Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 11, No. 1, 2024

Building Damage Degree Recognition Based on Temporal Attention Features

Zhenzhao Jiang*

School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, Henan, China
* Corresponding author: Zhenzhao Jiang (Email: jiangzhenzhao9527@gmail.com)

Abstract: Natural disasters pose significant harm to society. As an important place for social activities and economic development, the degree of damage to building areas is directly related to disaster loss assessment and emergency rescue. Remote sensing image data, characterized by its wide coverage and multi-temporal features, provides important data support for post-disaster loss assessment. However, imaging differences caused by factors such as shooting time, imaging angle, and different sensors can interfere with the extraction of damage features and loss assessment. This paper proposes a Dual-Exchange-Attention U-Net (DERU-Net) model, which transforms the identification of building damage levels into intra-class semantic change detection. The DFMA feature attention fusion module is introduced to enhance the ability of dual-temporal feature extraction and achieve end-to-end assessment of building damage. The proposed method is comprehensively evaluated and tested on the xBD dataset. Experimental results show that compared with other methods, the DERU-Net proposed in this paper exhibits better stability and evaluation accuracy in assessing the degree of building damage.

Keywords: Structure; damage assessment; semantic change-detection.

1. Introduction

Natural disasters pose tremendous harm to society. As an important venue for social activities and economic development, building areas require accurate and immediate responses from humanitarian assistance and disaster response (HADR) efforts when natural disasters occur. The degree of damage to buildings is directly related to the assessment of disaster losses[1-3]。The assessment of building damage serves as a crucial data support for post-disaster emergency rescue and reconstruction. The objective is to segment specific instances of buildings and then label the degree of damage for each instance. Understanding and grasping the quantity and extent of damaged buildings is vital, as it directly determines the needs and priorities of rescue efforts[4]。Traditional emergency response planning measures rely heavily on ground-based assessment reports and statistics, requiring rescue personnel to conduct field investigations and evaluations in disaster-stricken areas in order to obtain information about disaster losses and rescue needs[5]，this is often highly dangerous and time-consuming, and lacks objectivity[6].

With the continuous development of remote sensing technology, high-resolution satellite image data, characterized by its wide coverage and long-term temporal sequence, has become an important data support for assessing the degree of building damage. As shown in Figure 1, although remote sensing image data provides rich information and data, there are limitations in the slow extraction of information by manual analysts, making it difficult to distinguish between certain similar categories. Additionally, existing research is often limited to a single disaster type, and there are no unified standards for disaster loss assessment[7-10], the lack of accurate guidelines and comprehensive databases for assessing the degree of building damage across different disaster types presents a significant challenge. Therefore, utilizing artificial intelligence technology to accelerate the analysis and interpretation of satellite images, reduce the subjectivity of visual interpretation, and achieve quantitative calculations has become an important direction for improving the accuracy and automation of obtaining building damage information. This approach can help determine damaged areas more quickly, which is crucial for enhancing the efficiency of disaster response[6].

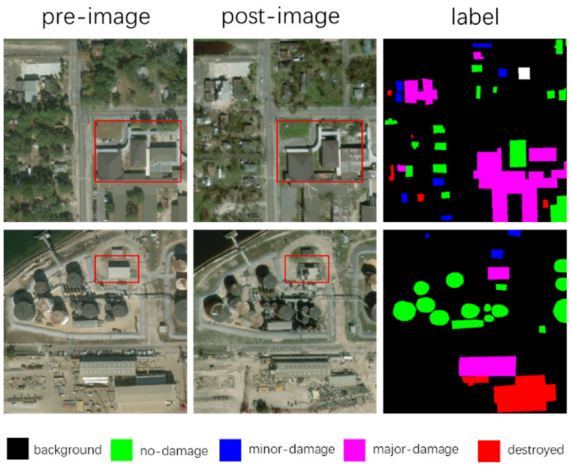

Figure 1. A pair of images and their annotations from the xBD dataset.

However, in practical applications, there exist issues of imaging differences due to inconsistent dual-temporal imaging conditions. As shown in Figure 1, two pairs of image data from the xBD dataset exhibit problems such as angle shifts and color differences in buildings within the imaging space due to inconsistent imaging conditions between pre- and post-disaster dual-temporal images. Additionally, due to the limitations of image resolution, it is difficult to accurately assess the degree of damage that exhibits visual similarity solely through visual interpretation.

In summary, for the interference in the extraction of

damage features caused by imaging differences due to differences in shooting time, shooting angle, sensors, etc. between pre-disaster and post-disaster images in the method of identifying the degree of building damage based on dual-temporal phase, this paper constructs a dual cross-attention model (Dual-Exchange-Attention U-Net, DERU-Net), to achieve end-to-end evaluation of building damage, the specific contributions are summarized as follows:

To mitigate the issue of different imaging styles in dual-temporal data caused by varying imaging conditions, an attempt was made to employ Fast Fourier Transform (FFT) for style unification, aiming to reduce such impacts.

