remote sensing

MDPI

Article

Real-Time Ground-Level Building Damage Detection Based on Lightweight and Accurate YOLOv5 Using Terrestrial Images

Chaoxian Liu [1], Haigang Sui [1,*], Jianxun Wang [1], Zixuan Ni [1] and Liang Ge [2]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; cx_leo@whu.edu.cn (C.L.); wangjianxun@whu.edu.cn (J.W.); zixuan.ni@uq.net.au (Z.N.)

[2] Tianjin Institute of Surveying and Mapping Company Limited, No. 9 Changling Road, Liqizhuang, Tianjin 300381, China; geliang0021@126.com

* Correspondence: 00201543@whu.edu.cn; Tel.: +86-027-6877-8229

Abstract: Real-time building damage detection effectively improves the timeliness of post-earthquake assessments. In recent years, terrestrial images from smartphones or cameras have become a rich source of disaster information that may be useful in assessing building damage at a lower cost. In this study, we present an efficient method of building damage detection based on terrestrial images in combination with an improved YOLOv5. We compiled a Ground-level Detection in Building Damage Assessment (GDBDA) dataset consisting of terrestrial images with annotations of damage types, including debris, collapse, spalling, and cracks. A lightweight and accurate YOLOv5 (LA-YOLOv5) model was used to optimize the detection efficiency and accuracy. In particular, a lightweight Ghost bottleneck was added to the backbone and neck modules of the YOLOv5 model, with the aim to reduce the model size. A Convolutional Block Attention Module (CBAM) was added to the backbone module to enhance the damage recognition effect. In addition, regarding the scale difference of building damage, the Bi-Directional Feature Pyramid Network (Bi-FPN) for multi-scale feature fusion was used in the neck module to aggregate features with different damage types. Moreover, depthwise separable convolution (DSCONV) was used in the neck module to further compress the parameters. Based on our GDBDA dataset, the proposed method not only achieved detection accuracy above 90% for different damage targets, but also had the smallest weight size and fastest detection speed, which improved by about 64% and 24%, respectively. The model performed well on datasets from different regions. The overall results indicate that the proposed model realizes rapid and accurate damage detection, and meets the requirement of lightweight embedding in the future.

Keywords: building damage; terrestrial image; ghost bottleneck; CBAM; Bi-FPN; YOLOv5

Citation: Liu, C.; Sui, H.; Wang, J.; Ni, Z.; Ge, L. Real-Time Ground-Level Building Damage Detection Based on Lightweight and Accurate YOLOv5 Using Terrestrial Images. *Remote Sens.* **2022**, *14*, 2763. https://doi.org/10.3390/rs14122763

Academic Editor: Andrea Ciampalini

Received: 12 May 2022
Accepted: 6 June 2022
Published: 8 June 2022

1. Introduction

Rapid building damage detection provides timely and detailed emergency information [1]. Owing to the advantage of data acquisition and observation techniques, such as different platforms and Very-High-Resolution (VHR) observation, remote sensing (RS) technology plays a crucial role in damage detection [2,3]. The damage types caused by earthquakes to buildings are very complex, which is not only reflected in the change in the overall structure, such as debris and collapse, but also in the partial damage of some components, such as foundation settlement or wall damage [4]. Current building damage detection uses three types of input data at the satellite level, drone level, and ground level [5]. Although great progress has been made in building damage detection on satellite and drone platforms, accurate minor damage detection is still difficult to meet the needs of practical applications, such as post-disaster claims and reconstruction [6,7]. Moreover, damaged targets with different granularity from satellites, drones, and the ground show significant characteristic differences in RS images, as shown in Figure 1. Generally, satellite images are useful for large-scale severe damage detection with relatively low resolution,

and drone footage is convenient for façade and roof structure damage detection with oblique observation [8,9]. In contrast, terrestrial images with high resolution and lower cost focus on the damage details, including minor damage, such as cracks and spalling [10]. Overall, most studies focus on damage extraction with a single type or granularity, lacking systematic research on multi-granularity damage targets.

(**a**) (**b**) (**c**)

Figure 1. The difference in multi-granularity damage characteristics from satellite images (**a**), drone footage (**b**), and terrestrial images (**c**).

Generally, building damage detection based on macro-level images, i.e., satellite and UAV images, is appropriate for rough assessment [11,12]. However, data collection and damage analysis based on macro-level images are costly and rely on favorable weather. Especially in the practical application of field investigation, macro-level images are not very suitable for comprehensive and accurate damage detection. Much more detailed damage in building façades needs to be identified based on high-resolution images. The fast evolution of sensor technology (including smartphones and cameras), particularly its imaging capability, indicates that post-disaster terrestrial image capture has become a viable alternative, with VHR images, convenient collection, and real-time collection. Compared to macro-level images, the damage detection approach based on terrestrial images is inexpensive and independent of weather conditions. The extent of damage to buildings can be easily detected and site-specific information can be acquired instantly using social media. Most importantly, detailed building damage information can be obtained. Some studies have demonstrated the usability of ground-level images for effective damage analysis [13,14].

With the development of artificial intelligence and computer vision, object detection methods based on deep convolutional neural networks (CNNs) have been widely used for identifying building damage [15,16]. According to the stage of realization, the existing approaches include target detection frameworks based on the region extraction method and regression methods [17]. The region-extraction-based target detection framework is based on the Regional CNN (R-CNN), such as fast R-CNN and mask R-CNN. It firstly locates objects in the candidate regions, and then determines the class and location by sliding the window. It has good flexibility and precision advantages. The regression-based target detection framework is dominated by the "you only look once" (YOLO) series and the single-shot multibox detector (SSD), which can predict multiple box locations and classes much more quickly by simplifying the feature extraction process. However, in general, deep learning methods require adequate computing resources and a long training time owing to their complex structure. Therefore, they are time-consuming and difficult to be used in embedded devices. To solve these problems, some compressed and lightweight models have attracted increased attention for their flexibility and maintainability. However, they are rarely applied in damage detection due to the complex diversity of building damage [18].

When a disaster occurs, witnesses can take a large number of terrestrial images in a short time, thereby collecting damage information quickly and accurately, which directly

affects the subsequent quantitative analysis [19]. Combining terrestrial images with deep learning methods may be an effective way forward. However, we understand that only a few studies have been conducted in the field of real-time damage detection so far, which considered the diversity of damage targets or the limitations of computer hardware in analysis. The following challenges may be the reason: (1) the lack of appropriate ground-level damage datasets makes it impossible to apply deep learning methods to damage detection. Recent approaches have, therefore, focused on satellite-level and UAV-level dataset construction [20]. (2) Owing to the complex background and diverse damage targets, traditional approaches to damage detection are prone to missing information or to acquiring fake data [21]. (3) More attention has been paid to improving damage detection accuracy using a complex CNN structure, which requires substantial computing power and is difficult to embed into a mobile device, and causes insufficient timeliness.

A lot of progress has been made in building damage detection. As for the dataset collection of building damage, a small number of datasets have been accumulated in some experimental areas around specific disaster types. However, these datasets focus on satellite or drone-level information, and the involved damage type is very limited. For example, the commonly used xBD is a representative satellite-level damage dataset, and it is used mainly in severe damage detection and hard to apply to minor damage detection [22]. For multi-type damage detection, more studies have focused on macro assessment of different damage levels, lacking the identification of damage details. For example, some studies have classified the damage levels referring to the most common standard European Macroseismic Scale 1998 (EMS-1998), but they lacked concrete discrimination in the damage details, such as cracking and spalling [23,24]. In terms of improving the practicability of the proposed model on the detection approach, both the detection accuracy and efficiency must be jointly considered. Object detection methods, such as the R-CNN family and YOLO series, have been used in damage detection, but they cannot balance between accuracy and efficiency [25]. Therefore, some research has concentrated on the optimization adjustment of the CNN structure and parameters to build the detection model due to the effectiveness of deep learning methods [26]. Specifically, many researchers choose to add more layers to the CNN model or combine it with some other CNN structures to enhance its ability to learn more image characteristics [27]. In addition, some researchers have used module adjustment to improve the detection results, such as the optimized activation function [28]. However, the calculation efficiency and hardware requirement are rarely taken into account in the methods which have been mentioned previously. To address this shortcoming, structure optimization, model pruning, model quantification, and knowledge distillation are the most commonly used [29,30]. For instance, ShuffleNet and MobileNet are some representative accelerated network architectures [31]. Quantization acceleration can not only reduce the network storage, but also greatly accelerate the detection speed. Furthermore, cutting those unimportant network connections or replacing some complex operations are also commonly used methods [32]. However, it should be emphasized that the majority of these changes do not account for accuracy loss. Despite a small number of lightweight approaches, there is little substantial research on balancing the detection speed and accuracy. It is still extremely difficult for the existing approaches to meet the post-earthquake emergency response requirements.

To overcome such challenges, in our study, we attempted to transfer YOLOv5 to the detection of damaged parts using terrestrial images. This thesis commenced a study on balancing detection speed and accuracy. We compiled a Ground-level Detection in Building Damage Assessment (GDBDA) dataset and proposed a lightweight and accurate detection model (LA-YOLOv5), which uses a YOLOv5 model enhanced by Ghost bottleneck, Convolutional Block Attention Module (CBAM), Bi-Directional Feature Pyramid Network (Bi-FPN,) and Depthwise Separable Convolution (DSConv). In summary, the main contributions of this study are as follows:

(1) A ground-level building damage dataset of considerable data volume was created from terrestrial images, which cover a wide variety of types of building damage so as to facilitate future detailed damage analysis of buildings.
(2) Ghost bottleneck and CBAM modules are introduced into the backbone of YOLOv5, and DSConv and BiFPN are introduced into the neck module of YOLOv5 to accelerate the damage detection efficiency and enhance the damage features.
(3) One prototype system is designed and implemented based on the proposed lightweight LA-YOLOv5 model. It can be used for real-time damage detection from smartphone or camera images. Importantly, the proposed model can be embedded into smartphones or other ground terminals in the future, which is convenient for ground investigators to conduct building damage investigation.

