BD3: Building Defects Detection Dataset for Benchmarking Computer Vision Techniques for Automated Defect Identification

Praveen Kottari
praveenkotta@iisc.ac.in
Robert Bosch Centre for Cyber Physical Systems
Indian Institute of Science
Bangalore, India

Pandarasamy Arjunan
samy@iisc.ac.in
Robert Bosch Centre for Cyber Physical Systems
Indian Institute of Science
Bangalore, India

Abstract

The current manual visual inspection of built environments is time-consuming, labor-intensive, prone to errors, costly, and lacks scalability. To address these limitations, automated building inspection techniques have emerged in recent years, leveraging low-cost computer vision systems, drones and mobile robots. However, the practical implementation of these systems is hindered by the lack of robust and generalizable models trained on comprehensive defect image datasets. In this paper, we present *BD3: Building Defects Detection Dataset*, a comprehensive image dataset designed to benchmark computer vision techniques aimed at improving the robustness and generalizability of automated building inspection systems. The BD3 dataset contains 3,965 high-quality, manually collected, and annotated images. Unlike other datasets that primarily focus on crack and non-crack images, BD3 includes images of six distinct building defects (algae, major crack, minor crack, peeling, spalling, and stain), as well as images representing normal building conditions. We benchmarked the BD3 using five state-of-the-art computer vision models to classify defect and normal images. The experimental results indicate that the Vision Transformer (ViT) model achieved the highest F1-scores of 0.9342 and 0.9879 on the original and augmented datasets, respectively. The BD3 dataset and its accompanying reproducible codebase are publicly available for benchmarking other defect detection algorithms.

CCS Concepts

• **Computing methodologies → Computer vision**.

Keywords

Building Inspection, Building Defects, Defect Identification, Deep Learning, Computer Vision, Building Defect Dataset.

ACM Reference Format:
Praveen Kottari and Pandarasamy Arjunan. 2024. BD3: Building Defects Detection Dataset for Benchmarking Computer Vision Techniques for Automated Defect Identification. In *The 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '24), November 7–8, 2024, Hangzhou, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3671127.3698789

BuildSys '24, November 7–8, 2024, Hangzhou, China

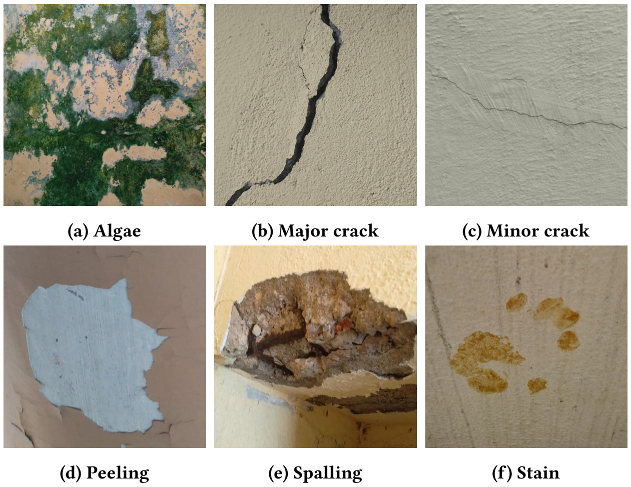
ACM ISBN 979-8-4007-0706-3/24/11
https://doi.org/10.1145/3671127.3698789

1 Introduction

Maintaining built environments, particularly older buildings with defects, and ensuring their structural integrity presents significant challenges. Currently, manual visual inspection methods, which involve skilled professionals, are widely employed. These methods also utilize handheld gadgets, digital meters, and other tools to facilitate the inspection process. While manual inspection techniques can be comprehensive and flexible, they are often time-consuming, labor-intensive, costly, and susceptible to human error [3, 16]. Additionally, they face limitations in accessibility, particularly in high-rise buildings, and scalability, making them less effective for the proactive maintenance of built environments. In response, the recent development of low-cost computer vision technologies, together with mobile robots and drones, presents promising opportunities for automating and modernizing defect inspection methods. These technologies offer a more scalable approach to defect detection and analysis. They can support or complement manual methods by addressing expertise gaps and enabling more detailed, consistent inspection results [3, 22].

