International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 4 ǁ April 2024 ǁ PP. 130-138

Concrete crack image detection based on optimized lightweight convolutional neural network MobileNetV2

Fujie Yang

*[*1](University of Shanghai for Science and Technology， Shanghai 200093, China)*

Abstract: *In the realm of urban infrastructure management, the prompt and accurate detection of concrete cracks is vital for ensuring safety and extending the lifespan of constructions. Traditional methods, while effective, often involve excessive costs and intense labor. This study introduces a cutting-edge approach that utilizes Convolutional Neural Networks (CNNs), specifically adapting the efficient MobileNetV2 architecture, to autonomously identify concrete cracks using high-resolution imagery. Utilizing the "Concrete Crack Images for Classification" dataset—which includes 40,000 diverse images—our modified CNN model has been optimized for not only high-accuracy desktop analysis but also for mobile and drone-based applications, enhancing its utility in complex environments. Demonstrating a remarkable validation accuracy of 99.23%, our approach outstrips conventional methods, offering a scalable, cost-efficient solution for real-time, comprehensive infrastructure monitoring. This breakthrough serves as a foundation for future advancements in intelligent infrastructure assessment, driving towards more automated and precise monitoring techniques.*

Keywords: *MobileNetV2, Convolutional Neural Networks, Concrete Crack Detection, Deep Learning, Image Processing, Real-Time Analysis.*

Date of Submission: 02-04-2024 Date of acceptance: 10-04-2024

I. INTRODUCTION

In the global context, health monitoring of infrastructure has progressively become a pivotal task in urban management, particularly for aging bridges and roads. Timely and accurate detection of concrete cracks is crucial not only for extending their service life but also for preventing tragic incidents. Although traditional crack detection techniques such as acoustic and electromagnetic testing are widely employed, they often rely on labor-intensive operations and incur high costs, with their applicability being significantly limited in complex environments. To address these limitations, this study explores a revolutionary approach: using Convolutional Neural Networks (CNNs) for image-based automatic detection of concrete cracks.

With the rapid advancement of computer vision and machine learning technologies, particularly deep learning, CNNs have demonstrated superior performance in image recognition and classification across various fields. Compared to traditional methods, CNN-based crack detection not only significantly enhances the automation and accuracy of the detection process but also reduces operational costs. This paper presents breakthrough developments in this technology, including how deep learning models process and analyze large-scale crack image datasets and their practical applications in real-world infrastructure monitoring.

Furthermore, this study details the training and validation process of CNN models using the "Concrete Crack Images for Classification" dataset, which includes nearly 40,000 high-resolution images of cracked and non-cracked surfaces under various surface treatments and lighting conditions. By innovatively adjusting the MobileNetV2 architecture, an innovative network structure, we have developed a lightweight yet efficient model. This model is not only suitable for large-scale computations on servers but can also be deployed on edge computing devices such as mobile devices and drones, significantly expanding its application scenarios and practicality.

This paper not only elucidates an efficient and cost-effective solution for crack detection but also opens new perspectives and possibilities for future research in the domain of infrastructure maintenance and safety monitoring. As technologies continue to evolve and their applications expand, future infrastructure monitoring is expected to become more intelligent and automated, and this research is a part of that evolutionary process.

II. DEVELOPMENT OF CONCRETE CRACK DETECTION METHODS

2.1 Traditional Methods of Concrete Crack Detection

This section discusses traditional methods for detecting concrete cracks, broadly categorized into acoustic wave-based methods and image vision-based methods.

2.1.1 Acoustic Wave-Based Detection Methods

In Japan, Acoustic Emission (AE) technology has been widely applied since the 1990s to assess the integrity of concrete structures. Researchers like Ohno and Ohtsu[1] have used AE to differentiate types of concrete cracks, and Aggelis[2] has further refined AE parameters to improve crack distinction. However, despite its effectiveness, AE requires sophisticated equipment and is prone to noise interference, which limits its utility in complex environments.

