AUTOMATION ASPHALT PAVEMENTS DISTRESS DETECTION USING DEEP LEARNING-BASED RETINANET

TỰ ĐỘNG HÓA PHÁT HIỆN VẾT NỨT TRONG MẶT ĐƯỜNG NHỰA SỬ DỤNG THUẬT TOÁN HỌC SÂU RETINANET

[1*]Van Phuc Tran, [2]Thai Son Tran, [3]Van Phuc Le, [4]Hyun Jong Lee

[1,2,4] Department of Civil and Environmental Engineering, Sejong University, South Korea

[3]Faculty of Civil Engineering, University ofTransport and Communications, Hanoi, Vietnam

[1]vanphuctran053@gmail.com*

Abstract: *In this paper, the author proposed a supervised machine learning network to identify and classify different crack types in asphalt-surface pavements. A laser camera captured surface images from pavement surface then classified them into 3 classes following the manual of pavement distress identification by the Federal Highways Administration (FHWA). These classes are three different crack types: alligator (fatigue), longitudinal, and transverse cracks. The training database was collected from 1,000 images with the original size of 3,704x10,000 pixels. These images then were divided into 20,000 smaller images of 1,852x1,000 pixels size. The images data are labeled based on the nine crack types and trained using a deep learning algorithm called RetinaNet. The trained model is verified using 2,400m of pavement surface images obtained from Seoul city urban road. The results have shown that the trained network model has an accuracy of around 85% for crack detection and classification.*

Keywords: *Automated crack detection, asphalt crack detection, deep learning, RetinaNet.*

Classification number: *11.2*

Tóm tắt: *Trong bài báo này, nhóm tác giả đề xuất một mạng máy học có giám sát để xác định và phân loại các dạng vết nứt khác nhau trên mặt đường nhựa. Laser camera được sử dụng để chụp lại hình ảnh mặt đường nhựa, sau đó các ảnh này được phân thành ba loại theo hướng dẫn phân loại vết nứt trên mặt đường của Cục Quản lý Đường cao tốc Liên bang Mỹ (FHWA). Các dạng phá hoại này được phân thành ba loại vết nứt khác nhau: Vết nứt thành lưới (mỏi), vết nứt dọc và vết nứt ngang. Dữ liệu sử dụng để huấn luyện mạng học sâu được thu thập từ 1.000 hình ảnh với kích thước ban đầu là 3.704 x 10.000 pixels. Những hình ảnh này sau đó được chia thành 20.000 hình ảnh nhỏ hơn có kích thước 1.852 x 1.000 pixels. Những ảnh này được gắn nhãn dựa trên chín loại vết nứt và được huận luyện dựa trên thuật toán học sâu gọi là RetinaNet. 2.400m, ảnh mặt đường đô thị khảo sát tại thành phố Seoul được dùng để kiểm tra độ chính xác của mô hình huấn luyện. Kết quả cho thấy phương pháp đề xuất có độ chính xác khoảng 85% trong việc phát hiện và phân loại vết nứt.*

Từ khóa: *Phát hiện vết nứt tự động, phát hiện vết nứt nhựa đường, học sâu, RetinaNet.*

