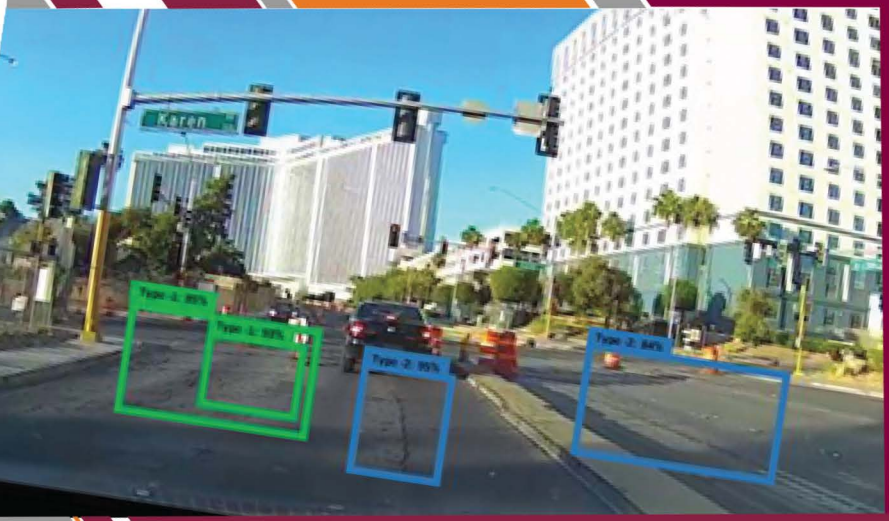
Detecting Pavement Distresses Using Crowdsourced Dashcam Camera Images

August 2021 | Final Report



SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

VT Transportation Institute
Virginia Tech

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. TTI-TTI-Student – 07	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Detecting Pavement Distresses Using Crowdsourced Dashcam Camera Images		5. Report Date May 2021 6. Performing Organization Code:
7. Author(s) Bahar Dadashova (TTI/TAMU)*) Chiara Silvestri Dobrovolny (TTI/TAMU) Mahmood Tabesh (TTI/TAMU-Student)		8. Performing Organization Report No. Report TTI-Student-07
9. Performing Organization Name and Address: Safe-D National UTC Texas A&M Transportation Institute		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/TTI-Student-07
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report 14. Sponsoring Agency Code
15. Supplementary Notes This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.		

16. Abstract

Pavements play a vital role in the transportation infrastructure in the United States. Damage to public road transportation infrastructure causes roadways to fail to perform as intended and increases crash risks. Road damage must be detected quickly and accurately in order to maintain roads and effectively allocate repair money. In this study, four deep-learning object detection models were used for detecting five types of pavement damages using Nexar Dashcam images. The single-shot multi-box detector (SSD) and faster region-based convolutional neural networks (Faster R-CNN) object detection models using MobileNet and Inception techniques were applied to the selected and processed images. An attempt was made to detect alligator cracking, longitudinal cracking,transverse cracking, and patching on the roadway surface using the developed model. A total number of 6,326 images with a total number of 8,970 damage instances were selected from the provided data. Faster R-CNN models showed higher recall and precision compared to the SSD models. Faster R-CNN Inception Resnet showed the highest average precision of 56%, followed by Faster R-CNN Inception with an average precision of 50%. Recall values showed almost the same trend as precision. Faster R-CNN showed the highest average recall of 43%. The results showed that Type-I and Type-II damages were detected with the highest recall and precision compared to other damages. Type-I damage was detected with an average precision of 81% and recall of 70 % in all of the models.

17. Key Words Pavement Damage Detection, Machine Vision, Object Detection, SSD, Faster RCNN, Dashcam video Image	18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 22	22. Price $0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

Pavements play a vital role in the transportation infrastructure in the United States. Damage to public road transportation infrastructure causes roadways to fail to perform as intended and increases crash risks. Road damage must be detected quickly and accurately in order to maintain roads and effectively allocate repair money. In this study, four deep-learning object detection models were used for detecting five types of pavement damages using Nexar Dashcam images. The single-shot multi-box detector (SSD) and faster region-based convolutional neural networks (Faster R-CNN) object detection models using MobileNet and Inception techniques were applied to the selected and processed images. An attempt was made to detect alligator cracking, longitudinal cracking, transverse cracking, and patching on the roadway surface using the developed model. A total number of 6,326 images with a total number of 8,970 damage instances were selected from the provided data. Faster R-CNN models showed higher recall and precision compared to the SSD models. Faster R-CNN Inception Resnet showed the highest average precision of 56%, followed by Faster R-CNN Inception with an average precision of 50%. Recall values showed almost the same trend as precision. Faster R-CNN showed the highest average recall of 43%. The results showed that Type-I and Type-II damages were detected with the highest recall and precision compared to other damages. Type-I damage was detected with an average precision of 81% and recall of 70 % in all of the models.

Acknowledgments

The authors acknowledge the help and feedback of Elinor Swery, Brandon Long, and Jake Halu of Nexar for participating in this project and providing access to their products. The authors acknowledge the help and feedback of Subasish Das for reviewing this report.

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

Table of Contents

List of Figures

List of Tables

Introduction

Pavements play a vital role in the transportation infrastructure in the U.S. According to the Bureau of Transportation Statistics (2019), about 80% of passenger and 68% of cargo transportation was via roads and highways. Road infrastructure must be regularly monitored and maintained to sustain and maximize its benefits. A 2015 Highway Performance Monitoring System report shows that 19% of federal-aid highways in the U.S. are in poor condition (Federal Highway Administration).

Poor road conditions can cause multiple problems for drivers, such as decreased driver comfort and increased crash risk and vehicle maintenance costs (Islam & Buttlar, 2012). Pavement distress has been found to be one of the influencing factors in increasing crash rates (Lee et al., 2015). Chan et al. (2010) used the historical crash data and pavement condition index from the pavement management system database in Tennessee to develop crash frequency prediction models. The authors found that pavement roughness can increase crash severity, particularly on high-speed roads (Lee et al., 2015). Elghriany et al. (2016) found a clear correlation between crash rate and pavement conditions, indicating that deteriorating pavement conditions can have a negative effect on traffic safety in a normal driving environment. Chen et al. (2019) studied the relationship between pavement roughness and the safety of different road types. The results indicate that the pavement roughness has a significant impact on the safety of multi-lane roadways. Overall, poor pavement surface was found to be associated with high numbers of total rear-end, same-direction sideswipe, and single-vehicle crashes (Lee et al., 2020).