1) To mitigate the issue of different imaging styles in dual-temporal data caused by varying imaging conditions, an attempt was made to employ Fast Fourier Transform (FFT) for style unification, aiming to reduce such impacts.

2) A dual-temporal feature fusion module (DFMA) is proposed to effectively integrate pre-disaster and post-disaster feature maps. By fusing these dual-temporal feature maps, the feature representation is enhanced.

3) An end-to-end convolutional neural network is proposed for evaluating the extent of building damage and locating the buildings.

2. Related Work

2.1. A Comprehensive Review of Building Damage Assessment

Current methods for identifying the degree of building damage can be divided into two categories based on whether pre-disaster images are referenced: methods for identifying building damage degree based on single-temporal images and methods for identifying building damage degree based on dual-temporal images:

(1) The method for identifying the degree of building damage based on single-temporal images adopts the idea of semantic segmentation, using high-resolution remote sensing (Very High Resolution, VHR) image data after disasters with detailed texture information and contextual information as the input of the network. By detecting various features in the images such as the changed spectrum, texture, edges, spatial relationships, structure, shape, and shadows of buildings due to damage, the identification of the degree of building damage is achieved[11, 12]. Sumer ed.al[13] the buildings are extracted from the post-disaster images using building vector maps. Then, the damage status of the buildings is determined based on the pixel value differences and gradient direction differences between intact buildings and damaged buildings.Ye et al[14] first determined the locations of building objects in the post-disaster images using building vector maps. Then, they considered both the boundary integrity index (ES) and the roof integrity index (LOE) of the building objects. Finally, a building object was only determined as damaged when both indices indicated damage.Rudner et al[15] proposed a method for extracting damaged buildings based on a semantic segmentation model, which can directly obtain pixel-level interpretation results of damaged buildings from the input post-disaster images. However, in the case of large-scale disasters, where buildings are severely damaged to the point that boundary information is no longer present, these buildings appear as weakly defined targets with blurred boundaries that are easily drowned out in complex background environments. Therefore, the rates of false positives and missed detections are relatively high.

(2) The method of identifying the degree of building damage based on double temporal phases adopts the research idea of change detection, usually using the method of image enhancement and post-classification comparison to obtain the state differences of buildings in different temporal phases, thus analyzing and identifying the degree of building damage[16, 17]. The most commonly used approach for conducting building damage assessment tasks is based on temporal data. Compared to single-temporal phase methods for building damage identification, the dual-temporal phase method not only requires identifying changed areas (i.e., damaged buildings) but also non-changed areas (i.e., undamaged buildings), resulting in a higher model complexity。FC-Siam-diff[18] to extract temporal features from two temporal phase image data, a symmetric network is used, and the differences are subtracted to obtain a change map. While the difference map is the most intuitive feature for interpreting changes, it is prone to false detections due to differences in imaging conditions between the two temporal phases, such as viewpoint changes, occlusions, shadows, and other factors[19]; Xiao et al[20]proposed the DCFNet framework, which consists of two modules: DCF (Dynamic Cross-task Fusion) and TSH (Task-shared Head). It supports adaptive feature selection between multi-level features to optimize detection performance.。Fu et al[21] proposed an end-to-end framework for building damage detection (BDD) based on the Super-Resolution Generative Adversarial Network (SRGAN) and U-Net convolutional network, which restores high-resolution building damage results from low-resolution images. While the accuracy of building damage assessment methods based on dual-temporal satellite image data is relatively higher, the availability of pre-disaster image data is difficult to guarantee due to the suddenness of natural disasters and the limitations of the revisit cycle of remote sensing satellite data. At the same time, there remains a significant challenge in how to reasonably establish a mapping relationship model between dual-temporal data with different imaging conditions, in order to generate a building damage localization attribute map.

2.2. Attention Model

The attention mechanism of deep learning is inspired by the attention thinking mode of human vision, forming an attention focus[22].By incorporating an attention module, the network can focus more on the feature information of the image. In SE-Net[23]. SPA-Net[24] the separate utilization of channel attention and spatial attention mechanisms has achieved great success in remote sensing image processing tasks. The subsequently proposed CBAM[25] combines channel and spatial attention modules in series, allowing for the consideration of information interaction in both channel and spatial dimensions, thereby generating more expressive feature maps. Currently, attention-based models are less frequently applied in assessing the degree of building damage [26]. In the literature[27], a non-local attention model[28] was used to capture remote spatial information from pre-disaster and post-disaster images. This is because the non-local mechanism requires calculating attention maps based on high-resolution features, which means that a large number of pixels and higher-dimensional data need to be processed, leading to higher computational costs. Although the non-local mechanism can provide more comprehensive global information, its high computational cost needs to be weighed against practical considerations.

