Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Deep Learning Approaches for Autonomous Crack Detection in Concrete Wall, Brick Deck and Pavement

Fethi ŞERMET[1*], İshak PAÇAL[2]

[1]Igdir University, Civil Engineering Department, fethi.sermet@igdir.edu.tr, Orcid No: 0000-0001-8221-689X
[2]Igdir University, Computer Engineering Department, ishak.pacal@igdir.edu.tr, Orcid No: 0000-0001-6670-2169

ARTICLE INFO

Article history:

Received 10 March 2024
Received in revised form 20 April 2024
Accepted 29 April 2024
Available online 30 June 2024

Keywords:

Crack detection, structural cracks, deep learning, image processing, CNN

Doi: 10.24012/dumf.1450640

* Sorumlu Yazar

ABSTRACT

Detecting cracks is vital for inspecting and maintaining concrete structures, enabling early intervention and preventing potential damage. The advent of computer vision and image processing in civil engineering has ushered in deep learning-based semi-automatic/automatic techniques, replacing traditional visual inspections. These methods, driven by autonomous diagnosis, have applications across various sectors, fostering rapid progress in civil engineering. In this study, we present an approach that combines vision transformers and convolutional neural networks (CNN) for autonomously diagnosing cracks in bridges, roads, and walls. Performance enhancement was achieved through transfer learning, data augmentation, and optimized hyperparameters, utilizing popular CNN and vision transformers (ViT) architectures. The proposed method was tested on the SDNET2018 dataset, comprising over 56,000 images. Experimental results demonstrated the approach's effectiveness, achieving high accuracy in detecting road cracks at 96.41%, wall cracks at 92.76%, and bridge cracks at 92.81%. These findings highlight the promising potential of deep learning in this field.

Introduction

Cracks are a common issue observed on various man-made structures such as pavements, bridges, nuclear power plant walls, and tunnel ceilings. Cracking occurs when a structural element separates into distinct pieces, representing a mechanism to relieve stress when concrete is subjected to forces beyond its tensile capacity [1]. It's a symptom of deterioration processes that weaken concrete or subject it to excessive stresses, causing it to lose its integrity [2]. When cracking happens, the tensile stresses perpendicular to the crack are eliminated [3]. Due to concrete's heterogeneous material structure and brittle behavior, it's widely accepted that cracks will eventually appear during the structure's lifespan. Building codes explicitly acknowledge this, ensuring that structures can endure loads over their intended service life despite crack formation. Concrete cracks can lead to severe consequences, such as reduced strength and stiffness, diminished aesthetics, shorter durability, and compromised waterproofing [4]. The loss of stiffness due to cracks results in additional deformations and displacements in structural elements.

Timely detection and repair of cracks are crucial for maintaining infrastructure health and preventing further damage. Developing a fast, reliable, and cost-effective algorithm to identify surface defects is a top priority for robust infrastructure management systems [5, 6, 7]. The rising demand for computer-aided intelligent infrastructure monitoring/inspection systems, such as pavement surface inspection [8], underground pipeline inspection [9], bridge crack monitoring [10], and railway track assessment [11], has led to a growing interest in automatic crack detection in recent years. Deep learning algorithms, a subset of artificial intelligence, have been prominently featured in this trend and have demonstrated successful applications across various fields.

In recent years, deep learning has seen widespread applications in various fields such as industry, healthcare, natural language processing, and autonomous vehicles, especially within engineering disciplines [12, 13, 14, 15]. In this context, numerous studies have explored the use of deep learning techniques in crack detection. Loverdos and Sarhosis [16], curated a comprehensive dataset consisting of images of stacked brick walls with diverse colors, textures, and sizes. They discovered that employing image-based techniques and machine learning for brick segmentation produced superior results com

pared to conventional image processing methods. In a study conducted by Ali R. et al. [17] focusing on crack detection in structures, applications of CNN outperformed traditional image processing techniques and other machine learning methods in crack classification and segmentation. Xu Z. et al. [18] introduced a locally developed transformer network (LETNet) method capable of effectively detecting road surface cracks, achieving high accuracy in crack detection. They highlighted LETNet's exceptional performance in identifying various surface cracks under diverse road and weather conditions. Additionally, Chaiyasarn, K. et al. [19] reported remarkable results in crack predictions using CNN, achieving high accuracy (99.88%), precision (82.2%), recall (90.2%), and F1 score (86.01%).

Yu, Y. et al. [20], an approach utilizing advanced pre-trained CNN was proposed to predict cracks on concrete structures, leveraging a dataset comprising 41,780 images. While the method proved to be effective and robust in detecting concrete cracks, the authors acknowledged its limitation in characterizing the size and type of small cracks. On a parallel note, Ma, D. et al. [21] introduced a technique for asphalt road crack detection based on a CNN with multiple feature layers. The model extracted multi-scale features to enhance accuracy in road crack recognition. Following hyperparameter tuning, the model achieved an impressive accuracy of 98.217% and a crack detection rate of 96.6 frames per second (FPS). Müller, A. et al. [22], a machine learning-based approach was developed to automatically detect the onset of fractures using a dataset of over 30,000 images derived from material characterization experiments. The researchers mentioned that their classifier model could identify both the initiation and propagation of cracks. While their study concentrated on analyzing ductile fracture images, they noted that the methodology they proposed could readily be adapted for brittle fracture problems as well. Moreover, Fang et al. [23] introduced a fatigue crack growth prediction method that employed machine learning model correction to mitigate errors arising from uncertain factors in crack growth. Hamidia, M. [24], an extensive database comprising 264 surface crack models corresponding to 61 non-ductile reinforced concrete moment frame (RCMF) was created. These specimens were tested under various shear displacement rates. The researcher proposed a machine learning-based procedure to automate the characterization of the damage state of non-ductile reinforced concrete moment frames. This characterization was based on visual indices of crack patterns observed on concrete surfaces. Interestingly, the study revealed that predictions relying on models utilizing compressive strength information did not significantly enhance accuracy. This suggests that the surface crack models of RCMF provide adequate information to estimate the maximum shear displacement ratio experienced by moment frames during seismic vibrations. In essence, the research indicated that the crack pattern model could be effectively utilized to predict the maximum shear displacement ratio sustained by damaged non-ductile RCMF during seismic events.

