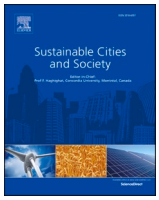
Sustainable Cities and Society 92 (2023) 104467

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs



Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images

Ziyao Xing [a,b], Shuai Yang [a,b], Xuli Zan [c], Xinrui Dong [a,b], Yu Yao [a,b], Zhe Liu [a,b], Xiaodong Zhang [a,b,*]

[a] College of Land Science and Technology, China Agricultural University, Beijing 100193, China
[b] Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing 100193, China
[c] Beijing Water Science & Technology Institute, Beijing 100048, China

ARTICLE INFO

Keywords:
Sreet view image
Remote sensing
Flood
Vulnerability assessment
Deep learning
Semantic segmentation

ABSTRACT

Urban flood risk management requires an extensive investigation of the vulnerability characteristics of buildings. Large-scale field surveys usually cost a lot of time and money, while satellite remote sensing and street view images can provide information on the tops and facades of buildings respectively. Thereupon, this paper develops a building vulnerability assessment framework using remote sensing and street view features. Specifically, a UNet-based semantic segmentation model, FSA-UNet (Fusion-Self-Attention-UNet) is proposed to integrate remote sensing and street view features and the vulnerability information contained in the images is fully exploited. And the building vulnerability index is generated to provide the spatial distribution characteristics of urban building vulnerability. The experiment shows that the mIoU of the proposed model can reach 82% for building vulnerability classification in Hefei, China, which is more accurate than the traditional semantic segmentation models. The results indicate that the integration of street view and remote sensing image features can improve the ability of building vulnerability assessment, and the model proposed in this study can better capture the correlation features of multi-angle images through the self-attention mechanism and combines hierarchy features and edge information to improve the classification effect. This study can support for disaster management and urban planning.

1. Introduction

With the continuous climate change, flood disaster risks are increasing worldwide, and nearly 1000 large flood events occurred between 2000 and 2018(Jongman, 2021; Tellman et al., 2021). The urbanization process along the coastline, river basins, and flood plains increases the exposure risk of urban flooding (Neumann et al., 2015; Kundzewicz et al., 2014), leading to enormous losses of property and life. This urgently requires to assessing and predicting flood risks for risk mitigation and urban planning (Wu et al., 2020).

The flood risk is the result of the joint action of the hazard factor and vulnerability of the hazard-affected body. The extent and intensity of the floods depend on topographic, soil and climatic factors (Ahmadlou et al., 2022), while the vulnerability of the hazard-affected body is the inherent sensitivity and natural attribute of the hazard-affected body responding to the strike of disasters (Bin et al., 2010). Vulnerability can be defined as the susceptibility of a hazard-affected body to damage or injury in a given disaster event (Timmerman, 1981). Buildings are one of the important hazard-affected bodies in urban disasters. Flood vulnerability assessment of buildings and related risk management had been proven to be an effective way to increase the flood resilience of the whole city (Stephenson, 2014). Especially in households with low income, it is crucial to mitigate and assess the damage caused by floods to buildings, which can effectively reduce economic losses (Chen et al., 2021).

The common methods of assessing the flood vulnerability of buildings can be divided into two categories: data-driven methods and indicator-based methods. The data-driven methods require multiple

Abbreviations: DCNNs, deep convolutional neural network; DSM, digital surface model; FN, false negative number; FSA-UNet, fusion-self-attention-UNet; GPU, graphics processing unit; mIoU, mean Intersection over union; P, precision; R, recall; SAR, synthetic aperture radar; SBD, semantic boundaries dataset; TP, true positive number; UAV, unmanned aperture radar.

* Corresponding author.
E-mail address: Zhangxd@cau.edu.cn (X. Zhang).

https://doi.org/10.1016/j.scs.2023.104467
Received 14 November 2022; Received in revised form 12 February 2023; Accepted 12 February 2023
Available online 23 February 2023

historical flood case data to calculate the damage rate through the existing disaster property loss data, or fit the vulnerability function of buildings under the different intensity of disaster factors (Papathoma--Kohle et al., 2022; Komolafe et al., 2019), etc. Data like flood hazard features, building characteristics, and reported losses should be collected (Amadio et al., 2019). Pham et al.(2022) assessed buildings' flood exposure based on flood information extracted from remote sensing, but didn't consider the possible damage degree of buildings. Marín-García et al.(2022) estimate the damage that a building could suffer under different environmental factors by machine learning. Although the results obtained by this method are easy to apply, there may sometimes be cases where it is difficult to comprehensively collect the building damage under different flood intensities in practice, resulting in certain uncertainties by using this kind of methods. Since physical vulnerability depends on building characteristics, the use of vulnerability indexes may be more helpful for qualitative assessment of vulnerability in areas where information is sparse. The research on index-based methods is continuously increasing (Papathoma-Köhle et al., 2019). Index-based methods need to obtain detailed information about the structure and materials of buildings and combine expert knowledge to construct an index system (Park, 2021; Ciurean et al., 2016), which may mean a lot of field work. Detailed field surveys and analysis building-by-building are expensive in terms of labor and time, limiting the ability to assess building vulnerability on a large scale, so finding reliable data sources the focus of future studies.

Remote sensing images can be used to quickly browse large areas of buildings and are less affected by ground conditions (Nex et al., 2019). In recent years, with the rapid development of high-resolution remote sensing technology, satellites have been able to provide sub-meter spatial resolution images. These images are available for free or at a low cost from several online platforms like Google Earth and Tianditu. Segmentation of these images can extract building footprints and the derived attribute layers by using deep learning methods (Jing et al., 2022; Abdollahi et al., 2020). However, conventional remote sensing images can only provide building roof information and cannot characterize detailed vulnerability characteristics like structure. Public street view images are increasingly available in cities and can be updated frequently, which can be used in solar radiation mapping (Deng et al., 2021), greenspace analysis (O'Regan et al., 2022), building height estimation (Hu et al., 2020), and so on. Street view images show the facade features of buildings from a human-like perspective, which has been applied to the vulnerability assessment of buildings in earthquake disasters by scholars (Pelizari et al., 2021). At present, the street view or remote sensing images are often used alone in disaster research, which is not enough to fine building vulnerability assessment. Integrating remote sensing and street view for flood vulnerability assessment of buildings is rare. How to overcome the effect of multi-angle images on the model remains to be studied.

The objectives of this study are as follows: First, to explore the feasibility of combining street view and remote sensing images instead of field surveys to assess the building vulnerability in flood disasters. Second, to construct a semantic segmentation model integrating the visual features of street view and remote sensing images, which improves the accuracy of the assessment. This paper constructs a vulnerability grading criteria based on the characteristics of the building itself. The buildings are divided into five levels according to the characteristics which can be acquired from images such as structure, condition, and the number of storeys. Moreover, a UNet-based model(FSA-UNet) is proposed, using ResNet-50 as the encoder and adding edge features and a self-attention block to fuse street view and remote sensing features into the decoder. Because Hefei, Anhui Province, China has suffered many times from floods, this paper take Hefei as an example to demonstrate the proposed method has a better segmentation effect than the traditional method with a mIoU of 82%. In addition, an urban building vulnerability index is constructed to measure the distribution of vulnerable buildings in different areas of the city.

The novelties of this work are listed as follows: (1) Both street view and remote sensing are used for flood vulnerability assessment of urban buildings, and the building vulnerability characteristics in the data are excavated and summarized. (2) A new model (FSA-UNet) is proposed to integrate street view and remote sensing images for building vulnerability assessment. The model can extract key features and improve the representational power of street view features in semantic segmentation for remote sensing image, and has the potential to be applied to other multi-angle image fusions. (3) The framework of building vulnerability assessment based on remote sensing and street view images is provided, and based on this, the footprint and spatial distribution characteristics of building vulnerability can be extracted with large-area coverage in a low-cost way. The research can help design sustainable and resilient urban buildings and disaster risk management.

2. Relate work

Remote sensing plays an increasingly important role in building identification, disaster risk, and rapid damage assessment with the characteristics of wide coverage and easy access. Many studies (Pham et al., 2022; Armenakis et al., 2017; Arabameri et al., 2019;) focused on building exposure research and remote sensing was used to obtain environmental information while the characteristics of the buildings themselves were ignored. The building vulnerability assessment requires the reconstruction and characterization of each building, with expert knowledge to determine vulnerability levels based on building attributes like size and roof type (Pelizari et al., 2021; Angela et al., 2013). Extracting building material types from remote sensing images is usually based on roof color and does not consider other parts of building like wall materials. However, many types of building roofs are similar, and vulnerability assessment based on remote sensing images alone is prone to be inaccurate. For example, Mück et al.(2013) used Quickbird data to derive building vulnerability characteristics like height, size, and shape, but the assessment was not automated enough.

Many researchers tried to integrate data from other sources to improve the accuracy. Barbierato et al. (2020), Ruggieri et al.(2021) and Geiss et al. (2014) combined the field survey data with information from remote sensing data to assess the vulnerability of buildings. In these studies, remote sensing often provides only the footprint of buildings. However, field data collection is often expensive, and data quality and evaluation criteria vary, making it unsuitable for large-scale applications. Researchers are constantly looking for data sources that replace or reduce field surveys. Polli et al. (2011) also used Synthetic Aperture Radar (SAR) data to obtain the height information of buildings and combined it with the roof information of optical images to assess the vulnerability to earthquakes. Wu et al.(2022) integrated multi-source remote sensing images, DSM, and traffic features to estimate building height. Geiss et al. (2015) used the support vector machine and random forest to supervise the classification of multi-source remote sensing images and DSM data, mainly considering the structural features of buildings. Moreover, non-orthophoto images of UAVs (Unmanned Aerial Vehicle) can offer more reliable proxies for building material assessment from building facades (Ilehag et al., 2017). Previous studies have proved the availability of comprehensive building shape, material, and height information in building vulnerability extraction, but the accuracy in some categories is still low. Besides, obtaining high-precision DSM data or UAV data is not easy, so it is urgent to find low-cost information acquisition methods.

