

CMPE 257 - Project Report

**Potholes Detection using Machine Learning**

**Submitted By:**

**Group 12**

**Deepali Upadhyay (011809694)**

**Mayank Dhingra(013773942)**

**Shivani Parate(013729612)**

**Under the guidance of:**

**Dr. Prof Jerry Gao**

**Table of Contents**

[**Chapter 1 Introduction**](#_yxyoa6jqerl1) **3**

[1.1 Overview of current scenario](#_4mh7q01i2siy) 3

[1.2 Objectives of our project](#_efbxaity4dai) 4

[1.3 Difference between our work and existing works](#_h192lsm7p097) 5

[1.4 Report Organization](#_brrxfgwbhswm) 5

[**Chapter 2 Related Work**](#_o7h5q3pwti6r) **6**

[**Chapter 3 Training and Test Data Preparation**](#_wb6phxakfa7c) **8**

[3.1 Dataset Statistics](#_5fmrfqbuqr8o) 8

[3.2 Preparing the training Dataset for model creation](#_lt91pj42z3vj) 8

[**Chapter 4 Machine Learning Models and AI Algorithms**](#_ea7ix3irhpv6) **9**

[4.1 Overview](#_nw3nv72gjb6n) 9

[4.2 Models](#_29wpodg1oq4t) 10

[4.2.1 You Only Look Once Algorithm](#_xk9nnnsinrpx) 10

[4.2.2 Single shot detector Algorithm](#_sixpxvoji9jc) 14

[4.2.3 Regional Convolutional Neural Networks](#_nb56xuqhgj5f) 15

[**Chapter 5 Implementation**](#_9u3st13iwjmj) **17**

[**Chapter 6 Results**](#_m2u7buq6u9ki) **17**

[**Chapter 7 Comparison of the models used**](#_rn2emjjpjpxh) **17**

[**Chapter 8 References**](#_resu3yej00xe) **18**

# Chapter 1 Introduction

## 1.1 Overview of current scenario

Potholes are a structural damage to the road with hollow which is usually caused by poor drainage system on the roads which allows water to go under the soil structure of the road. Due to unpredictable weather these days, the damage to the roads is increasing. According to the statistics, one-third of the accidents are caused by bad road conditions. According to American Automobile association, almost 16 million drivers have suffered some kind of damage due to potholes like tire puncture, bent wheels, damage to suspensions and shockers. The bicyclist face greatest dangers from these as they can easily lose balance and get hit by incoming traffic. So, it's really important to have a reliable pothole detection system to correctly identify the pothole which can later be used to warn the drivers.



*Fig: 1 Example of a pothole*

Current pothole detection methods fall in three important categories:

a) *Vibration Based Methods* - They use accelerometers to sense the pothole. The accelerometer senses the force on the sensor. The advantage of this method is low cost maintenance of the equipment and the data captured can be directly used to know the existence of pothole. The disadvantages are the low response time and the need to go through the pothole to get the readings.

b) *3D Reconstruction* - They use 3D laser scanners to create a 3D image of the pothole. The advantage of this methods is that potholes can be detected in real time and the data captured is quite accurate. The only disadvantage is the high cost of the equipment used which is a challenge.

c) *Vision Based Methods* - These methods use expensive equipment like cameras to capture the images in real time. The images are then processed using complex algorithms to get the desired data. The disadvantage of this method is high maintenance cost of the hardware. The advantage of this method is that characterization of the pothole can be done in addition to its detection. Also there is no need to go over the pothole to capture the data.

## 

## 1.2 Objectives of our project

We plan to develop a system based on vision which can detect the potholes with at least 85% accuracy. Additionally, we plan to measure the depth of the pothole. We also plan to connect our dataset with the google api to capture the latitude and longitude of the potholes which can later be used to know the location of the pothole on the map.

## 1.3 Difference between our work and existing works

Traditionally, the pothole detection project only deals with the identification of the pothole on the street. No studies have been performed to get the statistics of the pothole in terms of size and depth. We plan to include the statistics as well along with the detection of the pothole.

## 1.4 Report Organization

We have organized our report in eight chapters.

Chapter 1 deals with the objective and motivation of our project.

Chapter 2 summarizes the research work already done in this area.

Chapter 3 talks about the dataset used and the statistics of the data.

Chapter 4 summarizes the different machine learning models used to solve the problem. Also tells about the advantages and disadvantages of all the different methods.

Chapter 5 lists down the implementation method of the algorithm and the model. Also it tells about how the performance of the model was improved.

