ISSN: 2710-0588

Enhancing Concrete Crack Image Detection Using MobileNetV2 and Transfer Learning

Fujie Yang

University of Shanghai for Science and Technology, Shanghai, 200093, China

Abstract

The safety of civil engineering structures such as bridges and buildings is paramount, necessitating the early detection and repair of cracks to prevent structural failures. Traditional crack detection methods, including manual inspections and basic image processing, are often hampered by slow detection speeds and limited accuracy. This study explores the integration of advanced deep learning techniques, specifically MobileNetV2, enhanced through transfer learning, to improve the efficiency and accuracy of crack detection. The "Concrete Crack Images for Classification" dataset was employed, featuring 40,000 images from real-world scenarios. Our approach leverages the lightweight architecture of MobileNetV2 and the efficiency of transfer learning to create a robust model that not only reduces computational demands but also achieves high accuracy, with a validation accuracy rate of 99.38%. This paper details the model training, the innovative use of MobileNetV2, and comparative analyses with other deep learning models, demonstrating significant improvements over traditional methods. The findings underscore the potential of MobileNetV2 and transfer learning in practical applications, offering a promising avenue for future research in automated crack detection across various engineering fields.

Keywords

MobileNetV2; Concrete Crack Detection; Convolutional Neural Networks; Deep Learning; Transfer Learning.

1. Introduction

Crack detection is a crucial technology for ensuring the safety of engineering structures. Whether it involves roads, bridges, buildings, or pipeline systems, the early detection and repair of cracks are essential to prevent structural failure, ensure safety for individuals, and reduce maintenance costs. With the advancement of automation and digitization, the use of image processing techniques for crack detection has become a focal point of research, extending its applications across civil engineering, aerospace, and mechanical manufacturing sectors[1].

Despite significant advancements in crack detection technologies, traditional methods such as manual inspection and basic image processing often face limitations in detection speed and accuracy. Manual inspections are time-consuming and labor-intensive, while conventional automated methods struggle to cope with complex environmental changes and diverse types of cracks. The emergence of deep learning technologies has provided a new solution for crack detection. Particularly, the application of transfer learning has been able to significantly enhance training efficiency and generalization ability of models, while lightweight deep learning models such as MobileNetV2 have been widely adopted in mobile devices and edge computing due to their efficient processing capabilities. By combining these two approaches, not only can the challenges of data scarcity and computational resource limitations be overcome, but also an efficient and accurate method for crack detection can be provided[2].

ISSN: 2710-0588

2. Related Work

In recent years, the rapid development of deep learning technologies, particularly the application of Convolutional Neural Networks (CNNs) in image processing, has become a focal point of research. In the field of concrete crack detection, the use of CNNs for automated detection has proven to be both effective and accurate. Optimized lightweight networks, such as MobileNetV2, are particularly suited for deployment in mobile devices or edge computing environments due to their computational efficiency and reduced resource consumption.

MobileNetV2[3] is a lightweight deep convolutional neural network that achieves higher efficiency through an inverted residual structure, making it highly effective in image classification and detection tasks. Specifically for crack detection, it achieves high accuracy and speed with reduced computational resources, which is crucial for real-time detection tasks.

Recent studies, such as those by Wu et al. [4], have utilized an enhanced MobileNetV2 combined with the DeepLabV3 algorithm for bridge crack detection, demonstrating significant improvements in processing speed while maintaining high detection accuracy. This method leverages the efficiency of MobileNetV2 and the precise segmentation capabilities of DeepLabV3 to effectively enhance crack detection performance.

Furthermore, research by Gan et al. [5] has shown the potential of combining Faster R-CNN with Building Information Modeling (BIM) to improve the precision and recall rates in crack detection. Although their study did not directly use MobileNetV2, the potential application of this network in crack detection is evident.

In practical applications, lightweight network models like MobileNetV2 not only reduce computational and storage requirements but also accelerate the detection process without sacrificing much accuracy, making these technologies more suitable for use in resource-limited environments, such as drone-mounted or smartphone applications.

In summary, the use of optimized lightweight CNN models like MobileNetV2 in concrete crack detection not only enables efficient real-time detection but also demonstrates outstanding performance and potential for application in various practical scenarios. These studies provide important references and foundations for future technological advancements and applications in this field.

