Performance evaluation of an improved deep CNN-based concrete crack detection algorithm

Sanjeetha Pennada[a], Marcus Perry[a], Jack McAlorum[a], Hamish Dow[a], Gordon Dobie[b]
[a]Dept. of Civil and Environmental Engineering, University of Strathclyde, Glasgow, G1 1XJ, UK.
[b]Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XJ, UK.

ABSTRACT

This study uses a novel directional lighting approach to produce a computationally efficient five-channel Visual Geometry Group-16 (VGG-16) convolutional neural network (CNN) model for concrete crack detection and classification in low-light environments. The first convolutional layer of the proposed model copies the weights for the first three channels from the pre-trained model. In contrast, the additional two channels are set to the average of the existing weights along the channels. The model employs transfer learning and fine-tuning approaches to enhance accuracy and efficiency. It utilizes variations in patterned lighting to produce five channels. Each channel represents a grayscale version of the images captured using directed lighting in the right, below, left, above, and diffused directions, respectively. The model is evaluated on concrete crack samples with crack widths ranging from 0.07 mm to 0.3 mm. The modified five-channel VGG-16 model outperformed the traditional three-channel model, showing improvements ranging from 6.5 to 11.7 percent in true positive rate, false positive rate, precision, F1 score, accuracy, and Matthew's correlation coefficient. These performance improvements are achieved with no significant change in evaluation time. This study provides useful information for constructing custom CNN models for civil engineering problems. Furthermore, it introduces a novel technique to identify cracks in concrete buildings using directed illumination in low-light conditions.

Keywords: Multi-channel neural network, fine-tuning, convolutional neural networks, transfer learning, crack detection, directional lighting, deep learning, structural health monitoring.

1. INTRODUCTION

Structural Health Monitoring (SHM) is a technique, used to ensure the safety and longevity of aerospace, mechanical, and civil engineering infrastructures.[1] In civil engineering, continuous monitoring of the structures is essential in order to identify defects like cracks at an early stage.[2] Non-Destructive Testing (NDT) techniques identify the state of the concrete structures without causing any permanent damage.[3] The use of technology, such as robots and cameras (visual inspection), to capture images[4] is an effective non-destructive evaluation (NDE) approach for automatic detection of cracks in concrete structures;[5] also it is faster and more accurate than the traditional manual inspection.[5,6] This led to the development of two image defect identification systems: white-box (edge detectors and thresholding[7]) and black-box (machine learning, artificial neural networks[8]) techniques. Black-box methods are considered better for crack identification in concrete structures.[9] The deep convolutional neural network (CNN) architecture, Visual Geometry Group (VGG) has been proven useful in detecting and classifying cracks in concrete structures.[10]

Existing research on automated crack detection in low-light environments has focused solely on diffused light sources to highlight the scenes captured in images.[11] However, the CNN technique is also developed in other experiments, allowing it to detect defects even under difficult light conditions.[12] This study aimed to automate the process of concrete crack detection in low-light environments, with the goal of replacing the human inspectors. To achieve this objective, a customized five-channel VGG-16 CNN model is implemented using a robot equipped with a machine-vision camera that captures images by directing light onto a concrete surface from multiple angles and directions, thereby improving the models accuracy. The model uses transfer learning and fine-tuning[13] approaches. Compared to the standard three-channel method, this technique greatly enhances crack identification and is a potential option for automated crack detection in real-world applications.

Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2023,
edited by Zhongqing Su, Branko Glisic, Maria Pina Limongelli, Proc. of SPIE Vol.
12486, 1248615 · © 2023 SPIE · 0277-786X · doi: 10.1117/12.2657723

2. RESEARCH DESIGN

The proposed method involves capturing images of the concrete slab under five lighting conditions: right, below, left, above and diffused. These images are cropped into a several blocks equal to the standard size of 224×224 pixels and further combined block-wise to form a single five-channel TIFF image, where each channel represents an image taken under a specific lighting direction. The pre-trained VGG-16 model is improved by utilizing transfer learning and fine-tuning methodologies to handle multi-channel input and detect cracks. Each stage of the approach is discussed in detail in the relevant sections.

