buildings

MDPI

Article

An Integrated Method Using a Convolutional Autoencoder, Thresholding Techniques, and a Residual Network for Anomaly Detection on Heritage Roof Surfaces

Yongcheng Zhang [1], Liulin Kong [2], Maxwell Fordjour Antwi-Afari [3] and Qingzhi Zhang [2,*]

[1] Department of Construction Management, Nantong Institute of Technology, Nantong 226001, China; china.20246020@ntit.edu.cn
[2] Faculty of Engineering, China University of Geosciences, Wuhan 430070, China; kongliulin@cug.edu.cn
[3] College of Engineering and Physical Sciences, Aston University, Birmingham B4 7ET, UK; m.antwiafari@aston.ac.uk
* Correspondence: 4457581@gmail.com

Abstract: The roofs of heritage buildings are subject to long-term degradation, resulting in poor heat insulation, heat regulation, and water leakage prevention. Researchers have predominantly employed feature-based traditional machine learning methods or individual deep learning techniques for the detection of natural deterioration and human-made damage on the surfaces of heritage building roofs for preservation. Despite their success, balancing accuracy, efficiency, timeliness, and cost remains a challenge, hindering practical application. The paper proposes an integrated method that employs a convolutional autoencoder, thresholding techniques, and a residual network to automatically detect anomalies on heritage roof surfaces. Firstly, unmanned aerial vehicles (UAVs) were employed to collect the image data of the heritage building roofs. Subsequently, an artificial intelligence (AI)-based system was developed to detect, extract, and classify anomalies on heritage roof surfaces by integrating a convolutional autoencoder, threshold techniques, and residual networks (ResNets). A heritage building project was selected as a case study. The experiments demonstrate that the proposed approach improved the detection accuracy and efficiency when compared with a single detection method. The proposed method addresses certain limitations of existing approaches, especially the reliance on extensive data labeling. It is anticipated that this approach will provide a basis for the formulation of repair schemes and timely maintenance for preventive conservation, enhancing the actual benefits of heritage building restoration.

Keywords: heritage buildings; roof damage; detection and evaluation; UAV; computer vision

Citation: Zhang, Y.; Kong, L.; Antwi-Afari, M.F.; Zhang, Q. An Integrated Method Using a Convolutional Autoencoder, Thresholding Techniques, and a Residual Network for Anomaly Detection on Heritage Roof Surfaces. *Buildings* **2024**, *14*, 2828. https://doi.org/10.3390/buildings14092828

Academic Editor: Giuseppina Uva

Received: 27 July 2024
Revised: 31 August 2024
Accepted: 5 September 2024
Published: 8 September 2024

1. Introduction

Heritage buildings serve not only as historical artifacts but also as cultural icons. Each structure embodies the architectural techniques, artistic styles, and social characteristics of its respective era, encapsulating a wealth of historical and cultural information [1]. Preserving these buildings is vital to safeguarding historical knowledge and cultural heritage. As these structures increasingly attract global tourism [2], the urgency of effective preservation and restoration has heightened. However, many heritage buildings face risks from extreme weather events [3,4] such as hurricanes and blizzards, natural disasters like earthquakes and hailstorms, as well as challenges such as inadequate maintenance and deliberate damage. These factors present significant hurdles to the conservation of heritage buildings.

Roof issues are particularly prominent in the field of heritage building preservation. The roof serves not only as the primary protective layer of a structure but also plays a crucial role in maintaining the historical and esthetic value of the building. Traditional materials, often used in the roofs of heritage buildings, are durable but susceptible to

wear and damage over time due to prolonged exposure to the elements. Maintaining and repairing these roofs are essential for structural stability, but the diversity of materials and damage types requires highly skilled personnel with extensive technical knowledge. The unique spatial position of roofs within building structures complicates inspection, and the inability to simultaneously detect and repair damage often leads to incomplete hazard assessments and delayed repairs. Moreover, roof repairs in heritage buildings often rely on costly manual labor. These characteristics make roof inspection and maintenance a critical and challenging aspect of heritage building preservation. Efficient roof detection measures, coupled with timely repairs, can prevent further structural damage and ensure the long-term stability of these structures.

Tile roofing is widely adopted globally due to its benefits such as lightweight construction, excellent thermal insulation properties, and straightforward installation. For instance, tile roofing is prevalent in medieval and Renaissance architecture across France, Italy, and Germany, as well as in various historic complexes in China including palaces, temples, and residential buildings. Similarly, mosques and palaces in Iran and Turkey also extensively utilize tile roofing. However, tile roofs have drawbacks such as limited durability and susceptibility to damage, which increases maintenance requirements compared to other roofing materials. Additionally, the intricate structure of tile roofs, involving components such as mortar backing, mud backing, tile mud, and individual tile pieces, complicates manual inspection and reduces the efficiency and accuracy of damage detection.

Given the critical role of roofing issues in the preservation of heritage buildings and the widespread use of tile roofs, developing more efficient damage detection methods for these structures is crucial for advancing conservation technologies. In recent years, the integration of artificial intelligence (AI) has brought about significant advancements in this field. AI techniques, particularly machine vision combined with neural network technologies, simulate human visual functions. Deep learning algorithms, leveraging the capabilities of deep neural networks, excel at identifying patterns in data and performing inference tasks, making them ideal for practical applications in detection, measurement, and control.

Previous research has applied many innovative and effective methods to building preservation, including aerial diagnostic technologies like satellite imaging, GIS, and certain deep learning techniques [**?**] for the protection of civil buildings. However, these studies have not focused on roof types made of specific materials in historic buildings, and the deep learning methods used have been relatively isolated, lacking effective integration. For the first time, we have introduced deep learning techniques into the niche area of detecting and identifying damage in tiled roofs of historic buildings. Our convenient and efficient approach fills a significant research gap in this field and contributes to the advancement of historic building preservation.

This study employed unmanned aerial vehicles (UAVs) to capture image data of heritage building roofs and integrates deep learning-based visual algorithms with statistical methods to propose a streamlined and effective strategy for diagnosing anomalies in these structures. By leveraging convolutional autoencoders and thresholding methods, the approach enables efficient anomaly detection and extraction of damaged areas from images. Subsequently, convolutional neural networks were deployed to classify types of damage, facilitating the comprehensive assessments of tile roof conditions and the formulation of timely maintenance plans. By introducing AI-based machine vision algorithms to this domain, the study aims not only to enhance the efficiency and precision of damage detection but also to reduce the time and costs associated with manual inspection, thereby offering a more intelligent and practical approach to restoration efforts in heritage architecture.

2. Literature Review

In heritage building preservation, manual visual inspections conducted by construction and maintenance teams have traditionally been a primary method for assessing

structural health. However, these inspections are often time-consuming, subjective, and their quality can vary due to environmental constraints and human error.

Computer vision, emerging in the 1960s with techniques like edge detection [6], corner detection [7], and template matching [8], evolved significantly with advancements in image processing and hardware in the 1980s. Methods such as Support Vector Machines (SVMs) [9] and Histogram of Oriented Gradients (HOG) [10] improved visual system performance around the year 2000. The breakthrough of deep learning, highlighted by AlexNet [11]'s success in the 2012 ImageNet competition, revolutionized computer vision. convolutional neural networks (CNNs) like Visual Geometry Group (VGG) [12], GoogLeNet [13], and ResNet [14] excelled in image classification and feature extraction, while algorithms such as Single Shot MultiBox Detector (SSD) [15], Region-based Convolutional Neural Network (R-CNN) series [16–18], and You Only Look Once (YOLO) [19] advanced object detection speed and accuracy. Instance and semantic segmentation further enhanced capabilities with algorithms like Mask R-CNN [20] and U-Net [21], while Generative Adversarial Networks (GANs) [22] showed promise in image generation and enhancement. These advancements have driven computer vision applications across various fields including autonomous driving, medical imaging, and intelligent surveillance.

Computer vision technology, which has proven successful in diverse fields, is increasingly being applied to the preservation of heritage buildings. Unlike 3D point clouds and hyperspectral data requiring specialized equipment, RGB images offer the advantage of easy capture with portable devices. Consequently, current research predominantly focuses on 2D visual data. Scholars have applied algorithms to key computer vision tasks—image classification, object detection, and image segmentation—in various contexts within heritage building preservation.

In the realm of image classification, the primary aim is to identify and categorize surface cracks and damage on these structures. For instance, Chaiyasarn et al. [23] introduced a convolutional neural network-based algorithm for crack detection, pioneering deep learning applications in this field. Although achieving higher accuracy than traditional methods, its applicability remains constrained to specific crack types. Dais et al. [24] utilized convolutional neural networks and transfer learning to enhance efficiency in crack classification and segmentation on masonry surfaces.