Transportation agencies benefit from having an up-to-date road conditions database, as this allows them to maintain ideal road conditions, schedule timely road surface maintenance, improve transportation safety, and reduce transportation costs for vehicle owners. Traditionally, and still in most developing countries, pavement assessment is done manually by transportation agencies. Manually conducted field inspections are highly inefficient due to the time-consuming and labor-intensive processes necessary to conduct measurements, record information, and process data (Livneh, 1994). This approach is sensitive to the subjectivity/biases of the technicians who are responsible for carrying out these inspections (Guan et al., 2015).

Furthermore, it exposes the inspectors to dangerous working conditions on highways (Ragnoli et al., 2018). Thus, automatic road damage recognition and classification techniques have become of increasing interest to transportation agencies. Most of the previous research has focused on detecting and classifying images for a limited number of defect types. Also, much of the research has used high-quality images captured perpendicularly from the surface of the road using special equipment. Collecting this type of high-resolution imagery data can be costly and is therefore generally limited to a few sites.

Emerging innovations such as dashboard cameras, or dashcams, have made it easier to capture high-quality data from cars. Dashcams installed on vehicles have a great potential for detecting pavement damage. They can cover larger areas and do not require special equipment. However,

this approach has not been sufficiently explored in the literature. Applying object detection techniques on crowdsourced dashcam imaginary data would open the opportunity to use these techniques over a wider area and at a lower cost than using dedicated vehicles. This project addresses these limitations and explores the potential of detecting pavement distress using crowdsourced imagery data.

In this project, the research team used Nexar Dashcam (Nexar Inc, 2020) images to develop deep learning (DL) models to detect five major pavementdamage types: (Type I) Alligator Cracking, (Type Longitudinal Cracking, (Type III) Transverse Cracking, (Type IV) Patching, and (Type V) Potholes. The researchers used the data collected in Las Vegas, Nevada, in May 2020 from work zones since the pavement in work zones is assumed to have many complications and cracks, which can help to improve the image detection task. Each image was labeled manually for the presence of pavement defects by specifying the type and location of the distress. Two object detection models—single shot multi-box detector (SSD) Inception V2 and faster region-based convolutional neural networks (R-CNN) Inception V2 from TensorFlow object detection API (TensorFlowAPI 2021)—were then applied to identify the pavement distress types. The results of this project can be used for developing a system to detect pavement damage in real-time and to create a database of roadway distresses. The results show that Type-I and Type- II damages were detected with the highest recall and precision compared to other damages. Type-I damage was detected with an average precision of 81% and recall of 70% in all of the models. Type-II damage average precision and recall were 79% and 49%. Type-III damage had intermediate average precision and recall equal to 47% and 45%. Type-IV and Type-V damage had the lowest average precision and recall at less than 20%.

The rest of this report is organized as follows:

1) **Background**: discusses the available literature on pavement distress data collection methods and DL techniques for detecting pavement distress.
2) **Method**: describes the object detection algorithms and feature extraction modules.
3) **Empirical Study**: describes data sources and image data processing methods used in the study.
4) **Results and Discussion**: introduces the experiment set-up and presents the evaluationresults.
5) **Conclusions and Recommendations**: presents the conclusions and recommendations for future research

Background

Pavement Distress Data Collection Methods

Transportation agencies benefit from having an up-to-date road conditions database, which allows them to maintain optimum road conditions, schedule timely road surface maintenance, improve transportation comfort and safety, and lower transportation costs for vehicles by reducing vehicle maintenance needs. Many attempts have been made to inspect asphalt road conditions effectively in response to the demand for robust tools to support road damage inspection tasks. Pavement evaluation has traditionally been performed manually by transportation authorities. Manual field inspections are highly inefficient due to the time-consuming and labor-intensive processes necessary to take measurements, record information, and process data (Livneh, 1994). Furthermore, manual inspections are susceptible to the subjectivity and biases of the technicians in charge of these inspections (Guan et al. 2015). Technicians are also put in dangerous working conditions on highways (Ragnoli et al., 2018). As a result, transportation authorities are becoming increasingly interested in automated road damage identification and classification techniques.

To address the limitations of the subjective visual assessment process, many attempts have been made to create automated devices for pavement distress acquisition and detection. These revolutionary approaches provide objective, highly efficient, and reliable data collection, allowing for the development of quantitative pavement analysis. Automated approaches are also cost-effective and can help reduce field inspection costs considerably. The latest systems consist of one or more acquisition devices and post-processing software for automatic data extraction procedures using computer vision and image processing algorithms (Coenen & Golroo 2017). Pavement damage detection employs cutting-edge technology such as digital cameras, line scan cameras, 3D laser imaging, and terrestrial laser scanners (Ragnoli et al., 2018), the latter two methods enabling detection of road surface damages in three dimensions (3-D) with high accuracy (Hou et al. 2007, Li et al., 2009, Yu & Salari, 2011).

Another system that is widely used for detecting distress is a digital camera, which captures two-dimensional (2-D) photographs from the pavement's surface. Changes in image processing techniques have resulted in major improvements in automated pavement crack detection models in recent years. In these models, 2-D images of road pavement are analyzed precisely to extract and determine useful information regarding existing cracks, as well as to label relevant images for further analysis. Most studies based on 2-D images have achieved over 95% accuracy in the identification and classification of cracks, which is a considerable success in pavement defect detection (Zakeri et al., 2017). Despite these achievements, digital cameras have been restricted in practice because it is difficult to take high quality images from an entire section of pavement due to illumination and speed variations, and it is expensive to operate dedicated vehicles to capture new road picture data on a regular basis. Moreover, the photos gathered are generally insufficient to cover the entire road segment (Cao et al., 2020). Dashcams installed on vehicles can be useful sources of information for detecting pavement damage over the entire roadway segment. Being

able to use these images in pavement distress detection models effectively will have a huge impact on the role of maintenance. Some dashcam producers have already begun to create a picture and video archive from the captured data (Nexar Inc. 2020). These archives are already huge and can be used for producing insights from roadways. In addition, installing dashcams is cheap and feasible for any type of car. Since the dashcam image is captured randomly from different roadways and is geospatially tagged, a comprehensive database of the damages from the entire pavement network, including low traffic and minor arterials, can be developed. One caveat is that the roadway surface images may include other objects in addition to the cracks on the pavement, which could complicate roadway damage detection and classification (Cao et al. 2020). This limitation can be addressed using DL tools, as described in the following section.