2.1.2 Image Vision-Based Detection Methods

Over the past few decades, image vision technology has undergone significant development. Initially, these methods relied primarily on subjective and time-consuming manual visual inspections; Now, it's turning to automated machine vision technology. The Digital Image Correlation (DIC) technique is widely used in the study of concrete fractures. Ziou and Abbou's study[3] points out that despite advances in image processing algorithms, problems such as noise interference remain a challenge.

2.2 Current Research in Deep Learning-Based Crack Detection

The advent of computer vision and machine learning has ushered in a new era of automated crack detection. Deep learning, particularly using Convolutional Neural Networks (CNNs), has revolutionized this field with its robust feature extraction capabilities.

Initially, traditional image processing methods such as edge detection and image binarization were commonly used. These techniques, while straightforward, were often hindered by image noise and other algorithmic limitations.

As the technology evolved, more sophisticated methods like DIC and wavelet transforms were incorporated to enhance the sensitivity and reliability of detection systems. Contemporary research now leverages complex models like CNNs, which can autonomously learn intricate features from extensive datasets of crack images, thus substantially improving both accuracy and efficiency. For instance, CNNs developed by Ciresan[4] et al. have excelled in image recognition tasks, and models by Cha[5] et al. have shown remarkable adaptability and precision in detecting concrete cracks.

Moreover, the integration of deep learning methods with drone technology allows for automated inspections in inaccessible areas[6], reducing manual inspection risks and improving the thoroughness and systematization of the detection processes.

Despite these advancements, deep learning-based crack detection still faces significant challenges, including a high dependence on image quality and a limited ability to recognize various crack types. Future research will need to focus on algorithm optimization, enhancing data diversity, and developing real-time detection systems to overcome these limitations.

III. DATA PREPARATION AND RESEARCH DESIGN

3.1 Dataset and Preprocessing

3.1.1 Dataset Overview

This study utilizes the "Concrete Crack Images for Classification" dataset[7~9], curated by Çağlar Fırat Özgenel and publicly available on Mendeley Data since 2019. The dataset encompasses a total of 40,000 images, systematically partitioned into two distinct categories, each comprising 20,000 images. Figure 1 illustrates the positive class images, characterized by the presence of cracks, while Figure 2 presents the negative class images, which are devoid of any cracks. The images are standardized at a resolution of 227x227 pixels in RGB color mode. These images are derived from high-resolution original images that capture a wide range of crack scenarios on various surfaces within the Middle East Technical University campus. The dataset's diversity in surface treatment and lighting conditions is meticulously maintained by extracting standardized image sets from these high-resolution originals without any form of data augmentation such as random rotations or flips. This approach preserves the original characteristics and conditions of the images, highlighting the dataset's value for training and evaluating crack detection models. To broaden the dataset's diversity and enhance the model's generalizability, strategies like image rotation, flipping, scaling, and the combination of minimal field data acquisition are employed. This meticulous preparation aims to bolster the dataset's comprehensiveness, ensuring robust model training and evaluation.

3.1.2 Data Set Partitioning

To ensure the high accuracy of the concrete crack detection model, this study meticulously organized and divided the dataset using automated scripts. Initially, foundational structures were established within the project directory, and existing data directories were automatically cleared to eliminate interference. Subsequently, the script categorized and copied data to the training set directory based on file name identifiers. Approximately 20% of the images were randomly selected and moved to the validation set to ensure data diversity and model generalization capabilities. Moreover, to prevent data leakage, the validation set data was strictly used for performance evaluation. This systematic data management strategy significantly enhanced processing efficiency and provided a solid foundation for model training and accuracy assessment.