2.1 Image Acquisition

The proposed model employs a unique inspection device comprising a camera and four angle-adjustable arms fitted with LED lighting strips. The inspection process is conducted in a dark environment. The LED lighting strips are attached to the angle-adjustable arms positioned at 50 degrees relative to the concrete surface to increase crack visibility by magnifying the shadows they cast. In this study, four directional lighting images are obtained by projecting light from the right (R), below (B), left (L), and above (A) onto the scene. In addition, using diffused (D) lighting, a fifth image is also captured. These images serve as the basis for the proposed crack detection method. A visual representation of these images is shown in Figure 1.

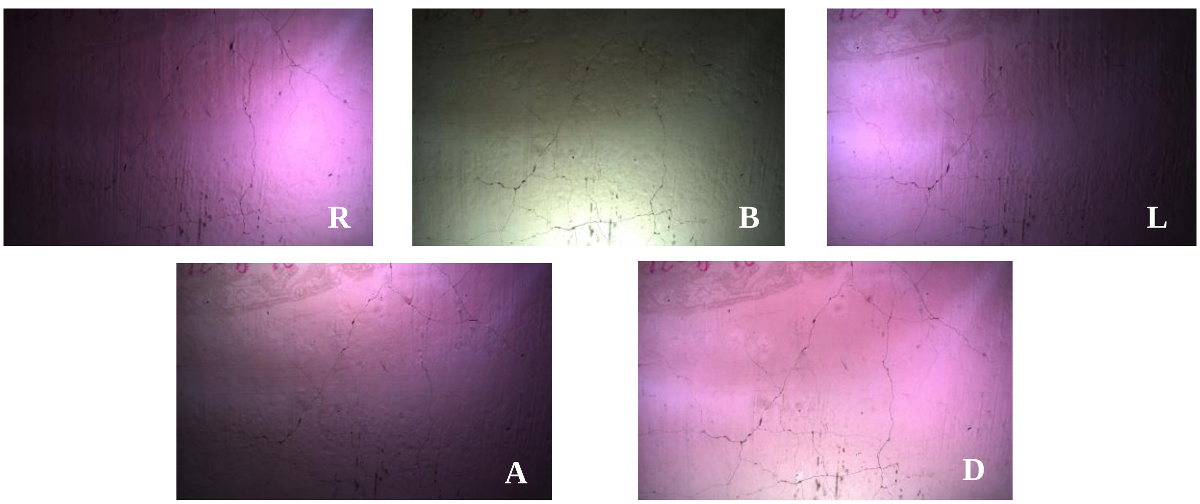

Figure 1. Captured concrete slab areas with lighting from Right (R), Below (B), Left (L), Above (A) and Diffused (D) directions, respectively.

2.2 Image Preprocessing

The VGG-16 model accepts an input image of size 224×224 pixels.[14] Therefore, to ensure that the images obtained during the crack detection process are compatible with the VGG-16 model, the large image in .jpg format (3458×5429 pixels) undergoes a preprocessing step where the images are cropped into several blocks equal to the standard input size of 224×224 pixels in .jpg format, before being used as an input to the proposed model.

2.3 Formation of Five-Channel Tag Image File Format (TIFF) image

The VGG-16 network expects an input image of size (224, 224, 3), where the three channels are red, green, and blue, respectively.[14] This specific input size works optimally with the VGG-16 architecture. Tag Image File Format (TIFF) is a popular file format for storing high-quality, high-resolution photos or images with a deep color depth.[15] The proposed deep learning model uses a five-channel TIFF image as an input, which is obtained by combining the grayscale versions of the original images captured in five different lighting conditions (R, B, L, A, and D). The final image format is TIFF and has dimensions of (224, 224, 5).

2.4 Modifying the pre-trained VGG-16 model for Binary classification

VGG-16 is a well-known convolutional neural network trained on more than one million images from the ImageNet database.[14,16] When the number of channels in the input image increases from three to five, it only affects the trainable parameters in the first convolutional layer, leaving the subsequent convolutional layers unchanged. The weights of all convolutional layers except the first one are transferred to the customized VGG-16 model. The weight values for the first three channels of the first convolutional layer are obtained from the original VGG-16 model and the weights for the extra two channels are calculated by taking the average across the existing channel weights.[17] To improve the performance of the model in classifying the five-channel dataset (as crack or no crack), transfer learning and fine-tuning approaches are employed. Specifically, the pre-trained VGG-16 model weights are leveraged, and then fine-tuned to adjust to the characteristics of the new dataset, allowing the model to quickly learn the features of new dataset.[13] The step-by-step implementation of the customized five-channel model is explained in detail in Section 3.3.