3. Data Preparation and Research Design

3.1. Dataset and Preprocessing

This study utilizes the "Concrete Crack Images for Classification" dataset[6~8], provided by Çağlar Fırat Özgenel and made available on Mendeley Data since 2019. The dataset comprises 40,000 images, evenly distributed between cracked (positive) and uncracked (negative) categories, with 20,000 images in each. Images are standardized to a resolution of 227x227 pixels in RGB format, directly extracted from diverse crack scenarios across the Middle East Technical University campus, ensuring authenticity and complexity of the scenes. To preserve the originality of the images, conventional data augmentation techniques such as rotation and flipping were not employed. Instead, data diversity was enhanced through minimal on-site data collection and selective image processing, thereby strengthening the model's generalization capabilities.

To ensure the integrity and accuracy of the detection model, a stringent data segmentation strategy was implemented. Initially, a hierarchical file system was established to optimize data management; subsequently, data cleansing was conducted to eliminate potential conflicts. Images were automatically assigned to training and validation directories using a script based on file name patterns, with a random 20% subset allocated to the validation set to assess the

ISSN: 2710-0588

model's performance across varied samples, ensuring diversity and representativeness of the validation set.

Effective image preprocessing is critical for enhancing the accuracy of crack detection. Preprocessing aims to augment the model's ability to extract features from images and to foster generalization:

Training Set Preprocessing: A suite of data augmentation techniques such as Random Cropping, Rotation, Horizontal and Vertical Flipping, and Color Adjustment were applied to simulate different crack scenarios, increase data diversity, and ensure uniformity of input data format. The processed training set is depicted below.

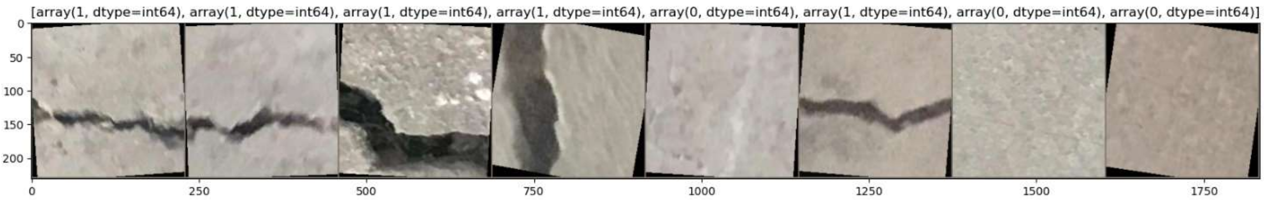

Figure 1. The processed training set.

Validation Set Preprocessing: Simplified preprocessing was implemented, involving Image Resizing and Center Cropping to ensure consistency and fairness in evaluation.

Both training and validation images were converted into tensors and normalized using the mean values (0.485, 0.456, 0.406) and standard deviations (0.229, 0.224, 0.225) derived from the ImageNet dataset, to optimize training speed and model convergence.

These meticulously designed preprocessing steps ensure high quality and standardized image data during training and validation, laying a solid foundation for the effective training of the deep learning model and precise evaluation of its accuracy.