In recent years, there has been growing interest in using Artificial Intelligence (AI) and computer vision techniques to automate building inspection methods [8, 17]. However, developing efficient computer vision models presents multiple challenges related to the availability of datasets for training [6]. Firstly, the quality and quantity of available image datasets for training these models are significant concerns. Computer vision models generally require large samples of images per class that include variations; without sufficient data, their performance may suffer when deployed. Secondly, the diversity of defects poses a challenge; existing public datasets contain only a limited number of defect types, making it difficult for models to accurately identify various building issues. Finally, the variability in building types, materials, and environmental conditions complicates the training process and can negatively impact the performance of computer vision models in real-world scenarios. To address these challenges, there is a pressing need for the development of large datasets containing images of multiple defects collected from different buildings under varying conditions. Such datasets would enable the training of AI-based building inspection systems that are more accurate, robust, and generalizable [6].

In this paper, we present *BD3: Building Defects Detection Dataset*, a manually collected and annotated dataset of 3,965 RGB images for training and evaluating the robustness of computer vision techniques to improve building inspection methods. The BD3 dataset includes images of six different defect types (algae, major crack, minor crack, peeling, stain, spalling), as well as images of defect-free buildings, collected from diverse building types. It also includes



(a) Algae **(b) Major crack** **(c) Minor crack**

(d) Peeling **(e) Spalling** **(f) Stain**

Figure 1: Sample images of 6 defect classes in our BD3 dataset.

an augmented version of the original dataset, containing 14,000 samples that feature multiple variants of the original images to enhance the dataset's generalizability. These variants include changes in lighting conditions, image angles, scales, and other alterations to simulate defects in various real-world built environments. Furthermore, we evaluate the utility of BD3 by implementing five state-of-the-art computer vision models: Vision Transformer (ViT) [4], VGG16 [19], ResNet18 [7], AlexNet [9], and MobileNet-V2 [18], and we compare their performance using various evaluation metrics. Our experimental results showed that ViT models achieved the highest F1-scores of 0.9342 and 0.9879 on the original and augmented datasets, respectively. The BD3 dataset and the accompanying reproducible benchmarking code are released as open-source for community use[1].

2 Related Works

There are several studies in the literature that discuss the need for high-quality annotated datasets and the implementation of vision-based defect detection systems [13, 20]. These studies also emphasize the necessity for datasets to be recorded in a standardized manner. A comparison of existing datasets for image-based structural inspection is presented in [1]. In [2], authors introduced a dataset containing images of cracks in masonry structures. In [22], authors presented a wall defect dataset with four classes: cracks, chalk, joints, and normal surfaces. Whereas, in [10], authors presented an aggregated dataset from various open source containing 9000 samples. The SDNET2018 [3] dataset includes 56,000 samples of crack of walls and pavements augmented from only 230 original images and compared the model performance in both fully trained and transfer learning settings. In [5], authors presented a study focusing on the visual tracking of cracks in historical buildings facilitating early identification of structural health issues through architectural examinations. Furthermore, in [15], authors conducted a multidimensional performance analysis on several pre-trained networks assessing factors such as training dataset size, network depth and adaptability to various building materials.

In summary, most existing datasets contain images of only two classes: crack and no crack, as shown in Table 1. These datasets were primarily collected from outdoor structures such as bridges, pavements, and buildings. While a few studies have focused on more detailed defect categories like corrosion, stains, and spalling, prevalent in many structures and crucial for thorough building inspections, the availability of comprehensive datasets with diverse defects remains lacking. Our BD3 dataset addresses this gap.

3 The BD3: Building Defects Detection Dataset

3.1 Data collection and Annotation

Our data collection process began by visiting a diverse range of buildings in Bangalore, India, and collecting their basic attributes such as building type, materials, and age. We subsequently selected 20 buildings, with ages ranging from 10 to over 60 years, various architectural styles and materials, including stone and brick structures. Next, we captured images of different defects and normal conditions from both interior and exterior surfaces using a high-resolution camera at different times during the day time. This approach ensured that the dataset reflects variations in environmental exposure, lighting conditions, and weathering effects.

