# Deep Learning Networks for Detection, Classification and Analysis of Car Damage

MSc Research Project
Master of Science in Data Analytics


Shubham Chaudhari
Student ID: x20160836

School of Computing
National College of Ireland


Supervisor: Prof. Mohammed Hasanuzzaman

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

| | | | |
|---|---|---|---|
| **Student Name:** | Shubham Sarjerao Chaudhari | | |
| **Student ID:** | x20160836 | | |
| **Programme:** | MSc in Data Analytics | **Year:** | 2021-2022 |
| **Module:** | MSc Research Project | | |
| **Supervisor:** | Prof. Mohammed Hasanuzzaman | | |
| **Submission Due Date:** | 31th January 2022 | | |
| **Project Title:** | Deep Learning Networks for Detection, Classification and Analysis of Car Damage | | |
| **Word Count:** | **6721** | **Page Count**: | **21** |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** Shubham Sarjerao Chaudhari

**Date:** 31th January 2022

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Deep Learning Networks for Detection, Classification and Analysis of Car Damage

Shubham Chaudhari
x20160836

**Abstract**

Cars are incredibly essential in the contemporary generation, and the ability to detect car damage autonomously is a feature that the automobile insurance industry is quite interested in. Assigning an object to each of these groups visually is known as classifying it. Machine learning as well as computer vision are used in visual image recognition technology. Research started by training a fundamental CNN model. Despite this, it fails to perform well due to a small and unbalanced data sample. The effect of model pretraining proceeded by fine model tuning is then investigated. Finally, several deep learning models are used to apply ensemble learning and transfer learning to the data given. According to research, transfer learning surpasses domain-specific models. According to the findings, MobileNet and InceptionResnetV2 could obtain training accuracy of around 93%.

## 1 Introduction

Cars represent a shift in mobility and comfort, enabling us to easily move from one point to another. The societal benefits (and also the costs) are immense. Vehicles have become a necessary mode of mobility in modern daily lives. The automobile, too, is considered as a tool towards presenting one's financial situation or as a representation of our socioeconomic status in numerous locations. Moreover, for many car enthusiasts around the world, the vehicle has become a source of interest. Consumer desire for automobiles has developed over time to encompass not only economy and durability, but also outstanding convenience as well as elegance. Cars are a diverse object class with such a wide range of designs as well as models, which could lead to more complicated and dependable computer vision algorithms as well as approaches.

Detecting flaws in car images is a particular example of an image-labeling task, even in the most foundational sense. Allocating a photo to a specific class or groups of subclasses is the first step in detecting image damage. The ability to detect car damage on a real - time basis, on the other hand, is a topic of research with a lot of operational ramifications. Automobiles are routinely subjected to damage inspections, both of which are inconvenient for consumers and expensive for businesses.

Inside this section, we'll examine common types of damage including dents, cracks, as well as scratches. Because of the brightly gleaming metallic frames of cars as well as the fact that photos are acquired in an unconstrained environment, traditional machine learning algorithms are difficult to employ in this scenario. This inspired us to investigate other methods for automating this procedure using recent Deep Learning algorithms. I will detail the building of a conceptual framework capable of recognizing and maybe locating faults in automotive images, and also performance when compared to those of other systems, all through this research. To train the system and evaluate its effectiveness, a dataset collection will be obtained. This dataset would be created using images from websites and other publicly available sources. Figure 1 shows sample dataset used by K. Patil et al., 2017.

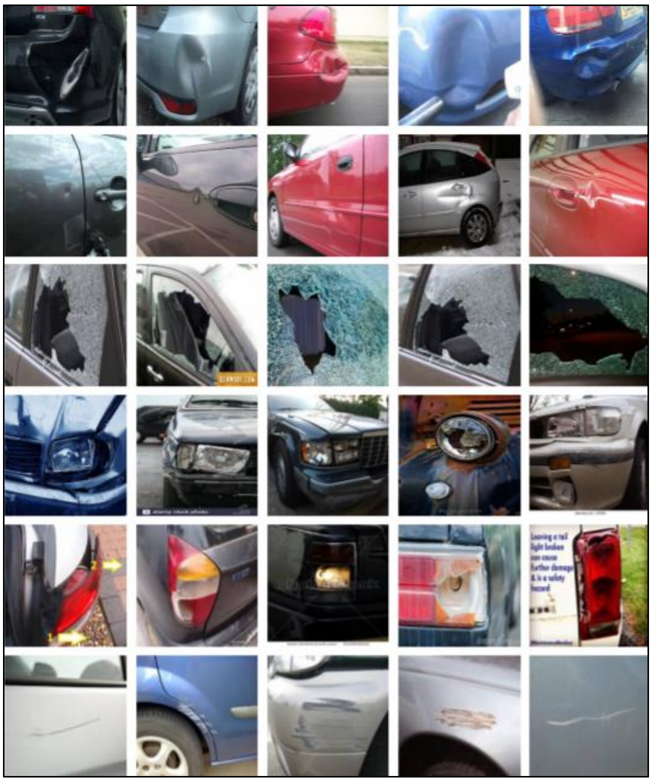

**Figure 1: Car Damage Images (K. Patil et al., 2017)**

## 1.1 Research Questions:

- **"How efficient is the object detection algorithms when detecting vehicle damages?":** Object detection, as we all know, comes before object classification, and our goal is to improve object detection efficiency before training the classifier. Our research will demonstrate how effective our object detector is in detecting various sorts of damage in car images.
- **"How efficient is the deep neural network when classifying vehicle damages?":** Deep neural networks are mostly used for computer vision challenges, and our goal is to identify different forms of vehicle damage using several deep learning models. This research question will justify how accurately and efficiently these deep learning models will perform.
- **"How will the pre-trained models work on smaller dataset to detect damage efficiently?":** From the perspective of deep learning models, it has to be made sure that sufficient data is available for training the models. Training of smaller datasets can lead towards model overfitting. The question above examines the application of pre-trained deep learning models to the dataset provided and analyzes the outcomes in perspective of the use case.

## 1.2 Aims and Objectives:

The purpose of this research is to use Deep Neural Networks to tackle the problem of auto insurance companies automatically identifying and finding vehicle damage from images of automobiles so that they can submit claims faster.

The following are the goals and objectives of the car damage detection and classification research project:

- **To create the dataset, collect image data:** As a publicly available dataset is not enough for model training, we have to scrap data from Google for building a dataset for car damage detection and classification.
- **Image Detection:** To achieve damage detection objectives, images need to be annotated with different car damage categories. Bounding boxes would be used to represent such outcomes
- **Image Classification:** To meet damage classification goals, image processing techniques such as Image Augmentation may be used to increase the number of images and to analyze and compare alternative deep learning models on the same dataset.
- **Use of pre-trained deep learning models:** Pre-trained models are used on a smaller dataset to assess and compare different models depending on the results obtained. The goal was to choose between various models derived from the literature review and use cases provided.
- **To perfection, configure the model's setup using hyper parameters, as well as the learning process parameters:** For getting the highest accuracy and better prediction results for damage detection deep learning models, hyperparameter tunning will be performed. Parameters like epochs, batch size, learning rate, dropout layers, optimizer, etc. will be tuned.
- **Model evaluation with various matrices:** After completion of model training, models need to evaluate on the basis of different evaluation matrices. Matrices like confusion matrix, precision, recall, f1-score, support and model’s training, testing and validation accuracy. These matrices will give a clear picture of the model’s performance and ability to predict correctly.

The major contribution of this study is a thorough exploration and evaluation of the performance of a deep neural network neural network model reinforced with a forefront propagation modules in the domain of automobile damage detection and classification in a deep neural network scenario.

