www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

OPTIMISED DENOISING-BASED DEEP LEARNING CLASSIFICATION FOR EVALUATING CONCRETE SURFACE CRACKS

SUBMITTED: June 2025
REVISED: September 2025
PUBLISHED: September 2025
EDITOR: Mahesh Babu Purushothaman, Ali GhaffarianHoseini, Amirhosein Ghaffarianhoseini, Farzad Rahimian
DOI: 10.36680/j.itcon.2025.064

Sandra Matarneh, Associate Professor
Department of Architectural Engineering, United Arab Emirates University, Al Ain, UAE
ORCID: https://orcid.org/0000-0002-1351-2780
sandra.matarneh@uaeu.ac.ae

Faris Elghaish, Lecturer
School of Natural and Built Environment, Queen's University Belfast; UK
ORCID: https://orcid.org/0000-0002-7558-6291
F.Elghaish@qub.ac.uk

Farzad Pour Rahimian, Professor
School of Computing, Engineering and Digital Technologies, Teesside University; UK
ORCID: https://orcid.org/0000-0001-7443-4723
F.Rahimian@tees.ac.uk

Essam Abdellatef, Assistant Professor
Faculty of Computer Science and Engineering, Alamein International University, Matrouh; Egypt
essam.abdellatef@Aiu.edu.eg

Algan Tezel, Associate Professor
College of Engineering & Technology, University of Doha, Qatar
algan.tezel@udst.edu.qa

Abdul-Majeed Mahamadu, Associate Professor
The Bartlett School of Sustainable Construction, University College London, UK
ORCID: https://orcid.org/0000-0001-7757-8562
a.mahamadu@ucl.ac.uk

Mohammed Abdelmegid, Lecturer
School of Civil Engineering, University of Leeds, UK
ORCID: https://orcid.org/0000-0001-6205-570X
M.Abdelmegid@leeds.ac.uk

SUMMARY: *Conventional visual inspections of concrete structures are hazardous, time-consuming, and prone to subjectivity, which has accelerated the adoption of automated image-based techniques for structural health monitoring. Deep learning methods, particularly convolutional neural networks (CNNs), offer significant potential for crack detection, yet their accuracy is often compromised by image noise arising from environmental conditions, sensor artefacts, and preprocessing. This study systematically evaluates the integration of five state-of-the-art denoising approaches (HRL, SANet, ADNet, SW-CNN, CDNet) with six pre-trained CNN architectures (AlexNet, VGG19, GoogLeNet, ShuffleNet, ResNet-101, and Xception) to enhance concrete crack detection and classification. The research is structured into three methodological stages. First, the baseline classification performance of the six pre-trained CNN architecturesis evaluated using a dataset of 40,000 concrete surface images, evenly divided between cracked and non-cracked samples. Second, five state-of-the-art denoising methods are applied as a preprocessing step to mitigate noise effects prior to classification. Third, the impact of each* *denoising approach is quantitatively assessed using accuracy, sensitivity, and F1-score metrics. The integration of denoising techniques led to substantial performance improvements across all models. For instance, AlexNet's F1-score increased from 53.31% to 71.19%, while Xception achieved the highest overall F1-score of 97.72% and accuracy of 97.7% following denoising. ResNet-101 similarly improved to 96.3% accuracy and 96.27% F1-score. Lightweight models such as ShuffleNet also demonstrated excellent gains, reaching 90.5% accuracy and 89.58% F1-score when paired with SW-CNN. Notably, SW-CNN yielded the most consistent performance, achieving the highest F1-score in four of the six models, while CDNet and ADNet were especially effective in boosting sensitivity metrics. Efficiency analysis further highlighted practical deployment trade-offs: ShuffleNet+SW-CNN achieved 3.7 ms/image latency, ~270 images/s throughput, and an 18 MB model size, making it suitable for edge devices, whereas Xception+SW-CNN, though heavier (228 MB, 11.2 ms/image), maximized accuracy for server-class monitoring. These results underline the importance of balancing performance and efficiency in real-world applications. On average, the application of denoising methods resulted in F1-score improvements of 13–15%, underscoring the effectiveness of preprocessing in enhancing model reliability. These findings highlight the critical role of image denoising in improving the performance of deep learning-based crack detection systems. Moreover, the combination of efficient CNN architectures with robust denoising offers promising pathways for both edge deployment and server-based structural monitoring solutions. This research demonstrates that coupling CNNs with denoising substantially enhances crack detection robustness and reliability, contributing to safer and more scalable structural health monitoring systems. Future work should validate the pipeline on external datasets, perform controlled noise-stress testing, and integrate domain-specific augmentations to ensure generalizability across diverse materials and field conditions.*

KEYWORDS: *structural health monitoring, concrete crack detection, convolutional neural networks (CNNs), image denoising, deep learning, model efficiency, digital construction.*

REFERENCE: *Sandra Matarneh, Faris Elghaish, Farzad Pour Rahimian, Essam Abdellatef, Algan Tezel, Abdul-Majeed Mahamadu & Mohammed Abdelmegid (2025). Optimised denoising-based deep learning classification for evaluating concrete surface cracks. Journal of Information Technology in Construction (ITcon), Special issue: 'Smart and Sustainable Built Environment (SASBE 2024)', Vol. 30, pg. 1573-1594, DOI: 10.36680/j.itcon.2025.064*

1. INTRODUCTION

Human-performed conventional visual assessments of buildings are regarded to be dangerous and frequently yield inconsistent outcomes (Chan et al., 2015; S. Chen et al., 2019). To improve both safety and efficiency, efforts have been made to automate the process of physically examining structures (Chow et al., 2021) as well as the detection and classification of any damage (Ali et al., 2021). Human analysis effort is being replaced by techniques, including traditional image processing methods (referred to as white-box techniques), less transparent artificial neural networks (often known as black-box techniques) (Hsieh & Tsai, 2020), and the advent of deep learning in defects detection and segmentation (Chow et al., 2021).

Currently, the structural health monitoring community heavily relies on the use of deep learning techniques (Zhao et al., 2019), including image-based damage and crack detection (Fan et al., 2020; Jang et al., 2019; Xu et al., 2019). The automatic tuning of network parameters using backpropagation algorithms (Géron, 2022; Goodfellow, 2016) is an intrinsic feature of deep neural networks, particularly convolutional neural networks (CNNs). This capability offers an alternative methodology for conducting automated defect inspection using images in a wide range of conditions. One common utilisation of this approach was in the identification of defects, with most of the research focused on the binary classification of cracks (Chen & Jahanshahi, 2018; Dung & Le Duc, 2018; Flah et al., 2020; Jang et al., 2019; Li & Zhao, 2019; Ni et al., 2019).

Defect classification can be achieved on either the whole image or on patches (cropped parts of an image). In the case of patches, the output results are stitched together to form a final comprehensive prediction. Subsequently, deep learning models were applied in the domain of object detection, where bounding boxes of different hues were used to define different types of defects (Kang et al., 2020; Y. Zhang et al., 2020). Deep learning models were also used to generate pixel-wise predictions of defects on various civil infrastructure such as road pavements (Huyan et al., 2020; Liu et al., 2019; W. Song et al., 2019; Zhou & Song, 2021; Zou et al., 2018); tunnels (Ren et al., 2020; Q. Song et al., 2019), bridges (Alipour et al., 2019; Li & Zhao, 2020), and dams (Feng et al., 2020). In conclusion, the use of deep learning models in image-based defect detection can differ depending on the unique inspection needs and the availability of annotated datasets for training the models.

Tolerance and acceptable levels of error vary greatly depending on the application. Blurring, area thresholding, and gap connection errors, for example, are acceptable in the identification of large road cracks since cracks smaller than these tolerances are insignificant. In contrast, ensuring that the reduction of noise pixels does not impair the image's real positive pixels is of the utmost importance when it comes to water-retaining buildings built of concrete, which must have a crack detection limit of 0.1 mm (Dow et al., 2023). It is possible to reduce noise in binary images through adjustments to the hardware and post-processing methods employed during image acquisition. On the other hand, algorithms are utilised to eliminate residual noise and environmental noise (Zhang et al., 2019).

Despite ongoing advancements in deep learning and computer vision, achieving precise and resilient crack segmentation in concrete images remains a persistent challenge. While image denoising methods have shown promise across diverse domains, the unique characteristics of concrete surfaces such as heterogeneous textures, shadows, and environmental artefacts pose additional complexities that general-purpose approaches often fail to address. These conditions amplify the difficulty of distinguishing subtle cracks from noise, thereby undermining the accuracy and generalisability of automated inspection systems.

Although several studies have explored CNN-based crack detection, only limited efforts have systematically evaluated the integration of denoising methods specifically tailored to concrete surface images. Consequently, the comparative benefits and trade-offs of different denoising strategies remain unclear in this context, leaving practitioners without a clear framework for selecting model-denoising combinations.

Therefore, this study addresses the following problem: how can advanced CNN-based denoising methods be systematically integrated with pre-trained deep learning architectures to improve the accuracy, precision, and recall of automated concrete crack detection?

To address this problem, this study investigates a range of advanced denoising techniques as a preprocessing step to improve the robustness of crack classification in concrete surface images. The proposed approach integrates several CNN-based denoising methods—HRL, SANet, ADNet, SW-CNN, and CDNet—into the image classification pipeline to enhance the performance of six widely used CNN architectures, including AlexNet, VGG19, GoogLeNet, ShuffleNet, ResNet-101, and Xception. The methodology is empirically validated using a balanced dataset of 40,000 RGB images, demonstrating average performance improvements of 13–15% in F1-score following denoising.

The primary objective of this research is to address two core challenges: the accurate detection and classification of concrete surface cracks, and the mitigation of diverse noise distortions that impair visual inspection accuracy. By systematically evaluating the integration of denoising models with CNN architectures, this study aims to establish a reliable and scalable framework for enhancing structural health monitoring systems.

In doing so, the research contributes both practical solutions for real-world deployment and theoretical insights into the role of image preprocessing in deep learning–based visual inspection. The findings serve as a foundation for more robust, efficient, and generalisable concrete defect detection systems.

