Deep Learning based Pothole Monocular Depth Estimation and Segmentation Using 3D Scanner-Derived Depth Maps

RAHMAT ALI and YOUNG-JIN CHA

ABSTRACT

Potholes pose significant safety risks to drivers and cause damage to vehicles. This paper modified a novel approach called the monocular depth estimation and segmentation (modified 3DPredicNet) network [1] to accurately estimate depth maps and segment potholes. To facilitate model training and evaluation, a comprehensive dataset consists of RGB images captured using a DSLR camera and corresponding 3D scan data for generating depth maps. The depth maps derived from the 3D scans are utilized for pothole depth estimation, while masks are used for pothole segmentation. The evaluation results reveal the model's ability to accurately predict and segment potholes in RGB images, achieving a minimum absolute relative error (ARel) of 0.062, square relative error (SRel) of 0.011, and root mean square error (RMSE) of 0.118 when tested on the newly developed dataset. Moreover, when tested on the newly developed dataset, the model demonstrates good pothole segmentation performance, attaining a high mean intersection over union (mIoU) of 81.05. Furthermore, when utilizing the publicly available dataset, the modified 3DPredicNet achieved accurate depth estimation with ARel of 0.093 and SRel of 0.033.

INTRODUCTION

Road infrastructure plays a vital role in our everyday lives, exerting a profound influence on the development and well-being of city dwellers [1]. The quality of pavement holds immense significance in this context. As time progresses, the pavement undergoes distress, encompassing a range of issues including potholes. These deteriorations stem from a multitude of factors, including but not limited to, traffic loads, weather conditions, construction anomalies, subgrade soil properties, and insufficient maintenance measures. Among these concerns, potholes emerge as a particularly vexing problem, arising from a complex interplay of circumstances [2]. When moisture infiltrates beneath the pavement, it undergoes expansion and contraction during freezing and thawing cycles. Consequently, weakened sections cave under the

Rahmat Ali, Ph.D. Candidate, Dept. of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada
Young-Jin Cha, Associate Professor, Dept. of Civil Engineering, University of Manitoba, Winnipeg, MB, Canada

weight of passing vehicles, giving rise to potholes, making driving and pedestrian movement uneasy, exposing vehicles to potential damage, and serious accidents [3]. Therefore, regular inspections are crucial to ensure road safety by promptly identifying and addressing potential hazards.

Despite the drawbacks, many transportation departments still rely on traditional methods for pavement inspection. In recent years, extensive research has been dedicated to the development of deep learning-based methods for damage detection in the field of structural health monitoring (SHM). Cha et al. (2017) proposed a convolutional neural network (CNN)-based approach for crack detection, demonstrating promising outcomes [4]. Subsequently, they extended their methodology to detect multiple types of damage using a faster region-based convolutional neural network [5]. Kang and Cha (2018) employed autonomous unmanned aerial vehicles for concrete damage detection, expanding the applicability of the approach [6]. Building on this achievement, they further refined the method and successfully applied it to the assessment of steel bridges and parkade structures [7, 8]. In addition, other research endeavors have concentrated on investigating diverse methodologies for pixel-level damage detection, such as SDDNet [9] and STRNet [10] for surface damage detection, and IDSNet for subsurface damage detection [11-13]. Moreover, thermography was integrated with deep learning techniques to effectively detect subsurface damage in steel bridges [11], and concrete structures [12], as demonstrated in several recent studies as well as for crack in tunnel structures. Additional studies have been conducted in the field of deep learning, focusing on various applications in different environments. Specifically, one study [14] explored deep learning-based damage detection in complex environments, while other studies [15, 16] investigated deep learning techniques for noise control in noisy environments. Furthermore, researchers developed a dual encoder-decoder-based deep polyp segmentation network specifically designed for colonoscopy images [17]. In a separate study, an unsupervised deep auto-encoder combined with a one-class support vector machine was proposed for damage detection [18]. Lastly, volumetric damage quantification was addressed in multiple studies [19, 20].

