IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 10 July 2024, accepted 7 August 2024, date of publication 13 August 2024, date of current version 21 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3442926

RESEARCH ARTICLE

Adaptive and Explainable Deep Learning-Based Rapid Identification of Architectural Cracks

JIANG-YI LUO[1] **AND YU-CHENG LIU**[2]
[1]School of Civil Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China
[2]School of Mathematics, Chengdu University of Technology, Yibin 610059, China

Corresponding author: Yu-Cheng Liu (ho8lxa@163.com)

ABSTRACT Concrete architectural structures are widely used in urban construction, making the health diagnosis and maintenance of these structures increasingly essential and urgent. Crack identification is crucial for maintaining the structural integrity and safety of buildings. Traditional methods rely on manual inspection, which is plagued by low accuracy, inefficiency, and safety hazards. This paper proposes a technique combining an attention-based SqueezeNet network with Gradient-weighted Class Activation Mapping (Grad-CAM) for automatically recognizing and visually explaining building cracks. By integrating the Squeeze-and-Excitation (SE) attention mechanism with the lightweight SqueezeNet network, this method can adaptively adjust the importance of feature channels by learning global information, effectively improving the network's accuracy and efficiency. The experimental results show that the Att-SqueezeNet model achieved a high precision of 0.995, a training time of only 133 seconds, and a model size of 4.9M, significantly outperforming models such as SqueezeNet, RF, CNN, VGG-19, and B-CNN. This demonstrates its robustness, rapid identification and suitability for practical applications and building crack identification. Moreover, the utilization of Grad-CAM for visualization not only offers an intuitive explanation of the model's decision-making process but also provides a more comprehensible understanding of crack detection results. This is crucial for advancing building maintenance automation, reducing reliance on manual labor, and increasing the precision and reliability of detection tasks.

INDEX TERMS Crack identification, SqueezeNet network, SE attention mechanism, Grad-CAM, model interpretability.

I. INTRODUCTION

With the development of China's modernization, the construction of residential and commercial buildings plays a vital role in economic growth. However, these structures are often compromised by the widespread presence of structural cracks, which can be attributed to various factors, including material defects, structural design issues, and environmental influences [1] Cracks serve as critical indicators for assessing the safety and stability of structures, playing an essential role in maintaining the long-term usability of buildings and ensuring the safety of their occupants [2], [3] Within the field of structural engineering, cracks not only signify the natural consequences of material aging and wear but can also indicate improper design, construction quality issues, or the impacts of external environmental factors. As buildings age, the potential for cracks to expand and compromise the structure's overall stability increases, potentially endangering the safety of residents. Thus, the timely detection and accurate assessment of the nature of these cracks are crucial steps in preventing potential structural failures. This proactive approach towards identifying and evaluating cracks emphasizes the importance of regular inspections and the use of advanced diagnostic tools, ensuring that early signs of deterioration are addressed promptly to safeguard the integrity and longevity of structural systems [4]

Crack detection and analysis are indispensable parts of structural safety evaluations, encompassing a variety of techniques from visual inspections to cutting-edge non-destructive testing methods [5] These approaches not only

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

aid in identifying the existence of cracks but also in understanding their progression and potential impacts on the structure [6] In structural engineering, traditional methods of crack identification predominantly involve manual inspections, inherently limited by their dependence on human observation [7] This methodology is fraught with subjectivity, potentially leading to oversights and inaccuracies in identifying structural defects. Moreover, the effectiveness of manual inspections largely hinges on the inspector's initiative and diligence. Factors such as hazardous working conditions and inaccessible inspection environments significantly undermine the reliability of crack inspections. While technologies like ultra-sonic testing and magnetic particle inspection can provide more accurate results, their high equipment costs, complex operation, and limitations in applicability on large-scale or hard-to-reach structures restrict their efficiency and accuracy in crack monitoring [8] These limitations highlight the necessity for more advanced, technology-driven solutions to ensure accurate, safe, and efficient assessments of structural cracks. The adoption of such sophisticated methods could mitigate the challenges posed by traditional techniques, offering a more reliable and comprehensive approach to structural integrity evaluation.

With the rapid advancement of deep learning technology, crack identification methods based on deep learning have demonstrated significant advantages. The application of deep learning in crack identification marks a substantial shift from traditional approaches to new technology, eliminating the need for predefined crack features or extensive image preprocessing [9] Instead, deep learning techniques learn from a vast array of samples and automatically extract features, thus facilitating the recognition and extraction of cracks [10] Compared to traditional methods, deep learning can not only process a large volume of data but also enhance identification accuracy through learning, especially in detecting cracks under complex backgrounds or in blurred conditions [11] A notable advancement in this field is the use of convolutional neural networks, which effectively address the issue of input image size limitation inherent in traditional neural networks. These networks are capable of processing images of varying dimensions, making them more adaptable to different scenarios in structural health monitoring. Moreover, the high degree of automation in deep learning methods significantly reduces the workload of manual inspections, offering a more efficient and reliable solution for crack monitoring. However, deep learning models with complex architectures typically comprise numerous parameters, demanding high computational complexity and storage capacity. The lightweight approaches of deep learning models aim to reduce model parameters, lower computational complexity, decrease storage requirements, get rapid identification while maintaining high performance, and enhance the model's generalization ability. These advancements contribute to a more practical and effective framework for detecting and monitoring structural cracks, facilitating their adoption in real-world applications [12]

Moreover, neural networks are often criticized due to their lack of interpretability. Functioning as a ''black box'', these models do not readily provide an understanding of the logic behind their decisions, making it challenging to assess the accuracy of the structure and diagnose the reasons for incorrect outcomes. Research into the interpretability of neural networks is crucial for enhancing the trust and reliability of decision-making. In short, this study introduces a method that combines the SqueezeNet network with Grad-CAM technology for crack identification and interpretation. SqueezeNet is a lightweight deep learning model with a compact structure and high efficiency, particularly suited for applications requiring rapid processing and low resource consumption. By incorporating an attention mechanism into SqueezeNet, our method focuses more on the crack regions within images, improving identification accuracy [13] At the same time, the use of Grad-CAM technology for visual interpretation of the identification results not only enhances the model's transparency but also provides an intuitive understanding of crack detection, aiding professionals in further analysis and decision-making. This approach, which combines efficient identification and interpretative capabilities, can make an adaptive network and offer new research directions and practical application prospects in the field of crack monitoring.

The remainder of the article is as follows: Section II provides the literature review. Section III describes the detailed methodology. Section IV conducts the validation of the approach through a case study. Section V outlines the conclusions and directions for future research.