3. Proposed Method

Inspired by the literature[29], this paper constructs DERU-Net. DERU-Net adopts a dual encoder-decoder backbone network, including two different encoders, a building localization decoder, and a building damage degree decoder. The dual weight-shared encoders are responsible for extracting the feature information of dual-temporal images and enhancing the ability of the network model to extract the boundary of damaged buildings with blurred boundaries after disasters through parameter sharing. Then, the channel switching module is used to perform feature exchange to ensure that each encoder contains the characteristics of dual-temporal image data. In the T1 branch, the building localization decoder integrates the T1 decoder features and the building encoder features through skip connections to accurately locate the building area and generate a building localization mask file. Similarly, for post-disaster image data, the same decoder is used for auxiliary positioning. In the building damage degree decoder, the temporal fusion attention module (DFMA) is used to accurately fuse the dual-temporal features to generate the final building damage degree localization map. In the final output stage of the model, the building localization mask and the building damage degree localization map are fused through element-wise multiplication to generate the final instance building damage degree localization map.

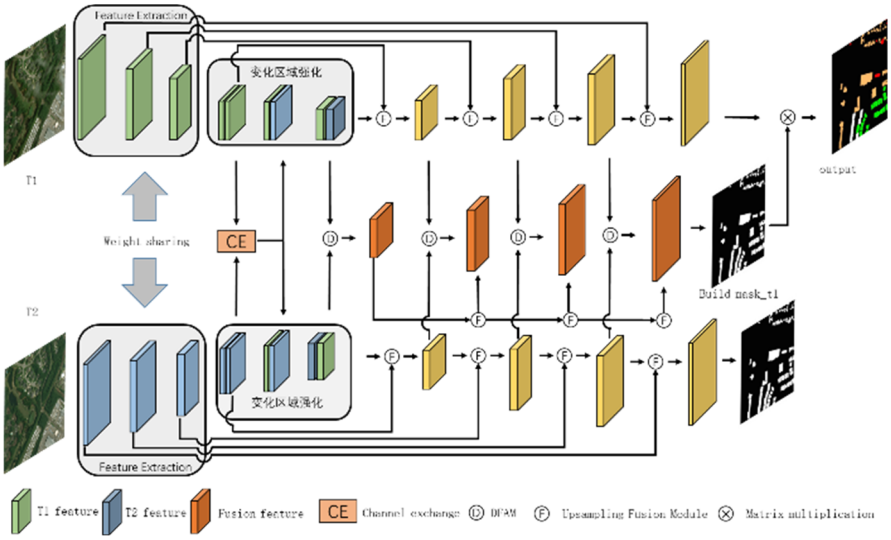

Figure 2. Overview of the proposed framework (DERU-Net)

3.1. Encoder Part

In this study, U-Net[30] is selected as the backbone network. Compared to traditional CNN-based network architectures, U-Net introduces downsampling and upsampling to aggregate features at multiple scales. Through skip connections, it integrates low-level and high-level features to enhance the network's learning capability [31]. However, the U-Net network structure is relatively simple, and feature extraction primarily relies on the double convolutional layers in the encoder. While the double convolutional layers can extract rich image features and increase the model complexity, they also increase computational costs and are more prone to overfitting. Additionally, the problem of gradient disappearance in deep networks may also affect the effective feature learning of deeper convolutional layers. To address these issues in the encoder part, the following improvements are made in this paper:

(1) In the downsampling part of the model, a residual structure is chosen to mitigate the issue of information loss during forward propagation. A convolutional layer with a stride of 2 is used to extract information from the feature map passed from the upper layers of the network. Simultaneously, the Haar wavelet transform downsampling module[32] is employed to reduce the spatial resolution of the feature map while preserving as much information as possible for the residual connection part.

(2) To reduce the model parameters and computational complexity, and achieve fast and effective feature extraction, this paper employs the HCU (Half Convolutional Unit) proposed in the literature[29] to replace the original double convolutional layers. The feature map passed from the downsampling part is divided in the channel dimension, with one being passed to the convolutional layer for information extraction and the other serving as residual features. Meanwhile, to maximize the retention of information, the CBAM module is selected as the residual feature of the entire feature map, which is used to enhance the weights of the input feature map. Ultimately, the feature extraction of a single encoder part is completed.