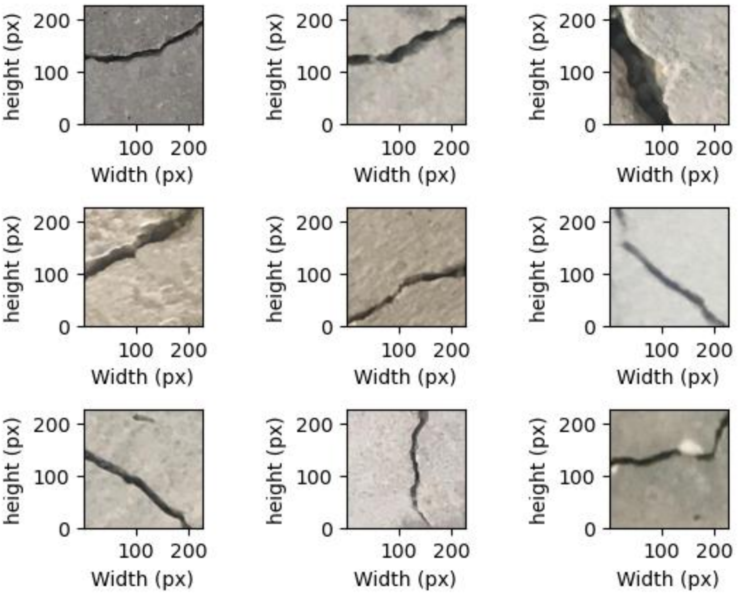

Figure 1: Schematic of Images with cracks in the Dataset.

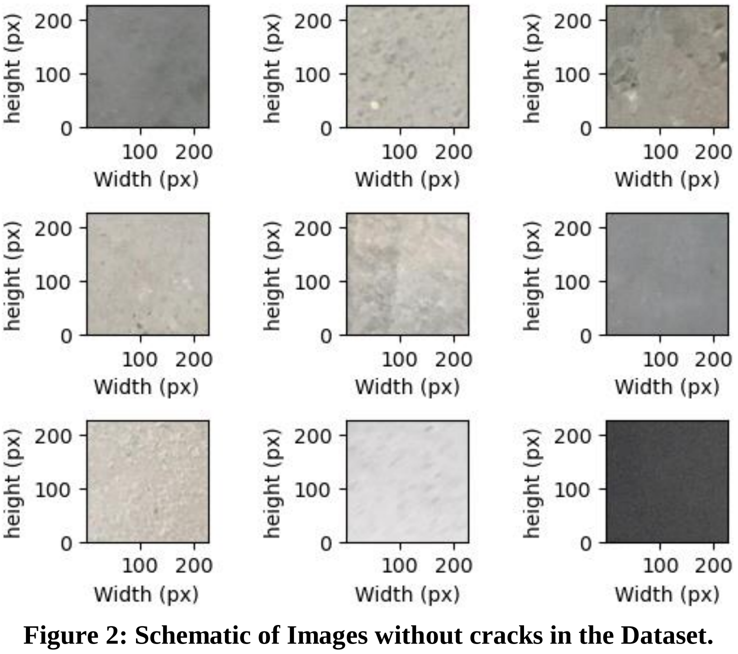

Figure 2: Schematic of Images without cracks in the Dataset.

3.1.3 Data Preprocessing

In this study, effective preprocessing of concrete crack images is pivotal for achieving high accuracy in crack detection. Preprocessing enhances the model's generalization capability and aids in better feature extraction from images. Our preprocessing strategy involves:

Training Set Preprocessing: For the training set, a series of data augmentation techniques simulate various shooting conditions and crack appearances, thereby increasing data diversity. The operations include Random Resized Crop, Random Rotation, Random Horizontal Flip, Random Vertical Flip, and Color Jitter, ensuring uniform input data format for the model.

Validation Set Preprocessing: The validation set undergoes simpler preprocessing methods to ensure fairness and consistency in evaluation. This involves Resize and Center Crop to standardize image sizes inputted into the model.

Image Standardization: Both training and validation images, after undergoing the transformations, are converted into tensors and standardized. Using preset mean values [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225] derived from ImageNet dataset statistics, images are normalized to expedite the training process and enhance model convergence speed.