II. LITERATURE REVIEW

A. CNNs FOR CRACK IDENTIFICATION

Convolutional Neural Networks (CNNs) have revolutionized the field of crack identification and analysis, offering significant advantages over traditional methods. By inputting a crack image into a CNN, the network can learn and extract crucial feature representations through a series of convolution and pooling operations. This process enables the CNN to detect local patterns, edges, and textures within the crack image, which are essential for accurately classifying and locating cracks [14] One of the primary strengths of CNNs lies in their adaptability and generalization capability [5] Through multiple layers of convolution and pooling, CNNs can learn complex characteristics of fractures, allowing them to handle cracks of varying sizes, shapes, and orientations effectively. The back-propagation algorithm further optimizes network parameters during training, enhancing the model's ability to generalize and adapt to new samples. Moreover, CNNs can improve crack identification performance through data augmentation and transfer learning. Data augmentation manipulates crack images by translating, rotating, scaling, and flipping, creating a more diverse training dataset that increases the model's robustness and disturbance resistance [15] Transfer learning leverages pre-trained CNN models on the crack recognition task, extracting and fine-tuning features to speed up model training and enhance accuracy [16]

Despite the significant advantages of CNNs, such as their powerful capabilities in image recognition and classification, they face several challenges in application. These challenges primarily stem from their dependence on large volumes of labeled training data and the substantial computational resources required to train complex models. The larger models and increased training parameters lead to higher computational complexity and storage needs, potentially limiting the application of CNNs in scenarios with limited computational capacity [17] In response to these challenges, SqueezeNet was developed. This innovative CNN architecture is designed to significantly reduce the model's parameters and computational resource demands without compromising performance [18] It achieves this goal through unique design strategies, such as employing 1 × 1 convolution filters to reduce the number of parameters, enhancing data flow efficiency with densely connected architectures, and reducing computational burden with delayed activation. These strategies collectively enable SqueezeNet to drastically reduce the model size while maintaining or surpassing larger models' performance [19]

The success of SqueezeNet not only lies in its performance but also in paving the way for new directions in research on efficient neural network architectures. Subsequent research and development have drawn on the design principles of SqueezeNet, further exploring how to reduce computational resource requirements while maintaining or enhancing model performance [18] These studies are significant for advancing the application of deep learning technology in resource-constrained environments, such as mobile devices, embedded systems, and IoT devices. In summary, SqueezeNet represents a significant breakthrough in deep learning, offering valuable design insights and inspiration for future researchers and developers. It demonstrates the feasibility of achieving efficient and powerful deep-learning models under strict resource limitations [20] This achievement further emphasizes the importance of considering model efficiency in designing neural network architectures, pushing towards more intelligent and sustainable advancements in artificial intelligence technology.

B. ATTENTION-BASED MECHANISM

To further enhance the ability to identify cracks in images with noise, increase recognition rates, and improve the accuracy and stability of Convolutional Neural Network (CNN) models, this chapter delves into the innovative integration of attention mechanisms in the crack detection process. The introduction of attention mechanisms is based on their demonstrated potential in artificial intelligence, especially in image processing, where they significantly boost the model's focus on critical areas of the image, thereby improving the efficiency of recognizing details in complex scenes and subtle features [21] By integrating attention mechanisms into the CNN model for crack detection, our goal is to effectively distinguish between noise and actual cracks in images, overcoming the limitations of traditional methods in handling noise interference and enhancing the model's precision and robustness in crack detection [22].

The Squeeze-and-Excitation (SE) attention mechanism innovatively enhances the performance of Convolutional Neural Networks (CNNs) by meticulously adjusting the significance of each channel in the feature maps produced by convolutional layers [23] This mechanism consists of two pivotal steps: Squeeze and Excitation. The Squeeze step aggregates global spatial information for each channel through global average pooling, thus generating a global context representation for each channel. Following this, the Excitation step leverages this global information to learn the recalibration of channel weights, dynamically adjusting the response intensity of each channel through a simple fully connected layer network, enabling the model to focus more on informative feature channels [24] This mechanism effectively improves the model's efficiency in utilizing information across different feature channels, allowing CNN to capture details and global contextual information in images more flexibly. By recalibrating features at the channel level, the SE mechanism enhances the model's understanding of complex visual patterns, leading to significant improvements in performance across various computer vision tasks such as image classification, object detection, and image segmentation [25].

Notably, the design of the SE module allows for its easy integration into existing CNN architectures without imposing excessive computational overhead. This modular and efficient approach boosts the model's accuracy and maintains computational efficiency. Thus, the Squeeze-and-Excitation attention mechanism is not merely an effective means to enhance CNN performance; it represents a significant innovation in improving models' capability to understand and process images. Intelligently recalibrating feature channels offer a powerful tool for deep learning models to achieve efficient and precise visual recognition across a broader range of application scenarios.

C. MODEL EXPLAINABILITY

In the application of deep learning, especially in Convolutional Neural Networks (CNNs), model interpretability has emerged as a critical area of research. These interpretability methods aim to unveil how models learn features from input data and make decisions, providing researchers and practitioners with profound insights. Understanding the decision-making process of models is crucial for validating their accuracy and uncovering any potential biases. As deep learning models increasingly grow in complexity and level of abstraction, ensuring their transparency and interpretability becomes more important than ever. Moreover, model interpretability not only aids in enhancing models' scientific and practical design but also plays a crucial role in boosting user trust in model predictions. We can identify and correct biases within models through visualization techniques and interpretative frameworks, optimize performance, and ensure fairness and transparency. In many domains, especially in high-stakes fields like healthcare, finance, and legal, model interpretability is an indispensable part of decision support systems, crucial for promoting societal acceptance and responsible use of technology [26].

Among interpretability methods for models, Grad-CAM (Gradient-weighted Class Activation Mapping) is recognized as a notably effective and popular technique. Grad-CAM, a visualization tool, utilizes the model's gradient information to underscore the image areas most crucial to the model's predictions [27] This approach is especially suited for elucidating the workings of Convolutional Neural Networks (CNNs) in tasks like image recognition and classification, producing "heatmaps" that delineate the image sections critical for making certain category decisions. The merits of Grad-CAM encompass (1) Intuitiveness: The heatmaps generated enable a visual comprehension of the segments within an image that significantly impacts the model's decision-making process, thus augmenting the decision-making transparency of the model; (2) Versatility: Applicable to any CNN architecture without necessitating structural alterations to the model, Grad-CAM showcases exceptional flexibility and broad applicability; (3) Explanatory Power: Grad-CAM assists both researchers and practitioners in enhancing their understanding of the model's operational mechanisms, fostering trust in the model's predictive capabilities [27].

Overall, the Grad-CAM approach seamlessly connects the high-level abstractions characteristic of CNNs with the demand for practical insights. By offering transparent visualizations of how various features influence the output, it elucidates the often enigmatic processes within deep neural networks. The method provides a powerful means to augment deep learning models, enhancing the model's transparency and facilitating a deeper understanding of the decision-making process. It serves as an effective tool for improving model interpretability.