(3) Channel Exchange Module. A channel exchange module [29] is introduced in the Encoder3 part, which passes the dual-temporal encoder features into the CE module for feature half-exchange operations, promoting the fusion of dual-temporal features and refining the change regions. This can be expressed as the following formula(1):

$$T_1', T_2' = M * T_1 + (1 - M) * T_2 \quad (1)$$

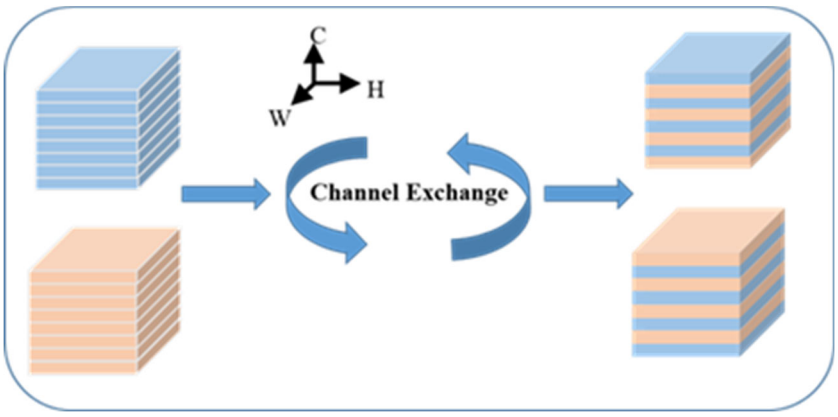

Figure 3. Channel Exchange Module

3.2. Temporal Fusion Attention Module (DFMA)

Methods for fusing dual-temporal features can be divided into three categories: simple fusion, convolution enhancement, and attention enhancement[33-37]. Simple fusion methods[33] directly perform element-wise algebraic operations on dual-temporal features for fusion, which are prone to noise interference and difficult to achieve effective feature fusion. Convolution enhancement methods[35] use convolution operations to reduce noise and enhance features before fusion, but they mainly focus on feature enhancement before fusion and ignore the temporal information between dual-temporal features. Attention enhancement methods utilize the attention mechanism to achieve effective feature fusion, but this approach mainly focuses on feature enhancement after simple fusion and also neglects the temporal information between dual-temporal features.

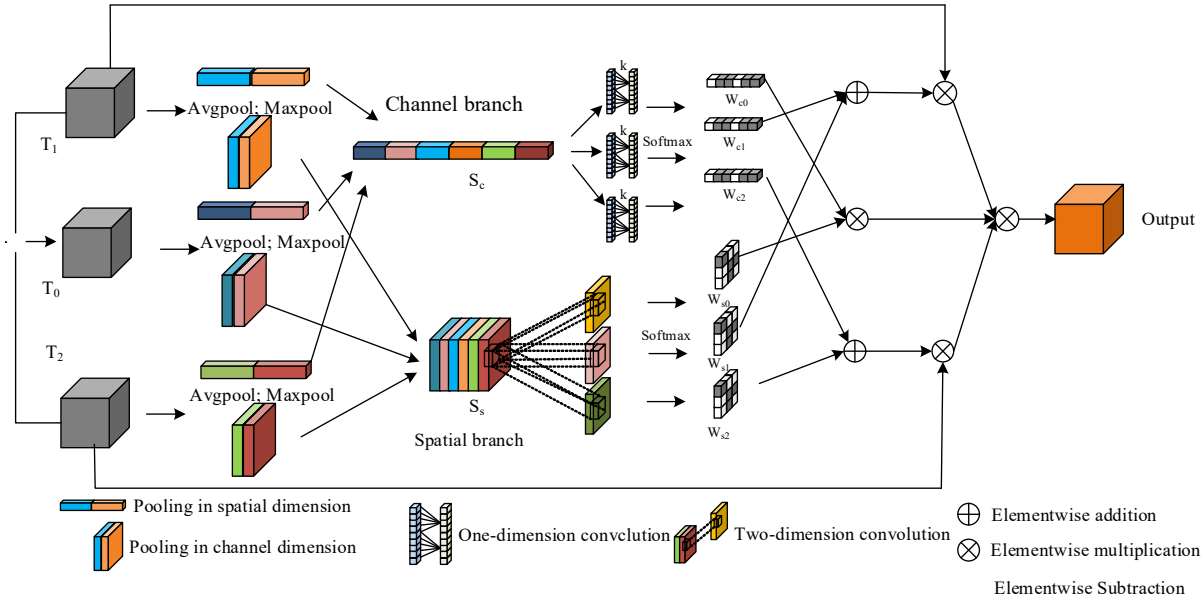

Figure 4. Illustration of the structure of DFAM.