Deep Learning Methods for Pavement Image Analysis

Recent improvements in the capability of processing units have led image processing techniques to be commonly used for damage detection in civil engineering (Maeda et al., 2018). Machine learning methods, including artificial neural network (ANN), support vector machine (SVM), classification and regression tree (CART), and DL-based image processing methods, are primarily used in published studies on detecting and classifying road damage (Ouma & Hahn 2016, Hoang, 2018, Butcher et al., 2014, Hoang et al., 2018). Following the extraction of features from images using image processing techniques, machine learning-based approaches are used to determine if these features represent a particular form of damage (Cha et al., 2017). The various distress forms are segregated and categorized based on a number of factors, including the width, propagation direction, and pixel color variation.

DL-based models have recently been established as the most effective method for detecting features in images of structural concrete cracks and road damage (Jeong 2020, Cha et al., 2017). Among the DL algorithms, Faster Region-based Convolutional Neural Networks (Faster RCNN) (Ren et al., 2017), Region-based fully Convolutional Networks (R- FCN), Single Short Multi-box Detector (SSD) (Liu et al., 2016), and You Only Look Once (YOLO) (Redmon et al., 2016) methods have dramatically improved object detection accuracy. Integrating these algorithms with feature extraction systems such as Inception (Ioffe & Szegedy, 2015), Residual Networks (ResNet) (He et al., 2016), and MobileNet (Howard et al., 2017) enhances the object detection accuracy and speed. DL-based models have been used to classify road damage in a number of studies with impressive results. To recognize cracks/non-cracks in road pictures in a study by Zhang et al. (2017), a CNN model was built using 500 images of 3,264 x 2,448 pixels, reaching an accuracy of 90%. Yusof et al. (2019), using CNN filters, classified images based on the existing type of distress with an accuracy of 94.5%.

Object detection in roadway images is difficult and time-consuming, necessitating the use of state-of-the-art DL methods. Jo and Ryu (2015) created a new pothole detection algorithm that works with the embedded computing environments of black-box cameras and allows for real-time pothole detection. Nienaber et al. (2015) applied similar shapes geometry, collinear point

investigation, and camera matrix calibration methods on dashcam images for detecting potholes on the road surface. Maeda et al. (2018) used SSD Inception V2 and SSD MobileNet for identifying eight types of damages with a precision of 77% and recall of 71% from images captured by a smartphone installed on a dashboard. Chun and Ryu (2019) applied modified dashcam image database by segmentation and augmentation techniques and used a Fully Convolutional Network for detecting road damage. A combination of Faster R-CNN and Structured Random Forest Edge Detection (SRFED) was usedby Kalfarisi et al. (2020) in detecting cracking on structural surfaces. Crack, road marking, and intact area detection was performed using a combination of segmentation and classification modules by Park et al. (2019). Several other studies used Fully Convolutional Networks (FCNN) in the crack segmentation and detection process (Huang et al., 2018, Bang et al., 2019, Dung & Anh, 2019).

Method

This section describes the main components of object detection models. Two main object detection algorithms—Faster-RCNN and SSD—are described, and feature extraction paradigms of MobileNet, Inception, and Resnet are introduced. Following that is a description of the image processing process and experiment set-up.

Object Detection Algorithms

Single Shot Multi-Box detection (SSD)

The SSD model, first introduced by Liu et al. (2016), is one of the object detection algorithms that performs relatively quickly with high accuracy. Initial instances of the bounding boxes are processed by a feed-forward CNN and the final detection is optimized using a non-maximum suppression method.

A base network and a sequence of multiscale feature blocks are the core components of the model. The base network generates features of the image through a deep CNN. The feature matrix achieves a higher depth and lower size by going deep into the network. Shallow layers are associated with smaller objects, and deeper layers are associated with bigger objects. The SSD model is a multiscale-scale object detection model that can incorporate the ResNet and MobileNet algorithms. SSD defines anchor boxes at each level of the feature layers and compares them with the ground truth boxes in each image. If the corresponding default boundary box has an intersection over union greater than 0.5 with the ground truth, it will be fed into the loss function. The objective function in the SSD model is a combination of the localization loss and confidence loss presented in Equation (1). The mismatch between the ground truth box and the anticipated boundary box is known as the localization loss. Class prediction in each default boundary box can be scored by the confidence loss. The losses for each positive match prediction were penalized based on the confidence level of the relevant class.

$$LL = \frac{1}{NN}\left(LL_{cccccccc} + \alpha\alpha\ LL_{llcccc} \right) \quad (1)$$

Where L is the total loss, L_{loc} is a summation of the location loss of all default boundary boxes, L_{Conf} is the summation of the softmax loss over multiple classes, N is the number of matched default boxes, and α is the weight factor for the localization loss.

Faster Region-based Convolutional Neural Networks (Faster R-CNN)

Generally, Faster R-CNN (see Figure 1) is composed of two parts: (1) a region proposal network (RPN) for producing region suggestions and (2) a network for detecting objects utilizing these proposals (Ren et al., 2017). The primary difference between Fast R-CNN and this method is that the latter employs selective search to generate region suggestions. RPN has a substantially lower time cost for generating region recommendations than selective search, and RPN shares the most processing with the object detection network. RPN ranks region boxes (also known as anchors) and suggests the ones that are most likely to contain target objects. An RPN's output is a collection of boxes/proposals that will be analyzed by a classifier and regressor to check for the presence of objects.

