

Article

Deep Learning-Based Automated Detection of Cracks in Historical Masonry Structures

Kemal Hacıfendioğlu ¹ , Ahmet Can Altunışık ^{1,2,3,*} and Tuğba Abdioğlu ¹

¹ Department of Civil Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey; kemalhaciefendioğlu@ktu.edu.tr (K.H.); abdioglutugba@gmail.com (T.A.)

² Earthquake and Structural Health Monitoring Research Center, Karadeniz Technical University, 61080 Trabzon, Turkey

³ Dynamic Academy Software, Construction Ind. Trade. Co., Ltd. Şti., 61081 Trabzon, Turkey

* Correspondence: ahmetcan@ktu.edu.tr

Abstract: The efficient and precise identification of cracks in masonry stone structures caused by natural or human-induced factors within a specific region holds significant importance in detecting damage and subsequent secondary harm. In recent times, remote sensing technologies have been actively employed to promptly identify crack regions during repair and reinforcement activities. Enhanced image resolution has enabled more accurate and sensitive detection of these areas. This research presents a novel approach utilizing deep learning techniques for crack area detection in cellphone images, achieved through segmentation and object detection methods. The developed model, named the CAM-K-SEG segmentation model, combines Grad-CAM visualization and K-Mean clustering approaches with pre-trained convolutional neural network models. A comprehensive dataset comprising photographs of numerous historical buildings was utilized for training the model. To establish a comparative analysis, the widely used U-Net segmentation model was employed. The training and testing datasets for the developed technique were meticulously annotated and masked. The evaluation of the results was based on the Intersection-over-Union (IoU) metric values. Consequently, it was concluded that the CAM-K-SEG model exhibits suitability for object recognition and localization, whereas the U-Net model is well-suited for crack area segmentation.

Keywords: deep learning; convolutional neural network; crack; Grad-CAM; U-Net; K-Mean clustering



Citation: Hacıfendioğlu, K.; Altunışık, A.C.; Abdioğlu, T. Deep Learning-Based Automated Detection of Cracks in Historical Masonry Structures. *Buildings* **2023**, *13*, 3113. <https://doi.org/10.3390/buildings13123113>

Received: 12 November 2023

Revised: 5 December 2023

Accepted: 12 December 2023

Published: 15 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The majority of masonry structures are categorized as “Registered Structures” and represent a significant part of our “Cultural Heritage”, thereby necessitating their preservation and appreciation [1]. Unfortunately, numerous historical buildings in our nation are in a state of disrepair. Earthquakes, ground-related issues, fires, and degradation resulting from environmental factors have notably impacted both the structural integrity and aesthetics of these constructions. The occurrence of cracks in many historical buildings is often attributable to irregularities and disconnections within their foundations, potentially leading to partial or complete collapse. Rather than resorting to “demolition and reconstruction” approaches, there is a pressing requirement for “repair and maintenance” strategies when dealing with historical buildings. Presently, conventional techniques such as visual inspections and manual surveys are employed to assess the existing masonry structures for signs of damage, including cracking and fragmentation. However, these methods are labor-intensive and prone to individual errors [2,3]. Previous attempts to repair cracks and the structural systems of these buildings have proven ineffective due to a lack of investigation into the underlying causes of such damages. Prior to strengthening a damaged historical building, a comprehensive evaluation of its ground properties, load-bearing system, utilized materials, and historical background is imperative. These investigations should be conducted meticulously to determine the appropriate interventions

and strengthening measures. Neglecting the essential investigative procedures may result in unintended damage during preservation efforts. Hence, a conscious effort must be made to safeguard and fortify historical structures before it becomes too late.

The visual inspection process is commonly employed to examine and assess the present condition of historical buildings. However, this method entails challenges in terms of its laborious and time-consuming nature, necessitating inspectors with expertise and specialized knowledge to evaluate structural conditions based on visual observations. Furthermore, the frequency of inspections is limited due to the associated high labor costs, proneness to human error, and difficulties in accessing inspection sites. Recognizing the constraints inherent in manual inspections, the use of computer vision-based assessments has become increasingly prominent in the field of engineering [4–8]. Notably, the application of computer vision techniques for crack detection has garnered substantial attention from researchers. Image-based crack detection emerges as an exemplary non-destructive evaluation approach, particularly valuable for historic structures subject to stringent regulations that limit even minor interventions mandated by conservation authorities.

In recent times, deep learning (DL), a branch of artificial intelligence focusing on leveraging image-based inspection techniques, has emerged as a powerful method for analyzing structural damage. One prominent tool within DL is the convolutional neural network (CNN), which proves effective in this context. Unlike traditional machine learning approaches, DL possesses the capability to autonomously learn features and automatically detect damages without the need for manual feature definition. In the crack detection phase, various photos serve as input, and the network outputs the identified cracks without any manual intervention [9–11]. Particularly, studies conducted on classifying damage types in historical monuments and structures, as well as determining the affected areas, have significantly elevated the relevance of deep learning methodologies in the present day [12–17].

Segmentation methods represent an advanced approach for examining cracks in structures. Within this domain, classic networks, convolutional neural networks (CNN), and fully connected networks (FCN) are noteworthy examples of deep learning-based segmentation methods. To achieve satisfactory accuracy in their results, CNN and FCN architectures necessitate a well-trained, extensively labeled, or annotated dataset [11,18–25]. Nonetheless, these architectures exhibit a high potential for detecting intricate features. Another technique called DL-based transfer learning can be employed, entailing the pre-training of a CNN model using a larger dataset, subsequently enhancing performance when training with smaller datasets. Transfer learning is applicable to any CNN and FCN (fully convolutional networks) architecture [26]. Deep learning-based automated crack detection systems have been developed for masonry structures [27,28]. These systems utilize deep convolutional neural networks and image-processing techniques to detect cracks on concrete and masonry surfaces [29,30]. The models are trained on labeled and augmented image datasets, achieving high accuracy in crack classification. The proposed methods have been benchmarked on historical masonry structures, such as the Khaju Bridge in Iran, and have demonstrated accurate identification of both major and minor crack patterns. These automated crack detection systems provide a key step towards automated damage inspection and health evaluation for infrastructure, enabling reliable retrofitting interventions and conservation strategies. Cardellicchio et al. [31] focused on combining visual inspections with automated recognition of bridge defects using deep learning techniques. Their aim is to develop non-invasive investigation protocols for effective defect control and interpretation of results for bridge inspections. Cardellicchio et al. [32] studied using YOLOv5, a single-stage object detection model, for automatic defect detection on existing bridges. A database of typical defects was gathered and labeled by domain experts, and YOLOv5 was trained, tested, and validated. The results showed good effectiveness and the accuracy of the proposed methodology, demonstrating the potential of artificial intelligence for automatic defect detection on bridges.

Recovering compressed images for automatic crack segmentation can be achieved using generative models. Generative adversarial networks (GANs) have been used in several papers to address this problem. One approach replaces the sparsity regularization with a generative model that effectively captures a low-dimensional representation of targeted images [33]. Another paper proposes a road crack segmentation method based on GANs, where the generator generates fake crack images similar to real ones, and the discriminator distinguishes between real and fake crack images [34]. Furthermore, a GAN-based neural network called CrackSegAN is proposed for automatic pavement crack segmentation, achieving high F1 scores on different datasets [35]. Finally, an automatic concrete infrastructure crack semantic segmentation method is proposed using deep learning, which shows optimal performance for crack image segmentation [36].

U-Net represents a deep fully convolutional network (FCN) specifically designed for biomedical image segmentation and exhibits favorable performance when applied to datasets with limited width [37]. Leveraging its proficiency in detecting fine edges, U-Net has emerged as a benchmark for image segmentation across various domains and has found extensive utility in structural examinations [38]. Notably, U-Net has been successfully employed for segmenting brick or stone elements within masonry structures. The U-Net architecture has shown exceptional capabilities in segmenting bricks and detecting defects in masonry structures, showcasing its effectiveness in these tasks. Moreover, the U-Net architecture proves to be a suitable choice for automatic documentation of masonry structures through digital images and structural inspections, fulfilling the necessary algorithmic requirements [26,39].

In this research, a novel crack detection system is proposed for the segmentation of cracks found in historical masonry structures. The system utilizes a pre-trained convolutional neural network (CNN) model for deep learning. The model is developed by integrating Grad-CAM visualization [40] and K-Mean clustering methods [41] with pre-trained CNN models, resulting in the CAM-K-SEG segmentation model. A dataset comprising photographs captured from numerous historical buildings is utilized to train the model. The performance of the proposed method was gauged through a comparative analysis with the U-Net segmentation model.

2. Methodology

Figure 1 depicts the detailed structure of the proposed CAM-K-SEG model [42], which combines the class activation map (CAM) method and U-Net model for crack segmentation. Initially, transfer learning was employed to perform classification analysis on image data categorized as cracked or non-cracked using pre-trained models. The CAM technique was utilized to identify the crack class within the dataset and highlight distinctive image regions. This enabled a deeper understanding of the areas crucial for image classification with the employed deep learning model. Several pre-trained CNN models, including VGG-16 and VGG-19 [43], Inception-V3 [44], Xception [45], and ResNet50 [46], alongside a custom CNN model, were utilized in this study, and their performance metrics were compared. The best-performing model was selected for subsequent analyses. Using this model, a deep learning-based segmentation model, referred to as the CAM-K-SEG model, was developed to detect crack areas in historical masonry images. The model was obtained by combining the Grad-CAM visualization and K-Mean clustering methods with pre-trained CNN models. For the segmented crack areas, bounding boxes were generated and compared with ground truth bounding boxes, and the IoU (Intersection-over-Union) metric values were computed. Additionally, the U-Net method was employed to detect cracks in the walls of historical masonry buildings. This approach involved manual labeling for crack area detection and semantic segmentation tasks on crack damage images. The dataset was partitioned into sets for training and testing, ensuring that the test images were not employed during the training phase. After training, predictions were compared with ground truth masks, and IoU values were obtained. The obtained IoU performance metrics between U-Net and CAM-K-SEG were subsequently compared.

