International Journal of GEOMATE, Aug., 2025 Vol.29, Issue 132, pp.172-181
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2025.132.4897
Geotechnique, Construction Materials and Environment

DAMAGE DETECTION IN HISTORICAL STRUCTURES USING FASTER REGION-BASED CONVOLUTIONAL NEURAL NETWORK

Phromphat Thansirichaisree[1], *Onpriya Klincharoen[1], Apichat Buatik[1], Sireethorn Potipipit[1] and Nakhorn Poovarodom[1]

[1]Thammasat Research Unit in Infrastructure Inspection and Monitoring, Repair and Strengthening (IIMRAS), Faculty of Engineering, Thammasat School of Engineering, Thammasat University Rangsit, Klong Luang, Pathumthani 12120, Thailand

*Corresponding Author, Received: 31 Jan. 2025, Revised: 02 April 2025, Accepted: 03 April 2025

ABSTRACT: Historic sites in Thailand, many of which are recognized by UNESCO, hold immense historical and cultural significance. However, these invaluable structures are increasingly at risk of deterioration due to aging, environmental factors, vibrations, and natural disasters such as floods. Cracks and structural damage are common challenges, further aggravated by the inefficiencies and limitations of traditional inspection methods, which are labor-intensive, prone to human error, and often restricted by the inaccessibility of certain areas. To address these challenges, this study proposes an automated damage detection system specifically designed for heritage masonry structures using Faster Region Convolutional Neural Networks (FRCNN). The system efficiently detects and localizes structural damage with high precision, providing a reliable and cost-effective alternative to manual inspections. The system's performance was evaluated using five Faster R-CNN models with different backbone architectures: VGG16, VGG19, ResNet50, ResNet101, and ResNet152. In this comparison, the ResNet152 model achieved the highest accuracy of 81.29% and a recall value of 81.13%, demonstrating its capability to effectively detect cracks in historical structures.

Keywords: Object Detection, Automatic Detection, Masonry Structures, Faster Region Convolutional Networks

1. INTRODUCTION

Thailand's cultural heritage sites, such as Wat Si Phichit Kirati Kanlayaram in Sukhothai (Fig. 1), hold immense value in preserving the nation's historical and cultural identity. However, these structures are increasingly vulnerable to deterioration due to aging, human activities, and natural disasters such as floods and earthquakes. The ability to detect and monitor structural damage is crucial for maintaining their integrity and long-term conservation.

Currently, visual inspection is the primary method used to assess the condition of historical structures. However, this approach is time-consuming, costly, and susceptible to human error. Moreover, accessing certain sites, particularly tall structures like stupas, poses significant challenges. Traditional defect detection methods often rely on handcrafted features, which have limited accuracy and adaptability. These methods depend heavily on the expertise of researchers and struggle with complex images where distinguishing between defects and the background becomes challenging.

Most traditional damage detection techniques involve two main steps: (1) feature extraction using techniques like gray-level co-occurrence matrices (GLCM) [1], edge detection [2], multi-feature approaches [3], and principal component analysis (PCA) [4]; and (2) pattern recognition through classifiers such as Support Vector Machines (SVM)

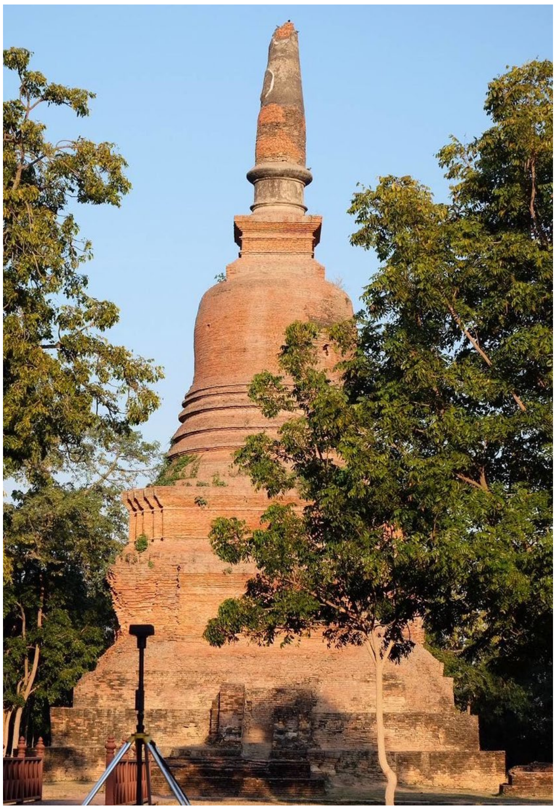

Fig. 1 Wat Si Phichit Kirati Kanlayaram, a temple in the historic province of Sukhothai, Thailand

[5] and Neural Networks [6]. For example, Shahid Kabir et al. (2010) [7] employed GLCM features and Artificial Neural Networks (ANN) to identify damages caused by alkali-aggregate reactions (AAR). Similarly, Abdul-Qadir (2003) [8] applied edge detection techniques for crack identification in bridges. However, edge detection algorithms often perform poorly on noisy images, as they are optimized for noise-free conditions [9]. Nishikawa (2012) [10] proposed sequential image filtering for crack detection in concrete, and German et al. (2012) [11] used machine vision for post-earthquake safety assessments, with a focus limited to concrete spalling. Yuem (2015) [12] developed an automated crack detection system for bridge inspections, while Cha (2016) [13] proposed a vision-based method to detect loosened bolts using Hough transform features and a Support Vector Machine classifier. Similarly, Zalama (2014) [14] utilized Gabor filters for feature extraction in road pavement damage detection. While these methods have been enhanced in studies such as those by Liao (2016) [15] and Chen (2012) [16], they often require computationally intensive pre- and post-processing steps, limiting their efficiency. To address these challenges, deep learning methods have emerged as a superior alternative.