These preprocessing steps ensure the model receives consistently high-quality, uniformly formatted image data throughout training and validation, laying a solid foundation for effective deep learning model training and accurate evaluation.

3.2 The Architecture of MobileNetV2 Network

3.2.1 Overview of Network Architecture

MobileNetV2 is an advanced neural network architecture specifically designed for mobile and edge devices, which operate in environments where computational resource efficiency is crucial. This architecture introduces an innovative structure called the Inverted Residual Block, which not only integrates linear bottleneck layers but also employs a streamlined version of depth wise separable convolutions, optimizing the trade-off between latency and accuracy. These technological innovations make MobileNetV2 not only lightweight but also high-performing, particularly suitable for applications requiring real-time processing and sensitivity to power consumption.

Key features of MobileNetV2 include the use of ReLU6 activation functions, which are particularly suitable for fixed-point implementations and optimize computational efficiency; and the removal of non-linearities in the final layers, which helps maintain higher representational accuracy at lower computational costs. These designs make MobileNetV2 extremely efficient in terms of memory usage and processing speed, significantly enhancing its practicality for real-time applications.

Moreover, the lightweight characteristics of MobileNetV2 make it an ideal choice for mobile vision applications, such as real-time object recognition and image segmentation. Especially on devices like drones, this architecture allows for rapid image processing and decision feedback without sacrificing accuracy, which is critical for performing dynamic tasks and working in resource-limited environments.

With these innovations, MobileNetV2 not only meets the demands of edge computing devices for efficiency and performance but also provides a reliable way to maintain high precision and fast responsiveness in complex environments. Therefore, it is one of the few architectures that achieves a good balance between high efficiency and precision, making it especially suitable for mobile applications and devices with stringent requirements on speed and power consumption.

3.2.2 Optimization Adjustments for MobileNetV2

To adapt the MobileNetV2 model for the task of concrete crack detection, several modifications were implemented, primarily focusing on the final layer's configuration, hyperparameter adjustments, and the incorporation of regularization techniques.

(i) Final Layer Modification:

Originally designed for a 1000-class classification task, the original output layer of MobileNetV2 features 1000 output nodes. To tailor the model for binary classification pertinent to crack detection, this layer was substituted with a new fully connected layer comprising two output nodes, representing "crack present" and "crack absent" respectively.

(ii) Hyperparameter Adjustments:

Learning Rate: Initially set at 0.001, a decay strategy was applied to the learning rate, reducing it by an order of magnitude every three epochs. This approach aids in fine-tuning the model as it approaches a best solution.

Batch Size: A smaller batch size of 8 was utilized during training to enhance the model's generalization capability on training data. Conversely, a larger batch size of 256 was employed during validation to expedite the evaluation process.

Optimizer: The Adam optimizer was selected for its capability to adjust learning rates adaptively, thereby optimizing the training process and accelerating convergence.

(iii) Application of Regularization Techniques:

Dropout: A dropout layer was incorporated into the fully connected layer of the model, with a dropout rate of 0.4. This method randomly nullifies the output of a subset of neurons, reducing overfitting and enhancing the model’s performance on unseen data.

These modifications were driven by the goal of enhancing the practicality and efficiency of the model, ensuring not only robust performance on training datasets but also high accuracy and reliability in practical applications. Through these optimizations, the MobileNetV2 model has been effectively adapted for crack detection tasks, showing its advantages in lightweight and efficient operation, particularly in resource-constrained environments.

3.3 Training Strategy

The training strategy employed in this study includes the implementation of a staircase learning rate decay policy, starting from a first value of 0.001 and reducing by 80% every five epochs, to finely regulate the model training process. Additionally, dropout and L2 regularization were applied after the convolutional layers in the model to control model complexity and suppress overfitting. Furthermore, to enhance the accuracy of crack detection, most of the convolutional layers were frozen to preserve extensive image recognition capabilities, while focusing training on the final classification layer.