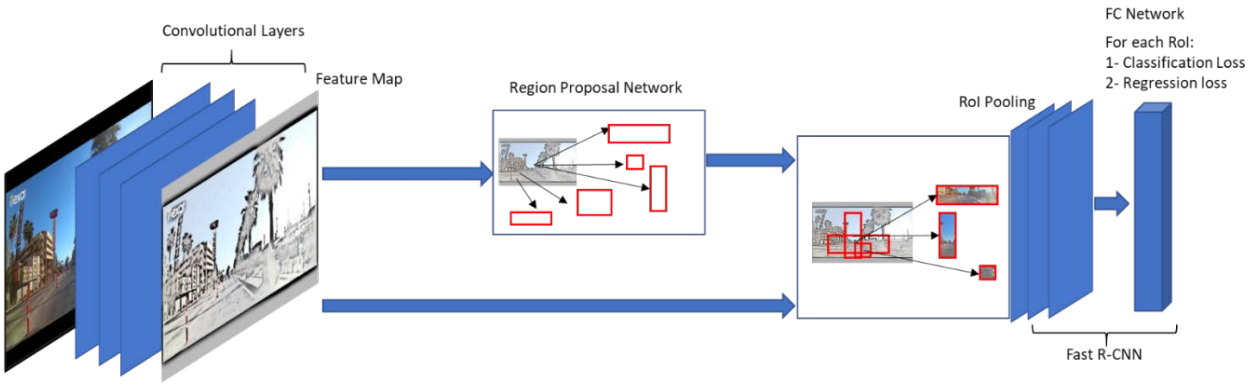


Figure 1. Architecture of a Faster R-CNN object detection model.

The RPN's overall loss is made up of both the classification and regression losses. The classification loss is associated with the detection of the object in the anchors and the regression loss is associated with the location of the anchors relative to the ground truth region. RPN provides suggested regions of various sizes. RPN's box suggestions are used to crop areas from the feature map. Items are then inputted into the feature extractor to select a class and refine the anchors' boxes in those regions. This second-stage box classifier's loss function has the same format as the RPN part. The computation cost of this stage highly depends on the number of RPN-proposed regions. Faster R-CNN has an almost lower speed compared to the alternative object detection models; however, it gives more accurate results. As a result, the performance of the Faster R-CNN detector on the road damage detection task is worth exploring.

Feature Extraction Modules

High-level convolutional extractors can be used with object detection algorithms to retrieve high-level features. Inception, ResNet, and MobileNet are the CNNs that are commonly used for an image classification task.

Inception

Inception was built with the goal of minimizing the computing cost of deep neural networks while will achieving high performance (Szegedy et al., 2015). Researchers have developed different versions of Inception that can also be combined with other feature extraction algorithms. The Inception family of feature extractors is based on the increasing width of the feature map. The Inception module runs numerous transformations in parallel on the same input map and then combines the results into a single output. Each layer of the input does a 1x1, a 5x5, and a 3x3 convolution, and a max-pooling, which is concatenated in the output of the Inception. Thus, the model has different levels of information from the image that can be used in further object detection algorithms. Inception quickly established itself as a model architectural standard. Some versions of the Inception are benefitted from batch-normalization-based modifications,elimination of Dropouts, and increased learning rates (Szegedy et al., 2016).

Residual Network (ResNet)

While Inception focuses on computational cost, ResNet emphasizes computational accuracy (He et al., 2016). Generally, increasing the depth of the network decreases the accuracy of the prediction. Deep networks suffer from a phenomenon known as vanishing gradient. Weights of the earlier layers in deep networks cannot be trained well because the effect of the gradient will perish in the back propagation stage of the training. The ResNet algorithm can help reduce the effect of the vanishing gradient using the concept that direct mappings are hard to learn. In this algorithm, weights of the earlier layers will be combined with the output of the deeper layers and fed into the next module. The gradient signal can flow back to early layers using this shortcut mechanism in ResNet, allowing for the creation of many layers of the network without sacrificing accuracy.

MobileNet

The goal of MobileNet is to construct lightweight deep neural networks by using depth-wise separable convolutions (Howard et al., 2017). MobileNet increases the speed of the feature extraction by applying a depthwise convolution and a pointwise convolution. In a standard convolution, the convolution kernel is applied to all the channels of the input image by making a weighted sum of the input pixels with the filter and then sliding to the next input pixels across the pictures. In depth-wise convolution, a discrete operation is performed on each channel; thus, the number of output layers is the same as the number of input channels. The pointwise convolution is the following phase, which is similar to a conventional convolution but with a 1x1 filter. The goal of pointwise convolution is to combine the depthwise convolution's output channels to create additional features. This approach lowers both the computing cost and the number of parameters.

Empirical Study

Nexar Images

This section describes the dataset used for pavement damage detection, preprocessing, and the annotation process performed before model training. The image data for this study was provided by Nexar (Nexar Inc. 2020). The images were captured from Nexar users' dashcams. A total of 18,962 images, sized at 1,280 × 720 pixels, were collected randomly from Las Vegas, NV, in May 2020. The images were taken at 2-minute intervals at different vehicle speeds. Image locations cover various road types, road configurations, lighting, pavement conditions, traffic loads, and speed limits. This prevents the developed model from overfitting a special type of road and/or condition.

Pavement Distress Categories

Figure 2 presents the damage types that were identified in the damage detection models. In this project, each damage type is represented with a class name:

1) Type-I pavement damage is top-down fatigue cracking (Alligator Cracking), which is the most common distress in asphalt pavements. This type of damage size can vary from 6 inches to 3 feet across. The size of damage that is discernible in the dashcam images has medium to high severity.

2) Type-II damage is longitudinal cracking that propagates parallel to the pavement centerline. Generally, high severity cracks with crack openings greater than a half-inch can be discerned in the images.

3) Type-III damage is transverse cracking that propagates perpendicular to the centerline of the roadway.

4) Type-IV consists of patching areas that have a high potential for crack initiation and distress growing and should be monitored regularly.