D. RESEARCH GAPS

In summary, the literature review primarily focuses on the current status and development within the field of crack detection and analysis using deep-learning neural networks. The main research gaps identified in this study are twofold: (1) At the level of methodology, although the combination of SqueezeNet, SE attention mechanism, and Grad-CAM techniques has shown tremendous potential in improving the efficiency and accuracy of crack detection, there is still a lack of a unified framework that can fully leverage the synergistic benefits of these technologies. (2) At the level of application, the quality of crack images is severely affected by environmental factors, such as noise and lighting conditions, posing significant challenges to the performance of image-based crack detection systems [28] Current methods still require improvements, especially in recognizing crack images under low-light conditions. Considering the above-mentioned research gaps, this paper proposes to develop an innovative framework that integrates the SqueezeNet architecture, SE attention mechanism, and Grad-CAM technique, thereby achieving significant improvements in the efficiency, accuracy, and interpretability of crack detection. We believe that through fine-tuning the SqueezeNet model, efficient learning of crack features can be achieved; meanwhile, the introduction of the SE attention mechanism will further enhance the model's focus on crucial crack features, [29] improving recognition precision; finally, by employing the Grad-CAM technique, we aim to enhance the interpretability of the model's decision-making process, providing a more intuitive explanation of the results to the end-users.

In light of the research gaps, we intend to develop a novel framework by integrating the SqueezeNet, SE attention mechanism, and Grad-cam technique, which can potentially yield significant value in improving the efficiency, accuracy, and explainability of crack identification [30] The main research questions to be solved contain: (1) How to effectively select and optimize the SqueezeNet model to meet the specific requirements of crack detection; (2) How to improve and optimize the SE attention mechanism to enhance the model's sensitivity and recognition ability towards key crack features; (3) How to utilize the Grad-CAM technique to improve the explainability of the model's predictions, ensuring that the decision-making process is transparent and understandable to end-users. Specifically, it's crucial to delve into the specifics of the chosen CNN architecture for image recognition, providing justification based on its suitability for crack detection [31] The architecture must be selected with careful consideration of its ability to recognize and accurately classify crack features under various conditions. Additionally, it is essential to elaborate on the selected attention mechanism, outlining the rationale for its choice. The selection should be justified by its ability to focus the model learning on the most relevant features of crack images, thereby improving detection accuracy. Furthermore, it is crucial to discuss the methods employed for interpreting the prediction results. This involves highlighting the importance of ensuring a comprehensive and understandable explanation of the decision-making process. Techniques such as Grad-CAM provide visual explanations for the predictions made by CNNs, highlighting the areas of the image that were most influential in the decision-making process.

This comprehensive approach not only addresses the identified research gaps but also contributes significantly to advancing the field of crack identification using deep learning techniques. By focusing on the optimization and interpretability of the model, alongside the careful selection of the CNN architecture and attention mechanism, this research aims to push the boundaries of what is currently achievable in crack detection, leading to more reliable, efficient, and explainable deep learning solutions.

III. METHODOLOGY

The proposed framework in this paper comprises three main components: data collection and processing, Att-SqueezeNet network for crack identification, and Grad-CAM for visual

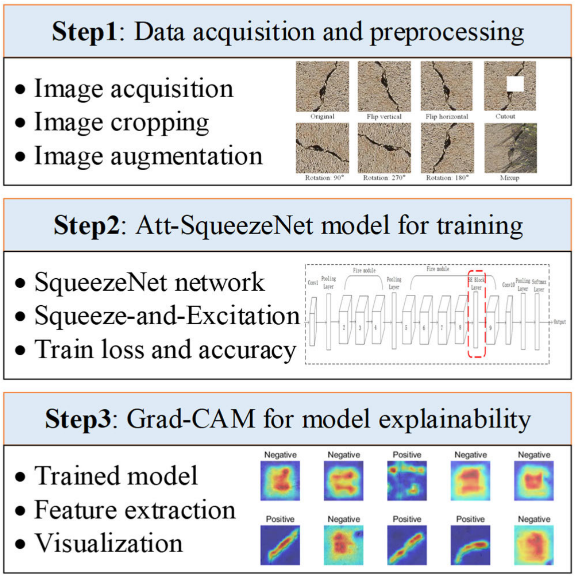

FIGURE 1. The flowchart of the proposed methodology.

explanation of results, as shown in Fig. 1. Initially, crack image datasets are captured using high-speed cameras and cropped to specified dimensions, followed by geometric transformations such as rotation and scaling of the images. Subsequently, these processed images are fed into an Attention-based SqueezeNet (Att-SqueezeNet) network model for model training. Finally, Grad-CAM technology is utilized to produce heat maps that visually highlight the areas where the model focuses during crack detection, providing an intuitive demonstration of the model's attention areas.

A. DATA AUGMENTATION OF CRACK IMAGE

In the field of crack detection, the collection and enhancement of data play crucial roles in improving models' accuracy and generalization ability by simulating various real-world scenarios. The data collection process emphasizes the importance of gathering crack images from diverse angles and lighting conditions to ensure the dataset's variety. This involves collecting images of cracks with varying widths, lengths, shapes and those set against different background conditions. Additionally, the data pre-processing phase involves basic image processing operations such as resizing images and normalizing pixel values to standardize the format and quality of training data.

Advanced data augmentation techniques, including traditional methods like flipping and rotating, as well as sophisticated approaches like Cutout and Mixup, further enhance the model's adaptability to complex and variable real-world environments. Cutout simulates the obstruction of cracks by randomly erasing parts of the image, thus improving the adaptability to environmental changes. On the other hand, Mixup creates virtual training samples by linearly mixing different images and their labels, enhancing the generalization ability toward unseen samples by making the discrete sample space more continuous. The combined application of these augmentation techniques significantly boosts the performance of crack detection models in complex settings, forming an essential strategy for achieving high precision in crack detection. Fig. 2 presents the combined application of the common data augmentation techniques.

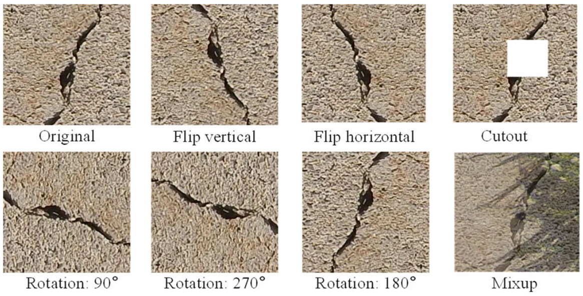

FIGURE 2. Combined application of the common data augmentation techniques.

