sustainability

MDPI

Article

Building Extraction from Unmanned Aerial Vehicle (UAV) Data in a Landslide-Affected Scattered Mountainous Area Based on Res-Unet

Chunhai Tan [1,2], Tao Chen [1,2], Jiayu Liu [1,2], Xin Deng [1,2], Hongfei Wang [1,2] and Junwei Ma [1,2,3,*]

[1] Badong National Observation and Research Station of Geohazards, China University of Geosciences, Wuhan 430074, China; tanch@cug.edu.cn (C.T.); chent@cug.edu.cn (T.C.); liujiayu@cug.edu.cn (J.L.); dengxin@cug.edu.cn (X.D.); whf6812@163.com (H.W.)
[2] Three Gorges Research Center for Geo-Hazards of the Ministry of Education, China University of Geosciences, Wuhan 430074, China
[3] Hubei Key Laboratory of Operation Safety of High Dam and Large Reservoir, Yichang 431000, China
* Correspondence: majw@cug.edu.cn

Abstract: Building extraction in landslide-affected scattered mountainous areas is essential for sustainable development, as it improves disaster risk management, fosters sustainable land use, safeguards the environment, and bolsters socio-economic advancement; however, this process entails considerable challenges. This study proposes a Res-Unet-based model to extract landslide-affected buildings from unmanned aerial vehicle (UAV) data in scattered mountain regions, leveraging the feature extraction capabilities of ResNet and the precise localization abilities of U-Net. A landslide-affected, scattered mountainous region within the Three Gorges Reservoir area was selected as a case study to validate the model's performance. Experimental results indicate that Res-Unet displays high accuracy and robustness in building recognition, attaining accuracy (ACC), intersection-over-union (IOU), and F1-score values of 0.9849, 0.9785, and 0.9892, respectively. This enhancement can be attributed to the combined model, which amalgamates the skip connections, the symmetric architecture of U-Net, and the residual blocks of ResNet. This integration preserves low-level detail during recovery at higher levels, facilitating the extraction of multi-scale features while also mitigating the vanishing gradient problem prevalent in deep network training through the residual block structure, thus enabling the extraction of more complex features. The proposed Res-Unet approach shows significant potential for the accurate recognition and extraction of buildings in complex terrains through the efficient processing of remote sensing images.

Keywords: Res-Unet; building extraction; scattered mountainous area; unmanned aerial vehicle (UAV) data

Citation: Tan, C.; Chen, T.; Liu, J.; Deng, X.; Wang, H.; Ma, J. Building Extraction from Unmanned Aerial Vehicle (UAV) Data in a Landslide-Affected Scattered Mountainous Area Based on Res-Unet. *Sustainability* **2024**, *16*, 9791. https://doi.org/10.3390/su16229791

Academic Editor: Maxim A. Dulebenets

Received: 23 October 2024
Revised: 7 November 2024
Accepted: 7 November 2024
Published: 9 November 2024

1. Introduction

The significance of landslide risk management has markedly increased due to the escalating frequency of landslides. According to statistics from the EM-DAT global catastrophe database [1], landslides have led to over 32 million fatalities, impacted approximately 1 billion individuals, and incurred economic damages estimated at around 3.58 trillion US dollars since the early 1900s. Statistics from the Ministry of Natural Resources, as reported in the National Geological Disaster Bulletin, indicate that from 2005 to 2022, landslides were the most prevalent geohazard in China, comprising 71.5% of all incidents. Building extraction in landslide-affected mountainous regions is essential for sustainable development, as it enhances disaster risk management, promotes sustainable land use, protects the environment, and supports socio-economic development. In alignment with the mandates of Sustainable Development Goal (SDG) 11 [2] for sustainable cities and communities, the production of a building inventory enhances comprehension of the spatial distributions of

structures in mountainous regions. This empowers authorities to perform comprehensive disaster assessments, devise efficient early warning systems, and formulate effective emergency response strategies, thus ensuring the affordability and sustainability of housing and services, ultimately augmenting the community's resilience and safety. Moreover, in accordance with the protection principles of SDG 15 pertaining to terrestrial ecosystems, precise building mapping enables the identification and management of buildings in hilly regions. This directs land use decisions to circumvent high-risk areas, enforces construction restrictions, and enhances infrastructure development, ultimately mitigating damage to natural ecosystems and fostering sustainable land use [3]. This consequently facilitates the attainment of both SDGs 11 and 15 [4]. Moreover, understanding the distribution and characteristics of buildings aids in assessing the potential impacts of landslides, facilitating more efficient emergency response and resource allocation. Ultimately, this enhances the safety and resilience of populations in landslide-prone areas.

While field surveys for building inventories provide comprehensive data, they also present significant drawbacks. Implementing large-scale landslide disaster scenarios necessitates substantial human resources and effort, rendering these surveys labor-intensive and time-consuming to conduct [5,6]. Moreover, carrying out field surveys may expose individuals to potential hazards, particularly in unstable or risky environments. These constraints underscore the urgent need for alternative methodologies, such as remote sensing technology, to effectively acquire data on elements at risk [7,8].

