Computers, Materials & Continua
DOI: 10.32604/cmc.2022.029544
Article

Tech Science Press

A Deep Learning-Based Approach for Road Surface Damage Detection

Bakhytzhan Kulambayev[1,*], Gulbakhram Beissenova[2,3], Nazbek Katayev[4], Bayan Abduraimova[5], Lyazzat Zhaidakbayeva[2], Alua Sarbassova[6], Oxana Akhmetova[7], Sapar Issayev[4], Laura Suleimenova[8], Syrym Kasenov[6], Kunsulu Shadinova[9] and Abay Shyrakbayev[10]

[1]International Information Technology University, Almaty, Kazakhstan
[2]M.Auezov South Kazakhstan University, Shymkent, Kazakhstan
[3]University of Friendship of People's Academician A. Kuatbekov, Shymkent, Kazakhstan
[4]Kazakh National Women's Teacher Training University, Almaty, Kazakhstan
[5]L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan
[6]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[7]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[8]South Kazakhstan State Pedagogical University, Shymkent, Kazakhstan
[9]Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan
[10]International Taraz Innovative Institute, Taraz, Kazakhstan
*Corresponding Author: Bakhytzhan Kulambayev. Email: bakhytzhankulambayev@gmail.com
Received: 06 March 2022; Accepted: 07 May 2022

Abstract: Timely detection and elimination of damage in areas with excessive vehicle loading can reduce the risk of road accidents. Currently, various methods of photo and video surveillance are used to monitor the condition of the road surface. The manual approach to evaluation and analysis of the received data can take a protracted period of time. Thus, it is necessary to improve the procedures for inspection and assessment of the condition of control objects with the help of computer vision and deep learning techniques. In this paper, we propose a model based on Mask Region-based Convolutional Neural Network (Mask R-CNN) architecture for identifying defects of the road surface in the real-time mode. It shows the process of collecting and the features of the training samples and the deep neural network (DNN) training process, taking into account the specifics of the problems posed. For the software implementation of the proposed architecture, the Python programming language and the TensorFlow framework were utilized. The use of the proposed model is effective even in conditions of a limited amount of source data. Also as a result of experiments, a high degree of repeatability of the results was noted. According to the metrics, Mask R-CNN gave the high detection and segmentation results showing 0.9214, 0.9876, 0.9571 precision, recall, and F1-score respectively in road damage detection, and Intersection over Union (IoU)-0.3488 and Dice similarity coefficient-0.7381 in segmentation of road damages.

Keywords: Road damage; mask R-CNN; deep learning; detection; segmentation

1 Introduction

Nowadays, the progress in computer vision is largely due to the appearance of a huge amount of labeled data. Autonomous driving systems related to the analysis of environmental images, detection and tracking of moving objects are being actively developed. Semantic segmentation datasets such as Citades [1], Wild duck [2] and Karlsruhe Institute of Technology in Technological Institute (KITTI) dataset [3] are used for training. Marking up such samples is carried out manually and costs a great deal of money and labor. The samples mainly contain instances of classes such as roadbed, pedestrian, vehicle, sky, road sign and other characteristic, common elements of the highway.

Due to the increasing demand of the road industry for computer processing of high-quality video data of highways, there is a call for development of an algorithm for automatic detection of defects of the roadway by image. The development of an effective algorithm for detecting defects of the roadway in images is an urgent task, since its results can be used both in road organizations and in unmanned vehicles.

The wearing of the road surface requires regular monitoring. Effective monitoring strategies allow timely detection of problem areas. This approach significantly increases the efficiency of road maintenance, reduces maintenance costs and ensures continuous operation. Technologies for detecting critical signs of the condition of the road surface have evolved from manual methods of photofixation to the use of high-speed digital technology [4].

The authors of this paper propose a new technological solution in the field of machine learning. Its implementation makes it possible to automate the process of assessing the quality of the road surface. For this purpose, a convolutional neural network is trained on data that marked up manually. Thus, the system learns to recognize and evaluate the main types of damage to control objects.

Reminder of this paper as follows: The following section provides an overview of the literature and related works, including the methods and datasets used. Section 3 explains methodology of this study that includes four subsections as data collection and preparation, the proposed model, computation resources, evaluation metrics. Section 4 demonstrates experiment results where we present road damage detection results, data markup, evaluation of the proposed method by classes. In Section 5, we conclude the paper by giving an accent to the proposed model and obtained results.

2 Related Works

2.1 Methods

Artificial neural networks (ANN) of the third generation with a special architecture as deep convolutional neural networks (CNN), are one of the most promising approaches for solving the problem of automated quality control of road surfaces [5,6]. The architecture of the convolutional neural network is based on the principles of the architecture of a multilevel neocognitron: based on low-level features within the same class, high-level ones are formed through the use of small-sized synaptic convolution nuclei. In contrast to the connection of neurons of two adjacent layers on the principle of "each with each", the speed characteristics of the detection process are improved. Convolutional neural networks are widely used in the automatic analysis of large volumes of images to accurately identify the distinctive features of both individual products and the system as a whole [7,8].

The growth of computing power of graphics processing units (GPU) allows the use of deeper architectures of machine learning models [9]. Thus, it has been made possible to avoid retraining [10], which has been facilitated by the development of such modern techniques as data augmentation, regularization, etc. The improvement of convolutional neural networks opens up the possibility of

more effective study and generalization of image features (for example, image classification [11], object search [12], vehicle detection [13]).

