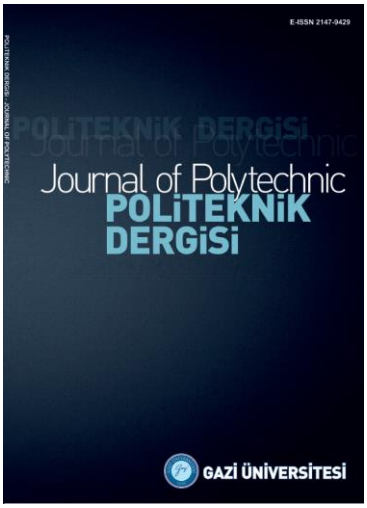


POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.org.tr/politeknik

Faster R-CNN structure for computer vision-based road pavement distress detection

Bilgisayarlı görü tabanlı yol kaplaması tehlikeleri tespiti için Faster R-CNN yapısı

***Yazar(lar) (Author(s))**: Furkan BALCI[1], Safiye YILMAZ[2]*

ORCID[1]: 0000-0002-3160-1517

ORCID[2]: 0000-0002-7836-4520

To cite to this article: Balcı F. ve Yılmaz S., "Faster R-CNN structure for computer vision-based road pavement distress detection", ***Journal of Polytechnic***, 26(2): 701-710, (2023).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Balcı F. ve Yılmaz S., "Faster R-CNN structure for computer vision-based road pavement distress detection", ***Politeknik Dergisi***, 26(2): 701-710, (2023).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: **10.2339/politeknik.987132**

Faster R-CNN Structure for Computer Vision-based Road Pavement Distress Detection

Highlights

- *Faster R-CNN structure were used for pavement crack/distress detection.*
- *Dataset was created for the detection of asphalt cracks.*
- *An accuracy of 93.2% was achieved in the detection and classification of asphalt cracks.*

Graphical Abstract

Using the Faster R-CNN structure, the automatic detectability of the detection of asphalt cracks was investigated. First of all, asphalt crack photos were taken for the dataset. Then, performance analyzes of the Faster R-CNN structure, which was created using the Python programming language and whose block diagram is shown below, were made on real data.

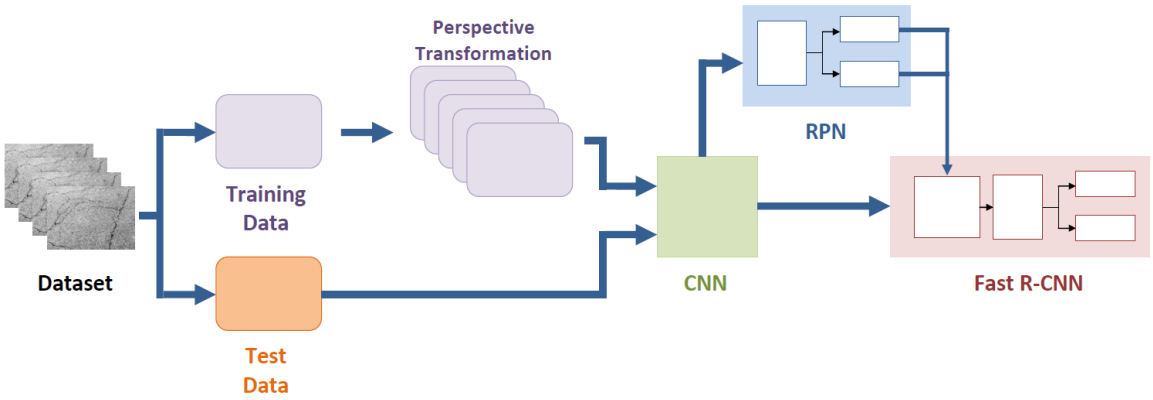


Figure. Block diagram of proposed approach

Aim

The aim of this study is to detect cracks on asphalt fully automatically using Faster R-CNN structure.

Design &Methodology

For this study, photographs of various asphalt cracks are used for the dataset. The size of the dataset is one of the factors that increase performance in deep learning methods. For this reason, the dataset has been expanded using the perspective transform. Then, Faster R-CNN structure was created using Python programming language. Performance analyzes were made with the data obtained as a result of the test.

Originality

Faster R-CNN structure was used to detect asphalt cracks in the study.

Findings

The average precision for all tags is greater than 0.90, of which Non-crack gets the highest score of 94% and the Longitudinal Crack gets the lowest score of 92%. Mean precision score (mAP) of all tag is 93.2%. Finally, the Kappa value is 0.915.

Conclusion

As a result of the study, it has been determined that the Faster R-CNN structure, which is one of the deep learning techniques, can be used in the detection of cracks on asphalt.

Declaration of EthicalStandards

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Faster R-CNN Structure for Computer Vision-based Road Pavement Distress Detection

Araştırma Makalesi / Research Article

Furkan BALCI[1*], Safiye YILMAZ[2]

[1]Faculty of Technology, Department of Electrical and Electronics Engineering, Gazi University, Turkey
[2]Faculty of Engineering, Department of Electrical and Electronics Engineering, Nuh Naci Yazgan University, Turkey

(Geliş/Received : 25.08.2021 ; Kabul/Accepted : 04.01.2022 ; Erken Görünüm/Early View : 14.01.2022)

ABSTRACT

Smart cities can be controlled in all aspects and it is desired to have a structure that is planned to have controllable feedback. Asphalt is generally used as pavement material on roads that provide transportation of vehicles such as cars and buses on the highway. Asphalt material is deformed due to weather conditions, heavy vehicle passage. In the smart city structure, similar deformations should be reported to the relevant unit. In this article, it was tried to determine the deteriorations on the asphalt by selecting the data set obtained from a region with image processing methods and deep learning technique. With the action camera placed in an automobile, a total of 4315 asphalt images with various distortions and without any deterioration were used as dataset. The dataset was classified using a pixel-based Faster Region-based Convolutional Neural Network. Accuracy, precision and sensitivity values were used to make the performance result obtained as a result of classification meaningful. With this proposed method, the average accuracy rate was 93.2%. With these results, an approach that can automatically detect asphalt deterioration in smart city structures has been developed.

Keywords: Deep learning, faster R-CNN, image processing, pavement distress detection.