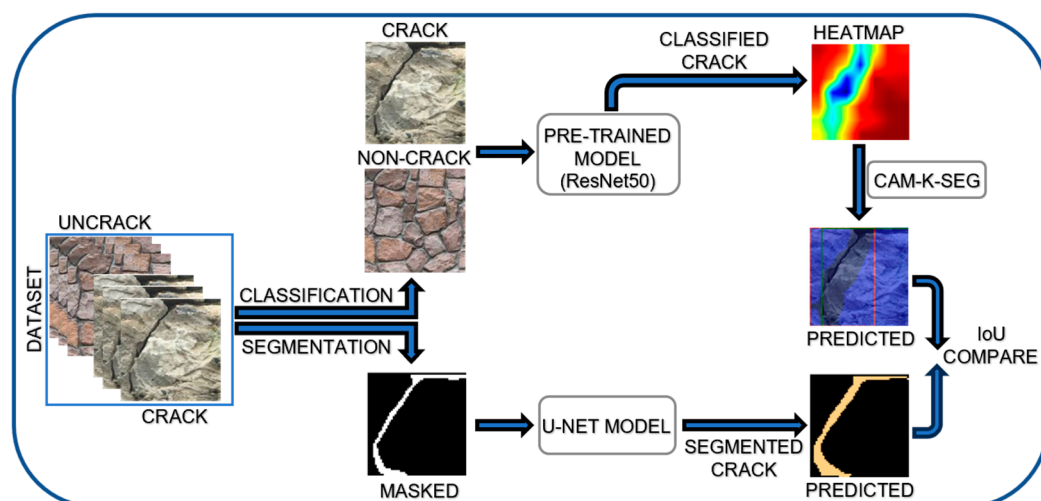


Figure 1. Methodology for the proposed and U-Net methods.

In addressing the methodologies employed in this study, it is pertinent to elucidate the distinct yet complementary roles of the CAM-K-SEG and U-Net models. While these models inherently serve different functions, their integration within the research framework is strategically aligned with the overarching objective of enhancing crack detection in masonry structure walls. The U-Net model, renowned for its acute sensitivity and precision in segmenting crack areas, excels in providing detailed pixel-level analysis. This granular approach is indispensable for a thorough assessment of crack severity and extent. Conversely, the CAM-K-SEG model, a novel development in the study, is designed to identify the presence and locations of cracks. Utilizing a rectangular area methodology to compute Intersection-over-Union (IoU) metrics, the CAM-K-SEG model adeptly pinpoints potential areas of structural concern, offering a rapid, macro-level perspective. This dual-model approach, encompassing both the U-Net and CAM-K-SEG models, thus furnishes a comprehensive toolkit for crack detection. It harmonizes the detailed segmentation capabilities of the U-Net model with the broader, location-focused analysis of the CAM-K-SEG model, thereby ensuring a holistic assessment of structural integrity in historical masonry structures.

This study's primary theoretical contribution lies in the development and validation of the CAM-K-SEG model, a novel approach that integrates class activation mapping with deep learning for the precise detection of cracks in historical masonry structures. This integration represents a significant advancement in the application of deep learning techniques to structural health monitoring, particularly in the nuanced field of heritage conservation.

Methodologically, the study distinguishes itself in several ways. Innovative Model Architecture: The CAM-K-SEG model's architecture, underpinned by the ResNet50 framework, is specifically tailored to enhance crack detection accuracy and efficiency, demonstrating a novel application in the realm of structural analysis.

Segmentation and Localization: A pivotal aspect of our methodological contribution is the innovative integration of K-Mean clustering with classification techniques using pre-trained models. This novel approach represents a significant methodological advancement in the field of structural analysis, specifically tailored for the identification and delineation of crack areas in complex masonry textures.

1. K-Mean Clustering for Enhanced Segmentation: The use of K-Mean clustering in our model facilitates the effective segmentation of crack areas. This technique allows for the precise grouping of similar pixels, aiding in the accurate identification of crack patterns amidst the diverse textures of masonry structures.
2. Classification with Pre-trained Models: The application of classification methodologies using pre-trained models, such as ResNet50, offers a robust framework for the localization and classification of identified crack areas. This integration enhances the

model's ability to accurately detect and categorize cracks, leveraging the strengths of advanced deep learning architectures.

Dataset Curation and Processing: The creation of a specialized dataset, which encompasses a diverse range of crack patterns under varying environmental conditions, and its meticulous preprocessing and annotation represent a methodological advancement in dataset preparation for structural analysis tasks.

Performance Benchmarking: The comparative analysis of the model against established models like U-Net in terms of segmentation capabilities provides a novel perspective on the efficacy of different deep learning approaches in the context of crack detection.

Practical Application and Scalability: The approach demonstrates practicality and scalability, emphasizing less labor-intensive methods and potential adaptability to various types of masonry and environmental conditions, thus broadening the scope of application in structural health monitoring.

2.1. Dataset

Dataset Description: This study employed a specialized dataset titled "High-Resolution Crack Imagery for Structural Analysis of Historical Masonry in Trabzon Province", meticulously assembled to support the detection and analysis of crack patterns on historical masonry walls. This section elaborates on the dataset's source, features, processing, and labeling.

Source and Composition of the Dataset: A collection of 502 high-resolution images was curated from various historical masonry buildings across Trabzon Province, Turkey, capturing a diverse array of crack instances. These images incorporate variability in seasonal weather and lighting conditions to ensure robustness and generalizability. Each photograph was taken using a 12-megapixel smartphone camera, providing the necessary detail for subsequent analysis.

Features and Processing: The dataset showcases a vast spectrum of crack patterns, essential for the comprehensive structural analysis aimed at preserving historical structures. To maintain consistency across the dataset, original images of varying sizes were resized to a uniform dimension of 128×128 pixels.

Labeling and Annotation: A meticulous manual labeling process was undertaken using Paint.net version 4.2.16, where each pixel was categorized as crack or non-crack, yielding a fully annotated set of images. For segmentation purposes, U-Net methodology was employed, with crack areas masked to generate a corresponding set of labeled images. Data augmentation techniques, including rotation, closure, and translation, were applied to enrich the dataset and bolster the model's predictive localization performance.

Ground Truth and Partitioning: For the purpose of crack damage detection, LabelImg version 1.8.3 was utilized to mark the images with ground truth bounding boxes, facilitating the identification of crack regions. The dataset was methodically divided into training, validation, and testing sets in a 7:2:1 ratio, providing a structured framework for training and evaluating the pre-trained models.

Evaluation Metrics: To assess the model's performance, a subset of 10 images was selected for in-depth analysis. The Intersection-over-Union (IoU) metric was employed to evaluate the degree of accuracy in predicted segmentation against the manual annotations, thus validating the model's efficacy in localizing and identifying crack areas.

This study utilizes a specialized dataset comprising a select number of high-resolution images, with each meticulously focused on crack areas within masonry walls. This dataset is characterized by its diversity in crack patterns and conditions, specifically chosen to enhance the accuracy of the deep learning model in crack detection. Concentrating exclusively on the cracks ensured clarity and detail in the features crucial for structural analysis, thereby eliminating background distractions. This targeted approach is instrumental in developing a reliable tool for identifying and assessing structural integrity in historical masonry structures, with potential future expansions to include broader structural contexts. Figure 2 illustrates examples of (a) a historical masonry wall and (b) a crack dataset.

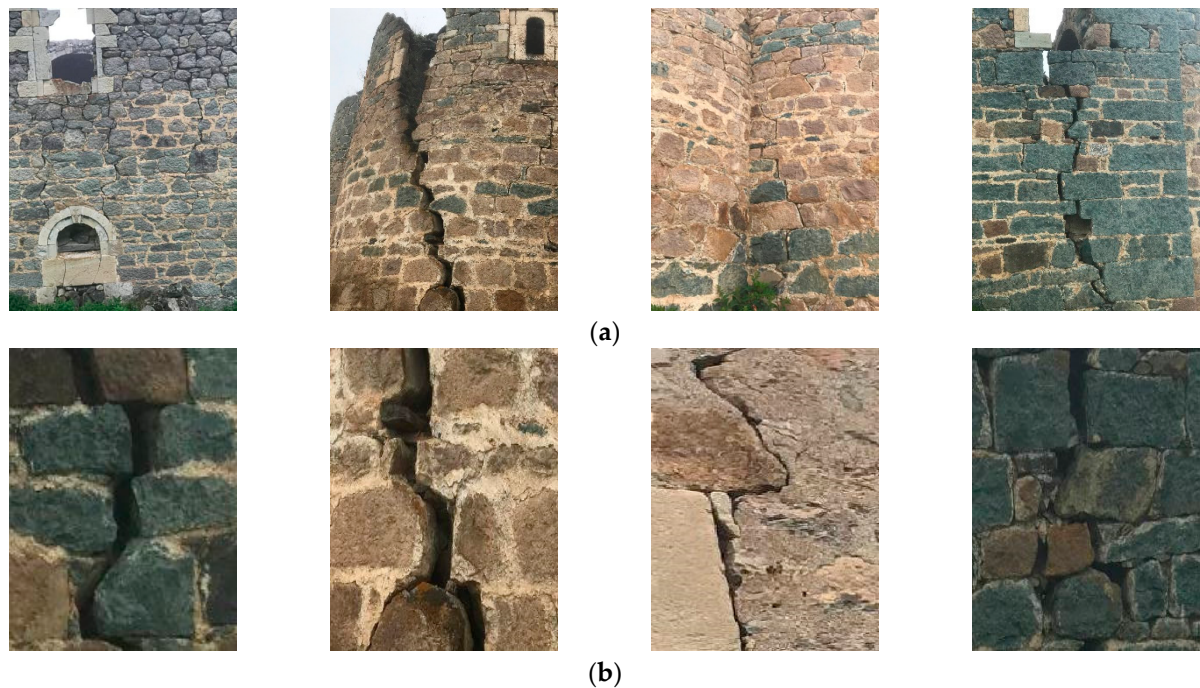


Figure 2. Examples of (a) historical masonry wall and (b) crack dataset.

2.2. Models and Computational Sources

This research aimed to compare the performances of several convolutional neural network (CNN) models, namely the custom CNN model, VGG-16, VGG-19, Inception-V3, Xception, and ResNet50. The custom CNN model was specifically designed for this study by sequentially stacking layers to form a neural network architecture. The main goal of this model is to diminish the dimensionality of the input images and to extract significant patterns pertinent to each class. To achieve this, the custom CNN model employed five convolutional blocks, each with an increasing number of filters compared to the previous block, facilitating finer-scale feature extraction. Within each block, a ‘Conv2D’ convolutional layer was employed, applying 16 filters of size 3-by-3 pixels to different regions of the input image. This process generated 16 feature maps, which represented the spatial distribution of specific features in the image. Furthermore, within each block of the model, a ‘MaxPooling2D’ layer was employed to downsample the feature maps. This layer consolidates 2-by-2-pixel grids into a single pixel, retaining the highest activation value from each grid. Consequently, the custom CNN model showcased its proficiency in capturing and representing critical features for the image analysis task.

