ILKOM Jurnal Ilmiah Vol. 18, No. 1, April 2026, pp. 58-68
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Enhancing Crack Detection on Levees with Synthetic Data Augmentation via ACGAN and Attention-Boosted Faster R-CNN

Saludin Muis [a,1,*]; Wiwit Priyadi [a,2]; Gita Puspa Artiani [b,3]; Risanto Darmawan [b,4]; Ali Khumaidi [a,5];
[a] *Universitas Bina Insani, Bekasi, 17114, Indonesia*
[b] *Universitas Krisnadwipayana, Jakarta, 13710, Indonesia*
[1] *saludin@binainsani.ac.id;* [2]*wiwit@binainsani.ac.id;* [3]*gita_artiani@unkris.ac.id;* [4]*risantodarmawan@unkris.ac.id;*
[5]*alikhumaidi@binainsani.ac.id*
** Corresponding author*

Article history: Received November 23, 2025; Revised December 16, 2025; Accepted February 15, 2026; Available online April 20, 2026

Abstract

This study introduces an innovative approach for detecting cracks on levee surfaces by integrating an Auxiliary Classifier Generative Adversarial Network (ACGAN) for data augmentation with a Faster R-CNN model enhanced by an attention mechanism. The ACGAN-based augmentation aims to generate synthetic images that enrich data variability in the original dataset. The attention-optimized Faster R-CNN is designed to improve detection precision, particularly for small objects and fine cracks that are difficult to distinguish from the background. Experimental results demonstrate that the incorporation of ACGAN improves detection performance, increasing both the mean Average Precision (mAP) and Average Recall (AR). The model achieved an mAP of approximately 0.56 at IoU = 0.50 and 0.34 at IoU = 0.75, while the AR (maxDets = 100) reached 0.55, indicating a strong capability in identifying most crack instances. When trained on the combined dataset of original and synthetic images, the Faster R-CNN model reached a precision of 0.92 for the severe crack class, while performance for minor cracks remained lower (precision 0.78). Adjusting the confidence threshold to 0.65 improved detection reliability by reducing noise and retaining high-confidence predictions. Improved performance in detecting severe cracks supports timely maintenance and repair decisions. This study demonstrates the effectiveness of GAN-based data augmentation and attention-enhanced object detection for automated structural health monitoring (SHM) of levee infrastructure.

Keywords: ACGAN; Faster R-CNN; Levee Crack Detection; Data Augmentation; Attention Mechanism.

Introduction

Levee infrastructure plays a crucial role in flood risk mitigation and water flow management in disaster-prone areas [1]. Structural damage to levee surfaces, particularly cracks, can lead to a reduction in load-bearing capacity and the potential for complete failure, which threatens public safety and the sustainability of such infrastructure [2]. Initially minor cracks are often undetected in their early stages, allowing further damage to escalate into more significant issues. Therefore, early detection of cracks in levees is essential to prevent more severe structural failures [3]. Conventional inspection methods, which rely on visual surveys by experts, are still predominantly used in levee monitoring practices [4]. However, this approach has significant limitations, including subjectivity, high labor demand, limited coverage of large areas, and restricted access to hazardous locations [5]. As a result, the development of automated damage detection systems based on computer vision has become an urgent need in the field of structural health monitoring (SHM).

Advancements in image processing and deep learning technologies have opened up new opportunities for automation in crack detection. Previous research has successfully implemented models based on Convolutional Neural Networks (CNN) and Faster R-CNN to detect cracks on road and bridge surfaces with high accuracy [6], [7]. However, these studies have identified key challenges in practical applications, including surface variability and limited diversity of training data [8]. In crack detection on infrastructures such as walls and asphalt, data augmentation using Generative Adversarial Networks (GAN) has proven effective in generating synthetic images that enhance data variability, thus improving model generalization across different crack patterns [9], [10], [11]. GAN has therefore emerged as a

promising solution to address data limitations in the vision domain by enriching dataset variation and improving detection robustness [12].

Despite the usefulness of GAN-based augmentation, crack detection on levee surfaces remains challenging due to complex textures and subtle visual characteristics, which often hinder models from identifying fine cracks [13]. Previous studies have also revealed that while object detection models such as Faster R-CNN are effective in detecting larger objects, additional architectural optimization is required to improve the detection of small or low-contrast cracks [14], [15], [16]. The unique challenges of levees include heterogeneous surface textures, dynamic environmental conditions, and complex crack patterns that differ from other man-made structures [17], [18]. Moreover, the scarcity of annotated levee-specific datasets, particularly for fine cracks, further limits the development of robust detection models.

To address these challenges, this study proposes an innovative framework that integrates two main components: (1) data augmentation using Auxiliary Classifier GAN (ACGAN) to generate high-quality synthetic samples that enrich the visual representation of various types of levee cracks, and (2) the implementation of an enhanced Faster R-CNN model with an attention mechanism to improve detection sensitivity to subtle crack features. The attention mechanism allows the model to adaptively focus on relevant regions in the feature map, thereby improving detection precision for fine cracks that were previously difficult to identify [19], [20].