Additionally, in a previous study, a novel approach was introduced, utilizing vehicle-mounted dual scanners to capture 3D surface data for effective pothole identification [21]. However, real-time detection capabilities were not achieved with this method. Alternatively, laser systems based on the time-of-flight principle have been employed for pothole detection, although these techniques can be costly and susceptible to environmental lighting variations [22]. Some studies have explored the combination of laser and image-based methods, but practical considerations such as expense and sensitivity to lighting conditions arise. Moreover, considerable research has been dedicated to pothole detection and segmentation, the inclusion of depth mapping is vital for accurately assessing the severity of damage. In this study, a diverse data was gathered using a DSLR camera and a 3D scanner [23] to effectively train the modified 3DPredicNet. The paper is structured as follows: Section 2 provides a comprehensive explanation of the proposed methodology. Section 3 encompasses the analysis and evaluation of the obtained results, shedding light on their significance. Finally, Section 4 concludes the study, offering insights into potential future opportunities of exploration and development.

METHODOLOGY

The proposed method [1] consists of several key steps for predicting pothole depth maps and performing pothole segmentation using RGB images. The overall proposed methodology is presented in Figure 1.

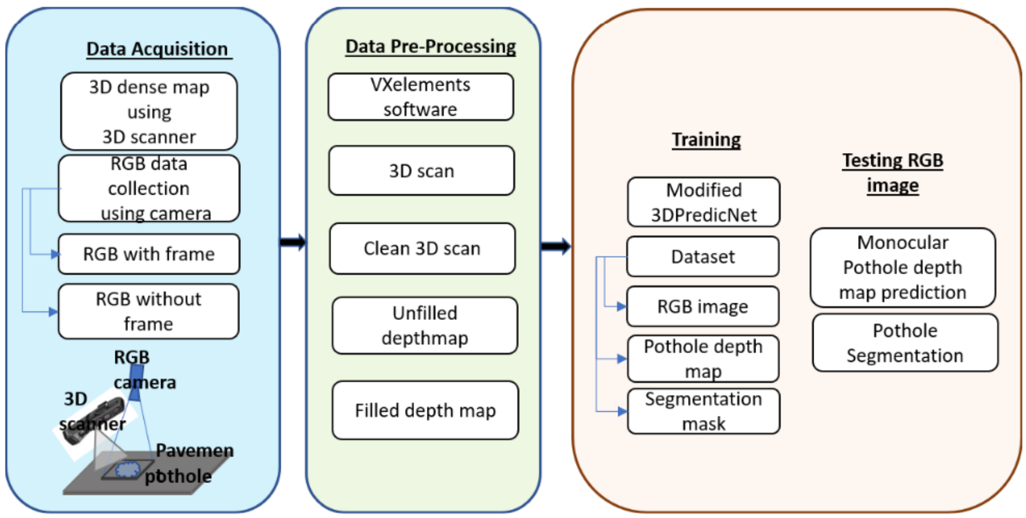

Figure 1. Flowchart of the proposed method

The first step is data acquisition, where a 3D scanner [23] is used to capture a 3D dense map of the environment, while an RGB camera collects RGB data with and without wooden frames. Following data acquisition, the collected data undergoes data pre-processing. The VXelements software is employed for this purpose. The 3D scans obtained from the 3D scanner are processed using software to obtain a clean and refined 3D scan. Additionally, the depth maps derived from the 3D scans are processed to generate depth maps. This pre-processing step ensures that the data is in an optimal form for model training. The next step in the methodology is training a newly developed model called modified 3DPredicNet. This model is trained using two datasets: a newly developed dataset, as shown in Table I, and the Pothole600 dataset, as shown in Table II. These datasets consist of RGB images, pothole depth maps, and segmentation masks.

TABLE I. NEWLY DEVELOPED DATASET

Dataset	Training	Validation	Test	Image size	Total
RGB	300	50	50	1200 × 800	400
Depth map	300	50	50	1200 × 800	400
Mask	300	50	50	1200 × 800	400

TABLE II. POTHOLE600 DATASET

Dataset	Training	Validation	Test	Image size	Total
RGB	240	180	180	400 × 400	600
Depth map	240	180	180	400 × 400	600
Mask	240	180	180	400 × 400	600

The RGB images provide visual information, while the pothole depth maps and segmentation masks serve as ground truth labels for the depth and segmentation predictions, respectively. Through the training process, the modified 3DPredicNet model learns the underlying correlations between RGB images and the corresponding pothole depth maps and segmentation masks. Once the training is completed, the trained modified 3DPredicNet model is ready for testing. In the testing phase, an RGB image is inputted to the model. The model then predicts the monocular pothole depth map, which estimates the depth information of the potholes present in the image. Additionally, the model performs pothole segmentation, accurately identifying and delineating the pothole regions within the image.