The flexibility and prospects of deep learning for the tasks of automatic detection of cracks in the pavement are shown in [14,15]. In [16], the use of neural networks for automatic detection and classification of cracks in asphalt is considered. The authors suggest using the average value and the variance of the values of shades of gray. Taking into account these indicators, the image is divided into fragments, after which each cell is classified as a crack. The expediency of using full weight deflectometers (FWD) to assess asphalt cracks is demonstrated. In 98% of cases, the system effectively detects a crack in the image.

In [17], the use of a neural network for detecting defects is investigated. The advantages of the method of clustering pixels as objects were found out. It allows you to increase the accuracy of identification and reduce noise. In [18], the authors used a deep learning architecture that includes the Visual Geometry Group-16 (VGG-16) model. It was previously trained to identify features that make it possible to distinguish between classes of images. The model demonstrated excellent recognition quality even during the work with images from areas unknown to it. Visual Geometric Group-16 Convolutional Neural Network (VGG-16 CNN) is used as a deep feature generator of road surface images. The authors trained only the last layer of the classifier. They conducted experiments with various machine learning models, showed their strengths and weaknesses.

Studies [19] have illustrated the success of using the described architecture in conditions of a limited amount of source data in various segmentation tasks due to a high degree of repeatability. A fragment of the image of the object of control is fed to the input of the neural network, and a map of the probabilities of the presence of a defect is compiled at the output. When the roadway is damaged, a large amount of noise and foreign objects appear in the images with a small gray range and a small difference between the background and the target object. Due to the allocation of appropriate classes, a trained and finely tuned CNN model allows the identification and evaluation of the main types of defects of different shapes and sizes in the images of the road surface. Tab. 1 demonstrates related work on the road surface damage detection problem, including the proposed approaches, features and results obtained.

Table 1: Related works in road damage detection

Reference	Year	Method	Feature	Result
[5]	2017	Convolutional neural network is built on the principles of the architecture of a multilevel neocognitron	Mean value and variance of grayscale values	-
[6]	2021	CNN-based road-surface crack detection model	Lighting conditions of the road surface	85% F1-score
[7]	2019	Sample and Structure-Guided Network (SGN)	Color and texture	87.92% accuracy
[8]	2021	Deep learning-based visual crack detection	Shapes of road images	90% accuracy
[11]	2018	Convolutional Neural Networks	Visual features	90.45% accuracy

(Continued)

Table 1: Continued

Reference	Year	Method	Feature	Result
[15]	2019	Encoder–decoder network for pixel-level road crack detection	Width, shape, and length	71.98% recall, 77.68% precision, 59.65% intersection of union
[16]	2021	Deep Neural Network	Visual featues	89.1% accuracy in crack detection
[17]	2021	Neural Networks	Clustering pixels as objects	-
[18]	2017	VGG-16	-	-
[19]	2019	Artificial Neural Network	-	-

2.2 Datasets

The most well-known available datasets related to road defects are considered:

1. German Asphalt Pavement Distress Dataset (GAPs dataset) [20]: 1969 grayscale coverage images from three German cities with a resolution of 1920 × 1080 pixels, divided into 64x64 pixel fragments that have a binary sign of cracks.
2. Crack500 dataset [21]: 500 red-green-blue (RGB) images of cracked asphalt pavement, with a resolution of approximately 2000 × 1500 pixels, obtained using a smartphone on the campus of Temple University. Each image is provided with a pixel-by-pixel binary mask belonging to the crack.
3. The Crack True 200 dataset [22]: 206 800 × 600 pixel coating images with various types of cracks, containing not only a uniform background texture, but also shadows. Each image is provided with pixel-by-pixel markup.
4. Computational Fluid Dynamics (CFD) dataset [23]: 118 480 × 320 crack images, semantically segmented, taken from above on Beijing city roads. They have shadows, oil stains and water stains.
5. The RoadDamageDataset dataset [8]: 9053 images from a smartphone mounted at the windshield of the car and aimed at shooting the general view ahead of the car. This set has eight types of destruction of the road surface, highlighted by rectangular bounding boxes. 15457 instances of destruction have been allocated, and the data set itself has the PASCAL Visual Object Classes (PASCAL VOC) structure.

The RoadDamageDataset dataset was recorded in seven cities in Japan, and includes eight types of pavement damage: five classes for cracks, two classes for marking wear and one class for potholes and subsidence. The data set has a PASCAL VOC structure and was presented at the Institute of Electrical and Electronics Engineers (IEEE) Big Data Cup forum in 2018.

The data set of Japanese scientists has revived an interest in solving the problem of automatic detection of defects using machine learning methods and, in particular, the use of convolutional neural networks. The advantage of the sample is its solid size, as well as the presence of other types of coating damage, not just cracks. The flipside is that it operates using the method of highlighting defects or the limiting frame, since due to the variety of shapes and sizes of defects with the help of the limiting frame, it is only possible to judge its presence in the image. For the purposes of assessing the quality of the highway, the best option is a pixel-by-pixel selection using a mask, which allows not only to accurately localize the defect, but also to estimate its area. Tab. 2 demonstrates datasets for road damage detection including damage types and features of the images to train neural networks.