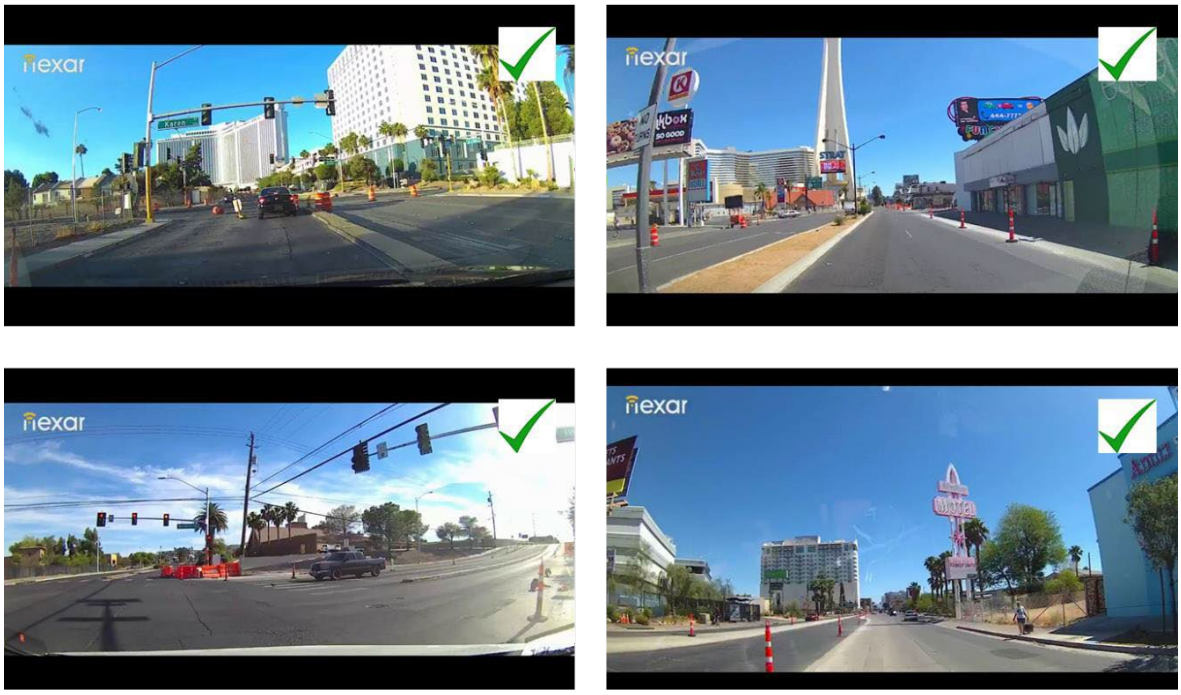
5) Type-V damage is a pothole, and is a common damage in asphalt and concrete pavements. Potholes are holes in the roadway that vary in size and shape and are the most dangerous distress and lead to the biggest increase in crash risks, especially on high-speed roads.

Figure 2. Pavement damage types, Type-I: Alligator Cracking, Type-II) Longitudinal Cracking, Type-III) Transverse Cracking, Type-IV) Patching, Type-IV) Potholes.

Image Review and Annotation

Images were taken from several dashcam cameras at different speeds. Some of the images were completely blurred, and some did not have a section of the road. Also, some of the images had very weak illumination that was not suitable for model training. In this step, the provided images were investigated, and unusable images were discarded. Figure 3 shows examples of acceptable and not acceptable images for damage detection purposes.

a) Discarded Images



b) Acceptable Images

Figure 3. Examples of the acceptable and discarded images for roadway damage detection.

After reviewing the images, the accepted images were annotated using the open-source application LabelImg V 1.8.5 (LabelImg, 2018). This application is a graphical image annotation tool written in Python with a graphical interface. Instances of damage types and the surrounding bounding box were determined manually. LabelImg saved annotations as XML files in PASCAL VOC format. Figure 4 shows the manual annotation of the images using LabelImg.

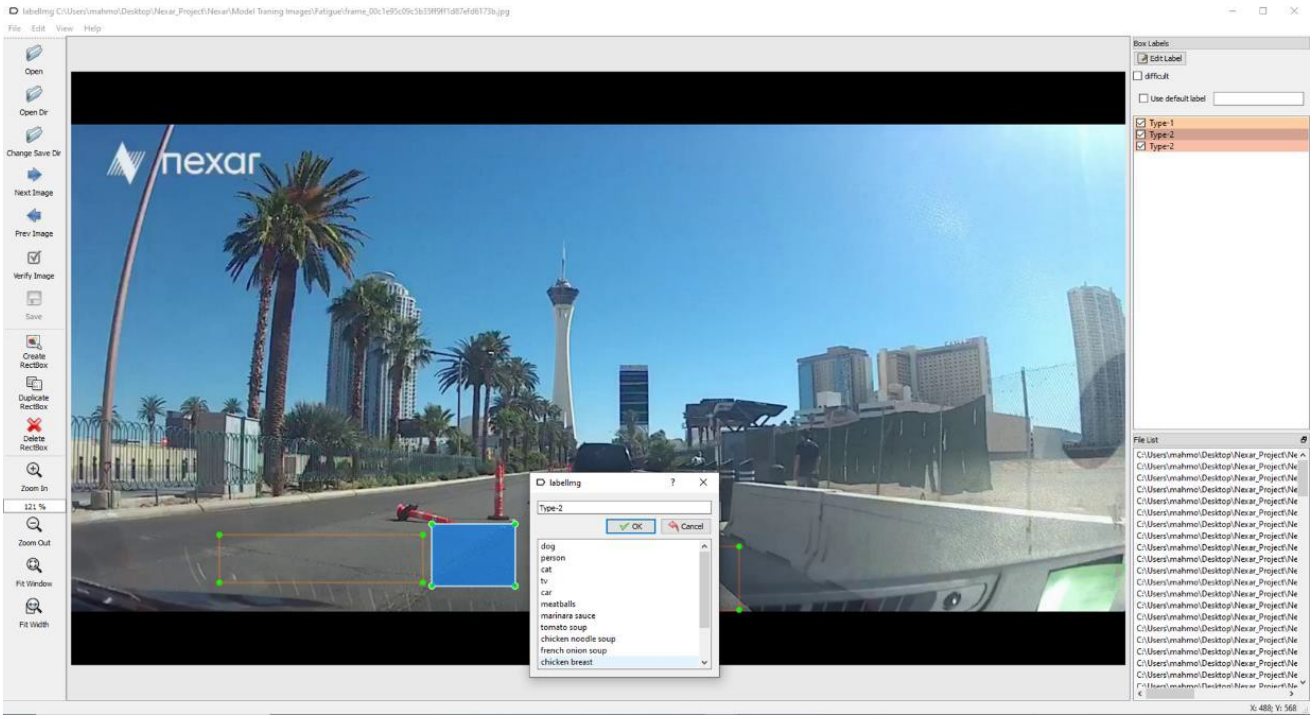


Figure 4. Data annotation using LabelImg.