In object detection, the emphasis lies on identifying and locating damage or cracks within buildings. Wang et al. [25] explored automatic damage detection in historic masonry buildings using mobile deep learning; however, devising a user-friendly mobile application that ensures consistency across various devices remains a challenge. Pathak et al. [26] proposed a method that integrates a 3D point cloud and 2D visual data for heritage damage detection, requiring substantial preprocessing and computational power due to the complexity of their algorithms. Mansuri and Patel [27] employed Faster R-CNN to detect defects in heritage buildings, such as cracks and exposed bricks, encountering difficulties in complex environments. Some scholars have applied YOLO series algorithms to detect and analyze surface damage in heritage buildings [28,29], either necessitating extensive data collection or requiring further validation for the method's applicability.

Image segmentation tasks involve the precise pixel-level delineation of damaged areas, which is crucial for quantitative assessment. Elhariri et al. [30] utilized the U-Net model for automatic pixel-level segmentation of historic surface cracks, showcasing efficiency and accuracy, with ongoing validation needed across different crack types. Makhanova et al. [31] introduced an enhanced deep residual network for crack detection in heritage buildings, maintaining high accuracy while reducing computational costs. Liu et al. [32] integrated semantic image segmentation and photogrammetry to monitor changes on heritage building facades to ensure accurate measurement of plant areas from diverse camera angles.

Additionally, some researchers have employed autoencoders in conjunction with anomaly detection to eliminate outliers by reconstructing normal class data. Deep learning applications in anomaly detection, based on autoencoders and Generative Adversarial

Networks (GANs), were reviewed by Pang et al. [33]. Advanced sensors have also been utilized by scholars to enhance data dimensions for improved detection outcomes. For example, Fais et al. [34] developed a non-destructive diagnostic approach for heritage building elements using 3D laser scanning and infrared thermography, achieving high accuracy and measurable results, despite challenges such as high equipment costs and limited generalizability of the method.

The concepts of digital twins and Heritage Building Information Modeling (HBIM) [35] are rapidly being implemented with the support of AI. Casillo et al. [36] proposed a method that uses IoT and deep learning for real-time monitoring and prevention of issues in cultural heritage buildings, helping to predict and prevent problems such as internal humidity. Pierdicca et al. [37] employed deep learning for the semantic segmentation of point cloud data to accelerate the modeling process of historic buildings. Ni et al. [38] introduced a framework utilizing digital twin technology and AI to optimize energy efficiency in historic buildings while preserving their structural and cultural integrity. These studies demonstrate the significant potential of AI technologies to enhance both the efficiency and accuracy of historic building preservation.

While existing research significantly contributes to the advancement of heritage building preservation, several challenges persist. Primarily, most current models rely on supervised learning paradigms. While neural networks excel in fitting large datasets and iteratively extracting features via gradient descent, the efficiency of manual annotation lags behind as data volumes and model complexities increase. Moreover, in the context of applying computer vision technology to heritage building preservation, the identified objects often exhibit shallow and anomalous features that are sporadic and highly diverse, posing challenges in constructing comprehensive databases necessary for effective training. Additionally, while integrating various deep learning methods or increasing image data dimensionality can enhance overall model performance, it also escalates training and deployment costs, limiting practical applicability in preservation efforts. Lastly, previous studies typically propose single-model solutions tailored to specific tasks, falling short of developing a holistic and efficient detection and evaluation system for systematic heritage building preservation.

To address these challenges, this study proposes a semi-supervised learning approach combining convolutional autoencoders, threshold methods, and ResNet for the detection and evaluation of tiled roofs. This method aims to offer a straightforward, efficient, and scalable solution for detecting and assessing heritage building roofs.

3. Methods

Figure 1 illustrates the comprehensive process of the tile damage recognition and evaluation system. As presented in Figure 1, UAV imagery data of tiled roofs from heritage buildings are captured along specific flight paths. These images are then fed into the tile damage recognition system. Based on quantifiable results provided by the recognition system, automatic diagnostics generate repair plans, facilitating simultaneous detection and restoration. This significantly enhances the convenience and timeliness of preservation efforts on heritage buildings.

For contemporary object detection or image segmentation models, the diverse damage types and sparse data of tiled roofs pose challenges for large-scale data annotation. Furthermore, the natural periodic arrangement of roof tiles allows neural networks to extract and reconstruct image features effectively, maintaining stable feature distribution before and after reconstruction. This enables efficient threshold-based segmentation of anomalous image blocks. Thus, we devised a multi-stage, integrated tile damage recognition system that employs convolutional autoencoders to reconstruct anomaly classes and utilizes threshold techniques based on statistical properties of reconstruction error distribution to extract anomalous image blocks. These blocks are subsequently classified using transfer learning to infer and diagnose damage types. By integrating various technologies, we have

developed a straightforward and efficient diagnostic system, implementing in Python and utilizing PyTorch as the deep learning framework.

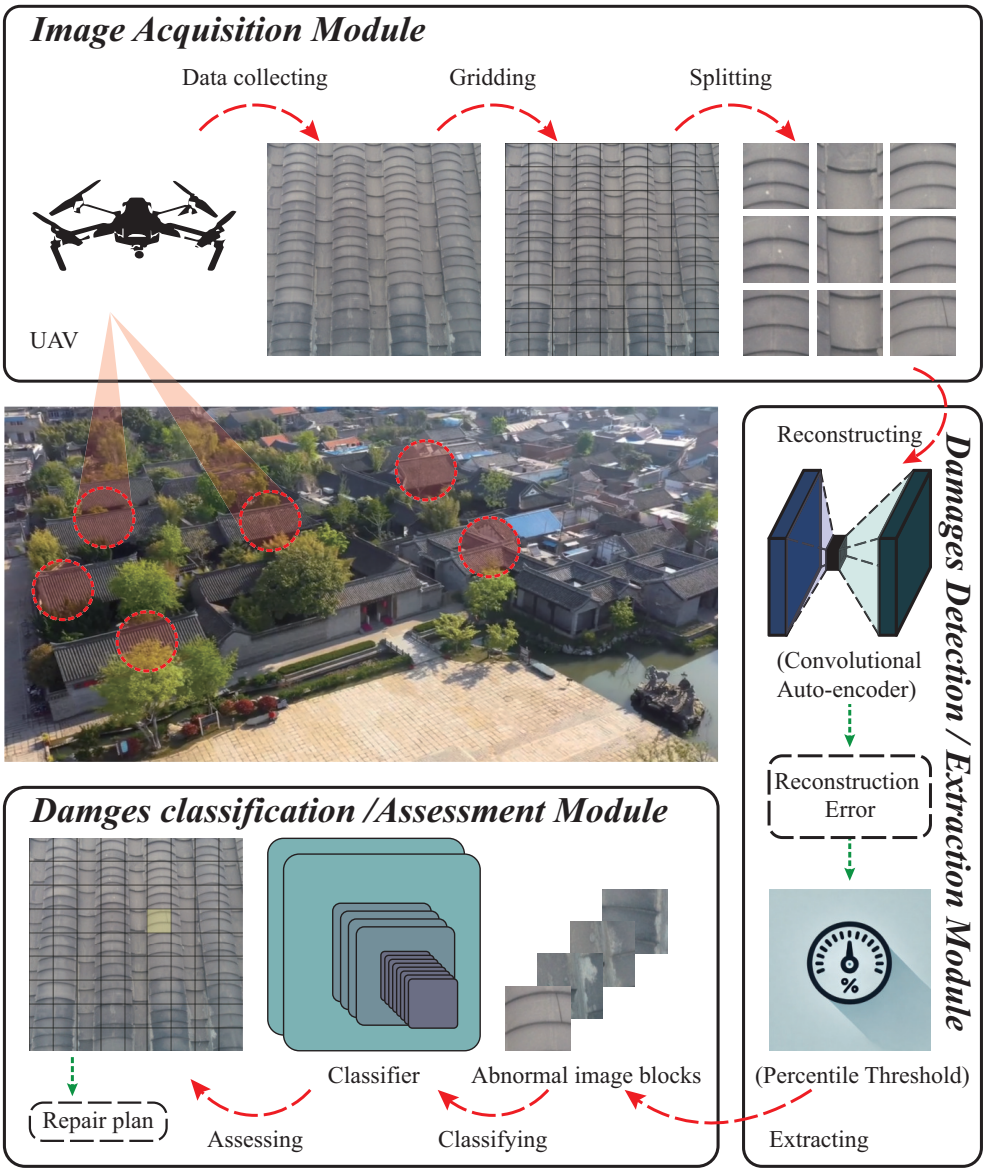

Figure 1. Damage recognition and evaluation system.