The raw dataset initially contained over 5,000 images with a resolution of 3024 x 4032 pixels in JPEG format. We carefully examined each image and omitted the poor-quality images. After this cleaning process, 3,965 images were left in our dataset. Next, we cropped the images to ensure that each one contains a single defect with clearly visible features and then downscaled all images to 512 x 512 pixels. We chose this moderate resolution because many computer vision models typically require square images, and it helps optimize both storage and computational resources needed for training and inference. Next, with the help of building inspection personnel, we annotated the collected samples into one of the six defect classes or normal. The names of the defect classes were chosen in consultation with the building inspection team, who recommended these defects as prevalent in many buildings. The entire data collection and annotation process took approximately 200 man-hours, and the final version of the dataset occupies 115 MB of disk space.

Figure 1 shows sample images of the six defects from the BD3 dataset. The following list provides the defect names, number of images, and descriptions:

- Algae (624): Existence of fungi that look like green, brown, black patches or slime on the surface.
- Major crack (620): Crack with visible gap.
- Minor crack (580): Crack with no gap.
- Peeling (520): Loss of outer covering of paint.
- Spalling (500): Surface break with visible inner material.
- Stain (521): Visible man-made or natural colour marks.
- Normal (600): Clean wall, no clue of existence of above defect classes.

3.2 Image Augmentation

Image Augmentation is one of the widely used techniques to generate additional images with different variations to enhance generalizability [11]. We applied different image augmentation techniques to the original dataset of 3,965 images and created an augmented version of BD3 dataset. Different variants of each raw

[1] https://github.com/samy101/bd3-building-defects-detection-dataset

Table 1: Comparison of existing building defect datasets.

Dataset	No. of images (original)	No. of images (augmented)	No. of classes	Image Resolution	Class names
Historical-crack18-19, Egypt [5]	40	3886	2	256 x 256	Crack, Non crack
SDNET 2018, Utah, USA [3]	230	56000	2	256 x 256	Crack, No crack
Masonry wall, Netherlands [2]	469	11491	2	224 x 224	Crack, No crack
Wall crack, Turkey [15]	500	40000	2	224 x 224	Crack, No crack
CSSC, China [21]	1232	-	2	130 x 130	Crack, Spalling
Building data-bank, China [12]	1250	60000	2	256 x 256	Crack, No crack
CODEBRIM, Germany [14]	1590	-	4	Varying	Crack, Corrosion Stain, Spalling, Other
Japan [22]	2000	17000	5	64 x 64	Crack, Chalk, Joint, Surface, Other
UK [16]	2622	-	4	224x224	Mould, Stain, Paint, Deterioration, No crack
BD3 dataset	**3965**	**14000**	**7**	**512 x 512**	**Algae, Major crack, Minor crack, Peeling, Stain, Spalling, Normal**

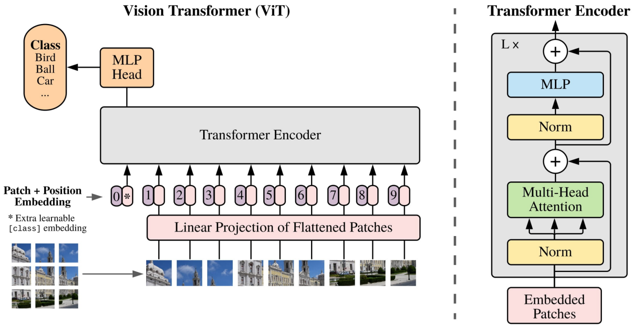


Figure 2: Vision Transformer Backbone [4]

image were generated by randomly applying geometric transformations such as rotations, vertical flips, and horizontal flips as well as color space adjustments to modify brightness, contrast, saturation, and hue. We generated 2,000 samples of augmented images for each of the seven classes resulting in a total of 14,000 samples. The augmented version of the dataset enhances generalizability by incorporating various variations of the original image samples, thereby improving the performance and robustness of computer vision models trained on this dataset.

