

Lightweight Deep Learning CNN Architecture Enhanced by Squeeze-and-Excitation Attention for Bridge Crack Image Analysis

MURAT UCAN¹, IBRAHIM BARAN KARASIN²

¹Department of Computer Technologies,
Diyarbakir Vocational School of Technical Sciences,
Dicle University,
Diyarbakir,
TURKEY

²Department of Construction Technologies,
Diyarbakir Vocational School of Technical Sciences,
Dicle University,
Diyarbakir,
TURKEY

Abstract: - Cracks that develop over time in bridges, which are an important part of transportation infrastructure, pose a structural safety risk when detected late and can lead to increased repair costs. This study aims to demonstrate the effectiveness of convolutional neural network (CNN) based deep learning architectures enhanced with attention mechanisms in the problem of classifying bridge crack images and to analyse whether these approaches offer a reliable alternative to traditional methods. In the proposed study, model training and testing processes were performed using an open-source dataset consisting of crack and crack-free images obtained from bridge surfaces. The proposed model achieved an accuracy rate of 97.36% in the predefined data splitting scenario, demonstrating higher classification performance than existing traditional methods. The results obtained demonstrate that the attention layer offers a robust solution not only in terms of high accuracy but also in terms of reliability and generalizability in critical infrastructure applications.

Key-Words: - Attention, Bridge, Crack Detection, CNN, Deep Learning, Image Classification, Mobilenet.

Received: March 16, 2026. Revised: May 11, 2026. Accepted: June 10, 2026. Published: September 30, 2026.

1 Introduction

Natural aging and environmental conditions can cause deterioration in bridges, which are an important part of transportation infrastructure. If cracks are not detected early, they can cause structural problems and, in some cases, create serious issues that can threaten human life, [1], [2]. Furthermore, early detection can be remedied with lower-cost solutions, while late detection leads to increased construction costs, [3].

Traditional methods of detecting bridge cracks rely solely on the observations of expert personnel. This is time-consuming, prone to human error, and far from being a sustainable solution for large-scale structures, [4]. In recent years, advances in machine learning and deep learning methods have enabled the successful analysis of images in many different disciplines. Image analysis using deep learning architectures has been proposed by researchers in

many fields as a fast, consistent, and sustainable solution.

In the field of computer vision, many classification and segmentation-based studies have been conducted by researchers on the autonomous analysis of bridge cracks, [5]. On the other hand, developed an end-to-end Convolutional Neural Network (CNN) model integrating the Atrous Spatial Pyramid Pooling (ASPP) module for the automatic detection of concrete bridge cracks, [6]. In the study, the depthwise separable convolution technique was used to reduce computational cost while aiming to preserve image resolution and extract multi-scale contextual information. The model developed by the researchers achieved an accuracy rate of 96.37%, demonstrating higher performance and computational efficiency compared to classical classification models such as ResNet and VGG.

Another model called SSENets based on the SSE module, which strengthens the connection between

shallow and deep layers, in order to improve accuracy in bridge crack detection and mitigate the vanishing gradient problem in deep networks, [7]. This proposed model achieved a success rate of 97.77%. As a comparative study, Neural Network models for the automatic detection of concrete cracks across three different datasets were analyzed 12 different transfer learning models, [8]. In this work, the researchers utilized contrast enhancement techniques to improve performance in situations where image quality was low. In their study analyzing bridge cracks without image enhancement, they achieved a classification success rate of 98.80%.

This study presents convolutional neural network (CNN) based deep learning architectures for the problem of classifying bridge crack images. The proposed method is developed using a lightweight MobileNet architecture enhanced with a Squeeze-and-Excitation (SE) attention mechanism to improve feature representation by emphasizing informative channels. The results obtained were evaluated comparatively with CNN-based and hybrid models commonly used in the literature. The main objective of the study is to reveal the extent to which attention-based architectures offer an effective and reliable solution for the bridge crack detection problem and to demonstrate that these approaches constitute a powerful alternative for structural health monitoring systems. In this respect, the proposed study aims not only to achieve high accuracy rates but also to develop a more stable, generalizable, and practically applicable automatic inspection approach.

2 Methodology

This section provides a detailed explanation of the dataset used in the study and the CNN-based architecture that forms the basis of the proposed approach. First, the structure of the dataset, class distribution, and preprocessing steps are presented, followed by an examination of the fundamental components of the model architecture and training process. Thus, the general framework of the proposed method and how the experimental process was designed are systematically presented in this section.