To overcome the challenges in dual-temporal feature fusion, this study proposes an attention-based fusion module called the DFMA mechanism, as shown in Figure 4. Let T1, T2 $\in R^{B\times C\times H\times W}$ be the input dual-temporal feature maps, where B represents the batch size, C represents the number of channels, and H and W are the height and width of the feature maps, respectively. The cosine similarity is calculated for T1 and T2, as shown in Equation 2. After obtaining the similarity matrix coefficient I, the change feature map Tc is computed as shown in Equation 3.

$$I = \frac{\sum_{i=0,j=0}^{n}(T_{1i,j}\times T_{2i,j})}{\sqrt{\sum_{i=0,j=0}^{n}(T_{1\,(i,j)})^2}\times\sqrt{\sum_{i=0,j=0}^{n}(T_{2\,(i,j)})^2}} \quad (2)$$

$$T_c = T_1 \times SoftMax((1-I)\times T_2) \quad (3)$$

In the DFMA module, T1, T2, and Tc are used to aggregate channel and spatial information to determine the important parts in the dual-temporal feature map. In the channel dimension, the input feature information is aggregated by global pooling across the spatial dimension. The aggregation process can be formulated as Equation 4, where Sc represents the aggregated spatial feature, Avg(.) and Max(.) respectively represent global average pooling and global maximum pooling across the spatial dimension. The channel weight coefficients are confirmed through one-dimensional convolution, as shown in Equation 5, where Conv1 represents one-dimensional convolution. The Softmax function is used

to determine the weights $W_{T_1}^C, W_{T_c}^C, W_{T_2}^C$ of different features in the channel dimension, thereby more effectively fusing these features, as shown in Equation 6. Spatial weights $W_{T_1}^S, W_{T_c}^S, W_{T_2}^S$ are determined in the same way in the spatial dimension to determine the important parts among spatial dimensional features. The dual-temporal channel weights and dual-temporal spatial weights are combined to determine the important parts between features, as shown in Equation 7.

$$S_c = Concat(Avg(T_1), Max(T_1), Avg(T_c), Max(T_c), Avg(T_2), Max(T_2)) \quad (4)$$

$$W_{T_1}, W_{T_c}, W_{T_2} = Conv_1(S_c), Conv_c(S_c), Conv_2(S_c) \quad (5)$$

$$W_{T_i}^C = \frac{e^{W_{T_i}}}{e^{W_{T_1}} + e^{W_{T_C}} + e^{W_{T_2}}} \quad (6)$$

$$output = \left(W_{T_1}^C + W_{T_1}^S\right) * T_1 + \left(W_{T_C}^C + W_{T_C}^S\right) * T_C + \left(W_{T_2}^C + W_{T_2}^S\right) * T_2 \quad (7)$$

3.3. Decoder Part

DERU-Net incorporates two independent decoders. As shown in the following figure, decoder a is primarily responsible for building localization. It integrates the features from the T1 decoder and the building encoder through skip connections to precisely locate the building area and generate a building localization mask file. Let the output of the building localization part be P_b ∈ R^(2×H×W), where H and W represent the height and width of the output, respectively. The building localization result can be expressed as Equation 8, where P_B ∈ {"0, 1"} ^(1×H×W) with 1 representing the building and 0 representing the background. Decoder b utilizes the Dual-temporal Fusion with Multi-Attention (DFMA) module to precisely fuse the dual-temporal features and generate the final building damage level localization map. Let the output of the damage assessment be $P_d \in R^{(C\times H\times W)}$, where C represents the number of damage levels. To guide the building location, the final output P is expressed as Equation 9, where · represents element-wise multiplication.

$$P_B = \arg\max(P_b) \quad (8)$$

$$output = \arg\max(P_B \cdot P_d) \quad (9)$$

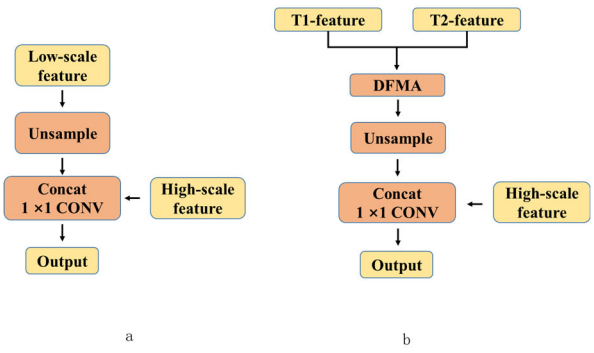

Figure 5. Diagram of the decoder

4. EXPERIMENTS

4.1. Dataset and Evaluation

To evaluate the effectiveness of the proposed method in this study, experiments were conducted using the xBD dataset. The xBD dataset is the largest and highest-quality natural disaster satellite image dataset available to date for assessing the degree of building damage. It contains a total of 19 natural disaster events from different regions, with annotations for over 800,000 buildings. The dataset consists of image pairs (pre-disaster and post-disaster) with a size of 1024 × 1024 pixels. In the image data, the degree of building damage is classified into four levels: no damage, minor damage, major damage, and destroyed.

