Department of Electrical and Computer Engineering

North South University

Directed Research - Report

Road Defect Detection Using Image Annotation

Md Sehabub Zaman Pranta	**1611251042**
Tanvir Ahmed	**1410982042**

Date: 19.01.2021

Course: CSE498R

Section: 9

Faculty Advisor:

Dr. Mohammad Ashrafuzzaman Khan

Assistant Professor

Department of Electrical & Computer Engineering

Fall 2020

ABSTRACT

Avoiding defective 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 have implemented a deep learning-based instance segmentation model with the transfer learning technique. The Mask R-CNN model with ResNet-101-FPN backbone has been trained and tested on our collected dataset with different batch sizes to detect road defects. After that, we evaluated the model using the AP matrix and compared their results. The model produced a satisfactory result as it can detect the defect's shape with a good confidence level and predict the damaged areas with different color annotations.

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 for supporting 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**
ACKNOWLEDGEMENT **3**
LIST OF FIGURES **5**
CHAPTER 1: INTRODUCTION **6**
1.1 IMAGE ANNOTATION 6
1.2 NEURAL NETWORK 6
1.3 TRANSFER LEARNING 6
1.4 INTELLIGENT SYSTEM 6
1.5 AVERAGE PRECISION (AP) 7
1.6 INTERSECTION OVER UNION (IOU) 7
1.7 PROJECT AIM AND OBJECTIVE 8
CHAPTER 2: RELATED WORK **9**
2.1 ROAD DAMAGE DETECTION AND CLASSIFICATION WITH DETECTRON2 AND FASTER R-CNN 9
2.2 ROAD DAMAGE DETECTION USING RETINANET 10
2.3 A DEEP LEARNING APPROACH FOR ROAD DAMAGE DETECTION FROM SMARTPHONE IMAGES 11
2.4 GENERATIVE ADVERSARIAL NETWORK FOR ROAD DAMAGE DETECTION 12
2.5 TRANSFER LEARNING-BASED ROAD DAMAGE DETECTION FOR MULTIPLE COUNTRIES 12
2.6 ROAD DAMAGE DETECTION AND CLASSIFICATION WITH FASTER R-CNN 13
2.7 DEEP LEARNING-BASED CRACK DETECTION USING MASK R-CNN TECHNIQUE 14
2.8 ASPHALT PAVEMENT POTHOLE DETECTION USING DEEP LEARNING METHOD BASED ON YOLO NEURAL NETWORK 14
2.9 DETECTION AND CLASSIFICATION OF ROAD DAMAGE USING R-CNN AND FASTER R-CNN: A DEEP LEARNING APPROACH 15
2.10 ROAD DAMAGE DETECTION AND CLASSIFICATION USING DEEP NEURAL NETWORKS WITH SMARTPHONE IMAGES 16
2.11 ROAD DAMAGE DETECTION AND CLASSIFICATION USING DEEP NEURAL NETWORKS (YOLOV4) WITH SMARTPHONE IMAGES 16
2.12 AUTOMATIC PAVEMENT CRACK DETECTION BASED ON STRUCTURED PREDICTION WITH THE CONVOLUTIONAL NEURAL NETWORK 17
2.13 ROAD CRACK DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORK AND ADAPTIVE THRESHOLDING 17
2.14 POTHOLE DETECTION IN ASPHALT PAVEMENT IMAGES 18
2.15 AN ASPHALT DAMAGE DATASET AND DETECTION SYSTEM BASED ON RETINANET FOR ROAD CONDITIONS ASSESSMENT 20
2.16 ROAD DAMAGE DETECTION AND CLASSIFICATION USING MASK R-CNN WITH DENSENET BACKBONE 20
2.17 A DEEP LEARNING APPROACH FOR STREET POTHOLE DETECTION 21
2.18 AN EFFICIENT AND RELIABLE COARSE-TO-FINE APPROACH FOR ASPHALT PAVEMENT CRACK DETECTION 22
2.19 MACHINE LEARNING ALGORITHMS APPLICATION TO ROAD DEFECTS CLASSIFICATION 22
2.20 AUTOMATIC CRACK DETECTION USING MASK R-CNN 23
CHAPTER 3: METHODOLOGY **24**
3.1 WORKFLOW 24
3.2 DATASET PREPROCESSING 24
3.3 NEURAL NETWORK MODELS 25
3.3.1 Mask R-CNN 25
3.3.1.1 Transfer Learning Implementation Details 26
3.3.1.2 Mask R-CNN Training Details 27
CHAPTER 4: RESULTS **29**
4.1 INFERENCE ON IMAGES 29
4.2 EVALUATION AND ANALYSIS 33
CHAPTER 5: CONCLUSIONS **34**
5.1 DISCUSSION 34
5.2 SUMMARY 34
5.3 FUTURE WORK 35
REFERENCES **36**

LIST OF FIGURES

FIGURE 1. IOU 7

FIGURE 2. DAMAGE TYPE DISTRIBUTIONS OVER THREE COUNTRIES (CZECH, INDIA, AND JAPAN) 9

FIGURE 3. ERROR IN ROAD DAMAGE TYPE PREDICTION BY A MACHINE LEARNING MODEL 10

FIGURE 4. RETINANET 10

FIGURE 5. A DEEP LEARNING APPROACH FOR ROAD DAMAGE DETECTION AND CLASSIFICATION 11

FIGURE 6. PREDICTED LABELS FOR SAMPLE IMAGES 13

FIGURE 7. EXAMPLES OF TRAINING IMAGES WITH CRACK GROUND TRUTH 14

FIGURE 8. LOSS OF YOLO V3, YOLO V3 TINY AND YOLO V3 SPP IN THE MODELING 15

FIGURE 9. DETECTION AND CLASSIFICATION RESULTS FOR EACH CLASS 16

FIGURE 10. EXAMPLES OF STRUCTURED PREDICTION BASED ON CNN 17

FIGURE 11. BILATERAL FILTERING AND IMAGE SEGMENTATION; (A) ORIGINAL POSITIVE IMAGE; (B) FILTERED POSITIVE IMAGE; (C) SEGMENTATION RESULT 18

FIGURE 12. POTHOLE DETECTION MODEL 19

FIGURE 13. RESULT OF IMAGE SEGMENTATION FOR TWO POTHOLES 19

FIGURE 14. THE PROPOSED MASK R-CNN FRAMEWORK WITH DENSENET 21

FIGURE 15. COMPARISON OF ACCURACY OF DIFFERENT MODELS 21

FIGURE 16. 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 ... 22

FIGURE 17. SAMPLE IMAGES FROM THE ANNOTATED CRACK DATASET 23

FIGURE 18. OVERALL WORKFLOW 24

FIGURE 19. ANNOTATING IMAGE WITH VGG 25

FIGURE 20. MASK R-CNN NETWORK ARCHITECTURE 26

FIGURE 21. OBJECT TO DEFECT DETECTION CONVERSION TECHNIQUE. 26

FIGURE 22. IMPLEMENTATION OF TRANSFER LEARNING. 27

FIGURE 23. TRAINING ACCURACY AND LOSS GRAPH. 28

FIGURE 24. INFERENCE ON IMAGES 32

CHAPTER 1: INTRODUCTION

1.1 IMAGE ANNOTATION

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, and regions is not easy 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 INTELLIGENT SYSTEM

