applied sciences

MDPI

Article

Enhancing 3D Reconstruction Model by Deep Learning and Its Application in Building Damage Assessment after Earthquake

Zhonghua Hong [1], Yahui Yang [1], Jun Liu [2,3,*], Shenlu Jiang [4,*], Haiyan Pan [1], Ruyan Zhou [1], Yun Zhang [1], Yanling Han [1], Jing Wang [1], Shuhu Yang [1] and Changyue Zhong [3]

1 College of Information Technology, Shanghai Ocean University, Shanghai 201306, China
2 National Earthquake Response Support Service, Beijing 100049, China
3 College of Civil Engineering and Architecture, Guizhou Minzu University, Guiyang 550025, China
4 School of Computer Science and Engineering, Faculty of Innovation Technology, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau SAR, China
* Correspondence: liujun_eq@sina.com (J.L.); shenlujiang@must.edu.mo (S.J.)

Abstract: A timely and accurate damage assessment of buildings after an earthquake is critical for the safety of people and property. Most of the existing methods based on classification and segmentation use two-dimensional information to determine the damage level of the buildings, which cannot provide the multi-view information of the damaged building, resulting in inaccurate assessment results. According to the knowledge of the authors, there is no related research using the deep-learning-based 3D reconstruction method for the evaluation of building damage. In this paper, we first applied the deep-learning-based MVS model to reconstruct the 3D model of the buildings after an earthquake using multi-view UAV images, to assist the building damage assessment task. The method contains three main steps. Firstly, the camera parameters are calculated. Then, 3D reconstruction is conducted based on CasMVSNet. Finally, a building damage assessment is performed based on the 3D reconstruction result. To evaluate the effectiveness of the proposed method, the method was tested in multi-view UAV aerial images of Yangbi County, Yunnan Province. The results indicate that: (1) the time efficiency of CasMVSNet is significantly higher than that of other deep learning models, which can meet the timeliness requirement of post-earthquake rescue and damage assessment. In addition, the memory consumption of CasMVSNet is the lowest; (2) CasMVSNet exhibits the best 3D reconstruction result in both high and small buildings; (3) the proposed method can provide detail and multi-view information of damaged buildings, which can be used to assist the building damage assessment task. The results of the building damage assessment are very similar to the results of the field survey.

Keywords: multi-view UAV images; deep learning; CasMVSNet; building damage classification

Citation: Hong, Z.; Yang, Y.; Liu, J.; Jiang, S.; Pan, H.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S.; et al. Enhancing 3D Reconstruction Model by Deep Learning and Its Application in Building Damage Assessment after Earthquake. *Appl. Sci.* **2022**, *12*, 9790. https://doi.org/10.3390/app12199790

Academic Editors: Anselme Muzirafuti, Giovanni Randazzo and Stefania Lanza

Received: 1 September 2022
Accepted: 25 September 2022
Published: 28 September 2022

1. Introduction

Earthquakes are one of the most serious natural disasters affecting humans. They cause many houses to be damaged and collapse, severely affecting the safety of both people and property. One of the key issues after an earthquake is the assessment of the damage of buildings. The results of the assessment can provide important information for disaster relief work. The timely and accurate assessment of damaged buildings is critical for rescues and consequential loss assessment.

Traditionally, post-earthquake building damage is evaluated and counted via manual field surveys, but this method is often time-consuming and laborious. Yamazaki et al. used the QuickBird satellite image after the Ms6.8 earthquake on the Mediterranean coast of Algeria and classified damaged buildings into five grades using the visual interpretation image method [1]. However, atmospheric conditions, such as cloud cover, will affect the image quality and lead to inaccurate evaluation.

With the development of artificial intelligence, machine-learning-related technologies have been gradually applied to the post-earthquake building damage assessment. Li et al. used remote sensing data before and after the earthquake through the decision tree method, in which the damaged buildings were divided into four grades [2]. The neural network of the genetic algorithm (GA) and the neural network composed of multi-layer perceptron (MLP) are used to predict the risk level of damage to reinforced-concrete (RC) structures [3,4]. The method achieves detailed investigation and inspection of buildings before the earthquake, reducing the loss of life and property. SMART SKY EYE (smart building safety assessment system using UAV) evaluates building wall cracks by analyzing natural factors based on machine learning methods such as random forest and support vector machine (SVM) [5]. However, these generalization capabilities are poor and the performance of the model is affected when the study area changes.

Compared with machine learning, convolutional neural networks (CNNs) based on deep learning have strong image processing abilities, strong feature learning and visual recognition abilities, and are widely used in building damage assessment. The dual-temporal methods use CNN to extract information on the characteristics of the images before and after the earthquake to determine the degree of damage to the building [6]. Ci et al. used deep-learning-based automatic detection and classification methods to evaluate and classify the loss levels of buildings in Ludian earthquake aerial images [7]. However, these methods can only achieve good performance when there are few categories, which cannot meet the needs for post-earthquake housing damage assessment. Ji et al. also combined machine learning and deep learning methods to evaluate five types of damage to buildings and improve the evaluation performance of damaged buildings using a combination of texture information from random forests and deep features extracted by CNN [8].

The above methods use two-dimensional semantic information to complete the damage assessment of the building, but they only contain damage information on one side of the building. Therefore, there is a big difference between the assessment results and the actual damage. On the contrary, three-dimensional semantic stereo information can provide structural features and height information of buildings. It is helpful to evaluate the damage grade of buildings after an earthquake. Mustafa et al. extracted the damaged information of buildings based on the differences in elevation between images before and after the earthquake [9]. However, the applicability of this method is limited due to the difficulty in obtaining pre-disaster and post-disaster digital elevation models (DEM). On the contrary, the 3D model efficiently reconstructed using the UAV can describe more detailed information of walls, beams, columns, and roofs from multiple angles [10,11]. Stepinac et al. used a laser scanner and a drone to generate 3D point clouds, after which the damage assessment of the building was performed by analyzing the three-dimensional structure of the building [12]. The scheme is expensive and is not suitable for large-scale 3D reconstruction. SMART SKY EYE used commercial software for 3D reconstruction and found defects in building structures using 3D models to complete the damage assessment of buildings [5]. However, the commercial software was developed from conventional methods [13–15], such as pix4d [16], smart 3d [17], and PhotoScan [18]. They improve the quality of 3D model, but the efficiency cannot meet the urgent needs of post-earthquake assessment.