Deep learning has been increasingly adopted for crack detection due to its ability to learn hierarchical features directly from images. Convolutional Neural Networks (CNNs) have demonstrated high accuracy in crack detection tasks. For instance, Cha (2017) [17] developed a CNN-based system for detecting cracks in concrete structures, while Zhang (2016) [18] applied deep convolutional networks to identify road cracks using smartphone images. However, while CNNs excel in multi-class classification, traditional sliding window techniques for damage localization remain inefficient due to variations in image size. To overcome these limitations, region-based CNNs (R-CNNs) were introduced. Girshick (2014) [19] pioneered R-CNN, which combined CNN for feature extraction with SVMs for classification and localization. However, R-CNN was computationally slow. He (2014) [20] improved this with Spatial Pyramid Pooling (SPP-net), increasing speed while maintaining accuracy. Girshick (2015) [21] then introduced Fast R-CNN, which streamlined feature extraction but still relied on selective search for object proposals, causing inefficiencies. Finally, Ren (2016) [22] developed Faster R-CNN, which integrated a Region Proposal Network (RPN) to significantly reduce computational costs while maintaining high detection accuracy. Hacıefendioğlu (2022) [23] further applied Faster R-CNN to crack detection on concrete roads under various environmental conditions, analyzing the effects of different factors and found that weather, illumination levels, distance, and image height significantly influenced crack detection accuracy.

Damage detection in historical structures presents additional challenges due to the delicate nature of materials and the varying surface textures of historical masonry. Research in this domain remains relatively limited compared to concrete-based crack detection studies. Wang (2019) [24] applied Faster R-CNN to historical brick structures but found that certain types of damage remained undetectable. More recently, Karimi (2024) [25] compared YOLOv5's performance in detecting cracks on historical materials such as stone, brick, cob, and tiles. The study concluded that YOLOv5 performed best on concrete, and background characteristics significantly influenced detection accuracy. In some cases, material joints were misclassified as cracks.

The objectives and motivations of this study are to leverage deep learning capabilities for accurately and efficiently detecting structural damage. Given the unique characteristics of historical surfaces, addressing these challenges is crucial for improving damage detection accuracy. Additionally, no prior research has specifically focused on Thailand's historical structures. To overcome these challenges, this study employs Faster R-CNN to enhance crack detection in Thailand's historical structures. The dataset for this study consists of images collected using DSLR cameras from various historical sites in Thailand. The proposed system enables near real-time damage detection, which is expected to be an efficient and reliable alternative to traditional inspection methods.

The paper is organized as follows: Section 2, research significance; Section 3, details the proposed methodology; Section 4, describes data generation and experimental results; Section 5, concludes and discusses the study; and Section 6, provides acknowledgments.

2. RESEARCH SIGNIFICANCE

This study presents an approach for detecting cracks in heritage masonry structures using Faster R-CNN, addressing a critical need in cultural heritage preservation. The proposed method enhances the reliability of structural assessments by providing a scalable, accurate, and cost-effective solution for automated structural damage detection. Its practical application supports long-term conservation efforts and advances the state-of-the-art in inspection technologies for cultural heritage preservation.

3. PROPOSED METHODOLOGY

The framework of the damage detection system is depicted in Fig. 2. It comprises three primary components: (1) image capture via DSLR, (2) training the Faster R-CNN model, and (3) testing the system. The system's output consists of images with damage regions clearly identified and localized.

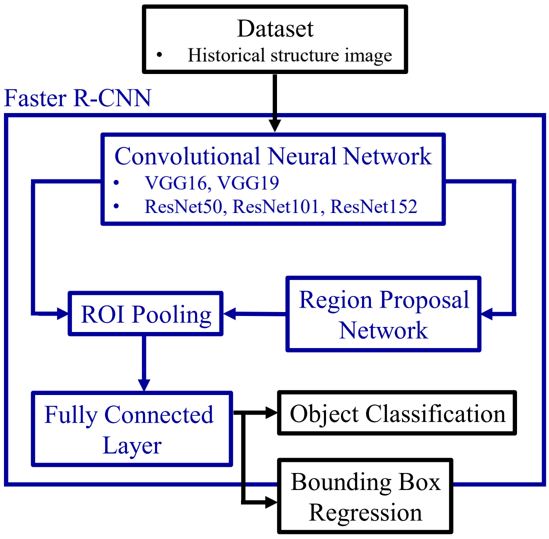

Fig. 2 The framework of the proposed system

3.1 Image Acquisition

In this study, images were captured using a DSLR camera to document the surrounding environment without approaching too close to the object of interest, as shown in Fig. 3. The original images were taken at a high resolution of 5184×3456 pixels. To ensure the dataset encompassed a variety of real-world conditions, images were captured under diverse lighting scenarios, including sunny and shaded environments, simulating challenges encountered during on-site inspections. The dataset includes images collected from multiple heritage sites in Thailand, featuring a range of materials and structural conditions, such as bricks and mixed surfaces. Each image was resized to a standardized 2K resolution (2560×1440 pixels) to ensure uniformity across the dataset. This 2K resolution is sufficient to provide high-quality data for analysis while maintaining computational efficiency and conserving computer resources. Additionally, bounding boxes were manually annotated to identify regions of interest, which were critical for training and evaluating the detection model. This carefully curated and diverse dataset serves as a robust foundation for achieving reliable and accurate damage detection outcomes.