Aravind N. et al. [25] concentrated on crack detection by employing image processing and fault pattern recognition techniques in conjunction with appropriate machine learning algorithms. They specifically targeted cracks in reinforced concrete beams subjected to bending loads and utilized six different classifiers for detection. The study highlighted that the support vector machine (SVM) classifier provided the most accurate results in predicting cracks. Han X. et al. [26] introduced a hybrid technique incorporating CNN and digital image processing to detect cracks in photographs. They proposed that by implementing transfer learning, the required volume of data and costs could be minimized without compromising accuracy. Meanwhile, Laxman, K. C. et al. [27] devised a comprehensive framework utilizing deep learning models for crack detection on concrete surfaces. Additionally, they focused on estimating crack depth using images captured from portable devices. Zhang et al. [28] introduced a crack detection model using Binary Level Sets (BLS), which effectively identifies crack images while being lightweight. Similarly, Martinez-Ríos et. al. [29] proposed utilizing Generalized Morse Wavelets (GMWs) in Continuous Wavelet Transform (CWT) to detect transverse cracks on sidewalks, employing spectrograms to fine-tune pre-trained CNN like GoogLeNet, SqueezeNet, and ResNet18. Among these, SqueezeNet demonstrated the highest average validation sensitivity. Meanwhile, Xu et al. [30] developed a real-time method for crack detection, segmentation, and parameter measurement, ensuring both accuracy and efficiency. Additionally, Yuan et al. [31] focused on machine learning models to predict self-healing concrete's ultimate crack width, successfully correlating input parameters like raw materials and pre-healing crack width with post-healing crack width, demonstrating accurate predictions of crack healing capacity. Iraniparast et al. [32] employed Deep Convolutional Neural Networks (DCNN) and transfer learning techniques to detect cracks in images of concrete structures. They incorporated multi-resolution image analysis through wavelet transformation for crack segmentation. Their DCNN classifier models demonstrated strong performance, with F1-scores ranging from 94.5% to 99.6%. Similarly, Katsigiannis et al. [33] introduced a deep learning method for crack detection in brick wall facades, using transfer learning with limited annotated data. They created a dataset of 700 brick wall facade images, using 500 for training, 100 for validation, and 100 for testing. Their approach proved highly effective, achieving accuracy and F1 scores of up to 100% during end-to-end training of the neural network.

When evaluating the studies in the literature, it is evident that deep learning methods are widely utilized for crack detection across various types of structures. These studies encompass a wide spectrum of buildings, ranging from brick walls to concrete structures, employing techniques such as CNN, transfer learning, and other machine learning methods. Particularly, experiments conducted on diverse datasets consistently demonstrate that deep learning-based approaches yield superior results compared to traditional image processing techniques. These studies underscore the effectiveness of deep learning in

crack detection, emphasizing its significant potential in the fields of civil engineering and structural maintenance in the future.

This study primarily focuses on crack detection using deep learning techniques. With the advancement of Artificial Intelligence (AI), deep learning has achieved significant success and is regarded as the most promising approach for crack detection. The main contributions of this study emphasize how crucial it is to detect cracks during the maintenance of concrete structures. Early crack detection is critical for ensuring the longevity and safety of these structures, allowing proactive measures to prevent potential damage. Additionally, this study demonstrates that the use of deep learning-based techniques instead of traditional visual inspection methods represents a significant transformation in the field of civil engineering. These techniques can be applied in various fields due to their autonomous diagnostic capability and will contribute to rapid advancements in civil engineering as well. Furthermore, this article presents the effectiveness of a crack detection approach developed using deep learning methods such as vision transformers and CNN. This method achieves high accuracy rates in autonomously diagnosing cracks in different types of structures, including bridges, roads, and walls. The dataset and experimental studies demonstrate the scalability of this approach and showcase the transformative impact of deep learning in civil engineering practice. Therefore, this study represents a significant advancement in the field of crack detection.

Material and Methods

Deep Learning

Deep learning is a type of machine learning that involves neural networks with at least three layers. These networks attempt to mimic the functioning of the human brain and have the ability to learn from large amounts of data. Additional hidden layers can enhance the accuracy of prediction. Deep learning technology empowers various artificial intelligence applications, such as digital assistants, fraud detection systems, and autonomous vehicles. Machine learning and deep learning models come in different types of learning, including supervised, unsupervised, and reinforcement learning [34,35]. Supervised learning works with labeled data, unsupervised learning detects patterns in unlabeled data and classifies them based on distinctive features, while reinforcement learning is used to improve a model's actions based on feedback.

Training deep learning models may take longer compared to other machine learning methods, but testing trials can be quicker. However, in cases where data is limited or better interpretability is required, traditional machine learning methods might be preferred and advantageous [36]. Deep learning is more effective in situations involving large datasets, complex problems, or feature extraction requirements. It finds applications in various fields, including civil engineering, and has become an effective tool for solving complex engineering problems. These applications include predicting structural responses, structural reliability and health monitoring, structural damage detection, estimating material properties, and providing decision support for intelligent transportation systems.

Convolutional Neural Networks (CNNs)

CNNs are a deep learning architecture widely used, especially in computer vision tasks. They were first proposed in the "Neocognitron" paper, based on Hubel and Wiesel's model of the visual system, and later optimized using backpropagation by Yann Lecun and colleagues [37]. CNN consist of three main types of layers: convolutional layers, which extract features using a sliding kernel; non-linear layers, applying activation functions to model non-linear relationships; and pooling layers, altering small regions of feature maps with statistical information. CNN layers have locally connected nodes and weight-sharing mechanisms through sliding kernels, significantly reducing parameters. Popular CNN architectures include VGGNet, ResNet, GoogLeNet, MobileNet, and DenseNet [38, 39].