2. RELATED WORK

2.1 Image denoising

In recent years, there has been a significant surge in the utilisation of images. Noise is a negative component that contributes to image corruption throughout the acquisition, compression, and transmission processes. The quality of images can be negatively affected by the presence of noise that is transmitted through several channels, including environmental and transmission channels. The notion of image noise within the field of image processing refers to the stochastic fluctuations in the level of a signal, which have the potential to impact the clarity of an image while trying to extract useful information. The presence of noise can significantly impair various image processing operations, such as video processing, image analysis, and segmentation, ultimately resulting in inaccurate diagnoses (Diwakar & Kumar, 2018). Therefore, the process of image denoising is of paramount importance in enhancing the image processing activities. The objective of denoising processes is to eliminate undesired noise and restore an image to its original, unaltered condition. Nevertheless, the task of distinguishing between noise, edges, and textures presents a considerable difficulty in image denoising due to the existence of high-frequency elements in each of these components.

Recently, there has been a significant increase in the advancement of convolutional neural network (CNN) methods, which have demonstrated remarkable efficacy in various low-level computer vision tasks (Kim et al., 2016; Nah et al., 2017). The application of a CNN for the purpose of image denoising may be traced back to the study conducted by Jain and Seung (Jain & Seung, 2008), in which a five-layer network architecture was proposed and implemented. Several CNN-based denoising approaches have been presented in the past several years (Chen & Pock, 2017; Cruz et al., 2018; Vincent et al., 2008; Zhang et al., 2017; Zhang et al., 2018). The performance of these approaches has shown significant improvement when compared to Jain and Seung (Jain & Seung, 2008). Moreover, denoising techniques based on CNNs may be categorised into two distinct groups: models employing multilayer perception (MLP) and methods utilising deep learning approaches.

2.2 Multilayer perception models

The multilayer perceptron (MLP) mimics the structure of the human brain. MLP is also synonymous with feed-forward artificial neural network (ANN). Between the input and output, MLP contains several hidden layers. The number of concealed layers is determined by the work at hand. Every neuron in the hidden layer communicates with neurons in the following layer. Weights are the connecting wires between neurons, and their values are adjusted with the help of the learning phase. The learning phase is performed indefinitely until the error value is less than the threshold level. The input layer consists of a combination of feature values. The classification will be predicted by the output layer based on the information received from the input layer. The classified output is compared to the observed output, and the error is computed. The adjustment of network weights occurs in a sequential manner, starting at the output layer and propagating towards the input layer, with the intermediate layer serving as the conduit for this process. The magnitude and direction of these weight changes are determined by the error seen during the training phase. The utilisation of interconnected weights, node values, and activation functions enables the computation of transmitted information (Kruse et al., 2022).

Various scholars have implemented multilayer perception-based image denoising models, such as Vincent et al. (2008) who developed deep architectures using auto-encoders, and Xie et al. (2012) who combined sparse coding and deep networks pre-trained with denoising auto-encoder (DA) to address low-level vision problems. A framework for image restoration known as trainable nonlinear reaction diffusion (TNRD) was introduced by Chen and Pock (2017). The suggested methodology attained cutting-edge results in various tasks, including image denoising, super-resolution, and JPEG deblocking, through the integration of learned filters and influence functions. Another method for despeckling synthetic-aperture radar (SAR) images utilising a MLP neural network was proposed by Tang et al. (2019). The MLP is trained using SAR image segments in order to acquire knowledge of the intensity attributes and autonomously establish thresholds and weights for despeckling. The approach demonstrated adequate efficacy in noise reduction and edge preservation in SAR images. Fociro et al. (2023) combined image processing and a multilayer perceptron (MLP) neural network to enable the automated classification of carbonate pebbles and identification of the Dunham texture. This methodology is predicated on the examination of grayscale images acquired from thin sections. Results showed that thin-section photographs achieved an accuracy of 91.3% and 90.5%, respectively, on two distinct test sets comprised of 348 and 250 photos. In their study, Ieracitano et al. (2021) introduced an innovative approach to classify automatically scanning electron microscope (SEM) images of nanofiber patches, including both homogeneous and non-homogeneous patches. To accomplish this goal, a classification system is developed that combines unsupervised and supervised machine learning methodologies. In particular, the system integrates MLP, which receives training via supervised learning, with an autoencoder, which is acquired through unsupervised learning. The results of the experimental simulations indicate that the hybrid technique outperforms traditional machine learning approaches, achieving an accuracy rate of as high as 92.5%.

2.3 Deep learning-based denoising methods

Recent sophisticated methods for denoising in the field of deep learning commonly rely on Convolutional Neural Networks (CNNs). CNNs were initially implemented in image denoising tasks by (Chiang & Sullivan, 1989). A neural network was utilised as a weighting factor in this research attempt to efficiently eliminate complex noise. Following this, a feedforward network (Hu et al., 2020) was implemented in order to attain a favourable compromise between the efficacy and performance of the resulting denoised images. In the early phases of CNN development, significant challenges such as the vanishing gradient problem, the choice of activation functions (namely sigmoid (Marreiros et al., 2008) and Tanh (Jarrett et al., 2009)), and the lack of compatible hardware platforms. However, the advent of AlexNet in 2012 (Krizhevsky et al., 2017) has substantially transformed the difficulties linked to the implementation of CNNs due to its capacity to attain exceptionally precise results on extremely complex datasets. Additional CNN architectures, like as VGG (Ha et al., 2018) and GoogLeNet (Tang et al., 2017), have been utilised in the field of computer vision for various tasks.

Table 1: Comparison of CNN denoising methods.

Author	Aim	CNN name	Noise type	Results
(Zhang et al., 2019)	To remove noise with unknown distribution	Dictionary learning model	Gaussian-mixed noise	Results demonstrated superior performance compared to previous denoising methods.
(Hong et al., 2019)	To tackle the extensive range of natural image patches encountered in the process of image denoising	Patch complexity local divide (PCLDCNet)	General noise	Results showed that the approach significantly enhances denoising performance compared to single network approaches, while requiring fewer training samples and parameters.
(Quan et al., 2021)	To investigate the capabilities of complicated valued CNNs in the context of image denoising	Complex-valued Denoising Network (CDNet)	General noise	CDNet has competitive performance when compared to real-valued CNNs, it demonstrates enhanced resilience to inconsistencies arising from variations in noise models between the training and test images.
(Xu et al., 2020)	To improve denoising performance	Bayesian deep matrix factorization network (BDMF)	General noise	The efficacy of BDMF has been demonstrated through both synthetic trials and real-world scenarios.
(Yin et al., 2020)	To filter images efficiently	Side Window Convolutional Neural Network (SW-CNN)	General noise	Results revealed that the approach attained better performance in comparison to the state-of-the-art networks, reducing learnable parameters by 96%, memory consumption by 50%, and running time by 50%.

(Sadrizadeh et al., 2022)	To eliminate impulsive noise from images	A blind CNN	Impulsive noise	The suggested methodology demonstrates superior performance compared to other techniques in both reconstruction quality and computational efficiency.
(Giannatou et al., 2019)	To mitigate noise and improve the precision of Line Edge Roughness (LER) metrology.	Scanning Electron Microscopy Denoising (SEMD)	General noise	The combined approach of SEMD and Power Spectral Density (PSD) methods improves LER parameter predictions.
(Shi et al., 2019)	To avoid network degradation and denoise images	Hierarchical residual learning	General noise	The method achieved superior performance compared to state-of-the-art methods in Gaussian denoising and single image super-resolution tasks
(Q. Zhang et al., 2020)	To remove non independent & identically distributed (non-i.i.d.) noise.	Deep spatio-spectral Bayesian posterior (DSSBPNe)	General noise	Results showed the efficiency of the approach in improving denoising performance compared to existing methods.
(Qi et al., 2022)	To denoise remote sensing images	Anisotropic weighted total variation feature fusion network (AWTVF2Net)	General noise	Results demonstrated that AWTVF2Net outperformed other existing algorithms.
(Solovyeva & Abdullah, 2022)	To improve accuracy and reduce network complexity in denoising and deblurring images.	Dual Autoencoder Network with Separable Convolutional Layers	Gaussian noise, Poisson noise, impulse noise, and speckle noise	The dual autoencoder achieved a high level of quality and stability through the utilisation of a separable CNN. This network effectively minimised the amount of learnable parameters and processing time, hence optimising the computational cost.
(S. H et al., 2023)	To address the issue of restoring noise in images that have been affected by additive white Gaussian noise	Edge-focused image denoising (EFID)	Gaussian noise	EFID achieved better preservation of textures and edges while eliminating noise compared to conventional methods.
(Lyu et al., 2020)	To eliminate mixed noise in images	Denoising generative adversarial network (DeGAN)	Mixed noise	The model outperformed state-of-the-art methods in removing mixed noise in three different scenarios
(Li et al., 2020)	To address the issue of detail loss in denoised images	A detail retaining convolutional neural network (DRCNN)	Gaussian noise	The suggested approach outperformed other denoising techniques in terms of image quality, generalization ability, and adaptability to different image restoration tasks.
(Guo et al., 2020)	To derive an estimation of the noise level distribution within a given area and then eliminate noise from the input image	Noise estimation and removal network (NERNet)	Realistic noise	NERNet achieves competitive results on both synthetic and realistic noisy images compared to state-of-the-art methods.