The custom CNN model employed in this study consists of 512 feature maps, each with a size of 5 by 5 pixels. In order to generate predictions, these feature maps underwent a process of dimensionality reduction through global averaged pooling (GAP). The GAP operation calculates the average activation value within each feature map, transforming them into a one-dimensional tensor. By applying the GAP layer, the feature extraction phase of the model is completed. The purpose of this layer is to ensure that the activation values across different layers are normalized and maintain a consistent scale. This normalization helps mitigate potential instability during the adaptation process, which can arise from the propagation of excessively large values throughout the network. Following the feature extraction, a dense layer was added to make predictions for the two classes, namely crack and non-crack. The Softmax activation function was employed in this layer. To optimize the custom CNN model and its hyperparameters specifically for the task at hand, the Adam optimizer was utilized with a batch size of 32. The hyperparameter optimization process involved conducting 30 objective function evaluations guided by empirical evidence [45,47]. The architecture of the custom CNN model is visually depicted in Figure 3.

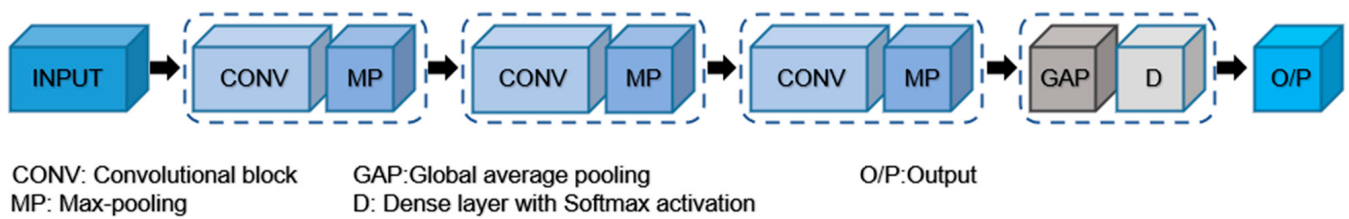


Figure 3. The architecture of the custom CNN model.

The pre-trained convolutional neural network (CNN) models in our study were initialized using weights from ImageNet. These models were then adapted by removing their fully connected layers and incorporating new layers, including global average pooling (GAP), a dropout layer with a ratio of 0.5, and a dense layer activated by a Softmax function. Additionally, the convolutional layer in these models uses zero-padding and 3×3 kernels and is designed with 1024 feature maps. Figure 4 in our manuscript illustrates these modified CNN architectures, highlighting their tailored configurations for our specific application.

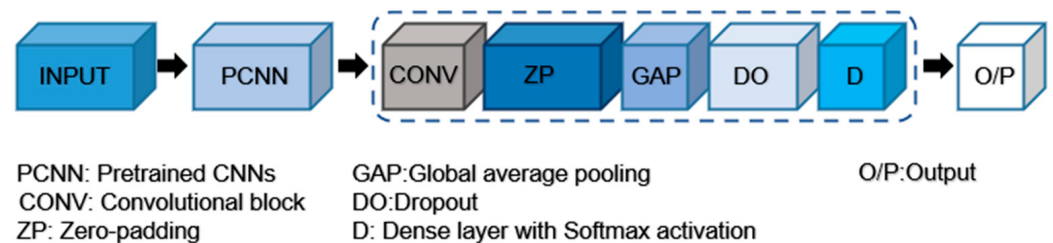


Figure 4. The architecture of the pre-trained CNNs.

In the selection of networks for the deep learning model, a strategic approach is employed based on several key criteria. These included the networks' proven performance in image recognition and feature extraction, their pre-training on extensive datasets like ImageNet, adaptability to transfer learning, computational efficiency, and previous success in similar applications such as structural health monitoring. Networks like ResNet50, VGG-16, VGG-19, Inception-V3, and Xception were chosen due to their robust capabilities in accurately identifying and analyzing diverse masonry textures and conditions. The decision was further reinforced by empirical evaluations on preliminary datasets, ensuring the networks' practical effectiveness in detecting and segmenting cracks in historical masonry structures.

Throughout the transfer learning and custom CNN training phases, a range of performance metrics, including accuracy, area under the curve (AUC), sensitivity, specificity, and the F measure, were used to evaluate the performance of the models. To expedite the computational processes, the CUDA/CUDNN libraries and the Keras API with a TensorFlow backend were employed, leveraging GPU acceleration. The training and validation of these models were carried out on a Windows 11 PC, which was equipped with 32 GB of RAM and an NVIDIA Quadro RTX 4000 GPU.

2.3. Transfer Learning

The use of pre-trained models as initial parameters for a different task is called transfer learning [36]. This method is frequently used in some deep learning problems. With the applied transfer learning method, designers have had the opportunity to both save time and obtain high accuracy rates. It is very difficult to obtain data and design complex models for different image processing problems. With the proposed transfer learning, it is possible to achieve higher performance with fewer data numbers.

Transfer learning, on the other hand, uses pre-trained models used in the solution of different problems as a starting parameter for the solution of the desired problems and provides solutions with faster and higher performance. Solving existing problems with deep learning methods requires a lot of data. For this reason, the number of data should be large to eliminate the overfitting problem. With the transfer learning method, transfer learning is used instead of training the network with random initial values [48,49]. With this method, training of convolutional neural network structures with less data is provided effectively.

In this study, first, the CNN model and then the transfer learning model were used for classification on the same data set. While 80% (± 2) success rate was achieved with CNN, 95% (± 1) success rate was obtained with the transfer learning model. The results are important in terms of showing that the transfer learning approach is useful. Figure 5 depicts a transfer learning architecture.

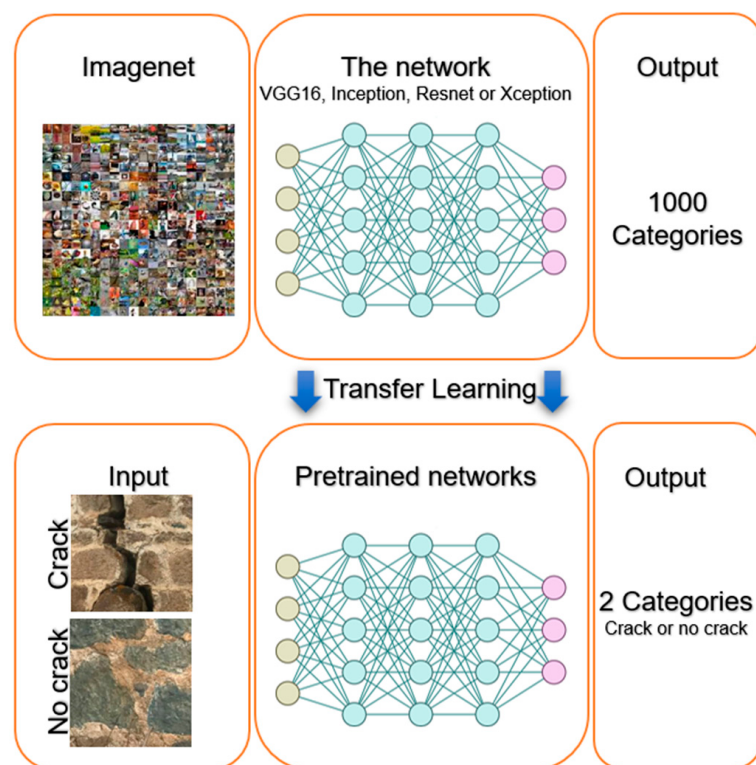


Figure 5. Transfer learning.

In proposed study, the weights of the pre-trained models are retained by setting the lower layers to non-trainable. This approach preserves the learned features from their original training. A global average pooling layer is integrated into the fully convolutional layer, calculating the mean of each feature map and creating a unique feature map for every class. These flattened feature maps are then processed through a series of layers—dense layers, dropout layers, and another dense layer—before entering the Softmax layer, which is crucial for multi-class classification.

The loss function used in this multi-class classification task is categorical cross-entropy, which was chosen for its effectiveness in such scenarios. The training process spans over 30 epochs with a batch size of 32, employing the Adam optimizer, which is known for its efficiency in handling large datasets and complex architectures, with a learning rate set at 10^{-4} . This particular learning rate was chosen for its balance between speed and stability in convergence.

Through the careful evaluation of various pre-trained models, it was found that the ResNet50 model stands out in terms of performance. This model's superior results demonstrate its robustness and suitability for the proposed specific application in identifying and classifying features within the dataset.

ResNet50

ResNet50 [46] consists of 152 layers and has approximately 23 million parameters. ResNet50 is a widely used image recognition and object classification architecture. The ResNet50 architecture, like other deep learning network architectures, aims to improve the congestion of deep network learning. This computer architecture, which resembles a regenerative neural network, works on the principle that the input is “residual” to the output of the next two convolutional layers. There is a “residual” feature block serving the ResNet architecture and a bottleneck occurrence in the ResNet50 architecture. Figure 6 depicts the ResNet architecture.

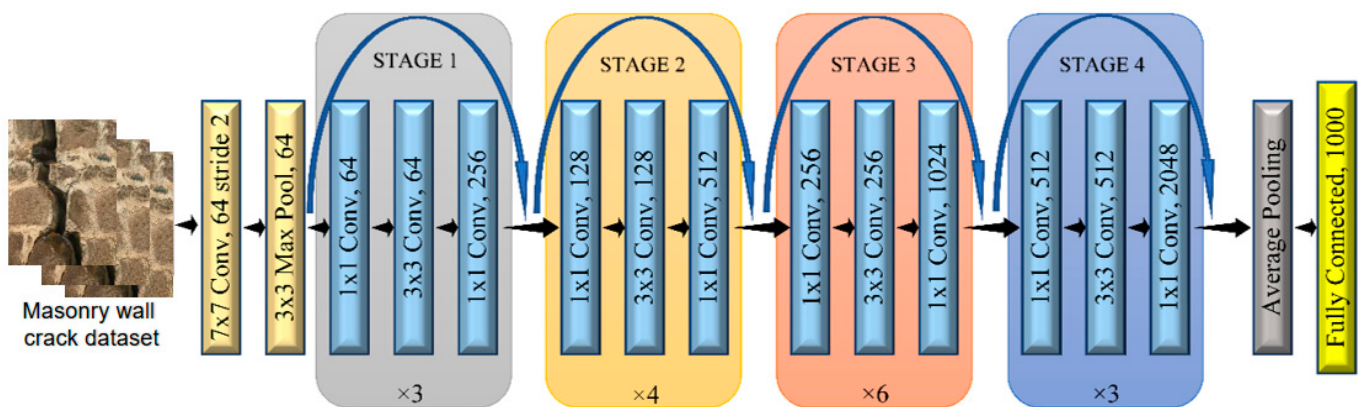


Figure 6. The network architecture of ResNet50.