The total number of damage instances is presented in Table 1. The number of instances for alligator Cracking (2,123) and longitudinal Cracking (5,362) were enough for training a damage detection model; however, potholes (112) were very rare in the provided images.

Table 1. Road Damage Types and Instances

Class Name	Damage Type	Number of Instances
Type-I	Alligator Cracking	2,123
Type-II	Longitudinal Cracking	5,362
Type-III	Transverse Cracking	650
Type-IV	Patching	723
Type-V	Potholes	112

Results and Discussion

In this study, a Faster R-CNN model combined with Inception V2 and Inception Resnet and an SSD model combined with MobileNet V2 and Resnet were trained, and the results were compared. A PC running OS Windows 10 with an Intel Core i7 CPU @ 1.8 GHz, 16 GB RAM, 256 GB SSD memory, and NVIDIA® GeForce RTX™ 3070 8GB GDDR6 were used for model training. Models were run on Google Colab IDE in a TensorFlow environment.

Model Parameters

The parameters in the object detection system utilizing DL that need to be tuned are large. There are important parameters that need to be selected carefully. The parameters that were selected in each model are as follows.

- **SSD**: the initial learning rate was 0.003 with a decay rate of 0.9 per 10,000 iterations. The batch size was set to 12. The input image size was maintained at the original 1,280 × 720. The non-linear activation function used was ReLU6. Initial weights were selected with a truncated normal distribution having a standard deviation of 0.03 and a mean of 0.

- **Faster R-CNN**: the batch normalization was applied on all of the feature extractors after the convolutional layers. The initial learning rate was 0.0002 with a decay rate of 0.9 per 10,000 iterations.

Training and Evaluation

The training and evaluation were conducted using Nexar dashcam images. A total number of 6,326 images with a total number of 8,970 damage instances were selected from the provided data. The training data were selected randomly at a ratio of 8:2. The rest of the images were used as an evaluation dataset. Also, 50% of the training data were flipped horizontally for data augmentation. In the evaluation, the intersection over union threshold was set to 0.5.

Recall and precision were used as evaluation metrics for the training models.

Evaluation Results

Figure 5 depicts a sample of detected damage types using the models described above. Different colors were used to represent the damage types. As observed, each pavement damage was detected separately in the provided images.

Figure 5. Examples of pavement damage detection in Nexar images.

Table 2 presents the detection and evaluation results for each class and training model. All the presented recall and precisions are at an intersection over union of 0.5. Faster R-CNN models show higher recall and precision than the SSD models regardless of the type of damage. Faster R-CNN Inception Resnet shows the highest average precision of 56%, followed by Faster R-CNN Inception with an average precision of 50%. The SSD model shows lower average precision of approximately 40%.

SSD mobileNet shows an average precision of 43%, and SSD Inception shows an average precision of 41%. Recall values show almost the same trend as precision. Faster R-CNN shows the highest average recall of 43%. Faster R-CNN Inception shows the second-highest recall of 41%. The average recall for the SSD model was approximately 34%.

The results show that Type-I and Type-II damages were detected with the highest recall and precision compared to other damages. Type-I damage was detected with an average precision of 81% and recall of 70% in all of the models. Type-II damage average precision and recalls were 79% and 49%. Type-III damage had intermediate average precision and recalls equal to 47% and 45%. Type-IV and Type-V damage had the lowest average precision and recall of less than 20%.

Table 2. Detailed Performance of Models (Intersection Over Union = 0.5) for Each Damage Type

Model	Metrics	Type -I	Type-II	Type-III	Type-IV	Type-V	Average
SSD InceptionV2	Recall	0.62	0.52	0.43	0.21	0.02	0.36
SSD InceptionV2	Precision	0.73	0.82	0.34	0.07	0.12	0.41
SSD MobileNet	Recall	0.67	0.42	0.38	0.15	0.05	0.33
SSD MobileNet	Precision	0.78	0.72	0.41	0.11	0.16	0.43
Faster R-CNN Inception	Recall	0.77	0.53	0.45	0.27	0.07	0.41
Faster R-CNN Inception	Precision	0.86	0.84	0.53	0.18	0.13	0.50
Faster R-CNN InceptionResnet	Recall	0.75	0.51	0.56	0.23	0.13	0.43
Faster R-CNN InceptionResnet	Precision	0.87	0.79	0.61	0.32	0.25	0.56

The operation speed of each model is presented in Figure 6. The speeds were recorded from the same PC with the same configuration. The SSD MobileNet had the highest speed among the object detection models. SSD Inception V2 was almost two times slower than SSD MobileNet. The Faster-RCNN models had the lowest operation speed. The Inception Resnet was is 1.5 times slower than the Inception algorithm in the Faster-RCNN model.