Zhang et al. (2017) introduced DnCNN, an image denoising deep convolutional neural network. The system integrates batch normalisation and residual learning, enabling it to effectively manage Gaussian denoising when the levels of noise are undetermined. A single model is utilised to train the network to perform multiple denoising tasks, including JPEG image deblocking, single image super-resolution, and Gaussian denoising. Experimental results showed that DnCNN outperformed existing state-of-the-art methods in terms of denoising efficacy. Another study conducted by Zhang et al. (2018) proposed FFDNet, a convolutional neural network designed for image denoising that is both quick and adaptable. By specifying a non-uniform noise level map, FFDNet is capable of removing spatially variant noise and can manage a broad spectrum of noise levels. In relation to both denoising performance and computational efficiency, it surpasses benchmark methods. Islam et al. (2018) proposed a method using CNN to reduce mixed Gaussian-impulse noise from images. Transfer learning is implemented in the CNN model to accelerate training and improve performance over existing techniques. The experimental outcomes demonstrate its superior precision and robustness. The attention-guided CNN (ADNet) was introduced by Tian et al. (2020) as a method for denoising images. The ADNet framework comprises an overall number of seventeen layers, each of which is structured into one of the following four blocks: attention (AB), feature enhancement (FEB), sparse block (SB), or reconstruction (RB). ADNet outperformed existing best practices in denoising, including blind denoising, on both synthetic and actual noisy images. A method named Multi-Level Information Fusion Convolutional Neural Network (MLIFCNN) was introduced by Xie et al. (2023) for image denoising. Four steps comprise the methodology: reconstruction (RB), multi-level information interaction (MIIB), coarse information refinement (CIRB), and fine information extraction (FIEB). FIEB employed parallel group convolutions to derive wide-channel information, whereas MIIB combined deep and wide-channel information via residual operations. RB obtained a clear image through the utilisation of a residual operation, while CIRB further refined the acquired information. The experimental findings provided evidence that the proposed method exhibited superior performance compared to alternative denoising methods. Table 1 shows a comparison between several studies that used different CNN denoising methods.

2.4 Deep learning-based concrete images denoising

CNNs are very effective at learning image characteristics utilising a simpler network topology than traditional machine learning approaches (Simard et al., 2003). By focusing on this positive features, algorithms based CNNs exhibit a high level of efficiency in detecting cracks, especially in scenarios involving multi-classification (Shin et al., 2016) and for large scale visual recognition (Yan et al., 2015). Furthermore, using transfer learning, existing CNN structures may be simply updated and used for crack detection (Radenović et al., 2016), increasing CNNs' adaptability for treating varied crack images. Several studies have focused on the concept of concrete crack detection (Cha et al., 2017; Fu et al., 2020). In these studies, the image was initially divided into sub-regions. From these sub-regions, crack characteristics were extracted in order to create the feature vectors. The feature vectors were utilised to train a CNN, which was subsequently employed to identify cracks within each sub-region of the entire image. The identification of the crack in the entirety of the image was ultimately achieved using the mix of the identification outcomes from each individual sub-region.

While image possessing inherent noise immunity, CNNs confront difficulty in accurately detecting concrete cracks (Zhang et al., 2017). This is because image background noise contains a high number of different features, complicating crack feature extraction and jeopardising the accuracy, efficiency, and versatility of CNNs. Unfortunately, as frequently experienced in practice, significant background noise originating in concrete surfaces from complex and diverse sources can hardly be avoided. With the development of deep learning methods, researchers started recently focusing on the combination of CNNs and image denoising methods for concrete images. For example, Fu et al. (2020) proposed a framework that combines conventional CNNs with a MLP strategy to enhance the accuracy and noise immunity of crack identification in concrete structures. The framework involved homomorphic filtering and the Otsu thresholding method to preprocess concrete surface images and extract crack features. The combination of CNNs and the MLP strategy improved the versatility of crack identification. Results demonstrated the effectiveness and efficiency of the proposed framework with increased accuracy of crack position detection by 3.1% under a moderate noise level. Another study conducted by GAN et al. (2023) proposed a collaborative denoising approach for Cone Beam Computed Tomography (CBCT) images through a combination of image segmentation and an unsupervised learning-based denoising algorithm. The method improves denoising by applying varying degrees of denoising to different regions of the CBCT image, guided by segmentation results. In their study, Flah et al. (2020) presented a novel technique for crack detection and quantification in concrete structures using deep learning image-based techniques. A CNN combined with improved Otsu image processing is used to classify cracks based on their orientation and quantify crack features such as length, width, and angle of orientation. Results showed that the proposed method achieved high accuracy in classifying cracks based on their orientation, achieving testing accuracies of over 96% for the different classifiers. Another study conducted by Jang et al. (2019) presented a deep learning-based technique for autonomous concrete crack detection using hybrid images obtained from vision and infrared thermography. The proposed method utilises a well-trained deep CNN (GoogLeNet) for crack identification and visualization while minimising false alarms and a statistical denoising process to remove undesired noise in concrete images. Experimental validation on lab-scale concrete specimens showed successful detection of macro- and microcracks.

Zhu and Song (2020) proposed a model that improves the VGG-16 CNN to accurately classify surface defects on cement concrete bridges. The model reduced the number of fully connected layers and replaced the Softmax classifier with a Softmax classification layer with seven defect tags. It used morphology-based weight adaptive denoising for image preprocessing and applied transfer learning by fine-tuning the pre-trained VGG-16 model.

Comparative experiments with other models, including neural networks (BPNN, SVM) and deep CNNs (AlexNet, GoogLeNet, ResNet), showed that the proposed model outperformed them in terms of mean detection accuracy and top-5 accuracy. The model effectively extracted multi-layer features from surface defect images, highlighting edges and textures. Xu et al. (2023) presented a method that involved acquiring large-scene images, denoising the background, using a maximum crack width calculation algorithm, and using the YOLOv5 algorithm for noise-resistant crack detection in bridges. The results showed improved detection efficiency and accuracy of up to 93.4%. Dow et al. (2023) introduced a new noise removal method, "Skele-Marker", for binary concrete crack images. The method achieved high recall (77%), precision (91%), intersection over union (72%), and F1 score (84%) in cracking detection. Finally, Fan et al. (2020) introduced a novel method for denoising vibration signals in structural health monitoring using a specialized Residual Convolutional Neural Network (ResNet), showcasing effective noise reduction and accurate modal identification even with different noise types and levels.

Table 2: Comparison of CNN based concrete images denoising methods.

Author	**Aim**	**CNN name**	**Noise type**	**Results**
(Fu et al., 2020)	To mitigate the impact of significant noise present in concrete surface images.	CNN, homomorphic filtering and the Otsu thresholding method	General noise	The MLP-CNN model demonstrated enhanced identification accuracy of 2.8% and 5.4% respectively when subjected to the presence of light spot and blur.
(Flah et al., 2020)	To classify and quantify cracks in concrete structures	CNN, OTSU, the non-linear filter and morphology for denoising.	General noise	The algorithm demonstrated high accuracy in classifying cracks based on their orientation, achieving testing accuracies of over 96% for the different classifiers.
(Jang et al., 2019)	To identify and visualize crack while minimizing false alarms.	GoogLeNet and statistical denoising process	Gaussian noise	Experimental validation on lab-scale concrete specimens shows successful detection of macro- and microcracks
(Zhu & Song, 2020)	To accurately identify the surface defects found on cement concrete bridges	VGG-16 and morphology-based weight adaptive denoising	Impulse noise	The proposed model successfully recovers multiple layers of features from surface defect images, effectively emphasising edges and textures.
(Xu et al., 2023)	To detect cracks utilizing large-scene images acquired by a UAV	YOLOv5 and a background denoising algorithm	General noise	The results showed that significant improvement in detection efficiency with accuracy reached up to 93.4%
(Dow et al., 2023)	To remove binary noise and segment noisy concrete crack images.	Skele-Marker method	General noise	The method achieved high recall (77%), precision (91%), intersection over union (72%), and F1 score (84%) in cracking detection.
(Fan et al., 2020)	To mitigate the impact of noise, particularly in challenging and adverse conditions encountered during structural health monitoring.	A vibration signal denoising approach based on a specialized Residual Convolutional Neural Networks (ResNet)	General noise	The ResNet extracts advanced features from the vibration signal and automatically grasps the structural modal information. As a result, it effectively retains the crucial vibration traits in the signals and helps differentiate the genuine physical modes from the false ones during structural modal identification.
(Dorafshan et al., 2018)	To evaluates how edge detectors stack up against DCNN in identifying cracks in concrete structures images.	AlexNet DCNN and edge detectors	Residual noise	The most effective technique, LoG, identified approximately 79% of damaged pixels accurately and identified cracks larger than 0.1 mm. In contrast, the most effective DCNN method, precisely identified 86% of damaged images and was able to detect cracks larger than 0.04 mm.

The above literature shows the efforts on image-based denoising by considering different techniques to improve the image quality for more accurate concrete crack detection. However, the existing research studies have primarily focused on enhancing image quality and clarity, very few studies were found focusing on denoising concrete images for accurate crack detection. Solving the ongoing challenge of image noise to increase the accuracy of crack detection in concrete will require substantial efforts. Hence, this study aims to make significant progress by addressing a fundamental issue in concrete image denoising. It focuses on reducing the influence of errors caused by noise during image acquisition and mitigating the influence of noise on crack detection accuracy. In addition, the existing studies focused on utilising one CNN denoising technique as shown in Table 2. Therefore, this study aims as well to compare the performance of different denoising CNN approaches on the collected dataset of concrete images.

3. METHODOLOGY

This study adopts a positivist philosophical stance and applies deductive reasoning to empirically evaluate and validate deep learning (DL) models within a deterministic research framework, consistent with the approaches discussed by Howden-Chapman et al. (2023) and Owusu-Manu et al. (2022). According to P. C. Chen et al. (2019), the development of machine learning and deep learning models typically follows a structured process, encompassing problem selection, data collection, model development, validation, impact assessment, and deployment.

The dataset used in this study was obtained from the work of Sorguç (2018) comprises 40,000 RGB images equally split into two subsets: 20,000 images of concrete surfaces with cracks (positive) and 20,000 without cracks (negative). Each image was resized to 227 × 227 pixels. These were derived from 458 high-resolution original photographs, each with a resolution of 4032 × 3024 pixels, taken from various buildings on the METU Campus. The source images captured a diverse range of surface conditions, including exposed concrete, plastered, and painted finishes. Although surface textures and conditions varied, the images were taken on the same day under consistent illumination, with the camera positioned approximately one metre from the surface and facing directly toward it. This consistency helped minimise variability due to environmental factors while preserving the diversity necessary for robust model training.