The contributions of this study are threefold: the development of an ACGAN-based data augmentation pipeline tailored to the characteristics of levee surfaces, the integration of an attention mechanism into the Faster R-CNN architecture to support multi-scale crack detection under complex surface conditions, and a comprehensive evaluation of the combined impact of synthetic data augmentation and attention mechanisms on detection performance. This research contributes to the advancement of automated SHM for levee infrastructure and demonstrates the relevance of attention-enhanced deep learning models for crack detection tasks under limited-data scenarios.

Method

A. Data Acquisition and Preprocessing

In this study, data were acquired through visual documentation of levee surfaces using a high-resolution drone mounted on an unmanned aerial vehicle (UAV). The data were collected along the Bekasi River in Indonesia, specifically in the Pondok Gede Permai area, which has potential levee damage risks. The data acquisition involved video footage with a resolution of 1280x720 pixels, which was then processed into individual image frames, each resized to 256×256 pixels. These images were then labeled into three categories: No Cracks, Minor Cracks, and Severe Cracks. Labeling was performed using the Roboflow platform to ensure consistent accuracy in labeling [21].

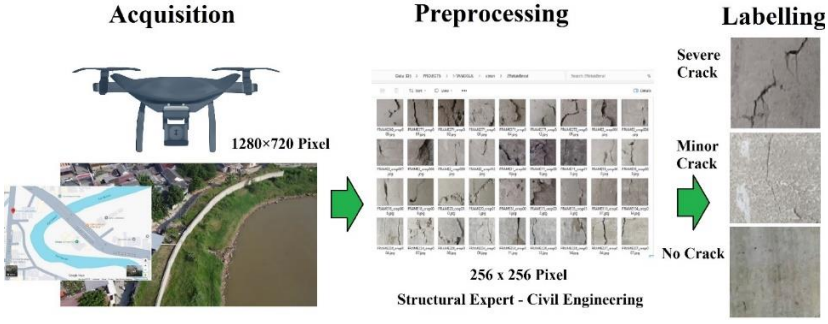

Figure 1. Data Acquisition and Preprocessing Workflow

Figure 1 illustrates the data acquisition and preprocessing workflow. After cropping, the images were further processed to ensure optimal visual quality. The preprocessing steps included resizing the images to 256x256 pixels to ensure consistent input size for the deep learning model, normalizing the pixel values to a range of 0-1 according to Equation (1), and applying median filtering for noise reduction to minimize interference in the images that could affect the model's accuracy, as described in Equation (2). The final dataset consisted of 2,970 images, which were systematically divided into training (70%), validation (20%), and testing (10%) subsets to ensure reliable model evaluation. The dataset follows the COCO JSON annotation format, enabling direct compatibility with object detection models such as Faster R-CNN.

$$I_{norm}(\mathrm{x},\mathrm{y}) = \frac{I(x,y) - I_{min}}{I_{max} - I_{min}} \quad (1)$$

$$I_{filtered}(\mathrm{x},\mathrm{y}) = Median\{I_{norm}(i,j)\}, \quad (\mathrm{i},\mathrm{j}) \in W(x,y) \quad (2)$$

B. Data Augmentation Using Generative Adversarial Network (GAN)

To address the issue of limited data, data augmentation was performed using the Auxiliary Classifier GAN (ACGAN) [22]. This GAN consists of two main components: the Generator, which produces synthetic images based on the training data distribution, and the Discriminator, which evaluates the authenticity of the images and classifies them (**Figure 2**). The Generator is trained to create synthetic images that closely resemble the original data, while the Discriminator assesses whether the generated images are real or fake [23].

During training, parameters such as image size (256×256 pixels), batch size, number of epochs, and Adam optimizer were configured to ensure model stability in generating new images. The output from ACGAN was additional synthetic images for each category, which were then evaluated for quality both visually and through the Fréchet Inception Distance (FID) calculation to ensure their similarity to the original data [24]. Each crack category was augmented with 460 synthetic images, resulting in 990 images per class and a total of 2,970 images. A visual selection process was applied to ensure that the generated images preserved realistic texture characteristics and crack morphology consistent with real levee surfaces.

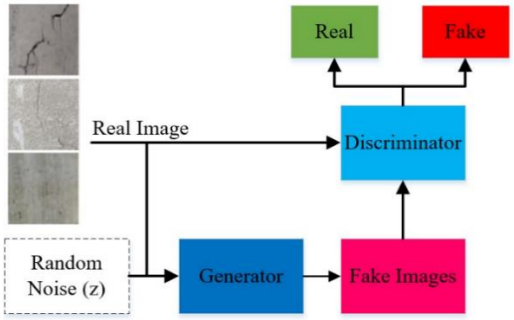

Figure 2. Data Augmentation Process Using ACGAN

C. Crack Detection Using Faster R-CNN with Attention Mechanism

The primary detection model used in this study is Faster R-CNN, a convolutional network architecture specifically designed for object detection in images [25]. Faster R-CNN relies on two key components: the Region Proposal Network (RPN), which generates candidate regions (bounding boxes) that are likely to contain cracks, and RoI Pooling, which classifies each proposal and accurately maps the crack locations [26].