Vision Transformers (ViT)

The Transformer is a deep learning model that uses self-attention mechanisms to evaluate the importance of each part of the input data. This model has gained significant attention, particularly in the field of natural language processing, replacing traditional models. ViT was introduced in 2021 for image recognition tasks. It treats images as sequences of patches, which are then transformed into vectors and processed by a standard Transformer encoder. ViT is pretrained on large datasets and can be fine-tuned for specific image classification tasks. This model has potential applications in various fields such as image recognition, object detection, and visual reasoning. However, the effectiveness of image processing models depends on factors such as the chosen optimization method, network depth, and specific hyperparameters of the dataset. Particularly, it has been observed that ViT has a more challenging optimization process compared to CNN. The Transformer generates a sequence of output tokens using the self-attention mechanism and maps these outputs to feature maps [40]. This approach allows the model to focus on crucial pixel-level information, reducing the number of tokens that need to be analyzed and significantly cutting down costs.

Dataset

The importance of the dataset in achieving the desired success of deep learning architectures is unquestionable. While classical machine learning approaches focus on manual feature extraction and small datasets, the key difference between the two approaches is that deep learning architectures require large-scale datasets and automatic feature extraction. In this study, the dataset created by Dorafshan et al. [41] was used. The SDNET2018 dataset, as seen in Figure 1, contains over 56,000 annotated images of cracked and non-cracked concrete surfaces, bridge decks, walls, and pavements.

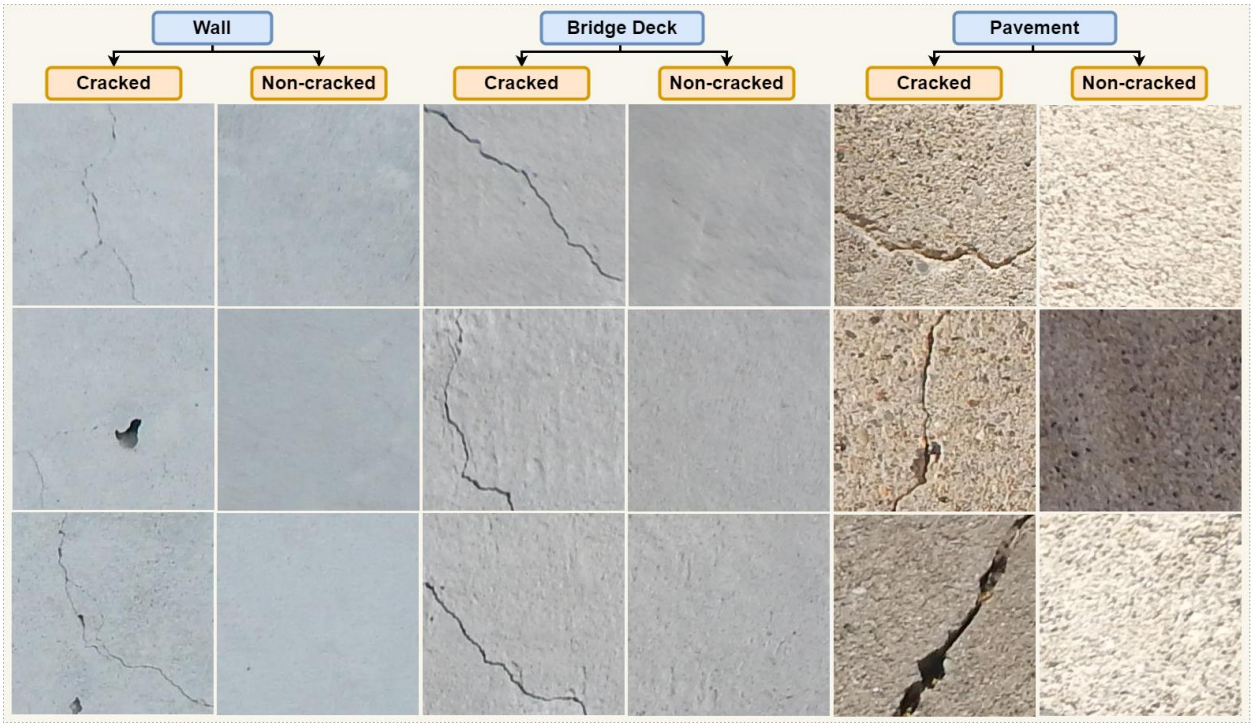

Figure 1. Some sample images from SDNET2018 dataset

Table 1. SDNET 2018 image dataset description and statistics.

	Number of images in the training set		Number of images in validation set		Number of images in the test set		
	Cracked	Non-cracked	Cracked	Non-cracked	Cracked	Non-cracked	Total
Wall	2695	10001	578	2143	578	2143	18138
Bridge Deck	1417	8115	304	1740	304	1740	13620
Pavement	1824	15208	392	3259	392	3259	24334

This dataset is designed for training, validation, and comparison of autonomous crack detection algorithms working with image processing, deep CNN architectures, and other techniques. As the popularity of such techniques in structural health monitoring increases, there is a need for a dataset containing various annotated images that were not previously available to continually improve crack detection algorithms.

The SDNET2018 dataset [41] has been enriched with images containing different types of cracks ranging from 0.06 mm in width to 25 mm. For instance, there are 3851 images of cracked walls, with 14287 images of non-cracked walls. Similarly, for bridge decks, there are 2025 cracked images and 11595 non-cracked images. As for pavement surfaces, there are 2608 cracked images and 21726 non-cracked images. The dataset has allocated 15% of cracked and non-cracked images for validation, 15% for testing, and 70% for training purposes. Statistical information about the SDNET2018 dataset is presented in Table 1.

