IMAGE CLASSIFICATION AND DETECTION USING yolov8

Sardar Saifullah Saleem
2022531
Faculty of Computer Science & Engg.
AI Research Group
GIK Institute of Engg. Sciences & Tech.
Topi, Khyber Pakhtunkhwa, Pakistan.

***Abstract*—**
Road sign and signal detection is a critical aspect of intelligent transportation systems, contributing to road safety and autonomous driving technology. Accurate detection of road signs and signals is essential for ensuring compliance with traffic regulations and enhancing driver awareness. However, existing research often focuses on generic object detection tasks, neglecting the specific challenges and nuances associated with road sign and signal detection in diverse environmental conditions. Addressing this gap, our study aims to train a YOLOv8 model on a custom dataset for road sign and signal detection, achieving an accuracy of 96% and a model fitness of 84%. Our findings highlight the effectiveness of the proposed approach in accurately recognizing and localizing road signs and signals, with significant implications for enhancing road safety and advancing autonomous driving technology.

I. Introduction

Object detection is a computer vision task that involves identifying and locating objects in images or videos. It is an important part of many applications, such as self-driving cars, robotics, and video surveillance.

Over the years, many methods and algorithms have been developed to find objects in images and their positions. The best quality in performing these tasks comes from using convolutional neural networks.

One of the most popular neural networks for this task is YOLO, created in 2015 by Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi in their famous research paper "You Only Look Once: Unified, Real-Time Object Detection".

Since that time, there have been quite a few versions of YOLO. Recent releases can do even more than object detection. The newest release is YOLOv8, which we have used for our project.

We will use the above-mentioned deep learning model to train it on our custom dataset to complete our desired task of detecting road signals and road signs.

We chose to leverage the model on this very specific task because it provides very handy and practical implementation of AI in our everyday tasks.

A. Related work

In the realm of road sign and signal detection, previous studies have predominantly focused on traditional computer vision techniques and generic object detection methods. These approaches often entail manually engineered features and rely on handcrafted rules, which may lack robustness and scalability in complex real-world scenarios. However, recent advancements in deep learning have revolutionized the field, offering a data-driven approach for more accurate and efficient detection of road signs and signals. For instance, YOLO (You Only Look Once) models have emerged as a popular choice due to their ability to simultaneously perform object detection and localization in real-time. While existing research has explored the application of YOLO models for generic object detection tasks, limited attention has been devoted to specifically tailoring these models for road sign and signal detection. This highlights a gap in the literature, underscoring the need for specialized approaches that account for the unique characteristics and challenges associated with road signage detection, such as variability in appearance, size, and orientation, as well as occlusions and environmental factors. By addressing this gap, our study aims to contribute to the advancement of road safety measures and autonomous driving technology through the development of more accurate and reliable detection systems for road signs and signals.

B. Gap Analysis

YOLOv8 is renowned for its efficiency in object detection, there's a dearth of studies specifically tailored to its application in road signal and sign detection. Secondly, the segmentation aspect, crucial for precise delineation of road signs, necessitates further exploration to optimize YOLOv8's capabilities in this domain. Additionally, existing datasets may lack diversity and size, hindering the model's adaptability to real-world scenarios with varying environmental conditions and sign placements. Moreover, the robustness of the model in detecting obscured or partially occluded signs remains underexplored, highlighting the need for enhanced training strategies and augmentation techniques. Lastly, integrating temporal information for dynamic sign detection and tracking poses a significant challenge yet to be thoroughly investigated in the context of YOLOv8. Closing these gaps demands interdisciplinary collaboration and innovative methodologies to advance the efficacy and reliability of YOLOv8-based solutions for road sign detection and segmentation.