Street view images, which are available for free and provide comprehensive information about buildings. Velez et al. (2021) proposed that street view images can collect a variety of flood vulnerability information such as building height, material, and condition. They used crowdsourcing to collect people's evaluation of building features in street view images, which provided ideas for subsequent research. Pelizari et al. (2021) used Deep Convolutional Neural Networks (DCNNs) to classify the structural vulnerability of buildings in street view images for

earthquake disasters. Pittore et al. (2012) combined the location gathered from remote sensing images with the height information of buildings obtained from street view images to assess the earthquake vulnerability of buildings. The street view has some limitations as well. Because of the occlusion of trees and other things, street view images do not completely cover every building. Building locations cannot be precisely determined by using street view alone (Zhao et al., 2019). Furthermore, street view images may be missing in areas that are inaccessible to traffic. Therefore, a single data source cannot meet the requirements of building vulnerability assessment. The current works have not established a completely automatic assessment model of building vulnerability integrating street view and remote sensing, which often requires tedious and split models. So our work will explore how to fuse data from multiple perspectives for reducing the difficulty of building vulnerability assessment.

Existing studies have proved that integrating complementary features of remote sensing and street view images can improve the recognition accuracy of multiple ground objects (Barbierato et al., 2020; Cao et al., 2018), which provides some reference for our research. Most studies used feature-level fusion for image classification tasks. For example, Chen et al. (2022a) and Fan et al.(2022) extracted features from different angles of street view images and remote sensing images to identify urban villages. Hoffmann et al.(2019) used the decision-level fusion method based on model fusion to train street view and UAV images respectively for building type classification. Most of the current models use image-level classification or scene classification. Per-pixel classification or semantic segmentation can obtain more information in the images at the same time, such as the location, area, and vulnerability level of buildings, which is more accurate than image-level classification. Multi-modal data such as optical image and SAR (Fan et al., 2022), optical and Lidar image fusion (Sun et al., 2021) for semantic segmentation have been proven to be possible. However, street view images and remote sensing images respectively cover information from different angles. Whether additional processing is needed for semantic segmentation needs further exploration.

3. Study area and data

The study area is located in the urban area of Hefei, Anhui Province, China (Fig. 1). Hefei has a subtropical monsoon climate with a relatively well-developed surface water system, and Chaohu Lake in its territory is one of the five major freshwater lakes in China.

The very-high resolution remote sensing images are needed for the fined assessment of building vulnerability. The remote sensing image of the Tianditu website (https://www.tianditu.gov.cn) is obtained from 91 Weitu Assistant for free. The images is 16-level with a spatial resolution of about 0.5 m, including R, G, and B bands, which is composed of multi-phase cloud-free images in 2021 collected by Worldview satellites.

The street view images come from Baidu Map taken in August 2020. As shown in Fig. 2, each street view picture is a 360-degree panorama using images taken by the street view collection vehicle from various angles. A total of 439,783 panoramas is downloaded, and the size of the picture is 4098×2048, which can completely cover the main streets of the Hefei urban area, and include some internal roads within residential areas, schools, etc. Street view image acquisition needs to use Baidu API which searches street view pictures according to latitude and longitude. So we first get the road network data of the study area through OpenStreetMap, turn the road network lines into sampling points with 50 m intervals, and then search for the nearest street view image id by the point one by one through Baidu Map API (https://mapsv0.bdimg.com/?qt=qsdata&x={x}&y={y}), according to which all the street view images on the whole road can be obtained.

4. Method

In this section, the proposed workflow for building vulnerability assessment to floods is presented (Fig. 3). First, a flood vulnerability grading criteria for urban buildings is constructed according to the related pieces of literature and building features. Second, this paper introduces the method of data collection and preprocessing. The preprocessing is the basis of the semantic segmentation model of data fusion. This paper then give a short general introduction to UNet, after which this paper go into the specifics of the proposed semantic segmentation model FSA-UNet architectures for building extraction and vulnerability grading. For areas without street image coverage, the outlines of the buildings are extracted by UNet without labeling for vulnerability grading. Then, a similarity calculation is conducted between the labeled and unlabeled buildings to give every building a vulnerability level. Finally, an urban building vulnerability index is developed based on the vulnerability grading result to understand the spatial distribution of building vulnerability.

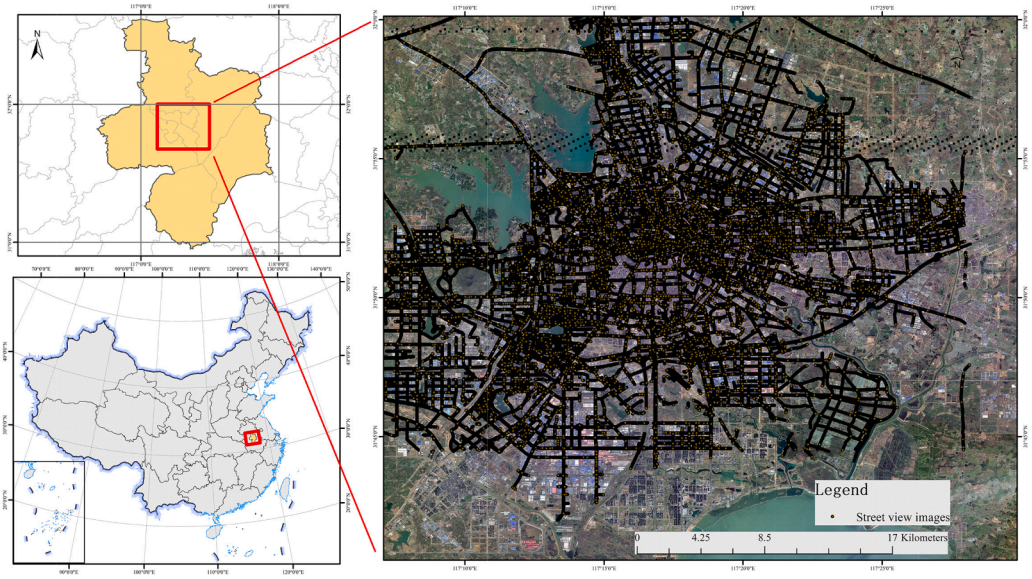


Fig. 1. Location of the study area and distribution of street view images.

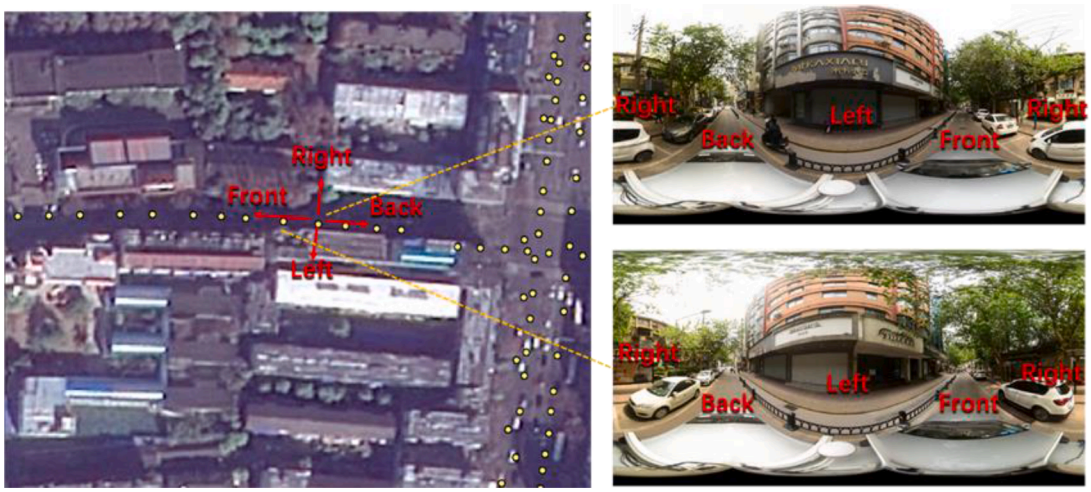


Fig. 2. Corresponding location of street view pictures and remote sensing images.

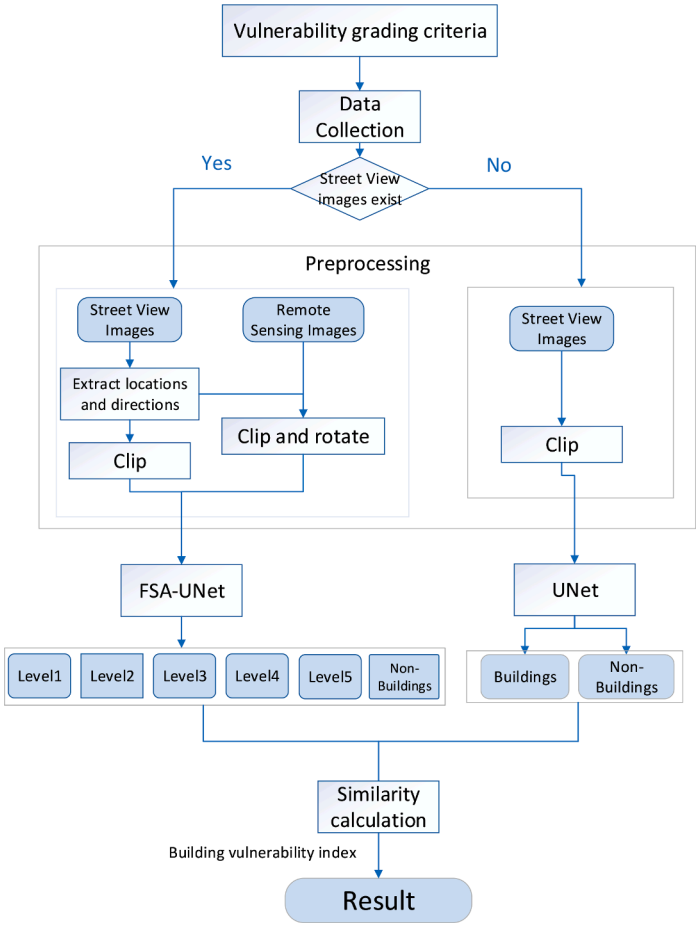


Fig. 3. building vulnerability assessment workflow.

4.1. Vulnerability grading criteria

Both the impact and immersion of flooding can cause damage to buildings. The ability of building materials and structures to resist flooding is important (Rana & Routray, 2016). For example, flood with higher velocity has stronger damage to wooden buildings, and the flood control ability of masonry buildings is higher than that of adobe buildings (de Ruiter et al., 2021). According to the damage investigation and research of Balasbaneh et al. (2019), concrete buildings have stronger flood control ability than masonry buildings and adobe buildings. The vulnerability of steel structures and reinforced concrete is similar, and they are usually classified into the same grade (Usman Kaoje et al., 2021; Miranda et al., 2019).

