Computer Fraud and Security
ISSN (online): 1873-7056

A Comprehensive Review of CNN-Based Methods for Earthquake-Induced Building Damage Detection Using Remote Sensing Imagery

Yubin Xu[1], Jun Liu[2]*, Ran Zou[1], Ruyan Zhou[1], Sha Chen[3], Changyue Zhong[4], Ying Wang[2], Delong Tang[5], Chunyan Han[2], Ke Han[2], Peilin Xu[1]
1. School of Information Science, Shanghai Ocean University, Shanghai, China
2. National Earthquake Response Support Service, Beijing, China
3. National Disaster Reduction Center, Beijing, China
4. Guizhou Minzu University, Guiyang Guizhou, China
5. Guizhou Earthquake Agency, Guiyang Guizhou, China
Corresponding Author's Email: liujun_nerss@sina.com

Abstract: Earthquakes are highly destructive and sudden natural disasters. The primary cause of casualties and major economic losses is the collapse of buildings due to ground shaking. Quick and accurate assessment and localization of damaged buildings is crucial for the rapid deployment of post-earthquake rescue operations and disaster reconstruction tasks. This paper focuses on building damage detection, discussing the application and current developments of deep learning Convolutional Neural Networks(CNNs) in this field. Based on dual-temporal and single-temporal image sources trained by networks, this study compares the characteristics and advantages of different network architectures. In terms of single-temporal imagery-based methods, this study individually examines the detection performance and application contexts for distinct types of detection tasks: semantic segmentation, object detection, and instance segmentation. Furthermore, this paper also analyzes and compares the main types of remote sensing data used in earthquake building damage detection tasks. Finally, this study summarizes the challenges faced by CNNs in the task of automatic building damage detection and the corresponding strategies, aiming to guide future improvement directions. This review offers researchers in the field of disaster assessment and emergency response post-earthquake a reference for decision-making solutions.

Keywords: CNN; building damage detection; earthquake disaster; remote sensing images; semantic segmentation

1. Introduction

Over the past few years, large-scale natural disasters have become more frequent across the globe. Among them, earthquakes are considered one of the most destructive and catastrophic natural disasters [1]. Turkey was hit by two powerful earthquakes on February 6, 2023, with magnitudes of 7.8 and 7.5, causing over 50,000 deaths, significant damage to nearly 430,000 buildings, and economic losses of 1.3 trillion US dollars [2]. A powerful 7.6-magnitude earthquake struck the Noto Peninsula in Japan on January 1, 2024, damaging over 80,000 buildings and leaving 15,920 families displaced [3]. Buildings, being the primary hubs for human production and daily life, are also the direct cause of casualties and economic losses when they suffer damage or collapse following earthquakes [4]. Rapid and accurate building damage detection and localization are crucial for reducing losses, enhancing rescue efficiency, and accelerating post-earthquake recovery and reconstruction [5].

Traditional building damage detection methods, such as field surveys and visual interpretation, are reliable, detailed, and accurate, but they are labor-intensive and time-consuming, and field investigations often carry significant risks [6]. Recently, re-searchers have focused on automating the detection and mapping of building damage, aiming to improve technology to match or even exceed the accuracy of manual assessments. Remote sensing data has long served as a fast, efficient, and reliable foundation for post-earthquake damage evaluation, becoming the preferred data source for such assessments [7]. The continuous advancement of remote sensing technology, provides higher-resolution and more accessible data support for disaster relief missions. For instance, the number of high-resolution optical satellites globally continues to in-crease [8], and the image resolution provided by commercial remote sensing satellite companies has surpassed 0.3 meters [9].

In recent years, deep learning (DL) has experienced both vigorous and explosive development, driven by the large-scale collection of datasets, iterative advancements in computational processors, the rapid evolution of algorithmic techniques, and the continuous refinement of model architectures. There is a growing interest in leveraging deep learning as an effective tool to address complex and challenging problems in earthquake engineering [10]. Convolutional Neural Networks (CNNs), as a subset of deep learning, are predominantly utilized in image processing applications. Due to their powerful automatic feature extraction and nonlinear feature mapping capabilities, CNNs have garnered widespread attention and played a significant role in numerous remote sensing applications. A large number of studies have shown that deep learning approaches, including CNNs, offer superior feature extraction compared to traditional methods. The experiment by Ji et al.

(2019) [11] involved extracting gray-level co-occurrence matrix (GLCM) texture features and CNN features from the 2010 Haiti earthquake data. Both traditional methods and CNN-based approaches were applied to damage detection and mapping tasks using a random forest classifier. The results showed that the CNN method achieved higher accuracy compared to the traditional approaches. Based on the data acquisition time, building damage detection methods can be classified into pre- and post-earthquake(dual-temporal) methods and post-earthquake (single-temporal) methods. In the field of dual-temporal studies, this review examines the research on Cascade Networks and Siamese Networks [12] from the perspective of CNN network architecture. The post -earthquake single-temporal approach mainly focuses on feature extraction. We conducted a review of the current mainstream networks and their improvements in building damage detection research, categorized according to the task types of semantic segmentation, object detection, and instance segmentation. This has certain reference significance for post-disaster emergency rescue and damage assessment analysis. However, there are still many challenges and limitations when applying CNN to building damage detection tasks. These challenges include the scarcity of publicly available datasets, the imbalance in building damage severity instances, and the limitations in the applicability of CNN models. These issues hinder the accuracy and efficiency of automated building damage detection models to some extent. Therefore, we review the relevant strategies to provide useful insights for addressing these challenges.

Using the Web of Science (WOS) multidisciplinary database, we performed an advanced search of relevant literature by titles, keywords, and abstracts. The search string used was as follows:

TS = ((building AND damage) AND ("deep learning" OR DL OR CNN OR FCN OR RNN OR "Deconvolutional Net" OR GAN OR Autoencoder OR "Transfer learning" OR "Supervised learning" OR "Unsupervised learning" OR "Reinforcement learning") AND (disaster OR earthquake)) AND LA= (English) AND PY= (2003-2024);

A total of 537 relevant pieces of literature were retrieved. After manual screening, literature meeting the following four criteria were excluded:

1. Not related to earthquake disaster research;
2. Did not utilize deep learning models;
3. Not focused on building damage detection;
4. Not driven by earthquake imagery data;

After screening, a total of 202 pieces of literature were obtained. Figure 1 shows the annual growth trend in research on the application of deep learning models for detecting collapsed and damaged buildings after earthquakes. Through word frequency analysis of literature titles, keywords, and abstracts, a word cloud was generated (see Figure 2). It can be observed that keywords such as "earthquake", "building damage detection", "CNN" and "deep learning" occupy a core position in the related research.

This review is composed of five main sections: the second section provides a comparative analysis of the application of different types of remote sensing datasets in earthquake disasters. The third section summarizes the application of CNN methods in dual-temporal studies. The fourth section provides an overview of the mainstream networks and their improvements within the context of different CNN task types and summarizes the research status of post-earthquake single-temporal studies. The fifth and sixth sections discuss the challenges and corresponding strategies encountered in building damage detection tasks caused by earthquakes and offer suggestions and prospects for the future application of CNN in this field.

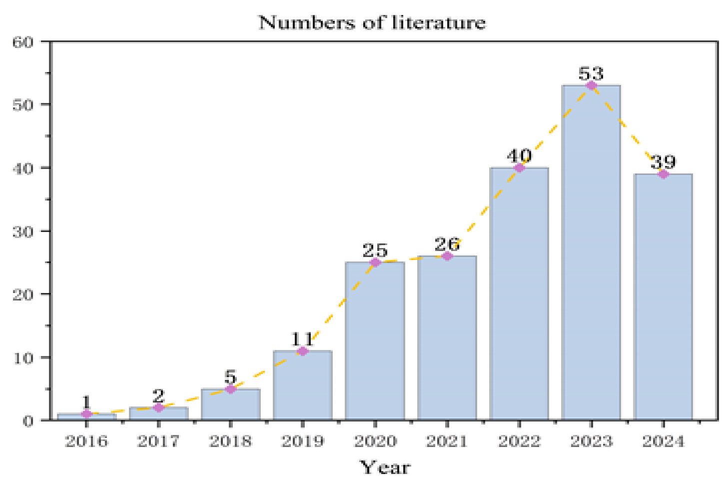

Figure. 1 The amount of deep learning literature on post-earthquake building damage detection (2016-2024).

