Indoor Defect Region Identification Using an Omnidirectional Camera and Building Information Modeling

Yonghan Kim[1] and Ghang Lee[1]

[1]Department of Architectural Engineering, Yonsei University, Republic of Korea
cgv03109@yonsei.ac.kr, glee@yonsei.ac.kr

Abstract –

An omnidirectional (i.e., 360°) camera is an efficient device that can capture the status of a room with a single shot. To detect objects in a spherical image captured by a 360° camera, the image should be flattened and divided into patches reflecting normal fields of view (NFoV). However, detecting indoor defects in omnidirectional camera images is difficult because they are relatively small and span multiple patches. Another challenge is to set the appropriate size for an NFoV patch. To overcome these challenges, this paper proposes a method to locate possible regions of indoor defects using building information modeling (BIM). The core idea is to subtract a 360° camera image from a photorealistically rendered BIM model image of the same location. Bounding boxes are generated around the areas where differences are detected. The proposed method was tested in a single room with artificially implanted cracks. In the experiments, two different omnidirectional cameras were used. The image classification algorithm was trained on open crack datasets. The results showed that the proposed method improved the F1-score from 0.15 to 0.39 and recall from 0.16 to 0.87. The proposed method could detect more cracks while reducing the number of patches needed for indoor crack inspection compared to the traditional method.

Keywords –

Defect Inspection; Omnidirectional Camera; Crack Detection; BIM.

1 Introduction

Indoor defect inspection is a crucial task for construction companies to ensure customer satisfaction and maintain the quality of work. Since a unit's condition directly impacts residents, defects in the interior can result in numerous complaints, claims, and even litigation [1]. These conflicts have negative impacts on construction companies' brand images, leading to financial damage [2]. Therefore, in industry and academia, methods for efficiently managing defects have been continuously required.

The methods of inspecting for indoor defects keep advancing, from visual inspection to increased reliance on low-cost, high-performance, easy-to-use devices and advanced technology. These technologies provide opportunities to tackle the challenges associated with the laborious, time-consuming, and error-prone methods of physical visual inspection [3–6]. However, detecting indoor defects in an equirectangular image is still difficult because indoor defects are relatively small compared to general objects.

This study proposes a method to identify indoor defect regions using pixel-wise subtraction of the on-site image from the photorealistically rendered building information modeling (BIM) image. Cracks were selected as the target defect type for inspection. The proposed method was validated through an experiment that involved implanting cracks of various sizes in a single room while capturing indoor scenes using two different off-the-shelf omnidirectional cameras. The results of detection performance and time for inference using the traditional method and the proposed method were compared.

2 Background and Related Studies

The omnidirectional (i.e., 360°) camera is an efficient and easy-to-use device that can capture a comprehensive view of an area in one shot. As such, 360° cameras are utilized for inspecting enclosed spaces, such as tunnels, pipes, and culverts [7–9], as well as collecting indoor scene data for real estate advertising purposes [10].

The 360° camera has also been used to capture facility conditions and to detect indoor defects by leveraging advanced image-processing technologies that can detect objects and classify patches. Humpe [11] has shown the captured visual feature from a 360° camera can be also utilized for autonomous crack inspection with similar

results to a standard high-definition camera when it is captured close enough. Chow et al. [4] used the 360° camera to capture defects on concrete surfaces and classify extracted patches with a deep learning-based image classification model. Luo et al. [12] proposed customized object detection models to detect defects on steel surfaces from extracted patches.

Equirectangular images, another form of a spherical image from a 360° camera, is the format used to not only store and transmit a spherical image but also widely adopted as an input format in related studies. There are three general approaches to detecting objects from the equirectangular image. Some studies [13–16] proposed methods to input the whole equirectangular image into the trained object detector. Other approaches [17,18] are to input cropped patches with a flat grid into the object detector or image classifier. The other approaches [4,13] are to input the normal field of view (NFoV) patches, which are divided based on a spherical grid and flattened. These studies have shown that the third approach presents a promising performance in detecting objects from a spherical image, including defects.

However, the major challenge in detecting indoor defects from equirectangular images using the third approach lies in determining the appropriate size for a NfoV patch. The conventional method of segmenting an equirectangular image into NFoV patches involves dividing the entire image into overlapping patches, with the size determined arbitrarily by the developer based on the target size. Moreover, indoor defects are relatively small compared to target objects in related studies, resulting in an excessive number of patches. Despite advances in computing capacity that allow for the inspection of numerous image patches, suitable patch sizes must still be identified to minimize redundancy and improve analysis speed. Consequently, a region identification method is required to localize defects and reduce the number of image patches that need to be examined by a trained classification model.

