Scientific paper

Crack Detection from a Concrete Surface Image Based on Semantic Segmentation Using Deep Learning

Tatsuro Yamane[1] and Pang-jo Chun[2*]

Received 11 May 2020, accepted 18 August 2020 doi:10.3151/jact.18.493

Abstract

Due to their wide applicability in inspection of concrete structures, there is considerable interest in the development of automated crack detection method by image processing. However, the accuracy of existing methods tends to be influenced by the existence of traces of tie-rod holes and formworks. In order to reduce these influences, this paper proposes a crack detection method based on semantic segmentation by deep learning. The accuracy of developed method is investigated by the photos of concrete structures with lots of adverse conditions including shadow and dirt, and it is found that not only the crack region could be detected but also the trace of tie-rod holes and formworks could be removed from the detection result with high accuracy. This paper is the English translation from the authors' previous work [Yamane, T. and Chun, P., (2019). "Crack detection from an image of concrete surface based on semantic segmentation by deep learning." *Journal of Structural Engineering*, 65A, 130-138. (in Japanese)].

1. Introduction

The issues seen in recent years have been the progressing deterioration of concrete structures such as bridges and tunnels. To use these structures safely, appropriate inspections are required, and a crack is one of the important items to be checked. A common inspection method to check the crack carried out currently is that an inspector visually inspects the crack and sketches it. However, since the number of concrete structures and the area that need to be inspected are large, manual inspections cannot keep pace with this, and there are calls for automated inspections (MLIT 2018; Chun *et al*. 2020a).

Under this circumstance, researches have widely been conducted into the possibility of photographing the surface of the concrete structures and detecting the crack via image analysis. The research work of Luxmoore (1973) is considered to be the one that marked the beginning of this particular research field, in which he discussed the crack detection method using holography along with its range of application. In these years, significant improvements in hardware such as photographic instruments and computers, as well as in software such as image processing algorithms have been observed, and inspections using various methods have been conducted, especially since 2000.

Abdel-Qader *et al*. (2003) demonstrated the superiority in detection accuracy of fast Haar wavelet transform by comparing analysis results on photo images of 50 bridges analyzed by the following four edge detection techniques: fast Haar wavelet transform, fast Fourier transform, Sobel operator, and Canny algorithm. Similar to this study, there were many studies on the crack detection using frequency analysis in the early 2000s. For example, Hutchinson and Chen (2006) used a Canny filter and the wavelet transform for the crack detection and estimated parameters. Ito *et al*. (2002) developed a system that was able to extract and analyze the crack on the concrete surface by combining several image processing techniques, including the wavelet transform, shading correction, and binarization. Although the number has been decreasing recently, there are still ongoing studies on the crack detection using the frequency analysis such as the work on Dual Tree Complex Wavelet Transform (Dixit and Wagatsuma 2018).

There are research projects focused on geometrical features as well. For example, Iyer and Sinha (2006) used a curvature feature evaluation with a mathematical morphology processing to detect the crack in contrast enhanced sewer pipeline images. The resulting images were used to measure the crack properties and determine pipe criticality levels. Sinha and Fieguth (2006) developed an algorithm based on the morphological operations to segment pipe cracks, holes, joints, laterals, and collapsed surfaces accurately, which is a crucial step in the classification of defects in underground pipes. This study is novel in that it takes a two-step approach to inspect the crack broadly and then locally. Hashimoto and his research team proposed a crack detection technique focused on crack's linearity and continuity using a gray-scale Hough transform and a percolation processing (Yamaguchi *et al*. 2008; Yamaguchi and Hashimoto 2008, 2009, 2010). As another research project focused on the geometrical features, Fujita *et al*. (2006) propose an extremely robust method for detecting the crack by effectively utilizing a probabilistic relaxation, a line ex-

[1]Doctoral Student, Department of International Studies, The University of Tokyo, Japan.
[2]Project Associate Professor, Department of Civil Engineering, The University of Tokyo, Japan.
*Corresponding author,
E-mail: chun@i-con.t.u-tokyo.ac.jp

tension processing, and a multi-scale line enhancement processing using Hessian matrices (Fujita *et al*. 2006; Fujita and Hamamoto 2009, 2011). There are much more projects such as the one by Amarasiri *et al*. (2009), in which they proposed a crack detection technique using the bidirectional reflection distribution function derived from the model based on the Gaussian reflectance model introduced by Ward (1992); a research work by Lee *et al*. (2013) that combined the binarization with a noise processing and a thinning processing; and a method proposed by Sohn *et al*. (2005), in which they used a three dimension analysis based on the photogrammetric methods combined with the Hough transform and a filter processing. The author's group also proposed a method to detect the crack by focusing on the geometrical features (Chun and Igo 2015; Chun *et al*. 2020b). There are various other methodologies. They are properly summarized in the several review papers (Koch *et al*. 2015; Zakeri *et al*. 2017; Mohan *et al*. 2018).

However, a common issue found in these studies is that the judgment results tend to be influenced by the presence of items other than the crack, such as traces of formwork or traces of tie-rod holes. The cause of this is that the methods used in these existing studies made the judgment using only the pixels of interest themselves or neighboring pixels, despite the fact that classification and determination are not possible for typical images, unless a broad judgment using an area of at least 100 × 100 pixels is made. It is considered that artificial intelligence techniques, such as deep learning, are useful as a method to resolve these issues.

The artificial intelligence technique including deep learning has gained popularity in recent years owing to its high performance to deal with the large amount of data and was used for civil engineering applications, for example, by Rafiei *et al*. (2017), Rafiei and Adeli (2018a, 2018b), Chun *et al*. (2019, 2020c, 2020d), Moon *et al*. (2019, 2020) and Okazaki *et al*. (2020). It is also used for the detection of the crack. Yokoyama and Matsumoto (2017) used a convolutional neural network (CNN), one of the deep learning techniques, and proposed a system that classified the concrete surface into five categories, i.e., cracked part, chalk letter part, joint part, surface part, and other parts, and displayed the cracked part as a rectangular area. In addition to displaying as the rectangular area, there is also a demand for extracting the crack pixel by pixel to evaluate the damage properly. As an example of such study, Dung (2019) first classified the existence of the crack within a small area using VGG, and then detected the crack pixel by pixel using fully convolutional network (FCN). Another study used an approach, in which screening was performed first using deep learning, and then the crack was detected using a random forest (Chun *et al*. 2017).