3. RESEARCH IMPLEMENTATION

The analysis of the suggested model consists of various procedures required to evaluate the models efficacy. The following are the crucial steps in evaluating the proposed models effectiveness and ability to improve automatic crack detection procedures.

1. Initially, combine the cropped images, directional-wise, into a single TIFF image with five-channels. This is used as an input to the proposed model. This multi-channel input improves crack detection, classification accuracy, and efficiency in concrete structures.
2. Create a dataset to train and test the models accuracy. The dataset should include positive and negative cases, as the model is a binary classifier.
3. Implement the five-channel VGG-16 model that takes advantage of the variations in patterned lighting to produce five channels.
4. Establish a benchmarking process to evaluate the proposed models performance. This compares the proposed model with the existing model regarding true positive rate, false positive rate, precision, F1 score, accuracy, and Matthew's correlation coefficient. Subsequent sections provide detailed explanations for each of these steps.

3.1 Creation of TIFF image

To produce a five-channel TIFF image, the first step is to convert the original three-channel images taken under five different lighting directions: right, below, left, above, and diffused into grayscale images. Once the grayscale images are created, they are combined to form a single five-channel image by stacking them along the third dimension. Each grayscale image is allocated its own channel, resulting in a multi-channel image that collects information from all five lighting directions. The final step involves saving the combined image in TIFF format, i.e. the multi-channel image is saved with .tiff extension, as shown in Figure 2. This format is very useful for the revised VGG-16 model, whose input shape is (224, 224, 5).

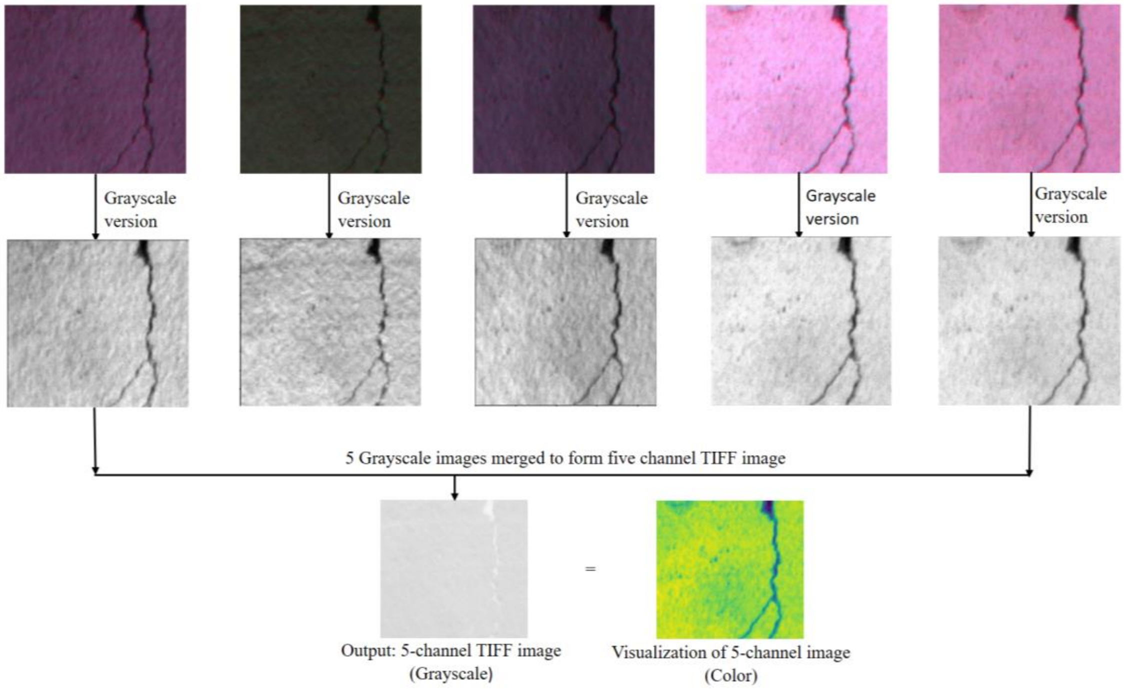

Figure 2. Summary results of the five-channel TIFF image obtained using directional lighting approach. The three-channel images captured using directional lighting are converted to corresponding grayscale images, and further combined to form a single five-channel TIFF image (grayscale). The visualization of the five-channel grayscale image is also shown.