The number of storeys in a building is also an important determinant of damage, as residents living in single-floor houses tend to suffer more damage than multi-floor houses, whereas residents living in multi-floor houses can easily removable items to higher storeys (Z. Li et al., 2022a). In addition, numerous studies have shown that older buildings have a

higher rate of damage at the same water depth (Dall'Osso et al., 2009). Based on this, this paper have summarized the building vulnerability indicators as shown in Table 1.

Remote sensing images can provide the roof structure and color, but some similar roof features of buildings are difficult to distinguish. Street view images can get material information through the features of exterior walls, and it is easier to see the extent of building damage. And the number of storeys of buildings is clearer in street view images. Combined with the building vulnerability indicator in Table 1 and the distinguishing ability of remote sensing and street view images, images are grouped according to the number of storeys first. In China, buildings with nine or more stories are mostly defined as high-rise buildings which are all reinforced concrete structures and built very recently. So the high-rise buildings are grouped into Level 1 with the lowest vulnerability. And then the mid-rise buildings with 3–8 floors are grouped together. Since there are very few single-story buildings in the study area, we grouped the story of 1–2 buildings into the same level. Next, the conditions of buildings are considered in groups of story numbers. Based on our observations, the damaged and dilapidated ones are the adobe or brick buildings with low height, constituting the highest vulnerability level (Level 5). We then classify other levels of vulnerability based on the structure and material of the building. Most residential and office buildings are mid-rise buildings with reinforced concrete or masonry-concrete structure and in good/excellent condition, classified into Level 2. The steel structure buildings in the study area are mainly industrial plants. Although steel structure is more resistant to flooding, these buildings have lower floors with moderate vulnerability (Level 3). Besides, there are many masonry-concrete buildings with 1–2 storeys classified as slightly higher vulnerability(Level 4). Finally, the buildings in the study area are divided into 1–5 levels as Table 2.Based on the criteria, buildings in the research area are labeled for semantic segmentation.

4.2. Image preprocessing

The information contained in the both remote sensing and street view images is relatively complex. To make it easier for the model to learn the features of buildings, remove irrelevant information, and match remote sensing and street view, image preprocessing is necessary. The steps are as follows (Fig. 4).

(1) Extract street view images.

Each street view image has the coordinates of the shooting position. The distance between the coordinates of every two images is affected by vehicle speed, and the average distance is about 6–12 m. To reduce the computational burden, this paper extract one street view image every 7 intervals.

(1) Identify the driving directions and trajectories of the street view collection vehicles. Clip and rotate the remote sensing images.

Table 1
building vulnerability indicators.

Vulnerability	Types	Condition	Number of story
High	Adobe building, temporary building, wooden building	Poor. The walls or roof are badly damaged and built over time	1
Median	Brick structure building, masonry-concrete, stone structure building	Good. The building's exterior structure is intact. No damage or slight damage to the wall.	2–3
Low	Reinforced concrete building, steel building	Excellent. The building is in good condition and built in a new age	>=4

The perspective of street view images is always centered on the road where the vehicles are driving. However, in remote sensing images, the location of the road is uncertain. To achieve the relative unity of the same ground object position in the two kinds of images, the remote sensing images need to be rotated according to the driving directions and trajectories. First, the azimuth angle between every two images is calculated according to the coordinates to obtain the driving azimuth and direction of the vehicle. The straight line where the coordinates of the two images are located is used as the driving trajectory of the vehicle.

Then, take the extracted street view images as the center point, and cut the remote sensing images as 712 × 712. Taking the trajectory as the axis, the remote sensing images are rotated according to the azimuth of street view images. The rotated remote sensing images take the driving trajectory of the vehicle as the central axis, and the upper part of remote sensing images always corresponds to the left side of the vehicle as shown in Fig. 4. Then each remote sensing image is cut into the upper and lower parts with a size of 512×256. Finally, the collection point of street view images in each remote sensing image is located at the middle point of the bottom edge of the remote sensing image.

(1) Clip street view images.

The size of a panoramic street view image is 4096 × 2048, in which the road, sky and other background information occupy a large part. The locations of buildings in about 200 images of different areas were counted, and then the top and bottom of the images were removed. Each image is re-spliced and clipped into two pictures of the left side of the vehicle and the right side of the vehicle. The final size of each street view image is 2000 × 1000.

4.3. Semantic segmentation with FSA-UNet (Fusion-Self-Attention-UNet)

4.3.1. Overview of FSA-UNet

As shown in Fig. 5, this paper propose an improved UNet model called FSA-UNet. UNet is one of the mainstream semantic segmentation models, which is derived from the fully convolutional network. It solves the semantic segmentation problem in an end-to-end way and has achieved good results in the fields of remote sensing image classification. FSA-UNet uses the backbone architecture of the UNet model, including two parts, namely the contraction path (encoder) and the expansion path (decoder). The encoder part extracts hierarchical features by gradually downsampling the spatial resolution of the feature map, while the decoder part learns more context information by gradually recovering the spatial resolution. Diakogiannis et al. (2020) pointed out that ResNet (He, 2016) can improve the performance when used for semantic segmentation of remote sensing data. The RestNet-50 is chosen as the encoder to extract low-level to high-level features of remote sensing and street view images respectively. ResNet-50 firstly reduces the original image to 256 × 256 × 64 by a 7 × 7 convolution and 3 × 3 maximum pooling to obtain the first-level feature map. Then the feature map is input into four stages of resblock. Each stage includes two 1 × 1 convolution kernels and a 3 × 3 convolution kernel for dimensionality reduction to downsample the feature map, and the feature maps of levels 2–5 are obtained.

The traditional UNet model uses the left-right symmetric U-shaped structure to directly fuse the deep features and shallow features through a layer-skipping connection, which may lead to the loss of boundary information. Therefore, an edge feature block is added in the last decode layer to strengthen the edge features. Moreover, to integrate remote sensing and street view images, this research designs a multi-source feature fusion self-attention block. In this block, remote sensing features are used to guide the weight calculation of street view features which are finally contacted with the remote sensing features in the decoder. For the balance between computing resource consumption and segmentation performance, the self-attention is only used for the deep

Table 2
flood vulnerability levels of buildings in the study area.

Level	Types	Condition	Number of story	Images
1	Reinforced concrete building	Excellent.	9>=	
2	Reinforced concrete or masonry-concrete building	Excellent or Good.	>=3 and <9	
3	Steel building	Excellent or Good.	1–2	
4	Masonry-concrete building	Good	1–2	
5	Adobe building or brick building	Bad	1–2	

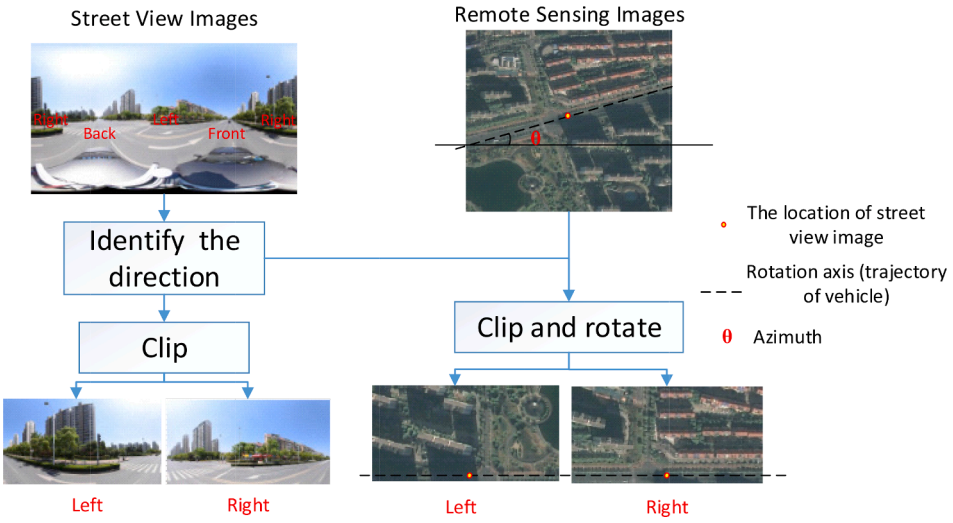


Fig. 4. Data preprocessing workflow.

features of the highest dimension. To solve the problem of large memory usage, images or features can be chunked and then input into the self-attention mechanism separately (Diaz et al., 2021). However, the corresponding areas of the remote sensing and street view images cannot be judged in advance, and there may be no corresponding effective feature in the same block after being chunked. Therefore, in the last layer, the original street view features are directly fused with remote sensing features and edge features. The original features of the lower layers can be preserved, which further improves the segmentation effect. The self-attention block and edge detection block will be introduced in the following.

4.3.2. *Multi-source feature fusion self-attention block*

A traditional self-attention mechanism (Vaswani, 2017) is used to learn the relationship between different parts of the input matrix to extract global context information. Some scholars used self-attention modules for multimodal interaction of verbal and visual content (Wei et al., 2020; Lu et al., 2019). Inspired by these models, a self-attention module for integrating remote sensing and street view images is constructed to process street view features.

The attention function can be described as mapping a query and a set of key-value pairs to an output, where the query Q, key K, value V, and output H are all vectors. The multi-source feature fusion self-attention is calculated by the following formulas (Vaswani, 2017).

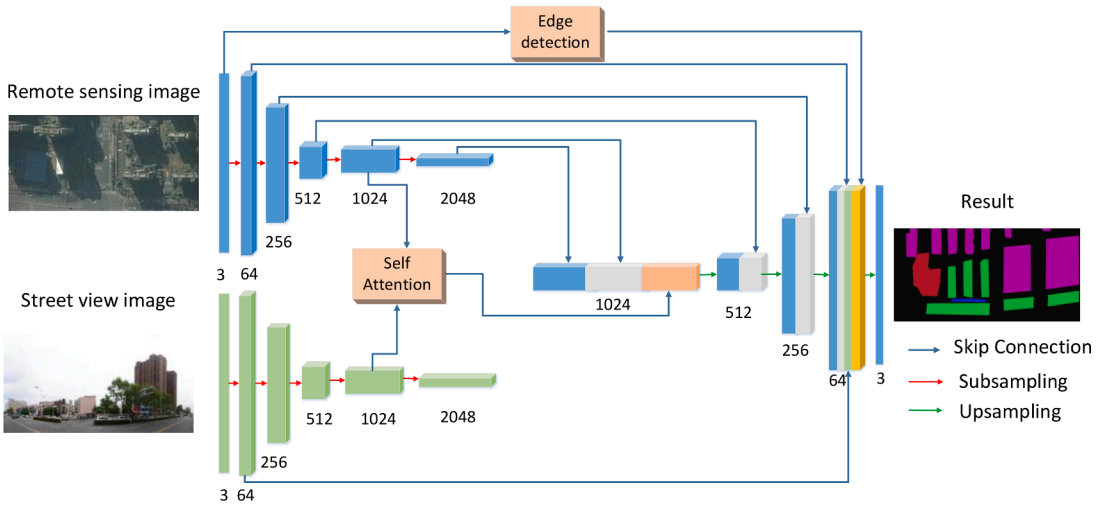


Fig. 5. The architecture of the proposed FSA-UNet.

$$H = \text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)\text{V} \tag{1}$$

where H is the output matrix, d_k is the dimension of the value matrix and key matrix, softmax is the function normalized by column. f_r is the remote sensing feature and f_s is the street view feature.