Additionally, as the extensions of the CNN, methods such as pix2pix (Kobayashi 2018), deep convolutional encoder-decoder network (Bang *et al*. 2019), method using fully convolutional network (Yang *et al.* 2018), CrackNet (Zhang *et al.* 2018a), CrackNet II and CrackNet-V proposed by improving original CrackNet (Zhang *et al.* 2018b; Fei *et al.* 2019), and DeepCrack (Liu *et al*. 2019) were recently proposed. There are also the studies by Minami *et al*. (2019a, 2019b), in which they improved the performance of photographic devices and performed the CNN. Other studies on crack detection using CNNs have also been conducted by, for example, Cha *et al.* (2017, 2018), Silva and Lucena (2018), Xue and Li (2018), Deng *et al.* (2019), Jiang and Zhang (2019), Li and Zhao (2019) and Zhang *et al.* (2019). Although several crack detection techniques using deep learning have been proposed so far as above, practical development is still in the early stage, since the technology itself is still immature. In this study, therefore, through a semantic segmentation using deep learning, the crack was detected pixel by pixel, while images were judged from a large perspective (Yamane and Chun 2019). Meanwhile, in existing studies, there is little mention of the trace of tie-rod hole or the trace of formwork. However, since these have the features of being dark, long, and thin, similar the crack, these are the main causes of false detection. This study detected these pixel by pixel to increase the accuracy of the crack detection.

2. Automatic crack detection method

The flow diagram for the automatic crack detection method proposed in this study is shown in **Fig. 1**. The details of each step are explained below.

2.1 Semantic segmentation

Semantic segmentation is a technology for extracting the region of an object in an image pixel by pixel, and various area extraction methods have been proposed up to this point, such as the Graph Cut and the Grow Cut

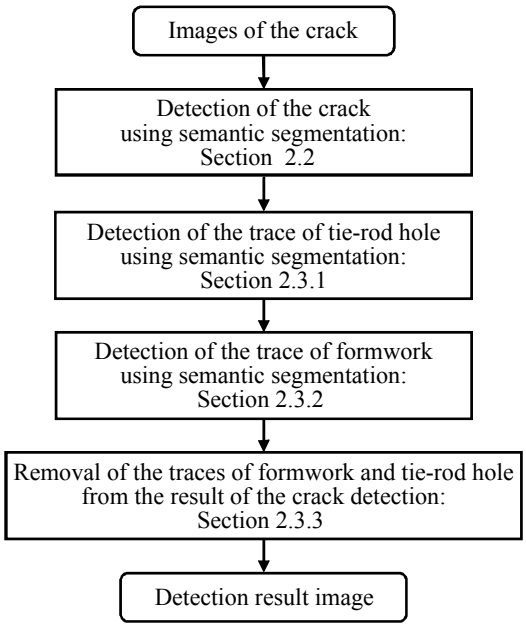


Fig. 1 Flow diagram of method proposed in this study.

(Blake *et al*. 2004; Vezhnevets and Konouchine 2005). Then with the development of deep learning technology in recent years, the accuracy of the extraction technology has improved rapidly.

In terms of the semantic segregation using deep learning, Long *et al*. (2015) proposed a FCN, which does not use a fully connected layer to output the feature map of the object area in the input image, but consists of convolution layer and pooling layer. Generally, the CNN outputs the classification results of the input image by eventually establishing the fully connected layer. On the other hand, the FCN outputs the feature map in relation to the input image by not using the fully connected layer. Then, this feature map becomes a probability map for the class to be labeled for each pixel. Note that, as the size of the feature map of an input image is reduced every time it passes through the convolution layer and the pooling layer, the final size of the feature map is extremely smaller than that of the input image. For this reason, in the FCN, the feature map eventually generated is enlarged (deconvolution process) to overlay it on the original image. However, as the information in the feature map that was enlarged in this way is rough compared to the original image, the FCN merges the feature map that is output in the intermediate layer at the final layer. Generally, as the feature map in the intermediate layer of the CNN has detailed features in a layer close to the input layer, it is possible to output a detailed feature map by using this information. However, with the FCN, the issue has been raised that high memory usage is required for merging the output feature map with the final layer.

To speed up the processing and save the memory usage, Badrinarayanan *et al*. (2017) proposed SegNet that has an encoder-decoder structure that does not need to memorize the feature map output in the intermediate layer. With the SegNet, the feature map is enlarged at the decoder side after the convolution and the pooling processing is carried out at the encoder side. At this time, as the feature map is enlarged based on the positional information selected using the pooling processing at the encoder side, it is possible to recover a detailed map while also saving memory.

Furthermore, as a method of carrying out the semantic segmentation at high accuracy, Mask R-CNN has recently been proposed based on Faster R-CNN that can detect rectangular areas of the object at high speeds (Ren *et al*. 2015; He *et al*. 2017). With the Faster R-CNN, the features of the input image are extracted first using the CNN, and then candidate areas on the object are selected. After judging whether the selected object candidate areas are objects or not, their classes are classified. For this process, the Faster R-CNN proposes a region proposal network (RPN) for making a selection of the object candidate area while learning, making a high-speed, and high-accuracy object selection possible. With the Mask R-CNN, similar to the Faster R-CNN, the features of the input image are extracted using the CNN, which is a backbone network. Then, the object candidate area is selected using the RPN, and the semantic segmentation is performed on the identified rectangular area, making a high accuracy area extraction possible (**Fig. 2**). As it was thought that, when actually performing the semantic segregation using deep learning, it was generally easier to use a method of analysis based on a method already established, a method of detecting the crack based on this Mask R-CNN was proposed in this study.

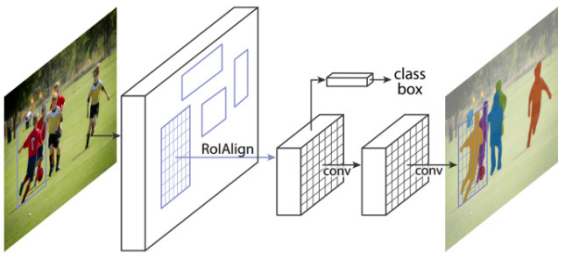


Fig. 2 Mask R-CNN structure (He *et al*. 2017).

The semantic segmentation using deep learning is frequently used in fields such as automated driving, factory inspections, etc., but only in its early stages in the civil engineering field. As stated in the introduction, whereas it has started to be used for the detection of the crack in concrete, as there is a wide divergence of situations and environments in which the crack exists in concrete, proposing an approach different from the existing methods of the crack detection is useful in raising the accuracy of the crack detection.

2.2 Crack detection using the semantic segmentation

In the semantic segmentation carried out in this study, a judgment on whether the pixel in question within the image is a crack or not is made using supervised machine learning. With the supervised machine learning, data that is prepared in advance by humans is applied as learning data, and by making it learn the relationship between the input and the output, the output is predicted against unknown input data. In this study, an algorithm for judgment is constructed by learning the crack image data of concrete surface prepared in advance for the input and data showing the position of the crack in each image data. By using this algorithm, it is possible to judge the position of the crack pixel by pixel even for unknown the crack that has not been used in the learning.