Recently, building extraction from remote sensing data, encompassing both optical and synthetic aperture radar (SAR) data, has garnered considerable attention due to its cost-efficiency and scalability in providing up-to-date information on urban structures. Initially, the development of building extraction techniques heavily relied on heuristic feature design approaches [9,10], which utilized geometric primitives; over-segmentation methods; and classifier-based techniques. These early methods employed geographical, spectral, and auxiliary data to hypothesize building locations [11,12]. However, they often encountered challenges related to scalability and robustness, primarily due to the complexities inherent in feature engineering.

Recent advancements in deep learning (DL) have significantly transformed the field of remote sensing image interpretation [13]. Convolutional Neural Networks (CNNs) have demonstrated superior feature extraction capabilities, outperforming traditional methods in terms of both efficiency and accuracy. For instance, Guo et al. [14] developed an integrated CNN model for extracting buildings in rural areas, achieving high accuracy and efficiency. Additionally, Zhang et al. [15] enhanced the Mask R-CNN architecture by integrating high-resolution remote sensing imagery with advanced DL techniques, introducing the Mask R-CNN fusion Sobel framework. This combination of CNNs and edge detection methods proved more efficient than traditional Mask R-CNN approaches in segmenting and extracting complex building structures.

The limitations of semantic segmentation network encoders in accurately capturing low-level feature representations result in diminished spatial information regarding building features, along with an excess of redundant information that does not effectively communicate the precise spatial details of the buildings. As a result, numerous researchers have implemented substantial enhancements to resolve these issues [16,17]. For example, Hui, et al. [18] proposed an enhanced U-Net model that incorporates the Xception module and a multi-task framework, aiming to extract robust features and improve spatial consistency using high-resolution remote sensing images. Wang and Miao [19] introduced the void space pyramid pool module within the U-Net structure, combining residual learning and pyramid pooling in empty spaces, resulting in improved accuracy and boundary definition for building extraction. Qiu et al. [20] further refined the U-Net model by integrating an optimized skip connection scheme, the void space convolutional pyramid pool module, and enhanced depth separable convolutional modules. Their new network, Refine-U-Net, significantly improved multi-scale and multi-level feature extraction capabilities, thereby enhancing the accuracy of building extraction. Nevertheless, as models have progressively

improved, the boosted depth of CNNs has resulted in challenges related to vanishing and bursting gradients. This study utilizes the U-Net architecture as the foundational framework to tackle the inadequate spatial details in multi-scale building extraction, substituting the U-Net encoder with ResNet to effectively capture multi-scale information and mitigate the issues of vanishing and exploding gradients.

Despite these advancements, extracting buildings in scattered mountainous areas continues to present unique challenges. The available satellite imagery often has relatively low resolution [21], complicating the distinction of small structures from their surroundings; even when conducting building extraction using higher resolution satellite imagery, the recognition accuracy remains limited. For instance, the Mask R-CNN algorithm employed by Raghavan et al. [22] achieved an accuracy of only 0.820, while the average precision of the object detection deep learning framework used by Nurkarim et al. [23] was merely 0.7466. In these regions, buildings, typically small houses, may occupy fewer than 100 pixels, complicating the extraction process. Additionally, the buildings may possess indistinct borders that are difficult to differentiate from the surrounding natural environment, such as agricultural land, meadows, or mountainous terrain [24]. The dataset is often imbalanced, with a notably low proportion of positive samples (buildings) relative to the background [25]. The diverse geographical settings in mountainous areas add to the complexity, while the sparse population density results in structures being widely dispersed rather than clustered, hindering the application of effective strategies in urban environments. To achieve accurate building extraction in such terrains, specialized methodologies and datasets tailored to the unique characteristics of these areas are essential.

A landslide-affected, scattered mountainous region within the Three Gorges Reservoir area was selected as a case study to validate model performance. The proposed models were compared against classical semantic segmentation models, specifically PSP-Net and DeepLabv3, to assess their efficacy in this challenging environment.

2. Materials and Methods

2.1. Study Area

The Three Gorges Reservoir area in China is characterized by complex geological conditions, making it highly susceptible to various geohazards that pose significant challenges for management and safety [26]. Among these, landslides represent one of the most critical threats, driven by steep slopes, rock fractures, and fluctuations in water levels [27]. They can be triggered by heavy rainfall, changes in reservoir water levels, and seismic activity [28,29]. Six typical landslide-affected areas were selected, as depicted in Figure 1. The distribution of buildings in these regions is markedly different from urban settings, characterized by scattered houses and significantly lower densities. This spatial distribution complicates building identification in mountainous and rural terrains, further challenging the application of conventional urban-centric segmentation models.

As one of the largest landslides, with a volume exceeding 10 million cubic meters in the Three Gorges Reservoir area [30], the Fanjiaping landslide was chosen as a case study. Located approximately 56 km northwest of the Three Gorges Reservoir Dam on the southern bank of the Yangtze River (see Figure 1 for the landslide location), it consists of two blocks: the Muyubao landslide and the Tanjiahe landslide. The Fanjiaping landslide has a total volume of over 106 million cubic meters and covers a planar area of approximately 1.96 million square meters. The thickness of the landslip varies between 40 and 139.16 m, with the material comprising loose accumulation layers at the surface and quartz sandstone and sandy conglomerate rocks of the Xiagxi Formation from the Lower Jurassic at the base.