Bilgisayarlı Görü Tabanlı Yol Kaplaması Tehlikeleri Tespiti için Faster R-CNN Yapısı

ÖZ

Akıllı şehirler tüm yönüyle kontrol altına alınabilir ve kontrol edilebilir geri bildirimleri olması planlanan yapıya sahip olması istenmektedir. Karayolunda otomobil, otobüs gibi taşıtların ulaşımını sağlayan yollarda kaplama malzemesi olarak genellikle asfalt kullanılmaktadır. Asfalt malzemesi hava koşulları, yoğun araç geçişi gibi sebeplerden deforme olmaktadır. Akıllı şehir yapısında buna benzer deformelerin ilgili birime iletilmesi gerekmektedir. Bu makalede görüntü işleme yöntemleri ve derin öğrenme tekniği ile bir bölgeden elde edilen veri seti seçilerek asfalt üzerinde bozulmalar tespit edilmeye çalışılmıştır. Bir otomobile yerleştirilen aksiyon kamerası ile çeşitli bozulmaların olduğu ve herhangi bir bozulmanın olmadığı toplamda 4315 adet asfalt görüntüsü veri seti olarak kullanılmıştır. Piksel tabanlı çalışan Daha Hızlı Bölge Tabanlı Evrişimli Sinir Ağı kullanılarak veri seti sınıflandırılmıştır. Sınıflandırma sonucunda elde edilen performans sonucunun anlamlandırılabilir olması için doğruluk, kesinlik ve duyarlılık değerleri kullanılmıştır. Önerilen bu yöntem ileortalama doğruluk oranı %93.2 elde edilmiştir. Bu sonuçlar ile akıllı şehir yapılarında asfalt bozulmaları için otomatik olarak tespit yapabilen bir yaklaşım geliştirilmiştir.

Anahtar Kelimeler: Derin öğrenme, faster R-CNN, görüntü işleme, kaplama tehlikesi tespiti.

1. INTRODUCTION

Road transportation is one of the most used transportation methods today. It plays an important role in both freight and passenger transport. In road transport, designated roads are used to provide transportation. Asphalt is used as the basic pavement material of these roads. Asphalt material deteriorates due to weather conditions, heavy vehicle passage, sewer collapse. These deteriorations pose a danger to drivers. In order to prevent these deteriorations, maintenance and repair processes should be checked frequently. Early detection of deterioration significantly reduces the high maintenance cost [1].

In addition to the studies in the literature, there are studies in some countries, such as the United States, in which asphalt deteriorations are actively examined manually on the image. As a result of the studies carried out, semi-automatic and automatic deterioration detection can be made with the data obtained from the cameras in various positions placed on the vehicles moving on the highways [2]. In order to detect deteriorations on asphalt in semi-automatic or classical analysis methods, reconnaissance and manpower are needed. Some types of cracks have been identified in the literature. Some of these crack types are shown in Figure 1. Various questionnaires are used for detection in classical detection methods. According to the results of this survey, the deteriorations on the road are detected and necessary actions are taken. However, it is argued that both the results of the survey are different due to the differences in the opinions of the experts and this determination made at the traffic areas is quite dangerous. For this reason, automatic systems are required for the detection of deterioration in asphalt [3].

*Sorumlu Yazar (Corresponding Author)
e-posta : furkanbalci@gazi.edu.tr

In recent years, the development of both imaging technology and artificial intelligence techniques, as in every field, there are various studies in the detection of deterioration on asphalt.

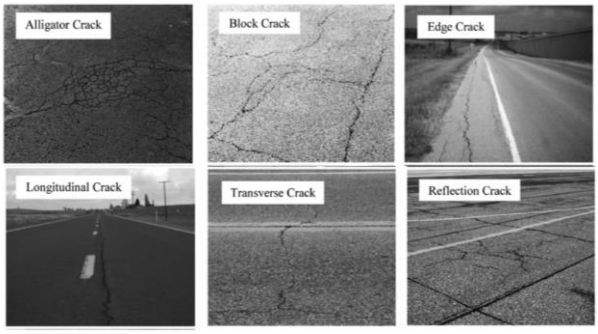


Figure 1. Common asphalt crack types in the literatüre [4]

Auto-detection studies are of three different types in the literature: 2D image processing, 3D cloud point scanning and 3D line scanning. These developed systems are designed to automatically detect and classify deteriorations on asphalt. The local image-processing methods, such as the intensity thresholding, edge detection and sub-window based hand-crafted feature extraction methods are widely used in practice. One of the main difficulties encountered in these studies is to extract the distortions and the other is to classify these distortions. In images with bad asphalt, the cracked pixel is darker than the two adjacent pixels. For this reason, histogram-based algorithms are widely used in the first studies encountered in the literature. In addition to this method, edge detection methods such as Sobel and Laplacian are also encountered. Wavelet transform, which is one of the basic transform techniques, is used to detect distortion in the frequency domain.

There are various approaches in the literature. Some of these approaches differ in the creation of datasets. Apart from the classical dataset creation methods, Majidifard followed a different method and carried out the training and testing process by collecting images via the Google Street View application [5]. In another study, a concept called Region of Belief (RoB) has been proposed. In this study, images were clustered according to pixel values and cracks were determined by applying the RoB method [6]. Some of the studies in the literature try to estimate the Asphalt Condition Index (PCI) value apart from classifying asphalt cracks [7]. Apart from artificial intelligence techniques, crack detection is also performed using various statistical methods [8].

The Faster Region-based Convolutional Neural Network structure is an architecture that can achieve high accuracy that can be used in image classification studies. In this study, a deep learning based prediction model is designed to detect deteriorations on asphalt structure. The dataset consists of 4436 images. Training data are tagged according to crack types. Tag information is the 4 types of asphalt deterioration frequently encountered in the literature. Performance analyzes of the Faster R-CNN structure were made according to these tags.

The rest of the paper is divided as follows. Detailed information about the dataset to be used under Materials and Method and general information about the Faster R-CNN structure are given. It contains information about the results of the Faster R-CNN structure under Results and Discussion and the discussions about them. Finally, under Conclusion, the results are summarized and information about future studies is given.

2. MATERIALS and METHOD

In this article, Faster R-CNN structure that can classify deteriorations on asphalt from asphalt images obtained from a determined region is proposed. This dataset, which was obtained during the training and testing phase of The Faster R-CNN structure, was used. The performance analyzes of the proposed system were made as a result of the tests performed on this dataset.

2.1. Dataset and pre-processing

In this article, an approach that automatically detects and classifies the deteriorations on asphalt has been tried to be designed. A dataset is needed to train and test this approach. Therefore, first of all, a region with various asphalt disturbances was determined.