To improve detection precision and speed, the Faster R-CNN model was enhanced with an attention mechanism [27]. This mechanism assigns greater weight to image areas relevant to cracks, allowing the model to focus on dominant features during detection and reduce attention to noise or irrelevant backgrounds [19]. The attention mechanism was formulated using Equation (3), where W represents the learnable weight matrix and f(x) denotes the extracted feature representation. Through this mechanism, the model improves sensitivity to fine cracks that are often overlooked in standard detection pipelines.

ResNet-50 was employed as the backbone network for feature extraction, with an additional Feature Pyramid Network (FPN) incorporated to support multi-scale crack detection [28]. **Figure 3** illustrates the overall detection architecture. The detailed training configuration of the attention-enhanced Faster R-CNN model, including optimizer settings, learning rate, batch size, and number of epochs, is summarized in **Table 1**.

$$A(x) = Softmax(W \,.\, f(x)) \quad (3)$$

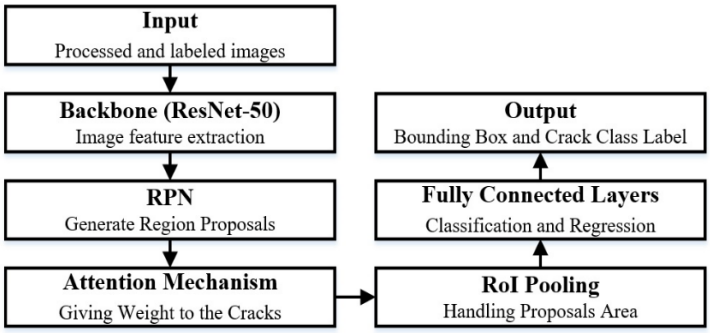

Figure 3. Faster R-CNN Architecture with Attention Mechanism

From a computational perspective, the use of ACGAN introduces additional overhead only during the data preparation stage, as synthetic image generation is performed offline prior to detector training. Consequently, ACGAN does not affect inference-time complexity. The Faster R-CNN model with an attention mechanism incurs a moderate increase in training-time computational cost due to additional feature weighting operations; however, this overhead remains manageable given the use of a ResNet-50 backbone, a limited batch size, and a fixed input resolution. At inference time, the proposed framework maintains practical efficiency, making it suitable for deployment scenarios where detection accuracy is prioritized over real-time processing.

Table 1. Hyperparameter Configuration of ACGAN and Faster R-CNN

Model	Hyperparameter	Value
ACGAN	Image size	256 × 256
	Optimizer	Adam
	Batch size	as implemented
	Epochs	as implemented
Faster R-CNN	Backbone	ResNet-50
	FPN	Enabled
	Optimizer	Adam
	Learning rate	0.001
	Batch size	4
	Epochs	50

D. Model Evaluation

Model evaluation was conducted to assess the performance of the crack detection model, trained using the dataset that includes both original and synthetic images generated by ACGAN. The evaluation aimed to measure the model's effectiveness in detecting various types of cracks on levee surfaces. The model's performance was evaluated using several standard metrics for object detection.

Mean Average Precision (mAP)

The mAP metric is used to measure the overall accuracy of object detection. mAP is calculated at various Intersection over Union (IoU) thresholds [29], specifically IoU = 0.50, IoU = 0.75, and IoU = 0.50:0.95. The Equation (4) for calculating mAP.

$$\text{mAP} = \frac{1}{\text{N}} \sum_{\text{i=1}}^{\text{N}} \frac{TP_i}{TP_i + FP_i} \tag{4}$$

Average Recall (AR)

The Average Recall (AR) metric measures the model's ability to detect most objects in an image [30]. AR is calculated at maxDets=1, maxDets=10, and maxDets=100, representing the number of successful detections at various maximum allowed detection limits per image. The Equation (5) for calculating AR.

$$\mathrm{AR} = \frac{1}{\mathrm{K}} \sum_{\mathrm{k}=1}^{\mathrm{K}} \frac{TP_k}{TP_k + FP_k} \tag{5}$$

Precision, Recall, and F1-Score

These metrics are used to measure the balance between true positive and negative predictions. Precision measures how accurately the model detects cracks, recall measures the model's ability to detect all relevant objects, while the F1-score provides a balanced measure between precision and recall. The Equation (6) for calculating precision, Equation (7) for calculating recall, and Equation (8) for calculating f1-score.

$$Precision = \frac{TP}{\mathrm{TP} + \mathrm{FP}} \tag{6}$$

$$Recall = \frac{TP}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{F1} = 2\ x\ \frac{Precision\ x\ Recall}{Precision\ +\ Recall} \tag{8}$$

Model Accuracy

Model accuracy is calculated to provide an overall picture of the detection performance by computing the ratio of true positive detections to the total number of detections. The Equation (8) for calculating accuracy. TP is the True Positive count, TN is the True Negative count, FP is the False Positive count, FN is the False Negative count.

$$Accuracy = \frac{TP\ +\ TN}{TP\ + TN + FP + \ FN} \tag{9}$$

In addition to the above metrics, a classification report was also used to analyze the model's performance based on each crack category: Severe Crack, Minor Crack, and No Crack. This evaluation aims to identify the strengths and weaknesses of the model, particularly in detecting minor cracks and distinguishing between smaller crack categories and areas without cracks. The results of this evaluation provide a clear picture of the effectiveness of data augmentation using ACGAN and the application of the Faster R-CNN model in detecting cracks on levee surfaces.