When recognizing the object area within the image using deep learning, as done in this study, recognition accuracy is greatly affected by the number of repetitions of the convolution layer and pooling layer, and the setting of parameters within each layer. In case of the Mask R-CNN used in this study, the CNN settings for performing the feature extraction of the input image are directly linked to recognition accuracy. The optimal solution for these settings differs depending on the problem, but in this study, ResNet-101 (He *et al*. 2016), for which the high recognition accuracy has been found in recent years, was used for the feature extraction and the structure after the feature extraction was set the same

as that shown in He *et al*. (2017).

Additionally, to make the procedure of setting these comparatively simple, it is possible to use an open source library such as Keras as a library for deep learning, and the analysis in this study was performed using the Mask R-CNN that was constructed using the Keras. He *et al*. (2017) indicated 0.02 as the value for learning rate, which is the hyperparameter that determines the update quantity of weights at the time of learning, but as weights was divergent if this was used as is, learning was performed in this study with the learning rate set to 0.001. Further, the optimization algorithm at this time used the stochastic gradient descent method. These settings were the same for learning the trace of tie-rod hole and the trace of formwork shown later.

In this study, 45 digital images (5184 × 3456 pixels) captured under various situations, such as the floor slabs and beams of bridges and the inner wall of buildings, were prepared for the use in image analysis, and performed validation after separating the prepared images into learning use and evaluation use.

When trying to detect extremely detailed objects such as the crack, prospects for obtaining high-accuracy recognition by learning the image as is were poor. This is due to the fact that the image is shrunk during the process of learning and information on the detailed areas of the image may be lost. Therefore, a portion of the image among the learning data prepared was cut out with the size large enough to recognize the crack even with human eyes, and by using these cutout images of small areas as teacher data, detection was performed with high accuracy. As shown in **Fig. 3**, learning was performed by cutting out an area of 384 × 256 pixels against an original image of 5184 × 3456 pixels. Additionally, in the same way as when cutting out small areas from the image for learning to use in learning, small areas were cut out from the image for evaluation and analyzed.

Then, after the analysis, an image of the original size was reconstructed by adding these together. Further, as shown in red in **Fig. 3**, a ground truth for the crack was set for each of the cutout images. The ground truth data for the crack in this study was set by judging crack areas from the images using human observation.

Additionally, when extracting the crack images, the cutout area is overlapped by a 1/2 width both vertically and horizontally, as shown in **Fig. 4**. Here in case the image cannot be overlapped and cut out, detection leaks may occur at the edges of each image after analysis, as shown in **Fig. 4(a)**. For this reason, by reconstructing an image with an original size that allows the area judged to be the crack in the image to be overlapped after analysis reduces the impact of detection leaks on the image edges, as shown in **Fig. 4(b)**.

In this study, multiple small areas were extracted in order from each image in this way, and a total of 676 images were generated from one image. Additionally, by dividing images into several small images and learning in this way, as image data for which there are few original images can be learned as data for many images, the aim is to be able to improve learning accuracy.

2.3 Detecting the trace of tie-rod holes and the trace of formwork

In terms of the crack detection, the existence of abnormality other than the crack such as the trace of tie-rod hole and the trace of formwork has been pointed out as a

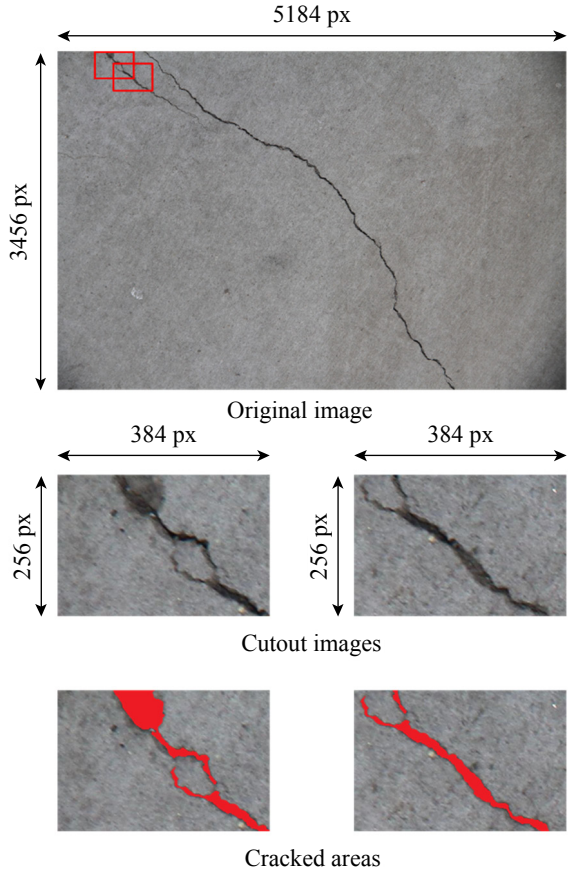
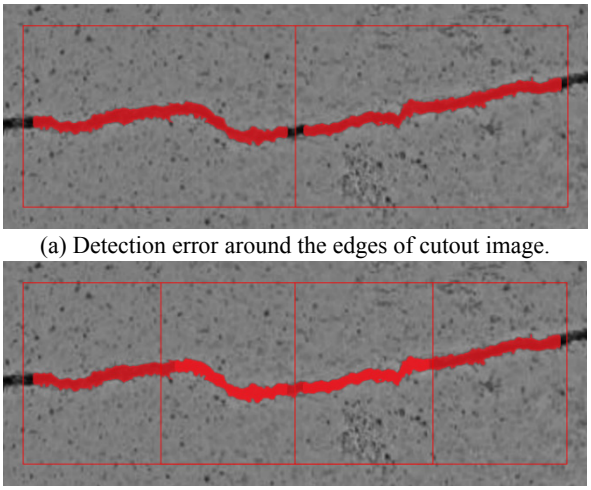


Fig. 3 Cutout position and enlarged image.



(a) Detection error around the edges of cutout image.

(b) Reducing the detection error by overlapping the image.

Fig. 4 Detection errors around the edge of cutout image and overlapping method to reduce them.

cause the false detection in many existing crack detection methods. Therefore, this study did not stop at carrying out the judgments made by learning the crack only, and by learning the trace of tie-rod holes and the trace of formwork, respectively to exclude these from the crack detection results, performed the crack detection with higher levels of accuracy. For these detections, the Mask R-CNN that used the ResNet-101 was used as in the case with the detecting the crack, and the setting of the hyperparameters at that time was the same as for when detecting the crack.