Equation (1) expresses the output of the residual building block.

$$H(x) = F(x) + x \quad (1)$$

Since there are so many layers in the network, learning $H(x)$ will be difficult as it progresses. This is why the term “skipping link” is used. We learned $F(x)$ as a direct x input as a final output. As a result, $F(x)$ is referred to as “residual” in Figure 7.

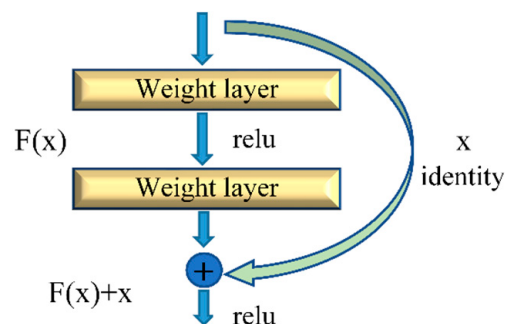


Figure 7. A residual block of deep residual network [46].

2.4. Segmentation Methods

2.4.1. CAM-K-SEG Method

Grad-CAM is a technique that combines class activation maps and saliency maps to highlight the importance of pixels with respect to a particular class label. It leverages the information present in the last convolutional layers to rank the significance of each pixel in relation to the target class. To achieve this, the gradient scores of each class are mapped onto the image features, and the associated gradient weights are computed using global average pooling (as described in Equation (2)).

$$W_m^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^m} \quad (2)$$

In Equation (2), the weight matrix, W_m^c , with dimensions of $n \times m$, is responsible for capturing the relevant feature map m for each class label c . The activation map A^m consists of Z pixels, and each of these pixels is assigned a gradient score y^c . After its computation, the weight matrix is multiplied by the output of the last convolutional layer. This product is then activated using the Rectified Linear Unit (*ReLU*) activation function [50]. This approach, consistent with the one employed in the earlier implementation of the ResNet-50 model [46], leverages the *ReLU* function's ability to introduce non-linearity, enhancing the model's learning capabilities.

$$L_{Grad-CAM}^c = Relu \left(\sum_m W_m^c A^m \right) \quad (3)$$

In Equation (3), $L_{Grad-CAM}^c$ denotes the *Grad-CAM* localization map tailored for class c , possessing dimensions of width u and height v . This map generates a heatmap of size $u \times v$, adeptly accentuating the pixels that are most relevant for the segmentation of the brick class. Integral to the *Grad-CAM* algorithm is the use of the *ReLU* function, which is pivotal in assigning positive weights to pixels identified as significant for the target class while effectively nullifying the influence of less important pixels by assigning them zero weights.

In proposed research, we developed and implemented a unique segmentation framework named CAM-K-SEG, designed to automatically generate accurately labeled masks for classified object images. This framework synergizes the *Grad-CAM* visualization technique and the K-Mean clustering algorithm with pre-trained classification models, facilitating the precise prediction of surface segmentation within various object classes. As detailed in Figure 8, the methodology for constructing the CAM-K-SEG model involved an initial application of the *Grad-CAM* method for the localization of target objects. This approach leveraged the outstanding classification performance of the ResNet50 pre-trained model, utilizing its advanced capabilities to enhance the effectiveness and accuracy of our segmentation process. Subsequently, the K-Mean clustering algorithm was applied to determine the optimal K value using the elbow method and identify the number of effective colors. Thresholding was then utilized to convert the heatmap image into a binary image based on a specific pixel intensity threshold, facilitating the extraction of dominant foreground and background objects. This approach resulted in the development of an algorithm that achieved optimal performance in object segmentation. The predicted crack areas were denoted by red rectangular boxes (predicted bounding boxes), while the actual crack areas were represented by green rectangular boxes (ground truth bounding boxes), which were added to the images. By comparing the coordinates obtained from these two bounding boxes, IoU metrics were computed for each image. The framework we have proposed is composed of three primary components: image classification for segmenting crack areas, class activation mapping (CAM) visualization, and K-Mean color clustering applied to the masked crack regions.

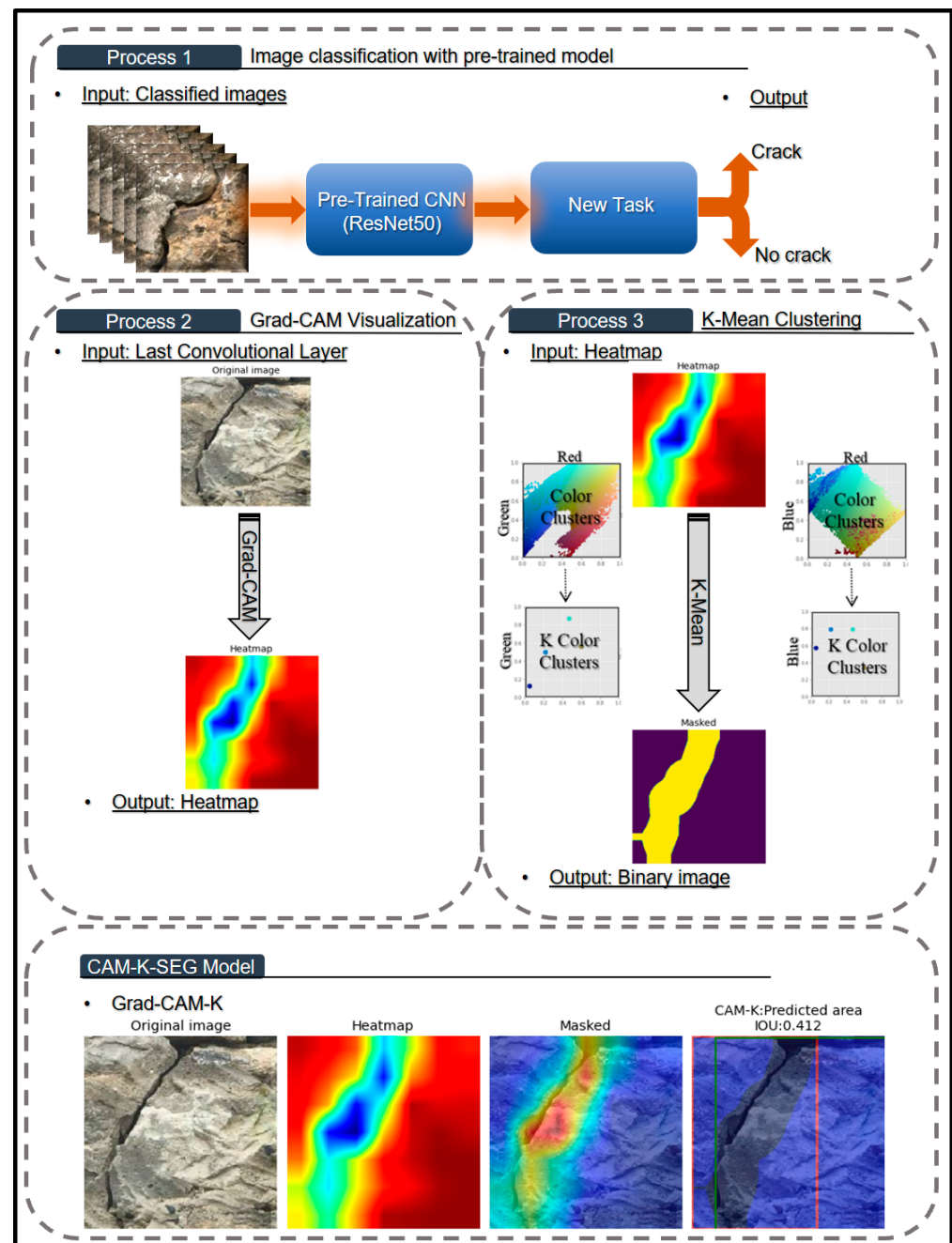


Figure 8. Methodology of CAM-K-SEG model.

2.4.2. U-Net Segmentation Model

U-Net, as conceptualized by Ronneberger et al. [37], is a convolutional neural network (CNN) architecture tailored for image segmentation. Characterized by its distinctive U-shape, depicted in Figure 9, the architecture features a symmetrical design with skip connections linking the downsampling and upsampling paths. U-Net is renowned for its various advantages, including reduced training time, simplicity, and a smaller parameter count compared to other networks, making it highly effective even with limited training data.

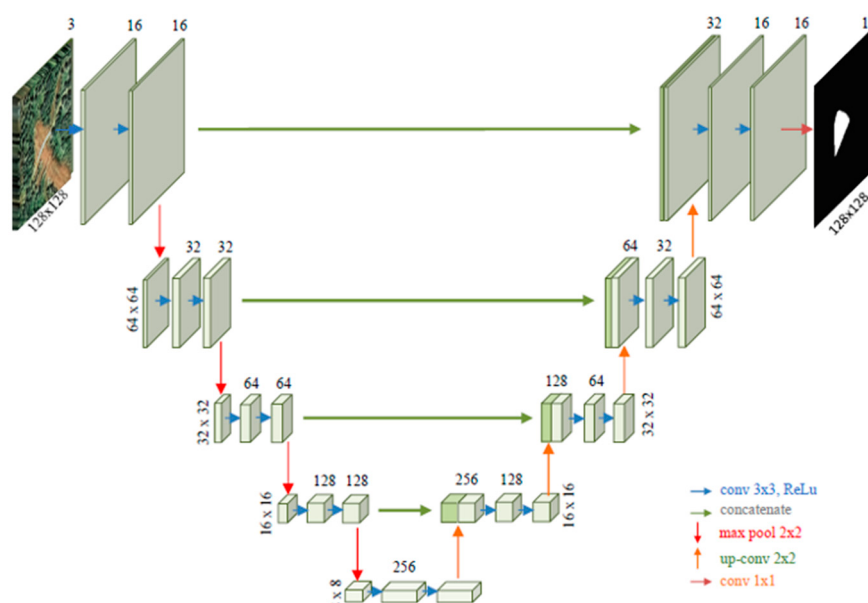


Figure 9. U-Net architecture [37].

The downsampling path on the left side of the U-shape comprises four blocks. Each block contains two 3×3 convolution layers, each followed by an activation function (including batch normalization) and a 2×2 max pooling layer. With each pooling step, the number of feature maps doubles, allowing the network to extract and process increasingly complex features from the input image. This path is crucial for capturing contextual information necessary for accurate segmentation.