Table 2: Comparison of datasets in road damage detection

Dataset	Feature	Damage type
Crack Images [16]	3704 × 10,000 sized 1000 images	Crack
GAPs [20]	1920 × 1080 sized 1969 images	Crack
Crack500 [21]	2000 × 1500 sized 500 images	Crack
CrackTree200 [22]	800 × 600 sized 206 images	Crack
CFD [23]	4480 × 320 sized 118 images	Crack
RoadDamageDataset [24]	9053 images	Eight types of pavement damages as D00: Linear crack, longitudinal, wheel mark part; D01: Linear crack, longitudinal, construction joint part; D10: Linear crack, lateral, equal interval; D11: Linear crack, lateral, construction joint part; D20: Alligator crack D40: Rutting, bump, pothole, separation; D43: Cross walk blur; D44: White line blur

3 Methodology

The analysis of the given thematic literature clearly demonstrated the exceptional advantages of deep convolutional neural networks and the validity of their use in the ongoing research. For starters, it is necessary to carry out segmentation of the roadway image with the allocation of appropriate classes, which will enable to detect a defect. For these purposes, specially designed CNN architectures like segmentation network (SegNet) [25] and U-Net [26] are currently being effectively used. The complexity of the task lies in the limited range of shades of gray in the images of the road surface, as well as in the slight difference between the target object and the background, the presence of noise and extraneous details. Due to the specificity of the processed images, segmentation is carried out using a fully convolutional neural network (FCNN) with an "en-encoder-decoder" structure, which allows to obtain a binary image at the output [27]. FCNN is formed by two parts — convolutional, which converts the input image into a multidimensional representation of features, and non-convolutional, which produces a segmented image based on these features. The first part is constructed by sequentially arranged five convolutional layers with sets of filters, followed by layers of sub discretization. Layers of increasing sampling together with convolutional layers allow you to restore the image size to the

original one after passing through these layers and form a probability map. The CrackForest dataset consists of 117 images, divided into training, test and validation (evaluating the quality) samples. 64×64 fragments are randomly selected from the training and test samples for each image. Gamma image correction improves the quality of the neural network. The optimal ratio of fragments with and without a defect was established at the level of 95% to 5%, taking into account defects occupying at least 5% of the image area. The ratio of the sample size of 15,200 fragments of the training and 3,968 test samples is optimal for the learning process and the operation of a deep neural network. Training and evaluation of the neural network takes place using metrics of intersection between two detections and an equivalent binary measure of similarity. Initialization of weights in FCNN layers is carried out by the Glorot method. When the input distributions of each layer are normalized, the internal covariance shift decreases, thereby achieving normalization of the batch. Stochastic optimization training is carried out using the Adam algorithm. It is established that the optimal number of epochs of neural network training is at the level of 25 (5 at the first stage and 20 at the second). The implementation of the built FCNN architecture is achieved by virtue of the Keras and Tensor-Flow frameworks. After training the ANN, it is checked and validated on test data. Each fragment of the image is fed to the input of the network, and a map of the probabilities of the presence of a defect is generated at the output.

In our case, an improved technique was used to train pre-trained Mask R-CNN models in TensorFlow Object Detection API to increase road damage detection performance. Afterwards the models are put to the test with the usage of sorted annotation data.

3.1 Data Collection and Preparation

Until now, images of the detection of damage on the road surface were either taken above the road surface or using on-Board cameras on vehicles. When models train with images that taken from above, the situations that can be applied in practice are limited considering the complexity of capturing such images. In contrast, when a model is built from images taken from the vehicle's onboard camera, these images can be easily applied to train the model for practical situations. For example, using an easily accessible camera, such as on smartphones and cars, anyone can easily detect road damage by running a model on a smartphone or transferring images to an external server and processing it on the server. Therefore, we created our own data set that includes six types of road damage. All images were annotated manually.

Fig. 1 gives samples of the various types of damage and their definition. In this project, each type of damage is represented by a class name, such as D20. Each type of damage is illustrated in the examples in the figures below. As you can see from the table, damage types are divided into six categories. First, the damage is classified as cracks or other damage. The cracks are then divided into linear cracks and alligator cracks. Other distortions include not only potholes and ruts, but also other road damage such as blurring of white lines. As far as we know, no previous study has covered such a wide variety of road injuries, especially in the case of image processing. For example, the method proposed by [28] detects only potholes in D40, and in Jana et al. [29] classifies damage types exclusively as longitudinal and lateral. In addition, previous research with the usage of deep learning [30–33] only detects the presence or absence of damage.