3.1. Image Anomaly Detection-Based Reconstruction

Anomaly detection constitutes a pivotal stage in identifying damaged tiles on the roofs of heritage buildings. Within computer vision, selecting an appropriate model tailored to the task type and object characteristics is critical for efficient detection. In the context of detecting surface damage on heritage building tiles, two primary considerations emerge.

Firstly, the tiles exhibit a distinct periodic pattern. Convolutional layers excel at extracting local features such as edges and textures, effectively capturing spatial characteristics of the input data. Convolution operations provide translation invariance, enabling shared weights of convolutional kernels to detect identical features at different positions within the input image. This architectural choice facilitates the learning of high-dimensional representations of tile arrangements, significantly reducing network parameters, lowering model complexity, conserving computational resources, and enhancing training efficiency.

Secondly, the diversity of surface anomalies on tiles and the scarcity of data for each anomaly type present significant challenges. Tile damage generally falls into two categories: detachment and cracking. Additionally, anomalies such as surface contamination and vegetation coverage exhibit even greater variability. In the long term, supervised learning paradigms necessitate extensive data collection and annotation across various anomalies, which proves impractical for widespread application. Drawing insights from manufacturing anomaly detection, autoencoders offer a more deployable and scalable solution. Autoencoders, a type of artificial neural network commonly employed for unsupervised learning and data dimensionality reduction, aim to reconstruct input data by learning low-dimensional representations. In this context, they effectively reconstruct normal tile surface images with minimal reconstruction error by using readily available normal data.

Considering the characteristics of tile surface damage detection, this study employed a convolutional autoencoder (CAE), which is designed and trained as an anomaly detector. The CAE framework utilized in this study is depicted in Figure 2, which constitutes an unsupervised learning framework built upon convolutional neural networks (CNNs), typically comprising an encoder and a decoder. The encoder compresses high-dimensional data into low-dimensional representations through a sequence of convolution and pooling operations, thereby extracting features from the input. Conversely, the decoder, mirroring the encoder, employs upsampling or transpose convolution operations to reconstruct the input data from the low-dimensional representations. When an autoencoder trained on normal data is given anomalous data as input, it typically yields a significant reconstruction error. By assessing the disparities between the input data and their reconstruction, anomalies can be promptly identified. In this study, Mean Squared Error (MSE) quantifies the reconstruction error, as expressed in Equation (1), where P_i denotes actual pixel values of the image, $\widehat{P}_i$ represents reconstructed pixel values by the decoder, and n signifies the total number of pixels in the image. The Adam optimizer, renowned for its efficient handling of sparse gradients and reduced sensitivity to raw data, is selected as the optimization algorithm.

$$MSE = \frac{1}{2n}\sum_{i=1}^{n}\left(\widehat{P}_i - P_i\right)^2 \quad (1)$$

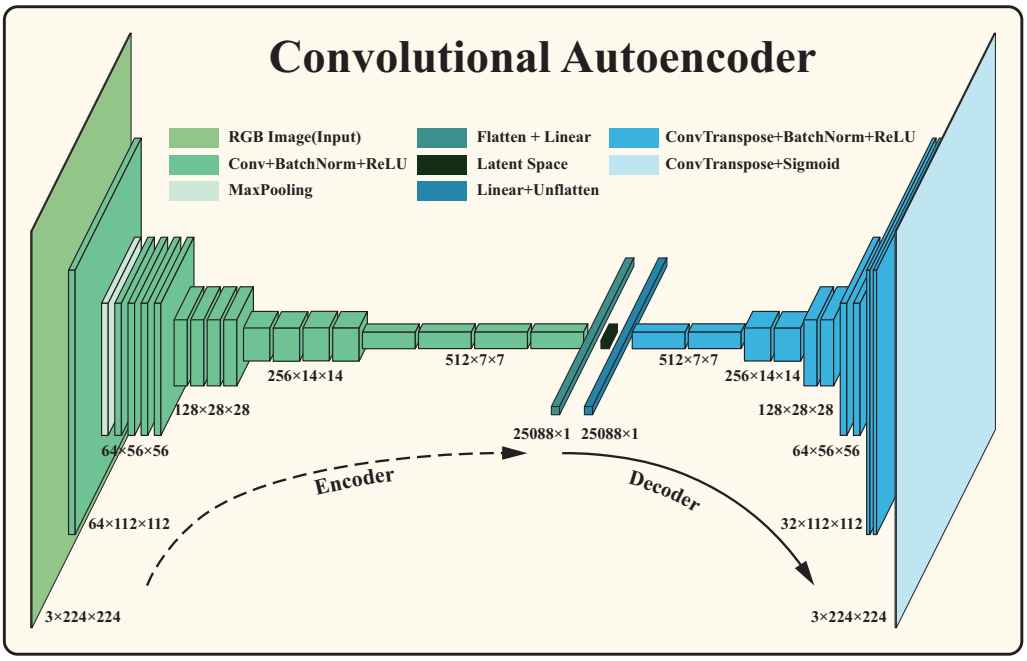

Figure 2. Convolutional autoencoder architecture.

3.2. Image Anomaly Extraction

Anomaly extraction constitutes a pivotal component of anomaly detection. Serving as a crucial component in our proposed evaluation system, we employed a global thresholding technique based on the percentile method to filter anomalies in the MSE of reconstructed image blocks.

Various threshold selection methods possess distinct advantages and limitations contingent upon the application context. The selection of the optimal method typically hinges on the specific task requirements, data characteristics, and desired outcomes. Common approaches encompass the mean and standard deviation method, Receiver Operating Characteristic (ROC) curve method, Precision–Recall (PR) curve method, cross-validation using a validation set, and adaptive methods. These methodologies may necessitate a known distribution of reconstruction errors, a validated set comprising positive and negative samples, or may involve substantial computational overhead.

The percentile method stood out for its efficiency in our evaluation system. Firstly, it operates effectively without requiring prior assumptions about the MSE distribution, making it adaptable to diverse data distributions. Secondly, in terms of computational complexity, the percentile method is straightforward, requiring only sorting and selecting a fixed percentile without iterative searches for an optimal threshold. Lastly, while the percentile method may not unfailingly eliminate all anomalies, our system's subsequent classification model effectively discerns normal images, thereby minimizing uncertainty associated with fixed percentile thresholds. The calculation formula for the percentile method is shown in Equation (2).

$$P = \frac{p}{100} \times (N + 1), \tag{2}$$

where P denotes the desired position (e.g., 25 for the 25th percentile and 95 for the 95th percentile), and N is the total number of data points in the dataset.

The sequence of reconstructed values followed a defined procedure. Initially, images of roof tiles captured by the UAV were resized using bilinear interpolation to ensure that they were multiples of 224, aligning with the input size required by the convolutional autoencoder (as illustrated in Figure 3). Subsequently, these images were partitioned into multiple 224 × 224 image blocks using a grid, and each block was fed into the trained convolutional autoencoder. The resulting reconstruction error for each image block within the original image was then calculated and recorded. The pixel value sequence of the original image block was denoted as $\{B_{\mathrm{k}}(x, y, z)\}$ and the sequence of the reconstructed image block was denoted as $\left\{\widehat{B_{\mathrm{k}}}(x, y, z)\right\}$. The reconstruction error for a single image block is defined as shown in Equation (3).

$$MSE_{\mathrm{k}} = \frac{1}{2mnl} \sum_{x=0}^{m} \sum_{y=0}^{n} \sum_{z=0}^{l} \left[B_{\mathrm{k}}(x, y, z) - \widehat{B_{\mathrm{k}}}(x, y, z)\right]^2, \tag{3}$$

where k represents the block number after the input image is segmented, and (x, y, z) denotes the coordinates of a pixel within the image block, $m = n = 224$ are the dimensions of the image block, and $l = 3$ is the number of channels in the image.