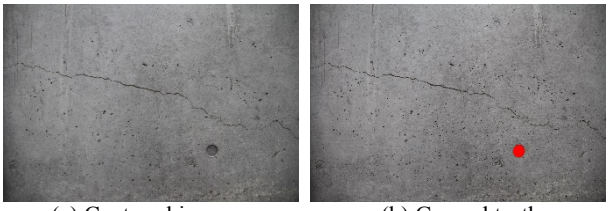
2.3.1 Detection of the trace of tie-rod holes

When detecting the trace of tie-rod holes, with the aim of increasing the accuracy of detecting the trace of tie-rod hole, 55 images (5056 × 3792 pixels) containing the trace of tie-rod holes were prepared, in addition to the 45 images used in the crack detection. In this study, we used a total of 100 images and used these for the detection of trace of tie-rod holes.

As the trace of tie-rod hole covers a relatively large area unlike the crack, the original image can be used as is without dividing the image into smaller regions as was performed when learning the crack. The trace of tie-rod holes used for learning and an example of a ground truth is shown in **Fig. 5**. The ground truth for the trace of tie-rod holes in this study are set, in the same way as with the crack, based on the judgment of the trace of tie-rod holes from images based on human observation.

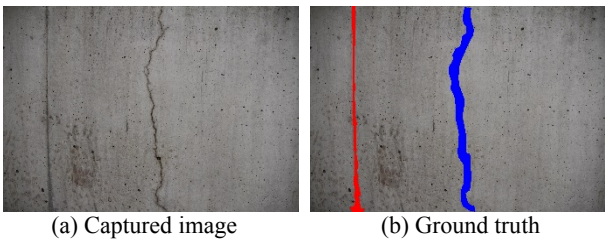
2.3.2 Detection of the trace of formwork

When detecting the trace of formwork, in the same way as when detecting the trace of tie-rod hole, since the aim is to increase the detection accuracy of trace of formwork, 60 images including trace of formwork (5056 × 3792 pixels) were also prepared in addition to the 45 images used for crack detection. When learning only the trace of formwork and performing the analysis, since the assumption is that the crack with a form close to a straight line, similar to the trace of formwork, may be falsely recognized as the trace of formwork, 45 images with linear cracks were also prepared in addition to these 105 images. Additionally, by making it possible to detect regions with both the trace of formwork and the crack, it is possible to ensure that the trace of formwork is not confused with the crack. In this study, a total of 150 of these images were used to perform the detection of the trace of formwork.



(a) Captured image (b) Ground truth

Fig. 5 Example of the images including the trace of tie-rod holes and their ground truth.



(a) Captured image (b) Ground truth

Fig. 6 Example of image and ground truth using trace of formwork.

Note that, the purpose of detecting the trace of formwork is to exclude the false detection area of the trace of formwork from the image of crack detection results. From this, for the detection of the trace of formwork, there is little necessity to make precise detection of the crack pixel by pixel, as shown in Section 2.2, and it is sufficient to be able to grasp the general location of the trace of formwork. Therefore, even when learning the trace of formwork, it is possible to learn using the original image in the same way as when learning the trace of tie-rod holes.

Meanwhile, the purpose when detecting the crack areas here is to avoid falsely recognizing the crack in linear form as the trace of formwork. For this reason, since it is not necessary to be able to perform detailed recognition in the order of a few pixels either, when setting the crack areas, the general area in which the crack exists is set as the ground truth. Additionally, by setting this kind of general area as the ground truth, the ground truth data set for each image can be simplified, it was confirmed that not only does this save time when creating used supervised data, but also the time required for one epoch of learning can be reduced by half.

Here, the trace of formwork images used for learning and an example of the ground truth are shown in **Fig. 6**. In **Fig. 6**, the ground truth for the trace of formwork is shown in red, and the crack ground truth is shown in blue. In these ground truths as well, they are set to be able to judge visually from the image, in the same way as for the trace of tie-rod holes.

2.3.3 Exclusion of falsely detected areas

In the areas in which the crack was detected in the 45 captured images analyzed in Section 2.2, there are areas in which the trace of tie-rod hole and the trace of formwork are falsely recognized as the crack. This is because, by dividing the original image into small areas, the trace of tie-rod hole and the trace of formwork in the images after division were seen in a form close to that of the crack.

Therefore, using the results of the detected trace of tie-rod hole and the trace of formwork, falsely detected trace of tie-rod hole and trace of formwork are excluded from the crack detection results; the crack detection accuracy is improved. Here, the areas judged as crack that exist in areas detected as being the trace of tie-rod hole and the trace of formwork areas were considered to

be falsely detected, and were all excluded from the crack detection results. When detecting the trace of formwork, in the same way as when detecting the trace of tie-rod hole, since the aim is to increase the detection accuracy of trace of formwork, 60 images including trace of formwork (5056 × 3792 pixels) were also prepared in addition to the 45 images used for crack detection. When learning only the trace of formwork and performing the analysis, since the assumption is that the crack with a form close to a straight line, similar to the trace of

Fig. 7 Example of the crack detection results.

formwork, may be falsely recognized as the trace of formwork, 45 images with linear cracks were also prepared in addition to these 105 images. Additionally, by making it possible to detect regions with both the trace of formwork and the crack, it is possible to ensure that the trace of formwork is not confused with the crack. In this study, a total of 150 of these images were used to perform the detection of the trace of formwork.

3. Results and discussion

3.1 Crack detection results

The effectiveness of the proposed method in this study was confirmed using K-fold cross-validation. With the K-fold cross-validation, all of the data sets are divided into K items, and learning is performed on K-1 items from among these, and the remaining data is used for evaluation purposes, repeating the process K times. Here validation is performed as K = 5. That is to say, when learning crack images, the 45 images prepared are divided into 4:1, and an equivalent of 36 were used for learning, and an equivalent of 9 were used for evaluation. At this time, as each image is divided into small regions, the 24336 subdivisions of the 36 images were used as images for learning, and the 6084 subdivisions of the 9 images were used as images for evaluation.

In **Fig. 7**, using the example of the six of the picked-up captured images, the captured image and ground truth image, as well as the crack detection results using the method shown in this study were shown. From the image shown in **Fig. 7**, it was confirmed that whereas all of the images contain many factors that may cause the false detection, such as shadow and dirt, the method of detecting the crack shown in this study is generally able to classify these well.

The first and second images from the top of **Fig. 7** are images that include many regions for which the brightness is low, due to shadow. However, using the method in this study, the crack could be detected with an extremely high level of accuracy. Additionally, in the second captured image, it was confirmed that there were coloring stains on the right side of the image, but these were not seen as affecting the detection results.