Mã phân loại: *11.2*

1. Introduction

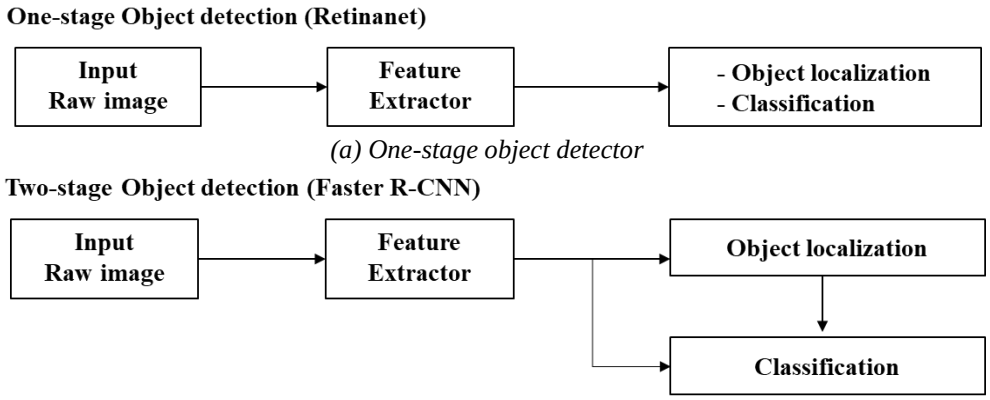
In a pavement management system (PMS), determination of pavement distress is essential to properly determine the appropriate rehabilitation method for the existing pavement condition. In asphalt pavement, the main distress necessary in calculating the Pavement Condition Index (PCI) includes rutting, cracking, patching, and pothole. Pavement rut depths can be obtained by analyzing captured rutting profiles from the vehicle-mounted 3D scanner. Meanwhile, other distresses can be determined by checking surveyed images manually. The current method of manually evaluating and analyzing the pavement condition in terms of distresses is uneconomical and time-consuming. With the huge demand for pavement rehabilitation due to the rapid development of urban cities, a new method of detecting pavement distresses that is fast, accurate, and reliable is significantly needed.

Currently, several studies have been conducted focusing on automated pavement distress detection systems. In the past decade, image-based algorithms of crack detection have been widely investigated; some techniques like thresholding [1], edge detection [2], and mathematical morphology [3] are the most popular approaches among the algorithms. One example is the study of

Koch et al. [4] that proposed an algorithm for automated pothole detection based on asphalt pavement surface images. This method utilizes image segmentation to divide the pavement surface image into the defect and non-defect regions. Then, the pothole location is figured based on the geometric properties of a defective region. Meanwhile, Ying et al. [5] used Beamlet transform-based technique for pavement crack detection, and classification wherein background pixels were utilized to identify uneven illumination by eliminating the crack and suspicious pixels from the process. The Beamlet transform is then used to extract the crack features. Finally, the extracted pavement cracks are linked together and classified. (Disadvantage of the current techniques).

Recently, Krizhevsky et al. [6] proposed a deep convolutional neural network architecture (AlexNet) that won ImageNet in 2012. Convolutional neural networks (CNN) have become the standard for image classification, wherein researchers applied this approach to detect distress in asphalt pavement. Fan et al. [7] proposed a method to detect pavement crack automatically based on a trained network using CNN; meanwhile, Li et al. [8] detect cracking in the original concrete surface by using the CNN network. In recent years, the CNN classification accuracy has improved, finally exceeding humans' capacity based on the ImageNet challenge. However, image classification is just a simple task compared to human visual awareness. For the real object classification, the task is to say what is the image label. On the other hand, object detection requires finding various objects in an image and classifying the object class. The limitation of traditional CNN is that its failure to handle images with multi-object. Because of this, Girshick et al. [9], in 2013, proposed a new algorithm called Regional CNN (R-CNN). This method can exactly detect the objects by applying bounding boxes even there are multiple and overlapping objects and complicated backgrounds in images. In 2015, the R-CNN was modified and improved by Ross Girshick, wherein a new version was developed called Fast R-CNN [10]. By 2016, this team proposed an updated version named Faster R-CNN [11]. Later, the Faster R-CNN is one of the most successful object detection algorithms with the highest precision. However, Faster R-CNN is quite slow in training and detecting objects in real-time because it is a two-stage detector algorithm. In the year 2018, the same team developed a better algorithm that can solve the training and detection issues of the previous algorithm called RetinaNet [12].

In this paper, the RetinaNet network was used to establish an automated detection model to identifying asphalt pavement from captured images. This method can classify the asphalt crack type of distress as described in the asphalt pavement crack identification manual from the Federal Highway Administration (FHWA) [13].