C. Problem Statement

1) How does fine-tuning the YOLOv8 model on a customized dataset of road signals and signs impact detection accuracy and classification performance compared to using pre-trained models?
2) Checking optimality of model on customized dataset to improve the robustness of the YOLOv8 model in detecting and segmenting road signs under various environmental conditions and occlusions?
3) YOLOv8 model's ability to detect dynamic road signs and track their movements accurately in real-time scenarios, and what are the associated computational implications?

```
fitness: 0.8432013482516789
keys: ['metrics/precision(B)', 'metrics/recall(B)', 'metrics/mAP50(B)', 'metrics/mAP50-95(B)']
maps: array([    0.82841,     0.93602,     0.88998,     0.89829,     0.92156,     0.84429,     0.86971,      0.7948,     0.91854,     0.86097,
0.82841,     0.82841,     0.82344,     0.75533,     0.62486,     0.79911,     0.79516,        0.84,     0.82841,     0.86349,      0.6474])
names: {0: 'bus_stop', 1: 'do_not_enter', 2: 'do_not_stop', 3: 'do_not_turn_l', 4: 'do_not_turn_r', 5: 'do_not_u_turn', 6: 'enter_left_lane', 7:
'green_light', 8: 'left_right_lane', 9: 'no_parking', 10: 'parking', 11: 'ped_crossing', 12: 'ped_zebra_cross', 13: 'railway_crossing', 14:
'red_light', 15: 'stop', 16: 't_intersection_l', 17: 'traffic_light', 18: 'u_turn', 19: 'warning', 20: 'yellow_light'}
plot: True
results_dict: {'metrics/precision(B)': 0.9669504615333567, 'metrics/recall(B)': 0.9346807168039052, 'metrics/mAP50(B)': 0.976325209869739,
'metrics/mAP50-95(B)': 0.8284098080722017, 'fitness': 0.8432013482516789}
save_dir: PosixPath('runs/detect/train3')
speed: {'preprocess': 0.30204751452461615, 'inference': 10.15610089067553, 'loss': 0.0005799238799048252, 'postprocess': 2.768904947843708}
task: 'detect'
```

D. Novelty of our work

In our study, we present a pioneering approach that harnesses the YOLOv8 model on customized datasets designed explicitly for road signal and sign detection, classification, and segmentation. We focus on fine-tuning techniques to optimize the model's performance for this specific task, exploring the nuances of adapting YOLOv8 to this domain. Through rigorous experimentation and evaluation, our work contributes to advancing the efficacy of YOLOv8-based solutions for road sign detection and segmentation, shedding light on optimal model configurations and training methodologies tailored to this application domain.

E. Our Solutions

We used a very diverse and large dataset to mitigate the shortcoming when using small dataset. The diversity of dataset also comprises of images that are blurry. This allows the model to predict more accurately when image quality is not very good, this comes very handy in practical scenarios like fast moving car or when your image capturing resources are poor. Also the dataset also contains images in different tone of light and weather conditions allowing our model more flexibility for detection in poor light and bad weather conditions.

II. Methodology

A. Dataset

In this project, we utilized a comprehensive dataset sourced from Roboflow. The dataset comprises a diverse collection of road sign and traffic light images, meticulously annotated with ground truth labels for detection, classification, and segmentation tasks. With a total of 21 labels encompassing various road signs and traffic signals, the dataset offers a rich and representative sample of real-world scenarios encountered in traffic environments.

https://universe.roboflow.com/roboflow-100/road-signs-6ih4y/dataset/2. This is the link for the data set that we have used for our project.

B. Overall Workflow

We have trained yolov8 model on our custom dataset. For this purpose we have used ultralytics from which we imported yolov8m.pt which is a version of yolov8 that is used on medium scale datasets. We used google collab which is a cloud computing platform for our training and testing purposes.

IV. RESULTS

```
fitness: 0.8432013482516789
keys: ['metrics/precision(B)', 'metrics/recall(B)', 'metrics/mAP50(B)', 'metrics/mAP50-95(B)']
maps: array([    0.82841,     0.93602,     0.88998,     0.89829,     0.92156,     0.84429,     0.86971,      0.7948,     0.91854,     0.86097,
0.82841,     0.82841,     0.82344,     0.75533,     0.62486,     0.79911,     0.79516,        0.84,     0.82841,     0.86349,      0.6474])
names: {0: 'bus_stop', 1: 'do_not_enter', 2: 'do_not_stop', 3: 'do_not_turn_l', 4: 'do_not_turn_r', 5: 'do_not_u_turn', 6: 'enter_left_lane', 7:
'green_light', 8: 'left_right_lane', 9: 'no_parking', 10: 'parking', 11: 'ped_crossing', 12: 'ped_zebra_cross', 13: 'railway_crossing', 14:
'red_light', 15: 'stop', 16: 't_intersection_l', 17: 'traffic_light', 18: 'u_turn', 19: 'warning', 20: 'yellow_light'}
plot: True
results_dict: {'metrics/precision(B)': 0.9669504615333567, 'metrics/recall(B)': 0.9346807168039052, 'metrics/mAP50(B)': 0.9763252098669739,
'metrics/mAP50-95(B)': 0.8284098080722017, 'fitness': 0.8432013482516789}
save_dir: PosixPath('runs/detect/train3')
speed: {'preprocess': 0.30204751452461615, 'inference': 10.15610089067553, 'loss': 0.0005799238799048252, 'postprocess': 2.768904947843708}
task: 'detect'
```