Modified 3DPredicNet

The modified 3DPredicNet [1] is specifically designed to predict the depth map and segmentation mask of potholes from a single RGB image. The overall architecture of the network is depicted in Figure 2. It incorporates several key components to effectively capture and utilize the relevant features within the image. The network architecture begins with a convolution block (Conv block) that applies convolutional layers to the input RGB image, resulting in feature maps with reduced spatial dimensions (H/2, W/2, D=16). The feature maps obtained from the convolution block are then passed through the residual intensive convolution modules (RICM) block. The RICM block further enhances the extracted features by leveraging residual connections and intensive convolution operations. This block helps in capturing both low level and higher-level representations and context information. After the RICM block, the feature maps are fed into a depth-wise separable convolution block (DSC).

The DSC block is specifically designed to reduce the number of parameters in the network while maintaining spatial information. It achieves this by separating the convolution process into depth-wise convolutions and point-wise convolutions, thereby optimizing the network's efficiency. To further enhance the network's performance, dual attention modules (DAM) are incorporated. The DAM module helps the network to focus on relevant spatial and channel-wise features by applying attention mechanisms. Both downsampling and upsampling operations are utilized in the network. Downsampling is performed to reduce the spatial dimensions of the feature maps, while upsampling is employed to restore the dimensions to the original size.

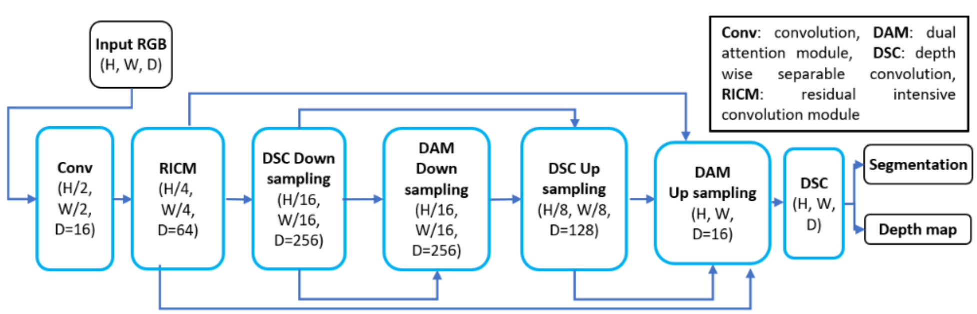

Figure 2. Modified 3DPredicNet

These operations help in capturing multi-scale information and preserving fine details during the prediction process. The network architecture involves the repetition of certain components to ensure effective feature extraction and information flow. Specifically, the feature maps pass through the DAM upsampling and DSC once again, allowing for additional refinement of the features. By incorporating these various components, the modified 3DPredicNet architecture is designed to effectively predict the depth map and segmentation mask of potholes from a single RGB image.

ANALYSIS

In this section, the performance of the proposed modified 3DPredicNet is assessed on both a newly developed pothole dataset as shown in Figure 3 and a publicly available dataset as shown in Figure 4 and 5 [24]. Figure 3 illustrates the input RGB image fed (Figure 3a) into the network, along with its corresponding segmentation (Figure 3b) and depth map (Figure 3c). The evaluation results demonstrate that the proposed method achieves a mean intersection over union (mIoU) of 81.05. Furthermore, the modified 3DPredicNet achieved a minimum absolute relative error of 0.062, a square relative error of 0.011, and a root mean square error of 0.118. These results show the effectiveness and reliability of the proposed methodology in accurately predicting and segmenting potholes in images.