## 1.3 Scope and Limitations

The goal of this research is to see if a deep neural network model for the detection and classification of vehicle damage can achieve equal or even superior accuracy rates in detecting and classifying vehicle damage. This study goes beyond earlier research in that it aims to exploit the substantial refers to a process of data annotations using bounding boxes and apply a bigger training dataset utilizing image augmentation techniques consisting mostly of labeled data to decrease the time and expense associated with data preparation.

Even if the data processing work is minimized, compiling raw datasets to an increased number of images that are made available for training still requires a significant amount of time. Furthermore, all the images need annotation, which takes a long time. Because the researcher has minimal annotation tool expertise and experience, a significant percentage of the remaining effort and time will be spent on the model's setup and configuration.
Because the project's overall timeline is restricted to 3 months, the overall analysis was limited for different detection and classification model architectures derived from previous literature. This choice of different models enables for straightforward comparisons to earlier research.

### 1.4 Project Plan:

Figure 2 shows breakdown of the research:

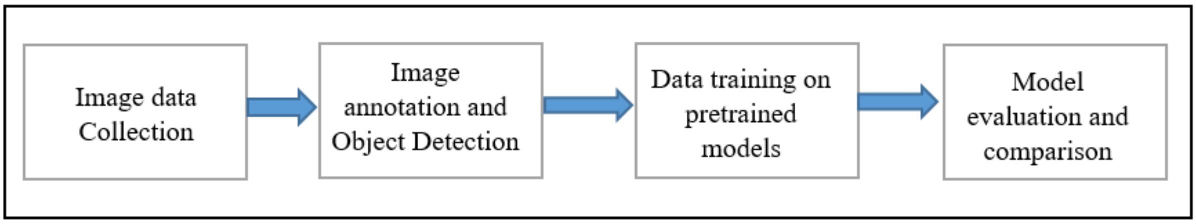


**Figure 2: Project Plan**

## 2 Related Work

Car accidents have become a major concern in recent years, leading to several vehicle damages. Considering the damage in the vehicles and its detection using deep learning in the images is a relatively specific study topic, and that as a whole, there hasn't been enough done within this area. This section will detail the research which has been performed in the previous few years.

### 2.1 Object Detection and Classification

Object detection performance on either a classic PASCAL VOC data has remained stable in recent years. In this paper, they (R.Girshick et al., 2014) describe a robust and simple detection approach that can improves mAP (mean average precision) by even greater than 30% well over previous best performance using VOC 2012, with such a mAP of 53.3 percent. Their approach incorporates two key concepts: (1) Using bottom-up selected features, higher-performing convolutional neural networks can indeed be utilised to recognise and split objects, and (2) When labelled training data is limited and costly, supervised pre-training for just a single additional task followed by fine-tuning yields an important result

Large neural network models have recently demonstrated good performance of the classifier in the ImageNet test (Krizhevsky et al., 2017). There is, however no clear understanding of why they work so well or how they may be improved. This research addresses both of these issues. They (M. Zeiler and R. Fergus, 2014) provide a novel visualization method for comprehending the functionality of convolutional feature regions and also the activity of a classifier. They show how their ImageNet improves decision making upon that Caltech-101 and Caltech-256 data, as well as on a variety of many other datasets. Their convnet model transitioned less well with the PASCAL data with no task adjustment, probably because to dataset biases, although it was within 3.2 percent of the greatest successful transaction.

In the year 2012, convolutional neural networks (CNNs) remerged (A. Krizhevsky et al., 2012). Given that a deep convolutional neural network could learn both stable and high-level visual properties from a source image, that's only natural to question if it can be used to distinguish objects. (R. Girshick et al., 2014, 2016) took the lead in resolving the deadlocks by proposing Regions with RCNN enabling object detection during 2014.

Object recognition is only one of the many sectors wherein computer vision has indeed been extremely successful. Object detection is the initial step in the development of self-driving cars and robotics They (A. Pathak et al., 2018) decoded the purpose of cnn-based deep learning techniques for object tracking. The study also goes through the different deep learning architectures as well as object identification services that are available. Object localization and identification datasets which have been published in international competitions are also addressed.

The operation time of the detecting network has been reduced by SPPnet and Fast R-CNN, showing region proposal computation as a bottleneck. In this work, they (S. Ren et al., 2015) offer a Region Proposal Network that exchanges full-image feature presentations with the detection technique, allowing for near-free sector recommendations. Their detection approach uses 300 proposals per image and operates at 5 frames a second (including all phases) on a GPU for an extraordinarily deep VGG-16 model, achieving state-of-the-art feature identification accuracy with PASCAL VOC (73.2%) & (70.4)

The Histogram of Oriented Gradients (HOG) feature detector was first proposed in 2005 by (S. Jayawardena, 2013). HOG represents a substantial development in the production of the complex structures transform (D. Lowe, 1999, 2004) along with shape limitations. HOG was created to solve the problem of recognising pedestrians, yet it can detect a wide set of objects. To recognise items of varied sizes, a HOG detector was employed to resize the image many times while keeping the border window size static.

Utilizing cascaded detection to detect objects is a common practise (P. Viola and M. Jones, 2001, F. Fleuret and D. Geman, 2001). Fundamental calculations are performed to filtering out the bulk of the essential backdrop frames, and afterwards complex calculations are utilised to evaluate the more demanding windows, according to a coarse-to-fine recognizing approach. Cascaded detecting has been applied to deep learning classifiers in past few decades, especially for "small item detection in huge scenes" scenarios such as facial recognition (H. Li et al., 2015).

Deep learning allows frameworks to learn extremely complex, sophisticated, and conceptual interpretations, leads to significant progress in fields such as object recognition, object recognition, virtual assistants, computer vision, natural language processing, drug discovery, medical image identification, and cell biology. DCNNs (Y. LeCun et al., 2015) are a type of deep learning model (DNN) that have improved image, video, speech, and audio processing. Object identification based on deep learning becomes a major study topic for many years because to its strong learning capabilities and benefits in coping with visibility, scale change, and background modifications. With varied degrees of R-CNN adaption, this paper (Z. Zhao et al., 2019) gives a complete review of object recognition system that resolve a range of subproblems, including interference, distortion, and low resolution. Finally, they suggest some intriguing potential avenues for gaining a complete understanding of the object detection context.

The great advantage of deep learning photo categorization has led to a lot of interest in object identification systems based on deep learning models in latest days. Throughout this article, they (X. Wu et al., 2020) present a complete summary of current breakthroughs in visual object recognition utilizing deep learning. They examine a large quantity of current relevant studies and divide the survey into three sections: I detection components, (ii) learning strategies, and (iii) implementations and benchmarks. In the research, they go over in depth a number of variables that affect detection accuracy, such as detector layouts, knowledge acquisition, proposed plan creation, sampling strategies, and so on.

For generalised object identification, coping with various degrees of variation in various object categories using tractable computations is a difficulty, necessitating informative and adaptable item presentations that are also efficient to assess for different locations. As a result, they (X. Wang et al., 2013) propose employing a cascaded improving classifier to create an object class by merging different types of features from competing local areas, known as regionlets. Upon that PASCAL VOC dataset, it achieves a recognition accuracy percentage of 41.7 percent over 20 item categories, and 39.7 percent upon that 2010 - VOC dataset.

Furthermore, research by (C. Szegedy et al, 2013), which again is ongoing concurrently with one another, reveals that this strategy may not even function in practise (Accuracy of 30.5% percent on VOC 2007). Another possibility is to build a sliding window analysis tool.

