**Department of Electrical and Computer Engineering**

**North South University**





**Senior Design Project**

**Road Defect Detection Using Image Annotation**

**Md Sehabub Zaman Pranta 1611251042**

**Tanvir Ahmed 1410982042**

**Sadia Ahmed Tandra 1631474042**

**Jannatul Ferdouse Onny 1631561042**

**Rokeya Akanda Sriti 1620165042**

**Date:** 30.09.2020

**Course:** CSE499B Senior Design II

**Section:** 9

**Faculty Advisor:**

**Dr. Mohammad Ashrafuzzaman Khan**

Assistant Professor

Department of Electrical & Computer Engineering

Summer 2020

# Abstract

Avoiding defect roads is very crucial for driving in an accident-prone country like Bangladesh. Also, in order to use maintenance resources correctly and efficiently, roads need to be continuously monitored. With the help of deep learning, it is possible to solve this problem. Using deep learning, a road defect detection model can precisely detect defects of the roads and alert the concerned authority to repair or save a careless driver from possible danger.

In this research, we developed a road defect detection model using different frameworks, then analyzed and compared their result. We started with researching image annotation. Then we studied many research papers related to the topic, defined our problem, and implemented a solution. Our approach includes various annotation tools, object detection algorithms, instance segmentation, and different frameworks for implementation. Our final version of the road defect detection model adopts the Mask R-CNN algorithm. We trained and tested the model with our collected data in different backbones, parameters, and batches configuration. Finally, we compared and analyzed the results to reach a conclusion.

# Acknowledgement

First and foremost, acclaims and gratitude to God, the Almighty, for His showers of blessings all through our research work to effectively finish the research.

We would like to thank our honorable faculty Dr. Mohammad Ashrafuzzaman Khan, Assistant Professor, Department of Electrical & Computer Engineering, North South University, for his insight and guidance throughout the semester, which helped us to do this project. We want to thank our family: our parents, our brothers, and our sisters, to support us spiritually throughout our life.

Finally, our thanks go to all the people who have supported us to complete the research work directly or indirectly.

Table of Contents

[Abstract 2](#_Toc52393896)

[Acknowledgement 3](#_Toc52393897)

[List of Figures 6](#_Toc52393898)

[Chapter 1: Introduction 7](#_Toc52393899)

[1.1 Image Annotation 7](#_Toc52393900)

[1.2 Neural Network 7](#_Toc52393901)

[1.3 Transfer Learning 7](#_Toc52393902)

[1.4 Darknet 7](#_Toc52393903)

[1.5 Generative Adversarial Network: 8](#_Toc52393904)

[1.6 Dense-Net 8](#_Toc52393905)

[1.7 Region Of Belief (ROB) 8](#_Toc52393906)

[1.8 Structured Prediction 8](#_Toc52393907)

[1.9 Adaptive Thresholding 8](#_Toc52393908)

[1.10 Bilateral Filter 9](#_Toc52393909)

[1.11 Retina-Net 9](#_Toc52393910)

[1.12 SSD (Single Shot MultiBox Detector) 9](#_Toc52393911)

[1.13 Intelligent System 9](#_Toc52393912)

[1.14 Project Aim and Objective 10](#_Toc52393913)

[CHAPTER 2: RELATED WORK 11](#_Toc52393914)

[2.1 Road Damage Detection and Classification in Smartphone Captured Images Using Mask R-CNN……………………………………………………………………………………………………………………………………………..11](#_Toc52393915)

[2.2 A Deep Learning Approach for Road Damage Detection from Smartphone Images 12](#_Toc52393916)

[2.3 Generative adversarial network for road damage detection 13](#_Toc52393917)

[2.4 Transfer Learning-based Road Damage Detection for Multiple Countries 14](#_Toc52393918)

[2.5 Road Damage Detection and Classification with Faster R-CNN 15](#_Toc52393919)

[2.6 Deep Learning-Based Crack Detection Using Mask R-CNN Technique 16](#_Toc52393920)

[2.7 Asphalt Pavement Pothole Detection using Deep learning method based on YOLO Neural Network 16](#_Toc52393921)

[2.8 Detection and Classification of Road Damage Using R-CNN and Faster R-CNN: A Deep Learning Approach 17](#_Toc52393922)

[2.9 Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images.…………………………………………………………………………………………………………………………………………18](#_Toc52393923)

[2.10 Road Damage Detection and Classification Using Deep Neural Networks (YOLOv4) with Smartphone Images 18](#_Toc52393924)

[2.11 Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network 20](#_Toc52393925)

[2.12 Road Crack Detection Using Deep Convolutional Neural Network and Adaptive Thresholding………………………………………………………………………………………………………………………………..20](#_Toc52393926)

[2.13 Pothole Detection in Asphalt Pavement Images 21](#_Toc52393927)

[2.14 Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone 23](#_Toc52393928)

[2.15 An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment 23](#_Toc52393929)

[2.16 Road Damage Detection and Classification Using Mask R-CNN with DenseNet Backbone 24](#_Toc52393930)

[2.17 A Deep Learning Approach for Street Pothole Detection 25](#_Toc52393931)

[2.18 An Efficient and Reliable Coarse-to-fine Approach for Asphalt Pavement Crack Detection 25](#_Toc52393932)

[2.19 Machine learning algorithms application to road defects classification 26](#_Toc52393933)

[2.20 Automatic Crack Detection using Mask R-CNN 27](#_Toc52393934)

[CHAPTER 3: METHODOLOGY 28](#_Toc52393935)

[3.1 Workflow 28](#_Toc52393936)

[3.2 Dataset Preprocessing 28](#_Toc52393937)

[3.3 Neural Network Models 29](#_Toc52393938)

[3.3.1 YOLO v3 29](#_Toc52393939)

[3.3.1.1 YOLO v3 Training Details 30](#_Toc52393940)

[3.3.2 Mask RCNN 31](#_Toc52393941)

[3.3.2.1 Mask R-CNN Training Details 31](#_Toc52393942)

[CHAPTER 4: Results 32](#_Toc52393943)

[4.1 Result of YOLO v3 32](#_Toc52393944)

[4.2 Result of Mask R-CNN 33](#_Toc52393945)

[4.3 Comparative Analysis 34](#_Toc52393946)

[CHAPTER 5: Conclusions 34](#_Toc52393947)

[5.1 Discussion 34](#_Toc52393948)

[5.2 Summary 34](#_Toc52393949)

[5.3 Future Work 35](#_Toc52393950)

[CHAPTER 6: REFERENCES 36](#_Toc52393951)

# List of Figures

[Figure 1. Detection Results from trained Mask R-CNN Model 11](#_Toc52394142)

[Figure 2. A Deep Learning Approach for Road Damage Detection and Classification 12](#_Toc52394143)

[Figure 3. F-measure for each dataset (SSD MobileNet) 13](#_Toc52394144)

[Figure 4. F-measure for each dataset (SSD Resnet50) 13](#_Toc52394145)

[Figure 5. Predicted Labels for Sample Images 14](#_Toc52394146)

[Figure 6. Examples of training images with crack ground truth 16](#_Toc52394147)

[Figure 7. Loss of Yolo v3, Yolo v3 Tiny and Yolo v3 SPP in the modeling 17](#_Toc52394148)

[Figure 8. Detection and classification results for each class 18](#_Toc52394149)

[Figure 9. mAP and Loss vs Iterations (Tiny-YOLOv3) 19](#_Toc52394150)

[Figure 10. mAP and Loss vs Iterations (YOLOv3) 19](#_Toc52394151)

[Figure 11. Examples of structured prediction based on CNN 20](#_Toc52394152)

[Figure 12. Bilateral filtering and image segmentation; (a) original positive image; (b) filtered positive image; (c) segmentation result 21](#_Toc52394153)

[Figure 13. Pothole detection model 22](#_Toc52394154)

[Figure 14. Result of image segmentation for two potholes 22](#_Toc52394155)

[Figure 15. The proposed Mask R-CNN framework with DenseNet 24](#_Toc52394156)

[Figure 16. Comparison of accuracy of different models 25](#_Toc52394157)

[Figure 17. Illustration of process of lane and sign markings removal. (a) the original pavement image, (b) bin map, (c) variance map, (d)fused binary image, (e) further result from morphological processing, (f) final result 26](#_Toc52394158)

[Figure 18. Sample images from the annotated crack dataset 27](#_Toc52394159)

[Figure 19. Overall Workflow 28](#_Toc52394160)

[Figure 20. Annotating image with VGG 29](#_Toc52394161)

[Figure 21. YOLO v3 Network Architecture 30](#_Toc52394162)

[Figure 22. Transfer Learning for defect detection 30](#_Toc52394163)

[Figure 23. Mask R-CNN network architecture 31](#_Toc52394164)

[Figure 24. Inference result of YOLO v3 32](#_Toc52394165)

[Figure 25. Inference result of Mask R-CNN 33](#_Toc52394166)

[Figure 26. Evaluation result of the validation set 33](#_Toc52394167)

# Chapter 1: Introduction

## 1.1 Image Annotation

The image annotation is the process by which a computer system automatically assigns metadata in the form of caption or keywords to a digital image. Image annotation for deep learning is mainly done for object detection with more precision. There are many types of image annotation techniques. For example: Bounding Boxes, Polygonal Segmentation, Line Annotation, Landmark Annotation, 3D Cuboids, Semantic Segmentation, etc. Among these types, Semantic segmentation suits best for this project. It is a form of image annotation that involves separating an image into different regions, assigning a label to every pixel in an image.