Chapter 6 shows the results achieved through each of the different models.

Chapter 7 compares the different models based on different performance factors.

Chapter 8 lists down the references used to create this project and the report.

# 

# Chapter 2 Related Work

There is a lot of research work done in this area. Below are some of the papers which we used to gather information on our topic of research.

There was a proposal to use SVM (Support Vector Machine) as in [2] , for pothole detection. The image region was extracted based on the histogram of the image and simple kernel SVM was used to locate the pothole. The target was well recognizable using this method.

As in paper [3] , CNN ( Convolutional Neural Network) based on deep learning is used to classify the potholes and cracks based on the images. They build a model which was not influenced by the noise due to incorrect illumination and shadows.

Hiroya Maeda, et al. [4] developed a system to detect the road damage using CNN methods on the images taken by phones. They gathered a huge dataset for pothole detection and applied deep learning algorithm to solve the problem and calculated its accuracy and the speed with which detection was made possible by the system.

Some other researchers have used binary classification [5] using deep neural networks to classify the road images whether they belong to normal road images or the ones with pothole. The features of the images need to feeded in the system before it can perform the classification.

A new neural model ‘Crack-net’ [6] was proposed for detecting the cracks on the road. The difference with the other neural models was that pooling layers are not included. This method was very efficient in detecting the cracks and uneven surfaces on the road.

In some experiments cheap sensors were used along with the deep CNN models like [7] for automatic crack detection. This papers presents a model which can automatically learn the feature without any feature extraction process.

Lastly, a real-time based pothole detection system for Android devices was proposed by A. Tedeschi & F. Benedetto [8].

# Chapter 3 Training and Test Data Preparation

## 3.1 Dataset Statistics

We found a dataset which was created by Electrical and Electronic Department, Stellenbosch University in 2015. The entire dataset consists of two different sets, one was considered to be simple and the other more complex. The dataset is collected by clicking pictures on smart phones by setting it up on the dashboard of a car.  
These datasets do share some files and there are a few instances where two different images would have the same name. Therefore, the appropriate measures need to be taken if the data is combined into one larger dataset.

Every folder contains 2 subfolders which contain the training and test data. Furthermore, the training data folder is divided into 2 more such subfolders namely positive data which contains the pictures of roads with potholes and negative data which consists of pictures of roads with no potholes.

## 3.2 Preparing the training Dataset for model creation

To create the training data, we need labelled images as labelled images contains the position and name of the object to be classified in the model. The images are labelled by creating a rectangular bounding box around the object manually on all the training images. Finding the exact position of these bounding box for all the training images could be tedious task. To overcome this, we’ll be using an Image labelling tool like labelme or labelImg. This tool makes labelling the potholes easier as an object can be labelled by just dragging a line across a pothole.

**Fig 2. Creation of bounding boxes using LabelImg tool**

Once all the images are labelled, a .xml is created for every image which contains the top-left and bottom-right coordinates of the bounding box. These coordinates are then fed to the model which will predict the potholes position.

# 

# Chapter 4 Machine Learning Models and AI Algorithms

## 4.1 Overview

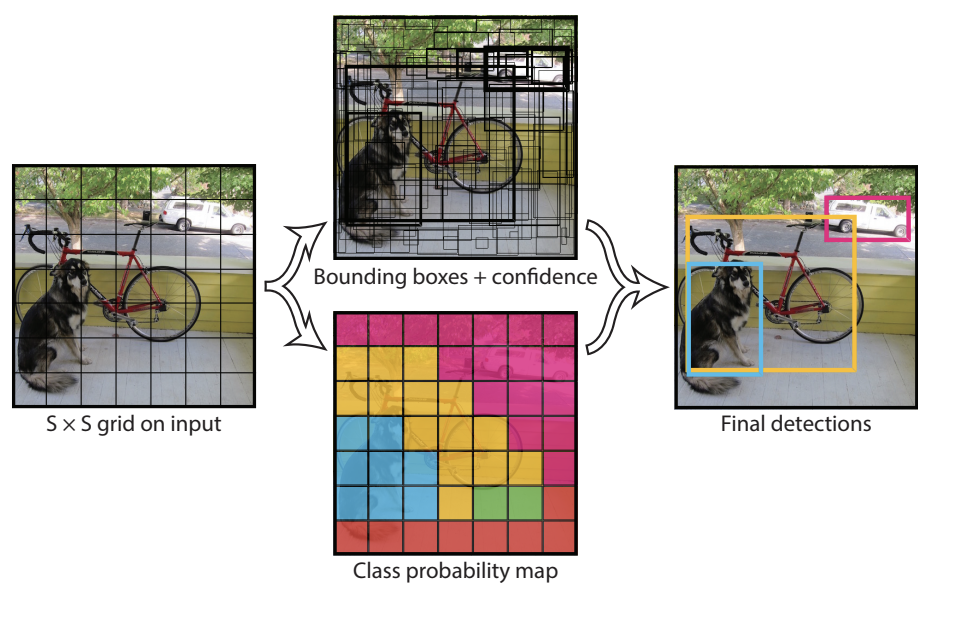
One of the most important research areas is Object recognition due to its wide range of applications. It is done by finding the coordinates of the bounding box which contains the object. These bounding boxes are the main reason why Object recognition is different from any other classification algorithm. Standard Convolutional Neural Networks can not be used to predict the coordinates of the bounding box due to the different length of last layer for various inputs.