In this paper, the low-level features of street view and remote sensing images are extracted respectively based on ResNet-50, and then input into the self-attention block. As shown in Fig. 6, firstly, the features of remote sensing images and street view images are normalized. Using remote sensing image feature as input, the query matrix Q is obtained through a linear transformation. The key matrix and value matrix are obtained from the street view feature. Then, the output vector H is obtained by using Formula (1).

In this block, street view and remote sensing features are interacted to capture the correlation between them and find the street view features that the model should pay more attention to, so as to enhance the representation ability of street view features.

4.3.2. Edge detection block

The UNet model has the problem of losing edge information, which leads to the inaccuracy of segmentation. This paper used the Sobel operator (Kanopoulos et al., 1988) commonly used in edge detection to extract edge information from the original remote sensing images. Compared with Robert, Prewitt, and other operators, Sobel has higher efficiency (Chaple et al., 2015), and it is found through experiments that it has a better effect on the edge extraction of buildings. Therefore, this proposed model use the Sobel operator to obtain the gradient information feature vector of the image target region, and its horizontal and vertical matrix G_x and G_y are as follows.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{3}$$

This block converts the Sobel operator into a convolution kernel as $\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$ that adapts the convolution operation. The original three-band remote sensing image is calculated and then input into the 2D convolution with the weight of this convolution kernel. Then, a 7×7 convolution kernel is used to reduce the dimension of the feature map with a size of (64,256,256) and input it into the FSA-UNet network to join with other features. The block can enhance the recognition ability of ground object edge, especially for the area where the building is shaded or the ground has similar features to the building.

4.3.3. Model training

Before model training, this paper divided the training dataset into a training set and a test set, with a ratio of about 9:1 to conduct 10–fold cross-validation, meaning that the whole dataset is broke into 10 equal parts (folds). Each time we train the model on 9 parts and validate on the remaining part for 10 times. We extract one street view image every 7 intervals from all images, and 5696 images were randomly selected to form the training set. To make up for the small number of training

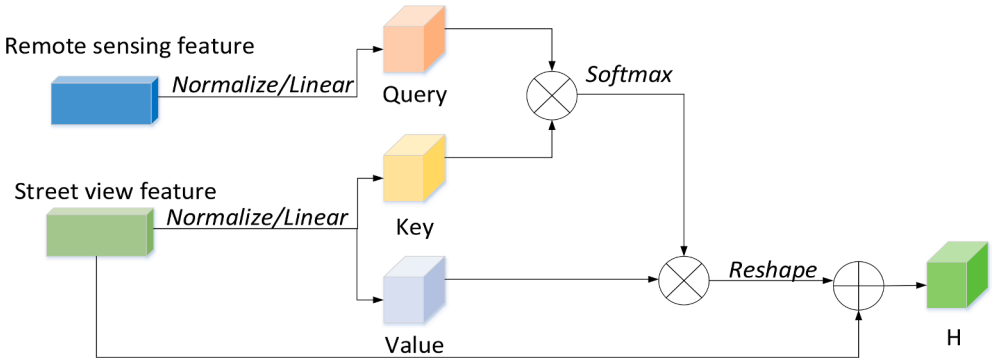


Fig. 6. Multi-source feature fusion self-attention block architecture.

samples, this paper used transfer learning, that is, used pre-trained models based on the PASCAL SBD dataset (VOC 2021 extended dataset) and VOC 2012 dataset. We trained for 150 epochs with a batch size of 8. The RAM of GPU allowed us to test a maximum batch size of 8, and we also test the batch size of 4, which has no effect on the accuracy of the model. All process is based on the PyTorch 1.7.0 and conducted on an Intel(R) Xeon(R) CPU E5–2678 v3, 64 GB of RAM, and two Nvidia RTX 2080 graphics processing unit (GPU).

4.4. Urban building vulnerability index

Based on the vulnerability levels of the buildings, this paper proposes an index to calculate the spatial distribution of building vulnerability in urban areas. In this paper, the study was divided into 500 m grids, and the index I_v was calculated for each grid to obtain the vulnerability degree of different regions.

$$I_v = \frac{\sum_{i=0}^{5} S_i W_i}{S_{all}} \tag{4}$$

where i stands for levels 1–5 of vulnerability. W_i is the weight of each level and levels 1–5 are assigned 1–5 respectively in this paper. S_i represents the area of buildings of level i in a grid, and S_{all} is the total area of buildings.

5. Results

5.1. Similarity calculation for expanding the vulnerability grading results

As street view images can only cover buildings near roads and the UNet model is not accurate in the classification of building vulnerability in areas without street view images, the UNet model is only used to classify buildings in areas without street view images into two categories: building and non-building, with the mIoU of 85%.

It is observed that the types of buildings in the study area tend to show aggregation characteristics, the types of buildings in a certain space are often the same, and the image features are quite similar. Therefore, the following steps are carried out:1.

(1) Obtaining the outlines of the buildings by UNet and FSA-UNet.2. (2) Search the nearest 8 marked buildings with the coordinates of the center point of each building without marked vulnerability levels as shown in Fig. 7. (3) According to the external rectangle of every building's outline, cut the original remote sensing image and a rectangle picture for each building can be obtained. Then reshape the 8 rectangle pictures of the marked buildings into the same size as the rectangle of the unmarked building. (4) Extract the RGB of each building picture, and calculate the similarity degree between the unmarked building and the marked building image by using the Euclidean distance one by one. (5) The vulnerability level of the most similar building is selected as the unlabeled building level, and the building vulnerability category of the whole study area was finally obtained.

Fig. 7. Schematic diagram of the calculation process. The green area represents the buildings that can be observed in street view images, which are classified by FSA-UNet. The red box is the building to be classified, and the green boxes are the buildings with which the similarities are calculated.

5.2. Model effectiveness analysis

5.2.1. Comparison of classification results

In this paper, we use Precision(P), $Recall(R)$, mean Intersection over Union (mIoU), and accuracy as the methods of accuracy evaluation. The formulas are referred to Bello et al., 2021; Shang et al., 2022.

$$\text{Precision(P)} = \frac{TP}{TP + FP} \tag{5}$$

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

$$\text{IoU} = \frac{TP}{TP + TN + FN} \tag{7}$$

$$\text{mIoU} = \frac{1}{n}\sum_{i=1}^{n} IoU_i \tag{8}$$

where TP is the true positive number, FP is the false positive number, and FN is the false negative number. n is the total number of categories. IoU_i is the cross-merge ratio of the category i.

The proposed FSA-UNet model is compared with the commonly used semantic segmentation models, including Deeplabv3+, PSPNet, HRNet, and UNetFormer (Table 3). DeepLab V3 + uses atrous convolution and combines the spatial pyramid pooling block and encoder-decoder structure to refine segmentation results. PSPNet is a pyramid scene parsing network exploiting global contextual information. HRNet changes the link between high and low resolution from series to parallel. These models were mainly chosen for comparison in many papers on semantic segmentation of remote sensing images (Z. Li et al., 2022b; Jeon et al., 2021; Liu et al., 2021; Wu et al., 2021). In particular, we compare the UNetFormer model proposed for urban image segmentation, which is based on ResNet-15 and the global-local attention mechanism. However, it does not achieve good results in this paper, we believe that the feature extraction capability of ResNet-15 is not strong enough to meet the classification requirements of multi-type buildings. FSA-UNet is the highest in mIoU, Recall, and Precision, followed by the original Res50-based UNet model, which shows that UNet performs better in the building classification compared to other traditional semantic segmentation models, and our improvement can further enhance the extraction of building vulnerability features.

According to the prediction results of each model in Fig. 8, it can be seen that the miss rate of DeeplabV3+ and PSPNet is very high. HRNet has difficulty correctly distinguishing the vulnerability types of buildings. The models of the UNet series are more accurate, but there are cases where pavements with certain distinct edge features are classified as buildings (e.g., the third set of figures).In addition, in the edge areas of some buildings (the first set of figures), UNetFormer incorrectly classifies them into other building types, but the FSA-UNet proposed in this paper can greatly reduce this error. Taking the second set of images as an example, the blue building (Level 3) on the upper left is very similar to Level 2 in the remote sensing images, but the difference can be seen in the street view images, which makes the classification result of adding the street view feature more accurate. As shown in the fifth group of figures, Level 3 is more difficult to distinguish from Level 4 and 5 from remote sensing images. FSA-UNet also has some misclassification in

Table 3
Accuracy comparison of different models.