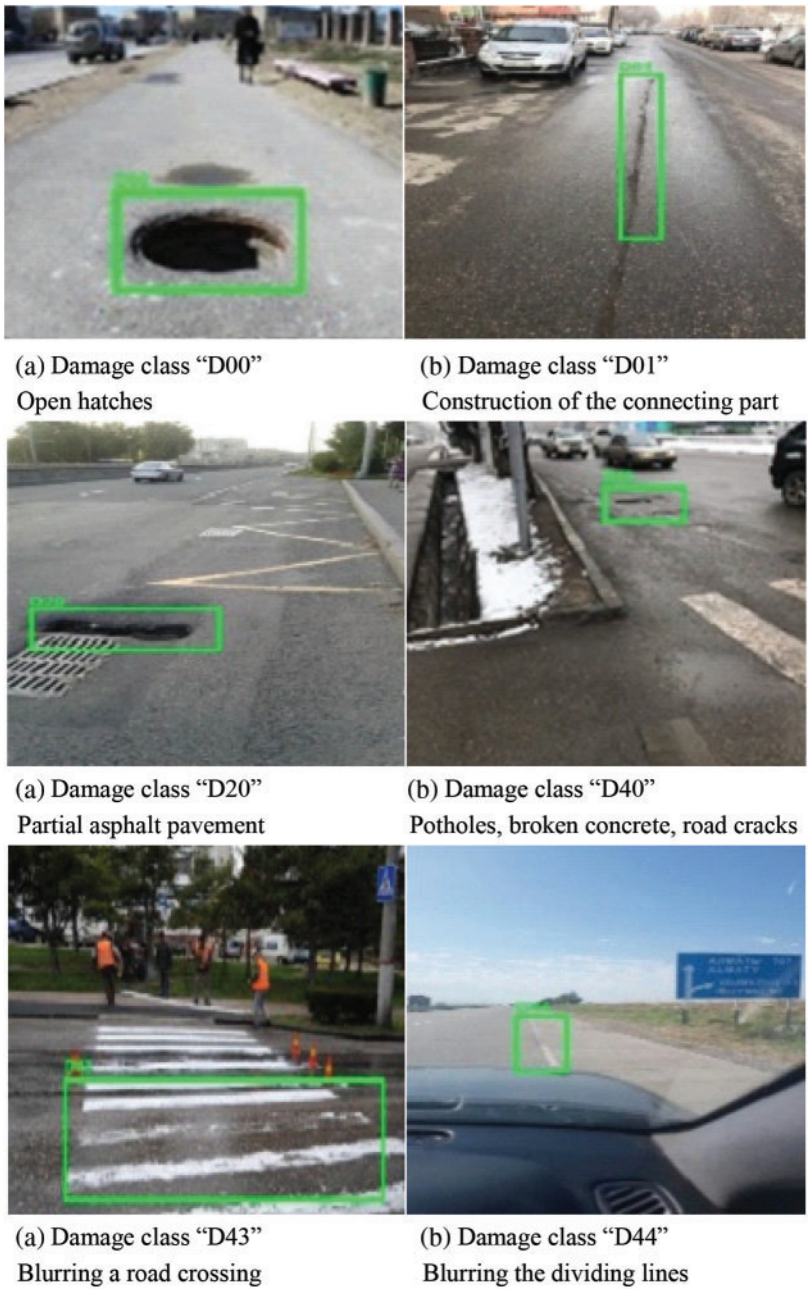

Figure 1: Road damage photos and classes for a model training

To distinguish such damages from others, the annotation data provides 12 classifications of road damages and cognitive items in the road photos. The Microsoft Visual Object Tagging Tool (VoTT) was used to annotate road color photos. In the bottom two-thirds of each picture, all visible items of the preset classes were segmented and labeled. Tab. 3 demonstrates the annotation data.

Table 3: Road images annotation data

Class ID	Classes	Training	Validation	Testing	Total
1	Linear crack	3080	660	660	4400
2	Grid crack	658	141	141	940
3	Pavement joins	854	183	183	1220
4	Patchings	448	96	96	640
5	Fillings	1344	288	288	1920
6	Pot-holes	406	87	87	580
7	Manholes	336	72	72	480
8	Stains	266	57	57	380
9	Shadow	1190	255	255	1700
10	Pavement markings	1414	303	303	2020
11	Scratches on markings	3360	720	720	4800
12	Grid crack in patchings	252	54	54	360
0	Total	13608	2916	2916	19440

The most segments were in the "Scratches on Markings" class, which had 3,360 in total. At 3,080 segments, "Linear Cracks" comes next. "Grid Cracks in Patchings" had the fewest segments (252), followed by "Stains", "Manholes" and "Pot-holes". At a ratio of 0.6:0.2:0.2, the segments of each class were separated into datasets for training, validating, validation, and testing.

3.2 The Proposed Model

To simultaneously solve the problem of crack detection and their pixel-by-pixel separation, it was decided to use the modern architecture of the Mask R-CNN convolutional network. Let's consider its structure and principle of operation. The Mask R-CNN architecture historically has the following number of predecessors based on the idea of processing small areas: Region-based Convolutional Neural Network (R-CNN), Fast R-CNN, Faster R-CNN.

Fig. 2 illustrates our Mask R-CNN architecture for road surface damage detection problem. The Mask R-CNN architecture has a complex block structure. Initially, the image is fed to the input of the neural network to highlight the feature map, which is often used as VGG-16, residual neural network with 50 layers (ResNet50) and residual neural network with 101 layers (ResNet101) with excluded layers responsible for classification. One of the improvements of this architecture compared to its predecessors is the use of the Feature Pyramid Network (FPN) approach, which extracts multi-scale feature maps. Successive layers of the SNA with decreasing dimension are considered as a hierarchical "pyramid" in which the maps of the lower levels have high resolution, and the maps of the upper levels have high generalizing, semantic ability.