## 1.2 Neural Network

A neural network is a series of algorithms that attempt to recognize the underlying relationships in a set of data through a process that mimics the way the human brain operates. The main problem with image classification is the difficulty of finding useful features. The manual handcraft of creating features from images like shapes, edges, regions is not an easy one even if there is significant progress in this field. However, a neural network, along with learning a model for classification is able to create and select automatically useful features.

## 1.3 Transfer Learning

Transfer learning is a research problem in machine learning that focuses on storing knowledge achieved while solving one problem and tries to apply this knowledge to a different but related problem. In this implementation, Transfer learning played a vital role. It is a machine learning method where a model has developed, could be re-used later. This approach is effective because it saves time and training cost-efficiently.

## 1.4 Darknet

Darknet is a neural network framework that is open source, written in C and CUDA. It features state-of-the-art real-time object detection algorithm YOLO ( You Only Look Once). Darknet can be used to classify images.

## 1.5 Generative Adversarial Network:

Generative modeling is an unsupervised learning task in machine learning that can imitate a given data. A Generative Adversarial Networks (GAN) is composed of two parts:

1. Generator: It generates new data.
2. Discriminator: It discriminates fake from real data.

## 1.6 Dense-Net

DenseNet is a modern CNN architecture that has achieved State-Of-The-Art (SOTA) results in fewer data classification (CIFAR, SVHN, ImageNet). It can be more depth and also quickly configured due to the recent use of residuals.

## 1.7 Region Of Belief (ROB)

A novel ROB extraction approach is proposed with a Region of Belief (ROB) definition identified for a hypothesis crack region with several reputation factors, followed by a new region increasing algorithm. Starting from a ROB as a seed, designating the search scope with a supposed code, and then searching and merging a ROB with different neighboring regions according to similarity criteria that implement numerous features to ensure the completeness of the cracks identified.

## 1.8 Structured Prediction

Structured prediction is a framework for solving classification or regression problems in which the output variables are mutually dependent or constrained. These dependencies and constraints reflect sequential, spatial, or combinatorial structure in the problem domain, and capturing such interactions is often as important as capturing input-output dependencies.

## 1.9 Adaptive Thresholding

Adaptive thresholding is the method where the threshold value is calculated for smaller regions, and therefore, there will be different threshold values for different regions. Thresholding is the simplest method of segmenting images. The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity is less than some fixed constant.

## 1.10 Bilateral Filter

A bilateral filter is a non-linear, edge-preserving, and noise-reducing smoothing filter for images. It replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels.

## 1.11 Retina-Net

RetinaNet uses a feature pyramid network to detect objects at multiple scales efficiently and introduces a new loss, the Focal loss function, to alleviate the severe foreground-background class imbalance problem.

## 1.12 SSD (Single Shot MultiBox Detector)

SSD (Liu et al., 2016) is an object detection framework that uses a single feed-forward convolutional network to directly predict classes and anchor offsets without requiring a second stage per-proposal classification operation. This framework's key feature is using multi-scale convolutional bounding box outputs attached to multiple feature maps at the top of the network.

## 1.13 Intelligent System

An intelligent system is a machine with an embedded, Internet-connected computer that can gather and analyze data and communicate with other systems. In modern days, the most needed thing is to enhance a model’s performance, in this proposed method preferred to use a pre-trained model for training. To build an intelligent system, it is essential to know how to use ML properly, update it rapidly, and deal with mistakes that make it smart enough to solve a problem by itself.

## 1.14 Project Aim and Objective

This project aims to implement an algorithm that can identify the defect from the road images. Our research has found that deep learning algorithms can be used to solve this problem. So, we have set our objective to develop a road defect detection model using deep learning.

To achieve this goal, we went through some steps, which includes -

1. Researching Image annotation technique, tools, algorithms.
2. Reviewing related works.
3. Studying various detection algorithms.
4. Collecting road images.
5. Working in different frameworks.
6. Implementing state-of-the-art object detection algorithm YOLO and Mask R-CNN
7. Experimenting with different backbones, parameters, and batches of the algorithm.
8. Analyzing and comparing the result to reach a conclusion.

All these steps are described in details with the results in this research.

# CHAPTER 2: RELATED WORK

## 2.1 Road Damage Detection and Classification in Smartphone Captured Images Using Mask R-CNN

The experiment in this paper was done as a part of the Road Damage Detection and Classification Challenge, 2018 IEEE International Conference on Big Data Cup. The authors used instance detection based on convolutional neural network and classification methods to solve the problem, which is Mask R-CNN. Not only was Mask R-CNN fast, but also it gave them significant results. They used an NVIDIA GeForce 1080Ti and achieved a mean F1 score of 0.528 at an IoU of 50% to detect and classify different types of road damage, where the images were captured using smartphones.

The implementation of Mask R-CNN was done using the Feature Pyramid Network and ResNet101 as the backbone of the network, and weights from training on the MS-COCO dataset were used here. Images were resized to 512×512 pixels, and augmentation was performed by horizontally flipping the images.

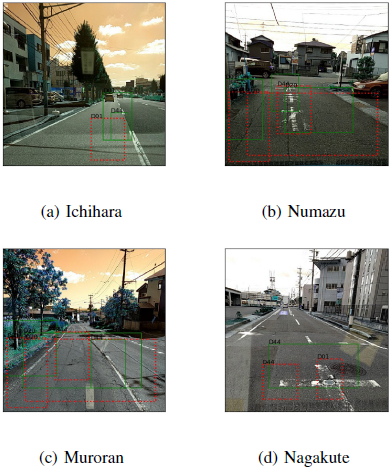


Figure 1. Detection Results from trained Mask R-CNN Model

The problem of the paper was that bounding boxes detected the damaged regions, but no exact shape of the regions was identified. In other words, they didn’t approach the instance segmentation method, which can precisely show the damaged regions. Bounding box annotation is faster in YOLO over Mask R-CNN.

## 2.2 A Deep Learning Approach for Road Damage Detection from Smartphone Images

This paper This paper describes the solution to road damage detection and classification challenge, IEEE Big Data Cup Challenge 2018. They used an object detection algorithm to train and detect damaged roads on different road damage types defined by Japan Road Association. The evaluation was done on various trained models and achieved an F1 score up to 0.62.

YOLO v3 with darknet43 backbone was used here as the object detection algorithm. The authors classified the road damages into eight types, generalized mainly into two categories: the first category is crack, and the second one is corruption. Python Augmentor was used to create synthesized images to solve the problem of class imbalance.

The dataset contained 7231 training images and 1813 testing images. All of the training images were annotated with one or more ground truth boxes corresponding to the eight types of road damages.

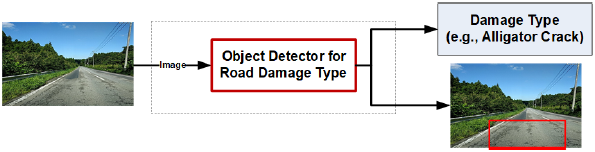


Figure 2. A Deep Learning Approach for Road Damage Detection and Classification

Although YOLO is fast, it uses bounding box annotation, which doesn’t detect the exact shape or detail information of the damaged regions. Also, the imbalance of different classes is a major problem here.

## 2.3 Generative adversarial network for road damage detection

In this paper, a progressive growing generative adversarial network with Poisson blending for artificially generating road damage images in improving performance was proposed. The paper claimed if the number of original images is small, then using that method F1 score can be improved by 5% and 2% for relatively large sample numbers.

