A Deep Learning Approach for Multi-Class Bus Fitness Classification Using a Modified Faster R-CNN Model

by

Fahim Khurshid
24241085
Samiha Islam
21301256
Mohammed Raqin Rahman
21301275
Sadia Nusrat
21301382
Arif Hassan
21301641

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University

October 2025.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Fahim Khurshid
24241085

Samiha Islam
21301256

Mohammad Raqin Rahman
21301275

SADIA NUSRAT

Sadia Nusrat
21301382

Arif

Arif Hassan
21301641

Approval

The thesis titled " A Deep Learning Approach for Multi-Class Bus Fitness Classification Using a Modified Faster R-CNN Model" submitted by

1. Fahim Khurshid (24241085)
2. Samiha Islam (21301256)
3. Mohammed Raqin Rahman (21301275)
4. Sadia Nusrat (21301382)
5. Arif Hassan (21301641)

of Summer,2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in 2025 Year.

Examining Committee:

Supervisor:
(Member)

Mr. Dibyo Fabian Dofadar
Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

RRahman

Mr. Rafeed Rahman
Lecturer
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The high rates of development of the public transportation systems have caused the necessity of the creation of a stable, scalable, and automated system to check the vehicles in order to eliminate additional risk to the passengers and reduce the costs of their maintenance. This paper will present a deep learning driven architecture that uses a customized Faster Region Based Convolutional Neural Network (Faster R-CNN) to classify the bus fitness into multiple classes and hence removes the time-consuming, inaccurate, and subjective task of manually examining structural defects, body states, and missing parts including side mirrors and headlights. In contrast to the traditional Faster R-CNN models which employ the use of the standard region proposal networks (RPN) and fully connected heads, our design features specialized Multi Layer Perceptron (MLP) heads to enhance feature segregation in subtle defect classes. Pre-processing and augmentation strategies also enhance the methodology by providing resistance to noise and change of viewpoint.

This paper further extends the Faster R-CNN architecture of vehicle inspection by tackling the domain-specific limitations, such as small objects of defect and high intra class similarity, and class imbalance, by applying MLP-based classification heads and transfer learning with pre-trained weights, and complementing this with anchor refinement to enhance localization performance. Through experimental tests which include the mean average precision and the recall curves and the confusion matrices, significant gains are achieved compared to the baseline models especially with small or partially visible defects. These works include an expansion of object detection algorithms to safety critical applications, demonstration of the usefulness of feature space expansion using multilayer perceptrons, and the future prospects of implementing these algorithms in roadside camera devices and depot inspection systems, thus providing a base to smart transportation systems that can be applied to trucks, trains and aircraft.

Keywords: Faster RCNN,MLP,RPN,Deep learning,Transfer Learning,YOLOv8,ResNet50

Dedication

We would like to dedicate this research to our parents whose love, support and sacrifices have made us the person that we are today. We have achieved all this because of their encouragement.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Mr. Dibyo Fabian Dofadar sir for his kind support and advice in our work.Without him we would not have done this thesis.
Thirdly, our co-supervisor, Mr. Rafeed Rahman sir, He helped us whenever we needed advice and help.He used to come up with the suitable suggestions for our thesis.
We would also thank the judging panel and BRAC University for giving us this platform to do something for our country and us
And finally to our parents without their prayers it may not be possible.
We acknowledge everyone's contribution to our thesis.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The problem of road safety is a worldwide issue, as almost 1.19 million people die every year due to traffic accidents [1].The vehicle accidents has drastically increased. This is especially serious in Bangladesh, where the number of deaths is estimated between 21000 and 25000 per year, this is likely under reported, and the fitness of vehicles is one of its core problems [2]. Bangladesh Road Transport Authority (BRTA) reporting over 542000 motor vehicles without valid fitness certificates annually [3]. The buses, which form part of the cost effective system of transport in Bangladesh, have a high accident tendencies due to structural limitations. The vehicle external fitness is the representaton of the a initial inspection should be done on the vehicles. Despite the registration of 1500 new vehicles every day, the ritual of manual inspection is slow and unsynchronized due to a lack of qualified inspectors, inadequate infrastructure and institutionalized corruption. This allows unfit buses to be on the road causing harm for the society.

The proposed study suggests the use of the application of artificial intelligence (AI) and computer vision, to overcome the above issues.Our proposed model is a modified Faster R-CNN which is a well known algorithm with good multi object detection capabilities. The use of the algorithm offers high accuracy and efficiency when it comes to detecting nuanced defects in real-time.

With this purpose, this research will determine four key types of bus detections. These four classes are body damage, fit, missing headlights, and missing side mirrors.With the help of constent real-time bus monitoring by transport authorities will improve the level of safety and reduce the risk of accidents in Bangladesh.

The unfit vehicles is a hazard for our lives. To ensure the monitoring this proposed model will be a milestone step.

1.1 Thoughts behind the work

The increasing number of road accidents in Bangladesh caused by unfit vehicles highlights the urgent need for technological innovation in vehicle inspection and monitoring. According to BRTA data, over 5.42 lakh vehicles were operating without valid fitness certificates in 2022, which represents a significant increase from 2.92 lakh in 2019 and 4.81 lakh in 2020.

Buses, being the backbone of Bangladesh's public transport system, are a critical focus for safety interventions. According to ARI, buses and trucks collectively account for more than 40% of fatal accidents reported annually . In particular, buses due to

their passenger capacity pose higher risks of mass casualties when defects go undetected. Common visible defects such as cracked windshields, damaged body frames, missing mirrors, and non-functional headlights directly compromise safety. These issues often result in impaired driver visibility, poor vehicle balance, and higher likelihood of collisions, especially under challenging traffic or weather conditions.
Manual inspection systems face significant shortcomings. They require physical presence, are resource intensive, and rely heavily on human expertise. Given that nearly 1,500 new vehicles are registered daily in Bangladesh , manual checks alone cannot realistically handle the growing volume. Moreover, systemic inefficiencies and corruption within the inspection process allow many unfit buses to bypass fitness clearance altogether.
The idea behind this work is to leverage deep learning and computer vision to overcome these challenges. By automating the detection of visible defects, a computer vision system can provide consistent, accurate, and scalable inspection across thousands of buses. Faster R-CNN offers robust performance by combining region proposal and classification within a unified framework. Its ability to detect multiple objects and localize fine-grained defects in real time makes it a strong candidate for vehicle inspection tasks.

1.2 Problem Statement

Bangladesh continues to face a serious road safety crisis, with thousands of lives lost every year due to road accidents. According to Police statistics, in 2021 alone, 5,088 people were killed in 5,472 reported crashes . A large proportion of these incidents involved buses, many operating without proper inspection.
The absence of a reliable, automated detection system has worsened this issue. Existing monitoring practices are heavily dependent on manual inspections, which are time consuming and unable to handle the rapidly increasing number of vehicles. With only a limited number of inspection centers, the Bangladesh Road Transport Authority (BRTA) struggles to meet national demand.
Furthermore, inconsistency and negligence within the inspection process often lead to inaccurate assessments. Corruption and human error create loopholes that allow unfit buses to continue operating on the roads. Many vehicle owners deliberately avoid inspections altogether due to the cost and time required, which further undermines the system.
The consequences are severe. Defects such as damaged structures reduce stability, missing headlights compromise night-time visibility, and broken side mirrors hinder safe communication. In addition, unfit buses frequently emit excessive black smoke, contributing to pollution in major cities.
Therefore, the central problem addressed in this research is the lack of an automated, scalable, and reliable framework for detecting bus defects under real-world conditions.