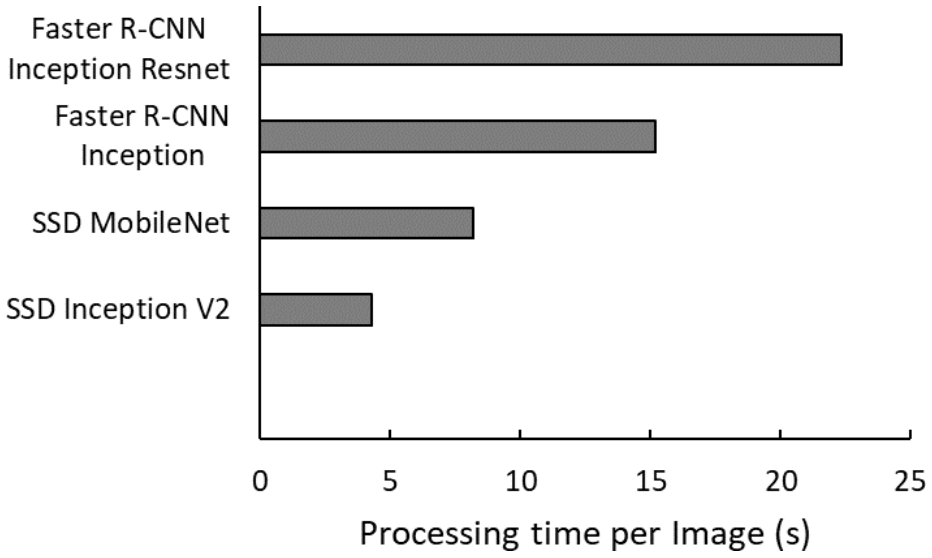


Figure 6. Processing time of detection models.

Finally, an error analysis was conducted on false positive (FP) and false negative (FN) instances. Examples of FP are shown in Figure 7. In some cases, the windshield stickers and reflections were the reason for the FP instances (Figure 7(a)). There were also instances where the side curb and their edges were detected as pavement damages (Figure 7(b)). In Figure 7(c) and Figure 7(d), a utility hole and a steel road plate were detected as pavement distress. In some cases, shadows on the ground were the source of the FN error. FP cases are defined as instances of pavement damage that the model was not able to detect. Patching was not detected in many cases. This could be because of the low number of patching instances in the training dataset.

Figure 7. Sample error of detection models.

Figure 7 (cont.). Sample error of detection models.

Conclusions and Recommendations

Pavements play a vital role in the transportation infrastructure in the U.S. Damage to public road transportation infrastructure causes roadways to fail to perform as intended and adds to traffic accidents. Road damage must be detected quickly and accurately in order to maintain roads and effectively allocate repair money.

Existing methods such as manual data collection and image data collection from special equipment have some limitations, in that they are very costly and can only collect data from a small sample of roadways. Dashcams installed on vehicles have a great potential for detecting pavement damage in a more cost-effective way. However, images collected from dashcams are low-quality and are subject to many errors. In this study, the research team applied four deep-learning object detection

models for detecting five types of pavement damages using Nexar Dashcam images. The SSD and Faster R-CNN object detection models using MobileNet and Inception techniques were applied to the selected and processed images.

Alligator cracking, longitudinal cracking, transverse cracking, patching, and potholes were detected on the roadway surface using the developed model. A total number of 6,326 images with a total number of 8,970 damage instances were selected from the provided data. Faster R-CNN models showed higher recall and precision compared to the SSD models. Faster R-CNN Inception Resnet showed the highest average precision of 56%, followed by Faster R-CNN Inception with an average precision of 50%. Recall values show almost the same trend as precision. Faster R-CNN shows the highest average recall of 43%. The results show that Type-I and Type-II damages were detected with the highest recall and precision compared to other damages. Type-I damage was detected with an average precision of 81% and recall of 70% in all of the models. Other types of damages were also detected with a lower precision. This could be due to the relatively small sample of images used for training. The object detection models show acceptable performance in detecting pavement damages. The developed models can be used in real-time damage detection models that can be implemented in Dashcam video cameras and improve drivers' safety.

The methodology and data used in this study have the potential to be improved by increasing the sample size of image data. This approach can be used by transportation agencies to detect pavement damages in a cost-effective way, without requiring manual labor or collecting image data with specialized equipment. By detecting the pavement damage from the widely-used dashcam images, transportation agencies can improve roadway conditions and thereby improve overall driver safety comfort.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project can be downloaded from the project page on the Safe-D website. The final project dataset is located on the Safe-D Dataverse.

Education and Workforce Development Products

This project resulted in one scientific paper, which will be submitted to Transportation Research Record to be presented at the Annual Meeting of Transportation Research Board.

Data Products

Data used in this study will be publicly available after the research team obtains permission from Nexar Inc.

References

1. Bang, S., S. Park, H. Kim and H. Kim. Encoder–decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering,* 34(8), 713-727, 2019.

2. Butcher, J. B., C. R. Day, J. C. Austin, P. W. Haycock, D. Verstraeten and B. Schrauwen. Defect Detection in Reinforced Concrete Using Random Neural Architectures. *Computer-Aided Civil and Infrastructure Engineering,* 29(3), 191-207, 2014.

3. Cao, M.-T., Q.-V. Tran, N.-M. Nguyen and K.-T. Chang. Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. *Advanced Engineering Informatics, 46*, 101182, 2020.

4. Cha, Y.-J., W. Choi and O. Büyüköztürk. Deep Learning-Based Crack Damage DetectionUsing Convolutional Neural Networks. *Computer-Aided Civil and Infrastructure Engineering,* 32(5), 361-378, 2017.

5. Chan, C. Y., B. Huang, X. Yan and S. Richards. Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS). *Journal of Advanced Transportation,* 44(3), 150-161, 2010.

6. Chen, S., T. U. Saeed, S. D. Alqadhi and S. Labi. Safety impacts of pavement surface roughness at two-lane and multi-lane highways: accounting for heterogeneity and seemingly unrelated correlation across crash severities. *Transportmetrica A: TransportScience,* 15(1), 18-33, 2019.

7. Chun, C. and S.-K. Ryu. Road Surface Damage Detection Using Fully Convolutional Neural Networks and Semi-Supervised Learning. *Sensors (Basel, Switzerland), 19*(24), 5501, 2019.

8. Coenen, T. B. J. and A. Golroo. A review on automated pavement distress detectionmethods. *Cogent Engineering, 4*(1), 1374822, 2017.

9. Dung, C. V. and L. D. Anh. Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction,* 99, 52-58, 2019.

10. Elghriany, A., P. Yi, P. Liu and Q. Yu. Investigation of the effect of pavement roughness on crash rates for rigid pavement. *Journal of Transportation Safety & Security, 8*(2), 164-176, 2016.