Results and Discussion

A. ACGAN Augmentation Model Evaluation

In the first evaluation stage, the ACGAN model was used to generate synthetic images as part of the dataset augmentation. These synthetic images aimed to enrich the training data variability, particularly to improve detection performance by increasing the variation across the three crack classes (No Crack, Minor Crack, and Severe Crack). Although the original dataset was class-balanced, augmentation was applied to enhance intra-class diversity, thereby improving model generalization.

The quality of the synthetic images was evaluated using visual inspection and the Fréchet Inception Distance (FID) metric. The evaluation results showed that the crack structure and patterns generated by ACGAN were similar to real images, with some variations in crack shapes not present in the original data. This contributed to increased variability in visual features used for training the detection model and enriched the representation of crack classes. ACGAN was able to generate synthetic images with quality very close to the original, which helped improve the model's detection performance, especially in recognizing various crack variations that might have been missed in the original dataset. These results are consistent with findings in previous studies that show GAN can enrich datasets and improve detection model performance [9], [12].

Figure 4 presents examples of synthetic Severe Crack images generated by ACGAN, demonstrating diverse crack morphologies while maintaining consistency with real crack patterns. A visual selection step was applied to ensure that synthetic images preserved realistic texture and crack characteristics, thereby preventing the inclusion of visually implausible samples.

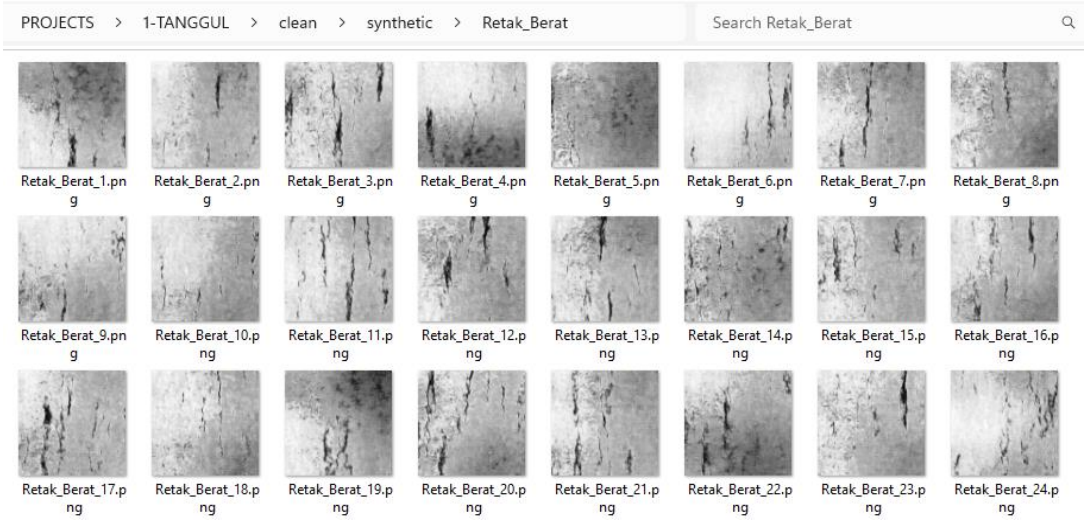

Figure 4. Sample Synthetic Image Generated by ACGAN

B. Faster R-CNN Detection Model Performance

In the subsequent evaluation stage, the performance of the Faster R-CNN model was assessed using a combined dataset consisting of original and ACGAN-generated images. Figure 5 summarizes the model performance across multiple mAP and AR metrics, providing an overview of detection effectiveness at different difficulty levels. The model achieved an mAP of approximately 0.56 at IoU = 0.50 and 0.34 at IoU = 0.75, with an mAP of 0.32 for IoU = 0.50:0.95. The Average Recall (AR) also showed a positive trend, with AR (maxDets = 100) approaching 0.55, indicating a strong ability to detect most crack instances. In general, higher mAP values at lower IoU thresholds indicate robust crack presence detection, whereas reduced performance at higher IoU thresholds suggests remaining challenges in precise boundary localization, particularly for small cracks. These results confirm that synthetic data augmentation contributes to improved detection performance across crack categories.

