

Received August 4, 2020, accepted August 28, 2020, date of publication September 10, 2020, date of current version September 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022786

Multiscale and Adversarial Learning-Based Semi-Supervised Semantic Segmentation Approach for Crack Detection in Concrete Structures

SEUNGBO SHIM^{1,2}, JIN KIM², GYE-CHUN CHO², AND SEONG-WON LEE¹

¹Future Infrastructure Research Center, Korea Institute of Civil Engineering and Building Technology (KICT), Goyang 10223, South Korea

²Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea

Corresponding author: Seong-Won Lee (swlee@kict.re.kr)

This research was supported by a grant (20CTAP-C157234-01) from the Technology Advancement Research Program (TARP) funded by the Ministry of Land, Infrastructure, and Transport of Korean government.

ABSTRACT Typically, the operational lifetime of underground concrete structures is several decades. At present, many such structures are approaching their original life expectancy. In this stage, the essential functionality of the structures may be considerably degraded, leading to various safety hazards such as collapse roof and tunnel flooding. In general, to overcome such problems, the maintenance of underground structures has been conducted through manual subjective inspections so far. However, recently, several objective inspection technologies have been actively developed by fusing artificial intelligence and imaging techniques recently. In particular, deep learning algorithms have been developed to detect concrete cracks, based on a large amount of data for supervised learning, including numerous labeled images. Such data acquisition requires considerable time and effort. To reduce these costs, in this study, multiscale and adversarial learning techniques were applied to realize crack detection. A total of 1,200 labeled data and 3,000 unlabeled data were used to implement and verify the proposed method. The multiscale segmentation neural network, discriminator neural network, and adversarial learning technique were used to realize accurate crack detection, enhance the learning performance, and ensure the efficiency of training data, respectively. The resulting algorithm had a pixel accuracy, mean intersection over union, frequency weighted intersection over union, and F1 score of 98.176%, 88.936%, 96.525%, and 88.789%, respectively. The proposed technique can be used to examine the conditions to ensure the safe maintenance of aging structures.

INDEX TERMS Adversarial learning, multiscale learning, crack detection, semantic segmentation, image processing.

I. INTRODUCTION

A. NECESSITY OF UNDERGROUND STRUCTURES MAINTENANCE

Underground concrete structures include various types of road, railway, and utility tunnels, most of which are designed to be utilized over several decades. In recent years, with the aging of many such underground structures worldwide, maintenance technology become essential for the public safety as old structures have potential risk. For example, in the United States, the majority of the road and rail tunnels were

constructed more than 50 years ago [1]. Japan is currently encountering similar situation [2]. According to Fujino *et al.*, the construction of Japanese highways was commenced in 1960s, and several such highways are no longer in a satisfactory operational condition. In Korea, the proportion of structures aged more than 30 years will increase to 33.7% by 2029, and thus, there is an increasing interest in ensuring the safety of existing structures [3].

The types of damage to underground concrete structures can be categorized as cracks, peeling, cavity, and efflorescence, among other phenomena. These aspects can degrade the functionality of the structure and cause safety problems that threaten the public. Consequently, prompt inspection and

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

repair must be performed. In general, the facility safety management agency examines the various types of concrete damage systematically through certain inspection processes [4]. However, the inspection is mainly performed visually and manually by using a crack microscope and crack rulers. Nevertheless, such approach involves a low objectivity and reliability because the inspectors determine the presence or absence of damage based on their experience and subjective decision making. To enable objective evaluation, the quantification and digitization of the information must be realized using equipment such as high-resolution image sensors [5].

B. RELATED RESEARCH

1) CRACK DETECTION WITHOUT DEEP LEARNING

Until now, various type of sensor technologies have been introduced to detect the different types of damage occurring in underground structures. Image and light detection and ranging (lidar) sensors have been used to detect the occurrence of cracks, spalling, corrosion, efflorescence, and other phenomena. In particular, the inherent filters and algorithms were applied to develop techniques to realize facility maintenance [5]–[9]. Among these aspects, the algorithm for crack detection has been most actively researched. According to the Korean government policy, the acceptable width of cracks in concrete structures is less than 0.3 mm [10]. Therefore, most of the existing studies were focused on detecting micro-cracks. To this end, image processing through a high-resolution image sensor, which employs a traditional technique involving pattern recognition for cracks, has been actively researched. One of the unique patterns of cracks is a boundary line. As this line appears to be created on the flat pavement surface, it is considered as an edge in the image. The cracked area is shaded owing to the lower reflection of light in this region than on the other surfaces, which leads to darkening. This difference in the brightness sharpens the crack border. Consequently, the edge of the crack can be detected clearly using several edge detection algorithms [11]–[14]. However, under unexpectedly bright natural sunlight conditions or low illumination, the detection performance of this approach is decreased.

Moreover, in the case of filter-based algorithms, a Gabor filter [15], [16], wavelet transform [17], [18], and Hough transform [19] are used to detect the cracks. The filter-based approach can be considered to a transformation method used to extract the unique feature of cracks through a region-based operation. Using the meaningful feature, the crack region is segmented after the foreground region is separated from the background by considering a certain threshold. However, the threshold can only be determined through many trial and error cycles and is highly sensitive to the detection environment, thereby requiring modifications according to the application field.

To improve the detection accuracy, machine learning techniques can be applied to identify the crack region by using the features determined from various filters. Such algorithms

extract the features from the input image, which are then used to train the machine learning model. The model extracts the actual crack from the components that involve cracks with a similar shape. A typical machine learning model is the support vector machine, which is used to enhance the detection precision [20], [21]. However, this machine learning method exhibits a deteriorated recognition performance when various types of cracks are input. In particular, the recognition result depends on the quality of the filter's output, and thus, the approach cannot be used in universal conditions.

2) CRACK DETECTION THROUGH DEEP LEARNING

In recent times, the concrete surface damage detection through an image processing technology utilizing deep learning has been actively examined. Feng *et al.* proposed an algorithm to detect cracks occurring on the surface of a concrete dam, by using a deep neural network with the auto-encoder technique consisting of encoder and decoder steps [22]. The researchers mounted a 20 MP camera on a drone and acquired the training image data recorded for 18 h. Based on the image data, a deep neural network was designed, and skeleton images were generated to measure the crack length. The resulting algorithm had a recall, precision, and F-measure of 80.45%, 80.31%, and 79.16%, respectively. In addition, a classification based algorithm was proposed to detect cracks as well as other damages occurring on concrete walls [23]. Cha *et al.* collected the image data recorded using a high-resolution DSLR camera and divided them into 256×256 frames to classify the damaged and intact images. These images were processed using the algorithm of deep neural networks for classification, and an accuracy of more than 90% was attained. A similar process was employed by Kim *et al.* [24]. First, the researchers collected images of cracks, joints, etc. through an internet web search using ScrapeBox [25]. Subsequently, a classification algorithm based on deep neural networks was developed. A high-resolution large image was cropped by a small window, and the cropped image was used as the input of the algorithm. The deep learning model determined the presence of cracks in the cropped area. In a further study, the researchers developed a technique to segment the crack region from the image at the pixel-level by using a mask region-based convolutional neural network [10], [26]. The resulting algorithm exhibited an accuracy of 95% with an intersection over union (IoU) of 50%.

