Automation in Construction 99 (2019) 52–58

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Autonomous concrete crack detection using deep fully convolutional neural network

Cao Vu Dung[a,*], Le Duc Anh[b]

[a] Advanced Research Laboratories, Tokyo City University, 8-15-1 Todoroki, Setagaya, Tokyo 158-0082, Japan
[b] NTT Hi-Tech Institute, Nguyen Tat Thanh University, 300A Nguyen Tat Thanh, Ward 13, District 4, Ho Chi Minh City, Vietnam

ARTICLE INFO

Keywords:
Concrete
Crack detection
Deep learning
Convolutional neural network
Semantic segmentation

ABSTRACT

Crack detection is a critical task in monitoring and inspection of civil engineering structures. Image classification and bounding box approaches have been proposed in existing vision-based automated concrete crack detection methods using deep convolutional neural networks. The current study proposes a crack detection method based on deep fully convolutional network (FCN) for semantic segmentation on concrete crack images. Performance of three different pre-trained network architectures, which serves as the FCN encoder's backbone, is evaluated for image classification on a public concrete crack dataset of 40,000 227 × 227 pixel images. Subsequently, the whole encoder-decoder FCN network with the VGG16-based encoder is trained end-to-end on a subset of 500 annotated 227 × 227-pixel crack-labeled images for semantic segmentation. The FCN network achieves about 90% in average precision. Images extracted from a video of a cyclic loading test on a concrete specimen are used to validate the proposed method for concrete crack detection. It was found that cracks are reasonably detected and crack density is also accurately evaluated.

1. Introduction

Infrastructures, such as bridges, roads, and dams have experienced accelerating deterioration due to environmental and loading effects. For instance, a large portion of bridges in Japan and the USA have been in service for more than 50 years [1,52]. Statistical analysis of bridge inspection data shows 46% of collapsed bridges were structurally deficient prior to the collapse, and age and structural deficiency, as well as structural deficiency and failure, are shown to be related [2]. The finding highlights the need for an efficient maintenance strategy for detecting early signs of structural weaknesses.

Digital technologies have been used to assist owners in appropriately planning monitoring and inspection activities. Koch et al. [3] reviewed the state-of-the-art computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure. In recent years, various vision-based methods have been proposed for concrete crack detection in civil infrastructures such as segmentation via Fuzzy C-means clustering [4], gray-scale histogram [5], V-shaped features [6], cascade features [7], spatial tuned-robust multifeatured (STRUM) [8] and spectral analysis [9]. Machine vision has also been used for structural load estimation from the surface crack patterns in reinforced concrete beams and slabs [10,11] and correlating crack patterns visually acquired using a multifractal analysis to structural integrity [12,13].

Recently, deep convolutional neural networks (CNN) have been developed for image classification and object detection in computer vision [14]. Inspired by such achievements, several recent studies have developed CNN-based algorithms for automated crack detection for road pavement, concrete structures, and nuclear power plants. Chen and Jahanshahi [15] proposed a detection method using CNN to analyse individual video frames for crack detection in combination with a Naïve Bayes data fusion scheme to aggregate the information extracted from each video frame to enhance the overall performance of the detection system. Fan et al. [16] trained a CNN model as a multi-label classifier for pavement crack detection. In this study, a strategy with modifying the ratio of positive to negative samples is proposed. Wang et al. [17] suggested a CNN consisting of three convolutional layers and two fully-connected layers with 1,246,240 parameters in total to recognize cracks on asphalt surfaces at subdivided image cells. Zhang et al. [18] proposed an automatic road crack detection method based on deep CNN trained to classify 99 × 99 × 3 image patches acquired by a low-cost smartphone sensor. Pauly et al. [19] created 99 × 99 × 3 image classifiers to demonstrate the effectiveness of using deeper CNN in improving the accuracy of pavement crack detection. Cha et al. [20] & Cha and Choi [21] proposed vision-based methods for detecting concrete crack using a 256 × 256 × 3 CNN classifier in combination

* Corresponding author.
E-mail address: caovu@tcu.ac.jp (C.V. Dung).

https://doi.org/10.1016/j.autcon.2018.11.028
Received 6 July 2018; Received in revised form 28 October 2018; Accepted 29 November 2018

with a sliding window technique. Cha et al. [22] proposed a damage detection method based on the faster region-based CNN (Faster R-CNN) [23] to provide quasi-real-time simultaneous detection of concrete cracks, steel corrosion, bolt corrosion, and steel delamination.

Using a pre-trained network, a saved network that is previously trained on a large dataset, typically on a large-scale image classification task, is a common and highly effective approach to deep learning on small datasets. CNN architectures such as AlexNet [24], VGG16 [25], InceptionV3 [26], and Resnet50 [27] were typically trained on a large image dataset, e.g., ImageNet [28], and have achieved the state-of-the-art results for general image classification. The spatial hierarchy of features learned by a deep CNN pre-trained on a large and general original dataset such as the ImageNet proves to be useful for many different computer vision problems, even for a new problem involving completely different classes than those of the original dataset from which those features were learned [29]. Gopalakrishan et al. [30] proposed different image classifiers by transferring the features learned from the ImageNet to detect cracks in Hot-Mix Asphalt (HMA) and Portland Cement Concrete (PCC). Maeda et al. [31] proposed a road damage detection method using the SSD Inception V2 and SSD MobileNet models to detect and classify eight different image types using the bounding box concept. Transfer learning has been proved to improve the efficiency and accuracy of a crack classifier [30,32].