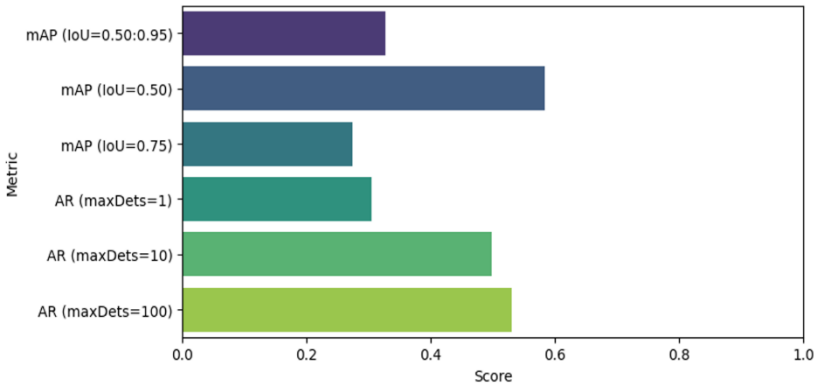

Figure 5. mAP and AR Metrics of the Final Faster R-CNN Model Evaluation (Test Dataset)

Table 2 presents class-wise precision, recall, F1-score, and accuracy results. The overall classification accuracy reached 85%, indicating stable detection performance. The Severe Crack class achieved the highest precision and recall values (0.92 and 0.90), demonstrating reliable identification of critical damage. The No Crack class also showed consistent performance, with an F1-score of 0.83, confirming the model's ability to distinguish intact surface regions. Performance for the Minor Crack class was comparatively lower (precision 0.78, recall 0.81), which can be attributed to visual similarity between minor cracks and non-crack regions, as well as variations in lighting and surface texture. Despite this limitation, the balanced precision–recall profile across classes indicates that the proposed model is suitable for automated crack detection on levee and similar infrastructure surfaces.

Table 2. Faster R-CNN Detection Performance Based on Precision, Recall, F1-Score, and Accuracy

Crack Class	Precision	Recall	F1-Score	Support
Severe Crack	0.92	0.90	0.91	99

Crack Class	Precision	Recall	F1-Score	Support
Minor Crack	0.78	0.81	0.80	99
No Crack	0.84	0.83	0.83	99
Accuracy	-	-	0.85	297
Macro avg	0.85	0.85	0.85	297
Weighted avg	0.85	0.85	0.85	297

C. Visualization and Confidence Threshold Analysis

The effect of confidence threshold selection on detection results was further analyzed. At a very low confidence threshold (0.05), the model generated a large number of overlapping bounding boxes, resulting in visually congested detections and increased noise. As shown in **Figure 6**, low-confidence predictions reintroduced previously missed cracks but substantially reduced detection reliability due to excessive false positives.

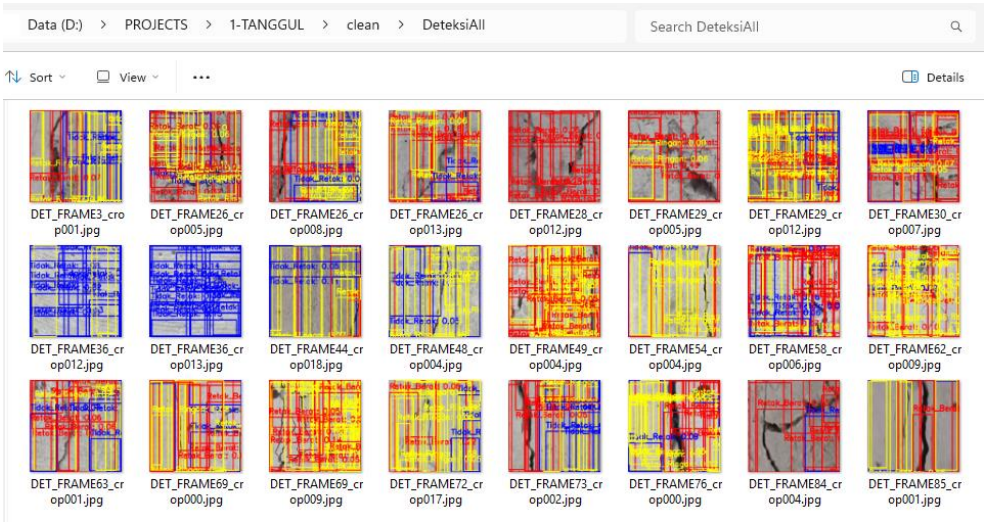

Figure 6. Detection Results at Very Low Confidence Threshold (0.05)

At a higher confidence threshold (0.65), detection outputs became significantly cleaner and more precise. **Figure 7** illustrates that noise was effectively suppressed, leaving only high-confidence crack detections. This adjustment improved precision and reduced false detections, although some low-confidence crack instances were excluded. These results highlight the importance of confidence threshold tuning to balance precision and recall in practical deployment scenarios.