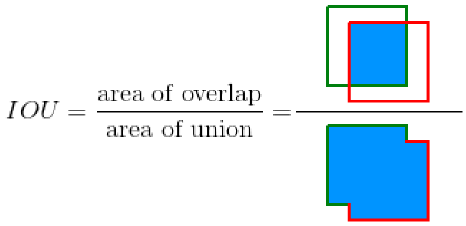
An intelligent system is a machine with an embedded, Internet-connected computer that can gather and analyze data and communicate with other systems. In modern days, the most needed thing is to enhance a model's performance, in this proposed method preferred to use a pre-trained model for training. To 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.

1.5 AVERAGE PRECISION (AP)

The average precision is an estimate that unites recall and precision for sequenced retrieval outcome. The precision is a measure of result relevancy, and the recall is a measure of how many actual relevant results are returned. For an information requirement, the average precision is the mean of the precision scores after each identical record is retrieved.

1.6 INTERSECTION OVER UNION (IOU)

It is a method to quantify the overlap between two areas. It can be used to evaluate an object detector. IoU of two areas A and B is calculated by the common area between the areas divided by the total area of them.



$$\mathrm{IoU(A,B)} = \frac{A \cap B}{A \cup B}$$

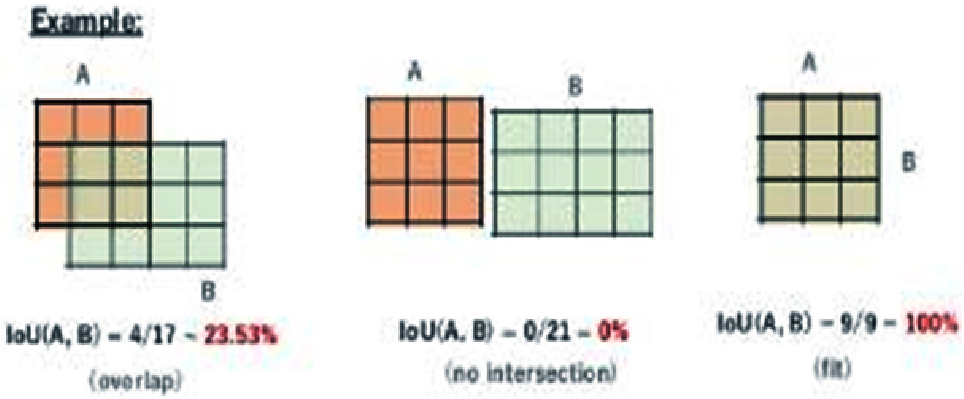


Figure 1. IoU

1.7 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 -

I. Researching Image annotation technique, tools, algorithms.
II. Reviewing related works.
III. Studying various detection algorithms.
IV. Collecting road images.
V. Working in different frameworks.
VI. Implementing state-of-the-art object detection algorithm Mask R-CNN.
VII. Experimenting with different backbones, parameters, and batches of the algorithm.
VIII. Analyzing and comparing the result to reach a conclusion.

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

CHAPTER 2: RELATED WORK

2.1 ROAD DAMAGE DETECTION AND CLASSIFICATION WITH DETECTRON2 AND FASTER R-CNN

The authors in this paper explored Detectron2's implementation of Faster R-CNN using different base models and configurations for detecting road damages. The detection is based on Object Recognition using Fast R-CNN. They also experimented with these models on the Global Road Damage Detection Challenge 2020 Dataset. There are 21,041 images (2,829, 7,706, and 10,506 for Czech, India, and Japan, respectively) in the training set. And test1 has 2,631, and test2 has 2,664 images. In total, the training dataset contains 34,702 annotated instances. Four types of annotated road damage types were classified, respectively Pothole, Alligator Crack, Transverse Crack, and Longitudinal Crack.

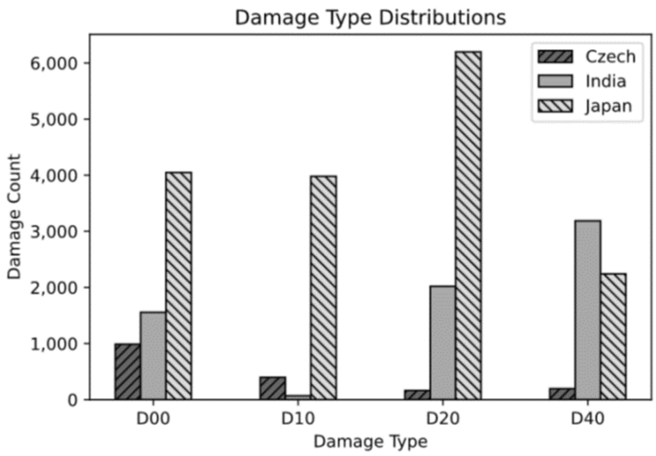


Figure 2. Damage type distributions over three countries (Czech, India, and Japan)

The X101-FPN base model for Faster R-CNN with Detectron2's default configuration gives them the most efficient result, which provides F1 scores of 51.0% and 51.4%, respectively, for the test1 and test2 sets of the challenge. They also evaluated the results against the existing annotations and came across some inconsistencies. Therefore, they suggested strategies to improve the labeling process for the dataset.

Nevertheless, the model can make a mistake in prediction like the below picture. Also, Faster R-CNN is not state-of-the-art in object detection, which cannot predict polygon shape. So, it doesn't show us the shape of the damage.

Figure 3. Error in road damage type prediction by a machine learning model

2.2 Road Damage Detection Using RetinaNet

In this paper, the authors trained and tested different deep learning models to find efficient models with high accuracy. They used faster one stage detectors, classification, and bounding-box regression rather than traditional slower two-stage detectors like R-CNN and Fast R-CNN. Among those one-stage models, they found that RetinaNet can detect road damages faster with higher accuracy.

Initially, they tested the Single Shot MultiBox Detector (SSD) backbone with DenseNet, ResNet, and VGG network. Then, they used RetinaNet. They trained and tested RetinaNet with various backbone neural networks such as VGG, ResNet, InceptionResNetV2, and DenseNet. They used a dataset containing 9,053 damage images and 15,435 damage bounding boxes, where the size of the images was 600x600 pixels.