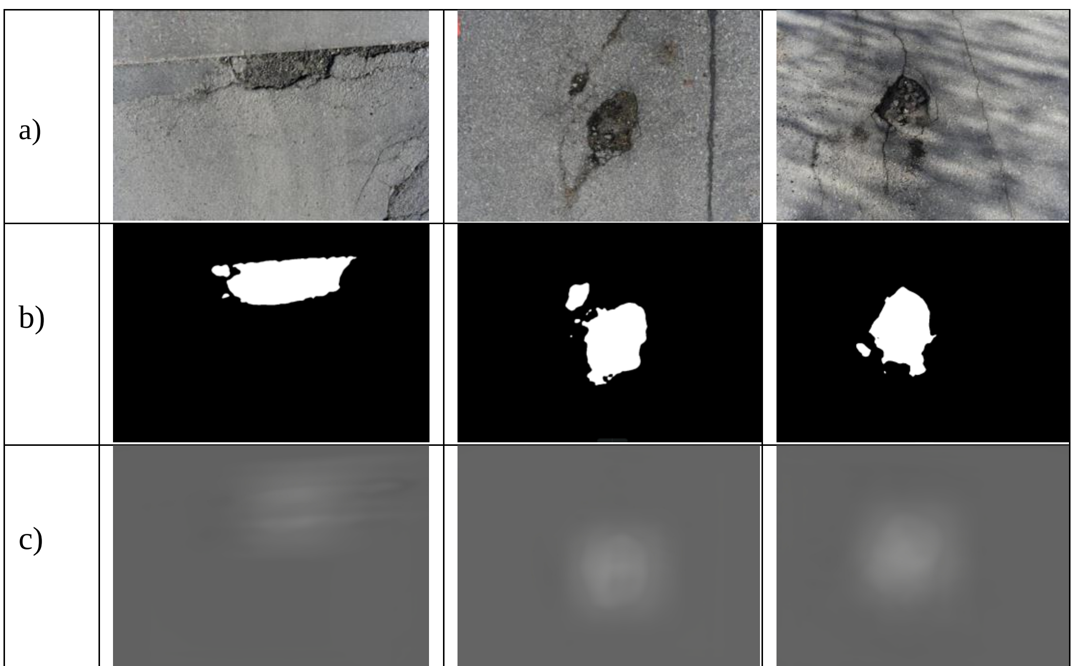

Figure 3. Pothole segmentation and depth map estimation using RGB image input of newly developed dataset

Figure 4 shows the outcomes of pothole segmentation using RGB image inputs (Figure 4a) from the Pothole600 dataset. The ground truth (Figure 4b) and segmentation output (Figure 4c) produced by the model are presented, illustrating the effectiveness of the proposed method. Three distinct RGB images depicting small potholes were fed into the trained model. The evaluation of the segmentation task on RGB input images resulted in an mIoU score of 71.90.