A camera was installed in a car to create the dataset. Images were obtained so that the camera can see the road completely to the bumper of the vehicle. General features of the camera used: 12 Megapixel CMOS sensor, 720P (1280x720 pixels) 120 FPS. While the data is being recorded, the car moves steadily at 30 km/h. In the dataset, the image size is cropped to 768x480 pixels. In Figure 2, there is a sample image group from the dataset.

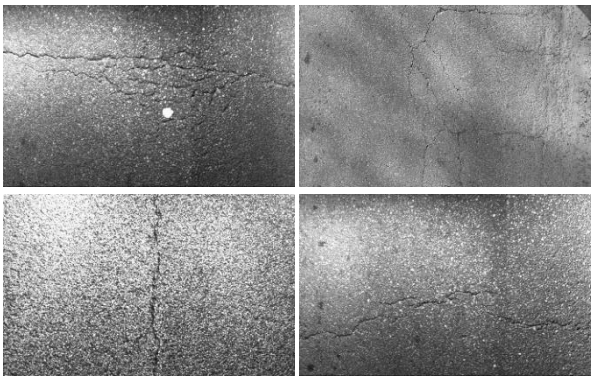


Figure 2. A subset of training dataset

4315 images were collected for the dataset. While these images contain various asphalt defects, some images consist of solid asphalt images. Various tags have been added to the properties of the images. This tag information is a six-dimensional vector. Since the Faster R-CNN structure to be used in the study is based on the CNN structure, increasing the training data slightly increases the accuracy rate, as in the CNN structure. For this reason, Affine transformation and perspective

transformation methods, which are widely used as dataset enlargement methods, were used in image processing studies. The affine transform is a linear transform technique. It provides reproduction of the image by performing some operations on the image: scaling, rotation, translation, flipping and shearing.

Perspective transformation is a 3D transformation technique as it changes the angle of the camera during the recording of the image. It affects some features of the camera with some 3D operations on the image: position, field of view and orientation. In this study, more perspective transformation has been applied in order to transform it into a study suitable for different camera angles. When expressing the perspective transformation mathematically, the initial camera position and the resulting camera image at the end of the transformation can be used [9]:

$$X_2 = HX_1, \quad X_1, X_2 \in \boldsymbol{R^3} \tag{1}$$

$$\lambda_1 x_1 = X_1, \lambda_2 x_2 = X_2 \tag{2}$$

$$\lambda_2 x_2 = H\lambda_1 x_1 \tag{3}$$

If Equation 1-3 is examined, H represents the homography matrix, λ_1 and λ_2 represent the scale coefficients. If the scale coefficients are neglected, it is seen that x_2 is equal to Hx_1 and there is a direct proportionality between these two expressions. From this ratio, the constraint equation is formed:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \tag{4}$$

If $x=x_2/z_2$, $y=y_2/z_2$ and $z_1=1$ in equation 4:

$$x = \frac{h_{11}x_1 + h_{12}y_1 + h_{13}}{h_{31}x_1 + h_{32}y_1 + h_{33}} \tag{5}$$

$$y = \frac{h_{21}x_1 + h_{22}y_1 + h_{23}}{h_{31}x_1 + h_{32}y_1 + h_{33}} \tag{6}$$

Now, the homography matrix has now become usable [9]. After applying the perspective transformation, 4 different transformed images were obtained from each image data. Close numbers of data were obtained for each label group. The characteristics of the data used for classification are summarized in Table 1.

Faster R-CNN structure. If this structure is examined in detail, the RPN structure can directly connect to the sampling layer. Thanks to this structure, Faster R-CNN architecture can detect and classify images more powerfully. Basically, the RPN structure is exactly like the CNN architecture. It uses the image data as input and detects the desired regions to be detected. It gives a score for each detected region along with this detection [13]. There are 3 different training models available for the Faster R-CNN structure. Four-Step Alternating Training method was chosen to be used in this study. As the name suggests, this method consists of four steps.

- Firstly, the training of the RPN structure takes place separately from the other parts. Two methods were used in the end-to-end training of the RPN structure: back-propagation and stochastic gradient descent. The CNN structure gets its initial weight from a previously trained structure, VGG16.
- Secondly, Fast R-CNN structure, like RPN structure, is trained separately from other parts. As with the RPN structure, the CNN structure uses VGG16 for its initial weights. However, the Fast R-CNN structure also uses the recommendations and weights from the RPN structure for training.
- Thirdly, RPN and Fast R-CNN made between fine-tuned for weights fixed structure of the joint layer RPN.
- Finally, in this step, the weights in the common layer are fixed and at the end of the third step, the new RPN structure and Fast R-CNN structure are fine-tuned.

The local proposition network utilizes anchor boxes to get the idea boxes of the element map. Faster-RCNN utilizes three arrangements of rectangular boxes with various length-width proportions (2:1, 1:1, 1:2), and each set uses three fixed-size rectangular boxes with various scales (128, 256, 512) to stack at every pixel on the component map, as displayed in Figure 4. Then, at that point, each crate was contrasted with ground truth box with ascertain the intersection over union (IOU). The crate with bigger IOU than the enormous preset limit is divided as frontal area, and the case with more modest IOU than the little preset edge is differentiated as

Table 1. Details of the dataset

Type of crack	Number of traning data	Number of augmented data	Number of testing data	Number of total data	Target values
Non-crack	135	675	200	875	$[1\ 0\ 0\ 0\ 0]^T$
Transverse Crack	125	625	200	825	$[0\ 1\ 0\ 0\ 0]^T$
Longitudinal Crack	139	695	200	895	$[0\ 0\ 1\ 0\ 0]^T$
Block Crack	134	670	200	870	$[0\ 0\ 0\ 1\ 0]^T$
Alligator Crack	130	650	200	850	$[0\ 0\ 0\ 0\ 1]^T$

2.2. Training and the proposed architecture of Faster R-CNN

ada, In this article, the Faster R-CNN architecture, which uses the CNN basis, is basically a combination of RPN and Fast R-CNN architectures [12]. Figure 4 shows the foundation. Both the cases with IOU between the two edges and the flood limit were straightforwardly disposed of. At last, we got the idea box to prepare the region proposal network. The anchor boxes enjoy brought many benefits. By setting various scales, every one of the

objectives can be covered beyond what many would consider possible, while decreasing the computation sum and significantly lessening the trouble of ensuing relapse calculation improvement.