In recent years, some multi-view stereo (MVS) networks based on deep learning have been widely used in 3D-reconstruction-related research [19]. The basic principle of MVS based on deep learning is to calculate the depth map of all images to complete the 3D reconstruction of the whole scene [20]. Based on the *DTU* dataset [21], MVSNet [20] completed the 3D reconstruction end-to-end for the first time. RMVSNet [22] used the GRU structure [23] to improve its regularization method, making large-scale 3D reconstruction possible. Subsequent improvements proposed by D^2HC-MVSNet [24], AA-RMVSNet [25], and CasMVSNet [26] have further improved network performance. Among them, CasMVSNet uses a multi-layer cascading method to compute the coarse-to-fine depth information [27,28], with higher computational efficiency and reconstruction quality.

However, all of the models mentioned above were tested in public datasets, which are mainly used to validate and evaluate different improved MVS networks. In addition, to the knowledge of the authors, there is little research that has used deep-learning-based MVS models to complete the damage assessment of buildings. The generalization ability of the model is the most important ability to move towards engineering applications, so it is necessary to use real post-earthquake image data analysis to evaluate the performance of the method. At the same time, these data also fill the gaps in the application of this method to the damage assessment of buildings after earthquakes [29,30]. Therefore, the objective of the study is to propose a deep-learning-based MVS method that is suitable for assessing building damage after an earthquake. Most importantly, the applicability of different 3D reconstruction models for post-earthquake building damage assessment is explored from the point view of time efficiency and the construction performance.

The remainder of this paper is organized as follows. Section 2 introduces the dataset used in this study. The methodology is presented in Section 3. In Section 4, the experimental details and results are shown. In Section 5, the performance of different methods is discussed. Finally, the conclusion is shown in Section 6.

2. Datasets

In the experiment, all MVS networks based on deep learning were trained on the public *DTU* dataset [21]. The data contained a wide range of scenarios, including housing models. The training data used 119 scenes, each containing 49 different view images with a pixel resolution of 640 × 512, with seven different intensity illuminations added to all images. The dataset was shot and calibrated by industrial manipulators, which can obtain high-precision camera parameters and improve the training of MVS networks. The dataset was downloaded from https://github.com/YoYo000/MVSNet (accessed on 29 August 2022) [20].

The dataset used in this study contained post-earthquake images of Yangbi County, Dali Prefecture, Yunnan Province, which occurred with an earthquake of magnitude 6.4 on 21 May 2021 with a focal depth of 8 km and an epicenter at 25.67 degrees north latitude and 99.87 degrees east longitude. As shown in Figure 1, a total of 411 UAV images with a pixel resolution of 5472 × 3648 (W × H) were obtained and 153 houses in the area were studied.

Figure 1. Multi-view UAV images of the earthquake area.

3. Methodology

The flow chart of the proposed method is shown in Figure 2, consisting of 3 steps: (1) calculation of camera parameters, (2) 3D reconstruction of MVS using deep learning method, (3) building damage assessment based on the result of the 3D reconstruction.

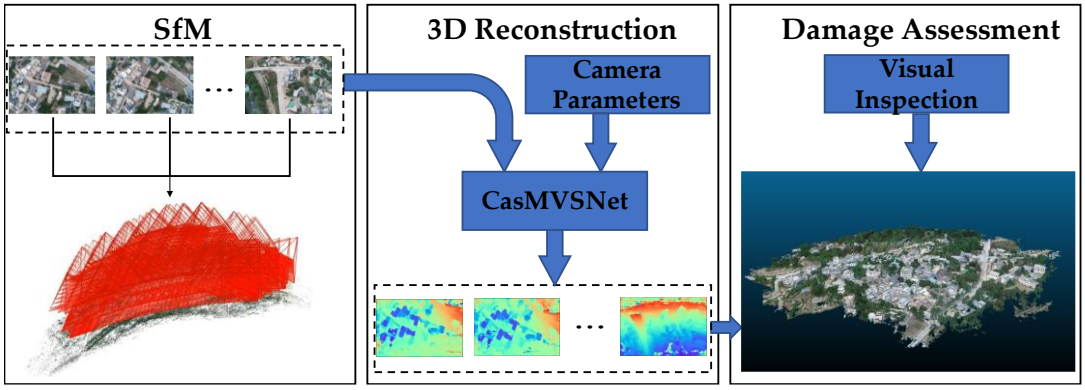


Figure 2. Workflow of building damage classification based on 3D model.

3.1. Calculation of Camera Parameters

When using the deep-learning-based MVS method to reconstruct the UAV image in the earthquake area, the COLMAP based on incremental SfM (structure-from-motion) [31] technology is used to complete the sparse reconstruction part to calculate camera parameters [20]. Incremental SfM is a processing method for sequential iterative reconstruction of 3D scenes. As shown in Figure 3, it usually starts with feature extraction and feature matching and then generates 3D models of scenes through geometric verification iteration. Next, the selected two-view is used as the basis of the model, and the registration of the new image is gradually added and the reconstruction is refined by triangulation and bundle adjustment (BA). After sparse reconstruction, the camera parameters of each image and the horizontal depth range of the model from the camera are output.