2. Study Area and Data Source

A violent Ms 7.9 Wenchuan earthquake on 12 May 2008 killed nearly 70,000 people, injured more than 370,000, and destroyed the majority of buildings [33]. In this study, the old town sites of Beichuan and Hanwang in Sichuan Province, which were completely preserved as the Site of Wenchuan Earthquake, were selected as the study areas. To evaluate the effectiveness of our proposed method, we mapped the earthquake ruins on the ground during 12–16 August 2019. Although these towns were severely damaged due to the significant earthquake, different types of building damage can still be found in these sites. Even if some building damage features are no longer significant as the initial damage at this site, conducting building damage research is still of significant value because of the rich and varied building damage types and damage samples. An overview of the study area, including spatial location and aerial images, is shown in Figure 2.

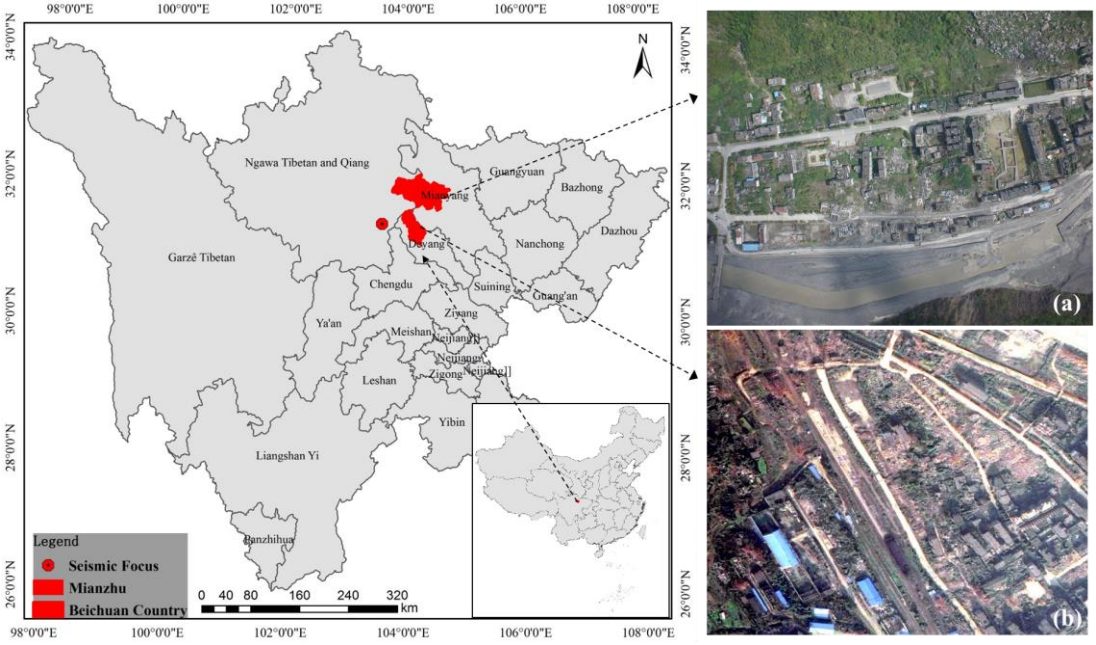


Figure 2. Location map of study area. (**a**) Beichuan site; (**b**) Hanwang site.

To evaluate the effectiveness of our proposed method, we conducted field investigations at two sites, Beichuan and Hanwang, of the 2008 Wenchuan earthquake in the Sichuan Province of China. Many terrestrial images of representative building damage were taken to test the proposed model. In addition, we also collected smartphone images pertaining the 2020 Croatia earthquake and the 2021 Luxian earthquake to verify the generalizability of the proposed model. In line with the common practice of building damage assessment, in this study, we focused on detecting debris, collapse, spalling, and cracking, as the blue, yellow, red, and green polygons denote in Figure 3. Here, debris refers to the ruins of

buildings, and collapse refers to partially only moderate damage when the overall structure of a building is still preserved, such as stairway damage. Spalling and cracking are defined as minor damage to wall surfaces. Damage labeling was conducted by using the LabelImg tool.

Figure 3. Visualization of our GDBDA dataset. The rectangles represent damaged bounding boxes.

Overall, 856 images with 3968 × 2976 pixels from smartphones and cameras were annotated for damage recognition, and 3918 cropped damage images with 800 × 800 pixels were marked. Data augmentation helped to expose the network to more diverse data, leading to a more generalizable output. Considering the image size and damage characteristics, the selected enhancement methods included image rotation, image flipping, color transformation, and image stretching. Then, a total of 8192 sample images were obtained, and 44,059 damage regions were marked. The dataset pertaining to the Wenchuan earthquake was subdivided into two groups: a training set for training the model, which accounted for 70% of the data, and a test set for evaluating the model, which accounted for the remaining 30%, as shown in Table 1. Since the number of damaged parts in different images varied from each other, the specific sample for training and testing was manually filtered.

Table 1. Dataset information of different damage types in GDBDA.

Disaster	Dataset Type	Dataset Number	Damage Type	Damage Number
Wenchuan Earthquake	Training dataset	5608 images	Debris	9565
			Collapse	9868
			Spalling	6167
			Crack	5243
	Testing dataset	2584 images	Debris	4098
			Collapse	4228
			Spalling	2643
			Crack	2247
Croatia Earthquake and Luxian Earthquake	Verifying dataset	148 images	Debris	188
			Collapse	231
			Spalling	159
			Crack	108

We also collected 148 ground-level images from the Croatia earthquake to verify any generalizability of the proposed model. The proportion of data from debris and collapse was higher in proportion to that of cracks and spalling. Figure 4a reports the number of object instances per class, and Figure 4b exhibits the distribution of instances with respect to size in our dataset.

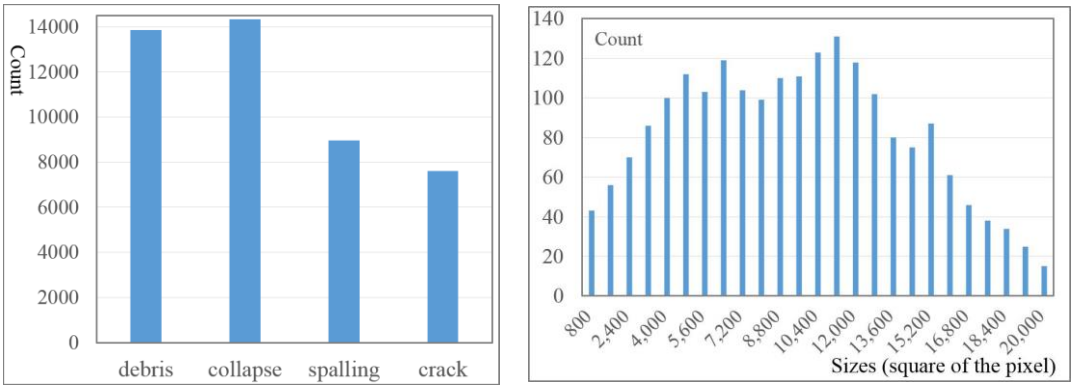


Figure 4. Statistics of different damage types (**a**) and sizes (**b**).

3. Methodology

3.1. Overview

The YOLOv5 series can be divided into YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x [34]. In this study, YOLOv5s was used as the basic network for damage detection as it had the smallest size and lowest number of model parameters. The architecture of the proposed LA-YOLOv5 is shown in Figure 5. To reduce parameter calculation and improve the overall network efficiency, as shown in Figure 5, we replaced BottleneckCSP and CONV with a Ghost bottleneck in the backbone, and used the Bi-FPN and depth-wise separable convolution (DSCONV) in the neck modules. For the Ghost module, a series of linear transformations was applied to generate the corresponding Ghost feature map in a more cheap way. CBAM was added to the backbone module to enhance the characteristics and CONV in the neck module was replaced with DSCONV to compress the parameters. In addition, Bi-FPN was used to optimize the multi-scale damage detection.

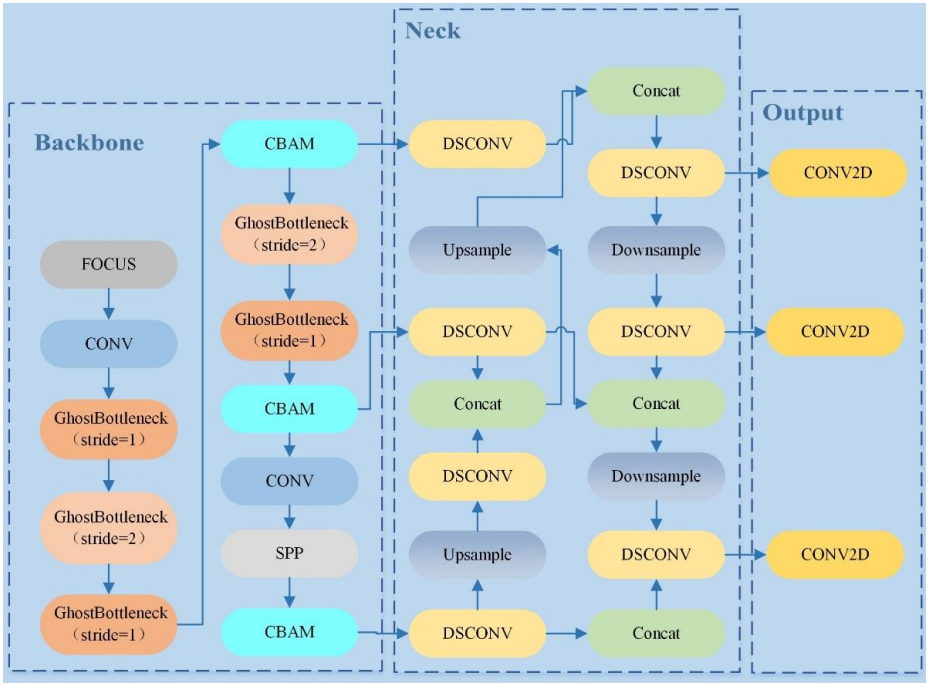


Figure 5. Architecture of the improved YOLOv5.