B. ATTENTION-BASED SQUEEZENET (ATT-SQUEEZENET) MODEL

To enhance the accuracy of crack detection, we employ the attention-based SqueezeNet (Att-SqueezeNet), a refined and efficient convolutional neural network modeled after the AlexNet algorithm. This network works by incorporating a deep compression technique to minimize the number of parameters, thus streamlining the model without compromising its accuracy significantly. Att-SqueezeNet's architecture cleverly incorporates several Fire modules alongside traditional convolutional network elements: convolutional, downsampling, and fully connected layers. As illustrated in Fig. 3, the Att-SqueezeNet structure includes pooling layers, Fire modules, an SE (Squeeze-and-Excitation) block layer, and a SoftMax layer. The maximum pooling layer is designed to reduce the computational demands of the model. Additionally, the Dropout layer selectively deactivates neurons to avert overfitting. The SoftMax classifier is then employed to determine the model's final output by selecting the class with the utmost probability. By integrating these key components, Att-SqueezeNet not only streamlines the crack detection process but also significantly optimizes the utilization of computational resources with its distinctive architectural design. Consequently, Att-SqueezeNet offers a precise and efficient methodology for crack detection.

The essence of SqueezeNet lies in its Fire module, a distinctive combination of Squeeze and Expand structures. Typically defined as Fire (M, N, E1, E2), where M and N represent the number of input and output channels for the Squeeze layer, respectively, and E1 and E2 denote the count of 1 × 1 and 3 × 3 convolutional kernels in the Expand layer. The Squeeze structure employs 1 × 1 convolutions to process the feature map output from the preceding layer, effectively reducing feature dimensions. This step is followed by activation using the ReLU function. Subsequently, the Expand structure, comprising both 1 × 1 and 3 × 3 convolutional kernels, sequentially enlarges the convolutional results.

After the convolution process, the ReLU activation function is applied, as illustrated in Fig 4. The calculation formula of the Squeeze layer is expressed as Eq. (1). Subsequently, the Expand layer extends the features using both 1 × 1 and 3 × 3 convolution kernels, as depicted in Eq. (2).

$$S(x) = f(W_s * x + b_s) \tag{1}$$

$$E(x) = f(W_{e1} * x + b_{e1}) + f(W_{e2} * x + b_{e2}) \tag{2}$$

where x is the input feature map; W_s represents the weight of the 1 × 1 convolution kernel; b_s is the bias term; $*$ denotes the convolution operation; f is the ReLU activation function.

In SqueezeNet, we choose to insert the SE (Squeeze-and-Excitation) module after Fire 8, which belongs to the output of the Fire module because the Fire module's output contains rich high-level semantic information. Calibrating it with an attention mechanism can effectively enhance the quality of feature representation. Furthermore, placing the SE module in the mid-to-late stages of the model allows leveraging higher-level information for more precise feature recalibration, thereby improving the final classification performance.

Specific reduction ratios and activation functions are chosen within the SE module to optimize performance. Typically, reduction ratios of 4 or 8 are used to reduce the parameter and computational overhead during the Excitation phase while maintaining adequate expressive capability. For activation functions, ReLU introduces non-linearity, and the final Sigmoid function confines weights to [0, 1], ensuring stability and interpretability in weight adjustments. These parameter choices balance the lightweight requirements of the model with the effectiveness of attention mechanisms.

Incorporating the Squeeze-and-Excitation (SE) attention mechanism into the model significantly enhances its capability to process complex image features. This mechanism operates by compressing the feature channels (Squeeze) to capture global information, followed by utilizing an excitation function (Excitation) to learn the importance of each channel, thus enabling adaptive recalibration of features. This advancement bolsters SqueezeNet's ability to handle intricate image characteristics, enhancing the model's precision and speed. The SE module compresses feature channels through global average pooling and then uses the weights of fully connected layers to learn the importance of each channel, as described in Eqs. (3)-(4). The original feature map is then element-wise multiplied by the importance weights for feature recalibration, as shown in Eq. (5). Through the unique structural design, Att-SqueezeNet not only simplifies the traditional crack detection process but also optimizes the use of computational resources through fine-grained feature adjustment, providing an accurate and efficient solution for crack detection. This network, which combines compact convolutional kernel design with an attention mechanism, is particularly suitable for high-performance crack detection tasks in resource-constrained environments.

$$z_c = \frac{1}{H \times W} \sum\nolimits_{i=1}^{H} \sum\nolimits_{j=1}^{W} x_c(i, j) \tag{3}$$

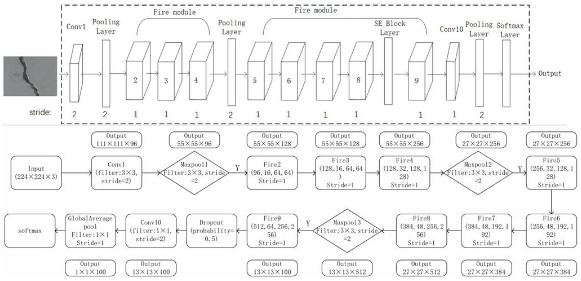
Fire module
Fire module
Conv1
Pooling Layer
Pooling Layer
SE Block Layer
Conv10
Pooling Layer
Softmax Layer
2
3
4
5
6
7
8
9
Output
stride: 2 2 1 1 1 2 1 1 1 1 1 1 2
Output 111×111×96
Output 55×55×96
Output 55×55×128
Output 55×55×128
Output 55×55×256
Output 27×27×256
Output 27×27×256
Input (224×224×3)
Conv1 (filter:3×3, stride=2)
Maxpool1 Filter:3×3, stride =2
Y
Fire2 (96, 16, 64, 64) Stride=1
Fire3 (128, 16, 64, 64) Stride=1
Fire4 (128, 32, 128, 1 28) Stride=1
Maxpool2 Filter:3×3, stride =2
Y
Fire5 (256, 32, 128, 1 28) Stride=1
softmax
GlobalAverage pool Filter:1×1 Stride=1
Conv10 (filter:1×1, stride=2)
Dropout (probability= 0.5)
Fire9 (512, 64, 256, 2 56) Stride=1
Y
Maxpool3 Filter:3×3, stride =2
Fire8 (384, 48, 256, 2 56) Stride=1
Fire7 (384, 48, 192, 1 92) Stride=1
Fire6 (256, 48, 192, 1 92) Stride=1
Output 1×1×100
Output 13×13×100
Output 13×13×100
Output 13×13×512
Output 27×27×512
Output 27×27×384
Output 27×27×384

FIGURE 3. The architecture of the proposed Att-SqueezeNet model.

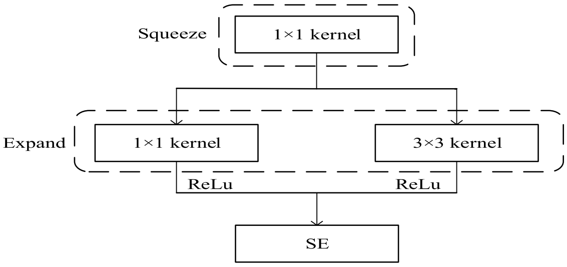

FIGURE 4. The structure of the fire module with SE attention mechanism.

$$s = \sigma(W_2 \delta(W z_1)) \tag{4}$$

$$\tilde{x}_c = s_c \cdot x_c \tag{5}$$

where z_c is the output of global average pooling; $x_c\,(i, j)$ represents the feature value at location $(i, j)\,H$ and W are the height and width of the feature map; W_1 and W_2 are the weights of the fully connected layers; δ is the ReLU function; σ is the sigmoid function; s represents the importance weights of each channel.