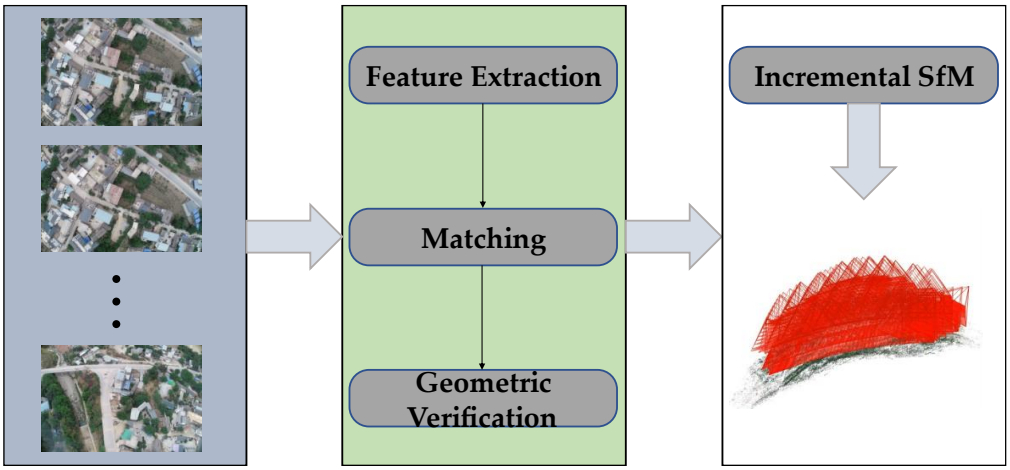


Figure 3. Principle of COLMAP [31].

The processed data are calibrated by the image and camera parameters, and there will be a slight deviation from the original image. In order to adapt to the input of each multi-view semantic stereo network, the image processed by COLMAP is preprocessed to a uniform size, and the image is restored to the original size. According to the image scaling ratio, the same scaling is performed on the camera internal parameters, including the main point offset coordinates and the camera focal length.

$$\frac{W}{W_0} = \frac{H}{H_0} = \frac{u}{u_0} = \frac{v}{v_0} = \frac{f}{f_0} \tag{1}$$

In Formula (1), W is the width of the image, H is the height of the image, u and v are the coordinates of the principal point of the image, and f is the focal length of the camera. Others are scaled corresponding parameters.

3.2. 3D Reconstruction Based on CasMVSNet

Considering the time cost and hardware equipment requirements for the three-dimensional modeling of UAV images, this study uses CasMVSNet [26] for 3D reconstruction due to its faster processing speed and higher accuracy. CasMVSNet extends the regularization of cost volume using 3D CNN [32] in MVSNet, which can better capture spatial feature information and use a multi-layer cascade method from coarse to fine. The multi-scale feature maps extracted from the feature extraction part are matched to construct cost volume with different resolutions. In the previous stage, the rough depth information was obtained via the calculation of the small-resolution feature map, and this is used to adaptively narrow the range of depth calculated by the higher resolution feature map in the next stage.

As shown in Figure 4, to obtain the multi-scale feature map and calculate the depth at different stages, the feature extraction part uses the feature pyramid network (FPN) method to extract the feature map with three scale resolutions, and their size is reduced by [33] times compared to the original image. Based on the above multi-scale feature map, the cost volume is constructed in stages from small to large. As shown in the green line in Figure 4, the depth information calculated by the cost volume with smaller resolution in the previous stage restricts the depth range of the homography transformation in the next stage to the depth value of this calculation. In fact, the depth interval of the previous stage is refined, and a smaller cost volume is constructed, which not only reduces the amount of calculation but also consumes explicit memory when calculating more accurate depth maps.

$$H_i^{(k+1)}\left(d^{(k)}+\Delta^{(k+1)}\right)=\left(d^{(k)}+\Delta^{(k+1)}\right)K_iT_iT_0^{-1} \tag{2}$$

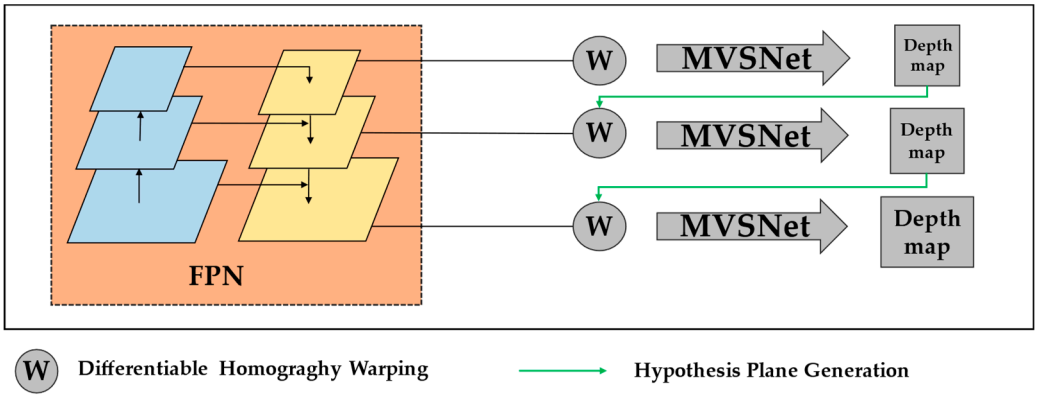


Figure 4. Coarse-to-fine computational deep MVS network.

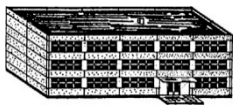
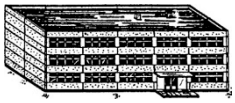
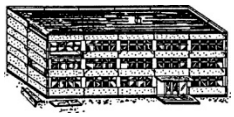
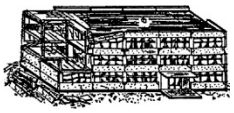
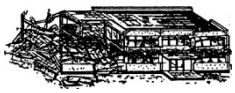
The differentiable homography [20] of CasMVSNet in different stages is as follows: The depth value of different stages is modified to $d = d^{(k)} + \Delta^{(k+1)}$, where k refers to the number of stages. The depth range for the next stage is calculated by adding the depth result $d^{(k)}$, calculated in the previous stage, to the residual depth $\Delta^{(k+1)}$. As shown in part W of Figure 4, with the change in the feature map, the intrinsic parameters of the camera are scaled equally. Using these parameters, the feature maps of the auxiliary view are warped into the reference image view space, and the cost volume is constructed by aggregating the cost matching the auxiliary view feature maps and the reference image feature maps.