3.2 Faster Region-based Convolutional Neural Network (Faster R-CNN)

The architecture of Faster R-CNN is illustrated in Fig. 4. In this model, the CNN's feature extraction process is utilized jointly by both the Region Proposal Network (RPN) and the Fast R-CNN module. A key advancement of Faster R-CNN compared to Fast R-CNN lies in its replacement of the selective search algorithm with the RPN, significantly enhancing processing speed and reducing computational demands. Moreover, the convolutional feature maps

Fig. 3 Sample images of structures captured using a DSLR camera

are shared between the RPN and the detection network, making the RPN computationally efficient and nearly cost-free.

3.2.1 Convolutional Neural Network

The damage detection system utilizing Faster R-CNN comprises two primary components: Fast R-CNN and the Region Proposal Network (RPN). Both components share the same Convolutional Neural Network (CNN) architecture. CNNs, a type of multilayer feedforward artificial neural network, excel at solving complex real-world problems by learning hierarchical features from input data. In this study, the CNN architectures evaluated include VGG16, VGG19, ResNet50, ResNet101, and ResNet152. VGG16 and VGG19 [26] are deep convolutional neural networks characterized by their simplicity and uniform architecture, with 16 and 19 layers, respectively. Both architectures consist of convolutional layers for feature extraction (13 in VGG16 and 16 in VGG19) followed by three fully connected (FC) layers for classification. ResNet50, ResNet101, and ResNet152 [27] belong to the Residual Networks (ResNet) family, which addresses the vanishing gradient problem through the use of skip connections, also known as residual connections. These connections facilitate the smooth flow of gradients throughout the network, enabling effective training in very deep architectures. ResNet50, ResNet101, and ResNet152 contain 50, 101, and 152 layers, respectively. Each ResNet architecture begins with a convolutional layer, followed by bottleneck blocks—comprising three convolutional layers each—and concludes with a fully connected layer for classification. Specifically, ResNet50 includes 16 bottleneck blocks, ResNet101 has 33, and ResNet152 incorporates 50 bottleneck blocks.

3.2.2 Region Proposal Network (RPN)

A Region Proposal Network (RPN) is a convolutional network that uses fully convolutional layers to generate object proposals in the form of rectangular bounding boxes from input images. The process begins with extracting feature maps using a CNN backbone. A sliding window is then applied

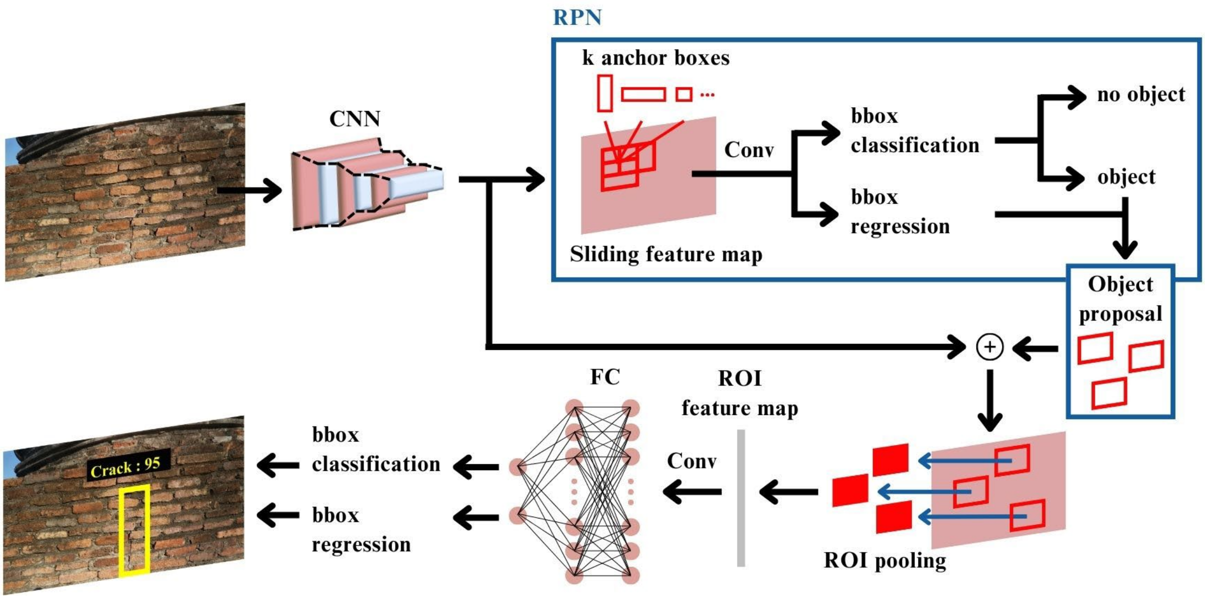

Fig. 4 The architecture of Faster R-CNN (The top portion of the figure represents region proposal network while the bottom portion represents Fast R-CNN)

across these feature maps, where each window is processed by a convolutional layer with ReLU activation. The resulting features are subsequently fed into a SoftMax layer and a set of bounding box regressors. The SoftMax layer determines whether each region contains an object or represents background, while the regressors adjust the bounding box coordinates for greater precision. These bounding boxes, referred to as anchors, are generated by combining various aspect ratios (e.g., 1:1, 2:1, and 1:2) with different scales (e.g., 128x128, 256x256, and 512x512). The RPN generates region proposals with associated probabilities and refined bounding box coordinates, which are subsequently passed to the detection network for further classification and precise localization. To provide a clearer understanding of the RPN architecture, Table 1 presents the specifications of the RPN layers used in this study, which are implemented across five backbone models: VGG16, VGG19, ResNet50, ResNet101, and ResNet152.

