Navigating Bangladesh's Roads: Co-DETR based approach for Enhanced Object Detection

Abdullah Ibne Hanif Arean
Dept. of Computer Science and Engineering
University of Dhaka
Dhaka, Bangladesh
abdullaharean2613@gmail.com

Mahmudul Hasan
Dept. of Computer Science and Engineering
University of Dhaka
Dhaka, Bangladesh
mahmudul.hhh@gmail.com

Nazmus Sakib Ahmed
Institute of Information Technology
University of Dhaka
Dhaka, Bangladesh
bsse1124@iit.du.ac.bd

Istiaq Ahmed Fahad
Institute of Information Technology
University of Dhaka
Dhaka, Bangladesh
bsse1204@iit.du.ac.bd

***Abstract*—In this paper, we present an effective approach for object detection in diverse driving environments on Bangladesh roads using the BadODD dataset and the Co-DETR model. Object detection in such scenarios poses unique challenges due to varying lighting conditions, different road types, and the presence of diverse vehicles. To address these challenges, we propose a collaborative hybrid assignments training scheme, Co-DETR, which enhances the performance of DETR-based detectors by improving feature learning in the encoder and attention learning in the decoder. Our approach leverages versatile label assignment manners and introduces multiple parallel auxiliary heads to provide more efficient and effective supervision during training. Additionally, we extract positive coordinates from these auxiliary heads to enhance training efficiency in the decoder. Through extensive experiments on DETR variants and Yolo models, conducted on the BadODD dataset, we demonstrate the effectiveness of our approach. Our method achieves better results, surpassing previous methods with reduced model sizes. Furthermore, our solution introduces more accuracy in diverse conditions, making it practical for real-world deployment. Overall, our work contributes to advancing autonomous navigation technology for Bangladesh roads and opens up new research avenues in the field of object detection for autonomous vehicles.**

***Index Terms*—Co-DETR, Yolo, Object Detection**

I. Introduction

Autonomous navigation technology has witnessed remarkable advancements in recent years, revolutionizing various industries and promising safer and more efficient transportation systems. In particular, the deployment of autonomous vehicles (AVs) holds immense potential for transforming the way we commute and travel. However, the successful realization of fully autonomous vehicles requires robust object detection systems capable of accurately identifying and localizing objects in complex and diverse driving environments.

Object detection serves as a fundamental component in the perception stack of autonomous vehicles, enabling them to understand and interact with their surroundings effectively.

Traditional object detection methods often rely on handcrafted features and complex pipelines, which may struggle to generalize across different environments and exhibit limited scalability. To overcome these limitations, recent research has shifted towards end-to-end trainable deep learning-based approaches, offering the promise of improved performance and adaptability.

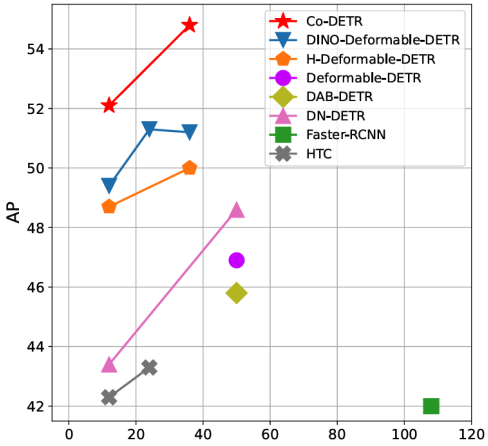


Fig. 1: Comparison analysis of Co-DETR among DETR-based Models

In this paper, we focus on addressing the unique challenges of object detection in diverse driving environments across Bangladesh roads. Bangladesh presents a distinctive setting for autonomous navigation, characterized by varying road conditions, diverse vehicle types, and challenging lighting conditions. To facilitate research and development in this domain, we introduce the BadODD [1] dataset, a comprehensive dataset specifically curated for object detection tasks in Bangladesh.

Our solution introduces the Collaborative DETR (Co-DETR [2]) model, which is a novel training scheme aimed at enhancing the performance of DETR [3]-based object detectors.

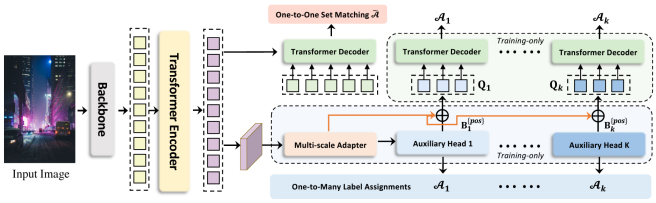


Fig. 2: Internal Workings of Co-DETR

We have evaluated this approach against YOLO [4]-based methods, demonstrating improvements in efficiency and effectiveness through versatile label assignment strategies.

It addresses issues related to sparse supervision and inefficient feature learning, thereby improving the model's ability to accurately detect objects in challenging scenarios. By combining the BadODD dataset with the Co-DETR model, we aim to provide a robust framework for object detection in Bangladesh's unique driving environments..