As powerful computer tech has already been enhanced, object identification technology based on neural networks has improving individual. Because of ultimate aim of the plan is to obtain higher precision as well as efficient detection systems, research teams (L. Jiao et al., 2019) have found a number of ways to pursue, including developing new architectural styles, attempting to extract advanced features, attempting to utilise good representations, attempting to improve computational power, trying to train from scrape, orientation methods, & attempting to solve advanced image issues (relatively small objects, occlusions objects), and also incentivized learning.

## 2.2 Car Damage Detection and Classification

Several damage detection systems for identifying automobile body damage have just been reported. (S. Jayawardena, 2013) proposes leveraging 3D CAD structure to perform automated photo-based automotive damage recognition. To find defects in the vehicle, they employ a database of undamaged 3D CAD replicas of cars as a testing sample data. Picture boundaries that do not visible in the 3D CAD architecture projection are defined as automobile injury.

In An Prevent Intrusion Detection System for Automobile Insurance Claim Related to Visual Proof, they presented a technique for creating robust image feature representation by appropriately finding the deficiencies, such as using YOLO to recognise the damage areas. To solve automotive body flaws, (Gontscharov et al., 2014) employs a variety of sensors data mining techniques.

(Cha. et al., 2017) employed a phase-based image features as well as extended Kalman filtering to determine defect damages after obtaining photos with the help of a web camera, then pre-processing wherein the received images are resized and divided, and then feature extraction for acquire the fault structure. A. Mohan and S. Poobal investigate and evaluate break detection through computer vision. Based on the empirical findings, they (A Mohan et al., 2018) assert that a greater number of studies have utilized a superior segmentation method to evaluate a webcam type image.

Damage detection methods based on image analysis have received a lot of attention. (Maeda et al. 2018) suggested using convolutional neural networks to construct a damage identification system for detecting damage to road surfaces. (S. Jayawardena, 2013) provided a method for identifying automobile scratches damage by constructing a 3D CAD modelling of an intact car using an image of a damaged car. (Kalpesh et al., 2017) proposed CNN for vehicle damage categorization, which achieved a performance of 89.5 percent by combining transfer learning with ensemble methods. It was discovered that it outperforms domain specialized training and just generating a CNN, but it falls short due to a small collection of supervised methods.

Image-based insurance processing is a vital industry with plenty of room for automated systems. They investigate the subject of automotive damage classification in (K. Patil et al., 2017), where some of the categories may be fairly fine-grained. They are investigating deep learning ways to achieve this goal. They start by retraining a CNN on their own. However, due to a small number of annotated photographs, it does not work well. Eventually, they put their beliefs to the test using transfer learning and ensemble learning. Transfer learning, according to research, outperforms domain-specific optimization. They employed a Convolutional Neural Network (CNN)-based technique to categorize various forms of vehicle damage throughout the investigation.

(K. He et al., 2016) propose an end-to-end method for automating that procedure benefits both the business and the customer. They used well-known extracting features models like the Mask R-CNN as well as an ensemble among these two, as well as a transfer learning focused VGG16 network, to complete many tasks of discovering and classifying various types of components and defects observed in the automobile.

In (H. S. Malik et al., 2020), they focused on the subject of vehicle damage classifications, which insurance companies may employ to expedite the process of completing auto insurance applications in a timely manner. They painstakingly acquired and labelled images of various types of vehicle damage from a variety of online sources. They have been able to achieve a high efficiency of 96.39 percent by using CNN architectures pre-trained upon that ImageNet dataset.

Claiming leakage is a severe problem for insurance providers that costs them a lot of money. Those deficits, as per experts (N. Dhieb et al., 2019), are the consequence of the firm's inaccurate billing procedure, fraud, and poor decision-making. They provide automated and effective deep learning-based solutions for locating and detecting automotive damages in vehicles (N. Dhieb et al., 2019). The proposed strategy combines deep convolutional neural network, instance segmentation, as well as transfer learning approaches to extract features and identify damage.

This study (Q. Zhang et al., 2020) proposes a vehicle damage detection and classification system that relies on transfer learning as well as an enhanced mask area-based CNN to quickly address road accident settlement difficulties. This study (Q. Zhang et al., 2020) improves the accuracy of car damage detection and characterization by adjusting the model framework of the bounding box as well as the loss function by restricting the number of layers upon layers within the recurrent neural network and making adjustments to the internal structure to reinforce the model's generalization ability and regularization.

Collecting higher quality datasets is an important aspect of constructing deep learning models because the algorithm can only be as efficient as the datasets it trains from. As a result, a significant amount of money and time will be spent on data analysis. Since there are no publicly available datasets for vehicle damage detection and classification with sufficient number of images for our purpose, we will create our own dataset using web scrapping technique. This will include images collected from various sources like Google Images and Yahoo Image. With the aid of scraping, several images of damaged vehicles will be acquired from the internet at first. The annotation tool like CVAT with annotation type as Bounding Box will be considered to manually annotate images specifying the damaged section.

In order to get better results from deep learning models, it is usually required to expand the number of images in the dataset utilizing image augmentation techniques. These methods are even more relevant when the dataset size is small. Not all methods should be utilized in every situation, and their usage might actually reduce the effectiveness of the models. Following with the splitting of dataset in training and testing sets, a variety of models depending on the data will be constructed, and the data modeling stage will commence.

# 3 Research Methodology

The effectiveness of damage classification and detection of deep learning models enhanced with hyper-parameter tuning is investigated in this research study. It expands on prior work that only used the underlying CNN algorithm in a supervised environment. Section 3.1 of this chapter begins with business understanding. Section 3.2 and Section 3.3 of this chapter introduces the data understanding and data pre-processing steps. These sections give a clear understanding of the data and it’s pre-processing required for this study. Section 3.4 focuses on the modeling segment, where deep learning models for detection and classification are discussed. Also, the architecture of these models has been discussed briefly. Whereas sections 3.5 and 3.6 give a glimpse of research evaluation and deployment.

## 3.1 Business Understanding

This is the first step in the research study, and it focuses on and reveals important components with respect to a business perspective. The following are the goals of the damage classification and detection system:

- To create the dataset, collect image data.
- Image Detection: Bounding boxes would be used to represent such outcomes.
- To perfection, configure the model's setup using hyper parameters, as well as the learning process parameters.
- Model evaluation with various matrices

## 3.2 Data Understanding

This research study is now in its second stage, which focuses on data collection, quality control, and pattern discovery in order to acquire insight into the data and generate hypotheses regarding missing information.

## 3.3 Data Pre-processing

The third step of the research study focuses on the identification and manipulation of the finished data collection. Selecting entries, tables, and attributes, as well as filtering and altering data, are all possible operations during this step. In this phase, we have performed various data preparation tasks like data cleaning, data augmentation, data annotation and data resizing.

## 3.4 Modeling

The fourth phase of the research study involves the evaluation and implementation of alternative modelling methodologies. Several parameters are chosen and different models are produced due to a similar data mining challenge. Before deciding on the deep learning detection and classification models for this project, a detailed examination of existing and reviewed literature was conducted.

### 3.4.1 Damage Detection

#### 3.4.1.1 RetinaNet

In computer vision, object detection is a significant topic. The algorithm is tasked with locating objects in a picture while also classifying them into various categories. Object detection models are divided into two types: single-stage & two-stage detectors. Two-stage detectors are usually more accurate, although they are slower. RetinaNet is a well-known single-stage detector that is both accurate and fast. RetinaNet employs a feature pyramid network to recognise objects at several scales and provides a new nonlinear function, the Focal nonlinear function, to address the excessive foreground-background imbalance problem.