3) CRACK DETECTION THROUGH SEMI-SUPERVISED LEARNING

As discussed, artificial intelligence methods represent a type of supervised learning approaches. Supervised learning can lead to an excellent performance in the presence of sufficient training data including annotation data. However, acquiring a large amount of training data is challenging, and in particular, a considerable amount of time is required to acquire the ground truth data for supervised learning. To overcome these limitations, several researchers have focused on the use

of semi-supervised learning. Zhan *et al.* used the mix-and-match technique in self-supervised learning methods [27], and Lee *et al.* proposed a method of randomly selecting hidden units and implementing weakly and semi-supervised learning [28]. This learning technique was fused with generative adversarial learning (GAN) technique, to derive a new form of semi-supervised learning [29]. When using the GAN theory, the time-consuming labeling work could be eliminated owing to the use of a deep neural network to acquire the new training image data [30], [31]. Li *et al.* attempted to apply this method to detect the cracks occurring in road pavements [32], and this approach could likely be used as a core technology for accurate diagnosis maintenance systems to realize the crack detection in concrete structures, which requires a highly precise inspection.

4) CONTRIBUTION OF THE PROPOSED APPROACH

The inspection technology to maintain concrete structures must have a sufficient resolution to detect micro-cracks. In this regard, the detection accuracy as well as the sensor resolution must be ensured. Considering these aspects, this article proposes a novel method to realize concrete crack detection by using multiscale and adversarial learning. The contribution of this work is three-fold which means the novelty of our method. First, to detect the concrete cracks, we developed a segmentation neural network that could realize highly accurate recognition; unlike the deep neural network of the auto-encoder type, the neural network of the proposed approach was trained through the multiscale learning method, which could improve the detection accuracy. Second, to enhance the training performance, a deep neural network based on the connectivity between the layers was proposed to generate new training data from the unlabeled image. Finally, a new adversarial learning structure was designed to train multiscale segmentation neural networks, and a new loss function was proposed to update the weights of the two neural networks, to reduce the amount of required training data.

To demonstrate these contributions, first, the training data acquisition and types of considered concrete cracks are described. Next, the adversarial learning method, including the structure of the segmentation neural network, discriminator neural network, and weight update strategy, is introduced. Furthermore, the experiment performed to compare the proposed method with the existing methods to evaluate its performance is described. Finally, the concluding remarks are presented.

II. CONCRETE CRACK IMAGE DATA FOR DEEP LEARNING

A. CRACK IMAGE

To develop a deep neural network algorithm to detect concrete cracks, image data are required for training. Crack image data can be simply recorded using cameras. However, because this approach requires considerable time and effort, we used the image data already collected in the previous studies.

For example, an image data set corresponding to a building in the Middle East Technical University (METU) [33] was used to develop an algorithm to identify the presence or absence of cracks. Among the 40,000 images, 20,000 images corresponded to cracked and intact images each. Another image data set provided by the Utah State University (USU) was used, the data in which were recorded uses a 16 MP Nikon camera [34]. The data consisted of 54 bridge images, 72 concrete wall images, and 104 pavement images. The images were divided to 256×256 and saved in the JPEG file format. A total of 1,196 images were selected from the two data sets, as shown in Fig. 1. The images corresponding to the METU and USU involved cracks occurring in buildings and bridges, respectively. In this study, these images were used to ensure the diversity of the training data set.

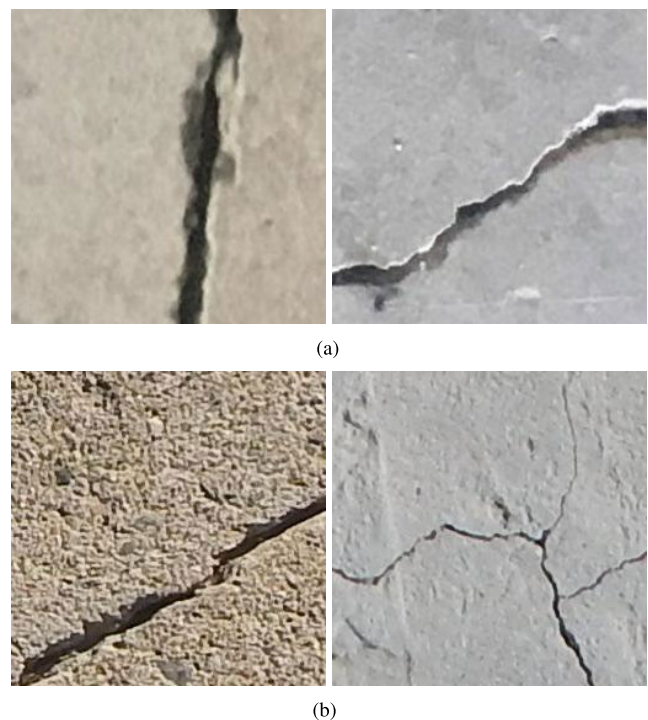


FIGURE 1. Concrete crack images from (a) METU, (b) USU.

B. LABELED IMAGE

To train the segmentation neural network to realize pixel-level classification, labeled images are required to identify the exact crack region in the image. The image data set provided by Dorafshan *et al.* could be used to develop a classification algorithm [34]; however, for the purpose of the present study, labeled images were required to be prepared. To this end, the LEAR Image Annotation Tool [35] was used, as shown in Fig. 2(a). The selected 1,196 images were labeled for training. Finally, we obtained the labeled data, consisting of crack and labeled images to realize supervised learning, as shown in Fig. 2(b). Among those images, 900 images were used for training, and the remaining 300 images were used for

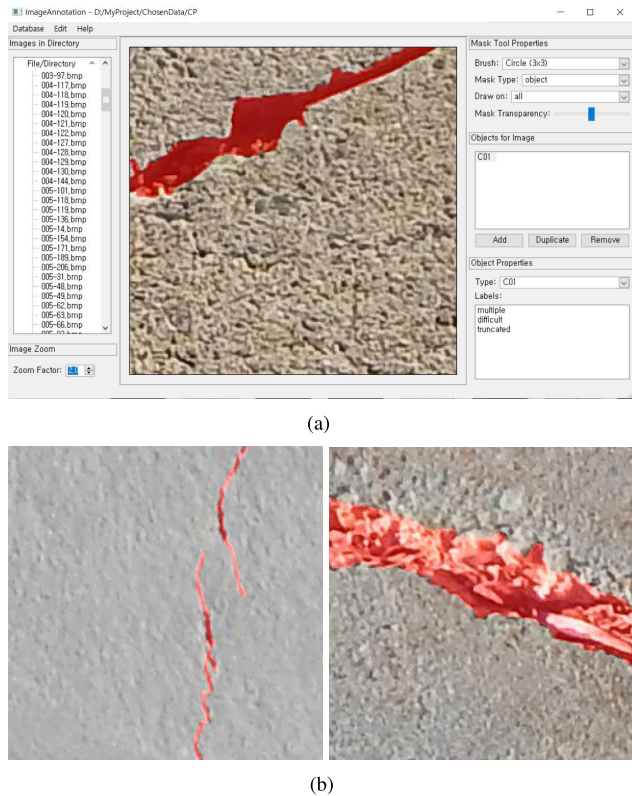


FIGURE 2. Image data annotation for segmentation: (a) Crack image labeling program, (b) Labeled images.

the validation. Moreover, 3000 crack images were added as unlabeled data to implement semi-supervised learning.