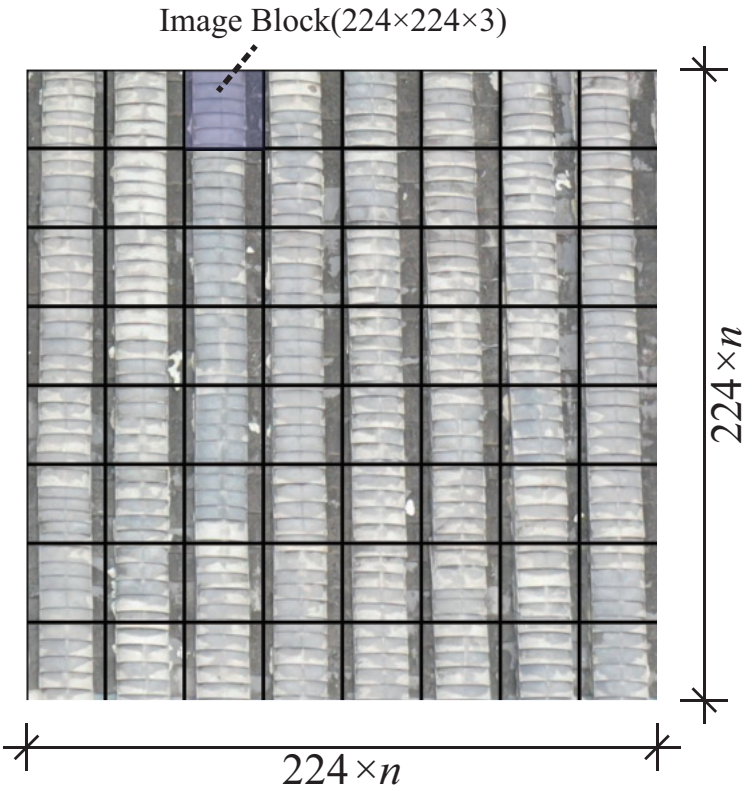

Figure 3. Grid illustration of image segmentation (224 × 224 × 3 blocks).

3.3. Damage Classification

3.3.1. Selection of Backbone Network

Damage classification plays a crucial role in the overall recognition and evaluation system. A high-performance classification model not only supports the anomaly detection and extraction module, which is based on image reconstruction and thresholding techniques, but also forms the basis for accurate evaluation within the system. Given potential deployment in resource-constrained environments such as mobile devices and embedded systems, lightweight networks serve as the backbone for damage classification.

According to research [39], there exists a positive correlation between the number of parameters and the consumption of computational resources. This study employs lightweight networks, including SqueezeNet, ShuffleNetV2, MobileNetV2, EfficientNet-B0, and ResNet-18, as backbone networks for damage classification. Table 1 provides the parameter counts and outlines the advantages of these networks. As shown in Table 1, SqueezeNet reduces the number of parameters through its fire module design, while maintaining high accuracy. ShuffleNetV2 enhances efficiency by using grouped convolutions and channel shuffle techniques. MobileNetV2 reduces parameters and complexity through an inverted residual structure and linear bottleneck, making it ideal for mobile and embedded systems. EfficientNet-B0 achieves a balance between performance and efficiency by simultaneously scaling depth, width, and resolution. ResNet-18, a lighter ResNet version, mitigates the vanishing gradient problem in deep networks via residual connections, effectively capturing deep features.

Table 1. Comparison of neural network models by parameter count and advantages.

Model	Parameters (M) [1]	Advantages
SqueezeNet [40]	1.2 M	Extremely low parameter count, comparable performance to AlexNet
ShuffleNetV2 [41]	2.3 M	Grouped convolutions and channel shuffle, highly efficient for mobile devices.
MobileNetV2 [42]	3.4 M	Efficient depthwise separable convolutions, suitable for mobile devices.
EfficientNet-B0 [43]	5.3 M	Achieves a balance of performance and efficiency through compound scaling
ResNet-18 [14]	11.7 M	A lighter variant of ResNet, suitable for resource-constrained scenarios

[1] The parameter count (M) refers to the number of parameters in millions (10^6).

3.3.2. Evaluation Metrics

Four commonly used classification metrics were employed to evaluate the performance of model. Accuracy represents the proportion of correctly predicted samples to the total number of samples, measuring overall model correctness, and is particularly effective for balanced datasets. Precision denotes the proportion of correctly predicted positive samples to the total number of samples predicted as positive. Recall signifies the proportion of correctly predicted positive samples to all actual positive samples. The $F1$ score, calculated as the harmonic mean of precision and recall, synthesizes these metrics into a single measure of performance. Together, these metrics offer a comprehensive and adaptable evaluation of the classification model's effectiveness. The Equations (4)–(7) for these metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \tag{7}$$

where TP (True Positive) represents correctly identified positive instances, FP (False Positive) denotes incorrectly identified positive instances, FN (False Negative) signifies incorrectly identified negative instances, and TN (True Negative) indicates correctly identified negative instances.

3.4. Damage Evaluation Strategy and Maintenance Plan Formulation

Tile damage on roofs can result from several factors. Thermal cracking occurs due to temperature fluctuations, which cause tile expansion and contraction. Foot traffic can damage tiles through direct pressure, while strong winds may lift or displace them. Hail impacts can lead to tile breakage, and long-term UV exposure can cause aging, discoloration, and brittleness, eventually resulting in cracking.

Based on the severity of damage impact on functionality and classification rationale, tile conditions are categorized into three states: undamaged, cracked, and detached, as illustrated in Figure 4. Detachment represents the final stage of cracking and has more complex causes and different functional impacts on buildings compared to cracks. Cracked tiles, even while remaining in place, significantly reduce waterproofing capabilities, potentially leading to leaks and further damage. Detached tiles expose the waterproofing layer directly, relying solely on tile protection. Prolonged exposure to elements can significantly

degrade waterproofing performance or even cause structural damage. Moreover, detached tiles can destabilize adjacent tiles and pose a risk of falling, creating significant hazards.

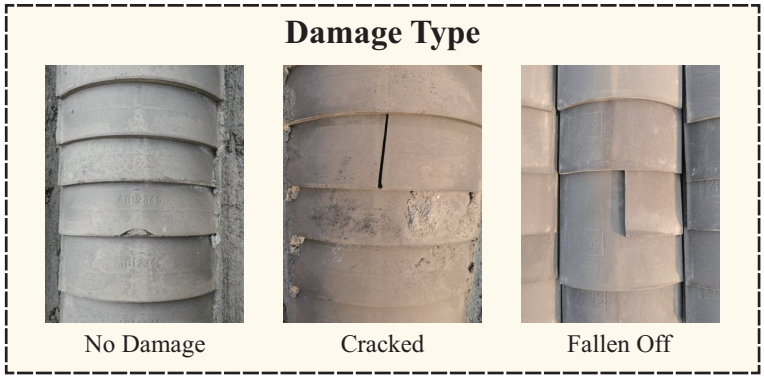

Figure 4. Classification of roof tile damage types.

This study defines the Damage Index (DI) to comprehensively assess the degree of tile damage. The DI values provide scientific guidance for repair prioritization. The formula is as follows (8):

$$\mathrm{DI} = \frac{1}{N}(\alpha D_{\mathrm{c}} + \beta D_{\mathrm{f}}) \tag{8}$$

where N is the number of image blocks divided from a single input image, D_{c} and D_{f} are the numbers of image blocks classified as cracked and fallen, respectively. α and β are dimensionless constants that serve as the importance coefficients for the respective damage types. Considering the impact range of damage types and the consequences of delayed handling, this study set $\alpha = 1$ and $\beta = 2$ to reflect the severity of various damage types.

Once the classification task was completed by the classifier, the damage assessment module overlaid the classification results on the original input image, allowing maintenance personnel to swiftly pinpoint damaged areas. To enhance the visibility of these areas, long-wavelength colors, such as red and yellow, were utilized to highlight fallen and cracked regions, respectively, while undamaged areas remained unmarked. Based on the types of damage identified in specific roof sections, targeted repair strategies were developed, as detailed in Table 2.

Table 2. Repair schemes for different types of roof tile damage.

Damage Type	No Damage	Cracked	Fallen Off
Repair Schemes	Always prioritize safety precautions	Seal cracks tightly with hemp mortar, level them, and then apply a layer of joint ridge mortar along both sides of the tile ridge. Finally, use a water trowel and brush to smooth the mortar.	Remove peeling and loose tiles, clean the base thoroughly, bind copper wires to the steel mesh, securely tie the tiles, and apply high-grade mortar for paving. Lastly, ensure proper cleanup and disposal of debris.

4. Roof Tile Restoration Case Study in Hexia Ancient Town

4.1. Roof Tile Image Data Collection in Hexia Ancient Town

When planning the UAV flight path for roof tile image collection, balancing cost and efficiency is essential. This balancing was achieved by carefully designing the flight route to select representative roof tiles, ensuring data completeness while effectively controlling costs. In this study, the roof tiles of five buildings in Hexia Ancient Town were chosen as the primary targets for aerial photography.