4.2. Implementation Details

(1)Data enhancement strategy. In the training stage, data enhancement techniques are applied to enhance the generalization ability of the model, including random flipping (probability = 0.5), transposition (probability = 0.5), random shifting (probability = 0.3), random scaling (probability = 0.3), and random rotation (probability = 0.3). In this paper, we use Albumentation [38] to implement all data enhancement methods with default settings. In addition, as mentioned in the table above, CutMix [39] improves the robustness and generalization ability of the model by fusing parts of images from different categories, enhances the diversity of training data, and effectively alleviates the problem of overfitting with a small sample size.

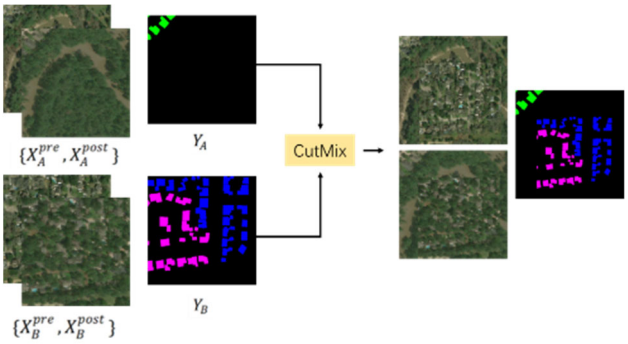

Figure 6. Diagram of CutMix

(2) Style-transfer: To address the issue of the difference in dual-temporal imaging before and after a disaster, this paper uses Fast Fourier Transform (FFT) to unify the styles of dual-temporal images (probability = 0.5). The spatial domain of an image is converted into a frequency domain, where the parts with significant changes in grayscale values correspond to high frequencies, and the rest correspond to low frequencies. High-frequency components mainly measure the edges and contours of an image, while low-frequency components mainly measure the overall intensity of the entire image. Therefore, by replacing the low-frequency part of the frequency domain of image B with the low-frequency part of the frequency domain of image A, and then converting it back to the spatial domain through Inverse Fast Fourier Transform (IFFT), the styles of images A and B can be unified.

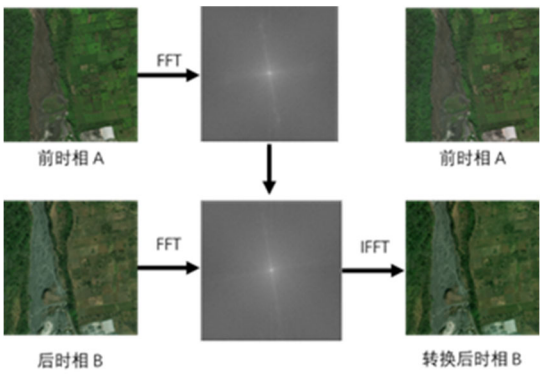

Figure 7. Fast Fourier Transform (FFT)

(3) Training and Inference: This paper uses PyTorch [40] to build DERU-Net and conducts experiments on a Lenovo P620 workstation equipped with an RTX A5000 GPU (24 GB memory). The image data is cropped to a size of 512 × 512 pixels for training, and the batch size of the network is set to 8. AdamW[41] is used as the optimizer, with an initial learning rate of 0.001 and a weight decay of 0.001. The learning rate is adjusted using a cosine annealing strategy.

4.3. Experimental results

As shown below, the comprehensive scores achieved by the proposed DEAU-Net instantiation model in the building damage assessment experiment are as follows: the $F_1^{\text{overall}}$is 70.6%, with F_1^{loc} and F_1^{dam} scores of 84.6% and 64.5% respectively. The F_1 scores for no damage (F_1^{no}), minor damage (F_1^{minor}), major damage (F_1^{major}), and complete destruction $F_1^{\text{destoryed}}$) are 82.5%, 46.6%, 67.6%, and 73.4% respectively. Compared to the MEDU-Net, which uses Unet as the feature extractor, the proposed DEAU-Net shows significant improvement in accuracy, with an overall F_1^{overall} lincrease of 7.2%. Notably, there are significant improvements in the accuracy of minor damage (F_1^minor), major damage (F_1^major), and complete destruction (F_1^destroyed). This result fully demonstrates that the proposed network model in this paper has high accuracy in determining the damage level of buildings with different degrees of damage.