(a) One-stage object detector

(b) Two-stage object detector

Figure 1. *The one-stage and two-stage detector processing: (a) One-stage object detector; (b) Two-stage object detector.*

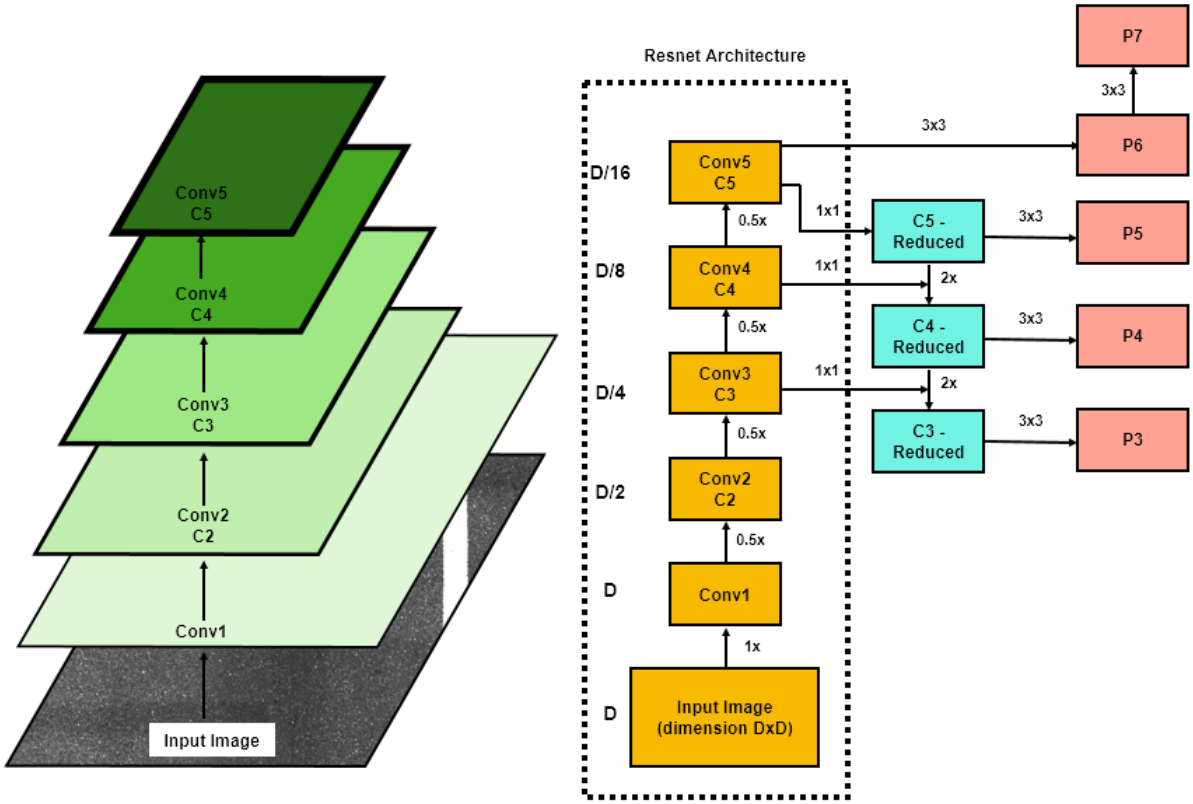


Figure 2. *The RetinaNet feature pyramid network.*

2. Background

The object detection algorithm can categorize into two groups: one-stage and two-stage algorithm, as presented in figure 1a and 1b. RetinaNet developed by Lin et al. [12], is a one-stage and state-of-the-art detection approach with significant high accuracy and testing speed compared with two-stage detectors (like Faster R-CNN [11]) approaches. Generally, one-stage detectors are applied over a regular, dense sampling of possible object locations having the potential to be faster and simpler but have trailed the accuracy of two-stage detectors. The authors of RetinaNet investigated this concern and improved the performance of RetinaNet by applying a novel loss function called Focal Loss which allows it to focus more on difficult samples. As shown in figure 2, RetinaNet is a unified network composed of a backbone network and two task-specific subnetworks. Backbone used the Feature Pyramid network architect built based on Resnet152 network [14] in figuring convolutional feature maps throughout an input image. The first sub-network, called the Class subnet, is used for calculating the probability of object presence at each location. It predicts KxN likely objects (N is the number of classes, and K is the entire quantity of anchors box). The second subnet is a box subnet that is utilized to predict the coordinates of a bounding box for a potential object. There are Kx4 (x1, y1, x2, y2 for each anchor box) coordinate values for the whole bounding box. The independent multiple networks are composed for classification and regression in RetinaNet's architecture cause RetinaNet is simpler than a two-stage object detector.