III. DESIGN OF MULTISCALE AND ADVERSARIAL LEARNING FRAMEWORK

A. MULTISCALE SEGMENTATION NEURAL NETWORK FOR CRACK DETECTION

A deep neural network was used to detect the concrete cracks in the images. Most of the existing algorithms for pixel-level segmentation use the auto-encoder method (e.g. SegNet [36], FRRN [37], LinkNet [38]), which involves a deep neural network to ensure a high recognition performance in road images. The pixel-level segmentation neural network consists of encoder and decoder stages. The encoder stage extracts a meaningful feature with several convolution operations and finally creates a feature map. The decoder stage localizes the target at the pixel-level through up-sampling operations.

Although this method has a reasonable performance, in this study, to enhance the recognition performance, a novel neural network structure was developed, which could generate the feature maps at several scales in the encoder stage and extract the positions of the objects from the maps. The proposed structure primarily included two stages, namely, the feature and localization stages, as shown in Fig. 3. The former stage was composed of dense and translation blocks, and the latter stage was composed of four link blocks. The neural network

in the feature stage was similar to DenseNet121 [39], which is composed of four dense blocks and three transition blocks. However, to reduce the number of parameters, we modified the original neural network in terms of the following four aspects: The initial number of features was changed from 64 to 16, and the growth rate was changed from 32 to 8. The max pooling operation was eliminated to reduce the size with transition blocks. The dense block values were changed from [6, 12, 24, 16] to [4, 3, 2, 1]. In addition, 1×1 convolutions were replaced by 3×3 convolutions.

The output of the four dense blocks in the feature stage was used as a feature map, along with the input of the link block in the localization stage. The four link blocks were similar to those in the LinkNet [38]. The original network structure consisted of four encoder and four decoder blocks. However, we incorporated three down-size operations in the encoder step and three up-size operations in the decoder step. The LinkNet demonstrates a high performance, even with a simple structure, owing to the presence of the residual block. Consequently, three skip connections were applied between the encoder and decoder steps. Through this configuration, the four link blocks operated simultaneously in the proposed network. The four sub-outputs obtained from the four link blocks were finally restored to a size of 256×256 . The sub-outputs were composed of two channels to ensure that the background and crack could be distinguished. The final output was summed according to the corresponding channel and divided by 4 to realize normalization.

B. DISCRIMINATOR NEURAL NETWORK

The discriminator neural network was used to determine the true and false aspects of the input image. The input of the discriminator neural network was a prediction image sized $256 \times 256 \times 2$, and the output was a confidence map sized $256 \times 256 \times 1$. The discriminator neural network employed in this work was a combination of the convolution-leaky ReLU (CL) blocks [40], as shown in Fig. 4. We used eight CL blocks in the discriminator neural network. The kernel size was 3, and the channels for the 8 CL blocks were set as 64, 64, 128, 128, 256, 256, 512, and 512. In addition, three dense connectivity [39] layers were used between the CL blocks to alleviate the vanishing gradient problem. The size of the features was reduced to 8×8 over five stages. When the prediction and labeled images were input, the corresponding outputs were considered as the false and true maps, respectively. In this manner, the discriminator neural network could identify the true labeled images.

C. ADVERSARIAL LEARNING METHOD

To perform adversarial learning, the training data set was categorized into two groups: In the group corresponding to the labeled data, the crack and labeled images were paired for the supervised learning, and in the other group, corresponding to the unlabeled data, only crack images were used for the semi-supervised learning. The two data set groups were used to simultaneously realize three types of learning: The first

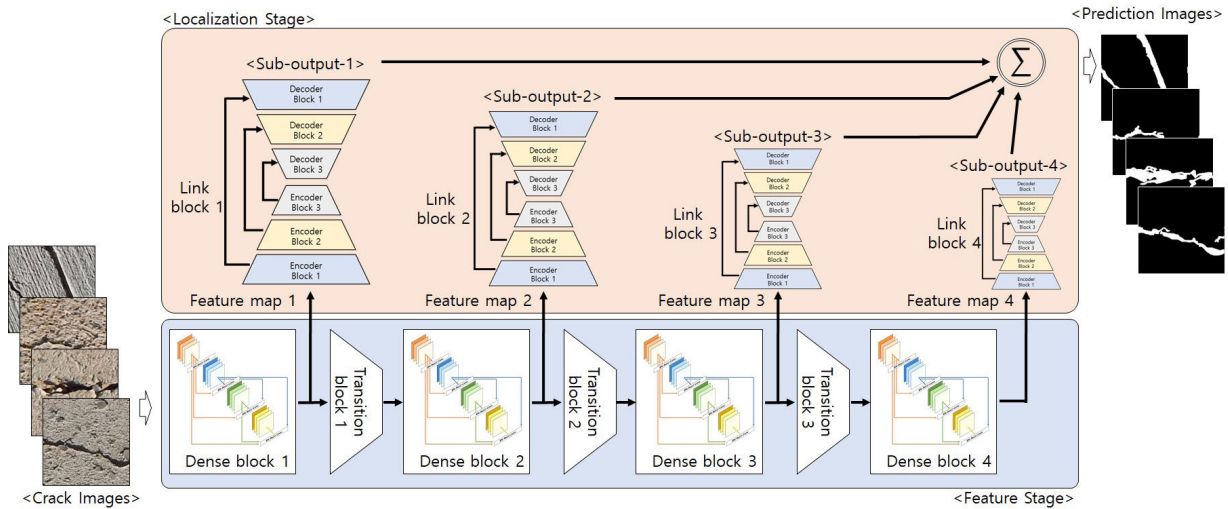


FIGURE 3. Structure of multiscale segmentation neural network.