4 Benchmarking

We evaluate BD3 using five contemporary and widely used computer vision algorithms: Vision Transformers (ViT), VGG16, ResNet18, AlexNet, and MobileNetV2. These algorithms represent different architectural approaches and varying levels of complexity. Our primary objective was to evaluate their performance on our dataset, gain insights into their detection efficiency for each defect class and identify any limitations.

4.1 Computer Vision Models

(1) **Vision Transformers (ViTs) [4]** have revolutionized computer vision by leveraging the power of the transformer architecture. They have demonstrated remarkable performance on various recognition tasks surpassing traditional convolutional neural networks (CNNs) in many applications. The original ViT architecture as shown in Figure 2 consists of a stack of 12 transformer layers. The input image is divided into 16x16 patches resulting in a length of 196. Absolute position embeddings are added and a classification token is prepended. The transformer encoder is then applied to the sequence and the final output is passed through a fully connected layer to obtain the image classification predictions.

(2) **VGG16** [19] has been widely adopted for various image classification tasks due to its effectiveness and relatively straightforward architecture. It consists of 16 layers including 13 convolutions followed by max pooling and 3 fully connected layers.

(3) **ResNet18** [7] utilizes residual learning through skip connections consists of 18 layers including convolutions layers, batch normalization and ReLU activation functions.

(4) **Alexnet** [9] is one of the first architectures that used GPUs for training and consists of 5 convolutional layers, 3 max pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer.

(5) **MobileNet-V2** [18] is a light-weight inverted residual structure consists of an initial convolution layer followed by a series of inverted residual blocks, global average pooling and a fully connected layers.

4.2 Experimental setup

We used the PyTorch framework to train all five computer vision models. We employed transfer learning by initializing the weights of the models based on ImageNet and added a custom fully connected layer to classify the seven classes: six types of defects and normal conditions. Separate defect classification models were developed using the original and augmented datasets, with a split of 60% for training, 20% for validation, and 20% for testing. All experiments were conducted on a server equipped with an NVIDIA A6000 GPU and 48 GB of memory. After training, we evaluated each model’s performance based on its predictions using standard metrics, including precision, recall, F1-score, and the confusion matrix.

Table 2: Comparison of performance metrics (precision, recall, and F1-score) for five defect classification models. The ViT model achieved the highest F1-scores of 0.9323 and 0.9879 on the original and augmented datasets, respectively.

Model	Original dataset			Augmented dataset		
	Precision	Recall	F1-score	Precision	Recall	F1-score
ResNet18	0.8320	0.8308	0.8301	**0.9915**	0.9516	0.9711
VGG16	0.8409	0.8359	0.8363	0.9066	0.9057	0.9056
MobileNetV2	0.8479	0.8422	0.8419	0.8756	0.8750	0.8746
AlexNet	0.8842	0.8801	0.8803	0.9399	0.9389	0.9391
ViTpatch16	**0.9342**	**0.9318**	**0.9323**	0.9880	**0.9879**	**0.9879**

Table 3: Comparison of the ViT model's performance across different defect types on the original and augmented datasets.

Class	Original dataset			Augmented dataset		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Algae	0.9915	0.9516	0.9711	1.0000	0.9975	0.9987
Major crack	0.8761	0.8534	0.8646	0.9794	0.9550	0.9670
Minor crack	0.8417	0.9435	0.8897	0.9612	0.9925	0.9766
Peeling	0.9595	0.9134	0.9359	0.9851	0.9925	0.9887
Spalling	0.9579	0.9100	0.9333	0.9875	0.9875	0.9324
Stain	0.9166	0.9519	0.9339	0.9950	0.9975	0.9962
Normal	1.0000	0.9916	0.9958	0.9974	0.9925	0.9949

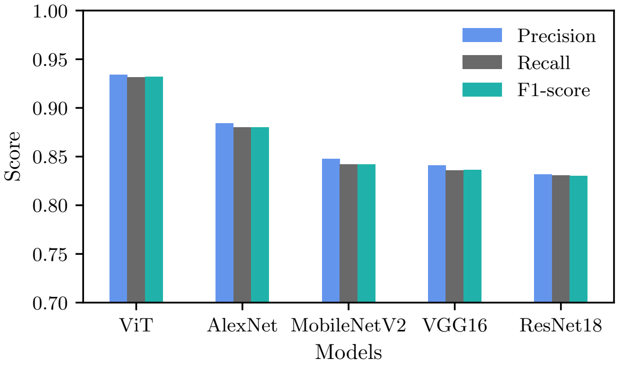


Figure 3: Comparison of model performance on the original dataset. ViT achieved the highest F1-score of 0.9323.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - score = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

where, True Positives (TP) refer to instances correctly predicted as positive, False Positives (FP) are instances incorrectly predicted as positive and False Negatives (FN) are instances incorrectly predicted as negative.