Introducing the SE (Squeeze-and-Excitation) attention mechanism significantly enhances the performance of the SqueezeNet model, primarily through feature recalibration, improved information flow, and lightweight advantages. The SE module reweights the feature maps of each channel based on global contextual information, effectively enhancing the representation of key features and significantly improving the model's ability to capture inter-layer information flow. Furthermore, the SE module has a small parameter count, which does not significantly increase the computational burden yet substantially enhances the model's representative capacity and classification performance dramatically. These improvements enable the model to achieve better classification results without increasing computational complexity.

C. GRAD-CAM FOR MODEL EXPLAINABILITY

Grad-CAM (Gradient-weighted Class Activation Mapping) is a visualization technique used to interpret decisions made by convolutional neural networks. It highlights crucial areas in an image that significantly influence the model's specific classification decision by generating heatmaps. Grad-CAM is a visualization technique that generates heat maps to reveal key regions for convolutional neural network (CNN) decisions. This method begins by extracting feature maps from a selected convolutional layer of the Att-SqueezeNet, capturing spatial information learned from the input image. The gradient of the network's prediction for a particular category relative to the feature maps is calculated. These gradients are then globally averaged to assign weights to each feature map, indicating their contribution to the final classification decision. By multiplying these weights with their corresponding feature maps and summing them up, a heatmap can be superimposed on the original image to highlight areas of the model's focus visually.

We conducted multiple case studies to show the Grad-CAM technique's practical contribution. The heatmap clearly demonstrates the model's focus on crack edges and critical regions for clear images. In images under low light conditions, the Grad-CAM technique shows the model's ability to identify cracks accurately even when the light is low. In images containing distractors, the heatmap proves that the model is able to ignore background interference and focus on the crack itself. However, in images with obstacles, Grad-CAM reveals the focus of the model when dealing with partially occluded cracks. Through these case studies, the Grad-CAM technique significantly enhances the interpretability of the model and enables users to gain insight into the model's decision-making process. This enhanced interpretability is crucial for practical engineering applications, helping users understand and trust the model's predictions.

In the computational process of Grad-CAM, let Y^c represent the network's predicted score for category c (before Softmax activation), and A^k denotes the $k-th$ feature map of a convolutional layer. The method first computes the gradient of Y^c with respect to A^k, represented as $\partial Y^c/\partial A^k$. This gradient is then subjected to global average pooling to determine the weight α_k^c of each feature map. The final heatmap $L_{\text{Gard-CAM}}^c$ is obtained using Eq. (6). The specific structure is shown in Fig. 5 below. This approach of Grad-CAM provides an intuitive means to understand and elucidate the decision-making process of models, especially highlighting areas focused on by the model in image classification tasks.

$$L_{\text{Gard-CAM}}^c = \text{ReLU}(\sum\nolimits_k \alpha_k^c A^k) \tag{6}$$

where the ReLU function is employed to isolate features contributing positively to category c.

To effectively integrate Grad-CAM technology into the Att-SqueezeNet network proposed in this study, a comprehensive set of steps has been developed: (1) the trained Att-SqueezeNet model and a diverse database of crack images are loaded to ensure the model's adaptability to varied real-world scenarios; (2) the network undergoes forward propagation to yield predictive outcomes for the input images, a crucial step that demonstrates the network's capability in crack feature recognition; (3) backpropagation is employed to accurately determine the gradient of the target class concerning the last convolutional layer of the model, a pivotal calculation for understanding the focal points of the model; (4) the Class Activation Map (CAM) is computed and merged with the original image, offering a visual representation that delineates the areas of primary focus in the crack identification task. This method of incorporating Grad-CAM not only enhances the accuracy of crack detection by the model but also significantly boosts its interpretability. Through visualization, it becomes evident how the model identifies cracks, providing deep insights into the underlying decision-making process. This enhancement in the model's credibility and explanatory power is vital for broader application and acceptance.

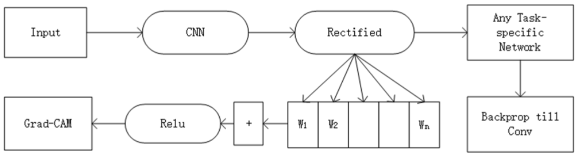

FIGURE 5. The flowchart of the Grad-CAM method for feature extraction.

IV. CASE STUDY

A. DATA ACQUISITION AND PREPROCESSING

The proposed framework is applied to actual engineering projects to identify crack-related diseases, verifying the model effectiveness in recognition tasks. This study employed a Canon camera (model: Canon DIGITAL IXUS 100 IS) as the image acquisition device. A total of 1,000 original images were collected, including a self-collected dataset of residential building cracks, including clear images, images with low lighting, images with distractions, and images with obstructions. Images from the Crack Forest Dataset (CFD) contain 118 images of urban concrete road cracks at approximately 480 × 320 pixels resolution. As shown in Fig. 6, the upper part shows images captured on-site, while the lower part depicts images from CFD. "Negative" denotes no cracks, and "Positive" denotes the presence of cracks. The dataset in this study is a fusion of actual field-collected data and existing databases, encompassing various types of cracks such as linear, grid-like, and patch-like cracks, observed in different environmental conditions. This effectively reflects the real crack environments in diverse building locations and make the model more adaptive. To reduce computational costs, original images were cropped to a resolution of 227 pixels by 227 pixels and standardized in size.

B. MODEL TRAINING AND ANALYSIS

The dataset comprises two folders: images with cracks and without. The folder containing cracked images encompasses various types and degrees of cracked images and diverse background noise possibilities. Such a dataset aids the model in better adapting to real-world scenarios and enhances its robustness. To ensure a sufficient sample size and prevent overfitting, image augmentation techniques such as mirroring, rotation, and Gaussian blur were applied, expanding the dataset by five times to a total of 5,000 images. The dataset was divided into training, validation, and test sets with a ratio of 70%:20%:10%, respectively. This information can be presented in TABLE 1. This comprehensive data collection and augmentation approach underlines the rigorous

TABLE 1. Experimental dataset information.

Crack Images	Negative	Positive
Training Set	1750	1750
Validation Set	500	500
Test Set	250	250
Total	2500	2500

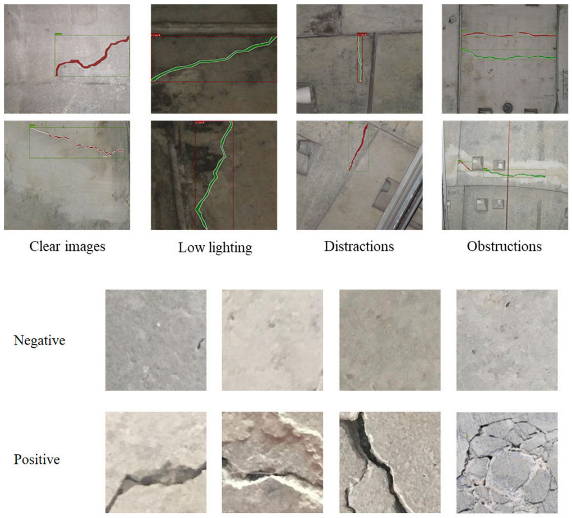

FIGURE 6. Partial training set images.

methodology employed to validate the effectiveness of the proposed algorithm in recognizing crack-related diseases in building structures.