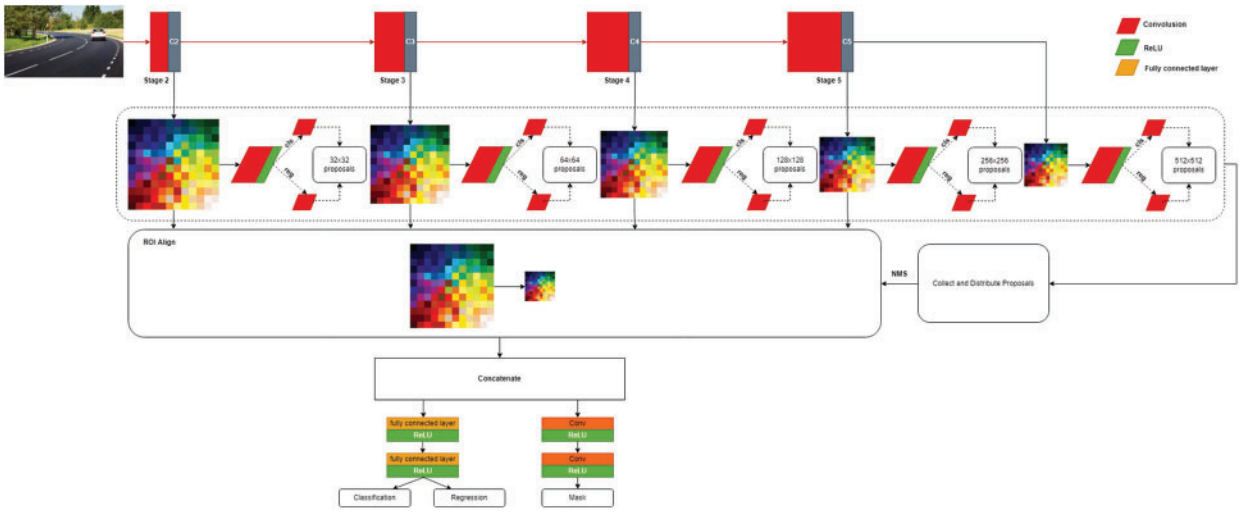

Figure 2: Architecture of mask R-CNN

The obtained feature maps are processed in the Region Proposals Network (RBN) block that has a task to generate the assumed regions in the image which contain objects. To do this, a neural network with a 3x3 window is slid along the feature map and an output is formed based on k anchors-the framework of a given dimension and position. For each anchor, RPN generates a prediction of the presence of an object, and a refinement of the coordinates of the bounding box of the object, if it has been detected. The purpose of this stage is to highlight regions of interest that may contain objects. At the end, duplicate regions are discarded due to the operation of non-maximum suppression.

Then, using the Region of Interest (ROI) Align operation, the values corresponding to the regions are selected from the feature maps and reduced to the same size. According to them, the final operations of classification, refinement of the coordinates of the bounding box and prediction of the mask are carried out. The mask at the output has a greatly reduced size, but contains real values. When the mask is scaled to the size of the selected object, it is possible to obtain sufficient accuracy.

3.3 Computation Resources

The tests were conducted on a machine with Intel Core i9-9900KF (8 cores/16 threads/3.60 GHz) central processing unit (CPU), 32GB CPU memory, Nvidia GeForce (Giga Texel Shader eXtreme) GTX 1080 GPU with with 8GB of graphics double data rate type 5X (GDDR5X) memory, with a 10Gbps memory speed, 256-bit memory interface and a memory bandwidth of 320GB/sec, 2560 compute unified device architecture (CUDA) cores, GPU clocks of 1607/1733 MHz. The language for programming is Python 3.6.9 using libraries from object detection application programming interface (API) version 1 on top of TensorFlow 1.15.0.

3.4 Evaluation Metrics

The various metrics used to evaluate the proposed model are the mean average precision (MaP) and average recall (AR) at various levels of intersection over union (IoU). In classification problems with localization and object detection, the ratio of the areas of the bounding boxes is most often used as a metric to determine the reliability of the location of the bounding box.

The Mask RCNN has a region proposal network layer that makes multiple inferences simultaneously on the class classification, the segmentation and the mask areas resulting six loss metrics. Besides above model-wise metrics, the average precisions and the average recalls at IoU = 0.5 are used for all twelve road object classes [34].

$$IoU = \frac{S(A \cap B)}{S(A \cup B)} \tag{1}$$

where A and B are the predicted bounding box and the current bounding box, respectively. IoU is zero in the case of disjoint bounding boxes and is equal to one in the case of a perfect overlap.

The goal of assessment is to identify as many instances as possible from a population for a screening method, hence false negatives should be kept to a minimum at the cost of increasing false positives. As a result, three primary metrics must be determined: true positive rate (TPR), false positive rate (FPR), and accuracy (ACC). In medical language, the first parameter is referred to as sensitivity (SEN) and is written as Eq. (2) [35]:

$$TPR = SEN = \frac{TP}{P} \tag{2}$$

where the number of true positive is TP, and the number of positive instances is P.

The estimation of the second term, false positive rate, expressed as Eq. (3) [36]:

$$FPR = \frac{FP}{N} \tag{3}$$

The population's cumulative number of negative occurrences is N, while the proportion of false positives is FP, and number of true negative samples is N. This statistic, on the other hand, is better understood as the ratio of genuine negatives to real negatives, known in medical language as the specificity (SPEC), which is given as Eq. (4) [37]:

$$TNR = SPEC = \frac{TN}{N} = 1 - FPR \tag{4}$$

Finally, accuracy determines the balance between real positives and true negatives. This may be a highly useful statistic when the number of positive and negative occurrences is not equal. This is expressed as Eq. (5) [38]:

$$ACC = \frac{TP + TN}{P + N} \tag{5}$$

4 Results

In this section, we divided the experimental results into two subsections. In the first subsection, we demonstrate results of road damage detection. In the further subsections, we present road damage segmentation results. In the second section, we demonstrate how the proposed model works in real time and show visual presentation. In addition, we indicate source images and marked up road images. In Subsection 3, we illustrate evaluation results of the proposed model by showing different evaluation parameters as precision, recall, f-score for each classified classes of road surface damages.

4.1 Road Damage Detection Results

The road damage detections system was developed using Mask R-CNN model. The proposed method might hide various sorts of cracks as well as spall within few moments from the photos

acquired using the camcorder in order to obtain the right form and amount of the damages. Tab. 4 shows the comparison of the results of road surface damage detection process by indicating precision, recall, F1-score as evaluation parameters.

Table 4: Evaluation of the proposed method by classes

Model	Precision	Recall	F1-score
Proposed model	0.9214	0.9876	0.9571
Fully convolutional encoder–decoder network [39]	0.9130	0.9410	0.9270
Deep learning-based semantic segmentation [40]	0.8340	0.6855	0.7524
UNet-based concrete crack detection CrackUnet19 [41]	0.9145	0.8867	0.9004
Two-step light gradient boosting machine [42]	0.6801	0.7578	0.6950
Semantic segmentation using deep learning [43]	0.4044	0.7847	0.4994
Automated vision-based detection [44]	0.9236	0.8928	0.9079

4.2 Data Markup

To determine the part of the image corresponding to the roadway, all pixels of the road mask are highlighted. Then, an algorithm for searching 8 connected regions is applied to the resulting binary mask. As a result, the area with the maximum number of pixels is taken as the coverage mask (highlighted in gray in Fig. 3).