Most of the previous studies have proposed methods based on image classification and/or object detection using bounding box. However, cracks typically appear as thin dark lines or strips with constantly varying angle and direction. Although the bounding box-based methods can detect cracks reasonably well, they do not provide precise information of crack path and density. Therefore, a pixel-based classification which distinguishes “crack” and “non-crack” pixels is desirable. In the present study, semantic segmentation is employed to obtain more precise information of crack path and density for accurate crack detection.

Inspired by the success of CNN for image classification task, researchers have adapted CNN to semantic segmentation as feature extraction. Several initial approaches employed CNN to classify pixels in an image by the sliding-window technique [33,34]. Before the deep learning era, traditional methods rely on the classification of pixels and superpixels, i.e., a group of connected pixels with similar colors or gray levels. Fulkerson et al. [35] proposed a method based on aggregating histograms in the neighborhood of each superpixel and then refined the result by a conditional random field on the superpixel graph. Russell et al. [36] proposed a hierarchical random field model, which integrates features from different levels of the quantization hierarchy. Arbelaez et al. [37] suggested an image segmentation method based on contour detection. A hierarchy graph is constructed to merge contours by similarity and position. Recently, fully convolutional networks (FCN) have been proposed for semantic segmentation. The FCN method employs deep learning for training an end-to-end pipeline [38–40]. FCN is combined with conditional random fields as post-processing to improve segmentation performance [41,42]. The FCN method in conjunction with residual networks currently achieves the state-of-the-art performance on semantic segmentation tasks. FCN for semantic segmentation has been applied to solve challenging problems in multi-disciplinary domains such as road detection to assist autonomous (self-driving) cars [43] and mapping the solar photovoltaic arrays in aerial imagery [44].

The present study proposes a FCN-based method for concrete crack detection. First, the performance of different pre-trained deep CNN architectures for image classification on a public concrete crack dataset is evaluated in order to select the best-performing architecture for the FCN encoder. The whole FCN network is then trained end-to-end for semantic segmentation on a subset of annotated crack images of the same dataset. Finally, the performance of the proposed method is verified using images extracted from a video of actual concrete crack opening under a cyclic loading test.

2. Experimental study

2.1. Methodology

An encoder-decoder FCN is trained end-to-end for the task of segmenting an image of concrete crack into “crack” and “non-crack” pixels for both crack detection and crack density evaluation. First, experiments are conducted to evaluate the performance of different pre-trained CNNs for the classification task on a public concrete crack image dataset. The selected pre-trained model will be used as the backbone of the FCN encoder. Next, the FCN is trained on a subset of the same dataset containing annotated crack images for the segmentation task.

2.1.1. Pre-trained convolutional neural networks for crack image classification

A CNN architecture typically consists of several convolutional blocks and a fully connected layer. Each convolutional block is composed of a convolutional layer, an activation unit, and a pooling layer. A convolutional layer performs convolution operation over the output of the preceding layers using a set of filters or kernels to extract the features that are important for classification. For example, LeNet-5 which is an early network architecture proposed for handwritten digit classification [45] has two convolutional blocks. Recently, “deeper” CNN architectures such as AlexNet, VGG16, Inception, and ResNet have improved the prior-art configurations by increasing the number of weight layers. Most previous studies have proposed crack detection methods using CNNs trained from scratch for classification. However, transfer learning has also been proved to enhance training efficiency and accuracy of a crack classifier. In the present study, three different pre-trained CNN models including VGG16, Inception, and ResNet are experimented to evaluate their performance as the encoder of the FCN. VGG16 is a widely used convolutional architecture pre-trained on ImageNet [25] whereas ResNet, having residuals nets with 152-layer depth which is eight times deeper than VGG models but with lower complexity, won the 1st place in classification competitions including ILSVRC 2015 and ILSVRC & COCO 2015 [27]. Meanwhile, the Inception V3 network benchmarked on the ILSVRC 2012 classification challenge validation set demonstrate substaintial gains over the state of the art [26]. Network parameter configuration of the VGG16, InceptionV3, and ResNet are illustrated in Table 1.

For transfer learning, a pre-trained model is first loaded. Only the convolutional part of the model up to the fully connected (FC) layers (i.e., the top FC layers are excluded) is initiated before running this model on the training and validation image data only once and saving the output of the last layer before the FC layer, i.e., the output features. Then a customized FC layer is trained on top of these output features. The output of the last convolutional layer is flattened and connected to the ReLU-activated units of the FC layer. The output layer consists of a single unit with sigmoid activation, which is frequently used for a binary classification network, because the sigmoid function mostly outputs a value close to 0 or 1 indicating “crack” or “non-crack”, respectively. The model is compiled using the binary cross-entropy loss. The rmsprop method is used as the optimizer. Training the crack classifiers is performed using Keras framework with TensorFlow backend, an open-source deep learning framework [29], on an Intel^{R} Xeon^{R} E5-2620V4 2@2.10 MHz-Processor CPU.

Table 1
Number of parameters of pre-trained networks.