3.2 Data Preparation

In our laboratory facilities, four large and thin concrete slabs of dimensions 500×500×10 mm were cast, and cracks are generated by applying forces at the edges. The crack widths on these samples range from 0.07 mm to 0.3 mm, making the dataset representative of the cracks typically observed in concrete structures. The dataset used in this study comprises high-resolution images of concrete surfaces with visible cracks captured when the lighting was directed at a 50-degree angle to the surface, resulting in five images. These five images are captured under various lighting conditions including Right, Below, Left, Above and Diffused; for each inspection area, using our directional lighting apparatus. The images are then pre-processed as detailed in Section 2.2 and combined into five-channel TIFF image as outlined in Section 2.3, to facilitate the input to customized VGG model explained in Section 3.3.

3.3 Model customization (Five-Channel VGG-16 CNN model)

In this implementation, we take a pre-trained VGG-16 model,[14] originally trained on the ImageNet[16] dataset with three input channels, and modify it to accept five input channels. The goal is to demonstrate how to fine-tune a pre-trained model for a specific binary classification task with five input channels. Figure 3 shows the process of implementation of five channel VGG-16 convolutional neural network model.

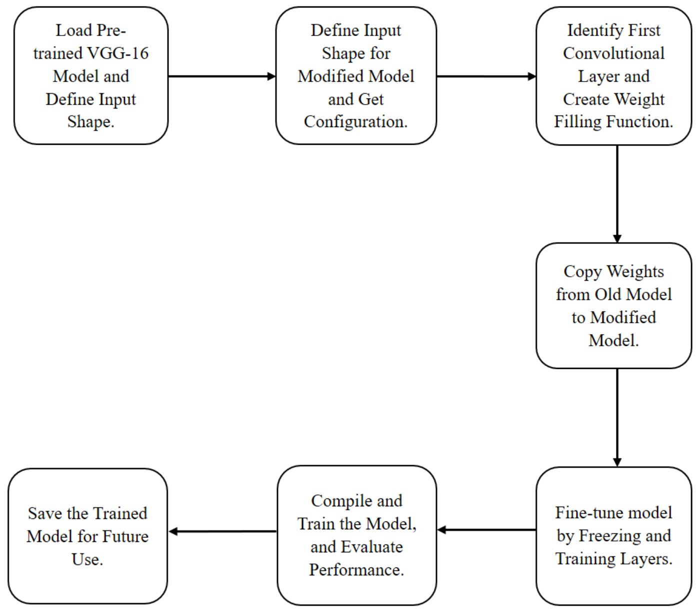

Figure 3. Implementation of five-channel VGG-16 convolutional neural network model.

3.4 Performance Evaluation

A training and testing dataset is used to evaluate the performance of the proposed five-channel VGG-16 model. First, train the modified model on the training dataset for 10 epochs, using a batch size of 32. Next, use the trained model to make predictions on the testing dataset and convert the predicted probabilities to binary class labels using a threshold of 0.5. Next, evaluate the models performance using a confusion matrix generated by comparing the predicted labels to the actual labels of the testing dataset. Next, calculate the values for true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values, which represent the number of correctly and incorrectly classified instances from the confusion matrix.[18] Finally, these values are utilized to calculate various evaluation metrics[18] shown in Table 1.

Table 1. Performance metrics for a classifier

Metric	Description	Equation
True positive rate (TPR) (recall)	The proportion of true positive predictions out of all actual positive instances.	$TPR = \frac{TP}{TP + FN}$
False positive rate (FPR)	The proportion of false positive predictions out of all actual negative instances.	$FPR = \frac{FP}{TN + FP}$
Positive predictive value (PPV) (precision)	The proportion of true positive predictions out of all positive predictions.	$PPV = \frac{TP}{TP + FP}$

F1 score (F1)	The harmonic mean of precision and recall.	$F1 = \frac{2 \times TPR \times PPV}{TPR + PPV}$
Accuracy	The measure of the models ability to correctly predict the outcomes.	$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$
Matthew's correlation coefficient (MCC)	The measure of the correlation between predicted and actual labels, accounting for true or false, both positive and negative predictions.	$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

4. EMPIRICAL ANALYSIS

The performance of the five-channel VGG-16 model is compared to the traditional three-channel VGG-16 model.[14] To evaluate the effectiveness of each model, we use performance metrics in Table 1. Figure 4 compares the performance metrics for the traditional VGG-16 three-channel model and the proposed VGG-16 five-channel model for crack detection and classification. It's observed that the five-channel model has a higher performance in several metrics than the three-channel model.