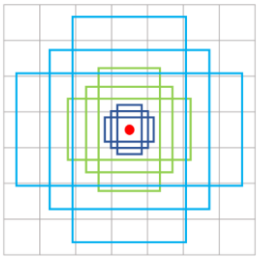


Figure 3. A subset of training dataset

Faster-RCNN can be isolated into four modules, including highlight extraction module, RPN module, ROI pooling, and target order and situating module. The general construction of the calculation is displayed in Figure 5. At present, the delivered code of Faster-RCNN typically takes the convolution-pooling part of VGG-16 as the feature extraction module.

tip RPN utilizes anchor boxes to make rectangular boxes for every pixel on the element map, appoints names through the IOU value determined with the ground truth boxes, and uses nonmaximum concealment calculation to take out covering boxes. The held closer view or foundation boxes will be allocated six factors, including four scale factors (x, y, w, h) and two name factors (Fg, Bg). The softmax calculation is then used to work out the objective likelihood score for each container. The crate relapse calculation utilizes four boundaries to finish the relapse of each anchor box, so the proposed box is nearer to the genuine position. The calculation cycle of RPN is displayed in Figure 7.

The ROI pooling coordinated feature maps and the proposition boxes produced by RPN, and proposition include maps were determined and created. As the size of the proposition box is unique, the feature of information factors ought to be fixed when the ensuing full association layer is utilized for classification. Subsequently, Faster-RCNN utilizes ROI pooling to separate the proposition highlight maps into seven equivalent parts on a level plane and in an upward direction, and plays out the greatest pool on each square, so the last size of every proposition include map is 7x7.

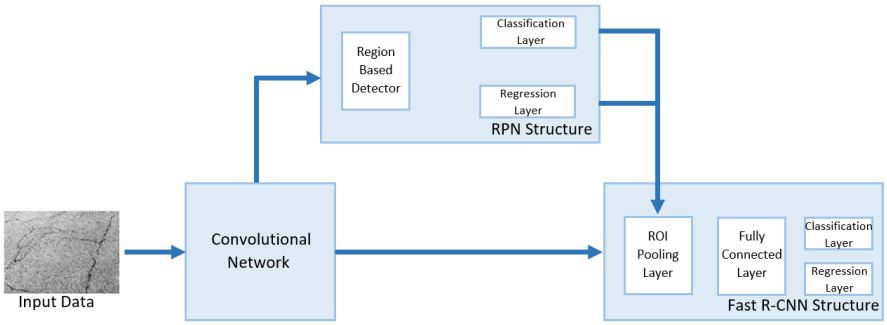


Figure 4. Faster R-CNN framework

The feature extraction module first and foremost resizes the information of any scale to a decent scale (the long side of the picture will not be more prominent than 1000, and the short side of the picture will not be more noteworthy than 600), then, at that point, separates the data features through a bunch of convolution-ReLU-pooling layers, and produces n-dimensional feature map (512-dimensional element map is created by VGG-16 organization). VGG-16 component extraction network is displayed in Figure 6.

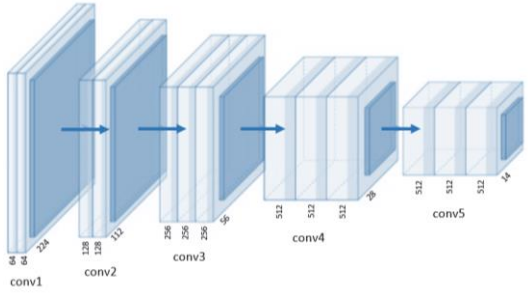


Figure 5. VGG-16 feature extract framework

The objective classification and situating module utilizes the full connection layer and softmax calculation to sort every proposition include guide, and result the likelihood vector. Simultaneously, bouncing box regression is utilized again to get the positional balanced of each recommended include map and to create a more precise recognition box.

The loss function of Faster-RCNN chiefly comprises of two sections, including the loss function of RPN and the loss function of RCNN, and every loss function work incorporates the grouping loss function and the regression loss. The classification loss $L_{cls}(*)$ was determined utilizing cross-entropy, and the regression loss $L_{reg}(*)$ was determined utilizing smooth-L1. As a result, it transfers the classification information and score value to the Fast R-CNN structure. There are two layers as output in Fast R-CNN structure. These layers are the Softmax classifier layer and the regression layer that gives the rate of detection accuracy. RPN loss function and RCNN loss function are shown in Equations 7 and 8.

$$L_{RPN}(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (7)$$

$$L_{RCNN}(p, u, t^u, v) = L_{cls}(p, u) + L_{reg}(t^u, v) \quad (8)$$

Coordinate data is needed for classification information. The classification layer checks whether the detected region exists or not, and outputs the coordinate information shown in Equation 9.

$$t_x = \frac{(x - x_a)}{w_a} t_y = \frac{(y - y_a)}{h_a}$$
$$t_w = \log\left(\frac{w}{w_a}\right) t_h = \log\left(\frac{h}{h_a}\right)$$
$$t_x^* = \frac{(x^* - x_a)}{w_a} t_y^* = \frac{(y^* - y_a)}{h_a}$$
$$t_w^* = \log\left(\frac{w^*}{w_a}\right) t_h^* = \log\left(\frac{h^*}{h_a}\right) \quad (9)$$

During the training of the Faster R-CNN structure, the photographs obtained from the selected region were used. Detailed information about the data used for training and testing purposes is shown in Table 1. The following three improvement strategies are proposed, including data augmentation, k-means clustering to generate anchor boxes, and transfer learning with ResNet-101.

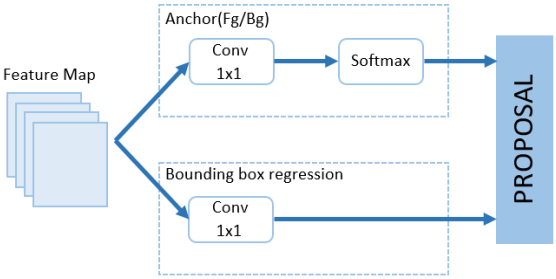


Figure 6. RPN framework

The size of anchor confines the Faster-RCNN calculation alludes to the objective size in VOC 2007 and VOC 2012 dataset. The produced boxes can cover the greater part of the objectives, and a more exact idea box can be acquired through regression. Albeit the authority anchor box can get a more precise idea box via preparing the regression coefficient, the change scale is excessively enormous, which isn't helpful for union. Consequently, k-means clustering is embraced in this paper to investigate all preparation information and get a bunch of anchor boxes with more fitting scale. K-means clustering pseudo code is displayed in Algorithm 1.