All experiments in this study were conducted using MATLAB software. The Adam optimizer was utilized for parameter updates, with an initial learning rate set at 2e-4. The network employs a cross-entropy loss function for training, as shown in Eq. (7). Model recognition accuracy was assessed by calculating the accuracy metric.

$$L = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{7}$$

where the ReLU function is employed to isolate features that contribute positively to category c.

In our study, to further enhance the recognition rate of the SqueezeNet model in crack detection tasks with noisy points, we introduced an attention mechanism by incorporating it into the fire8 layer of the SqueezeNet model. This attention module was inspired by the design of Squeeze-and-Excitation (SE) Blocks, where the weights learned through training are used to weight the feature maps. This allows the model to identify crack areas and extract crack features accurately.

Fig. 7 shows the loss and accuracy curves for both the training and validation processes throughout the model's training period. It is evident from the graph that the training loss experiences a rapid decline during the initial ten iterations, while the training accuracy sees a significant increase within the same timeframe. Following this initial period, both metrics reach a state of stable convergence, indicating the model's practical learning and capacity to generalize the training data well. The model achieves a final loss of 0.064 and an accuracy of 0.995, underscoring its high efficiency in recognizing crack images. This performance level indicates the model's robustness and potential applicability in real-world scenarios, where high precision in crack detection is paramount for ensuring structural safety. The convergence of the loss and accuracy to such optimal values reinforces the chosen architecture's effectiveness and the training regime's adequacy, including the optimization strategy and data augmentation techniques employed. Furthermore, the validation loss and accuracy curves closely mirror those of the training, with minimal deviation, suggesting that the model is not overfitting and maintains a good generalization on unseen data. The slight fluctuations in the validation curves towards the later iterations emphasize the importance of continuous monitoring and possibly implementing early stopping mechanisms to preserve the model performance.

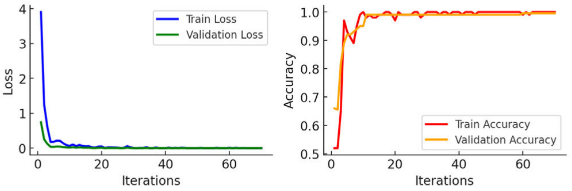

FIGURE 7. The variation of loss and accuracy in the training and validation process.

Fig. 8 utilizes the Grad-CAM-based visualization technique to depict the color map of crack identification, providing a clear visual representation of the areas focused on by the model during crack detection. In-depth analysis and visualization of the features from the intermediate layers of the model reveal a significant enhancement in the model's sensitivity and focus on crack regions upon integrating the attention mechanism. This shift in focus not only improves the accuracy of crack detection but also enhances the model's robustness in complex environments, maintaining high detection performance even under challenging conditions such as high noise levels or suboptimal lighting. The heatmaps generated by Grad-CAM serve as a powerful tool for validation, offering a direct visual insight into the primary areas considered by the model when making decisions. The highlighted regions in these heatmaps confirm that the model's decision-making logic is in line with expectations and demonstrate how the attention mechanism enables the model to focus more on the crucial crack features, thus enhancing the overall performance of crack detection.

In summary, these observations underscore the pivotal role of the attention mechanism in improving crack detection models, particularly in terms of enhancing the recognition of key features and the precision of decision-making. By combining advanced visualization techniques with cutting-edge attention mechanisms, we can deepen our understanding of how models operate and further refine model designs to tackle a variety of complex crack detection challenges.

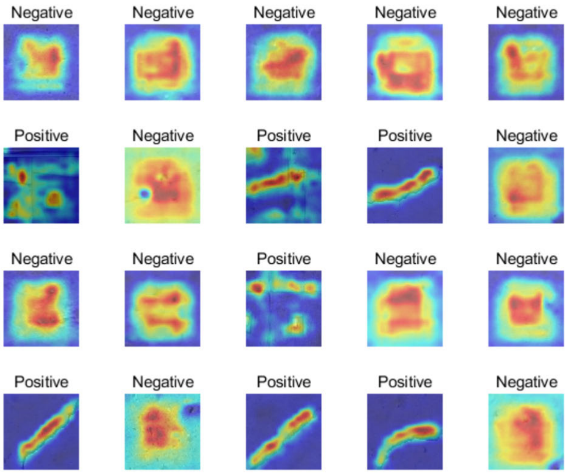

FIGURE 8. The visualization colormap of crack identification based on the Grad-CAM method. (Note: Positive=Crack; Negative=No crack.)

TABLE 2. Training configuration.

Candidate models	Learning Rate	Optimizer	Batch Size	Epochs
CNN	0.001	Adam	32	30
B-CNN	0.0005	Adam	32	30
VGG-19	0.0001	Adam	16	30
SqueezeNet	0.0002	Adam	100	30
Att-SqueezeNet	0.0002	Adam	100	30
RF	n_estimators=100, random_state=42, max_features=auto			

C. COMPARISON EXPERIMENTS

In our study, we conducted a comprehensive comparison between our proposed model and other state-of-the-art models, encompassing both traditional machine learning approaches like the Random Forest (RF) algorithm and advanced deep learning architectures, including Convolutional Neural Networks (CNN), Region-based Convolutional Neural Networks (R-CNN), VGG-19, and the compact SqueezeNet model. The specific parameters of each model are given in the following Table 2.

These models were evaluated on their ability to detect cracks, showcasing a spectrum of performance levels and computational complexities. To assess the efficacy of each model in crack detection, we utilized critical evaluation metrics, namely Precision (P), Recall (R), and F1-score ($F1$). The Equations are as follows:

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2PR}{P + R} \tag{10}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Table 3 depicts the comparison results with other state-of-the-art models. Fig. 9 presents the corresponding histograms of performance comparison of different models. In the comparative analysis of models for crack detection, the performance metrics indicate varied capabilities across the spectrum of traditional machine learning and advanced deep learning models. The Random Forest (RF) model, representing traditional algorithms, demonstrates respectable performance with balanced Precision and Recall. It indicates its competency in identifying relevant crack instances while maintaining moderate false positives. On the deep learning front, models like CNN and VGG-19 exhibit strong feature extraction capabilities, as reflected in their high Recall values, ensuring minimal misses in crack detection. Particularly noteworthy is the B-CNN model, which achieves near-perfect Precision and Recall, underscoring its exceptional accuracy in crack identification.