Conversely, the upsampling path also consists of four blocks. It involves deconvolution layers that merge with feature maps from the downsampling path, additional 3×3 convolution layers with activation functions (including batch normalization), and a final 1×1 convolution operation to produce the segmented image with the required number of channels.

The U-Net architecture's strengths lie in its ability to handle limited training data, recognize and fuse multi-scale features, maintain simplicity, and deliver high-quality pixel-level segmentation results. Since its introduction, U-Net has become highly popular in medical imaging and has seen various advancements and modifications, incorporating new methodologies and integrating different imaging techniques. Its application has extended beyond medical imaging, demonstrating its versatility and adaptability to new challenges and diverse image datasets. This adaptability of the enhanced U-Net model underscores its success in adjusting to various tasks and different image datasets.

3. Results

3.1. Performance Metrics Evaluation for CNN Models

Optimal hyperparameter values selected for both custom and pre-trained CNN models are detailed in Table 1. Table 2 presents the performance metrics achieved by various models in identifying crack classes, using a test set comprised of images labeled as 'crack' and 'no-crack'. A comparative analysis reveals that the pre-trained ResNet50 model surpasses others in key metrics such as accuracy, recall, F measure, and Matthews correlation coefficient. Except for precision, all metrics in Table 2 indicate the superior performance of the ResNet50 model. Consequently, the Grad-CAM visualization technique, integrated with the ResNet50 model, was chosen for the development of the CAM-K-SEG model. The receiver operating characteristic (ROC) curve and precision–recall graphs, displayed in Figure 10, further substantiate the outstanding performance of the ResNet50 model, as shown in Table 2.

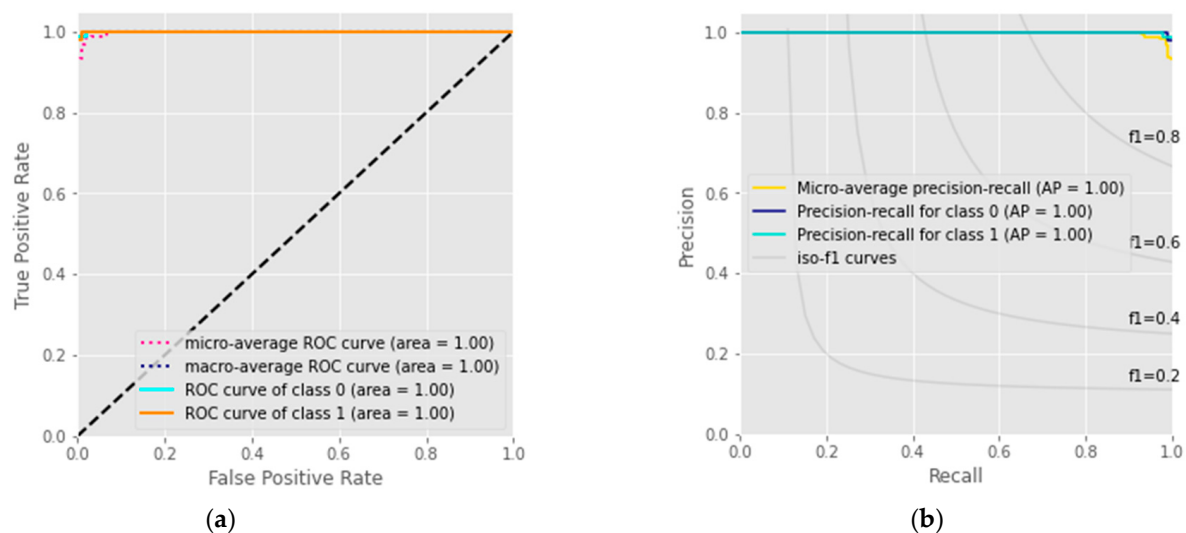
Table 1. Optimized hyperparameter values for the CNN models.

Model	Optimal Values			
	Learning Rate	Momentum	Depth	L2-Decay
Custom CNN	1×10^{-3}	0.90	1	1×10^{-6}
Pre-trained CNNs	1×10^{-4}	0.95	-	1×10^{-6}

Table 2. Performance metrics were achieved by the coarse models using the masonry test set.

Model	Accuracy	Precision	Recall	F Measure	AUC
Custom-CNN	%93.00	0.89	0.98	0.93	0.93
VGG-16	%86.46	0.82	0.93	0.87	0.89
VGG-19	%89.06	0.87	0.93	0.90	0.87
Inception-V3	%77.08	0.10	0.03	0.05	0.53
Xception	%96.35	0.94	0.99	0.97	0.54
ResNet50	%98.44	0.97	1.00	0.99	1.00

Notes: Bold values show superior performance.

**Figure 10.** Extension of (a) the receiver operating characteristic and (b) precision-recall curve to multi-class.

3.2. Visual Localization Evaluation CAM-K-SEG Model

The “conv5_block3_3_conv” layer of the ResNet50 model, essential for feature extraction, has exhibited outstanding performance in distinguishing between crack and no-crack image classes. To assess the precision of Grad-CAM in pinpointing crack areas, the ResNet50 model, which demonstrated the highest efficacy in crack detection, was applied to a test set specifically curated for crack classification. Figure 11 illustrates the localization outcomes achieved using the Grad-CAM technique with the ResNet50 model’s final convolutional feature extraction layer. This visualization method has been markedly successful in accurately identifying regions with cracks, thereby underlining its excellence in the field of crack detection.

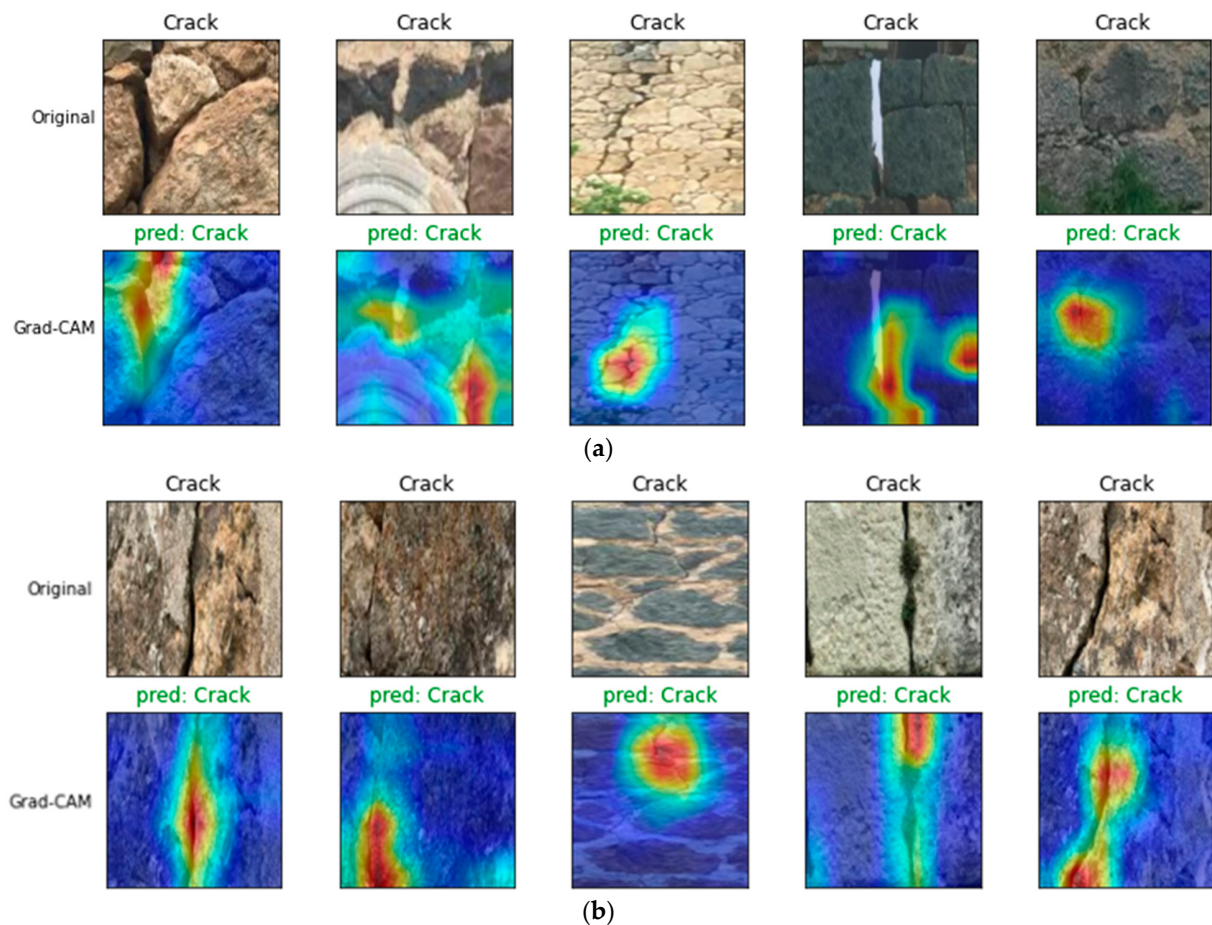


Figure 11. Grad-CAM visualization samples for crack areas, labeled as (a,b).

Figure 12 methodically illustrates the sequential steps involved in the CAM-K-SEG method. The process begins with Figure 12a the original input image, followed by Figure 12b the creation of a heatmap image using the Grad-CAM visualization technique based on the ResNet50 model. Subsequently, Figure 12d K-Mean clustering is applied to the Figure 12c composite image, which is formed by overlaying the original image with the heatmap to isolate the most effective colors. This is followed by converting the image Figure 12e to grayscale and applying filtering processes, which include the elimination of small, irrelevant objects. The image is then converted Figure 12f into a binary format. Finally, Figure 12g the image is segmented, and this segmented image is presented as the outcome of the CAM-K-SEG process.

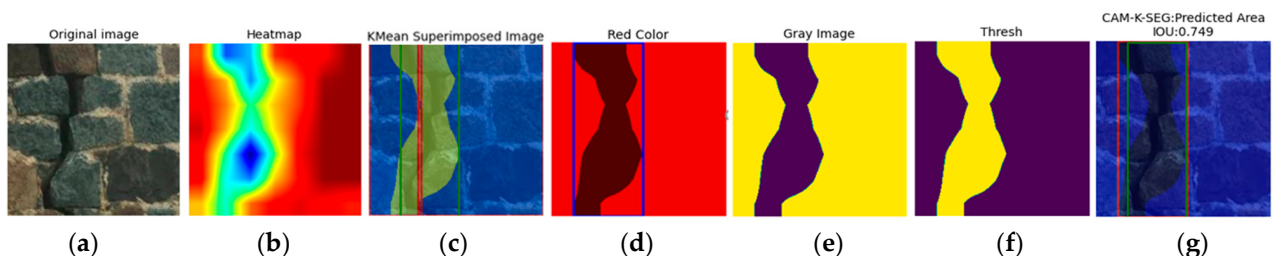


Figure 12. CAM-K-SEG Process Steps. (a) Original input image, (b) Heatmap image, (c) Composite image formed by overlaying the original image with the heatmap, (d) Application of K-Mean clustering to the composite image, (e) Conversion of the image to grayscale and application of filtering processes including the elimination of small, irrelevant objects, (f) Conversion of the image into a binary format, (g) Segmentation of the image, with the segmented image presented as the final outcome of the CAM-K-SEG process.