2.1 Dataset

The proposed study utilized a dataset containing bridge crack images shared from open-source accessible resources during the education, validation, and testing phases. The Bridge Crack Dataset (BCD), created and prepared with a focus on image analysis using deep learning methods, is one of the important

datasets in this field in the literature, [6]. The researchers created an original dataset using 2068 raw images with a resolution of 1024×1024 obtained with Phantom 4 Pro. During the preprocessing stage, the images were divided into 512×512 dimensions, and after filtering out blurry samples, a total of 6069 images were obtained, consisting of 4058 cracks and 2011 backgrounds. Of the images obtained, 4856 were labeled for training and 1213 for testing and were published publicly, [9]. The class labels of the images were shared in two separate list files, suitable for use in binary classification architectures. When sharing their images, the researchers resized them to 224×224 dimensions to make them suitable for use in deep learning architectures.

The dataset, created to reflect real-world problems, also includes scenarios involving environmental factors such as shadows, water stains, and edge cracks. Furthermore, the dataset contains examples where the crack covers only a small part of the image or its edge. Thanks to this feature of the dataset used in the study, the results obtained can be compared with real-life images, ensuring that the results are valid.

2.2 Attention-Based Hybrid MobileNet Architecture

In the proposed study, a hybrid CNN model that is lightweight and computationally efficient was used for the automatic classification of cracks forming on bridge surfaces. The proposed model uses the MobileNet architecture as its basic feature extractor. MobileNet, one of the popular deep learning architectures, uses fewer parameters compared to standard CNN architectures thanks to its depthwise separable convolution structure, [10]. This ensures that the architecture is computationally efficient and provides a structure suitable for real-time and field applications in particular, [11]. The developed architecture aims to achieve high accuracy with low computational load in practical applications such as bridge health monitoring systems. The general architectural block diagram and experimental workflow of the proposed LiteBridge-SE model are illustrated in Figure 1.

In the proposed architecture, a Squeeze-and-Excitation (SE) attention mechanism is integrated to enhance the classification performance of the deep feature maps obtained by MobileNet, [12]. SE layers help deep learning models recognize which features extracted autonomously are more meaningful, [13]. By highlighting more important features, these layers reduce the impact of information that contributes little to classification, thereby improving classification accuracy, [14].

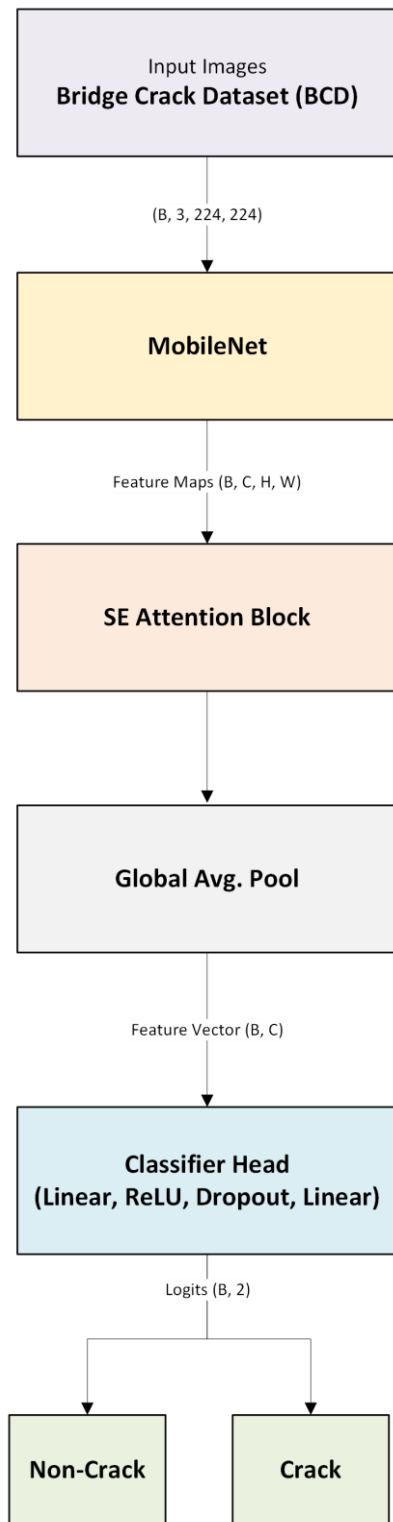


Fig. 1: General architectural block diagram and experimental workflow of the proposed LiteBridge-SE model

Source: created by the authors.