Also, the authors updated the Road Damage Dataset 2018 (Maeda et al., 2018) to the Road Damage Dataset 2019 and made it available publicly. They show that this study improves pothole detection accuracy.

This study shows that their PG-GAN with Poisson blending can improve the F1 score on detecting pothole on roads.

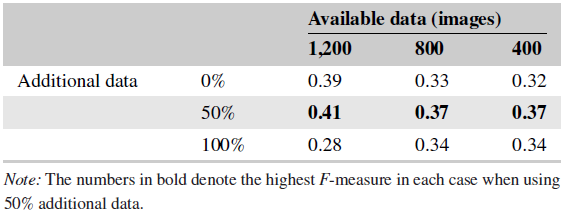


Figure 3. F-measure for each dataset (SSD MobileNet)

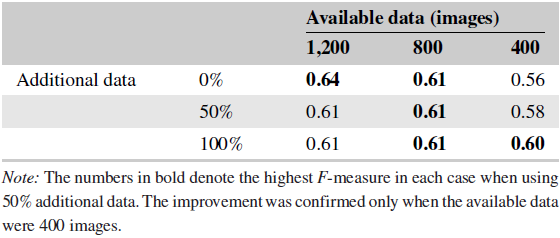


Figure 4. F-measure for each dataset (SSD Resnet50)

Generative Adversarial Network has some drawbacks as well.

1. It’s harder to train. User needs to give various kinds of data uninterruptedly to verify if it is working accurately or not.
2. Optimizing the loss function is very difficult.

## 2.4 Transfer Learning-based Road Damage Detection for Multiple Countries

This paper contributes in using the Japanese road damage detection model in other countries with transfer learning. The authors also provide a large road damage dataset of 26620 images. They introduced a generalized model for detecting and classifying damaged roads for many countries.

At first, they created a localized damaged road dataset by adding 3590 images from roads of Czech and 9892 images from roads of India together, which are captured using smartphones. Next, they labeled those images for crack and pothole and then mixed them with the Japanese road dataset. After that, the dataset was trained and evaluated on sixteen different deep neural network models in thirty scenarios on various combinations of the train and test datasets. Recommendations for other countries were provided based on the result from the evaluation.

SSD MobileNet is adopted in this paper as it is small, has low latency, and uses fewer resources. Using its convolutional feature extractor task of classification, detection and segmentation can be performed.



Figure 5. Predicted Labels for Sample Images

This study also uses bounding box annotation, which doesn’t give the precise shape or details of the damaged road.

## 2.5 Road Damage Detection and Classification with Faster R-CNN

The work on this paper was done as a solution to Road Damage Detection and Classification Challenge, 2018 IEEE International Conference on Big Data Cup. Their method can detect damages in road captured by smartphones. Faster R-CNN method and data augmentation techniques were used here. The detection model achieved an F1 score of 0.6255.

Fast R-CNN frame was used here as the detection method. Initially, images were processed by a feature extractor. The Region Proposal Network uses those feature maps as input and gives a group of rectangular object regions, including their scores. Then, from the feature maps, the Region of Interest (RoI) pooling layer extracts a fixed-length feature vector. Every feature vector is fed into a Fully Convolutional Network for classification and predicting bounding boxes.

For detection, Intersection over Union threshold was set to 0.5. For class imbalance issues, images were augmented by adjusting contrast, brightness, and Gaussian blur. Also, every image was Horizontal flipped during training. ResNet-152 was used as the backbone of Faster R-CNN and implemented on TensorFlow on a Linux PC with graphics card Nvidia GTX 1080Ti.

In the experiment, optimized parameters ResNet-152 gives F1 score of 0.6255, where ResNet-101 gave 0.6099.

The authors wanted to try other methods such as cascaded detection, multiscale inference, model ensembling for performance improvement. Also, state-of-the-art for instance segmentation, Mask R-CNN, can be used here for the exact shape of the damaged regions.

## 2.6 Deep Learning-Based Crack Detection Using Mask R-CNN Technique

This paper has also worked on the mask R-CNN with backbone Resnet-101 as the proposed method. This work got the data set of 352 crack images and divided the training, validation, and testing data. Here, the paper tried to build an automatic crack detector using the state-of-art technique. In the article, the obtained weights were from the pre-training model of the MSCOCO data set. The crack detection method in this paper was able to suppress noise and give an excellent result in real-time on-site.

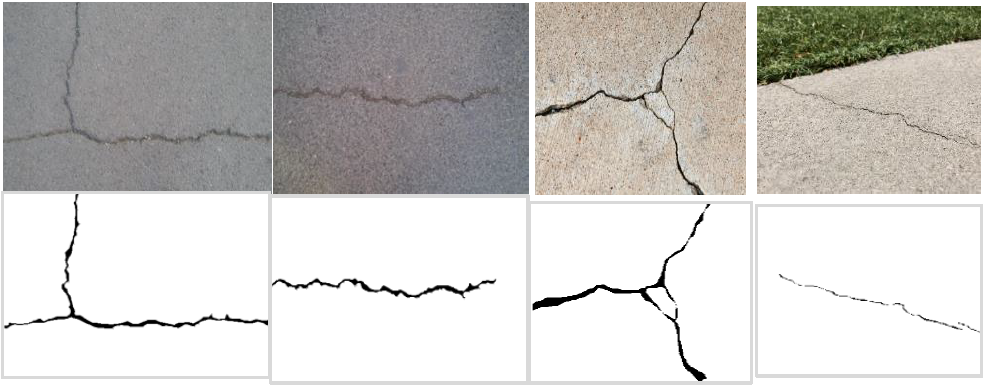


Figure 6. Examples of training images with crack ground truth

This method is used for crack detection, whether it's cracked or not, but our proposed method is to find different road damage types. But in the case of large defect detection, this approach may not work.

## 2.7 Asphalt Pavement Pothole Detection using Deep learning method based on YOLO Neural Network

There are three different architecture configurations in this article Yolo: YOLOv3, YOLOv3 tiny, and YOLOv3 SPP. In this paper, the pothole detection method was used on these three architectures. Firstly, research the data on, secondly, annotate and label the data, and build a model using Yolo V3 architecture. Finally, with the modeling phase's output on weight, the detection and area measurement with testing data was done. The modeling process was done with 10,000 iterations. The lost data of YOLOv3 tiny was higher than YOLOv3 and YOLOv3 SPP, but the lost data of YOLOv3 and YOLOv3 SPP was quite the same. By using spatial pyramid pooling in Yolo V3 architecture, the mean average precision(mAP) increased by 5.5%, which provided Yolo V3 SPP the best mAP in this experiment.

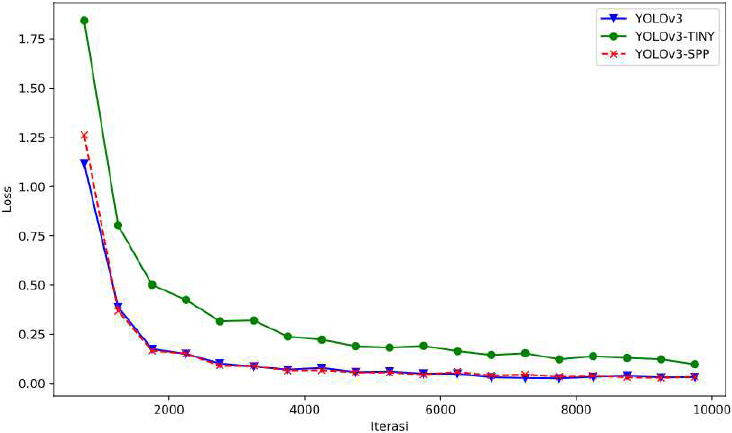


Figure 7. Loss of Yolo v3, Yolo v3 Tiny and Yolo v3 SPP in the modeling

In this article, the central concept was to detect potholes with different types of YOLOv3 architecture. As always, YOLOv3 gives a satisfactory result only with the potholes in the images where other types of damages exist. Here the approach of this article is different from our solution statement. Our solution is to find different types of damages, not only with the potholes.

## 2.8 Detection and Classification of Road Damage Using R-CNN and Faster R-CNN: A Deep Learning Approach

As monitoring manually in cities is time-consuming and lots of labor work, this paper proposed a model using R CNN and faster R-CNN to identify the road damage. In this model, 1100 images are resized and leveled with potholes, crack, etc. ran through R-CNN and Faster R-CNN for training. Here R-CNN built a massive integrated RPN and faster R-CNN with mutual convolutional feature levels or feature maps. ReLU activation function is implied between the convolutional and pooling layers. Adam Optimizer was used to minimize the loss. In this paper, Faster R-CNN worked better than R-CNN with an accuracy of 98.02% and validation accuracy of 99.80%, where R-CNN got the accuracy and validation accuracy of 71.44% and 76.01%.