Even under such controlled conditions, noise was present in the images from both real and synthetic sources. Real noise originated from sensor limitations such as electronic shot noise, compression artefacts from JPEG encoding, and subtle differences in illumination across surfaces. In addition, concrete itself introduced complexity through pores, efflorescence, paint irregularities, and micro-shadows, which can resemble or obscure cracks. Synthetic noise also arose during pre-processing, particularly when high-resolution images were down-sampled to 227 × 227 pixels, occasionally blurring fine crack boundaries. In this context, denoising is advantageous as it enhances the signal-to-noise ratio, suppresses irrelevant texture variation, and reduces compression artefacts while preserving crack features. This improves the discriminative ability of CNN models, lowering false positives from background textures and false negatives from faint cracks.

Notably, no data augmentation methods, such as rotation, flipping, or tilting, were employed, allowing the models to be tested against the raw features of the dataset.

The experimental phase involved the implementation and testing of six pre-trained Convolutional Neural Network (CNN) models: AlexNet, VGG19, GoogLeNet, ShuffleNet, ResNet-101, and Xception. Each model was assessed using standard performance metrics, namely accuracy, sensitivity, and F1-score. Following the initial evaluation, five advanced de-noising techniques were applied to each CNN model to enhance performance and reduce noise-related distortions. These de-noising methods included Hierarchical Residual Learning (HRL), Self-Attention Network (SANet), Adaptive de-noising Network (ADNet), Sliding Window CNN (SW-CNN), and Contextual Denoising Network (CDNet).

The application of these de-noising approaches led to measurable improvements in model performance. The minimum recorded improvement was 10.3%, observed in the Xception model, while the highest improvement was 17.7%, seen in the VGG19 model. On average, the de-noising strategies resulted in a 13.82% average point increase in F1-score across all CNN architectures tested. These findings provide strong validation for the incorporation of de-noising techniques in CNN-based image classification tasks, particularly for detecting cracks in concrete surfaces. The methodology and process flow used in evaluating these models are illustrated in Figure 1.

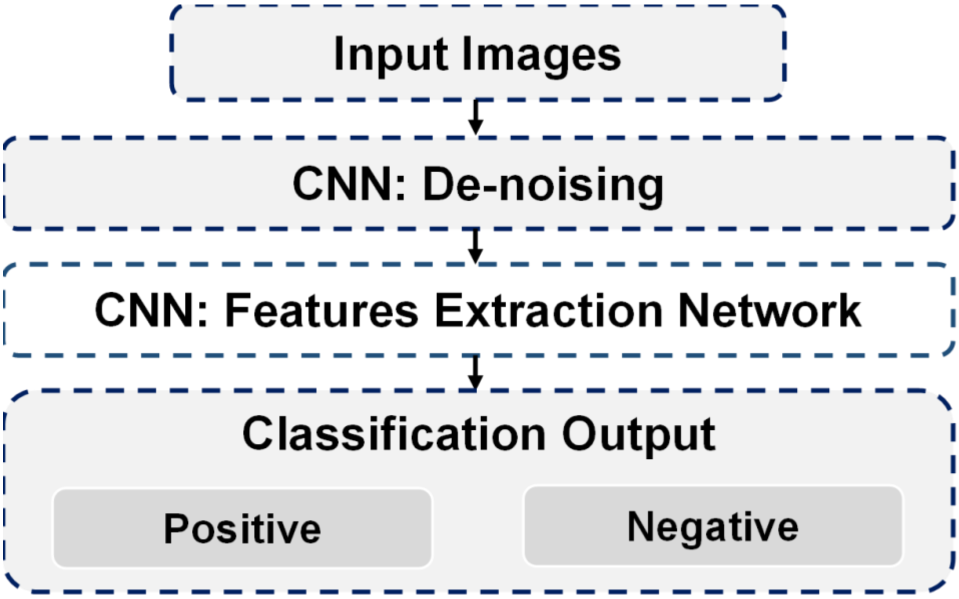

Figure 1: Classification methodology of concrete surface crack images using de-noising CNN approach.

Reproducibility and Experimental Setup

To ensure reproducibility, the dataset was divided using an 80/10/10 train/validation/test split with stratified sampling to preserve the cracked/non-cracked balance across sets. All experiments were run with a fixed random seed (42) and repeated three times to account for stochastic variation. Training was conducted using the Adam optimizer with an initial learning rate of 0.001 following a cosine annealing schedule. Each model was trained for 50 epochs with a batch size of 32. Only the final fully connected layers were fine-tuned, while convolutional backbones were frozen to leverage transfer learning effectively.

Experiments were executed on an NVIDIA RTX 3090 GPU with 24 GB VRAM and a host system equipped with 128 GB of RAM, running Ubuntu 22.04. The implementation used PyTorch version 2.0.1 with CUDA 11.8 and cuDNN 8.7. No data augmentation (e.g., rotation, flipping, tilting, brightness shifts) was applied in any experiment, ensuring consistency across all model evaluations.

4. RESULTS

Baseline CNN Performance The performance of six pre-trained CNN models is evaluated as shown in Table 3. Table 3 clearly shows that modern, deeper CNNs like ResNet-101 and Xception significantly outperform older models in the context of concrete crack classification with F1-score 84.58 and 88.17% respectively. While models like AlexNet and VGG19 provide baseline results, they are not sufficient for high-accuracy crack detection applications. Xception stands out as the most reliable choice when computational resources allow, whereas ShuffleNet offers a practical alternative for real-time embedded deployments.

4.1 Crack Classification Using Different Pre-Trained CNNs With Various De-Noising Techniques:

Table 4 provides a performance benchmark for six popular CNN architectures namely AlexNet, VGG19, GoogLeNet, ShuffleNet, ResNet-101, and Xception. Each architecture was tested in combination with five denoising CNN approaches: HRL, SANet, ADNet, SW-CNN, and CDNet. The performance metrics analyzed include accuracy (%), sensitivity (%), and F1-score (%), which are essential for evaluating the reliability of classification in concrete surface crack detection.

Denoising significantly compensates for AlexNet's limited representational depth, with SW-CNN offering optimal enhancement. SW-CNN achieved 68% accuracy, 79.01% sensitivity, and 71.19% F1-score, the highest across all metrics. Although CDNet yielded better sensitivity (89.08%) with VGG19, its F1-score was slightly lower due to likely higher false positives. However, SANet delivered the highest accuracy (83.7%) and F1-score (85.97%) when it comes to VGG19. GoogLeNet pairs best with CDNet or SANet, depending on whether F1-score or sensitivity is prioritized. CDNet achieved the highest F1-score (87.93%) and accuracy (86.4%), confirming its superior precision-recall balance. SANet had the highest sensitivity (98.54%), suggesting excellent detection of cracked regions, though possibly with reduced precision.

Table 3: Performance of various pre-trained CNN Models used for classification of concrete surface crack images.

CNN	Accuracy (%)	Sensitivity (%)	F1-score (%)
AlexNet	52.9	53.75	53.31
VGG19	66	75.53	68.97
GoogLeNet	75.18	54.2	68.59
ShuffleNet	77.1	78.06	77.32
ResNet -101	83.1	92.87	84.58
Xception	**87.4**	**93.69**	**88.17**

Notably, ADNet achieved higher sensitivity (96.53%) with ShuffleNet, suggesting it captures almost all crack cases, though potentially at the expense of precision. Meanwhile, CDNet is more balanced with 90.5% accuracy, 82.02% sensitivity, and 89.58% F1-score. Even for deep residual models, denoising networks like SW-CNN still significantly enhance learning from noisy input data in ResNet-101 yielding 96.3% accuracy, 96.78% sensitivity, and 96.27% F1-score. Xception is highly compatible with all five denoising networks, but SW-CNN offers the most optimal configuration. Again, SW-CNN leads with 97.7% accuracy and 97.72% F1-score, affirming its dominance in deep, efficient models.

Table 4: Performance of different CNNs with applying various de-noising approaches.

De-noising CNN Approaches	Accuracy (%)	Sensitivity (%)	F1-score (%)
AlexNet			
HRL	57.20	57.35	57.26
SANet	64.30	43.93	55.15
ADNet	59.90	44.16	53.19
SW-CNN	**68.00**	**79.01**	**71.19**
CDNet	56.80	63.33	59.46
VGG19			
HRL	80.80	93.86	83.02
SANet	**83.70**	**75.39**	**85.97**
ADNet	71.50	82.27	74.25
SW-CNN	80.60	85.81	81.59
CDNet	81.20	89.08	82.57
GoogLeNet			
HRL	84.80	93.03	86.6
SANet	84.20	98.54	86.16
ADNet	83.90	91.74	86.07
SW-CNN	84.30	92.34	86.44

CDNet	**86.40**	**92.03**	**87.93**
ShuffleNet			
HRL	88.00	92.97	88.58
SANet	85.60	94.41	87.34
ADNet	89.40	96.53	90.43
SW-CNN	89.30	84.4	90.35
CDNet	**90.50**	**82.02**	**89.58**
ResNet-101			
HRL	92.60	99.61	93.1
SANet	91.00	95.96	91.73
ADNet	91.30	93.05	91.92
SW-CNN	**96.30**	**96.78**	**96.27**
CDNet	93.50	88.56	93.93
Xception			
HRL	96.70	99.96	96.8
SANet	96.50	98.49	96.53
ADNet	97.20	99.84	97.23
SW-CNN	**97.70**	**95.55**	**97.72**
CDNet	97.40	99.84	97.49

Table 4 illustrates that SW-CNN emerges as the most universally effective denoising method, achieving top scores in both high-capacity (Xception, ResNet-101) and low-capacity (AlexNet) models. It also confirms the synergistic value of integrating denoising CNNs with mainstream deep learning architectures. The consistent improvements in accuracy, sensitivity, and F1-score reinforce that noise reduction is not merely optional but critical for optimizing classification in real-world crack detection applications.