The flight path is illustrated in Figure 5. As shown in Figure 5, the UAV commenced image capture from the east side of Building 1, starting with the third-floor tiles and systematically covering those on the third and fourth floors in the east, southeast, west, and north directions. After completing the imaging of Building 1, the UAV proceeded to systematically capture images from Buildings 2 to 5, as illustrated by the arrows. Due to the varying combinations of target objects and equipment, the flight parameters and imaging details had to be adjusted based on the actual conditions and experience. These adjustments are beyond the primary scope of this paper. However, to provide a reference, some parameters based on the equipment used in this study are presented. The UAV utilized in this study was equipped with a 12-megapixel visual sensor and a 14.66 mm focal length, which enabled it to provide a wide field of view and sufficient image detail at lower flight altitudes. The flight altitude was approximately 15 m, with a flight speed of about 1 m per second. This configuration ensured image clarity and detail capture, minimizing image blurring and instability during flight.

Figure 5. Schematic diagram of UAV flight path.

4.2. Image Reconstruction Based on Convolutional Autoencoder

4.2.1. Data Collection and Augmentation

In this study, UAVs captured a total of 120 RGB images of roof tiles over a large area. The images were collected at different times, under varying lighting conditions and shooting angles, to enable the model to learn a broader range of features and achieve better robustness. After manually removing images with damage or other anomalies, the remaining images were segmented into 224 × 224 × 3 blocks using Python's OpenCV library, resulting in 9842 image blocks. These blocks were augmented through rotations and flips, producing a total of 44,800 images (as shown in Figure 6). The normalized images were subsequently fed into the convolutional autoencoder.

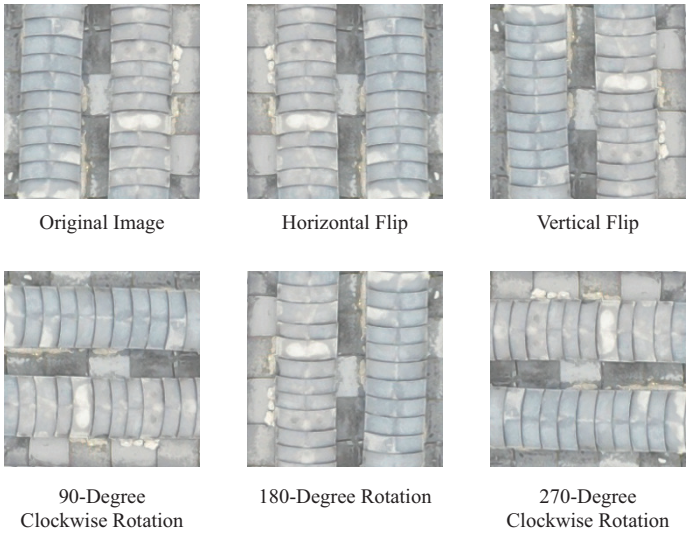

Figure 6. Overview of image augmentation.

4.2.2. Autoencoder Architecture and Training Strategy

The architecture of the convolutional autoencoder is depicted in Figure 2. Inspired by the feature extraction prowess of the VGG architecture, which utilizes block designs, our encoder incorporates similar blocks. Each block comprises a 3×3 convolutional layer, followed by a ReLU activation layer, and a 2×2 non-overlapping max pooling layer. By stacking four such blocks, the input is compressed from 224×224 to 7×7 feature maps, which are then flattened and projected into the latent space via a fully connected layer. The decoder mirrors the encoder's architecture, employing transpose convolutions for upsampling. Batch normalization is applied to ensure a consistent output distribution and improve training stability. The model was trained with a learning rate of 0.001, a batch size of 64, over 2000 epochs.

4.2.3. Training Results

The reconstruction results of the trained convolutional autoencoder for normal and anomalous images are depicted in Figure 7. All reconstructed images exhibit the typical features of the normal class, with red boxes highlighting the primary sources of reconstruction error, which are distinguishable through thresholding. Surface noise, such as cement dust, is effectively eliminated, showcasing the autoencoder's capability in image denoising. Although the anomalous images also exhibit some normal features in their reconstruction, a noticeable distortion in feature distribution leads to increased reconstruction errors, validating the effectiveness of the percentile-based thresholding technique.

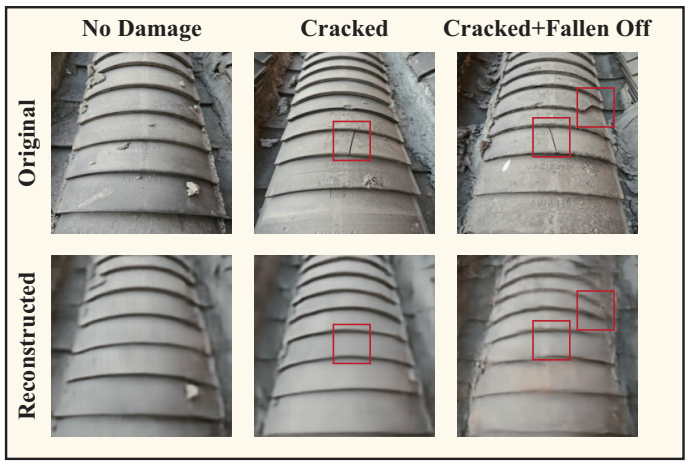

Figure 7. Demonstration of autoencoder reconstruction performance.

4.3. Anomalous Image Block Extraction Based on Thresholding Technique

A percentile-based thresholding technique was employed to extract anomalous blocks by analyzing the distribution of reconstruction error data. To ensure the appropriateness of the threshold setting, 10 randomly selected images of roof tiles captured by the UAV were utilized. These images were segmented into 224 × 224 blocks and fed into the trained convolutional autoencoder. The reconstruction errors obtained are depicted in Figure 8. It is observed that the majority of reconstruction errors for the image blocks fall within the range of 0.05 to 0.10, supporting the effectiveness of thresholding for anomaly detection.

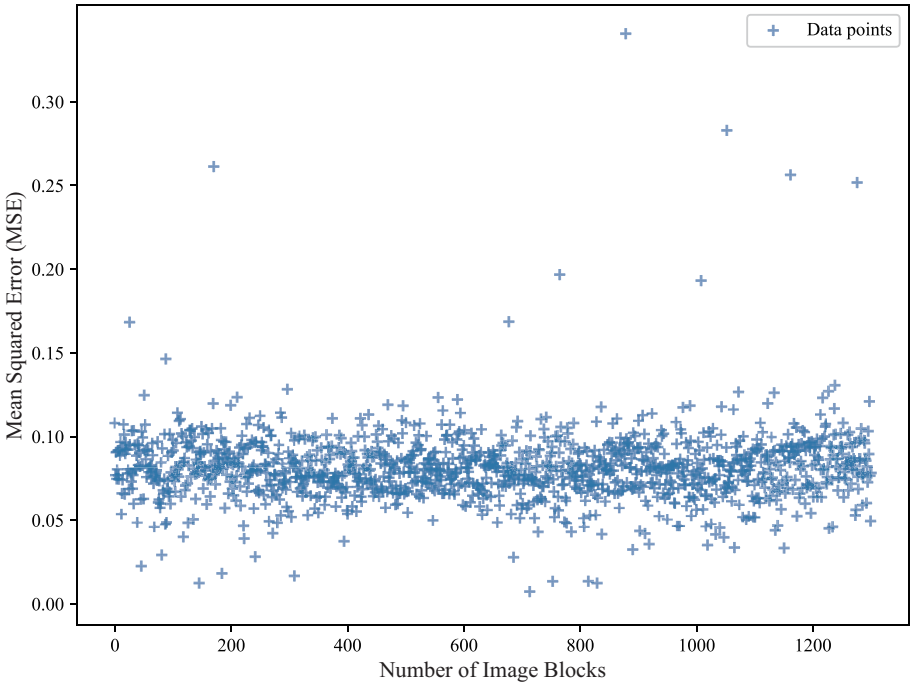

Figure 8. Scatter plot MSE distribution across image blocks.

The Central Limit Theorem states that the sum of a sufficiently large number of independent and identically distributed random variables tends to follow a normal distribution. In the context of an autoencoder model, if the reconstruction error of each input sample is considered as the aggregation of numerous independent small errors, the cumulative effect of these errors may follow a normal distribution. Assuming that the autoencoder accurately learns the underlying data distribution during training, and that its errors are unbiased and consistent, these errors may closely approximate a normal distribution. This study observes that a substantial portion of reconstruction errors in image block reconstructions exhibit characteristics akin to a normal distribution, reinforcing the efficacy of the autoencoder model's training.

After sorting of reconstruction errors from smallest to largest, the corresponding original images at selected percentiles (denoted as p) are displayed in Figure 9. Notably, from the 84th percentile onwards, image blocks manifest signs of tile damage. Additionally, some blocks devoid of abnormalities exhibit high reconstruction errors due to noise. To ensure the stable extraction of abnormal image blocks using the threshold, this study set the percentile at 80, thereby incorporating a heightened safety margin through judiciously lowered thresholds.