II. Dataset Description

A. Coverage of Bangladesh Districts

The dataset, named BadODD, covers 9 districts in Bangladesh: Sylhet, Dhaka, Rajshahi, Mymensingh, Maowa, Chittagong, Sirajganj, Sherpur, and Khulna. It includes various road scenarios such as urban and rural areas, highways, and expressways, providing a diverse representation of Bangladesh's road infrastructure. Data collection was done using smartphone cameras to ensure authenticity and to capture real-world driving conditions.

B. Image and Object Statistics

The dataset comprises 9,825 images capturing different lighting conditions and road scenarios, including day and nighttime datasets. There are a total of 78,943 objects annotated across the dataset, with 13 distinct classes representing various vehicles and objects commonly found on Bangladeshi roads. Frame selection was carefully planned to capture dynamic qualities of urban environments, with adaptive frame-rate sampling strategies implemented based on traffic densities.

C. Annotation Details

Classes were redefined based on vehicle characteristics rather than relying solely on local or globally acknowledged names. This approach aims to address the diverse types of vehicles encountered on Bangladesh roads and allows for scalability. The dataset's class distribution analysis highlights the imbalance in the occurrence of different classes, with some classes like Person, Autorickshaw, and Three Wheeler being more prevalent than others such as wheelchair, train, and construction vehicle. A comparative analysis of two object detection models, YOLOv5 and YOLOv8, was conducted, showcasing their performance in terms of mean Average Precision (mAP) scores on the dataset.

III. Experimental Setup

In our experimental setup, we utilized the Kaggle environment for training our YOLO models, leveraging the GPU P100. For the training of our DETR-based model, we utilized a high-performance workstation with the following specifications: Ubuntu 22.04 LTS, 16-core AMD Ryzen 9 5950x4 processor, 128GB of RAM, an NVIDIA RTX 3090 GPU with 24GB of memory, and a 2TB storage drive.

IV. Methodology

A. Data Preprocessing

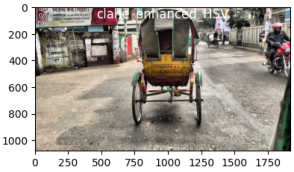
Before feeding our data into the model for training and inference, we performed preprocessing to enhance the quality of the images and facilitate better object detection. This involved applying various **Image Enhancement Techniques** to the image set of BadODD. These techniques aim to improve the visual quality of images, thereby enhancing their interoperability for both humans and machines. Specifically, we employed three different techniques:

- **Histogram Equalization:** A method used to improve the contrast of an image by redistributing pixel intensities.
- **Contrast Limited Adaptive Histogram Equalization (CLAHE):** An adaptive version of histogram equalization that limits the amplification of noise in regions with low contrast.
- **Gamma Correction:** A technique used to adjust the brightness and contrast of an image by modifying the gamma value.

These preprocessing steps were crucial for ensuring that the input data provided optimal conditions for object detection algorithms to perform effectively. Below are visual representations of these techniques applied to sample images from the BadODD dataset:

(a) Raw

(b) Histogram Equalization



(c) CLAHE

(d) Gamma Correction

Fig. 3: Different Image Enhancement Techniques

These enhanced images serve as the improved input data for our object detection models, contributing to more accurate and reliable detection results.

B. Model Selection

We adopted the YOLOv8 architecture for its advancements in object detection. YOLOv8 features an enhanced backbone network, adaptive training strategy, and improved object detection head. Its reputation for achieving superior accuracy while maintaining computational efficiency made it a compelling choice for our study.

We also explored the Co-DETR model, which represents a departure from traditional anchor-based approaches. Leveraging transformer architecture and a set prediction mechanism, Co-DETR captures spatial relationships effectively and predicts all object classes and their bounding boxes in a single pass. Its attention mechanisms enhance global context understanding, making it suitable for complex scenes. Our investigation into both YOLOv8 and Co-DETR aimed to provide a comprehensive comparative analysis of their respective strengths and weaknesses in object detection.

Our methodology entailed fine-tunig the Yolov8m model using the given dataset specifically collected for vehicle detection for Bangladesh.

C. Model Implementation

We opted for the YOLOv8 architecture due to its advancements in object detection. YOLOv8 features an enhanced backbone network, an adaptive training strategy, and an improved object detection head. Its reputation for achieving superior accuracy while maintaining computational efficiency made it a compelling choice for our study.

The training procedure involved fine-tuning the YOLOv8m model using the BadODD dataset specifically collected for vehicle detection on Bangladesh roads. We configured the training process with the following parameters:

write about codetr ..... about codetr 20 epoch mAp 0.638 and 9 epoch mAP 0.643

- **Model:** Yolov8m
- **Base Learning Rate:** 0.018
- **Momentum:** 0.933
- **Cosine learning rate:** True
- **Deterministic:** True
- **Seed:** 43
- **Evaluation Metric:** mAP

These parameters were chosen based on experimentation and best practices in the field of object detection for autonomous navigation tasks.

Before training the YOLOv8 model, we conducted exploratory data analysis (EDA) to understand the dataset better. We also applied image processing techniques to enhance visualization and improve object detection accuracy. Additionally, a validation set was created to ensure data integrity.