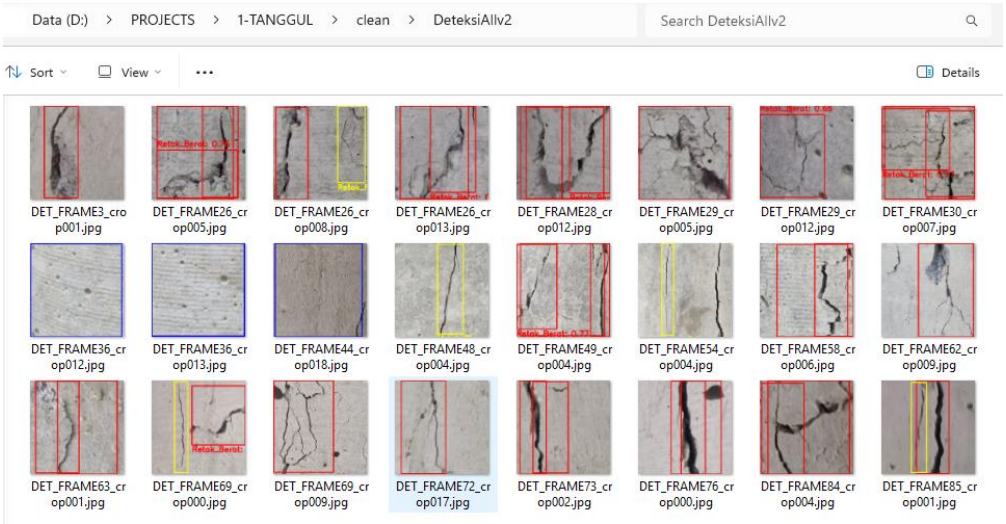

Figure 7. Detection Results at Logical Confidence Threshold (0.65)

D. Discussion and Implications

Impact of ACGAN on Detection Performance Improvement

In this study, the use of ACGAN for data augmentation significantly impacted the improvement of the Faster R-CNN detection model, especially in detecting various types of cracks on levee surfaces. Previously, the main challenge in crack detection on infrastructure was data limitations. Although the number of images for each class (minor cracks, severe cracks, and no cracks) was balanced, augmentation using ACGAN provided additional variability in the training data, enhancing the model's ability to recognize more diverse patterns.

The results of this study indicate that GAN can generate synthetic images that significantly improve detection performance on small and imbalanced datasets [12]. Furthermore, it was found that synthetic data from GAN could enrich the variability of objects in the dataset, improving mAP and precision in detecting small objects in images with complex variations, such as those on levee surfaces [9]. The integration of GAN with Faster R-CNN has been shown to provide much better performance compared to conventional detection approaches that only rely on original data.

Previous studies have also reported that ACGAN improves detection performance under heterogeneous textures and varying illumination conditions [31], [32]. In this study, the combined use of ACGAN and Faster R-CNN mitigated common failures in detecting small objects, such as minor cracks. However, residual performance gaps for minor cracks indicate the need for further optimization, potentially through richer augmentation strategies such as VAE-GAN or StyleGAN [33], [34].

Integration of ACGAN and Faster R-CNN in Crack Detection

The Faster R-CNN model demonstrated high effectiveness in detecting large crack objects, achieving a precision of 0.92 for severe cracks. Nonetheless, detecting fine cracks remains challenging due to subtle visual cues and complex surface textures. These findings align with prior work indicating that standard object detectors require additional architectural or parameter-level optimization to improve small-object detection performance [35], [36]. Adjusting the IoU threshold represents one potential approach to improving sensitivity for minor cracks. In this study, optimal performance was observed at IoU = 0.50, while performance decreased at IoU = 0.75, underscoring the trade-off between localization precision and detection sensitivity [37].

Implications for Early Detection System Development

Overall, the integration of ACGAN and Faster R-CNN produced robust and reliable crack detection performance, supporting its applicability as an early damage detection system. The proposed framework is not limited to levee infrastructure and can be extended to other structures such as bridges and retaining walls. By enabling earlier detection of critical damage, this approach supports preventive maintenance strategies that enhance infrastructure reliability and sustainability. Future extensions may incorporate additional high-resolution data sources to further improve model generalization under diverse field conditions.

Conclusion

This study successfully developed a crack detection system for levee surfaces using an approach based on ACGAN for data augmentation and Faster R-CNN enhanced with an attention mechanism. The evaluation results indicate that the combination of ACGAN and Faster R-CNN improves detection performance across varied image conditions and crack types. The use of ACGAN to generate synthetic images enhanced training data variability, contributing to improved mAP and recall and more stable detection of small objects, including minor cracks. The Faster R-CNN model trained with the combined dataset (original and synthetic images) demonstrated strong performance in detecting severe cracks, achieving high precision and recall (0.92 and 0.90). Nevertheless, detecting finer cracks remains challenging, indicating the need for further optimization. The confidence threshold was shown to significantly influence detection quality, where a higher threshold (0.65) effectively reduced noise and improved precision. Although the proposed framework shows promising results, performance gaps at higher IoU thresholds and for minor crack detection highlight existing limitations. Future improvements may be achieved through richer data variation, such as integrating alternative generative models (e.g., VAE-GAN or StyleGAN) and incorporating higher-resolution UAV imagery. Overall, this study contributes to the fields of computer vision and deep learning by demonstrating the effectiveness of combining GAN-based data augmentation and attention-enhanced object detection for automated structural health monitoring (SHM), with potential applicability to other infrastructure systems beyond levees.