Therefore, this study aims to identify and localize the region of defects from equirectangular images to reduce the number of image patches that need to be examined by a trained classification model, thereby ensuring efficient automated indoor defect inspection.

3 Methods and Implementations

A defect region is identified from the equirectangular images based on the proposed sequence of modularized processing methods as shown in Figure 1. The proposed method aims to improve the traditional method of extracting NfoVs, which uses a spherical grid for classification. The detailed implementation process consists of eight steps, as depicted and described in the following sub-sections. Most methods were implemented using Python with the OpenCV library.

3.1 View Generation

Revit, which was used to generate a BIM model, did not support viewpoint generation on exactly designated coordinates using a graphical user interface (GUI). Therefore, an application programming interface (API) that could generate a view of the exactly designated coordinates in the BIM model was used. In this study, the central coordinates of the target room and the height of each omnidirectional camera were used as parameters for view generation.

Enscape was plugged into Revit to generate 3-dimensional photorealistic rendering images directly

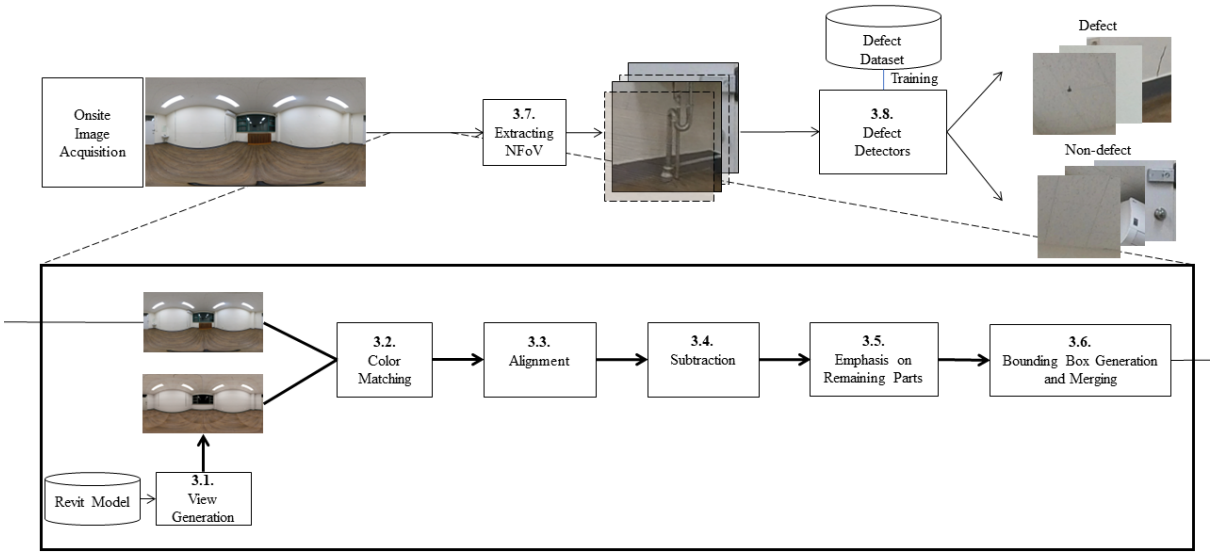

Figure 1. Overview of the proposed method.

from a Revit model. As Enscape was plugged into Revit, viewpoints from the Revit and Enscape environments were automatically synchronized. Enscape provides a feature for generating spherical images in equirectangular form by capturing multiple images at designated locations and stitching them.

3.2 Color Matching

Two images from different sources, regardless of how realistic and similar they are, inevitably have different distributions of pixel values because of different conditions, such as lighting and reflections. Therefore, color-matching methods were applied to reduce the gap between each pixel value. The pixel values of the original images were adjusted. Intensity refers to the pixel value distribution of the source image and was used to efficiently discard similar regions and emphasize distinct regions.

3.3 Alignment

Another merit of a 360° camera is that generated image can stand upright regardless of the pose of the device because of the inertial measurement unit (IMU) sensor. This, combined with the robust tripod, allows an on-site image to be easily aligned with the image from Enscape. However, the camera could be mislocated slightly away from the center of the target room or misdirected by errors from the sensors and a slanted floor.