4.3 Results

Table 2 and Figure 3 compare the performance of all five defect classification models. The ViT model achieved the highest F1-scores of 0.9323 and 0.9879 on the original and augmented datasets, respectively. AlexNet followed with F1-scores of 0.8803 and 0.9391. ResNet18 and MobileNetV2 had the lowest F1-scores among the five models on the original and augmented datasets, respectively. Notably, all models showed improved performance on the augmented dataset, as they could learn from a large number of images with various feature variations.

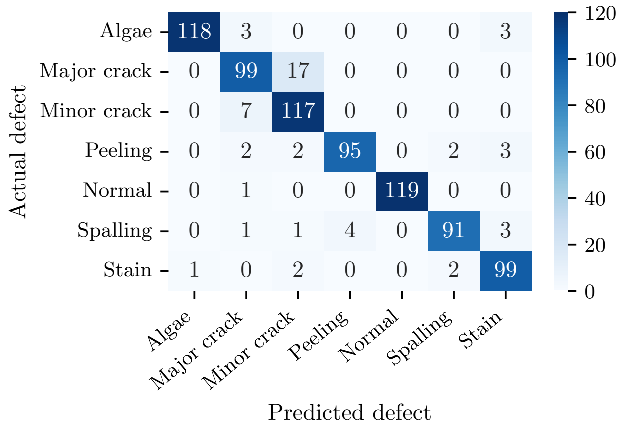


Figure 4: Confusion matrix of predictions made by the ViT model on the original dataset.

Table 3 provides a detailed comparison of class-wise performance of the ViT model on both the original and augmented datasets. From this table, we observe that in the original dataset, the F1-score for the normal class is the highest at 0.9958, followed by Algae at 0.9711, while Major Crack has the lowest score at 0.8646. Whereas, in the augmented dataset F1-score for Stain is the highest (0.9962), following Algae (0.9987), while Major Crack has the lowest (0.9670).

Figure 4 shows the confusion matrix for class-wise predictions of all 792 test samples from the original dataset. We observe that the F1-scores for Minor and Major cracks are the lowest due to the similarity between these two defect classes. Overall, these experimental results provide valuable insights into the performance of various defect classification models and the usability of the dataset.

5 Conclusions and Future works

Recently, many studies have proposed the development of computer vision techniques to automate the current manual visual inspection methods for built environments. However, the lack of comprehensive datasets needed to train efficient and robust defect classification models is one of the primary challenges in practically implementing them. To address these limitations, we developed BD3: Building Defects Detection Dataset, a comprehensive dataset containing 3,965 high-quality annotated images of six common defects and normal conditions collected from diverse building types in Bangalore, India. We benchmarked our dataset using five state-of-the-art image classification models and compared their performance. Our experimental results show that the ViT models trained on BD3 can be useful in classifying defects accurately and deployed into drones or mobile robots to localize the defects, which is the first step when conducting building inspection. It should be noted that a holistic building inspection requires detailed analysis of each defect, such as identifying the severity and depth. As future work, we plan to extend this dataset by segmenting the location of faults in each image and include additional context information, such as wall type and materials, to enable automated segmentation tasks, contributing to more precise building inspections.