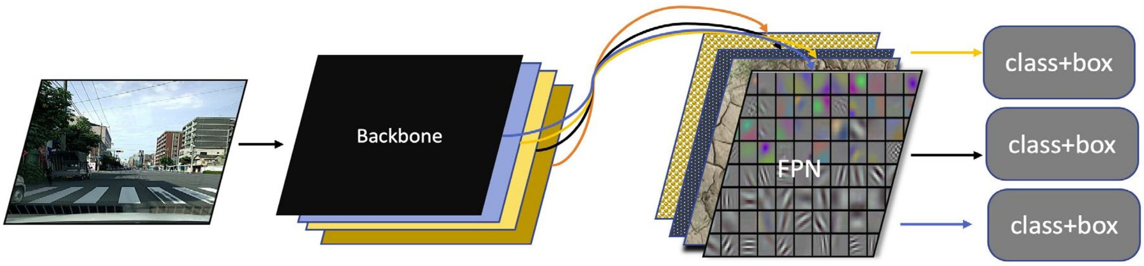


Figure 4. RetinaNet

SSD based models cannot achieve high accuracy because of imbalance class in the data samples. So, the authors trained and compared RetinaNet models with various backbones, including ResNet, VGG, DenseNet, and InceptionResNetV2, where DenseNet does not perform well. The authors concluded that their proposed RetinaNet based approach is efficient in detecting different types of road damage with high accuracy.

2.3 A DEEP LEARNING APPROACH FOR ROAD DAMAGE DETECTION FROM SMARTPHONE IMAGES

This paper describes the road damage detection and classification challenge's solution, 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.

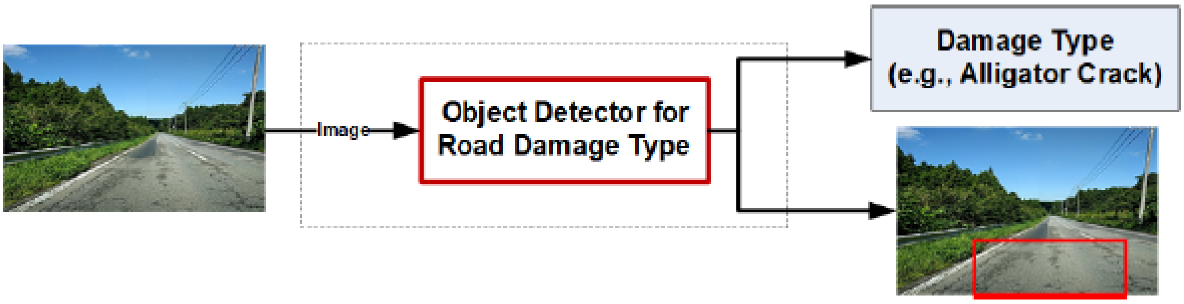


Figure 5. A Deep Learning Approach for Road Damage Detection and Classification

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.

2.4 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.

The authors also 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 potholes on roads.

Generative Adversarial Network has some drawbacks as well.

i. It's harder to train. The user needs to give various kinds of data uninterruptedly to verify if it is working accurately or not.
ii. Optimizing the loss function is very difficult.

2.5 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 provided 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 6. 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.

2.6 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.

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.

2.7 DEEP LEARNING-BASED CRACK DETECTION USING MASK R-CNN TECHNIQUE

The authors of this paper worked on Mask R-CNN with backbone Resnet-101 as the proposed method. They used a dataset consists of 352 crack images and divided them into training, validation, and testing data. Here, they tried to build an automatic crack detector using the state-of-art technique. In the paper, 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 gave an excellent result in real-time on-site.

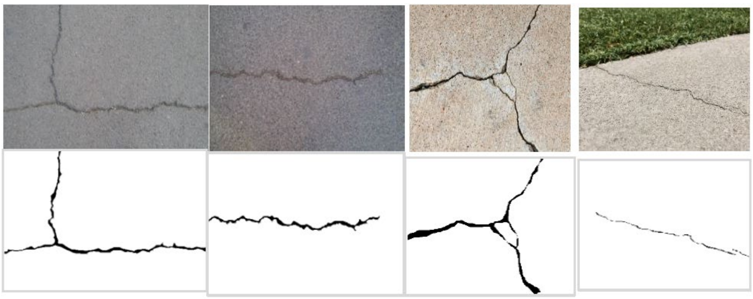


Figure 7. Examples of training images with crack ground truth

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

2.8 ASPHALT PAVEMENT POTHOLE DETECTION USING DEEP LEARNING METHOD BASED ON YOLO NEURAL NETWORK

There are three different architecture configurations of YOLOv3 in this paper: YOLOv3, YOLOv3 tiny, and YOLOv3 SPP. The pothole detection method was used on these three architectures. Firstly, the authors researched the data, secondly, annotated and labelled them, and built 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.

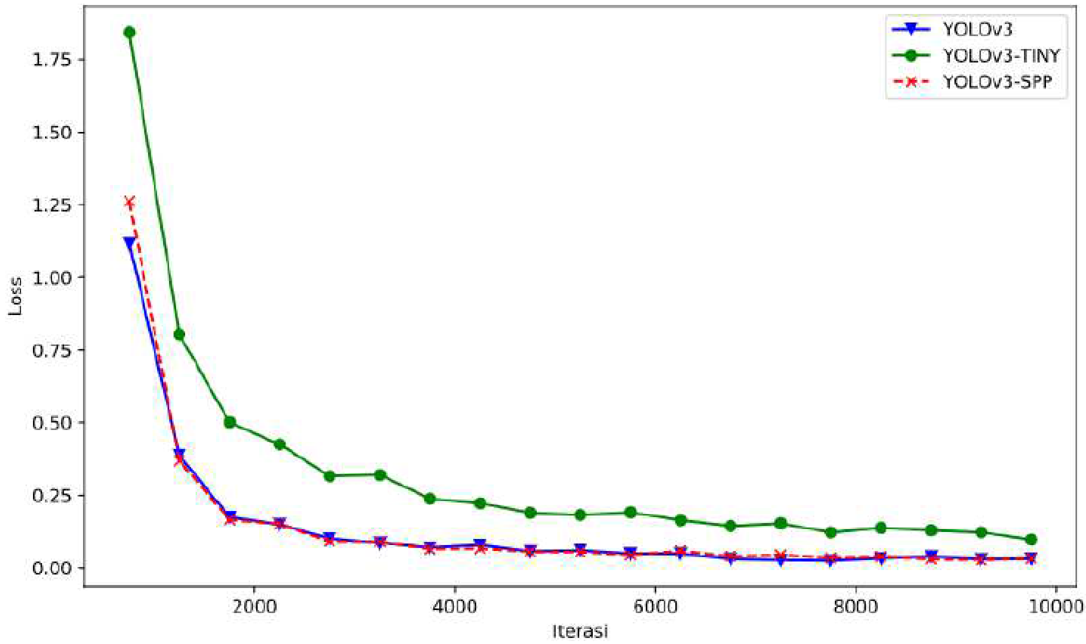


Figure 8. Loss of Yolo v3, Yolo v3 Tiny and Yolo v3 SPP in the modeling

In this paper, 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 does not exist. Here the approach of this method is different from our problem. Our goal is to find different types of damages including potholes.