Proposed Method and Process

In this section, the usage and validation of deep learning models in crack detection are explained. Deep learning models have shown effective results in various fields, along with their successes in crack detection in the literature. In this study, popular CNN-based algorithms such as ResNet [42], VGG [43], and MobileNet architectures [44] are used for crack detection. Additionally, recently emerging image transformer models in deep learning, such as MobileViT [45] and Multi-Scale Vision Transformers (MViT2) [46], are employed for crack detection. The effective training of deep learning models significantly impacts their generalization capabilities on test data. If effective training does not occur, deep learning models can yield unsuccessful results on test data, which can be interpreted as a result of overfitting or being trained in an uncontrolled manner. The proposed approach's architecture is illustrated in Figure 2.

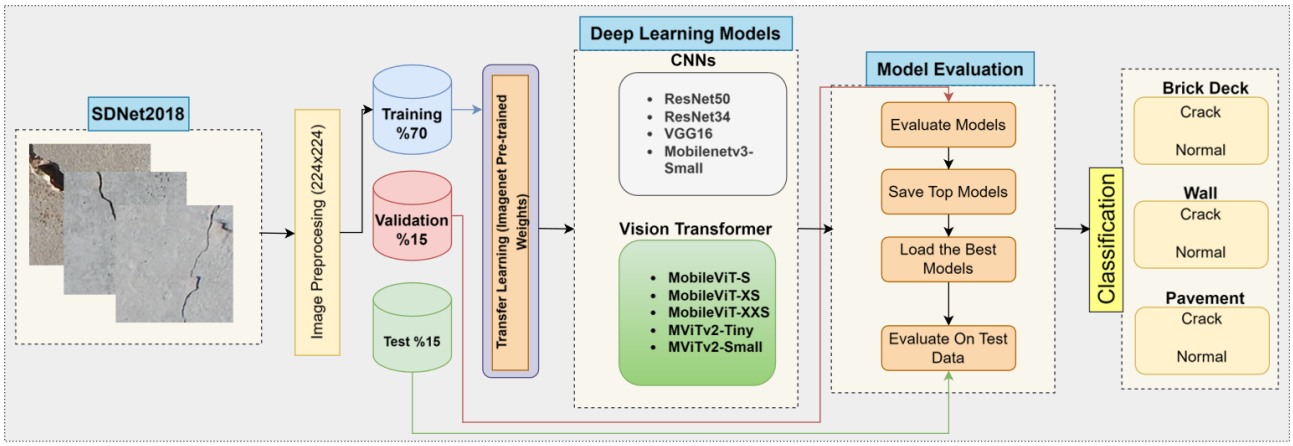

Figure 2. Stages of the proposed crack detection approach

Throughout the training process, deep learning models' accuracy and speed can be significantly enhanced through various techniques and parameters. Among the most effective methods are transfer learning and data augmentation. Furthermore, factors such as input image size, batch size, number of epochs, optimizer, learning rate, weight decay, decay rate, and warm-up augmentation play crucial roles in influencing the model's performance. In this study, essential data augmentation techniques such as scaling, smoothing, flipping, color jittering, and rotation were consistently applied across all models. Transfer learning was facilitated by utilizing weights from the ImageNet dataset, proving to be a valuable method for accelerating convergence and improving accuracy. The models were trained with optimized values for different parameters. For instance, the image resolution was set to 224x224, and other key parameters were configured as follows: step size: 0.000001, base step size: 0.1, momentum: 0.9, optimizer: sgd, weight decay: 2.0e-05, warm-up periods: 5, and warm-up learning rate: 1.0e-05.

Results and Discussion

Experimental Design

Typically, deep learning algorithms are trained using GPU (Graphics Processing Unit)-based graphics cards. These cards allow for faster processing of large datasets due to their parallel computing capabilities. Particularly, graphics cards using NVIDIA's CUDA architecture are widely preferred for training deep learning algorithms. The hardware components used in this study include the following: Linux-based Ubuntu 22.04 operating system, NVIDIA RTX 2080TI (11 GB GDDR6 and 4352 CUDA cores) graphics card, Intel Core i9 9900X (10 cores, 3.50 GHz, 19.25 MB Intel® Smart Cache) processor, and 32 GB DDR4 RAM. PyTorch was used as the deep learning library, and Python was the preferred programming language.

Evaluation Metrics

Evaluation metrics serve as tools to assess models from various perspectives. In the realm of object classification, fundamental metrics such as accuracy, precision, recall, and F1 score are commonly employed. These metrics are universally used in evaluating deep learning models in the existing body of literature. Their calculation hinges on knowing the values of true positive, true negative, false positive, and false negative, which typically constitute a confusion matrix. True positive signifies parts genuinely belonging to the positive class and accurately predicted. Conversely, true negative represents instances where parts rightfully belong to the negative class and are precisely predicted. False positive arises when parts truly belong to the positive class but are incorrectly predicted, while false negative denotes situations where the model inaccurately predicts the class.

$$Accuracy = \frac{True\ positive + True\ negative}{Total\ number\ of\ predictions}$$

$$Precision = \frac{True\ positive}{True\ positive + False\ pozitif}$$

$$Recall = \frac{True\ positive}{True\ pozitive + False\ negative}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

These formulas provide the mathematical foundation for evaluating the performance of object classification algorithms or architectures.

Experimental Results

In this section, the statistical results and evaluations of the achievements of each deep learning model used in this study on the SDNET2018 dataset are discussed. The SDNET2018 dataset comprises three main classes: bridge

decks, walls, and pavements. Each class contains both cracked and uncracked images. Therefore, the evaluations are presented in three tables, providing a more detailed comparison.

Experimental Results of Bridge Deck Classification

This section presents the experimental results for bridge decks (Table 2). It demonstrates the performance of the most popular deep learning architectures, namely CNN architectures, and the increasingly popular image transformer models in recent years. This study compares the results for both architectures, providing a comparison between CNN and image transformers. It evaluates the models' ability to generalize and perform on the test data in crack detection.