3.3. Assessment of Damaged Buildings

AeDES [34] provides five grades for damaged buildings. Each grade lists detailed close-range pictures of houses. The examples include detailed parameters, such as the width of the wall cracks and the internal structure of the damaged building. Therefore, it is more suitable for field survey. The European Macro-Earthquake Magnitude in 1998 (EMS-98) [35] also classifies damaged buildings into five grade levels in detail, and each level provides a schematic diagram of macroscopic structural damage. Therefore, this

sub-standard is more suitable for visual inspection of the 3D model reconstructed using the above method to obtain a post-earthquake damage assessment of buildings. Table 1 shows the detailed descriptions and examples of building damage classifications. Then, according to the EMS-98 standard and the result of 3D reconstruction, a visual interpretation was conducted to determine the damage level of the building.

Table 1. Detailed description of buildings with different damage level.

Reinforced Concrete	Masonry Buildings	3D Model	Classification of Damage
			**Grade0: Negligible to slight** damage (no structural damage, slight non-structural damage)
			**Grade1: Moderate damage** (slight structural damage, moderate non-structural damage)
			**Grade2: Substantial to heavy damage** (moderate structural damage, heavy non-structural damage)
			**Grade3: Very heavy damage** (heavy structural damage, very heavy non-structural damage)
			**Grade4: Destruction** (very heavy structural damage)

4. Experiment and Results

4.1. Experimental Details

The generalization ability of the model is the most important engineering application ability. The model trained on the high-precision camera parameter *DTU* dataset [21] was directly used to test the UAV image data of Jinniu Village in Yangbi, Yunnan. The experiment was carried out on a computer with an Intel Xeon (R) W-2295 3.00 GHz * 36 processor and a 24 GB GeForce RTX3090/PCIe/SSE2 graphics processor.

The network was trained for 16 epochs, with an initial learning rate of 0.001, which was reduced by a factor of 2 after 10, 14, and 16 epochs. The pixel resolution of the input image in the network was fine-tuned according to the original public test parameters. The *stage* of CasMVSNet was set to 3, *numdepth* was set to 192, *ndepths* was set to corresponded to {48, 32, 8} and *interval_scale* was set to 1.06.

4.2. Results

The efficiency of 3D reconstruction and the quality of the model are critical for damage assessment of buildings after an earthquake and also take into account the hardware requirements for implementing the work. On the basis of the above experiments, we quantitatively analyzed and compared the results of three indicators of statistical time consumption, video memory consumption, and visual modelling of different methods. For

AA-RMVSNet and D^2HC-MVSNet, *numdepth* was set to 192 and *interval_scale* was set to 1.06.

4.2.1. Time Consumption

Table 2 is the time comparison of different methods. We divided the reconstruction time into two parts: the time for calculation of camera parameters and the time of 3D reconstruction. As can be seen from the table, camera parameters were calculated for deep learning methods based on COLMAP, so the time for the first part was always 61 min. In terms of 3D reconstruction, we set up two sets of resolution experiments to compare the time of 3D reconstruction. When the resolution of the input image was 1184 × 800 (pixel), the time consumption of AA-RMVSNet and D^2HC-MVSNet was approximately 10 times and 7.8 times that of CasMVSNet, respectively. When the resolution of the input image increased to 2160 × 1440 (pixel), the time consumption of CasMVSNet was also significantly lower than the other methods. As indicated in the table, the 3D reconstruction time of AA-RMVSNet and D^2HC-MVSNet was about 9.6 times and 8.6 times that of CasMVSNet, respectively. The CasMVSNet method took the shortest time and had the highest efficiency.

Table 2. Time consumption of various methods.

Meth	Calculating Camera Parameters	3D Reconstruction (min)	
		1184 × 800 Pixel	2160 × 1440 Pixel
AA-RMVSNet	61	163	499
D^2HC-MVSNet	61	124	447
CasMVSNet	61	16	52

The reason for the time difference is that both D^2HC-MVSNet and AA-RMVSNet use a recurrent neural network to regularize the cost map at each depth in the cost aggregation part. Compared with D^2HC-MVSNet, AA-RMVSNet adds an Inter-view AA module to the feature extraction part for multi-scale feature fusion and adds the Inter-view AA module before aggregating cost volume, which takes the longest time. The CascMVSNet model uses a cascading approach to calculate the pixel depth from coarse to fine. The depth range is roughly divided at the initial stage, and the calculation result is then discretized into a depth range for the next stage. Given that the sum of its depth intervals at each stage is much smaller than the depth interval values of the above two methods, and CascMVSNet uses a 3D CNN method that is faster than the recurrent neural network for cost aggregation, the method is the shortest and most efficient.

4.2.2. Memory Consumption

Table 3 shows the results of the memory consumption of different deep-learning-based MVS networks. As can be seen from the table, the memory consumption of CasMVSNet was the lowest. When the resolution of the input image was 1184 × 800, the memory consumption of CasMVSNet was 9232 MiB and 500 MiB lower than that of AA-RMVSNet and D^2HC-RMVSNet, respectively. When the resolution of the input image increased to 2160 × 1440, the memory consumption between different models was more distinct. Compared with the memory of 9955 MiB of CasMVSNet, the memory consumption of D^2HC-RMVSNet was 4400 MiB higher. The memory consumption of AA-RMVSNet was the largest, which was about 2.3 times that of CasMVSNet.

Table 3. Memory consumption of various deep learning methods in different sizes of images.

Meth	1184 × 800 (MiB)	2160 × 1440 (MiB)
AA-RMVSNet	13659	22933
D^2HC-MVSNet	4907	14395
CasMVSNet	4407	9995

The reason for this phenomenon is that CasMVSNet narrows the depth range as the resolution of the feature map increases, building a smaller cost volume at each stage to lower the memory consumption compared to the other two deep learning methods. The D^2HC-MVSNet network using a recurrent neural network instead of 3D CNN for cost aggregation decomposes the whole cost volume into a cost map at each depth. The memory consumption is slightly higher than that of CasMVSNet in low-resolution performance, but with the increase of input image resolution, its consumption will be significantly higher than that of the CasMVSNet model. The AA-RMVSNet model is improved on the basis of the D^2HC-MVSNet model. The Inter-view AA module added to the model performs pixel-level weighted aggregation on the cost volume constructed from multiple perspectives, so the model is higher than the other two models in memory consumption.