Anchor boxes produced by k-means clustering are more in accordance with the genuine size of location targets contrasted and the preset anchor boxes dependent on experience, which is helpful for regression and acquiring an idea box with higher precision. As a rule, the bigger a k worth is, the more anchor boxes are produced and the higher the exactness is. Notwithstanding, when k worth is expanded partially, the exactness is fundamentally unaltered, while such a large number of idea boxes will extraordinarily decrease the computational productivity of the organization. The normal IOU of anchor boxes produced by k-implies grouping under various k qualities are displayed in Figure 8.

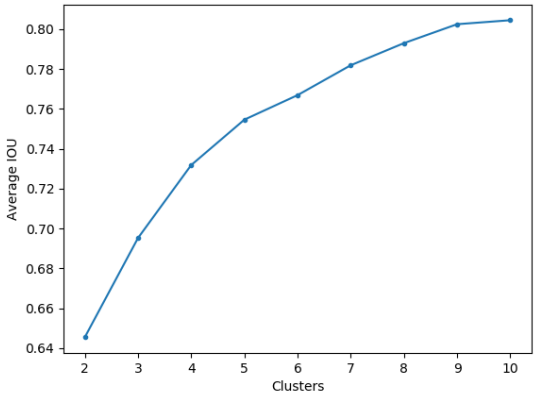


Figure 7. Average IOU and clusters relation

In the objective identification task, the profundity of the convolutional neural organization is normally up to handfuls or even many layers because of the intricacy of the shape, shading, and different qualities of the objective in the picture. Be that as it may, the further layer isn't helpful for the boundary streamlining of the organization, and slope blast or inclination vanishing are probably going to happen during the preparation interaction [25]. Despite the fact that clump standard, arbitrary slope plunge and different calculations can be utilized to accomplish a specific level of enhancement, the impact isn't self-evident. In this manner, in 2015, He proposed another organization structure unit, which is known as the lingering unit. In the mean time, in 2016, He worked on the remaining unit to make it more straightforward to prepare and expand its speculation capacity [26]. The excess units can be separated into 2 layers of residual

Algorithm 1. K-means clustering pseudo code.

Ground Truth Boxes: $\{(w_i^*, h_i^*)\}, i \in [1, N]$
Initialize k clustering centers: $\{(w_j, h_j),\ j \in [1, k]\}$
Iteration: stop if the change of new clustering centers is smaller than threshold.
for *i* in range(N):
  for *j* in range(k):
    $d_i(j) = 1 - IOU((w_i^*, h_i^*), (w_j, h_j))$
    $cls(i) = Index(\min d_i(j))$
  Calculate new k clustering centers: $w_j' = \frac{1}{N_j}\sum w_i^*(j), h_j' = \frac{1}{N_j}\sum h_i^*(j), i \in [1, N], j \in [1, k]$

units and 3 layers of outstanding units, as displayed in Figure 9. The profound convolutional network comprising of the excess units is called ResNet.

As of now, ResNet is generally utilized because of its momentous boundary advancement capacity in preparing [27–30]. In this paper, considering the precision of

Table 2. ResNet structures informations

Layer	Output Size	18-Layer	34-Layer	50-Layer	101-Layer	152-Layer
conv1	112x112	7x7, 64, stride 2				
conv2_x	56x56	3x3 max pool, stride 2				
		$\begin{bmatrix}3x3 & 64\\3x3 & 64\end{bmatrix}x2$	$\begin{bmatrix}3x3 & 64\\3x3 & 64\end{bmatrix}x2$	$\begin{bmatrix}1x1 & 64\\3x3 & 64\\1x1 & 256\end{bmatrix}x3$	$\begin{bmatrix}1x1 & 64\\3x3 & 64\\1x1 & 256\end{bmatrix}x3$	$\begin{bmatrix}1x1 & 64\\3x3 & 64\\1x1 & 256\end{bmatrix}x3$
conv3_x	28x28	$\begin{bmatrix}3x3 & 128\\3x3 & 128\end{bmatrix}x2$	$\begin{bmatrix}3x3 & 128\\3x3 & 64\end{bmatrix}x4$	$\begin{bmatrix}1x1 & 128\\3x3 & 128\\1x1 & 512\end{bmatrix}x4$	$\begin{bmatrix}1x1 & 128\\3x3 & 128\\1x1 & 512\end{bmatrix}x4$	$\begin{bmatrix}1x1 & 128\\3x3 & 128\\1x1 & 512\end{bmatrix}x8$
conv4_x	14x14	$\begin{bmatrix}3x3 & 256\\3x3 & 256\end{bmatrix}x2$	$\begin{bmatrix}3x3 & 256\\3x3 & 256\end{bmatrix}x6$	$\begin{bmatrix}1x1 & 256\\3x3 & 256\\1x1 & 1024\end{bmatrix}x6$	$\begin{bmatrix}1x1 & 256\\3x3 & 256\\1x1 & 1024\end{bmatrix}x23$	$\begin{bmatrix}1x1 & 256\\3x3 & 256\\1x1 & 1024\end{bmatrix}x36$
conv5_x	7x7	$\begin{bmatrix}3x3 & 512\\3x3 & 512\end{bmatrix}x2$	$\begin{bmatrix}3x3 & 512\\3x3 & 512\end{bmatrix}x3$	$\begin{bmatrix}1x1 & 512\\3x3 & 512\\1x1 & 2048\end{bmatrix}x3$	$\begin{bmatrix}1x1 & 512\\3x3 & 512\\1x1 & 2048\end{bmatrix}x3$	$\begin{bmatrix}1x1 & 512\\3x3 & 512\\1x1 & 2048\end{bmatrix}x3$
1x1	Average pool, 1000-d fc, softmax					
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9
Top-1 error		27.88%	25.03%	22.85%	21.75%	21.43%

The remaining unit changes the activity method of the conventional convolutional network, in which, the last worth acquired by convolutional layer and ReLU is displayed in Equation 10.

$$H(x) = F(x) + x \tag{10}$$

F(x) is the contrast between the result H(x) and the information x, in particular the leftover. In an ideal circumstance, when the organization arrives at a specific profundity, assuming the organization state is as of now ideal, F(x) ought to be set as 0, which is comparable to the result H(x) of the leftover unit as x, so the current profundity network doesn't debase, which guarantees the exactness of the profound organization model.