3.2. Improvement of Backbone Network

BottleneckCSP and CONV in the backbone module were changed to a Ghost bottleneck module. Specifically, the bottleneck of layer 3 was changed to a Ghost bottleneck with

Stride = 1, retaining the same output size. The CONV of the fourth and sixth layers was changed to a Ghost bottleneck with Stride = 2 to conduct sampling. The BottleneckCSP modules of layers 5 and 7 were replaced with two Ghost bottlenecks with Stride = 1. These adjustments aimed to achieve dimensionality reduction by reducing the number of parameters and preventing overfitting by expanding the receptive field. Moreover, we added a CBAM module after the feature map of each scale so that the weight parameters of the network were redistributed according to their importance

3.2.1. Ghost Bottleneck

The Ghost bottleneck module is a plug-and-play module that guarantees full appreciation of input data, which is mainly composed of two stacked ghost modules [35]. Based on a set of internal feature maps, simple linear transformations were applied to generate more ghost feature maps, which could fully reveal the information of the internal features. In this study, a ghost bottleneck is introduced into the backbone module. The lightweight design effectively reduced the overall size of the model without increasing the network parameters. This made the network better and faster to understand redundant information, and it improved the accuracy of the model. The Ghost bottleneck contained a bottleneck with Stride = 1 and Stride = 2, as shown in Figure 6. In the bottleneck with Stride = 1, the first Ghost module was used as the extension layer, increasing the number of channels. The second Ghost module reduced the number of channels required to match the shortcut path. In the bottleneck with Stride = 1, down sampling was achieved in this model, a fast path was realized, and a depth-wise convolution with Stride = 2 was inserted between two ghost modules for connection.

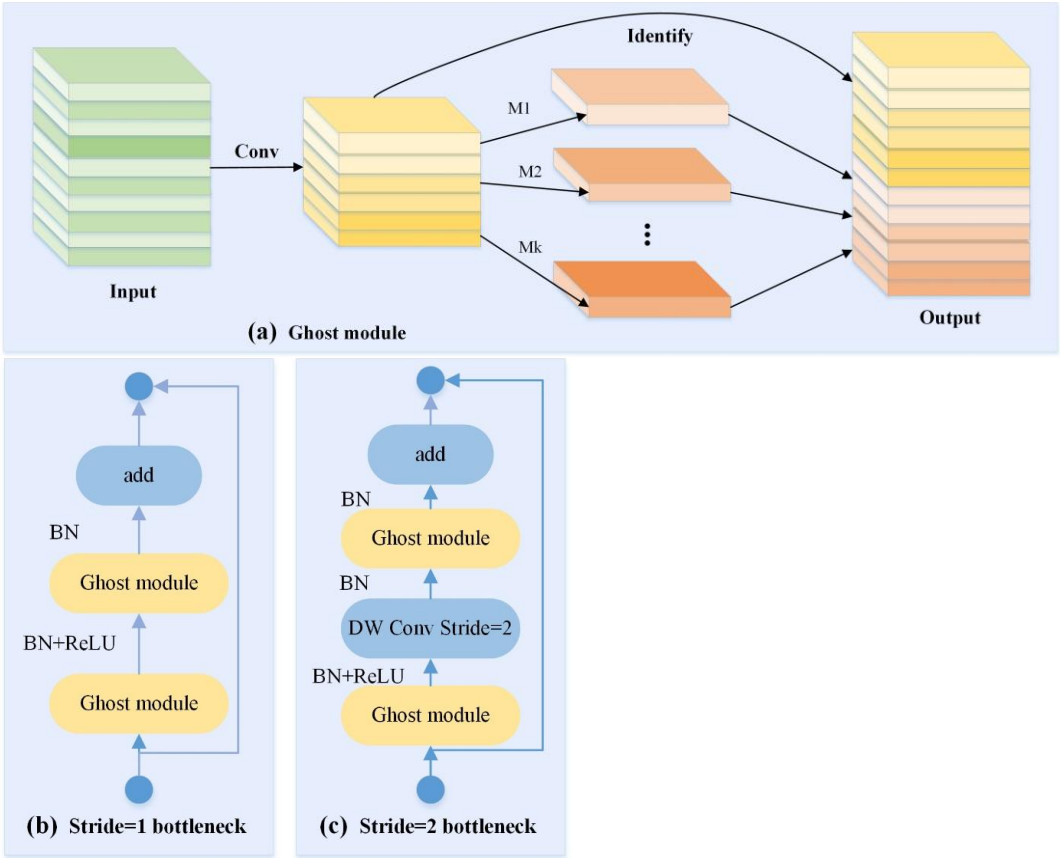


Figure 6. Diagram of the Ghost module (**a**) and the structure of the Ghost Bottleneck (**b**,**c**).

3.2.2. CBAM Module

The potential interference of windows, doors, and balconies requires the model to focus on specific damage information and reduce the loss of model accuracy caused by lightweight transformation. In this study, we inserted an effective CBAM module into the backbone network to enhance the feature representation of CNN networks [36]. The CBAM was composed of two complementary modules: channel attention and spatial attention, as shown in Figure 7. It could suppress the characteristics of complex backgrounds, highlight the characteristics of defects, and focus on the spatial location of cracks in photovoltaic cell images under complex backgrounds.

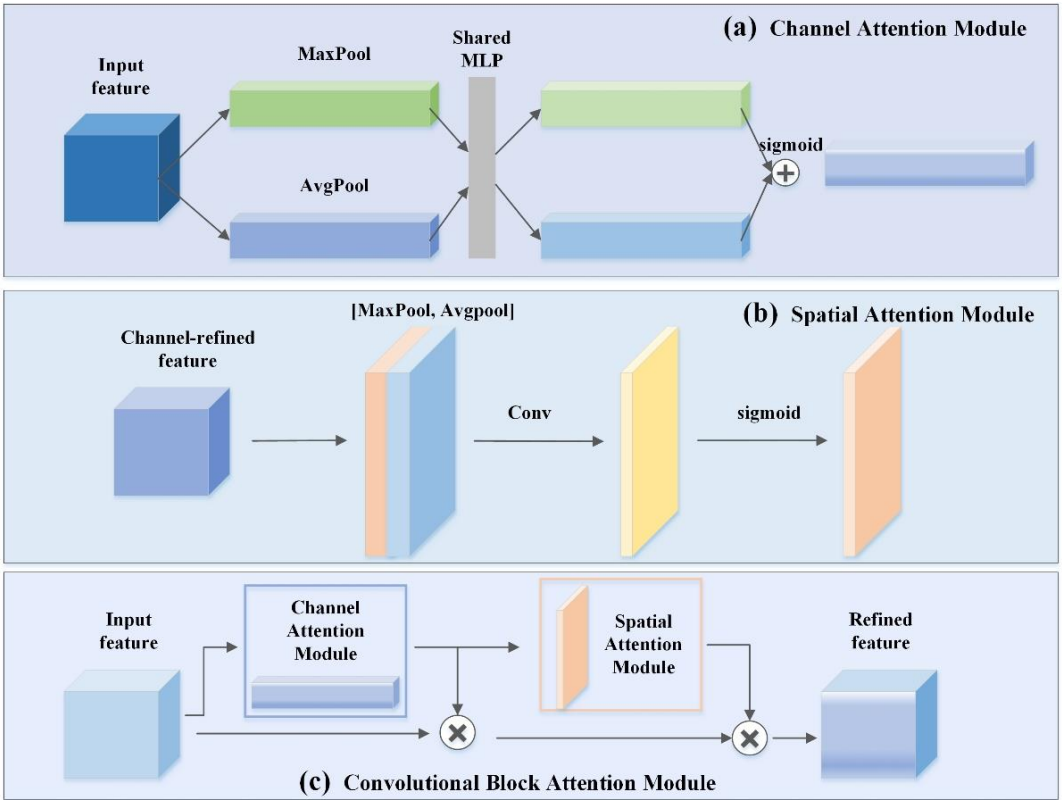


Figure 7. The overview of the CBAM and related attention module.

Among the latter, channel attention focused on "what is the target" [37]. By assigning greater weight to the channel containing more defect information and a smaller weight to the channel containing more background information, a channel containing useful defect feature information could be selected. Specifically, based on a VHR remote-sensing image, after some convolutional layers, we could obtain a multichannel feature map $F \in R^{C \times H \times W}$ (where C, H, and W are the number of channels, and the height and width of the feature map, respectively). The feature map of each channel focuses on different information. The purpose of channel attention is to use the relationship between each channel in the feature graph to learn the 1D weight, and then multiply it by the corresponding channel. In this way, it focuses on finding meaningful semantic information in the current task. To learn effective weight representation, we first aggregated spatial dimension information through global average pooling and global max-pooling, and generated two feature descriptors for each channel. Then, two feature descriptors were fed into a shared multilayer perceptron with a hidden layer to obtain a more representative feature vector. Then, we merged the output eigenvectors through the element summation operation. Finally, the final channel attention map could be obtained based on a sigmoid function. A flow chart of channel attention is the illustrated module in Figure 7a.