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

As monitoring roads manually in cities is time-consuming and requires lots of labor, the authors of this paper proposed a model using R-CNN and Faster R-CNN to identify road damages. In this model, 1100 images were resized and labelled with potholes, crack, etc. and 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 extents Faster R-CNN for pixel-level segmentation. It is quite obvious that Mask R-CNN is better than Faster R-CNN, and it also serves the purpose of our proposed method by doing instance segmentation.

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

The research in this paper was done 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. The dataset 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 class names of damages like D00 D01, D10, D11, D20, D40, D43, D44 were present and the accuracy were different from each other.

	D00	*D01*	*D10*	*D11*	*D20*	*D40*	*D43*	*D44*
Recall of SSD Inception V2	0.22	0.60	0.10	0.05	0.68	0.03	0.81	0.62
Precision of SSD Inception V2	0.73	0.84	0.99	0.95	0.73	0.67	0.77	0.81
Accuracy of SSD Inception V2	0.78	0.80	0.94	0.92	0.85	0.95	0.95	0.83
Recall of SSD MobileNet	0.40	0.89	0.20	0.05	0.68	0.02	0.71	0.85
Precision of SSD MobileNet	0.73	0.64	0.99	0.95	0.68	0.99	0.85	0.66
Accuracy of SSD MobileNet	0.81	0.77	0.92	0.94	0.83	0.95	0.95	0.81

Figure 9. 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 a 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.

2.11 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 dataset available. To train the model, a transfer learning technique was applied while using pre-trained weights. The dataset 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 at 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.

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 roads, and some of the classes are not even needed for damage segmentations.

2.12 AUTOMATIC PAVEMENT CRACK DETECTION BASED ON STRUCTURED PREDICTION WITH THE CONVOLUTIONAL NEURAL NETWORK

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.

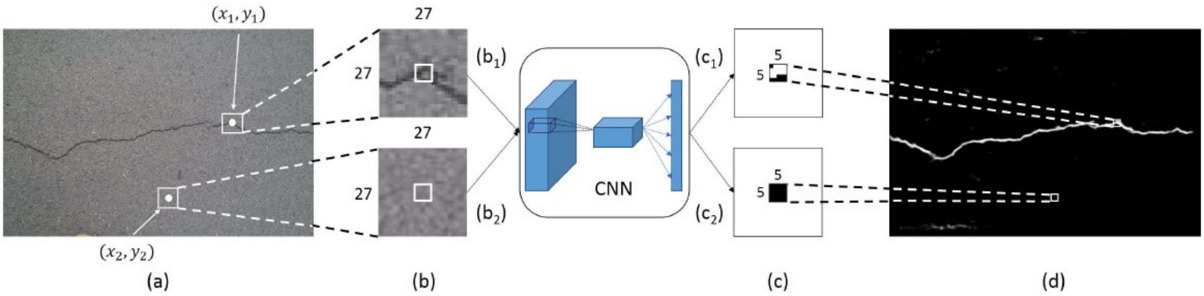


Figure 10. Examples of structured prediction based on CNN

This paper’s method can only find the cracks and determine whether the crack is present or not. However, our proposed method is to find any type of road damages, including their shape. So, some improvement is needed to achieve our solution.

2.13 ROAD CRACK DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORK AND ADAPTIVE THRESHOLDING

The work in this paper was done by using CNN to Adapt thresholding to detect the road crack. The work was 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's existence. One term is added: ReLU, representing a rectified linear unit, which is the most popular activation function for deep neural networks.

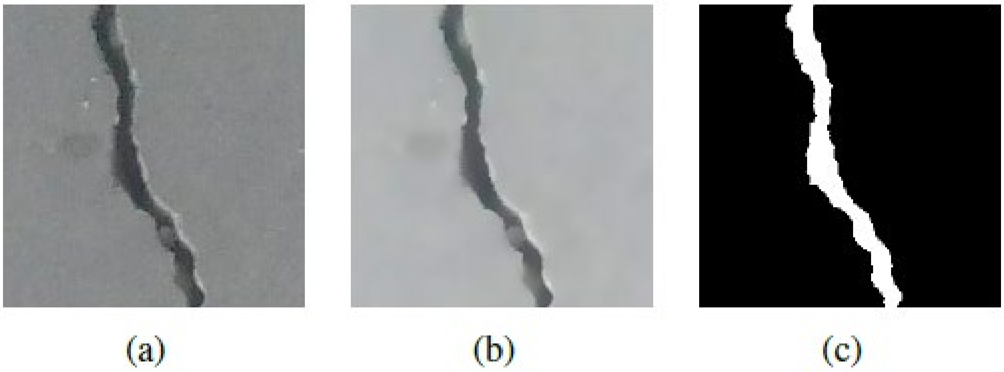


Figure 11. Bilateral filtering and image segmentation; (a) original positive image; (b) filtered positive image; (c) segmentation result

An adapting thresholding approach was used for the segmentation process. In the segmentation part, before approaching the adapting thresholding, a bilateral filter was 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, they formulated the thresholding problem as 2D. This 2D histogram thresholding formula can segment the crack images.

Here, this paper's method 2D histogram thresholding can only segment the crack-based areas, but our proposed model can detect the damaged road and segment its damaged part. For the paper's model, things need to be changed in the 2D histogram part because it only segments the crack areas.

2.14 POTHOLE DETECTION IN ASPHALT PAVEMENT IMAGES

The work in this paper is on asphalt pavement images to detect potholes. 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.

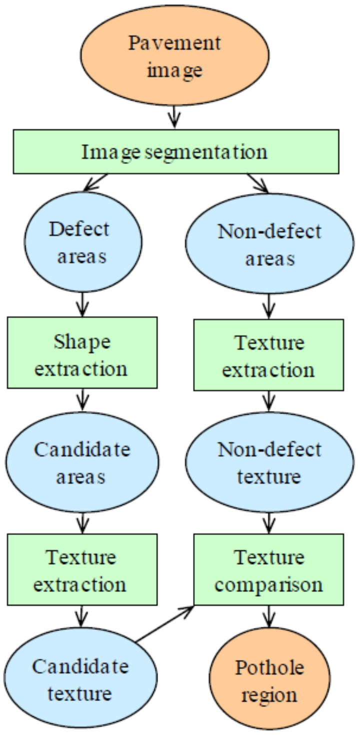


Figure 12. Pothole detection model

After segmenting the defect and non-defect parts, the shape extraction part works on the defected area. For measuring shade of the region, here it puts a threshold value and then shrinks the shade region to minimally connected skeleton where morphological thinning is used. The standard deviation of gray-level intensity values as a statistical measure. It is used to describe the texture of both the inside and the outside region. Finally, it works for both defect and non-defect region and compares the texture.