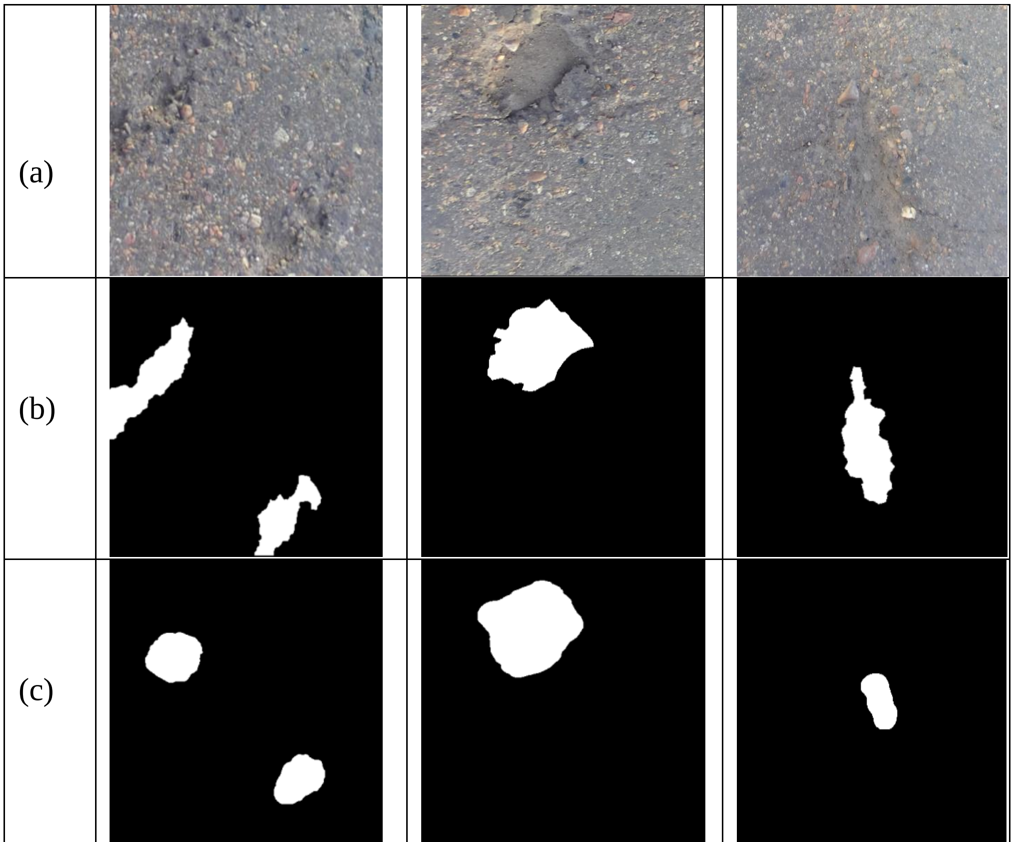

Figure 4. Pothole segmentation using RGB image input of Pothole600 dataset

Figure 5 presents the results of depth map estimation achieved by utilizing RGB images (Figure 5a) trained on the Pothole600 dataset. The depth maps generated by the model are compared with their corresponding ground truth data. The evaluation metrics used to assess the accuracy of the depth map estimation include an Absolute Relative Difference (ARel) of 0.093, a Scale Relative Difference (SRel) of 0.033, a Root Mean Squared Error (RMSE) of 0.191, and a logarithmic Root Mean Squared Error (RMSElog) of 0.0641. The presented results provide quantitative measures indicating the model's performance in estimating accurate depth maps. These findings validate the effectiveness of the proposed method for depth map estimation using RGB images from the Pothole600 dataset.

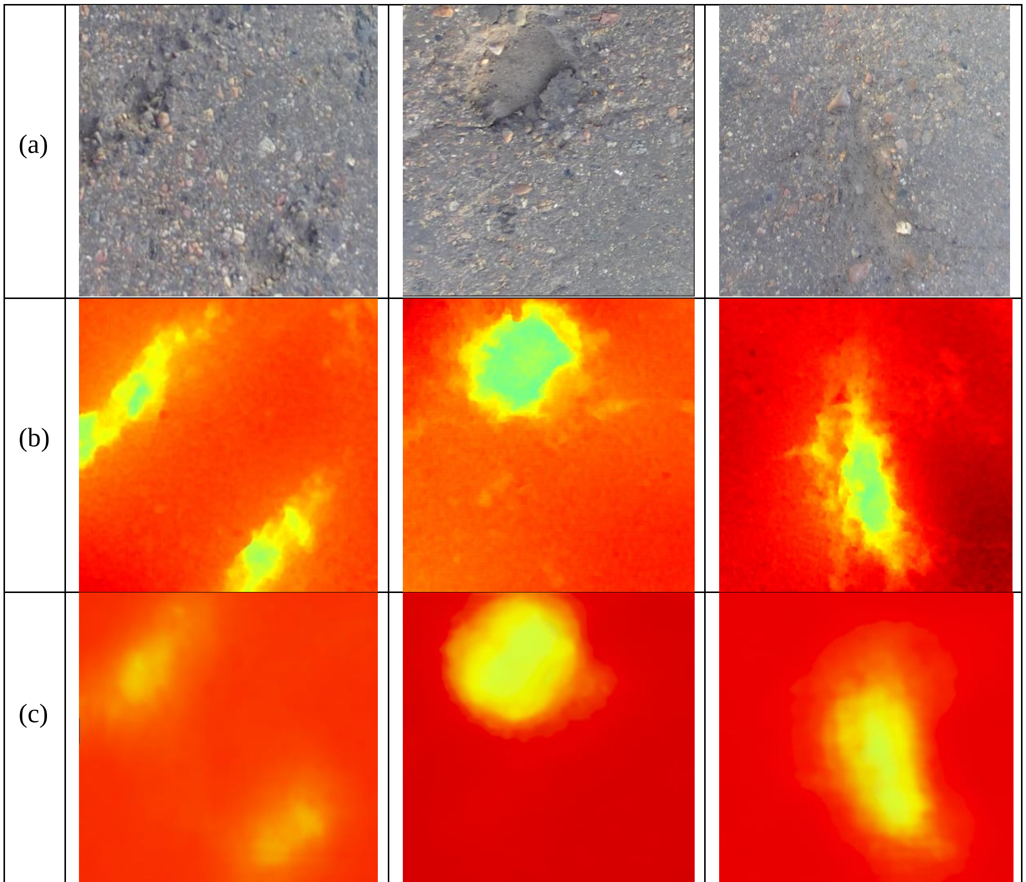

Figure 5. Depth map estimation using RGB image input from the Pothole600 dataset.