Table 2. Experimental results of the bridge deck

Model Name	Accuracy	Precision	Recall	F1-score
ResNet50	0.9227	0.8924	**0.7809**	0.8329
ResNet34	0.9178	0.9036	0.7508	0.8201
VGG16	0.9183	0.8949	0.7593	0.8215
MobileViT-S	0.9144	0.8595	0.7760	0.8156
MobileViT-XS	0.9217	0.8974	0.7721	0.8300
MobileViT-XXS	0.9203	0.8956	0.7672	0.8264
MobileNetv3-Small	0.9080	0.8776	0.7274	0.7955
MViTv2_Tiny	**0.9281**	**0.9328**	0.7732	**0.8455**
MViTv2_Small	0.9256	0.9160	0.7744	0.8393

Table 2 presents the experimental results for nine deep learning models related to bridge decks. Upon examining Table 2, it can be stated that all models exhibit high performance with an accuracy of over 90%. Only a few models surpass the 92% accuracy threshold, indicating their superior performance compared to other models. A detailed analysis of Table 2 showcases the performance of different deep learning models in the classification task. In light of the observed metrics, the MViTv2_Tiny model has the highest accuracy (0.9281) and precision (0.9328) scores, indicating its overall high ability for correct predictions and accurate positive predictions. However, this model has a lower sensitivity (0.7732) score compared to some other models, meaning the rate of true positives being predicted is slightly lower. On the other hand, the MobileNetv3-Small model stands out with lower accuracy (0.9080) and precision (0.8776) scores but might require more improvement in terms of sensitivity (0.7274). The confusion matrices for these models are shown in Figure 3 and Figure 4.

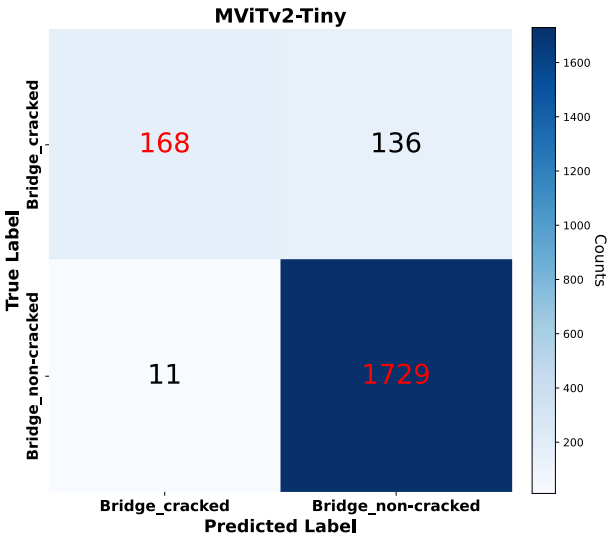

Figure 3. Confusion matrix of mvit2-tiny

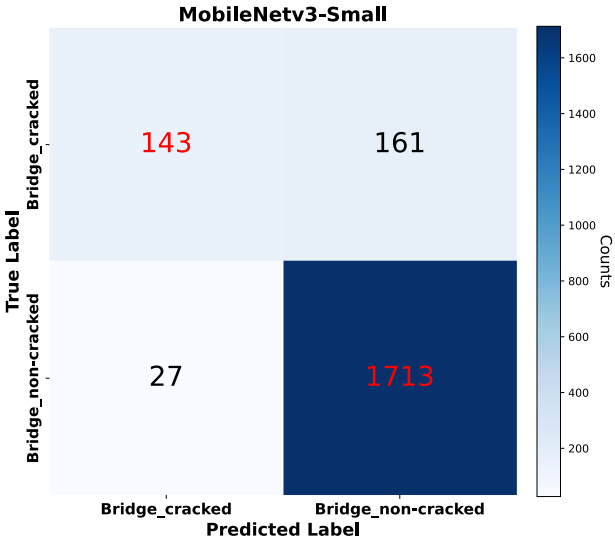

Figure 4. Confusion matrix for Mobilenetv3-small

Experimental Results of Wall Crack Classification

This section includes the experimental results related to wall crack detection in the wall class of the SDNNet2018 dataset (Table 3). Similar to the bridge decks, the performance of both CNN and vision transformer models in crack detection has been addressed. Additionally, a detailed comparison of the performance and metrics of both deep learning architectures and their respective models has been conducted to provide a more comprehensive analysis of crack detection.

Table 3. Experimental results of the wall element

Model Name	Accuracy	Precision	Recall	F1-score
ResNet50	0.8905	0.8955	0.7637	0.8244
ResNet34	0.9151	0.8847	0.8539	0.8690
VGG16	0.9199	0.8922	0.8613	0.8765
MobileViT-S	**0.9208**	**0.9113**	0.8665	0.8883
MobileViT-XS	0.9239	0.9031	0.862	0.8821
MobileViT-XXS	0.9236	0.8935	0.8731	0.8832
MobileNetv3-Small	0.8956	0.8559	0.82	0.8376
MViTv2_Tiny	**0.9276**	0.9035	0.8744	0.8887
MViTv2_Small	0.9236	0.8922	0.875	0.8835

Table 3 presents the experimental results for 9 deep learning models related to the wall class. Upon examining Table 3, it can be stated that all models except ResNet50 and MobileNetv3-Small achieved high accuracy, surpassing 90%. Only a few models demonstrated accuracy above 92%, indicating their superior performance compared to others. When considering all metrics and making a comprehensive evaluation, the highest performance was observed in the MViTv2 architecture, particularly in the models MViTv2-Tiny and MViTv2-Small, with MViTv2-Tiny achieving the highest overall accuracy. On the other hand, the lowest performance was attributed to the ResNet50 model. The confusion matrices for these models are shown in Figure 5 and Figure 6.