### 3.4.2 Damage Classification

#### 3.4.2.1 CNN

Convolutional Neural Networks (CNNs) are a popular form of deep neural network. Convolution is the name given to a statistically linear process among matrices. A completely connected layer, fully connected layer, a pooling layer, and a non-linearity layer are among the CNN layers. Pooling as well as non-linearity tiers aren't required to have features, but convolution & fully connected layers are. The CNN performs admirably when it comes to machine learning challenges. Image data advances, such as the world’s largest image classification data collection (Image Net), natural language processing and computer vision, were particularly notable.

#### 3.4.2.2 MobileNet

TensorFlow's first mobile computer vision architecture is the MobileNet framework, which as its name suggests, is designed for use in mobile apps. MobileNet employs depth-wise separated convnets. When compared to a solution with typical convolution layers of the very same depth in the systems, the total complication in terms of parameters is significantly reduced. As a result, lightweight neural networks have been developed. MobileNets is indeed a TensorFlow package of mobile workstation vision networks that optimise accuracy while taking into account the device's or connected app's scarce resources. MobileNets appear to be low-latency, low-power models which have been customised to match the resource constraints of different applications. On top of them, classification, identification, similarity measures, and fragmentation can all be built.

#### 3.4.2.3 VGG16

The VGG16 Convolutional Neural Network (CNN) is a simple and widely used CNN. The model exceeds 93 % of top-5 testing accuracy using ImageNet, a collection of data including over 14.5 million images belonging to 1000 categories. It built on AlexNet's concept by successively replacing large kernel-sized filters with multiple 3*3 kernel-sized filters. VGG16 is used in a number of deep learning algorithms for image categorization, and it is popular due to its ease of use.

#### 3.4.2.4 InceptionResnetV2

The Inception-ResNet-v2 structure is a CNN based on the Inception family of models, but with residual connections. Over a million pictures from the ImageNet collection were used to train the fully convolutional Inception-ResNet-v2. The 164-layer networks can classify pictures into 1000 different item categories, such as mouse, laptop, bottle, and other animals. As a result, the network was able to learn a wide range of complete feature predictions for a wide range of pictures.

## 3.5 Evaluation

This is the fifth stage of the research project, and it focuses on analysing the designs which have been created and identifying using the results. The algorithm determines how well the design is understood, allowing the models to be assessed to see whether they meet their goals. To precisely assess the performance of deep learning models, performance metrics such as train accuracy, validation accuracy, test accuracy, precision, f1-score, recall, support, and confusion matrix are used. The following is how accuracy and recall are calculated:

$$\textbf{Recall} = \textbf{TP} \div (\textbf{TP} + \textbf{FN})$$

$$\textbf{Precision} = \textbf{TP} \div (\textbf{TP} + \textbf{FP})$$

$$\textbf{F1-Score} = 2 * ((\textbf{Precision} * \textbf{Recall}) \div (\textbf{Precision} + \textbf{Recall}))$$

$$\textbf{Accuracy} = (\textbf{TP} + \textbf{TN}) \div (\textbf{TP} + \textbf{FP} + \textbf{FN} + \textbf{TN})$$

Where:

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

## 3.6 Deployment

This is the research study's sixth and last phase, and it focuses on selecting how to use the data and results acquired. This stage also focuses on data collection, analysis, and presentation. Figure 3, shows research process architecture conducted.

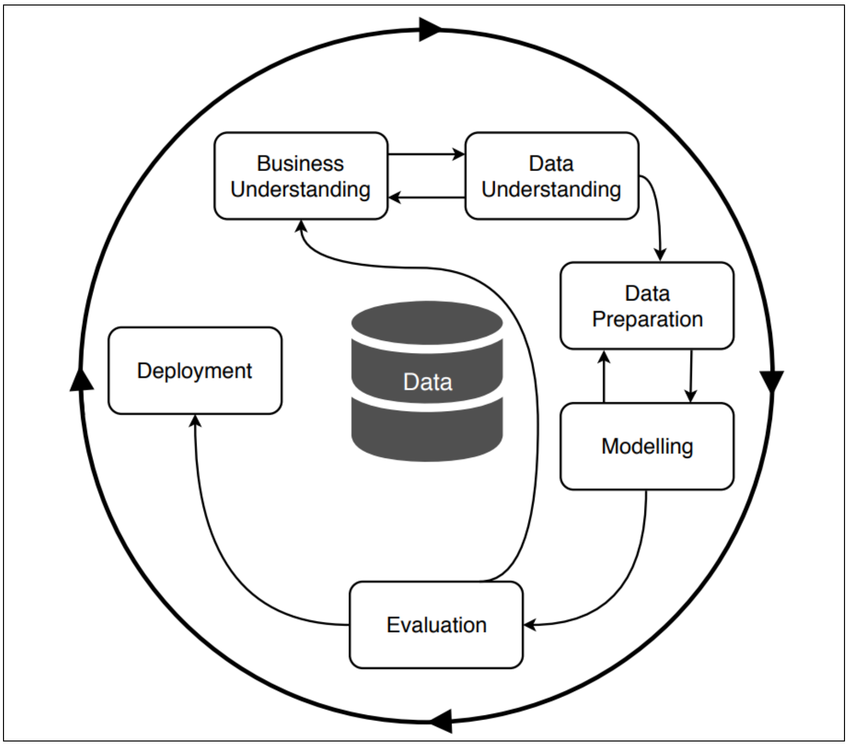


**Figure 3: Research process architecture (L. Contreras-Ochando et al., 2019)**

# 4 Implementation, Evaluation, and Results

## 4.1 Data Collection

- **Data Collection:** There are two phases to detecting and classifying car damage: Damage detection and classification.
  For building a classifier for damage detection and damage classification, I have used **SerpAPI** to scrap data from Google also downloaded Google chrome extension for downloading images.
- **Data Description:** In the following sections, a brief overview of obtained data will be provided. After data scrapping, all data was saved in a CSV file which consists of 4 different columns namely "image_name", "downloade_date", "imageurl", and "search_query". The number of images scraped from Google was divided into the following six categories:
  - Dent on the Bumper - 588
  - Dent on the Door - 611
  - Windshield - 772
  - Head Lamp - 720
  - Scratch - 691
  - Tail Lamp – 502

- SerpAPI was used to scrape images from Google Images and Yahoo Images, as well as images that were also downloaded from Google Chrome plugin to construct a dataset for vehicle damage detection and classification.

## 4.2 Data preparation

- **Data Cleaning:** Before we can use our dataset for training, we need to clean it up for better results. Let's talk about dataset two, which is Google data scraped. The problem faced while processing the data was that so many pictures from different categories were not connected to that particular category. As a result, the simplest choice for removing those unconnected photographs was to delete them manually. Similarly, the data cleaning process was finished by manually eliminating any images that were irrelevant, fuzzy, duplicated, or of a small scale. After data deletion, the total count of images was 3474.
- **Data Annotation for Object Detection Classifier:** The process of labelling data so that machines may utilise it is known as data annotation. CVAT is a web-based image and video annotation tool that is free as well as open source. It is used to label data for computer vision applications. Interpolation of patterns between key frames, semi-automatic labelling using deep neural networks, shortcuts for the most important activities, and a dashboard with such a list of annotation activities and assignments are just a few of CVAT's strong features. Data annotation are applied to all six categories. For image annotations, initially account will need to be created on CVAT for further processing. After creation of a particular task, we can manually annotate every image. In this, bounding boxes are used to allocate a particular type of damage. Figure 4 shows annotations of bumper dent and scratch damage types.