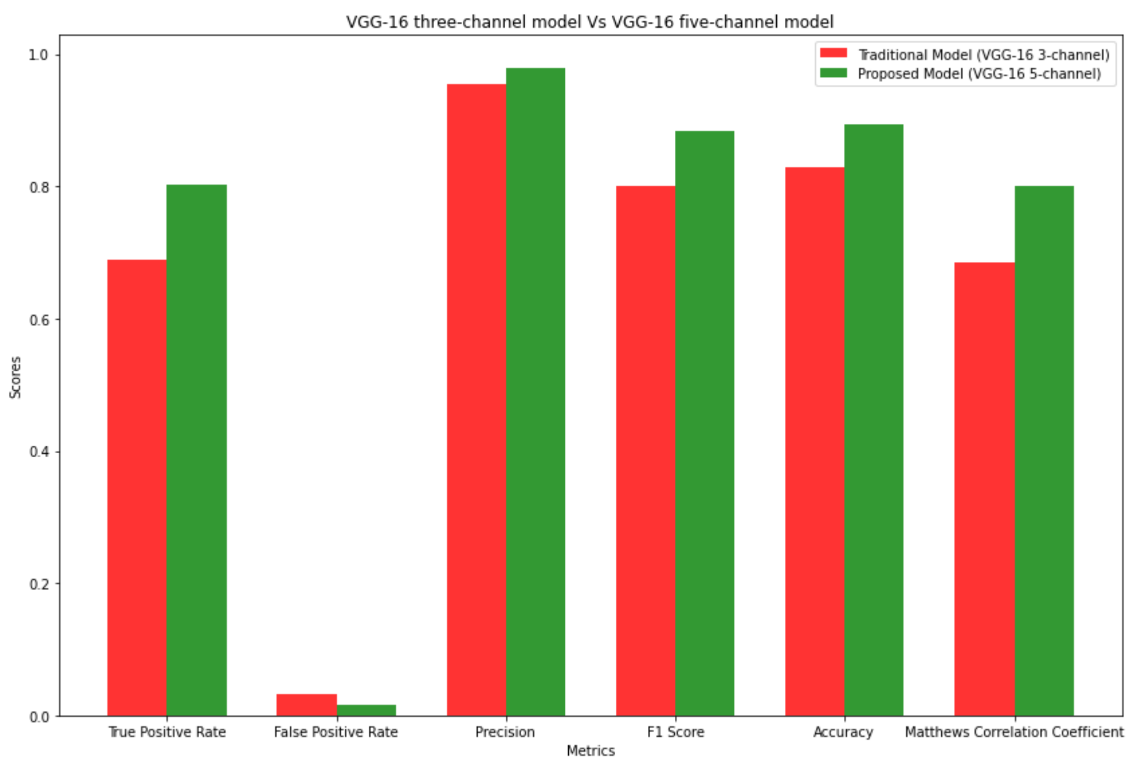

Figure 4. A comparison of the performance metrics between the existing method and the proposed method. The acronyms are explained in Table 1.

The five-channel model has a higher TPR of 0.80 than the three-channel model TPR of 0.69, indicating that the five-channel model is better at detecting positive cases. The five-channel model has a lower FPR of 0.02 compared to the three-channel model FPR of 0.03, indicating that the five-channel model is better at correctly identifying negative cases. The five-channel model has a higher Precision of 0.98 compared to the three-channel model Precision of 0.95 indicating that the five-channel model is better at avoiding false positive predictions. The five-channel model has higher F1 Score of 0.88 than the three-channel model F1 Score of 0.80, indicating that the five-channel model is better at balancing the trade-off between precision and recall. The five-channel model has a higher accuracy of 0.89 than the three-channel model accuracy of 0.83, indicating that the five-channel model is more accurate in correctly identifying positive and negative cases. Finally, the five-channel model has a higher MCC of 0.80 than the three-channel model MCC of 0.68, indicating that the five-channel model has no imbalance in true positive and false positive predictions. In conclusion, the five-channel model outperforms the three-channel model regarding TPR, FPR, precision, F1 score, accuracy, and MCC. These metrics suggest that the five-channel model is better at detecting positive cases, reducing false positive predictions, balancing precision and recall, and accurately identifying positive and negative cases than the traditional three-channel model.