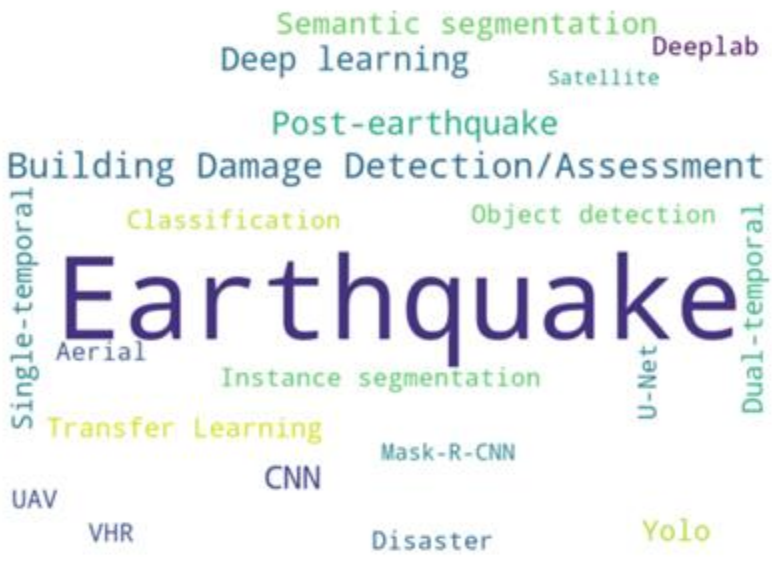

Figure. 2 The word cloud visualization of titles, abstracts, and keywords from 202 selected pieces of literature on post-earthquake building damage detection.

2. Remote Sensing Datasets for Earthquake Disasters

Remote sensing data, with its advantages of broad coverage, rapid response, large amounts of disaster-related information, and minimal impact from disasters on data collection, has been widely used for damage assessment in areas affected by cata-strophic events [[13],[14]].With the advancement of remote sensin g technology and the continuous updates and iterations of sensors, researchers have gained access to high-er-resolution and higher-quality remote sensing datasets. At the same time, the volume of data being acquired continues to increase. In building damage assessment tasks, remote sensing data can be classified based on sensor type, such as optical imagery, synthetic aperture radar (SAR), and light detection and ranging (LiDAR), and so on. As shown in Figure 3, the statistical distribution of the types of remote sensing imagery used in earthquake damage detection research indicates that optical imagery occupies a large proportion, with satellite imagery being the most frequently used.

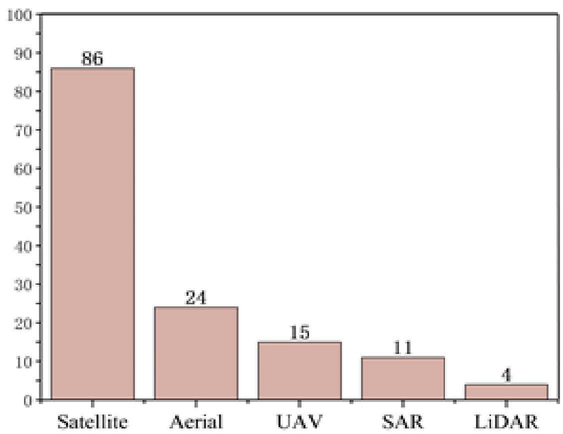

Figure. 3 Statistics on the usage of various types of remote sensing imagery data in earthquake building damage detection research.

2.1 Optical imagery

Optical imagery, a type of passive remote sensing technology, captures imaging in the visible and near-infrared spectra reflected from the ground using optical sensors. Optical imagery allows for easy extraction of features such as texture, geometry, and spectral reflectance. Compared to remote sensing imagery obtained from other sensors, optical imagery is more suitable for damage detection and mapping and is more favored by researchers. According to statistics, currently, more than half of the research on building damage detection using remote sensing imagery is based on optical imagery [15]. Optical imagery can be categorized into satellite, airborne, and low-altitude UAV imagery based on the data collection platform. Table 1 lists the publicly available optical imagery earthquake dataset.

2.1.1 Satellite imagery

Satellites such as GF-1, GF-2, GF-4, HJ-1A/B/C, Jilin-1, SuperView-1, Zhuhai-1, QuickBird, GeoEye-1, WorldView-2, WorldView-3, and Pleiades [[6],[16],[17]]are equipped with cameras for capturing optic al imagery, with a resolution range of 0.5m to 50m. Compared to aerial imagery from aircraft and UAVs, satellite optical imagery offers advantages such as wide imaging coverage and higher spatiotemporal resolution, making it the most widely used data source for building damage detection [[14],[15],[18]].

However, satellite imagery also has some limitations in post-earthquake building damage detection. For example, earthquakes are often accompanied by precipitation and secondary disasters such as landslides and mudslides, which significantly reduce the availability of optical satellite imagery [16]. Additionally, optical satellite remote sensing data typically only provide information on the building's rooftop, lacking details on façade damage, which affects the reliability of building damage detection.

2.1.2 Aerial imagery

Aerial imagery typically has a spatial resolution ranging from 0.1 to 0.3 meters. This higher resolution enables the provision of finer spatial details in building damage assessments, which is critical for enhancing the accuracy of evaluation results. Addition-ally, aerial remote sensing platforms offer flexible image capture angles, allowing for the collection of oblique views that facilitate the assessment of building damage [7]. How -ever, the acquisition of aerial imagery is costly, and collecting images in remote areas with manned aircraft is highly hazardous.

2.1.3 UAV imagery

Unmanned aerial vehicles (UAVs) typically capture ground imagery while hovering or cruising at low altitudes. Equipped with sensors, UAVs can collect exceptionally high-resolution optical imagery, with spatial resolution often below 0.1m. In recent years, research on building damage assessment using UAV imagery has increased dramatically. This surge is largely due to UAVs' flexibility in site deployment and flight planning, as well as their relatively low operational costs. Additionally, the ultra-high resolution of UAV imagery captures detailed damage features (such as cracks and de-bris), which are essential for detection tasks. Furthermore, compared to airborne aerial imagery, UAV imagery can provide a wider field of view, enabling the capture of higher-level building facade details to support high-precision damage assessment. However, the spatial coverage of UAV imagery is limited, and in sparsely built areas, a single image typically contains only 2 to 5 building instances.

Table 1 The public optical remote sensing imagery used in earthquake building damage detection

Dataset/Earthquake	Platform	Spatial Resolution	URL
xBD	WorldView-2 WorldView-3 GeoEye-1 QuickBird-2	< 0.8m	https://hyper.ai/datasets/13272 https://xview2.org/dataset
2010 Haiti Earthquake	Worldview-2	0.5m	https://drive.google.com/drive/folders /1Um9poJPwbrVRE1ge01prrKOONot0XfM6 https://www.haiti-patrimoine.org /PROJECTS/2010HAITI/haiti(3)maps.html
Yushu and Ludian datasets	UAV	-0.2,0.1m	https://github.com/city292/build_assessment
2023 Turkey Earthquake	UAV	---	https://www.kaggle.com/datasets/buraktaci/turkiye-earthquake-2023

2.2 LiDAR

LiDAR generates images by emitting laser pulses and calculating the time taken for their reflection, allowing for the precise acquisition of building height data and the creation of an accurate Digital Surface Model (DSM). In damage detection tasks, building height information is of significant value [19]. By comparing the changes in building height before and after the earthquake, the damage to the building can be visually assessed. However, research on automatic building damage detection using LiDAR data alone is limited, primarily due to the unavailability of pre-disaster LiDAR imagery [17]. Relying solely on post-disaster imagery's spatial and height features makes it difficult to effectively distinguish damaged buildings. In building damage detection tasks, LiDAR data is typically used to provide additional features (such as height and spatial characteristics) and is combined with other types of data (such as optical imagery) to facilitate the detection of building damage [[17],[19],[20]

2.3 SAR

SAR (Synthetic Aperture Radar), an active remote sensing technology, offers ad-vantages such as large coverage, rapid response, immunity to cloud cover and lighting conditions, and the ability to operate in all weather and all-time conditions. Due to these characteristics, researchers consider SAR imagery to be a more flexible and reliable data source for damage assessment in the early stages of a disaster [[21],[22]]

Researchers, based on the stability of texture features such as the regularity of building shapes in SAR imagery, have distinguished damaged buildings using only features like intensity, coherence, and polarization information from SAR data. In the study by Bai et al. (2017) [23], two ALOS2/PALSAR-2 dual-polarization SAR images collected from the earthquake-affected area were used to analyze spatial locations with high differences in SAR backscatter coefficients and low correlation coefficients to as-sess the damaged regions. In another study by Bai et al. (2017) [24], ALOS2/PALSAR-2 SAR images were combined with a machine learning framework, applying the k-nearest neighbors (K-NN) method to differentiate the damaged areas by analyzing the feature values from the pre-and post-earthquake differential images. In the study by Karimzadeh et al. (2017) [25], dual-polarization SAR data from Sentinel-1 (VV, VH) and ALOS-2 (HH, HV) satellites were used to analyze intensity, coherence, and polarization measurement information. This data, combined with optical image features, was utilized to detect post-earthquake damaged areas. In more recent research, SAR data has been used as a supplementary resource to optical imagery, offering its unique ad-vantages in feature extraction. Adriano et al. (2021) [16]used dual-temporal SAR data collected from sensors such as TerraSAR-X, COSMO-SkyMed, and ALOS-2, complementing optical data sets to create a global multi-temporal, multi-modal remote sensing dataset for damage detection tasks.