Network	No. of parameters
VGG16	200,721
InceptionV3	409,617
ResNet	16,401

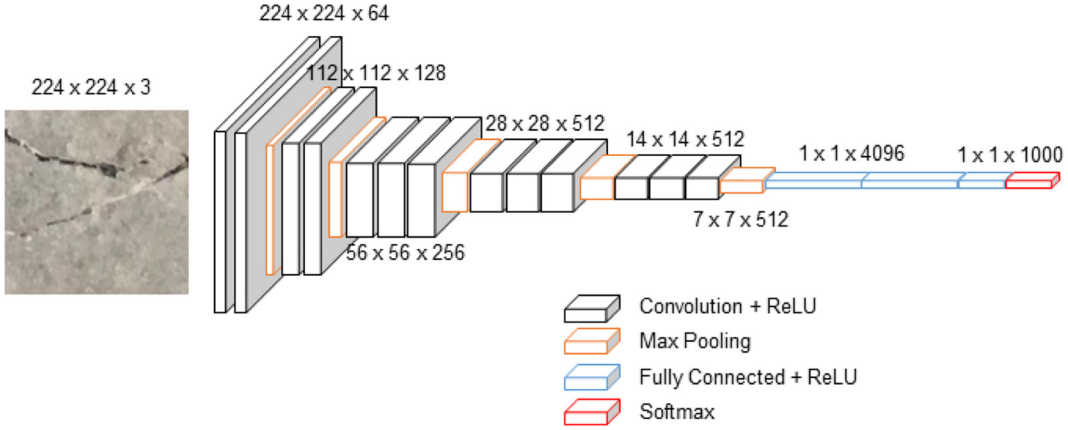


Fig. 1. Original network architecture of VGG16 for image classification.

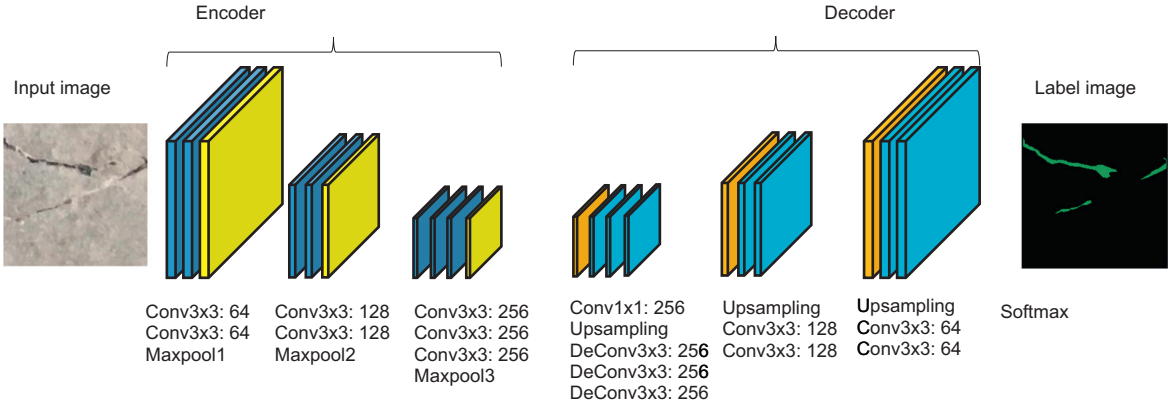


Fig. 2. Network architecture of FCN for semantic segmentation.

Table 2
Concrete crack image dataset.

Task	No. of images	Size (pixels)	Crack	Non-crack	Train set	Dev set	Test
Classification	40,000	227 × 227	20,000	20,000	32,000	4000	4000
Segmentation	600	227 × 227	600	0	400	100	100

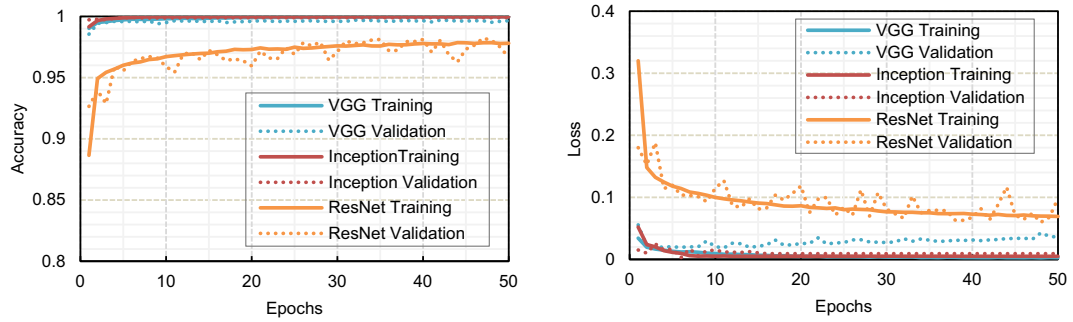
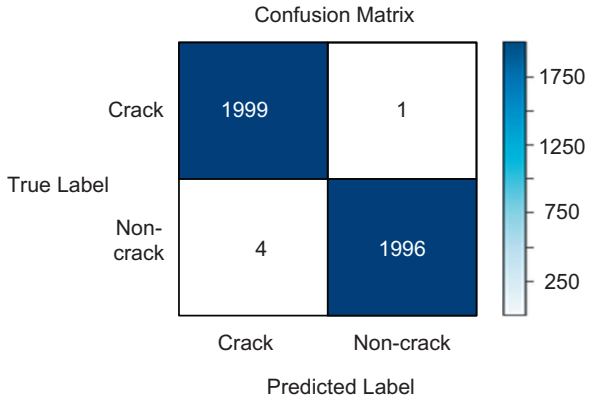