To mitigate these problems, an image alignment module was implemented. This module calculated the average of the subtracted pixel values and defined them as a target for minimization. The parameters to be optimized were pitch, yaw, and roll. The parameters could rotate on-site spherical images to align with the rendered images to minimize the difference between the two sets of images. The SciPy [19] library was utilized to implement the optimization process.

3.4 Subtraction

This section discusses the pixel-wise subtraction of two spherical images in equirectangular form. The normal subtraction operation, which subtracts the rendering image from the on-site image to negate visually similar parts, resulted in negative pixel values outside the valid range of 0 to 255. To handle negative pixel values, a unique operation was required. In this research, the absolute difference operation was used, which involves taking the absolute values of the subtracted pixel values. This operation ensured that the final values for the pixels were all positive and within the valid range. Using the absolute difference operation was particularly important because normal subtraction methods could result in different outcomes depending on the order of the minuend and subtrahend. By using the absolute values of the subtracted pixel values, a consistent measure of the differences between the rendered image and the on-site images, regardless of which was the minuend.

3.5 Emphasis on Remaining Parts

In ideal cases, the different parts will have high pixel values after subtraction while the same parts would have pixel values of approximately (0, 0, 0). Pixels that have values less than 5 are neglected by the thresholding operation. To emphasize the remaining features and denoise the subtracted result, canny edge detection was utilized [20]. Then, the parameter values of the low and high threshold for canny edge detection were heuristically set to 80 and 160 respectively.

3.6 Bounding Box Generation and Merging

The minimum bounding boxes of the emphasized edges can be generated using contour features from OpenCV. However, too many bounding boxes overlapped and were close to each other. Therefore, a distance-based bounding box merging algorithm was implemented with the expectation that each bounding box would cover a single object to be inspected. The distance between the bounding boxes was calculated based on the center coordinates of each box.

3.7 Extracting the Normal Field of View

Every pixel from an equirectangular image could be mapped on the spherical coordinate system based on the corresponding longitude and latitude. Therefore, considering the coordinates corresponding to the centers of the bounding boxes and the sizes of each box, the NFoV patches were extracted using flattening methods.

3.8 Defect Detector and Training Datasets

A module that can distinguish images with cracks and images without cracks was also implemented to validate the usage of the proposed method. The detector can be an object detection model, semantic segmentation model, or image classification model. Although segmentation models excel at detecting cracks, creating a dataset for indoor defects in the form of a segmentation or object detection task can be challenging due to the diverse range of defect shapes and sizes. Thus, this study chose an image classification model as a detector, which significantly reduced data preparation time and effort.

In this research, the image classification model VGG-19, a convolutional neural network (CNN)-based image classification algorithm that has demonstrated promising

performance for various classification tasks, was trained using open crack datasets [21,22]. The datasets for training were secured from publicly open datasets [21,23,24]. The main reason for selecting these datasets was to produce crack features for various backgrounds based on the on-site crack features.

4 Experiment

An experiment was conducted to determine the validity of the proposed and implemented methods in a controlled environment. In this section, the specifications of the 360° cameras, the as-designed BIM model, and the implanted cracks are discussed.

4.1 Devices Used to Acquire On-Site Images

A single room with artificially implanted cracks was inspected by two off-the-shelf omnidirectional cameras. Devices less than $600 were intentionally selected for their price, considering the accessibility and generalizability of each device. The Insta One X2 from Insta 360 was around $430, and the QooCam 8K was around $590. Two fish-eye lenses were applied to both devices to acquire more than each 180° scene and to stitch the two images together as one spherical image. The major difference between the two devices is the maximum resolution of the footage. The QooCam 8K can generate an image with an 8K (7680 × 3840) resolution, while the other can generate a 5.7K (6080 × 3040) resolution.

4.2 BIM Model Generation

The experiment was conducted in a real building, where a BIM model was generated using the as-designed conditions of the target room in real world. The BIM model was used to generate a reference spherical image, which was subtracted from an on-site image. Objects that were not described in the corresponding plan but exist in the physical room, such as the air conditioner, sink, pipes, and radiator, were neglected in the model. The detailed model featured objects such as power sockets, lights, windows, skirting, and a door. To create a photorealistic rendering as close to reality as possible, material mapping was conducted using manually acquired material libraries. Finally, in order to adjust the color and geometrical differences between the rendered image and the on-site image, ‘color matching’ and ‘aligning’ algorithms were utilized.