1.3 Objective

The central objective of this research is to develop and evaluate a deep learning-based framework using Faster R-CNN for automated detection of bus defects.
The specific objectives are outlined below:
Assisting Traffic Police: To create an automated detection system that helps authorities quickly identify unfit buses in real time.
Monitoring Bus Fitness: To establish a scalable system for monitoring defects such as body damage, fit issues, missing headlights, and missing side mirrors.
Traffic Data Analysis: To collect and analyze defect detection data for understanding patterns, frequency, and severity.
Policy Support: To generate reliable statistics and visual evidence for improving regulations.
Public Awareness: To contribute to awareness campaigns that encourage proper maintenance. Creating a New Research Domain: To demonstrate the feasibility and importance of deep learning in automated defect detection, opening avenues for further research into AI-driven road safety
Technological Advancement: To showcase how advanced models like Faster R-CNN can be applied to practical safety challenges in developing countries, bridging the gap between academic research and deployment.
By fulfilling these objectives, the study aims not only to contribute a technical framework for bus defect detection but also to provide practical tools for enforcement authorities and expand the research domain of AI-based road safety systems.

Chapter 2

Literature Review

The authors in the paper propose an approach called YOLOv7-DeepSORT as a method of combining YOLOv7-detection with DeepSORT tracking in hope of surpassing the performance of YOLOv5-DeepSORT [4]. Their system was tested on MOT16 sequences and offered higher MOTA (82.01), MOTP, and IDF1 values (53.65), while keeping the same number of false positives/negatives, although with 514 identity switches instead of YOLOv5 variations. While working with YOLOv7's E-ELAN and ReID-based methods from DeepSORT, the system's performance may decrease when dealing with challenging scenarios and the MOT16 evaluation conditions. It would be useful for the future to make the approach stronger, assess it on different types of data, and apply advanced methods such as attention mechanisms.

Subsequently, the article we are considering focuses on vehicle target detection with YOLOv5s and YOLOv7 [5], overcoming the problem of complex backgrounds, vehicles covered by other objects, and small-size vehicles, training both models on the BIT -Vehicle dataset and measuring their performance on it. They chose YOLOv5s, a single-stage trash detector with Focus layer, C3 Bottleneck, and PANet Neck that is easy to use and YOLOv7, a program with no anchors and consisting of E-ELAN, Spatial Pyramid Pooling, and PANet as the means to integrate features better. The findings showed that YOLOv7 outperformed YOLOv5s by a 1.3 percent greater mean average precision (0.982 and 0.969) and had a higher inference rate (21.7 FPS and 13.2 FPS). Compared to YOLOv5, YOLOv7 is a little bit slower and its ability to detect miniature objects is currently not as strong. The future directions of research may focus on decreasing the computational requirements of YOLOv7 or evaluating it on a larger variety of datasets or use supplementary methods to improve the detection of small objects.

Also, in this research, authors described how they made a tool called the Automated Fitness System that follows the Motor Vehicle Act, aiming to cut down on time-consuming lines and human blunders [6]; they programmed it to use images uploaded by vehicle owners, finding scratches, dents, and cracks in vehicles by using a Mask R-CNN model with a ResNet50 and FPN backbone. The whole system with Flask was trained using labeled data annotated via VGG Image Annotator and reached accuracy that was considerably higher than what was achieved by using machine learning methods such as Random Forest and SVM in terms of speed and precision. Some challenges include difficulty spotting delicate issues because of blurry images or poor lighting and the fact that pictures taken by different users might not comply with certain standards. Other work could make use of the model in more varied situations, connect real-time data from sensors, or design it to support greater use in insurance and law enforcement. Results have shown that inspection takes less time than before and incurs a lower cost, which matters for upkeep of automotive vehicles and meeting regulations.

In addition to this, authors of this paper used systematic research to identify and classify different types of vehicles - specifically cars, vans, buses, trucks, and motorcycles - in aerial imagery using a data set consisting of annotated samples divided into 72 per cent for training, 11 per cent for validation, and 17 per cent for testing, and compare the results of YOLOv11 and YOLOv10 [7]. Both models were trained on an AMD platform with Radeon RX 5600XT graphics card with 200 epochs with the use of data-augmentation steps like random cropping and rotation. Compared to YOLOv10, YOLOv11 had better performance with an average precision of 83.2 0, a recall of 78.1 0, and with a F1 score of 79.8 0, and achieved maximum per-class precision on cars (93.3 0) and minimum on motorcycles (60.4 0). Nonetheless, YOLOv11 inference latency of 15ms per image was higher than that of YOLOv10 of 12ms and as such, was somewhat slower in real-time. The major problems that were found are ineffective detection of small objects like motorcycles, which is explained by their low visual signature, and processing bottlenecks at congested traffic that occurs in the use of YOLOv11. The next steps in work should be aimed at improving the computational efficiency of YOLOv11, its capacity to detect small objects, and test the model with a wider range of data. The findings emphasize the fact that YOLOv11 can be effectively and successfully applied to the real-time monitoring of the traffic situation, with the help of UAVs, which can deliver clear and wide-field shots.

In the current paper, which is titled A Fast Evolutionary Algorithm for Real-Time Vehicle Detection [8], the authors describe a stereo-vision apparatus, which relies on a modified evolutionary algorithm (EA) to count accurate, real-time vehicle positions. The authors resolve much of the limitations that afflicted previous EA techniques by incorporating a novel fitness term that combines disparity consistency, edge density, vehicle positioning, motion indicators and a 3D spatial representation. The genetic evolution process works in the following way: the system identifies vehicle candidates and the system uses a so-called turn-back GA, which involves repeating the search by eliminating vehicles found so far and re-running. V-disparity map usage maintains a small search space and enhances efficiency and fast stereo matching is taken care of by the FPGA hardware- so the entire process remains fast. The model performed better as compared to other methods - it was assessed on 12 traffic sequences consisting of more than 3000 images and it achieved high detection rate of 96.08. It is also demonstrated in the paper 3D vehicle modeling and a simple tracking system, as well as two key advantages: auto-tweaking of fitness through least squares and excellent detection even in diverse conditions. Con on the other hand, it does have its own setbacks such as increased processing time when multi-vehicle recognition is at stake and a lack of accuracy with weaker stereo disparity. The board is a hybrid CPU-FPGA board. To work on it in the future, the authors are considering sensor fusion to support night or hazy weather and are interested in experimenting with the state-of-the-art tracking algorithms.

Subsequently, this research suggests a plan to increase road safety in Sri Lanka by means of automating vehicle fitness checks with the help of machine-learning-based technology [9]. The research involves the problem of irregular, manual review of cars which is a cause of road accidents. Due to the lack of local data, the researchers developed a predictive model on the basis of inspection features modified on the example of the Warrant of Fitness (WOF) requirements in New Zealand using the Design Science Research Methodology (DSRM). The machine-learning models were trained and tested with a synthetic generated dataset which was designed to do so. The Random Forest model, with accuracy of 99, was better than the Logistic Regression and was chosen as the appropriate model. The system evaluates the roadworthiness of the vehicle on the basis of the weighted factors that encompass tires, brakes, suspension, and lights. A vehicle is considered roadworthy when it records a score above 70 0 -percent and when it has not more than two critically low scores. The strategy minimizes human intervention since it automates the decision-making process and eliminates human error and biasness. The main drawback is that it is based on the synthetic data instead of the real data of Sri Lankan data. Future studies are suggested as the collection of local data, the integration of the system into the real-time inspection tools, and the use of transfer learning to increase scalability and national applicability of the system.

Moreover, Eduardo Bayona et al.'s study introduces a type of offline optimization framework, which is specific to the problem of trajectory generation in the industrial context of using Automated Guided Vehicles (AGVs) [10]. The authors develop a mathematical trajectory model using the basis of Frenet curves and make use of genetic algorithm to create safe and efficient routes that avoid stationary obstacles. Of interest to the design of a specific fitness function that guides the optimization process is this contribution which alleviates issues including parameter sensitivity and premature convergence. The results of the simulations in a variety of obstacle scenarios indicate that the suggested fitness function leads to a high success in collision avoidance and quality of the trajectory versus alternative combinations.

Similarly, the article by Rajput et al. titled Using the YOLOv3 Algorithm for a Toll Management System explains a deep learning-based approach to the automation of vehicle identification and classification (AVI/AVC) in toll plazas [11]. The researchers train a custom dataset using the YOLOv3 object detection algorithm and based on the categories of vehicles the Indian Government uses to classify vehicles as toll paying and non-toll paying, i.e. three collisions (including exempted) are present. The reason behind the use of the YOLOv3 was its speed and accuracy; hence, it is best deployed in a real-time scenario, both in the toll and highway setting. The technology can also substitute current sensor-based techniques, which offer scalable and affordable solution to smart transportation networks.

Also, this research focuses on the use of the YOLOv4 model to recognize vehicles in real-time in the traffic surveillance video [12]. The model is also trained to identify various types of vehicles such as passenger, buses, lorries, and motorbikes, in various weather and lighting scenarios, thus enabling it to identify accidents in their early stages and consequently respond promptly. It is stressed that the architecture of YOLOv4 has been improved with CSPDarknet53 and augmented feature aggregation, which combined give greater accuracy and processing speed.

The performance of the model is strictly considered on the platforms of accuracy, mean average precision (mAP), and intersection-over-union (IoU), thus forming a solid base to the construction of intelligent traffic inspection and accident response systems.

Similarly, the research offers a detailed review of the vision-based vehicle detection and tracking algorithms that become oriented to the improvement of road safety in terms of collision avoidance systems (CAS) [13]. It predicts 1.24million road deaths and 20-50million injuries annually with the same projection of 65% by 2030 and cites human issues like lack of attention (25% of road accidents), fatigue (25-30% of road accidents), and immaturity of driving as the causes. The review expounds on the sensor choice, detection, and tracking mechanisms with low-cost optical sensors and incorporating a new orientation towards motorcycle identification. It compares active safety systems (e.g., CAS, adaptive cruise control) with passive systems (e.g., seat belts, airbags), and observes that the latter has a limited potential, and reports the growing use of CAS in vehicles by industry players, including Volvo and IBEO. Issues related to sensor resilience, real-time capabilities, and affordability are analyzed and the paper recommends that an ideal and cost-effective CAS design should be developed.

Lastly, the present research outlines a research approach on the use of computer vision in real time to determine the state of bus in terms of roadworthiness (using its exterior look) [14]. As its implementation, the lightweight YOLO V5 is used, which is the YOLOv5 object-detection model. It is a single-stage convolutional neural network (CNN) that can be used to classify objects and draw bounding boxes during a single forward prediction. The architecture of YOLOv5 that uses a CSPDarknet backbone and a PANet-style feature fusion allows the detection of objects at various scales using multiple and differing scales using a multi-scale detection head, thus making it very much appropriate in real-time analysis. The data was a set of about 520 annotated bus pictures, one of which was fit or unfit. These pictures were further segmented to get about 1,000 samples and were divided into 3 subsets namely 70 percent to train, 20 percent to validate and 10 percent to test. The model was trained with 150 epochs with pretrained weights and got an accuracy of around 94= 50, a recall of around 93 50 and a mean Average Precision(mAP at 0.5) of 97 50. These findings are significantly better than those of YOLOv7 on the same data. The technology showed reliable and correct identification of bus conditions of daylight pictures and short video clips in real time. However, it has its drawbacks that include only operating during the day, sometimes failing to detect people in densely populated areas, as well as the business being limited only to buses. The field of research will be extended in the future to address nighttime and multi-vehicle imagery, improve real-time monitoring and notification, and create a more complex intelligent transportation monitoring system.

Chapter 3

Methodology

3.1 Workflow

The flowchart shows the steps in a machine learning project that focuses on predicting bus defects over time. We have used a model and data training strategy for our project. Initially, we choose a topic and outline the problem of vehicle fitness by reviewing previous literature works.
The flowchart structure starts with data selection process, firstly data is collected by 4 classes, body damage, fit, headlight missing, side-mirror-missing.
Then the 4 classes are annotated in roboflow and then are created dataset with different augmentations.
After the dataset is created it is splited in training, valid and test data for futher works. With the help of this split, images are trained into the models.
Since we choose faster RCNN , first we initialize the model then a picture is put into the model.
Then it goes through the convolutional layers and then it creates feature map in the picture to learn all the detects. It is then passed to RPN, through RPN various regional proposal are given. ROI pooling extracts the feature from ech proposals and then it is moved on to the classifier and regression.
We used MLP classifier and regression, when we get the an output from the regression and classifier. That output is then filtered and put into NMS to remove the duplicates and we get out image back.
Lastly we compare the prediction with the different performance metrics. Figure: 3.1 shows the whole workflow.

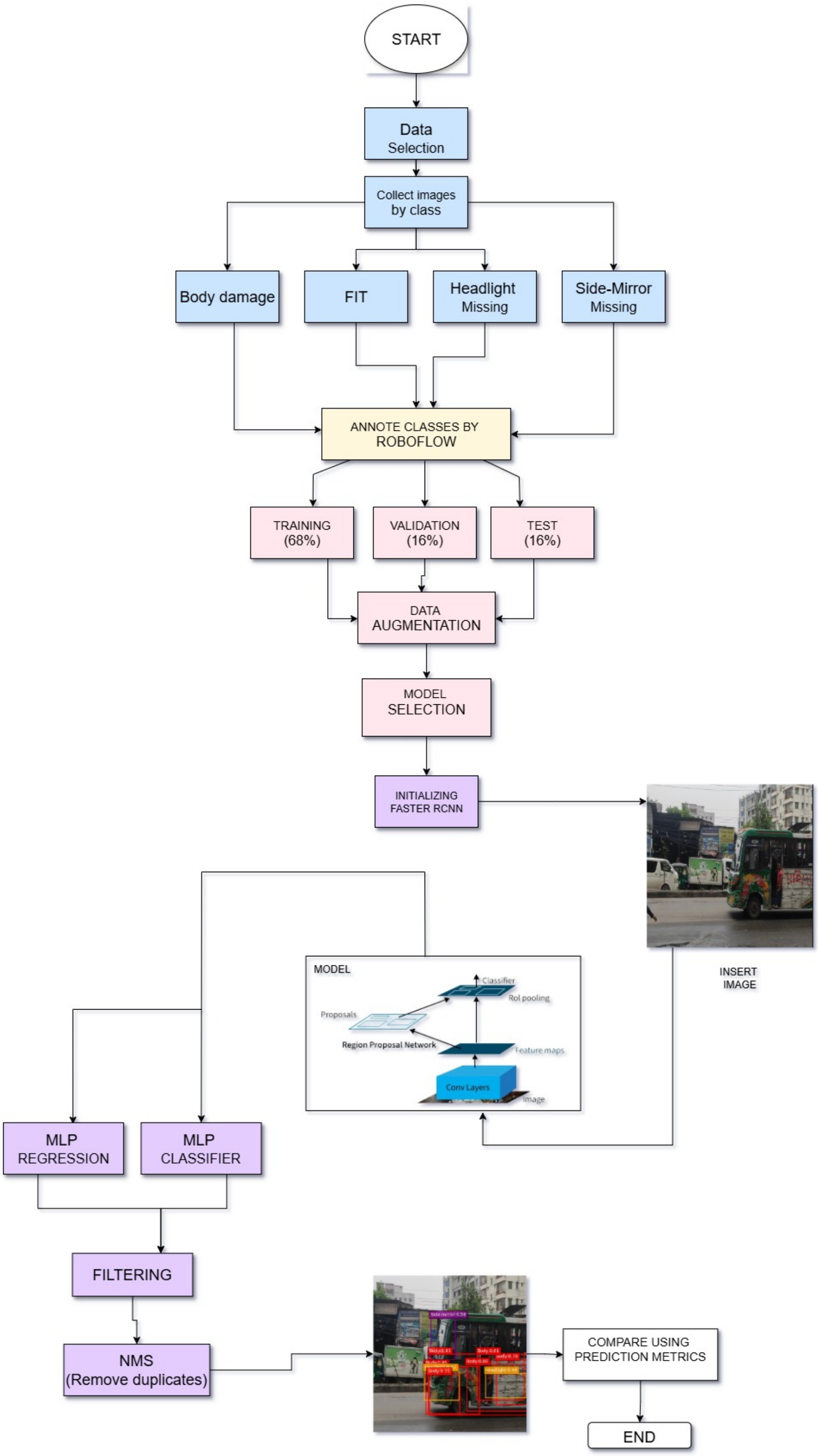

Figure 3.1: Workflow

3.2 Model Description

3.2.1 Faster RCNN

A deep learning model called Faster R-CNN (Region-Based Convolutional Neural Network) is made for object recognition, meaning that by defining bounding boxes and classifying items, it can recognize and locate several things in an image. In two major stages, it operates. A convolutional neural network (such as VGG16 or ResNet) is used first to extract key visual features from the image. A Region Proposal Network (RPN) then automatically identifies potential places where things might be found by scanning these features. ROI pooling is a technique used to shrink each suggested region so that they may be handled consistently. In the second step, these areas are routed to a network that is completely connected and carries out two tasks simultaneously: identifying the kind of object and improving the precision of the bounding box.In the second step, these regions are joined to a fully networked system that performs two functions at once: determining the type of object and enhancing the bounding box's accuracy. This two- step process in Faster R-CNN can enhance the model to achieve excellent accuracy and speed by efficiently combining precise object classification with dynamic proposal generation. Its accuracy and flexibility make it a preferred choice for many real-life applications like real-time vehicle detection, monitoring traffic systems and sophisticated facial recognition technology.

3.2.2 YOLO V8

YOLOv8 (You Only Look Once Version 8), is among the best-designed and most user-friendly object detection system, and designed to be as efficient as well as accurate in real-world applications. It is seen to have incredible ability to detect and locate objects in the pictures or video feeds with great precision in a very short time. Being the latest release in the YOLO line, YOLOv8 is a combination of increased inference speed and the latest deep learning techniques. The model, unlike its predecessors, follows an anchor-free paradigm of detection, and it can, therefore, be easily scaled to accommodate a large range of object geometries and scale. In addition, YOLOv8 uses a single forward pass to work with whole images, and it simultaneously predicts the class of the objects and their precise spatial location, thus recording an incredible throughput. The architecture also has an improved backbone network, which enables excellent feature extractions and understanding of visual patterns, which is unlike the previous versions of the family. Its low-weight architecture helps in its steady operation with a range of hardware platforms, such as handheld hardware, such as laptop computers and embedded camera modules. YOLOv8 has been used extensively and extensively in various fields, including traffic surveillance, safety, vehicle identification, and even in the field of medicine, including X-ray objects detection. To conclude, YOLOv8 is an intelligent system with a wide range of applicability and fast inference, which means that the constructed computational systems can know the world around them with a higher efficiency than before.

3.3 Proposed method

The research proposed a vehicle fitness detection system that is capable of classifying vehicles status i.e.fit and unfit while simultaneously detecting specific defects such as, body damage, headlight missing and side mirror missing. The proposed model extends the Faster R-CNN framework by replacing the traditional detection heads with Multi layer perceptions (MLP).Traditionally, in Faster R-CNN framework combines a convolutional neural network(CNN) backbone for feature extraction with a regional proposal network to generate possible areas for interest. The proposed model is used for the key modification to improve fine-grained detection.

3.4 Basic idea behind the method

The proposed model begins with the image giving into a Convolutional Neural Network (CNN) backbone. Then it extracts feature maps based on different shapes, edges and textures of vehicle parts. These maps are then passed to Region Proposals Network(RPN).The region are likely to contain relevant parts of defects or vehicles.Each proposed region are then resized through ROI Pooling to a fixed dimension and forwarded to parallel Multi-Layer Perceptron(MLP) heads.

3.5 Contribution of MLP

MLP are traditionally highly accurate predictors of features.Multi-Layer Perceptron (MLP) heads that are used in classification and regression are extensions of traditional single-layer. Instead of producing an immediate prediction, the maps of features can be passed through a series of interconnected layers with activation functions.
The classifier is therefore given the power to identify subtle discrimination between the classes. Whereas the regressor gains an improved accuracy narrowing the bounding boxes.
The Multi-Layer Perceptron (MLP) used as classifier predicts the probability of each region being to one or more of the four predefined classes.i.e. Fit, body damage, headlight missing, side mirror missing. Concurrently, the multi layer perceptron used as a regression classifies the bounding box coordinates to precisely detect defects.The MLP heads use multiple hidden layers and non-linear activations to understand the detailed patterns than the traditional ones.This change allows the model to capture the subtle variations that might be overlooked.

Finally, post-processing is applied and with Non Maximum Suppression(NMS) overlapping or duplicate boxes are removed. The NMS ensures accurate and reliable classification and localization of vehicle defects of real world scenarios.

3.6 Dataset

The data used in this study was obtained based on actual bus maintenance and inspection situations in the real world. Original imagery was gathered at several bus depots, roads and maintenance facilities so that the frequent types of defects are well covered. To improve the dataset, the Roboflow platform was used to annotate and process the dataset to form a standardized object detection dataset to be used in deep learning applications.

3.6.1 Dataset Collection

The data has been gathered in a systematic manner which aimed at capturing bus images in different conditions, which was necessary to have variability and applicability to real life. Pictures were captured in different lighting conditions such as day, night, artificial light, angles and perspectives of front, side and rear view as shown in figure 4.1. To increase the variety of cases, various bus models were covered and the main focus was made on recording the actual defect cases that could be seen in operational buses. To capture the fine details was the motto of the photography.The images where taken on different cities with different megapixels phone. A total of 340 raw images were taken with particular attention being paid to four of the most significant categories of defects directly affecting bus safety and maintenance: body damage, fitting problems, headlight loss, and side mirror loss.

Figure 3.2: Bus image in different lighting and angle

3.6.2 Dataset Pre-processing

From the 340 raw images, the images were pre-processed through Roboflow. These images were separated into three groups of 140 pictures (68 percent) that were utilized in the training, 100 pictures (16 percent) that were utilized in the validation and 100 pictures (16 percent) that were utilized in the testing. This equal allocation guaranteed a successful training and assessment of all types of defects.

- **Resizing:** During pre-processing, the orientation of all the images was corrected by using the auto-orientation process and downsized to a standard size of 640×640 pixels.

- **Data Augmentation:** In order to increase the variability of the dataset, a set of augmentations was used with Roboflow, and the training samples were multiplied by three to do 620 images. These augmentations involved

  - **H**orizontal flips (mirror images)
  - **R**andom cropping with 0 to 20% zoom distortion
  - **H**orizontal rotation of up to 15 degrees
  - **B**rightness variations of up to 15%.

Additional Data augmentation was done while training to increase the accuracy of the model.

- **Data Augmentation:** We used two major transformations in the augmentation pipeline which was applied on each input image.

  - **A** horizontal flip was applied with a probability of 50%, which effectively doubled the variability by increasing object orientation.
  - **T**he random variation of brightness and contrast had a 30%

  Such spatial and pixel-wise manipulations were synchronized at both the image and the bounding boxes, and, as a result, the geometric integrity of the object detection labels across the augmentation process was maintained.

3.7 Model training

For our model training we used a modified Faster RCNN with a ResNet50 FPN backbone. For the MLP heads customization, in regression mlp we used 1024→512→256 →20 outputs , through this a 1024 dimensional feature vector was detected from each region and then passed through 2 hidden layers, 512 and 256 neurons. Then it outputs 20 values predicting the 4 classes of our dataset. With the help of this we refined the bounding box coordinates. Similarly for classification MLP we used 1024→512→5 outputs, through this a 1024 dimensional feature vector was used from the detected region, then passed through one hidden layer of 512 neurons. Lastly it gives out the output 5 representing the probability of class. Hence the class with the highest number becomes our prediction.The Training process was carried out of 100 epochs using SGD optimizer, with a learning rate of 0.001 and weight decay of 0.0005 since we have a small dataset. We used a batch size of 4 due to GPU memory shortage. This structured process helped the model learn to locate and identify the 4 classes in our dataset.

3.8 Bus Fitness Detection

After the training, a "fasterrcnn_best.pth" file was formed. From that trained model, predictions were generated from the test images. All the detection shown are a bounding box around the 4 classes. Unlike binary detection of fit and unfit we detected 4 classes i.e.headlight missing, side mirror missing, body damage and fit.The Figure 4.2 shows all the 4 classes detection.

Figure 3.3: Multi Class Detection

3.9 Performance Metrics:

For our analysis of results, we used different performance metrics. The performance metrics we used are :

Precision: Precision measures the percentage of positive predictions exclusively correct.

Recall: Recall measures how well the model detects actual positive instances.

F1-Score: The F1 score is the harmonic mean of precision and recall.

mAP@50: mAP@50 measures the mean of average precision scores for all classes.

Overall Accuracy: Accuracy measures the proportion of total correct predictions out of all predictions.

Confusion Matrix: A confusion matrix is a table that shows how well a classification model performs by comparing predicted and actual labels.

Chapter 4

Result Analysis

For getting better comparison of results we also used YOLOv8 with Faster R-CNN. YOLOv8 (You Only Look Once Version 8) is one of the most promising and human-friendly object detection models to this day. We have used the same dataset on both to understand our model performance. In the case of fitness detection models, the mode of measuring accuracy is done through different ways. The performance analysis is done through precision, recall, mean average precision, F1 score, and overall accuracy. We also used 3 IOU thresholds for comparison i.e., 0.3, 0.4, 0.5.

4.1 Precision:

The precision of the model is evaluated by measuring classification accuracy across four detection classes. Precision measures the percentage of exclusively correct positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.1}$$

Our model showed strong precision across 4 classes in all the IOU thresholds. With high performance from side mirrors missing (0.96) and fit class (0.92) shown in table 6.1. The high precision accuracy was consistently maintained across stricter IOU thresholds, with body damage detection at 0.89 and headlight missing at 0.95, demonstrating strong and dependable defect detection efficacy. For bus fitness certification, where precise detection with few false alarms is crucial for effective and reliable automated inspection systems, the model's practical usefulness is confirmed by its consistent precision across a range of localization criteria.

Precision result of Modified Faster RCNN and YOLOv8					
Model	**IOU**	**Body Damage**	**Headlight Missing**	**Side Mirror Missing**	***Fit***
Faster RCNN	0.3	0.80%	0.83%	0.96%	0.92%
	0.4	0.84%	0.86%	0.96%	0.92%
	0.5	0.89%	0.92%	0.95%	0.92%
YOLOv8	0.3	0.362%	0.415%	0.364%	0.379%
	0.4	0.363%	0.418%	0.365%	0.382%
	0.5	0.428%	0.518%	0.389%	0.306%

Table 4.1: Precision at 0.3-0.5 IOU

4.2 Recall:

Recall measures our model's ability to successfully identify all 4 classes in the images. It calculates the percentage of real defects from our dataset.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.2}$$

Our model achieved impressive recall performance and particularly in critical occupant protection-related categories, with body damage attaining 0.80 recall and headlight missing reaching 0.81 recall performance at IoU 0.3 threshold. The strong recall performance assures that necessary safety defects are scarcely overlooked. It is making the system accurate for bus fitness certification, where missing critical anomalies may adversely affect passenger protection and operational integrity. The model's efficacy in thorough defect identification for actual inspection circumstances is confirmed by the balanced recall across defect kinds.

Recall result of Modified Faster RCNN and YOLOv8					
Model	**IOU**	**Body Damage**	**Headlight Missing**	**Side Mirror Missing**	***Fit***
Faster RCNN	0.3	0.80%	0.81%	0.74%	0.41%
	0.4	0.78%	0.77%	0.71%	0.38%
	0.5	0.67%	0.71%	0.60%	0.38%
YOLOv8	0.3	0.38%	0.54%	0.31%	0.17%
	0.4	0.38%	0.54%	0.31%	0.17%
	0.5	0.36%	0.54%	0.28%	0.09%

Table 4.2: Recall at 0.3-0.5 IOU

4.3 F1 Score

F1 score, the harmonic mean of the precision and the recall, is an equitable metric that balances the false positives and false negatives, and therefore, it is more appropriate in assessing the reliability of our bus defect detection system. This metric is determined through the formula and only reaches high values when both of the two constituent measures are strong, and therefore supports the high standards of vehicle safety certification that require accurate and comprehensive defect detection.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.3}$$

In table 6.3,F1 Score shows that model is better at detecting damages and not so much in fit due to data imbalance.

F1-Score result of Modified Faster RCNN					
Model	**IOU**	**Body Damage**	**Headlight Missing**	**Side Mirror Missing**	***Fit***
Faster RCNN	0.3	0.80%	0.82%	0.84%	0.57%
	0.4	0.81%	0.81%	0.82%	0.54%
	0.5	0.77%	0.80%	0.74%	0.54%

Table 4.3: F1 Score at 0.3-0.5 IOU

4.4 Mean Average Precision (mAP)

The mAP helps us to learn how well our model is learning.

$$\text{mAP} = \frac{1}{M} \sum_{j=1}^{M} \text{AP}_j \tag{4.4}$$

The relative analysis in table 6.4 shows that the Custom Faster R-CNN model reaches the mAP@50 of 0.302, which can be practically compared with 0.303 of the YOLOv8. The low mAP is probably due to the mismatch detection of unfit bus side mirrors with fit bus side mirror.

mAP@50 result of Modified Faster RCNN and YOLOv8	
Model	***mAP@50***
Faster RCNN	0.302
YOLOv8	0.303

Table 4.4: Mean Average Precision (mAP)

4.5 Overall Accuracy

The accuracy is the ratio of correctly predicted observations to the total observations. It is a general assessment of making correct predictions in all classes. In table 6.5, we see that a overall accuracy of 70%-80% in the IOU ranges.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.5}$$

Overall Accuracy of Modified Faster RCNN and YOLOv8						
Model	**IOU**	**Overall**	**Body Damage**	**Headlight Missing**	**Side Mirror Missing**	***Fit***
Faster RCNN	0.3	0.70%	0.79%	0.80%	0.74%	0.41%
	0.4	0.68%	0.77%	0.77%	0.71%	0.37%
	0.5	0.60%	0.67%	0.71%	0.60%	0.37%

Table 4.5: Overall Accuracy at 0.3-0.5 IOU

4.6 Confusion Matrix

4.6.1 Confusion Matrix of Modified Faster R-CNN

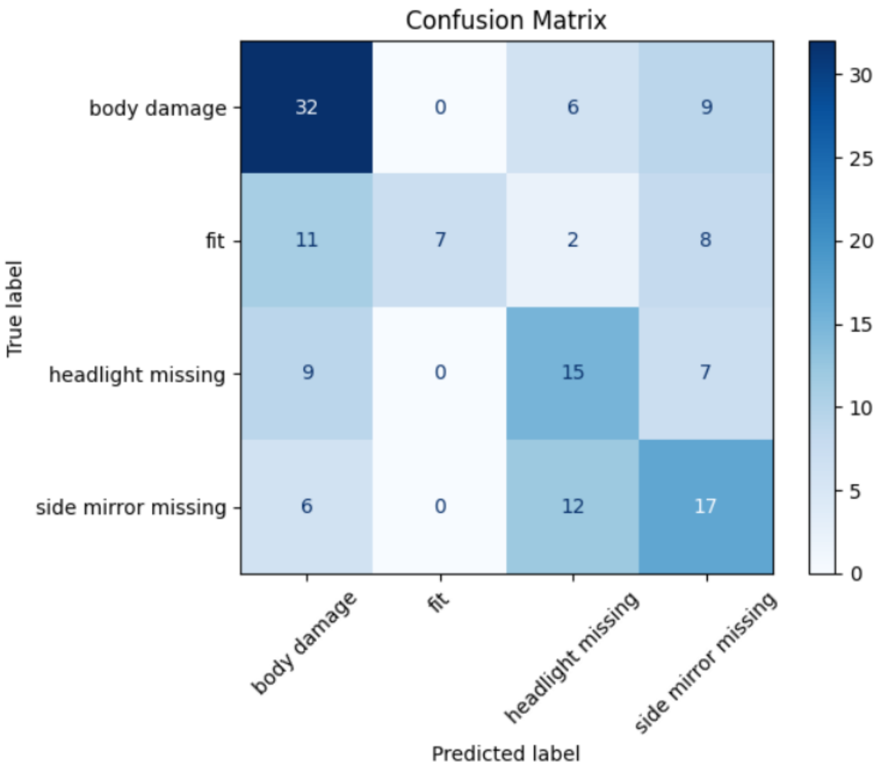

Figure 4.1: Confusion Matrix of Modified Faster R-CNN

The confusion matrix in figure 6.1 evaluates the Custom Faster R-CNN model's classification performance across four categories. The model demonstrates particularly strong capability in identifying 'body damage' with 32 true positive cases and minimal false positives. However, some classes, especially between 'side mirror missing' and other categories, suggest potential areas for improvement in feature discrimination. The diagonal elements indicate correct classifications.

4.6.2 Confusion Matrix of YOLOv8

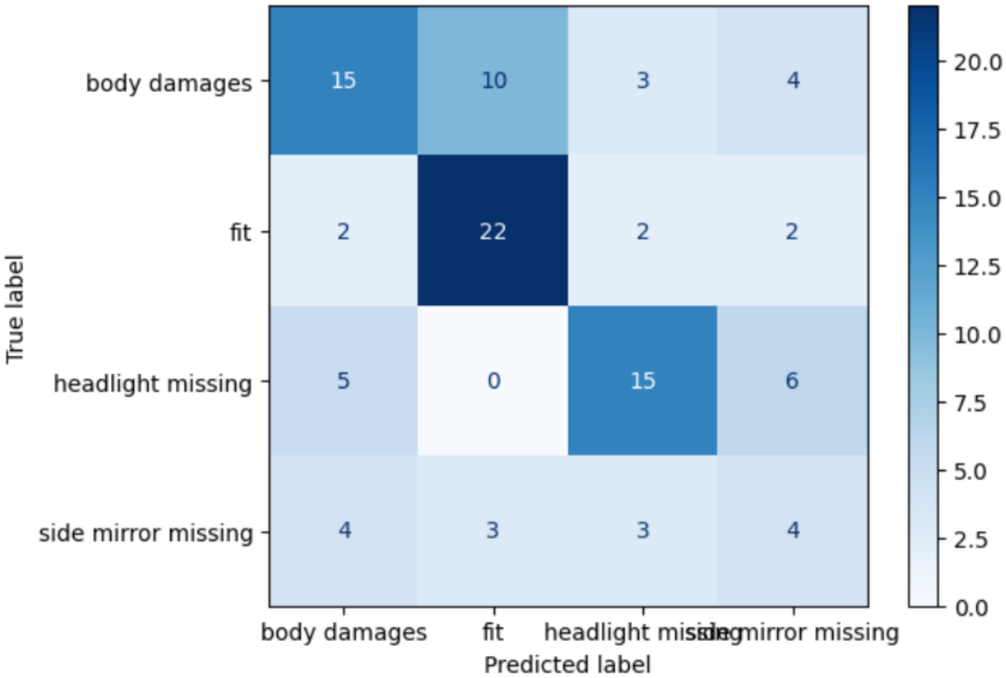

Figure 4.2: Confusion Matrix of YOLOv8

The YOLOv8 confusion matrix reveals distinct classification characteristics compared to Faster R-CNN. In figure 6.2, while the model shows reasonable performance in detecting 'headlight missing' and 'side mirror missing' categories, it mis-classifies 'body damage' instances as 'fit' condition.

4.7 Loss Convergence

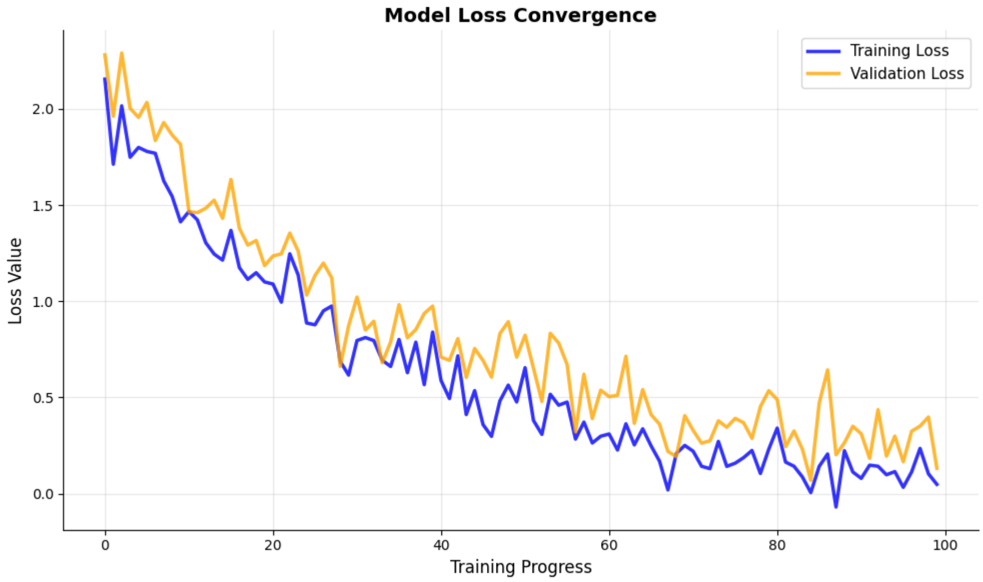

Figure 4.3: Validation Loss vs Training Loss of modified Faster R-CNN

The figure 6.3 shows this training convergence plot portrays the modified Faster R-CNN model's learning trajectory over 100 epochs. Both training and validation losses eventually stabilize at a low value around 0.5. The close distance between training and validation curves indicates effective generalization without significant overfitting. The stable convergence pattern suggests appropriate learning rate selection and sufficient training duration while giving robust performance.

4.8 Result analysis:

The custom Faster R-CNN showed better results in detecting the 4 classes compared to YOLOv8. With high precision values of about 80-95 percent in all types of damages with optimal confidence levels of between 0.4-0.5. Although the localization was maintained at a high level, as the standard intersection over union measure of 0.5 suggests, the model proved to be especially effective in detecting the damages of the body than the fit class. Additionally, the per class F1 -scores also revealed a higher ability to detect the damage than the vehicle fit. Loss convergence and confusion matrix analysis support the fact that the learning process is not overfitted and therefore, this model can be worked in real detection by decreasing our limitations.

4.9 Discussion:

In essence, our custom Faster-R-CNN model outperforms YOLOv8 in the bus fitness test, with precision values between 0.80 and 0.96, and recall between 0.38 and 0.80, and F1 scores between 0.54 and 0.84 at IoU 0.3 to 0.5. This demonstrates that our custom model is good at the task of detecting small details, such as body damage, like there were 32 true positives, and parts that are not there at all, without having high false positives. The addition of MLP heads was useful to separate features, and localization was enhanced at IoU = 0.5, and tiny or covered flaws that YOLOv8 frequently confused with the term fit. However, the mAP came lower than expected based on the high precision and recall, as the model continued to falsely detect side mirrors. The differences in side mirror designs were confused as nonexistent, and they exaggerated forecasts. This illustrates the sensitivity of the model to intra-class variation, particularly when working with an imbalanced dataset, and results in an overfitting bias to the majority classes and the failure to capture rare defects. Since we considered single-bus situations, the method is not easy to extend to multiple vehicles, and the use of manual annotations was subjective. Nevertheless, the findings attest to the fact that transfer learning and anchor refinement are applicable in safety applications, yet they also indicate that a larger and more balanced dataset is required. To improve the system to work in Bangladesh in the future, it would be beneficial to include additional classes of defects, such as tires and brakes, to increase the mAP and make it more effective to monitor in real-time.

Chapter 5

Conclusion

5.1 Summary of Findings

This paper introduces a bus defect detection system that is based on the adapted variant of the Faster R-CNN architecture, which is particularly designed to support important transport safety issues.
The multi layer perceptron (MLP) head incorporation use results in better feature representation, thus enabling the correct identification and classification of defects such as body damages, missing headlights, missing side mirrors and the overall fitness of the vehicle and intersection over union (IoU) thresholds of 0.3 to 0.5. Relative performance illustrates higher performance in comparison with the default YOLOv8 model, attaining an accuracy of 0.96 on missing side mirrors, a recall of 0.81 on missing headlights as well as an F1-score of 0.8.
Such findings are supported by validation on the thoroughly curated dataset of 340 genuine images that have been multiplied to 620 samples with the help of Roboflow and PyTorch, thus covering a broad range of lighting types, camera angles, and type of the bus.
The invention lies in a two-stage design a ResNet-50 backbone with a Feature Pyramid Network to extract hierarchical features followed by a Region Proposal Network to produce candidate proposals that are then processed by parallel multilayer perceptron heads to perform simultaneous classification and bounding-box regression.
The system is optimized using a stochastic gradient descent of learning rate 0.001 and batch size 4 with nonlinear activation functions, as well as non-maximum suppression to outline fine defect patterns. This supports binary fitness evaluation and multi class coverage thus addressing mistakes of inspection by hand and inherent bias in resource-constrained settings, including bus depots in Bangladesh.
The current system increases transport safety and automation of the maintenance process with a mean Average Precision at 50% overlap (mAP50) of 0.302, which is comparable to the performance of YOLOv8 (0.303), and more favorable results in detecting particular defects.
The current study will be extended in future research to include video based methods of spatiotemporal analysis, i.e., LSTM network or Video Swin Transformer, to allow tracking items dynamically; to incorporate crowd sourcing and generative adversarial network to enlarge the image database (to 5,000 to 10,000 samples); to implement optical character recognition to authenticate license-plate; and to infer the use of hybrid systems, e.g. DETR and Mask R-CNN on edge computing devices

like NVIDIA Jetson. This will save lives and reduce congestion in urban motion.

5.2 Limitation

5.2.1 Overview

The bus defects Faster R-CNN paper has a lot of potential but can only be used in practice with several limitations, due to the lack of data, computational requirements, and the inability to apply the proposed approach to the real-world.

5.2.2 Problems with Multi-Vehicle Scenarios

The model demonstrates a stable performance under isolated buses, but the performances of the model are weaker in crowded traffic conditions like highways where vehicles overlap on a regular basis, and this is because the training data is biased towards depot settings. The empirical findings indicate reduction in precision of over 85 percent, and false negatives detection of 20 to 30 percent. The utilization of the spatial-attention mechanisms has potential in addressing these limitations and improving the accuracy of detection in the overcrowded vehicular environment.

5.2.3 Limited Categories of defects

Only four types of defects are included in the dataset on the basis of BRTA/ARI statistics of 620 images, which excludes over 60 WHO violations like tire and brake deficiencies. The effect of this omission is the under-reporting of fitness, creation of an imbalance in classes, and encouragement of over-fitting. In turn, the number of classes should be increased to 15 which would demand significantly greater amounts of data, especially when deployment is limited.

5.2.4 Dataset Scale and Diversity Issues

The dataset contains 620 images, 68% of which is used in the training division and 32% in validation and test divisions. This fairly small sample size is prone to overfitting, as well as causing misclassifications in the edges of objects. Even in the case of data-augmentation techniques. Models conditioned on such data, therefore, have poor generalizability outside of their direct geographic setting. One other expected source of variability is the manual annotations, which can create some subjectivity and labeling variability.

5.2.5 False Positive on Original Designs

A side mirror, such as the ones those are Fit, is mistakenly treated as missing by the system since it has no adequate training data representing a fit side mirror missing. Due to this problem even our mAP is also less.

5.2.6 Future Development Implication

The mentioned constraints, which are associated with the detection scope, data gaps, computational barriers, and evaluation restrictions, are typical of the initial stage of the project. They shed light on the barriers to scaling laboratory results to real world application and hence steer the creation of more resilient and embracing evolutionary approaches.

5.3 Future Work

5.3.1 Preparation of Future Extensions:

Through a Faster R-CNN bus defect detection architecture, the study is an early effort in AI-enhanced road safety in Bangladesh. Future studies must also seek to address the limitations identified, introduce new technologies, and align the approach with policy frameworks, and thus be drawn to the United Nations Sustainable Development Goal 3.6. The general goal is to achieve a 50% decrease in road carnage across the globe by 2030. However, our mail goal is to surpass, one by one, all the limitations our paper has. Specially concentrating more on our dataset and results.

5.3.2 Transition to Video Analysis:

In the current research, the substitution of stationary images with moving video records makes it possible to capture the appearance of defects in a progressive manner, especially the damage of the fissure. Detection on video. Video can be acquired using dashboard mounted cameras, closed circuit television systems, and aerial platforms that have low sampling rate integrated LSTM-augmented Faster R-CNNs or Video Swin Transformers to provide strong per-frame tracking. A pilot test on benchmark datasets, consisting of a total of more than 3,000 clips (BDDO100K and BRTA) it is shows that the motion seen artefacts are reduced, real time anomaly warning is enabled, and the total inspection times are halved.

5.3.3 Generalization through Expansion of Data:

The existing 620 images dataset would be susceptible to not give a proper accuracy, and it would not be able to reflect the entire range of real-world variability in vehicular imagery. We would recommend expanding the data set to some 5,000-10,000 images representing various seasons and locations, e.g., Dhaka and Sylhet, to overcome this limitation. It will be expanded with the help of crowd sourced contributions through applications, such as Pathao, the synthesis of weather-condition scenarios with the use of Generative Adversarial Networks (GANs), and the incorporation of a huge variety of car models. The cooperation with the Bangladesh Road Transport Authority (BRTA) will make it possible to include more than fifteen other labeling classes, along with such granular features as tire wear. It is expected that utilizing transfer learning based on the COCO benchmark and incorporating federated learning protocols will increase gains in mean average precision (mAP) by 10-15 per cent and make a significant improvement in the generalization ability of the model.

5.3.4 Vehicle Identification Integration:

When optical character recognition techniques such as EasyOCR and Tesseract, used on R-CNN region proposals, are utilized, about ninety percent accuracy in license plate character recognition is achieved; it can also be improved by video averaging. These plates are automatically recorded into the database of the Bangladesh Road Transport Authority (BRTA) by the system, which proves to be accurate, thus allowing enforcement of fines and recalls automatically and aiding in predictive maintenance by age and mileage information.

5.3.5 Increasing the Impact of Societies through Alerts:

The suggested framework will convert the model outputs into SMS, API, and GIS notification to enable police notifications, and deploy low-latency barrier detection on edge platforms like Jetson Nano with quantized models. Cooperation will be created with the Associate Research Institute of Bangladesh University of Engineering and Technology to incorporate the Grad-CAM based explainability. The AI proposed system will be empirically validated by deploying A/B trials along the routes of Dhaka to Chittagong, allowing to compare the suggested AI-based violation detection with the traditional approach involving manual detection.

5.3.6 Discovering Hybrid Methodologies:

The proposed study suggests the use of a hybrid architecture, as it combines Faster R-CNN and YOLOv8/RT-DETR and MobileNetV3 backbones to ensure edge inference times are less than fifty milliseconds and the overall accuracy is higher than eighty-five percent. We also use DETR attention to increase the detection of minor defects and use Mask R-CNN ensembles to conduct pixel-level structure, thus giving accurate instructions on the further repair efforts.

Overall Vision:

The above mentioned paths, including those that run between video and big data analytics, hybrid systems and alert systems, are the drivers of the creation of scalable safety solutions. The respectively introduced approaches have a potential of implementing a national instrument, thus illustrating life saving results and the revolution of mobility systems in Bangladesh and beyond.

References

[1] World Health Organization. *Road traffic injuries*. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. Fact sheet, Accessed: 2024-10-11. Dec. 2023.

[2] Dhaka Tribune. *25,000 lives lost in road accidents every year*. Update: 22 Oct 2023. Oct. 2023. URL: https://www.dhakatribune.com/bangladesh/328677/25-000-lives-lost-in-road-accidents-every-year (visited on 10/11/2024).

[3] The Daily Star. *55,000 without fitness clearance*. Frontpage section, Article ID: 1559608. Apr. 2018. URL: https://www.thedailystar.net/frontpage/55000-without-fitness-clearance-1559608 (visited on 10/11/2024).

[4] Feng Yang, Xingle Zhang, and Bo Liu. "Video object tracking based on YOLOv7 and DeepSORT". In: *arXiv preprint arXiv:2207.12202* (2022).

[5] Chenyu Gu et al. "Comparison and Application of Vehicle Target Detection Methods Based on YOLOv5s and YOLOv7". In: *2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE. 2024, pp. 745–749.

[6] Rabel Guharoy et al. "Evaluation of Vehicle Fitness Using Convolutional Neural Networks". In: *2024 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*. IEEE. 2024, pp. 1–5.

[7] Murat Bakirci et al. "Multi-class vehicle detection and classification with YOLO11 on uav-captured aerial imagery". In: *2024 IEEE 7th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD)*. IEEE. 2024, pp. 191–196.

[8] Vinh Dinh Nguyen et al. "A fast evolutionary algorithm for real-time vehicle detection". In: *IEEE Transactions on Vehicular Technology* 62.6 (2013), pp. 2453–2468.

[9] DGTL Geethmal and Mohamed Sapraz. "Designing an Automated Vehicle Fitness Evaluation System by Integrating Machine Learning to Achieve Unbiased Results in Warrant of Fitness Assessments". In: *Transformative Applied Research in Computing, Engineering, Science and Technology*. CRC Press, 2024, pp. 55–62.

[10] Eduardo Bayona et al. "In search of the best fitness function for optimum generation of trajectories for Automated Guided Vehicles". In: *Engineering Applications of Artificial Intelligence* 133 (2024), p. 108440.

[11] Sudhir Kumar Rajput et al. "Automatic vehicle identification and classification model using the YOLOv3 algorithm for a toll management system". In: *Sustainability* 14.15 (2022), p. 9163.

[12] Vibhanshu Singh Sindhu. “Vehicle identification from traffic video surveillance using YOLOv4”. In: *2021 5th international conference on intelligent computing and control systems (ICICCS)*. IEEE. 2021, pp. 1768–1775.

[13] Amir Mukhtar, Likun Xia, and Tong Boon Tang. “Vehicle detection techniques for collision avoidance systems: A review”. In: *IEEE transactions on intelligent transportation systems* 16.5 (2015), pp. 2318–2338.

[14] Md Mohiuddin Khan, Dewan Mamun Raza, and Md Aynul Hasan Nahid. “Real-time vehicle fitness detection using yolov5”. In: *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE. 2023, pp. 1–7.