3. Dual-Temporal CNN Method

The research on building damage based on dual-temporal images involves using pre- and post-earthquake imagery as inputs to train deep learning models for accurate prediction of building damage. Recently, researchers have designed various automated detection CNNs using dual-temporal imagery as input. These networks can be classified based on their network structure into cascade (see Figure 4) and Siamese (see Figure 5) architectures. Cascade architectures include models like SegDetector [26], while Sia-mese architectures primarily include TDA-Net [27], BDA-Net [28], SDCI-Net [4], BD-SKU-Net [29], MS4D-Net [30], and U-BDD++ [31].

3.1 Cascade network architecture

The cascade network architecture stacks pre- and post-earthquake images as input to the CNN model. A simple cascade network consists of basic convolutional layers and fully connected layers. More complex cascade networks, however, do not simply stack dual-temporal images as input. Instead, they break the task into two main components: identifying building regions and classifying damage levels. The first component focuses on assigning semantic labels to each pixel to detect building areas, while the second component labels each building with a unique damage level [28], reflecting the extent of the damage. In cascade networks, some studies use the same baseline network and fi-ne-tune the network structure based on the differences in the localization and classification tasks, while other studies use different networks to achieve more accurate de-tection.

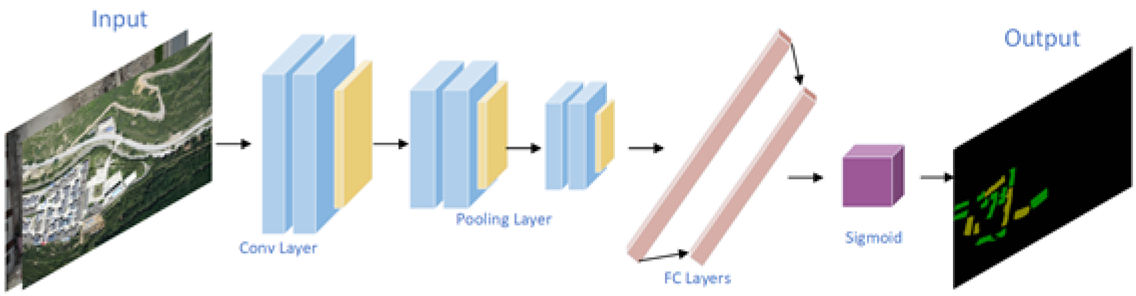

Figure. 4 Dual-temporal Cascade Network Architecture.

Early cascade network architectures typically stacked pre- and post-disaster images as input in a simple manner [11]. This approach failed to fully exploit the potential of CNN models to extract the feature differences between pre- and post-disaster imagery, limiting the model's performance in building damage detection. Later studies focused more on exploring different backbone network models or improving the models to achieve higher detection accuracy. Moradi et al. (2020) [32]stacked pre- and post-disaster images as input for the cascade network and replaced the max-pooling layer of the Deep Residual U-Net with stride convolution layers for training the network model. This method successfully performed building localization and damage classification for damaged buildings. Compared to the baseline Deep Residual U-Net, this approach improved the overall accuracy by 7.21%. Miyamoto et al. (2021) [33]extended the time dimension in a standard 2-D CNN based on the temporal characteristics of dual-temporal imagery, constructing a 3-D CNN with two degrees of freedom in the spatial direction and one degree of freedom in the time direction, and used it as the feature extraction network in the cascade structure. Additionally, the model incorporated building structural information (such as building age and structure type) to build a multimodal cascade network. This network structure significantly improved detection accuracy. Recent research has shifted toward a two-stage strategy for building assessment: the first stage focuses on building localization, i.e., building image segmentation, while the second stage performs damage classification on the segmented results. In this context, we can selectively choose the best-performing segmentation and classification CNN models to achieve optimal detection results. Chen (2021) [34]designed an image seg-mentation network based on Unet+VGG16_BN for building localization. In the damage classification stage, it combined the ResNet18 pre-trained on ImageNet with the U-Net network trained in the localization stage. Similarly, Bouchard et al. (2022) [35]used Attention-U-Net as the building localization network and Siamese-ResNet architecture as the damage classification network, with both networks combined in a cascade manner to jointly complete the building damage classification task.

3.2 Siamese network architecture

The Siamese network architecture parallelly inputs dual-temporal imagery into two branch networks, which separately extract feature representations from pre- and post-earthquake images. Some feature extraction branch networks use the same convolutional layer structure and share weights, aiming to bridge the knowledge gap between localization and classification tasks. This network structure is called a "Siamese" model [36]. Other studies have designed completely different localization and damage classification branch networks, which are then fused to complete the building damage detection task. This design fully combines the advantages of network structures that perform well in localization and classification tasks, significantly improving the overall detection accuracy of the model. This is referred to as a "fusion" model [36].

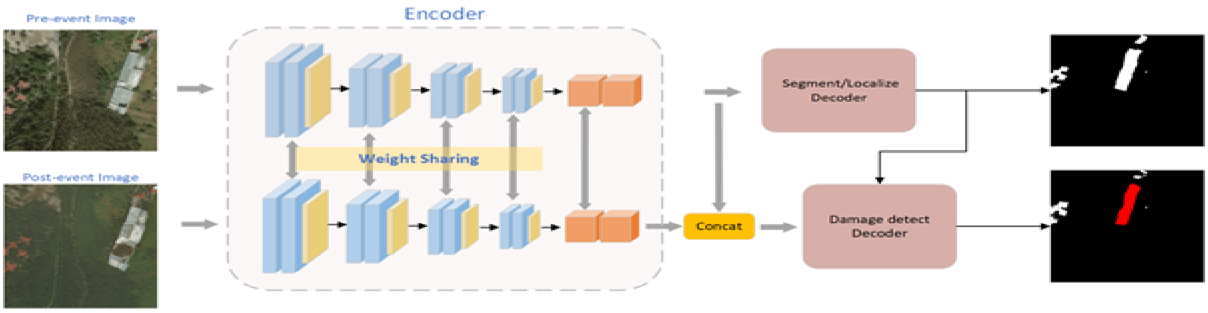

Figure. 5 Dual-temporal Siamese Network Architecture.

The initial research on dual-temporal building damage detection based on the Siamese network architecture primarily focused on integrating the features extracted by the two branch networks. Weber et al. (2020) [37]designed a Siamese network architecture where both branches utilized a ResNet50 backbone pre-trained on ImageNet with shared weights, combined with a Feature Pyramid Network (FPN) for feature ex-traction. May et al.(2022) [38]also adopted the Siamese network architecture and com-pared the performance differences between UNetEfficientNet-B0 and UNetEfficient-Net-B5 in building damage assessment tasks. In their experiment, the encoder networks are identical and share weights. The outputs from the skip connections are generated in parallel, then concatenated and used as input to the decoder. Adriano et al. (2021) [16]used a Siamese U-Net architecture with an added attention mechanism. This architecture also shares the weights of the encoder networks for both pre- and post-disaster images and connects the corresponding encoder and decoder blocks through skip connections to restore the original pixel resolution of the input images, while passing the learned information from the encoder to the decoder. Recently, some researchers have leveraged the flexibility and scalability of the Siamese network architecture by first training a building localization network using pre-disaster images and then fusing the damage classification features extracted by the Siamese network with the building localization results to perform building damage detection [[28],[39]]. Zheng et al.(2021) [12]propos ed a Siamese network architecture based on deep object semantic change detection (ChangeOS), which consists of two main parts: the encoder and the multi-task decoder. The encoder is composed of a shared-weight ResNet and a task-aware context encoder. The branch decoder, which takes pre-disaster images as input, is responsible for building localization. It sends the extracted deep object features, along with features from the post-disaster images, to the damage classification decoder. The final detection results undergo an object-based post-processing stage, and when compared to the xView2 baseline method and the top solution of xView2, ChangeOS achieves the highest F1 scores in both localization and classification tasks.