4.2.3. Result of 3D Reconstruction

As shown in Figure 5, the visualization results of different 3D reconstruction methods are displayed in CloudCompare [36]. As can be seen from the figures, all the methods exhibited relatively good results and were able to reflect the multidimensional details of the damaged building. Compared to the higher part of the building, the UAV obtained more detailed building information and could obtain more detailed stereo semantic information in the modeling process to complete stereo matching. Therefore, all of the methods performed well in this type of building. The Intra-view AA added in the feature extraction part of AA-RMVSNet maintained the correlation between the original geometric features of the image and the Inter-view AA added in the cost aggregation part on the basis of fusing multi-scale features. Compared to D^2HC-MVSNet, both improved the accuracy of the 3D model. Unlike AA-RMVSNet and D^2HC-MVSNet, CasMVSNet uses a multi-layer cascaded and gradually refined depth calculation method, which can better reflect the advantages of finer division and calculation of depth information when dealing with UAV image reconstruction. The building wall information in the 3D model reconstructed by this method was complete and delicate.

4.2.4. Result of the Evaluation

In this study, the point cloud generated by the 3D reconstruction was converted to 3DTiles format and imported into a seismic information visualization system, which was developed based on Cesium. In Figure 6, the model is marked with a Web page as a carrier to visually view the results of the assessment of the disaster situation. Address at www.peteralbus.com:8085 (accessed on 29 August 2022).

Based on the results of the above comparison, in this experiment, the reconstruction results of the CasMVSNet network were selected to evaluate the damage grade of 153 houses in the Jinniu Village area according to the EMS-98 standard. Furthermore, the proportion of the number of damaged houses at each grade to the total number of houses was calculated and compared with the evaluation results of other methods in this area. In total, two comparison methods were involved. The first was the result obtained by Zhang et al. In this research, visual interpretation was conducted on orthophotos of UAV images of the old street near the Yunlong Bridge in Yangbi County [37]. Another method was visual interpretation using spliced orthophotos. The comparison results are shown in Table 4. As can be seen in the table, our results are very close to those of the other two methods. Compared to the results of the field survey, the relative error was 1.3%, 1.0%, 0.6%, 1%, and 0.6% for G0, G1, G2, G3, and G4, respectively, which indicates the effectiveness of the proposed method.

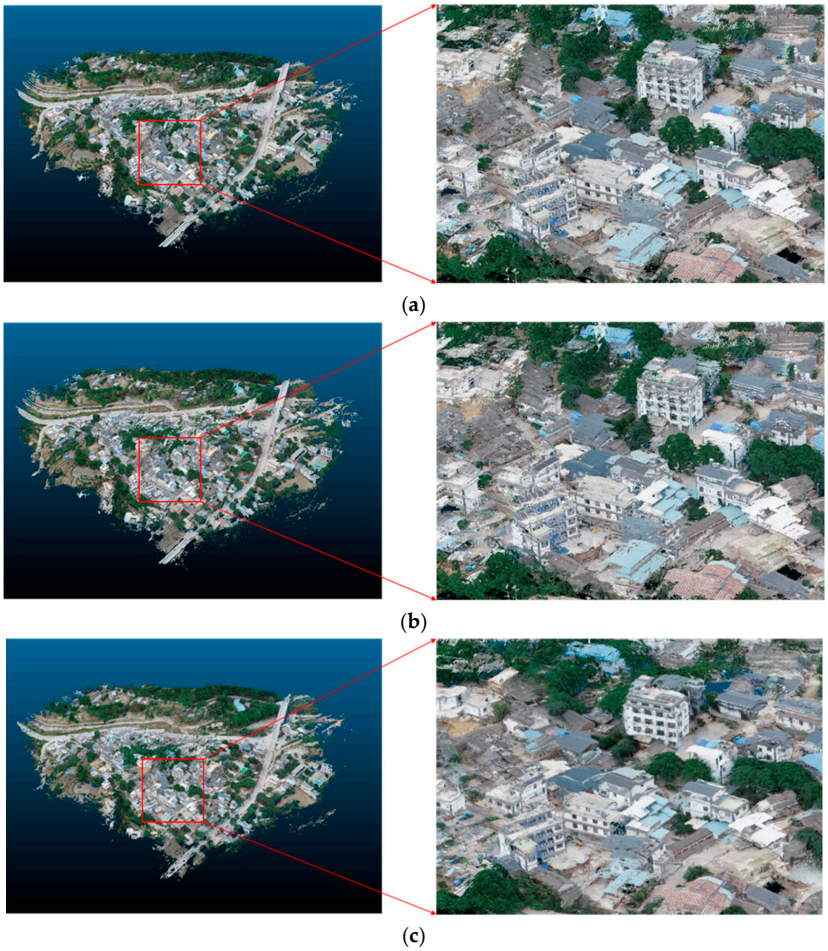


Figure 5. Visualization results of various 3D reconstruction methods. (**a**) D^2HC-MVSNet; (**b**) AA-RMVSNet; (**c**) CasMVSNet.

Figure 6. Results of the damaged assessment in Jinniu Village, Yangbi.

Table 4. Comparison of the results of the 2D and 3D assessment ratio and field survey.