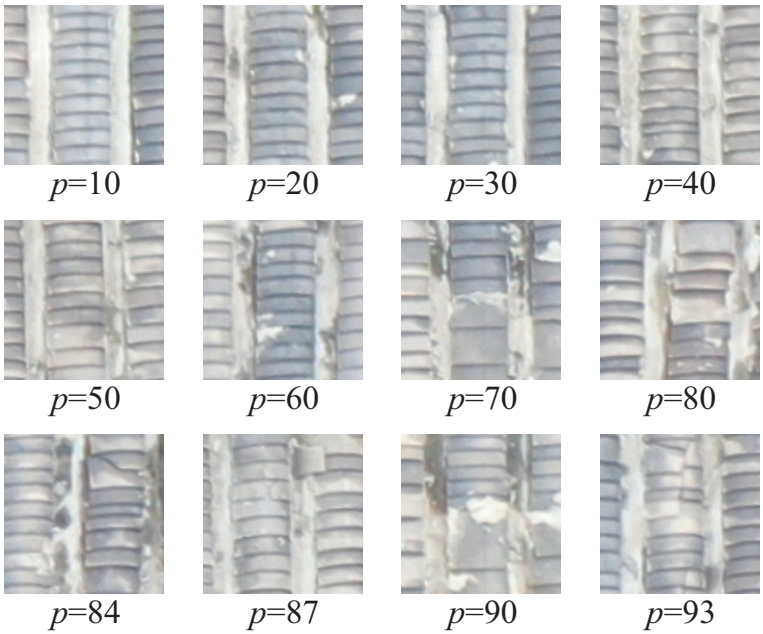

Figure 9. Original images selected at different percentiles based on reconstruction error.

4.4. Construction of Damage Classification Model

4.4.1. Dataset Partitioning

The images in the constructed dataset were categorized into three types: undamaged, cracked, and detached. The dataset contained a total of 349 defective samples, evenly distributed between cracked and detached tiles. To maintain balance, the number of undamaged tile images was restricted to 175 samples. Given the modest size of the dataset, k-fold cross-validation was employed to maximize data utilization.

Figure 10 illustrates the dataset's division into k subsets. In each iteration, k-1 subsets were utilized for training, while the remaining subset served as the validation set. This process iterated k times, ensuring that each subset served as the validation set once. The average accuracy across validation sets served as the evaluation metric, negating the necessity for a separate test set. In this study, k was set to 5. Model convergence was determined if validation loss did not decrease over 30 consecutive epochs. The model exhibiting the minimum validation loss was identified as the optimal model for this study.

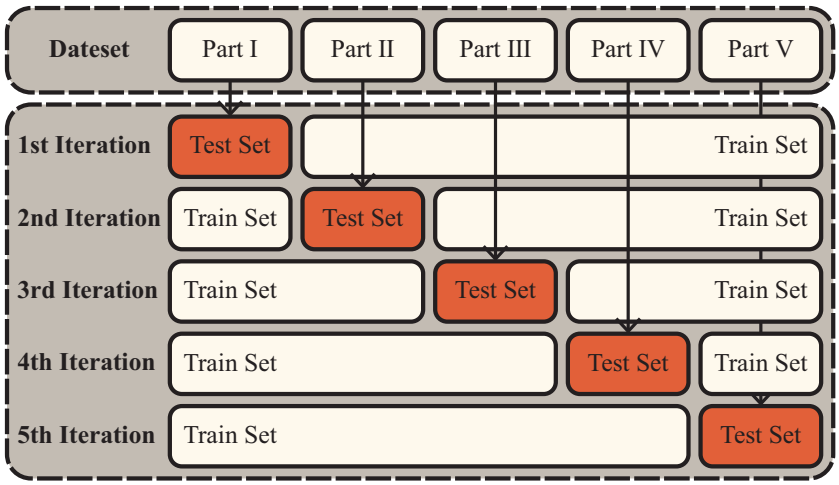

Figure 10. K-fold cross-validation.

4.4.2. Model Training Configuration

Given the task's limited number of defect samples, this study employed transfer learning to enhance model performance. Transfer learning, a machine learning technique, boosts learning efficiency and effectiveness in new tasks by leveraging features learned from extensive datasets like ImageNet. This approach enables the reuse of common image features and structural hierarchies across smaller datasets. To tailor the model to the specific task, we fine-tuned the final layers to match the image categories. Transfer learning efficiently leverages prior knowledge, reduces computational demands compared to training from scratch, accelerates model convergence, improve performance on new tasks, and mitigates overfitting risks. Table 3 details the hyperparameter settings for each model type.

Table 3. Model hyperparameter settings.

Model	Learning Rate	Batch Size	Epochs
SqueezeNet	0.001	32	300
ShuffleNetV2	0.01	32	400
MobileNetV2	0.001	32	200
EfficientNet-B0	0.001	32	400
ResNet-18	0.001	32	200

In this study, the image damage classification model was trained using the Adam optimization method. Adam is renowned for its rapid iteration speed and robustness, facilitating efficient convergence of the training process. The loss function employed was categorical cross-entropy (CCE), defined as follows (9):

$$CCE = -\sum_{l=0}^{2} y_l \log(p_l) \tag{9}$$

4.4.3. Training Process

Through hyperparameter selection and comparison, the performance of models under optimal settings is illustrated in Figure 11. All five pre-trained models converge swiftly on the dataset, achieving consistent accuracy. This study integrated standard model selection criteria, evaluating performance across training and inference stages. During training, the emphasis was placed on convergence speed and model stability.

ShuffleNet (b), leveraging grouped convolutions and channel shuffle strategies, offers high computational efficiency but exhibits slower convergence with higher learning rates across early, middle, and late training stages. EfficientNet-B0 (d) and SqueezeNet (a) demonstrate comparable performance during training, each balancing model stability and convergence speed. Their lightweight architectures, however, compromise feature extraction, resulting in variable loss curves.

In contrast, MobileNetV2's (c) depthwise separable convolutions and ResNet18's (e) residual connections achieve rapid and stable iteration during early training stages. This is evident from the swift decline in training loss within initial epochs, followed by stable oscillations. MobileNetV2 and ResNet18 are selected as candidate models, and further evaluated for classification performance.