The third image from the top in **Fig. 7** is an image that includes an exceptionally large number of detailed the crack. The width of these cracks are extremely fine, between 1 and 3 pixels, but it was seen that these kinds of the cracks too are extremely well detected. Additionally, in the central area of this diagram, there are thin water stains in the vertical direction, but these were also not falsely detected.

The fourth image from the top in **Fig. 7** is an image containing a large amount of dirt. In this image as well, whereas several partially blackened areas were detected as the crack, many other areas where there was a lot of dirt were not falsely detected, and it was confirmed that the areas where there are the crack were detected well.

The 5th and 6th captured images from the top in **Fig. 7** are images that include the trace of tie-rod hole and the trace of formwork. As in the diagrams shown up to this point, the crack area can be detected well, but the trace of tie-rod hole and the trace of formwork are misdetection as the crack. Therefore, the elimination of these false detections caused by the detection of the trace of tie-rod hole and the trace of formwork is shown in the following.

3.2 Trace of tie-rod hole and trace of formwork detection results

When learning the trace of tie-rod hole and the trace of formwork as well, as when learning the crack, learning was perfumed using 80% of the prepared images, and overall analysis was performed by repeating analysis of the remaining 20% of images five times.

In **Fig. 8** three images among the captured images of the trace of tie-rod hole were used as examples, and those captured images and the detection results of using the method described in this study are shown. It can be seen that the traces of tie-rod holes were detected with high level of accuracy in all images shown in **Fig. 8**.

In the second captured image from the top, there are multiple instances of the trace of tie-rod holes, but all of these traces of tie-rod holes were detected with high level of accuracy.

Furthermore, in the third captured image from the top, only a part of the trace of tie-rod hole is shown on the bottom right on the image, but this could also be detected here with good accuracy.

Next, three of the captured images are shown as the examples of the trace of formwork detection results in **Fig. 9**. The part shown with red is the area judged to be the trace of formwork, From **Fig. 9**, it can be seen that the area in which the trace of formwork exists can be detected accurately nearly all of the time. Furthermore,

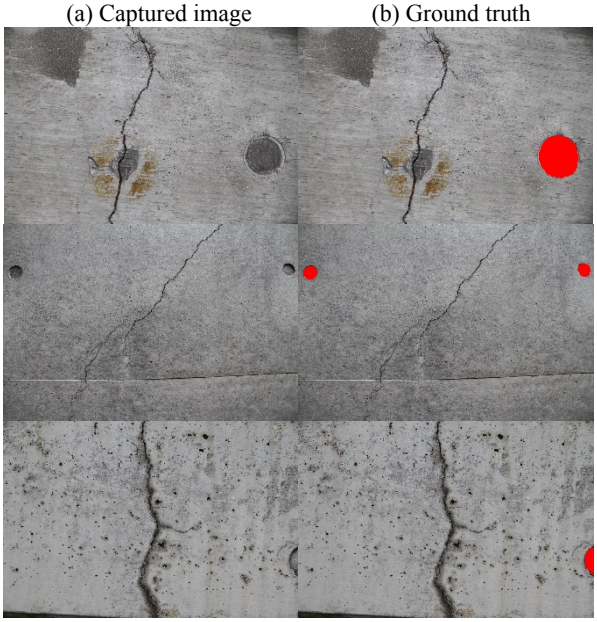


Fig. 8 Example of the trace of tie-rod hole detection results.

there is a tendency in the detection results to recognize the trace of formwork to be rather large, but if it is considered that the aim is to eliminate misdetection areas, it can be said that it is providing effective detection.

In the top image in **Fig. 9**, there is a crack in the vertical direction near the center. It was confirmed that this was not recognized as the trace of formwork, and only the trace of formwork on the left was correctly detected. This is because not only the trace of formwork, but also the crack areas are detected at the same time, as considered in Section 2.3.2. From this, its effectiveness can be confirmed. Further, in the third image from the top, the trace of formwork exists in the extremely dark area of the image, but this is also detected without any problem.

In the second and third image, there are various areas where the crack are not detected, but this is caused by the fact that most of the images of the crack used when learning the trace of formwork were linear the crack for preventing false detection at the time of form detection. It is considered that the detection accuracy could be improved by increasing the types of crack image learning data, but as the objective is clearly to detect the trace of formwork, it is considered that the current detection accuracy is not problematic. Additionally, even in regard to the trace of formwork that are partially undetected, it is considered, in the same way that the detection accuracy could be further improved by increasing the learning data.

3.3 Results excluding falsely detected areas in the crack results

Three examples, from the 45 images used when detecting the crack, of falsely detected the trace of tie-rod hole and the trace of formwork excluded from the crack detection results are shown in **Fig. 10**. From **Fig. 10**, it can be seen that virtually all of the areas where the trace of tie-rod hole and the trace of formwork were falsely detected could be excluded, and that this greatly contributes to the detection accuracy for the crack.

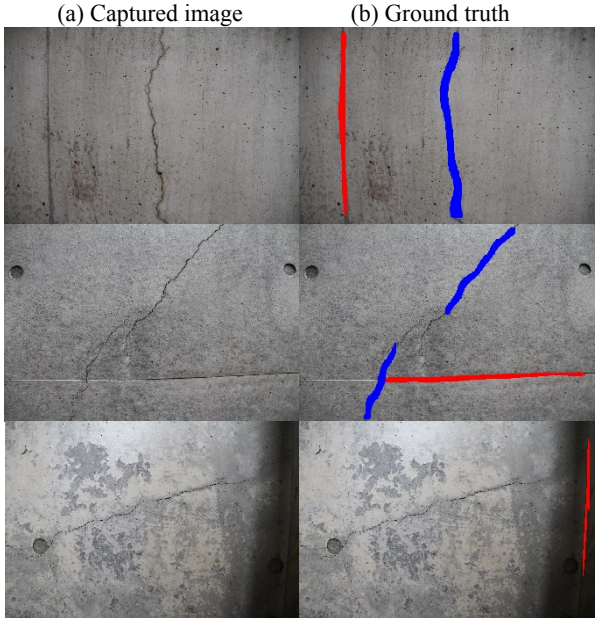


Fig. 9 Example of trace of formwork detection results.

Table 1 Class classification confusion matrix.

Detected results \ Ground truth	Cracked	Not cracked
Cracked	TP	FP
Not cracked	FN	TN

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TF} + \text{TN} + \text{FN}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{F-measure} = \frac{2 \times \text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \quad (5)$$

Furthermore, the crack detection accuracy before and after the exclusion of the falsely detected areas was evaluated based on five indicators, i.e., accuracy, sensitivity, specificity, precision, and F-measure. These indicators are defined according to equations (1) through (5) shown below. In the formulas, TP, TN, FP, and FN are abbreviations for true positive, true negative, false positive, and false negative, respectively, and are the number of pixels is when the pixels were classified based on **Table 1**.