Figure 3: Marked up road images

To evaluate the effectiveness of the developed approach to the identification of defects, a small data set of 50 real images with cracks on the road was manually prepared. Fig. 4 shows the results of human selection of cracks and segmentation process using the proposed neural network pixel-by-pixel selection in a real image.

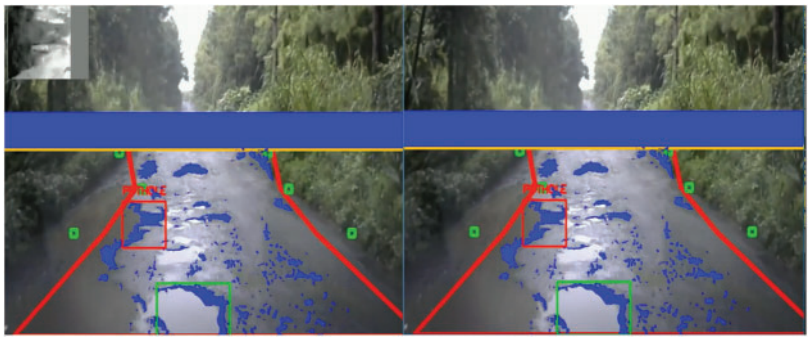

Figure 4: Road surface damage segmentation process

Fig. 5 demonstrates results of the model testing on 100 epochs. Fig. 5a shows accuracy and validation accuracy of the proposed model. From there, we can conclude that our model can gate approximately 90% accuracy in about 60 epochs. This shows the stability and acceptability of the proposed model in real life. Fig. 5b illustrates training and validation loss of the proposed method. As we can see from the figure, we have a minimum loss, and it means the proposed model can get minimum errors in practice.

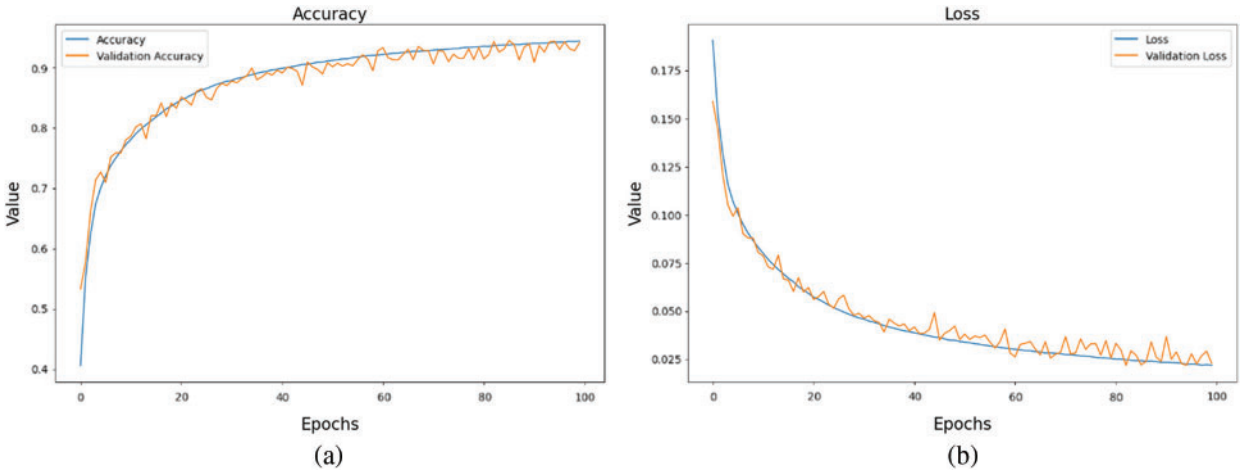

Figure 5: Model testing on 100 epochs. (a) Accuracy (b) Testing and validation loss

There are different approaches that uses deep learning techniques for safety in the roads. Recent literatures proposes interesting solutions for this problem [45]. For example, [46] Vehicle Re-Identification method to solve the problem of attributable to the large intra-class differences caused by different views of vehicles in the traveling process and obvious inter-class similarities caused by similar appearances. Our model directed to detect road surface damages using smartphone cameras or any other equipment that can fix real-time road videos. In the result of the provided experiments, we can conclude that deep learning methods can be successfully applied in the problems for safety and security in the roads.

4.3 Evaluation of the Proposed Model

On the bounding boxes and segmentation masks, Tab 5 displays several model metrics. On bounding boxes, the values of mAP (IoU = 50:.05:.95), mAP (IoU = 50), and mAP (IoU = 75) are 0.2432, 0.4382, and 0.2482, respectively, while on segmentation masks, they are 0.1600, 0.3257, and 0.1279, with a significant drop in those metrics. When compared to the Precision mAP (big) of large objects and medium objects, the Precision mAP (small) for tiny objects = 0.0365 and 0.0133 on bounding boxes and segmentation masks, respectively, are noticeably smaller. On boundary boxes, the Average Recall for small, medium, and big objects is 0.1166, 0.3132, and 0.4717, respectively, while on segmentation masks, it is 0.1021, 0.2528, and 0.2732. Our target damage classes of linear cracks (Crack1), grid cracks (Crack2), potholes, scratches on markings, and grid cracks in patchings have detection precisions of 0.4085, 0.4958, 0.5714, 0.5934, and 0.4000 at IoU = 50, respectively.