In our proposed method, the use of mask R-CNN that extents faster R-CNN for pixel-level segmentation. It is quite obvious that mask R-CNN is the better version of Faster R-CNN and it also serves the purpose of our proposed method by doing instance segmentation.

## 2.9 Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images

The research was on damage detection or automatic road surface inspection with a large scale of 9053 damaged road images and 15,435 instances of road surface damage. In this article, the data set was accessible by the public for privacy matters. SSD's setup uses Inception V2 and SSD using mobile-net, where the initial learning rate is 0.003 with a learning rate decay of 0.95 every 10000 iterations. Different types of the class name of the damages like D00 D01, D10, D11, D20, D40, D43, D44, and the accuracy were different from each other.

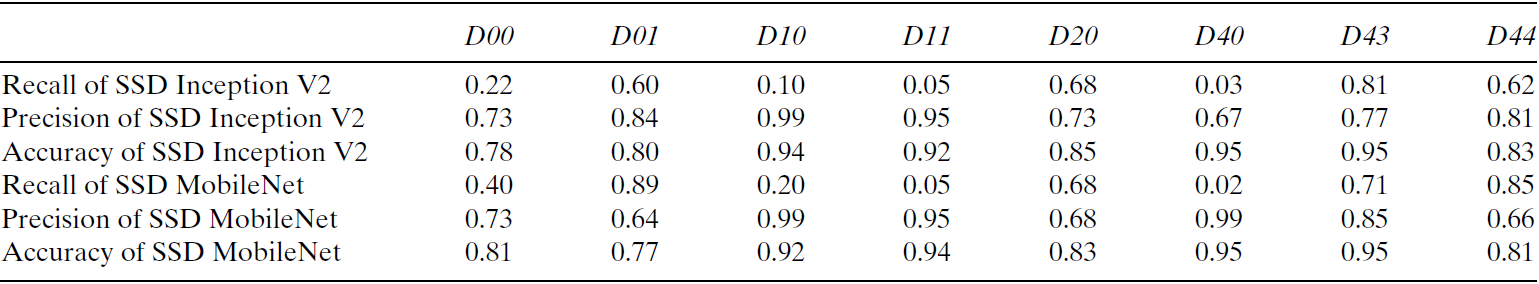


Figure 8. Detection and classification results for each class

* Here, MobileNet is a neural network used for classification and recognition, and SSD is a framework to detect some object. Both of them work for object detection.

Though the work is excellent, where the result returns the bounding box, and the classes are too limited because there might be different types of damages with different shapes, this will not handle all of them with different instances. That is where this model lacks.

## 2.10 Road Damage Detection and Classification Using Deep Neural Networks (YOLOv4) with Smartphone Images

This paper worked on three versions of YOLO with the IEEE big data Cup challenge, where the most extensive road damage data set available. To train the model transfer learning technique was applied while using pre-trained weights. The data set was divided to 80% to the training set,10% to validation, and 10% image for testing to perform each version.

With version 2, YOLO performed very poorly even with the Adam optimizer. Here, the Confidence of object detection was only 0.25 or higher, and the threshold was set on 0.2 in this model. With version 3, the mean average precision was supposed to increase, but it decreased slightly. However, Tiny-YOLOv3, with the best weights achieved at 14,400 iterations, gives the highest map means it worked better than YOLOv3. Now, the YOLOv4 training was stopped because of the time restriction after 6400 iterations. Though increasing the network resolution might increase precision. Figures of YOLO V3 and V4 are below.

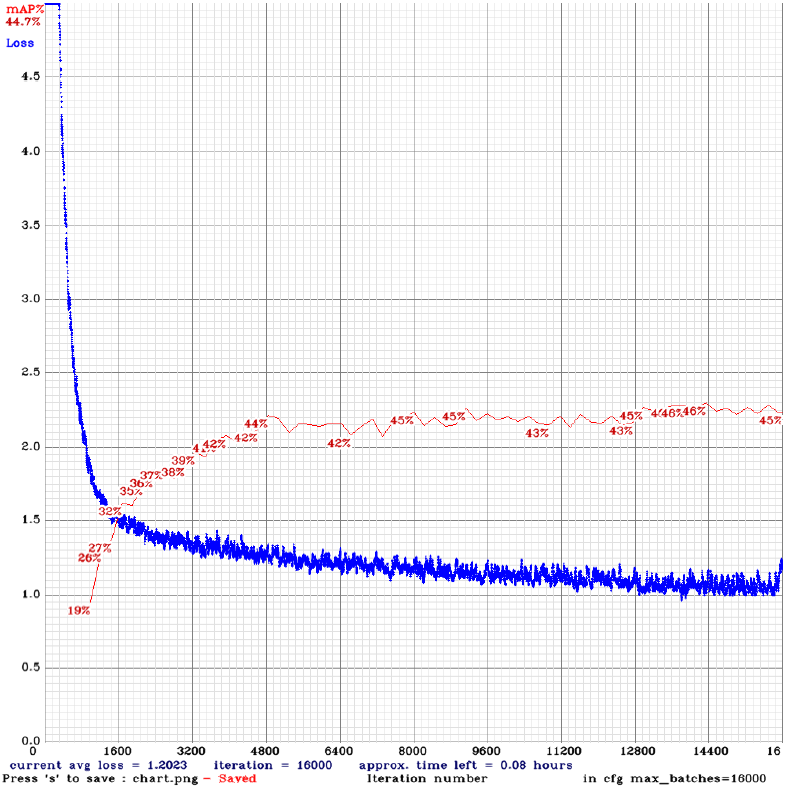


Figure 9. mAP and Loss vs Iterations (Tiny-YOLOv3)

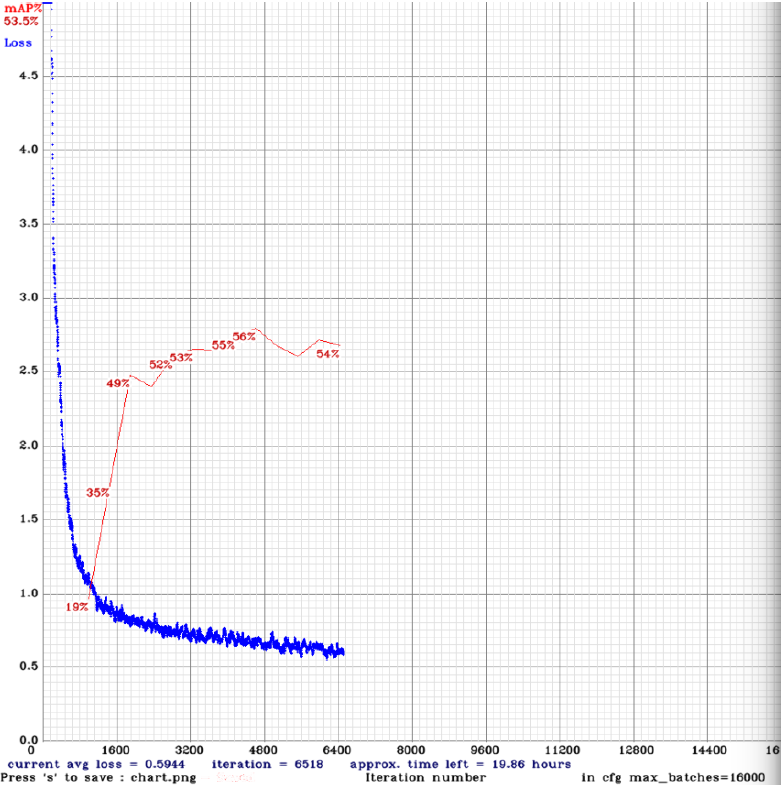


Figure 10. mAP and Loss vs Iterations (YOLOv3)

Here the problem is almost the same as mentioned before. The outcome might be great in this article, but the model's classes are too limited to predict when it comes to detecting the damages of the road, and some of the classes are not even needed for damage segmentations.

## 2.11 Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network

In this paper, proposed a method for pavement crack detection based on structured prediction with CNN. This method is trained and tested on CFD with RGB images and AigleRN with gray-level images and creates the network with CNN's help. Since CNN can extract useful features from raw data, this paper added a structured prediction based on the CNN method to learn a small patch's crack structure within an image to find the full crack on pixel level without preprocessing.