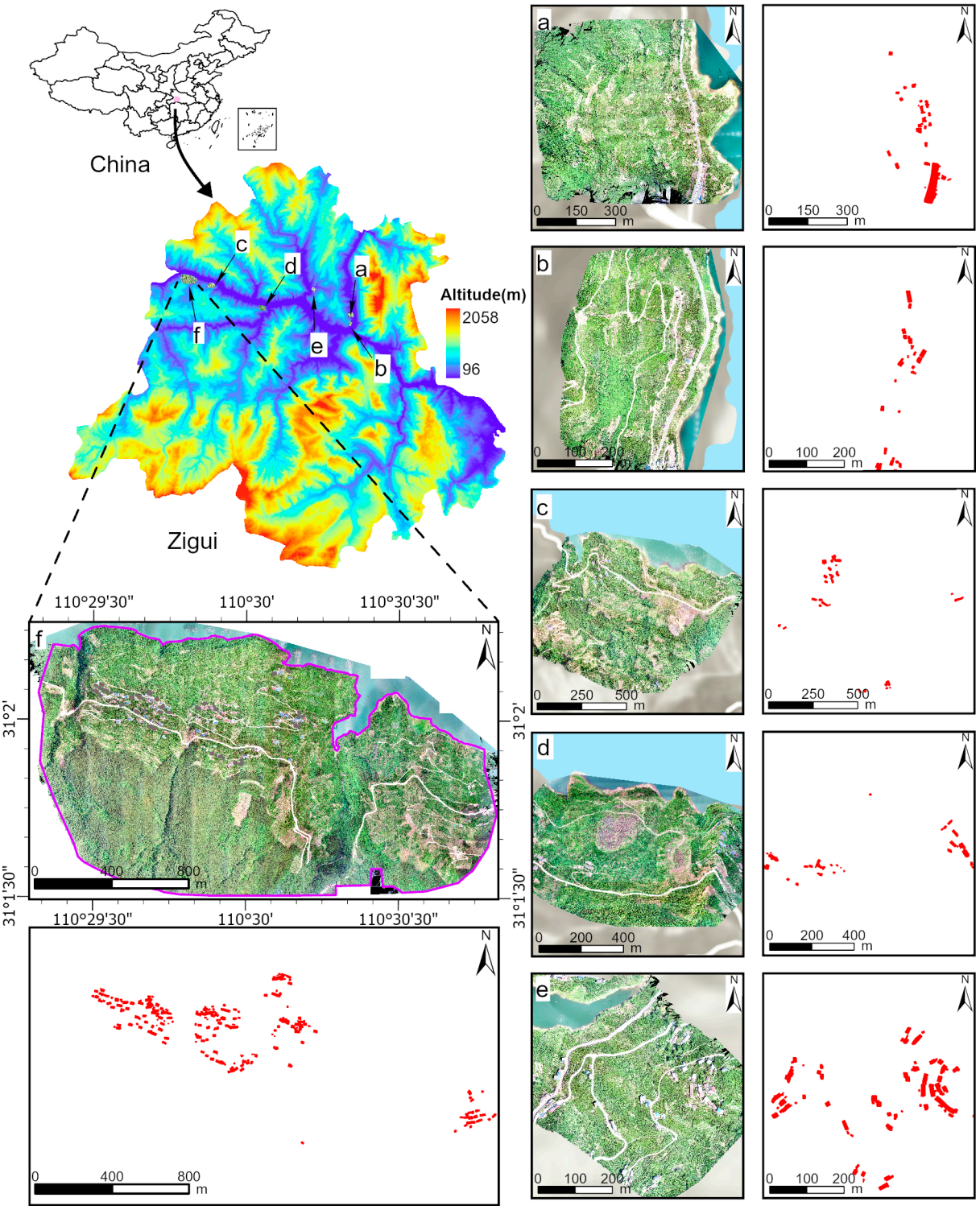

Figure 1. The location and an unmanned aerial vehicle (UAV) image of the typical landslide-affected areas in the Three Gorges Reservoir: (**a**) Baijiabao landslide; (**b**) Bazimen landslide; (**c**) Baishuihe landslide; (**d**) Shuping landslide; (**e**) Majiagou landslide; (**f**) Fanjiaping landslide.

The Fanjiaping landslide is classified as an old landslide that has been reactivated due to the filling of the Three Gorges Reservoir. During the 156 m water storage period, cracks appeared at the rear edge of the Tanjiahe landslide, leading to the relocation of nearby residents. Between June and August 2007, the area experienced continuous heavy rainfall. In September 2007, a crack approximately 30 m long and 20 cm wide, with a scarp height

of 25 cm, was identified at the rear of the landslide at an elevation of 420 m. By April 2012, the eastern boundary of the landslide, from the rear edge at an elevation of 400 m to the middle section near Shahuang Road, exhibited feather-like, intermittently connected cracks extending approximately 200 m. These cracks, with a strike direction of about 30°, had new openings ranging from 1 to 10 cm in width [31]. Small-scale collapses were observed along the crack zone, and the continuous deformations posed significant threats to local residents.