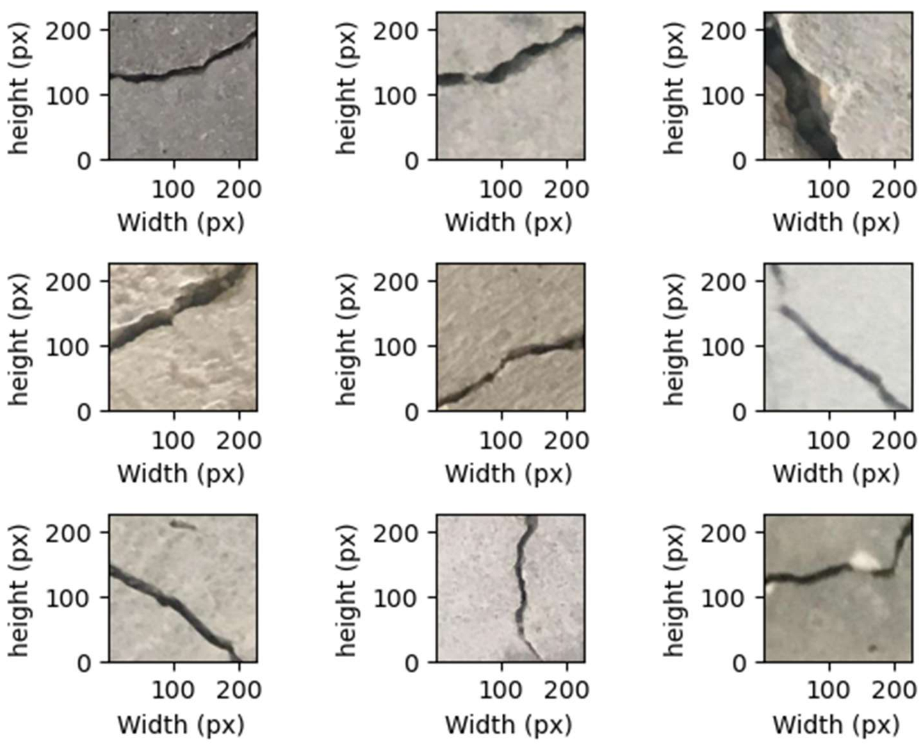

Figure 2. Schematic of Images with cracks in the Dataset.

ISSN: 2710-0588

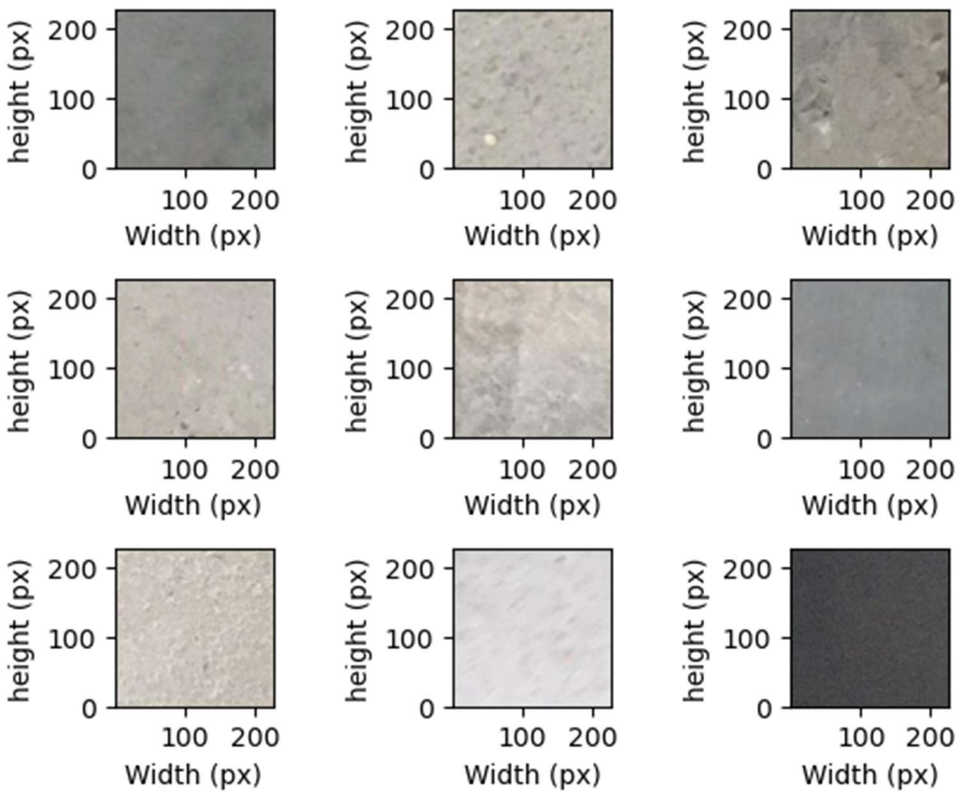

Figure 3. Schematic of Images with cracks in the Dataset.

3.2. Transfer Learning Strategy

To effectively utilize MobileNetV2 for crack detection, this study employs a transfer learning approach. Transfer learning allows us to expedite the learning process for new tasks by leveraging models pre-trained on large-scale datasets, thereby saving time and improving efficiency. The following details the transfer learning strategy used:

In this research, the pre-trained MobileNetV2 model serves as the foundational architecture for addressing the task of crack detection. MobileNetV2 is an advanced neural network architecture designed specifically for mobile and edge computing devices, known for its efficient use of computational resources. The model optimizes depth-wise separable convolutions by introducing Inverted Residual Blocks and linear bottleneck layers, effectively balancing latency and accuracy, making it particularly suitable for real-time processing and energy-sensitive applications.