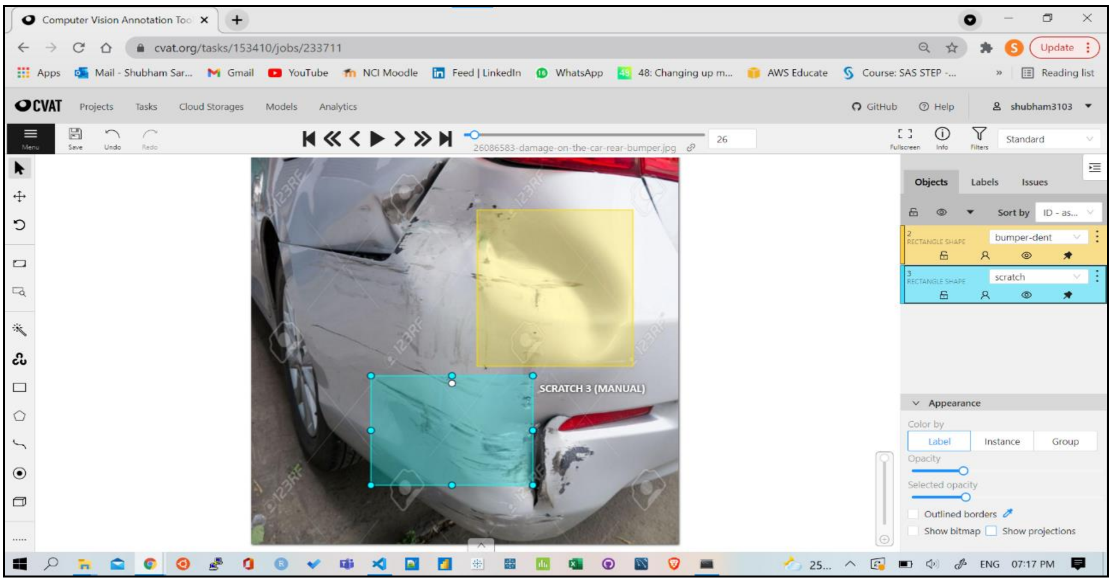


**Figure 4: Annotation of specified task**

- **Data Augmentation for image Classification Classifier:** The process of increasing the volume of data necessary to train a classifier is known as data augmentation. A large training dataset is sometimes required for deep learning models, which is not always available. As a consequence, existing data is added to create a more complete model. Because of deletion, the amount of data was decreased after the data cleaning stage. Unbalanced data was the result of this. I have to execute data augmentation across all categories to eliminate data unbalancing. Below is the image count I utilised for further processing following data augmentation:

- Dent on the Bumper - 1500
- Dent on the Door - 1500
- Windshield - 1500
- Head Lamp - 1500
- Scratch - 1500
- Tail Lamp - 1500

- **Image Resizing:** Image compression is the process of increasing or decreasing the size of an image without removing any content. You are simply changing the size of the file whenever you resize an image. Because there are times when size matters. Your image may need to be resized to fit into a certain region on a display, including a blog or social media post.

  Our acquired car image data was of varying sizes; each image's pixel was unique, necessitating resizing to the same scale. As a consequence, all of the images were converted to a fixed size format of 1024 by 1024 pixels. These scaled photos were then utilised for colour conversion and other post-processing.
- Images were labelled with several car damage classes to meet the damage detection goal. Bounding boxes were selected as the annotation type for various damage categories.

## 4.3 Evaluations

The evaluation section focuses on analysing the solutions that have been developed and determining how to use the outcomes. The models determine how the results are interpreted, allowing them to be assessed to see if they meet their goals. Section 3.5 focuses on the evaluation parameters and their formulas, which were used to decide the effectiveness of deep learning algorithms. Various parameters like train accuracy, validation accuracy, test accuracy, precision, f1-score, recall, support and confusion matrix are used for model evaluation.

The purpose of this study was to compare various deep learning models in order to categorise car damage categories. Convolutional Neural Networks (CNN), MobileNet, VGG16, and InceptionResnetV2 were the deep learning methods employed. The training dataset was divided into three parts: training, testing, and validation, in the proportion of 70:15:15. The dataset has a total of 6 classes.

### 4.3.1 RetinaNet

RetinaNet employs a feature pyramid network to recognise objects at several scales. To achieve damage detection objective, RetinaNet is used for implementing damage detection. Figure 5 shows number of trainable as well as non-trainable parameters. Damage detection was implemented using the RetinaNet model. The model has been trained over 20 epochs with 36,486,682 parameters in total.

```
========================================
Total params: 36,486,682
Trainable params: 36,380,442
Non-trainable params: 106,240
________________________________________
```

**Figure 5: Total trained parameters**

Figure 6 depicts result of damage detection class with probability of success. From figure, we can say that RetinaNet performing well as it detected scratch and head lamp damage class perfectly. The figure shows a sample of two categories, head-lamp and scratch were taken to test the RetinaNet model's performance. Model achieved an accuracy of 92% with the average time 0.46 sec.

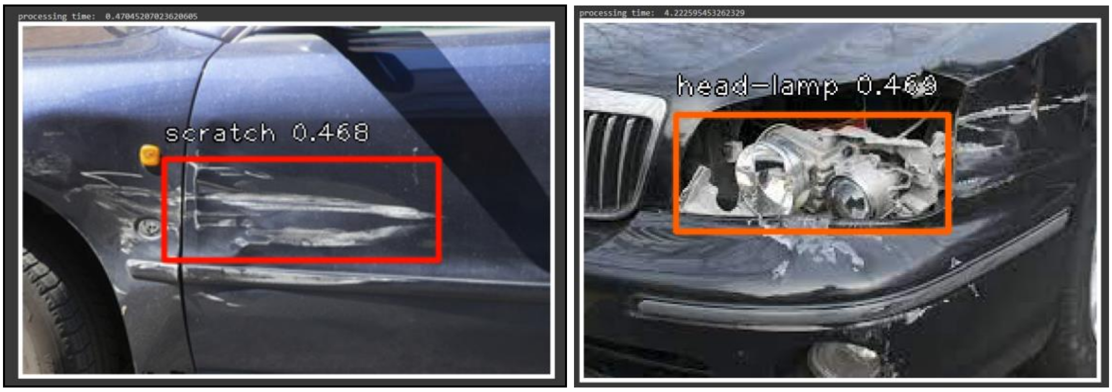


**Figure 6: Damage Detection Result**

### 4.3.2 CNN

A convolutional neural network comprising of 32 feature maps, as well as a (3*3) tiny filter were used to create the CNN model for categorization. The layer was then followed by a 2x2 scaled pooling layer, in which maximum pooling has been used to minimise testing without sacrificing information, preventing over-fitting. A dropout layer with a value of 0.5 is utilised. Further on, this model design includes a flattening layer, that ensures that the features are mapped into something like a single-dimensional matrix in order to gain superior features, which are then provided as feed to CNN. Figure 7 shows Model Accuracy & Model Loss of CNN.

The CNN classifier is trained with 40 epochs on the training sample and produced a validation accuracy of around 35 percent, while testing data yielded an accuracy of around 36 percent and training accuracy of 47 percent. However, the acquired findings are unsatisfactory since the chart doesn't really depict optimal training accuracy.