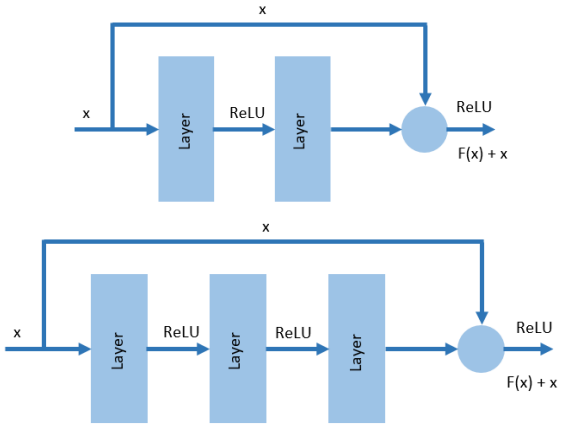


Figure 8. 2 and 3 layers residual units

ResNet has an assortment of organization structures, which can be isolated into 18, 34, 50, 101 and 152 layers as indicated by the convolutional layer profundity. The remaining units contained in various organization structures are somewhat unique, as shown in Table 2. ResNet-18 and ResNet-34 are made out of 2-layers lingering units, while ResNet-50, ResNet-101 and ResNet-152 are made out of 3-layers remaining units.

imperfection recognition and real figuring capacity, the convolution part of ResNet-101 was chosen as the component extraction module, and the first VGG-16 was supplanted to work on the exactness of the last classification results.

3. RESULTS and DISCUSSION

In this study, the Faster R-CNN structure was created and its performance in detecting cracks on asphalt was analyzed. While performing the experiments, a computer with the following features was used: Intel i7-9750H processor of 4.5 GHz and NVIDIA GTX 1650 graphicscard. The programming language used in this study is Python. Various libraries have been used: TensorFlow, Pandas, Numpy. In this study, a dataset containing 6141 asphalt images was created and the images were labeled with four types of cracks frequently encountered in the literature. These tags are: non-crack, transverse crack, longitudinal crack, block crack and alligator crack. In order to reproduce the training data, the dataset was expanded by applying perspective transformation to the data other than the test data. 5 different image samples with perspective transformation and the homography matrices of these samples are shown in Figure 5 and Equation 11.

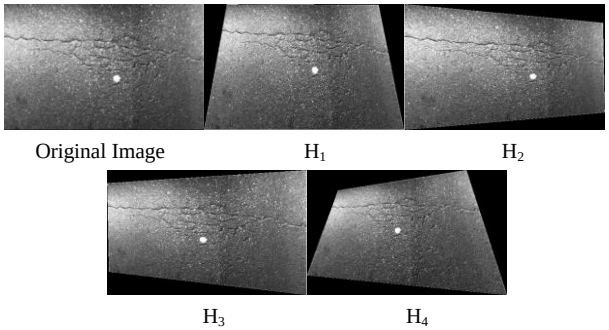
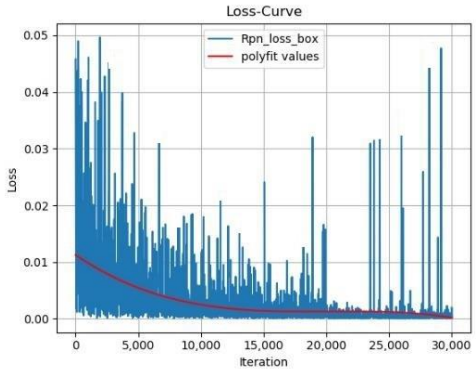


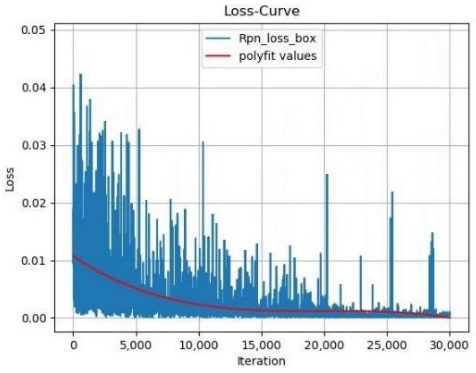
Figure 9. Augmented images by perspective transformation

$$H_1 = \begin{bmatrix} 0.73 & -0.13 & 38.9 \\ 0 & 0.7 & 13.8 \\ 0 & 0.001 & 1 \end{bmatrix}$$
$$H_2 = \begin{bmatrix} 1.65 & 0 & -44.8 \\ 0.23 & 1.2 & -22.4 \\ 0.0002 & 0 & 1 \end{bmatrix}$$
$$H_3 = \begin{bmatrix} 0.6 & 0.6 & 50 \\ -0.15 & 0.8 & 49.8 \\ -0.0008 & 0.0006 & 1 \end{bmatrix}$$
$$H_4 = \begin{bmatrix} 0.45 & 0.21 & 73.2 \\ -0.17 & 0.545 & 57.9 \\ -0.0007 & -0.0007 & 1 \end{bmatrix} \quad (11)$$

In this study, 9 fixed anchor boxes are used in Faster R-CNN structure. These anchor boxes used can create position information on asphalt images to cover cracks, if any. However, due to the size of the offset value in the regression structure, it causes the anchor boxes to detect incorrect locations. In order to prevent this, anchor boxes were created using the k-means clustering structure. By dint of this structure, it has been determined that an improvement is achieved in the regression loss curve graph. If Figure 11 is examined, the curve shown in red is the data fitting curve. Two different loss curve graphs that emerged as a result of iterations are shown. By using the K-means clustering structure, the loss average was reduced from 1.902x10-4 to 1.124x10-4.



(a)



(b)

Figure 10. RPN loss curves. (a) Standart method for setting anchor boxes; (b) Anchor boxes that created using the k-means clustering structure

In addition to the standard Faster R-CNN structure, performance measurements were carried out using different methods such as VGG-16 and ResNet-101. The performance value increased as a result of using the ResNet-101 structure as a feature extraction module is shown in Table 3 comparatively. Brief information about the methods applied using the same data set and the obtained mean accuracy precision metric results are listed in Table 3.