11. Federal Highway Administration. *Status of the Nation's highways, bridges, and Transit: Conditions & Perfomance.* Washington, DC: Federal Highway Administration, 2015.

12. Guan, H., J. Li, Y. Yu, M. Chapman, H. Wang, C. Wang and R. Zhai. Iterative TensorVoting for Pavement Crack Extraction Using Mobile Laser Scanning Data. *IEEE Transactions on Geoscience and Remote Sensing, 53*(3), 1527-1537, 2015.

13. He, K., X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition. ed. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30, 770-778, June 2016.

14. Hoang, N.-D.. An Artificial Intelligence Method for Asphalt Pavement Pothole Detection Using Least Squares Support Vector Machine and Neural Network with Steerable Filter-Based Feature Extraction. *Advances in Civil Engineering,* 7419058, 2018,.

15. Hoang, N.-D., Q.-L. Nguyen and D. Tien Bui. Image Processing–Based Classification of Asphalt Pavement Cracks Using Support Vector Machine Optimized by Artificial Bee Colony. *Journal of Computing in Civil Engineering,* 32(5), 04018037, 2018.

16. Hou, Z., K. C. P. Wang and W. Gong,. Experimentation of 3D Pavement Imaging through Stereovision. *International Conference on Transportation Engineering 2007.* 376-381, 2007.

17. Howard, A. G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861,* 2017.

18. Huang, H.-w., Q.-t. Li and D.-m. Zhang. Deep learning based image recognition for crackand leakage defects of metro shield tunnel. *Tunnelling and Underground Space Technology, 77*, 166-176, 2018.

19. Ioffe, S. and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. ed. *International conference on machine learning*, 448-456, 2015.

20. Islam, S. and W. G. Buttlar. Effect of Pavement Roughness on User Costs. *Transportation Research Record, 2285*(1), 47-55, 2012.

21. Jeong, D., Road Damage Detection Using YOLO with Smartphone Images. ed. *2020 IEEE International Conference on Big Data (Big Data)*, 5559-5562, 10-13 Dec. 2020.

22. Jo, Y. and S. Ryu. Pothole Detection System Using a Black-box Camera. *Sensors, 15*(11), 29316-2933, 20151.

23. Kalfarisi, R., Y. Wu Zheng and K. Soh. Crack Detection and Segmentation Using Deep Learning with 3D Reality Mesh Model for Quantitative Assessment and Integrated Visualization. *Journal of Computing in Civil Engineering, 34*(3), 04020010, 2020.

24. LabelImg, 2018. [online]. Available from: https://github.com/tzutalin/labelImg.

25. Lee, J., M. Abdel-Aty and E. Nyame-Baafi. Investigating the Effects of Pavement Roughness on Freeway Safety using Data from Five States. *Transportation ResearchRecord, 2674*(2), 127-134, 2020.

26. Lee, J., B. Nam and M. Abdel-Aty. Effects of Pavement Surface Conditions on TrafficCrash Severity. *Journal of Transportation Engineering, 141*(10), 04015020, 2015.

27. Li, Q., M. Yao, X. Yao and B. Xu. A real-time 3D scanning system for pavement distortion inspection. *Measurement Science and Technology, 21*(1), 015702, 2009.

28. Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, SSD: Single shot multibox detector. ed. *European conference on computer vision* , 21-37, 2016.

29. Livneh, M., Repeatability and reproducibility of manual pavement distress survey methods. ed. *Proc., 3rd Int. Conf. on Managing Pavements*, 279-289, 1994.

30. Maeda, H., Y. Sekimoto, T. Seto, T. Kashiyama and H. Omata. Road Damage Detectionand Classification Using Deep Neural Networks with Smartphone Images. *Computer- Aided Civil and Infrastructure Engineering, 33*(12), 1127-1141, 2018.

31. NexarInc., 2020. [online]. Available from: https://www.getnexar.com/.

32. Nienaber, S., M. J. Booysen and R. Kroon. *Detecting Potholes Using Simple ImageProcessing Techniques and Real-World Footage,* 34th Annual Southern African Transport Conference SATC, 2015.

33. Ouma, Y. O. and M. Hahn. Wavelet-morphology based detection of incipient linear cracksin asphalt pavements from RGB camera imagery and classification using circular Radon transform. *Advanced Engineering Informatics, 30*(3), 481-499, 2016.

34. Park, S., S. Bang, H. Kim and H. Kim. Patch-Based Crack Detection in Black Box ImagesUsing Convolutional Neural Networks. *Journal of Computing in Civil Engineering, 33*(3), 04019017, 2019.

35. Ragnoli, A., M. R. De Blasiis and A. Di Benedetto. Pavement Distress Detection Methods:A Review. *Infrastructures, 3*(4), 58, 2018.

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

37. Ren, S., K. He, R. Girshick and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis andMachine Intelligence, 39*(6), 1137-1149, 2017.

38. Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the inception architecture for computer vision. ed. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818-2826, 2016.

39. Szegedy, C., L. Wei, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions. ed. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9, 7-12 June 2015.

40. TensorFlowAPI, 2021. *TensorFlow 2 Detection Model Zoo* [online]. Available from: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md.

41. Yu, X. and E. Salari, Pavement pothole detection and severity measurement using laser imaging.ed. *2011 IEEE International Conference on Electro/Information Technology*, 1-5, 15-17 May 2011.

42. Yusof, N. A. M., A. Ibrahim, M. H. M. Noor, N. M. Tahir, N. M. Yusof, N. Z. Abidin and M. K.Osman. Deep convolution neural network for crack detection on asphalt pavement. *Journal of Physics: Conference Series,* 1349, 012020, 2019.

43. Zakeri, H., F. M. Nejad and A. Fahimifar. Image Based Techniques for Crack Detection, Classification and Quantification Asphalt Pavement: A Review. *Archives of Computational Methods in Engineering, 24*(4), 935-977, 2017.

44. Zhang, A., K. C. P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J. Q. Li and C. Chen. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using aDeep-Learning Network. *Computer-Aided Civil and Infrastructure Engineering, 32*(10), 805-819, 2017.

45. NexarInc., 2020. *CityStream* [online]. Available from: https://data.getnexar.com/product/citystream/.