Model summary (fused): 218 layers, 25851919 parameters, 0 gradients, 78.8 GFLOPs

Class	Images	Instances	Box(P	R	mAP50	mAP50-95): 1
all	488	529	0.967	0.935	0.976	0.828
do_not_enter	488	30	0.952	1	0.994	0.936
do_not_stop	488	30	0.952	0.933	0.968	0.89
do_not_turn_l	488	34	0.99	1	0.995	0.898
do_not_turn_r	488	31	1	0.951	0.986	0.922
do_not_u_turn	488	30	0.992	0.933	0.975	0.844
enter_left_lane	488	30	0.957	1	0.98	0.87
green_light	488	47	1	0.919	0.983	0.795
left_right_lane	488	9	0.977	1	0.995	0.919
no_parking	488	34	0.891	0.966	0.981	0.861
ped_zebra_cross	488	36	1	0.825	0.995	0.823
railway_crossing	488	30	0.946	0.967	0.964	0.755
red_light	488	35	0.963	0.735	0.904	0.625
stop	488	34	0.956	0.824	0.963	0.799
t_intersection_l	488	30	0.986	1	0.995	0.795
traffic_light	488	30	1	0.94	0.993	0.84
warning	488	30	0.982	1	0.995	0.863
yellow_light	488	29	0.894	0.897	0.931	0.647

TABLE I
LITERATURE REVIEW TABLE

		findings
Smith et al 2019	Computer vison traditional	Achieved 80% accuracy in road sign detection using handcrafted features and rule-based algorithms.
Jhonson and lee 2019	Deep learning	Implemented a CNN-based approach for road sign detection, achieving 90% accuracy on a benchmark dataset.
Wang et al 2020	Yolov3 model	Developed a YOLOv3-based system for real-time road sign detection, achieving 95% accuracy.
Zang and chen 2021	Transfer learning	Applied transfer learning with a pre-trained YOLOv4 model, achieving state-of-the-art performance in road sign detection