Table 3. Mean precision value information of models

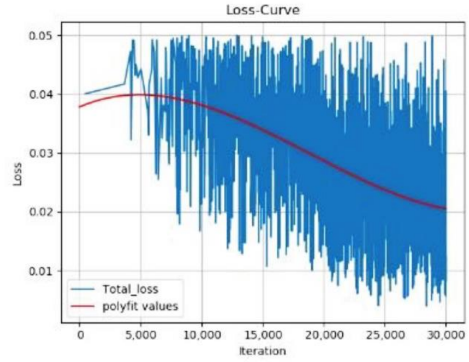
Training Model	**mAP**
VGG-16	0.8235
VGG-16 & Augmented Data	0.8423
VGG-16 & Augmented Data & Clustering	0.8645
ResNet-101 & Augmented Data	0.9134
Proposed Method	0.939

The dataset to be used for training consists of augmented data, while the data reserved for testing is not expanded. That is, the dataset used for testing was limited to the images in the original dataset. Approximately 70% of the 1280x720 images were used as training data. During the training phase, the steps of the Four-Step Alternating Training method were followed sequentially. Initially, the training of the RPN structure was reduced from 0.001 in 2/3 of the iteration to 0.0001 in the remaining 1/3. Then the same values were applied in the Fast R-CNN structure. Then, fine-tuning is done for the RPN and Fast R-CNN structure, respectively. The Faster R-CNN structure was designed and the performance analysis of the system was measured with the dataset reserved for testing.

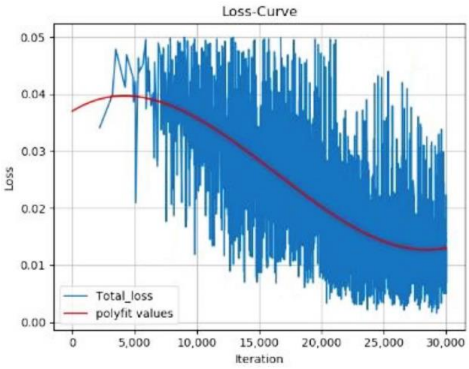
In order to make distinctions related to detection and classification, various colors are used to enable the distinction to be visually understood. Visual separation is enhanced by using the #B5E51E color code for the Transverse crack, the #3F47CC color code for the Block crack, the #FEF200 color code for the Alligator crack, and the #FF528B color code for the Longitudinal crack. 200 tests were performed for each piece of dataset. A complexity matrix was created in order to show the results of the tests performed in a meaningful way. The test dataset was randomly selected according to the labels. Figure 13 shows some results of the tests performed on the image. The average duration of each test is 0.05 seconds. Some scales are used to make test results meaningful for each tag: Recall, precision, overall accuracy and kappa [14, 15]. Detailed information about these values is shown in Table 4. The average precision for all tags is greater than 0.90, of which Non-crack gets the highest score of 94% and the Longitudinal Crack gets the lowest score of 92%. Mean precision score of all tag is 93.2%. Finally, the Kappa value is 0.915. Total loss graphs are shown in Figure 12 to determine the difference between the Standard Faster R-CNN and the approach proposed in this study. If these graphs are examined, it is seen that the polyfit value shown in red has decreased.

Table 4. Detection and classifications results

		Truth Data						
		N-c	TC	LC	BC	AC	**Test**	**Precision**
Classifier Results	N-c	**188**	5	4	2	1	200	94%
	TC	4	**187**	4	3	2	200	93.5%
	LC	2	5	**184**	7	2	200	92%
	BC	3	2	5	**186**	4	200	93%
	AC	0	3	4	6	**187**	200	93.5%
	Truth Overall	197	202	201	204	196		
	Recall	95.431%	92.574%	91.542%	91.176%	96.408%		



(a)



(b)

Figure 11. Total loss curve. (a) Faster R-CNN without data augmentation or another processes; (b) Proposed method

Unlike the studies conducted with Faster R-CNN in the literature, this study gained originality by classifying according to different cracks, not whether there are cracks or not [16]. These results show that the Faster R-CNN structure is an impressive method for the detection of cracks on asphalt. In order to compare the studies in the literature and this new approach proposed, Table 5 lists some results. When these results are examined, it is seen that the results that can be included in the literature are obtained both in terms of performance and classification diversity. When the studies in the literature are examined, the accuracy rate increases when the number of classifications decreases. It is seen that the proposed approach in this study has an optimal result in terms of both the number of classifications and the accuracy rate.

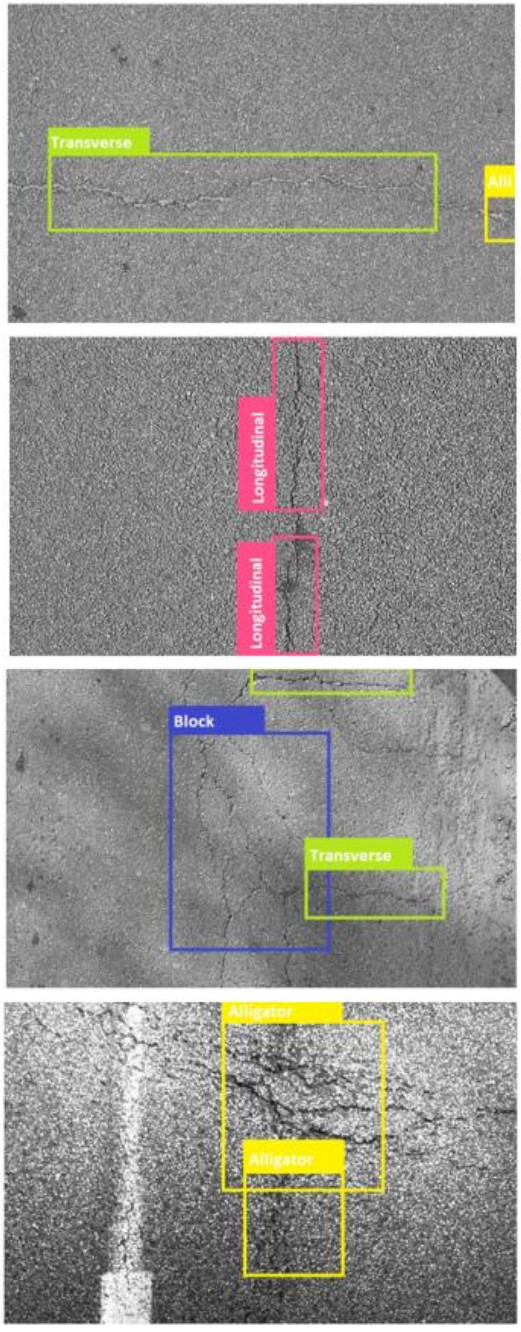


Figure 12. Detection examples from proposed method

Table 5. Comparative summary table of the studies in the literature

Study	Method	Remarks
Du et al. [17]	YOLO network. 7 types pavement distress are classified.	Accuracy: 73.64%
Gao et al. [18]	Faster R-ConvNet. 3 types pavement distress are classified.	Precision: 89.13%
Mei et al. [19]	5 Layer fully convolutional network. Crack or Non-crack classification.	Precision: 80.88% Recall: 76.64%
Huidrom et al. [20]	CDDMC Algorithm. 6 types pavement distress are classified.	Accuracy: 97%, 94% and 90% Precision: 95%, 93% and 8.5%
Ibragimov et al. [21]	R-CNN. 3 types pavement distress are classified.	Precision: 38.15%, 78.53% and 84%
Cha et al. [22]	CNN. Crack or Non-crack classification.	Accuracy: 97.42%
Proposed method	Faster R-CNN &ResNet-101 & K-Means Clustering. 5 types pavement distress are classified.	Accuracy: 93.2%