IV. EXPERIMENTATION AND ANALYSIS

4.1 Training Environment Configuration

The key hardware components utilized in this study include: an NVIDIA GeForce RTX 3060 Ti GPU, which provides robust image processing capabilities; an Intel Core i9-12900K CPU, facilitating high-speed data processing; a storage solution combining a 2 TB SSD with a 1 TB HDD, along with 48 GB of DDR4 RAM, ensuring efficient data handling. The cooling system employed is the Thermaltake X360 liquid cooler, with a Gigabyte Z690 UD DDR4 motherboard and a 750W GAMEMAX RGB 750 PRO power supply.

Regarding the software environment, the system runs on Windows 11, with programming conducted in Python 3.11. The use of CUDA 12.1 and CuDNN 8.9.6 enhances the computational performance of deep learning tasks. These configurations provide a solid foundation for training and testing deep learning models, supporting complex computational demands.

4.2 Calculation processes

Table 1 delineates the computational architecture of the MobileNetV2 network. This table primarily displays the model’s layers along with their respective parameter values, and it includes configurations of hyperparameters such as activation functions. A variant of the ReLU activation function, ReLU6, is used within the model. Batch normalization (BN) layers are strategically placed after the convolutional layers and before the activation functions to perfect the training process. To clarify the computational workflow of Table 1, Figure 3 presents a flowchart of the model's computational architecture, with each segment of the flowchart denoting the layer name and corresponding layer number.

Table 1： Calculation processes.

Seq. No.	**Layer (type)**	**Output Shape**	**Param (pcs)**
1	Conv2d	[-1, 32, 112, 112]	864
2	BatchNorm2d	[-1, 32, 112, 112]	64
3	ReLU6	[-1, 32, 112, 112]	0
4	Conv2d	[-1, 32, 112, 112]	288
5	BatchNorm2d	[-1, 32, 112, 112]	64
6	ReLU6	[-1, 32, 112, 112]	0
7	Conv2d	[-1, 16, 112, 112]	512
8	BatchNorm2d	[-1, 16, 112, 112]	32
9	Inverted Residual	[-1, 16, 112, 112]	0
10	Conv2d	[-1, 96, 112, 112]	1536
11	BatchNorm2d	[-1, 96, 112, 112]	192
12	ReLU6	[-1, 96, 112, 112]	0
13	Conv2d	[-1, 96, 56, 56]	864
14	BatchNorm2d	[-1, 96, 56, 56]	192
15	ReLU6	[-1, 96, 56, 56]	0
16	Conv2d	[-1, 24, 56, 56]	2304
17	BatchNorm2d	[-1, 24, 56, 56]	48
18	Inverted Residual	[-1, 24, 56, 56]	0
...	...	...	...
154	Conv2d	[-1, 320, 7, 7]	307,200
155	BatchNorm2d	[-1, 320, 7, 7]	640
156	Inverted Residual	[-1, 320, 7, 7]	0
157	Conv2d	[-1, 1280, 7, 7]	409,600
158	BatchNorm2d	[-1, 1280, 7, 7]	2,560
159	ReLU6	[-1, 1280, 7, 7]	0
160	Dropout	[-1, 1280]	0
161	Linear	[-1, 128]	163,968
162	ReLU	[-1, 128]	0
163	Dropout	[-1, 128]	0
164	Linear	[-1, 2]	258