Table 1. RPN layer specifications

Layer Name	Filter size	Depth	stride
conv1	3x3	64	2
conv2_x	3x3	128	2
Conv3_x	3x3	256	2
Conv4_x	3x3	512	2
rpn_conv	3x3	512	1
rpn_cls	1x1	9	1
rpn_bbox	1x1	36	1

3.2.3 Fast Region-based Convolutional Neural Network (Fast R-CNN)

The Fast R-CNN processes object proposals generated by the RPN by using a CNN to extract feature maps from these proposals, which correspond to specific regions of the original input image. The region proposals are mapped onto the feature maps, with the extracted features referred to as Regions of Interest (RoIs). During the RoI pooling process, max-pooling is applied to the RoIs to produce fixed-size feature vectors for each region. These feature vectors are then passed through fully connected (FC) layers, followed by a softmax layer for classification. The softmax layer computes the probability of damage within the image, while a regression layer refines the bounding box coordinates and dimensions for precise localization.

3.2.4 Loss Function and Optimization

During the training process, loss functions are used to compute errors for different components of the model. Smooth L1 Loss (L_{loss}) is applied for regression loss in both the Region Proposal Network (RPN) and the classifier, as defined in Equation (1). For classification loss in the RPN, Binary Cross Entropy Loss (L_{BCE}) is used, as defined in Equation (2). The classifier utilizes Categorical Cross Entropy Loss (L_{CCE}) for classification loss, as defined in Equation (3). Additionally, Mean Absolute Error (MAE) is employed to compile the entire model, as defined in Equation (4). The computed losses are backpropagated through the network, allowing iterative parameter updates to enhance model performance.

$$L_{loss}(\Delta) = \begin{cases} 0.5 \text{ x } (\Delta)^2 & \text{if } |\Delta| < 1 \\ |\Delta| - 0.5 & \text{if } |\Delta| > 1 \end{cases} \quad (1)$$

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left(\text{y}\log(y_{pred}) + (1-\text{y})\log(1-y_{pred})\right) \quad (2)$$

$$L_{CCE} = -\frac{1}{N}\sum_{i=1}^{N} \text{ylog}\,(y_{pred}) \quad (3)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y - y_{pred}| \quad (4)$$

$\Delta = y - y_{pred}$, where N is number of samples in the dataset, y is true label (either 0 or 1) and y_{pred} is predicted probability (between 0 and 1). The optimization process leverages the Adam optimizer [28], which incorporates bias correction to stabilize gradient updates and improve convergence speed toward the optimal solution.

3.2.5 Intersection over Union (IoU)

The Intersection over Union (IoU) metric is employed to evaluate target detection performance by measuring the proximity of the predicted bounding box to the ground-truth bounding box. The area of overlap is defined as the intersection between the ground-truth bounding box and the predicted bounding box, while the area of union represents the combined area covered by both bounding boxes. The IoU value is computed as described in Eq. (5).

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \quad (5)$$

3.2.6 Confusion matrix

The performance of the proposed method was assessed using precision and recall metrics, as defined in Eq. (6) and (7). These metrics were calculated based on True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True Positives (TP) refers to Instances where the model correctly identifies an image as containing a crack, and the image indeed contains a crack. True Negatives (TN) refers to Instances where the model correctly identifies an image as not containing a crack, and the image indeed does not contain a crack. False Positives (FP) refers to Instances where the model incorrectly identifies an image as containing a crack, but the image does not actually contain one. False Negatives (FN) refers to Instances where the model fails to identify an image as containing a crack, even though the image actually contains one. False Negatives are particularly concerning, as they may result in the oversight of significant structural damage, posing potential risks.

4. EXPERIMENTS AND RESULTS

4.1 Training dataset

A total of 598 images from various temples, including Wat Mahathat Ayutthaya, Wat Mahathat Sukhothai, Wat Son Khao, Wat Si Chum, and Wat Sri Sawai, were collected to form the training dataset. The dataset was divided into a training set of 478 images (80%) and a validation set of 100 images (20%). For testing purposes, 68 images from Wat Si Phichit Kirati Kanlayaram were gathered, as detailed in Table 2. All images were captured using a DSLR camera at a resolution of 5184×3456 pixels. The training images were resized to a uniform 2K resolution (2560×1440 pixels), and bounding boxes were annotated on each image, as illustrated in Fig. 5.