In dual-temporal CNN-based research methods, these two network architectures each have their advantages. In terms of feature extraction efficiency, the Siamese net-work, where pre- and post-event images are input separately, has a significant ad-vantage over the cascade network, which stacks the dual-temporal images as input. Xu et al. (2019) [40]and Kalantar et al. (2020) [36]compared the performance of the tw o methods in post-earthquake building damage detection. The findings indicated that the Siamese network surpassed the cascade network in both overall accuracy and model robustness, excelling particularly in distinguishing undamaged categories from dam-aged details, such as building debris. In the context of splitting the earthquake building damage detection task into two stages—localization and damage classification—the cascade network is more suitable for emergency rescue situations and assessment tasks [35]. This is because cascade networks typically do not require shared weights and can be trained and deployed independently, resulting in lower computational costs and faster convergence. On the other hand, the Siamese structure, which eliminates the knowledge gap through shared network weights, requires both sub-networks to be trained simultaneously to update the weights together. This imposes higher computational resource demands and may lead to longer training times.

4. Single-temporal CNN Method

Single-temporal methods use only post-event images as input to the model, automatically extracting building textures, background features, and other detailed in-formation through CNNs for building damage detection [22]. With the continuous iteration and improvement of convolutional neural network models, many models with higher accuracy and faster detection speeds have been proposed, greatly advancing the development of post-earthquake single-temporal research. Depending on the task type, post-earthquake building damage detection can be categorized into three main types: semantic segmentation, object detection, and instance segmentation. Figure 6 summarizes the CNN models used in earthquake damage detection.

Figure. 6 The three main types of convolutional neural network models used for building damage detection tasks.

4.1 Semantic Segmentation Network

Semantic segmentation involves assigning a semantic label to each pixel in an image, and it is a fundamental task in computer vision. In the research on post-earthquake building damage detection, commonly used semantic segmentation models include U-Net [41], SegNet [42], and DeepLab series [43]

4.1.1 U-Net

U-Net, an enhanced version of the Fully Convolutional Network (FCN), addresses the issue of the sliding-window approach, which requires large patches for training by using a sliding window to provide pixels. This also causes redundancy due to overlap-ping patches, resulting in overfitting as the same features are trained multiple times. Its network structure is shown in Figure 7. U-Net achieves precise segmentation even with a small training dataset, offering advantages such as accurate target localization and fast network speed. On a single GPU core, it takes less than one second to segment a 512×512 image. This feature has made U-Net and its improved versions widely used in post-earthquake building damage detection. Additionally, the structure has a strong ability to extract local information, making it effective for building localization and classification in images [44].

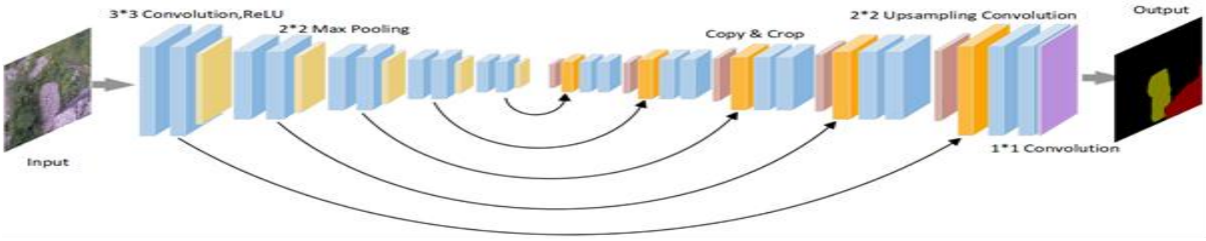

Figure. 7 Basic U-Net Network Structure Diagram.

Due to regional differences in building structures and materials, the characteristics of houses vary, and the damage to buildings becomes more complex after an earthquake. The internal feature extraction mechanism of U-Net is relatively weak in capturing complex image features, which can lead to the loss of accurate earthquake damage information. Khankeshizadeh et al. (2024) [45]designed a U-Net-based model that integrates weighted ResNet34, Vgg16, and InceptionV3 model weights to overcome the issue of U-Net losing damage feature information and further improve the model's generalization ability. Wang et al. (2021) [46]integrated an Object Contextual Attention (OCR) module into the fourth layer shortcut connection of U-Net to consolidate pixel-level context, thereby enhancing pixel representations. This approach helps the model recover lost damaged details of buildings in post-earthquake scenes, where surface features of collapsed, damaged, and intact buildings are mixed.

4.1.2 DeepLab

The DeepLab series of networks has evolved from V1, V2, and V3 to the latest DeepLabV3+. Figure 8 shows the network structure of DeepLabV3+. Its architecture includes a depthwise separable convolution module in the encoder part, an ASPP (Atrous Spatial Pyramid Pooling) module, and a decoder module for fusing shallow and deep features.

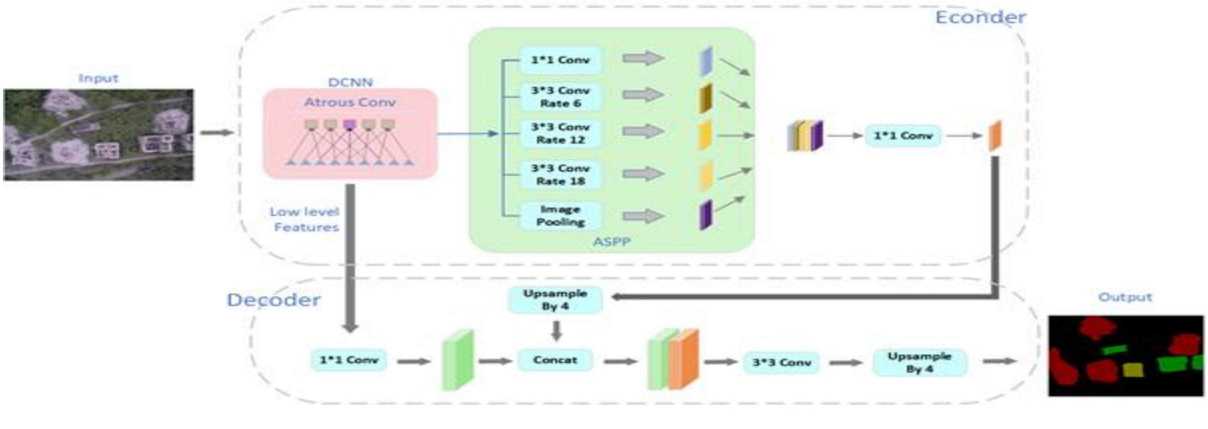

Figure. 8 DeepLabV3+ Network Structure Diagram.

The DeepLab series networks enhance their ability to capture contextual information and effectively preserve spatial features by introducing dilated convolutions to expand the receptive field. DeepLabV3+ further optimizes the ASPP (Atrous Spatial Pyramid Pooling) module from DeepLabV3 to improve global context extraction and incorporate depthwise separable convolutions to boost computational efficiency. It us-es an encoder-decoder architecture to fuse low-level features extracted by depthwise separable convolutions with high-level features extracted by the ASPP module. The DeepLab network architecture is highly effective in capturing detailed damage features in complex post-earthquake scenes. Its powerful multi-scale information extraction ability allows it to support mixed training of remote sensing data from various sources in earthquake-affected areas.

The large number of parameters in deep neural networks results in higher time and space overheads for the DeepLab network compared to other networks. However, earthquake rescue tasks are time-sensitive, and rapid building damage detection is crucial for emergency response. Therefore, when applying the DeepLab network or its improved versions for post-earthquake building damage detection, it is essential to find a balance between time overhead and detection accuracy. In addition, dilated convolutions sample pixels with gaps by skipping, which can easily result in a significant loss of boundary information when segmenting buildings. Song et al. (2020) [47]proposed a method based on DeepLabV2, incorporating superpixel segmentation (SLIC) to accurately extract the regional boundaries of earthquake-damaged buildings. Wang et al. (2021) [46]integrated the object context attention (OCR) module into the DeepLabV3+ decoder, after fusing shallow and deep features, to enhance the feature representation of object regions. They designed a new boundary enhancement loss function (BE Loss) aimed at further refining the boundary segmentation of damaged buildings and reducing the loss of boundary information. This method effectively reduced the false detection rate of collapsed buildings while significantly improving the extraction of building contours and the accurate localization of boundaries.