Table 5: Evaluation of the proposed method by classes

Classes	Precision @ 0.5 IoU (Bounding box)	Recall @ 0.5 IoU (Bounding box)	Recall @ 0.5 IoU (Segmentation)	Recall @ 0.5 IoU (Segmentation)
Linear crack	0.5383	0.3847	0.3583	0.2639
Grid crack	0.6256	0.7140	0.5920	0.6744
Pavement joins	0.4900	0.5179	0.2498	0.2531
Patchings	0.7644	0.5584	0.8161	0.5843
Fillings	0.6071	0.4667	0.3040	0.2528
Pot-holes	0.7012	0.4155	0.7012	0.4155
Manholes	0.9596	0.8798	0.9596	0.8798
Stains	0.1798	0.1484	0.1191	0.1282
Shadow	0.5273	0.4317	0.3285	0.2713
Pavement markings	0.7522	0.7460	0.5065	0.5002
Scratches on markings	0.7232	0.7531	0.4863	0.4944
Grid crack in patchings	0.5298	0.2474	0.7298	0.3063

5 Conclusion

As part of the task of automatic detection of roadway defects, Mask R-CNN was introduced to detect cracks and their segmentation at the pixel level. The completed work has shown that the use of such architectures can be successful with a small amount of source data.

An analytical review of this area has shown that crack detection studies are limited, since automatic crack detection at the pixel level remains a difficult task due to the heterogeneous pixel intensity, complex crack topology, different lighting and noisy coating texture.

The contribution of this work consists of three parts. Initially, we reviewed and analyzed previous work and identified the advantages and disadvantages of existing approaches. Secondly, data was collected that contains 12 classes of road damage. Thirdly, we have developed a deep learning model based on the RCN Mask architecture for detecting and segmenting road damage. The results obtained

allow us to judge the applicability of the training approach on a synthetic sample, which enables us to get better results compared to using a small data set marked up manually. The proposed model showed IoU-0.3488, Dice-0.7381, which demonstrates applicability in practice.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest of to report regarding the present study.

References

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler *et al.,* "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, The US, pp. 3213–3223, 2016.

[2] O. Zendel, K. Honauer, M. Murschitz, D. Steininger and G. Dominguez, "Wilddash-creating hazard-aware benchmarks," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 402–416, 2018.

[3] J. Zhang, Y. Sun, H. Liao, J. Zhu and Y. Zhang, "Automatic parotid gland segmentation in MVCT using deep convolutional neural networks," *ACM Transactions on Computing for Healthcare*, vol. 3, no. 2, pp. 1–15, 2021.

[4] S. Chen, Y. Zhang, Y. Zhang, J. Yu and Y. Zhu, "Embedded system for road damage detection by deep convolutional neural network," *Mathematical Biosciences and Engineering: MBE*, vol. 16, no. 6, pp. 7982–7994, 2019.

[5] K. Gopalakrishnan, S. Khaitan, A. Choudhary and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, no. 1, pp. 322–330, 2017.

[6] T. Lee, Y. Yoon, C. Chun and S. Ryu, "CNN-Based road-surface crack detection model that responds to brightness changes," *Electronics*, vol. 10, no. 12, pp. 1402–1412, 2021.

[7] S. Wu, J. Fang, X. Zheng and X. Li, "Sample and structure-guided network for road crack detection," *IEEE Access*, vol. 7, no. 1, pp. 130032–130043, 2019.

[8] M. Maniat, C. Camp and A. Kashani, "Deep learning-based visual crack detection using google street view images," *Neural Computing and Applications*, vol. 33, no. 21, pp. 14565–14582, 2021.

[9] D. Dewangan and S. Sahu, "RCNet: Road classification convolutional neural networks for intelligent vehicle system," *Intelligent Service Robotics*, vol. 14, no. 2, pp. 199–214, 2021.

[10] M. Masud, M. Hossain, H. Alhumyani, S. Alshamrani, O. Cheikhrouhou *et al.,* "Pre-trained convolutional neural networks for breast cancer detection using ultrasound images," *ACM Transactions on Internet Technology*, vol. 21, no. 4, pp. 1–17, 2021.

[11] S. Bang, S. Park, H. Kim, Y. Yoon and H. Kim, "A deep residual network with transfer learning for pixel-level road crack detection," *Network*, vol. 93, no. 84, pp. 89–03, 2018.

[12] Y. Chen, H. Wang, W. Li, C. Sakaridis, D. Dai *et al.,* "Scale-aware domain adaptive faster r-cnn," *International Journal of Computer Vision*, vol. 129, no. 7, pp. 2223–2243, 2021.

[13] D. Quang and S. Bae. "A hybrid deep convolutional neural network approach for predicting the traffic congestion index," *Promet-Traffic & Transportation*, vol. 33, no. 3, pp. 373–385, 2021.

[14] N. Safaei, O. Smadi, B. Safaei and A. Masoud, "Efficient road crack detection based on an adaptive pixel-level segmentation algorithm," *Transportation Research Record*, vol. 2675, no. 9, pp. 370–381, 2021.

[15] S. Bang, S. Park, S., Kim and H. Kim, "Encoder–decoder network for pixel-level road crack detection in black-box images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 8, pp. 713–727, 2019.

[16] V. Tran, T. Tran, H. Lee, K. Kim, J. Baek *et al.,* "One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects," *Journal of Civil Structural Health Monitoring*, vol. 11, no. 1, pp. 205–222, 2021.