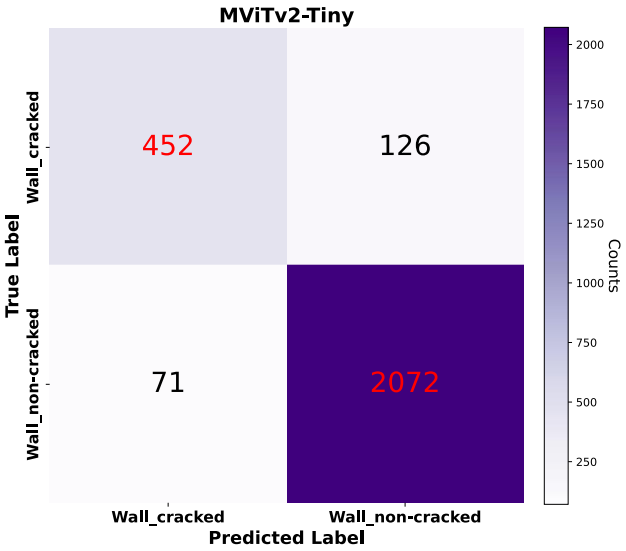

Figure 5. Confusion matrix for MViTv2-Tiny

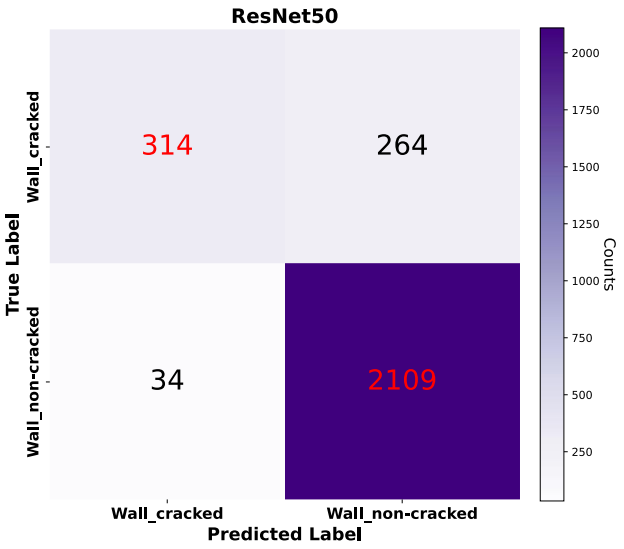

Figure 6. Confusion matrix for ResNet50

A detailed analysis of the table reveals that the MViTv2-Tiny model stands out with high accuracy (0.9276) and precision (0.9035), indicating its excellent classification ability. However, other models also exhibited strong performance. The VGG16 and MobileViT-S models are notable for their high precision and recall values, reflecting their success in minimizing false positive and false negative results. Figure 5. Confusion matrices of MViTv2-Tiny

Experimental Results of Pavement Classification

This section covers the experimental results related to the road surface (Table 4). It demonstrates the performance of deep learning architectures, including popular CNN architectures and models based on image transformation, which have become popular in recent years. This study compares and evaluates the generalization capabilities and performances of both deep learning architectures on the test data, focusing on crack detection.

Table 4. Experimental results of pavement

Model Name	Accuracy	Precision	Recall	F1-score
ResNet50	0.9584	0.9272	0.8443	0.8838
ResNet34	0.9551	0.8989	0.857	0.8775
VGG16	0.9554	0.9235	0.8302	0.8744
MobileViT-S	0.9606	0.9416	0.8433	0.8897
MobileViT-XS	0.9619	0.9221	0.871	0.8958
MobileViT-XXS	0.9622	0.9360	0.8576	0.8951
MobileNetv3-Small	0.9543	0.9074	0.8409	0.8729
MViTv2-Tiny	**0.9641**	0.9392	0.8654	0.9008
MViTv2-Small	0.9636	0.9459	0.8562	0.8988

Table 4 presents the experimental results for the road surface with 9 deep learning models. When examining Table 4, it can be stated that all models achieve accuracy

above 95%, indicating high performance. Upon considering all metrics and making a general evaluation, the highest performance is achieved by the MViTv2 architecture, particularly the MViTv2-Tiny and MViTv2-Small models. Among these, the MViTv2-Tiny model stands out with the highest performance. On detailed inspection of Table 4, the MViTv2-Tiny model demonstrates the highest accuracy rate (96.41%), emphasizing its overall high performance. It also exhibits successful performance in terms of precision (93.92%) and F1 score (90.08%), highlighting its ability to make accurate positive predictions. On the other hand, ResNet50 and ResNet34 models have high accuracy rates (95.84% and 95.51%, respectively), but their precision (92.72% and 89.89%) and F1 scores (88.38% and 87.75%) are slightly lower compared to MViTv2-Tiny. These models seem to yield robust results in a large dataset. The MobileViT series models (MobileViT-S, MobileViT-XS, MobileViT-XXS) also have high accuracy rates, but their precision and F1 scores are more uneven when compared to some other models. Especially noteworthy are the precision of MobileViT-S (94.16%) and MobileViT-XXS (93.60%). The confusion matrices for these models are shown in Figure 7 and Figure 8.