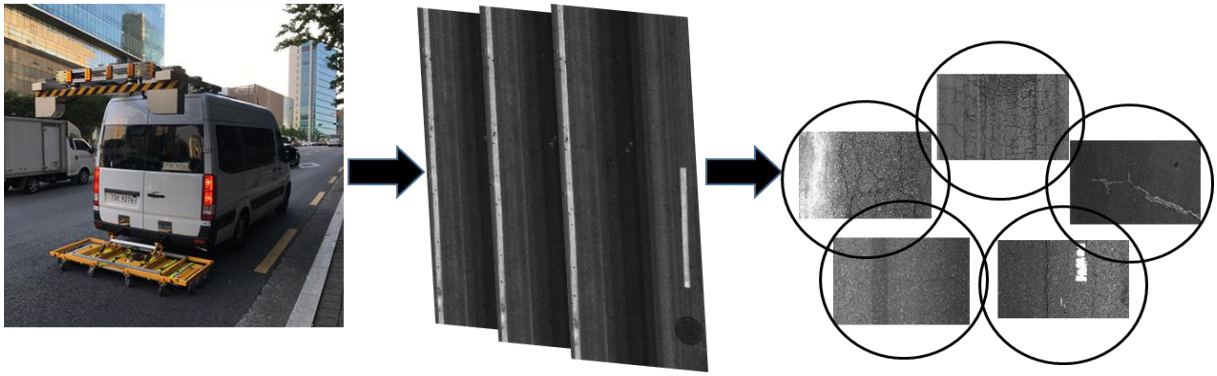


Figure 3. *Obtaining and processing data: (a) scanner vehicle with CMOS laser sensor camera; (b) surface images obtained from the laser camera; and (c) split images for establishing dataset.*

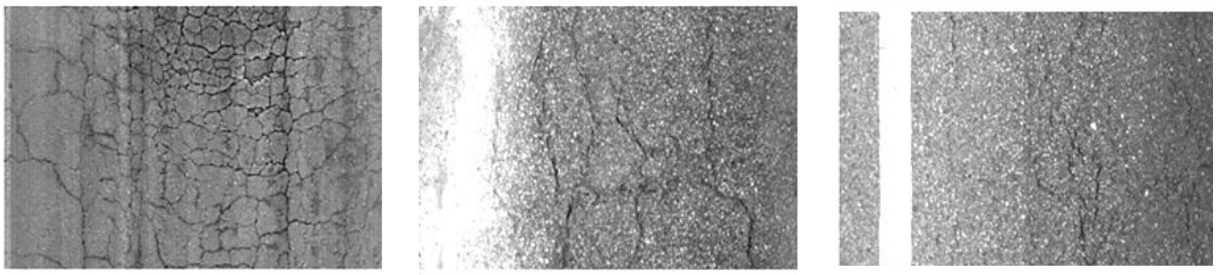
3. Methodology

In this research, a laser camera was mounted on an automated survey vehicle to collect asphalt surface images, as shown in figure 3. As the car travels on the path, a pair of cameras captured surface images. The camera has a CMOS laser sensor with a maximum resolution of 3,704 x 10,000 pixels (that represents 3.704 x 10 meters of the road) in which the captured images can be processed in both daylight and night conditions. All image datasets were collected in the Seoul City area from 2019 to 2020. These captured images were used in developing the training database with a resolution size of 3,704x10,000 pixels (width x height), wherein 1 mm is equal to 1 pixel. Each big image is split into 20 small images with 1,852 x 1,000 pixels size (figure 3c), then an image annotation tool is used to create image labels for object detection. The small images are labeled based on different crack types such as transverse, longitudinal, and fatigue cracking with three levels: high, medium, and low severity. To validate the training network model, different road sections in Seoul City urban road with a total of 2.4 km length and 3.740 meters width were collected by the survey vehicle. After data processing, the camera export one image for 10 meters. Therefore, 240 images (with resolution 3,740 x 10,000 pixels) were used to test these road sections. After that, these images were analysed by the proposed network model.