4. CONCLUSION

Deterioration occurs on asphalt over time due to various reasons. Detection of these deteriorations is in smart city planning. For this reason, studies on automatic deterioration detection are important. In this study, Faster R-CNN structure was created that performs the classification of asphalt degradation. The average accuracy rate of the method used in this study was 93.2%. However, there were some points that were determined during the study. The most important of these is the detection of pits formed without cracks. For future studies, the compression distances that occur in the shock absorbers of the vehicles where the camera is placed can be used as a separate feature, so that these deteriorations can be detected. With similar studies, the accuracy rate can be increased and the maintenance and repair process can be done in a short time by placing it on public transport vehicles and making instant detections.

NOMENCLATURE

CNN: Convolutional Neural Network

RPN: Region Proposal Network

Fast R-CNN: Fast Region-based Convolutional Neural Network

Faster R-CNN: Faster Region-based Convolutional Neural Network

ReLU: Rectified Linear Unit

ROI: Region of Interest

N-c: Non-crack

TC: Transverse Crack

LC: Longitudinal Crack

BC: Block Crack

AC: Alligator Crack

DECLARATION OF ETHICAL STANDARDS

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

AUTHORS' CONTRIBUTIONS

Furkan BALCI: Performed the experiments and analyse the results.

Safiye YILMAZ: Performed the experiments and analyse the results. Wrote the manuscript.

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] Gopalakrishnan K., Khaitan S. K., Choudhary A. and Agrawal A., "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection", ***Construction and Building Materials***, 157: 322-330, (2017).

[2] Bello-Salau H., Aibinu A. M., Onwuka E. N., Dukiya J. J., Onumanyi A. J. and Ighagbon A. O., "Development of a laboratory model for automated road defect detection", ***Journal of Telecommunication, Electronic and Computer Engineering***, 8: 97-101, (2016).

[3] Shi Y., Cui L., Qi Z., Meng F. and Chen Z., "Automatic road crack detection using random structured forests", ***IEEE Transactions on Intelligent Transportation Systems***, 17: 1-12, (2016).

[4] Li B., Wang K. C. P., Zhang A., Yang E. and Wang G., "Automatic classification of pavement crack using deep convolutional neural network", ***International Journal of Pavement Engineering***, 21: 457-463, (2020).

[5] Majidifard H., Jin P., Adu-Gyamfi Y. and Buttlar W. G., "Pavement image datasets: a new benchmark dataset to classify and densify pavement distresses", ***Transportation Research Record***, 2674: 328-339, (2020).

[6] Zhang D., Li Q., Chen Y., Cao M., He L. and Zhang B., "An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection", ***Image and Vision Computing***, 57: 130-146, (2017).

[7] Shahnazari H., Tutunchian M. A., Mashayekhi M. and Amini A. A., "Application of soft computing for prediction of pavement condition index", ***Journal of Transportation Engineering***, 138: 1495-1506, (2012).

[8] Xu W. and Tang Z., "Pavement crack detection based on saliency and statistical features", ***IEEE International Conference on Image Processing***, Melbourne Australia, 175–198, (2013).

[9] Dubrofsky E., "Homography Estimation",***Master of Science Thesis***, the University of British Columbia, (2009).

[10] Simonyan K. and Zisserman B., "Very deep convolutional networks for large-scale image recognition", ***the International Conference on Learning Representations***, San Diego USA, 1-15, (2015).

[11] Xie D., Zhang L. and Bai L., "Deep learning in visual computing and signal processing", ***Applied Computational Intelligence and Soft Compututing***, 2017: 1–13, (2017).

[12] Girshick R., "Fast R-CNN", ***the IEEE International Conference on Computer Vision***, Santiago Chile, 1440-1448, (2015).

[13] Ren S., He K., Girshick R. and Sun J., "Faster R-CNN: Towards real-time object detection with region proposal networks", ***the Advances in Neural Information Processing Systems***, 91-99, (2015).

[14] Everingham M., Van Gool L., Williams C. K., Winn J. and Zisserman A., "The pascal visual object classes (VOC) challenge", ***International Journal of Computer Vision***, 88: 303-338, (2010).

[15] Turpin A. and Scholer F., "User performance versus precision measures for simple search tasks", ***the ACM International Conference on Research and Development in Information Retrieval***, Washington USA, 11-18, (2006).

[16] Song L. and Wang X., "Faster region convolutional neural network for automated pavement distress detection", ***Road Materials and Pavement Design***, 22: 23-41, (2021).

[17] Du Y., Pan N., Xu Z., Deng F., Shen Y. and Kang H., "Pavement distress detection and classification based on YOLO network", ***International Journal of Pavement Engineering***, in press.

[18] Gao J., Yuan D., Tong Z., Yang J. and Yu D., "Autonomous pavement distress detection using ground penetrating radar and region-based deep learning", ***Measurement***, 164: 108077, (2020).

[19] Mei Q. and Gül M., "A cost effective solution for pavement crack inspection using cameras and deep neural networks", ***Construction and Building Materials***, 256:119397, (2020).

[20] Huidrom L., Das L. K. and Sud S. K., "Method for automated assessment of potholes, cracks and patches from road surface video clips", ***Procedia-Social and Behavioral Sciences***, 104, 312-321, (2013).

[21] Ibragimov E., Lee H. J., Lee J. J. and Kim N., "Automated pavement distress detection using region based convolutional neural networks", ***International Journal of Pavement Engineering***, 1-12, (2020).

[22] Cha Y. J., Choi W. And Büyüköztürk O., "Deep learning-based crack damage detection using convolutional neural networks", ***Computer-Aided Civil and Infrastructure Engineering***, 32(5): 361-378, (2017)