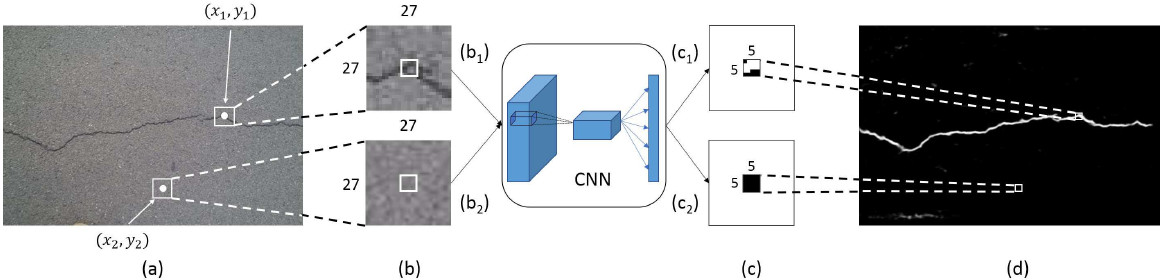


Figure 11. Examples of structured prediction based on CNN

This paper method can find only the cracks specifically and determine whether the crack has present or not. However, our proposed method is to find the different road damage types to find the crack's shape, so some improvement is needed to achieve our solution.

## 2.12 Road Crack Detection Using Deep Convolutional Neural Network and Adaptive Thresholding

This paper has worked by using CNN to Adapt thresholding to detect the road crack. The work has proceeded in two steps: image classification and the other is image segmentation. Here CNN is being used mainly for the image classification part as it works as a feature extractor and determines the crack existence. One term is added: ReLU, representing a rectified linear unit, which is the most popular activation function for deep neural networks.

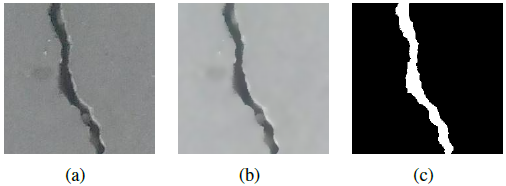


Figure 12. Bilateral filtering and image segmentation; (a) original positive image; (b) filtered positive image; (c) segmentation result

An adapting thresholding approach has been used for the segmentation process. In the segmentation part, before approaching the adapting thresholding, a bilateral filter has been used to smooth the input images. Here the primary technique of this work is adapting thresholding for segmentation. This thresholding method hypothesizes that the filtered image comprises two parts: foreground (cracks) and background (road surface). To find the best threshold δ, formulate the thresholding problem as a 2D. This 2D histogram using k means clustering and divided into four regions, and these four regions contain different vectors of foreground and background. Based on this, the 2D histogram thresholding formula can segment the crack images.

Here, this paper's method 2D histogram thresholding can only segment the crack-based area, but our proposed model can detect the damaged road and segment its damaged part. For the proposed model, things need to change in the 2D histogram part because its work segmented the crack area only.

## 2.13 Pothole Detection in Asphalt Pavement Images

This paper has worked on asphalt pavement images to detect pothole. The work is done in three steps: 1) image segmentation, 2) shape extraction, and 3) texture extraction and comparison. A histogram shape-based thresholding algorithm has been used to separate the defected region from the background by transforming original color images into gray-scale images in the segmentation part.

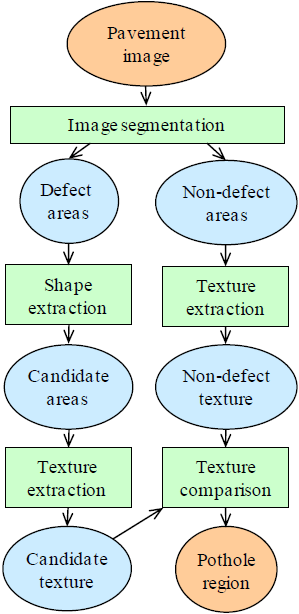


Figure 13. Pothole detection model

After segmented the defect and non-defect part, the shape extraction part has worked on the defected area, for measuring shade of the region here puts a threshold value and then shrinks the shade region to minimally connected skeleton morphological thinning is used. The standard deviation of gray-level intensity values as a statistical measure. it has been used to describe the texture of both the inside and the outside region. In third, it has worked for both defect and non-defect region and compares the texture.

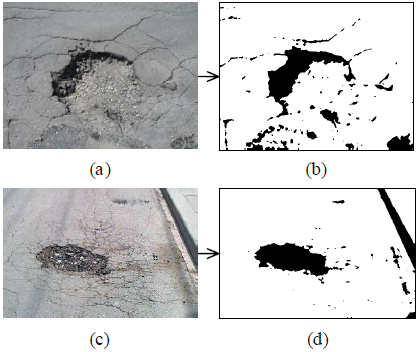


Figure 14. Result of image segmentation for two potholes

In this paper, the segmentation part is related to this work because it aims to segment the damaged area. However, this paper’s method has segmented both defects and non-defect areas, and our proposed method is to segment the only damage region. Here, the method of shape extraction and texture extraction can be useful by improving, which can be for better output.

## 2.14 Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone

Transfer learning is the use of the knowledge gained while solving one problem and applying it to a different but related problem. In this paper, we developed a new large-scale dataset for road damage detection and classification. This dataset is composed of 9,053 road damage images captured with a smartphone installed in a car. Damage roads have been annotated in the form of a bounding box. For object detection methods based on CNN, there are four types of systems used for object detection, like **Faster R-CNN, YOLO, R-FCN, and SSD**. In all these object detection systems, a convolutional feature extractor as a base network is applied to the input image to obtain high-level features. Six features extractors are darknet-19, VGG-16, Resnet, Inception V2, Inception Resnet, MobileNet.

Here this paper has mentioned some object detection systems useful for work. YOLO had chosen for work. YOLO can predict the region and class of objects with a single CNN in bounding box shape. However, for segmentation, YOLO is not the correct system and can focus on R-CNN. Some features extractors can be used from this like VGG, Resnet; these are relatable for our work.

## 2.15 An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment

Transfer learning is the use of the knowledge gained while solving one problem and applying it to a different but related problem. This paper introduced a new and very large asphalt dataset, containing damages that are not present in previous studies. For object detection, different types of models have been trained for mobile implementation. In this work, the RetinaNet object detector has been used. This system can detect different asphalt structural damages from video with high accuracy and low inference time. An additional advantage of RetinaNet is that it presents less jitter in the detection, owing to improved non‐maximum suppression strategies and better performance.

Here, RetinaNet is used mainly for smartphone implementation. It can detect the damaged object on the asphalt road, but our proposed method detects every road damage. It is here gaining some advantages of different types of feature extractors like VGG, Resnet. These can be used to improve further if work with a smartphone, then this RetinaNet will be useful.

## 2.16 Road Damage Detection and Classification Using Mask R-CNN with DenseNet Backbone

This paper proposed a new network named DenseNet, linked with convolution networks into the Mask R-CNN framework. This method provides additional features such as alleviating the problem of vanishing gradients, increasing the proliferation of features, and promoting reuse of features. In a feed-forward fashion, DenseNet attaches every layer of the convolutional layers to each layer. This network is a region proposal network for region proposal generation. In this DenseNet, three neural network headers are used for road damage recognition, bounding box idea enhancements, and street defect classifying. Road damage can also be segmented at the pixel level. This paper added the new method, and It detects road damages precisely and can also segment the road damage mask correctly.

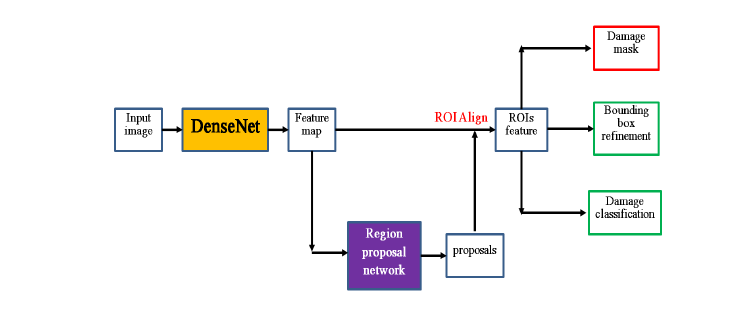


Figure 15. The proposed Mask R-CNN framework with DenseNet

This method has some errors in finding out cracks. Only the longitudinal linear crack damage is detected by this method. It is not feasible to work with goal project. The proposed project is not only to detect longitudinal linear crack damages. So, after analysis of this work, it needed more improvement for aim.