(1) Layer Freezing and Fine-Tuning

Freezing Layers: The initial layers of MobileNetV2 are frozen as they typically capture basic image features such as edges and textures, which are common across various vision tasks. Freezing these layers utilizes already learned features, reducing training complexity.

Fine-Tuning Layers: Subsequent layers are unfrozen and fine-tuned to allow the learning of more task-specific advanced features. This step is crucial for adapting the model to recognize specific characteristics of cracks.

(2) Output Layer Reconstruction

Modifying the Output Layer: The original output layer of MobileNetV2, designed for 1,000 class classification, is replaced with a new fully connected layer with just two output nodes, representing "crack present" and "crack absent." This modification tailors the model for binary classification tasks.

ISSN: 2710-0588

3.3. Training Process

The training process is critical to ensure the performance and accuracy of the model. The following are the detailed methods and strategies employed in training the crack detection model:

(1) Learning Rate Strategy

Dynamic Adjustment: The training begins with a low learning rate (0.001) and incorporates a dynamic decay strategy that reduces the learning rate by 10% after every fixed period. This helps in fine-tuning the model and avoiding local optima.

(2) Batch Size Settings

Training Phase: A smaller batch size (8) is used during training to enhance the model's robustness when facing varied data; this helps improve generalization.

Validation Phase: A larger batch size (256) is employed during validation to increase processing efficiency and stability of results.

(3) Optimizer Selection

Adam Optimizer: The Adam optimizer is chosen for its ability to combine momentum with adaptive learning rate adjustments, facilitating optimized training and faster model convergence.

(4) Regularization Techniques

Dropout: A dropout rate of 0.4 is implemented in the fully connected layer to mitigate the risk of overfitting and to enhance the model’s generalization capability on unseen data.

These strategies not only enhance the model's training efficiency and generalization capabilities but also ensure high performance and accuracy across various operating environments. The training and adjustment measures make the MobileNetV2 model particularly suitable for crack detection on resource-constrained mobile devices.

4. Experimental Design and Results

To validate the effectiveness and performance of the model, we established a set of specific experimental parameters and evaluation criteria. Here are the key components of the experimental design:

Training Epochs: The model underwent fifty training epochs to ensure sufficient iterations for stable convergence.

Data Split: The "Concrete Crack Image Classification" dataset was divided into 80% for training and 20% for validation. Additionally, an independent test set was reserved for final performance evaluation.

Evaluation Metrics: The primary metrics used to measure the model's performance were accuracy, recall, and F1 score. Accuracy assesses the overall correctness of the model in identifying cracks. Recall evaluates the model's ability to identify all actual crack cases, while the F1 score is the harmonic mean of accuracy and recall, used to balance both metrics.

In this study, the pre-trained MobileNetV2 model, enhanced through transfer learning, was utilized for detecting concrete cracks, demonstrating exceptional performance. As illustrated in Figure 4, the classification results closely matched actual observations, indicating high accuracy. Table 1 reveals that the total training time for this model across 50 epochs was only 2497 seconds, underscoring its efficiency for real-time applications. Additionally, the model achieved a high accuracy rate of 99.38% on the validation set, proving its efficacy and reliability in crack detection tasks.