Acknowledgement

This research was funded by the 2025 Research Grant from the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, under contract number 0070/C3/AL.04/2025. We express our sincere gratitude for the support and funding provided, which has made this research possible. We would also like to extend our thanks to all those who have contributed valuable input and provided facilities that have supported the smooth progress of this study.

References

[1] M. Farid et al., "Development of an Empirical Flood Fragility Curve for Levee Failure and Its Application in Probabilistic Flood Risk Assessment: A Case Study of Citeureup Village, Indonesia," Int. J. Disaster Risk Reduct., p. 105793, Sep. 2025, doi: 10.1016/j.ijdrr.2025.105793.

[2] A. Eskandarinejad, R. Nazari, M. R. Nikoo, D. Arellano, S. Pezeshk, and S. H. Ghasemi, "A comprehensive review of geotechnical implications of floods and water-driven disasters," Sci. Total Environ., vol. 985, p. 179731, Jul. 2025, doi: 10.1016/j.scitotenv.2025.179731.

[3] Z. Qu, Y. Zhang, Z. Liu, R. Si, and J. Wu, "A review on early-age cracking of concrete: Causes and control," Case Stud. Constr. Mater., vol. 21, p. e03848, Dec. 2024, doi: 10.1016/j.cscm.2024.e03848.

[4] S. J. H. Rikkert, M. Kok, K. Lendering, and R. Jongejan, "A pragmatic, performance-based approach to levee safety assessments," J. Flood Risk Manag., vol. 15, no. 4, Dec. 2022, doi: 10.1111/jfr3.12836.

[5] Q. Guo, "Strategies for a resilient, sustainable, and equitable Mississippi River basin," River, vol. 2, no. 3, pp. 336–349, Aug. 2023, doi: 10.1002/rvr2.60.

[6] G. Yu, X. Zhou, and X. Chen, "VDCrackGAN: A Generative Adversarial Network with Transformer for Pavement Crack Data Augmentation," Appl. Sci., vol. 14, no. 17, p. 7907, Sep. 2024, doi: 10.3390/app14177907.

[7] T. S. Tran, S. D. Nguyen, H. J. Lee, and V. P. Tran, "Advanced crack detection and segmentation on bridge decks using deep learning," Constr. Build. Mater., vol. 400, p. 132839, Oct. 2023, doi: 10.1016/j.conbuildmat.2023.132839.

[8] H. Kaveh and R. Alhajj, "Recent advances in crack detection technologies for structures: a survey of 2022-2023 literature," Front. Built Environ., vol. 10, Jul. 2024, doi: 10.3389/fbuil.2024.1321634.

[9] G. Chen, S. Teng, M. Lin, X. Yang, and X. Sun, "Crack Detection Based on Generative Adversarial Networks and Deep Learning," KSCE J. Civ. Eng., vol. 26, no. 4, pp. 1803–1816, Apr. 2022, doi: 10.1007/s12205-022-0518-2.

[10] S. Lie et al., "High Quality Training Set Development for Road Crack Detection Using Progressive Data-Driven Models with Integrated Identification," Case Stud. Constr. Mater., p. e05416, Oct. 2025, doi: 10.1016/j.cscm.2025.e05416.

[11] L. Pei, Z. Sun, L. Xiao, W. Li, J. Sun, and H. Zhang, "Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network," Eng. Appl. Artif. Intell., vol. 104, p. 104376, Sep. 2021, doi: 10.1016/j.engappai.2021.104376.

[12] E. Branikas, P. Murray, and G. West, "A Novel Data Augmentation Method for Improved Visual Crack Detection Using Generative Adversarial Networks," IEEE Access, vol. 11, pp. 22051–22059, 2023, doi: 10.1109/ACCESS.2023.3251988.

[13] D. P. Yadav, B. Sharma, S. Chauhan, and I. Ben Dhaou, "Bridging Convolutional Neural Networks and Transformers for Efficient Crack Detection in Concrete Building Structures," Sensors, vol. 24, no. 13, p. 4257, Jun. 2024, doi: 10.3390/s24134257.

[14] Q. Wu and R. Han, "Deep Learning Based Crack Detection Method for Bridges," in 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE), IEEE, Mar. 2025, pp. 449–455. doi: 10.1109/EESPE63401.2025.10987561.

[15] Y. Sun, Z. Sun, and W. Chen, "The evolution of object detection methods," Eng. Appl. Artif. Intell., vol. 133, p. 108458, Jul. 2024, doi: 10.1016/j.engappai.2024.108458.

[16] C. Cao et al., "An Improved Faster R-CNN for Small Object Detection," IEEE Access, vol. 7, pp. 106838–106846, 2019, doi: 10.1109/ACCESS.2019.2932731.

[17] D. Dais, İ. E. Bal, E. Smyrou, and V. Sarhosis, "Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning," Autom. Constr., vol. 125, p. 103606, May 2021, doi: 10.1016/j.autcon.2021.103606.