The confusion matrices presented in Figure 2 provide a comparative performance analysis of six pre-trained CNN models, each integrated with their respective optimal denoising approaches for concrete surface crack classification. As demonstrated in Figure 2(a), while SW-CNN enhances AlexNet's performance, the model's detection capabilities prove insufficient for crack detection, exhibiting a substantial false negative rate of approximatly 21 % (sensitivity 79.01%) and modest precision (~64.8%). With an accuracy of only 68%, AlexNet demonstrates the lowest performance among all evaluated models. In contrast, VGG19 paired with SANet (Figure 2(b)) achieves superior results, though it still fails to detect 25% of cracks. Notably, this configuration achieves zero false positives, rendering it particularly suitable for scenarios where false alarms must be minimized. Figure 2(c) illustrates the effectiveness of CDNet in enhancing GoogLeNet 's performance, yielding 99% precision with merely 1% false positives. Meanwhile, ShuffleNet combined with CDNet (Figure 2(d)) attains an exceptionally high recall of 98.7%, though at the expense of an elevated false positive rate (18%). ResNet-101 integrated with SW-CNN (Figure 2(e)) exhibits a balanced trade-off between precision and recall, achieving 95.8% recall and 96.8% precision (3.2% false positives), positioning it as a leading CNN model for reliable crack detection. Similarly, Xception coupled with SW-CNN (Figure 2(f)) demonstrates exceptional performance, attaining almost 100% precision (no false positives) and a recall of 95.5%, marginally lower than that of ResNet-101.

The results present various efficacy of denoising approaches: SW-CNN consistently enhances recall (e.g., ResNet-101 improves from 92.87% to 95.8%), CDNet proves particularly effective with lightweight architectures (e.g., ShuffleNet, GoogLeNet), and SANet prioritizes precision (e.g., VGG19 achieves zero false positives).

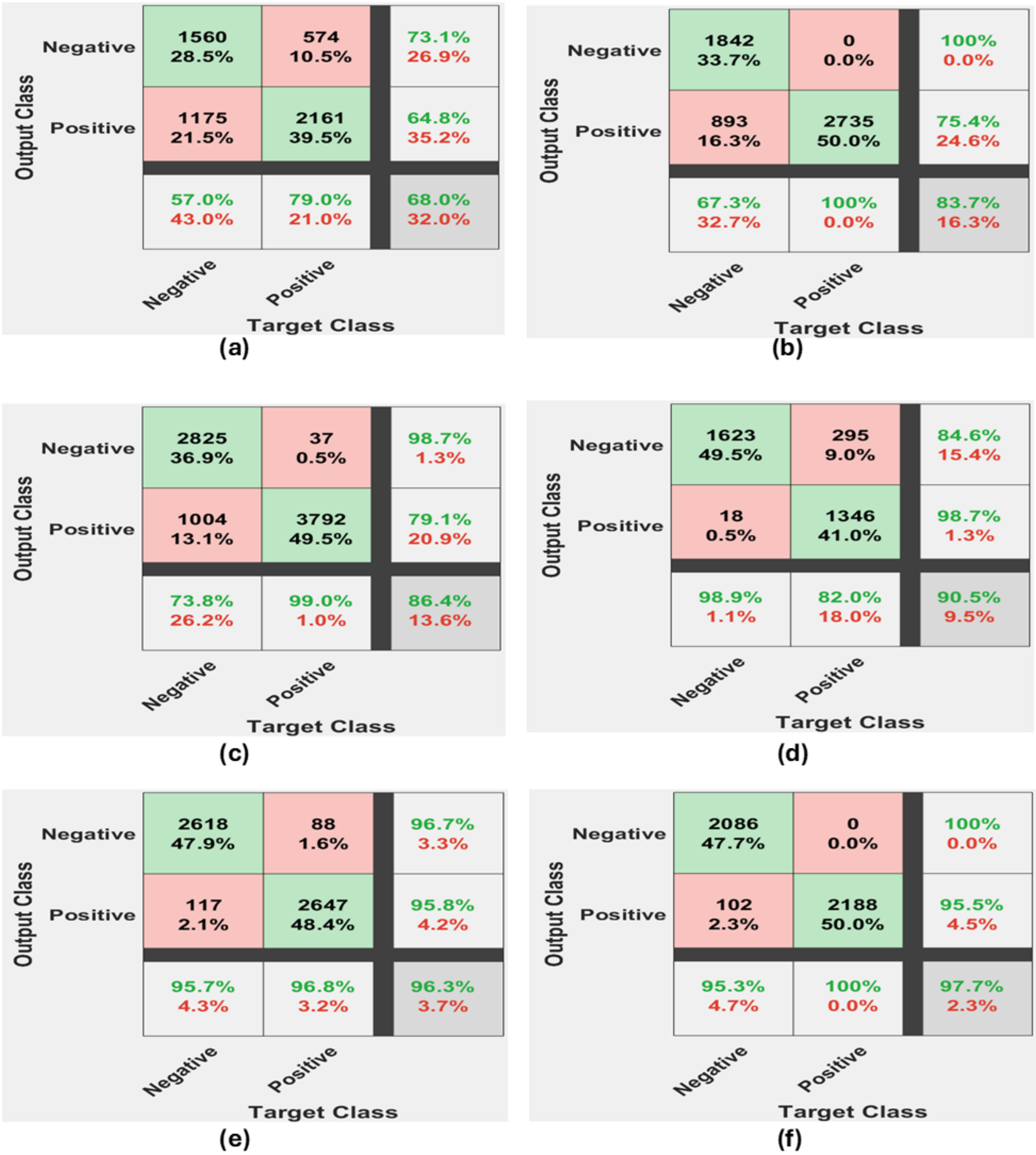

Figure 2: The confusion matrices of six pre-trained CNN models paired with their best-performing de-noising approaches, (a) AlexNet – SW-CNN, (b) VGG19- SANet, (c) GoogleNet-CDNet, (d) ShuffleNet - CDNet, (e) ResNet-101 – SW-CNN, and (f) Xception- SW-CNN.

4.2 Crack Classification Using Different Pre-Trained CNNs With Various De-Noising Techniques:

A comprehensive before-and-after analysis of six popular CNN architectures: AlexNet, VGG19, GoogLeNet, ShuffleNet, ResNet-101, and Xception is illustrated in Figure 3 showing their performance in detecting concrete surface cracks with and without the application of denoising techniques. Every CNN model shows a substantial performance increase in all three metrics (accuracy, sensitivity, and F1-score) after applying denoising. This clearly indicates that noise reduction is critical for enhancing feature extraction and classification reliability in surface crack detection tasks.

Figure 3 clearly demonstrates that applying denoising techniques before feeding images into CNNs significantly enhances classification performance for concrete crack detection tasks. All CNN architectures (regardless of depth or complexity) showed notable improvements, making a strong base for incorporating denoising into preprocessing techniques. For maximum performance, Xception or ResNet-101 with denoising are the best performers with F1-score 97.72 and 96.27% respectively and sensitivity 95.55 and 96.78% respectively.

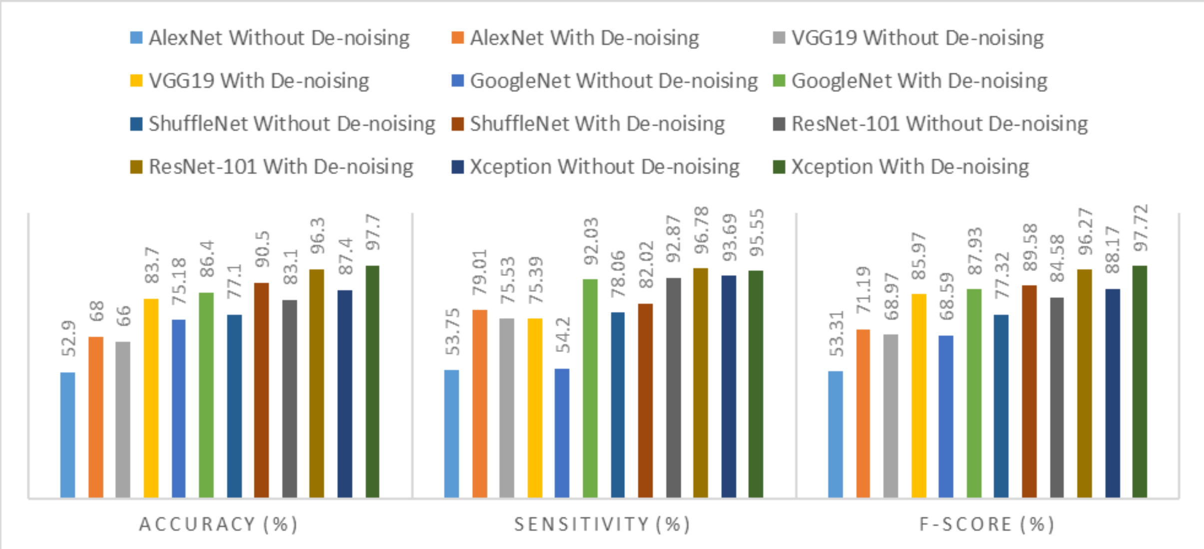

Figure 3: A Comparison between different CNN models with and without applying de-noising approach.

4.3 Efficiency Analysis:

To complement accuracy-based evaluation, a quantified computational efficiency for representative model–denoising pairings was peformed. On an NVIDIA RTX 3090 GPU, Xception+SW-CNN (server-oriented) achieved an average inference latency of 11.2 ± 0.3 ms/image, corresponding to a throughput of 89 images/s, with a model size of 228 MB. By contrast, ShuffleNet+SW-CNN (edge-oriented) achieved 3.7 ± 0.2 ms/image (270 images/s) with a compact model size of 18 MB. These results confirm that ShuffleNet is more suitable for embedded or edge deployments where efficiency is critical, while Xception maximizes accuracy for server-class environments. A summary is provided below in Table 5.

Table 5: Efficiency metrics (latency, throughput, and model size) for representative CNN–denoising pairings (ShuffleNet+SW-CNN on edge device; Xception+SW-CNN on server).