4.3 Implanting Cracks

Since the target room did not have cracks that could be detected by the trained classifier, 19 cracks of various sizes were manually implanted. The sizes of the implanted cracks were intentionally determined based on the width and length of each crack from small to large. The small cracks were less than 100 mm^2, the medium-sized cracks were less than 1,000 mm^2, and the large-sized cracks were over 1,000 mm^2.

5 Results

Following the traditional method, an image from each device generated 1,028 extracted NFoV patches, meaning each patch overlapped approximately 30% of the neighboring image. The Qoocam, which can generate images with a higher resolution, had the best detection results, securing an F1 score of 0.218 by detecting the largest number of cracks.

After applying the proposed methods, the number of patches to be inspected decreased from 1,028 to 187 and 118 patches, respectively, leading to increased recall and precision values, indicating a reduction in the number of false positive and false negative detection cases. The actual number of detected cracks using Insta One X2 footage increased from 4 to 7 but decreased from 9 to 7 in the case of the Qoocam images. The reason for the decreased number of detected cracks is that the traditional method gives a model multiple opportunities to classify the same cracks from a different perspective, while the proposed method only provides a single opportunity. Another reason is that even though overall detection performance was high with the high-resolution camera, some crack features are neglected through the ‘emphasis on remaining parts’.

Considering the effectiveness that can be quantified as F1-score, the results indicate that the proposed method is valid, even though it sacrifices some number of true positive detection results. The detailed detection results are shown in Table 1. Processing time refers to the overall duration taken for classifying the extracted patches. With the VGG-19 model, 1028 patches from the traditional method were processed in 14 seconds while the reduced number by the proposed method lead to decreasing

Table 1. Overall Experimental Results

Condition	Device (Resolution)	No. of patches	Processing Time (Sec)	F1 Score	Detected Cracks (detected/total)
Traditional method	Qoocam (8K)	1028	13.82	0.218	9/19
	InstaOne X2 (5.7K)	1028	8.81	0.145	4/19
Proposed method	Qoocam (8K)	187	2.51	0.263	7/19
	InstaOne X2 (5.7K)	118	1.59	**0.387**	6/19

processing time to 3 seconds on an RTX 3090 GPU. The result signifies a substantial decrease in processing time, with the total time for processing reduced to approximately 18% of the processing time incurred by the traditional method.

Five types of cracks that could not be detected before could be detected using the proposed methods. However, small cracks, shorter than 100 *mm*, could not be detected by any device using any of the methods.

Figure 2 and Figure 3 show the samples of true positive cases, while the red boxes indicate the cracks that were not detected by the traditional method but detected when adjusting the proposed method.

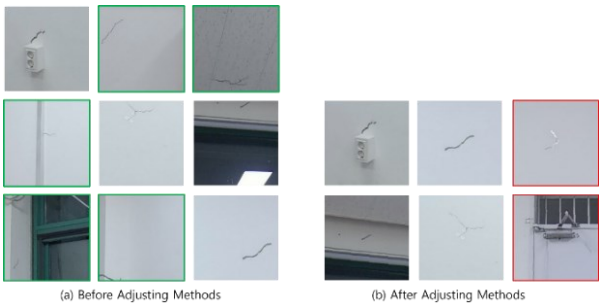

Figure 2. True positive cases and after adjusting the methods based on InstaOne X2 images.

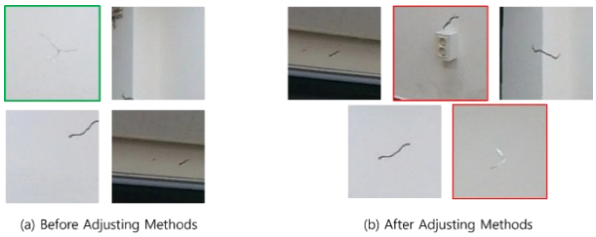

Figure 3. True positive cases before and after adjusting the methods based on Qoocam images.

6 Discussion

The main purpose of this research, to reduce the number of redundant patches, was achieved using the proposed method. Even though it primarily inspected single patches for one crack feature, the proposed method could detect new crack features that formal methods could not detect. This phenomenon happens because the patches' sizes and locations were determined to have similar visual features based on the training datasets. The major parts of the training images had centralized crack features.