Table 2. Summary of the crack image dataset

Task	Total of images	Training dataset	Validation dataset
Training	598 (100%)	478 (80%)	120 (20%)
Testing	68	-	-

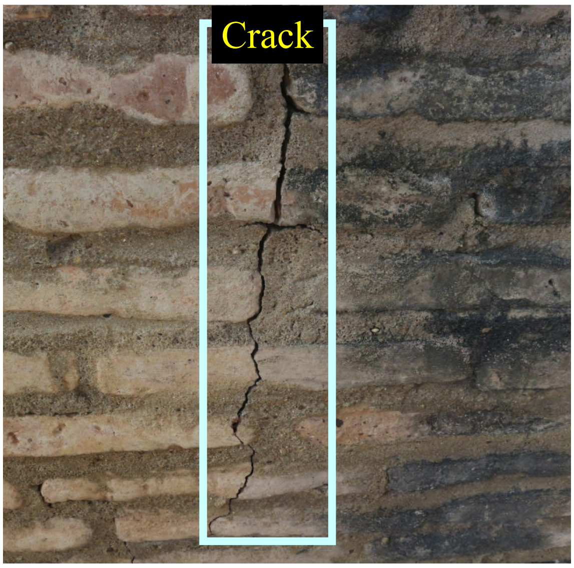

Fig. 5 Sample bounding boxes on images for training

4.2 Training models

The dataset was partitioned into three subsets: training, validation, and testing. The training process spanned 70 epochs, with each epoch comprising 1,000 iterations and a learning rate of 0.00001. The CNN backbones evaluated in this study included VGG16, VGG19, ResNet50, ResNet101, and ResNet152. The training RPN classification accuracy graphs for all five models converge toward 1, indicating strong performance, as illustrated in Fig. 6. Meanwhile, the training loss graphs for all five

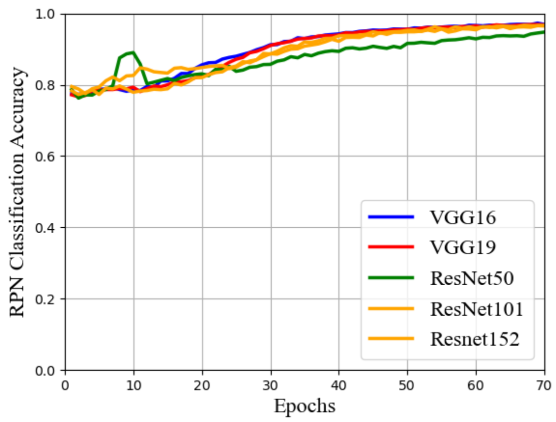

Fig. 6 The training RPN classification accuracy graph

models consistently approach 0, demonstrating effective learning capabilities. However, the RPN classification loss does not converge to 0 as effectively as the other losses, suggesting a reduced ability to predict object presence in certain regions, as illustrated in Fig. 7.

4.3 Crack Detection

The CNN backbones evaluated in this study included VGG16, VGG19, ResNet50, ResNet101, and ResNet152. Among them, ResNet152 outperforms the others, demonstrating superior accuracy and efficiency in detecting structural damage. The testing results for these models are illustrated in Fig. 8. The accuracy, mean IoU, precision, and recall metrics for all five backbones indicate that ResNet152 achieves the highest performance, with an accuracy of 81.29, a mean IoU of 55.91, a precision of 63.24, and a recall of 81.13. These results, which are the highest among all models, are summarized in Table 3.

Table 3. Summary of the confusion matrix results for the testing dataset

Model	Accuracy	Average IoU	precision	recall
VGG16	10.38	3.29	1.47	1.64
VGG19	9.06	4.09	1.47	1.61
ResNet50	29.15	10.07	7.35	9.80
ResNet101	59.86	26.01	23.53	39.02
ResNet152	**81.29**	**55.91**	**63.24**	**81.13**

5. CONCLUSION AND DISCUSSIONS

This study evaluates five Faster R-CNN models with different backbone architectures: VGG16, VGG19, ResNet50, ResNet101, and ResNet152. The accuracy and loss graphs exhibit similar trends across all models. The training RPN classification accuracy graphs for all five models converge toward 1, indicating strong performance, while the training loss graphs consistently approach 0, demonstrating effective learning capabilities. However, the RPN classification loss does not converge to 0 as effectively as the other losses. Among all models, the RPN classification loss of ResNet152 achieves the closest convergence to 0, yet still indicates a reduced ability to predict object presence in certain regions. This issue may stem from challenges in data collection, as the complex and uneven surfaces of historical structures pose significant difficulties in achieving optimal detection accuracy and reliability, particularly in the real world. These challenges are worsened by variations in lighting conditions, textured surfaces, and obstructions like debris or dirt, all of which hinder detection performance. To improve the system's accuracy and applicability in practical scenarios, expanding the training dataset is essential, as existing datasets for historical structures remain limited. Additionally, each structure exhibits unique variations in materials and construction methods, introducing further complexity and necessitating data collection from multiple locations to capture diverse structural characteristics and environmental conditions.

The testing results indicate that ResNet152 outperforms the other models in crack detection. However, some areas remain undetected, with certain cracks falling outside the rectangular frame, as shown in Figure 7, Column 4, Row 6, suggesting that its performance is still insufficient for practical implementation. ResNet152 achieves the highest accuracy, mean IoU, precision, and recall values of 81.29, 55.91, 63.24, and 81.13, respectively. Despite its superior performance, the mean IoU and precision remain relatively low, suggesting that further improvements in model architecture are required to enhance computational efficiency and detection accuracy. Recent studies [29] has proposed a new methodology, demonstrating that incorporating an improved attention-based backbone structure can enhance precision, leading to more reliable crack detection results. By implementing these improvements, the system's capability to detect cracks in historical structures is expected to be significantly enhanced, improving its reliability, accuracy, and practicality for conservation and restoration efforts.