Model	mIoU	Recall	Precision
Deeplabv3+	64.15	73.7	79
PSPNet	68.6	79.31	83.85
UNet	75.8	83.95	87.6
HRNet	69.47	78.92	83.85
UNetFormer	72.44	74.40	89.86
FSA-UNet	**82.23**	**88.43**	**92.07**

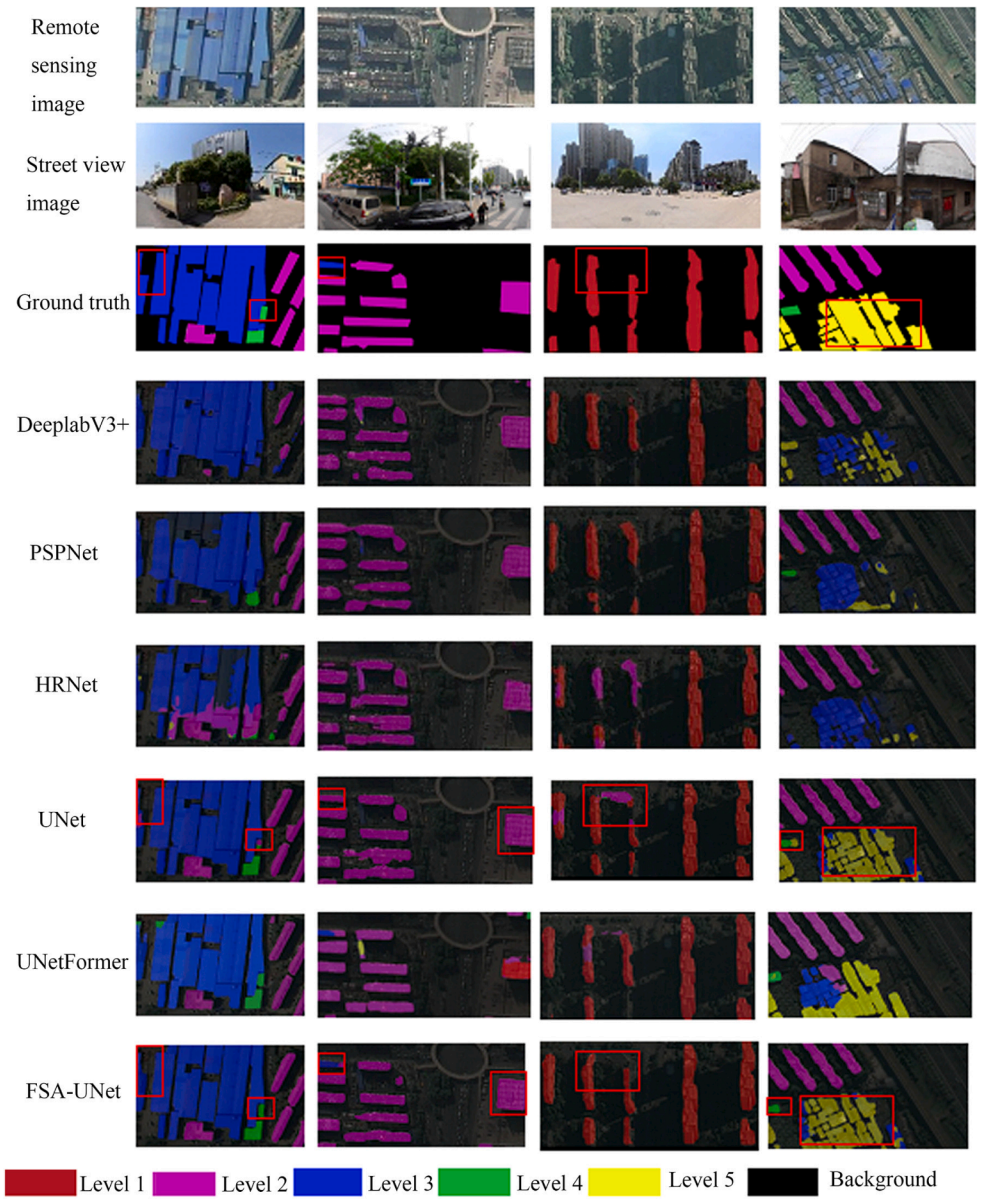


Fig. 8. Comparison with the common models in semantic segmentation.

distinguishing between Level 3 and Level 5, but the results are better than the other models.

We plot the mIoU and loss curve of the validation set during the training of FSA-UNet (Fig. 9). The model stops improving its accuracy around epoch 130. From the confusion matrix (Fig. 10), it can be seen that the classification accuracy of Level 2 and Level 5 exceeds 91%. Level 4 has the worst classification effect. The buildings in Level 4 are mainly low-rise buildings, generally small in size, with colors more similar to bare ground and roads, which are easily mixed with the background. Additionally, the high-rise buildings of Level 1 are mainly affected by the shadows of the buildings and are easy to be misclassified. The fusion of street view images has improved the differentiation ability

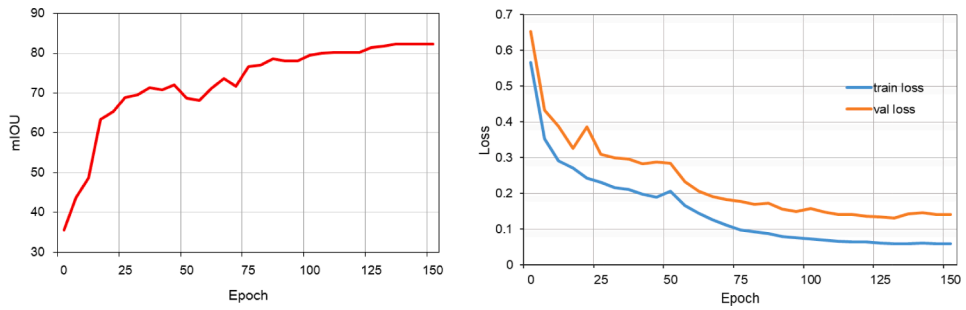


Fig. 9. mIoU and loss curve.

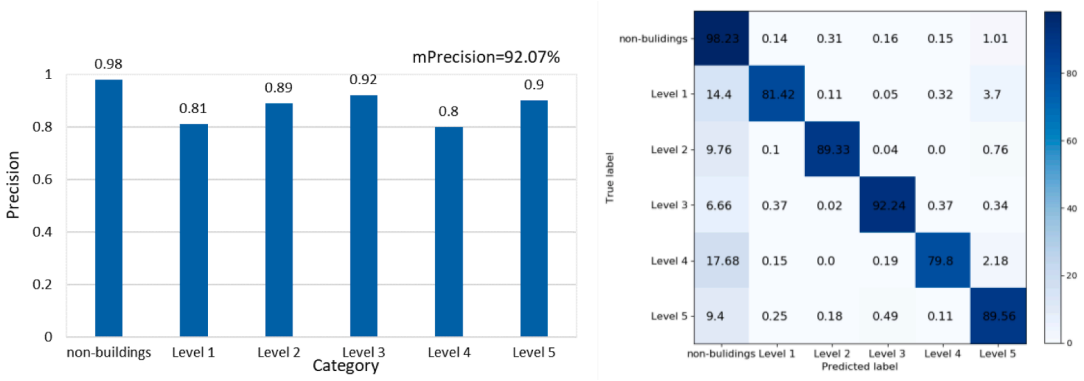


Fig. 10. Precision of categories and confusion matrix.

between Level 3, Level 4, and Level 5 to some extent, but the classification accuracy is still low because low-rise buildings are more likely to be obscured in street view images.

5.2.2. *Ablation experiment*

We designed an ablation experiment using the ResNet-50-based UNet as a baseline to determine the role of each block (Table 4).

Compared with the baseline (Experiment 1), adding the multi-source feature fusion self-attention block separately (Experiment 2) can improve the mIoU by about 2%, and then adding the original low-level features (Experiment 3) can improve the accuracy by 6%. We also compared the direct concatenation of remote sensing with the street view image at the same location (Experiment 4), that is, without the multi-source feature self-attention block, directly concatenating the low-level features of the first layer and the high-level features of the fourth layer from the street view and remote sensing, and the mIoU is 78.05%, which is a 3.7% decrease. This illustrates that lower-level street view features are more important in our task, and the self-attention mechanism does facilitate the alignment of higher-level street view and remote sensing features. The information related to buildings in street view images can be given greater weight to supplement the information lacking in remote sensing.

The edge detection block is added to the baseline model in experiment 5, and the mIoU is increased by 1.36%. This block strengthens the edge features of objects and makes up for the loss of edge information with the deepening of the network. Finally, after adding the edge detection block based on multi-source feature self-attention block and low-level street view features, which is the proposed FSA-UNet, the mIoU and Recall are increased but accuracy is decreased compared to experiment 5. Although the improvement was not significant, the effect was slightly better when predicting in the study area. Some broken areas were reduced, especially in Level 3. Industrial buildings of Level 3 tend to have large areas, and the interior or edge of buildings will be divided into the background.

Table 4
Accuracy comparison of different blocks.

Experiment		mIoU	Recall	Accuracy	Training time per epoch (min)
1	Baseline	75.8	83.95	94.92	7
2	+multi-source feature self-attention block	77.12	85.01	95.16	17
3	+high and low level street view features	78.05	85.99	95.75	17
4	+multi-source feature self-attention block and low-level street view features	81.77	88.31	96.14	17
5	+edge detection block	77.16	84.99	95.18	8
6	FSA-UNet	82.23	88.65	96.03	17

In addition, compared to the baseline model, FSA-UNet takes more than twice as long to train, mainly because it takes a longer time to load street view data and extract features using ResNet-50. The time loss caused by the self-attention mechanism is almost negligible.

5.3. *Spatial distribution of building vulnerability*

We applied the proposed method in Hefei, and the spatial distribution of various vulnerability levels of buildings is shown in Fig. 11. Most of the high-rise buildings of reinforced concrete structures with the lowest vulnerability are located around the Second Ring road. The construction age is relatively new, which is related to the development of the city. The area surrounded by Huancheng Park Road is almost all multi-story and low-rise concrete buildings with moderate and low vulnerability. While industrial buildings are widely distributed outside the main urban area. For example, buildings of Level 3 which are mostly steel structures can be seen in the north and south of the study area, but with the low story and medium vulnerability. Level 4 buildings are few and scattered all over the city, these buildings are low and often single, and most of them are small commercial buildings along the streets on the outskirts of residential. Most of the areas with the highest vulnerability are located in the northeastern part of the East 2nd Ring Road and the eastern part of Feidong County. These buildings are usually densely distributed with little green space in the surrounding, which is not conducive to flood drainage.

By calculating the vulnerability index of buildings (Fig. 12), it can be found that the vulnerability of Baohe District, Luyang District, Shushan District, and Yaohai district is increasing from inside to outside, with the old city at the intersection as the center. Among them, the buildings in Changfeng County and Feidong County are the most vulnerable. Feidong, Feixi and Changfeng County are important parts connecting urban and rural areas. To further expand the urban area, it is necessary to further improve the fragile degree of buildings, especially in the area adjacent to Nanfei River, Chaohu Lake, and other river systems, the overflow of rivers will lead to more serious building damage.