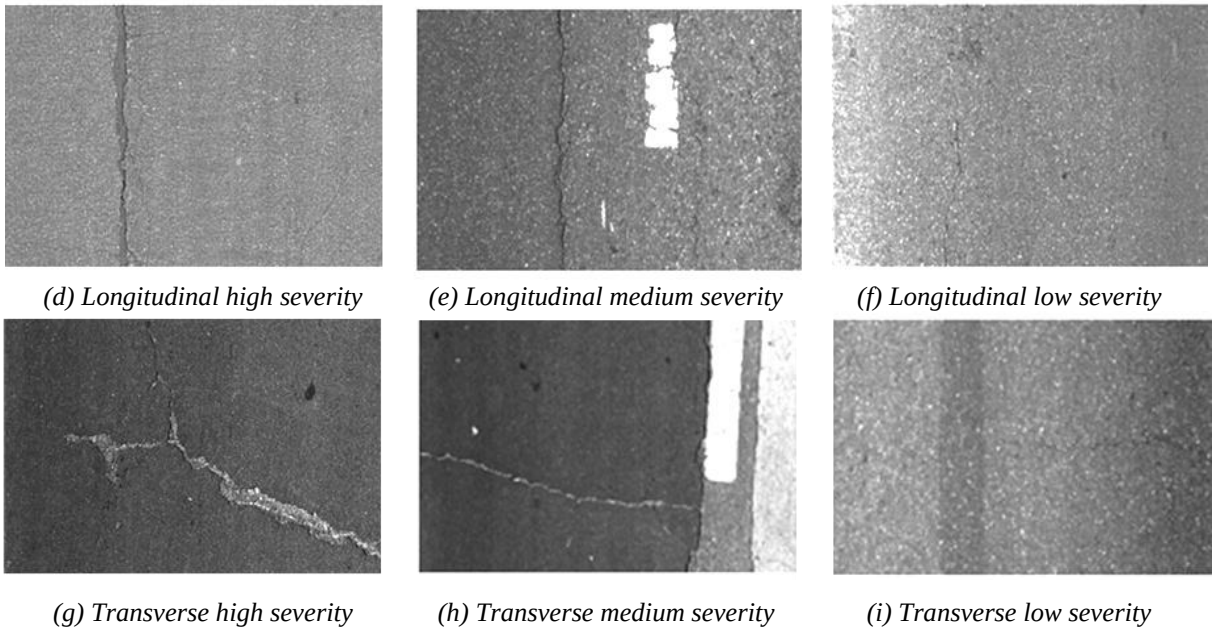
4. Training and Testing Process

4.1. Training Process

The images obtained from road survey sections were divided into nine groups then labeled to establish a training database, as shown in Figure 4. Figures 4a, 4b, and 4c shown fatigue cracks with low, medium, and high severity levels. Figure 4d to figure 4i shows longitudinal and transverse cracking examples with three levels: high, medium, and low severity.



(a) Fatigue high severity *(b) Fatigue medium severity* *(c) Fatigue low severity*



(d) Longitudinal high severity *(e) Longitudinal medium severity* *(f) Longitudinal low severity*

(g) Transverse high severity *(h) Transverse medium severity* *(i) Transverse low severity*

Figure 3. *Training crack pattern examples.*

The obtained image with the size of 3,704 x 10,000 pixels from the survey vehicle matches a pavement area of 3.704 x 10 meters. Considering the image width of 3.7 m is broader than an individual lane with less than 3.6 m, it is important to recognize the lane markers to reduce the cracks outside of the lane markers area. Moreover, the road width may not be consistent, especially on urban roads. Thus, lane markers are considered objects needed to be distinguished from calculating cracking percent in one lane more precisely. Figures 4c, 4e, and 4h exhibit lane marker example images used for training.