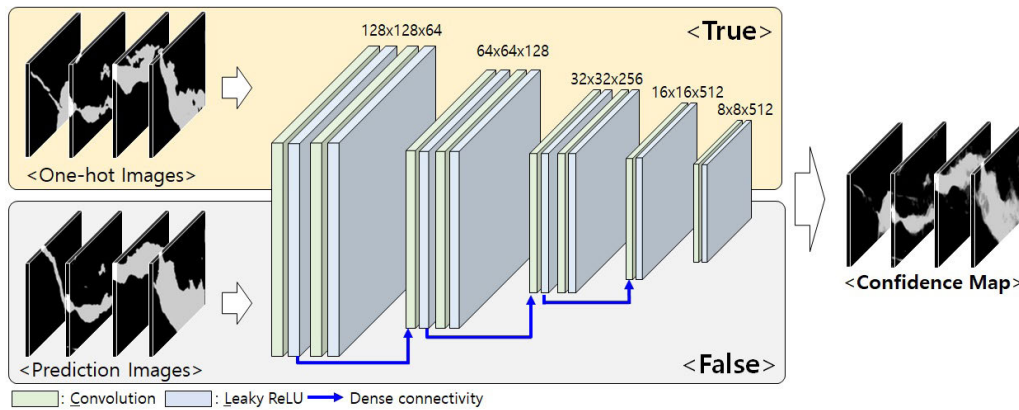


FIGURE 4. Structure of the discriminator neural network.

type corresponded to the training of the discriminator neural network, for which the input included prediction images and one-hot images. The prediction image was the output image obtained when the crack images of the labeled and unlabeled data were input to the segmentation neural network. The one-hot image was the normalized image, in which the crack and background regions were marked as 1 and 0 in the labeled image, respectively. The two sets of images were input to the discriminator neural network, and a confidence map was output. This confidence map was considered to be true and false for the results pertaining to the one-hot image and prediction image, respectively. However, because it was impossible to create a virtual image with such honest learning, cheat learning was incorporated. In particular, when a confidence map was produced by the prediction image obtained from the crack image of unlabeled data, it was regarded as true. Using this duping method, the discriminator could be trained under a mutually competitive frame.

Next, we trained the segmentation neural networks through supervised learning. To update the weight of the proposed segmentation neural network, the image resize operation was performed. For multiscale learning, the size of the four sub-outputs and corresponding size of the labeled images was required to be the same. Therefore, the labeled image sized 256×256 was reduced to images sizes 128×128 , 64×64 , and 32×32 . Each resized labeled image was connected to the sub-output corresponding to its size and later used to update the weight of the segmentation neural network.

Finally, semi-supervised learning was performed to train the segmentation neural network by using the crack image of the unlabeled data. The confidence map, which was the output of the adequately trained discriminator neural network, was considered as the labeled image. This confidence map was subjected to the image resize operation to generate four labeled images. These labeled images were matched to the sub-output of the segmentation neural network, and its weight

was updated in the same manner as in the supervised learning process.

D. LOSS FUNCTION OF MULTISCALE AND ADVERSARIAL LEARNING

The loss function of adversarial learning was applied to the two deep neural networks simultaneously. The total loss was the sum of the losses of the discriminator and segmentation network, as indicated in Eq. (1). The total loss value, and the loss values of the discriminator and segmentation network were denoted as L_{all} , L_D and L_{Seg} , respectively.

$$L_{all} = L_D + L_{Seg} \quad (1)$$

The loss function of the discriminator neural network can be defined as in Eq. (2), where X and Y denote the crack and labeled images, respectively. $S(\cdot)$ and $D(\cdot)$ denote the prediction image and confidence map resulting from the segmentation and discriminator neural networks, respectively. y denotes the weight balance, whose value is 0 or 1 when a prediction or labeled image is input, respectively. In addition, i denotes the position of a pixel in the image, and N denotes the total number of pixels.

$$L_D = - \sum_i^N (1 - y_i) \log(1 - D(S(X))_i) + y_i \log(D(Y)_i) \quad (2)$$

The loss function of the segmentation neural network can be defined as in Eq. (3), and it consists of three loss functions. The first loss function is the cross-entropy function (Eq. (4)) used when implementing supervised learning. $Sub(\cdot)$ denotes the sub-output of a multiscale segmentation neural network, and $N(s)$ indicates the total number of pixels of the sub-output, corresponding to the scale of s . Y denotes the pixel value in a one-hot image. The second loss function, L_{adv} , corresponds to the discriminator neural network, and can be defined as in Eq. (5). In particular, this parameter is a function of $D(\cdot)$ when $S(\cdot)$ is input. The third loss function corresponds to semi-supervised learning and can be defined as in Eq. (6). In this equation, T_{semi} is set as 0.3, and I is composed of zeroes or ones as the identity matrix. Moreover, \hat{Y} , which is a binary digit, is the pixel value of the image predicted as a crack image by the segmentation neural network.

$$L_{Seg} = L_{CE} + L_{adv} + L_{semi} \quad (3)$$

$$L_{CE} = - \sum_{s=1}^4 \sum_i^{N(s)} Y_{s,i} \log(Sub(X)_{s,i}) \quad (4)$$

$$L_{adv} = - \sum_i^N \log(D(S(X))_i) \quad (5)$$

$$L_{semi} = - \sum_{s=1}^4 \sum_i^{N(s)} I(D(X)_{s,i} > T_{semi}) \times \hat{Y}_{s,i} \log(Sub(X)_{s,i}) \quad (6)$$

IV. EXPERIMENT DETAILS

A. METHODOLOGY

Four experiments were conducted to evaluate the recognition performance of the proposed method. First, the performance of two segmentation neural networks, namely, the auto-encoder type and multiscale type networks, was compared. To enable a fair comparison, only labeled data was used, and the obtained accuracy was evaluated under the same conditions. The second experiment was performed to evaluate the performance of the proposed discriminator neural network and examine the amount of improvement in the learning performance when implementing dense connectivity. The findings were compared with those of the Fully Convolutional Network (FCN) [41] to clarify the difference between the two discriminators. In the third experiment, the adversarial learning method was considered. The efficiency of the approach was clarified by simultaneously applying the proposed segmentation neural network and discriminator neural network and evaluating the improvement in the recognition performance compared to that of the original method. Finally, the training parameters, which influence the learning performance, were analyzed to optimize the process.

B. IMPLEMENTATION AND COMPUTING ENVIRONMENT

For supervised learning, the initial values were set following Glorot and Bengio [42]. ADAM was used as the optimization function. Its the learning rate, beta-1, and beta-2 were set as 0.001, 0.9, and 0.999, respectively [43]. The number of epochs was 500, and the batch size was 4 for every supervised learning process.

For adversarial learning, ADAM was used as the optimization function for the segmentation and discriminator neural networks. Most of the optimization parameters were the same as those for the supervised learning. However, the learning rate was set as 0.0025 and 0.001 for the segmentation and discriminator neural networks, respectively. The batch size was 4 and the number of iterations was 30,000. Until 5,000 iterations, the weight of the segmentation neural network was updated through the labeled data, and subsequently, the weight was updated using the results of the adversarial learning.

The algorithm was implemented under the Ubuntu 18.04 framework. The deep learning framework was Pytorch, all the experiments were performed under the same learning environment, and the development PC specifications were as follows: i10-109800, 128 GB RAM, and NVIDIA-RTX 2080TI.