Model + Denoising	Accuracy (%)	F1-score (%)	Latency (ms/image)	Throughput (images/s)	Model Size (MB)
ShuffleNet + SW-CNN (edge-oriented)	89.3 ± 0.3	90.4 ± 0.3	3.7 ± 0.2	~270	18
Xception + SW-CNN (server-oriented)	97.7 ± 0.2	97.7 ± 0.2	11.2 ± 0.3	~89	228

5. DISCUSSION

This study explored the effectiveness of integrating various denoising convolutional neural network (CNN) approaches with pre-trained models for the classification of concrete surface crack images. The comparative analysis, as illustrated in Table 4, underscores a consistent and significant performance improvement across all evaluated CNN architectures when denoising techniques were applied prior to classification. Key performance indicators: accuracy, sensitivity, and F1-score collectively confirmed the added value of incorporating image pre-processing techniques to overcome irrelevant noise and enhance feature extraction.

5.1 Effectiveness of De-noising on Baseline CNN Models

The baseline performance without denoising revealed that traditional architectures like AlexNet and VGG19 struggled to achieve satisfactory classification scores. Specifically, AlexNet exhibited an accuracy of 52.9% and a F1-score of 53.31%, while VGG19 achieved a modest 66% accuracy. However, with the integration of denoising CNNs, AlexNet's accuracy improved to 68% and its F1-score rose sharply to 71.19%, indicating a 34% improvement in sensitivity (Table 4). This pattern was mirrored in VGG19, which achieved an accuracy of 83.7% and F1-score of 85.97% when paired with SANet, thereby validating that denoising significantly augments low-capacity models by boosting their generalization ability to identify crack patterns more effectively (Table 4).

5.2 Comparative Performance of Advanced CNN Architectures

Advanced architectures such as GoogLeNet, ShuffleNet, ResNet-101, and Xception demonstrated inherently higher baseline performance, yet all exhibited further improvements when integrated with denoising methods. For instance, ResNet-101 achieved 83.1% accuracy without denoising, which increased to 96.3% when combined with SW-CNN (Table 4). Similarly, Xception's performance peaked with an F1-score of 97.72% using SW-CNN (Table4), indicating that even state-of-the-art CNNs benefit from noise reduction techniques, particularly when distinguishing subtle surface anomalies.

5.3 Comparative Analysis of De-noising Methods

Among the denoising techniques examined, SW-CNN consistently provided the most balanced improvement across architectures. It achieved the highest F1-score in four of the six models (AlexNet, ResNet-101, Xception, and VGG19) and led to the highest overall accuracy in Xception (97.7%). While CDNet and ADNet also delivered competitive results, especially in terms of sensitivity. Their performance occasionally lagged in achieving optimal F1-scores due to potential trade-offs in precision.

HRL, on the other hand, emerged as the most recall-intensive method, often achieving the highest sensitivity scores (e.g., 99.96% with Xception, Table 4). This makes HRL suitable for use in safety-critical applications where missing a crack instance is unacceptable. However, its relatively lower F1-score and accuracy suggest potential over-classification (false positives), which must be addressed in precision-sensitive deployment scenarios.

5.4 Performance Improvement Trends

The heatmap in Figure 4 presents the F1-score (%) (a harmonic mean of precision and recall) for six CNN models (rows) paired with five denoising methods (columns) in the context of concrete crack detection. Xception consistently achieves the highest F1-scores (96.53–97.72%), demonstrating its robustness across all denoising methods. ResNet-101 and ShuffleNet also perform well, with F1-scores more than 90% for most methods. AlexNet and VGG19 exhibit the lowest F1-scores (53.19–85.97%), highlighting their limitations for cracks detection. Meanwhile, SW-CNN emerges as the most effective denoising method, significantly improving weaker models (e.g., AlexNet) and optimizing strong ones (e.g., ResNet-101, Xception). Its success may stem from adaptive noise reduction that preserves crack features. Xception's F1-scores vary minimally (±1.2%) across methods, indicating inherent robustness to noise.

In summary, the heatmap reveals that model architecture choice has a greater impact on performance than the denoising method. However, SW-CNN is the most effective denoising approach overall, particularly for state-of-the-art models like Xception and ResNet-101.

A cross-model analysis shows that the average performance improvement in F1-score due to denoising ranged from 13–15% across all models. This supports the hypothesis that image noise significantly impairs the ability of CNNs to detect fine-grained surface damage, and that denoising serves as an essential preprocessing step, particularly in real-world conditions where lighting, shadows, and environmental factors introduce unpredictable variance in input images.

Finaly, beyond accuracy, efficiency is critical for deployment. Our analysis shows that ShuffleNet+SW-CNN, with 3.7 ms/image latency and 18 MB model size, offers a lightweight edge-oriented solution. Conversely, Xception+SW-CNN, while requiring more computational resources (11.2 ms/image, 228 MB), delivers the highest accuracy and F1-scores, making it ideal for server-based monitoring. These findings highlight the trade-offs between efficiency and performance that must be considered in real-world structural health monitoring deployments.

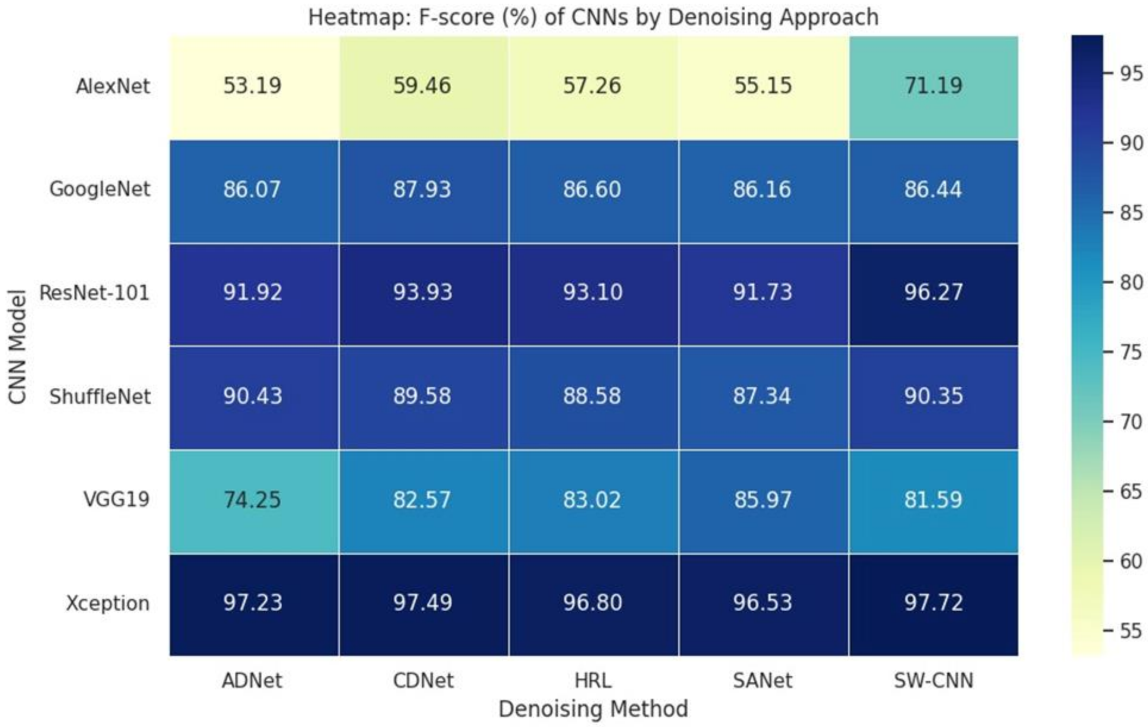

Figure 4: Heatmap: F1-Score (%) for six CNNs models paired with five de-noising approaches.

5.5 Implications for Practice and Research

From a practical standpoint, these findings are pivotal for infrastructure monitoring systems that rely on automated image-based defect detection. Lightweight models like ShuffleNet, when combined with CDNet or SW-CNN, offer a feasible solution, balancing computational efficiency and classification performance. Meanwhile, ResNet-101 and Xception, when paired with SW-CNN, provide high-accuracy solutions ideal for crack inspections requiring high precision.

From a research perspective, this work emphasizes the need for further exploration of hybrid denoising strategies, which may adaptively preserve crack-relevant textures while suppressing non-informative regions. Additionally, the integration of domain-specific augmentations, such as synthetic noise patterns mimicking site conditions, may help improve model robustness.

6. CONCLUSION

This study systematically investigated the integration of advanced denoising techniques with pre-trained convolutional neural networks (CNNs) for automated detection of cracks in concrete structures. By evaluating the performance of five denoising methods—HRL, SANet, ADNet, SW-CNN, and CDNet—across six widely used CNN architectures (AlexNet, VGG19, GoogLeNet, ShuffleNet, ResNet-101, and Xception), the research demonstrated that robust image preprocessing can substantially enhance the accuracy and reliability of crack classification.

Empirical results confirmed that denoising consistently improves performance metrics across all tested models. For example, AlexNet improved from 52.9% to 68% in accuracy and from 53.31% to 71.19% in F1-score after applying denoising. VGG19 saw its F1-score rise from 68.97% to 85.97%, and GoogLeNet achieved an F1-score increase from 68.59% to 87.93%. The highest accuracy and F1-score were obtained using the Xception model, improving to 97.7% and 97.72%, respectively, while ResNet-101 closely followed with 96.3% accuracy and

96.27% F1-score. Lightweight models such as ShuffleNet also showed significant gains, increasing from 77.32% to 89.58% in F1-score, making them ideal for resource-constrained edge deployments.

Among the denoising techniques, SW-CNN delivered the most consistent improvements, achieving the highest F1-score in four out of six models and providing the best overall performance. CDNet and ADNet were also effective, particularly in improving sensitivity, demonstrating their strength in detecting true positive cases. On average, denoising techniques led to 13–15% improvements in F1-score, highlighting their value in enhancing deep learning model performance for structural defect detection.