ISSN: 2710-0588

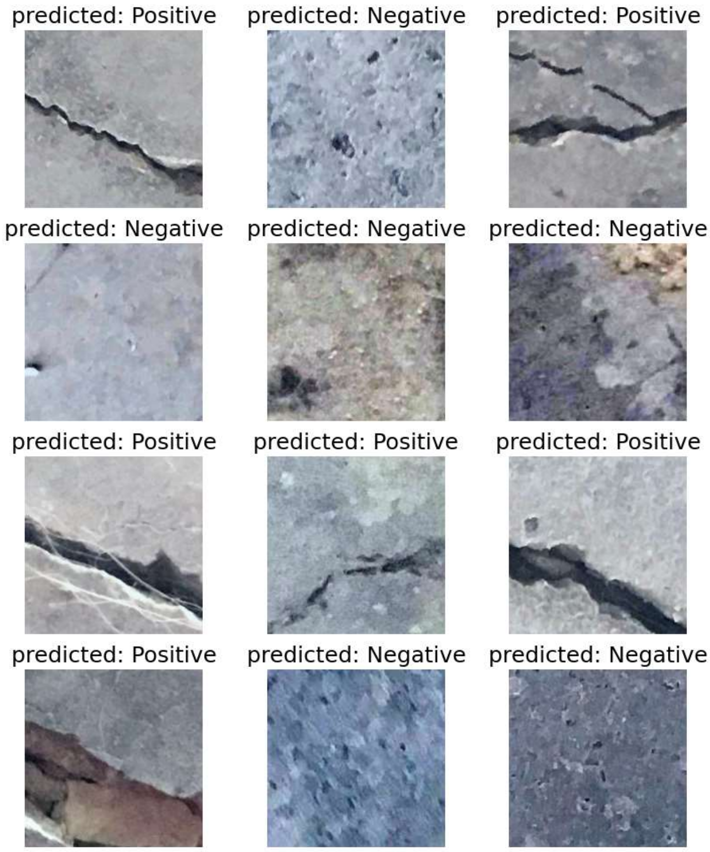

Figure 4. Results of model classification.

Table 1. Summary of Results.

Name	Value
Total training time (s)	2497
Best val Acc (%)	99.38
Average Accuracy (%)	99.04

5. Analysis of Experimental Results

Based on the data from the training and validation period shown in Table 2, combined with Figure 5 and Figure 6, we can perform the following analysis to evaluate the performance and trends of the model:

(1) Training and Validation Loss

The training loss (Train Loss) gradually decreased from 0.2835 to 0.1665, indicating continuous performance improvement of the model over time. Similarly, the validation loss (Val Loss) decreased from 0.0816 to 0.0229, demonstrating enhanced generalization ability of the model on unseen data over time.

Although the overall trend is downward, significant fluctuations in validation loss were observed in certain epochs, such as the 5th and 19th epochs. This could be an early sign of overfitting or a result of adjustments in the learning rate and the randomness of data batches during training.

(2) Training and Validation Accuracy

The training accuracy (Train Acc) improved from 88.96% to 93.37%, while the validation accuracy (Val Acc) increased from 97.94% to 99.35%. This indicates that the model's ability to correctly identify and classify answers is progressively strengthening.

ISSN: 2710-0588

Table 2. Training Process Loss and Accuracy Statistics.

Epoch	Train Loss(%)	Train Acc(%)	Val Loss(%)	Val Acc(%)
1	28.35	88.96	8.16	97.94
2	21.13	91.71	5.28	98.42
3	20.49	91.83	4.80	98.56
4	19.00	92.28	4.78	98.56
5	18.83	92.45	4.19	98.73
6	18.25	92.64	5.78	98.26
7	18.23	92.87	3.73	98.76
8	18.94	92.52	5.08	98.63
9	18.15	92.85	4.21	98.91
10	17.84	92.98	3.38	98.89
11	17.73	92.96	4.31	98.64
12	18.45	92.62	3.04	99.02
13	17.97	92.77	3.35	99.12
14	18.74	92.38	3.93	98.88
15	17.58	92.86	3.11	99.09
16	17.28	93.25	2.91	99.04
17	17.79	92.93	2.70	99.14
18	18.19	92.74	3.35	99.09
19	17.41	93.03	3.25	99.00
20	17.93	92.79	5.17	98.50
21	17.13	93.11	3.29	99.14
22	17.31	93.13	2.76	99.31
23	17.42	93.08	2.50	99.17
24	17.38	93.10	3.46	99.12
25	17.33	93.16	2.94	99.12
26	16.79	93.26	2.71	99.22
27	17.29	93.03	2.69	99.25
28	16.53	93.41	2.73	99.18
29	17.90	92.90	2.99	99.18
30	17.47	93.07	3.05	99.18
31	17.33	93.17	3.06	99.23
32	17.08	93.16	4.34	98.98
33	17.32	93.09	2.41	99.24
34	17.04	93.36	2.38	99.26
35	16.95	93.17	3.14	99.12
36	17.31	93.31	2.71	99.27
37	17.04	93.29	2.51	99.17
38	17.38	93.15	2.47	99.31
39	17.32	93.01	3.03	99.14
40	16.93	93.12	2.86	99.18
41	17.19	93.33	2.98	99.23
42	17.12	93.19	2.78	99.21
43	16.83	93.34	3.07	99.18
44	17.01	93.32	2.72	99.36
45	16.33	93.69	2.85	99.22