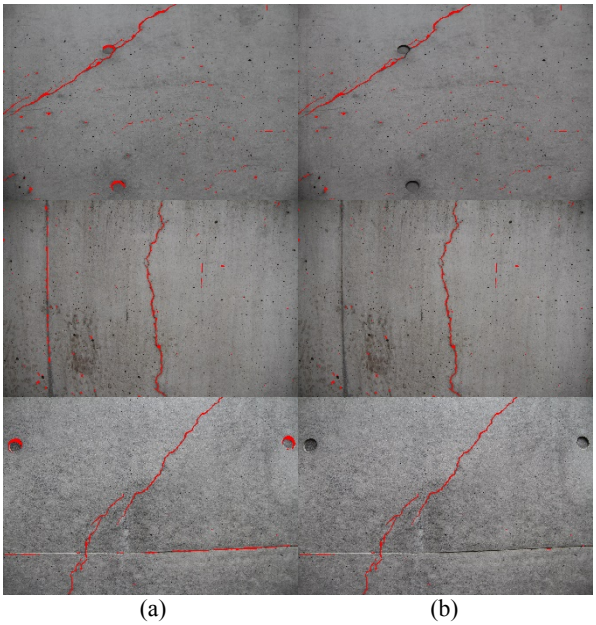


Fig. 10 (a) Images before excluding falsely detected areas; (b) Images after excluding falsely detected areas.

Table 2 Quantitative evaluation of the crack detection results.

	Before excluding false detection	After excluding false detection
Accuracy	0.9915	0.9921
Sensitivity	0.7881	0.7847
Specificity	0.9927	0.9933
Precision	0.3924	0.4044
F-measure	0.4862	0.4994

Table 3 Quantitative evaluation of the detection results of the trace of tie-rod holes and the trace of formwork.

	Trace of tie-rod hole	Trace of formwork
Accuracy	0.9990	0.9909
Sensitivity	0.9404	0.4549
Specificity	0.9998	0.9974
Precision	0.9638	0.7837
F-measure	0.9563	0.6614

That is to say, the accuracy is the ratio of pixels for which, in relation to the total number of pixels for the image as a whole, have the same labels in the analysis results and the ground truth data. Additionally, sensitivity is the ratio, in relation to the pixels in the images of the crack among the ground truth images, recognized as the crack in the analysis results as well. The specificity is ratio, in relation to the pixels in the images of not of the crack among the ground truth images, recognized as not of the crack in the analysis results as well. The precision is ratio, in relation to the pixels in the images of the crack among the analysis results, recognized as the crack in the ground truth images as well. Further, the F-measure evaluates both the precision and sensitivity, which generally are in a trade-off relationship.

Table 2 takes the 45 images used for crack detection and shows the mean value for each indicator before and after excluding the falsely detected areas. From **Table 2**, it can be seen that these values are improved for virtually all indicators. With regard to the fact that the sensitivity value is slightly decreased, this is mainly due to the fact that crack images adjacent to the trace of tie-rod holes and the crack occurring inside the trace of tie-rod holes are excluded. However, from the fact that the F-measure that is the harmonic mean of the sensitivity and precision is increased, it may be considered that the detection accuracy as a whole is increased. Furthermore, the level of improvement in the detection accuracy is affected by the number of the trace of tie-rod holes and the trace of formwork included in the image. In this study, of the 45 images, there were 15 images including the trace of tie-rod holes, and 6 images including the trace of formwork prepared, but in images where there were many trace of tie-rod holes and trace of formwork, it is considered that there is a greater level of detection accuracy improvement due to the greater false detection area exclusion.

Table 3 shows the various values when evaluating the same indicators for detection accuracy of the trace of tie-rod hole and the trace of formwork. Here the 100 images for which the trace of tie-rod hole analysis was performed and 105 images for which the trace of formwork analysis was performed included images for which the calculation equations (1) through (5) could not be applied. For example, in the case of images that did not contain the trace of tie-rod hole or the trace of formwork that were targeted, the sensitivity, precision, and F-measure could not be calculated. For this reason, **Table 3** shows the mean values for only the evaluation values that could be calculated. From **Table 3**, regarding the trace of tie-rod hole, it can be seen that the detection occurs with an exceptionally high level of accuracy. As for the trace of formwork, the level of detection accuracy was not as high as that of the trace of tie-rod holes. However, the F-measure indicated that the detection accuracy was greater than that of the crack, and it can be said that the method in this study is sufficiently effective in the removal of the false detection.

4. Conclusion

In this study, we proposed a method of detecting the crack in concrete surfaces using semantic segmentation through deep learning, and verified its detection accuracy. In addition to the crack, we also detected the trace of tie-rod holes and the trace of formwork that can easily become the cause of false detection, and by excluding these from the detection results, it was possible to detect the crack with greater accuracy.

Additionally, by applying the crack results using the methods shown in this study to the actual maintenance and management fields, it was not only possible to calculate the crack width based on the capture distance, but by thinning the crack detection results and creating a vector data, it is also considered possible to manage these as CAD data.

Some issues moving forward are described below. When detecting the crack, cutout images are overlapped to prevent detection leaks at the edges, but there is a possibility that a small number of edges of the original image may not be detected. This problem can be resolved if the images can be overlapped, so it is considered that one solution may be to pad the near of the boundary of the original image with a mirror image.

Additionally, by applying semantic segmentation in this study, it may be possible to propose a new direction in the evaluation of damage based on the image. For example, it is considered possible to obtain more information from images such as those after SIFCON blast obtained from the author's study (Chun *et al*. 2013), and we are currently prompting the study of these.

Acknowledgements

This work was partly supported by Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Infrastructure Maintenance, Renovation, and Management" (Funding agency: Japan Science and Technology Agency), and KAKENHI JP16K06440JP16K06440.

References

Abdel-Qader, I., Abudayyeh, O. and Kelly, M., (2003). "Analysis of edge-detection techniques for crack identification in bridges." *Journal of Computing in Civil Engineering*, 17(4), 255-263.

Amarasiri, S., Gunaratne, M. and Sarkar, S., (2009). "Modeling of crack depths in digital images of concrete pavements using optical reflection properties." *Journal of Transportation Engineering*, 136(6), 489-499.

Badrinarayanan, V., Kendall, A. and Cipolla, R., (2017). "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495.

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

Blake, A., Rother, C., Brown, M., Perez, P. and Torr, P., (2004). "Interactive image segmentation using an adaptive GMMRF model." In: T. Pajdla and J. Matas Eds. *Proceedings of the 8th European Conference on Computer Vision, Part 1*, Prague, Czech Republic 11-14 May 2004. Berlin and Heidelberg: Springer, 428-441.