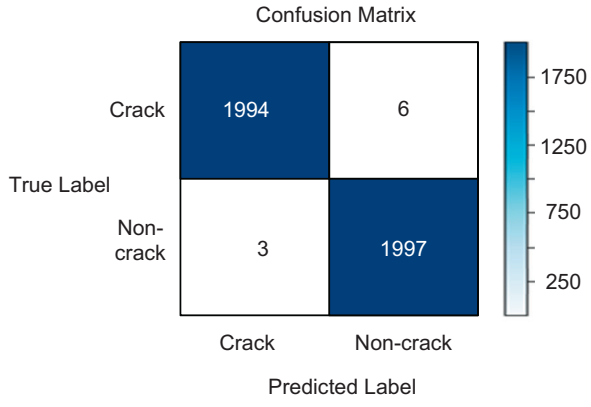
Fig. 3. Accuracy (left) and loss (right) during training and validation.

2.1.2. Fully convolutional network for semantic segmentation

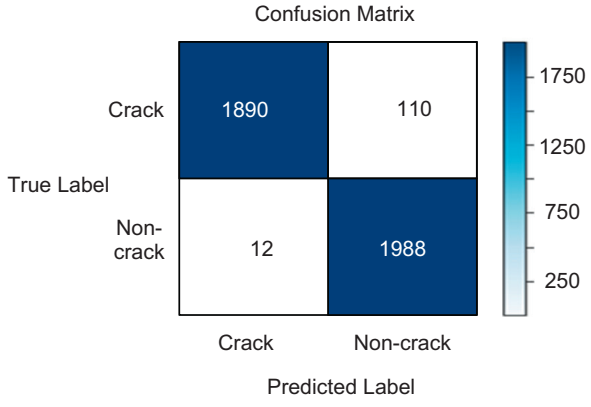
The KittiSeg network for road detection directly inspires the architecture of the FCN used in the present study [43]. The FCN model contains an encoder and a decoder (Fig. 2). The task of the encoder is to process an input image and extract features necessary for semantic segmentation. The encoder includes all the convolutional and pooling layers but discards the FC and softmax layers of VGG16 (Fig. 1). The weights of VGG16 pre-trained on ImageNet dataset is used for initialization [46]. The decoder uses deconvolution and upsampling layers to reconstruct the corresponding segmented image. Given the features created by the encoder, a 1×1 convolutional layer is employed to create a low-resolution segmentation. Then, the output is upsampled by the deconvolutional layers to extract high-resolution features. Each deconvolutional layer of the decoder is paired with a corresponding convolutional layer in the encoder. The upsampling layers use the max-pooling index from its corresponding layer in the encoder to construct the expanded feature map. This process creates a larger feature map from the output of the previous layer. The last layer



(a) VGG16



(b) InceptionV3



(c) ResNet

Fig. 4. Confusion matrices.

is a softmax layer used to classify each pixel into “crack” or “non-crack” classes.

2.2. Concrete crack image dataset

The open-sourced dataset of concrete crack images collected at various campus buildings of Middle East Technical University [47] is used for classification and segmentation (Table 2). The total 40,000 images of 227×227 pixels generated from 458 full photos of 4032×3024 pixels using the method proposed by Zhang et al. [48] were equally divided into “crack” and “non-crack” classes for the classification task. The full images have a high surface finish and

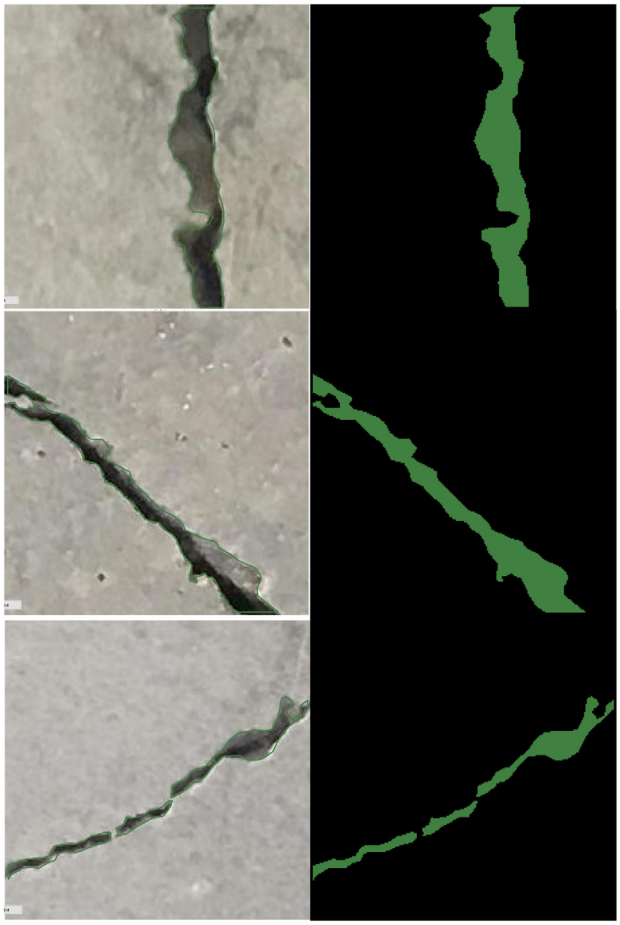


Fig. 5. Example of annotated concrete crack images.

Table 3
Performance of FCN segmentation method.