Spatial attention tells the network "where the point is" and helps the network to locate the defect in the feature map [38]. Hence, the CBAM module contributes to learning what and where to emphasize or suppress information and refine intermediate features effectively. Specifically, it focuses on where is valuable in the current task. In high-resolution images, the ground objects show various sizes and complex distribution. Therefore, using spatial attention is very useful for aggregating spatial information, especially for small targets. Spatial attention uses the relationship between different spatial positions to learn the 2D spatial weight map and then multiplies it by the corresponding spatial position to learn more representative features. In order to effectively learn the spatial weight relationship, we first generated two feature descriptors for each spatial location through the operation of global average pooling and global max-pooling. Then, we connected the two feature descriptors together through 7 × 7 convolution operation to generate a spatial attention graph. Finally, a sigmoid function was used to scale the spatial attention map to 0~1. The flow chart is illustrated in the spatial attention module of Figure 7b.

3.3. Improvement of Neck Network

The neck in YOLOv5 is used to generate feature pyramids and achieve the identification of the same object with different sizes. YOLOv5 uses the PA-Net as Neck to aggregate features. However, there is a failure in directly using the traditional model to detect multi-scale building damage. We should take into account the multi-scale damage and the large amount of calculation. The following methods mainly focused on both of them to improve the detection accuracy and efficiency.

3.3.1. Deep Separable Convolution

Some lightweight networks, such as mobilenet, use deep separable convolution to extract feature maps [39]. Compared with the conventional convolution operation, the number of parameters and operation cost are relatively low [40]. As shown in Figure 8, for the input with Dx × Dy × 3 feature map, the output with Dx × Dy × N feature map can be obtained after 3 × 3 convolution kernel. The standard convolution process combines N convolution kernels of 3 × 3 with each channel of the feature map to obtain the new feature map with N channels. The deep separable convolution firstly uses three convolution kernels of 3 × 3 to convolute with each channel of the input feature map to obtain a feature map in which the input channel is equal to the output channel. Then, N convolution kernels are used to convolute the feature map to obtain a new feature map with N channels. The number of parameters using different convolutions can be calculated using the following equations.

$$P1 = Dx \times Dy \times 3 \times N \quad (1)$$

$$P2 = Dx \times Dy \times 3 + Dx \times Dy \times 1 \times N \quad (2)$$

$$P2/P1 = (3 + N)/(3N) << 1, (N >> 3) \quad (3)$$

where P1 and P2 represent the number of parameters using standard convolution and deep separable convolution separately. Dx and Dy are the length and width of the input feature map, respectively. N is the number of convolution kernels. It can be seen that the P2/P1 result was far lower than 1. After using deep separable convolution, the effect was similar to that of standard convolution, and the number of parameters used in convolution can be greatly reduced. Specifically, for the Neck module, the original convolution was changed to DSCONV with the aim of realizing a lightweight model design. By using DSCONV in feature map extraction, the number of parameters as well as operation costs significantly decreased [41]. The DSCONV is composed of a depthwise (DW) layer and a pointwise (PW) layer [42]. DW uses a 3 × 3 convolution kernel, and PW uses a 1 × 1 convolution kernel. Each convolution result is processed using a batch normalization algorithm and a rectified linear unit. The BN algorithm adjusts the data distribution, which avoids the disappearance of the gradient and the setting of complex parameters.

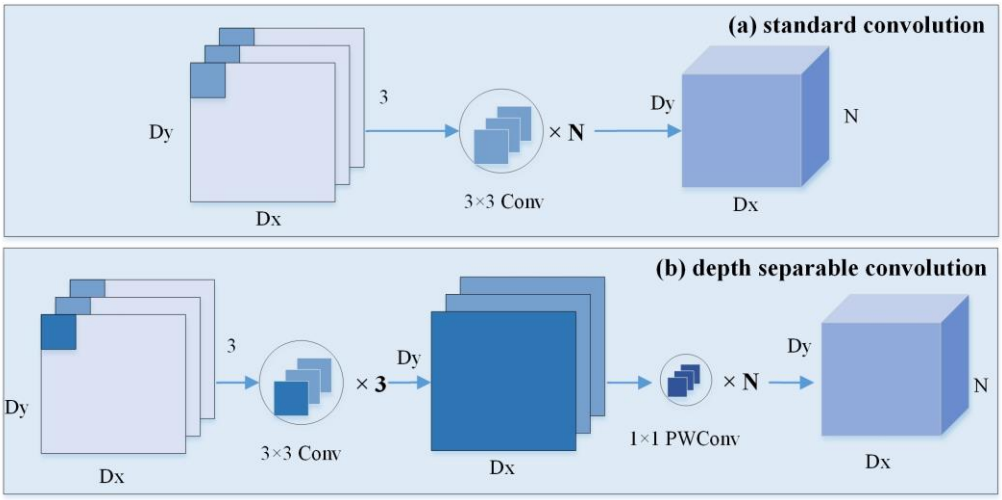


Figure 8. Comparison of standard convolution and depth separable convolution.

3.3.2. Multi-Scale Feature Fusion Using Bi-FPN

Complex and diverse damage targets are apparently different between different scales. How to combine multi-scale features in the proposed approach is of great importance in damage detection. The latest algorithms based on deep learning can obtain low-level and high-level features from high-resolution images through CNNs [43,44]. Therefore, we need to fuse these multi-scale features in the networks. Although PANet in YOLOv5 also can achieve different feature fusions through up-sampling and down-sampling, it has a large amount of calculation [45]. In contrast, the Bi-FPN used in EfficientDet is a typical complex bidirectional feature fusion FPN (feature pyramid networks) structure, as shown in Figure 9 [46]. On the basis of the traditional bidirectional feature fusion FPN (such as panet), Bi-FPN removes the two intermediate nodes of the highest viterbilt feature layer and the lowest viterbilt feature layer entering the FPN structure, and adds a residual edge connecting the input feature map and the output feature map to each feature layer in the middle.

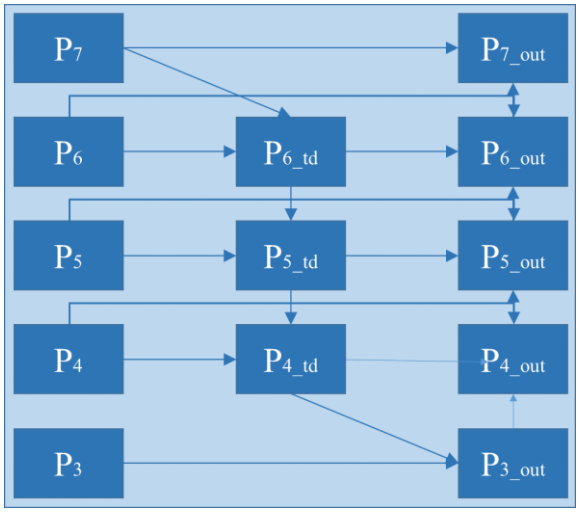


Figure 9. Bi-FPN structure for multi-scale feature fusion.

Since the contribution weights of different feature map inputs to the final output should be different at each node for feature fusion, Bi-FPN introduces a weight for training to adjust the contribution of different inputs to the output feature map. In terms of weight selection, Bi FPN uses a fast normalized fusion strategy, which improves the speed by 30% when the effect is similar to that of softmax-based optimization [47]. After several rounds

of training, at each feature fusion node, the input feature map will obtain the weight that results in the best effect of the target detection algorithm. On the whole, Bi-FPN achieves easy and fast multi-scale feature fusion, and the structure of FPN is simplified to a certain extent. In this study, Bi-FPN was used to improve the model efficiency.

4. Experimental Materials

All experiments were carried out on a PC with 40 GB memory, Intel Xeon E5 CPU with 2.7 GHz, and Tesla K80 GPU with 24 GB. Some basic parameters of the proposed model are as follows: the pre-trained weights using ImageNet, a batch size of 4, an initial learning rate of 0.01, and iterations of 300 epochs. The detection algorithms are often evaluated in terms of accuracy and speed. The accuracy is measured by precision, recall average, F1-score@0.75, and the mean average precision (mAP). Speed is measured by the training time, parameter size, weight size, and inference time for each image.

4.1. Migration Network Initialization

Transfer learning is a common machine learning method, whose key is to transfer the knowledge that has been trained in a certain field to another new field [48]. As for this paper, it concerns the completion of the model's pre-training. The results will be migrated to the YOLO v5 network of kiwi flaw detection to help the training of the detection model. To initialize the model parameters of a small training set, a pre-trained network model is selected with a good learning ability to complete. Since the damaged image samples in this paper are limited and few, migration learning was also chosen to initialize the parameters of the YOLO v5 network, which can ensure the successful migration of the learned knowledge and the capability to make the new network capable to learn quickly. In this way, the overfitting caused by insufficient damage samples can be improved to a certain degree. At the same time, the generalization ability of building damage detection can also be improved correspondingly so that the recognition model can be facilitated. Even under complex natural conditions, the model had a good recognition ability to perform migration learning.