A key contribution of this work lies in empirically validating that denoising not only improves model accuracy but also enhances sensitivity and precision, offering a practical preprocessing solution for noisy visual data in structural health monitoring. This supports informed selection of denoising methods based on task priorities—whether accuracy, computational efficiency, or deployment environment. From a practical perspective, the findings provide a roadmap for deploying automated inspection systems in real-world settings. Pairings such as ShuffleNet with SW-CNN balance performance and efficiency for edge applications, while Xception with SW-CNN offers a high-precision option for server-based systems where accuracy is paramount.

This study has several limitations that should be acknowledged. First, the dataset used was collected under relatively controlled capture conditions on a single campus, with consistent illumination and viewpoints. While this helped minimize variability, it may also restrict the diversity of environmental conditions such as lighting, weather, and camera perspectives that occur in real-world inspections. Second, reliance on a single benchmark dataset poses risks of overfitting and limited generalizability to other concrete structures, materials, or geographical contexts. Third, no data augmentation (e.g., rotations, flips, or brightness shifts) was applied, which could have improved robustness against natural variability in crack orientation and surface textures. Fourth, although performance was reported as mean ± standard deviation over repeated runs, no formal statistical significance testing was conducted between methods, which limits the ability to confirm whether observed improvements are statistically robust. Finally, while this paper makes use of a widely recognized benchmark dataset (Sorguç, 2018), code and data used in this study will be made available upon reasonable request from the corresponding author to facilitate transparency and reproducibilityFuture research should explore more diverse and realistic noise scenarios, evaluate generalisability across other material types such as metals and masonry, and optimise models for energy efficiency and real-time deployment. Additionally, integrating these enhanced models within Industry 4.0 frameworks, such as Digital Twins, could further strengthen the robustness and responsiveness of automated structural health monitoring systems.

REFERENCES

Ali, L., Harous, S., Zaki, N., Khan, W., Alnajjar, F., & Jassmi, H. (2021). Performance Evaluation of different Algorithms for Crack Detection in Concrete Structures. https://doi.org/10.1109/ICCAKM50778.2021.9357717

Alipour, M., Harris, D. K., & Miller, G. R. (2019). Robust Pixel-Level Crack Detection Using Deep Fully Convolutional Neural Networks. Journal of Computing in Civil Engineering, 33(6), 04019040. https://doi.org/doi:10.1061/(ASCE)CP.1943-5487.0000854

Cha, Y.-J., Choi, W., & Büyüköztürk, O. (2017). Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Computer-Aided Civil and Infrastructure Engineering, 32(5), 361-378. https://doi.org/https://doi.org/10.1111/mice.12263

Chan, B., Jo, J., & Blumenstein, M. (2015). Towards UAV-based bridge inspection systems: A review and an application perspective. Structural Monitoring and Maintenance, 2, 283-300. https://doi.org/10.12989/smm.2015.2.3.283

Chen, F. C., & Jahanshahi, M. R. (2018). NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion. IEEE Transactions on Industrial Electronics, 65(5), 4392-4400. https://doi.org/10.1109/TIE.2017.2764844

Chen, P. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. Nat Mater, 18(5), 410-414. https://doi.org/10.1038/s41563-019-0345-0

Chen, S., Laefer, D. F., Mangina, E., Zolanvari, S. M. I., & Byrne, J. (2019). UAV Bridge Inspection through Evaluated 3D Reconstructions. Journal of Bridge Engineering, 24(4), 05019001. https://doi.org/doi:10.1061/(ASCE)BE.1943-5592.0001343

Chen, Y., & Pock, T. (2017). Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1256-1272. https://doi.org/10.1109/TPAMI.2016.2596743

Chiang, Y. W., & Sullivan, B. J. (1989, 14-16 Aug. 1989). Multi-frame image restoration using a neural network. Proceedings of the 32nd Midwest Symposium on Circuits and Systems,

Chow, J. K., Liu, K.-f., Tan, P. S., Su, Z., Wu, J., Li, Z., & Wang, Y.-H. (2021). Automated defect inspection of concrete structures. Automation in Construction, 132, 103959. https://doi.org/https://doi.org/10.1016/j.autcon.2021.103959

Cruz, C., Foi, A., Katkovnik, V., & Egiazarian, K. (2018). Nonlocality-Reinforced Convolutional Neural Networks for Image Denoising. IEEE Signal Processing Letters, 25(8), 1216-1220. https://doi.org/10.1109/LSP.2018.2850222

Diwakar, M., & Kumar, M. (2018). A review on CT image noise and its denoising. Biomedical Signal Processing and Control, 42, 73-88. https://doi.org/https://doi.org/10.1016/j.bspc.2018.01.010

Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Construction and Building Materials, 186, 1031-1045. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2018.08.011

Dow, H., Perry, M., McAlorum, J., Pennada, S., & Dobie, G. (2023). Skeleton-based noise removal algorithm for binary concrete crack image segmentation. Automation in Construction, 151, 104867. https://doi.org/https://doi.org/10.1016/j.autcon.2023.104867

Dung, C., & Le Duc, A. (2018). Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction, 99, 52-58. https://doi.org/10.1016/j.autcon.2018.11.028

Fan, G., Li, J., & Hao, H. (2020). Vibration Signal Denoising for Structural Health Monitoring by Residual Convolutional Neural Networks. Measurement, 157, 107651. https://doi.org/10.1016/j.measurement.2020.107651

Feng, C., Zhang, H., Wang, H., Wang, S., & Li, Y. (2020). Automatic Pixel-Level Crack Detection on Dam Surface Using Deep Convolutional Network. Sensors, 20(7), 2069. https://www.mdpi.com/1424-8220/20/7/2069

Flah, M., Suleiman, A. R., & Nehdi, M. L. (2020). Classification and quantification of cracks in concrete structures using deep learning image-based techniques. Cement and Concrete Composites, 114, 103781. https://doi.org/https://doi.org/10.1016/j.cemconcomp.2020.103781

Fociro, O., Fociro, A., Muci, R., Skrame, K., Pekmezi, J., & Mezini, M. (2023). Carbonate texture identification using multi-layer perceptron neural network. Open Geosciences, 15, 20220453. https://doi.org/10.1515/geo-2022-0453

Fu, R., Xu, H., Wang, Z., Shen, L., Cao, M., Liu, T., & Novák, D. (2020). Enhanced Intelligent Identification of Concrete Cracks Using Multi-Layered Image Preprocessing-Aided Convolutional Neural Networks. Sensors, 20(7), 2021. https://www.mdpi.com/1424-8220/20/7/2021

GAN, J., Yu, N., Qian, G., & He, N. (2023). Concrete learning method for segmentation and denoising using CBCT Image Proceedings of the 2023 4th International Conference on Control, Robotics and Intelligent System, Guangzhou, China. https://doi.org/10.1145/3622896.3622903

Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media. https://books.google.jo/books?id=V5ySEAAAQBAJ

Giannatou, E., Papavieros, G., Constantoudis, V., Papageorgiou, H., & Gogolides, E. (2019). Deep learning denoising of SEM images towards noise-reduced LER measurements. Microelectronic Engineering, 216, 111051. https://doi.org/https://doi.org/10.1016/j.mee.2019.111051

Goodfellow, I. (2016). Deep learning. In: MIT press.

Guo, B., Song, K.-C., Dong, H., Yan, Y., Zhibiao, T., & Zhu, L. (2020). NERNet: Noise Estimation and Removal Network for Image Denoising. Journal of Visual Communication and Image Representation, 71, 102851. https://doi.org/10.1016/j.jvcir.2020.102851

Ha, I., Kim, H., Park, S., & Kim, H. (2018). Image retrieval using BIM and features from pretrained VGG network for indoor localization. Building and Environment, 140, 23-31. https://doi.org/https://doi.org/10.1016/j.buildenv.2018.05.026

Hong, I., Hwang, Y., & Kim, D. (2019). Efficient deep learning of image denoising using patch complexity local divide and deep conquer. Pattern Recognition, 96, 106945. https://doi.org/https://doi.org/10.1016/j.patcog.2019.06.011

Howden-Chapman, P., Bennett, J., Edwards, R., Jacobs, D., Nathan, K., & Ormandy, D. (2023). Review of the Impact of Housing Quality on Inequalities in Health and Well-Being. Annual Review of Public Health, 44(Volume 44, 2023), 233-254. https://doi.org/https://doi.org/10.1146/annurev-publhealth-071521-111836

Hsieh, Y.-A., & Tsai, Y. (2020). Machine Learning for Crack Detection: Review and Model Performance Comparison. Journal of Computing in Civil Engineering, 34, 04020038. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000918

Hu, J., Wang, X., Shao, F., & Jiang, Q. (2020). TSPR: Deep network-based blind image quality assessment using two-side pseudo reference images. Digital Signal Processing, 106, 102849. https://doi.org/https://doi.org/10.1016/j.dsp.2020.102849

Huyan, J., Li, W., Tighe, S., Xu, Z., & Zhai, J. (2020). CrackU-net: A novel deep convolutional neural network for pixelwise pavement crack detection. Structural Control and Health Monitoring, 27, e2551. https://doi.org/10.1002/stc.2551

Ieracitano, C., Paviglianiti, A., Campolo, M., Hussain, A., Pasero, E., & Morabito, F. C. (2021). A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers. IEEE/CAA Journal of Automatica Sinica, 8(1), 64-76. https://doi.org/10.1109/JAS.2020.1003387

Islam, M. T., Mahbubur Rahman, S. M., Omair Ahmad, M., & Swamy, M. N. S. (2018). Mixed Gaussian-impulse noise reduction from images using convolutional neural network. Signal Processing: Image Communication, 68, 26-41. https://doi.org/https://doi.org/10.1016/j.image.2018.06.016

Jain, V., & Seung, H. (2008). Natural Image Denoising with Convolutional Networks.