## 2.17 A Deep Learning Approach for Street Pothole Detection

In this paper, for street pothole detection author proposed a method based on deep learning. There are four models trained and checked with a pre-trained dataset like YOLO V3, SSD, HOG, SVM, and Faster R-CNN. The appropriate data is collected and then transformed the labeled image file to a train used by the models as an input. Images were trained and labeled by creating a rectangular bounding box around the item on all of the training photos using the LabelMe tool. Hyperparameters are calibrated, and the size estimation of potholes is considered for more precise detection outcomes.

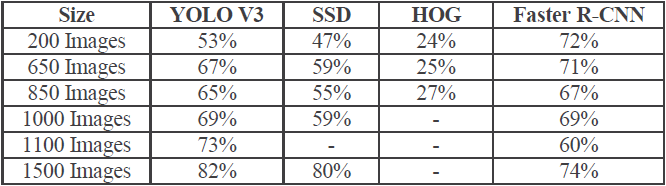


Figure 16. Comparison of accuracy of different models

There was a significant decrease in localization errors. The paper presented that YOLO V3 architecture has more speed than other models. YOLO V3 has super speed, and accuracy rates are very high. Nevertheless, there were some errors in the detection of small objects. This work needs more improvement for the aimed project.

## 2.18 An Efficient and Reliable Coarse-to-fine Approach for Asphalt Pavement Crack Detection

This research paper presented an exceptionally effective pavement crack detection system which is efficient and reliable. There are four prominent features, such as a new explanation of the cracks focused on the pixels with identical grey-level. The adaptive threshold approach for image segmentation considers the geographic variation, atmospheric conditions, geometric characteristics of cracks, a new concept termed Region of Belief (ROB), and crack detection, which begins with a ROB seed, a novel region growing algorithms proposed. For image segmentation, an enhanced adaptive thresholding algorithm is presented. In order to promote identification, a new design Region of Belief (ROB) has been implemented. A novel region growing algorithm was suggested for the identification of defects.

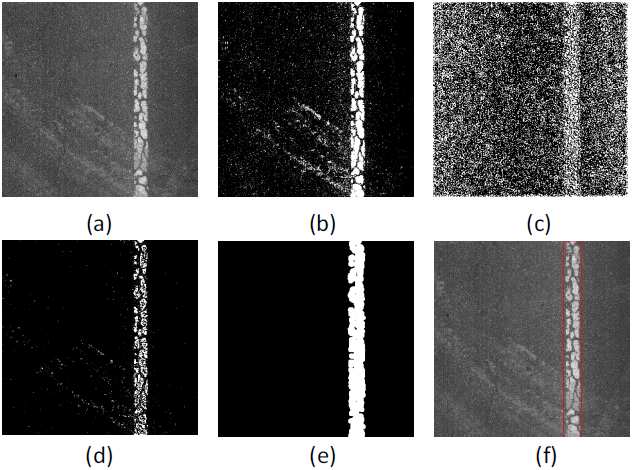


Figure 17. Illustration of process of lane and sign markings removal. (a) the original pavement image, (b) bin map, (c) variance map, (d)fused binary image, (e) further result from morphological processing, (f) final result

The CrackTree and VCrack algorithm were used in this paper. These algorithms detect cracks on the path. However, in our aimed project, it is needed to segment all damages of roads.

## 2.19 Machine learning algorithms application to road defects classification

This research paper indicated a new approach to automated identification and analysis of road defects based on machine research algorithms. The road defects are analyzed based on shape and texture feature analysis—the presented paper implemented on MATLAB with the Random Forest algorithm and boosting algorithm.

The boosting algorithm is being used to develop the classification models, and both sets have data showing the accuracy of the proposed system. These algorithms are required to identify road damages by following the Random Forest algorithm accurately. It is also recommended using the graph cutting method and Marcov algorithm that enhances image segmentation efficiency.

The author used the boosting algorithm, Marcov algorithm, graph cutting method, and Random Forest algorithm, but it was sensitive to noise during implementation and difficult to adjust. For our project, it seemed problematic and needed for more advancement.

## 2.20 Automatic Crack Detection using Mask R-CNN

This paper presented that R-CNN masks can identify cracks on concrete surfaces and get correlating masks to help isolate other properties useful for analysis. The paper's author developed a ground-truth mask database on images from a regular crack dataset for train Mask R-CNN for crack detection.

Several literature methods, such as detection of saliency, texture analysis, transforming wavelets, minimum direction tracking, and machine learning. This paper adapted Mask R-CNN to simplify crack detection on concrete surfaces to the current state-of-the-art detection model.

Multi-class analysis for other components and defects of infrastructure will minimize defects and provide efficient inspection means of civil structures. The results are an accuracy value of 93.94% and a recall of 77.5%.

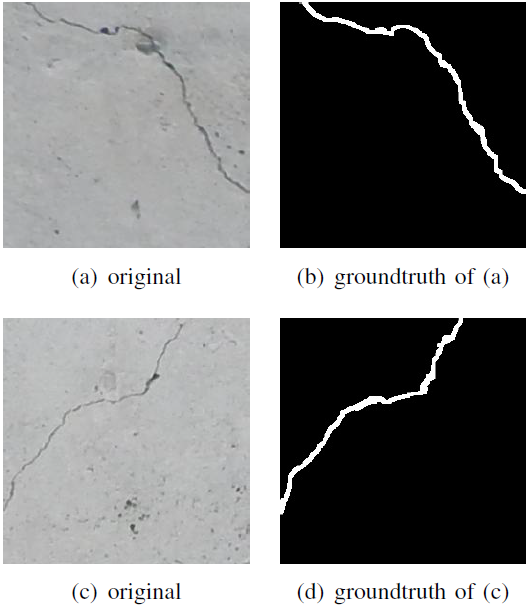


Figure 18. Sample images from the annotated crack dataset

This is only for crack detection in many things like whether it has been cracked or not, but the proposed method is to find different road damage types. So, things needed to change in order to do so.

# CHAPTER 3: METHODOLOGY

## 3.1 Workflow

Behind every step of the workflow, some research was conducted to find the optimal way to do the task. There are lots of methods and techniques available in this field now. So, it is very challenging to find the best one and implementation of the selected technique.

In this section, the comprehensive analysis will focus on the image annotation and instance segmentation models. One object detection and another instance segmentation algorithm have been used to solve the given problem. The problem statement has been solved by gradually moving from the YOLO algorithm to the Mask R-CNN algorithm. A complete workflow diagram of the solution is shown in Figure.

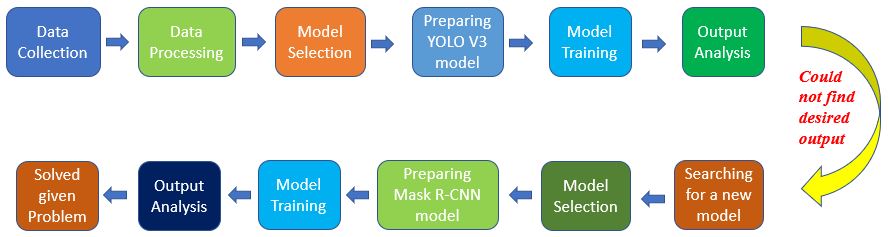


Figure 19. Overall Workflow

## 3.2 Dataset Preprocessing

The data set contains only the road images with roughly 700 samples. The image samples were collected from the internet, from some different persons, and manually by us. The dataset contains different resolutions of pictures, and the pixel value ranges from 300 to 1200 pixels.

The next step was to annotate the image samples to make training and a validation set. To complete this task, all of the image samples were annotated by the annotation tool. The annotation tool successfully marked the image's defected area, and each image needed multiple annotations to draw the different defect patters.

During the annotation, the total amount of samples has divided among five group members to annotate samples individually. The VGG image annotator version 2.0.10 has been used to complete this annotation task. After completing the annotation, each group member returned their task with a JSON file. Later all of these JSON files integrated and formed the main dataset. The annotation demo is shown in the figure.