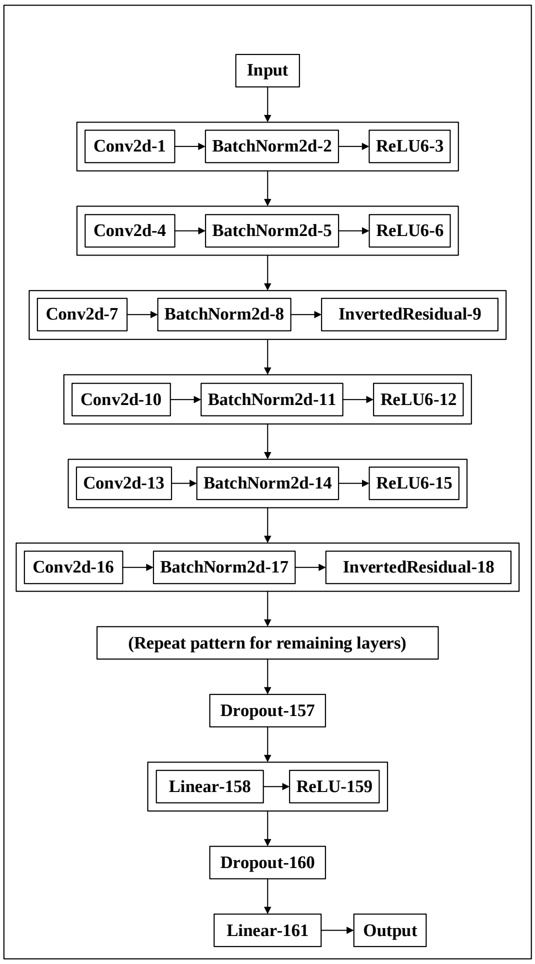

Figure 3 Flowchart of the computational architecture of the model.

4.3 Summary of Results

In this study, the MobileNetV2 model employed demonstrated exceptional performance in the task of concrete crack detection. Figure 4 illustrates the outcomes of the classification, showing that the predicted values are nearly identical to the actual values, indicating highly accurate classification results. As illustrated in Table 1, The total training time of the model for 10 cycles was only 1058 seconds, which further confirmed the applicability of the model for real-time applications. Moreover, the model not only achieved a high accuracy of 99.23% on the verification set, but also maintained a stable index of 98.92% in terms of average accuracy, recall rate, F1 score and overall accuracy, highlighting the efficiency and reliability of the model in crack identification.

These results collectively highlight the profound capabilities of the specifically enhanced MobileNetV2 model in handling complex image recognition tasks, particularly in engineering applications where rapid and accurate structural health monitoring is essential. Future work may explore additional optimization techniques and algorithmic enhancements to further improve the model's performance and broaden its application scope.

Table 2 Summary of Results.

Name	Value
Total training time (s)	1058
Best val Acc (%)	99.23
Average Precision (%)	98.92
Average Recall (%)	98.92
Average F1 Score (%)	98.92
Average Accuracy (%)	98.92
Total params (pcs)	2,388,098
Trainable params (pcs)	164,226
Non-trainable params (pcs)	2,223,872

Input size (MB)	0.57
Forward/backward pass size (MB)	152.86
Params size (MB)	9.11
Estimated Total Size (MB)	162.55