3.3. Comparisons

The present study employed two distinct approaches for the detection and localization of cracks in historical masonry structures that develop over time due to external factors. The first method, known as CAM-K-SEG, employs bounding boxes to locate the crack regions in the model, allowing for a subsequent comparison with the ground truth bounding boxes. This comparison enables the computation of the Intersection-over-Union (*IoU*) metric values using Equation (4). On the other hand, the U-Net segmentation model was utilized for the precise segmentation of crack areas. To facilitate this, the actual crack regions were masked, resulting in the creation of a new training dataset for the model. In both of these approaches, the *IoU* metric, a commonly used evaluation metric for assessing segmentation algorithms, was employed in this study. The *IoU* metric measures the similarity between two masks, namely y_{true} , y_{pred} , and is calculated accordingly.

$$IoU = \frac{y_{true} \cap y_{pred}}{y_{true} \cup y_{pred}} \quad (4)$$

In Figure 13, analyses were performed on five different cracked wall samples. On the samples, the green bounding box marks the actual crack area, while the red bounding box represents the estimated crack areas. For each sampling performed with CAM-K-SEG, it has been observed that the segmented areas and the bounding boxes surrounding them do not have the desired level of *IoU* scores, although they have real crack areas and are included in ground truth bounding boxes. In some samples (Figure 13b), it is understood that not all cracks can be detected correctly due to the Grad-CAM method's feature of estimating only the weighted regions.

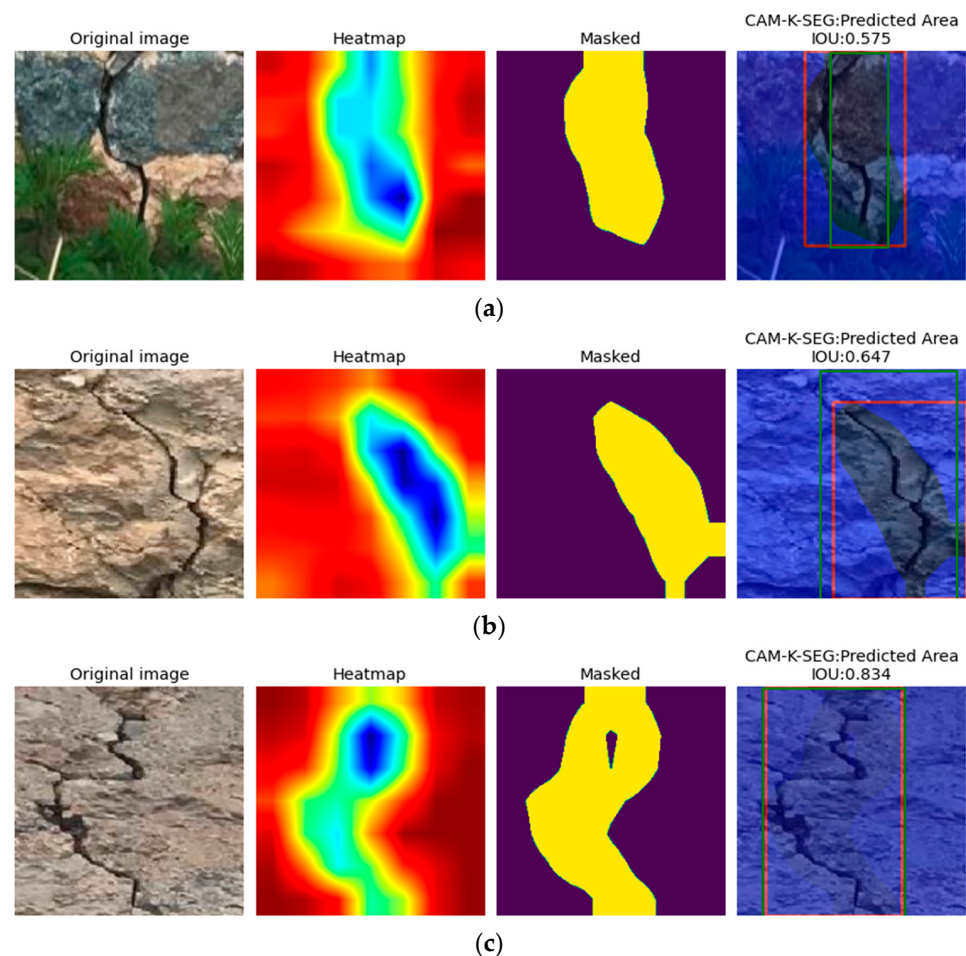


Figure 13. Cont.

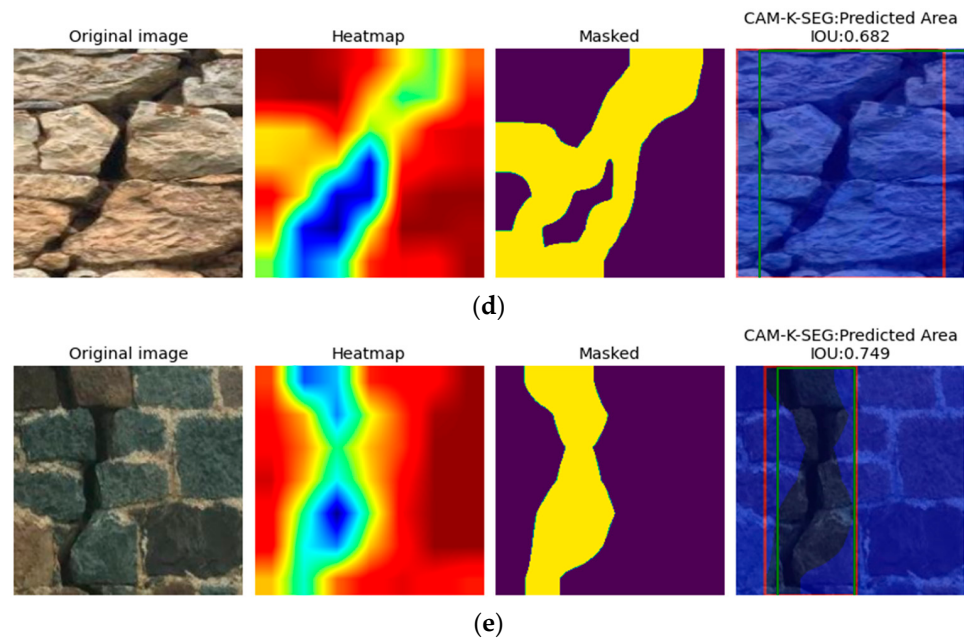


Figure 13. Examples of using CAM-K-SEG to determine the five different cracked wall samples, labeled as (a–e).

The U-Net architecture was implemented in its original form to train the model using the provided dataset. The training process was conducted over 100 epochs, with a batch size of 16, where each iteration of the training data in the corresponding dataset was considered as one epoch. The loss and accuracy values achieved were calculated as 0.036 and 0.95, respectively. The evaluation of the loss and accuracy for the U-Net architecture with a resolution of 128×128 is visualized in Figure 14. The training accuracy is represented by the orange curves, while the training loss rate is depicted by the blue curves.

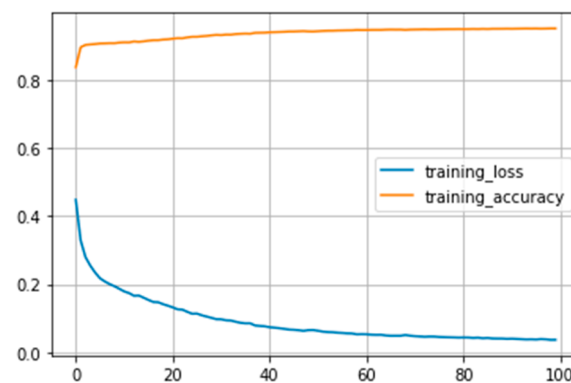


Figure 14. Loss assessment and accuracy results of U-Net architecture.

Figure 15 shows the crack areas and IoU metric values obtained with the U-Net model. The same crack images were taken into account for comparison with the CAM-K-SEG model. It is clear that IoU metric values underperform in these samples. However, since the crack areas are very small, a small, wrong area scanning makes a big difference in the estimation of the segmented areas. This results in low IoU values. Depending on the IoU metric values, it is seen that the CAM-K-SEG model performs better than the U-Net model. However, the point to be noted here is that the comparison is made according to the IoU values determined by different methods.