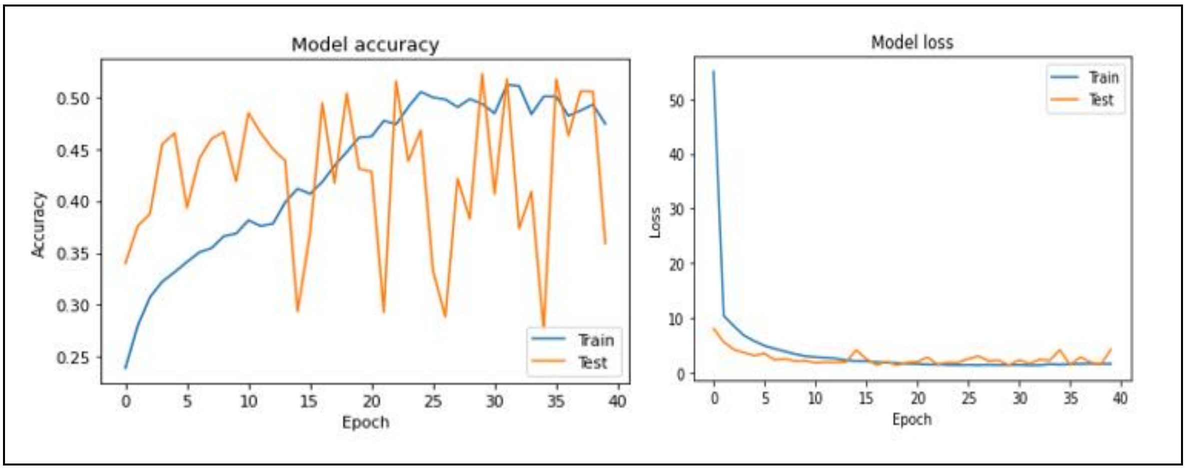


**Figure 7: Model Accuracy & Model Loss of CNN**

Figure 8 represents classification results of CNN model, where scratch damage is classified correctly but another scratch is classified as headlamp damage that is inaccurate.

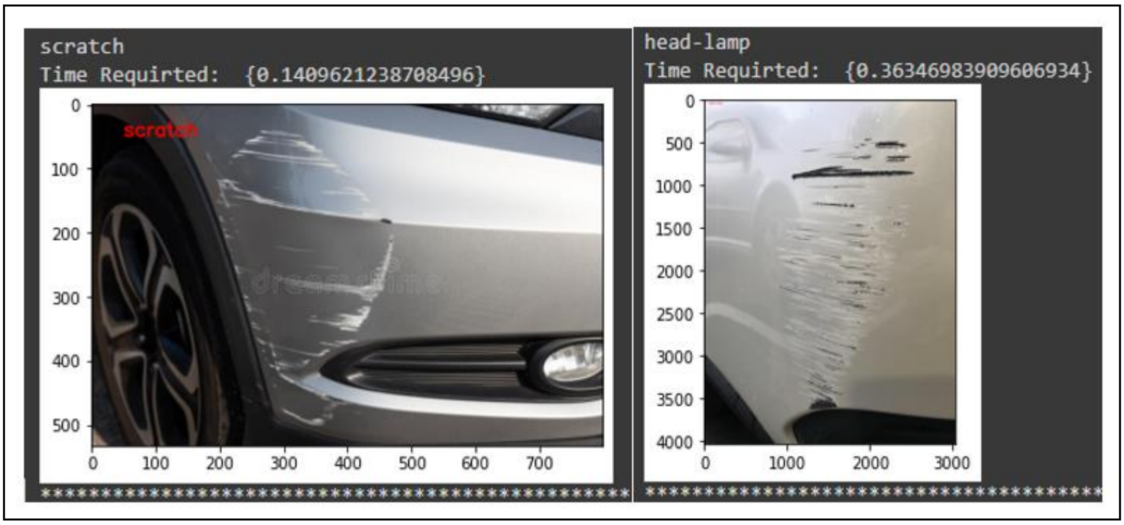


**Figure 8: Classification results of CNN Model**

### 4.3.3 MobileNet

Other model employed was MobileNet, that is a depth wise convolutional method that has been tweaked. The classifier was evaluated with ImageNet parameters; however, the upper layer was not taken into account. This classifier was trained with both a total of 20 epochs and a batch size of 40.

The MobileNet model's performance was deemed satisfactory, as evidenced by the figure 9, which show that both the network loss after each epoch and the model's precision are steadily growing. Accuracy rate of validation was 84 percent, whereas accuracy rate of training was 93 percent and evaluate the testing accuracy of precisely 84%.

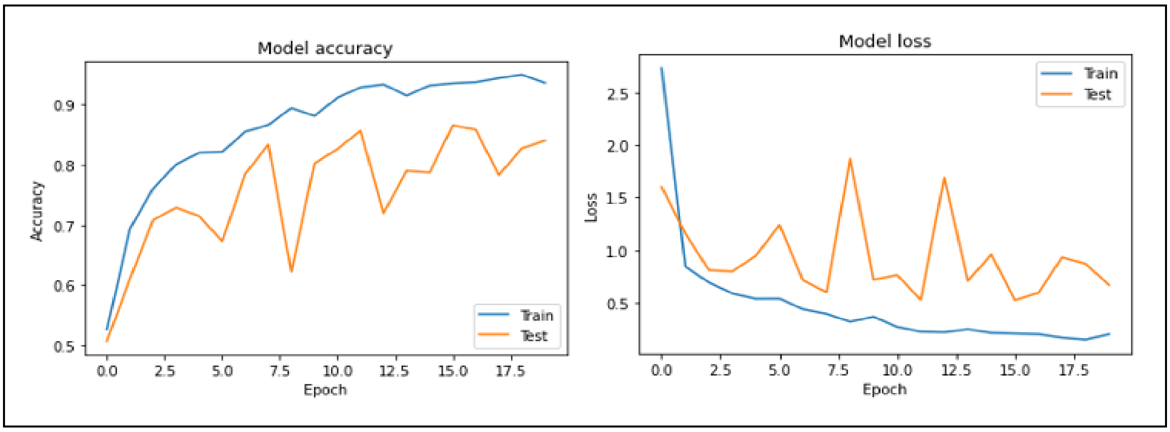


**Figure 9: Model Accuracy & Model Loss of MobileNet**

Figure 10 represents classification results of MobileNet model. As a result of better accuracy than CNN and VGG16 model, MobileNet model Classified both damages accurately.

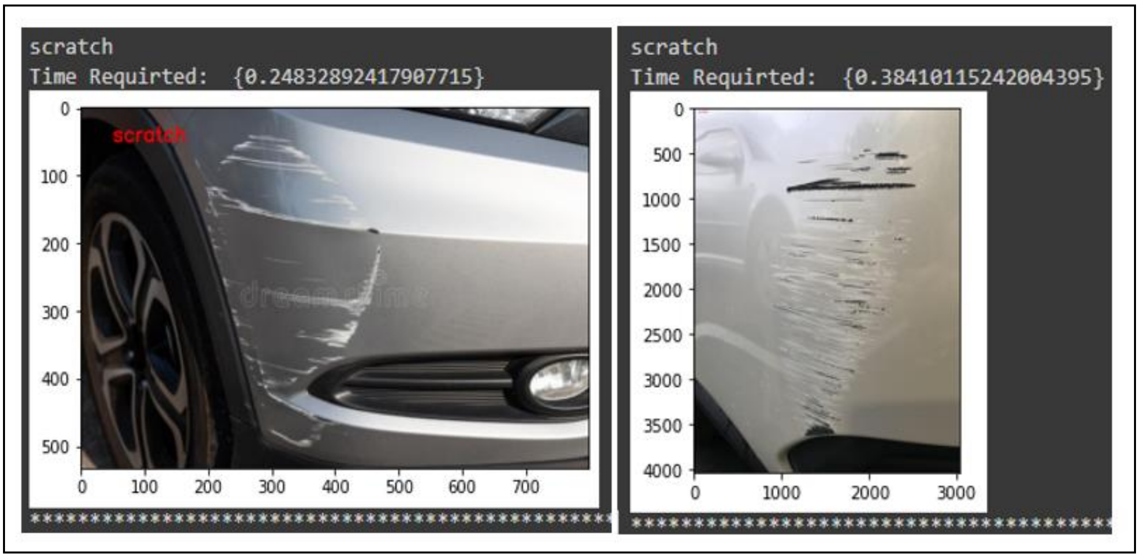


**Figure 10: Classification results of MobileNet Model**

### 4.3.4 VGG16

VGG16 was also one of those classification techniques that utilised the 'ImageNet' parameters. As the name implies, the architecture was made up of 16 layers. This classifier was trained with both a total of 20 epochs and a batch size of 50. “categorical_crossentropy” is used as a loss function.