This is one of the main reasons why specific object detection algorithms are made. We will be working on the following three approaches to develop our models.

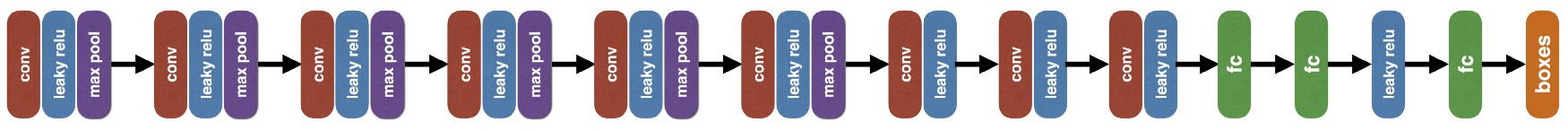
## 4.2 Models

## 4.2.1 You Only Look Once Algorithm

YOLO (You Only Look Once), is an object detection algorithm that uses just a single neural network for object detection. The object detection task is divided into two parts. The first part is determining the location of the objects present on the image and the second part is classifying that object. All the previous algorithms like R-CNN use a pipeline to perform these simple tasks in multiple steps which was slow to run and hard to optimize.

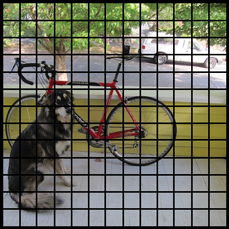


YOLO made it easier by using a single neural network for both classification and localising the object using bounding boxes to solve this problem. This is the architecture of YOLO :



**Fig 3. YOLO Architecture**

The end result is a tensor value of 13\*13\*30 where 13\*13 are the number of grids on the image and 30 is the size of a single grid in this case.



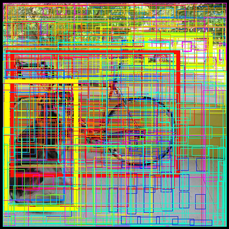
**Fig 4. Image divided into grids**

Each of these cells is responsible for predicting 5 rectangles that encloses an object. It also gives the confidence score that tells how certain is that the predicted object location of bounding box actually encloses some object. The image after predicting all the boundary boxes may look something like this.



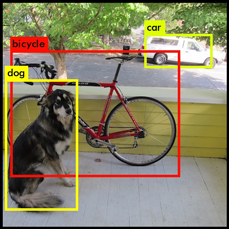
**Fig 5. Figure containing all the bounding boxes**

The confidence value can be determined by the thickness of the bounding box. This grids not only predicts the bounding boxes but also predicts a class. A final value is calculated by combining the class prediction and confidence score of the bounding boxes which gives us the probability of the presence of a class label in that bounding box. For example, the confidence score of the thick yellow bounding box is 85% that it contains the object “dog”



**Fig 6. Figure containing all the coloured bounding boxes for all the objects**

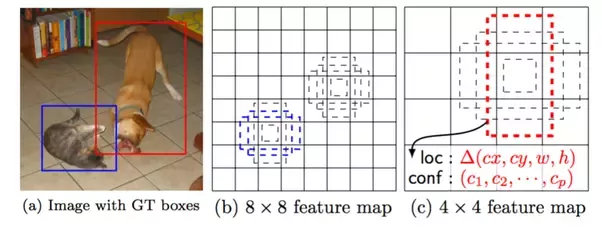
Out of all the 169 grid cells or 845 bounding boxes, we get only three bounding boxes if we set the threshold confidence score to 0.3 which means we only get those bounding boxes whose confidence score is greater than 0.3. The final bounding boxes are:



**Fig 7. Left bounding boxes after setting the threshold to 0.3**

## 4.2.2 Single shot detector Algorithm

This algorithm does the localization and classification in single forward pass and is similar to that of YOLO as it divides the image into grids of equal sizes



**Fig 8 . Single Shot detector**

This above image explains how Single Shot Detector algorithm (SSD) works. The main task of SSD is to match labels with default boxes of different aspects as dashed rectangles. A number is associated with every element of the feature map. Something is considered a match if the IOU value of any default box crosses 0.5. The 4 by 4 box is matched with the object dog and 8 by 8 box is matched with the object cat. Objects are identified with the help of 6 different feature maps.