The Fanjiaping landslide is situated in a hilly canyon landscape [32], where the variable topography can result in sudden alterations in lighting conditions. Cloud cover and mountain shadows influence imaging quality, while reflecting objects like concrete surfaces and pebbles can provide intense sunlight reflections, leading to distorted building outlines. This may result in overexposed or underexposed photos, with overexposure leading to detail loss and underexposure obscuring building characteristics, thu complicating recognition. Consequently, the selection of Res-Unet, which has multi-scale feature extraction capabilities, can proficiently tackle these issues, thereby enhancing the accuracy and robustness of building extraction.

2.2. UAV Data Collection and Processing

Orthophoto images of the Fanjiaping landslide were acquired using a DJI Phantom 4 RTK UAV (Shenzhen, China), conducting terrain-following flights at a height of 100 m. The DJI Phantom 4 RTK is a compact multi-rotor drone designed specifically for low-altitude photogrammetry, equipped with a high-definition aerial survey camera and a GNSS positioning system that provides centimeter-level accuracy. The specifications of the DJI Phantom 4 RTK UAV are detailed in Table 1.

Table 1. The specifications of the DJI Phantom 4 RTK UAV.

Camera Sensor	Field of View	Max Image Size	Effective Pixels	Focal Length	Positioning ACC (RTK-Enabled and Functioning Properly)
1″ CMOS	84°	5472 × 3648 (3:2)	20 M	24 mm	Horizontal: ±0.1 m Vertical: ±0.1 m

The surveyed area spanned longitudes from 110°29′20″ to 110°30′50″ east and latitudes from 31°1′30″ to 32°2′15″ north, covering approximately 4 km^2. Aerial photography was conducted on 17 July 2023, under sunny weather conditions. The UAV photogrammetry operated at a relative flying altitude of 100 m, with forward and side overlap rates established at 80%. This study employed Pix4D desktop software (https://www.pix4d.com/) to process the UAV image data, which included initial processing, point cloud densification, and the generation of Digital Surface Models (DSM) and orthophotos. In the initial phase, the images underwent a process to extract specific feature points, known as keypoints. Subsequently, image matching techniques were utilized to identify these keypoints across additional images, with the effectiveness of matching closely tied to the degree of overlap achieved during the flight. The keypoints were then used to construct a sparse 3D point cloud through automatic aerial triangulation and bundle block adjustment. The sparsely distributed 3D point cloud was aligned with geographic coordinates using GPS and inertial measurement unit data from the UAV. By incorporating ground control point data, the model was re-optimized, resulting in the enhanced positional accuracy of the sparse 3D point cloud. In Stage 2, the point cloud was made denser using multi-view stereo algorithm [33]. Then, in Stage 3, the dense 3D point cloud was utilized to create DSM and orthophotos. The resulting orthophoto image boasts a high resolution of 0.36 cm and is shown in Figure 1.

2.3. Methodology

2.3.1. ResNet

ResNet introduced by He et al. [34] is a prominent deep neural network architecture designed to effectively address the vanishing gradient problem encountered in very deep networks. The core innovation of ResNet lies in its use of residual blocks, which incorporate shortcut connections (or skip connections) that bypass one or more layers. This design allows the network to learn residual functions relative to the inputs of those layers. Rather than learning direct mappings, ResNet learns the differences (residuals) between inputs and outputs, facilitating improved training for significantly deep networks [35]. A notable architecture within the ResNet family is ResNet-50, which comprises 50 layers featuring multiple residual blocks that include convolutional layers, batch normalization, and rectified linear unit (ReLU) activations. The skip connections promote efficient gradient flow through these deep layers, enhancing training efficacy and enabling the construction of networks with over a hundred layers. ResNet's architecture has achieved exceptional performance across various computer vision tasks, setting benchmarks in competitions such as ImageNet. It has become a foundational model in deep learning, inspiring numerous subsequent innovations and variants in the field due to its effective balance of depth and training efficiency.

2.3.2. U-Net

U-Net is a CNN architecture specifically designed for biomedical image segmentation, initially proposed by [36]. Named for its U-shaped symmetric structure, U-Net features a contracting path that serves as a mechanism for feature extraction. This path consists of successive convolutional layers with small 3×3 kernels and ReLU activations, followed by max-pooling layers for down-sampling. The expansive path is responsible for up-sampling to reconstruct the image resolution for pixel-wise classification through up-convolution layers. A key aspect of U-Net's architecture is the incorporation of skip connections in the expansive path, which concatenate feature maps from corresponding contracting path layers. This design ensures precise localization by merging low-level and high-level features, thereby capturing fine details essential for accurate segmentation. Notably, U-Net performs well with relatively small training datasets, making it particularly advantageous for medical imaging tasks where annotated data are often limited. It has demonstrated state-of-the-art performance in various segmentation challenges, including cell tracking [37] and brain tumor segmentation [38]. Beyond biomedical applications, U-Net's success has extended to other domains such as satellite imagery analysis [39], road segmentation [40], and object detection [41], showcasing its robustness and versatility in diverse image processing tasks.