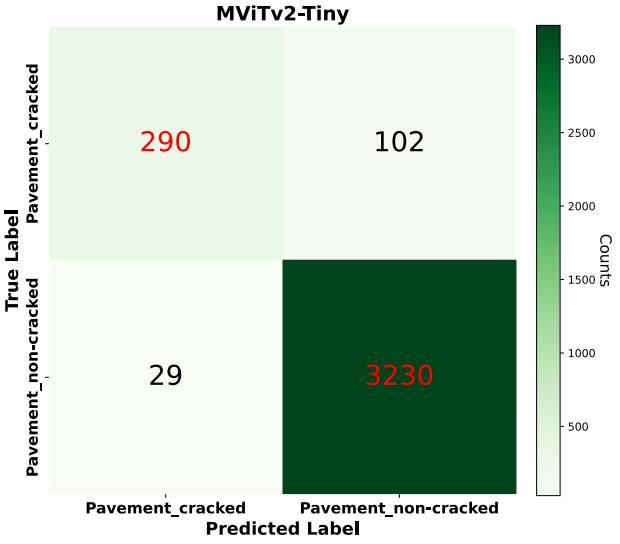

Figure 7. Confusion matrix for MViTv2-Tiny

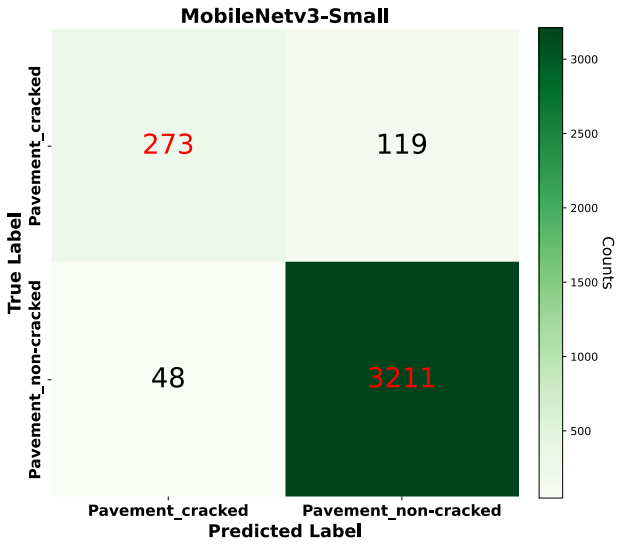

Figure 8. Confusion matrix for MobileNetv3-Small

Conclusions

Detecting cracks in structures is challenging due to lighting conditions, complex backgrounds, lighting effects, and low contrast between crack and non-crack areas. Deep learning has addressed these challenges by automatically extracting features from images and detecting cracks. This study introduces an innovative approach for autonomous crack diagnosis in various structures such as bridges, roads, and walls, using vision transformers and CNN. To enhance crack detection performance, the study combines deep CNN and ViT architectures by employing transfer learning, data augmentation, and hyperparameter optimization. Impressive results were obtained by applying this approach to the extensive SDNET2018 dataset, consisting of more than 56,000 images. In trial runs, this approach achieved remarkable accuracy rates, such as 96.41% in pavement crack detection, 92.76% in wall crack detection, and 92.81% in bridge crack detection. These findings demonstrate the promise of deep learning in crack detection and its transformative impact on civil engineering applications.

Consequently, deep learning techniques have the potential to be applied in structural analysis and detection tasks like identifying concrete cracks. However, their effectiveness relies on several factors, including the availability of accurate data, suitable algorithms, and appropriate application conditions.

Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared

There is no conflict of interest with any person / institution in the article prepared

Authors' Contributions

Şermet F: Study conception and design, interpretation of data, drafting of manuscript.

Pacal İ: Study conception and design, visualization, analysis, and interpretation of data, drafting of manuscript

References

[1] Kovler, K., & Chernov, V. (2009). Types of damage in concrete structures. In N. Delatte, *Failure, distress and repair of concrete structures* (pp. 32-56). Boca Raton: Woodhead Publishing Limited.

[2] Larosche, C. J. (2009). Types and causes of cracking in concrete structures. In N. Delatte, *Failure, distress and repair of concrete structures* (pp. 57-83). Boca Raton: Woodhead Publishing Limited.

[3] Ghali, A., Favre, R., & Elbadry, M. (2002). Concrete Structures- Stresses and Deformation. *Spon Press.*

[4] ACI Committee 201. (2001). Guide to Durable Concrete. In ACI Manual of Concrete Practrice Part 1 -Materials and General Properties of Concrete (pp. 20 1.2Rl-20 1.2R41). *Farmington Hills: American Concrete Institute.*

[5] Daghighi, A. (2020). Full-Scale Field Implementation of Internally Cured Concrete Pavement Data Analysis for Iowa Pavement Systems. *Creative Components*. 638. https:// lib. dr. iasta te. edu/ creat iveco mpone nts/ 638.

[6] Hosseini, S., & Smadi, O. (2020). How prediction accuracy can affect the decision-making process in pavement management system. *Infrastructures.* https:// doi. org/ 10. 31224/ osf. io/ t28ue.

[7] Abukhalil, Y. B. (2019). Cross asset resource allocation framework for pavement and bridges in Iowa. *Graduate Theses and Dissertations.* 16951. https:// lib. dr. iasta te. edu/ etd/ 16951.

[8] N. F. Hawks, and T. P. Teng. (2014). Distress identi_cation manual for the long-term pavement performance project. *National academy of sciences*.

[9] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart. (2016). Automatic crack detection on two-dimensional pavement images: an algorithm based on minimal path selection," *IEEE Trans. Intell. Transp. Syst., vol. 17*, no. 10, pp. 2718-2729.

[10] I. Abdel, O. Abudayyeh, and M. E. Kelly. (2003). Analysis of edge-detection techniques for crack identi_cation in bridges," *J. Computer Civil Eng., vol. 17*, no. 4, pp. 255-263.

[11] P. Lad, and M. Pawar. (2016). Evaluation of railway track crack detection system," in Proc., *IEEE ROMA*, pp. 1-6.

[12] E. Aslan ve Y. Özüpak, "Classification of Blood Cells with Convolutional Neural Network Model", *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, c. 13, sy. 1, ss. 314–326, 2024, doi: 10.17798/bitlisfen.1401294.

[13] Lubbad, M. A., Kurtulus, I. L., Karaboga, D., Kilic, K., Basturk, A., Akay, B., ... & Pacal, I. (2024). A Comparative Analysis of Deep Learning-Based Approaches for Classifying Dental Implants Decision Support System. *Journal of Imaging Informatics in Medicine*, 1-22.

[14] Kunduracioglu, I., & Pacal, I. (2024). Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. *Journal of Plant Diseases and Protection*, 1-20.

[15] Pacal, I. (2024). A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *International Journal of Machine Learning and Cybernetics*, 1-19.

[16] Loverdos, D., & Sarhosis, V. (2022). Automatic image-based brick segmentation and crack detection of masonry walls using machine learning. *Automation in Construction 140*, 104389.

[17] Ali, R., Chuah, J-H., Abu Talip, M-S., Mokhtar, N., Shoaib, M-A. (2022). Structural crack detection using deep convolutional neural networks. *Automation in Construction 133*, 103989.