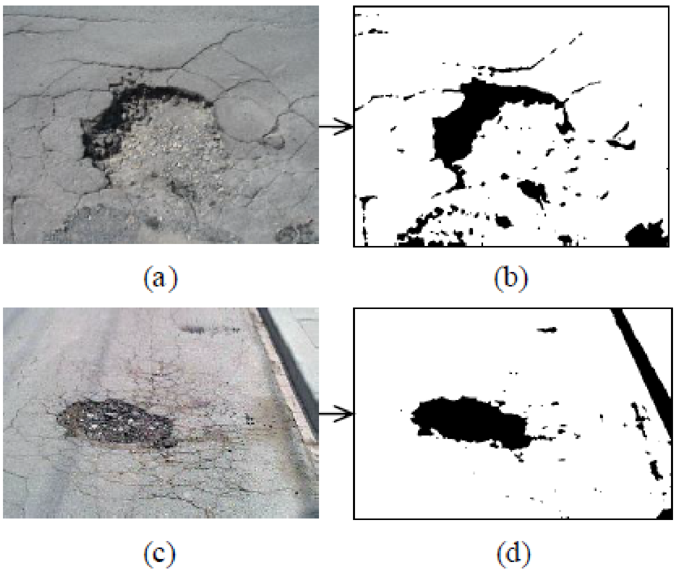


Figure 13. Result of image segmentation for two potholes

This paper's method has segmented both defect and non-defect areas, and our proposed method is for segmenting the damaged region only. Here, the method of shape extraction and texture extraction can be useful by improving for better output.

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 were trained for mobile implementation. In this work, the RetinaNet object detector was 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 smartphones. It can detect damaged objects on the asphalt road, but our proposed method detects any road damage.

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 a new method, and it detects road damages precisely and can also segment the road damage mask correctly.

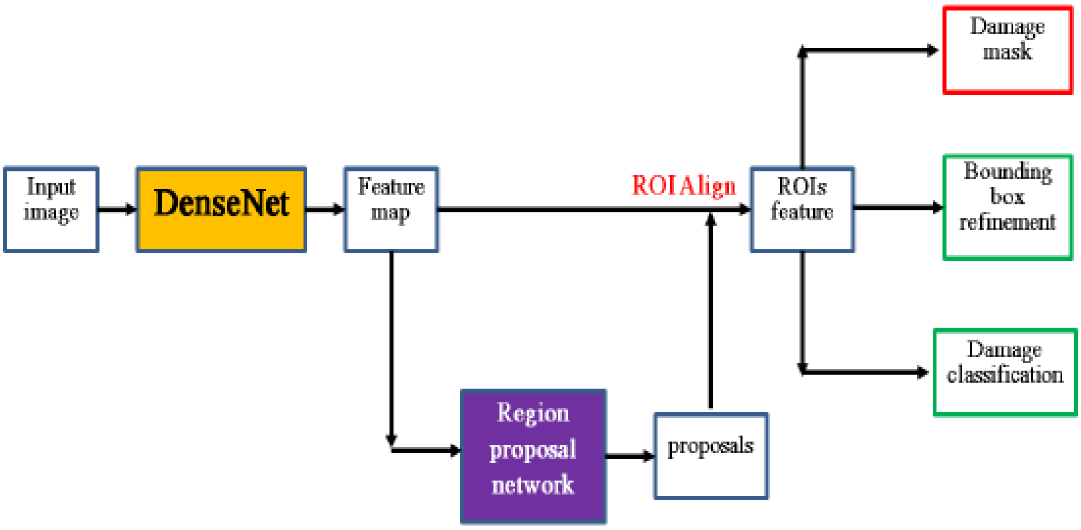


Figure 14. The proposed Mask R-CNN framework with DenseNet

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 our project's goal.

2.17 A Deep Learning Approach for Street Pothole Detection

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

Size	YOLO V3	SSD	HOG	Faster R-CNN
200 Images	53%	47%	24%	72%
650 Images	67%	59%	25%	71%
850 Images	65%	55%	27%	67%
1000 Images	69%	59%	-	69%
1100 Images	73%	-	-	60%
1500 Images	82%	80%	-	74%

Figure 15. 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. 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 that 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 was presented. In order to promote identification, a new design Region of Belief (ROB) was implemented. A novel region growing algorithm was suggested for the identification of defects.

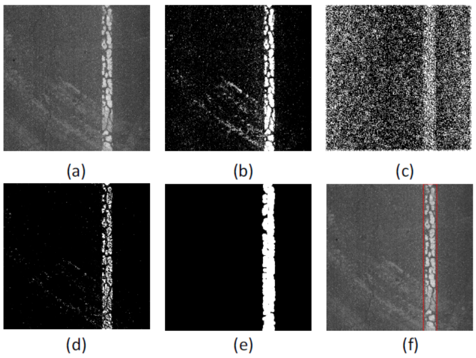


Figure 16. Illustration of process of lane and sign markings removal. (a) the original pavement image, (b) bin map, (c) variance map, (d)fused binary image, (e) further result from morphological processing, (f) final result

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 introduced a new approach for automated identification and analysis of road defects based on machine research algorithms. The road defects were 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 was used to develop the classification models, and both sets had data showing the accuracy of the proposed system. These algorithms were required to identify road damages by following the Random Forest algorithm accurately. It was also recommended using the graph cutting method and Marcov algorithm that enhances image segmentation efficiency.

The authors 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 seems problematic and needs 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 authors developed a ground-truth mask dataset on images from a regular crack dataset to train Mask R-CNN for crack detection. 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 were an accuracy value of 93.94% and a recall of 77.5%.

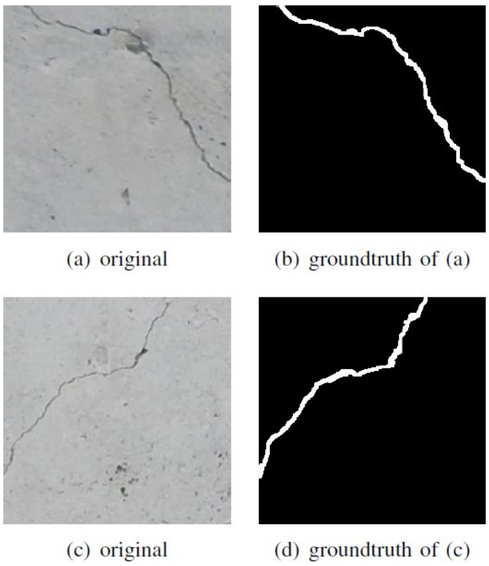


Figure 17. Sample images from the annotated crack dataset

This method is only for crack detection, but our proposed method is to find different road damage types. So, things need to be changed in order to do so.

CHAPTER 3: METHODOLOGY

3.1 WORKFLOW

In every step of the workflow, we have conducted some research to find the optimal method to do the task. There are lots of methods and techniques available in this field at this moment. So, we have selected the best approach and implemented it accordingly to solve the problem statement.

