Vehicle Detection using YOLOv8

1. Introduction

Vehicle detection plays a significant role in applications such as autonomous driving, traffic management, and security systems. In this project, we aim to improve the performance of the YOLOv8 model for vehicle detection. We compare the base YOLOv8 model with a modified version, where certain architectural enhancements are introduced to increase accuracy and robustness. These modifications mainly focus on the neck section, with no changes made to the backbone or head of the model.

2. Base Model

The base model used in this study is the **YOLOv8** architecture, a well-established model for object detection. It is designed to balance accuracy and speed, making it suitable for real-time applications.

Base Model Architecture

- **Backbone**: The backbone of the base model remains unchanged and is responsible for extracting features from input images using standard convolutional layers.
- **Neck**: The neck aggregates the features extracted by the backbone and prepares them for the final detection process.
- **Head**: The head of the model predicts bounding boxes and class probabilities. In this project, the head of the model remains the same in both the base and modified versions.

3. Modified Architecture Changes

The **modified YOLOv8 model** introduces specific changes primarily in the **neck** section. These modifications are aimed at improving the feature aggregation process, which is critical for object detection performance. However, no changes were made to the **backbone** or **head** of the model.

Changes in the Neck:

1. **EMAFeatureLayer**: The modified model introduces an **Exponential Moving Average (EMA) feature layer** in the neck section. This layer helps stabilize the feature representations by maintaining a running average of the features over time. It reduces noise and ensures that the model learns more stable and robust features during training.

   The EMA feature layer accumulates weighted averages of the feature maps, which is particularly useful in improving the model's ability to handle fluctuations in the input data.

2. **GhostConv and C3Ghost Layers**: These custom layers were introduced in the neck section to optimize the feature extraction and reduce the computational load. The GhostConv layers are designed to create more feature maps with fewer computations, improving both the speed and efficiency of the model.

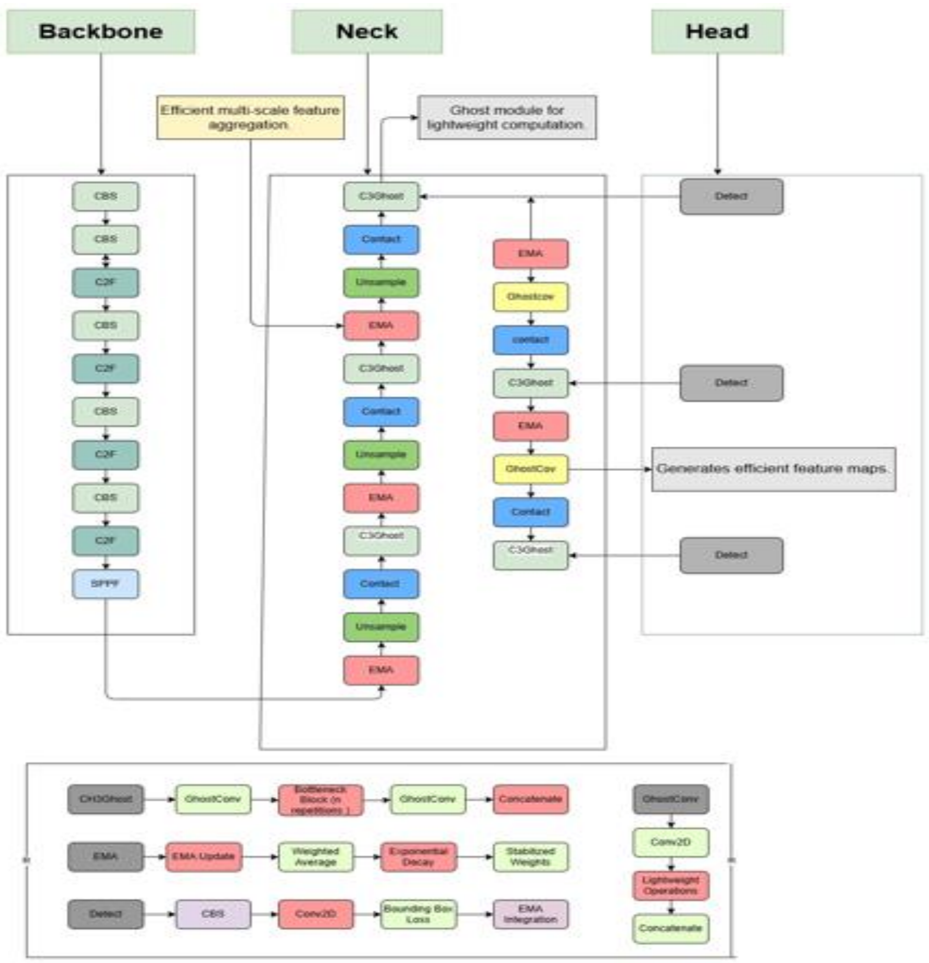
   The C3Ghost layer adds another convolutional operation, further enhancing the feature extraction process while using a dropout technique to reduce overfitting.

By modifying the neck, the model becomes better at aggregating and refining the features before passing them to the head for final prediction.

Unchanged Backbone and Head:

- **Backbone**: The backbone remains unchanged, continuing to use the original YOLOv8 architecture for feature extraction.
- **Head**: The head also remains the same, maintaining the original detection and classification layers.

5.Architecture Diagram



5. Comparison of Results

To evaluate the effectiveness of the architectural changes, the performance of the base model and the modified model were compared using key object detection metrics: **Precision**, **Recall**, **mAP50**, and **mAP50-95**.

Results Comparison

Metric	Base Model	Modified Model	Percentage Improvement
Precision	0.6467	0.7557	16.91%
Recall	0.5653	0.6488	14.77%
mAP50	0.6258	0.6769	8.18%
mAP50-95	0.4011	0.4478	11.69%

The **modified model** demonstrates improvements across all metrics:

- **Precision** increased by **16.91%**, indicating that the modified model is more accurate in detecting vehicles.
- **Recall** improved by **14.77%**, suggesting the model is better at detecting all vehicles in the dataset.
- **mAP50** showed an **8.18%** improvement, reflecting better performance at the standard Intersection over Union (IoU) threshold of 0.5.
- **mAP50-95** improved by **11.69%**, indicating that the modified model performs better across a range of IoU thresholds.

These results confirm that the modifications to the neck section, including the EMAFeatureLayer and GhostConv layers, led to significant improvements in the model's performance.

5. Conclusion

The modifications made to the YOLOv8 model, particularly in the neck section, have resulted in substantial improvements in vehicle detection performance. By incorporating the EMAFeatureLayer and GhostConv layers, the modified model achieved higher precision, recall, and mAP scores compared to the base model.

These results highlight the importance of optimizing the feature aggregation process in the neck section of the model. The modifications contribute to more stable and robust feature representations, leading to better detection accuracy and efficiency. Further research could focus on testing additional enhancements or experimenting with other hyperparameters to further improve model performance.