Accuracy metrics	Training	Validation	Test
Max F1 (%)	91.9	89.6	89.3
Average Precision (AP) (%)	90.9	89.9	89.3

illumination condition variance. Data augmentation using random rotation and flipping was not applied. For segmentation, 600 crack images are randomly selected from the total 20,000 crack-labeled images of the dataset and annotated using the lightweight MATLAB^R tool LIBLABEL created by Geiger et al. [49] for semantic/instance annotation (Fig. 5). The training, validation, and test sets for the segmentation task contain 400, 100, and 100 images, respectively.

3. Results and discussions

3.1. Crack image classification using different pre-trained networks

The classifiers are trained for 50 epochs with a batch size of 16. During training, the VGG16 and InceptionV3-based classifiers achieve almost 0.999 while the ResNet-based classifier produces the maximum of 0.975 (Fig. 3). This trend is also observed during testing. The VGG16 and Inception V3-based classifiers achieve almost perfect classification scores with only 6 or less false positives (FP) or false negatives (FN) over the total 4000 test images of the test set (Fig. 4a,b). Meanwhile, the ResNet-based classifier achieves lower accuracy than the two other classifiers, at about 110 FPs and 12 FNs (Fig. 4c).

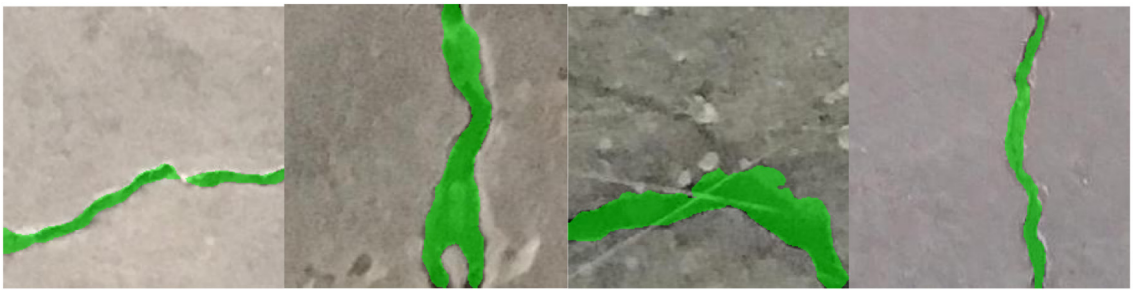


Fig. 6. Examples of segmentation results for test images.

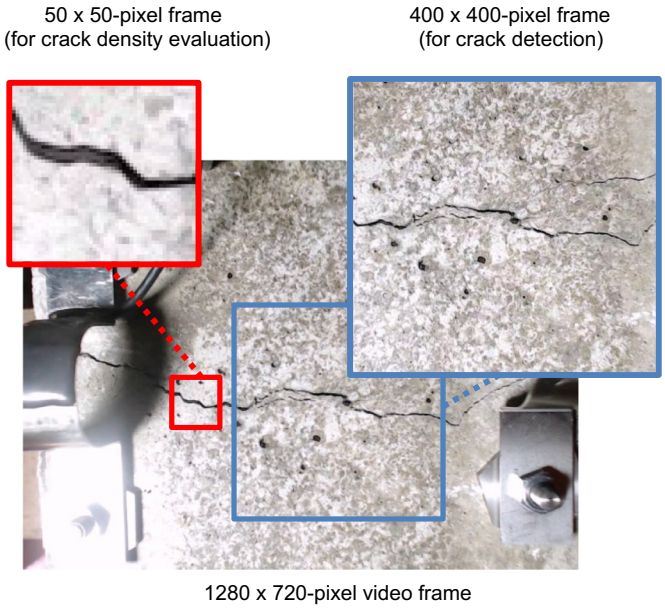


Fig. 7. Video frame of cyclic crack opening test.

3.2. End-to-end training of fully convolutional network

The weights of all the layers of the FCN encoder are initialized using the pre-trained weights of VGG16. Next, the final FC layers are removed and replaced by the decoder. Finally, the encoder-decoder FCN network is trained end-to-end using the 500 images of the train and validation data annotated for the segmentation task. The softmax cross-entropy loss function is employed as the objective function to train the

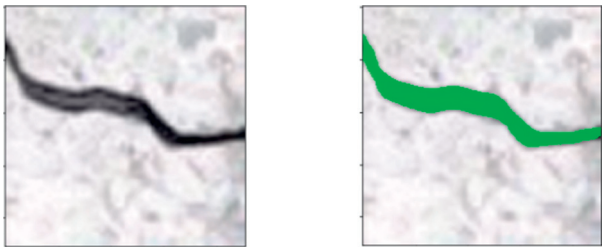


Fig. 9. Segmentation results for 50 × 50-pixel frame. (a) Original image; (b) Segmented image.

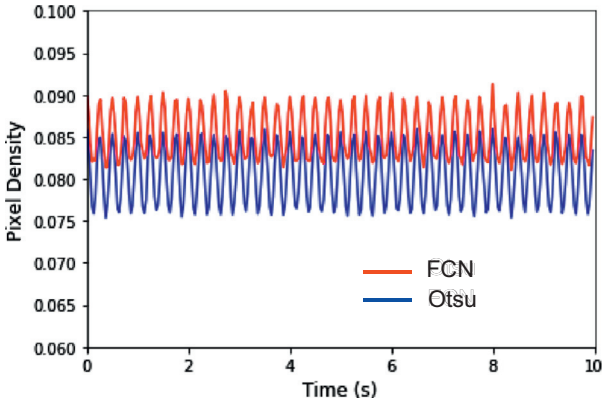


Fig. 10. Crack pixel density evaluation for 50 × 50-pixel frame.