[17] Z. Lingxin, S. Junkai and Z. Baijie, "A review of the research and application of deep learning-based computer vision in structural damage detection," *Earthquake Engineering and Engineering Vibration*, vol. 21, no. 1, pp. 1–21, 2022.
[18] H. Li, Z. Todd, N. Bielski and F. Carroll, "3D lidar point-cloud projection operator and transfer machine learning for effective road surface features detection and segmentation," *The Visual Computer*, vol. 38, no. 5, pp. 1759–1774, 2022.
[19] S. Patra, A. Middya and S. Roy, "PotSpot: Participatory sensing based monitoring system for pothole detection using deep learning," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 25171–25195, 2021.
[20] T. Rateke and A. Von Wangenheim, "Road surface detection and differentiation considering surface damages," *Autonomous Robots*, vol. 45, no. 2, pp. 299–312, 2021.
[21] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei *et al.,* "Feature pyramid and hierarchical boosting network for pavement crack detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1525–1535, 2019.
[22] Q. Zou, Y. Cao, Q. Li, Q. Mao and S. Wang, "CrackTree: Automatic crack detection from pavement images," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227–238, 2012.
[23] Y. Shi, L. Cui, Z. Qi, F. Meng and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.
[24] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.
[25] H. Afify, K. Mohammed and A. Hassanien, "An improved framework for polyp image segmentation based on SegNet architecture," *International Journal of Imaging Systems and Technology*, vol. 31, no. 3, pp. 1741–1751, 2021.
[26] B. Omarov, A. Tursynova, O. Postolache, K. Gamry, A. Batyrbekov *et al.,* "Modified UNet model for brain stroke lesion segmentation on computed tomography images," *CMC-Computers, Materials & Continua*, vol. 71, no. 3, pp. 4701–4717, 2022.
[27] D. Laredo, S. Ma, G. Leylaz, O. Schütze and J. Sun, "Automatic model selection for fully connected neural networks," *International Journal of Dynamics and Control*, vol. 8, no. 4, pp. 1063–1079, 2020.
[28] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto and H. Omata, "Generative adversarial network for road damage detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 47–60, 2020.
[29] S. Jana, S. Thangam, A. Kishore, V. Sai Kumar and S. Vandana, "Transfer learning based deep convolutional neural network model for pavement crack detection from images," *International Journal of Nonlinear Analysis and Applications*, vol. 13, no. 1, pp. 1209–1223, 2022.
[30] B. Kim, N. Yuvaraj, K. Sri Preethaa and R. Arun Pandian, "Surface crack detection using deep learning with shallow CNN architecture for enhanced computation," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9289–9305, 2021.
[31] L. Zhang, Z. Wang, L. Wang, Z. Zhang X. Chen *et al.,* "Machine learning-based real-time visible fatigue crack growth detection," *Digital Communications and Networks*, vol. 7, no. 4, pp. 551–558, 2021.
[32] E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis and T. Stathaki, "Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing," *Applied Intelligence*, vol. 49, no. 7, pp. 2793–2806, 2019.
[33] D. Carvalho, E. Pereira and J. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, pp. 832, 2019.
[34] D. Russo, K. Zorn, A. Clark, H. Zhu and S. Ekins, "Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction," *Molecular Pharmaceutics*, vol. 15, no. 10, pp. 4361–4370, 2018.
[35] V. Thambawita, D. Jha, H. Hammer, H. Johansen, D. Johansen *et al.,* "An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification," *ACM Transactions on Computing for Healthcare*, vol. 1, no. 3, pp. 1–29, 2020.

[36] B. Omarov, A. Batyrbekov, K. Dalbekova, G. Abdulkarimova, S. Berkimbaeva *et al.,* "Electronic stethoscope for heartbeat abnormality detection," in *5th Int. Conf. on Smart Computing and Communication (SmartCom 2020)*, Paris, France, pp. 248–258, 2020.

[37] B. Omarov, N. Saparkhojayev, S. Shekerbekova, O. Akhmetova, M. Sakypbekova *et al.,* "Artificial intelligence in medicine: Real time electronic stethoscope for heart diseases detection," *CMC-Computers, Materials & Continua*, vol. 70, no. 2, pp. 2815–2833, 2022.

[38] S. Guillon, F. Joncour, P. Barrallon and L. Castanié, "Ground-truth uncertainty-aware metrics for machine learning applications on seismic image interpretation: Application to faults and horizon extraction," *the Leading Edge*, vol. 39, no. 10, pp. 734–741, 2020.

[39] M. Islam and J. Kim, "Vision-based autonomous crack detection of concrete structures using a fully convolutional encoder–decoder network," *Sensors*, vol. 19, no. 19, pp. 4251, 2019.

[40] T. Yamane and P. Chun, "Crack detection from a concrete surface image based on semantic segmentation using deep learning," *Journal of Advanced Concrete Technology*, vol. 18, no. 9, pp. 493–504, 2020.

[41] L. Zhang, J. Shen and B. Zhu, "A research on an improved Unet-based concrete crack detection algorithm," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1864–1879, 2021.

[42] P. Chun, S. Izumi and T. Yamane, "Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 61–72, 2021.

[43] D. Lee, J. Kim and D. Lee, "Robust concrete crack detection using deep learning-based semantic segmentation," *International Journal of Aeronautical and Space Sciences*, vol. 20, no. 1, pp. 287–299, 2019.

[44] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, no. 10, pp. 3452, 2018.

[45] F. Bi, X. Ma, W. Chen, W. Fang, H. Chen *et al.,* "Review on video object tracking based on deep learning," *Journal of New Media*, vol. 1, no. 2, pp. 63–74, 2019.

[46] X. R. Zhang, X. Chen, W. Sun, X. Z. He, "Vehicle Re-identification model based on optimized DenseNet121 with joint loss," *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3933–3948, 2021.