2.3.3. Res-Unet

In this study, we proposed a novel approach for building extraction by integrating the strengths of U-Net and ResNet architecture into a fused model called Res-Unet. This integration can be achieved through two primary methods: either replacing all plain blocks of U-Net with residual blocks from ResNet [42,43] or substituting U-Net's encoder structure with a specific ResNet network [44,45]. Residual blocks increase the depth and learning capacity of existing layers, while a ResNet encoder provides a more robust and sophisticated feature representation from the outset. Incorporating a specific ResNet encoder typically demands greater computational power and memory compared to merely adding residual blocks. Additionally, substituting the encoder with ResNet inherently facilitates transfer learning through pretrained weights, a benefit not directly achieved by simply integrating residual blocks into U-Net. We adopted the latter approach, utilizing the ResNet-34 structure to replace the entire encoder of U-Net. This approach entirely substituted the encoder component, enhancing feature extraction efficacy and network representational strength, hence yielding substantial performance enhancements, particularly in intricate tasks and extensive datasets. The proposed Res-Unet model, as illustrated in Figure 2, began with

a 7 × 7 convolutional layer (stride 2) and a 3 × 3 max-pooling layer (stride 2), followed by four stages of residual modules cycling 3, 4, 6, and 3 times, respectively. To maintain consistent input and output feature matrix shapes, the first residual blocks in the 2nd, 3rd, and 4th stages were modified with stride 2 convolutions. Each residual structure incorporated batch normalization and ReLU activation functions, addressing internal covariate shift, accelerating convergence, mitigating gradient vanishing, enhancing robustness to changes in input data distribution, and introducing necessary nonlinearity for complex feature representation. The decoder structure retains U-Net's original design, employing transposed convolutions for up-sampling and successive 3 × 3 convolutions to transform low-level features into higher-level semantic information. This fusion of U-Net and ResNet architecture leverages their respective strengths, potentially offering improved performance in landslide hazard-bearing body identification compared to traditional methods.

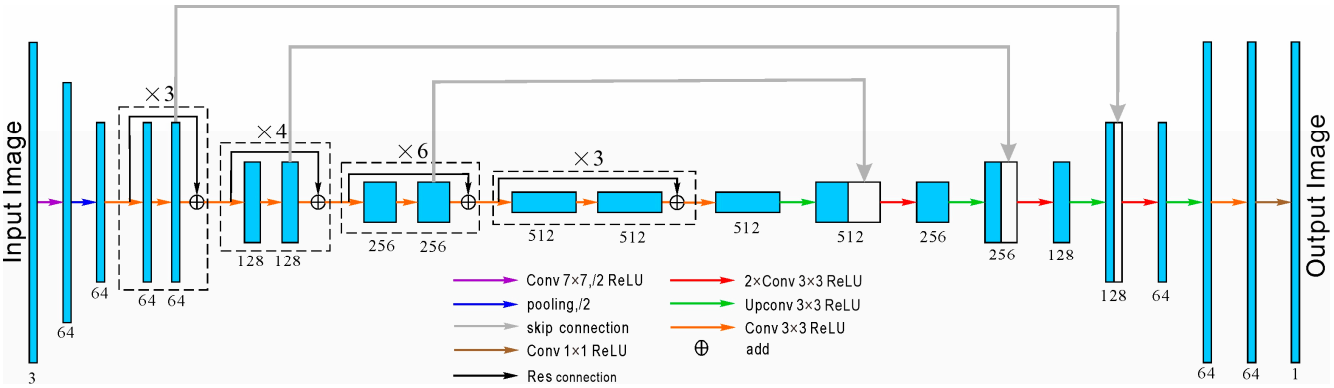

Figure 2. Res-Unet structure.

2.3.4. Res-Unet-Based Building Extraction from UAV Data

In this study, we developed a robust DL model for landslide image classification by implementing a comprehensive data preparation and training process, as illustrated in Figure 3 to create an effective training set. We applied various augmentation techniques to the original orthophoto image, including translations with a 128 × 128 pixel stride, 256 × 256 pixel cropping, and 90-degree rotations, resulting in 1249 images of 256 × 256 pixel resolution. We also digitized sample annotations, yielding a sample collection of 66 samples, and extracted a comprehensive set of features, including pixel values, positions, dimensions, and colors, resulting in 1458 feature samples. To enhance model robustness, 10% of the training data were set aside as a validation set. The model architecture combined the U-Net model with the ResNet network, replacing the U-Net encoder with ResNet residual unit blocks. The training process involved feeding the prepared dataset into the model, computing the loss, and updating model parameters through backpropagation. This iterative process continued until the loss function converged or the desired ACC was achieved. Finally, the trained model was applied to the entire orthophoto of the landslide to generate a classification map, serving as a basis for further analysis and discussion. This structured approach enabled the model to achieve high ACC in classifying landslide images, providing valuable insights for future research and applications in the field.

The Res-Unet model was performed on a Windows 11 Professional operating system, employing a 14th-generation Inter(R) Core (TM) i9-14900K 3.20 GHz and an NVIDIA GeForce RTX 4090 D graphics card with 64 GB of RAM. The batch size was set to 8, with a total of 50 iterations, to optimize model performance and obtain optimal parameters through the system's training process.