The VGG16 model's performance was deemed satisfactory than basic CNN but performed poor as compared to MobileNet, as evidenced by the figure 11, which show that both the network loss after each epoch and the model's precision are steadily growing. Accuracy rate of validation was 60 percent, whereas accuracy rate of training was 56 percent and evaluate the testing accuracy of precisely 61%.

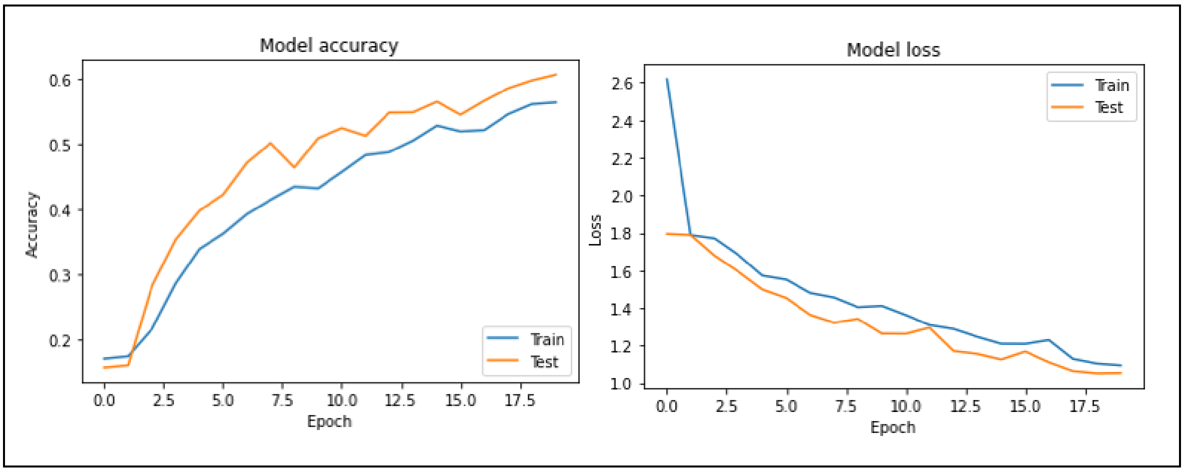


**Figure 11: Model Accuracy & Model Loss of VGG16**

As the VGG16 model's performance was deemed satisfactory than basic CNN but performed poor as compared to MobileNet. Classification results shown in figure 12 represents same, i.e., second image of door dent is classified correctly but first image of scratch is classified as door dent.


**Figure 12: Classification results of VGG16 Model**

### 4.3.5 InceptionResnetV2

The InceptionResnetV2 is combination of pretrained Inception and Resnet model, which was again trained on imagenet dataset. This classifier was trained with both a total of 20 epochs and a batch size of 50. “categorical_crossentropy” is used as a loss function. Following are the number of parameters used while model training:

- Total parameters: 64,168,934
- Trainable parameters: 64,108,390
- Non-trainable parameters: 60,544

The inceptionResnetV2 model's performance was satisfactory than basic CNN and VGG16 as well as performed approximately similar as compared to MobileNet, as evidenced by the figure 13, which show that both the network loss after each epoch and the model's precision are steadily growing. Accuracy rate of validation was 83 percent, whereas accuracy rate of training was 95 percent and evaluate the testing accuracy of precisely 84%.

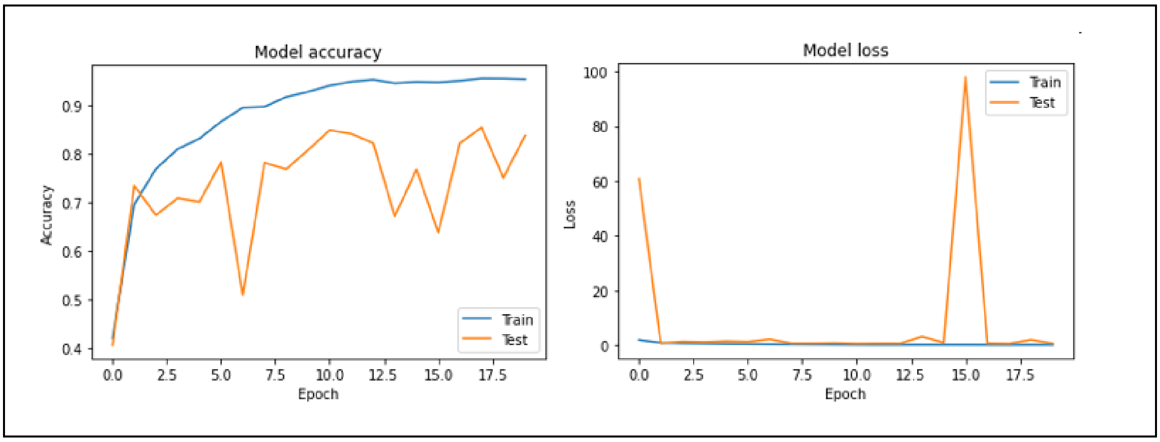


**Figure 13: Model Accuracy & Model Loss of inceptionResnetV2**

Figure 14 represents classification results of inceptionResnetV2 model. As a result of better accuracy than CNN and VGG16 model, MobileNet and inceptionResnetV2 shown approximately similar performance.