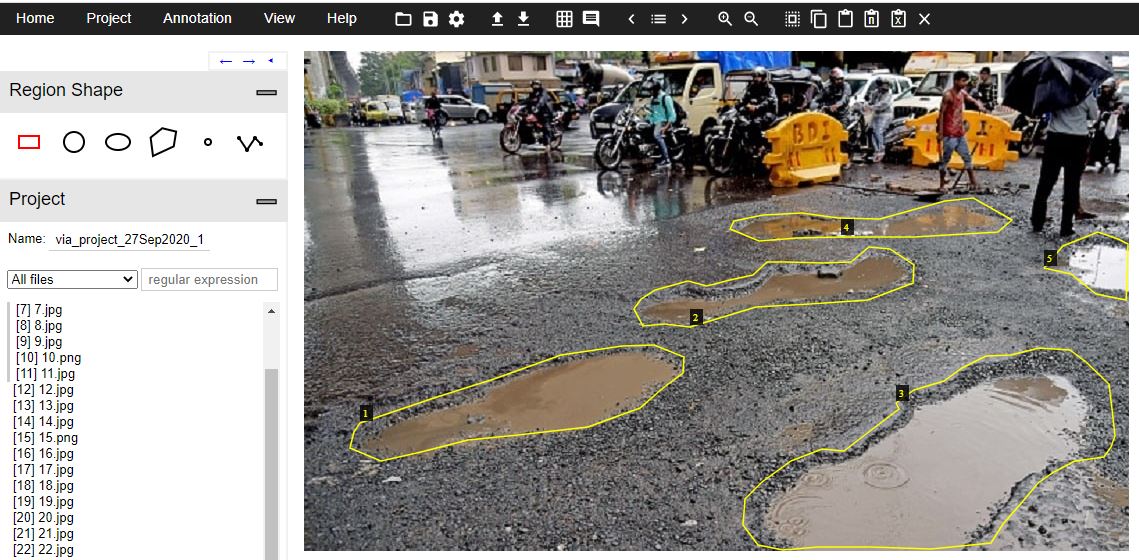


Figure 20. Annotating image with VGG

## 3.3 Neural Network Models

There are many deep learning models available in this field. The maximum real-world problem could be solved only by using some pre-trained model or some already defined ones. Our research has found two popular models for annotation and segmentation. One is YOLO V3, and another one is Mask R-CNN.

### 3.3.1 YOLO v3

In this implementation, a pre-trained YOLO model has been used with the technique of transfer learning. YOLO V3 is an evolution over past YOLO identification algorithms. Contrasted with earlier forms, it highlights multi-scale annotation, a more advanced extractor network, and a few loss functions changes. Therefore, this model would now be able to identify a lot more objects in the images from very little to large scale. Furthermore, much like other detector models, YOLO V3 also runs very quickly and makes real-time inference. The network architecture has shown in the figure.

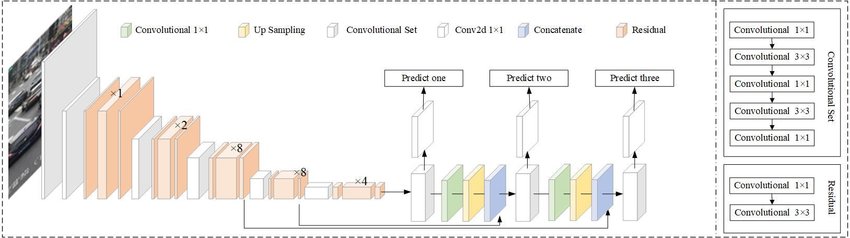


Figure 21. YOLO v3 Network Architecture

Now the most important task is to configure the model for our problem. The default configuration of the model contained 80 classes. The technique transfer learning has used to make the prediction only for one class named “bad.” The main aim was to convert the network from object detection to defect detection. The process has shown in the figure.

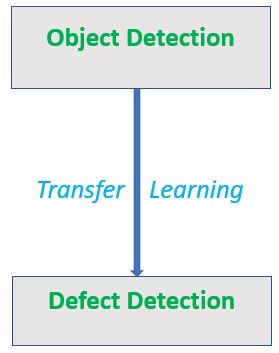


Figure 22. Transfer Learning for defect detection

#### 3.3.1.1 YOLO v3 Training Details

The training phase of this algorithm was so manageable and comfortable for us. This network structure was straight forward and responded quickly, even in a standard configuration of a machine. The intended layer modification was also very easy for its transparent and user-friendly structure.

The Darknet framework was used to complete this implementation task, and all the required libraries were imported from this framework. The network was configured with a batch size of 32, the learning rate of 0.001, class of 1, filters of ((classes+5) \* 3), IOU threshold of 75%, and 1000 iteration. The other hyperparameters have been kept in default condition, which has given in the Darknet documentation.

### 3.3.2 Mask RCNN

The Mask RCNN is a deep neural network algorithm designed to solve the instance segmentation problem in the field of AI. This algorithm's main power is that it can isolate various objects in a picture or a video. There are two phases of Mask RCNN. To start with, it produces proposals about the areas where there may be an instance dependent on the sample picture. Second, it predicts the instance's class, adjusts the bounding box, and produces a mask based on the instance's pixel-level dependent on the first stage proposal. The two phases are associated with the backbone structure. The network architecture has shown in figure.

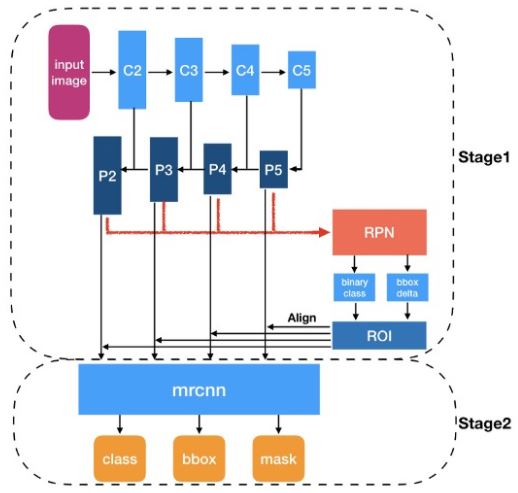


Figure 23. Mask R-CNN network architecture

#### 3.3.2.1 Mask R-CNN Training Details

At first, the TensorFlow framework has used to implement this task. The network was configured with the backbone of ResNet-101, NMS threshold of 70%, Image minimum dimension of 800, a maximum dimension of 1024, the epoch of 200, the learning rate of 0.001, and the number of classes are 1. But the problem was raised with the version conflict of this implementation’s library. Some of the methods were obsolete, and the real performance was not achieved. Then the framework has been changed from TensorFlow to PyTorch.

In the PyTorch framework, the problem statement has been solved successfully. The training was conducted with different variations of backbone and batch size. Some of them are ResNet-101 backbone with 32 batch size, ResNet-50 backbone with 64 batch size, ResNet-101 backbone with 64 batch size, and ResNet-50 backbone with 512 batch size. Among them, the ResNet-50 with a high batch size of 512 has given the best outcome.

1

# CHAPTER 4: Results

The Deep Learning model is nowadays generating excellent outputs in the field of AI. In our case, also, the result was satisfactory. In this section, the results of the two implemented models have been demonstrated.

## 4.1 Result of YOLO v3

This algorithm produced acceptable results within a short training time. After 900 iterations, this model generated a loss value of 0.14. The image defect was labeled as “Bad” for the initial experiment. This model was evaluated on some test images of broken roads, and it annotated the fault correctly. But for some of the information shortage, we had to change the model. Some testing samples have been given below.

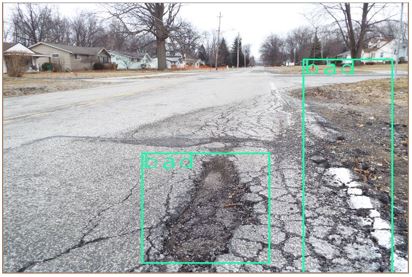
****

Figure 24. Inference result of YOLO v3

## 4.2 Result of Mask R-CNN

This model was tested in two ways. First, it was used to inference on the validation set, and second, it was evaluated using AP matrix. The training result was 96% accuracy, with a loss of 1.5. During the inference, the threshold was set to 70%, and the model segmented the images correctly. Some testing samples have been given below.

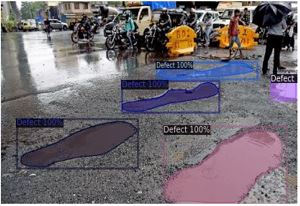


Figure 25. Inference result of Mask R-CNN

To find the validation set's real performance, we have used the inference technique with different training weights. Now to evaluate the validation set with AP matrix, different backbone and batch size have been applied. The table of the AP matrix has shown below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Iteration | Backbone | Batch Size | AP | AP50 | AP75 | APS | APM | APL |
| 50k | ResNet-50 | 32 | 15.737 | 38.906 | 9.061 | 0.480 | 9.413 | 21.526 |
| 50k | ResNet-50 | 64 | 17.016 | 39.925 | 9.624 | 0.000 | 11.517 | 23.012 |
| 50k | ResNet-50 | 512 | 17.138 | 41.726 | 11.744 | 0.000 | 11.623 | 23.264 |
| 50k | ResNet-101 | 64 | 19.693 | 46.633 | 13.456 | 0.000 | 14.357 | 26.079 |

Figure 26. Evaluation result of the validation set

## 4.3 Comparative Analysis

This comparative analysis has been based on model accuracy, single image classification, model size, training time, average loss, and other factors.