Future research will focus on improving crack detection in historic structures by expanding the dataset with contributions from multiple sources and environments, including images captured in low-light conditions, while also incorporating a broader range of materials. Additionally, we plan to refine anchor box sizes to enhance detection precision.

Furthermore, an enhanced Faster R-CNN model with attention will be developed to improve both detection accuracy and on-site inspection efficiency. Prior research, such as [29], has shown that incorporating attention mechanisms can enhance crack localization accuracy. Additionally, studies like [30] have demonstrated that YOLO achieves significantly higher processing speeds than Faster R-CNN but at the cost of lower detection accuracy. While manual inspection remains a widely used method, it is often time-consuming and prone to human error, particularly in hard-to-access locations. Previous studies, such as [31][32], have highlighted the advantages of automated methods in addressing these challenges—particularly in reducing inspection time, minimizing human error, and improving accessibility to difficult-to-reach areas. Given these considerations, this study aims to further evaluate the effectiveness of different AI-driven models. A comparative analysis of Faster R-CNN, YOLO, and an attention-enhanced Faster R-CNN will be conducted to assess their performance in terms of detection accuracy, processing speed, and real-world applicability. These improvements are expected to enhance the system's resilience and reliability, ultimately supporting the long-term preservation of historic structures.

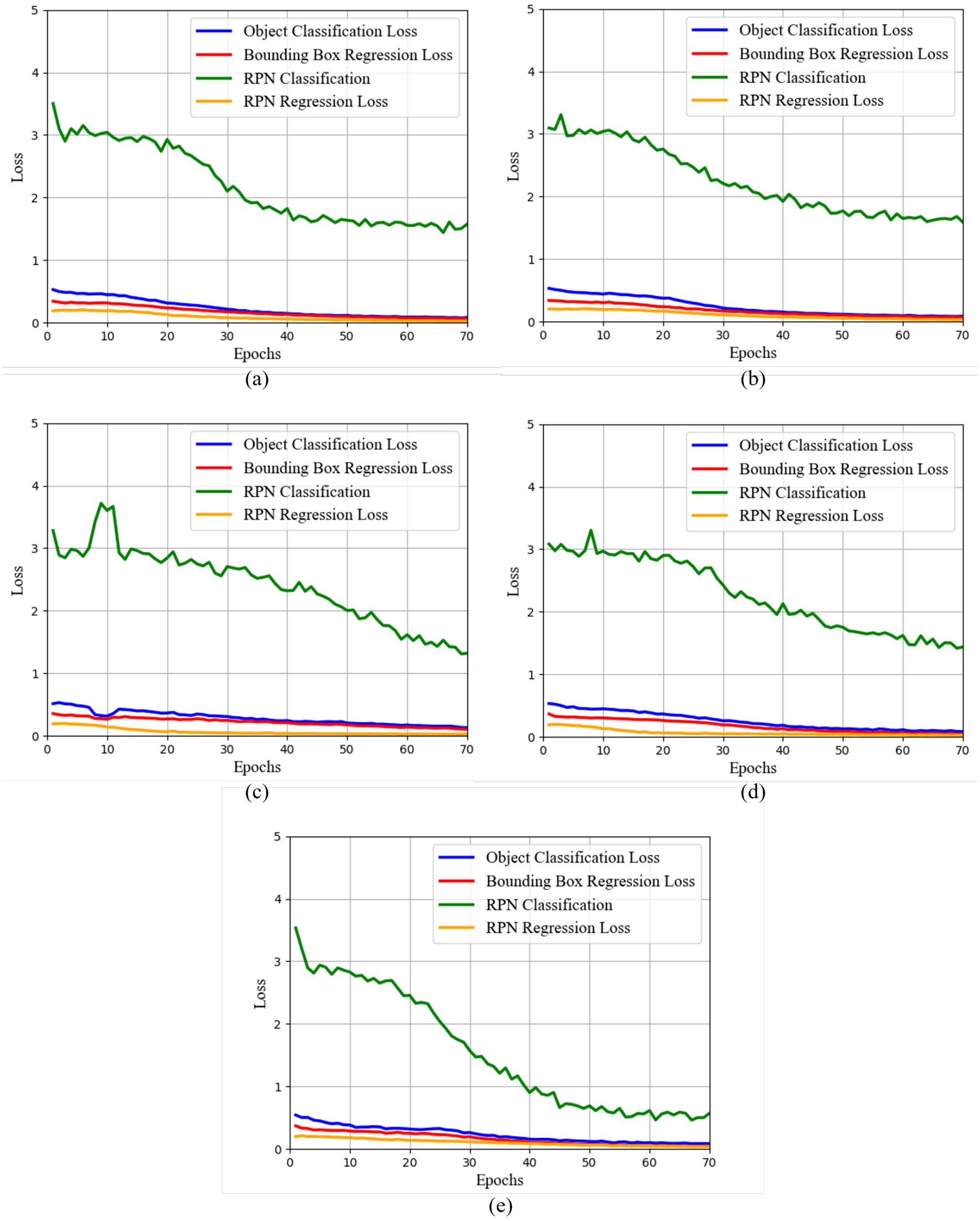

Fig. 7 Training loss graphs for the proposed system using (a) VGG16, (b) VGG19, (c) ResNet50, (d) ResNet101, and (e) ResNet152