This mechanism was chosen for our study because it can provide significant advantages, particularly in distinguishing cracks with fine and irregular structures from background textures.

Following the SE block, the model classifies images with and without cracks into two categories through global average pooling and fully connected layers, [15]. Thus, the proposed attention-supported CNN architecture can effectively capture local features and offer higher classification performance by prioritizing features at the channel level. The results obtained in this study demonstrate that this hybrid structure offers an effective solution for critical structural health monitoring applications such as bridge crack detection.

In the proposed LiteBridge-SE architecture, the Squeeze-and-Excitation attention mechanism has been used to model channel dependencies between feature maps and to emphasize more informative channels. The SE block summarizes the feature maps obtained from the convolution layer using global average pooling and learns the channel weights by transforming them through nonlinear activation functions. The obtained weights enable the rescaling of features by multiplying the original feature maps on a channel basis. The general mathematical expression of the SE block is given in Equation (1).

$$\tilde{U} = \sigma \left(W^2 \delta \left(W^1 \text{GAP}(U) \right) \right) \odot U \quad (1)$$

In Formula 1, GAP represents the global average pooling operation, W_1 and W_2 represent the fully connected layer weights, δ represents the ReLU activation function, σ represents the sigmoid function, and $\odot$ represents the channel-wise multiplication operation. This process allows the deep learning architecture to enhance more informative and discriminative feature channels while reducing the impact of less important channels. Especially in bridge crack images, where fine and irregular crack structures can be difficult to distinguish from background textures, the SE mechanism enhances the quality of feature representation by adaptively learning channel importance. This attention-based hybrid reweighting structure contributes to more effectively capturing the distinguishing features between cracks and background textures while preserving the lightweight and computationally efficient structure of the MobileNet architecture.

3 Results and Discussions

All experimental processes conducted in this study were performed using the platform's GPU on the Pro+ infrastructure, which is the paid version of the Google Colab development environment. The developed model was trained for 10 epochs with a

batch size of 16. The input size of the images used in the architecture was set to 224x224.

The dataset used in the study was used together with the original training and test splits defined in CSV files, and a separate experimental scenario was created while maintaining this structure. Thus, the model performance could be directly observed under the officially defined split structure of the dataset.

Comprehensively evaluate the classification performance of the proposed LiteBridge-SE architecture within the scope of the study; confusion matrices and classification performance evaluation metrics commonly used in the literature have been employed. The MobileNet model achieves a lighter computational complexity by replacing standard convolutions with depthwise separable convolutions. This structural change reduces the computational cost and the number of parameters compared to traditional networks. Model performance evaluation includes not only overall accuracy but also evaluation metrics such as precision, recall, and F1-Score. Since the open-source dataset used in the study consists of 4058 crack samples and 2011 background images, exhibiting a significant class imbalance, accuracy alone may not sufficiently reflect the model's prediction reliability. The F1-Score is an evaluation metric that mathematically accounts for false positives and false negatives between classes. The F1-Score evaluates the architecture's effectiveness in distinguishing fine, irregular crack structures from background tissues. Precision is another important evaluation metric that shows how many of the examples classified as positive by the model are actually positive. In other words, it is a metric that controls the false positive rate and is particularly important in cases where mistakenly classifying background tissues as cracks is critical. Recall, on the other hand, is a classification performance evaluation metric that shows how many of the true positive examples are correctly identified by the model. The recall metric is of critical importance, especially in terms of not missing fine and irregular crack structures, and it directly reflects the false negative rate. All the mathematical expressions of the basic classification performance metrics used in the study are provided in formulas 2, 3, 4, and 5. The TP given in the formulas indicates the correct classification of a truly positive example as positive by the model. TN refers to the correct prediction of a negative example as negative. In contrast, FP indicates that a negative example has been incorrectly classified as positive. FN, on the other hand, defines the misclassification of a positive example as negative by the model.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (2)$$

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = TP / (TP + FN) \quad (4)$$

$$F1 - Score = 2TP / (2TP + FP + FN) \quad (5)$$

The success of the proposed system was measured by calculating evaluation metrics such as precision, recall, F1-score, and accuracy. Table 1 shows the class-based evaluation metric scores for two separate experiments conducted on the original split data sets. The results in the table demonstrate that the method proposed in our study can detect cracks in bridge images with high success.