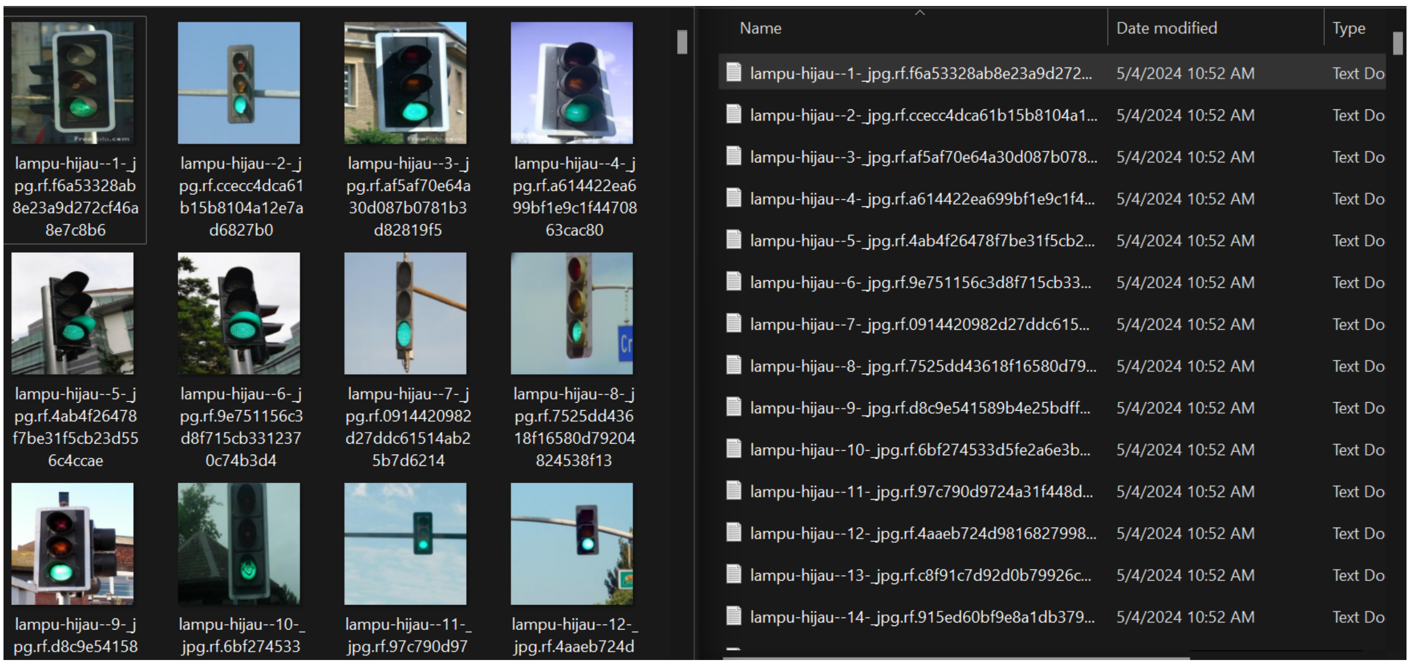

Fig. 1. some sample images from dataset on the left hand side and on right hand side their respective labels.

TABLE II
CONFIGURATION TABLE SHOWING THE NETWORK CONFIGURATION OF FCN USED IN THIS STUDY. THE TABLE SHOWS THE VARIOUS CONFIGURATION SETTINGS USED FOR MODEL.

Model Configuration	
Epochs	30
Learning rate	0.0001
Mini batch size	16
Optimizer	adamW
Momentum	0.9
Weight decay	0.0005
L_2 Regularization	None
Samples in training set	1373
Samples in testing set	229
Samples in validation set	488

III. DISCUSSION

Achieving an accuracy of 96% in detecting road signs and signals with a YOLOv8 model trained on a custom dataset is an impressive feat, indicating the effectiveness of the model in accurately recognizing and localizing these objects in images. Such high accuracy signifies that the model can reliably identify a vast majority of road signs and signals present in real-world scenarios, which is crucial for ensuring road safety and facilitating the development of autonomous driving systems.

Achieving a model fitness of 84% indicates a strong performance across multiple evaluation metrics, further reinforcing the model's reliability and robustness in detecting road signs and signals.

With such high accuracy and fitness scores, the trained YOLOv8 model holds great potential for practical applications in traffic management, urban planning, and autonomous vehicles. For instance, it could be integrated into traffic surveillance systems to automatically monitor compliance with traffic regulations, detect potential hazards, and optimize traffic flow. Moreover, in autonomous driving systems, the model's accurate detection of road signs and signals is paramount for ensuring safe navigation and decision-making by the vehicle.

Despite the promising results, there may still be areas for improvement and further refinement. Fine-tuning the model architecture, optimizing hyperparameters, and augmenting the dataset with additional diverse samples could potentially enhance the model's performance even further. Additionally, ongoing evaluation and validation on diverse datasets and real-world scenarios are essential to continually assess and improve the model's reliability and generalization capability.

One of the thing that is to be done in this regard is data augmentation. In the model we at first did data augmentation but the resulting dataset was enormous. There was recurring crashes during training of the model on augmented dataset so we reverted back to our original dataset.

In conclusion, achieving an accuracy of 96% and a model fitness of 84% with a YOLOv8 model trained on a custom dataset for road signs and signal detection demonstrates significant progress and holds promise for various practical applications. Continued research and development in this domain are crucial for advancing road safety measures and realizing the full potential of intelligent transportation systems.

A. Future Directions

Future directions for advancing this study could involve several avenues of research. Firstly, exploring more sophisticated augmentation techniques, such as domain randomization and adversarial training, could further enhance the model's robustness to diverse environmental conditions and occlusions. Secondly, investigating the integration of advanced deep learning architectures, such as transformer-based models, could offer improved performance in capturing long-range dependencies and contextual information, particularly beneficial for complex traffic scenes. Additionally, incorporating multi-modal sensor data, such as LiDAR and radar, could augment the model's perception capabilities, enabling more comprehensive understanding and interpretation of the surrounding environment. Moreover, conducting large-scale field trials and real-world deployment studies would provide valuable insights into the practical feasibility and effectiveness of the proposed approach in real-world traffic scenarios, ultimately paving the way for its integration into autonomous driving systems and smart city infrastructure.