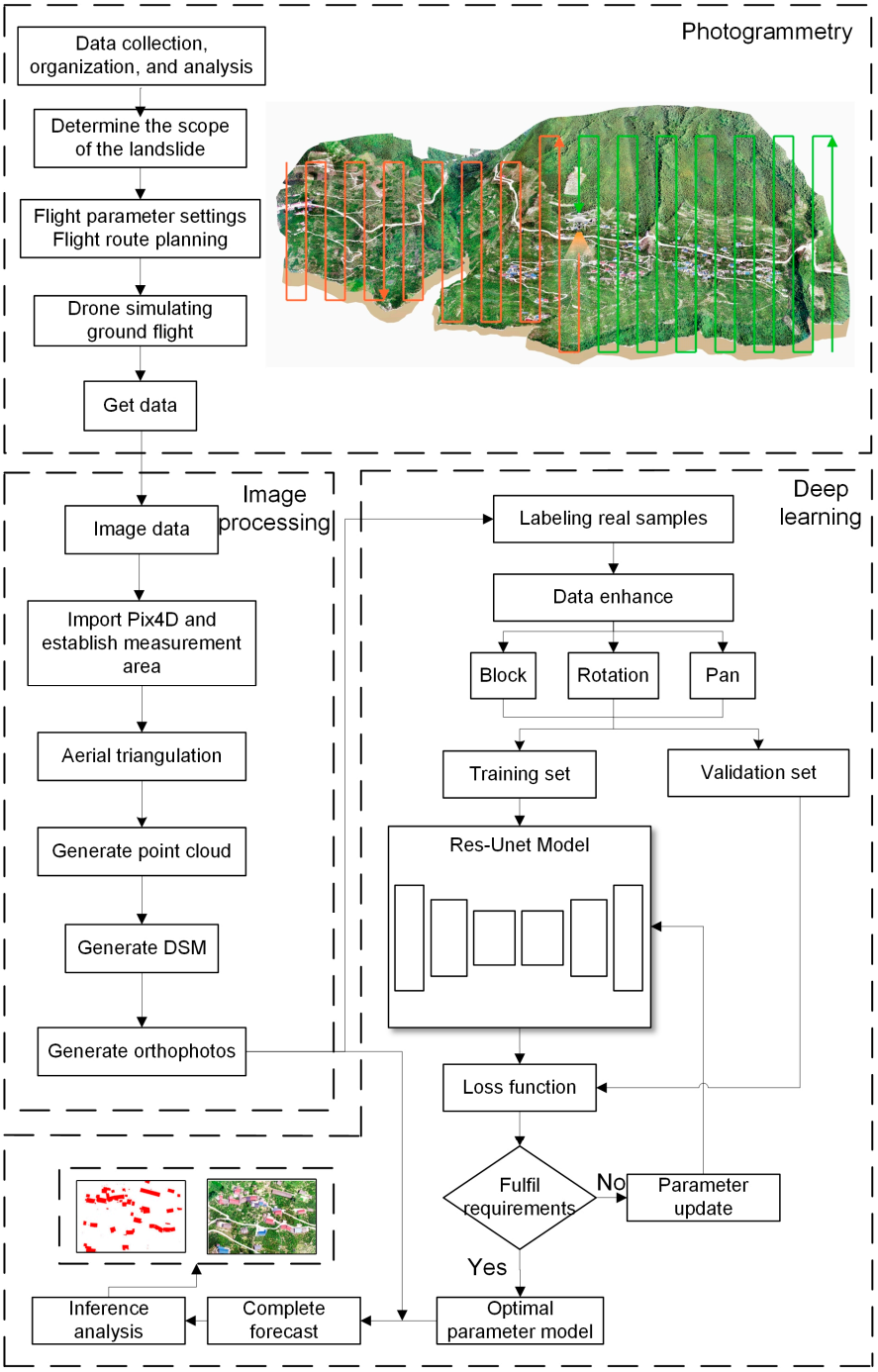

Figure 3. The overall workflow for building extraction from UAV data in a landslide-affected scattered mountainous area based on Res-Unet.

2.4. Loss Function

In this study, we employed Cross-Entropy Loss (CE Loss) and dice loss [46] to evaluate the performance of the Res-Unet-based model for building extraction from UAV data. CE Loss, which represents the distance between two distributions, was utilized to characterize the difference between predicted values and label values. This loss function was particularly effective in quantifying the discrepancy between the model's output and the ground

truth, providing a robust measure of classification ACC. The CE Loss was mathematically expressed as follows:

$$\text{CE Loss} = -\sum_{i=1}^{N} y_i \log p_{y_i'} \tag{1}$$

where y_i and y_i' represent the label value and the predicted value, respectively, and $p_{y_i'}$ represents the probability of the predicted value.