The YOLOv8 model was fine-tuned using the diverse BadODD dataset, covering various driving environments in Bangladesh. Leveraging transfer learning, the pre-trained YOLOv8m model was adapted to the specific task of vehicle detection on Bangladeshi roads. Although the YOLOv8 model showed competitive performance, we opted to explore the Transformer-based model, **Collaborative Detection Transformer (Co-DETR)**, for potentially higher efficiency and accuracy.

Co-DETR is a cutting-edge approach that combines transformer architecture with set prediction mechanisms, making it adept at understanding complex scenes like those on Bangladesh's roads. With its ability to capture global context, Co-DETR seemed like a promising alternative to traditional models like YOLOv8, especially in environments with diverse road conditions.

In our study, we decided to give Co-DETR a shot. We fine-tuned it using a diverse dataset that covered a wide range of driving scenarios in Bangladesh. We carefully tweaked its hyperparameters, including **backbone depths** and **query head settings**, to optimize its performance for vehicle detection. Our goal was simple: to leverage Co-DETR's advanced capabilities to improve object detection accuracy and efficiency in Bangladesh's dynamic driving environments.

By exploring Co-DETR, we aimed to push the boundaries of what's possible in object detection. Through rigorous experimentation and evaluation, we sought to validate its effectiveness in addressing the unique challenges posed by Bangladesh's roads.

D. Comparative Analysis

Table I presents the evaluation scores for YOLOv8m and Co-DETR models at different epochs.

TABLE I: Evaluation Score

Model	Epoch	Public Score	Private Score
YOLOv8m	50	0.26491	0.38444
YOLOv8m	70	0.27067	0.38853
Co-DETR	4	0.28271	0.41787
Co-DETR	9	0.2981	0.43806
Co-DETR	20	0.28929	0.42931

After testing multiple Confidence Thresholds including 0.4, 0.35, and 0.5, we found that the best results were achieved under a threshold of **0.4**. Across all epochs, Co-DETR consistently outperforms YOLOv8m, demonstrating its superior accuracy in object detection. The performance disparity between the two models increases with additional training epochs, underscoring the effectiveness of Co-DETR's architecture in capturing complex object representations over time. Particularly, the performance of Co-DETR peaks at **9 epochs** according to the provided charts.

In summary, Co-DETR exhibits promising potential for autonomous vehicle detection in real-world scenarios, exhibiting superior performance compared to YOLOv8m on the evaluated dataset.

V. Results and Discussion

Based on the analysis of the entire document, the results and discussion section encapsulate the effectiveness of the proposed approach for object detection on Bangladesh roads. Leveraging the BadODD dataset and employing both

YOLOv8m and **Co-DETR** models, the study demonstrates significant advancements in autonomous navigation technology. Through iterative model refinement and experimentation, Co-DETR emerges as a superior choice for accurately detecting objects in diverse driving environments. For instance, Co-DETR achieves a peak performance with a public score of **0.2981** and a private score of **0.43806** at **9 epochs**, surpassing YOLOv8m's performance. YOLOv8m achieves a public score of **0.26491** and a private score of **0.38444** at **50 epochs**. These findings signify a substantial step forward in enhancing road safety and efficiency, offering practical implications for autonomous vehicle development and deployment in Bangladesh and beyond.

Additionally, it's noteworthy that the study utilized 20% of the dataset for validation purposes, ensuring that images appearing in the training set do not overlap with those in the validation set, and vice versa. This meticulous validation process enhances the reliability and integrity of the model evaluation, contributing to the robustness of the findings.

These results underscore the significance of **transformer-based approaches like Co-DETR** in addressing the complex challenges of object detection, paving the way for safer and more efficient autonomous navigation systems.

VI. Kaggle Competition Results

Our team took part in the DL Enigma Challenge hosted on Kaggle. Our performance on the public leaderboard on the given data resulted in a public score of **0.43806** and in private score of **0.29810.**

VII. Conclusion

Based on our analysis, we advocate for the adoption of **Co-DETR** in object detection for Bangladesh's roads. Leveraging the **BadODD** dataset, our study demonstrates Co-DETR's superiority over traditional methods like **YOLOv8m**. With a peak **public score** of **0.2981** and **private score** of **0.43806** at **9 epochs**, Co-DETR outperforms YOLOv8m, showcasing its efficacy in real-world scenarios. By employing transformer-based architectures, Co-DETR effectively addresses the challenges posed by diverse road conditions and lighting variations. Additionally, the use of 20% of the dataset for validation ensures robust model evaluation and generalizability of findings. Overall, our study advocates for Co-DETR as a transformative approach for advancing autonomous navigation technology, offering practical implications for safer and more efficient transportation systems.

References

[1] M. Ataullha, M. N. Baig, R. Alienware, and T. I. Fahim, "Dl enigma 1.0 - sust cse carnival 2024," 2024. [Online]. Available: https://kaggle.com/competitions/dl-enigma-10-sust-cse-carnival-2024

[2] Z. Zong, G. Song, and Y. Liu, "Detrs with collaborative hybrid assignments training," 2022.

[3] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[4] J. Redmon, S. Divvala, and R. Girshick, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.