C. PERFORMANCE EVALUATION METRICS

To evaluate the accuracy of the segmentation neural network, four indices, namely, the pixel accuracy ($Pix.Acc.$), mean IoU ($M.IoU$), frequency weighted IoU ($F.IoU$) and F1 score ($F1$) were used, as described in Eqs. (7), (8),(9), and (10), respectively. In Eq. (7), $n_{i,j}$ indicates the number of pixels in

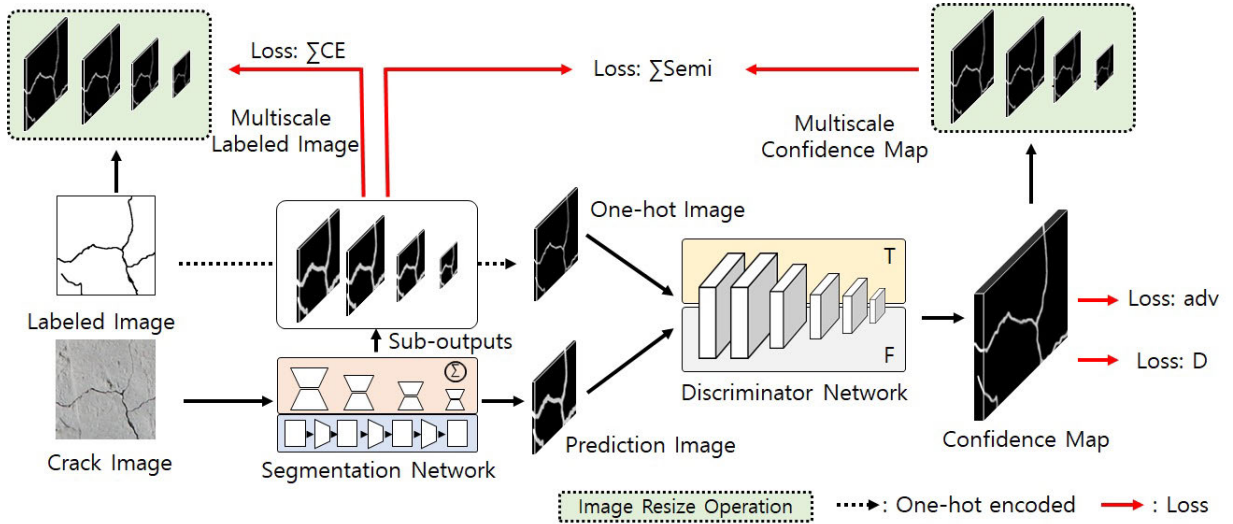


FIGURE 5. Adversarial learning framework.

class i predicted to belong to class j . In Eq. (8), n_{cl} denotes the number of classes, and t_i indicates the number of all pixels belonging to class i . Eq. (7) indicates the accuracy in pixel units, and Eqs. (8) and (9) indicate the accuracy in area units. The F1 score is the harmonic mean of the precision (Pr) and recall (Re). These indices are used to assess the crack prediction accuracy of the neural network model.

$$Pix.Acc. = \frac{\sum_i n_{ii}}{\sum_i t_i}, \quad \text{where } t_i = \sum_j n_{ij} \quad (7)$$

$$M.IoU = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \quad (8)$$

$$F.IoU = \left(\sum_k t_k \right)^{-1} \sum_i \frac{t_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \quad (9)$$

$$F1 = \frac{2 \times Pr \times Re}{Pr + Re} \quad (10)$$

D. RESULTS

1) SEGMENTATION NEURAL NETWORK

Three different deep neural networks were considered for the performance comparison with the segmentation neural network, as indicated in Table 1. All the networks were of the auto-encoder type, and their mean IoU and F1 scores are nearly equivalent with values of more than 80%. The mean IoU of the FRRN and LinkNet was more than 86%, and the F1 score of LinkNet was more than 86%. Nevertheless, the proposed multiscale segmentation neural network denoted as Ours(S) outperformed the other three auto-encoder type algorithms in terms of all the indicators. In particular, the mean IoU and F1 score of Ours(S) were 87.252% and 86.877%, which were 0.266% and 0.328% higher than those of the LinkNet, respectively.

TABLE 1. Results of the segmentation.

Model	Evaluation metrics			
	Pix. Acc.	M.IoU	F.IoU	F1
SegNet	97.498	84.894	95.261	84.035
FRRN	97.747	86.197	95.697	85.589
LinkNet	97.846	86.986	95.911	86.549
Ours(S)	97.868	87.252	96.011	86.877

TABLE 2. Results for the discriminator.

Model		Evaluation metrics			
Disc.	Seg.	Pix.Acc.	M.IoU.	F.IoU	F1
FCD	SegNet	97.538	84.978	95.310	84.118
	FRRN	97.760	86.502	95.755	85.981
	LinkNet	97.882	87.128	95.968	86.706
	Ours(S)	98.009	87.721	96.180	87.374
Ours(D)	SegNet	97.655	85.447	95.494	84.660
	FRRN	97.887	87.115	95.969	86.684
	LinkNet	97.903	87.169	95.995	86.742
	Ours(S)	98.176	88.936	96.525	88.789

2) DISCRIMINATOR NEURAL NETWORK

To assess the performance of the discriminator network, 900 and 3,000 images corresponding to the labeled and unlabeled data were used, respectively. The testing was performed by connecting different discriminator neural networks to the four segmentation neural networks. For the discriminator neural network, the FCD tested by Hung *et al.* and the proposed Ours(D) network were used. After the training completion, the segmentation neural network model was applied to the validation data and its performance was compared with that of supervised learning. The corresponding results are presented in Table 2. It could be noted that the performance of all the eight models improved in terms of all the indicators, regardless of the type of discriminator used. Though, the comparison of the two discriminators indicated that the Ours(D) network outperformed the FCD in terms of all the

indicators. In particular, when the FCD was connected to Ours(S), the mean IoU and F1 score increased by 0.469% and 0.497%, respectively. In addition, when Ours(D) was connected to Ours(S), the mean IoU and F1 score increased by 1.684% and 1.912%, respectively. The performance increment in the case of Ours(D) was higher than that when using FCD. Thus, it was concluded that the learning performance of the proposed discriminator was higher than that of the FCD.

3) ADVERSARIAL LEARNING

The influence of adversarial learning was evaluated according to the amount of labeled data used. In particular, the performance was evaluated considering the labeled data ratio, corresponding to 1/2, 1/4 and 1/8 of 900 images, along with the labeled images. For adversarial learning, 3,000 additional unlabeled images were used. The results presented in Table 1 were defined as the baseline, and the data ratio required to cross the relevant values was identified as shown in Table 3. When 1/2 of the data was applied, the frequency weighted IoU of SegNet was 95.241%, 0.02% less than one of SegNet's baseline. When 1/4 of the data was used, the learning results of FRRN and LinkNet did not exceed the baseline in all the indicators. Finally, when 1/8 of the data was used, the frequency weighted IoU and F1 score of Ours(S) were 95.999% and 86.862%, which were less than Ours(S)'s baseline by 0.012% and 0.015%, respectively. In summary, the amount of data used in semi-supervised learning was not reduced when using the SegNet. The FRRN and LinkNet required only 1/2 of the supervised learning data to cross the baseline. The proposed Ours(S) and Ours(D) networks were the most efficient neural network, as it could achieve the target performance using only 1/4 of the data required for the supervised learning.