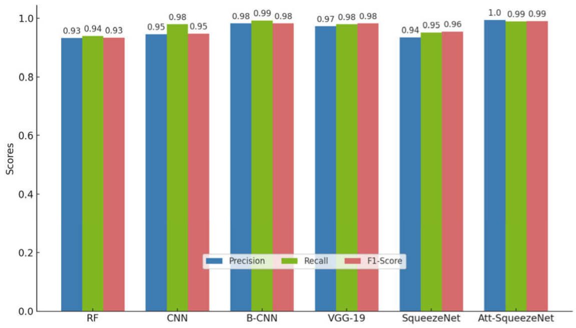

FIGURE 9. The histograms of performance comparison of different models.

TABLE 3. The comparison results with other state-of-the-art models.

Candidate models	Size	Parameter counts	Precision	Recall	F1-Score	Training time(s)
RF	4.05M	1.02M	0.933	0.940	0.934	30
CNN	98.46M	25.64M	0.946	0.980	0.948	272
B-CNN	1.06G	276.73M	0.983	0.993	0.983	402
VGG-19	548.20M	143.67M	0.973	0.980	0.983	306
SqueezeNet	4.80M	1.24M	0.935	0.952	0.955	96
Att-SqueezeNet	5.00M	1.25M	0.995	0.990	0.991	133

The proposed Att-SqueezeNet method integrates an attention mechanism with the efficient architecture of SqueezeNet, achieving unparalleled Precision and Recall among the evaluated models. This enhancement allows the model to focus more precisely on crucial features for crack detection, significantly reducing false positives and ensuring comprehensive coverage of true positives. The resulting F1-Score of 0.991 for Att-SqueezeNet confirms its balanced performance and highlights its superiority in accurately detecting cracks. This makes Att-SqueezeNet particularly advantageous for applications where accuracy, efficiency, and reliability are critical, establishing it as a leading method in the field of crack detection.

Moreover, the training time presented in the table further highlights the advantage of Att-SqueezeNet in quickly identifying cracks. Att-SqueezeNet's training time is only 133 seconds, much faster than many other complex deep learning models such as VGG-19 (306 seconds) and B-CNN (402 seconds). This indicates that Att-SqueezeNet not only surpasses other models in terms of accuracy but also excels in training efficiency and has rapid identification. Furthermore, Att-Squeezenet has a size of only 5.00M, close to RF's 4.05M (the smallest size among the model candidates). Moreover, Att-Squeezenet has 1.25M parameters, which is relatively small among all candidate models, indicating its small footprint, ease of embedding, high efficiency, and fast computation and communication speeds. Att-SqueezeNet offers a fast and reliable solution for large-scale crack detection applications requiring rapid deployment and efficient processing. This makes it particularly advantageous for applications where accuracy, efficiency, and reliability are critical, establishing its leading position in the field of crack detection.

V. CONCLUSION

This study proposes a novel framework that integrates the SE attention mechanism with the SqueezeNet network and employs Grad-CAM technology specifically designed for the task of crack identification in buildings. In short, the research mainly contributes to two aspects: (a) From the state of knowledge, this study advances the theoretical understanding of integrating attention mechanisms with the SqueezeNet model for visual tasks. The SE attention mechanism dynamically captures channel feature responses, enhancing the model's focus on critical crack features. The Grad-CAM technology further contributes by offering a novel approach to visualizing and interpreting the model's internal workings. (b) From the state of practice, this framework demonstrates effective practical benefits for building maintenance. The enhanced Att-SqueezeNet significantly improves detection accuracy and efficiency by focusing more effectively on crack features. Grad-CAM visualization offers intuitive explanations of the model's decision-making process, facilitating a clearer understanding of detection results. This advancement not only automates building maintenance by reducing reliance on manual labor but also increases the precision and reliability of detection tasks, ultimately ensuring the structural safety of buildings through more effective and reliable crack identification. The main results and conclusions of this work are summarized as follows:

(1) The integration of the SE attention mechanism enables the Att-SqueezeNet network to achieve higher accuracy while effectively avoiding the problem of low crack recognition caused by complex backgrounds and uneven sunlight exposure around the cracks. This enhancement improves the model's recognition accuracy in complex environmental conditions, making it more adept at handling various challenging scenarios in real-world construction crack detection tasks and improving its adaptability.

(2) Through comprehensive experiments across the collected dataset, Att-SqueezeNet was benchmarked against a range of models, including RF, CNN, B-CNN, VGG-19, and the original SqueezeNet. Evaluation metrics such as precision, recall, and F1-score were employed for this comparison. Att-SqueezeNet outperformed the other models with an impressive accuracy of 0.995 and an F1-score of 0.991, the highest among the compared models, indicating a highly effective tool for crack identification.

(3) The combination of Att-SqueezeNet and Grad-CAM significantly bolsters the model's interpretability, facilitating the accurate detection of cracks and providing a transparent understanding of the decision-making process. Such clarity is instrumental in enabling stakeholders to take informed actions based on the finding results. Their combination can act as a pivotal advancement in construction engineering crack detection, offering high accuracy and actionable insights.

(4) Att-SqueezeNet's faster training speed compared to other complex deep learning models provides a significant advantage in its application to crack detection. The rapid training speed not only shortens the model development and optimization time but also reduces deployment and maintenance costs, allowing more practical engineering applications to benefit from efficient crack detection technology. This high-efficiency training capability enables Att-SqueezeNet to quickly adapt to and handle diverse types of cracks in complex building structures. As technology advances and application scenarios expand, the model's fast training speed provides crucial support for real-time or near-real-time crack detection, further driving the development and application prospects in this field.

Some limitations need to be further improved. For example, when compared to SqueezeNet, integrating an attention mechanism in Att-SqueezeNet results in increased parameters, potentially leading to decreased operational efficiency. The image dataset may also be limited, failing to encompass all unique scenarios. Future research could focus on refining the network structure and attention mechanism to accommodate better the diverse challenges presented by complex crack detection contexts.

A series of improvements can be carried out to overcome the existing method limitations. Firstly, in terms of optimizing the network structure and attention mechanism, parameter pruning and quantization techniques are considered to reduce the number of parameters of the model to improve the computational efficiency while maintaining the model's accuracy as much as possible. Furthermore, exploring more efficient attention mechanisms, such as lightweight SE modules or other variants, can reduce computational overhead and maintain high performance. At the same time, using more advanced network architecture design, such as EfficientNet, can improve computational efficiency while maintaining high performance.

Secondly, expanding and diversifying the dataset is the key to improving the model's generalization ability. Using data augmentation techniques such as rotation, scaling, cropping, and color transformation can increase the data diversity and enhance the model adaptability under different environmental conditions. At the same time, generating synthetic data to simulate various environmental conditions and fracture types helps to enrich the dataset. Data sets from other fields are combined to further improve the model adaptability.