Table 1. Model Accuracy Comparison Table

model	F_1^{overall} %	F_1^{loc} %	F_1^{dam} %
DESDU-Net	50.4	69.6	42.1
MESDU-Net	63.6	82.4	56.9
MEDU-Net	67.5	83.8	60.5
DEAU-Net	70.6	84.6	64.5

As stated in the table above, this paper verifies the building damage networks of different frameworks under the same conditions. Based on the experimental results, the following analysis can be drawn:

(1) Comparing the accuracy of DESDU-Net and MESDU-Net, it can be seen that the adoption of the Siamese network structure in the encoder part of MESDU-Net allows the sharing of weight parameters in the feature extraction branches of pre-disaster and post-disaster images. With the help of the intact boundaries of pre-disaster images and the features of undamaged buildings as prior information, a significant improvement in the F_1^{loc} score can be observed when comparing the experimental results of the two networks. The experiment proves that introducing the feature information of intact pre-disaster buildings can effectively improve the positioning information of buildings.

(2) Comparing the accuracy of MESDU-Net and MEDU-Net, it is evident that using attention mechanisms to enhance the fusion of dual-temporal feature data results in higher extraction accuracy than feature map concatenation or feature algebraic operations. Especially in the improvement of F_1^{minor} and F_1^{major}. Compared to simple subtraction operations, using attention mechanisms for different temporal feature map fusion can more effectively capture temporal change information, suppress the interference of irrelevant features, and obtain more accurate and robust change detection results through the advantages of adaptive weight allocation, long-range dependency modeling, feature selectivity, and end-to-end trainability.

(3) Comparing the accuracy of MEDU-Net and DEAU-Net, it can be seen that by using the channel shuffle module in the encoder part, each branch contains dual-temporal information after channel shuffle. This means that the dual-temporal features are interconnected, and each branch can determine the change area independently. Based on the change area, features that only contain single-temporal information before channel shuffle contain rich spatial features, which can be used to refine the change area and accurately locate the changed objects in the same temporal phase. Finally, by fusing the dual-temporal features, all changed objects can be precisely located.

A comprehensive analysis of all models shows that although improving the network structure and adding different modules do enhance the detection results, compared to comparing with the actual ground distribution Label files obtained through field investigations, in cases of large-area damage, the existing models tend to have problems in judging the damage of small individual buildings. The judgment of the damage degree of buildings with smaller areas is easily influenced, leading to misjudgments.

5. Conclusion

This paper proposes an instantiated building damage assessment model for identifying and evaluating the degree of building damage using the DEAU-Net, which is suitable for multiple disaster types. Based on U-Net, the model integrates a feature extraction structure with residual structure to optimize the encoding part of U-Net, enhancing feature transmission and acquiring more semantic information from images. A dual attention module, consisting of channel attention and spatial attention, is introduced in the skip connection part of U-Net to emphasize the semantic and spatial features of buildings while suppressing irrelevant background information. The model fully utilizes the weight-sharing characteristics of the siamese network to share the image features of buildings before and after disasters. Experimental results on the xBD extraction dataset show that DEAU-Net achieves optimal results in various evaluation metrics for building extraction experiments, demonstrating its ability to perform cross-disaster assessment of building damage degrees.References

References

[1] ENGEL C B, JONES S D, REINKE K. A seasonal-window ensemble-based thresholding technique used to detect active

fires in geostationary remotely sensed data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(6): 4947-56.

[2] PIERDICCA N, ANNIBALLE R, NOTO F, et al. Triple collocation to assess classification accuracy without a ground truth in case of earthquake damage assessment [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(1): 485-96.

[3] YAMAGUCHI T, MIZUTANI T, TARUMI M, et al. Sensitive damage detection of reinforced concrete bridge slab by "time-variant deconvolution" of SHF-band radar signal [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(3): 1478-88.

[4] RITWIK G. xbd: A dataset for assessing building damage from satellite imagery [J]. arXiv preprint, 2019.

[5] GUPTA R, SHAH M. Rescuenet: Joint building segmentation and damage assessment from satellite imagery; proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), F, 2021 [C]. IEEE.

[6] WEBER E, KANé H. Building disaster damage assessment in satellite imagery with multi-temporal fusion [J]. arXiv preprint arXiv:200405525, 2020.

[7] SU J, BAI Y, WANG X, et al. Technical solution discussion for key challenges of operational convolutional neural network-based building-damage assessment from satellite imagery: Perspective from benchmark xBD dataset [J]. Remote Sensing, 2020, 12(22): 3808.

[8] CHEN S A, ESCAY A, HABERLAND C, et al. Benchmark dataset for automatic damaged building detection from post-hurricane remotely sensed imagery [J]. arXiv preprint arXiv:181205581, 2018.