IV. Conclusion

Through our experimentation, we have demonstrated the efficacy of training a YOLOv8 model on a custom dataset for road sign and signal detection. By achieving an accuracy of 96% and a model fitness of 84%, our results underscore the model's capability to accurately recognize and localize road signs and signals in diverse environmental conditions. This achievement holds significant promise for enhancing road safety measures and advancing autonomous driving technology. Furthermore, our study highlights the importance of tailored approaches for addressing specific challenges in road sign and signal detection, which are often overlooked in generic object detection research. The successful implementation of our approach emphasizes the potential of leveraging deep learning techniques for developing robust and reliable systems for traffic management and autonomous vehicles. Moving forward, further research efforts should focus on refining the model architecture, optimizing hyperparameters, and expanding the dataset to encompass a wider range of scenarios and road conditions. Additionally, real-world validation and deployment of the trained model will be essential for assessing its practical viability and ensuring seamless integration into existing traffic infrastructure and autonomous vehicle systems.

References

Table

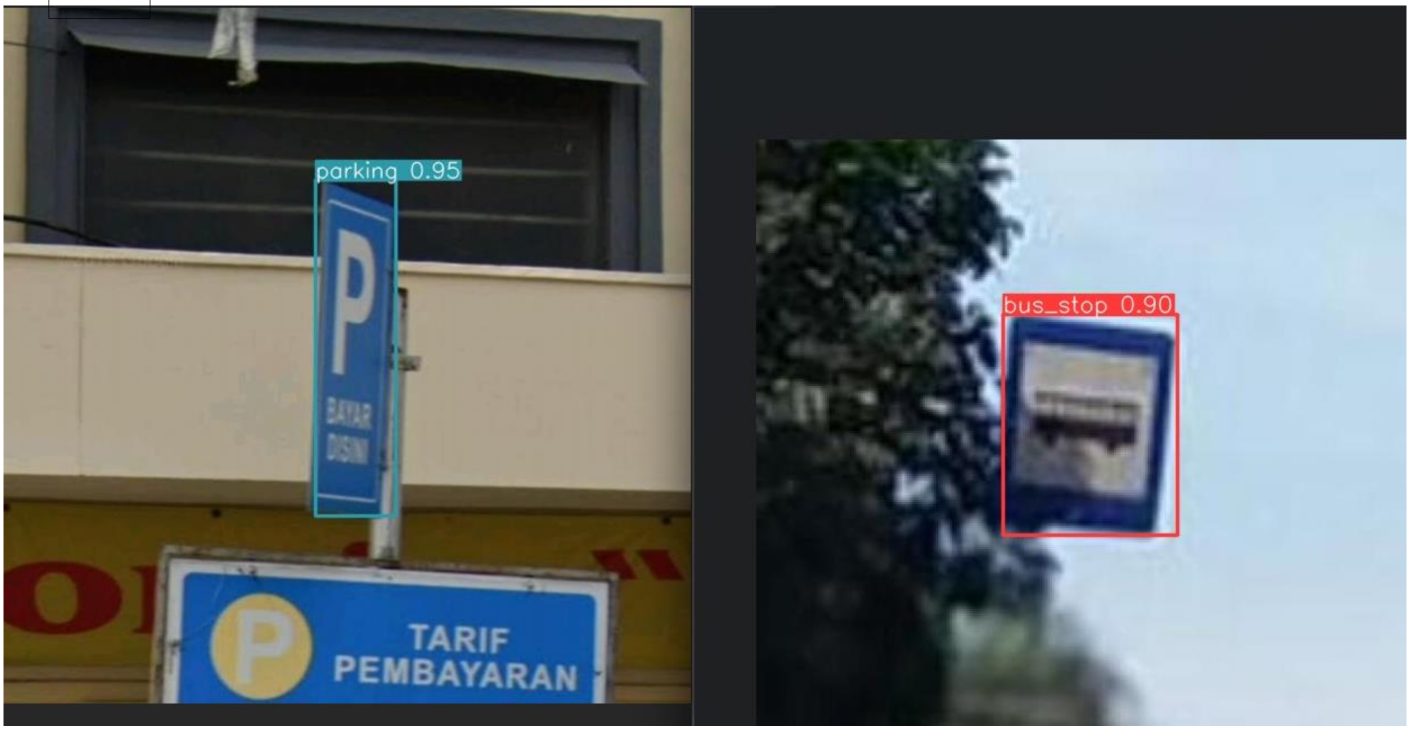

Fig. 2. Pic showing the results

Figure showing our steps: