VEHICLE DETECTION AND CLASSIFICATION USING A DEEP LEARNING APPROACH

By

SARAF BINTE SARWAR

ID: 182-15-994

This thesis is submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mohammad Monirul Islam
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

FEBRUARY 04, 2023

APPROVAL

This Thesis titled **"Vehicle Detection and Classification Using A Deep Learning Approach"**, submitted by **SARAF BINTE SARWAR**, ID No: **182-15-994** to the Department of Computer Science and Engineering, Daffodil International University has been approved as to its style and substance and acknowledged as adequate for the partial completion of the criteria for the degree of B.Sc. in Computer Science and Engineering. The presentation has been held on 04/02/2023.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. S. M. Aminul Haque
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dewan Mamun Raza
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

4/2/2023

Dr. Shamim H Ripon
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

This research was carried out under the direction of **Mohammad Monirul Islam, Assistant Professor** of CSE Daffodil International University. We also certify that neither this thesis nor any portion of this research has been submitted to any other institution for the award of a degree or diploma.

Supervised by:

Mohammad Monirul Islam
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Saraf Binte Sarwar

SARAF BINTE SARWAR
ID: 182-15-994
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, we express our heartfelt thanks and gratitude to Almighty God for His wonderful gift, which enabled us to successfully finish the final year project/internship.

First and foremost, we express our heartfelt thanks and gratitude to Almighty God for His wonderful gift, which enabled us to successfully finish the final year project/internship.

We are really grateful and desire to express our heartfelt gratitude to **Mohammad Monirul Islam, Assistant Professor**, Department of CSE, Daffodil International University, Dhaka. Our supervisor's deep knowledge and significant interest in the topic of "Deep Learning" are required to complete this assignment. His unending patience, intellectual direction, continuous encouragement, consistent and energetic supervision, constructive criticism, good counsel, reading numerous poor draughts, and editing them at all stages enabled this project to be completed.

We would like to express our heartiest gratitude to **Prof. Dr. Touhid Bhuiyan** and Head, Department of CSE, for his kind help to finish our project and to other faculty members and the staff of the CSE department of Daffodil International University. We would like to thank our entire course mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must express our gratitude for our parents' unwavering support and patience.

ABSTRACT

This research proposes a mechanism to detect and classification if a vehicle will be a car or not car that is the mean of vehicle or non-vehicle. The focus of this research vehicle detection and help for traffic violations. The ability to recognize automobiles in traffic scenes enables the analysis of driver behavior as well as the detection of traffic violations and accidents. Due to the variety of vehicle types and weather and light circumstances, detecting and classifying cars is a difficult operation. Feature extraction methods and Neural Networks are used in a number of solutions. Convolutional neural networks, on the other hand, have been shown to be possibly more effective. We describe a CNN that has been trained to categorize and recognize automobiles from diverse angles in this thesis. In addition, the Keras model is employed for data preprocessing. Our dataset has consist of a total of 3,026 images and txt datasets. For image classification, the Convolutional Neural Network (CNN) is used. On the other hand, to compare with CNN, we have used other algorithms like Inception V3 and AlexNet. I evaluated the Precision, Recall, and F1-score for performance. A vehicle detection system's principal work is to locate one or more automobiles in input photos. CNN beat Inception V3, AlexNet, and had adequate accuracy for vehicle detection and classification tasks, according to the results. This model acquired 97.03% accuracy in classifying vehicles through Convolutional Neural Network (CNN).

TABLE OF CONTENTS

CONTENTS	PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Since the inhabitants and transportation systems are growing at the same time, the demand for management is growing as well. The planet is rapidly becoming populous. As a result, the total number of machinery, and vehicles, grew at the same time. However, additional challenges like traffic, accidents, and many more must be addressed. It is challenging to oversee them utilizing the former ways, recent fads, and advances that have been found and delivered to manage every single achievement that humankind is endeavoring to reach. One of these issues is gridlock on roadways and in urban communities. To deal with this situation, a variety of measures such as traffic lights, signage, and other devices are used.

These choices appear to be insufficient or ineffective on their own. New technologies like object identification and tracking are created in order to leverage automated video surveillance to deliver information that can make sense to a decision-making process. This phenomenon has been used on a variety of topics. Object detection and tracking are two of the numerous parts of the upcoming Intelligent Transportation System (ITS). This technology detects vehicles, lanes, traffic signals, or when a vehicle detects something. The capacity to recognize and classify vehicles allows us to enhance traffic flows and roadways, avoid accidents and track traffic offenses and violations. Humans can easily recognize cars in photographs and distinguish between different types of vehicles. It is largely dependent on the types of data in machine learning and programs. Some obstacles, such as the weather or the amount of light available, can make the procedure easier or more difficult. At the same time, we have cars of all types and shapes. More than that, a new problem could be identifying moving objects in real-time photos where their size and shape differ. Vehicle detection and categorization can be accomplished using a variety of approaches and procedures.

Convolutional Neural Network (CNN), Inception V3, AlexNet, Support Vector Machine (SVM), Decision Tree, Recurrent Neural Network (RNN), and other algorithms are examples of these techniques. Since the industry is centered on this system or Computer visionary, the sector is continually evolving. In this thesis, we look at three algorithms: CNN, Inception V3, and AlexNet, to see how they may be used in practice and which one performs better.

1.2 The Problem Statement

Object detection has recently piqued the interest of the research community. Researchers are attempting to learn more about the subject in order to arrive at a degree of accuracy that is acceptable. Detecting objects is done via machine learning. There are numerous strategies for achieving this, however, in order to identify the best model among the suggested models, this thesis will investigate how to detect objects while comparing two suggested approaches and recommending the best one that produces the maximum accuracy and performance. Because it is a classification task, the task of recognizing and classifying objects in images is ideally suited to machine learning. The reason for this is because of the dataset's complicated properties. Many fields, in particular, rely on the system. The method is useful in a variety of sectors, including transportation and vehicle detection. A large dataset was used for training and analysis, which is detailed in full in Chapter 4.

1.3 Motivation

Since industrialization, the number of automobiles has steadily increased. These are some of the world’s largest new challenges is traffic congestion. People in cities spend the majority of their time stuck in traffic trying to get somewhere. It is critical for all countries throughout the world to have a digitalized traffic system that operates 24 hours a day, seven days a week and, makes activities simple and efficient.

As a result, a computerized traffic system cannot function without a reliable vehicle detection system. The system clearly has an impact on the economy, citizens' lives, industry, and so on. That stated, we must continue to contribute to the topic until the algorithms are capable of reliably detecting all types of objects or vehicles.

1.4 Objectives of this Research

The purpose of this research is to create a (CNN) that can recognize and classify vehicles in images with and without backgrounds. The following are the objectives in further detail:

- Create a classifier that can correctly classify images into automobiles or non-vehicles.
- Create a vehicle detection that must predict CNN vehicle extractions.
- When preprocessing data, use the Keras model.
- After that, model development and data separation for training and testing.
- We must then train our model with a training dataset after separating the data.
- After that, the model is put to the test using test data.
- After that, we'll utilize a confusion matrix, precession, recall, and F1-score to assess our model's correctness.
- We compared three algorithms to see which one was the most accurate.
- Examine whether CNN, Inception V3, and AlexNet help or hurt the developed solution's accuracy.

1.5 Aim of the Study

The goal of this thesis is to create three algorithms to recognize automobiles in image datasets: CNN, Inception V3, and AlexNet. The three techniques are compared using the same dataset and preprocessing procedures. Finally, suggest a method that has a high level of accuracy and performance. The thesis will implement three classifiers that can predict the image's class, whether it is Vehicles or Non-Vehicles. The object coordinates of the vehicles will also be predicted by the vehicle detector.

1.6 Contribution

Vehicle detection and classification for CNN is one of the contributions. In addition, look into the impact of the Keras model on the input data. The following procedures must be taken to accomplish this:

- Locating the datasets
- Using preprocessing method on the data that has been obtained. The experimental pre-processing with the Keras model is included in this stage.
- Creating a CNN architecture that is appropriate for the dataset and the thesis's purpose. To get a good model, you must experiment with different hyperparameters.
- Teaching the CNN how to classify and detect vehicles.
- Putting the final solution to the test.

1.7 Thesis Layout

There are five chapters in this thesis. The remaining sections of the thesis are organized as follows:

Chapter 1 Introduction: Provides the summary of this research-based thesis. Along with, what motivated me to do such a research-based thesis is explained well in this chapter too. The most important part of this chapter is the objective of the Study. Then motivation and then, what are the problems that I will face are also discussed in this chapter? The first chapter provides a general overview and background on the subject. This chapter discusses the thesis's purpose, contribution, significance, and research overview.

Chapter 2 Background and Literature Review: This chapter covers the discussion about what has already been done in this domain before. Last portion of this chapter discourse about the classifying methodologies and required tools. Chapter two previous types of research on the topic are reviewed in this chapter, together with their techniques and outcomes.

Chapter 3 Research Methodology: This chapter is about the theoretical discussion of this research. Besides, it has the flow charts or architecture of my proposed work.

Chapter 4 Results and Evaluation: This chapter is related to the outcome of this thesis. Have some theories about the models that I used to classify the news labels. The last portion has some pictures and figures for a better understanding of the results.

Chapter 5 Conclusion and Future Work: This is based on the conclusion topics of the thesis. And it closes with future studies of this thesis.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

2.1 Background

Detection and classification of objects can be utilized for a variety of reasons. Vehicle and pedestrian detection, traffic sign and lane detection, vehicle make detection, and object detection are all employed in the Intelligent Transportation Systems (ITS) sector. The capacity to identify or categorize traffic-related elements enables further enhancements of the road and traffic flow, the avoidance of fatal traffic collisions, and the reporting of traffic infractions and crimes like speeding or car theft. [1]. Given the rising number of individuals who drive passenger automobiles, this is particularly important. Furthermore, the idea of driverless automobiles has recently gained traction. Humans can easily recognize vehicles in photographs and differentiate between different car types. The data format has a significant impact on how challenging it is for a computer programme to identify and classify vehicles. The two most challenging elements to overcome are lighting and the environment, not to mention the overall caliber of either shot. Vehicles come in a range of sizes, hues, and even identical models in certain cases. Additionally, it is significantly more challenging to identify several moving objects in real-time. Nonetheless, there are a variety of approaches for detecting and classifying vehicles. A lot of work handles these issues by first extracting important characteristics and then making subsequent decisions using a classifier like CNN, Inception V3, or AlexNet.

2.2 Related Works

Vehicle detection and tracking have piqued experts' interest in recent decades. The subject drew a lot of interest. Various sensory modalities have been utilized to detect objects, particularly cars. Computer vision is one of these modalities. The appeal stems from tremendous advances in picture processing. The earliest signs and models of image processing are from the 1960s and 1970s, after which a variety of approaches and techniques were devised and proposed. This chapter briefly reviews recent relevant studies in the field of vehicle detection and classification.

In 2018, Viktoria Plemakova, by Vehicle Detection Based on CNN - He concentrated on the categorization of Neural Networks. Feature extraction methods and SVM classifiers are used in several solutions. Convolutional neural networks, on the other hand, have been shown to be possibly more effective. He describes a convolutional neural network that has been trained to categorize and recognize automobiles from diverse angles in this thesis. Furthermore, during data preprocessing, the Fast Fourier Transform is performed. The impact of such preprocessing on the generated vehicle classifier and detector is investigated. He received the highest accuracy score of 96 [1]. This paper has a few lacking like comparison each other algorithms, lacking efficiency, etc.

In 2019 Huansheng Song, Haoxiang Liang, Huaiyu Li, Zhe Dai, and Xu Yun, By Vision-based vehicle detection and counting system using deep learning in highway scenes - In this paper, Clever vehicle recognition and including are turning out to be progressively significant in the field of thruway the executives. Nonetheless, because of the various sizes of vehicles, their location stays a test that straightforwardly influences the precision of vehicle counts. To resolve this issue, this paper proposes a dream-based vehicle identification and counting framework. Another top quality interstate vehicle dataset with an aggregate of 57,290 clarified occasions in 11,129 pictures is distributed in this review. Contrasted and the current public datasets, the proposed dataset contains comments on little items in the picture, which gives the total information establishment to vehicle discovery in view of profound learning. In the proposed vehicle discovery and counting framework, the roadway street surface in the picture is first extricated and isolated into a distant region and a proximal region by a recently proposed division technique; the strategy is urgent for further developing vehicle recognition. Then, at that point, the over two regions are set into the YOLOv3 organization to distinguish the sort and area of the vehicle [2]. At last, the vehicle directions are acquired by the Sphere calculation, which can be utilized to pass judgment on the driving bearing of the vehicle and get the number of various vehicles. A few interstate observation recordings in view of various scenes are utilized to confirm the proposed techniques. The test results confirm that utilizing the proposed division technique can give higher identification exactness, particularly for the location of little vehicle objects. Besides, the original system depicted in this article performs outstandingly well in passing judgment on driving bearing and counting vehicles. This paper has general down-to-earth importance for the administration and control of parkway scenes. This paper has a few lacking like comparison each other algorithms, lacking efficiency, real-time detection problem, small size datasets, etc. The accuracy was 92.3% for detection and 93.2% for counting.

Liu et al. [13] by The Generative Antagonistic Nets (GANs) were developed to organise automobiles using records of traffic observation. Three levels of in-vehicle characterisation are included in the created technique. In order to generate unfavourable instances for the interested classes, GAN was initially trained on a collected traffic dataset. The next step was to create an outfit-based Convolutional Neural Organization (CNN) using the uneven dataset, and then test selection was carried out to exclude the antagonistic cases with lesser quality. Finally, the group model on the larger dataset was improved using the selected antagonistic cases. Broad research revealed that the developed GAN technique, as measured by the Cohen kappa score, mean review, accuracy, and mean accuracy, successfully executed in-vehicle grouping on MIO-TCD.

Fu et al. [14] fostered another vehicle arrangement procedure based on a progressive multi-SVM (multi-Backing Vector Machine) classifier. At first, the frontal area objects were removed from the reconnaissance recordings, and afterward, the progressive multi-SVM procedure was produced for vehicle characterization. Moreover, a democratic-based amendment approach was utilized to follow the ordered vehicles for the presentation assessment. In this writing review, a pragmatic framework was created in light of the progressive multi-SVM method for hearty vehicle characterization in a weighty rush hour gridlock scene. Henceforth, the created strategy is incapable in essentially jam-packed rush hour gridlock scenes, because of the various perspectives, shadows, and weighty impediments.

Further, Şentaş et al. [15] utilized the little Just go for it with the SVM classifier for vehicle location and arrangement. In the test portion, the exhibition of the created model was approved on the Piece Vehicle Dataset considering accuracy and review. The consequence of the analysis affirms that the created model fundamentally arranges the vehicle type's continuously real-time traffic recordings. Notwithstanding, SVM was a twofold classifier, which upholds just paired grouping that was a significant limit in this review.

Wang et al. [16] fostered a vehicle type order framework in light of the quicker R-CNN strategy. The exhibition of the created procedure was assessed on a continuous dataset that contains genuine scene pictures caught at the intersection. As a future upgrade, an original strategy is expected to work on the capacity to identify a vehicle that is impeded because of various light circumstances, points, and sizes of the pictures.

Zhuo et al. [17] fostered a CNN model for vehicle characterization which incorporates two significant advances adjusting and retraining. In the retraining step, AlexNet was applied on the ImageNet Huge Scope Visual Acknowledgment Challenge 2012 (ILSVRC2012) dataset to get the underlying model with association weight. In the tweaking step, the got introductory model was adjusted on the vehicle dataset to accomplish the last grouping. In this writing study, the gathered interstate observation recordings incorporate six vehicle classes like van, minibus, truck, transport, vehicle, and cruiser. In the test stage, the presentation examination was done on the vehicle dataset through precision. In any case, the created CNN model is computationally costly and has a significant issue of "overfitting."

Murugan and Vijaykumar [18] fostered another system for vehicle type order that incorporates six primary stages, for example, information preprocessing, identification of the vehicles, vehicle following, underlying coordinating, extraction of the highlights, and vehicle arrangement. Subsequent to gathering the traffic observation recordings, information preprocessing was achieved by utilizing commotion expulsion and shading transformation. Further, the Otsu thresholding calculation and foundation deduction were utilized to identify the vehicles. Then, at that point, vehicle following was achieved utilizing the Kalman channel to follow the moving vehicles. Furthermore, the log Gabor channel and the Harrish corner locator were utilized to remove the component vectors, and afterward, the got highlights were taken care of to the Fake Neural Fluffy Induction Framework (ANFIS) for the characterization of the vehicles. Broad trials showed that the created system accomplished huge execution in-vehicle characterization considering blunder rate and precision. The created structure expands the dimensionality issue that records the model’s intricacy.

Dong et al. [19] carried out another semi-directed CNN design for vehicle type order. In the created design, a meager Laplacian channel was applied to separate the rich and discriminative data of the vehicles. In the result layer, a softmax classifier was prepared by performing multiple tasks learning for vehicle type arrangement. In this writing study, the highlights learned by the semi-managed CNN engineering were discriminative to work essentially in complex scenes. Broad tests were assessed on the Piece Vehicle Dataset and a public dataset to investigate the productivity of the created design considering characterization exactness. The semi-directed CNN design incorporates a few layers, so the preparation cycle consumes additional time.

Before the carried out another semi-regulated model for vehicle type characterization was based on Head Part Investigation Convolutional Organization (PCN). In the created model, convolutional channels were used to separate the various leveled and discriminative elements. The reproduction result showed that the created model got better execution progressively applications, because of its strength against commotion defilements, brightening conditions, pivot, and interpretation. The created PCN model contains a more prominent number of preparing boundaries that lead to an overfitting issue [21].

Awang et al. [22] fostered the Scanty Sifted CNN with Layer Skipping (SF-CNNLS) approach for vehicle type arrangement. In this writing study, three channels of the SF-CNNLS approach were applied to extricate discriminant and rich vehicle highlights. Moreover, the worldwide and nearby elements of the vehicles were removed from the three channels of a picture in light of their shading, splendor, and shape. In the Exploratory Outcomes and Conversation, the exhibition of the created SF-CNNLS approach was approved on a benchmark dataset. At last, the softmax relapse classifier was utilized to arrange the vehicle types like a truck, minivan, transport, traveler, taxi, vehicle, and SUV. The created softmax relapse classifier incorporates more elevated level layers; notwithstanding, by installing lower-goal vehicle pictures, there might be a deficiency of vehicle type data.

Nasaruddin et al. [23] fostered a consideration-based methodology and a profound CNN strategy for lightweight moving vehicle characterization. In this writing, the created model exhibition was approved on a continuous dataset through explicitness, accuracy, and - score.

In any case, the created model presentation was restricted in such conditions as a pattern, camera jitter classes, and terrible climate. The techniques are attempted, datasets, benefits of involving the created strategies in-vehicle type grouping, and disservices of the strategies are obviously given for every writing paper. To resolve the above-expressed issues, another gathering profound learning procedure is proposed in this exploration paper to further develop vehicle-type characterization.

A multi-label classification system called CNN-RNN [14] combines a recurrent neural network with a convolutional neural network. The goal of multi-label classification is to identify all of the objects in a single image's labels. The authors claim that the accuracy was increased by including RNN in the initial network. Long short-term memory units make up this network's recurrent units. The network's VGGNet-based CNN component is used to extract semantic data from pictures. The relationship between labels and pictures is the focus of the RNN section.

As a result, numerous researchers have preferred CNNs to the previously described feature extractors in recent years. A number of CNN architectures are used in the associated study, demonstrating that CNN is a versatile tool in the field of computer vision.

2.3 Challenges

The biggest challenges for this thesis are data set collection and selection for an appropriate algorithm besides the CNN algorithm for comparison. It's very difficult to get the collection of various vehicle pictures from the Google platform. Several transportation authorities are increasingly concentrating more on vehicle categorization data collection to meet the expanding information demands for their IT system, traffic planning, and transportation analysis initiatives. Applications and concerns of vehicle classification detection are explored, as well as the growing demands for data flow, sensing technologies for vehicle classification, and challenges to enhancing the quality of classification data. Despite the necessity of gathering vehicle classification data and the capabilities of various sensors to recognize vehicle types.

2.4 Artificial Intelligence (AI)

The process of transmitting data, knowledge, and human cognition to robots is known as artificial intelligence or AI. The creation of autonomous systems capable of thinking and acting like people is the primary goal of artificial intelligence. These machines can do tasks and emulate human behaviour by learning how to overcome obstacles. Most AI systems mimic human intellect to address challenging issues. Artificial intelligence (AI), machine learning (ML), and deep learning (DL) have become the most talked-about technology in today's business sector as companies use these advancements to create smarter goods and apps. Despite the fact that business, talks throughout the world frequently employ these terms.

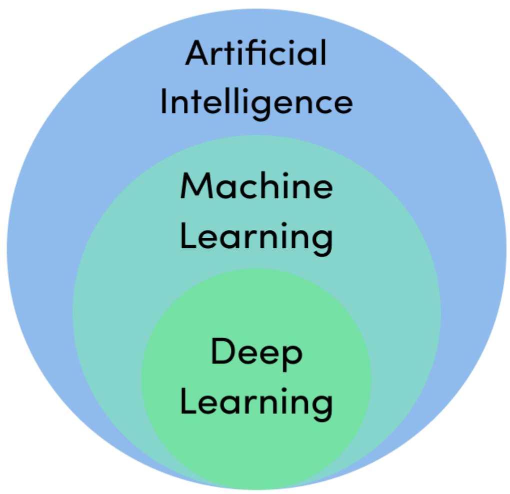

Figure 2.1 Relation of AI, Machine Learning, and Deep Learning [7]

Building intelligent machines with independent thought is known as artificial intelligence (AI). A subfield of artificial intelligence called machine learning supports the creation of AI-powered software. Deep learning is a branch of ML that uses sophisticated techniques and many data to train models.

2.4.1 Machine Learning Techniques

Artificial intelligence, which is the study of computer systems that can learn and evolve on their own via use and data, includes (ML), which is a subset of AI. Machine learning data analysis automates the creation of analytical models. It is an area of artificial intelligence that is based on the notion that machines can learn from data, see patterns, and make judgements with little to no assistance from humans.

Types of Machine Learning (ML)

Supervised Learning: Here, training samples are labeled, and when applied in the correct scenario, it is incredibly strong.

Unsupervised Learning: In this method, training data are left unlabeled, and the algorithm creates clusters or natural groupings of input samples into a small number of classes.

Reinforcement Learning: also known as behavioral machine learning is a type of reinforcement learning. It's comparable to supervised learning, however, it doesn't need sample data to train. It has a self-improving algorithm that learns from new scenarios using a trial-and-error approach.

2.4.2 Deep Learning Techniques

A more recent improvement on artificial neural networks is called deep learning, which makes use of readily available low-cost computing. Artificial neural networks, which are algorithms modelled after the structure and operation of the brain, are the subject of the machine-learning field known as deep learning.

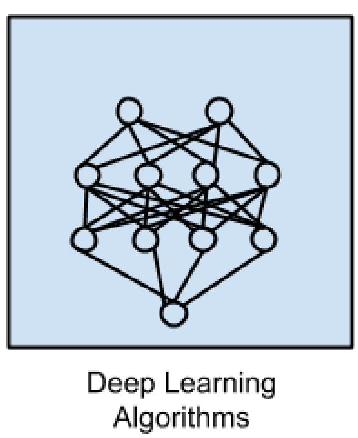

Figure 2.2 Deep Learning Algorithm [8].

You could be perplexed if you are inexperienced with deep learning or have already dealt with neural networks. Like many of my coworkers and friends who researched and utilised neural networks in the 1990s and early 2000s, I was first perplexed. Deep learning is explained by professionals and industry leaders, whose varying and nuanced points of view provide a wealth of knowledge on the topic.

As mentioned above, many of the algorithms work with very large datasets of tagged analogue data, including picture, text, audio, and video. They are concentrating on creating far bigger and more intricate neural networks.

The most popular deep learning algorithms are:

- CNN
- RNNs
- LSTMs, etc

Robots that can do complicated tasks are supposed to be clever enough to use deep learning algorithms and methods, which are based on the structure and operation of the human brain. Deep learning approaches build models with several hidden layers of neural networks to produce precise predictions.

2.5 Classifying Methodologies

This section presents an introduction to the algorithms I used in my thesis work. I used Convolutional Neural Network (CNN), Inception V3, and AlexNet as algorithms for the experiment in my system.

2.5.1 Convolutional Neural Network (CNN)

A CNN (ConvNet/CNN) is a Deep Learning system that can analyze an image as input, assign pertinent weights and biases to various traits and objects within the image, and then distinguish between them. Comparing a ConvNet to other classification methods, the amount of pre-processing needed is significantly less. ConvNets can learn these filters and properties with enough training, unlike simple systems that need hand engineering of filters.

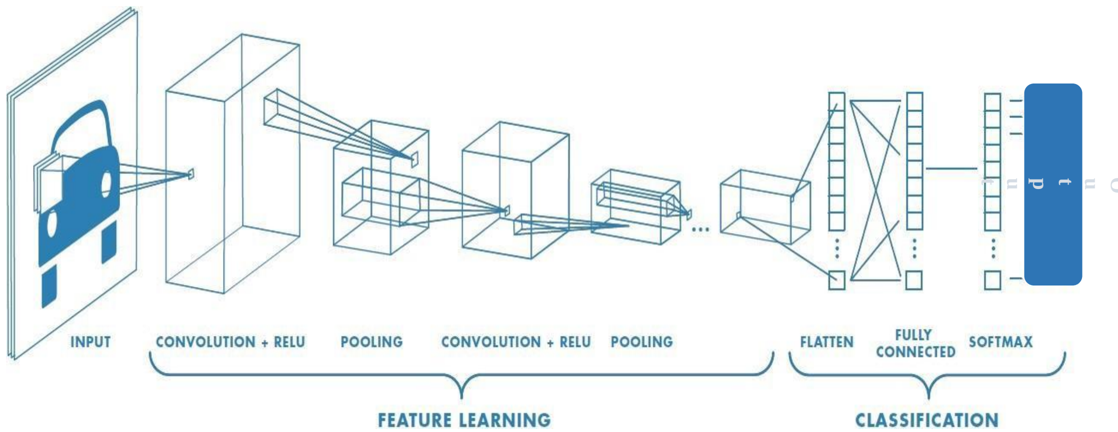

Figure 2.3 Understand the Architecture of CNN.

The design of a ConvNet is inspired by the architecture of the visual brain and mirrors the connections between neurons in the human brain. Individual neurons are only able to react to stimuli in a small portion of the visual field known as the Adaptable Field.

2.5.2 Inception V3

Inception v3 is a convolutional neural network that was developed as a GoogleNet plugin to aid with object and picture identification. The third generation of the Google Inception Convolutional Neural Network, which made its debut at the ImageNet Identify the effect, has now been released.

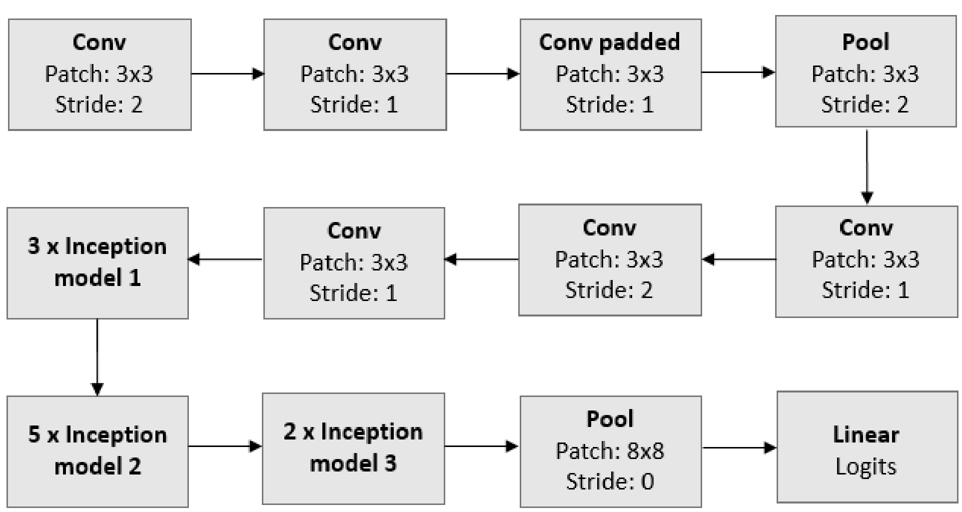

Figure 2.4 Basic Architecture of Inception V3 [11].

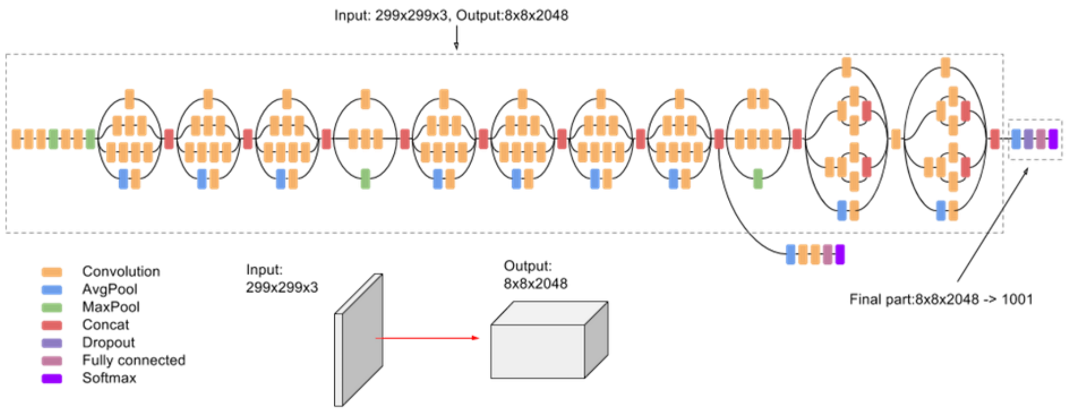

Figure 2.5 InceptionV3 model of Schematic Diagram [13].

A deep learning model for categorising images based on convolutional neural networks is called the Inception V3. The Inception V3 was created by updating the fundamental model Inception V1, which was first introduced as GoogLeNet in 2014. It was created by a Google team, as the name would imply.

Label Smoothing, Factorized 7 x 7 convolutions, and the addition of an auxiliary classifier to transfer label information farther down the network are advancements included in the Inception-v3 convolutional neural network design.

2.5.3 AlexNet

Alex Krizhevsky and associates developed deep neural network AlexNet in 2012. It won first place in the ImageNet LSVRC-2010 competition for identifying photographs [1]. Additionally, it worked with a variety of GPUs. AlexNet was really created as a deep convolutional neural network to handle huge colourful pictures (227x227x3). In all, it had more than 62 million trainable parameters. The details 11 layer of AlexNet is seen in figure 2.6.

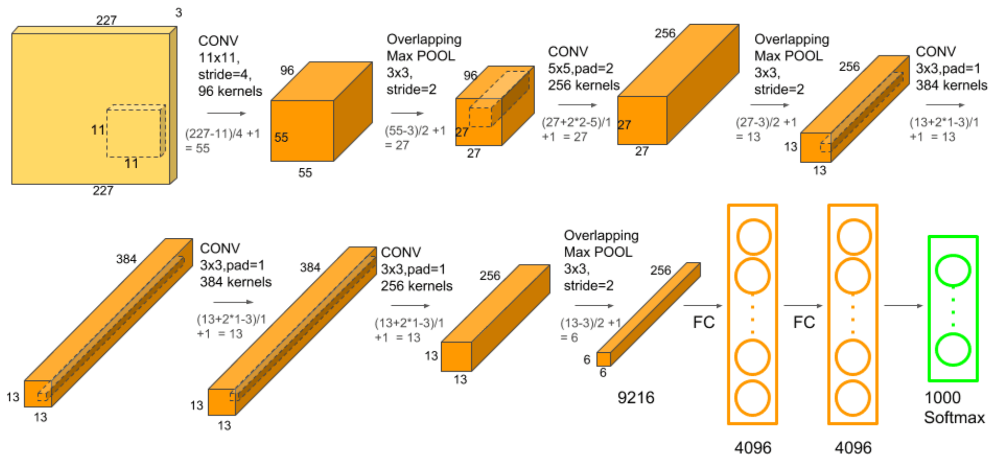

Figure 2.6 The Architecture of AlexNet CNN Diagram [14].

Convolutional neural networks like AlexNet have significantly influenced machine learning, especially when it comes to using deep learning for machine vision. It is well known for having decisively won the 2012 ImageNet LSVRC-2012 competition (15.3 percent mistake rates versus 26.2 percent error rates in second place). The network's design was relatively similar to Yann LeCun et alLeNet, .'s although it was more intricate, had more filters per layer, and was built with stacked convolutional layers. Convolutions, maximum pooling, dropout, and data augmentation were all included, ReLU activations, and SGD with momentum. ReLU activations were introduced after each convolutional and fully-connected layer.

CHAPTER 03

RESEARCH METHODOLOGY

3.1 Introduction

The chapter of this study explains how the workflow is maintained. It will provide a clear picture of the research. This part covers data collecting, data preprocessing, system workflow, feature extraction, tools and technologies, and other aspects of my thesis. That is to say, all approaches and technologies for detecting and tracking cars will be explored. The tools used in the research are outlined in detail on how they are employed in the context right from the start. Data preparation, cleansing, and algorithm development then start the implementation phase; a quick synopsis of these issues is offered at the last chapter.

3.2 Required Tools Used

This thesis, like all other research and studies, used some software and tools to create models and tests. Such research requires a variety of tools, such as the dataset containing a large number of vehicle and non-vehicle photos used for training, the many Python packages that are necessary or helpful in order to generate ML models, or the author's use of Python as the model development programming language. Each instrument that will be used in the thesis is introduced in this chapter.

3.2.1 Python

Python is a widely used programming language created by Guido van Rossum. It is used for many different purposes, including arithmetic, computer GUIs, the web, and a vast array of complex logical applications. Making programming easy was the main purpose when creating Python, so that anybody on Earth could write code. It is known for being straightforward because of this. Python is skilled at disentangling. Despite the fact that Python was designed for beginners, it currently outperforms all other programming languages in a variety of areas, which is astounding and in some ways excellent. Python has all the same capabilities as other programming languages. Python has gained popularity in artificial intelligence (AI) applications over the past several years, from the web to computations to desktop apps, and so on, thanks to its wealth of useful modules that streamline and accelerate work. Experts entering this sector come from many backgrounds. The most straightforward and helpful language to start with as they don't have a programming background is python. Despite the aforementioned qualities, the analyst chose Python as the programming language for the model turn of events because of their knowledge and sense of responsibility. The following are the libraries that were utilized in this postulation.

3.2.1.1 NumPy

A Python package called NumPy is used while working with exhibitions. Additionally, it has the ability to operate in the lattice, Fourier, and polynomial math fields. 2005 saw the creation of NumPy by Travis Oliphant. Since it is an open-source project, you can use it without restriction. For Mathematical Python, use NumPy. An open-source library called Numpy does registration with the aid of intricate frameworks and clusters. It has a variety of capabilities that make working with this sort of information straightforward. We truly want to use displays in information investigation to make it quick and effective. As a result, this library enables information researchers to process large amounts of information more quickly. Typically, the model's capacity demands manifest as boundaries to work rapidly and decrease the preparation forecast time during discovery and estimation.

3.2.1.2 Matplotlib

A Python tool called Matplotlib makes it possible to make interactive, animated, and static visualisations. Matplotlib is one of the most well-known Python data visualisation libraries. A cross-platform library uses array data to produce 2D charts. It works in Jupyter notebooks, Python shells, and Python shells. Plotting is becoming more popular in all industries as a way to see and analyze data. As a result, Matplotlib is used as a plotting library to construct various graphs and figures for various purposes. Matplotlib has the advantage of producing nice plots and graphs with only a few lines of code. To extract color characteristics and construct histograms, matplotlib is utilized.

3.2.1.3 Seaborn

Seaborn is a library in Python transcendently utilized for making measurable designs. A matplotlib-based information perception package called Seaborn is tightly integrated with Python's Pandas data structures. The main component of Seaborn is representation, which aids in the analysis and understanding of data. Python's Seaborn package allows users to create quantifiable graphics. It builds upon matplotlib and integrates well with Panda's data structures. Seaborn aids with your research and understanding of your data. In order to produce effective plots, plotting capabilities must operate on information edges and clusters including whole datasets and therein play out the crucial semantic planning and quantifiable total. Instead of focusing on the finer points of how to build your plots, its dataset-situated, definite programming interface enables you to focus on what the various plot components actually signify.

3.2.1.4 Scikit-Learn

Scikit-learn is a free Python ML library. It's utilized in a variety of algorithms, including CNN, SVM, KNN, and Decision Tree, among others. NumPy and SciPy, two Python numerical and scientific libraries, are supported. The most practical and reliable machine-learning package for Python is called Scikit-learn (SKlearn). For a number of effective machine learning and statistical modelling techniques, including classification, regression, clustering, and image segmentation, it offers a consistent Python interface.

3.2.1.5 Sklearn.metrics

The sklearn.metrics module offers several loss, score, and utility techniques to measure categorization success. Certain metrics may need binary decision values, confidence values, or positive class probability estimations.

3.2.1.6 Keras Model

An effective and user-friendly deep learning model construction and evaluation framework is available for Python called Keras. With just a few lines of code, you can test and train neural network models using Theano and TensorFlow, two effective numerical computation frameworks.

3.2.3 Computer

The following are the characteristics of the PC that is being used to train and test the models:

- Model: DCL
- RAM: 4 GB
- Processor: Core i3, GEN-8
- Graphic: Intel HD 4 GB
- Windows: 11

3.3 Work Overview

First, I gathered a dataset because there was none on the internet concerning "Vehicle Detection and Classification Using CNN" that may aid me in my research. I gather information from Google and the UCI Platform. Then I used Python to put my ideas into action. I begin by importing my data. Then I used the Keras packages to preprocess my data. After that, I labeled and resized my dataset using a classifier-based strategy. The data was then divided into 80-20 divisions. This means that 80% of the data will be useful for the training set and 20% will be useful for the testing set. I then used all of the machine learning methods.

3.4 System Architecture

To build a model, I need to describe the nature of the work. Here to implement the model, I followed these steps of machine learning. They were:

- Collection Data
- Pre-processing Data
- Model Train by Training Dataset
- Applying Algorithm
- Testing
- Result

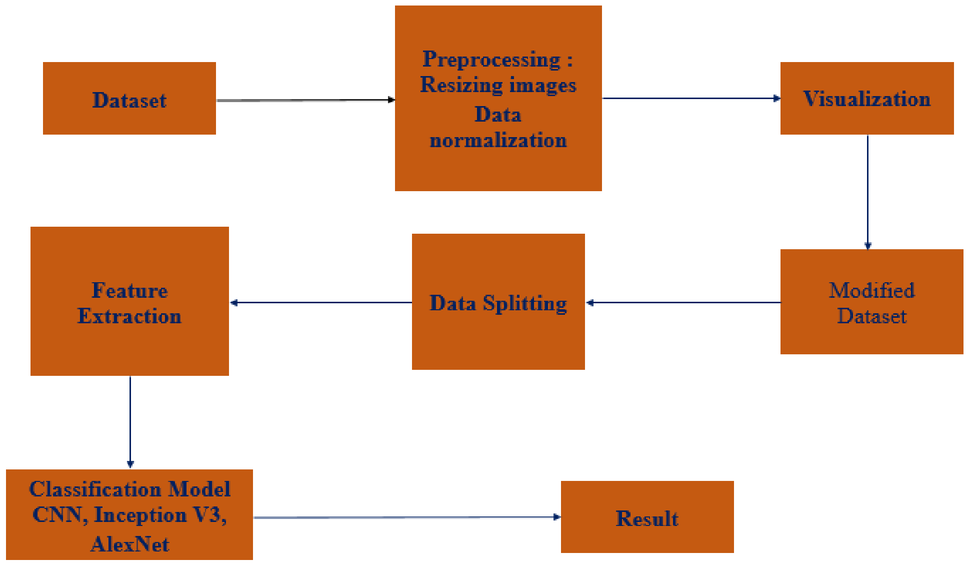

Figure 3.1 Basic System Architecture

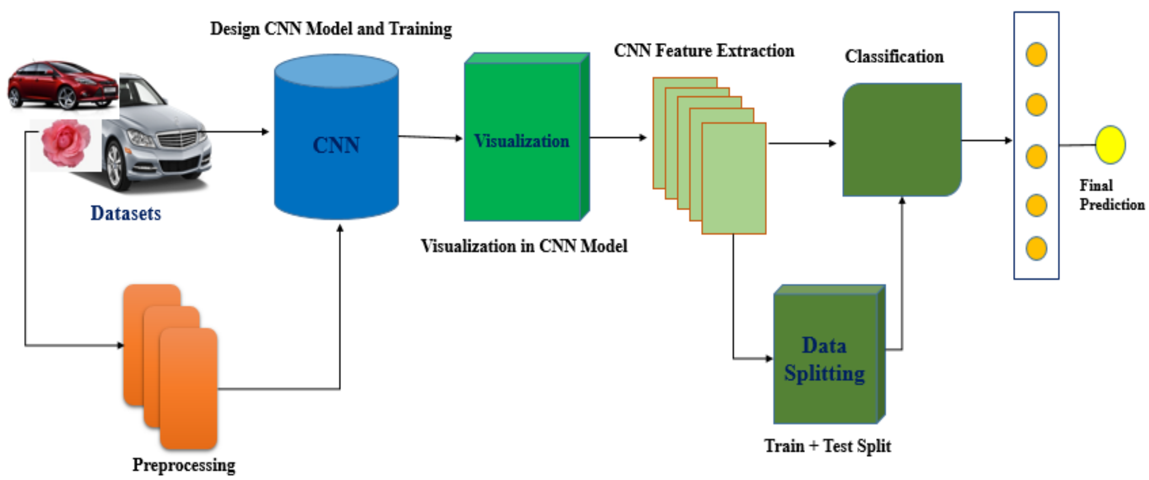

Figure 3.2 Developed System Architecture of CNN

3.4.1 Datasets

I could not discover any unique data on this topic to use for this project. As a result, I get all data from the Google platform and the UCI platform. Students can learn how to train classifiers by visiting the UCI website. As illustrated in figure 3.2, data from cars and non-vehicles was used for training. The data was obtained from the Google website. These examples were created using a combination of the GTI car picture collection and other sources.

Figure 3.3 Vehicle and Non-Vehicle images

These datasets were collected from various sources from the google platform for vehicle images and non-vehicle images from the UCI platform. This dataset consists of 3,026. There are 1513 images dataset and 1513 txt data sets.

3.4.2 Data Pre-processing

This section describes the final solution's data preparation steps. This involves creating more training data, normalising the data, scaling the data, applying three algorithms to photos, and normalising the data.

3.4.2.1 Resizing Images

Images of varying sizes are included in the dataset utilized in this thesis. This is because the final dataset was created by mixing multiple existing photographs. Because CNN’s typically need equal-sized input photos, the vehicle and background images have all been reduced to 50 by 50 pixels. Despite the fact that it is not required, most well-known CNN designs appear to prefer square-shaped input images. After the input photos are downsized, the folder must be scaled to obtain correct coordinates.

3.4.2.2 Data Normalization

The standardization procedure is used to normalize all of the photos. It is accomplished by subtracting the mean from each characteristic before diving by standard deviation:

$$z = \frac{x - \mu}{\sigma}$$

Where µ is the standard deviation and is the mean. The training set's mean and standard deviation are determined, and the same values are utilized to normalize the test set [37]. After normalising, the features are zero-centered and have a standard deviation of one.

3.4.2.3 Data Augmentation

Data augmentation is the process of adding preparatory information to a dataset that is already generally available. The amount of data in a dataset greatly influences how accurate a model is, and often, more data results in better models. The problem with little datasets is that models overfit more quickly [30]. This problem may be solved via regularisation, for instance by using dropout or L1/L2 regularisation in a CNN. An additional viable solution to this problem is to expand information [31, 32].

The changes that are utilized incorporate zooming in and out, moving both in an upward direction and evenly, and shearing. Flipping vehicles and streets in an upward direction would not check out and even flip is superfluous since, for instance, the vehicles are as of now portrayed from various points.

3.4.2.4 Train and Test Split

The StandardScaler() method assumes that data within each feature is regularly distributed and scales it so that it is now centered around 0 with a standard deviation of 1. The function transforms the 'x' values to produce the scaled X output. There are a few libraries that can assist you to partition the dataset. One of these is the 'sklearn' function 'train test split,' which helps split the dataset into train and test data for the classifier.

3.5 Developed CNN Algorithm

The conventional CNN model of an artificial neural network is a sophisticated and high-potential variation. It is made to manage escalating complexity, preprocessing, and data assembling. It is based on how neurons are organised in the visual cortex of an animal's brain. One of the best models for both image and non-image data specialisation is CNN.

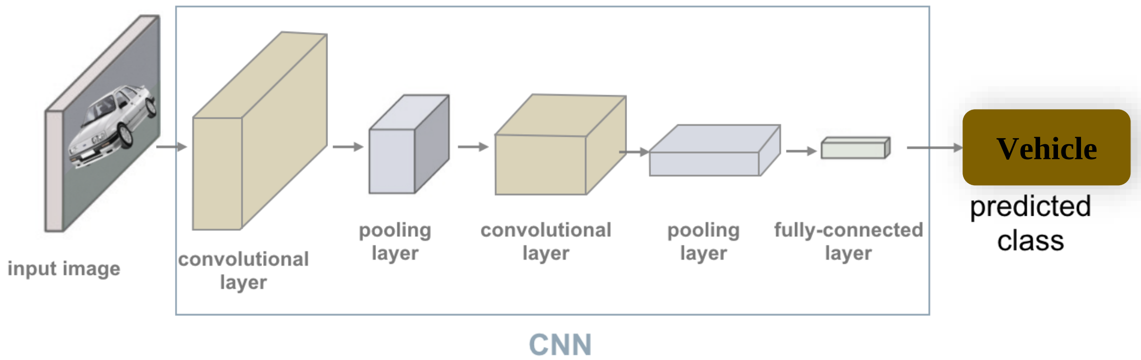

Figure 3.4 CNN Architecture Layer Based on Datasets

3.5.1 Activation Function

In deep learning models, activation functions introduce nonlinearity. An activation function is utilised in CNN models after each convolutional layer. Most modern CNN designs employ rectified linear units (ReLUs):

$$f(x) = max(0,x)$$

The sigmoid and hyperbolic tangents are two other possible activation functions, as shown in Equations (3) and (4), respectively.

$$f(x) = \sigma(x) = \frac{1}{1+e^{-x}} \qquad (3)$$

$$f(x) = tanh(x) = \frac{2}{1+e^{-2x}} - 1 \qquad (4)$$

3.5.2 Convolutional Layer

Feature extraction and detection are done at the convolutional layer. Each convolutional layer contains a unique set of kernels that are combined with the layer's input to create a number of feature maps. Let xli be the input to the ith feature map of layer L, with f the activation function, and the kernel, then

$$x_i^l = f(\sum_j w_{ij}^l * x_j^{l-1})$$

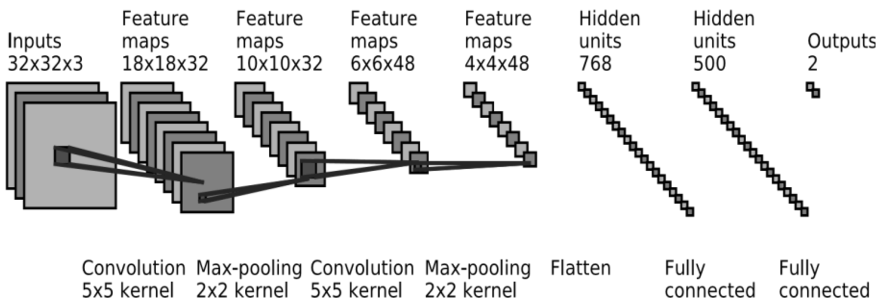

Figure 3.5 Two-Layer Architecture of CNN

This is the convolutional neural network's basic layer. Its settings are made up of a series of filters. These filters are modest, yet they cover the entire volume range of the input. The extraction of high-level features is the primary function of the convolutional layer. The first (as seen in the figure above) is in charge of extracting low-level information such as color and edges. Following convolutional layers, the high-level characteristics are removed, resulting in a comprehensive acceptance of the image.

3.5.3 Pooling layer

If a pooling layer is employed, it is normally applied after the convolutional layer. It does not matter where each recognised feature is placed directly, according to LeCun et al. [8]. The presence of a certain feature is significant. Therefore, superfluous data from feature maps is removed using pooling layers. A pooling function, such as max pooling, average pooling, or L2-norm pooling, is used to achieve this [33]. Even though the original LeNet-5 [8] employed average pooling, max pooling recently overtook it as the most used pooling method.

Feature map areas of size n x n are subjected to the pooling procedure. For example, from each region, a maximum value is picked to be included in the resulting feature map. Depending on the stride size, the zones may or may not overlap. Figure 3.6 depicts the concept of maximum pooling. Similar to the convolution, if the stride size S is higher than 1 then $W_o = \frac{W-W_p}{S} + 1$ and $H_o = \frac{H-H_p}{S} + 1$.

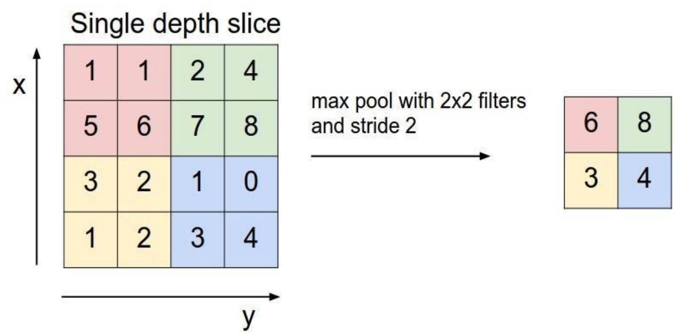

Figure 3.6 Example of 2×2 max pooling with a stride of 2 [47].

Although pooling layers are incorporated in the majority of CNN designs, an all-convolutional architecture has been shown [34]. The authors discovered that skipping max-pooling layers has no effect on accuracy when the stride of convolutional layers is increased. By lowering the number of parameters and the size of the feature maps, pooling layers, on the other hand, can aid in decreasing overfitting. The developers of AlexNet, for instance, demonstrated that overlapping pooling aids in reducing overfitting.

3.5.4 Fully-connected layer

The final layers used by CNN are completely linked layers. The last layer outputs the estimates. The last layer of multi-class networks use the softmax classifier, whereas binary classification employs the sigmoid function. Similar to conventional neural networks, fully-connected layers' computational units are interconnected to every unit in the layer below. Contrary to convolutional and pooling layers, fully-connected layers are one-dimensional.

3.6 Proposed Network Architecture

The Keras [35] deep learning framework with a Python and Tensorflow backend is used to create the CNN architecture. With modest output layer modifications, the same network is then utilised for vehicle classification and detection. Six convolutional layers and five maximum pooling layers are included in the suggested CNN. The last layer, which comes after the last convolutional layer, is the sole fully connected layer. The input layer accepts batches of images with a size of 128x128x3 pixels. In both classification and detection networks,

While sigmoid activation is used in the output layers, ReLU is used between the hidden layers. The first convolutional layer contains 32 filters with a 33-by-32-pixel dimension. It is followed by a 22-size pooling layer that uses the max-pooling technique. There are 64 filters overall in the second convolutional layer, with each filter being 33% in size. A max-pooling layer with a 22 kernel size follows it. This resembles the first convolutional layer in several ways. 128 filters with the same kernel size as the first layer's kernel make up the third convolutional layer. Once more, the maximum pooling of 22 is utilised. The third and fourth convolutional layers are identical, however the fourth does not employ maximum pooling. The same max-pooling layer as the preceding convolutional layers comes after the final two convolutional layers, which have 256 sizes 33 filters. The parameters of the specified layers are also included in Table 1.

	Kernel size	Number of kernels	Stride size	Output size
Convolution 1	3x3	32	1	50 x 50 x 32
Max pool 1	2x2	-	2	25 x 25 x 32
Convolution 2	3x3	64	1	25 x 25 x 32
Max pool 2	2x2	-	2	13 x 13 x 32
Convolution 3	3x3	128	1	28 x 28 x 128
Max pool 3	2x2	-	2	14 x 14 x 128
Convolution 4	3x3	128	1	12 x 12 x 128
Convolution 5	3x3	256	1	10 x 10 x 256
Max pool 4	2x2	-	2	5 x 5 x 256
Convolution 6	3x3	256	1	3 x 3 x 256
Max pool 5	2x2	-	2	1 x 1 x 256

Table 3.1. Parameters of convolutional and pooling layers.

The key approaches and procedures of implementing vehicle recognition and classification were discussed in this chapter's summary. Obtaining a suitable dataset is the first step in the workflow. The photos are then all preprocessed. Resized and normalized images are used. For the classification problem, a data augmentation strategy is employed to gain more training data. In addition, when the previous preprocessing procedures have been completed, keras is applied to the input photos. Finally, the preprocessed data is loaded into the CNN that has been proposed. Depending on the task, the output is either a label or four test image data.

CHAPTER 04

RESULT AND EVALUATION

4.1 Overview

In this section, I offer the accuracy, classification report, and confusion matrix of this system in several types of models that I utilized for my study, as well as a commentary of the results. In addition, we will test our model using performance metrics in this chapter. Using multiple classification methods such as CNN, Inception V3, and AlexNet, as well as performance measures, we can determine which one is best for our model. The confusion measures also indicate precision, recall, and f1 score.

4.2 Performance Evaluation

Performance evaluation is a quantified expression of the values that are used to fulfill the organization's objectives. To assess the performance of our suggested model, I used a variety of performance metrics. The confusion matrix, accuracy score, precision score, recall score, and f1 score were all completed.

Confusion Matrix

The Confusion Matrix is a fantastic resource for analysing the behaviour and comprehending the efficacy of a binary or categorical classifier. The confusion matrix is a two-dimensional array that contrasts the real label with the expected category labels. Binary categorization is indicated by the terms True Positive, True Negative, False Positive, and False Negative.

Precision Score

What percentage of the expected positive label is actually positive is known as precision.

$$Precision = \frac{True\ Positive}{(True\ Positive\ + False\ Positive)}$$

Recall Score

The fraction of actual positive labels that are accurately anticipated as positive is measured by the recall.

$$Recall = \frac{True\ Positive}{(True\ Positive\ + False\ Negative)}$$

F1 Score

Another effective performance matrix that makes use of both the accuracy and recall matrices is F1-score. By combining accuracy and recall, the "Harmonic Mean" may be used to determine the F1-score. Contrary to recall and accuracy, which are primarily concerned with false negatives, the F1-score concentrates on both false positives and false negatives.

$$F1\_Score = 2 \times \frac{(Precision \times Recall\ )}{(Precition\ + Recall)}$$

4.3 Experimental Results

To assess the performance of our suggested model like Convolution Neural Network (CNN), Inception V3, AlexNet. I used a variety of performance metrics. The confusion matrix, accuracy score, precision score, recall score, and f1 score were all completed. And also compared each other algorithms for knowing better accuracy.

4.3.1 Convolution Neural Network Classifier (CNN)

I got 97% accuracy by using the CNN model or classifier.

Train Model Accuracy

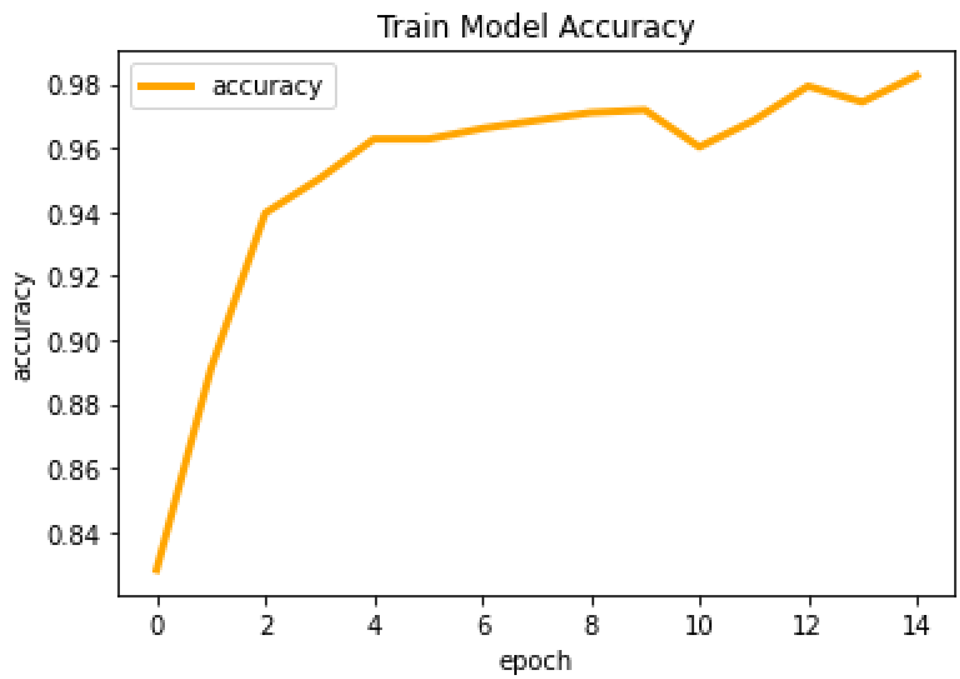

Figure 4.1 Train Model Accuracy

Train Model loss

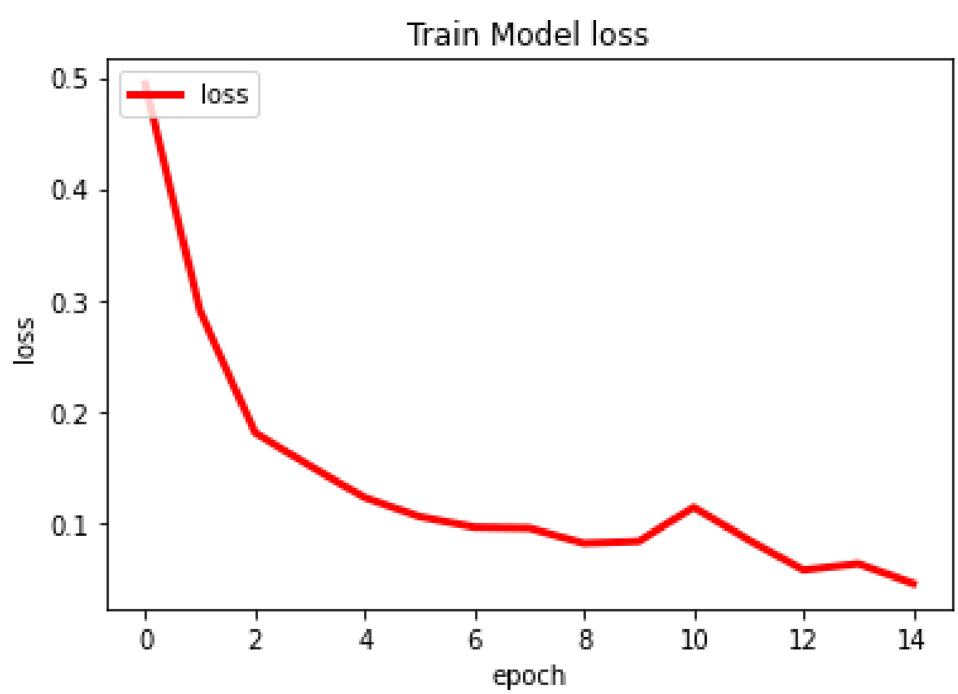

Figure 4.2 Train Model loss Accuracy

Actual test labels

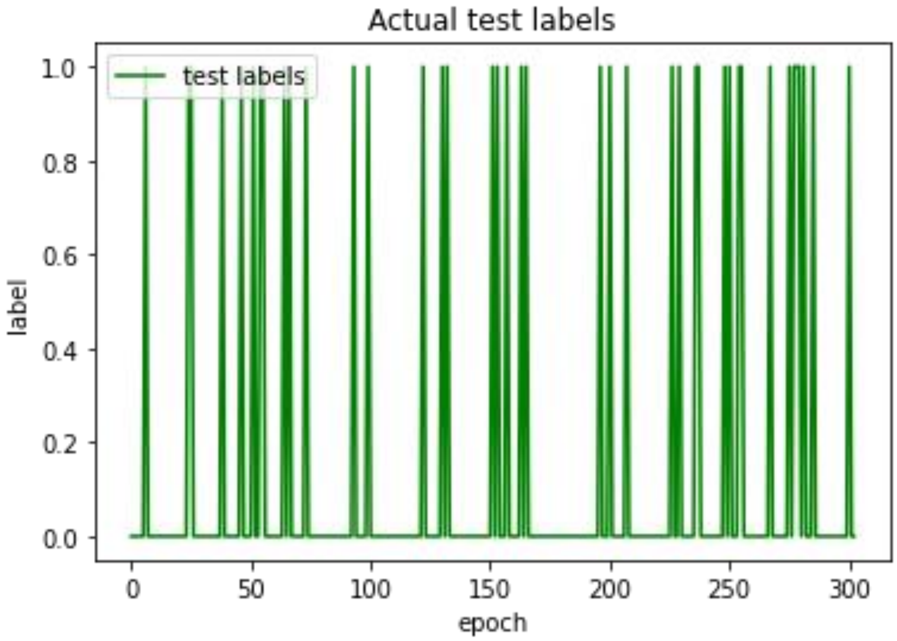

Figure 4.3 Actual test labels

Predicted test labels

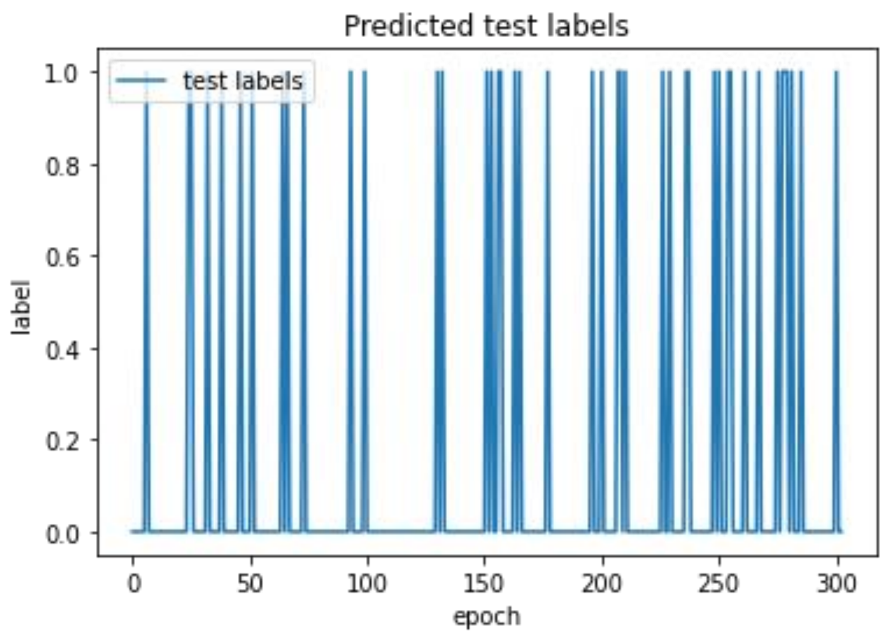

Figure 4.4 Predicted test labels

Final Accuracy

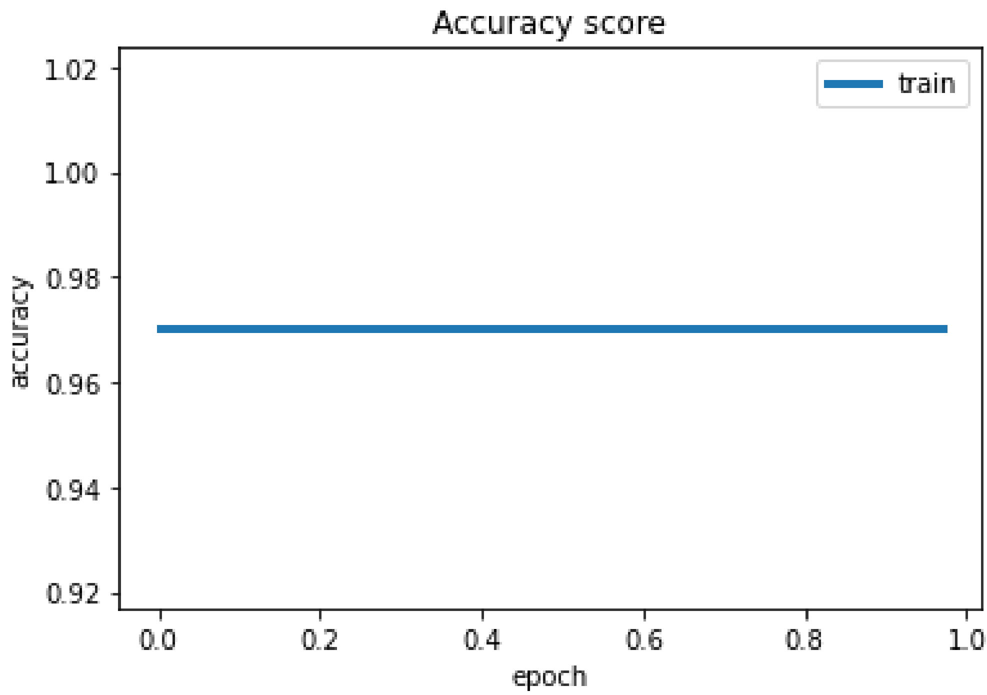

Figure 4.5 Final Accuracy of deployed CNN model

Classification Report

```
Confusion matrix :
 [[ 37   3]
 [  6 257]]
Outcome values :
 37 3 6 257
Classification report :
               precision    recall  f1-score   support

           1       0.86      0.93      0.89        40
           0       0.99      0.98      0.98       263

    accuracy                           0.97       303
   macro avg       0.92      0.95      0.94       303
weighted avg       0.97      0.97      0.97       303
```

Figure 4.6 CNN Classification Report

Above we can see that the precision, recall, f1-score is 0.92, 0.95, 0.97 respectively.

Confusion Matrix

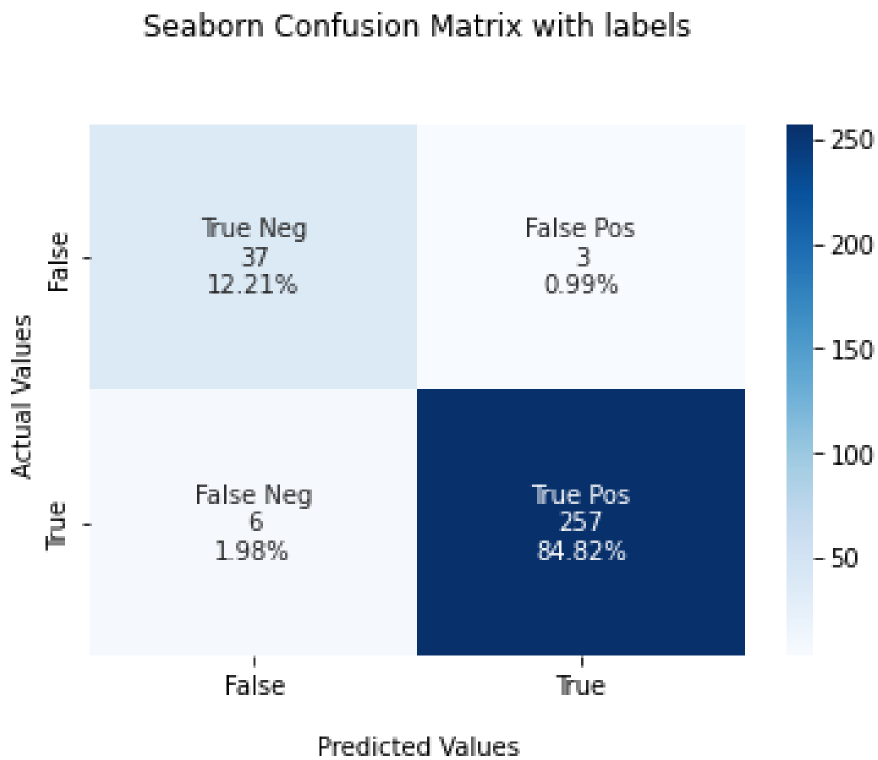

Figure 4.7 CNN Confusion Matrix

Above we can see that, the diagonal values of all classes increased very highly.

4.3.2 Inception V3

I got 92% accuracy by using the Inception V3 model. Now I see the developed model of every plot diagram.

Train Model Accuracy

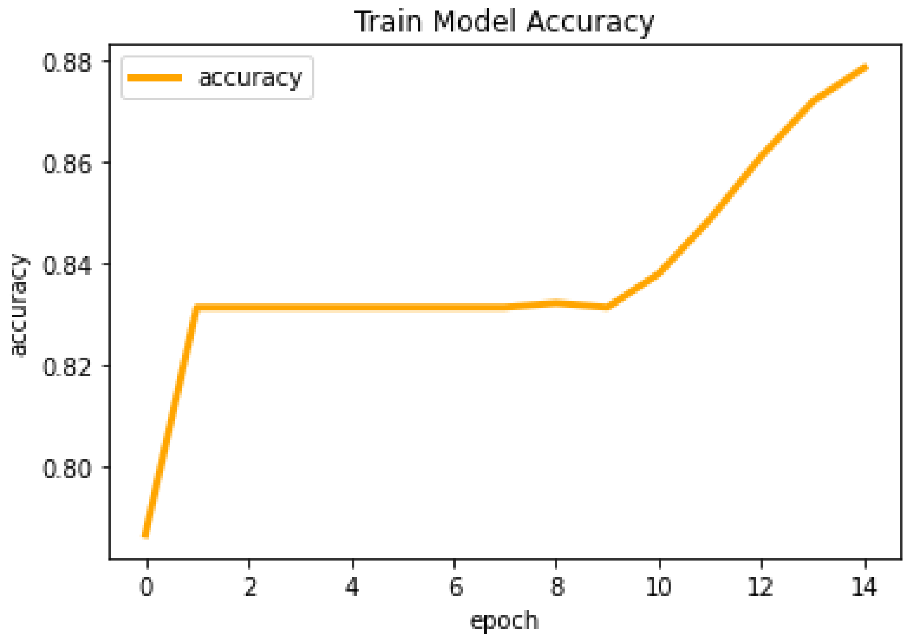

Figure 4.8 Train Model Accuracy of Inception V3

Train Model loss

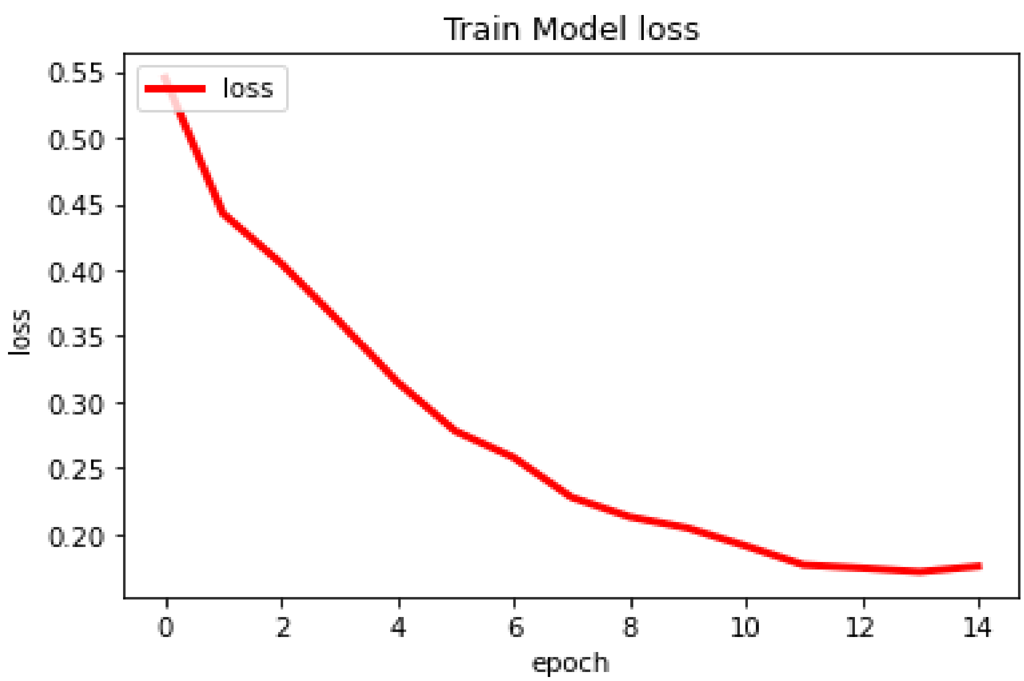

Figure 4.9 Train Model loss Accuracy I. V3

Actual test labels

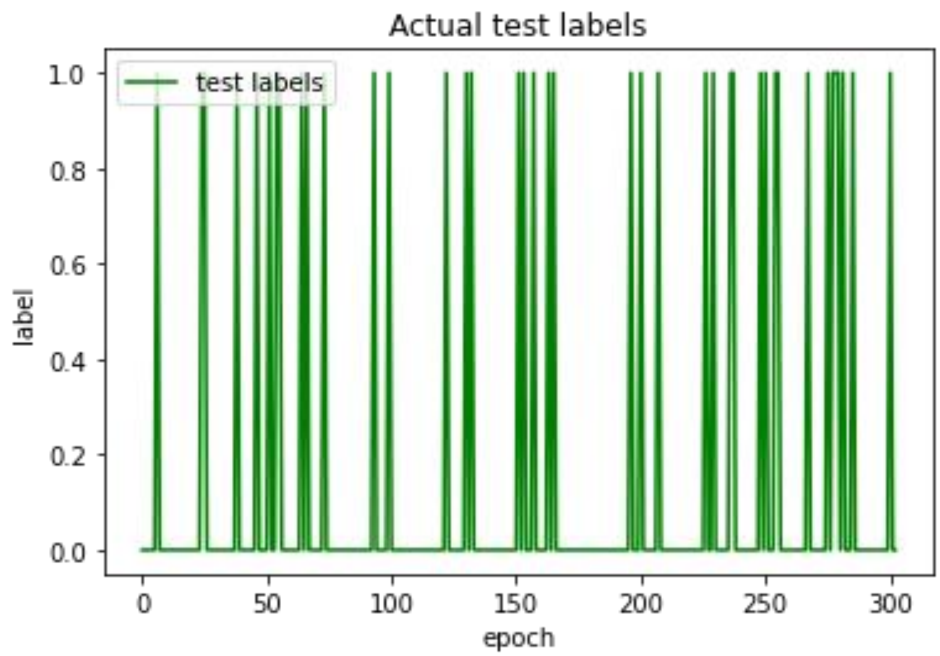

Figure 4.10 Actual test labels of I. V3

Predicted test labels

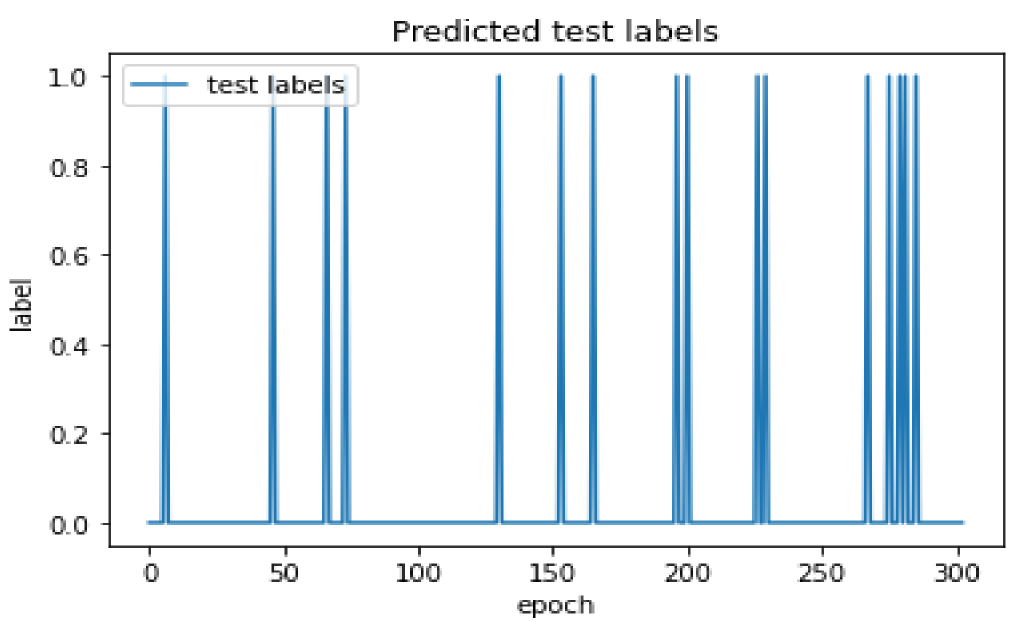

Figure 4.11 Predicted test labels of I. V3

Final Accuracy

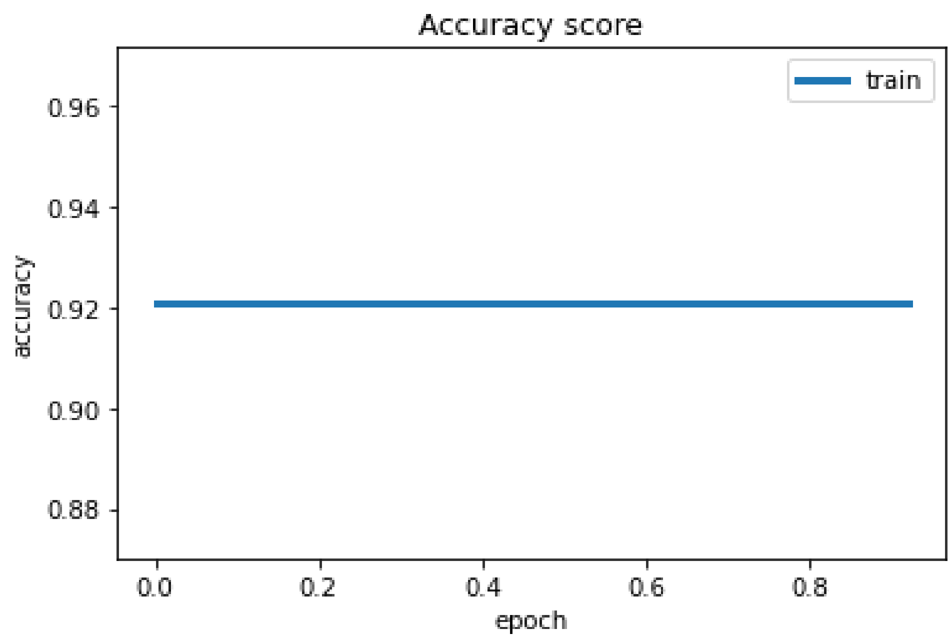

Figure 4.12 Final Accuracy of the deployed Inception V3 model

Classification Report

```
Confusion matrix :
 [[ 16  24]
 [  0 263]]
Outcome values :
 16 24 0 263
Classification report :
               precision    recall  f1-score   support

           1       1.00      0.40      0.57        40
           0       0.92      1.00      0.96       263

    accuracy                           0.92       303
   macro avg       0.96      0.70      0.76       303
weighted avg       0.93      0.92      0.91       303
```

Figure 4.13 Inception V3 Classification Report

Above we can see that the precision, recall, f1-score is 0.96, 0.70, 0.92 respectively.

Confusion Matrix

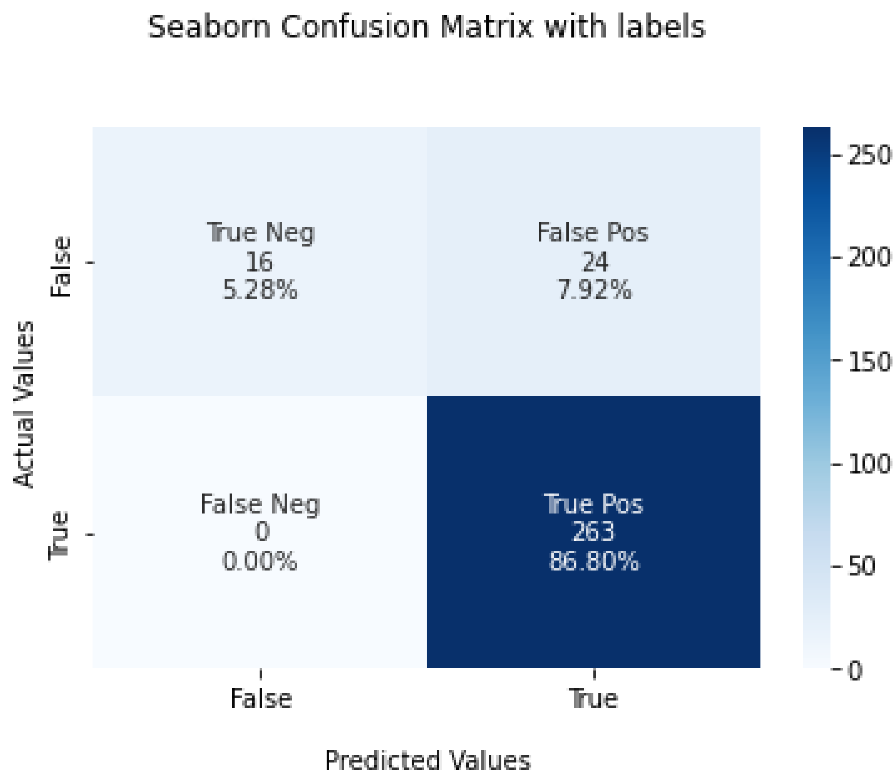

Figure 4.14 Inception V3 Confusion Matrix

Above we can see that, the diagonal values of all classes increased very highly.

4.3.3 AlexNet

AlexNet is a model or architecture of Convolution Networks. This model from, I got the same accuracy as Inception V2. Its accuracy is almost 92%.

Train Model Accuracy

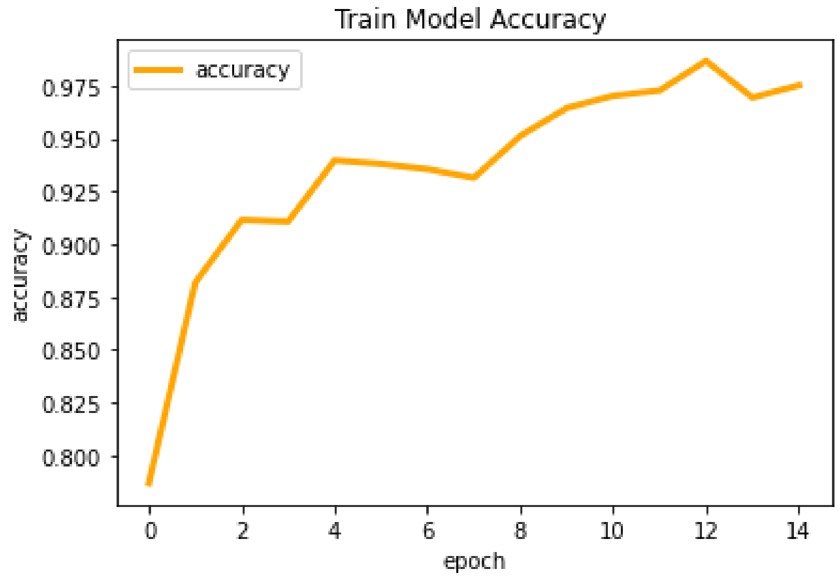

Figure 4.15 Train Model Accuracy of AlexNet

Train Model loss

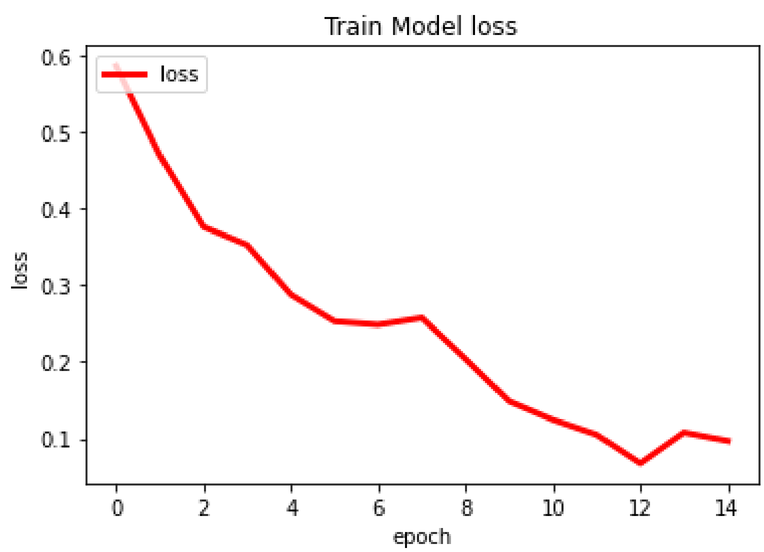

Figure 4.16 Train Model loss Accuracy of AlexNet

Actual test labels

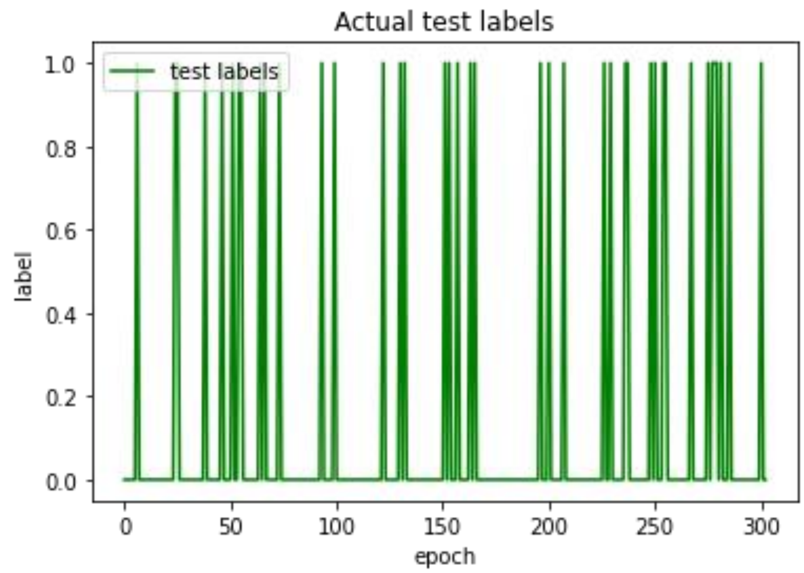

Figure 4.17 Actual test labels of AlexNet

Predicted test labels

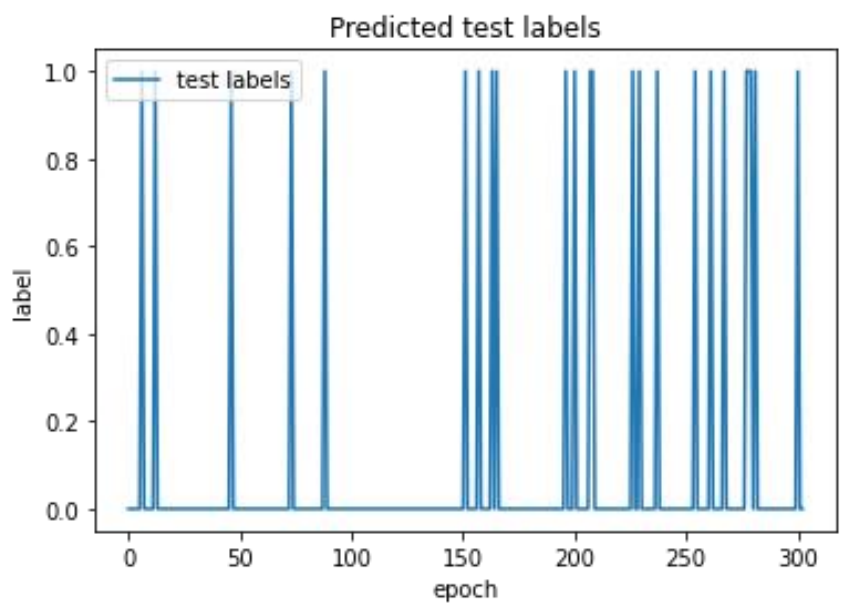

Figure 4.18 Predicted test labels of AlexNet

Final Accuracy

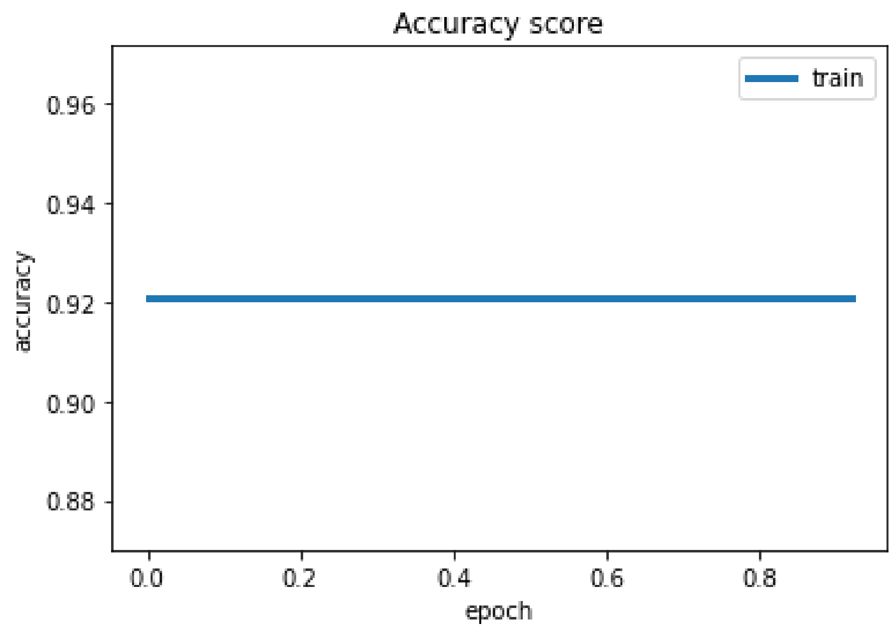

Figure 4.19 Final Accuracy of deployed AlexNet model

Classification Report

```
Confusion matrix :
 [[ 20  20]
 [  4 259]]
Outcome values :
 20 20 4 259
Classification report :
               precision    recall  f1-score   support

           1       0.83      0.50      0.62        40
           0       0.93      0.98      0.96       263

    accuracy                           0.92       303
   macro avg       0.88      0.74      0.79       303
weighted avg       0.92      0.92      0.91       303
```

Figure 4.20 AlexNet Classification Report

Above we can see that the precision, recall, f1-score is 0.88, 0.74, 0.92 respectively.

Confusion Matrix

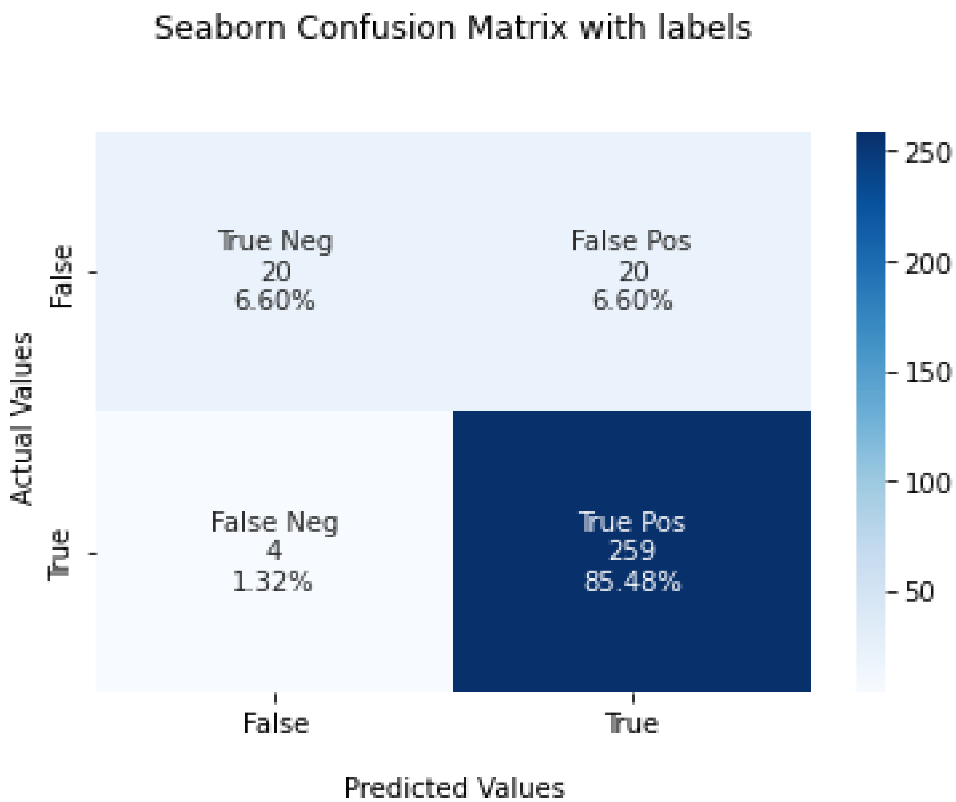

Figure 4.21 AlexNet CNN Confusion Matrix

Above we can see that, the diagonal values of all classes increased very highly.

4.4 Comparisons of Different Algorithm

At the end of this chapter, we can see that convolution neural networks perform better accuracy, better result, and better scorer for all algorithms. Now we can compare each other algorithms:

CNN – 97%

Inception V3 – 92%

AlexNet – 92 %

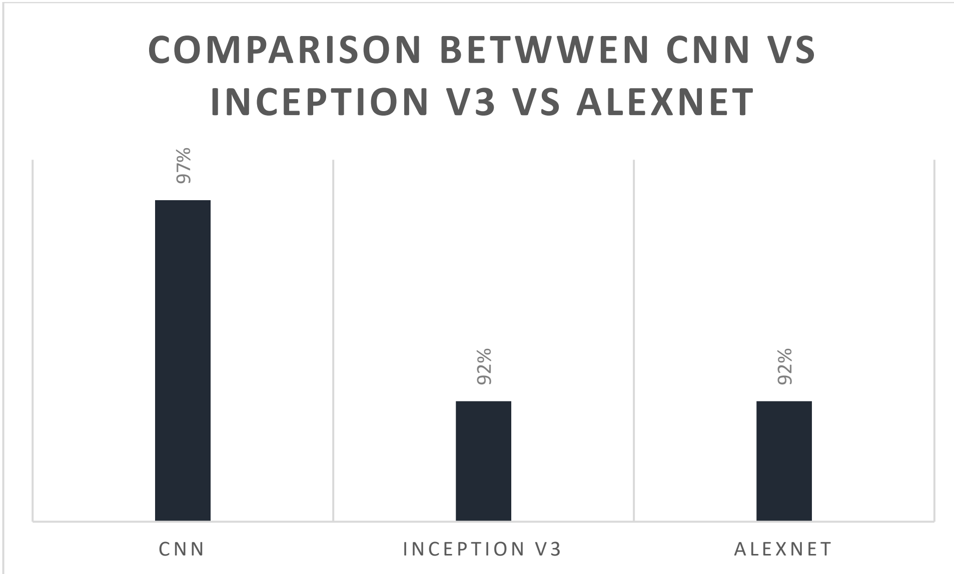

Figure 4.22 Comparison of CNN, Inception V3 and Alexnet

These three algorithms comparatively do better perform. Basically, Inception V3 and AlexNet accuracy were almost the same, But Convolution Neural Networks (CNN) given the high accuracy, high efficiency, and better result than the other two algorithms.

Three assessment matrices with varying levels of accuracy are shown in the picture. The accuracy level for CNN is 97%, as you can see in the image, while it is lower for I. V3 and AlexNet and higher for the other two models, whose accuracy levels are high (92%). This demonstrates how effective CNN is at performing vehicle detection and classification tasks.

CHAPTER 05

IMPACT ON SOCIETY

5.1 Impact on Society

A project's influence on society may be positive if it has been properly carried out. The three parts of this chapter's description of the impact of the vehicle detection and categorization project. It makes life easy for us. This project's impact on society and on people in general will be profound. Then there are ethical issues to think about. In order to understand how this project might assist automobiles and the traffic sector of a country, the ethical component of the study was carefully investigated. The project's long-term viability was also taken into account. Overall, it offers a simple solution to the issues brought on by the present traffic congestion and illegal cars. a foundation upon which future substantial study can be built.

5.2 Ethical Aspects

Traffic police in Bangladesh still identify vehicles and their associated information manually. We can only hope that this type of machine learning initiative will eventually eliminate it. A machine is incapable of making accurate predictions. The process will take some time. The model will be more robust when the database has millions of entries, and it is believed that over time, it will be able to make accurate predictions. If future artificial intelligence and deep learning can be connected to a database, there will come a day when not a single automobile will be missed it.

CHAPTER 06

CONCLUSION AND FUTURE WORK

5.1 Conclusion

With fast improvement in vehicle and traffic ventures, the development of the populace on the planet brought the requirement for various devices and methods, particularly innovative arrangements to oversee traffic in urban communities and populated regions. In the interim, object recognition can be utilized in different fields to assist people with residing effectively with solace and make the world a superior spot to live in. Object recognition can be utilized in enterprises, digitized urban areas, government, research, the scholarly community, climate, and so on Vehicle detection and classification is essential for object identification which is utilized in rush hour gridlock, urban communities, and so forth the significance of the subject is becoming bigger. That being said, this examination is expected to add to the improvement of the exactness of these calculations and models by means of accessible procedures and apparatuses. This Proposal created two classifier calculations to recognize and group vehicles. These three models are CNN, Inception V3, and AlexNet. The calculation determination depended on different investigations in the writing survey. The most proposed models by different analysts were these three models. Thusly, the creator chose to pick these models and contrast them all together to indicate the best model among these three. Numerous strategies have been conveyed to build the exactness level and to make the ideal outcome. The models are prepared with the equivalent dataset and the assessment result showed that CNN performs better compared to the next two algorithms. The consequence of the models is introduced in picture designs in which the framework distinguishes the vehicles that are passing from the screen and tracks them.

5.2 Future Works

Thinking back to the restrictions of the review, there are undertakings and choices which can be added to this examination or perhaps to chip away at it independently. The subject is under the consideration of scientists and works on step by step. Hypothetical advancement is required to have been followed and when any hypothetical advancement is distributed and accomplished, the scientists ought to use those ideas basically involving calculations to further develop the precision level of the location and following interaction.

The dataset can be improved, future work can be trying these models utilizing a superior and bigger dataset with a monstrous number of vehicle and non-vehicle pictures from better places, holy messengers, vehicles, streets, cameras, distances, and so forth besides, different models can be added to the examination rundown of models to make the correlation more dependable and tremendous.

As a result, the algorithm requires more work in order to become more efficient. Furthermore, the thesis was only concerned with detecting a single-vehicle. As a result, the existing CNN might be upgraded in the future to recognize several automobiles in images or videos.

References

[1] J. Zhang, F. Y. Wang, K. Wang, W. H. Lin, X. Xu, and C. Chen, "Data-Driven Intelligent Transportation Systems: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 1624–1639, Dec 2011.

[2] Baha, N. (2014). Real-Time Obstacle Detection Approach using Stereoscopic Images. International Journal of Information Engineering and Electronic Business, 6(1), 42.

[3] Kalghatgi, M. P., Ramannavar, M., & Dr. Sidnal, N. S. (2015). A Neural Network Approach to Personality Prediction based on the Big-Five Model. International Journal of Innovative Research in Advanced Engineering, 2(8), 56–63

[4] Kanistras, K., Martins, G., Rutherford, M. J., & Valavanis, K. P. (2015). Survey of unmanned aerial vehicles (UAVs) for traffic monitoring. Handbook of unmanned aerial vehicles, 2643-2666.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.

[6] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and R. V. Babu, "A Taxonomy of Deep Convolutional Neural Nets for Computer Vision," *CoRR*, vol. abs/1601.06615, 2016.

[7] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, pp. 396–404, 1990.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[9] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR*, vol. abs/1311.2524, 2013.

[10] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," *CoRR*, vol. abs/1312.6229, 2013.

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

[12] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3367–3375, 2015.

[13] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "Cnn-rnn: A unified framework for multi-label image classification," in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, pp. 2285–2294, IEEE, 2016.

[14] M. A. Manzoor and Y. Morgan, "Vehicle Make and Model classification system using bag of SIFT features," in *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual*, pp. 1–5, IEEE, 2017.

[15] M. C. Narhe and M. Nagmode, "Vehicle classification using SIFT," *International Journal of Engineering Research and Technology ESRSA Publications*, 2014.

[16] X. Ma and W. E. L. Grimson, "Edge-based rich representation for vehicle classification," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2, pp. 1185–1192, IEEE, 2005.

[17] J.-W. Hsieh, L.-C. Chen, and D.-Y. Chen, "Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition," *IEEE Transactions on intelligent transportation systems*, vol. 15, no. 1, pp. 6–20, 2014.

[18] X. Li and X. Guo, "A HOG feature and SVM based method for forward vehicle detection with single camera," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on*, vol. 1, pp. 263–266, IEEE, 2013.

[19] S. Sivaraman and M. M. Trivedi, “Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, 2013.

[20] F. Han, Y. Shan, R. Cekander, H. S. Sawhney, and R. Kumar, “A two-stage approach to people and vehicle detection with HOG-based SVM,” in *Performance Metrics for Intelligent Systems 2006 Workshop*, pp. 133–140, 2006.

[21] P. Negri, X. Clady, S. M. Hanif, and L. Prevost, “A cascade of boosted generative and discriminative classifiers for vehicle detection,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 136, 2008.

[22] Z. Dong, Y. Wu, M. Pei, and Y. Jia, “Vehicle type classification using a semisupervised convolutional neural network,” *IEEE transactions on intelligent transportation systems*, vol. 16, no. 4, pp. 2247–2256, 2015.

[23] Z. Dong, M. Pei, Y. He, T. Liu, Y. Dong, and Y. Jia, “Vehicle Type Classification Using Unsupervised Convolutional Neural Network,” in *2014 22nd International Conference on Pattern Recognition*, pp. 172–177, Aug 2014.

[24] C. M. Bautista, C. A. Dy, M. I. Mañalac, R. A. Orbe, and M. Cordel, “Convolutional neural network for vehicle detection in low resolution traffic videos,” in *Region 10 Symposium (TENSYMP), 2016 IEEE*, pp. 277–281, IEEE, 2016.

[25] H. Huttunen, F. S. Yancheshmeh, and K. Chen, “Car type recognition with deep neural networks,” in *Intelligent Vehicles Symposium (IV), 2016 IEEE*, pp. 1115– 1120, IEEE, 2016.

[26] Y. Zhou, H. Nejati, T.-T. Do, N.-M. Cheung, and L. Cheah, “Image-based vehicle analysis using deep neural network: A systematic study,” in *Digital Signal Processing (DSP), 2016 IEEE International Conference on*, pp. 276–280, IEEE, 2016.

[27] J. E. Espinosa, S. A. Velastin, and J. W. Branch, “Vehicle Detection Using Alex Net and Faster R-CNN Deep Learning Models: A Comparative Study,” in *International Visual Informatics Conference*, pp. 3–15, Springer, 2017.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[29] Q. Fan, L. Brown, and J. Smith, "A closer look at Faster R-CNN for vehicle detection," in *Intelligent Vehicles Symposium (IV), 2016 IEEE*, pp. 124–129, IEEE, 2016.

[30] M. Ozuysal, V. Lepetit, and P.Fua, "Pose Estimation for Category Specific Multiview Object Localization," in *Conference on Computer Vision and Pattern Recognition*, (Miami, FL), June 2009.

[31] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 416–431, 2006.

[32] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, (Madison, Wisconsin), pp. 264–271, June 2003.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A largescale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255, IEEE, 2009.

[35] Razakarivony, S., & Jurie, F. (2016). Vehicle detection in aerial imagery: A small target detection benchmark. Journal of Visual Communication and Image Representation, 34, 187-203.

Plagiarism Report

Vehicle Detection and Classification Using A Deep Learning Approach

ORIGINALITY REPORT

20%	11%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	5%
2	Preetha Jagannathan, Sujatha Rajkumar, Jaroslav Frnda, Parameshachari Bidare Divakarachari, Prabu Subramani. "Moving Vehicle Detection and Classification Using Gaussian Mixture Model and Ensemble Deep Learning Technique", Wireless Communications and Mobile Computing, 2021 Publication	5%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
4	www.coursehero.com Internet Source	1%
5	Submitted to Singapore Institute of Technology Student Paper	1%
6	Submitted to University of Hertfordshire Student Paper	1%