6. Discussion

Since both remote sensing and street view have strengths and weaknesses, integrating them is demonstrated to be a great way to improve the assessment. In this paper, a building vulnerability assessment framework using remote sensing and street view features is developed, which provides a workflow including vulnerability grading criteria, image preprocessing, semantic segmentation and so on. The proposed model FSA-UNet can fusion the multi-angle features and uses street view to supplement for the lack of building facade information in remote sensing. To our knowledge, this is the first attempt to integrate remote sensing and street view in building vulnerability assessment for floods, and its feasibility effectiveness is proved. The principal findings,

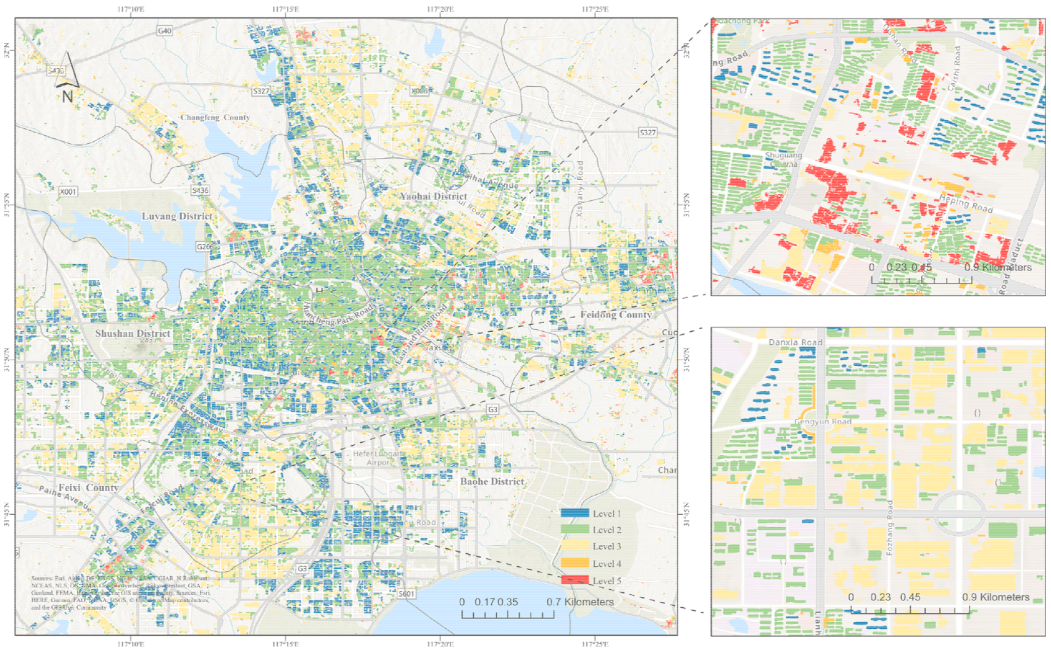


Fig. 11. Spatial distribution of building vulnerability outputs from FSA-UNet.

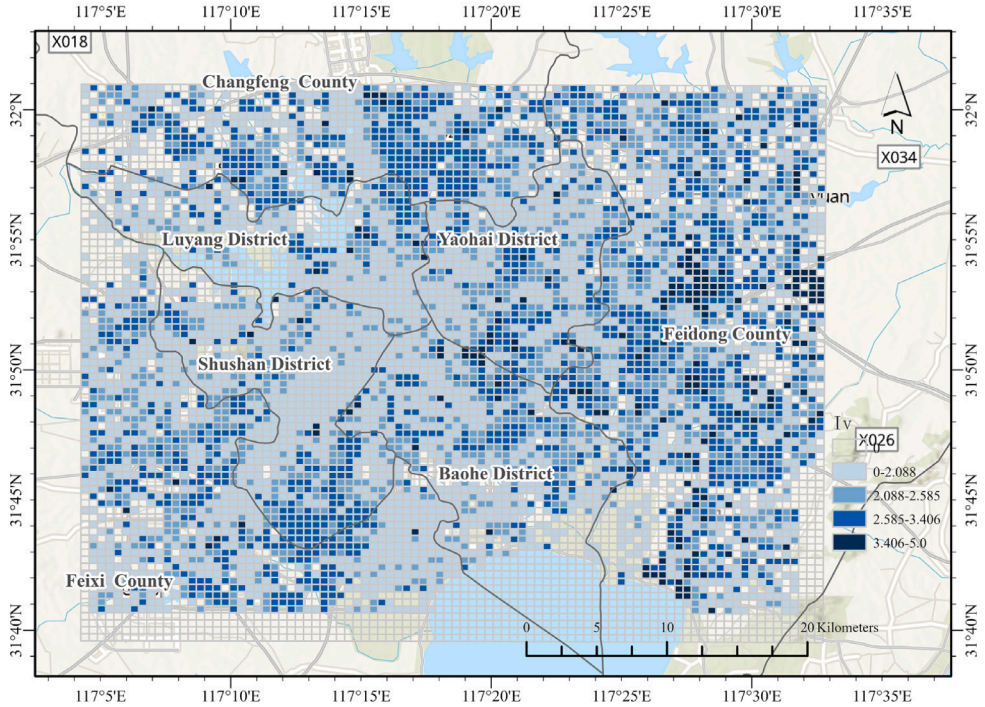


Fig. 12. Spatial distribution of the building vulnerability index in 500 m grids.

limitations, and further studies are discussed in the following.

6.1. Principal findings and comparison with previous researches

6.1.1. Flood vulnerability assessment of buildings

Buildings are the primary shelter for people in floods, so understanding and reducing the vulnerability of buildings is critical. In previous research, street view images are only used to aid the field surveys in flood vulnerability assessment (D'Ayala et al., 2020; Velez et al., 2021).This paper attempts to use street view and remote sensing at the same time to get the distribution of building vulnerability directly and proves that integrating remote sensing and street view images can improve the ability of vulnerability assessment. The advantage of street view images is that it directly reflects physical characteristics like building facade material, condition, and height. Remote sensing images provide building footprints and roof information in a large scale. Through the experiments in this paper, the most vulnerable buildings tend to be dilapidated and densely packed, clustered in some urban villages. The flood vulnerability of urban villages has been widely studied. These areas lack appropriate planning with high vulnerability and low-income residents (Marko et al., 2019; Erena et al., 2019). Accurate understanding of the distribution of these areas and the

characteristics of buildings is very important for urban planning, construction and disaster management.

On the other hand, the findings of this paper and previous research (Chen et al., 2022b) indicate that the flood vulnerability is different in urban centers, suburbs and rural areas, which can be attributed to the urban overutilization in China (Song et al., 2019). In this paper, the eastern part of Hefei has relatively old and fragile buildings, suggesting that relevant departments should strengthen the spatial management of flood risk and improve the waterlogging prevention ability of buildings.

6.1.2. The key to integrating remote sensing and street view images

It is worth discussing how to fuse remote sensing and street view images in the model. Common approaches to multimodal data fusion can be divided into three types: feature-level fusion, image-level fusion, and model-level (decision-level) fusion. Hoffmann's experiments (Hoffmann et al., 2019) of fusing street view and aerial images concluded that decision-level fusion is usually better than feature-level fusion. However, Chen et al. (2022a) and Fan et al. (2022) both chose the feature-level fusion. The decision-level fusion directly combines the softmax probabilities or the classification labels. However, the location of the same ground object in the remote sensing and street view image is completely different, so it is difficult to achieve good results by connecting the labels directly in the semantic segmentation task. And producing labels of both remote sensing and street images will consume a lot of labor. We have also tried image-level fusion. The RGB of street view and remote sensing images are directly concatenated into 6 channels matrix as the input of the model. Without proper pre-training, the model is difficult to converge. In experiment, the proposed model in this paper proved that the feature-level fusion is effective. At present, there are still few papers on the fusion of remote sensing and street view for semantic segmentation, but the same result can be proved in other papers of semantic segmentation based on multi-modal data such as optical and LiDAR (Iyer et al., 2020), Optical and SAR (Zhang et al., 2020) and hyperspectral and lidar data(Man et al., 2015).

According to our research, the biggest problem in feature-level fusion is the feature alignment between remote sensing and street view images. Laumer et al. (2020) proposed a global optimization method to locate detected trees in street view images, which combines multi-view detections per tree into a single representation. Ning et al. (2021) use a projection approach where the street scene is first semantically segmented and then projected to an ortho-image based on area. But building shadows or vegetation occlusion may lead to inaccurate projection of object location.

To solve this problem, we first rotate the remote sensing images during preprocessing, so that the roads are always under the images, and the left and right ground objects in street view images are the same as the left and right objects in remote sensing images. The subsequent model training results show that processing such as rotation can initially enable the model to match the remote sensing image with the street view image, which is considered simple and effective compared to traditional methods such as the projection of features. We try to use remote sensing images without rotation directly, which leads to low accuracy. The self-attention mechanism in FSA-UNet further solves the problem of feature alignment. The self-attention mechanism makes the street view features related to remote sensing features given more weight, and the irrelevant parts of street view features are not retained, and this block can learn the direct correlation between street view and remote sensing, which improves the usability of street view features.

In addition, this paper calculates the similarity between remote sensing features of buildings to supplement the evaluation results of the area without street view images, while almost all the existing studies only give the result near the streets (Pelizari et al., 2021, Chen et al., 2022a). Moreover, low buildings may be entirely obscured by trees on both sides of the street. In practical applications, street view images collected in winter can be chosen as much as possible to reduce the influence.

6.2. Limitations and future research

First, the vulnerability assessment criteria in this paper needs to incorporate more expert knowledge. Besides, it is more qualitative than quantitative now. The height information in the street view images can be further quantified, for example, by the location of the roof line to measure the height of the building (Zhao et al., 2019). The height of doors and windows of buildings in street view images can be used to evaluate whether buildings are prone to water intake, which can be estimated by object detection of doors, windows, and stairs. Moreover, this paper only puts the emphasis on the building's own characteristics. The environment is considered to be an important factor for the vulnerability of disaster-bearing bodies in a broad sense. Environmental information such as DEM, building density, and impervious water surface can be further considered. In addition, the building characteristics in the dataset are still relatively limited. By collecting more samples, the proposed method can be applied in more areas and can identify diverse building vulnerability information.