Jang, K., Kim, N., & An, Y.-K. (2019). Deep learning–based autonomous concrete crack evaluation through hybrid image scanning. Structural Health Monitoring, 18, 1722-1737. https://doi.org/10.1177/1475921718821719

Jarrett, K., Kavukcuoglu, K., Ranzato, M., & LeCun, Y. (2009, 29 Sept.-2 Oct. 2009). What is the best multi-stage architecture for object recognition? 2009 IEEE 12th International Conference on Computer Vision,

Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y.-J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. Automation in Construction, 118, 103291. https://doi.org/https://doi.org/10.1016/j.autcon.2020.103291

Kim, J., Lee, J. K., & Lee, K. M. (2016, 27-30 June 2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Commun. ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., & Steinbrecher, M. (2022). Multi-layer Perceptrons. In R. Kruse, S. Mostaghim, C. Borgelt, C. Braune, & M. Steinbrecher (Eds.), Computational Intelligence: A Methodological Introduction (pp. 53-124). Springer International Publishing. https://doi.org/10.1007/978-3-030-42227-1_5

Li, S., & Zhao, X. (2019). Image-Based Concrete Crack Detection Using Convolutional Neural Network and Exhaustive Search Technique. Advances in Civil Engineering, 2019, 1-12. https://doi.org/10.1155/2019/6520620

Li, S., & Zhao, X. (2020). Automatic Crack Detection and Measurement of Concrete Structure Using Convolutional Encoder-Decoder Network. IEEE Access, 8, 134602-134618. https://doi.org/10.1109/ACCESS.2020.3011106

Li, X., Xiao, J., Zhou, Y., Ye, Y., Lv, N., Wang, X., Wang, S., & Gao, S. (2020). Detail retaining convolutional neural network for image denoising. Journal of Visual Communication and Image Representation, 71, 102774. https://doi.org/https://doi.org/10.1016/j.jvcir.2020.102774

Liu, Y., Yao, J., Lu, X., Xie, R., & Li, L. (2019). DeepCrack: A Deep Hierarchical Feature Learning Architecture for Crack Segmentation. Neurocomputing, 338, 139-153. https://doi.org/10.1016/j.neucom.2019.01.036

Lyu, Q., Guo, M., & Pei, Z. (2020). DeGAN: Mixed noise removal via generative adversarial networks. Applied Soft Computing, 95, 106478. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106478

Marreiros, A. C., Daunizeau, J., Kiebel, S. J., & Friston, K. J. (2008). Population dynamics: Variance and the sigmoid activation function. NeuroImage, 42(1), 147-157. https://doi.org/https://doi.org/10.1016/j.neuroimage.2008.04.239

Nah, S., Kim, T. H., & Lee, K. M. (2017, 21-26 July 2017). Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

Ni, F., Zhang, J., & Chen, Z. (2019). Pixel-level crack delineation in images with convolutional feature fusion. Structural Control and Health Monitoring, 26(1), e2286. https://doi.org/https://doi.org/10.1002/stc.2286

Owusu-Manu, D.-G., Quaigrain, R. A., Edwards, D. J., Hammond, M., Hammond, M., & Roberts, C. (2022). Energy conservation literacy among households in Sub-Sahara Africa. International Journal of Energy Sector Management, 16(6), 1130-1149. https://doi.org/10.1108/IJESM-09-2021-0010

Qi, H., Tan, S., & Li, Z. (2022). Anisotropic Weighted Total Variation Feature Fusion Network for Remote Sensing Image Denoising. Remote Sensing, 14(24), 6300. https://www.mdpi.com/2072-4292/14/24/6300

Quan, Y., Chen, Y., Shao, Y., Teng, H., Xu, Y., & Ji, H. (2021). Image denoising using complex-valued deep CNN. Pattern Recognition, 111, 107639. https://doi.org/https://doi.org/10.1016/j.patcog.2020.107639

Radenović, F., Tolias, G., & Chum, O. (2016, 2016//). CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples. Computer Vision – ECCV 2016, Cham.

Ren, Y., Huang, J., Hong, Z., Lu, W., Yin, J., Zou, L., & Shen, X. (2020). Image-based concrete crack detection in tunnels using deep fully convolutional networks. Construction and Building Materials, 234, 117367. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2019.117367

S. H, K., Park, N., & Lee, B. (2023). EFID: Edge-Focused Image Denoising Using a Convolutional Neural Network. IEEE Access, 11, 9613-9626. https://doi.org/10.1109/ACCESS.2023.3239835

Sadrizadeh, S., Otroshi-Shahreza, H., & Marvasti, F. (2022). Impulsive noise removal via a blind CNN enhanced by an iterative post-processing. Signal Processing, 192, 108378. https://doi.org/https://doi.org/10.1016/j.sigpro.2021.108378

Shi, W., Jiang, F., Zhang, S., Wang, R., Zhao, D., & Zhou, H. (2019). Hierarchical residual learning for image denoising. Signal Processing: Image Communication, 76, 243-251. https://doi.org/https://doi.org/10.1016/j.image.2019.05.007

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging, 35(5), 1285-1298. https://doi.org/10.1109/TMI.2016.2528162

Simard, P., Steinkraus, D., & Platt, J. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. https://doi.org/10.1109/ICDAR.2003.1227801

Solovyeva, E., & Abdullah, A. (2022). Dual Autoencoder Network with Separable Convolutional Layers for Denoising and Deblurring Images. Journal of Imaging, 8(9), 250. https://www.mdpi.com/2313-433X/8/9/250

Song, Q., Wu, Y., Xin, X., Yang, L., Yang, M., Chen, H., Liu, C., Hu, M., Xuesong, C., & Li, J. (2019). Real-Time Tunnel Crack Analysis System via Deep Learning. IEEE Access, PP, 1-1. https://doi.org/10.1109/ACCESS.2019.2916330

Song, W., Jia, G., Jia, D., & Zhu, H. (2019). Automatic Pavement Crack Detection and Classification Using Multiscale Feature Attention Network. IEEE Access, PP, 1-1. https://doi.org/10.1109/ACCESS.2019.2956191

Sorguç, A. G. (2018, 2018/07/22). Performance Comparison of Pretrained Convolutional Neural Networks on Crack Detection in Buildings Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC),

Tang, P., Wang, H., & Kwong, S. (2017). G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition. Neurocomputing, 225, 188-197. https://doi.org/https://doi.org/10.1016/j.neucom.2016.11.023

Tang, X., Lei, Z., & and Ding, X. (2019). SAR image despeckling with a multilayer perceptron neural network. International Journal of Digital Earth, 12(3), 354-374. https://doi.org/10.1080/17538947.2018.1447032

Tian, C., Xu, Y., Li, Z., Zuo, W., Fei, L., & Liu, H. (2020). Attention-guided CNN for image denoising. Neural Networks, 124, 117-129. https://doi.org/https://doi.org/10.1016/j.neunet.2019.12.024

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders Proceedings of the 25th international conference on Machine learning, Helsinki, Finland. https://doi.org/10.1145/1390156.1390294

Xie, J., Xu, L., & Chen, E. (2012). Image Denoising and Inpainting with Deep Neural Networks. Advances in Neural Information Processing Systems, 1.

Xie, S., Song, J., Hu, Y., Zhang, C., & Zhang, S. (2023). Using CNN with Multi-Level Information Fusion for Image Denoising. Electronics, 12(9), 2146. https://www.mdpi.com/2079-9292/12/9/2146

Xu, S., Zhang, C., & Zhang, J. (2020). Bayesian deep matrix factorization network for multiple images denoising. Neural Networks, 123, 420-428. https://doi.org/https://doi.org/10.1016/j.neunet.2019.12.023

Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. Structural Control and Health Monitoring, 26, e2313. https://doi.org/10.1002/stc.2313

Xu, Z., Wang, Y., Hao, X., & Fan, J. (2023). Crack Detection of Bridge Concrete Components Based on Large-Scene Images Using an Unmanned Aerial Vehicle. Sensors, 23(14), 6271. https://www.mdpi.com/1424-8220/23/14/6271

Yan, Z., Zhang, H., Piramuthu, R., Jagadeesh, V., DeCoste, D., Di, W., & Yu, Y. (2015, 7-13 Dec. 2015). HD-CNN: Hierarchical Deep Convolutional Neural Networks for Large Scale Visual Recognition. 2015 IEEE International Conference on Computer Vision (ICCV),

Yin, H., Gong, Y., & Qiu, G. (2020). Fast and efficient implementation of image filtering using a side window convolutional neural network. Signal Processing, 176, 107717. https://doi.org/https://doi.org/10.1016/j.sigpro.2020.107717

Zhang, J., Luo, H., Hui, B., Chang, Z., & Zhang, X. (2019). Unknown noise removal via sparse representation model. ISA Transactions, 94, 135-143. https://doi.org/https://doi.org/10.1016/j.isatra.2019.03.028

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Transactions on Image Processing, 26(7), 3142-3155. https://doi.org/10.1109/TIP.2017.2662206

Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising. IEEE Transactions on Image Processing, 27(9), 4608-4622. https://doi.org/10.1109/TIP.2018.2839891

Zhang, Q., Yuan, Q., Li, J., Sun, F., & Zhang, L. (2020). Deep spatio-spectral Bayesian posterior for hyperspectral image non-i.i.d. noise removal. ISPRS Journal of Photogrammetry and Remote Sensing, 164, 125-137. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.04.010

Zhang, Y., Sun, X., Loh, K. J., Su, W., Xue, Z., & Zhao, X. (2020). Autonomous bolt loosening detection using deep learning. Structural Health Monitoring, 19(1), 105-122. https://doi.org/10.1177/1475921719837509

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115. https://doi.org/10.1016/j.ymssp.2018.05.050

Zhou, S., & Song, W. (2021). Deep learning–based roadway crack classification with heterogeneous image data fusion. Structural Health Monitoring, 20(3), 1274-1293. https://doi.org/10.1177/1475921720948434

Zhu, J., & Song, J. (2020). An Intelligent Classification Model for Surface Defects on Cement Concrete Bridges. Applied Sciences, 10(3), 972. https://www.mdpi.com/2076-3417/10/3/972

Zou, Q., Zhang, Z., Li, Q., Qi, X., & Wang, Q. (2018). DeepCrack: Learning Hierarchical Convolutional Features for Crack Detection. IEEE Transactions on Image Processing, 28, 1498-1512. https://doi.org/10.1109/TIP.2018.2878966