Future research should focus on the following aspects. On the one hand, the network structure and attention mechanism are optimized to further reduce the number of parameters and computational complexity and improve the model operation efficiency. Through more efficient network design and attention mechanism, the consumption of computational resources is reduced while maintaining or even improving the model performance. On the other hand, the dataset was expanded and diversified, and more crack images under different environmental conditions were collected and introduced to enhance the generalization ability and adaptability of the model. The rich data set will help the model maintain efficient crack recognition performance in a broader range of practical application scenarios.

REFERENCES

[1] M. Mishra, P. B. Lourenço, and G. V. Ramana, "Structural health monitoring of civil engineering structures by using the Internet of Things: A review," *J. Building Eng.*, vol. 48, May 2022, Art. no. 103954.

[2] R. Ali, J. H. Chuah, M. S. A. Talip, N. Mokhtar, and M. A. Shoaib, "Structural crack detection using deep convolutional neural networks," *Autom. Construct.*, vol. 133, Jan. 2022, Art. no. 103989.

[3] C. V. Dung and L. D. Anh, "Autonomous concrete crack detection using deep fully convolutional neural network," *Autom. Construct.*, vol. 99, pp. 52–58, Mar. 2019.

[4] C. M. Yeum and S. J. Dyke, "Vision-based automated crack detection for bridge inspection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 30, no. 10, pp. 759–770, Oct. 2015.

[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jul. 2014.

[6] E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis, and T. Stathaki, "Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing," *Int. J. Speech Technol.*, vol. 49, no. 7, pp. 2793–2806, Jul. 2019.

[7] W. Ye, J. Ren, C. Lu, A. A. Zhang, Y. Zhan, and J. Liu, "Intelligent detection of fastener defects in ballastless tracks based on deep learning," *Automat. Construct.*, vol. 159, no. 1, p. 105280, 2024, Art. no. 105280.

[8] L. Zhang, L. Zhang, and H. Du, "Crack detection and measurement with computer vision," *Robot. Comput.-Integr. Manuf.*, vol. 39, pp. 57–69, Jul. 20116.

[9] K. Li, R. Wan, Y. Wang, Y. Liang, and Q. Chen, "A comparative study of deep learning in crack detection," *Neurocomputing*, vol. 324, pp. 103–116, Jul. 2019.

[10] J. Fareleira, M. Melo, and F. Simaes, "Deep learning techniques for automatic crack detection and classification," *Appl. Sci.*, vol. 11, no. 2, p. 692, 2021.

[11] Y. Zhu and H. Tang, "Automatic damage detection and diagnosis for hydraulic structures using drones and artificial intelligence techniques," *Remote Sens.*, vol. 15, no. 3, p. 615, Jan. 2023.

[12] Y. Xu, Y. Bao, J. Chen, W. Zuo, and H. Li, "Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images," *Struct. Health Monitor.*, vol. 18, no. 3, pp. 653–674, May 2019.

[13] L. Mingyue, H. Lesheng, and G. Youmei, "Fine-grained image classification model based on attention feature fusion with SqueezeNet," *Natural Science*, vol. 43, no. 5, pp. 868–876, 2021.

[14] J. Zhou, Y. Zhang, J. Yan, and L. Quan, "Crack classification with deep convolutional neural networks for bridge inspection," *Sensors*, vol. 19, no. 8, pp. 1–20, 2019.

[15] Y. N. Silva and U. J. Nunes, "Crack detection in pavements based on pretrained deep convolutional neural networks," *Expert Syst. Appl.*, vol. 100, pp. 250–259, Jul. 2018.

[16] S. Sony, K. Dunphy, A. Sadhu, and M. Capretz, "A systematic review of convolutional neural network-based structural condition assessment techniques," *Eng. Struct.*, vol. 226, Jan. 2021, Art. no. 111347.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[18] W. Jixiao, L. Yang, and Z. Yangshuo, "Light-weight image fusion method based on SqueezeNet," *J. Comput. Appl.*, vol. 40, no. 3, pp. 837–841, 2020.

[19] E. Akleman, "Deep learning," *Computer*, vol. 53, no. 9, pp. 1–17, Sep. 2020.

[20] W. Wenxiu, Z. Peng, and Z. Jiaqi, "Rods surface defect identification based on improved SqueezeNet," *J. Electron. Meas. Instrum.*, vol. 37, no. 4, pp. 240–249, 2023.

[21] Y.-J. Xiong, Y.-B. Gao, H. Wu, and Y. Yao, "Attention U-Net with feature fusion module for robust defect detection," *J. Circuits, Syst. Comput.*, vol. 30, no. 15, pp. 1–29, Dec. 2021.

[22] C. Haihan, W. Guodong, and T. Hong, "Research advances on deep learning recommendation based on attention mechanism," *Comput. Eng. Sci.*, vol. 43, no. 2, pp. 370–380, 2021.

[23] M. H. Guo, T. X. Xu, J. J. Liu, Z. N. Liu, P. T. Jiang, T. J. Mu, S. H. Zhang, R. R. Martin, M. M. Cheng, and S. M. Hu, "Attention mechanisms in computer vision: A survey," *Comput. Visual Media*, vol. 8, no. 3, pp. 331–368, 2022.

[24] Y. Zhang and K. Yuen, "Crack detection using fusion features-based broad learning system and image processing," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 12, pp. 1568–1584, Dec. 2021.

[25] G. Guangshang, "Survey on attention mechanisms in deep learning recommendation models," *Comput. Eng. Appl.*, vol. 58, pp. 9–18, Aug. 2022.

[26] S. Wang and Y. Zhang, "Grad-CAM: Understanding AI models," *Comput., Mater. Continua*, vol. 76, no. 2, pp. 1321–1324, 2023.

[27] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020.

[28] Y. Yu, B. Samali, M. Rashidi, M. Mohammadi, T. N. Nguyen, and G. Zhang, "Vision-based concrete crack detection using a hybrid framework considering noise effect," *J. Building Eng.*, vol. 61, Dec. 2022, Art. no. 105246.

[29] Z. Chenjia, Z. Lei, and Y. Lu, "Review of attention mechanism in convolutional neural networks," *Comput. Eng. Appl.*, vol. 57, no. 20, pp. 64–72, 2021.

[30] Z. Liu, Y. Cao, Y. Wang, and W. Wang, "Computer vision-based concrete crack detection using U-net fully convolutional networks," *Autom. Construct.*, vol. 104, pp. 129–139, Aug. 2019.

[31] F. Liu and L. Wang, "UNet-based model for crack detection integrating visual explanations," *Construct. Building Mater.*, vol. 322, Jul. 2022, Art. no. 126265.

JIANG-YI LUO is currently pursuing the bachelor's degree with Xi'an University of Architecture and Technology. His research interests include green building and concrete damage monitoring.

YU-CHENG LIU is currently pursuing the bachelor's degree with the School of Mathematical Sciences, Chengdu University of Technology. His research interest includes machine learning.