4.2 Object Detection Network

In building damage assessment tasks, object detection models determine the precise location of buildings by predicting bounding boxes and labeling the damage level of each building. Object detection models such as Faster R-CNN [48], SSD [49]and the YOLO series [50]are widely used in these task

4.2.1 Faster R-CNN

The Faster R-CNN network structure consists of the following components: (1) Backbone: Convolutional layers for feature extraction, responsible for generating multi-level feature maps from the image. (2) RPN: Generates candidate detection boxes. (3) ROI: Filters candidate regions that may contain the target and extracts fixed-size features from the candidate boxes. (4) Classification: Classifies the candidate regions, predicts their category, and performs regression on the detection boxes to precisely adjust their positions.

The classic two-stage object detection network, Faster R-CNN, generates high-quality object region proposals through the Region Proposal Network (RPN), which predicts bounding boxes. It performs exceptionally well in target localization and classification accuracy. Faster R-CNN can run at near-real-time frame rates, which is of significant practical importance for quickly detecting earthquake-damaged buildings. Additionally, the backbone network of Faster R-CNN is highly flexible, allowing for the use of various pre-trained networks as feature extractors, thereby speeding up the training process and saving valuable time in earthquake damage assessment tasks [51].

The RPN network of Faster R-CNN is used to generate object region prediction bounding boxes. By default, it uses anchors with 3 scales and 3 aspect ratios, producing 9 rectangular object region proposal boxes. However, this multi-scale and multi-aspect ratio anchor setting can lead to overlapping issues between the candidate boxes. When collapsed or damaged buildings cover a large area with unclear boundaries, it is easy for multiple detection boxes to overlap in the target regions. Ding et al. (2020) [52]proposed the Intersection-over -Object ratio (IPO) to replace the traditional IoU and used this strategy to improve the Non-Maximum Suppression (NMS) algorithm. This was applied to Faster R-CNN to accurately and efficiently detect collapsed buildings in post-earthquake drone images. Experimental results demonstrate that this approach effectively minimizes the occurrence of overlapping bounding boxes.

4.2.2 YOLO Series

The real-time object detection system YOLO (You Only Look Once) has under-gone continuous iterations and optimizations since the release of its first version in 2015, and as of October 2024, it has evolved to YOLOv11. With outstanding real-time detection performance and high accuracy, the YOLO series has maintained strong momentum in its updates and has become a core mainstream model in the field of object detection [53]. As shown in Figure 9, the YOLOv5 network structure consists of four parts: (1) Input: Supports data augmentation techniques such as rotation, random cropping, brightness and contrast adjustment, and Mosaic to enhance the model's generalization ability. (2) Backbone: Based on the Focus module and CSPDarknet53 structure, the backbone network is used to extract deep features. (3) Neck: Uses SPP and

FPN+PAN structures to fuse multi-scale features, improving object detection performance. (4) Output: Outputs the predicted results, including the confidence score, class probabilities, and anchor box coordinates.

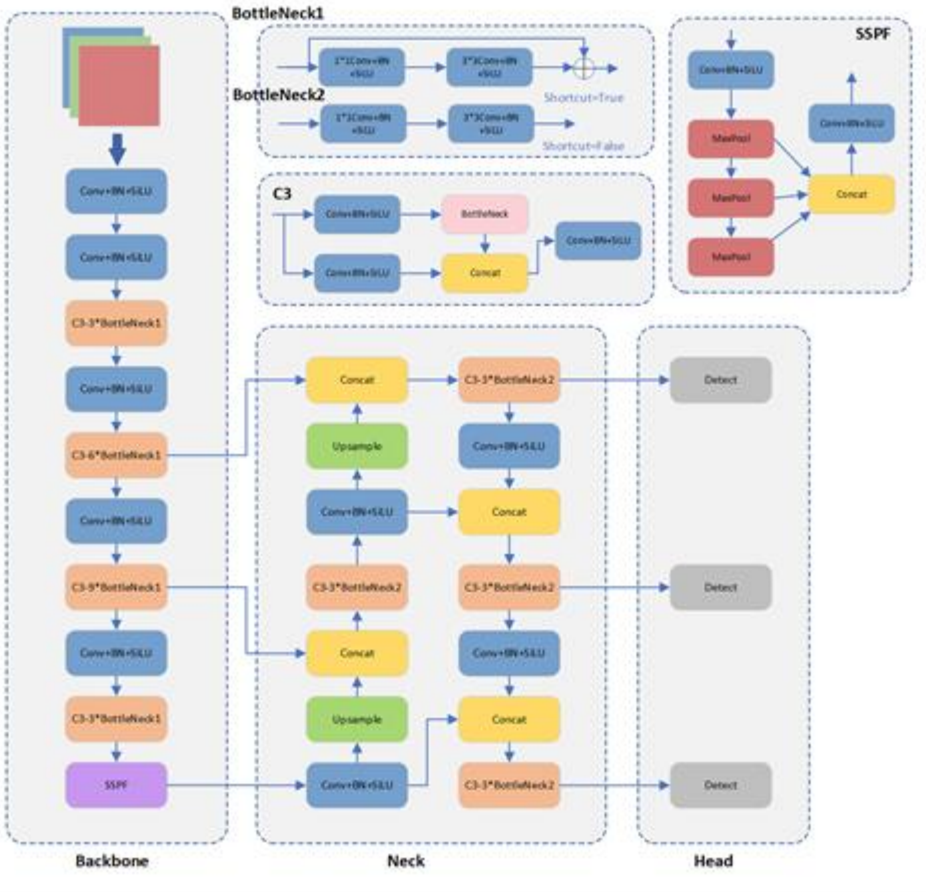

Figure. 9 Yolov5 Network Structure Diagram.

Unlike traditional object detection models that need several evaluations of an im-age to detect objects, YOLO series networks use a single-pass detection method, significantly improving detection speed. Due to their strong real-time processing and object detection capabilities in dynamic environments, YOLO networks are widely used in post-earthquake emergency response and building damage detection research. Wang et al.(2023) [54]applied the improved YOLOv5s model to an embedded system for drone imagery, enabling real-time detection of damaged building areas. However, in the complex scenarios of post-earthquake areas and multi-target building damage detection, factors such as significant differences in building sizes and a large number of targets present important challenges when applying object detection models. YOLOv3 and its subsequent networks, with their Anchor mechanism, global regression methods, and multi-detection head design, exhibit outstanding multi-scale and multi-target detection capabilities, giving them a clear advantage over many other object detection networks.

While YOLO prioritizes detection speed and real-time performance, it compromises detection accuracy. Despite its fast detection speed, its localization accuracy is lower than two-stage object detection methods, particularly in detecting small and densely located damaged buildings in post-earthquake scenarios [55]. In earthquake damage detection, optimizing the balance between YOLO's detection accuracy and speed is a key direction for improving detection efficiency. Ma et al. (2019) [56]replaced the YOLOv3 backbone feature extraction network with the lighter ShuffleNet v2, which improves both detection accuracy and speed. Shi et al. (2021) [57]used ResNext as the backbone of YOLOv4, reducing network complexity and improving feature ex -traction capabilities. They introduced a new Focal_EIOU_loss function to optimize bounding box regression. Jing et al. (2022) [58]adopted a bidirectional feature pyramid network (BiFPN), while Wang et al. (2023) [54]introduced the BDCAM attention mod-ule to replace the PANet structure in YOLOv5. Both improvements significantly reduced the computational complexity of multi-scale feature fusion while effectively balancing computational overhead and accuracy. For small object detection, Zou et al. (2024) [59]based their work on the YOLOv5s-seg network and introduced the Attention Fusion Sequence (ASF) method, which improves the

Neck component to enhance the detection of small building targets in post-earthquake images. Ablation experiments showed that the ASF method improved the overall detection accuracy of the model.

4.3 Instance Segmentation Network

4.3.1 Mask R-CNN

Instance segmentation tasks can be viewed as a combination of semantic segmentation and object detection. The goal is to precisely segment each building instance in an image, locate its position using bounding boxes, and assign a damage degree label to each building. Classic instance segmentation models commonly used for detecting earthquake-damaged buildings include Mask R-CNN [60]and the Yolo-seg series.