ISSN: 2710-0588

Table 2. (Continued Table)

Epoch	Train Loss(%)	Train Acc(%)	Val Loss(%)	Val Acc(%)
46	16.79	93.50	2.53	99.34
47	17.09	93.13	2.54	99.27
48	16.62	93.43	2.50	99.29
49	16.89	93.25	2.49	99.38
50	16.65	93.37	2.29	99.35

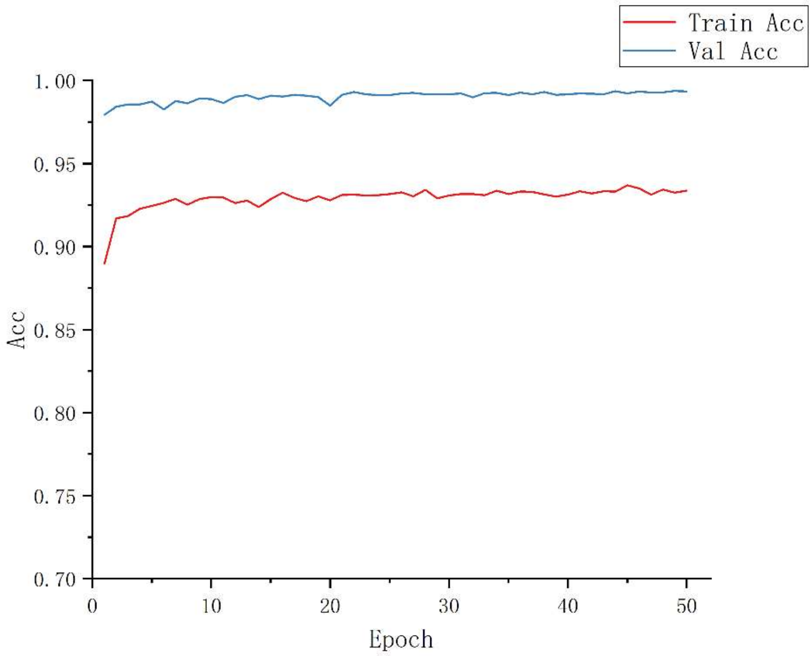

Figure 5. Training Process Accuracy.

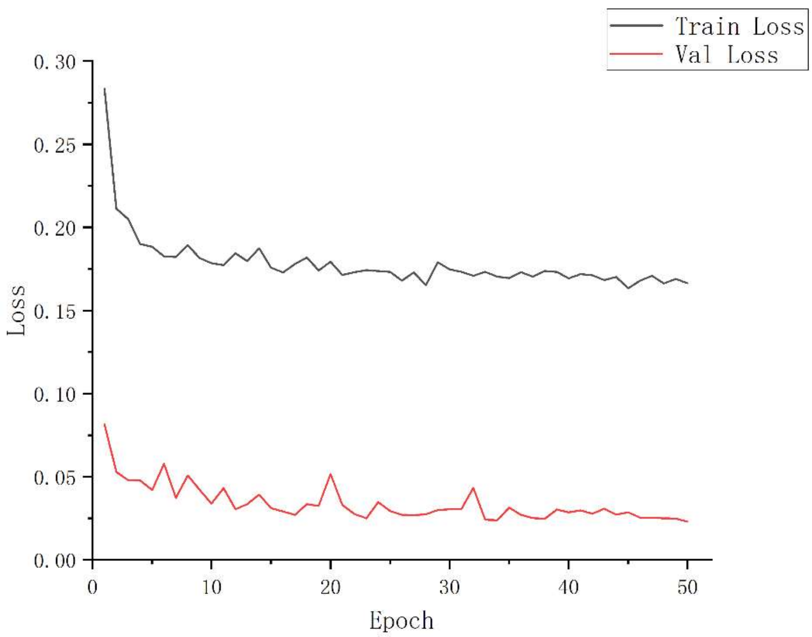

Figure 6. Training Process Loss.