Assessment of Damage	Orthophoto	Field Survey	3D Models
G0	22.7%	20.3%	21.6%
G1	37.7%	37.7%	36.7%
G2	21.6%	26.1%	25.5%
G3	13.2%	10.1%	11.1%
G4	4.8%	5.8%	5.2%
Total	100%	100%	100%

5. Discussion

Timeliness is the most important factor for the damage assessment of buildings after an earthquake. Therefore, the applicability of different MVS models for post-earthquake building damage assessment is discussed from the point of view of the network structures and the robustness. As shown in Figure 7, similar to CasMVSNet, deep-learning-based multi-view stereo networks, such as D^2HC-MVSNet and AA-RMVSNet, have mostly been improved on the basis of MVSNet. The initial MVSNet uses the 3D CNN method to regularize the cost volume and generate the probability. The soft argmin [38] operation calculates the depth value for each pixel in a winner-take-all manner and estimates the initial pixel-level depth map. Finally, the reference image is used to refine the depth map to improve the accuracy of the boundary region and complete refined pixel-level depth map estimation. However, as the input image size increases, the parameters of the model increase exponentially, so the method requires higher memory consumption.

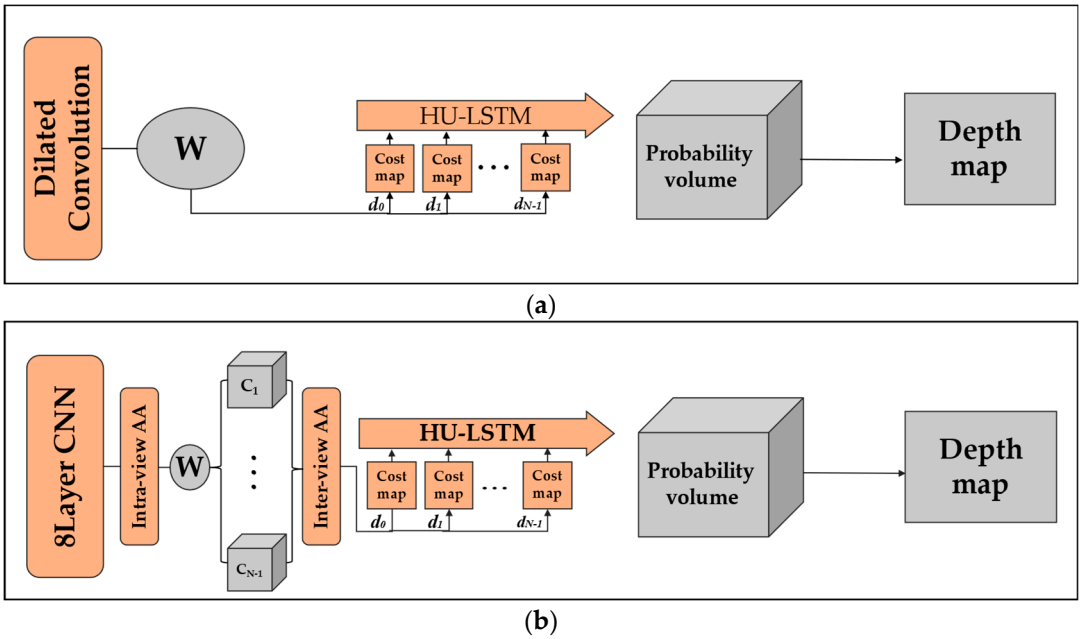


Figure 7. Schematic diagram of various deep learning MVS networks. (**a**) D^2HC-MVSNet; (**b**) AA-RMVSNet.

5.1. Network Structure

D^2HC-MVSNet improves the R-MVSNet GRU gating to some extent, with more powerful loop convolution units and dynamic consistency checking strategies. Firstly, the network uses dilated convolution [39] to obtain a larger range of feature information in the 2D feature extraction part, connects the feature output via different convolutional layers, and aggregates context feature information without losing resolution. D^2HC-MVSNet introduces a cyclic encoding–decoding HU-LSTM structure [24] to regularize the cost volume along the depth direction in the cost aggregation part, and realizes the connection memory of the same size features along the depth direction on the feature map of each scale

in the regularization process. This method not only aggregates the spatial geometric context information of the cost map but also preserves the cost aggregation output of the original resolution size with lower memory consumption. However, this method decomposes the cost volume into cost maps at each depth and processes them sequentially using a recurrent neural network, resulting in a slower computational efficiency of the network. When used to assist in assessing the damage level of houses in post-earthquake areas, it cannot meet the urgent needs of this work.

AA-RMVSNet is improved on the basis of D^2HC-MVSNet. The feature extraction part uses the common CNN to obtain the high-dimensional information of the image. The last two downsamplings output the feature maps of the original image with 1/4 and 1/16 size 32 channels, respectively. One of the innovations of this method is that an Intra-view AA module composed of deformable convolution [40,41] is added to the feature extraction part, which is used to adaptively aggregate the features of different scales and regions with different texture richness. This module processes the feature maps of the last three layers through deformable convolution and then upsamples the output of the last two layers after processing, and integrates with the previous layer as the final output of the feature extraction part, maintaining the geometric shape of the object in the image to the greatest extent when extracting features. In addition, when the network aggregates the cost volume of multiple perspectives, the Inter-view AA module is added to suppress the mismatched pixels by pixel-level weighting, and the pixels with higher matching correlations are given greater weight, rather than the matching results of all perspectives being treated equally. In short, the network fully retains the original geometric information of the object in the image and pays attention to the correlation problem after matching each perspective, thus improving the quality of 3D reconstruction. However, this method also uses the recurrent neural network, joining the above two optimization methods. Therefore, the quality of 3D reconstruction is guaranteed but the time efficiency is not well balanced when used in the post-earthquake building damage assessment work, and the added Inter-view AA module makes the network's memory consumption higher, resulting in higher GPU hardware requirements when carrying out this work.

5.2. Robustness

Through the MVS network modeling channel based on deep learning, the post-earthquake housing damage assessment work is completed—that is, from the public close-range experimental data migration to the real image of the UAV in the post-earthquake area and reconstruction of the 3D model—so whether the model has good robustness is extremely important. MVS network modeling based on deep learning completes 3D modeling in the form of calculated depth maps, so higher-resolution depth maps can reconstruct finer 3D models. The dilated convolution used by D^2HC-RMVSNet takes into account the features of a larger field of view, and AA-RMVSNet refines feature extraction and filter matching results, both of which improve when rebuilding 3D models. However, when used for high-resolution post-earthquake regional UAV image reconstruction, the series of models obtain higher-resolution depth maps by fine-tuning the max_w and max_h of the input image. However, from the experimental results, both methods have large fluctuations in time consumption and memory consumption. However, the CasMVSNet method constructs a smaller cost volume, so it has better stability at this time.