4.2. Evaluation Metric

For the evaluation of the classification, the results were compared to the manually labeled reference data based on field survey data and visual interpretation results using terrestrial images. In this study, the Precision (Pre.), Recall (Rec.), Overall Accuracy (OA), and F1-score were used to evaluate the extraction and classification performance for each scene, as shown in Equations (4)–(7). OA can provide the under- or over-estimation of damage classification in uneven positive and negative samples. Thus, Pre. was used to evaluate the false detection ratio, Rec. was used to evaluate the missed detection ratio, and F_1-score denoted the comprehensive evaluation index of Pre. and Rec. Here, TP denoted the true positive, which meant the number of positive samples that were correctly classified as positive. FP stood for the false positive, which meant the number of negative samples that were incorrectly classified as positive. FN was the false negative, which represented the number of positive samples that were incorrectly classified as negative. TN denoted the true negative, which was the number of negative samples that were correctly classified as negative. It should be noted that TP not only relied on the classification accuracy, but also considered the size of the detected bounding box. Specifically, the bounding box should be completely included within the actual damage. Otherwise, the detected damage's bounding box should have a similar size with actual damage.

$$\text{Pre.} = \text{TP}/(\text{TP} + \text{FP}) \tag{4}$$

$$\text{Rec.} = \text{TP}/(\text{TP} + \text{FN}) \tag{5}$$

$$\text{OA} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{6}$$

$$F_1\text{-score} = (2 \times \text{Pre.} \times \text{Rec.})/(\text{Pre.} + \text{Rec.}) \tag{7}$$

5. Result and Analysis

5.1. Detection Assessment for Different Damage Types

We chose four representative damage scenarios involving four damage targets to show the detection accuracy. The final corresponding accuracy indicators of the four damage targets are listed in Table 2. The blue, yellow, red, and green boxes represent debris, collapse, spalling, and crack labels, as shown in Figure 10. The number next to the label represents the confidence in a predicted type of damage. High confidence means that building damage could be detected accurately and performed relatively reliably. High confidence meant that the model accurately identified damage, and very little building damage was missed. In all cases of high confidence in damage recognition, confidence ranged above a value of 0.86. The highest confidence score of up to 0.98 was achieved when detecting debris and collapse. The confidence in detecting spalling and cracks was comparatively lower, which explained the small size and scale of the damage.

Table 2. Identification result of different damage types.

Types	Precision (%)	Recall (%)	mAP (%)	F_1-Score (%)
Debris	95.56	92.95	94.26	92.42
Collapse	91.35	90.93	93.53	91.86
Spalling	89.82	90.18	91.28	89.28
Crack	87.91	89.58	90.63	90.59
Average	91.16	90.91	92.43	91.06

Figure 10. Different damage recognition in GDBDA dataset. Multi-class target detection results for images at near and far distances, where different colors represent different targets. (Note: The Chinese term in each image refers to the sign or slogan on the building's exterior).

As shown in Table 2, the average precision, recall, F1 score, and mAP values for the four damage types reached 91.16%, 90.91%, 92.43%, and 91.06%, respectively. Interpreting the results, debris and collapse recognition worked the best, and recognition of spalling and cracks was less accurate. Overall, the amount of available damage samples had the biggest impact on the outcome, while damage features, such as shape and size, were also important. Apart from a few exceptions, almost every index reached more than 90%, which confirmed the good performance of the model and met the accuracy required for detailed damage detection.

5.2. Ablation Experiments

In this study, we firstly conducted ablation analysis for different methods, including YOLOv5, Ghost-bottleneck-based YOLOv5 (G-YOLOv5), CBAM-based YOLOv5 (C-YOLOv5), Ghost-CBAM-based YOLOv5 (GC-YOLOv5), BiFPN-based YOLOv5 (B-YOLOv5), Ghost-BiFPN-based YOLOv5 (GB-YOLOv5), and our proposed lightweight and accurate YOLOv5 (LA-YOLOv5). Figure 11 shows the curves of precision and recall using different models. Compared with using other methods, the curve of the proposed model gradually rose with the most minor oscillations. However, the amplitudes changed notably in the rising limb of the curves of the other models, and this phenomenon continued to the end. The overall results showed good damage recognition performance using the proposed method. Especially for the original YOLOv5, its curve had some large random fluctuations. All of the optimized models avoided these severe oscillations, to a certain extent.

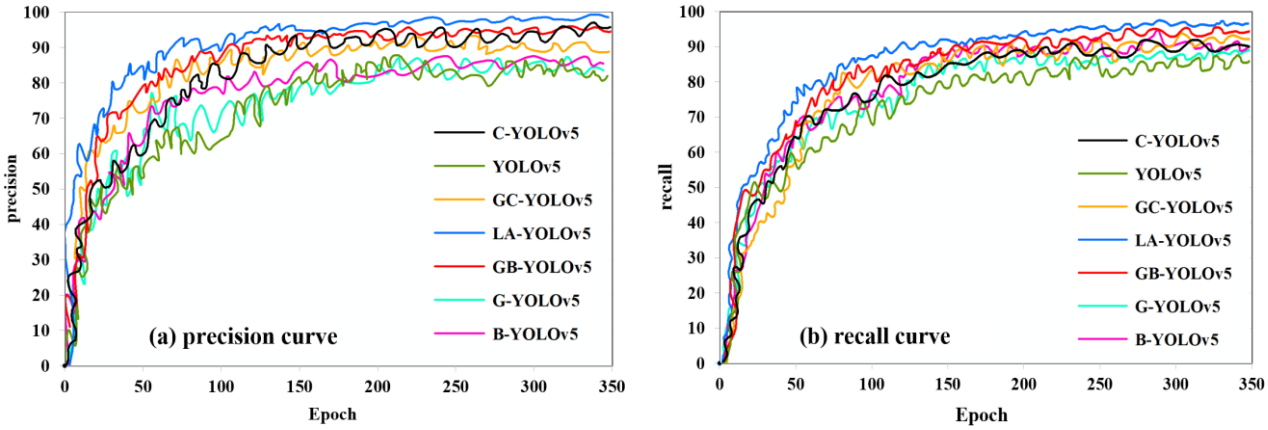


Figure 11. Evaluation comparison using different models.

As Table 3 shows, the ablation analysis demonstrated that our model achieved the superior results in terms of detection accuracy and speed after stepwise optimization. For LA-YOLOv5 and GB-YOLOv5, by deleting the CBAM module, the detection efficiency rose slightly, but the detection accuracy decreased above 2%. It can be found that GB-YOLOv5 outperformed GC-YOLOv5, which illustrated that the Bi-FPN was more effective than CBAM in terms of damage detection. The detection result using B-YOLOv5 and C-YOLOv5 also demonstrated this point. On the whole, the LA-YOLOv5 model required the least amount of time for sample training. The weights file of the LA-YOLOv5 model was only 7.51 MB, which was approximately one-third of that using YOLOv5; additionally, the model had the lowest number of parameters. Additionally, the inference time (detection of single image time) was significantly less than that of the other methods. The detection results of the LA-YOLOv5 model were better, and the detection speed of LA-YOLOv5 was the fastest, which can meet real-time requirements. In conclusion, for accelerating the detection efficiency and improving detection accuracy, every optimization approach in our LA-YOLOv5 model was effective. Moreover, the proposed lightweight model with its high accuracy and small size could easily be transplanted onto embedded devices to improve the damage detection efficiency.

Table 3. Comparison of building damage detection in ablation experiments.

Model	Training Hours	Weight Size (MB)	Parameter Size	Inferences (s)	Precision	Recall	mAP (%)	F_1-Score (%)
LA-YOLOv5	8.43	7.51	3.18×10^6	0.033	92.47	91.39	93.43	92.26
GB-YOLOv5	8.29	6.92	2.97×10^6	0.030	90.25	89.78	91.86	90.08
GC-YOLOv5	8.63	8.55	3.42×10^7	0.039	89.95	89.08	89.15	88.57
C-YOLOv5	9.26	10.37	4.29×10^6	0.048	88.25	88.46	87.12	87.48
G-YOLOv5	8.49	7.58	3.13×10^6	0.035	87.92	87.63	86.64	86.29
B-YOLOv5	8.85	8.73	3.62×10^6	0.037	88.38	88.72	87.52	87.86
YOLOv5	9.19	20.62	3.95×10^7	0.043	85.96	84.27	84.63	85.54

5.3. Comparison Analysis Using Different Models

In recent studies, there were some lightweight CNN models in the target detection field. Thus, we compared the model with other classic target detection algorithms, such as Ghost-bottleneck-based YOLOv5 with Squeeze-and-Excitation module (GS-YOLOv5), YOLOv4, and Faster R-CNN. To illustrate the advantage of the proposed LA-YOLOv5, we first conducted the comparison analysis with some representative lightweight methods, including MobileNet-SSD, Nanodet, and MobileDets. Different from the YOLO series, Single-Shot MultiBox Detector (SSD) is another typical target-detection algorithm. Combining SSD and MobileNet is a common lightweight method. Nanodet is a real-time anchor-free detection model for a mobile terminal. It is hoped that it will provide performance no less than that of the YOLO series, and it is also convenient for training and transplantation. For MobileDets, by merging regular CNN in the search space and directly optimizing the network architecture of target detection, a series of target detection models can be obtained. All the above-mentioned approaches can accelerate the target detection and obtain the lightweight model. In addition, to illustrate the advantages compared with traditional methods, Ghost-bottleneck-based YOLOv5 with the Squeeze-and-Excitation module (GS-YOLOv5), YOLOv4, and Faster R-CNN were used. Among them, GS-YOLOv5 was used to compare the effectiveness of the proposed CBAM module. Based on the provided damage dataset, the final damage detection results using different detection methods are shown in Figure 12 and Table 4. It should be noted that all of the comparative experiments are based on the same training dataset and test dataset.