TABLE 3. Results for adversarial learning.

Data amount	Model		Evaluation metrics			
	Disc.	Seg.	Pix.Acc.	M.IoU	F.IoU	F1
1/2		SegNet	97.645	85.862	95.241	85.223
	FCD	FRRN	97.797	86.260	95.755	85.635
		LinkNet	97.847	87.050	95.922	86.631
	Ours(D)	Ours(S)	98.100	88.772	96.427	88.639
1/4		SegNet	97.244	83.151	94.753	81.824
	FCD	FRRN	97.467	84.497	95.171	83.518
		LinkNet	97.512	84.891	95.273	84.021
	Ours(D)	Ours(S)	98.074	88.303	96.330	88.068
1/8		SegNet	96.379	78.605	93.248	75.814
	FCD	FRRN	97.345	83.856	94.959	82.731
		LinkNet	97.159	82.851	94.628	81.470
	Ours(D)	Ours(S)	97.894	87.256	95.999	86.862

4) HYPERPARAMETERS

For adversarial learning, optimal parameters must be set to update the weight of the neural network. In this study, we optimized the parameters as indicated in Table 4 to attain the highest accuracy. Half of the labeled data and 3,000 unlabeled images were used for training. Moreover, λ_{adv} was set as 0.01. The change in the mean IoU value was determined

TABLE 4. Results for the hyperparameters.

Data amount	λ_{adv}	λ_{semi}	T_{semi}	M.IoU
1/2	0.01	0.01	0.2	88.289
	0.01	0.05	0.2	88.689
	0.01	0.10	0.2	88.536
1/2	0.01	0.05	0.2	88.689
	0.01	0.05	0.3	88.776
	0.01	0.05	0.4	88.431

according to the variation in λ_{semi} and T_{semi} . To determine the optimal λ_{semi} , it was varied from 0.01 to 0.10 under a fixed T_{semi} . The highest mean IoU was attained when λ_{semi} was 0.05. Next, T_{semi} was varied from 0.2 to 0.4 under a fixed λ_{semi} of 0.05. The highest mean IoU was 88.776% when T_{semi} was 0.3. According to these findings, the optimal parameters corresponded to a λ_{semi} and T_{semi} of 0.05 and 0.3, respectively.

5) CRACK DETECTION

The resulting images for the concrete crack detection are presented in Table 5. The first column shows the crack images containing the damaged spot, and the second column presents the ground truth indicating the damaged area. The subsequent columns show images from the segmentation neural network, obtained after training with all the labeled and unlabeled images. It can be noted that the results obtained using the proposed Ours(S) and Ours(D) networks are more similar to the ground truth, compared to the other resulting images. As indicated in the first and second rows, Ours(S) and Ours(D) detect micro-cracks more clearly and precisely compared to the other networks. Moreover, as indicated in the fifth to seventh rows, the pepper noise is reduced in the detected area with the concrete crack. In conclusion, the proposed model outperforms the other models in terms of the detection performance.


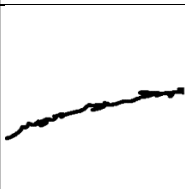
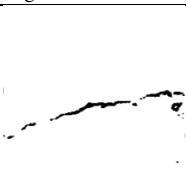
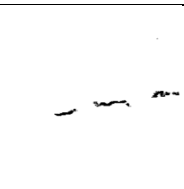
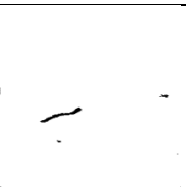
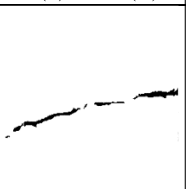









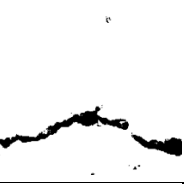
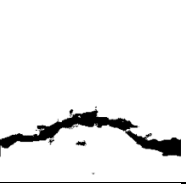


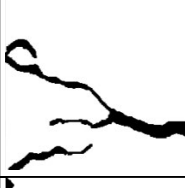







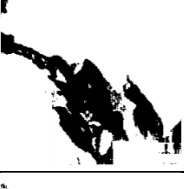





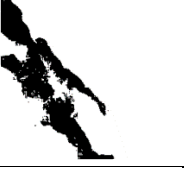



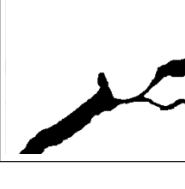

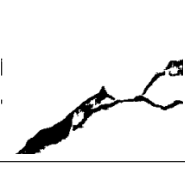
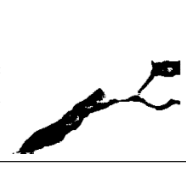
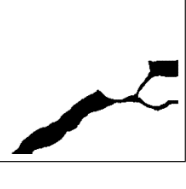
V. DISCUSSION

A. PROPOSED NEURAL NETWORK

Micro-cracks in concrete hinder the structural safety, and thus, the objective of crack detection is to ensure the safety of concrete structures and prolong their lifetime. Therefore, ensuring the detection accuracy is paramount when evaluating the structural safety. To this end, the development of detection technology using deep learning has been actively realized in recent years. A typical example of this technique is the use of segmentation neural networks based on the auto-encoder type algorithm. In this approach, the number of layers is increased to create more unique and meaningful features. Moreover, to alleviate the vanishing gradient problem, many methods combining encoder and decoder networks have been suggested.

Nevertheless, the multiscale segmentation neural network is based on the assumption that a sufficiently meaningful feature is already generated in the encoder, and a simple neural network is applied to estimate the location of the object.

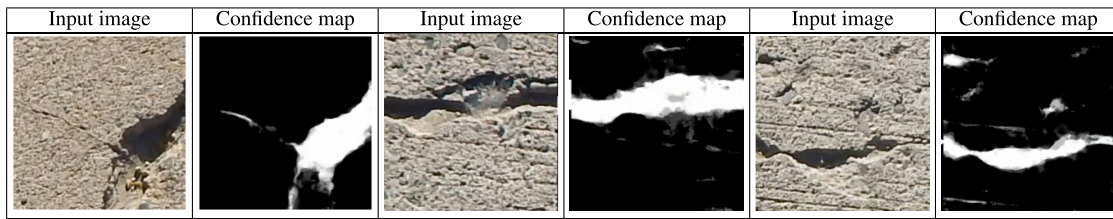
TABLE 5. Results for the concrete crack detection.

Original	Ground truth	SegNet + FCD	FRRN + FCD	LinkNet + FCD	Ours(S) + Ours(D)
					
					
					
					
					
					
					

Considering this aspect, the Ours(S) network was designed to consist of a feature stage and localization stage. In general, as the neural network deepens, the spatial information is lost. To minimize this loss, the information generated in the feature stage can be transferred to the final layer by using a link block. Furthermore, to enable the application of the proposed approach to an embedded system, the number of parameters in Ours(S) was designed to be 12.19 M. Among the other networks, the SegNet, FRRN, and LinkNet networks have 29.44 M, 23.57 M, and 11.53 M parameters, respectively.