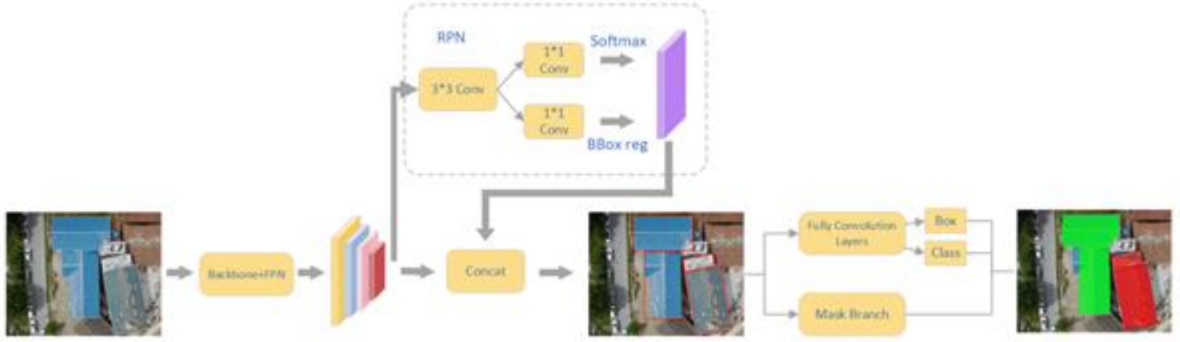

Figure. 10 Mask-R-CNN Network Structure Diagram.

Mask R-CNN is an enhanced version of the Faster R-CNN object detection net-work. Its network architecture is similar to Faster R-CNN, as shown in Figure 10. Mask R-CNN introduces the RoIAlign layer to replace the RoIPool layer in Faster R-CNN, aiming to achieve more precise feature alignment for the candidate proposal regions generated by the RPN. In addition, the Mask R-CNN model adds a parallel branch to the classification and bounding box prediction branches of Faster R-CNN to predict object masks.

Mask R-CNN has excelled in instance segmentation tasks and has surpassed the champions of the COCO 2015 and COCO 2016 Segmentation Challenges since its debut. This model is not only fast and accurate but also capable of simultaneously performing both object detection and semantic segmentation tasks, providing more comprehensive disaster information for the assessment of damaged buildings, such as the degree of damage, the location of the damage, and the area of the damage. Yildirim et al. (2023) [1]used Mask R-CNN to identify collapsed buildings in remote sensing data. By optimizing the backbone network and training parameter combinations, the best model achieved detection results of AP = 81.28% for intact buildings and AP = 69.26% for col-lapsed buildings. Zhan et al.(2022) [61]improved Mask R-CNN by replacing the FPN feature pyramid structure with a Path Aggregation Network (PANet), aiming to reduce the loss of useful information in multi-scale feature extraction, thereby improving detection accuracy. Additionally, they utilized the NMS algorithm to effectively sup-press the generation of overlapping bounding boxes.

5. Challenges and Strategies

With the rapid development and continuous iteration of CNNs in computer vision and disaster image processing, researchers are expected to keep focusing on applying CNN methods to post-earthquake building damage detection tasks. However, there are still several unique challenges in this task, such as small dataset sizes, low-quality training datasets, class imbalance in classification instances, and the fusion of features from images at different scales. Therefore, there remains significant room for improvement in applying CNNs to building damage detection tasks.

5.1 Dataset scarcity

In building damage detection tasks, there is limited availability of publicly accessible image data from earthquake-affected areas [62], and labeled image datasets are extremely scarce. The generalization ability and robustness of deep learning models are often affected by the diversity and quantity of the training datasets [63]. A lack of training data may result in poor CNN model performance and an increased risk of overfitting. Data augmentation and transfer learning can help alleviate the problems to some extent, improving the model's performance.

5.1.1 Data augmentation

Data augmentation is an efficient and practical method for expanding datasets. Common techniques include simple translations, scaling, rotations, and horizontal or vertical flips applied to the existing training datasets.

These operations encourage the network model to learn repeatedly from the limited data, improving the model's detection accuracy for buildings and their damaged features. Data augmentation methods such as brightness transformations, uniform or non-uniform fogging, and noise reduction can effectively simulate common scenarios found in disaster zone imagery, including cloud cover, harsh weather conditions, and extreme lighting situations (either too bright or too dim). These methods not only help expand the dataset but also significantly enhance the model's exposure robustness under varying lighting conditions, improving its ability to detect in complex environments [64]. Moreover, recent innovative data augmentation methods in image classification and object detection re-search, such as CutMix [65]—where a random region is cut from one image and pasted into a specified location on another image to generate a new image, GridMask [66]—where a set of evenly distributed square regions is removed from the original training image, and Random Erasing [67]—where a random rectangular region is selected, and its pixel values are replaced with random values or the region's average value, offer fresh approaches to address the challenge of limited earthquake image datasets and demonstrate promising research and application prospects.

5.1.2 Transfer learning

Transfer learning seeks to enhance performance or reduce the need for labeled samples in the target domain by utilizing knowledge from the source domain [68]. In transfer learning, the use of pre-trained models is the most common approach. Utilizing pre-trained models not only saves training time but also significantly enhances the model's generalization ability. On the other hand, fine-tuning with a smaller set of disaster zone image data enables the rapid development of an efficient damage detection model, effectively overcoming the challenges posed by limited datasets and avoiding the difficulty of training a high-accuracy model due to insufficient data. Wu et al. (2019) [69]and Nag et al. (2020) [70]fine-tuned the VGG16 model pre-trained o n ImageNet with disaster-related images, and the resulting collapsed building detection models showed improvements in overall accuracy and classification capability. Yang et al. (2021) [71]fine-tuned the ImageNet pre-trained DenseNet121 with the xBD dataset, and the results indicated that the model had strong geographical transferability and performed best in detection. This demonstrates the effectiveness of transfer learning in addressing challenges such as overfitting and poor generalization caused by limited disaster zone datasets. During post-earthquake response, rescue evaluation, and dam-age detection phases, pre-trained models can not only help save valuable rescue time but also significantly reduce the workload of damage detection.

5.2 Class imbalance

In the task of post-earthquake building collapse and damage detection, most datasets are characterized by a higher number of intact buildings than damaged ones. This causes the trained models to favor detecting intact buildings, leading to significant false positives and reducing the classification accuracy for damaged buildings. Common solutions to address the class imbalance problem include data-level methods and algorithm-level methods.

At the data processing level, oversampling strategies can be used to increase the number of damaged building instances or undersampling strategies can reduce the number of intact building instances to balance the frequency of training on features of buildings with different damage levels. This helps the model better focus on the features of damaged buildings, improving both the balance and accuracy of classification. At the algorithmic level, cost-sensitive learning assigns different loss weights to different classes in the dataset based on the importance or imbalance of the damage categories. This adjustment of the loss function allows damaged and collapsed features to contribute more significantly during the network weight update process [[72],[73]] . Ensemble learning optimizes the model's ability to recognize and detect minority classes by integrating multiple base classifiers and combining strategies like resampling and weighting, which can effectively alleviate the issue of class imbalance [74]. Learning strategy design methods can also be applied to address class imbalance by improving the learning approach and optimizing the learning process [75], aiming to balance the training of different class instances input into the model [76].

5.3 Adaptability of CNN models

Most mainstream CNNs are not fully adapted for post-earthquake building dam-age detection tasks due to the specific nature of the damage features. On one hand, from a size perspective, this is a small object detection task. On the other hand, the geometric features of intact buildings are generally more regular and have clear boundaries. However, after an earthquake, damaged buildings may exhibit various degrees of de-formation, with debris such as broken stones and falling building fragments scattered around undamaged units. By constructing CNN models based on the special characteristics of detection tasks and applying improved modules to mainstream models, promising results have been achieved, showing great potential for future development.

To address the irregularity of collapsed buildings, deformable convolution (DCN) can be used to replace traditional convolution layers, enhancing the extraction of irregular features from collapsed buildings [52]. The attention module shifts the model's focus to the most important regions of the image (damaged buildings), which can effectively improve the detection performance for densely packed and smaller targets. Modules like ASPP, SPPF [77], and dilated convolution operations can effectively enlarge the receptive field of the CNN model, enhancing its ability to capture multi-scale contextual information in feature maps [43]. In the task of detecting damaged buildings, embedding these modules into the network can further improve the model's ability to identify subtle damage features. Particularly in a four-class classification scenario, this method helps to enhance the recognition accuracy for the categories of lightly damaged and severely damaged buildings, thus improving the overall detection accuracy of the model.

6. Conclusions

This review provides an overview of the current progress in earthquake-damaged building detection research based on CNNs. It systematically reviews aspects such as datasets, models, and methods, while highlighting the limitations of CNNs in this task and the various challenges they face. The review also proposes strategies to address these challenges, offering important references for future earthquake rescue, disaster area assessment, and post-disaster reconstruction efforts.

With the continuous upgrade of sensor devices and the significant improvement in computing power, deep learning CNNs will demonstrate even greater potential in the future. The combination of remote sensing imagery and CNNs for earthquake-damaged building detection still has substantial room for improvement.