Unlike the traditional methods, the proposed method could also extract the region that is different from the clean image from the as-designed BIM model. The patches generated using traditional methods consist of many meaningless patches from floor and ceiling images, but the proposed methods can generate a suspected region that is different from the as-design image.

Despite the achievements of this research, the results have several limitations. First, the trained classifier showed that the proposed method cannot be applied for practical defect inspection. Even though the detector was trained using numerous crack images with various backgrounds, the crack images from the indoor spherical images were unseen data. Therefore, future research is needed to find an appropriate detector that can distinguish real indoor defects on a practical level with securing real indoor defect datasets.

Second, small crack features could be neglected by the proposed method. For example, green boxes in Figure 2 and Figure 3 indicate cases where the traditional method successfully detected defects that the proposed method could not detect. These small crack features were neglected while thresholding or emphasizing the remaining parts after image subtraction, depending on user-defined parameters for the canny edge detection algorithm. Therefore, the optimal parameters for each indoor scene condition must be determined in future research.

Third, the location of the camera for on-site image acquisition may not have been in the exact center of the target room. Although several modules were devised to mitigate this issue, images acquired from a completely misplaced device cannot be assessed using this method. In future research, the center points must be accurately secured by leveraging additional methods, such as simultaneous localization and mapping (SLAM).

This method was developed aiming at detecting cracks and other types of defects during construction or the pre-occupancy inspection before furniture, home appliances, and electrical fixtures are installed. However, any visual inspection using cameras has a disadvantage in that it can only work on visible things. The proposed vision-based method has this inherited limitation.

7 Conclusion

An omnidirectional camera is an efficient device for the inspection of rooms because it can capture the whole scene of the target space at one time. However, traditional methods for detecting defects using spherical images have been difficult to implement because a crack occupies a small portion of an image and because there is ambiguity regarding the appropriate size of the NFoV patches. Therefore, a method to extract regions suspected of having defects was proposed as a sequence of modules. The proposed method was validated through a lab-conditioned experiment. The results showed that the proposed method could effectively reduce the number of NfoV patches to be inspected while ensuring the same or even better detection performance than the traditional

method. This method is expected to decrease the number of false positive cases and reduce the overall inspection time. Ultimately, the proposed method is expected to offsets the time and effort required to create a rendering image when it applied to defect inspection of numerous units of identical housing types, such as apartment units.

Our future research will focus on determining proper detectors and ways to train them to identify patent indoor defects. Additional methods must be applied to ensure that the camera is in the correct location and position. Finally, experiments for validation must be conducted at a site featuring real defects.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT, MSIT) (No. 2021R1A2C3008209)

References

[1] Kim, D.-H., Lee, D., Lee, H.-J., Min, Y.-G., Park, I., Cho, H. Analysis of importance by defect type in apartment construction. *Journal of the Korea Institute of Building Construction,* 20, 357–365, 2020. https://doi.org/10.5345/JKIBC.2020.20.4.357.

[2] Othman, A.A.E. An international index for customer satisfaction in the construction industry. *International Journal of Construction Management* 15: 33–58, 2015. https://doi.org/10.1080/15623599.2015.1012140.

[3] Perez, H. and Tah, J.H.M. Deep learning smartphone application for real-time detection of defects in buildings. Structural Control and Health Monitoring, 28: e2751, 2021. https://doi.org/10.1002/stc.2751.

[4] Chow, J.K., Liu, K., Tan, P.S., Su, Z., Wu, J., Li, Z., Wang, Y.-H. Automated defect inspection of concrete structures. *Automation in Construction* 132: 103959, 2021. https://doi.org/10.1016/j.autcon.2021.103959.

[5] Filgueira, A., Arias, P., Bueno, M., Lagüela, S. Novel inspection system, backpack-based, for 3D modelling of indoor scenes. International Conference on Indoor Positioning and Indoor Navigation, page 4, 2016.

[6] Du, H., Henry, P., Ren, X., Cheng, M., Goldman, D.B., Seitz, S.M., Fox, D. Interactive 3D modeling of indoor environments with a consumer depth camera. *Proceedings of the 13th International Conference on Ubiquitous Computing,* pages 75–84, 2011. https://doi.org/10.1145/2030112.2030123.

[7] Li, D., Xie, Q., Gong, X., Yu, Z., Xu, J., Sun, Y., Wang, J. Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. *Advanced Engineering Informatics* 47: 101206, 2021. https://doi.org/10.1016/j.aei.2020.101206.