CONCLUSION

The modified 3DPredicNet [1] was tested further to measure its performance on additional data. The model was tested on both a newly developed pothole dataset and a publicly available dataset, demonstrating its effectiveness in various scenarios. The evaluation results revealed that the modified 3DPredicNet with 2.79 million paramters achieved mIoU of 81.05, indicating accurate segmentation performance when tested on newly developed dataset. Moreover, the model achieved a minimum ARel of 0.062, SRel of 0.011, and RMSE of 0.118. These results indicate the model's ability to accurately predict and segment potholes in RGB images. The evaluation of the depth map estimation using RGB images trained on the Pothole600 dataset further showcased the capabilities of the modified 3DPredicNet. The model achieved a ARel of 0.093 and SRel of 0.033 demonstrating accurate depth estimation.

REFERENCES

1. Ali. R., and Cha. Y.J., (2023) Monocular computer vision-based pothole segmentation and 3D volume prediction using advanced deep learning. Automation in construction (under review).
2. U.S. Department of transportation Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13092/13092.pdf

3. Miller, J. S., & Bellinger, W. Y. (2003). Distress identification manual for the long-term pavement performance program (No. FHWA-RD-03-031). United States. Federal Highway Administration. Office of Infrastructure Research and Development.
4. Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, *32*(5), 361-378.
5. Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering*, *33*(9), 731-747.
6. Kang, D., & Cha, Y. J. (2018). Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. *Computer-Aided Civil and Infrastructure Engineering*, *33*(10), 885-902.
7. Ali, R., Kang, D., Suh, G., & Cha, Y. J. (2021). Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures. *Automation in Construction*, *130*, 103831.
8. Waqas, A., Kang, D., Cha, Y. J. (2023), Deep learning–based obstacle-avoiding autonomous UAVs with fiducial marker–based localization for structural health monitoring, Structural Health Monitoring, SAGE.
9. Choi, W., & Cha, Y. J. (2019). SDDNet: Real-time crack segmentation. *IEEE Transactions on Industrial Electronics*, *67*(9), 8016-8025.
10. Kang, D. H., & Cha, Y. J. (2022). Efficient attention-based deep encoder and decoder for automatic crack segmentation. Structural Health Monitoring, 21(5), 2190-2205.
11. Ali, R., & Cha, Y. J. (2019). Subsurface damage detection of a steel bridge using deep learning and uncooled micro-bolometer. *Construction and Building Materials*, *226*, 376-387.
12. Ali, R., Zeng, J., & Cha, Y. J. (2020, April). Deep learning-based crack detection in a concrete tunnel structure using multispectral dynamic imaging. In *Smart Structures and NDE for Industry 4.0, Smart Cities, and Energy Systems* (Vol. 11382, pp. 12-19). SPIE.
13. Ali, R., & Cha, Y. J. (2022). Attention-based generative adversarial network with internal damage segmentation using thermography. *Automation in Construction*, *141*, 104412.
14. Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y. J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. *Automation in Construction*, *118*, 103291.
15. Mostafavi, A., & Cha, Y. J. (2023). Deep learning-based active noise control on construction sites. Automation in Construction, 151, 104885.
16. Cha, Y. J., Mostafavi, A., & Benipal, S. S. (2023). DNoiseNet: Deep learning-based feedback active noise control in various noisy environments. Engineering Applications of Artificial Intelligence, 121, 105971.
17. Lewis, J., Cha, Y. J., & Kim, J. (2023). Dual encoder–decoder-based deep polyp segmentation network for colonoscopy images. Scientific Reports, 13(1), 1183.
18. Wang, Z., & Cha, Y. J. (2021). Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage. Structural Health Monitoring, 20(1), 406-425.
19. Beckman, G. H., Polyzois, D., & Cha, Y. J. (2019). Deep learning-based automatic volumetric damage quantification using depth camera. Automation in Construction, 99, 114-124.
20. Yuan, C., Xiong, B., Li, X., Sang, X., & Kong, Q. (2022). A novel intelligent inspection robot with deep stereo vision for three-dimensional concrete damage detection and quantification. *Structural Health Monitoring*, *21*(3), 788-802.
21. Tsai, Y. C., & Chatterjee, A. (2018). Pothole detection and classification using 3D technology and watershed method. *Journal of Computing in Civil Engineering*, *32*(2), 04017078.
22. Mathavan, S., Kamal, K., & Rahman, M. (2015). A review of three-dimensional imaging technologies for pavement distress detection and measurements. IEEE Transactions on Intelligent Transportation Systems, 16(5), 2353-2362.
23. Go! SCAN 3D. Available at: https://www.creaform3d.com/en/handheld-portable-3d-scanner-goscan-3d.
24. Pothole600 dataset Available athttps://sites.google.com/view/pothole-600/dataset.