In this section, the main focus is on the instance segmentation model. We have used state of the art Mask R-CNN algorithm to solve the given problem. The problem statement has been solved by applying the transfer learning technique to the Mask R-CNN algorithm. A complete workflow diagram of the solution is shown in Figure.

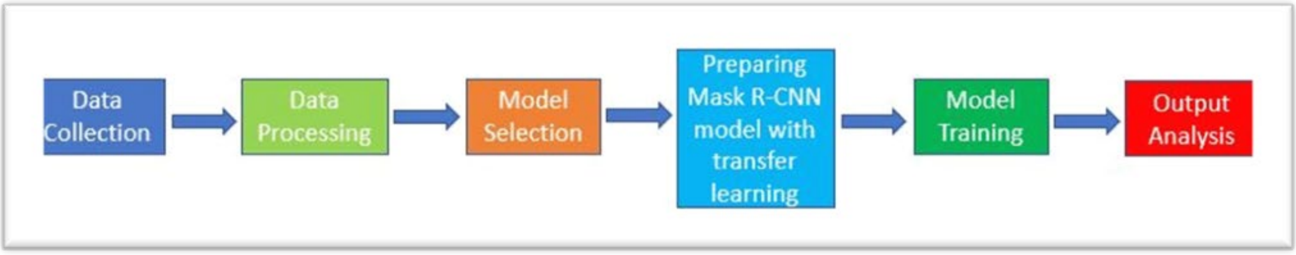


Figure 18. Overall Workflow

3.2 DATASET PREPROCESSING

The data set comprises only the road images with roughly 1300 samples. It contains different resolutions of pictures. The image samples were collected from the internet, from different persons, and manually by us.

The next step was to annotate the image samples to make training and a validation set. All of the image samples were annotated with an annotation tool's help to complete the annotation task. The annotation tool successfully marked the image's defective area, and each image needed multiple annotations to draw the different defect patterns. The total amount of image samples has been divided between two group members to annotate samples individually in the annotation time. The VGG image annotator version 2.0.10 has been used to complete this annotation task. After completing the annotation, all of the JSON files were integrated and formed the trainable dataset. The annotation demo is shown in the figure.

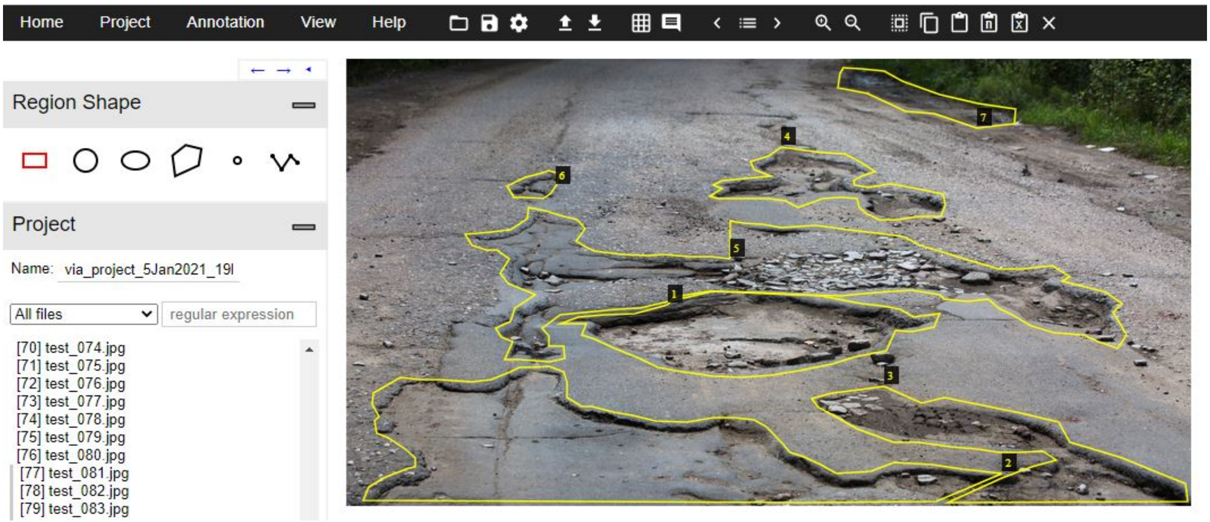


Figure 19. Annotating image with VGG

3.3 Neural Network Models

Many neural network models are working well for object annotation these days in this road defect detection problem. But the case of instance segmentation, very few models have a good performance. So, we have selected the state-of-the-art Mask R-CNN model with the technique of transfer learning to solve our problem.

3.3.1 Mask R-CNN

The Mask RCNN is a deep neural network algorithm designed to solve the instance segmentation problem in the field of deep learning. This algorithm's main ability is that it can separate various objects in a picture or a video. There are two phases of Mask RCNN. First of all, 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, which extracts the image samples feature maps. The network architecture has shown in the figure.

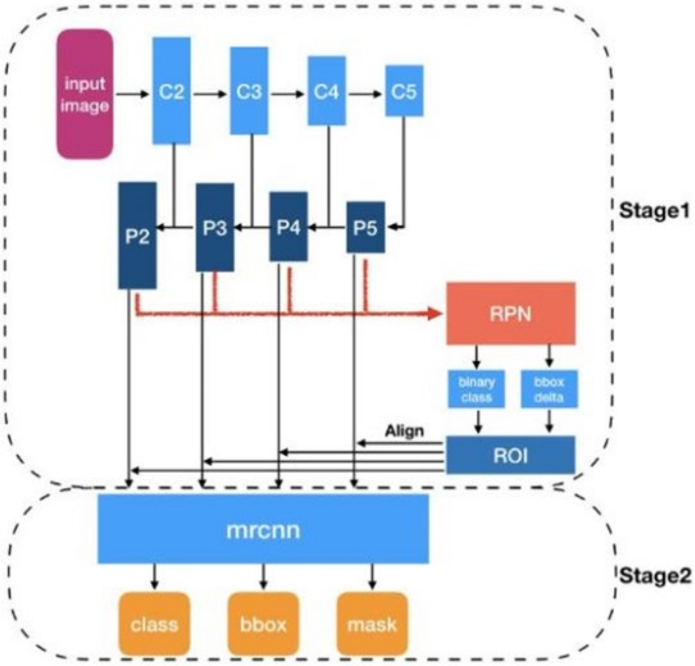


Figure 20. Mask R-CNN Network Architecture

3.3.1.1 Transfer Learning Implementation Details

The main idea is to take advantage of a state-of-the-art algorithm. Since we already have a very robust existing model, we have not designed any new model. The Mask R-CNN model, which was mainly developed for object detection, converted this model to road defect detection using the transfer learning technique. The abstract view of our idea has shown in the figure below.

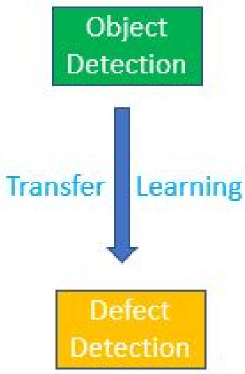


Figure 21. Object to Defect Detection Conversion Technique.