Table 1. Results of the Bridge Crack Detection Architecture Based on Class in the Data Set

Class	Precision	Recall	F1-Score
0	83.07	1.00	90.75
1	1.00	96.97	98.46

Source: created by the authors.

An important representation used to measure classification performance in machine learning and deep learning methods is the confusion matrix. Using confusion matrices, researchers reveal how well or poorly their developed models perform on a class-by-class basis. Confusion matrix representation was also used to measure the effectiveness of the proposed model. Figure 2 shows the confusion matrices created for both experiments conducted in the proposed study.

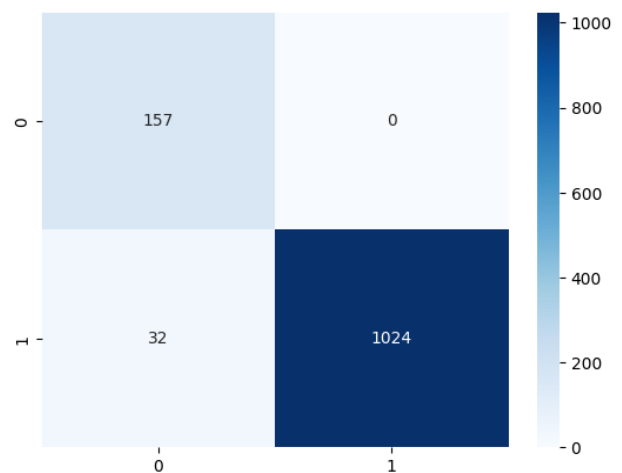


Fig. 2: Confusion matrix of the proposed attention-based deep learning architecture

Source: created by the authors.

Table 2 presents performance comparisons of different deep learning architectures for the bridge crack detection problem. According to Table 2,

among classical CNN-based approaches, the VGG19 and ResNet50 models exhibit relatively lower performance with accuracy rates of 85.41% and 89.45%, respectively, while the depthwise separable convolutional network (Depthwise Separable CNN) architecture provides a significant improvement with 96.37% accuracy

To ensure accurate comparisons in Table 2, only studies from the literature that used similar approaches. Our proposed LiteBridge-SE model achieved a high classification success rate of 97.36% in the autonomous analysis of bridge cracks. These results demonstrate that the proposed approach offers stronger generalization capabilities compared to existing methods in the literature and constitutes a reliable alternative for crack detection in bridge structures.

The experimental findings of this study yield important results for the field of autonomous bridge crack detection. Traditional inspection methods, which rely heavily on manual visual assessments, are labor-intensive, time-consuming, and prone to human error. The LiteBridge-SE architecture has demonstrated that it can achieve significantly better results than traditional methods, with a classification success rate of 97.36%. This innovative approach facilitates the early and accurate detection of structural defects. It also has the potential to significantly reduce long-term maintenance and repair costs.

Table 2. Results of the Bridge Crack Detection Architecture Based on Class in the Data Set

Model	Accuracy
Vgg19	85.41%
Resnet50	89.45%
ASPP-Based CNN, [6]	96.37%
Our (LiteBridge-SE)	97.36%

Source: created by the authors.

4 Conclusion

This study focuses on solving the problem of detecting cracks in bridge structures quickly, without human error, and in a sustainable manner. Within the scope of this study, a hybrid deep learning architecture called LiteBridge-SE was developed. The fast and lightweight structure of the MobileNet architecture was enhanced with attention blocks and used in the hybrid architecture. Comprehensive experimental analyses were conducted on an open-source bridge crack image dataset. The LiteBridge-SE architecture has achieved reliable results,

obtaining a classification success rate of 97.36% accuracy.

The main limitation of the study is that the proposed work focuses only on classifying the presence of cracks. The study did not analyze other structural defects in bridge infrastructures. In future studies, it is planned to train the proposed architecture as a multi-class model that can also detect different types of damage. Furthermore, it is intended to use the proposed architecture in examining structural damage through real-time systems by integrating it with systems such as unmanned aerial vehicles.

Acknowledgement:

This study was supported under the Dicle University Research Projects Support Program with the number 24.006

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author(s) utilized artificial intelligence (AI) technologies solely for the purpose of linguistic refinement and grammatical correction. Following the application of these tools, the author(s) comprehensively reviewed, verified, and edited the resulting text. The author(s) maintain full accountability and responsibility for the accuracy, integrity, and final content of the publication.