[9] GUPTA R, GOODMAN B, PATEL N, et al. Creating xBD: A dataset for assessing building damage from satellite imagery; proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, F, 2019 [C].

[10] FOULSER-PIGGOTT R, SPENCE R, SAITO K, et al. The use of remote sensing for post-earthquake damage assessment: lessons from recent events, and future prospects; proceedings of the Proceedings of the Fifthteenth World Conference on Earthquake Engineering, F, 2012 [C].

[11] DONG L, SHAN J. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2013, 84: 85-99.

[12] 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 [J]. Remote Sensing, 2019, 11(23): 2858.

[13] SUMER E, TURKER M. Building damage detection from post-earthquake aerial imagery using building grey-value and gradient orientation analyses; proceedings of the Proceedings of 2nd International Conference on Recent Advances in Space Technologies, 2005 RAST 2005, F, 2005 [C]. IEEE.

[14] YE X, LIU M, WANG J, et al. Building-based damage detection from postquake image using multiple-feature analysis [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(4): 499-503.

[15] RUDNER T G, RUßWURM M, FIL J, et al. Rapid Computer Vision-Aided Disaster Response via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery; proceedings of the Proceedings of the First Workshop on AI for Social Good Neural Information Processing Systems (NIPS-2018), Montreal, QC, Canada, F, 2018 [C].

[16] ZHU Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 370-84.

[17] TEWKESBURY A P, COMBER A J, TATE N J, et al. A critical synthesis of remotely sensed optical image change detection techniques [J]. Remote Sensing of Environment, 2015, 160: 1-14.

[18] BLASCHKE T, KELLY M, MERSCHDORF H. Object-based image analysis: Evolution, history, state of the art, and future vision [M]. 2015.

[19] ZAGORUYKO S, KOMODAKIS N. Learning to compare image patches via convolutional neural networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2015 [C].

[20] XIAO H, PENG Y, TAN H, et al. Dynamic cross fusion network for building-based damage assessment; proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), F, 2021 [C]. IEEE.

[21] FU X, KOUYAMA T, YANG H, et al. Toward Faster and Accurate Post-Disaster Damage Assessment: Development of End-to-End Building Damage Detection Framework with Super-Resolution Architecture; proceedings of the IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, F, 2022 [C]. IEEE.

[22] FILIPE S, ALEXANDRE L A. From the human visual system to the computational models of visual attention: a survey [J]. Artificial Intelligence Review, 2013, 39(1): 1-47.

[23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].

[24] GUO J, MA X, SANSOM A, et al. Spanet: Spatial pyramid attention network for enhanced image recognition; proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), F, 2020 [C]. IEEE.

[25] WOO S, PARK J, LEE J-Y, et al. Cbam: Convolutional block attention module; proceedings of the Proceedings of the European conference on computer vision (ECCV), F, 2018 [C].

[26] SHEN Y, ZHU S, YANG T, et al. Bdanet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.

[27] HAO H, BAIREDDY S, BARTUSIAK E R, et al. An attention-based system for damage assessment using satellite imagery; proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, F, 2021 [C]. IEEE.

[28] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].

[29] ZHAO S, ZHANG X, XIAO P, et al. Exchanging Dual-Encoder–Decoder: A New Strategy for Change Detection With Semantic Guidance and Spatial Localization [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-16.

[30] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation; 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, F, 2015 [C]. Springer.

[31] LIU L, CHENG J, QUAN Q, et al. A survey on U-shaped networks in medical image segmentations [J]. Neurocomputing, 2020, 409: 244-58.

[32] [XU G, LIAO W, ZHANG X, et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation [J]. Pattern Recognition, 2023, 143: 109819.

[33] DAUDT R C, LE SAUX B, BOULCH A. Fully convolutional siamese networks for change detection; proceedings of the 2018 25th IEEE international conference on image processing (ICIP), F, 2018 [C]. IEEE.

[34] FANG S, LI K, SHAO J, et al. SNUNet-CD: A densely connected Siamese network for change detection of VHR images [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.

[35] CHEN P, ZHANG B, HONG D, et al. FCCDN: Feature constraint network for VHR image change detection [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 187: 101-19.

[36] ZHANG C, YUE P, TAPETE D, et al. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 166: 183-200.

[37] ZHANG L, HU X, ZHANG M, et al. Object-level change detection with a dual correlation attention-guided detector [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 177: 147-60.

[38] BUSLAEV A, IGLOVIKOV V I, KHVEDCHENYA E, et al. Albumentations: fast and flexible image augmentations [J]. Information, 2020, 11(2): 125.

[39] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features; proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, F, 2019 [C].

[40] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library [J]. Advances in neural information processing systems, 2019, 32.

[41] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:14126980, 2014.