of urban driving scenarios.
- **Real-time detection capability**, suitable for urban driving scenarios.
- **Lightweight and efficient**, enabling training and inference on limited compute resources.
- **COCO-pretrained weights** support faster convergence on vehicle detection tasks
- **Anchor-based design** provides stable learning and performs well in crowded scenes with overlapping objects.

Architectural Considerations

- **Backbone:** Extracts rich **semantic and spatial features** from the input image.
- **Neck:** Aggregates features across multiple scales to combine **low-level spatial detail** with **high-level semantic information**.
- **Detection Head:** Predicts **bounding box coordinates**, **object confidence**, and **class scores** for each detected object.

---

5. Model Evaluation Metrics (Reasoning)

Chosen Metrics

- **mAP@0.5:** Measures detection quality under a relaxed overlap constraint.
- **mAP@0.5:0.95:** Evaluates robustness across a range of IoU thresholds.
- **Precision and Recall:** Capture false positive and false negative behavior.

Evaluation Strategy

- **Metrics:** Use **mAP@0.5 and mAP@0.5:0.95** to evaluate detection accuracy and localization quality.
- **Validation:** Evaluate on **clear-weather validation data** to establish a stable baseline.
- **Testing:** Test on **mixed/adverse-weather data** to measure generalization and domain shift.
- **Training setup:** Train for a **fixed number of epochs and batch size** to ensure stable convergence.

- **Initialization:** Use **COCO-pretrained YOLOv8 weights** for faster and more reliable training.
- **Analysis:** Use **precision–recall curves** to analyse confidence behaviour and class-wise performance.

Parsed epochs: 30

Final Epoch: 30

Final Precision: 0.806

Final Recall: 0.604

Final mAP@0.5: 0.7

Final mAP@0.5:0.95: 0.422