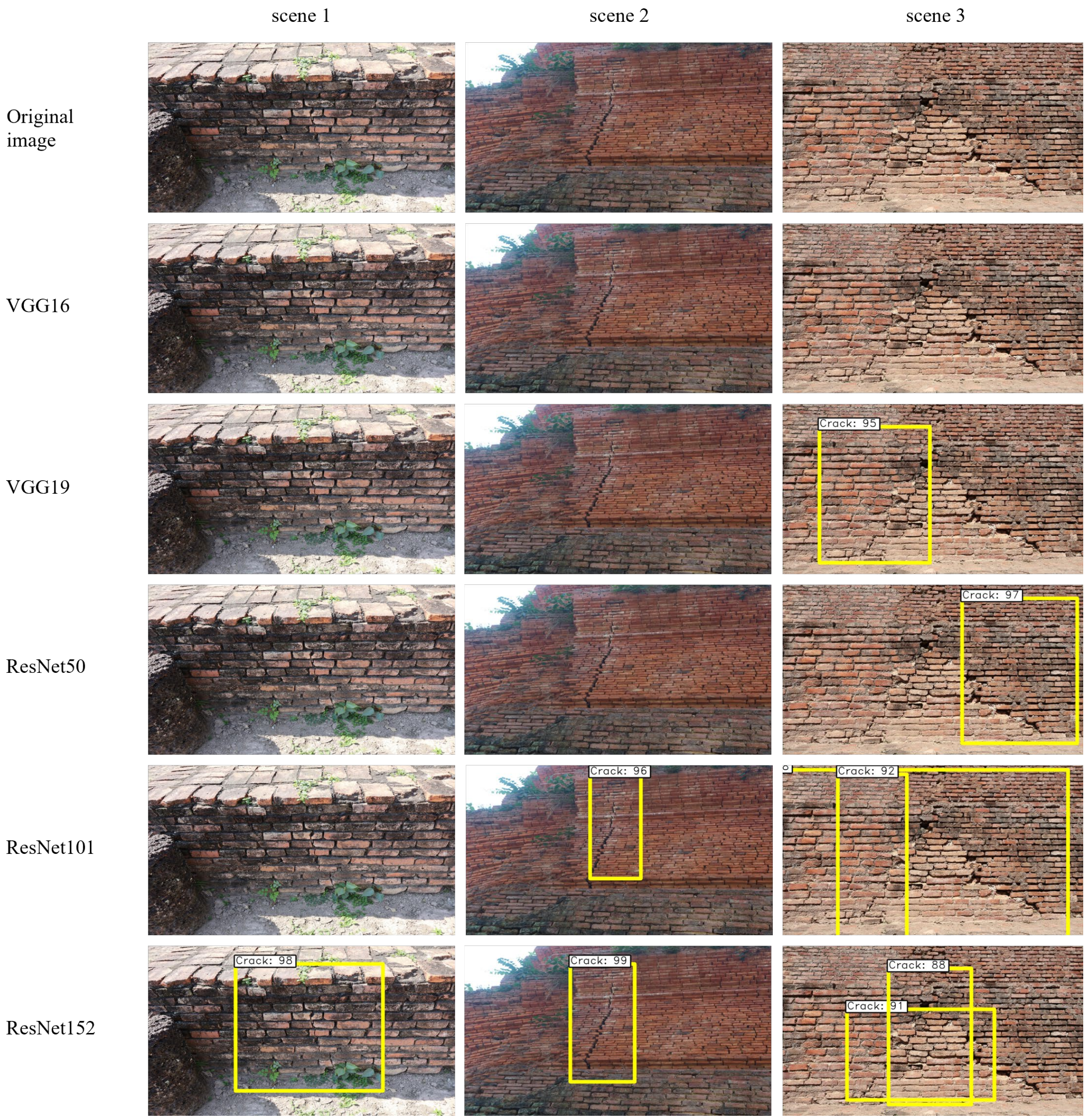

Fig. 8 The results of the testing data of the proposed system

6. ACKNOWLEDGMENTS

This research was funded by the Thammasat School of Engineering (TSE), Thammasat University, Thailand, which contributed to the successful completion of this research project.

7. REFERENCES

[1] Hu, Y., Zhao, C. X., and Wang, H. N., Automatic pavement crack detection using texture and shape descriptors. IETE Technical Review, Vol. 27, Issue 5, 2010, pp. 398-405.

[2] Abdel-Qader, I., Abudayyeh, O., and Kelly, M. E., Analysis of edge-detection techniques for crack identification in bridges. Journal of Computing in Civil Engineering, Vol. 17, Issue 4, 2003, pp.255-263.

[3] Prasanna, P., Dana, K. J., Gucunski, N., Basily, B. B., La, H. M., Lim, R. S., and Parvardeh, H., Automated crack detection on concrete bridges. IEEE Transactions on automation science and engineering, Vol. 13, Issue 2, 2016, pp.591-599.

[4] Abdel-Qader, I., Pashaie-Rad, S., Abudayyeh, O., and Yehia, S., PCA-based algorithm for unsupervised bridge crack detection. Advances in Engineering Software, Vol. 37, Issue 12, 2006, pp.771-778.

[5] Suykens, J. A., and Vandewalle, J., Least squares support vector machine classifiers.

Neural processing letters, Vol. 9, Issue 3, 1999, pp.293-300.

[6] Kaseko, M. S., and Ritchie, S. G., A neural network-based methodology for pavement crack detection and classification. Transportation Research Part C: Emerging Technologies, Vol. 1, Issue 4, 1993, pp.275-291.