The dice loss coefficient was mathematically equivalent to the intersection-over-union ratio between the predicted result area and the ground truth area. The dice loss offers several advantages in our context: it directly employs the segmentation effect evaluation index as the loss function to supervise the network, and it mitigates the issue of imbalance between positive and negative samples by disregarding a substantial number of background pixels during the intersection-over-union ratio calculation. Consequently, this approach facilitates rapid convergence, making it particularly suitable for our landslide image classification task. The dice loss is mathematically expressed as follows:

$$\text{Dice Loss} = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{2}$$

where p_i represents the predicted probability of pixel i belonging to the foreground class, g_i represents the ground truth label of pixel i (1 for foreground, 0 for background), and N represents the total number of pixels in the image.

2.5. Validation Metrics

In the context of building extraction from UAV data, we employed a comprehensive set of evaluation metrics to assess our model's performance. These metrics are widely used to evaluate the effects of researchers' own semantic segmentation models [47–50]. Among them, precision quantifies the ratio of genuine positive samples to all predicted positives, indicating the model's accuracy in identifying the positive class. Recall denotes the ratio of true positives accurately detected, demonstrating the model's capacity to recognize pertinent events. Accuracy is the proportion of right predictions to the total instances, offering a comprehensive evaluation; yet, it may be deceptive in situations of class imbalance. The F1-score is the harmonic mean of precision and recall, effectively balancing the significance of both measurements. Intersection over union (IOU) evaluates the precision of segmentation tasks by measuring the intersection between the predicted and actual regions. Collectively, these indicators provide a thorough foundation for assessing model performance.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{7}$$

where true positives (TP) represent pixels correctly identified as buildings, true negatives (TN) are pixels accurately classified as non-building areas, false positives (FP) occur when non-building pixels are erroneously identified as buildings, and false negatives (FN) represent building pixels incorrectly classified as non-building areas.

3. Results and Discussions

In this study, we conducted a comparative analysis of three prominent image segmentation models: our proposed Res-Unet model, Deeplabv3 [51], and PSP-Net [52]. The selection of DeepLabv3 and PSP-Net as baseline models is attributed to their exceptional efficacy in semantic segmentation tasks, sophisticated network designs, and multi-scale feature extraction abilities. These models have attained superior performance on multiple benchmark datasets, including Pascal VOC and Cityscapes, and are extensively utilized and esteemed in numerous research publications. The experiment utilized orthophoto data from the Fanjiaping landslide (Figure 1f) in the Three Gorges Reservoir area as the training input. As evidenced by Table 2, the Res-Unet model consistently outperformed its counterparts across all evaluation metrics. Notably, Res-Unet achieved superior F1-score and IOU score values of 0.9892 and 0.9785, respectively, demonstrating its exceptional ACC and coverage capabilities. The higher F1-score indicates Res-Unet's ability to effectively balance precision and recall, while the elevated IOU score suggests a more precise delineation of target areas. Furthermore, Res-U-Net's improved recall rate signifies a reduced incidence of missed extractions, a critical factor in comprehensive target area identification.

Table 2. Performance comparison of Res-Unet, Deeplabv3, and PSP-Net for building extraction from UAV data. The best values are highlighted.

Model	Precision	Recall	ACC	F1	IOU
Deeplabv3	0.9854	0.9808	0.9760	0.9831	0.9668
PSP-Net	0.9809	0.9680	0.9643	0.9744	0.9500
Res-Unet	**0.9903**	**0.9881**	**0.9849**	**0.9892**	**0.9785**

Figure 4 presents the loss function and ACC index across iterations for three models, clearly demonstrating the superior performance of Res-Unet in terms of model optimization, convergence speed, and feature extraction capabilities. The performance of PSP-Net exhibits significant fluctuations during the iterative process, with its ACC and loss on the validation set failing to show a continuous improvement trend. This indicates that while PSP-Net undergoes continuous optimization and approaches an optimal solution, its parameter adjustment and training process are less stable. In contrast, Res-Unet achieves a lower final loss value compared to Deeplabv3, highlighting its efficient feature extraction that not only accelerates the learning process but also enables the model to quickly identify and utilize effective features, thereby enhancing overall performance. This streamlined training process substantially reduces computational resource requirements and time costs, underscoring Res-Unet's considerable potential for practical applications in various domains requiring efficient and accurate image segmentation.

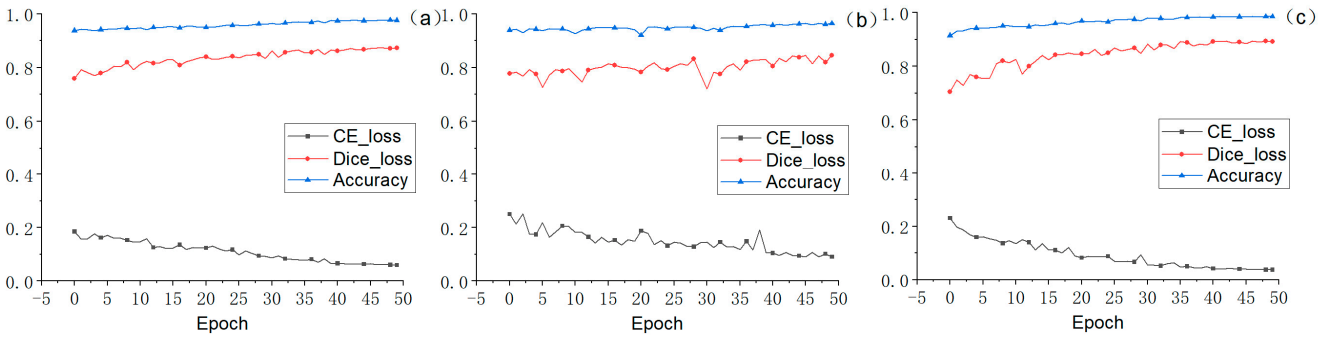

Figure 4. CE Loss, dice loss, and ACC for Deeplabv3 (**a**), PSP-Net (**b**), and Res-Unet (**c**) during the the iteration process.