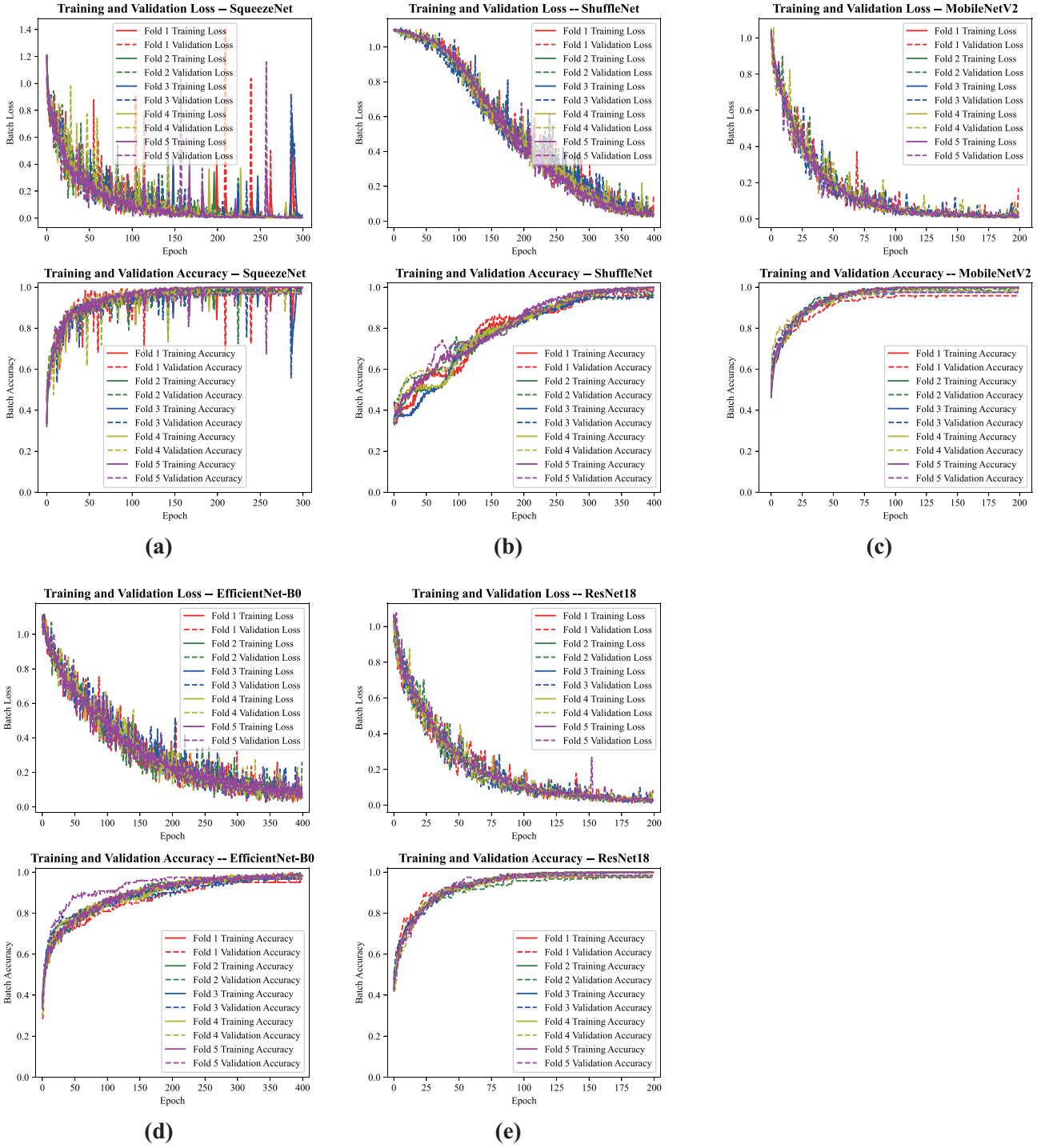

Figure 11. Training and validation performance of five classification models: loss and accuracy curves.

4.4.4. Model Performance Evaluation

In this experiment, the performance of five mainstream models (SqueezeNet, ShuffleNet, MobileNetV2, EfficientNet-B0, and ResNet18) was compared on image classification tasks. These models were evaluated using four metrics: accuracy, precision, recall, and F1 score. Table 4 presents each metric as "mean ± standard deviation", where the mean represents the average across multiple experimental results, and the standard deviation indicates result variability.

Table 4. Comparative performance analysis of deep learning models for image classification.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SqueezeNet	97.7 ± 1.8	97.6 ± 1.9	97.6 ± 2.1	97.6 ± 2.0
ShuffleNet	96.5 ± 3.0	96.6 ± 2.9	96.3 ± 2.5	96.4 ± 2.7
MobileNetV2	97.8 ± 3.0	97.9 ± 2.9	97.6 ± 3.3	97.7 ± 3.1
EfficientNet-B0	96.8 ± 2.5	97.0 ± 2.2	96.5 ± 3.0	96.7 ± 2.7
ResNet18	$\mathbf{98.0 \pm 0.9}$	$\mathbf{98.1 \pm 0.9}$	$\mathbf{98.0 \pm 1.1}$	$\mathbf{98.0 \pm 1.0}$

* The bolded ResNet18 model in the table demonstrates optimal performance across all performance metrics.

The experimental results highlight ResNet18's exceptional performance across all evaluation metrics. Specifically, ResNet18 achieves an accuracy of 98.0 ± 0.9 %, precision of 98.1 ± 0.9%, recall of 98.0 ± 1.1 %, and F1 score of 98.0 ± 1.0 %. These findings not only underscore ResNet18's superior average performance and high stability, evidenced by its smallest standard deviation among all models, but also its minimal result fluctuation and robustness.

Although other models demonstrate comparable performance in certain metrics, they generally exhibit lower overall performance and stability. For instance, MobileNetV2 achieves an accuracy of 97.8 ± 3.0 %, indicating significant result variability compared to ResNet18 despite its higher average value.

Therefore, based on both average performance and result stability, ResNet18 is defined as the optimal choice. Its deep network structure and effective residual learning mechanism contribute to its enhanced representation and generalization capabilities for handling complex image features. In conclusion, after detailed analysis of experimental data, ResNet18 is unequivocally selected as the superior model.

4.5. Roof Damage Assessment Case

Different types of damage on the original image are delineated in various colors by the damage assessment module. Integration with location ID data from the collection points of original images enables swift localization of roof defects. By evaluating quantifiable damage indices (DI values), maintenance personnel can efficiently prioritize repair sequences. Furthermore, incorporating tailored repair strategies for each damage type allows for the proactive preparation of necessary tools and materials. Figure 12 illustrates clearly marked areas corresponding to different damage types, with DI values indicating higher urgency for repairing areas 2-17 compared to areas 1-14. Adjustments in material ratios and repair plans, informed by the number and types of damaged areas, along with construction experience, ensure the expedient completion of repair tasks.

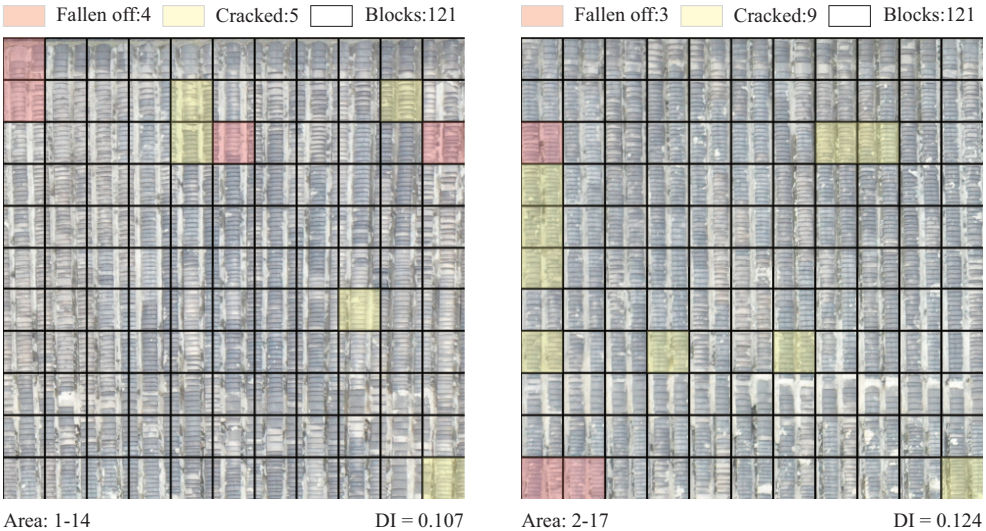

Figure 12. Roof damage assessment case study in heritage structures.

5. Discussion

Compared to traditional image segmentation or object detection methods, significant advantages are offered by the proposed method in terms of data requirements, model complexity, and scalability, especially within the specific application scenario of detecting damage to heritage building roof tiles. Specifically, the semi-supervised architecture based on convolutional autoencoders reduces reliance on extensive datasets, particularly excelling in scenarios with limited anomalous data. Whereas traditional methods like U-Net and YOLO typically demand substantial annotated datasets for training, autoencoders can achieve training proficiency using unannotated normal data alone.

Additionally, the combination of autoencoders and thresholding yields a lightweight model that circumvents the need for intricate segmentation or detection networks. This allows for swift training and inference even in resource-constrained environments such as UAVs or mobile devices, unlike traditional vision methods that often require more computational resources and longer training periods [44]. Moreover, the simplicity of the autoencoder and thresholding techniques facilitates straightforward deployment and scalability, enabling rapid integration into existing systems. In contrast, traditional vision methods pose greater challenges in deployment and maintenance. Finally, the interconnected techniques employed in this detection and evaluation model bolster its robustness.

However, this study harbors certain limitations. Primarily, the data diversity is constrained as the image dataset primarily originates from Hexia Ancient Town, potentially influencing the model's generalization capabilities due to the limited variety of tile types. Secondly, the study does not encompass a fully end-to-end application system and still requires manual intervention for threshold configuration, limiting system automation. Furthermore, the evaluation methods remain qualitative and lack precise quantitative assessments. To address these limitations, future research could be enhanced and expanded in several directions:

- **Broaden data collection scope:** encompass a wider range of heritage building tile types to enhance the model's generalization.
- **Extend data dimensions:** introduce multi-sensor data like 3D point clouds and infrared imaging, employing multi-source data fusion strategies to further refine the evaluation system's quantifiability.
- **Develop a comprehensive end-to-end application system:** formulate a system integrating data collection, processing, damage assessment, and maintenance recommendations to enhance automation and practicality.
- **Promote methodology in other preservation applications:** apply this approach to additional facets of heritage building preservation and explore its viability in tasks such as bridge and road inspections.