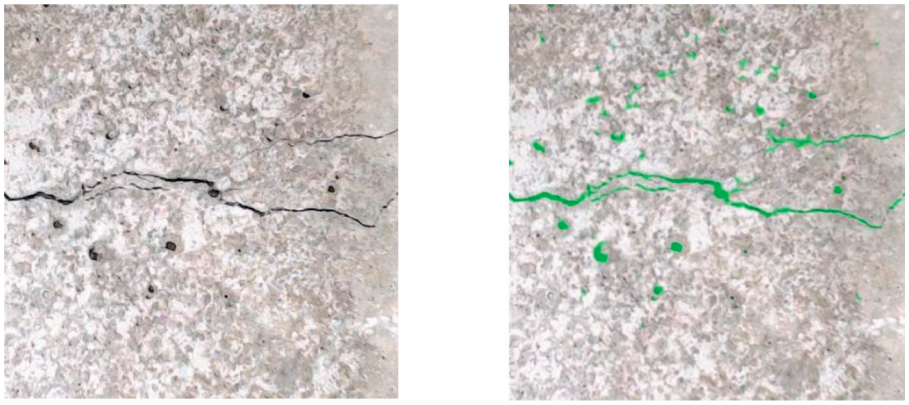


Fig. 8. Crack detection by segmentation method for 400 × 400-pixel frame.

segmentation network. The Adam optimizer with a learning rate of 10^{-5} and a weight decay of 5×10^{-4} is used. For evaluating semantic segmentation, pixel-based metrics including the maximum F1 score (Max F1) and averaged precision (AP) are used. The F1 score is defined as the harmonic mean of precision and recall by the following equation:

$$F1 = \frac{2^{*}Precision^{*}Recall}{Precision + Recall} \quad (1)$$

For classifying pixels, the classification threshold t is chosen to maximize the F1 score:

$$Max\ F1 = argmaxF1(t) \quad (2)$$

To provide insights into the performance over the full recall range, the AP metric similar to that used in PASCAL VOC challenges [50] is computed for different recall values of "crack" and "non-crack" classes:

$$AP = \frac{1}{2}\sum_{r=0}^{1} \underset{r':r'>r}{max}\ Precision(r') \quad (3)$$

The semantic segmentation algorithm achieves the max F1 and AP of approximately 90% on training, validation, and testing sets (Table 3) at the speed of about 13.8 frames per second (fps) on GPU Nvidia GeForce GTX 1070. Fig. 6 shows the prediction results by the segmentation network on the test set. Detailed information of the crack properties including crack path and density seems to be accurately predicted by the FCN segmentation method proposed in the present study.

3.3. Fully convolutional network's performance verification

The performance of the proposed semantic segmentation method is verified for both crack detection and crack density evaluation tasks. A ten-second video captured during a crack opening test on a concrete specimen subjected to a cyclic loading at 4 Hz are is used for verification (Fig. 7). The video was taken at 30 fps. For crack detection, the segmentation method was applied to detect cracks in a video frame captured during the cyclic test (Fig. 8a). The 400 × 400 cropped image is splitted into 8 × 8 cells of 50 × 50 pixels, and the segmentation algorithm is applied separately to each cell. Since the FCN contains symmetric layers of convolution/deconvolution and max-pooling/upscale, it can process images in different sizes. The network processes an input image and produces an output image that has the same size. The prediction results show that all the cracks in the test image were detected with an accuracy almost equal to that obtained during training. However, the segmentation algorithm still makes several minor false predictions for the dark dot-like features induced by the concrete surface's imperfection (Fig. 8b).

In a further attempt to evaluate crack density, the proposed segmentation algorithm is applied to 300 cropped video frames of 50 × 50 pixels (Fig. 7). Crack density is evaluated using pixel density which is defined as the ratio of the pixels predicted as "crack" over the total number of pixels of the cropped frame. The standard Otsu's thresholding technique [51] is employed to obtain the "ground truth" to avoid introducing bias and error associated with manual segmentation at pixel level. There is about 5.8% difference in average density between the FCN method proposed in the present study and the Otsu method (Figs. 9 and 10).

For the simple background of the 50 × 50 cropped frame in the above evaluation, both the Otsu thresholding method and the proposed segmentation method appears to perform well. However, for a more complicated background, the proposed method can still work reasonably well because the segmentation network has learned to automatically distinguish "crack" and "non-crack" pixels from training images. The segmentation network utilizes both information of the pixel and its neighbor pixels for distinguishing "crack" and "non-crack" pixels. Binarization methods such as the Otsu thresholding method may not work well because a single threshold is determined for classifying "crack" and "non-crack" pixels.