The MVS series networks based on deep learning calculate the pixel depth information based on the assumption of discrete depth intervals. Therefore, when using the UAV image data of large scenes to reconstruct the 3D model of the post-earthquake region, the single-stage networks of D^2HC-RMVSNet and AA-RMVSNet can fine-tune and refine the discrete depth intervals. However, the above two networks use the RNN method, so it will increase the workload of the recurrent neural network when fine-tuning the assumed discrete intervals. CasMVSNet can not only improve the robustness of 3D reconstruction by refining the discrete depth intervals of each stage of the network, but also refine the depth map output in the final stage by increasing the level of the network. In this way,

CasMVSNet realizes the 3D reconstruction of UAV image data in large scenes after the earthquake. Whether the number of stages of the network is adjusted or the discrete depth interval is refined, it will refine the calculation of depth information and provide better stability.

6. Conclusions

In this work, a multi-view stereo (MVS) method based deep learning was first applied to assist in assessing the damage level of buildings in post-earthquake areas. The method was tested in aerial UAV images from Yangbi County, Yunnan Province. The time consumption, memory consumption, and the performance of 3D reconstruction of different models were compared. In addition, the applicability of different 3D reconstruction models was discussed. A number of conclusions can be made as follows: (1) the time efficiency of CasMVSNet is significantly higher than that of other deep learning models, and the memory consumption is the lowest, which can meet the timeliness requirement of post-earthquake rescue and damage assessment; (2) CasMVSNet exhibited the best 3D reconstruction result in both high and small buildings; (3) the deep-learning-based 3D reconstruction method can provide the detail and multi-view information of damaged buildings, which can be used to assist the building damage assessment task. The assessment results were very close to the results of the field survey.

The main contributions of our study can be summarized as follows: (1) we first attempted to use deep-learning-based MVS to UAV aerial 3D reconstruction and building damage assessment research, which can provide 3D information for post-earthquake rescue and loss assessment; (2) the applicability of different MVS models for 3D reconstruction of UAV images have been analyzed and discussed.

The limitation of the proposed method is that the damage level of the buildings is determined by visual interpretation. In the future, we will devote efforts to construct a network to realize the automatic classification of buildings, based on the results of the 3D reconstruction.

Author Contributions: Conceptualization, Z.H., J.L. and S.J.; data curation, Y.Y., Z.H., J.L. and C.Z.; methodology, Z.H., Y.Y., S.J. and J.L.; validation, Y.Y., H.P., R.Z., Y.Z., Y.H., J.W. and S.Y.; formal analysis, Z.H. and Y.Y.; investigation, Y.Y. and C.Z.; writing—original draft preparation, Z.H., Y.Y. and H.P.; writing—review and editing, Z.H., Y.Y., H.P., J.L. and S.J.; supervision, J.L. and S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key R&D Program of China, grant number 2017YFC1500906; the National Natural Science Foundation of China, grant number 41871325, 42061073; and the Natural Science and Technology Foundation of Guizhou Province under Grant [2020]1Z056.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The *DTU* datasets are freely available online, and can be found at https://github.com/YoYo000/MVSNet (accessed on 29 August 2022).

Acknowledgments: We thank our team's graduate Master Jinmeng Gao for providing UAV images so that our work could be successfully completed.

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

Abbreviations

UVA	unmanned air vehicle
GA	genetic algorithm
MLP	multi-layer perceptron
RC	reinforced-concrete
SMART SKY EYE	smart building safety assessment system using UAV
CNN	convolutional neural networks
DEM	digital elevation models
MVS	multi-view stereo
SfM	structure-from-motion
BA	bundle adjustment
FPN	feature pyramid networks
EMS-98	European Macro-Earthquake Magnitude in 1998