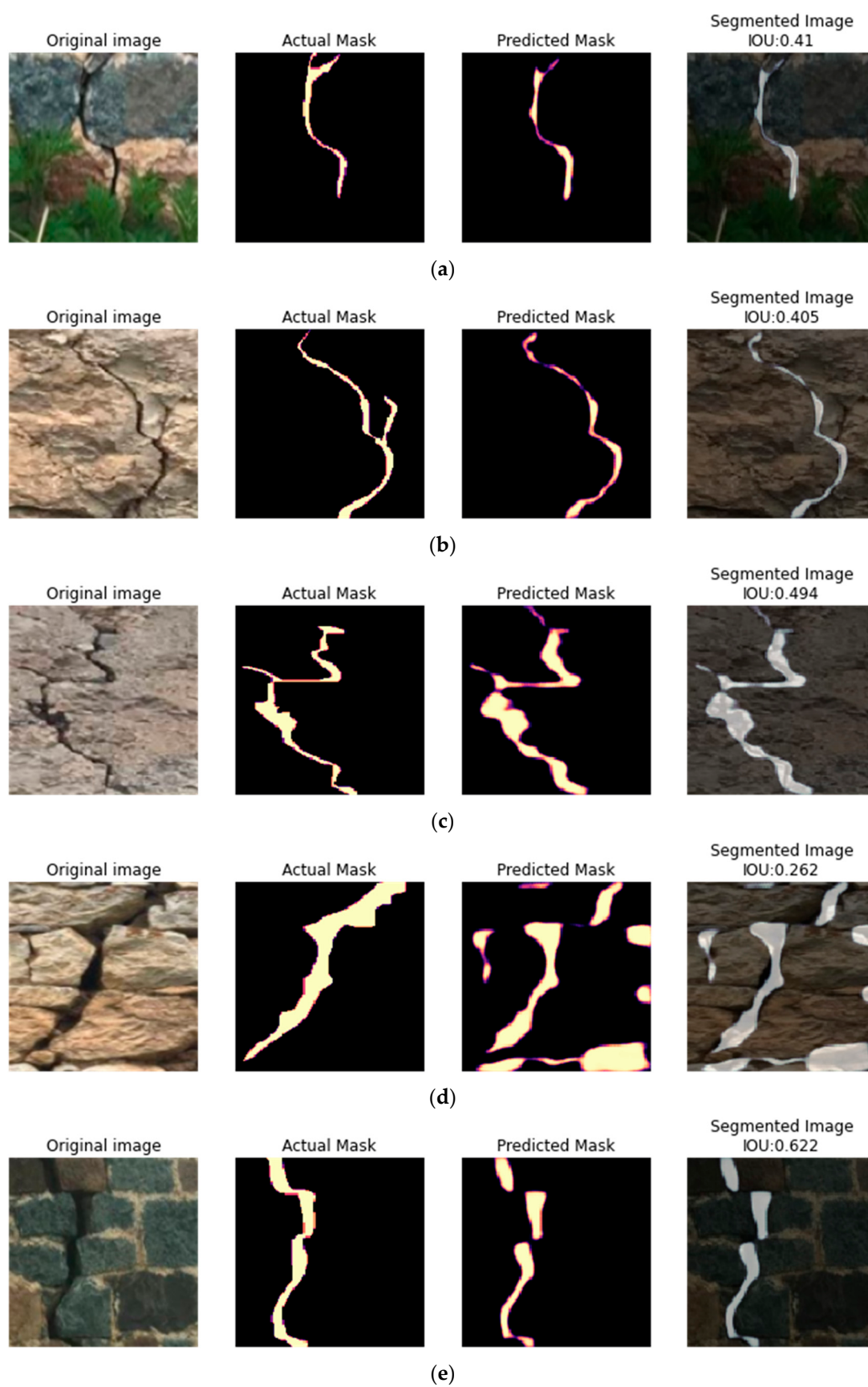


Figure 15. Examples of five different crack areas (labeled as a–e) along with their respective IoU metric values as predicted by the U-Net model.

3.4. Discussion

In the comparative analysis, the performance of the CAM-K-SEG model was compared with U-Net to verify the effectiveness of the proposed model. This comparison was critical to demonstrate that while the U-Net model demonstrated high performance on segmentation tasks, the CAM-K-SEG model exhibited comparable, if not superior, capabilities, particularly in the context of segmentation of crack areas within masonry structures. However, it showed superior performance in detecting crack locations.

Therefore, the CAM-K-SEG model is not only validated by this comparison but was also demonstrated to be a special solution well suited for the subtle task of crack detection in the wall. Its development was driven by the need to create a model for structural analysis that could not only segment but also effectively localize and classify crack areas.

In this study, the CAM-K-SEG model, leveraging the ResNet50 architecture, has shown a superior ability in object identification and localization of cracks, particularly when compared to traditional methods that largely depend on manual inspection and basic image processing. The model's average IoU value of 0.70 is a testament to its effectiveness in accurately locating crack areas, a significant achievement given the complex textures and patterns often found in historical masonry structures. However, it is important to note that while the CAM-K-SEG model excels in identifying the location of cracks, it may not always precisely define the exact boundary regions of these cracks.

In contrast, the U-Net model, renowned for its image segmentation capabilities, demonstrates a high success rate in segmenting crack areas, achieving an average IoU value of 0.43. This performance is particularly notable when compared to other deep learning models used in similar applications. The U-Net model's proficiency in segmenting small and intricate crack patterns effectively makes it an invaluable tool for detailed analysis of structural integrity.

Both the CAM-K-SEG and U-Net models stand out for their practicality, requiring less labor and proving to be cost-effective during the estimation and localization phases of crack areas. This represents a significant advancement over existing methods, which often necessitate extensive manual labor and struggle with the complex nature of historical masonry structures.

Additionally, the approach of utilizing cellphone images for crack detection introduces an element of accessibility and ease of use, which is not commonly found in many advanced methods. This adaptability makes the method more suitable for regular monitoring and assessment of historical structures, a critical aspect in their preservation and maintenance.

However, during the initial training phase of the CAM-K-SEG model, the focus was primarily on more pronounced cracks, leading to a gap in performance on subtler crack patterns. Recognizing this limitation, the importance of incorporating a dataset with finer cracks in future studies is emphasized. This will necessitate modifying the CAM-K-SEG model to enhance its sensitivity and accuracy in detecting these less obvious cracks. By expanding the scope of the dataset and refining the model accordingly, this paper aims to improve the model's applicability in real-world scenarios, particularly in the context of historical structure preservation, where early detection of minor cracks is crucial. This planned enhancement of the CAM-K-SEG model will not only address its current limitations but also significantly contribute to the field of structural health monitoring by providing a more comprehensive and reliable tool for crack detection.

In conclusion, this comparative study with state-of-the-art methods underscores the effectiveness of the proposed models. The CAM-K-SEG model excels in identifying and localizing crack areas, while the U-Net model is highly effective in the detailed segmentation of cracks. Both models offer significant improvements over traditional methods, providing more accurate, efficient, and cost-effective solutions for crack detection in historical masonry structures.

4. Conclusions

This research presents a comprehensive analysis of crack detection in historical masonry structures using deep learning methods. This study introduces the CAM-K-SEG model, which leverages the class activation mapping technique to detect cracks induced by environmental effects with a high degree of success. The model's strength lies in its rapid training capabilities and minimal labor requirements for operation, significantly streamlining the process of crack detection and damage mapping for engineers.

The CAM-K-SEG model, underpinned by the ResNet50 architecture, has demonstrated exemplary performance in the experiments. It not only detects but also localizes the crack areas with remarkable accuracy, making it a potent tool for structural analysis. The model's effectiveness is rooted in its ability to facilitate rapid, accurate damage assessment, which is crucial for the maintenance and preservation of heritage structures.

Concurrently, the U-Net model was deployed to assess its segmentation prowess on the same task. While the U-Net showed remarkable success in object segmentation, indicating its suitability for high-performance segmentation jobs, the findings suggest that the CAM-K-SEG model is more adept at object identification and localization. This distinction underscores the CAM-K-SEG model's utility in scenarios where precise crack localization is paramount.

This study's results indicate that the CAM-K-SEG model, using ResNet50, exhibits the best performance metrics among the models tested, specifically in the segmentation of crack areas based on pixels. Meanwhile, the U-Net model's high performance in segmentation tasks is noted, reinforcing the model's applicability in scenarios that demand detailed segmentation over simple localization.

In synthesizing these findings, it is asserted that the CAM-K-SEG model is a significant contribution to the field of crack detection in masonry structures. It is a method that not only facilitates rapid and accurate crack area detection but also embodies the potential for future applications across a broader spectrum of structural health monitoring tasks.

This research introduces a transformative approach to detecting cracks in historical masonry structures, leveraging the CAM-K-SEG deep learning model. This method significantly surpasses traditional inspection techniques in accuracy and efficiency, offering a cost-effective, accessible solution for structural health monitoring. By utilizing cellphone images, it enables easy and widespread application, even in challenging locations. Crucially, it facilitates early detection of structural issues, essential for the preventive maintenance and preservation of heritage buildings. The quantitative data provided by the proposed model also support informed, data-driven decision-making in repair and maintenance strategies, ensuring effective resource allocation and extending the lifespan of culturally significant structures.

Future work will focus on enhancing the model's precision in detecting finer cracks, broadening its scope to a wider range of structures, and incorporating advanced data sources like drone imagery and 3D scanning. We also aim to explore semi-supervised learning to reduce manual labeling efforts, develop real-time crack detection applications, and enhance the interpretability of the model's decisions. These efforts are geared towards refining our approach, making it more scalable and universally applicable in the field of structural health monitoring and the preservation of global cultural heritage.

Author Contributions: Conceptualization, K.H. and T.A.; Methodology, K.H., A.C.A. and T.A.; Software, K.H. and T.A.; Validation, A.C.A.; Formal analysis, A.C.A.; Investigation, A.C.A. and T.A.; Resources, T.A.; Writing—original draft, K.H.; Writing—review & editing, A.C.A.; Visualization, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data, models, and/or codes that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