Broader integration of space-air-ground data is a developing trend in urban planning, disaster management and other research. The proposed FSA-UNet provides a method for semantic segmentation of multi-angle data, which is not only applicable to remote sensing and street view images. The images of video surveillance or mobile phone on social media can also be combined with remote sensing images. Many studies have shown that social media data has potential for assisting urban flood monitoring and risk assessment (Li et al., 2023; Ouyang et al., 2022). For example, in the extraction of urban flood extent, floods in remote sensing images are easy to be blocked by buildings and trees. Combined with real-time images, water features can be better extracted. In the future, the FSA-UNet model can be applied to other kinds of image fusion to further assist urban planning and construction.

7. Conclusion

Street view and remote sensing images provide feature information such as buildings from different perspectives, which provides a data basis for the large-scale investigation of building vulnerability to flood disasters. The research in this paper shows that the structure, material, number of storeys, and condition contained in street view and remote sensing images can be used to assess the building vulnerability. In addition, the proposed FSA-UNet model can capture the relevant and complementary features among multi-source data by using a multi-source feature fusion self-attention block, edge detection block, and multiple-level features. The proposed model revealed considerable capability for the fusion of street view and remote sensing multi-angle features.

This study proves that the inclusion of street view images can improve the capability of traditional remote sensing-based building semantic segmentation. The method is adapted to building vulnerability assessment of other hazards such as earthquakes and typhoons and also can be generally applied for other segmentation of integrating remote sensing and street view images. Based on the results of the assessment, urban buildings can be planned and constructed to promote the sustainable development of the city. The research in this paper can provide a reference for urban planning and disaster risk assessment management. In practice, many circumstances intervene in disasters that sometimes mean that the real results do not coincide with the predictions, so integrating multiple sources of data like hydrologic and land cover information may provide a more specific estimation of the flood risks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was supported by National Key Research and Development Program of China grant number 2018YFC1508901.

References

Abdollahi, A., Pradhan, B., & Al-amri, A. (2020). An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images. *Geocarto International, 37*, 3355–3370. https://doi.org/10.1080/10106049.2020.1856199

Ahmadlou, M., Gajari, Y. E., & Karimi, M. (2022). Enhanced Classification and Regression Tree (CART) by Genetic Algorithm (GA) and Grid Search (GS) for Flood Susceptibility Mapping and Assessment. *Geocarto International*. https://doi.org/10.1080/10106049.2022.2082550

Amadio, M., Scorzini, A. R., Carisi, F., Essenfelder, A. H., Domeneghetti, A., Mysiak, J., et al. (2019). Testing empirical and synthetic flood damage models: The case of Italy. *Natural Hazards and Earth System Sciences*. https://doi.org/10.5194/NHESS-19-661-2019

Angela, B., Norbert, H., & Jochen, S. (2013). Building extraction from remote sensing data for parameterising a building typology: A contribution to flood vulnerability assessment. *Joint Urban Remote Sensing Event, 2013*, 147–150. https://doi.org/10.1109/JURSE.2013.6550687

Arabameri, A., Rezaei, K., Cerda, A., Lombardo, L., & Rodrigo-Comino, J. (2019). GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *The Science of the total environment, 658*, 160–177. https://doi.org/10.1016/j.scitotenv.2018.12.115

Armenakis, C., Du, E.X., Natesan, S., Persad, R.A., & Zhang, Y. (2017). Flood risk assessment in urban areas based on spatial analytics and social factors.https://doi.org/10.3390/GEOSCIENCES7040123.

Balasbaneh, A. T., Bin Marsono, A. K., & Gohari, A. (2019). Sustainable materials selection based on flood damage assessment for a building using LCA and LCC. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2019.03.005

Barbierato, E., Bernetti, I., Capecchi, I., & Saragosa, C. (2020). Integrating remote sensing and street view images to quantify urban forest ecosystem services. *Remote Sensing, 12*, 329. https://doi.org/10.3390/rs12020329

Bello, R.-. W., Sufril Azlan Mohamed, A., & Zawawi Talib, A. (2021). Enhanced Mask R-CNN for herd segmentation. *International Journal of Agricultural and Biological Engineering*. https://doi.org/10.25165/j.ijabe.20211404.6398

Bin, Z., Hongyong, Y., Quanyi, H., Renqiang, W., & Junqiang, G. (2010). Research on fine spatial quantitative model about vulnerability of hazard-affected bodies. *International Journal of Digital Earth, 3*(4), 395–405. https://doi.org/10.1080/17538947.2010.496497

Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., et al. (2018). Integrating Aerial and Street View Images for Urban Land Use Classification. *Remote Sensing, 10*, 1553. https://doi.org/10.3390/rs10101553

Chaple, G. N., Daruwala, R. D., & Gofane, M. S. (2015). Comparisions of Robert, Prewitt, Sobel operator based edge detection methods for real time uses on FPGA. In *2015 International Conference on Technologies for Sustainable Development (ICTSD)* (pp. 1–4). https://doi.org/10.1109/ICTSD.2015.7095920

Chen, Y., Liu, T., Ge, Y., Xia, S., Yuan, Y., Li, W.-. J., et al. (2021). Examining social vulnerability to flood of affordable housing communities in Nanjing, China: Building long-term disaster resilience of low-income communities. *Sustainable Cities and Society, 102939*. https://doi.org/10.1016/J.SCS.2021.102939

Chen, B., Feng, Q., Niu, B., Yan, F., Gao, B., Yang, J., et al. (2022a). Multi-modal fusion of satellite and street-view images for urban village classification based on a dual-branch deep neural network. *International Journal of Applied Earth Observation and Geoinformation, 109*, Article 102794. https://doi.org/10.1016/j.jag.2022.102794

Chen, Y., Liu, H., Ye, Z., Zhang, H., Jiang, B., & Zhang, Y. (2022b). Social justice in urban–rural flood exposure: A Case Study of Nanjing, China. *Land, 11*(9), 1588. https://doi.org/10.3390/land11091588

Ciurean, R., Hussin, H., Westen, C. J.v., Jaboyedoff, M., Nicolet, P., Chen, L., et al. (2016). Multi-scale debris flow vulnerability assessment and direct loss estimation of buildings in the Eastern Italian Alps. *Natural Hazards, 85*, 929–957. https://doi.org/10.1007/s11069-016-2612-6

Dall'Osso, F., Gonella, M., Gabbianelli, G., Withycombe, G., & Dominey-Howes, D. (2009). A revised (PTVA) model for assessing the vulnerability of buildings to tsunami damage. *Natural Hazards and Earth System Sciences, 9*, 1557–1565. https://doi.org/10.5194/NHESS-9-1557-2009

de Ruiter, M. C., de Bruijn, J. A., Englhardt, J., Daniell, J. E., de Moel, H., & Ward, P. J. (2021). The Asynergies of Structural Disaster Risk Reduction Measures: Comparing Floods and Earthquakes. *Earth's Future, 9*. https://doi.org/10.1029/2020EF001531

Deng, M., Yang, W., Chen, C., Wu, Z. S., Liu, Y., & Xiang, C. (2021). Street-level solar radiation mapping and patterns profiling using Baidu Street View images. *Sustainable Cities and Society, 75*, Article 103289. https://doi.org/10.1016/J.SCS.2021.103289

D'Ayala, D., Wang, K., Yan, Y., Smith, H., Massam, A., Filipova, V., Pereira, J. J., et al. (2020). Flood vulnerability and risk assessment of urban traditional buildings in a heritage district of Kuala Lumpur. *Malaysia. Natural Hazards and Earth System Sciences, 20*(8), 2221–2241. https://doi.org/10.5194/nhess-20-2221-2020

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ArXiv*. https://doi.org/10.1016/j.isprsjprs.2020.01.013. abs/1904.00592.

Diaz, D. H., Qin, S., Ingle, R. R., Fujii, Y., & Bissacco, A. (2021). Rethinking text line recognition models. *ArXiv*. https://doi.org/10.48550/arXiv.2104.07787. abs/2104.07787.

Erena, S. H., & Worku, H. (2019). Urban flood vulnerability assessments: The case of Dire Dawa city, Ethiopia[J]. *Natural Hazards, 97*, 495–516. https://doi.org/10.1007/s11069-019-03654-9

Fan, R., Li, J., Li, F., Han, W., & Wang, L. (2022). Multilevel spatial-channel feature fusion network for urban village classification by fusing satellite and streetview images. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–13. https://doi.org/10.1109/TGRS.2022.3208166

Geiss, C., Geiss, C., Pelizari, P. A., Marconcini, M., Sengara, W., Edwards, M., et al. (2015). Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *Isprs Journal of Photogrammetry and Remote Sensing, 104*, 175–188. https://doi.org/10.1016/J.ISPRSJPRS.2014.07.016

Geiss, C., Taubenböck, H., Tyagunov, S., Tisch, A., Post, J., & Lakes, T. (2014). Assessment of seismic building vulnerability from space. *Earthquake Spectra, 30*, 1553–1583. https://doi.org/10.1193/121812EQS350M

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/cvpr.2016.90

Hoffmann, E. J., Wang, Y., Werner, M., Kang, J., & Zhu, X. (2019). Model fusion for building type classification from aerial and street view images. *Remote Sensing, 11*, 1259. https://doi.org/10.3390/RS11111259

Hu, C., Zhang, F., Gong, F.-. Y., Ratti, C., & Li, X. (2020). Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. *Building and Environment*. https://doi.org/10.1016/j.buildenv.2019.106424

Ilehag, R., Schenk, A., & Hinz, S. (2017). Concept for Classifying Façade Elements Based on Material, Geometry and Thermal Radiation Using Multimodal UAV Remote Sensing. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 145–151. https://doi.org/10.5194/ISPRS-ARCHIVES-XLII-2-W6-145-2017

Iyer, G., Chanussot, J., & Bertozzi, A. (2020). A Graph-Based Approach for Data Fusion and Segmentation of Multimodal Images. *IEEE Transactions on Geoscience and Remote Sensing, 59*, 4419–4429. https://doi.org/10.1109/TGRS.2020.2971395

Jeon, E.-i., Kim, S., Park, S., Kwak, J., & Choi, I. (2021). Semantic segmentation of seagrass habitat from drone imagery based on deep learning: A comparative study. *Ecological Informatics*. https://doi.org/10.1016/j.ecoinf.2021.101430

Jing, W., Lin, J., Lu, H., Chen, G., & Song, H. H. (2022). Learning holistic and discriminative features via an efficient external memory module for building extraction in remote sensing images. *Building and Environment*. https://doi.org/10.1016/j.buildenv.2022.109332

Jongman, B. (2021). The fraction of the global population at risk of floods is growing. *Nature, 596 7870*, 37–38. https://doi.org/10.1038/d41586-021-01974-0

Kanopoulos, N., Vasanthavada, N., & Baker, R. L. (1988). Design of an image edge detection filter using the Sobel operator. *IEEE Journal of Solid-state Circuits, 23*, 358–367. https://doi.org/10.1109/4.996

Komolafe, A. A., Herath, S., & Avtar, R. (2019). Establishment of detailed loss functions for the urban flood risk assessment in Chao Phraya River basin, Thailand. *Geomatics, Natural Hazards and Risk, 10*, 633–650. https://doi.org/10.1080/19475705.2018.1539038

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., et al. (2014). Flood risk and climate change: Global and regional perspectives. *Hydrological Sciences Journal, 59*, 1–28. https://doi.org/10.1080/02626667.2013.857411

Laumer, D., Lang, N., Doorn, N. S.v., Mac Aodha, O., Perona, P., & Wegner, J. D. (2020). Geocoding of trees from street addresses and street-level images. *ArXiv*. https://doi.org/10.1016/j.isprsjprs.2020.02.001. abs/2002.01708.