References:

- [1] H. Inam, N. U. Islam, M. U. Akram, and F. Ullah, "Smart and Automated Infrastructure Management: A Deep Learning Approach for Crack Detection in Bridge Images," *Sustainability*, vol. 15, no. 3, p. 1866, Jan. 2023, doi: 10.3390/su15031866.
- [2] N. Alfaz, A. Hasnat, A. MD. R. N. Khan, N. S. Sayom, and A. Bhowmik, "Bridge Crack Detection Using Dense Convolutional Network (DenseNet)," in *Proceedings of the 2nd International Conference on Computing Advancements*, New York, NY, USA: ACM, Mar. 2022, pp. 509–515. doi: 10.1145/3542954.3543027.
- [3] H. S. Munawar, A. W. A. Hammad, S. T. Waller, and M. R. Islam, "Modern Crack Detection for Bridge Infrastructure Maintenance Using Machine Learning," *Human-Centric Intelligent Systems*, vol. 2, no. 3–4, pp. 95–112, Sep. 2022, doi: 10.1007/s44230-022-00009-9.

- [4] J. Zhang, S. Qian, and C. Tan, "Automated Bridge Surface Crack Detection And Segmentation Using Computer Vision-Based Deep Learning Model," *Eng. Appl. Artif. Intell.*, vol. 115, p. 105225, Oct. 2022, doi: 10.1016/j.engappai.2022.105225.
- [5] 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.
- [6] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "Automatic Bridge Crack Detection Using a Convolutional Neural Network," *Applied Sciences*, vol. 9, no. 14, p. 2867, Jul. 2019, doi: 10.3390/app9142867.
- [7] Li, H., Xu, H., Tian, X., Wang, Y., Cai, H., Cui, K., Chen, X., "Bridge Crack Detection Based on SSENets," *Applied Sciences*, vol. 10, no. 12, p. 4230, Jun. 2020, doi: 10.3390/app10124230.
- [8] Bhalaji Kharthik, K.S., Onyema, E.M., Mallik, S., "Transfer Learned Deep Feature Based Crack Detection Using Support Vector Machine: A Comparative Study," *Sci. Rep.*, vol. 14, no. 1, p. 14517, Jun. 2024, doi: 10.1038/s41598-024-63767-5.
- [9] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "Crack-detection," 2026, [Online]. <https://github.com/tjdxhy/crack-detection/tree/master> (Accessed Date: January 10, 2026).
- [10] A. G. Howard, Z. Menglong, B. Chen. D. Kalenichenko, W. Wang, T. Weyand, M. Adreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. <http://arxiv.org/abs/1704.04861> (Accessed Date: January 10, 2026).
- [11] V.-T. Huynh, T.-T. Nguyen, T. V. Nguyen, and M.-T. Tran, "MobileNet-SA: Lightweight CNN with Self Attention for Sketch Classification," *Image and Video Technology. PSIVT 2023. Lecture Notes in Computer Science*, vol 14403. Springer, Singapore. 2024, pp. 110–123. doi: 10.1007/978-981-97-0376-0_9.
- [12] J. Hu, L. Shen, and G. Sun, "Squeeze-And-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition*, 2018, pp. 7132–7141, [Online]. https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html (Accessed Date: January 10, 2026).
- [13] A. Agarwal and S. Susan, "Attention-Augmented Squeeze-And-Excitation Enhanced Mobile Network for Occluded Facial Expression Recognition in Resource-Constrained Environments," *Signal Image Video Process.*, vol. 19, no. 9, p. 687, Sep. 2025, doi: 10.1007/s11760-025-04278-2.
- [14] S. Ucan, M. Ucan, and M. Kaya, "Deep Learning Based Approach with EfficientNet and SE Block Attention Mechanism for Multiclass Alzheimer's Disease Detection," in *2023 4th International Conference on Data Analytics for Business and Industry (ICDABI)*, IEEE, Bahrain, Oct. 2023, pp. 285–289. doi: 10.1109/ICDABI60145.2023.10629380.
- [15] R. D, S. K. Manigandan, P. K. Kartthega, P. K. Roshegaa, S. Keerthana, and J. Velmurugan, "Enhanced Classification Model for Plant Disease Prediction Using MobileNet with Squeeze-and-Excitation and Channel Attention Mechanism," in *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, IEEE, Greater Noida, India, Nov. 2025, pp.1-6. doi:10.1109/ICCCA66364.2025.11325202.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This study was supported under the Dicle University Research Projects Support Program with the number 24.006

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0 https://creativecommons.org/licenses/by/4.0/deed.en_US