Currently, relying solely on 2D information to assess building damage has significant limitations. However, the integration of deep learning-based methods such as Multi-View Stereo (MVS) [78], NeRF [79], and 3 D Gaussian Splatting [80], along with low-altitude drone oblique multi-view images and multi-view satellite remote sensing imagery for 3D reconstruction in disaster zones, will enable more detailed and com-prehensive building damage assessments in the future. This approach will become a key development direction for earthquake-damaged building detection. The Vision Trans-former (ViT), based on the self-attention mechanism, has proven to surpass traditional CNNs in large-scale training [81]. Its attention modules can completely replace convolutional operations. Applying ViT or its attention-based modules to post-earthquake building damage detection tasks can offer new insights for the field. Additionally, the self-supervised pre-training phase of ViT on large-scale datasets can reduce the work-load of manual annotations and save rescue time [82]. Fine-tuning the network on medium and small-sized datasets can also address the challenge of scarce building data in earthquake-affected areas. Moreover, the rapid development of large models in areas like natural language processing, computer vision, and speech recognition has showcased their strong learning and generalization capabilities. These models can accurately predict unseen data and learn complex high-level features from vast amounts of data. The construction of large models for disaster prediction, post-disaster rescue, damage assessment, and building damage classification detection has broad future development prospects.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

The authors acknowledge the National Key Research and Development Programs of China (Grant: 2022YFC3004405), the National Key Research and Development Programs of China (Grant: 2022YFC3004405).

References

[1] Yildirim, E.; Kavzoglu, T. Detection of collapsed buildings from post-earthquake imagery using mask region-based convolutional neural network. Intercontinental Geoinformation Days 2023, 7, 119-122.
[2] Tasci, B.; Acharya, M.R.; Baygin, M.; Dogan, S.; Tuncer, T.; Belhaouari, S.B. InCR: Inception and concatenation residual block-based deep learning network for damaged building detection using remote sensing images. International Journal of Applied Earth Observation and Geoinformation 2023, 123, 103483.
[3] Wiguna, S.; Adriano, B.; Vescovo, R.; Mas, E.; Mizutani, A.; Koshimura, S. Building Damage Mapping of the 2024 Noto Peninsula Earthquake, Japan Using Semi-supervised Learning and VHR Optical Imagery. IEEE Geoscience and Remote Sensing Letters 2024.
[4] Zhang, H.; Ma, G.; Fan, H.; Gong, H.; Wang, D.; Zhang, Y. SDCINet: A novel cross-task integration network for segmentation and detection of damaged/changed building targets with optical remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing 2024, 218, 422-446.

[5] Qing, Y.; Ming, D.; Wen, Q.; Weng, Q.; Xu, L.; Chen, Y.; Zhang, Y.; Zeng, B. Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level. International Journal of Applied Earth Observation and Geoinformation 2022, 112, 102899.

[6] Ritwik, G.; Richard, H.; Sandra, S.; Nirav, P.; Bryce, G.; Jigar, D.; Eric, H.; Howie, C.; Matthew, G. xbd: A dataset for assessing building damage from satellite imagery. arXiv preprint 2019, 1-9.

[7] Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. Remote sensing 2018, 10, 1636.

[8] LI, D. From the luojia series satellites to the oriental smart eye constellation. Geomatics and Information Science of Wuhan University 2023, 48, 1557-1565.

[9] Li, D.; Wang, M.; Yang, F.; Dai, R. Internet intelligent remote sensing scientific experimental satellite LuoJia3-01. Geo-Spatial Information Science 2023, 26, 257-261.

[10] Xie, Y. Deep Learning in Earthquake Engineering: A Comprehensive Review. arXiv preprint arXiv:2405.09021 2024.

[11] 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 Sensing 2019, 11, 1202.

[12] Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A.; Zhang, L. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. Remote Sensing of Environment 2021, 265, 112636.

[13] Potlapally, A.; Chowdary, P.S.R.; Shekhar, S.R.; Mishra, N.; Madhuri, C.S.V.D.; Prasad, A. Instance segmentation in remote sensing imagery using deep convolutional neural networks. In Proceedings of the 2019 International Conference on Contemporary Computing and Informatics (IC3I), 2019, 117-120.

[14] Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G. Satellite image classification of building damages using airborne and satellite image samples in a deep learning approach. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2018, 4, 89-96.

[15] Matin, S.S.; Pradhan, B. Challenges and limitations of earthquake-induced building damage mapping techniques using remote sensing images-A systematic review. Geocarto International 2022, 37, 6186-6212.

[16] Adriano, B.; Yokoya, N.; Xia, J.; Miura, H.; Liu, W.; Matsuoka, M.; Koshimura, S. Learning from multimodal and multitemporal earth observation data for building damage mapping. ISPRS Journal of Photogrammetry and Remote Sensing 2021, 175, 132-143.

[17] Wang, X.; Li, P. Extraction of urban building damage using spectral, height and corner information from VHR satellite images and airborne LiDAR data. ISPRS Journal of Photogrammetry and Remote Sensing 2020, 159, 322-336.

[18] Nex, F.; Duarte, D.; Tonolo, F.G.; Kerle, N. Structural building damage detection with deep learning: Assessment of a state-of-the-art CNN in operational conditions. Remote sensing 2019, 11, 2765.

[19] Amini Amirkolaee, H.; Arefi, H. CNN-based estimation of pre-and post-earthquake height models from single optical images for identification of collapsed buildings. Remote Sensing Letters 2019, 10, 679-688.

[20] Ural, S.; Hussain, E.; Kim, K.; Fu, C.-S.; Shan, J. Building extraction and rubble mapping for city port-au-prince post-2010 earthquake with GeoEye-1 imagery and lidar data. Photogrammetric Engineering & Remote Sensing 2011, 77, 1011-1023.

[21] Shi, W.; Zhang, M.; Zhang, R.; Chen, S.; Zhan, Z. Change detection based on artificial intelligence: State-of-the-art and challenges. Remote Sensing 2020, 12, 1688.

[22] Ge, P.; Gokon, H.; Meguro, K. A review on synthetic aperture radar-based building damage assessment in disasters. Remote Sensing of Environment 2020, 240, 111693.

[23] Bai, Y.; Adriano, B.; Mas, E.; Gokon, H.; Koshimura, S. Object-based building damage assessment methodology using only post event ALOS-2/PALSAR-2 dual polarimetric SAR intensity images. Journal of Disaster Research 2017, 12, 259-271.

[24] Bai, Y.; Adriano, B.; Mas, E.; Koshimura, S. Machine learning based building damage mapping from the ALOS-2/PALSAR-2 SAR imagery: Case study of 2016 Kumamoto earthquake. Journal of Disaster Research 2017, 12, 646-655.

[25] Karimzadeh, S.; Mastuoka, M. Building damage assessment using multisensor dual-polarized synthetic aperture radar data for the 2016 M 6.2 Amatrice Earthquake, Italy. Remote Sensing 2017, 9, 330.

[26] Yu, Z.; Chen, Z.; Sun, Z.; Guo, H.; Leng, B.; He, Z.; Yang, J.; Xing, S. Segdetector: A deep learning model for detecting small and overlapping damaged buildings in satellite images. Remote Sensing 2022, 14, 6136.

[27] Zhang, H.; Wang, M.; Zhang, Y.; Ma, G. TDA-Net: A novel transfer deep attention network for rapid response to building damage discovery. Remote Sensing 2022, 14, 3687.