In this regard, the Ours(S) network is similar to the LinkNet, and it is expected that field application can be realized without any difficulties. Consequently the multiscale segmentation neural network outperforms the other models, and its number of parameters is relatively small.

Adversarial learning is a type of semi-supervised learning, and it is applied to generate a new labeled image from an unlabeled image. The labeled image generated using a discriminator neural network is known as a confidence map, as indicated in Table 6. In this study, we developed a novel

TABLE 6. Visualization of the confidence maps.

discriminator neural network, Ours(D), to create a more accurate confidence map than that obtained using the FCD. Ours(D) has a more complex structure than that of the FCD and includes 6.23 M parameters, more than two times that in the FCD. Nevertheless, Ours(D) is a superior method in terms of its learning performance. In general, a discriminator neural network is used to increase the accuracy of the segmentation neural network by creating a precise confidence map. Moreover, most of these tasks are performed on the learning server device instead of on the application site, and thus, sufficient computational resource support is available. Therefore, when designing the proposed network, the focus was on enhancing the performance, instead of optimizing the weight or complexity of the discriminator. Consequently, the accuracy of the proposed network was considerably enhanced compared to that of the FCD.

B. MULTISCALE AND ADVERSARIAL LEARNING

As the learning method for the proposed network, the adversarial learning was combined with multiscale segmentation neural networks. In particular, Ours(S) produced four sub-outputs. To train this network, the labeled image was resized to match the size of the sub-outputs. This resizing process was also applied in the adversarial learning process. The confidence map generated from the unlabeled image was connected to the sub-output. Through this approach, the spatial information generated in the confidence map updated weights at various scales. Owing to this updating method, the semi-supervised learning has more effective for the multiscale segmentation neural networks than the auto-encoder neural networks. The difference in the effect was evaluated by connecting LinkNet and FCD and connecting Ours(S) and FCD, as indicated in Table 2. Comparing the performance of each combination with the results of supervised learning, it was noted that LinkNet increased the mean IoU and F1 score by 0.142% and 0.157%, respectively. Furthermore, Ours(S) increased the mean IoU and F1 score by 0.469% and 0.497%, respectively, corresponding to an increase amount that is three times higher those of LinkNet. When the discriminator was replaced by Ours(D), the amount of increase was more than nine times. LinkNet led to an increase of 0.183% and 0.193% in the mean IoU and F1 score, respectively, and the corresponding increase for the Ours(s) network was 1.684% and 1.912%, respectively. It can be concluded that even when using the same discriminator, the multiscale adversarial learning leads to an enhanced performance. In fact,

this phenomenon became more pronounced when the data amount ratio of the labeled data was reduced. By comparing the results of Ours(S) and Ours(D) at 1/8 the data amount, as indicated in Table 3, with those of all the other auto-encoder models, as indicated in Table 1, it could be noted that the Ours(S) and Ours(D) networks outperformed the auto-encoder models. Consequently, the most efficient learning technique corresponded to a combination of the multiscale and adversarial learning approaches.

VI. CONCLUSION

In this study, we developed a detection algorithm and learning method to segment crack areas, which may occur in underground concrete structures, through images. In the existing studies, deep neural networks were developed through supervised learning to detect the damaged areas. However, in this algorithm, a large amount of image data is required, and considerable time and effort are required to generate labeled images for training. To reduce the required time and effort, adversarial learning is a promising technique. Considering this aspect, we proposed a method to create a new precise labeled image from unlabeled images, while improving the recognition performance of segmentation neural networks. When our method is compared with that of the auto-encoder type algorithm, the detection accuracy of the multiscale segmentation neural network was higher. Moreover, the use of dense connectivity was effective in enhancing the learning performance. In addition, it was noted that multiscale and adversarial learning could be combined to reduce the required amount of data. When diagnosing the status of a concrete structure, it is necessary to ensure a high detection accuracy because the safety grade depends on the width of the micro-cracks. The crack characteristics are used to decide whether a structure requires reinforcement or repair. Generally, the crack's physical dimension can be measured by triangulation or photogrammetry methods but basic technology is crack detection algorithm like ours. In this regard, the proposed highly accurate and efficient crack detection algorithm can be used to improve the reliability of structural safety diagnosis.

REFERENCES

- [1] FTA (Federal Transit Administration). (1997). *Inspection Policy and Procedures for Rail Transit Tunnels and Underground Structures*. Transportation Research Board National Research Council, TRCP Synthesis. [Online]. Available: <http://onlinepubs.trb.org/onlinepubs/trcp/tsyn23.pdf>