Figure 12. Comparison results of building damage detection using different methods. (Note: The Chinese term in each image refers to the sign or slogan on the building's exterior).

Table 4. Comparison of building damage detection using different methods.

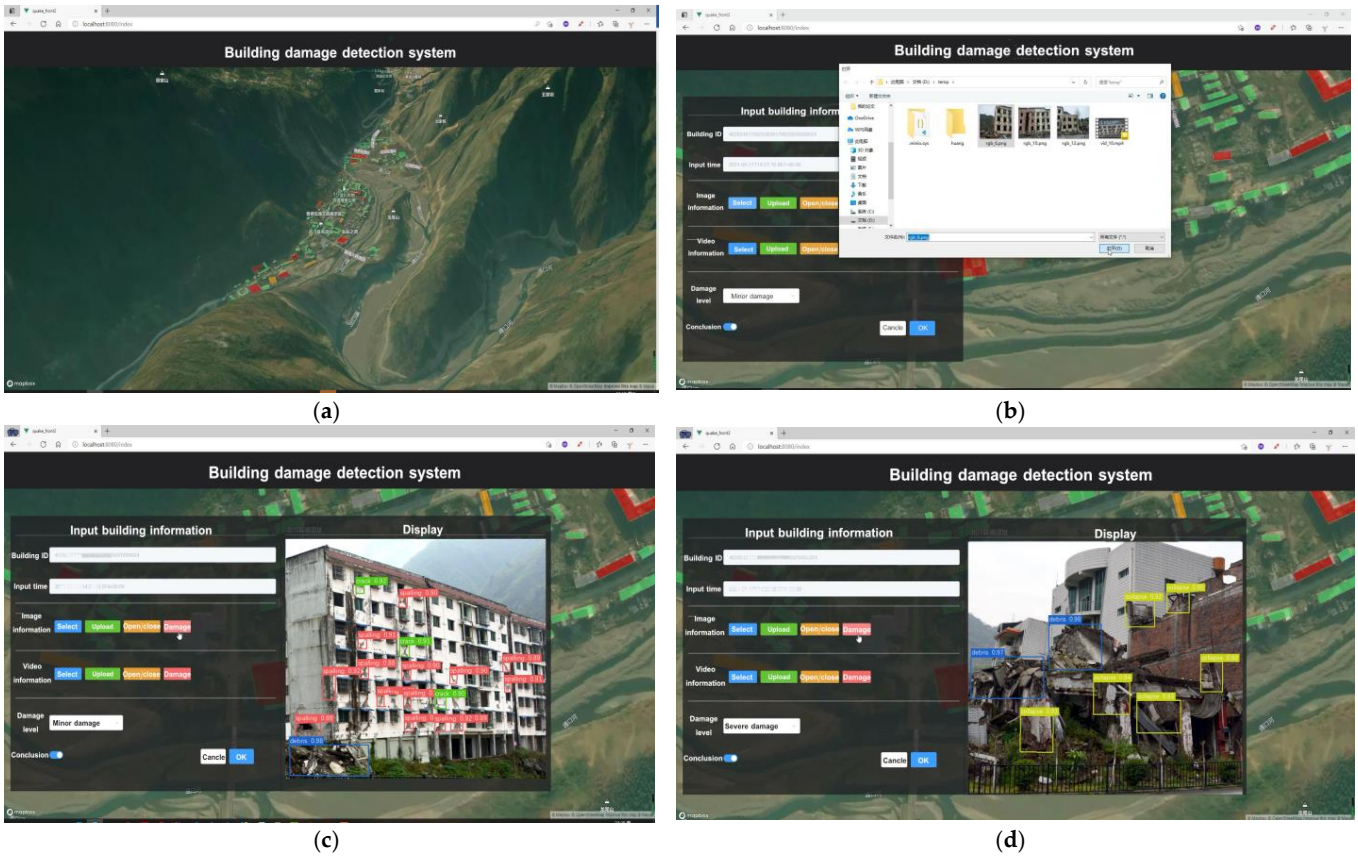
Model	Training Hours	Weight Size (MB)	Parameter Size	Inferences (s)	Precision	Recall	mAP (%)	F_1-Score (%)
LA-YOLOv5	8.43	7.51	3.18×10^6	0.033	91.16	91.29	92.43	91.36
MobileNet-SSD	10.15	26.59	6.47×10^6	0.059	88.37	87.42	87.92	86.47
Nanodet	7.14	7.28	3.73×10^6	0.027	86.92	84.63	84.14	83.29
MobileDets	6.85	6.83	2.12×10^6	0.022	84.73	83.12	84.52	82.26
GS-YOLOv5	8.82	8.72	3.32×10^6	0.036	90.25	89.78	89.86	89.04
YOLOv4	11.63	123.55	9.72×10^7	0.106	88.95	87.08	87.15	83.57
Faster RCNN	12.36	328.62	4.65×10^7	0.228	79.82	83.33	81.57	82.39

As Table 4 shows, the detailed results and thorough ablation analysis showed that our model outperformed the other methods by balancing the damage detection accuracy and speed as much as possible. For different lightweight models, MobileNet-SSD had the potential to reduce the calculation amount, but the improvement in efficiency and accuracy were lower than those using the proposed method in this paper. Nanodet and MobileDets greatly reduced the calculation amount, which was better than the proposed method. However, the damage detection accuracy was obviously poor, which may have been limited by the size of the damage dataset. For traditional target detection methods, GS-YOLOv5 obtained a relatively satisfactory result, but it was slightly worse than the proposed GC-YOLOv5. YOLOv4 and Faster RCNN required greater computational cost and obtained lower accuracy. Specifically, the inference time (detection of single image time) was significantly lower than that of the other methods, which demonstrated a good detection efficiency. In addition, the LA-YOLOv5 model achieved a higher mAP value for the building damage dataset with the smallest weight size. Moreover, our results indicate that the detection results of the LA-YOLOv5 model were better, and the weights and parameters of the model were smaller than those of other approaches before improvement. Furthermore, the detection speed of LA-YOLOv5 was the fastest, and it could meet real-time requirements. In conclusion, our method differed from most current methods because these methods merely focused on improving the detection accuracy. The proposed LA-YOLOv5 balances the detection speed and accuracy to the greatest extent.

5.4. Validation Using a Prototype System

Eventually, we verified the feasibility of the proposed framework by implementing a prototype system. The system was designed with a B/S architecture, and the overall framework was implemented with Springboot + Vue by decomposing the system into front-end and back-end. Specifically, taking the earthquake site of Beichuan County in Sichuan Province as the research area, the prototype system of post-earthquake building damage detection was tested. In this section, ground photos and videos of each building collected in the study area were used as test data to verify the function of the prototype system. The prototype system mainly included five functions: 3D scene rendering, building data uploading, image damage extraction, video damage detection, and building damage conclusion judgment. The satellite map, 3D terrain data, and building vector of the study area could all be successfully superimposed and visually displayed in the system's main interface, as shown in Figure 13a. We uploaded the building image or video and visually displayed the building video using the "open/close" button, as shown in Figure 13b. After successfully uploading the building image/video data, the "damage" button could be used to conduct rapid building damage detection, as shown in Figure 13c,d. After finishing the background execution, the detection results of building damage would be automatically returned to the front display for users. Finally, based on the detailed detection results of each building, the concrete damage level could be concluded. After verification, the prototype system could effectively realize the intelligent and integrated technical process from data acquisition and uploading, to background computing and processing, and then to front-end visual feedback, so as to provide a certain reference for the software and hardware

equipment design of post-disaster command and decision-making and emergency rescue in the future.



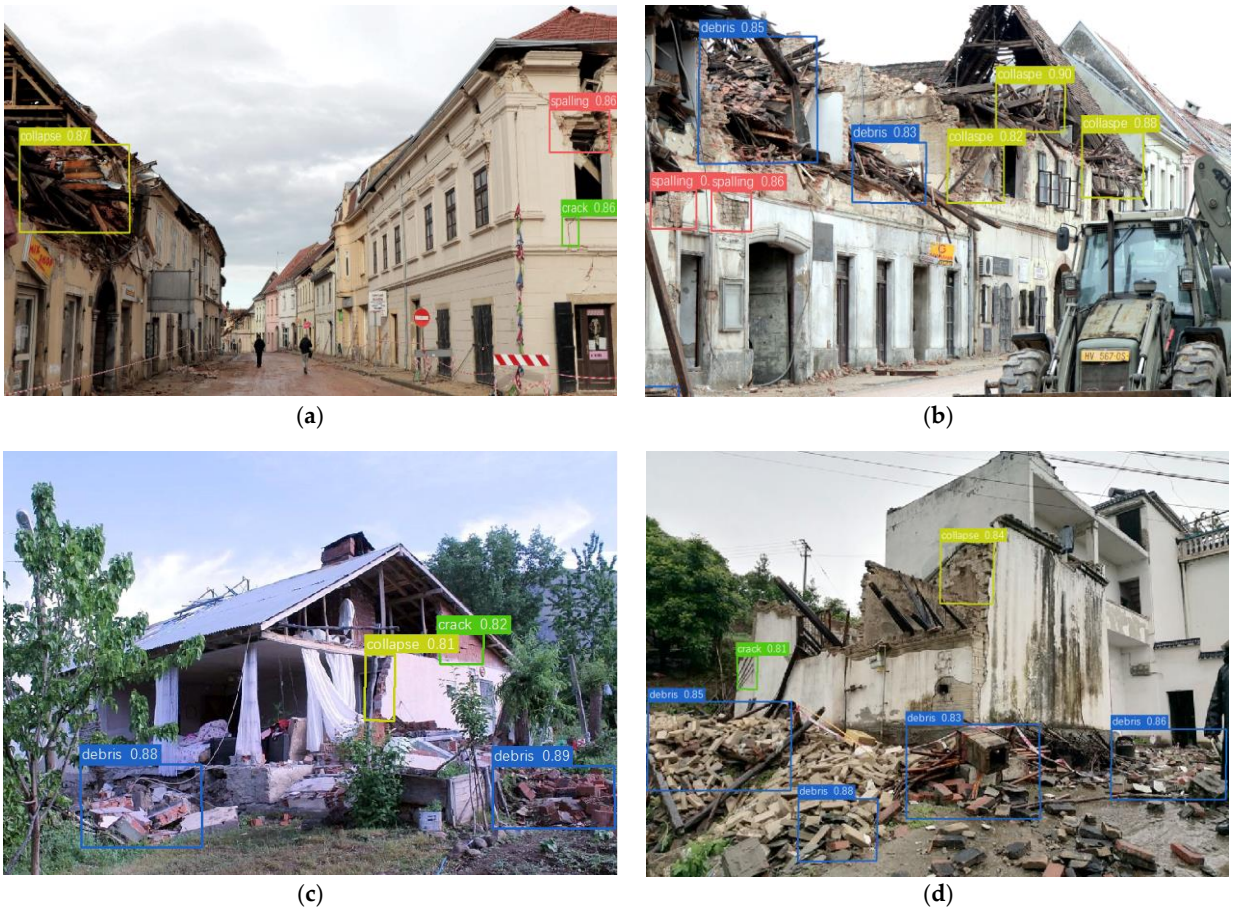
(**a**) (**b**)