References

[1] Eric Bianchi and Matthew Hebdon. 2022. Visual structural inspection datasets. *Automation in Construction* 139 (2022), 104299. https://doi.org/10.1016/j.autcon.2022.104299

[2] Dimitris Dais, İhsan Engin Bal, Eleni Smyrou, and Vasilis Sarhosis. 2021. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. *Automation in Construction* 125 (2021), 103606. https://doi.org/10.1016/j.autcon.2021.103606

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

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs.CV] https://arxiv.org/abs/2010.11929

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

[6] Jingjing Guo, Pengkun Liu, Bo Xiao, Lu Deng, and Qian Wang. 2024. Surface defect detection of civil structures using images: Review from data perspective. *Automation in Construction* 158 (2024), 105186.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs.CV] https://arxiv.org/abs/1512.03385

[8] Zhili He, Wang Chen, Jian Zhang, and Yu-Hsing Wang. 2023. Infrastructure Crack Segmentation: Boundary Guidance Method and Benchmark Dataset. arXiv:2306.09196 [cs.CV] https://arxiv.org/abs/2306.09196

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger (Eds.), Vol. 25. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[10] Shreyas Kulkarni, Shreyas Singh, Dhananjay Balakrishnan, Siddharth Sharma, Saipraneeth Devunuri, and Sai Chowdeswara Rao Korlapati. 2022. CrackSeg9k: A Collection and Benchmark for Crack Segmentation Datasets and Frameworks. arXiv:2208.13054 [cs.CV] https://arxiv.org/abs/2208.13054

[11] Teerath Kumar, Alessandra Mileo, Rob Brennan, and Malika Bendechache. 2023. Image Data Augmentation Approaches: A Comprehensive Survey and Future directions. arXiv:2301.02830 [cs.CV] https://arxiv.org/abs/2301.02830

[12] Shengyuan Li and Xuefeng Zhao. 2018. Convolutional neural networks-based crack detection for real concrete surface. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018*, Vol. 10598. SPIE, 955–961.

[13] Kangcheng Liu, Guidong Yang, Jihan Zhang, Zuoquan Zhao, Xi Chen, and Ben M. Chen. 2022. Datasets and Methods for Boosting Infrastructure Inspection: A Survey on Defect Segmentation and Detection. In *2022 IEEE 17th International Conference on Control & Automation (ICCA)*. 23–30. https://doi.org/10.1109/ICCA54724.2022.9831925

[14] Martin Mundt, Sagnik Majumder, Sreenivas Murali, Panagiotis Panetsos, and Visvanathan Ramesh. 2019. *CODEBRIM: COncrete DEfect BRidge IMage Dataset*. https://doi.org/10.5281/zenodo.2620293

[15] Ç F Özgenel and A Gönenç Sorguç. 2018. Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In *Isarc. proceedings of the international symposium on automation and robotics in construction*, Vol. 35. IAARC Publications, 1–8.

[16] Husein Perez, Joseph H.M. Tah, and Amir Mosavi. 2019. Deep Learning for Detecting Building Defects Using Convolutional Neural Networks. (Aug. 2019). https://doi.org/10.20944/preprints201908.0068.v1

[17] Luis Perez and Jason Wang. 2017. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv:1712.04621 [cs.CV] https://arxiv.org/abs/1712.04621

[18] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2019. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381 [cs.CV] https://arxiv.org/abs/1801.04381

[19] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV] https://arxiv.org/abs/1409.1556

[20] Guidong Yang, Kangcheng Liu, Jihan Zhang, Benyun Zhao, Zuoquan Zhao, Xi Chen, and Ben M. Chen. 2022. Datasets and processing methods for boosting visual inspection of civil infrastructure: A comprehensive review and algorithm comparison for crack classification, segmentation, and detection. *Construction and Building Materials* 356 (2022), 129226. https://doi.org/10.1016/j.conbuildmat.2022.129226

[21] Liang Yang, Bing Li, Wei Li, Zhaoming Liu, Guoyong Yang, and Jizhong Xiao. 2017. A robotic system towards concrete structure spalling and crack database. In *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. 1276–1281. https://doi.org/10.1109/ROBIO.2017.8324593

[22] Suguru Yokoyama and Takashi Matsumoto. 2017. Development of an Automatic Detector of Cracks in Concrete Using Machine Learning. *Procedia Engineering* 171 (2017), 1250–1255. https://doi.org/10.1016/j.proeng.2017.01.418 The 3rd International Conference on Sustainable Civil Engineering Structures and Construction Materials - Sustainable Structures for Future Generations.