[7] Kabir, S., Rivard, P., and Ballivy, G., Detection of AAR deterioration patterns in concrete using wavelets for multiscale texture analysis. Special Publication, Vol. 234, 2006, pp.127-146.

[8] Abdel-Qader, I., Abudayyeh, O., and Kelly, M. E., Analysis of edge-detection techniques for crack identification in bridges. Journal of Computing in Civil Engineering, Vol. 17, Issue 4, 2003, pp.255-263.

[9] Bennamoun, M. (1997, October). Edge detection: Problems and solutions. IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation , Vol. 4, 1997, pp.3164-3169

[10] Nishikawa, T., Yoshida, J., Sugiyama, T., and Fujino, Y., Concrete crack detection by multiple sequential image filtering. Computer-Aided Civil and Infrastructure Engineering, Vol. 27, Issue 1, 2012, pp.29-47.

[11] German, S., Brilakis, I., and DesRoches, R., Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments. Advanced Engineering Informatics, Vol. 26, Issue 4, 2012, pp.846-858.

[12] Yeum, C. M., and Dyke, S. J., Vision-based automated crack detection for bridge inspection. Computer-Aided Civil and Infrastructure Engineering, Vol. 30, Issue 10, 2015, pp.759-770.

[13] Cha, Y.-J., You, K. and Choi, W., Vision-based detection of loosened bolts using the Hough transform and support vector machines, Automation in Construction, Vol. 71, Issue 2, 2016, pp.181– 88.

[14] Zalama, E., Gómez-García-Bermejo, J., Medina, R., & Llamas, J., Road crack detection using visual features extracted by Gabor filters. Computer-Aided Civil and Infrastructure Engineering, Vol. 29, Issue 5, 2014, pp.342-358.

[15] Liao, K. W., and Lee, Y. T., Detection of rust defects on steel bridge coatings via digital image recognition. Automation in Construction, Vol. 71, 2016, pp.294-306.

[16] Chen, P. H., Shen, H. K., Lei, C. Y., and Chang, L. M., Support-vector-machine-based method for automated steel bridge rust assessment. Automation in Construction, Vol. 23, 2012, pp.9-19.

[17] Cha, Y. J., Choi, W., & Büyüköztürk, O., Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Computer-Aided Civil and Infrastructure Engineering, Vol. 32, Issue 5, 2017, pp.361-378.

[18] Zhang, L., Yang, F., Zhang, Y. D., and Zhu, Y. J., Road crack detection using the deep convolutional neural network. IEEE international conference on image processing (ICIP), 2016, pp.3708-3712.

[19] Girshick, R., Donahue, J., Darrell, T., and Malik, J., Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp.580-587.

[20] He, K., Zhang, X., Ren, S., and Sun, J., Spatial pyramid pooling in deep convolutional networks for visual recognition. In European conference on computer vision, 2014, pp.346-361.

[21] Girshick, R., Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, 2015, pp.1440–1448.

[22] Ren, S., He, K., Girshick, R., and Sun, J., Faster R-CNN: towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, Vol. 39, Issue 6, 2017, pp.1137-1149.

[23] Hacıefendioğlu, K., and Başağa, H. B., Concrete road crack detection using deep learning-based faster R-CNN method. Iranian Journal of Science and Technology, Transactions of Civil Engineering, Vol. 46, Issue 2, 2022, pp.1621-1633.

[24] Wang, N., Zhao, X., Zhao, P., Zhang, Y., Zou, Z., and Ou, J., Automatic damage detection of historic masonry buildings based on mobile deep learning. Automation in Construction, Vol. 103, 2019, pp.53-66.

[25] Karimi, N., Mishra, M., and Lourenço, P. B., Automated surface crack detection in historical constructions with various materials using deep learning-based YOLO network. International Journal of Architectural Heritage, 2024, pp.1-17.

[26] K. Simonyan and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". arXiv, arXiv: 1409.1556, Sep 4 2014, doi: https://doi.org/10.48550/arXiv.1409.1556.

[27] Wu, S., Zhong, S., and Liu, Y., Deep residual learning for image steganalysis. Multimedia tools and applications, Vol. 77, 2018, pp.10437-10453.

[28] Kingma, D. P., and Ba, J. (2014). "Adam: A Method for Stochastic Optimization". arXiv, arXiv:1412.6980, Dec 22 2014, doi: https://doi.org/10.48550/arXiv.1412.6980.

[29] Qu, Z., Chen, W., Wang, S. Y., Yi, T. M., and Liu, L.,A crack detection algorithm for concrete pavement based on attention mechanism and

multi-features fusion. IEEE Transactions on Intelligent Transportation Systems, Vol. 23, Issue 8, 2021, pp.11710-11719.

[30] Nie, M., and Wang, C., Pavement Crack Detection based on yolo v3. In 2019 2nd international conference on safety produce informatization (IICSPI), 2019, pp. 327-330.

[31] Nandini, K. A., Dharshan, S. S., and Bandaragoda, T., Automated Crack Analysis and Reporting in Civil Infrastructure using Generative AI. In IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society, 2024, pp. 1-6.

[32] Zhou, Q., Ding, S., Qing, G., and Hu, J., UAV vision detection method for crane surface cracks based on Faster R-CNN and image segmentation. Journal of Civil Structural Health Monitoring, Vol. 12, Issue 4, 2022, pp.845-855.

Copyright © Int. J. of GEOMATE All rights reserved, including making copies, unless permission is obtained from the copyright proprietors.