Li, Z., Wang, L., Shen, J., Ma, Q., & Du, S. (2022a). A Method for Assessing Flood Vulnerability Based on Vulnerability Curves and Online Data of Residential Buildings—A Case Study of Shanghai. *Water*. https://doi.org/10.3390/w14182840

Li, Y., Osei, F. B., Hu, T., et al. (2023). Urban flood susceptibility mapping based on social media data in Chengdu city, China[J]. *Sustainable Cities and Society, 88*, Article 104307. https://doi.org/10.1016/j.scs.2022.104307

Liu, M., Fu, B., Fan, D., Zuo, P., Xie, S., He, H., et al. (2021). Study on transfer learning ability for classifying marsh vegetation with multi-sensor images using DeepLabV3+ and HRNet deep learning algorithms. *International Journal of Applied Earth Observation and Geoinformation, 103*, Article 102531. https://doi.org/10.1016/j.jag.2021.102531

Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Paper presented at the NeurIPS*. https://doi.org/10.48550/arXiv.1908.02265

Man, Q., Dong, P., & Guo, H. (2015). Pixel- and feature-level fusion of hyperspectral and lidar data for urban land-use classification. *International Journal of Remote Sensing, 36*, 1618–1644. https://doi.org/10.1080/01431161.2015.1015657

Marín-García, D., Rubio-Gómez-Torga, J., Pinheiro, M. D., & Moyano, J. J. (2022). Simplified automatic prediction of the level of damage to similar buildings affected by river flood in a specific area. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2022.104251

Miranda, F. N., & Ferreira, T. M. (2019). A simplified approach for flood vulnerability assessment of historic sites. *Natural Hazards*, 1–18. https://doi.org/10.1007/s11069-018-03565-1

Mück, M., Taubenböck, H., Post, J., Wegscheider, S., Strunz, G., Sumaryono, S., et al. (2013). Assessing building vulnerability to earthquake and tsunami hazard using remotely sensed data. *Natural Hazards, 68*, 97–114. https://doi.org/10.1007/s11069-012-0481-1

Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. *PloS one, 10*. https://doi.org/10.1371/journal.pone.0118571

Nex, F., Duarte, D., Steenbeek, A., & Kerle, N. (2019). Towards real-time building damage mapping with low-cost UAV solutions. *Remote Sensing, 11*, 287. https://doi.org/10.3390/rs11030287

Ning, H., Ye, X., Chen, Z., Liu, T., & Cao, T. (2021). Sidewalk extraction using aerial and street view images. *Environment and Planning B: Urban Analytics and City Science, 49*, 7–22. https://doi.org/10.1177/2399808321995817

O'Regan, A. C., Byrne, R., Hellebust, S., & Nyhan, M. M. (2022). Associations between google street view-derived urban greenspace metrics and air pollution measured using a distributed sensor network. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4179128

Ouyang, M., Kotsuki, S., Ito, Y., et al. (2022). Employment of hydraulic model and social media data for flood hazard assessment in an urban city. *Journal of Hydrology: Regional Studies, 44*, Article 101261. https://doi.org/10.1016/j.ejrh.2022.101261

Papathoma-Köhle, M., Schlögl, M., Dosser, L., Roesch, F., Borga, M., Erlicher, M., et al. (2022). Physical vulnerability to dynamic flooding: Vulnerability curves and vulnerability indices. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2022.127501

Papathoma-Köhle, M., Schlögl, M., & Fuchs, S. (2019). Vulnerability indicators for natural hazards: An innovative selection and weighting approach. *Scientific Reports, 9* (1), 1–14. https://doi.org/10.1038/s41598-019-50257-2

Park, K., Choi, S.-. H., & Yu, I. (2021). Risk type analysis of building on urban flood damage. *Water*. https://doi.org/10.3390/w13182505

Pelizari, P. A., Geiss, C., Aguirre, P., María, H. S., Peña, Y. M., & Taubenböck, H. (2021). Automated building characterization for seismic risk assessment using street-level imagery and deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2021.07.004

Pham, Q. B., Ali, S. A., Bielecka, E., Całka, B., Orych, A., Parvin, F., et al. (2022). Flood vulnerability and buildings' flood exposure assessment in a densely urbanised city: Comparative analysis of three scenarios using a neural network approach. *Natural Hazards, 113*, 1043–1081. https://doi.org/10.1007/s11069-022-05336-5

Pittore, M., & Wieland, M. (2012). Toward a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing. *Natural Hazards, 68*, 115–145. https://doi.org/10.1007/s11069-012-0475-z

Polli, D.A., & Dell'acqua, F. (2011). Fusion of optical and SAR data for seismic vulnerability mapping of buildings. https://doi.org/10.1007/978-3-642-14212-3_15.

Rana, I. A., & Routray, J. K. (2016). Actual vis-à-vis perceived risk of flood prone urban communities in Pakistan. *International journal of disaster risk reduction, 19*, 366–378. https://doi.org/10.1016/j.ijdrr.2016.08.028

Ruggieri, S., Cardellicchio, A., Leggieri, V., & Uva, G. (2021). Machine-learning based vulnerability analysis of existing buildings. *Automation in Construction*. https://doi.org/10.1016/j.autcon.2021.103936

Shang, H., Sun, C., Liu, J., Chen, X., & Yan, R. (2022). Deep learning-based borescope image processing for aero-engine blade in-situ damage detection. *Aerospace Science and Technology*. https://doi.org/10.1016/j.ast.2022.107473

Marko, K., Kusratmoko, E., Tambunan, M. P., et al. (2019). A Spatial Approach in Assessing Flood Losses in Floodplain Area of Pesanggrahan River (Case Study on Ulujami and Cipulir Urban Villages, South Jakarta)[C]. *IOP Conference Series: Earth and Environmental Science. IOP Publishing, 338*(1), Article 012030. https://doi.org/10.1088/1755-1315/338/1/012030

Stephenson, V., & D'Ayala, D (2014). A new approach to flood vulnerability assessment for historic buildings in England. *Natural Hazards and Earth System Sciences, 14*, 1035–1048. https://doi.org/10.5194/NHESS-14-1035-2014

Sun, Y., Fu, Z., Sun, C., Hu, Y., & Zhang, S. Z. (2021). Deep Multimodal Fusion Network for Semantic Segmentation Using Remote Sensing Image and LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–18. https://doi.org/10.1109/TGRS.2021.3108352

Song, J., Chang, Z., Li, W., Feng, Z., Wu, J., Cao, Q., et al. (2019). Resilience-vulnerability balance to urban flooding: A case study in a densely populated coastal city in China. *Cities (London, England), 95*, Article 102381. https://doi.org/10.1016/j.cities.2019.06.012

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C., Brakenridge, G. R., et al. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature, 596 7870*, 80–86. https://doi.org/10.1038/s41586-021-03695-w

Timmerman, P. (1981). *Vulnerability, resilience and the collapse of society. a review of models and possible climatic appli-cations*. Toronto, Canada: Institute for Environmental Studies, University of Toronto.

Usman Kaoje, I., Abdul Rahman, M.Z., Idris, N.H., Razak, K.A., Wan Mohd Rani, W.N.M., Tam, T.H., et al. (2021). Physical flood vulnerability assessment using geospatial indicator-based approach and participatory analytical hierarchy process: A Case Study in Kota Bharu, Malaysia. https://doi.org/10.3390/w13131786.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Paper presented at the NIPS*. https://doi.org/10.48550/arXiv.1706.03762

Velez, R., Calderon, D., Carey, L., Aime, C., Hultquist, C., Yetman, G., et al. (2021). Advancing data for street-level flood vulnerability: extraction of variables from google street view in Quito. *Ecuador. ArXiv*. https://doi.org/10.48550/arXiv.2108.05489. abs/2108.05489.

Wei, X., Zhang, T., Li, Y., Zhang, Y., & Wu, F. (2020). Multi-modality cross attention network for image and sentence matching. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10938–10947). https://doi.org/10.1109/CVPR42600.2020.01095

Wu, H., Liang, C., Liu, M.-. S., & Wen, Z. (2021). Optimized HRNet for image semantic segmentation. *Expert System with Application, 174*, Article 114532. https://doi.org/10.1016/j.eswa.2020.114532

Wu, X., Ou, J., Wen, Y., Liu, X., He, J., & Zhang, J. (2022). Developing a data-fusing method for mapping fine-scale urban three-dimensional building structure. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2022.103716

Wu, Z., Shen, Y., Wang, H., & Wu, M. (2020). Urban flood disaster risk evaluation based on ontology and Bayesian Network. *Journal of Hydrology, 583*, Article 124596. https://doi.org/10.1016/j.jhydrol.2020.124596

Zhang, R., Tang, X., You, S., Duan, K., Xiang, H., & Luo, H. (2020). A Novel Feature-Level Fusion Framework Using Optical and SAR Remote Sensing Images for Land Use/Land Cover (LULC) Classification in Cloudy Mountainous Area. *Applied Sciences, 10*, 2928. https://doi.org/10.3390/app10082928

Zhao, Y., Qi, J., & Zhang, R. (2019). CBHE: Corner-based building height estimation for complex street scene images. In *The World Wide Web Conference*. https://doi.org/10.1145/3308558.3313394