[28] Shen, Y.; Zhu, S.; Yang, T.; Chen, C.; Pan, D.; Chen, J.; Xiao, L.; Du, Q. Bdanet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images. IEEE Transactions on Geoscience and Remote Sensing 2021, 60, 1-14.
[29] Ahmadi, S.A.; Mohammadzadeh, A.; Yokoya, N.; Ghorbanian, A. BD-SKUNet: Selective-kernel UNets for building damage assessment in high-resolution satellite images. Remote Sensing 2024, 16, 182.
[30] He, Y.; Wang, J.; Liao, C.; Zhou, X.; Shan, B. MS4D-Net: Multitask-Based Semi-Supervised Semantic Segmentation Framework with Perturbed Dual Mean Teachers for Building Damage Assessment from High-Resolution Remote Sensing Imagery. Remote Sensing 2023, 15, 478.
[31] Zhang, Y.; Wang, Z.; Luo, Y.; Yu, X.; Huang, Z. Learning Efficient Unsupervised Satellite Image-based Building Damage Detection. In Proceedings of the 2023 IEEE International Conference on Data Mining (ICDM), 2023, 1547-1552.
[32] Moradi, M.; Shah-Hosseini, R. Earthquake damage assessment based on deep learning method using VHR images. Environmental Sciences Proceedings 2020, 5, 16.
[33] Miyamoto, T.; Yamamoto, Y. Using 3-D convolution and multimodal architecture for earthquake damage detection based on satellite imagery and digital urban data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2021, 14, 8606-8613.
[34] Chen, X. Using satellite imagery to automate building damage assessment: A case study of the xbd dataset. In Proceedings of the IISE 2020 Annual Meeting, 2021.
[35] Bouchard, I.; Rancourt, M.-È.; Aloise, D.; Kalaitzis, F. On transfer learning for building damage assessment from satellite imagery in emergency contexts. Remote Sensing 2022, 14, 2532.
[36] Kalantar, B.; Ueda, N.; Al-Najjar, H.A.; Halin, A.A. Assessment of convolutional neural network architectures for earthquake-induced building damage detection based on pre-and post-event orthophoto images. Remote Sensing 2020, 12, 3529.
[37] Weber, E.; Kané, H. Building disaster damage assessment in satellite imagery with multi-temporal fusion. arXiv preprint arXiv:2004.05525 2020.
[38] May, S.; Dupuis, A.; Lagrange, A.; De Vieilleville, F.; Fernandez-Martin, C. Building damage assessment with deep learning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2022, 43, 1133-1138.
[39] Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A.; Zhang, L. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. Remote Sensing of Environment 2021, 265, 112636.
[40] Xu, J.Z.; Lu, W.; Li, Z.; Khaitan, P.; Zaytseva, V. Building damage detection in satellite imagery using convolutional neural networks. arXiv preprint arXiv:1910.06444 2019.
[41] 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, October 5-9, 2015, proceedings, part III 18, 2015, 234-241.
[42] Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 2017, 39, 2481-2495.
[43] Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 2017, 40, 834-848.
[44] Salem, M.; Gomaa, A.; Tsurusaki, N. Detection of earthquake-induced building damages using remote sensing data and deep learning: A case study of mashiki town, japan. In Proceedings of the IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium, 2023, 2350-2353.
[45] Khankeshizadeh, E.; Mohammadzadeh, A.; Arefi, H.; Mohsenifar, A.; Pirasteh, S.; Fan, E.; Li, H.; Li, J. A novel weighted ensemble transferred U-net based model (WETUM) for post-earthquake building damage assessment from UAV data: A comparison of deep learning-and machine learning-based approaches. IEEE Transactions on Geoscience and Remote Sensing 2024.
[46] Wang, C.; Qiu, X.; Huan, H.; Wang, S.; Zhang, Y.; Chen, X.; He, W. Earthquake-damaged buildings detection in very high-resolution remote sensing images based on object context and boundary enhanced loss. Remote Sensing 2021, 13, 3119.
[47] Song, D.; Tan, X.; Wang, B.; Zhang, L.; Shan, X.; Cui, J. Integration of super-pixel segmentation and deep-learning methods for evaluating earthquake-damaged buildings using single-phase remote sensing imagery. International Journal of Remote Sensing 2020, 41, 1040-1066.
[48] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence 2016, 39, 1137-1149.
[49] 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, October 11–14, 2016, Proceedings, Part I 14, 2016, 21-37.

[50] Varghese, R.; Sambath, M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024, 1-6.
[51] Hacıefendioğlu, K.; Başağa, H.B.; Demir, G. Automatic detection of earthquake-induced ground failure effects through Faster R-CNN deep learning-based object detection using satellite images. Natural Hazards 2021, 105, 383-403.
[52] Ding, J.; Zhang, J.; Zhan, Z.; Tang, X.; Wang, X. A precision efficient method for collapsed building detection in post-earthquake UAV images based on the improved NMS algorithm and faster R-CNN. Remote Sensing 2022, 14, 663.
[53] Vijayakumar, A.; Vairavasundaram, S. Yolo-based object detection models: A review and its applications. Multimedia Tools and Applications 2024, 1-40.
[54] Wang, Y.; Feng, W.; Jiang, K.; Li, Q.; Lv, R.; Tu, J. Real-time damaged building region detection based on improved YOLOv5s and embedded system from UAV images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2023, 16, 4205-4217.
[55] Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. Proceedings of the IEEE 2023, 111, 257-276.
[56] Ma, H.; Liu, Y.; Ren, Y.; Yu, J. Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3. Remote Sensing 2019, 12, 44.
[57] Shi, L.; Zhang, F.; Xia, J.; Xie, J.; Zhang, Z.; Du, Z.; Liu, R. Identifying damaged buildings in aerial images using the object detection method. Remote Sensing 2021, 13, 4213.
[58] Jing, Y.; Ren, Y.; Liu, Y.; Wang, D.; Yu, L. Automatic extraction of damaged houses by earthquake based on improved YOLOv5: A case study in Yangbi. Remote Sensing 2022, 14, 382.
[59] Zou, R.; Liu, J.; Pan, H.; Tang, D.; Zhou, R. An improved instance segmentation method for fast assessment of damaged buildings based on post-earthquake uav images. Sensors 2024, 24, 4371.
[60] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017, 2961-2969.
[61] 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 Sensing 2022, 14, 1002.
[62] Liu, C.; Sepasgozar, S.M.; Zhang, Q.; Ge, L. A novel attention-based deep learning method for post-disaster building damage classification. Expert Systems with Applications 2022, 202, 117268.
[63] Luo, Z.; Yang, W.; Yuan, Y.; Gou, R.; Li, X. Semantic segmentation of agricultural images: A survey. Information Processing in Agriculture 2023.
[64] Cui, L.; Jing, X.; Wang, Y.; Huan, Y.; Xu, Y.; Zhang, Q. Improved swin transformer-based semantic segmentation of postearthquake dense buildings in urban areas using remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2022, 16, 369-385.
[65] Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019; pp. 6023-6032.
[66] Chen, P.; Liu, S.; Zhao, H.; Wang, X.; Jia, J. Gridmask data augmentation. arXiv preprint arXiv:2001.04086 2020.
[67] Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2020, 13001-13008.
[68] Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. Proceedings of the IEEE 2020, 109, 43-76.
[69] Wu, F.; Wang, C.; Zhang, B.; Zhang, H.; Gong, L. Discrimination of collapsed buildings from remote sensing imagery using deep neural networks. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, 2646-2649.
[70] Nag, S.; Pal, T.; Basu, S.; Das Bit, S. CNN based approach for post disaster damage assessment. In Proceedings of the Proceedings of the 21st International Conference on Distributed Computing and Networking, 2020, 1-6.
[71] Yang, W.; Zhang, X.; Luo, P. Transferability of convolutional neural network models for identifying damaged buildings due to earthquake. Remote Sensing 2021, 13, 504.
[72] Ghosh Mondal, T.; Jahanshahi, M.R.; Wu, R.T.; Wu, Z.Y. Deep learning-based multi-class damage detection for autonomous post-disaster reconnaissance. Structural Control and Health Monitoring 2020, 27, e2507.
[73] Ji, M.; Liu, L.; Buchroithner, M. Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 Haiti earthquake. Remote Sensing 2018, 10, 1689.

[74] Fernández-Gómez, M.J.; Asencio-Cortés, G.; Troncoso, A.; Martínez-Álvarez, F. Large earthquake magnitude prediction in Chile with imbalanced classifiers and ensemble learning. Applied Sciences 2017, 7, 625.
[75] Ge, J.; Tang, H.; Ji, C. Self-incremental learning for rapid identification of collapsed buildings triggered by natural disasters. Remote Sensing 2023, 15, 3909.
[76] Wang, Y.; Chew, A.W.Z.; Zhang, L. Building damage detection from satellite images after natural disasters on extremely imbalanced datasets. Automation in construction 2022, 140, 104328.
[77] Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 2020.
[78] Hong, Z.; Yang, Y.; Liu, J.; Jiang, S.; Pan, H.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S. Enhancing 3D reconstruction model by deep learning and its application in building damage assessment after earthquake. Applied Sciences 2022, 12, 9790.
[79] Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM 2021, 65, 99-106.
[80] Yu, Z.; Chen, A.; Huang, B.; Sattler, T.; Geiger, A. Mip-splatting: Alias-free 3d gaussian splatting. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, 19447-19456.
[81] Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. ACM computing surveys (CSUR) 2022, 54, 1-41.
[82] Saad, O.M.; Chen, Y.; Savvaidis, A.; Fomel, S.; Chen, Y. Real-time earthquake detection and magnitude estimation using vision transformer. Journal of Geophysical Research: Solid Earth 2022, 127, e2021JB023657.