(**c**) (**d**)

Figure 13. The main function of the building damage detection system. (**a**) 3D scene rendering, (**b**) building data uploading, (**c**) detection example of minor damage, and (**d**) detection example of severe damage.

5.5. Analysis of Generalization Ability

Although our algorithm achieved a high-quality performance using our test dataset, the robustness and generalizability needed to be verified. To do so, we collected images available on the Internet pertaining to other disasters as the verification dataset. In this section, based on the designed prototype system, we used 86 terrestrial images from the 2020 Croatia Earthquake and 62 terrestrial images from the 2021 Luxian Earthquake to verify the model. Despite the different image resolutions, the damage details could be observed from the selected high-resolution image, as shown in Figure 14. Judging from the final results, the detection effect in the Croatia dataset and Luxian dataset did not show a significant difference. The majority of damage types could be accurately detected with confidence > 80% based on the pre-trained model. According to the evaluation metrics in Table 5, the proposed method also performed well in terms of inference speed and accuracy. Owing to differences in the image resolution and light conditions, there were inevitable differences in the damage detection efficiency. Although the final accuracy was obviously lower than that of the testing dataset, the average accuracy was also above 83%, which met the basic requirements for damage detection in field investigations.



(**a**) (**b**)

(**c**) (**d**)

Figure 14. Damage recognition based on the verification dataset. (**a**,**b**) 2020 Croatia Earthquake; (**c**,**d**) 2021 Luxian Earthquake.

Table 5. Validation results of building damage using different datasets.

Dataset	2020 Croatia Earthquake				2021 Luxian Earthquake			
	Inference (s)	Precision	Recall	mAP (%)	Inference (s)	Precision	Recall	mAP (%)
Verifying	0.042	81.52	82.08	82.57	0.046	80.74	82.15	82.16
Testing	0.039	91.16	91.29	91.03	0.042	90.33	91.33	92.34

6. Discussion

6.1. Analysis of Importance of Lightweight Model in Damage Detection

Timeliness is key in post-earthquake damage detection. Airborne and spaceborne remote sensing techniques have been widely used in severe damage detection, but the ground-level approach is still of great importance in efficiently and accurately identifying minor damage. Especially after an earthquake, people affected by the earthquake can obtain a large number of the spot videos or photos in real-time. How to timely and effectively distinguish the damage footage and detect the damage targets makes a significant contribution to the post-earthquake emergency. In the post-earthquake field investigation, this issue still remains to be settled. As shown in our proposed LA-YOLOv5, for a lightweight model, the amount of calculation was greatly reduced. The requirement for hardware and other computing resources was no longer demanding. Moreover, the lightweight model could be used in embedded devices, which significantly enhanced the practicality of damage detection. In future studies, we will focus on the application in embedded devices, such as smartphones or other handheld terminals, based on the proposed model.

6.2. Performance Analysis Based on Different Depths and Widths of the Network

The YOLOv5 series include YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, where s, m, l, and x represent the increase in model depth in turn. Compared with the other three models, YOLOv5s was the best network with the smallest depth and feature map width. Depth controls the network depth and width controls the network width. In this study, YOLOv5s was used as the basic network for damage detection as it had the smallest size and lowest number of model parameters. To illustrate the impact of different network depths and widths, we explored the performance of wider and deeper networks for building damage detection using the YOLOv5 series and proposed method. As shown in Figure 15, small denoted the baseline network used in this study, middle had a depth and width of (0.65, 0.75), large had a depth and width of (1, 1), and x-large had a depth and width of (1.35, 1.25), respectively. The final detection accuracy showed that the baseline network used in this study outperformed other models, which indicated the superiority of the proposed method.

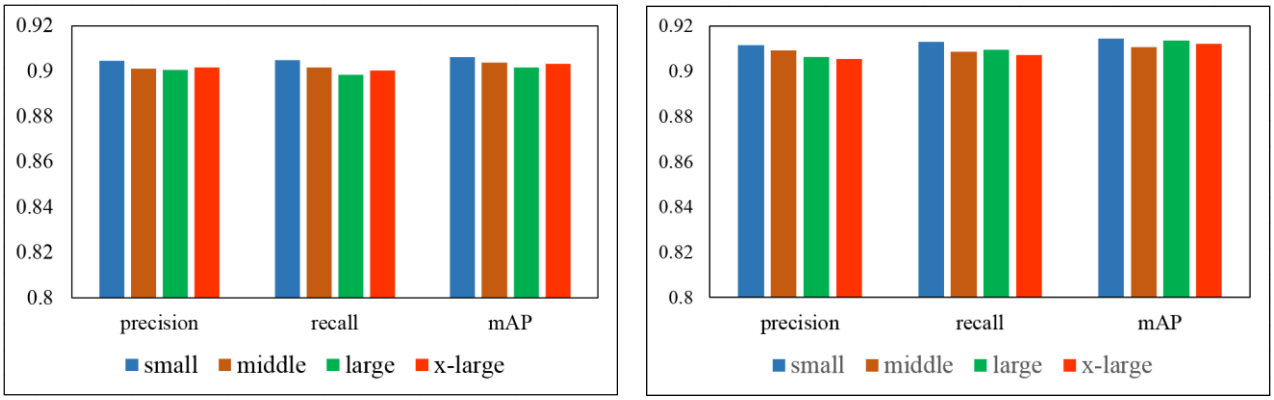


Figure 15. Comparison of detection accuracy based on networks with different depths and widths. (**a**) YOLOv5 series; (**b**) proposed methods.

7. Conclusions

The study was carried out on the high-accuracy and semi-automated assessment method of damage to buildings based on the integration of two-dimensional and three-dimensional features. The entire process was systematically conducted, including model implementation of building point extraction, sample set construction, and model implementation of damage extraction. In addition, the proposed damage detection framework was compared with other commonly used approaches and verified through transferability analysis in another scene. The strengths and potential of the proposed building damage extraction framework were fully reflected.

The greatest challenge now for ground-level building damage assessment is how to achieve sufficiently accurate, real-time detection for diverse damage types with as little calculation as possible. To that end, we created a GDBDA dataset from smartphone images covering some representative types of damage, such as debris, collapse, spalling, and cracks, and proposed a lightweight model using LA-YOLOv5 for real-time and accurate damage recognition. Ghost bottleneck and DSCONV were used to reduce the calculation parameters and the overall weight size. The CBAM module was introduced into the backbone module to make the network focus on the detailed damage characteristics and improve the damage detection accuracy. The final results realized ground-level damage detection with a low time cost, smaller weight size, and higher detection accuracy. Thus, the proposed real-time detection method can contribute to damage assessment in post-disaster emergencies.

Future Developments

In future studies, we first need to expand the data source from different post-disaster areas to further promote the application scope in damage detection. Images from different observation platforms also need to be combined together. Based on more damage samples, not only can we verify the further transferability of the proposed method, but also explore more types of building damage characteristics, which will help to determine the specific damage level. In terms of specific methods, further improving the model's applicability to identifying damage targets in complex scenes will become one of the main tasks. In addition, the proposed model will be applied to embedded devices, such as smartphones, with limited computing power and real-time calculation requirements.

Author Contributions: C.L. and H.S. conceived and designed the experiments; J.W. and L.G. collected terrestrial images and conducted data processing; C.L. conducted the damage detection and Z.N. analyzed the results; H.S. and L.G. revised the paper; and C.L. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Major Program (42192580 and 42192583), National Natural Science Foundation of China (41771457), and the Guangxi Science and Technology Major Project (AA22068072).

Data Availability Statement: Please follow the web site link of our research team: http://www.lmars.whu.edu.cn/suihaigang/index, accessed on 11 May 2022.

Acknowledgments: The authors would like to thank Chengdu University of Technology, and Hanwang and Beichuan Country Earthquake Sites for data support.