The network outputs for each default box are following-

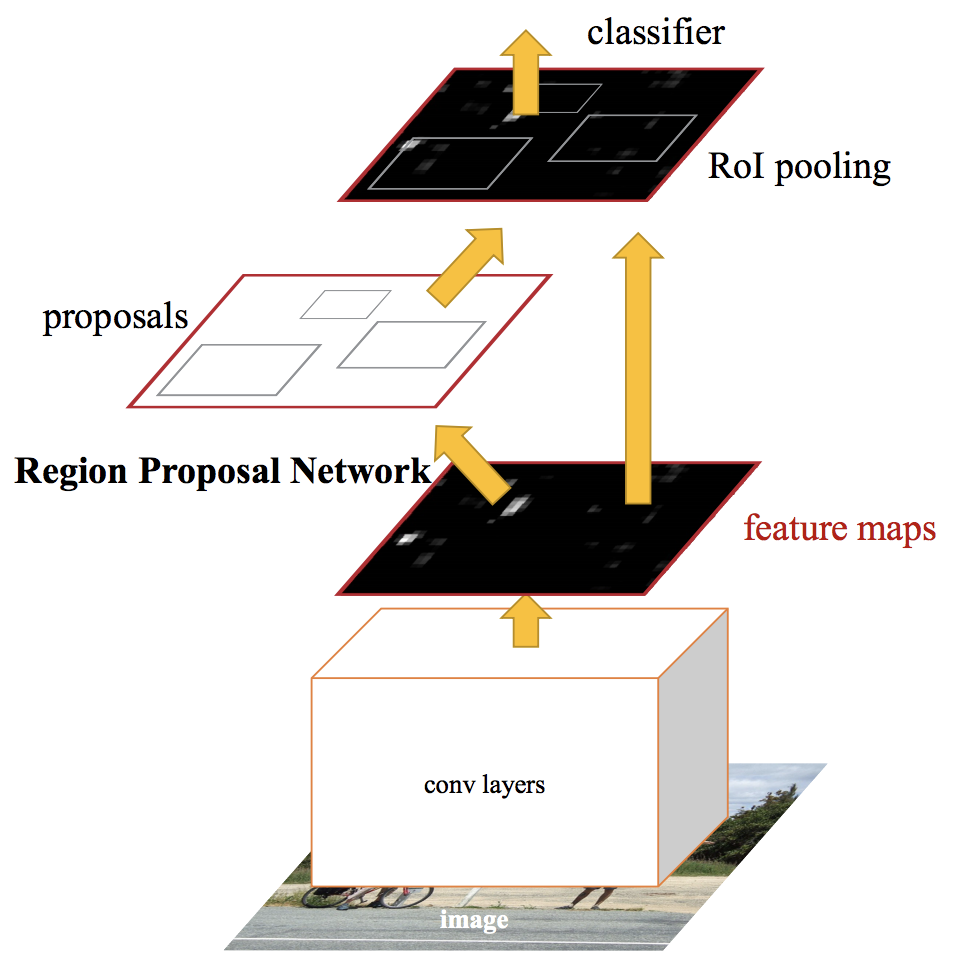
1. A probability vector of size equal to the number of classes which represents the probability of the presence of the object in the box.
2. An offset vector which contains the predicted offsets given in the format (cx, cy, w, h) – centre x, centre y, and width & height offsets, and are only meaningful if there actually is an object contained in the default box.

## 4.2.3 Regional Convolutional Neural Networks

Faster R-CNN (Regional Convolutional Neural Networks) consists of two parts which consists RPN (Region Proposal Network) for generating region proposals and another network for detecting labels using these proposals.

How it works-

* A feature map is obtained by running the image through a CNN
* Interesting boxes or regions are obtained by running the Activation Map through RPN
* These boxes use several fully connected layer to output class + Bounding Box coordinates.



**Fig 9. Faster R-CNN**

# 

# 