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

Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S. and Büyüköztürk, O., (2018). "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types." *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 731-747.

Chun, P., Lee, S. H., Cho, S. H. and Lim, Y. M., (2013). "Experimental study on blast resistance of SIFCON." *Journal of Advanced Concrete Technology*, 11(4), 144-150.

Chun, P. and Igo, A., (2015). "Crack detection from image using random forest." *Journal of Japan Society of Civil Engineers, Ser. F3 (Civil Engineering Informatics)*, 71(2), I_1- I_8. (in Japanese)

Chun, P., Shimamoto, Y., Okubo, K., Miwa, C. and Ohga, M., (2017). "Deep learning and random forest based crack detection from an image of concrete surface." *Journal of Japan Society of Civil Engineers, Ser. F3 (Civil Engineering Informatics)*, 73(2), I_297-I_307. (in Japanese)

Chun, P., Yamane, T., Izumi, S. and Kameda, T., (2019). "Evaluation of tensile performance of steel members by analysis of corroded steel surface using deep learning." *Metals*, 9(12), 1259.

Chun, P., Dang, J., Hamasaki, S., Yajima, R., Kameda, T., Wada, H., Yamane, T., Izumi, S. and Nagatani, K., (2020a). "Utilization of unmanned aerial vehicle, artificial intelligence, and remote measurement technology for bridge inspections." *Journal of Robotics and Mechatronics*. (in press)

Chun, P., Izumi, S. and Yamane, T., (2020b). "Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine." *Computer-Aided Civil and Infrastructure Engineering*. (in press)

Chun, P., Ujike, I., Mishima K., Kusumoto, M. and Okazaki, S., (2020c). "Random forest-based evaluation technique for internal damage in reinforced concrete featuring multiple nondestructive testing results." *Construction and Building Materials*, 253(30), Article ID 119238.

Chun, P., Yamane, T., Izumi S. and Kuramoto N., (2020d). "Development of a machine learning-based damage identification method using multi-point simultaneous acceleration measurement results." *Sensors*, 20(10), 2780.

Dixit, A. and Wagatsuma, H., (2018). "Comparison of effectiveness of dual tree complex wavelet transform and anisotropic diffusion in MCA for concrete crack detection." In: L. O'Conner, Ed. *Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2018)*, Miyazaki, Japan 7-10 October 2018. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 2681-2686.

Dung, C. V., (2019). "Autonomous concrete crack detection using deep fully convolutional neural network." *Automation in Construction*, 99, 52-58.

Deng, J., Lu, Y. and Lee, V. C. S., (2019). "Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network." *Computer-Aided Civil and Infrastructure Engineering*, 35(4), 373-388.

Fei, Y., Wang, K. C., Zhang, A., Chen, C., Li, J. Q., Liu, Y., Yang, G. and Li, B., (2019). "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V." *IEEE Transactions on Intelligent Transportation Systems*, 21(1), 273-284.

Fujita, Y., Mitani, Y. and Hamamoto, Y., (2006). "A method for crack detection on a concrete structure." In: B. Werner, Ed. *Proceedings of the 18th International*

Conference on Pattern Recognition (ICPR'06), Hong Kong, China 20-24 August 2006. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 3, 901-904.

Fujita, Y. and Hamamoto, Y., (2009). "A robust method for automatically detecting cracks on noisy concrete surfaces." In: B. C. Chien, T. P. Hong, S. M. Chen and M. Ali, Eds. *Proceedings of the 22nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Tainan, Taiwan 24-27 June 2009. Berlin: Springer, 76-85.

Fujita, Y. and Hamamoto, Y., (2011). "A robust automatic crack detection method from noisy concrete surfaces." *Machine Vision and Applications*, 22(2), 245-254.

He, K., Zhang, X., Ren, S. and Sun, J., (2016). "Deep residual learning for image recognition." In: L. O'Conner, Ed. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA 27-30 June 2016. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 770-778.

He, K., Gkioxari, G., Dollár, P. and Girshick, R., (2017). "Mask R-CNN." In: L. O'Conner, Ed. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy 22-29 October 2017. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 2961-2969.

Hutchinson, T. C. and Chen, Z., (2006). "Improved image analysis for evaluating concrete damage." *Journal of Computing in Civil Engineering*, 20(3), 210-216.

Ito, A., Aoki, Y. and Hashimoto, S., (2002). "Accurate extraction and measurement of fine cracks from concrete block surface image." In: *Proceedings of the 28th Annual Conference of the Industrial Electronics Society (IECON 02)*, Sevilla, Spain 5-8 November 2002. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 3, 2202-2207.

Iyer, S. and Sinha, S. K., (2006). "Segmentation of pipe images for crack detection in buried sewers." *Computer-Aided Civil and Infrastructure Engineering*, 21(6), 395-410.

Jiang, S. and Zhang, J., (2019). "Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system." *Computer-Aided Civil and Infrastructure Engineering*, 35(6), 549-564.

Kobayashi, T., (2018). "Spiral-Net with F1-based optimization for image-based crack detection." In: C. V. Jawahar, H. Li, G. Mori and K. Schindler, Eds. *Proceedings of the 14th Asian Conference on Computer Vision*, Perth, Australia 2-4 December 2018. Berlin: Springer, Part I, 88-104.

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B. and Fieguth, P., (2015). "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure." *Advanced Engineering Informatics*, 29(2), 196-210.

Lee, B. Y., Kim, Y. Y., Yi, S. T. and Kim, J. K., (2013). "Automated image processing technique for detecting and analysing concrete surface cracks." *Structure and Infrastructure Engineering*, 9(6), 567-577.

Li, S. and Zhao, X., (2019). "Image-based concrete crack detection using convolutional neural network and exhaustive search technique." *Advances in Civil Engineering*, 2019(2), 1-12.

Liu, Y., Yao, J., Lu, X., Xie, R. and Li, L., (2019). "DeepCrack: A deep hierarchical feature learning architecture for crack segmentation." *Neurocomputing*, 338, 139-153.

Long, J., Shelhamer, E. and Darrell, T., (2015). "Fully convolutional networks for semantic segmentation." In: S. Lakušić, Ed. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA 8-10 June 2015. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 3431-3440.

Luxmoore, A., (1973). "Holographic detection of cracks in concrete." *Non-Destructive Testing*, 6(5), 258-263.

Minami, T., Urata, W., Fujiu, M., Fukuoka, T., Sagae, M., Suda, S. and Takayama, J., (2019a). "Development of automatic concrete cracks detection system using average shifted mesh." *Journal of the Eastern Asia Society for Transportation Studies*, 13, 1560-1570.