Now in the implementation, we have used the Detectron2 platform, which is built using the PyTorch framework. Since this Mask R-CNN algorithm doesn't have any model class, we had to override the class methods from the Detectron2 module, which contains the classes MetadataCatalog and DatasetCatalog. We have overridden the "register" and "get" methods from those two classes to register our custom dataset and set the class labels. After changing these settings, we have successfully implemented the transfer learning technique. Finally, the Mask R-CNN algorithm's prediction layer has been converted according to our custom dataset's requirement with a single class label, which was initially structured with the COCO datasets format with eighty class labels. The transformation of the implementation has shown below.

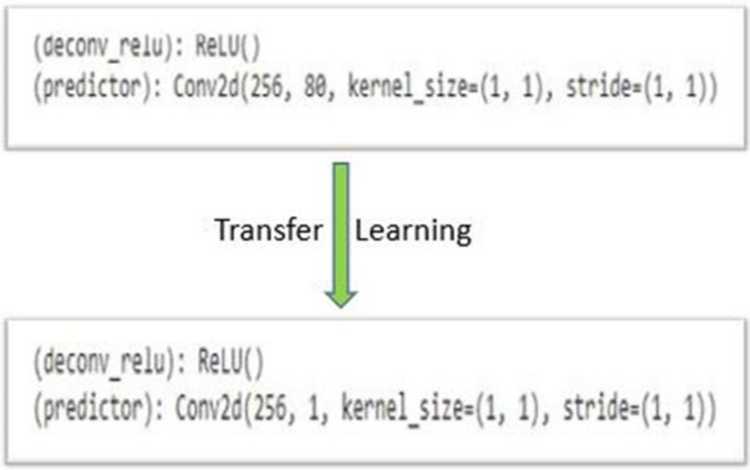


Figure 22. Implementation of Transfer Learning.

3.3.1.2 Mask R-CNN Training Details

We have used the Google Colab Pro with NVIDIA Tesla P100 GPU as our environment to train the model faster.

In the training time, we have directly imported the model from the Detectron2 model zoo and started the training using the COCO weights instead of training from scratch. We have kept the hyperparameters setting in default except for the epoch, batch size, backbone, and learning rate.

The training was conducted with different variations of backbone and batch size. They are ResNet-101-FPN backbone with 64 batch size, ResNet-101-FPN backbone with 32 batch size, ResNet-101-FPN backbone with 16 batch size, ResNet-50-FPN backbone with 16 batch size, ResNet-101-FPN backbone with 8 batch size, and ResNet-101-FPN backbone with 4 batch size. Among them, the ResNet-101-FPN with a batch size of 4 has produced the best outcome. The figure below shows the best training results among all the experiments.

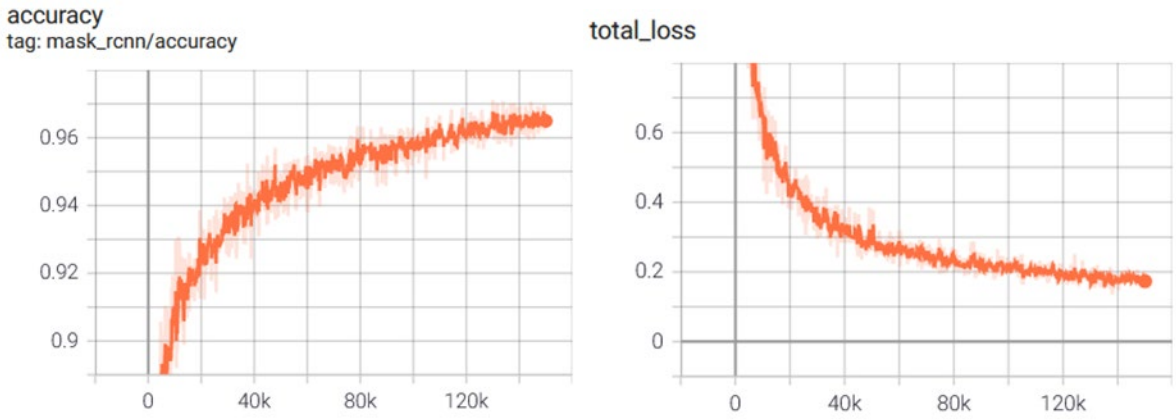


Figure 23. Training Accuracy and Loss Graph.

CHAPTER 4: RESULTS

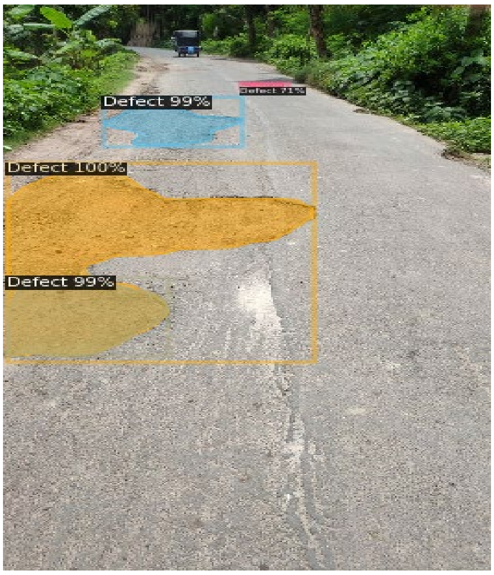
The Deep Learning model is nowadays generating excellent outputs in the field of AI. In our case, also, the result was satisfactory. In this section, the results of the two implemented models have been demonstrated.

4.1 INFERENCE ON IMAGES

In the time of inference, we have set the testing threshold to 70% and get the best outcome. Some test images have shown below.