4. Conclusions

In the present study, a vision-based method for concrete crack detection and density evaluation using FCN is proposed. The backbone of the FCN encoder was selected as VGG16, which performed better than InceptionV3 and ResNet for crack image classification. Next, the whole encoder-decoder FCN architecture was trained end-to-end on a subset of crack images of the same dataset and reached approximately 90% for both the max F1 and AP scores on training, validation, and test sets. For verification, the proposed method was employed to detect and evaluate crack density in a video of crack opening. The crack path could be accurately identified using the FCN-based segmentation method. Furthermore, the crack density variation was also accurately captured using the pixel density ratio. Therefore, applications of the proposed FCN method for crack segmentation in structural health monitoring for concrete structures should be further investigated. Although the proposed method has reasonably captured crack path, it is still challenging to quantify crack size autonomously, especially when a test image has many noisy crack-like features. Therefore, future studies should focus on how to improve the proposed method to make autonomous crack density evaluation more robust.

Acknowledgements

The authors would like to thank Prof. Takuyo Konishi at the Advanced Research Laboratories, Tokyo City University for kindly providing the verification data. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] ASCE, American Society of Civil Engineers (ASCE), Infrastructure Report Card, 2017, https://www.infrastructurereportcard.org/, (2017) (Last access: 26 Oct 2018).
[2] W. Cook, P.J. Barr, Observations and trends among collapsed bridges in New York state, J. Perform. Constr. Facil. 31 (4) (2017) 04017011, https://doi.org/10.1061/(ASCE)CF.1943-5509.0000996.
[3] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure, Adv. Eng. Inform. 29 (2) (2015) 196–210, https://doi.org/10.1016/j.aei.2015.01.008.
[4] Y. Noh, D. Koo, Y.M. Kang, D. Park, D. Lee, Automatic crack detection on concrete images using segmentation via fuzzy C-means clustering, in: May (Ed.), Applied System Innovation (ICASI), 2017 International Conference on, IEEE, 2017, pp. 877–880, , https://doi.org/10.1109/ICASI.2017.7988574.
[5] T.H. Dinh, Q.P. Ha, H.M. La, Computer vision-based method for concrete crack detection, in: November (Ed.), 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE, 2016, pp. 1–6, , https://doi.org/10.1109/ICARCV.2016.7838682.
[6] Y. Sato, Y. Bao, Y. Koya, Crack detection on concrete surfaces using V-shaped features, World Comp. Sci. Inform. Technol. J. 8 (1) (2018).
[7] R. Ali, D.L. Gopal, Y.J. Cha, Vision-based concrete crack detection technique using cascade features, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, Vol. 10598 International Society for Optics and Photonics, 2018, March, p. 105980L, , https://doi.org/10.1117/12.2295962.
[8] P. Prasanna, K.J. Dana, N. Gucunski, B.B. Basily, H.M. La, R.S. Lim, H. Parvardeh, Automated crack detection on concrete bridges, IEEE Trans. Autom. Sci. Eng. 13 (2) (2016) 591–599, https://doi.org/10.1109/TASE.2014.2354314.
[9] R.S. Adhikari, O. Moselhi, A. Bagchi, Image-based retrieval of concrete crack properties for bridge inspection, Autom. Constr. 39 (2014) 180–194, https://doi.org/10.1016/j.autcon.2013.06.011.
[10] R. Davoudi, G.R. Miller, J.N. Kutz, Computer vision based inspection approach to predict damage state and load level for RC members, Structural Health Monitoring 2017, (shm), 2017, https://doi.org/10.12783/shm2017/14225.
[11] R. Davoudi, G.R. Miller, J.N. Kutz, Structural load estimation using machine vision and surface crack patterns for shear-critical RC beams and slabs, J. Comput. Civ. Eng. 32 (4) (2018) 04018024, , https://doi.org/10.1061/(ASCE)CP.1943-5487.0000766.
[12] A. Ebrahimkhanlou, A. Farhidzadeh, S. Salamone, Multifractal analysis of two-dimensional images for damage assessment of reinforced concrete structures, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2015, Vol. 9435 International Society for Optics and Photonics, 2015, March, p.