There are 20,000 images with 1,852x1,000 pixels size prepared for training crack and lane marker detection. The amounts of the image database for each object as follows:

- Fatigue crack high level: 1,526 objects;
- Fatigue crack medium level: 2,563 objects;
- Fatigue crack low level: 2,852 objects;
- Longitudinal crack high level: 1,625 objects;
- Longitudinal crack medium level: 2,924 objects;
- Longitudinal crack low level: 2,755 objects;
- Transverse crack high level: 1,011 objects;
- Transverse crack medium level: 2,646 objects;
- Transverse crack low level: 2,688 objects;
- Lane marker: 3,564 objects;

4.2. Testing process

The raw image with 3,704 x 10,000 pixels size is split into 20 small images with 1,852 x 1,000 pixels size during the testing of the trained network model. Lane markers and cracks surface distress from the split images are then recognized using the trained model. Following testing, the 20 split images were merged to their initial 3,704 x 10,000 pixels size. The whole process shows in figure 5.

Figure 6 presents a sample of image results after applying the proposed network for lane marker and crack type detection. As observed in the figure, each box with a particular color describing a specific crack type (i.e., green, yellow, or red colors represent transverse, longitudinal, or fatigue cracks). Each crack type's severity levels are represented by the thickness and type of the line used in each rectangle box. The dashed line describes a medium level, the thin and thick solid lines represent low and high levels.

Still, in figure 6, each bounding box's diagonal distance approximately expresses the particular linear crack length. The entire length of one linear crack class is determined by getting the sum of the individual bounding box's diagonal distance corresponding to the linear cracks' specific class. Likewise, the whole area of the fatigue crack is computed by taking the bounding area of the fatigue cracks box. The network used in this paper spends about three hours training 20,000 images for each epoch (The computer used for analysis is a CPU Intel®Core(TM) i7-8700K @ 3.70 GHz with Graphic card GeForce GTX 1080). Meanwhile, it takes six seconds to test one image using the original size of 3,704 x 10,000 pixels corresponding to a 10m long section of pavement. It means that the network model can analyze approximately 5 km of road surface image per hour.