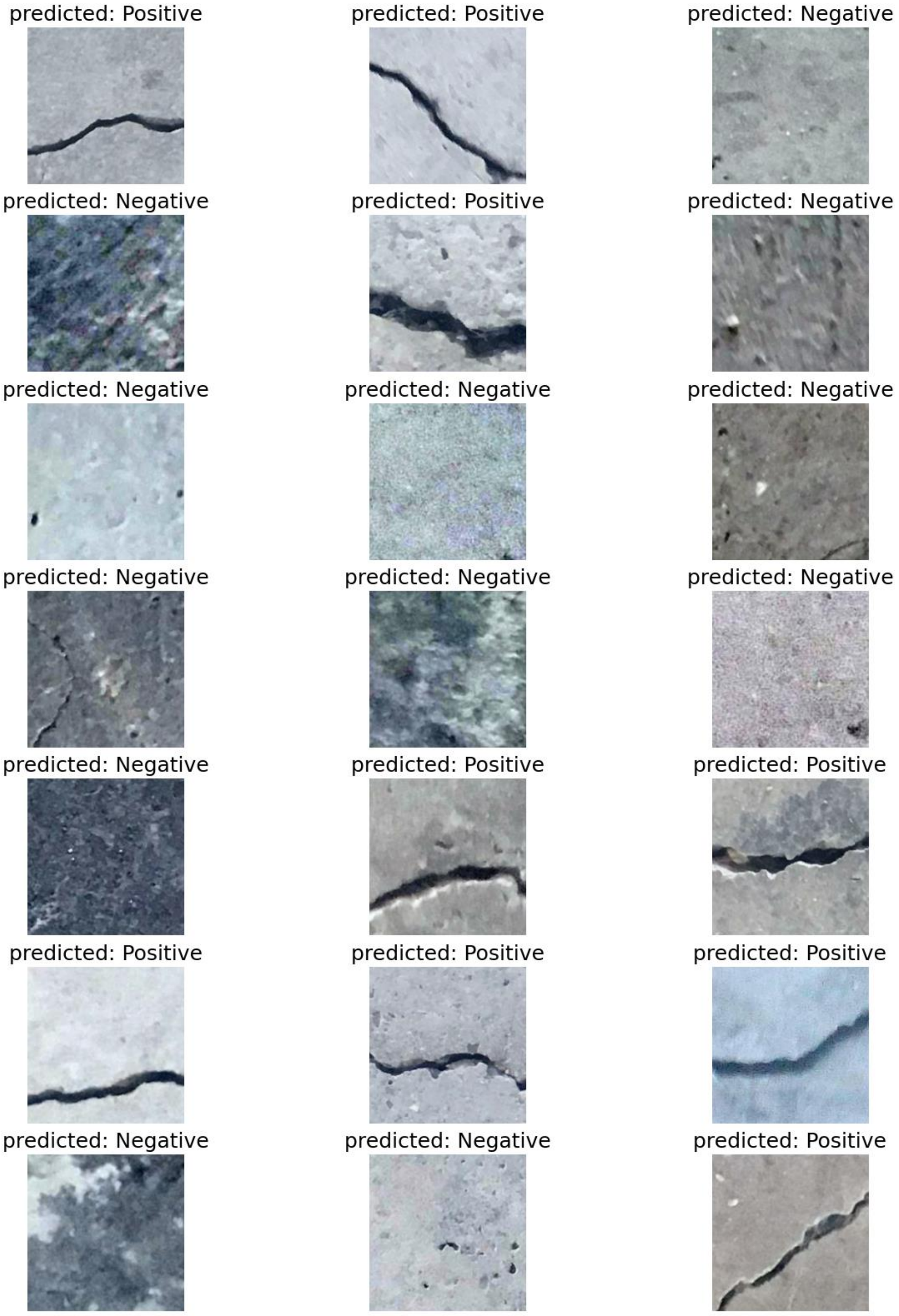

Figure 4: Results of model classification.

4.4 Analysis of Experimental Results

According to the data presented in Table 3, the MobileNetV2 model demonstrates a notable computational time advantage among various deep learning algorithms. In comparison to other commonly used convolutional neural networks such as GoogleNet, the ResNet series, and the VGG series, MobileNetV2 requires only 106 seconds per epoch, significantly less than the 1227 to 3789 seconds required by the other models. This result underscores the efficiency of MobileNetV2 in processing time, making it particularly suitable for deployment on resource-constrained mobile devices.

As shown in Table 4, the model underwent 10 training epochs, during which both training and validation loss and accuracy were monitored. Training loss decreased from 19.78% to 16.97%, while training accuracy improved from 92.06% to 93.21%. Concurrently, accuracy on the validation set steadily increased, reaching a peak of 99.23%. These metrics indicate significant enhancements in the model's generalization performance as training progressed, demonstrating its robust learning capabilities and adaptability.

The experimental outcomes of this study highlight the MobileNetV2 model's efficient computational performance and superior image processing capabilities, particularly applicable to real-time and on-site crack detection applications. The high generalizability and accuracy of the model ensure reliability across various settings, positioning it as a potent tool in infrastructure maintenance. Future research should explore the real-time deployment of the model and its potential applications in a broader range of scenarios.