5. CONCLUSION

The proposed five-channel VGG-16 convolutional neural network model demonstrated significant improvements over the traditional three-channel model in various metrics, including TPR (recall), FPR, precision, F1 score, accuracy, and Matthew’s correlation coefficient (MCC). These results indicate the potential of the five-channel model in accurately detecting and classifying cracks in concrete structures in low-light environments, with widths ranging from 0.07 mm to 0.3 mm. The use of directional lighting, transfer learning and fine-tuning approaches significantly enhanced the performance of the proposed model. The methodology used in this study has potential applications in other image-based classification problems, such as object detection, segmentation, and tracking in fields such as medical imaging, autonomous driving, and surveillance systems. The evaluation time for the three-channel and five-channel models are comparable, and Graphics Processing Unit (GPU) can further improve processing time. Future work will involve comparing deep learning models and architectures by incorporating additional input channels to enhance the models accuracy. The study develops a bespoke deep learning model that analyzes the images captured under different lighting conditions to detect cracks that are not visible to the naked eye. This improves the accuracy and efficiency of SHM and supports civil engineers in identifying potential issues early, lowering the risk of failure and increasing the life span of structures.

ACKNOWLEDGEMENTS

The University of Strathclyde’s Advanced Nuclear Research Centre supported this work.

REFERENCES

[1] Farrar, C. R. and Worden, K., “An introduction to structural health monitoring,” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **365**(1851), 303–315 (2007).

[2] Zhang, B., Xia, K., Li, X., Yan, G. and Zhang, K., “Smart film for crack monitoring of concrete bridges,” Structural Health Monitoring **10**, 275–289 (2011).

[3] Saleem, M. and Gutierrez, H., “Using artificial neural network and non-destructive test for crack detection in concrete surrounding the embedded steel reinforcement,” Structural Concrete **22**(5), 2849–2867 (2021).

[4] Lins, R. G. and Givigi, S. N., “Automatic Crack Detection and Measurement Based on Image Analysis,” IEEE Transactions on Instrumentation and Measurement **65**(3), 583–590 (2016).

[5] Tian, Y., Chen, C., Sagoe-Crentsil, K., Zhang, J. and Duan, W., “Intelligent robotic systems for structural health monitoring: Applications and future trends,” Automation in Construction **139**, 104273 (2022).

[6] Agnisarman, S., Lopes, S., Madathil, K. C., Piratla, K. and Gramopadhye, A., “A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection,” Automation in Construction **97**, 52–76 (2019).

[7] Golding, V. P., Gharineiat, Z., Munawar, H. S. and Ullah, F., “Crack Detection in Concrete Structures Using Deep Learning,” Sustainability **14**(13) (2022).

[8] Sony, S., Dunphy, K., Sadhu, A. and Capretz, M., “A systematic review of convolutional neural network-based structural condition assessment techniques,” Engineering Structures **226**, 111347 (2021).

[9] Dorafshan, S., Thomas, R. J. and Maguire, M., “Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete,” Construction and Building Materials **186**, 1031–1045 (2018).

[10] SeolDong-Hyeon, OhJi-Hoon, and KimHong-Jin., “Comparison of Deep Learning-based CNN Models for Crack Detection,” Journal of the Architectural Institute of Korea Structure & Construction **36**(3), 113–120 (2020).

[11] Yangí, L., Yang, G., Liu, Z., Chang, Y., Jiang, B., Awad, Y. and Xiao, J., “Wall-climbing robot for visual and GPR inspection,” 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 1004–1009 (2018).

[12] Cha, Y.-J., Choi, W. and Büyüköztürk, O., “Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks,” Computer-Aided Civil and Infrastructure Engineering **32**(5), 361–378 (2017).

[13] Yosinski, J., Clune, J., Bengio, Y. and Lipson, H., “How transferable are features in deep neural networks?,” Advances in neural information processing systems **27** (2014).

[14] Simonyan, K. and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556 (2014).

[15] Desk, A. D., “Tiff” (1986).

[16] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L., “ImageNet: A large-scale hierarchical image database,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255 (2009).

[17] Oguntayo, S., “Implementing Transfer Learning from RGB to Multi-channel Imagery,” Towards Data Science, 23 March 2021, <https://towardsdatascience.com/implementing-transfer-learning-from-rgb-to-multi-channel-imagery-f87924679166>.

[18] Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D. and Lestantyo, P., “Cross-validation Metrics for Evaluating Classification Performance on Imbalanced Data,” 2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 14–18 (2019).