[18] Xu, Z., Guan, H., Kang, J, Xiangda, L., Ma, L., Yu, Y., Chen, Y., Li, J. (2022). Pavement crack detection from CCD images with a locally enhanced transformer network. *International Journal of Applied Earth Observations and Geoinformation*, *110*,102825.

[19] Chaiyasarn, K., Buatik, A., Mohamad, H., Zhou, M., Kongsilp, S., Poovarodom, N. (2022). Integrated pixel-level CNN-FCN crack detection via photogrammetric 3D texture mapping of concrete structures. *Automation in Construction*, *140*, 104388 Available.

[20] Yu, Y., Samali, B., Rashidi, M., Mohammadi, M., Nguyen, T.N., Zhang, G. (2022). Vision-based concrete crack detection using a hybrid framework considering noise effect. *Journal of Building Engineering*, *61*, 105246.

[21] Duo Ma, Hongyuan Fang, Niannian Wang, Binghan Xue, Jiaxiu Dong & Fu Wang (2022). A real-time crack detection algorithm for pavement based on CNN with multiple feature layers. *Road Materials*

and Pavement Design, 23:9, 2115-2131, DOI: 10.1080/14680629.2021.1925578.

[22] Müller, A., Karathanasopoulos, N., Roth, C.C., Mohr, D. (2021). Machine Learning Classifiers for Surface Crack Detection in Fracture Experiments. *International Journal of Mechanical Sciences 209*, 106698.

[23] Fang, X., Liu, G., Wang, H., Xie, Y., Tian, X., Leng, D., Mu, W., Ma, P., Li, G. (2022). Fatigue crack growth prediction method based on machine learning model correction. Ocean Engineering 266, 112996.

[24] Hamidia, M., Mansourdehghan, S., Asjodi, A-H., Dolatshahi, K-M. (2022). Machine learning-based seismic damage assessment of non-ductile RC beam-column joints using visual damage indices of surface crack patterns. *Structures 45*, 2038–2050.

[25] Aravind, N., Nagajothi, S., Elavenil, S. (2021). Machine learning model for predicting the crack detection and pattern recognition of geopolymer concrete beams. *Construction and Building Materials 297*, 123785.

[26] Han, X., Zhao, Z., Chen, L., Hu, X., Tian, Y., Zhai, C., Wang, L., Huang, X. (2022). Structural damage-causing concrete cracking detection based on a deep-learning method. *Construction and Building Materials 337*, 127562.

[27] Laxman, K. C., Tabassum, N., Ai, L., Cole, C., Ziehl, P. (2023). Automated crack detection and crack depth prediction for reinforced concrete structures using deep learning. *Construction and Building Materials 370*, 130709.

[28] Zhang, J., Cai, Y-Y., Yang, D., Yuan, Y., He, W-Y., Wang, Y-J. (2023). MobileNetV3-BLS: A broad learning approach for automatic concrete surface crack detection. *Construction and Building Materials 392*, 131941.

[29] Martinez-Ríos, E.A., Bustamante-Bello, R., Navarro-Tuch, S.A. (2023). Generalized Morse Wavelets parameter selection and transfer learning for pavement transverse cracking detection. *Engineering Applications of Artificial Intelligence 123*, 106355.

[30] Xu, G., Yue, Q., Liu, X. (2023). Deep learning algorithm for real-time automatic crack detection, segmentation, qualification. *Engineering Applications of Artificial Intelligence 126*, 107085.

[31] Yuan, X., Cao, Q., Amin, M-N., Ahmad, A., Ahmad, W., Althoey, F., Deifalla, A-F. (2023). Predicting the crack width of the engineered cementitious materials via standard machine learning algorithms. *Journal of Materials Research and Technology, 24*: 6187 – 6200.

[32] Iraniparast, M., Ranjbar, S., Rahai, M., Nejad, F-M. (2023). Surface concrete cracks detection and segmentation using transfer learning and multi-resolution image processing. *Structures 54*, 386–398.

[33] Katsigiannis, S., Seyedzadeh, S., Agapiou, A., Ramzan, N. (2023). Deep learning for crack detection on masonry façades using limited data and transfer learning. *Journal of Building Engineering 76*, 107105.

[34] Pacal, I. (2024). Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. *Expert Systems with Applications*, 238, 122099.

[35] E. Aslan, M.A. Arserim, A Uçar. (2023). Development of Push-Recovery control system for humanoid robots using deep reinforcement learning. *Ain Shams Engineering Journal*. doi: https://doi.org/10.1016/j.asej.2023.102167.

[36] Lubbad, M., Karaboga, D., Basturk, A., Akay, B. A. H. R. İ. Y. E., Nalbantoglu, U., & Pacal, I. (2024). Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. *Neural Computing and Applications*, 1-25.

[37] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE*, *86*, 2278–2324.

[38] Kurtulus, I. L., Lubbad, M., Yilmaz, O. M. D., Kilic, K., Karaboga, D., Basturk, A., ... & Pacal, I. (2024). A robust deep learning model for the classification of dental implant brands. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 101818.

[39] Karaman, A., Pacal, I., Basturk, A., Akay, B., Nalbantoglu, U., Coskun, S., ... & Karaboga, D. (2023). Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC). *Expert systems with applications*, 221, 119741.

[40] Pacal, I. (2024). MaxCerVixT: A Novel Lightweight Vision Transformer-Based Approach for Precise Cervical Cancer Detection. *Knowledge-Based Systems*: 111482.

[41] Dorafshan, S., Thomas, R.J., Maguire, M. (2018). SDNET2018: An Annotated İmage Data Set for non-contact concrete crack detection using deep convolutional neural networks. *Data in Brief, 21*,1664–1668.

[42] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[43] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*:1409.1556.

[44] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1314-1324).

[45] Mehta, S., & Rastegari, M. (2021). Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv*:2110.02178.

[46] Li, Y., Wu, C. Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., & Feichtenhofer, C. (2022). Mvitv2: Improved multiscale vision transformers for classification and detection. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4804-4814).