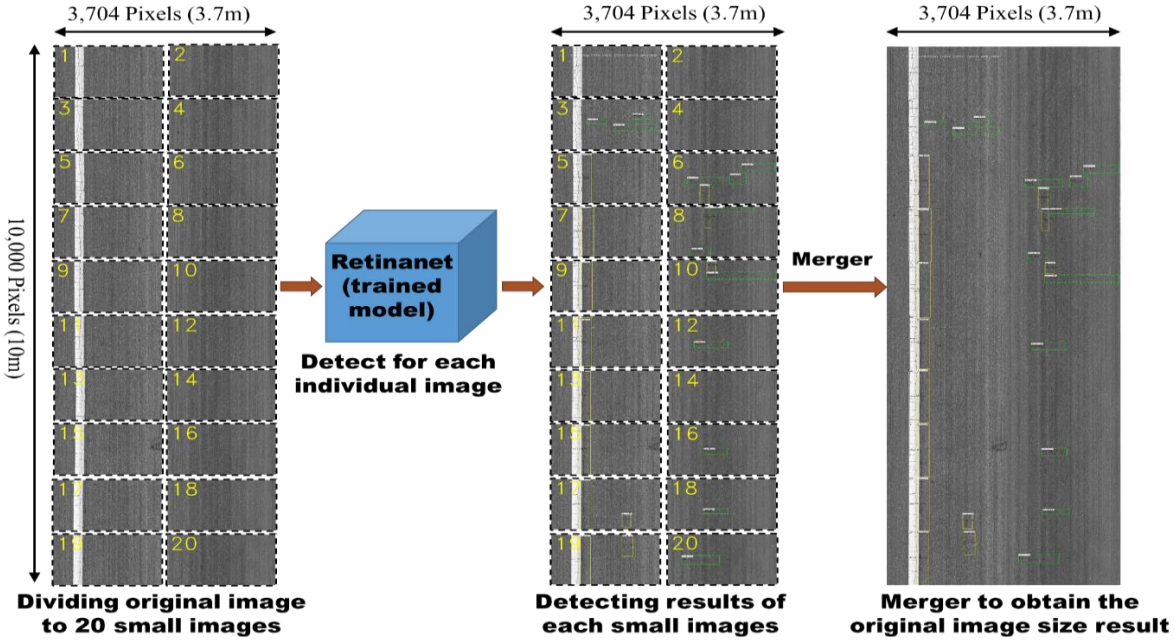


Figure 4. *Example of divide image, testing and merge an image.*

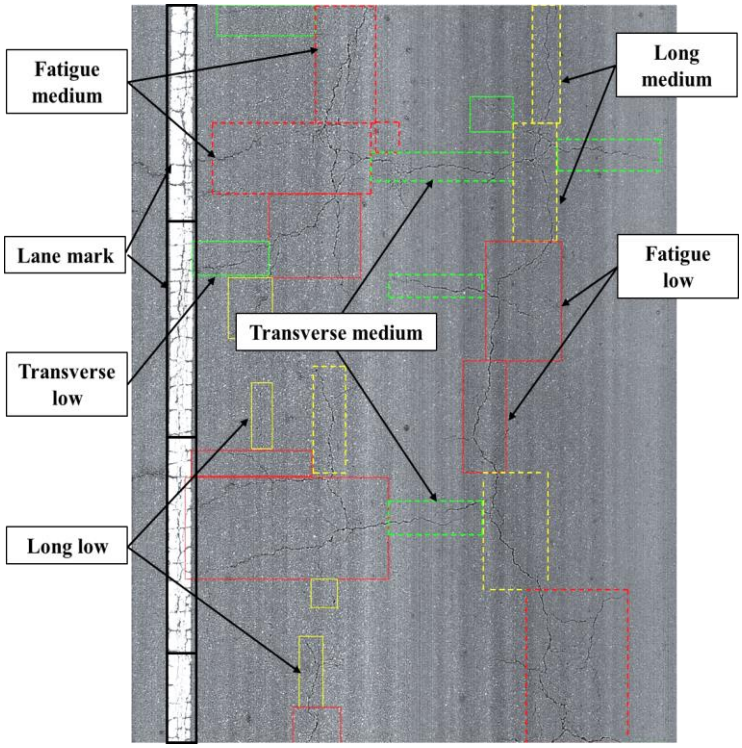


Figure 5. *Example of detection results from the proposed network model.*

5. Validation of the trained network model

Two road sections (with 2,400 m) were collected from Seoul city urban road to validate the accuracy of the proposed network model. A total of 240 images are selected. Firstly, these images were analyzed using the developed, trained model to detect lane markers and crack distress automatically. Secondly, those images were checked by manual to compare the accuracy of the trained model. Table 1 summarises the detection accuracy for cracks considering both crack type and its severity level. As observed from the table, the average accuracy of the proposed model is 85%. It could be primarily due to the complexity of object classification because of the increased number of objects to be detected when considering the crack type and the severity level. It was seen from the manual checking results that when the linear crack width is close to the value between low and medium severity levels, most of the misclassifications happened. Moreover, since the number of training databases of high severity cracks was comparatively lower than the medium and low severity levels, that is why they create more detection errors in terms of the high severity level, as exhibited in the table. It is also noted that the residual errors were significantly reduced compared to the absolute errors.

Because the residual errors of the model seem to be acceptable. Therefore, the detection accuracy of the proposed network model in this study can be useful to the PMS network level as the evaluated pavement section length to be long enough.

6. Conclusion

In this study, a one-stage object detector called RetinaNet was used for pavement distress detection from scanned pavement images. A network based on RetinaNet is trained to detect different types of cracking such as fatigue, longitudinal and transverse wherein its severity was also classified as low, medium, and high. The comparison between the results from the trained network and manual checking showed a good correlation between the two methods.

From several pieces of training done in this study, it was found that the two main factors influencing the accuracy of detection are the number of datasets used for training and the quality of the dataset. In general, higher accuracy of detection is obtained as the number of datasets increases. However, the collection of datasets for training demands a lot of effort and is time-consuming therefore choosing the appropriate number of datasets for training is essential.