- [2] Y. Fujino and D. M. Siringoringo, "Recent research and development programs for infrastructures maintenance, renovation and management in japan," *Struct. Infrastruct. Eng.*, vol. 16, no. 1, pp. 3–25, Jan. 2020.
- [3] C.-H. Park and H.-I. Lee, "Future trend of capital investment for Korean transportation infrastructure," in *Proc. Construct. Economy Res. Inst. Korea Seoul*, South Korea, 2016, pp. 6–14.
- [4] M.-K. Kim, H. Sohn, and C.-C. Chang, "Localization and quantification of concrete spalling defects using terrestrial laser scanning," *J. Comput. Civil Eng.*, vol. 29, no. 6, Nov. 2015, Art. no. 04014086.
- [5] T. Dawood, Z. Zhu, and T. Zayed, "Machine vision-based model for spalling detection and quantification in subway networks," *Autom. Construct.*, vol. 81, pp. 149–160, Sep. 2017.
- [6] S. German, I. Brilakis, and R. DesRoches, "Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments," *Adv. Eng. Informat.*, vol. 26, no. 4, pp. 846–858, Oct. 2012.
- [7] B. Valeti and S. Pakzad, "Automated detection of corrosion damage in power transmission lattice towers using image processing," in *Proc. Struct. Congr.*, Apr. 2017, pp. 474–482.
- [8] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyükoztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 731–747, Sep. 2018.
- [9] H. González-Jorge, I. Puente, B. Riveiro, J. Martínez-Sánchez, and P. Arias, "Automatic segmentation of road overpasses and detection of mortar efflorescence using mobile LiDAR data," *Opt. Laser Technol.*, vol. 54, pp. 353–361, Dec. 2013.
- [10] B. Kim and S. Cho, "Image-based concrete crack assessment using mask and region-based convolutional neural network," *Struct. Control Health Monitor.*, vol. 26, no. 8, Jun. 2019, Art. no. e2381.
- [11] S.-N. Yu, J.-H. Jang, and C.-S. Han, "Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel," *Autom. Construct.*, vol. 16, no. 3, pp. 255–261, May 2007.
- [12] H. Zhao, G. Qin, and X. Wang, "Improvement of canny algorithm based on pavement edge detection," in *Proc. 3rd Int. Congr. Image Signal Process.*, Oct. 2010, pp. 964–967.
- [13] R. Salim Lim, H. M. La, Z. Shan, and W. Sheng, "Developing a crack inspection robot for bridge maintenance," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 6288–6293.
- [14] R. S. Lim, H. M. La, and W. Sheng, "A robotic crack inspection and mapping system for bridge deck maintenance," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 367–378, Apr. 2014.
- [15] S. Chanda, G. Bu, H. Guan, J. Jo, U. Pal, Y.-C. Loo, and M. Blumenstein, "Automatic bridge crack detection—A texture analysis-based approach," in *Proc. Artif. Neural Netw. Pattern Recognit. (ANNPR)*, Montreal, QC, Canada, Oct. 2014, pp. 193–203.
- [16] R. Medina, J. Llamas, E. Zalama, and J. Gomez-Garcia-Bermejo, "Enhanced automatic detection of road surface cracks by combining 2D/3D image processing techniques," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 778–782.
- [17] P. Subirats, J. Dumoulin, V. Legeay, and D. Barba, "Automation of pavement surface crack detection using the continuous wavelet transform," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 3037–3040.
- [18] X. Wang and X. Feng, "Pavement distress detection and classification with automated image processing," in *Proc. Int. Conf. Transp., Mech., Electr. Eng. (TMEE)*, Dec. 2011, pp. 1345–1350.
- [19] S. Mathavan, K. Vaheesan, A. Kumar, C. Chandrakumar, K. Kamal, M. Rahman, and M. Stonecliffe-Jones, "Detection of pavement cracks using tiled fuzzy Hough transform," *J. Electron. Imag.*, vol. 26, no. 5, Sep. 2017, Art. no. 053008.
- [20] M. O'Byrne, B. Ghosh, F. Schoefs, and V. Pakrashi, "Regionally enhanced multiphase segmentation technique for damaged surfaces," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 29, no. 9, pp. 644–658, Oct. 2014.
- [21] W. Zhang, Z. Zhang, D. Qi, and Y. Liu, "Automatic crack detection and classification method for subway tunnel safety monitoring," *Sensors*, vol. 14, no. 10, pp. 19307–19328, Oct. 2014.
- [22] C. Feng, H. Zhang, H. Wang, S. Wang, and Y. Li, "Automatic pixel-level crack detection on dam surface using deep convolutional network," *Sensors*, vol. 20, no. 7, p. 2069, Apr. 2020.
- [23] Y.-J. Cha, W. Choi, and O. Büyükoztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017.
- [24] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, no. 10, p. 3452, Oct. 2018.
- [25] *ScrapeBox—The Swiss Army Knife of SEO!*. Accessed: Jun. 29, 2020. [Online]. Available: <http://www.scrapebox.com>
- [26] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [27] X. Zhan, Z. Liu, P. Luo, X. Tang, and C. C. Loy, "Mix-and-match tuning for self-supervised semantic segmentation," in *Proc. 32nd AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 1–9.
- [28] J. Lee, E. Kim, S. Lee, J. Lee, and S. Yoon, "FickleNet: Weakly and semi-supervised semantic image segmentation using stochastic inference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5267–5276.
- [29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [30] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," 2018, *arXiv:1802.07934*. [Online]. Available: <http://arxiv.org/abs/1802.07934>
- [31] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5688–5696.
- [32] G. Li, J. Wan, S. He, Q. Liu, and B. Ma, "Semi-supervised semantic segmentation using adversarial learning for pavement crack detection," *IEEE Access*, vol. 8, pp. 51446–51459, 2020.
- [33] Ç. F. Özgenel and A. G. Sogucu, "Performance comparison of pretrained convolutional neural networks on crack detection in buildings," in *Proc. 35th Int. Symp. Autom. Robot. Construct. (ISARC)*, Jul. 2018, pp. 1–8.
- [34] S. Dorafshan, R. J. Thomas, and M. Maguire, "SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks," *Data Brief*, vol. 21, pp. 1664–1668, Dec. 2018.
- [35] M. Marszalek and C. Schmid, "Accurate object localization with shape masks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [36] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [37] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4151–4160.
- [38] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.
- [39] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*. [Online]. Available: <http://arxiv.org/abs/1608.06993>
- [40] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Atlanta, GA, USA, vol. 30, no. 1, Jun. 2013, p. 3.
- [41] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, May 2010, pp. 249–256.
- [43] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, May 2015, pp. 1–15.



SEUNGBO SHIM was born in Donghae, South Korea, in 1982. He received the B.S. degree in electrical engineering from Pusan National University, Busan, South Korea, in 2009, and the M.S. degree in electronic and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 2011. He is currently pursuing the Ph.D. degree with the Korea Advanced Institute of Science and Technology. From 2010 to 2017, he was a Research Engineer with Samsung Heavy Industries. Since 2017, he has been a Research Specialist with the Korea Institute of Civil Engineering and Building Technology. His current research interests include machine vision, deep learning, automatic measurement systems, sensor networks, and infrastructure maintenance.



JIN KIM was born in Chuncheon, South Korea, in 1994. She received the B.S. degree in civil engineering from Korea University, Seoul, South Korea, in 2019. She is currently pursuing the M.S. degree with the Korea Advanced Institute of Science and Technology. Her current research interests include deep learning, image processing, and infrastructure maintenance.



GYE-CHUN CHO was born in South Korea, in 1970. He received the B.S. degree in civil engineering and the M.S. degree in geotechnical engineering from Korea University, Seoul, South Korea, in 1994 and 1996, respectively, and the Ph.D. degree in geotechnical engineering from the Georgia Institute of Technology, GA, USA, in 2001. He was an Assistant Professor with Pennsylvania State University, PA, USA, from 2001 to 2002, and the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, from 2002 to 2005, where he was an Associate Professor, from 2005 to 2012, and a Visiting Professor with the Georgia Institute of Technology, from 2009 to 2010. Since 2012, he has been a Professor with the Department of Civil and Environmental Engineering, KAIST. His current research interests include infrastructure maintenance, tunnel excavation, biopolymers, and soil behavior analysis.



SEONG-WON LEE was born in Seoul, South Korea, in 1967. He received the B.S. degree in mining and mineral engineering and the M.S. degree in geophysical exploration engineering from Inha University, Incheon, South Korea, in 1991 and 1993, respectively, and the Ph.D. degree in geotechnical engineering from Korea University, Seoul, in 2006. He has been working at the Korea Institute of Civil Engineering and Building Technology (KICT), since 1993. He was a Visiting Research Fellow with the Department of Mining Engineering, University of New South Wales, in 2008, and a Research Fellow with the Research and Development Planning Division, Korea Agency for Infrastructure Technology Advancement, from 2012 to 2014. He was the Director of the Geotechnical Research Division, KICT, from 2016 to 2018. His current research interests include infrastructure maintenance, tunnel excavation, underground construction measurement, and geotechnical exploration.

...