[8] Meegoda, J.N., Kewalramani, J.A., Saravanan, A. Adapting 360-degree cameras for culvert inspection: Case study. *Journal of Pipeline Systems Engineering and Practice* 10: 05018005, 2019. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000352.

[9] Karkoub, M., Bouhali, O., Sheharyar, A. Gas pipeline inspection using autonomous robots with omni-directional cameras. *IEEE Sensors J*, 21: 15544–15553, 2021. https://doi.org/10.1109/JSEN.2020.3043277.

[10] Sulaiman, M.Z., Aziz, M.N.A., Bakar, M.H.A., Halili, N.A., Azuddin, M.A. Matterport: Virtual tour as a new marketing approach in real estate business during pandemic COVID-19. Atlantis Press, 221–226, 2020. https://doi.org/10.2991/assehr.k.201202.079.

[11] Humpe, A. Bridge inspection with an off-the-shelf 360° camera drone. *Drones* 4: 67, 2020. https://doi.org/10.3390/drones4040067.

[12] Luo, C., Yu, L., Yan, J., Li, Z., Ren, P., Bai, X., Yang, E. Liu, Y. Autonomous detection of damage to multiple steel surfaces from 360° panoramas using deep neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 36, 1585–1599, 2021. https://doi.org/10.1111/mice.12686.

[13] Chou, S.-H., Sun, C., Chang, W.-Y., Hsu, W.-T., Sun, M., Fu, J. 360-Indoor: Towards learning real-world objects in 360° indoor equirectangular images. *IEEE Winter Conference on Applications of Computer Vision (WACV*), pages 834–842, 2020 https://doi.org/10.1109/WACV45572.2020.9093262.

[14] Wang, K.-H., Lai, S.-H. Object Detection in Curved Space for 360-Degree Camera. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3642–3646, 2019 https://doi.org/10.1109/ICASSP.2019.8683093.

[15] Coors, B., Condurache, A.P., Geiger, A. SphereNet: Learning spherical representations for detection and classification in omnidirectional images. *ECCV 2018.* 518–533.

[16] Lee, Y., Jeong, J., Yun, J., Cho, W., Yoon, K.-J. SpherePHD: Applying CNNs on 360 Images With Non-Euclidean Spherical PolyHeDron Representation. *IEEE Trans. Pattern Anal. Mach. Intell.,* pages 834–847, 2022 https://doi.org/10.1109/TPAMI.2020.2997045.

[17] Turečková, A., Tureček, T., Janků, P., Vařacha, P., Šenkeřík, R., Jašek, R., Psota, V., Štěpánek, V., Komínková Oplatková, Z. Slicing aided large scale tomato fruit detection and counting in 360-degree video data from a greenhouse. *Measurement.* 204: 111977, 2022 https://doi.org/10.1016/j.measurement.2022.111977.

[18] Hirabayashi, M., Kurosawa, K., Yokota, R., Imoto, D., Hawai, Y., Akiba, N., Tsuchiya, K., Kakuda, H., Tanabe, K., and Honma, M. Flying object detection system using an omnidirectional camera. *Forensic Science International: Digital Investigation* 35, 301027, 2020 https://doi.org/10.1016/j.fsidi.2020.301027.

[19] SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, (n.d.). Online: https://www.nature.com/articles/s41592-019-0686-2.

[20] Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-8* 679–698, 1986. https://doi.org/10.1109/TPAMI.1986.4767851.

[21] Ãzgenel, Ã.F. and SorguÃ§, A.G. Performance comparison of pretrained convolutional neural networks on crack detection in buildings, *ISARC Proceedings,* 693–700, 2018.

[22] Xu, H., Su, X., Wang, Y., Cai, H., Cui, K., Chen, X., Automatic Bridge Crack Detection Using a Convolutional Neural Network*, Applied Sciences.* 9 2867, 2019. https://doi.org/10.3390/app9142867.

[23] Dorafshan, S., Thomas, R.J., and Maguire, M. SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. *Data in Brief*, 21: 1664–1668, 2018. https://doi.org/10.1016/j.dib.2018.11.015.

[24] Elhariri, E., El-Bendary, N., and Taie, S.A. Historical-crack18-19: A dataset of annotated images for non-invasive surface crack detection in historical buildings. *Data in Brief*, 41: 107865, 2022. https://doi.org/10.1016/j.dib.2022.107865.