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

References

1. Çömert, R.; Matci, D.K.; Avdan, U. Detection of Collapsed Building from Unmanned Aerial Vehicle Data with Object Based Image Classification. *Eskisehir Technol. Univ. J. Sci. Technol. B Theor. Sci.* **2018**, *6*, 109–116. [CrossRef]
2. Ge, P.; Gokon, H.; Meguro, K. A review on synthetic aperture radar-based building damage assessment in disasters. *Remote Sens. Environ.* **2020**, *240*, 111693. [CrossRef]
3. Foulser, R.; Spence, R.; Eguchi, R.; King, A. Using remote sensing for building damage assessment: GEOCAN study and validation for 2011 Christchurch earthquake. *Earthq. Spectra* **2016**, *32*, 611–631. [CrossRef]
4. Suppasri, A.; Mas, E.; Charvet, I.; Gunasekera, R.; Imai, K.; Fukutani, Y. Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami. *Nat. Hazards* **2013**, *66*, 319–341. [CrossRef]
5. Dong, L.; Shan, J. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS J. Photogramm. Remote Sens.* **2013**, *84*, 85–99. [CrossRef]
6. Adriano, B.; Xia, J.; Baier, G.; Yokaya, N.; Koshimura, S. Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 sulawesi earthquake and tsunami in Palu, Indonesia. *Remote Sens.* **2019**, *11*, 886. [CrossRef]
7. Liu, C.; Sui, H.; Huang, L. Identification of building damage from UAV-based photogrammetric point clouds using supervoxel segmentation and latent dirichlet allocation model. *Sensors* **2020**, *20*, 6499. [CrossRef]
8. Kaya, G.; Taski, K.; Musaoglu, N.; Ersoy, O. Damage assessment of 2010 Haiti earthquake with post-earthquake satellite image by support vector selection and adaptation. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 1025–1035. [CrossRef]
9. Kerle, N.; Nex, F.; Gerke, M.; Duarte, D.; Vetrivel, A. UAV-Based Structural Damage Mapping: A Review. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 14. [CrossRef]
10. Xu, Z.; Lu, X.; Cheng, Q.; Guan, H.; Deng, L.; Zhang, Z. A smart phone-based system for post-earthquake investigations of building damage. *Int. J. Disaster Risk Reduct.* **2017**, *27*, 214–222. [CrossRef]
11. Tu, J.; Sui, H.; Feng, W.; Sun, K.; Hua, L. Detection of damaged rooftop areas from high-resolution aerial images based on visual bag-of-words model. *IEEE Geosci. Remote Sens. Lett.* **2016**, *3*, 1817–1821. [CrossRef]
12. Janalipour, M.; Mohammadzadeh, A. Building damage detection using object-based image analysis and ANFIS from high-resolution image (case study: BAM earthquake, Iran). *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2019**, *9*, 1937–1945. [CrossRef]
13. Liu, C.; Zhou, L.; Wang, W.; Zhao, X. Concrete Surface Damage Volume Measurement Based on Three-Dimensional Reconstruction by Smartphones. *IEEE Sens. J.* **2021**, *21*, 11349–11360. [CrossRef]
14. Zhai, W.; Peng, Z. Damage assessment using google street view: Evidence from hurricane Michael in Mexico beach, Florida. *Appl. Geogr.* **2020**, *123*, 102252–102266. [CrossRef]
15. Ji, M.; Liu, L.; Du, R.; Buchroithner, M.F. A comparative study of texture and convolutional neural network features for detecting collapsed buildings after earthquakes using pre- and post-event satellite imagery. *Remote Sens.* **2019**, *11*, 1202. [CrossRef]

16. Vetrivel, A.; Gerke, M.; Kerle, N.; Nex, F.; Vosselman, G. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 45–59. [CrossRef]
17. 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.
18. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
19. Pham, T.H.; Apparicio, P.; Gomez, C.; Weber, C.; Mathon, D. Towards a rapid automatic detection of building damage using remote sensing for disaster management: The 2010 Haiti earthquake. *Disaster Prev. Manag.* **2014**, *23*, 53–66. [CrossRef]
20. Adams, S.M.; Levitan, M.L.; Friedland, C.J. High Resolution Imagery Collection Utilizing Unmanned Aerial Vehicles (UAVs) for Post-Disaster Studies. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 1161–1168. [CrossRef]
21. Ye, X.; Liu, M.; Wang, J.; Qin, Q.; Ren, H.; Wang, J.; Hui, J. Building-based damage detection from postquake image using multiple-feature analysis. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 499–503. [CrossRef]
22. Wu, C.; Zhang, F.; Xia, J.; Xu, Y.; Liu, R. Building damage detection using u-net with attention mechanism from pre- and post-disaster remote sensing datasets. *Remote Sens.* **2021**, *13*, 905. [CrossRef]
23. Schwarz, J.; Raschke, M.; Maiwald, H. Comparative seismic risk studies for German earthquake regions on the basis of the european macroseismic scale ems-98. *Nat. Hazards* **2006**, *38*, 259–282. [CrossRef]
24. Orlando, M.; Salvatori, L.; Spinelli, P.; De Stefano, M. Displacement capacity of masonry piers: Parametric numerical analyses versus international building codes. *Bull. Earthq. Eng.* **2016**, *14*, 2259–2271. [CrossRef]
25. Zhan, Y.; Liu, W.; Maruyama, Y. Damaged Building Extraction Using Modified Mask R-CNN Model Using Post-Event Aerial Images of the 2016 Kumamoto Earthquake. *Remote Sens.* **2022**, *14*, 1002. [CrossRef]
26. Chen, F.; Yu, B. Earthquake-induced building damage mapping based on multi-task deep learning framework. *IEEE Access* **2019**, *7*, 181396–181404. [CrossRef]
27. Li, Y.; Hu, W.; Dong, H.; Zhang, X. Building damage detection from post-event aerial imagery using single shot multibox detector. *Appl. Sci.* **2019**, *9*, 1128. [CrossRef]
28. Cheng, C.S.; Behzadan, A.H.; Noshadravan, A. Deep learning for post-hurricane aerial damage assessment of buildings. *Comput. Aided Civ. Infrastruct. Eng.* **2021**, *36*, 695–710. [CrossRef]
29. Zhang, X.; Zou, J.; He, K.; Sun, J. Accelerating very deep convolutional networks for classification and detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 1943–1955. [CrossRef]
30. Guo, J.M.; Yang, J.S.; Seshathiri, S.; Wu, H.W. A Light-Weight CNN for Object Detection with Sparse Model and Knowledge Distillation. *Electronics* **2022**, *11*, 575. [CrossRef]
31. Gao, C.; Yao, R.; Zhou, Y.; Zhao, J.; Fang, L.; Hu, F. Efficient lightweight video person re-identification with online difference discrimination module. *Multimed. Tools Appl.* **2022**, *81*, 19169–19181. [CrossRef]
32. Liu, R.; Jiang, D.; Zhang, L.; Zhang, Z. Deep depthwise separable convolutional network for change detection in optical aerial images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 1109–1118. [CrossRef]
33. Xu, G.; Fang, W.; Shi, P.; Yuan, Y. The fast loss assessment of the Wenchuan earthquake. *J. Earthq. Eng. Eng. Vib.* **2008**, *28*, 74–83.
34. Zhu, L.; Geng, X.; Li, Z.; Liu, C. Improving YOLOv5 with Attention Mechanism for Detecting Boulders from Planetary Images. *Remote Sens.* **2021**, *13*, 3776. [CrossRef]
35. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 1580–1589.
36. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
37. Zhang, Y.; Yi, P.; Zhou, D.; Yang, X.; Yang, D.; Zhang, Q.; Wei, X. CSANet: Channel and spatial mixed attention CNN for pedestrian detection. *IEEE Access* **2020**, *8*, 76243–76252. [CrossRef]
38. Lu, X.; Ji, J.; Xing, Z.; Miao, Q. Attention and feature fusion ssd for remote sensing object detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5501309. [CrossRef]
39. Li, L.; Zhang, S.; Wu, J. Efficient object detection framework and hardware architecture for remote sensing images. *Remote Sens.* **2019**, *11*, 2376. [CrossRef]
40. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
41. Ma, S.; Liu, W.; Cai, W.; Shang, Z.; Liu, G. Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions. *IEEE Access.* **2019**, *7*, 57023–57036. [CrossRef]
42. Xiao, Z.; Zhang, Z.; Hung, K.W.; Lui, S. Real-time video super-resolution using lightweight depthwise separable group convolutions with channel shuffling. *J. Vis. Commun. Image Represent.* **2021**, *75*, 103038. [CrossRef]
43. Ma, J.; Yuan, Y. Dimension reduction of image deep feature using PCA. *J. Vis. Commun. Image Represent.* **2019**, *63*, 102578. [CrossRef]
44. Xiao, Y.; Tian, Z.; Yu, J.; Zhang, Y.; Liu, S.; Du, S.; Lan, X. A review of object detection based on deep learning. *Multimed. Tools Appl.* **2020**, *79*, 23729–23791. [CrossRef]

45. Mekhalfi, M.L.; Nicolo, C.; Bazi, Y.; Rahhal, M.; Alsharif, N.A.; Maghayreh, E.A. Contrasting yolov5, transformer, and efficientdet detectors for crop circle detection in desert. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 3003205. [CrossRef]
46. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787.
47. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. Available online: https://arxiv.org/abs/2004.10934 (accessed on 28 June 2020).
48. Pan, B.; Tai, J.; Zheng, Q.; Zhao, S. Cascade convolutional neural network based on transfer-learning for aircraft detection on high-resolution remote sensing images. *J. Sens.* **2017**, *2017*, 1796728. [CrossRef]