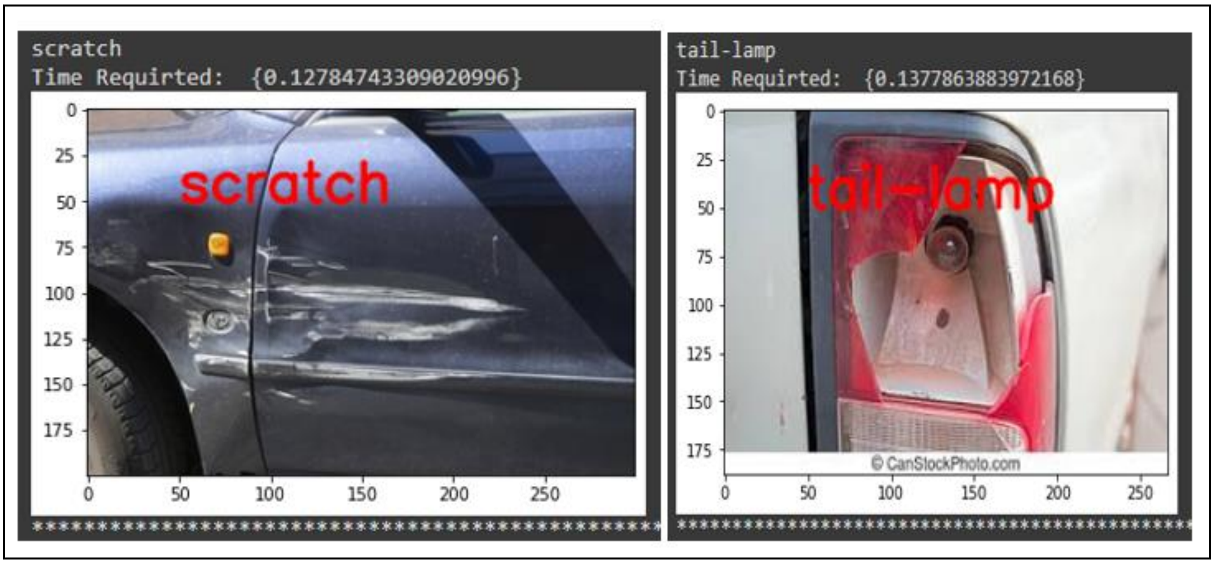


**Figure 14: Classification results of inceptionResnetV2 Model**

Hyperparameter tuning was used to get the maximum accuracy for car damage detection models. Batch size, epochs, learning rate, optimizer, dropout layer, and other parameters were tweaked. As a result, these models' configurations were optimized for the parameters that provided the best accuracy.

Following the completion of the damage detection model training, the models then evaluated using several evaluation matrices. Precision, confusion matrix, recall, support, f-1 score, and model testing, validation and training accuracy are examples of matrices. These matrices provide a clear view of both the model's performance and predictive capacity.

## 4.4 Results

Table 1 depicts accuracies and used parameters of all trained models.

**Table 1: Accuracy and parameters of all trained models**

| Trained Model | Training Accuracy | Validation Accuracy | Testing Accuracy | Batch Size | Epochs |
|---|---|---|---|---|---|
| CNN | 47.41% | 35.93% | **36%** | 50 | 40 |
| VGG16 | 56.44% | 60.70% | **61%** | 50 | 20 |
| MobileNet | 93.59% | 84% | **84%** | 50 | 20 |
| InceptionResnetV2 | 95.27% | 83.67% | **84%** | 50 | 20 |

Using the transfer learning technique, MobileNet and InceptionResNetV2 achieved the maximum accuracy on the testing, training, and validation sets with number of epochs as 20 and batch sizes of 50 Basic CNN, on the other hand, performed poorly on the all the three sets, with the epochs of 40 and batch size of 50. Except for the pretrained models, which had their epochs reduced to 20, all of the damage detection models were trained for the same batch size. To establish a baseline for accuracy, Basic CNN was trained on pre-processed dataset for 40 epochs.

# 5 Conclusion and Future Work

Based on thoroughly conducted research, all the research questions were addressed. The project's main goal was to identify and categorize different types of automotive damage using different deep learning models. Number of damage categories were also increased with respect to previous research. Detection algorithm named RetinaNet was used which successfully detected the damage with 92% accuracy with a time of 0.46 secs to detect. This concludes the first research question. Multiple pre-trained models were analysed to get best evaluating parameters which indent was achieved from MobileNet and InceptionResnet50 deep learning pre-trained models. Accuracy of 84% with classification time of 0.25 secs, was achieved on testing dataset. This concludes the second research question. Looking at the dataset with 3474 images, which intends to be very less for the research study. Implementing different techniques to increase the dataset size to 9000 images was a challenging task. Four different models were evaluated with ImageNet as pre-trained weights. Accuracy of 93% can be seen on training dataset by MobileNet and InceptionResnet50. Whereas basic the CNN model achieved just 40% accuracy. This concludes pre-trained models performed efficiently, accurately and precisely. This concludes the third research question.

In recent years, significant advances has been made in the fields of computer vision, particularly picture categorization. In these breakthroughs, Deep Convolutional Neural Nets were extensively employed, but not solely. Some of these advancements were featured in this project. These breakthroughs were utilised to determine the extent of damage as well as its location in autos. This and several other activities that rely on them are now done entirely by hand, even without the aid of machinery. The data provided in this research led to the conclusion that technology like those examined here, could be used to aid people in doing specific tasks.

The study described here just provides the groundwork for additional in-depth research in this area. A more accurate assessment of human being performance in some of these jobs might also be beneficial. What constitutes a satisfactory result on a categorization job is frequently determined by the challenge itself. As a result, maintaining a human being performance benchmark could be very helpful in determining how effective such models are. Although expanding the dataset size might potentially enhance the findings, this would contradict the primary goal of reducing data pre-processing effort. Also, annotating images with different annotations techniques will enhance model accuracy. Alternative model optimization techniques, such as the better tuning of hyperparameters, should be investigated in future studies.

# 6 References


Cha, Y.J., Choi, W. and Büyüköztürk, O., 2017. Deep learning-based crack damage detection using convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering, 32(5), pp.361-378.

Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramırez-Quintana, M.J. and Flach, P., 2019. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories.

Dhieb, N., Ghazzai, H., Besbes, H. and Massoud, Y., 2019, December. A very deep transfer learning model for vehicle damage detection and localization. In 2019 31st International Conference on Microelectronics (ICM) (pp. 158-161). IEEE.

Dwivedi, M., Malik, H.S., Omkar, S.N., Monis, E.B., Khanna, B., Samal, S.R., Tiwari, A. and Rathi, A., 2021. Deep learning-based car damage classification and detection. In Advances in Artificial Intelligence and Data Engineering (pp. 207-221). Springer, Singapore.

Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Gontscharov, S., Baumgärtel, H., Kneifel, A. and Krieger, K.L., 2014. Algorithm development for minor damage identification in vehicle bodies using adaptive sensor data processing. Procedia Technology, 15, pp.586-594.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z. and Qu, R., 2019. A survey of deep learning-based object detection. IEEE access, 7, pp.128837-128868.

Jayawardena, S., 2013. Image based automatic vehicle damage detection.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.

Li, H., Lin, Z., Shen, X., Brandt, J. and Hua, G., 2015. A convolutional neural network cascade for face detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5325-5334).

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T. and Omata, H., 2018. Road damage detection using deep neural networks with images captured through a smartphone. arXiv preprint arXiv:1801.09454.

Mohan, A. and Poobal, S., 2018. Crack detection using image processing: A critical review and analysis. Alexandria Engineering Journal, 57(2), pp.787-798.

Pathak, A.R., Pandey, M. and Rautaray, S., 2018. Application of deep learning for object detection. Procedia computer science, 132, pp.1706-1717.

Patil, K., Kulkarni, M., Sriraman, A. and Karande, S., 2017, December. Deep learning based car damage classification. In 2017 16th IEEE international conference on machine learning and applications (ICMLA) (pp. 50-54). IEEE.

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

Szegedy, C., Toshev, A. and Erhan, D., 2013. Deep neural networks for object detection.

Viola, P. and Jones, M., 2001, December. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). Ieee.

Wang, X., Yang, M., Zhu, S. and Lin, Y., 2013. Regionlets for generic object detection. In Proceedings of the IEEE international conference on computer vision (pp. 17-24).

Wu, X., Sahoo, D. and Hoi, S.C., 2020. Recent advances in deep learning for object detection. Neurocomputing, 396, pp.39-64.

Zeiler, M.D. and Fergus, R., 2014, September. Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

Zhang, Q., Chang, X. and Bian, S.B., 2020. Vehicle-damage-detection segmentation algorithm based on improved mask RCNN. IEEE Access, 8, pp.6997-7004.

Zhao, Z.Q., Zheng, P., Xu, S.T. and Wu, X., 2019. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 30(11), pp.3212-3232.