94351A, , https://doi.org/10.1117/12.2084052.
[13] A. Ebrahimkhanlou, A. Farhidzadeh, S. Salamone, Multifractal analysis of crack patterns in reinforced concrete shear walls, Struct. Health Monit. 15 (1) (2016) 81–92, https://doi.org/10.1177/1475921715624502.
[14] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: a comprehensive review, Neural Comput. 29 (9) (2017) 2352–2449, https://doi.org/10.1162/neco_a_00990.
[15] F.C. Chen, M.R. Jahanshahi, NB-CNN: deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion, IEEE Trans. Ind. Electron. (2017), https://doi.org/10.1109/TIE.2017.2764844.
[16] Z. Fan, Y. Wu, J. Lu, W. Li, Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network, (2018) (arXiv preprint arXiv:1802.02208).
[17] K.C. Wang, A. Zhang, J.Q. Li, Y. Fei, C. Chen, B. Li, Deep learning for asphalt pavement cracking recognition using convolutional neural network, Airfield and Highway Pavements 2017, 2017, pp. 166–177, , https://doi.org/10.1061/9780784480922.
[18] L. Zhang, F. Yang, Y.D. Zhang, Y.J. Zhu, Road crack detection using deep convolutional neural network, Image Processing (ICIP), 2016 IEEE International Conference on, IEEE, 2016, September, pp. 3708–3712, , https://doi.org/10.1109/ICIP.2016.7533052.
[19] L. Pauly, D. Hogg, R. Fuentes, H. Peel, Deeper networks for pavement crack detection, Proceedings of the 34th ISARC. 34th International Symposium in Automation and Robotics in Construction, 28 Jun - 01 Jul 2017, IAARC, Taipei, Taiwan, 2017, July, pp. 479–485.
[20] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Comput. Aided Civ. Inf. Eng. 32 (5) (2017) 361–378, https://doi.org/10.1111/mice.12263.
[21] Y.J. Cha, W. Choi, Vision-based concrete crack detection using a convolutional neural network, Dynamics of Civil Structures, Vol. 2 Springer, Cham, 2017, pp. 71–73, , https://doi.org/10.1007/978-3-319-54777-0_9.
[22] Y.J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, O. Büyüköztürk, Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, Comput. Aided Civ. Inf. Eng. (2017), https://doi.org/10.1111/mice.12334.
[23] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031.
[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Proces. Syst. (2012) 1097–1105 http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (Last access: 26 Oct 2018).
[25] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks For Large-scale Image Recognition, (2014) (arXiv preprint arXiv:1409.1556).
[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 2818–2826 https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf (Last access: 26 Oct 2018).
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 770–778 https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html (Last access: 26 Oct 2018).
[28] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, June, pp. 248–255, , https://doi.org/10.1109/CVPR.2009.5206848.
[29] F. Chollet, Keras. Github, https://github.com/fchollet, (2018) Last access: 26 Oct 2018).
[30] K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, A. Agrawal, Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection, Constr. Build. Mater. 157 (2017) 322–330, https://doi.org/10.1016/j.conbuildmat.2017.09.110.
[31] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone, (2018) (arXiv preprint arXiv:1801.09454).
[32] K. Zhang, H.D. Cheng, B. Zhang, Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning, J. Comput. Civ. Eng. 32 (2) (2018) 04018001, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000736.
[33] A. Giusti, D.C. Ciresan, J. Masci, L.M. Gambardella, J. Schmidhuber, Fast image scanning with deep max-pooling convolutional neural networks, in: September (Ed.), Image Processing (ICIP), 2013 20th IEEE International Conference on, IEEE, 2013, pp. 4034–4038, , https://doi.org/10.1109/ICIP.2013.6738831.
[34] H. Li, R. Zhao, X. Wang, Highly Efficient Forward and Backward Propagation of Convolutional Neural Networks For Pixelwise Classification, (2014) (arXiv preprint arXiv:1412.4526).
[35] B. Fulkerson, A. Vedaldi, S. Soatto, Class segmentation and object localization with superpixel neighborhoods, Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, September, pp. 670–677, , https://doi.org/10.1109/ICCV.2009.5459175.
[36] C. Russell, P. Kohli, P.H. Torr, Associative hierarchical crfs for object class image segmentation, in: September (Ed.), Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 739–746, , https://doi.org/10.1109/ICCV.2009.5459248.
[37] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 898–916, https://doi.org/10.1109/TPAMI.2010.161.
[38] V. Badrinarayanan, A. Kendall, SegNet, R. C, A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, arXiv preprint arXiv:1511.00561 (2015), https://doi.org/10.1109/TPAMI.2016.2644615.
[39] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2015) 3431–3440 https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html (Last access: 26 Oct 2018).
[40] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, Cham, 2015, October, pp. 234–241, , https://doi.org/10.1007/978-3-319-24574-4_28.
[41] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, https://doi.org/10.1109/TPAMI.2017.2699184.
[42] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, ... P.H. Torr, Conditional random fields as recurrent neural networks, in: December (Ed.), Computer Vision (ICCV), 2015 IEEE International Conference on, IEEE, 2015, pp. 1529–1537, , https://doi.org/10.1109/ICCV.2015.179.
[43] M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, R. Urtasun, Multinet: real-time joint semantic reasoning for autonomous driving, 2018 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2018, June, pp. 1013–1020, , https://doi.org/10.1109/IVS.2018.8500504.
[44] J. Camilo, R. Wang, L.M. Collins, K. Bradbury, J.M. Malof, Application of a Semantic Segmentation Convolutional Neural Network for Accurate Automatic Detection and Mapping of Solar Photovoltaic Arrays in Aerial Imagery, (2018) (arXiv preprint arXiv:1801.04018).
[45] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436, https://doi.org/10.1038/nature14539.
[46] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, ... A.C. Berg, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[47] Çağlar Fırat Özgenel, "Concrete Crack Images for Classification", Mendeley Data, v1, (2018), https://doi.org/10.17632/5y9wdsg2zt.1.
[48] H. Zhang, J. Tan, L. Liu, Q.J. Wu, Y. Wang, L. Jie, Automatic crack inspection for concrete bridge bottom surfaces based on machine vision, Chinese Automation Congress (CAC), 2017, IEEE, 2017, October, pp. 4938–4943, , https://doi.org/10.1109/CAC.2017.8243654.
[49] A. Geiger, M. Lauer, C. Wojek, C. Stiller, R. Urtasun, 3d traffic scene understanding from movable platforms, IEEE Trans. Pattern Anal. Mach. Intell. 36 (5) (2014) 1012–1025, https://doi.org/10.1109/TPAMI.2013.185.
[50] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338, https://doi.org/10.1007/s11263-009-0275-4.
[51] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66 https://pdfs.semanticscholar.org/fa29/610048ae3f0ec13810979d0f27ad6971bdbf.pdf (Last access: 26 Oct 2018).
[52] Road Bureau, Ministry of Land, Infrastructure, Transportation, and Tourism, Roads in Japan, Retrieved from: http://www.mlit.go.jp/road/road_e/index_e.html.