6. Conclusions

This paper proposed a comprehensive solution for identifying and assessing defects in tiled roofs, specifically for heritage building preservation. The solution effectively addresses the challenges associated with tiled roof inspections. By combining convolutional autoencoders, thresholding techniques, and ResNet neural networks, we developed a semi-automated, cost-effective, and efficient system for defect detection and assessment. This system utilizes UAVs equipped with visual sensors for data collection and adopts a semi-supervised learning paradigm to minimize the need for extensive data annotation. Our approach fills a gap in the application of artificial intelligence technologies for the inspection of tiled roofs in heritage building preservation.

However, several limitations still exist. Since the system is designed for a specific task, it currently cannot integrate an automated process to recognize a broader range of defect types. Unrecognized defects may impact the accuracy of assessment results from a different perspective, despite the high performance of our classification model. Additionally, the system cannot infer potential invisible defects, such as predicting future damage types based on latent features or combining meteorological data to forecast damage occurrence.

Expanding the detection model's cognitive abilities to address these advanced issues in artificial intelligence is a worthwhile avenue for future exploration.

Furthermore, to maintain the efficiency of the assessment system and keep computational resource consumption within reasonable limits, a trade-off was made by allowing a certain degree of precision to be compromised. We only qualitatively analyzed the degree of damage to different roofs, while some more expensive models can achieve pixel-level segmentation of cracks and, when combined with photogrammetry, obtain real-world crack dimensions for the quantitative analysis of damage severity and progression stages. Achieving these objectives within controlled cost parameters will be a focus of future research.

Moreover, the primary advantage of using UAVs with traditional 2D visual sensors is their low cost. However, this approach may not be optimal for all detection tasks. For instance, satellite imaging combined with a Geographic Information System (GIS) could be employed for long-range tasks to assess roof damage across large areas, or a versatile detection model capable of accommodating various task types could be developed. Additionally, radar or LiDAR sensors offer superior capabilities for capturing detailed damage features. Although higher-dimensional data increase computational burdens at each stage, these sensors play a dominant role in integration with HBIM (Historic Building Information Modeling) or digital twin technologies. Promisingly, recent advances in machine vision methods, such as Neural Radiance Fields (NeRFs), 3D Gaussian Fields, and their variants, are overcoming these barriers, advancing the integration of machine vision technology within HBIM and digital twins.

Overall, our research has practical applications and has shown promising results in detecting tiled roof damage in Hexia Ancient Town, although significant limitations remain. Nevertheless, from another perspective, our model can integrate new technologies and concepts, potentially achieving higher levels of automation and intelligence. This research area holds substantial potential for development and further study, and we will continue to advance our work in the field of heritage building preservation.

Author Contributions: Conceptualization, Y.Z. and Q.Z.; methodology, Y.Z.; software, Q.Z.; validation, M.F.A.-A. and L.K.; formal analysis, Q.Z.; investigation, L.K.; resources, Y.Z.; data curation, Q.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, L.K. and M.F.A.-A.; visualization, Q.Z.; supervision, Q.Z.; project administration, L.K.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the "National Natural Science Foundation of China (NNSFC)" under Grant number 72301256.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Taher Tolou Del, M.S.; Saleh Sedghpour, B.; Kamali Tabrizi, S. The semantic conservation of architectural heritage: The missing values. *Herit. Sci.* **2020**, *8*, 70. [CrossRef]
2. Kostopoulou, S. Architectural heritage and tourism development in urban neighborhoods: The case of upper city, Thessaloniki, Greece. In *Conservation of Architectural Heritage*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 139–152.
3. Li, Y.; Du, Y.; Yang, M.; Liang, J.; Bai, H.; Li, R.; Law, A. A review of the tools and techniques used in the digital preservation of architectural heritage within disaster cycles. *Herit. Sci.* **2023**, *11*, 199. [CrossRef]
4. Aboulnaga, M.; Abouaiana, A.; Puma, P.; Elsharkawy, M.; Farid, M.; Gamal, S.; Lucchi, E. Climate Change and Cultural Heritage: A Global Mapping of the UNESCO Thematic Indicators in Conjunction with Advanced Technologies for Cultural Sustainability. *Sustainability* **2024**, *16*, 4650. [CrossRef]
5. Braik, A.M.; Koliou, M. Automated building damage assessment and large-scale mapping by integrating satellite imagery, GIS, and deep learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2024**, *39*, 2389–2404. [CrossRef]
6. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698. [CrossRef]
7. Harris, C.; Stephens, M. A combined corner and edge detector. In *Alvey Vision Conference*; The Plessey Company: London, UK, 1988.

8. Briechle, K.; Hanebeck, U.D. Template matching using fast normalized cross correlation. In Proceedings of the Optical Pattern Recognition XII, Orlando, FL, USA, 16–17 April 2001; Volume 4387, pp. 95–102.
9. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
10. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
12. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
13. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
16. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
17. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
18. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef]
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
21. 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; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015, pp. 234–241.
22. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
23. Chaiyasarn, K.; Sharma, M.; Ali, L.; Khan, W.; Poovarodom, N. Crack detection in historical structures based on convolutional neural network. *GEOMATE J.* **2018**, *15*, 240–251. [CrossRef]
24. Dais, D.; Bal, I.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]
25. 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]
26. 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]
27. Mansuri, L.E.; Patel, D. Artificial intelligence-based automatic visual inspection system for built heritage. *Smart Sustain. Built Environ.* **2022**, *11*, 622–646. [CrossRef]
28. Karimi, N.; Mishra, M.; Lourenço, P.B. Deep learning-based automated tile defect detection system for Portuguese cultural heritage buildings. *J. Cult. Herit.* **2024**, *68*, 86–98. [CrossRef]
29. Yan, L.; Chen, Y.; Zheng, L.; Zhang, Y. Application of computer vision technology in surface damage detection and analysis of shedthin tiles in China: A case study of the classical gardens of Suzhou. *Herit. Sci.* **2024**, *12*, 72. [CrossRef]
30. Elhariri, E.; El-Bendary, N.; Taie, S.A. Automated pixel-level deep crack segmentation on historical surfaces using U-Net models. *Algorithms* **2022**, *15*, 281. [CrossRef]
31. Makhanova, Z.; Beissenova, G.; Madiyarova, A.; Chazhabayeva, M.; Mambetaliyeva, G.; Suimenova, M.; Shaimerdenova, G.; Mussirepova, E.; Baiburin, A. A Deep Residual Network Designed for Detecting Cracks in Buildings of Historical Significance. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*. [CrossRef]
32. Liu, Z.; Brigham, R.; Long, E.R.; Wilson, L.; Frost, A.; Orr, S.A.; Grau-Bové, J. Semantic segmentation and photogrammetry of crowdsourced images to monitor historic facades. *Herit. Sci.* **2022**, *10*, 1–17. [CrossRef]
33. Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep learning for anomaly detection: A review. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [CrossRef]
34. Fais, S.; Casula, G.; Cuccuru, F.; Ligas, P.; Bianchi, M.G. An innovative methodology for the non-destructive diagnosis of architectural elements of ancient historical buildings. *Sci. Rep.* **2018**, *8*, 4334. [CrossRef] [PubMed]

35. Dore, C.; Murphy, M. Integration of Historic Building Information Modeling (HBIM) and 3D GIS for recording and managing cultural heritage sites. In Proceedings of the 2012 18th International Conference on Virtual Systems and Multimedia, Milan, Italy, 2–5 September 2012; pp. 369–376.
36. Casillo, M.; Colace, F.; Gupta, B.B.; Lorusso, A.; Marongiu, F.; Santaniello, D. A deep learning approach to protecting cultural heritage buildings through IoT-based systems. In Proceedings of the 2022 IEEE International Conference on Smart Computing (SMARTCOMP), Helsinki, Finland, 20–24 June 2022, pp. 252–256.
37. Pierdicca, R.; Paolanti, M.; Matrone, F.; Martini, M.; Morbidoni, C.; Malinverni, E.S.; Frontoni, E.; Lingua, A.M. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sens.* **2020**, *12*, 1005. [CrossRef]
38. Ni, Z.; Eriksson, P.; Liu, Y.; Karlsson, M.; Gong, S. Improving energy efficiency while preserving historic buildings with digital twins and artificial intelligence. *Proc. Iop Conf. Ser. Earth Environ. Sci.* **2021**, *863*, 012041. [CrossRef]
39. Kaplan, J.; McCandlish, S.; Henighan, T.; Brown, T.B.; Chess, B.; Child, R.; Gray, S.; Radford, A.; Wu, J.; Amodei, D. Scaling laws for neural language models. *arXiv* **2020**, arXiv:2001.08361.
40. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
41. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 116–131.
42. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
43. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
44. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep Learning vs. Traditional Computer Vision. In *Advances in Computer Vision*; Arai, K., Kapoor, S., Eds.; Springer: Cham, Switzerland, 2020; pp. 128–144.