To demonstrate the advantages of the method proposed in this study over other existing technologies, we compared it against several traditional crack detection methods:

ISSN: 2710-0588

Traditional Image Processing Methods: Compared to traditional methods that utilize thresholding and edge detection, our model not only improves accuracy but also better adapts to changes in lighting and background conditions.

Other Deep Learning Models: According to the data presented in Table 3, our method, which utilizes MobileNetV2 for transfer learning, shows higher efficiency and similar or improved performance in resource-limited environments compared to unoptimized general deep learning models. The MobileNetV2 model exhibits a significant computational time advantage over various deep learning algorithms. For instance, it only requires 50 seconds per epoch, which is lower than the 1227 to 3789 seconds required by other commonly used convolutional neural networks like GoogleNet, the ResNet series, and the VGG series. This marked reduction in processing time is attributed to the use of transfer learning, which enhances the efficiency of MobileNetV2, making it especially suitable for deployment on resource-constrained mobile devices.

Table 3. Comparisons of computational time for per round.

CNN algorithm	Time(s)
MobileNetV2	50
GoogleNet	1,227
ResNet50	1,666
ResNet101	2,447
ResNet152	3,789
VGG16	2,827
VGG19	2,943

6. Conclusion

This study has demonstrated the efficacy of MobileNetV2, a lightweight deep convolutional neural network, enhanced through transfer learning for the task of concrete crack detection. The choice of MobileNetV2 was guided by its architectural advantages, particularly its efficient use of computational resources facilitated by inverted residual structures and depth-wise separable convolutions. These features make it an ideal candidate for real-time and energy-sensitive applications such as mobile and edge computing devices.

The use of a transfer learning strategy was critical in leveraging the pre-trained capabilities of MobileNetV2, significantly reducing the model's training time while maintaining high levels of accuracy and efficiency. This approach allowed the model to effectively learn task-specific features from the "Concrete Crack Images for Classification" dataset, while bypassing the extensive computational cost typically associated with training deep neural networks from scratch.

Experimental results underscored the performance of the MobileNetV2 model, which achieved a validation accuracy of 99.38%, confirming its potential for practical deployment. The dynamic adjustment of learning rates and the strategic combination of freezing and fine-tuning layers were pivotal in optimizing the training process, further enhancing the model's ability to generalize well on unseen data.

Comparative analysis with traditional image processing methods and other deep learning models illustrated the advantages of using MobileNetV2. Not only did it outperform traditional methods by adapting more robustly to varying conditions, but it also exhibited superior computational efficiency compared to more complex models such as GoogleNet, ResNet series, and VGG series.

ISSN: 2710-0588

In conclusion, the successful application of transfer learning in this context highlights its significance as a method for enhancing the performance of deep learning models in specialized tasks. The findings from this study provide a solid foundation for future research and practical applications in crack detection and other related fields, where efficiency and accuracy are paramount.

References

[1] Ali S B, Wate R, Kujur S, et al. Wall crack detection using transfer learning-based CNN models[C]//2020 IEEE 17th India Council International Conference (INDICON). IEEE, 2020: 1-7.

[2] Su C, Wang W. Concrete cracks detection using convolutional neuralnetwork based on transfer learning[J]. Mathematical Problems in Engineering, 2020, 2020: 1-10.

[3] Sandler M, Howard A, Zhu M, et al.: MobileNetV2: Inverted Residuals and Linear Bottlenecks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018), p.4510-4520.

[4] Wu Z, Tang Y, Hong B, et al. Enhanced precision in dam crack width measurement: Leveraging advanced lightweight network identification for pixel-level accuracy[J]. International Journal of Intelligent Systems, 2023.

[5] Gan L, Liu H, Yan Y, et al.: Bridge Bottom Crack Detection and Modeling Based on Faster R-CNN and BIM, IET Image Processing, Vol. 18 (2024) No.3, p.664-677.

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

[7] Özgenel Ç F, Sorguç A G.: Performance Comparison of Pretrained Convolutional Neural Networks on Crack Detection in Buildings, ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (2018), Vol. 35, p.1-8.

[8] Zhang L, Yang F, Zhang Y D, et al.: Road Crack Detection Using Deep Convolutional Neural Network, 2016 IEEE International Conference on Image Processing (ICIP) (2016), p.3708-3712.