According to the research, the YOLO V3 could not solve the entire problem statement. It only solved a portion of the problem. Though the training cost was less compared to Mask-RCNN, it was only able to detect the square region of the object. On the other hand, the Mask R-CNN solved the entire problem with a good precision though the training cost was very high compared to YOLO V3. The Mask R-CNN was able to mark the object's exact shape with different color labels and confidence. After observing the overall performance and solution statement, the Mask R-CNN was state of the art.

# CHAPTER 5: Conclusions

## 5.1 Discussion

In this project, the main aim was to predict road conditions. The output was produced based on the image input that has given satisfactory results. The implementation of this project started with image classification, then gradually implemented the instance segmentation model.

The implementation of the solution of the problem statement was very challenging. Due to this global pandemic situation, it was difficult to manage the workflow like data collection, group members meeting, elaborate discussion with faculty, and other difficulties.

This project was not involved with any external hardware instrument, which was the main advantage of the implementation. We had not done any hardware experiments and implemented the entire project based on software tools only. Though the situation was adverse, the group members' collaboration and faculties regular instruction made it possible to complete the project.

## 5.2 Summary

This research started from scratch and finally reached the phase of instance segmentation, which can identify the defects from road images.

The most prominent object detection model YOLO V3 did a great job on this project data. It was able to identify the defects successfully. The main drawbacks of this algorithm were, it could not determine the exact shape of a fault. After further research, another state of art algorithm was found to detect the precise shape of a defect from a road image.

The Mask R-CNN played a vital role in this field. The algorithm was designed so that it can predict the class, mask, bounding box, and confidence level at the same time. The algorithms only downside was its training cost. It requires a very powerful GPU and long training time to produce the expected output. Once the training becomes complete, then it shows the impressive outcomes. It successfully managed the solution of this problem statement.

## 5.3 Future Work

Training with more samples could be an important experiment in the future, which was not possible for this global pandemic. The image annotation cost could be reduced using active learning or finding a more precise method by research. The training time reduction could be another focus and needs further research to find a clue.

Some other algorithms also could be used to solve this problem, but it must be selected based on comparative performance. Here also needs a concern on the image samples preparation because more accurate the annotated samples will produce a more precise result.

This project could be deployed on a car with a camera to detect the defect of roads from the live videos or images taken by the camera. Worldwide auto driven cars are using the similar kind of technology. So, live defect detection is also possible with this project by some more improvement.

# CHAPTER 6: REFERENCES

1. Singh, J., & Shekhar, S. (2018). Road Damage Detection And Classification In Smartphone Captured Images Using Mask R-CNN. ArXiv, abs/1811.04535.
2. A. Alfarrarjeh, D. Trivedi, S. H. Kim and C. Shahabi, "A Deep Learning Approach for Road Damage Detection from Smartphone Images," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5201-5204, doi: 10.1109/BigData.2018.8621899.
3. Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., & Omata, H. Generative adversarial network for road damage detection. Computer‐Aided Civil and Infrastructure Engineering.
4. Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2020). Transfer Learning-based Road Damage Detection for Multiple Countries. arXiv preprint arXiv:2008.13101.
5. Wang, W., Wu, B., Yang, S., & Wang, Z. (2018, December). Road damage detection and classification with Faster R-CNN. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 5220-5223). IEEE.
6. Tan, C., Uddin, N., & Mohammed, Y. M. (2019, August). Deep Learning-Based Crack Detection Using Mask R-CNN Technique. In 9 th International Conference on Structural Health Monitoring of Intelligent Infrastructure.
7. Ukhwah, E. N., Yuniarno, E. M., & Suprapto, Y. K. (2019, August). Asphalt Pavement Pothole Detection using Deep learning method based on YOLO Neural Network. In 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 35-40). IEEE.
8. Arman, M. S., Hasan, M. M., Sadia, F., Shakir, A. K., Sarker, K., & Himu, F. A. (2020, February). Detection and Classification of Road Damage Using R-CNN and Faster R-CNN: A Deep Learning Approach. In International Conference on Cyber Security and Computer Science (pp. 730-741). Springer, Cham.
9. Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. Computer‐Aided Civil and Infrastructure Engineering, 33(12), 1127-1141.
10. Faramarzi, M. (2020). Road Damage Detection and Classi cation Using Deep Neural Networks with Smartphone Images. Available at SSRN 3627382.
11. Fan, Z., Wu, Y., Lu, J., & Li, W. (2018). Automatic pavement crack detection based on structured prediction with the convolutional neural network. arXiv preprint arXiv:1802.02208.
12. Fan, R., Bocus, M. J., Zhu, Y., Jiao, J., Wang, L., Ma, F., ... & Liu, M. (2019, June). Road crack detection using deep convolutional neural network and adaptive thresholding. In 2019 IEEE Intelligent Vehicles Symposium (IV) (pp. 474-479). IEEE.
13. Koch, C., & Brilakis, I. (2011). Pothole detection in asphalt pavement images. Advanced Engineering Informatics, 25(3), 507-515.
14. Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection using deep neural networks with images captured through a smartphone. arXiv preprint arXiv:1801.09454.
15. Ochoa-Ruiz, G., Angulo-Murillo, A. A., Ochoa-Zezzatti, A., Aguilar-Lobo, L. M., Vega-Fernández, J. A., & Natraj, S. (2020). An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment. Applied Sciences, 10(11), 3974.
16. Chen, Q., Gan, X., Huang, W., Feng, J., Shim, H. (2020). Road Damage Detection and Classification Using Mask R-CNN with DenseNet Backbone. CMC-Computers, Materials & Continua, 65(3), 2201–2215.
17. Ping, Ping & Yang, Xiaohui & Gao, Jerry. (2020). A Deep Learning Approach for Street Pothole Detection. 10.1109/BigDataService49289.2020.00039.
18. D. Zhang, Q. Li, Y. Chen, M. Cao, L. He, and B. Zhang, “An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection,” Image and Vision Computing, vol. 57, pp. 130–146, Jan. 2017, doi: 10.1016/j.imavis.2016.11.018.
19. Thu Huong, Nguyen & The Long, Nguyen & Sidorov, Denis & Dreglea, Aliona. (2018). Machine learning algorithms application to road defects classification. Intelligent Decision Technologies. 12. 1-8. 10.3233/IDT-170323.
20. L. Attard, C. J. Debono, G. Valentino, M. Di Castro, A. Masi and L. Scibile, "Automatic Crack Detection using Mask R-CNN," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 2019, pp. 152-157, doi: 10.1109/ISPA.2019.8868619.
21. Yanjia Li, E., 2019. Dive Really Deep Into YOLO V3: A Beginner’S Guide. [online] Medium. Available at: <https://towardsdatascience.com/dive-really-deep-into-yolo-v3-a-beginners-guide-9e3d2666280e> [Accessed 26 September 2020].
22. Q. Mao, H. Sun, Y. Liu and R. Jia, "Mini-YOLOv3: Real-Time Object Detector for Embedded Applications," in IEEE Access, vol. 7, pp. 133529-133538, 2019, doi: 10.1109/ACCESS.2019.2941547.
23. Zhang, X., 2018. Simple Understanding Of Mask RCNN. [online] Medium. Available at: [https://medium.com/@alittlepain833/simple-understanding-of-mask-rcnn-134b5b330e95](https://medium.com/@alittlepain833/simple-understanding-of-mask-rcnn-134b5b330e95#:~:text=Mask%20RCNN%20is%20a%20deep,two%20stages%20of%20Mask%20RCNN) [Accessed 27 September 2020].
24. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
25. ROSENFELDER, M., 2020. Detectron2 - How To Use Instance Image Segmentation For Building Recognition. [online] Python Tutorials for Machine Learning, Deep Learning and Data Visualization. Available at: <https://rosenfelder.ai/Instance_Image_Segmentation_for_Window_and_Building_Detection_with_detectron2/> [Accessed 27 September 2020].
26. SHARMA, P., 2019. Image Segmentation. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/> [Accessed 27 September 2020].
27. Brownlee, J., 2019. A Gentle Introduction To Transfer Learning For Deep Learning. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/transfer-learning-for-deep-learning/> [Accessed 27 September 2020].