In this study, the trained network can detect not only the type of crack but also the severity level. Crack detection with severity was done by building and training a dataset from the captured images adapting the FHWA guidelines for classifying cracks and their severity. The crack's severity classification is done by counting the number of pixels on the crack area of the image. Initially, the original image with a size of 3,704 x 10,000 is divided into smaller images of 1,852 x 1,000 for crack and severity classification and then is merged after to its original size.

The classified objects are then bounded by boxes that are labeled depending on the type of crack and its severity. The trained network then was tested by surveying a 2.4 km of road and applying the detection tool. It was found that the trained network based on RetinaNet performs well when detecting cracks and determining their severity with an accuracy of 84.9%. It can be concluded that the trained network performs well in detecting crack and classifying its severity❑

Table 1. *Prediction performance of the trained network model regarding both crack type and severity level.*

Object type	Longitudinal (m)			Transverse (m)			Fatigue (m^2)		
	low	**medium**	**high**	**low**	**medium**	**high**	**low**	**medium**	**high**
Quantity	535.	293.15	35.35	208.	176.05	30.65	270.1	96.25	22.7
Error (%) (+)	7.6	3.15	5.3	4.4	5.85	2.85	5.3	6.55	4.85
Error (%) (-)	-8.3	-9.4	-13.4	-8.2	-8.3	-	-7	-7.65	-
Residual Error (%)	3.35	6.25	8.05	3.8	3.65	10	2.4	3.4	10.1
Absolute Error (%)	15.9	12.6	18.65	12.6	14.1	15.7	12.3	14.2	19.85
Accuracy (%)			84.3			85.9			84.5
Average Accuracy (%)									84.9

References

[1] H. Oliveira and P. L. Correia (2009), *Automatic road crack segmentation using entropy and image dynamic thresholding*, in 2009 17th European Signal Processing Conference, 2009, pp. 622-626: IEEE;

[2] H. Zhao, G. Qin, and X. Wang (2010), *Improvement of canny algorithm based on pavement edge detection*, in 2010 3rd International Congress on Image and Signal Processing, 2010, vol. 2, pp. 964-967: IEEE;

[3] N. Tanaka and K. Uematsu (1998), *A Crack Detection Method in Road Surface Images Using Morphology*, MVA, vol. 98, pp. 17-19, 1998;

[4] C. Koch and I. Brilakis (2011), *Pothole detection in asphalt pavement images*, Advanced Engineering Informatics, vol. 25, no. 3, pp. 507-515, 2011;

[5] L. Ying and E. Salari (2010), *Beamlet transform-based technique for pavement crack detection and classification*, Computer-Aided Civil and Infrastructure Engineering, vol. 25, no. 8, pp. 572-580, 2010;

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton (2012), *Imagenet classification with deep convolutional neural networks*, in Advances in neural information processing systems, 2012, pp. 1097-1105;

[7] Z. Fan, Y. Wu, J. Lu, and W. Li (2018), *Automatic pavement crack detection based on structured prediction with the convolutional neural network*, arXiv preprint arXiv:1802.02208, 2018;

[8] S. Li and X. Zhao (2018), *Convolutional neural networks-based crack detection for real concrete surface*, in Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, 2018, vol. 10598, p. 105983V: International Society for Optics and Photonics;

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik (2014), *Rich feature hierarchies for accurate object detection and semantic segmentation*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587;

[10] R. Girshick (2015), *Fast r-cnn*, in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448;

[11] S. Ren, K. He, R. Girshick, and J. Sun (2015), *Faster r-cnn: Towards real-time object detection with region proposal networks*, Conference on Neural Information Processing Systems, 2015;

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár (2017), *Focal loss for dense object detection*, in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988;

[13] J. S. Miller and W. Y. Bellinger (2014), *Distress identification manual for the long-term pavement performance program*, United States, Federal Highway Administration, Office of Infrastructure 2014;

[14] K. He, X. Zhang, S. Ren, and J. Sun (2016), *Deep residual learning for image recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

Received: April 6, 2021
Reviewed: April 9, 2021
Revised: May 1, 2021
Accepted: May 7, 2021