Minami, T., Fujiu, M., Takayama, J., Suda, S. and Okumura, S., (2019b). "A study on image diagnostic technology for bridge inspection using ultra high resolution camera." In: *Proceedings of the 5th International Conference on Road and Rail Infrastructure*, Zadar, Croatia 17-19 May 2019. Alexandria, VA, USA: International Road Federation, 79-85.

MLIT, (2018). "*Roads in Japan 2018*." Tokyo: Ministry of Land, Infrastructure, Transport and Tourism, Japan.

Mohan, A. and Poobal, S., (2018). "Crack detection using image processing: A critical review and analysis." *Alexandria Engineering Journal*, 57(2), 787-798.

Moon, H. S., Ok, S., Chun, P. J. and Lim, Y. M., (2019). "Artificial neural network for vertical displacement prediction of a bridge from strains (Part 1): Girder bridge under moving vehicles." *Applied Sciences*, 9(14), 2881.

Moon, H. S., Chun, P. J., Kim, M. K. and Lim, Y. M., (2020). "Artificial neural network for vertical displacement prediction of a bridge from strains (Part 2): Optimization of strain-measurement points by a genetic algorithm under dynamic loading." *Applied Sciences*, 10(3), 777.

Okazaki, Y., Okazaki, S., Asamoto, S. and Chun, P., (2020). "Applicability of machine learning to a crack model in concrete bridges." *Computer-Aided Civil and Infrastructure Engineering,* 35(8), 1-18.

Rafiei, M. H., Khushefati, W. H., Demirboga, R. and Adeli, H., (2017). "Supervised deep restricted Boltzmann machine for estimation of concrete." *ACI Materials Journal*, 114(2), 237-244.

Rafiei, M. H. and Adeli, H., (2018a). "A novel unsupervised deep learning model for global and local health condition assessment of structures." *Engineering Structures*, 156, 598-607.

Rafiei, M. H. and Adeli, H., (2018b). "Novel machine-learning model for estimating construction costs considering economic variables and indexes." *Journal of Construction Engineering and Management*, 144(12), Article ID 04018106.

Ren, S., He, K., Girshick, R. and Sun, J., (2015). "Faster R-CNN: Towards real-time object detection with region proposal networks." In: C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, Eds. *Proceedings of 28th Conference on Advances in Neural Information Processing Systems (NIPS 2015)*, Montreal, Canada 7-12 December 2015. San Diego, CA, USA: Neural Information Processing Systems Foundation, 91-99.

Silva, W. R. L. D. and Lucena, D. S. D., (2018). "Concrete cracks detection based on deep learning image classification." In: D. V. Hemelrijck, D. Aggelis, N. D. Belie, F. Delaunois, T. Geernaerts, P. Guillaume, A. M. Habraken, P. Hendrick, E. Reynders, A. Simar and S. Vanlanduit, Eds. *Proceedings of the 18th International Conference on Experimental Mechanics (ICEM18)*, Brussels, Belgium 1-5 July 2018. Basel, Switzerland: Multi-Digital Publishing Institute, 2(8), 489.

Sinha, S. K. and Fieguth, P. W., (2006). "Segmentation of buried concrete pipe images." *Automation in Construction*, 15(1), 47-57.

Sohn, H. G., Lim, Y. M., Yun, K. H. and Kim, G. H., (2005). "Monitoring crack changes in concrete structures." *Computer-Aided Civil and Infrastructure Engineering*, 20(1), 52-61.

Vezhnevets, V. and Konouchine, V., (2005). "GrowCut: Interactive multi-label ND image segmentation by cellular automata." In: *Proceedings of 15th International Conference on Computer Graphics and Applications (GraphiCon 2005)*, Novosibirsk Academic Town, Russia 20-24 June 2005. Novosibirsk, Russia: Russian Foundation for Basic Research, 1(4), 150-156.

Ward, G. J., (1992). "Measuring and modeling anisotropic reflectance." *Computer Graphics*, 26(2), 265-272.

Xue, Y. and Li, Y., (2018). "A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects." *Computer-Aided Civil and Infrastructure Engineering*, 33(8), 638-654.

Yamaguchi, T., Nakamura, S., Saegusa, R. and Hashimoto, S., (2008). "Image based crack detection for real concrete surfaces." *IEEJ Transactions on Electrical and Electronic Engineering*, 3(1), 128-135.

Yamaguchi, T. and Hashimoto, S., (2008). "Improved percolation-based method for crack detection in concrete surface images." In: *Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08)*, Tampa, Florida, USA 8-11 December 2008. Piscatway, New Jersey, USA: Institute of Electrical and Electronics Engineers, 1-4.

Yamaguchi, T. and Hashimoto, S., (2009). "Practical image measurement of crack width for real concrete structure." *Electronics and Communications in Japan*, 92(10), 1-12.

Yamaguchi, T. and Hashimoto, S., (2010). "Fast crack detection method for large-size concrete surface images using percolation-based image processing." *Machine Vision and Applications*, 21(5), 797-809.

Yamane, T. and Chun, P., (2019). "Crack detection from an image of concrete surface based on semantic segmentation by deep learning." *Journal of Structural Engineering*, 65A, 130-138. (in Japanese)

Yang, X., Li, H., Yu, Y., Luo, X., Huang, T. and Yang, X., (2018). "Automatic pixel-level crack detection and measurement using fully convolutional network." *Computer-Aided Civil and Infrastructure Engineering*, 33(12), 1090-1109.

Yokoyama, S. and Matsumoto, T., (2017). "Development of an automatic detector of cracks in concrete using machine learning." *Procedia Engineering*, 171, 1250-1255.

Zakeri, H., Nejad, F. M. and Fahimifar, A., (2017). "Image based techniques for crack detection, classification and quantification in asphalt pavement: A review." *Archives of Computational Methods in Engineering*, 24(4), 935-977.

Zhang, X., Rajan, D. and Story, B., (2019). "Concrete crack detection using context aware deep semantic segmentation network." *Computer-Aided Civil and Infrastructure Engineering*, 34(11), 1-21.

Zhang, A., Wang, K. C. P., Fei, Y., Liu, Y., Chen, C., Yang, G., Li, J. Q., Yang, E. and Qiu, S., (2018a). "Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network." *Computer-Aided Civil and Infrastructure Engineering*, 34(3), 213-229.

Zhang, A., Wang, K. C. P., Fei, Y., Liu, Y., Tao, S., Chen, C., Li, J. Q. and Li, B., (2018b). "Deep learning-based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet." *Journal of Computing in Civil Engineering*, 32(5), Article ID 04018041.