Through these experiments and analyses, this study not only confirms the practicality of the MobileNetV2 model in concrete crack detection but also provides a methodological reference for other fields with similar needs. Additionally, considering the diversity in image quality and crack types, there is a need for future optimization of the algorithm and expansion of the dataset to address more complex detection environments.

Table 3: Comparisons of computational time for per round.

CNN algorithm	Time(s)
MobileNetV2	106
GoogleNet	1,227
ResNet50	1,666
ResNet101	2,447
ResNet152	3,789
VGG16	2,827
VGG19	2,943

Table 4: Training Process Loss and Accuracy Statistics.

Epoch	Train Loss (%)	Train Acc (%)	Val Loss (%)	Val Acc (%)
1	19.78	92.06	4.48	98.68
2	19.47	92.28	4.25	98.59
3	19.16	92.51	4.33	98.71
4	18.19	92.69	3.51	98.87
5	17.74	92.96	3.54	98.86
6	17.98	92.77	3.28	99.00
7	17.69	92.79	2.98	99.08
8	17.45	92.81	3.20	99.13
9	17.66	93.05	3.37	99.03
10	16.97	93.21	2.73	99.23

V. CONCLUSION

The research presented in this paper has shown the effectiveness of using MobileNetV2, a lightweight convolutional neural network, for the task of detecting concrete cracks in images. Through systematic experimentation and analysis, this study has shown that the MobileNetV2 model provides a high level of accuracy and efficiency, making it an ideal solution for real-time and on-site applications in infrastructure monitoring. The integration of deep learning techniques with crack detection significantly improves the process over traditional methods, by enhancing both the speed and reliability of detections.

Future work will focus on further optimizing the MobileNetV2 architecture and exploring its deployment across various mobile platforms. Additionally, efforts will be made to enhance the diversity and volume of the dataset to improve the model’s robustness and accuracy under different conditions. The potential for integrating this technology with drone and other smart device inspections will also be explored to broaden the scope of its applications and to provide more comprehensive solutions for infrastructure management.

This study not only contributes to the academic and practical understanding of concrete crack detection but also provides a valuable method for other domains where similar deep learning approaches can be utilized to enhance operational efficiency and accuracy.

REFERENCES

[1]. Ohno K, Ohtsu M. Crack classification in concrete based on acoustic emission[J]. Construction and Building Materials, 2010, 24(12): 2339–2346.

[2]. Aggelis G. Classification of cracking mode in concrete by acoustic emission parameters[J]. Mechanics Research Communications, 2011, 38(3): 153–157.

[3]. Ziou D, Tabbone S. Edge detection techniques-an overview[J]. International Journal of Pattern Recognition and Image Analysis, 1998, 8(4): 537–559.

[4]. Ciresan D, Meier U, Masci J, et al. Flexible, high performance convolutional neural networks for image classification[C]. in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona: AAAI, 2011: 1237–1242.

[5]. Cha Y-J, Choi W, Büyüköztürk O. Deep learning- based crack damage detection using convolutional neural networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361–378.

[6]. Ngo L, Xuan C L, Luong H M, et al. Designing image processing tools for testing concrete bridges by a drone based on deep learning[J]. Journal of Information and Telecommunication, 2023, 7(2): 227-240.

[7]. Özgenel Ç F. Concrete crack images for classification[J]. Mendeley Data, 2019, 2: 2019.

[8]. Özgenel Ç F, Sorguç A G. Performance comparison of pretrained convolutional neural networks on crack detection in buildings[C]. //Isarc. proceedings of the international symposium on automation and robotics in construction. IAARC Publications, 2018, 35: 1-8.

[9]. Zhang L, Yang F, Zhang Y D, et al. Road crack detection using deep convolutional neural network[C]. //2016 IEEE international conference on image processing (ICIP). IEEE, 2016: 3708-3712.