Conflicts of Interest: Author Ahmet Can Altunışık was the founder of Dynamic Academy Software, Construction Ind. Trade. Co., Ltd. Şti. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- McKibbins, L.; Melbourne, C.; Sawar, N.; Gaillard, C. *Masonry Arch Bridges: Condition Appraisal and Remedial Treatment*; Treatment (C656); CIRIA: London, UK, 2006; pp. 29–85.
- Phares, B.M.; Washer, G.A.; Rolander, D.D.; Graybeal, B.A.; Moore, M. Routine Highway Bridge Inspection Condition Documentation Accuracy and Reliability. *J. Bridge Eng.* **2004**, *9*, 403–413. [\[CrossRef\]](#)
- Eaton, J.; Edwards, M.; Crapper, M. *Heritage Railway Association: The Inspection and Maintenance of Civil Engineering Assets*; Ref No.: HGR-A0701; Heritage Railway Association (HRA): Edinburgh, UK, 2014.
- Spencer, B.F.; Hoskere, V.; Narazaki, Y. Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering* **2019**, *5*, 199–222. [\[CrossRef\]](#)
- Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P. A Review on Computer Vision Based Defect Detection and Condition Assessment of Concrete and Asphalt Civil Infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210. [\[CrossRef\]](#)
- Khuc, T.; Catbas, F.N. Computer Vision-Based Displacement and Vibration Monitoring without Using Physical Target on Structures. In *Bridge Design, Assessment and Monitoring*; Routledge: London, UK, 2018; pp. 89–100. [\[CrossRef\]](#)
- Feng, D.; Feng, M.Q. Computer Vision for SHM of Civil Infrastructure: From Dynamic Response Measurement to Damage Detection—A Review. *Eng. Struct.* **2018**, *156*, 105–117. [\[CrossRef\]](#)
- Hoskere, V.; Park, J.-W.; Yoon, H.; Spencer, B.F., Jr. Vision-Based Modal Survey of Civil Infrastructure Using Unmanned Aerial Vehicles. *J. Struct. Eng.* **2019**, *145*, 04019062. [\[CrossRef\]](#)
- Valero, E.; Forster, A.; Bosché, F.; Hyslop, E.; Wilson, L.; Turmel, A. Automated Defect Detection and Classification in Ashlar Masonry Walls Using Machine Learning. *Autom. Constr.* **2019**, *106*, 102846. [\[CrossRef\]](#)
- Brackenbury, D.; Brilakis, I.; Dejong, M. Automated Defect Detection for Masonry Arch Bridges. In Proceedings of the International Conference on Smart Infrastructure and Construction 2019, ICSIC 2019: Driving Data-Informed Decision-Making, Cambridge, UK, 8–10 July 2019; ICE Publishing: Miami, FL, USA, 2019; pp. 3–10.
- Wang, N.; Zhao, X.; Zhao, P.; Zhang, Y.; Zou, Z.; Ou, J. Automatic Damage Detection of Historic Masonry Buildings Based on Mobile Deep Learning. *Autom. Constr.* **2019**, *103*, 53–66. [\[CrossRef\]](#)
- Cintas, C.; Lucena, M.; Fuertes, J.M.; Delrieux, C.; Navarro, P.; González-José, R.; Molinos, M. Automatic Feature Extraction and Classification of Iberian Ceramics Based on Deep Convolutional Networks. *J. Cult. Herit.* **2020**, *41*, 106–112. [\[CrossRef\]](#)
- Navarro, P.; Cintas, C.; Lucena, M.; Fuertes, J.M.; Delrieux, C.; Molinos, M. Learning Feature Representation of Iberian Ceramics with Automatic Classification Models. *J. Cult. Herit.* **2021**, *48*, 65–73. [\[CrossRef\]](#)
- Pathak, R.; Saini, A.; Wadhwa, A.; Sharma, H.; Sangwan, D. An Object Detection Approach for Detecting Damages in Heritage Sites Using 3-D Point Clouds and 2-D Visual Data. *J. Cult. Herit.* **2021**, *48*, 74–82. [\[CrossRef\]](#)
- Mishra, M. Machine Learning Techniques for Structural Health Monitoring of Heritage Buildings: A State-of-the-Art Review and Case Studies. *J. Cult. Herit.* **2021**, *47*, 227–245. [\[CrossRef\]](#)
- Hatir, M.E.; Barstuğan, M.; İnce, İ. Deep Learning-Based Weathering Type Recognition in Historical Stone Monuments. *J. Cult. Herit.* **2020**, *45*, 193–203. [\[CrossRef\]](#)
- Hatir, E.; Korkanç, M.; Schachner, A.; İnce, İ. The Deep Learning Method Applied to the Detection and Mapping of Stone Deterioration in Open-Air Sanctuaries of the Hittite Period in Anatolia. *J. Cult. Herit.* **2021**, *51*, 37–49. [\[CrossRef\]](#)
- Nguyen, N.T. Automatic Assessment of Structural Damage of Masonry Structures by Visual Analysis of Surface Cracks. Unpublished Thesis, Centers & Institutes, Texas State University, San Marcos, TX, USA, 2021.
- Hallee, M.J.; Napolitano, R.K.; Reinhart, W.F.; Glisic, B. Crack Detection in Images of Masonry Using Cnns. *Sensors* **2021**, *21*, 4929. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bai, Y.; Sezen, H.; Yilmaz, A. Detecting Cracks and Spalling Automatically in Extreme Events by End-to-End Deep Learning Frameworks. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2021**, *V-2-2021*, 161–168. [\[CrossRef\]](#)
- Chaiyasarn, K.; Khan, W.; Ali, L.; Sharma, M.; Brackenbury, D.; Dejong, M. Crack Detection in Masonry Structures Using Convolutional Neural Networks and Support Vector Machines. In Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC2018), Berlin, Germany, 20–25 July 2018.
- Wang, C.; Antos, S.E.; Triveno, L.M. Automatic Detection of Unreinforced Masonry Buildings from Street View Images Using Deep Learning-Based Image Segmentation. *Autom. Constr.* **2021**, *132*, 103968. [\[CrossRef\]](#)
- Ye, X.W.; Jin, T.; Li, Z.X.; Ma, S.Y.; Ding, Y.; Ou, Y.H. Structural Crack Detection from Benchmark Data Sets Using Pruned Fully Convolutional Networks. *J. Struct. Eng.* **2021**, *147*, 04721008. [\[CrossRef\]](#)
- Musella, C.; Serra, M.; Menna, C.; Asprone, D. Building Information Modeling and Artificial Intelligence: Advanced Technologies for the Digitalisation of Seismic Damage in Existing Buildings. *Struct. Concr.* **2021**, *22*, 2761–2774. [\[CrossRef\]](#)

25. Chaayasarn, K.; Sharma, M.; Ali, L.; Khan, W.; Poovarodom, N. Crack Detection in Historical Structures Based on Convolutional Neural Network. *Int. J. Geomate* **2018**, *15*, 240–251. [\[CrossRef\]](#)
26. Dais, D.; Bal, İ.E.; Smyrou, E.; Sarhosis, V. Automatic Crack Classification and Segmentation on Masonry Surfaces Using Convolutional Neural Networks and Transfer Learning. *Autom. Constr.* **2021**, *125*, 103606. [\[CrossRef\]](#)
27. Latifi, R.; Hadzima-Nyarko, M.; Radu, D.; Rouhi, R. A Brief Overview on Crack Patterns, Repair and Strengthening of Historical Masonry Structures. *Materials* **2023**, *16*, 1882. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Saadatmorad, M.; Talookolaei, R.A.J.; Milani, G.; Khatir, S.; Cuong-Le, T. Crack Detection in Historical Masonry Structures Using Efficient Image Processing: Application on a Masonry Bridge in Iran. In Proceedings of the 2023 IEEE International Workshop on Metrology for Living Environment, MetroLivEnv 2023—Proceedings 2023, Milano, Italy, 29–31 May 2023; pp. 230–235. [\[CrossRef\]](#)
29. Wang, L. Automatic Detection of Concrete Cracks from Images Using Adam-SqueezeNet Deep Learning Model. *Frat. Ed Integrita Strutt.* **2023**, *17*, 289–299. [\[CrossRef\]](#)
30. Katsigiannis, S.; Seyedzadeh, S.; Agapiou, A.; Ramzan, N. Deep Learning for Crack Detection on Masonry Façades Using Limited Data and Transfer Learning. *J. Build. Eng.* **2023**, *76*, 107105. [\[CrossRef\]](#)
31. Cardellicchio, A.; Ruggieri, S.; Nettis, A.; Renò, V.; Uva, G. Physical Interpretation of Machine Learning-Based Recognition of Defects for the Risk Management of Existing Bridge Heritage. *Eng. Fail. Anal.* **2023**, *149*, 107237. [\[CrossRef\]](#)
32. Cardellicchio, A.; Ruggieri, S.; Nettis, A.; Mosca, N.; Uva, G.; Renò, V. On the Use of YOLOv5 for Detecting Common Defects on Existing RC Bridges. *Opt. Metrol.* **2023**, *15*, 126210G. [\[CrossRef\]](#)
33. Huang, Y.; Zhang, H.; Li, H.; Wu, S. Recovering Compressed Images for Automatic Crack Segmentation Using Generative Models. *Mech. Syst. Signal Process.* **2021**, *146*, 107061. [\[CrossRef\]](#)
34. Gao, Z.; Peng, B.; Li, T.; Gou, C. Generative Adversarial Networks for Road Crack Image Segmentation. In Proceedings of the International Joint Conference on Neural Networks 2019, Budapest, Hungary, 14–19 July 2019. [\[CrossRef\]](#)
35. Pan, Z.; Lau, S.L.H.; Yang, X.; Guo, N.; Wang, X. Automatic Pavement Crack Segmentation Using a Generative Adversarial Network (GAN)-Based Convolutional Neural Network. *Results Eng.* **2023**, *19*, 101267. [\[CrossRef\]](#)
36. Chen, B.; Zhang, H.; Wang, G.; Huo, J.; Li, Y.; Li, L. Automatic Concrete Infrastructure Crack Semantic Segmentation Using Deep Learning. *Autom. Constr.* **2023**, *152*, 104950. [\[CrossRef\]](#)
37. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015 18th International Conference, Munich, Germany, 5–9 October 2015; Volume 9351, pp. 234–241.
38. Hacıfendioglu, K.; Başağa, H.B.; Yavuz, Z.; Karimi, M.T. Intelligent Ice Detection on Wind Turbine Blades Using Semantic Segmentation and Class Activation Map Approaches Based on Deep Learning Method. *Renew. Energy* **2022**, *182*, 1–16. [\[CrossRef\]](#)
39. Loverdos, D.; Sarhosis, V. Automatic Image-Based Brick Segmentation and Crack Detection of Masonry Walls Using Machine Learning. *Autom. Constr.* **2022**, *140*, 104389. [\[CrossRef\]](#)
40. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2016**, *128*, 336–359. [\[CrossRef\]](#)
41. Ahmad, A.; Dey, L. A K-Mean Clustering Algorithm for Mixed Numeric and Categorical Data. *Data Knowl. Eng.* **2007**, *63*, 503–527. [\[CrossRef\]](#)
42. Hacıfendioglu, K.; Mostofi, F.; Toğan, V.; Başağa, H.B. CAM-K: A Novel Framework for Automated Estimating Pixel Area Using K-Means Algorithm Integrated with Deep Learning Based-CAM Visualization Techniques. *Neural Comput. Appl.* **2022**, *34*, 17741–17759. [\[CrossRef\]](#)
43. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
44. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE Computer Society: Washington, DC, USA, 2016; Volume 2016, pp. 2818–2826.
45. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2016; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; Volume 2017, pp. 1800–1807.
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE Computer Society: Washington, DC, USA, 2016; Volume 2016, pp. 770–778.
47. Rajaraman, S.; Kim, I.; Antani, S.K. Detection and Visualization of Abnormality in Chest Radiographs Using Modality-Specific Convolutional Neural Network Ensembles. *PeerJ* **2020**, *2020*, e8693. [\[CrossRef\]](#)
48. Hacıfendioglu, K.; Demir, G.; Başağa, H.B. Landslide Detection Using Visualization Techniques for Deep Convolutional Neural Network Models. *Nat. Hazards* **2021**, *109*, 329–350. [\[CrossRef\]](#)

49. Hacıfendioğlu, K.; Başağa, H.B.; Ayas, S.; Karimi, M.T. Automatic Detection of Icing Wind Turbine Using Deep Learning Method. *Wind Struct.* **2022**, *34*, 511–523.
50. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.