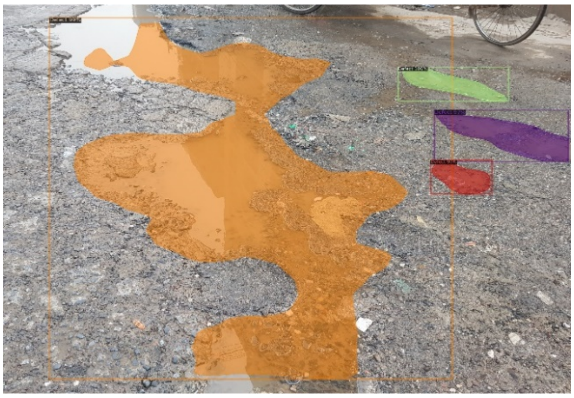
Base Image

After Annotation



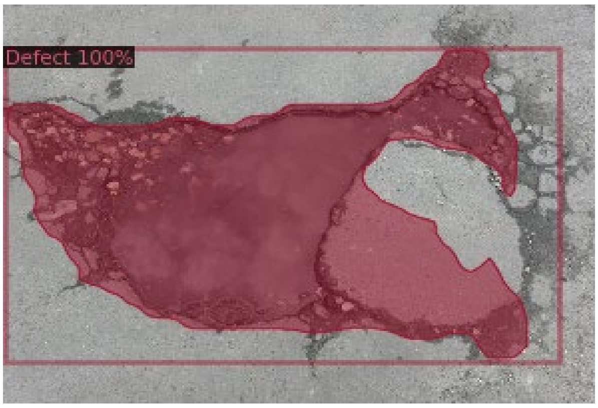
Base Image

After Annotation



Base Image

After Annotation

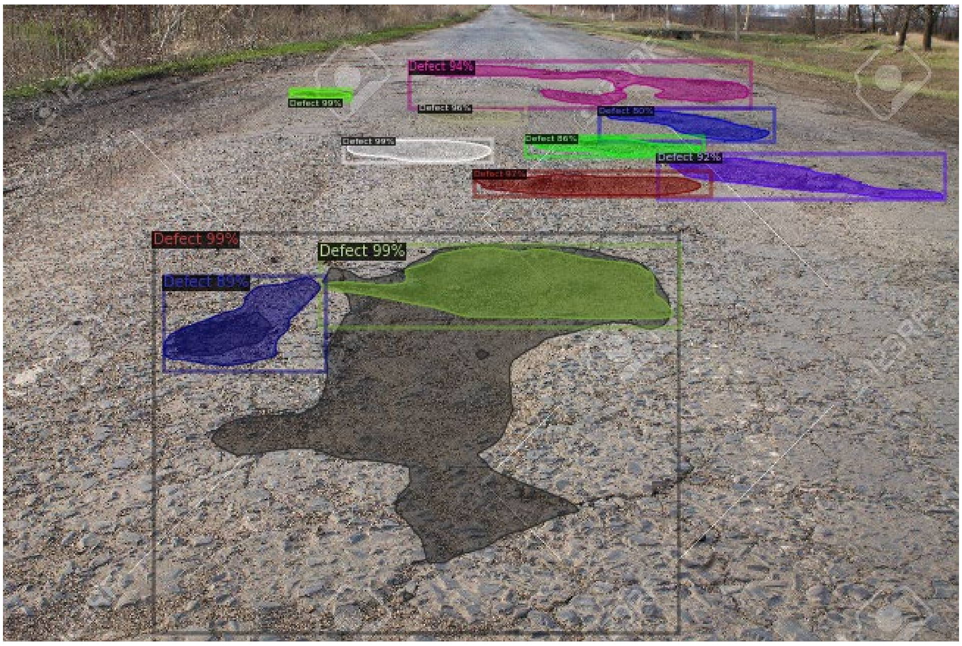


Base Image

After Annotation

Base Image

After Annotation



Base Image

After Annotation

Figure 24. Inference on Images

4.2 EVALUATION AND ANALYSIS

We have evaluated our testing results for different configurations using the AP matrix. The table below shows the results.

Batch size	**Backbone**	**AP**	**AP_{50}**	**AP_{75}**	**AP_{S}**	**AP_{M}**	**AP_{L}**
64	ResNet-101-FPN	25.700	50.884	21.666	12.453	24.968	29.970
32	ResNet-101-FPN	27.633	52.892	24.187	13.534	28.723	31.481
16	ResNet-101-FPN	27.120	52.725	22.992	12.958	27.004	31.596
8	ResNet-101-FPN	26.601	53.221	23.477	15.999	24.441	31.350
4	ResNet-101-FPN	**28.110**	**55.028**	**25.047**	**14.784**	**26.824**	**34.167**

Table 1. AP Matrix of Test Images

In the table above, the first rows bold titles indicate the batch size, model backbone, average precision, average precision for IOU threshold of 50%, average precision for IOU threshold of 75%, average precision for the small object, average precision for the medium object, and average precision for the large object.

The marked blue values are the maximum compared to other values for the small and large objects for batch size 32 and 8. Among the five different batch sizes, the batch size of 4 produces the overall best results those are indicated with green color.

CHAPTER 5: CONCLUSIONS

5.1 DISCUSSION

In this project, the main aim was to detect the broken parts of the roads. The segmentation result was generated based on the road image input. The implementation of this project started with binary classification, then gradually implemented the instance segmentation model.

The implementation of the solution of defect detection 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, computer resource management, discontinuous electricity, 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. The algorithm's only downside was its training cost. It requires a very powerful GPU and a long training time to produce the expected output. Once the training becomes complete, then it shows the impressive outcomes. Though the situation was adverse, the group member's collaboration and respected faculty's regular instruction made it possible to complete this project.

5.2 SUMMARY

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

State of the art Mask R-CNN model has been used to solve the defect detection problem. The algorithm was structured in a way so that it can predict the class, mask, bounding box, and confidence level at the same time. We have trained it with different batch sizes and backbone configurations. After finishing the training, we have tested the samples and calculated the AP matrix based on the test set. Among the different stages of experiments, this model worked very well on the small batch sizes. Finally, the Mask R-CNN model with transfer learning technique successfully managed the solution of the problem statement.

5.3 Future Work

In the future, training with more samples could be an important experiment, 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. To work with more data, increasing the computer's robustness by supporting more powerful hardware resources could be a big concern.

Some other algorithms also could be used to solve this problem by analyzing competitive 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 roads defect from live videos. Worldwide auto driven vehicles are using a similar kind of technology. So, defect detection from live videos is also possible with this project by some more improvement.

REFERENCES

[1] Pham, V., Pham, C., & Dang, T. (2020). Road Damage Detection and Classification with Detectron2 and Faster R-CNN. arXiv preprint arXiv:2010.15021.

[2] L. Ale, N. Zhang and L. Li, "Road Damage Detection Using RetinaNet," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5197-5200, doi: 10.1109/BigData.2018.8622025.

[3] 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.

[4] Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., & Omata, H. Generative adversarial network for road damage detection. Computer-Aided Civil and Infrastructure Engineering.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] Faramarzi, M. (2020). Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. Available at SSRN 3627382.

[12] 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.

[13] 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.

[14] Koch, C., & Brilakis, I. (2011). Pothole detection in asphalt pavement images. Advanced Engineering Informatics, 25(3), 507-515.

[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] Zhang, X., 2018. Simple Understanding Of Mask RCNN. [online] Medium. Available at: https://medium.com/@alittlepain833/simple-understanding-of-mask-rcnn-134b5b330e95 [Accessed 27 September 2020].

[22] 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).

[23] 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].

[24] 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].

[25] 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].

[26] Detectron2.readthedocs.io. 2019. Overview: Module Code — Detectron2 0.3 Documentation. [online] Available at: https://detectron2.readthedocs.io/_modules/ [Accessed 8 January 2021].

[27] Zhang E., Zhang Y. (2009) Average Precision. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_482 [Accessed 8 January 2021].

GitHub Link of the Research : https://github.com/sehab1611251/CSE-498R