References

1. Yamazaki, F.; Kouchi, K.I.; Matsuoka, M.; Kohiyama, M.; Muraoka, N. Damage detection from high-resolution satellite images for the 2003 Boumerdes, Algeria earthquake. In Proceedings of the 13th World Conference on Earthquake Engineering, International Association for Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004; p. 13.
2. Li, S.; Tang, H. Classification of Building Damage Triggered by Earthquakes Using Decision Tree. *Math. Probl. Eng.* **2020**, *2020*, 1–15. [CrossRef]
3. Bülbül, M.A.; Harirchian, E.; Işık, M.F.; Aghakouchaki Hosseini, S.E.; Işık, E. A Hybrid ANN-GA Model for an Automated Rapid Vulnerability Assessment of Existing RC Buildings. *Appl. Sci.* **2022**, *12*, 5138. [CrossRef]
4. Harirchian, E.; Jadhav, K.; Kumari, V.; Lahmer, T. ML-EHSAPP: A prototype for machine learning-based earthquake hazard safety assessment of structures by using a smartphone app. *Eur. J. Environ. Civ. Eng.* **2021**, *26*, 5279–5299. [CrossRef]
5. Bae, J.; Lee, J.; Jang, A.; Ju, Y.K.; Park, M.J. SMART SKY Eye system for preliminary structural safety assessment of buildings using unmanned aerial vehicles. *Sensors* **2022**, *22*, 2762. [CrossRef] [PubMed]
6. 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 Sens. Environ.* **2021**, *265*, 112636. [CrossRef]
7. Ci, T.; Liu, Z.; Wang, Y. Assessment of the degree of building damage caused by disaster using convolutional neural networks in combination with ordinal regression. *Remote Sens.* **2019**, *11*, 2858. [CrossRef]
8. 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]
9. Turker, M.; Cetinkaya, B. Automatic detection of earthquake-damaged buildings using DEMs created from pre-and post-earthquake stereo aerial photographs. *Int. J. Remote Sens.* **2005**, *26*, 823–832. [CrossRef]
10. Muzirafuti, A.; Cascio, M.; Lanza, S.; Randazzo, G. UAV Photogrammetry-based Mapping of the Pocket Beaches of Isola Bella Bay, Taormina (Eastern Sicily). In Proceedings of the 2021 International Workshop on Metrology for the Sea, Learning to Measure Sea Health Parameters (MetroSea), Reggio Calabria, Italy, 4–6 October 2021; pp. 418–422.
11. Randazzo, G.; Italiano, F.; Micallef, A.; Tomasello, A.; Cassetti, F.P.; Zammit, A.; D'Amico, S.; Saliba, O.; Cascio, M.; Cavallaro, F. WebGIS Implementation for Dynamic Mapping and Visualization of Coastal Geospatial Data: A Case Study of BESS Project. *Appl. Sci.* **2021**, *11*, 8233. [CrossRef]
12. Stepinac, M.; Lulić, L.; Ožić, K. The Role of UAV and Laser Scanners in the Post-earthquake Assessment of Heritage Buildings After the 2020 Earthquakes in Croatia. In *Advanced Nondestructive and Structural Techniques for Diagnosis, Redesign and Health Monitoring for the Preservation of Cultural Heritage*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 167–177.
13. Bleyer, M.; Rhemann, C.; Rother, C. Patchmatch stereo-stereo matching with slanted support windows. In Proceedings of the Bmvc, Dundee, UK, 29 August–2 September 2011; pp. 1–11.
14. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [CrossRef] [PubMed]
15. Tola, E.; Strecha, C.; Fua, P. Efficient large-scale multi-view stereo for ultra high-resolution image sets. *Mach. Vis. Appl.* **2012**, *23*, 903–920. [CrossRef]
16. Pix4D. Available online: https://www.pix4d.com/ (accessed on 29 August 2022).
17. ContextCapture. Available online: https://www.bentley.com/en/products/brands/contextcapture (accessed on 29 August 2022).
18. Agisoft. Available online: http://www.agisoft.com (accessed on 29 August 2022).
19. Zhu, Q.; Min, C.; Wei, Z.; Chen, Y.; Wang, G. Deep Learning for Multi-View Stereo via Plane Sweep: A Survey. *arXiv* **2021**, arXiv:2106.15328.
20. Yao, Y.; Luo, Z.; Li, S.; Fang, T.; Quan, L. Mvsnet: Depth inference for unstructured multi-view stereo. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 767–783.

21. Aanæs, H.; Jensen, R.R.; Vogiatzis, G.; Tola, E.; Dahl, A.B. Large-scale data for multiple-view stereopsis. *Int. J. Comput. Vis.* **2016**, *120*, 153–168. [CrossRef]
22. Yao, Y.; Luo, Z.; Li, S.; Shen, T.; Fang, T.; Quan, L. Recurrent mvsnet for high-resolution multi-view stereo depth inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5525–5534.
23. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
24. Yan, J.; Wei, Z.; Yi, H.; Ding, M.; Zhang, R.; Chen, Y.; Wang, G.; Tai, Y.-W. Dense hybrid recurrent multi-view stereo net with dynamic consistency checking. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 674–689.
25. Wei, Z.; Zhu, Q.; Min, C.; Chen, Y.; Wang, G. Aa-rmvsnet: Adaptive aggregation recurrent multi-view stereo network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 6187–6196.
26. Gu, X.; Fan, Z.; Zhu, S.; Dai, Z.; Tan, F.; Tan, P. Cascade cost volume for high-resolution multi-view stereo and stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 20–25 June 2020; pp. 2495–2504.
27. Tonioni, A.; Tosi, F.; Poggi, M.; Mattoccia, S.; Stefano, L.D. Real-time self-adaptive deep stereo. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 195–204.
28. Yin, Z.; Darrell, T.; Yu, F. Hierarchical discrete distribution decomposition for match density estimation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6044–6053.
29. Guptha, G.C.; Swain, S.; Al-Ansari, N.; Taloor, A.K.; Dayal, D. Evaluation of an urban drainage system and its resilience using remote sensing and GIS. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100601. [CrossRef]
30. Kazemian, I.; Torabi, S.A.; Zobel, C.W.; Li, Y.; Baghersad, M. A multi-attribute supply chain network resilience assessment framework based on SNA-inspired indicators. *Oper. Res.* **2022**, *22*, 1853–1883.
31. Schonberger, J.L.; Frahm, J.-M. Structure-from-motion revisited. In *Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; pp. 4104–4113.
32. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 221–231. [CrossRef] [PubMed]
33. Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. *Remote Sens.* **2018**, *10*, 1636. [CrossRef]
34. Baggio, C.; Bernardini, A.; Colozza, R.; Corazza, L.; Della Bella, M.; Di Pasquale, G.; Dolce, M.; Goretti, A.; Martinelli, A.; Orsini, G. *Field Manual for Post-Earthquake Damage and Safety Assessment and Short Term Countermeasures (AeDES)*; European Commission—Joint Research Centre—Institute for the Protection and Security of the Citizen: Ispra, Italy, 2007; pp. 1–100.
35. Grünthal, G. *European Macroseismic Scale 1998 (EMS-98)*; Conseil De L'europe: Strasbourg, France, 1998.
36. CloudCompare. Available online: http://www.danielgm.net/cc (accessed on 29 August 2022).
37. Zhang, L.; He, F.; Yan, J.; Du, H.; Zhou, Z.; Wang, Y. Quantitative Assessment of Building Damage of the Yangbi Earthquake Based on UAV Images. *South China J. Seismol.* **2021**, *41*, 76–81.
38. Kendall, A.; Martirosyan, H.; Dasgupta, S.; Henry, P.; Kennedy, R.; Bachrach, A.; Bry, A. End-to-end learning of geometry and context for deep stereo regression. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 66–75.
39. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
40. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
41. Zhu, X.; Hu, H.; Lin, S.; Dai, J. Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9308–9316.