[18] R. E. Philip, D. Andrushia A, N. Anand, M. E. Mathews, M. Z. Naser, and E. Lubloy, "Improved YOLOv5-based multi-crack detection in concrete wall surfaces," Appl. Eng. Sci., vol. 23, p. 100247, Sep. 2025, doi: 10.1016/j.apples.2025.100247.

[19] N. Zhang, X. Gao, X. Lai, H. Xu, Y. Liu, and H. Ma, "Research on crack detection method for shallow-buried underground compressed air energy storage cavern based on improved mask R-CNN model," Earth Energy Sci., vol. 1, no. 3, pp. 203–212, Sep. 2025, doi: 10.1016/j.ees.2025.07.001.

[20] Q. Li, X. Xu, J. Guan, and H. Yang, "The Improvement of Faster-RCNN Crack Recognition Model and Parameters Based on Attention Mechanism," Symmetry (Basel)., vol. 16, no. 8, p. 1027, Aug. 2024, doi: 10.3390/sym16081027.

[21] D. Matuzevičius, "A Retrospective Analysis of Automated Image Labeling for Eyewear Detection Using Zero-Shot Object Detectors," Electronics, vol. 13, no. 23, p. 4763, Dec. 2024, doi: 10.3390/electronics13234763.

[22] D. Ruan, X. Chen, C. Gühmann, and J. Yan, "Improvement of Generative Adversarial Network and Its Application in Bearing Fault Diagnosis: A Review," Lubricants, vol. 11, no. 2, p. 74, Feb. 2023, doi: 10.3390/lubricants11020074.

[23] H. Cang et al., "Jujube quality grading using a generative adversarial network with an imbalanced data set," Biosyst. Eng., vol. 236, pp. 224–237, Dec. 2023, doi: 10.1016/j.biosystemseng.2023.11.002.

[24] E. Thirumagal and K. Saruladha, "GAN models in natural language processing and image translation," in Generative Adversarial Networks for Image-to-Image Translation, Elsevier, 2021, pp. 17–57. doi: 10.1016/B978-0-12-823519-5.00001-4.

[25] X. Xu et al., "Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN," Sensors, vol. 22, no. 3, p. 1215, Feb. 2022, doi: 10.3390/s22031215.

[26] V. Sai Sandeep Velaga, "Object Detection Using Faster R CNN with ROI Alignment using Bilinear Interpolation," Int. J. Sci. Dev. Res., vol. 10, no. 9, 2025, doi: 10.56975/ijsdr.v10i9.304835.

[27] P. Fu and J. Wang, "Lithology Identification Based on Improved Faster R-CNN," Minerals, vol. 14, no. 9, p. 954, Sep. 2024, doi: 10.3390/min14090954.

[28] Q. Su, G. Zhang, S. Wu, and Y. Yin, "FI-FPN: Feature-integration feature pyramid network for object detection," AI Commun., vol. 36, no. 3, pp. 191–203, Aug. 2023, doi: 10.3233/AIC-220183.

[29] M. Trigka and E. Dritsas, "A Comprehensive Survey of Deep Learning Approaches in Image Processing," Sensors, vol. 25, no. 2, p. 531, Jan. 2025, doi: 10.3390/s25020531.

[30] A. Berger and S. Guda, "Threshold optimization for F measure of macro-averaged precision and recall," Pattern Recognit., vol. 102, p. 107250, Jun. 2020, doi: 10.1016/j.patcog.2020.107250.

[31] F. Liu, Y. Li, and Y. Zheng, "Time Series Data Generation Method with High Reliability Based on ACGAN," Entropy, vol. 27, no. 2, p. 111, Jan. 2025, doi: 10.3390/e27020111.

[32] G. Wang, Z. Yang, H. Sun, Q. Zhou, and Z. Yang, "AC-SNGAN: Multi-class data augmentation for damage detection of conveyor belt surface using improved ACGAN," Measurement, vol. 224, p. 113814, Jan. 2024, doi: 10.1016/j.measurement.2023.113814.

[33] Z. Sordo et al., "Synthetic Scientific Image Generation with VAE, GAN, and Diffusion Model Architectures," J. Imaging, vol. 11, no. 8, p. 252, Jul. 2025, doi: 10.3390/jimaging11080252.

[34] M. Ibrahim et al., "Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges," Comput. Biol. Med., vol. 189, p. 109834, May 2025, doi: 10.1016/j.compbiomed.2025.109834.

[35] W. A. Shobaki and M. Milanova, "A Comparative Study of YOLO, SSD, Faster R-CNN, and More for Optimized Eye-Gaze Writing," Sci, vol. 7, no. 2, p. 47, Apr. 2025, doi: 10.3390/sci7020047.

[36] H. Mahamivanan, J. Matthews, P. E. D. Love, and F. Nasirzadeh, "Toward accurate detection of small objects in rail construction: A deep learning perspective," Eng. Appl. Artif. Intell., vol. 160, p. 111977, Nov. 2025, doi: 10.1016/j.engappai.2025.111977.

[37] Q. Yuan, Y. Shi, and M. Li, "A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges," Remote Sens., vol. 16, no. 16, p. 2910, Aug. 2024, doi: 10.3390/rs16162910.