Figure 5 presents a qualitative comparison of three models, namely Deeplabv3, PSP-Net, and Res-Unet, applied to typical drone images of rural areas. The analysis reveals distinct performance characteristics for each model. Deeplabv3 exhibits limitations in processing semantic information of rural buildings, evidenced by the misclassification of roads and building-like structures as houses. While PSP-Net avoids misidentifying roads, it struggles with accurate boundary delineation, often classifying ground adjacent to houses as part of the structures. In contrast, Res-Unet demonstrates superior ACC in identifying disaster-prone objects and successfully differentiating between adjacent buildings. The figure highlights the challenges posed by rural environments, where disaster-prone objects often share similar colors and materials with the surrounding landscape, and lighting conditions can alter the appearances of roads and other features. In these complex scenarios, Res-Unet showcases its advanced capabilities, leveraging its unique architecture that combines residual network and U-Net structures. This design enables Res-Unet to effectively utilize both global and local image information, resulting in more precise segmentation. Furthermore, Res-Unet's training process facilitates learning features across diverse scenarios, enhancing its generalizability and ability to handle various complex terrains and lighting conditions. This comprehensive performance underscores Res-Unet's potential as a robust tool for accurate object identification in challenging rural and disaster-prone environments.

Figure 5. Extraction results of five typical images (**a**–**e**) of scattered mountain buildings affected by landslides obtained by different models.

In this study, we also analyze building extraction results from previous research for building extraction in scattered mountainous areas. As shown in Table 3, we illustrate the challenges associated with building extraction in these regions compared to urban areas. Most of the recognition accuracies in rural studies are relatively low. This is primarily due to the sparse and irregular distribution of buildings, complex and diverse terrain types, and high vegetation coverage in rural areas, which collectively increase the difficulty of building extraction. Notably, compared to the previous best model, the IOU score of our model is 4.37% higher, and the F1-score is approximately 13.31% higher. These improvements underscore the importance of enhancing the precision of building extraction in rural areas.

Table 3. A performance comparison of the method proposed in the current study with those reported in previous works for building extraction in scattered mountainous areas.

References	Model	IOU (%)	F1-Score (%)
Deng, et al. [53]	VGG-16 + U-Net	81.79%	/
Li, et al. [54]	Attention-enhanced U-Net	74.85%	85.61%
Wang, et al. [55]	ResNet152 + Mask R-CNN	63.6%	77.7%
Xue, et al. [56]	Dilated convolution + pyramid representation + VGG16	93.48%	/

The superior performance of Res-Unet in building extraction from UAV data can be attributed to its innovative synergistic architecture, which combines the strengths of ResNet and U-Net. This unique design facilitates efficient feature extraction and precise segmentation, crucial for processing complex UAV imagery. The residual connections in Res-Unet enable better gradient flow throughout the network, mitigating the vanishing gradient problem common in deep networks and allowing for more effective training of deeper architectures. The U-Net component enables multi-scale feature representation, capturing both local and global contextual information, which is particularly beneficial for building extraction. Enhanced information preservation through skip connections ensures accurate boundary delineation, while the residual learning framework promotes efficient learning of hierarchical features. The model's robustness to input variations, optimal depth without performance degradation, and efficient gradient flow during backpropagation contribute to its effectiveness in handling the complex and varied nature of UAV data. Furthermore, Res-Unet strikes a balance between capturing local details and global context, making it particularly well suited for UAV imagery analysis. Its adaptability to diverse scenarios, including varying building styles, urban densities, and environmental contexts, further enhances its utility. These fundamental characteristics collectively enable Res-Unet to effectively address the complexities and challenges associated with building extraction from UAV data, resulting in its superior performance compared to traditional models.

The Res-Unet model demonstrates significant advantages in building extraction from UAV data, consistently outperforming traditional models such as Deeplabv3 and PSP-Net across various evaluation metrics. Its unique architecture, combining residual networks and U-Net structures, enables rapid convergence, superior feature extraction, and accurate boundary delineation, even in complex rural environments. The model exhibits remarkable generalizability, effectively handling diverse terrains and lighting conditions while minimizing misclassifications. Res-Unet's efficient information flow, facilitated by residual connections, mitigates the vanishing gradient problem, leading to stable and efficient training.

Nonetheless, these benefits include specific trade-offs. The model's complexity may need greater computer resources for training and inference, and its performance is significantly influenced by the quality and amount of the available training data. Overfitting risks arise, especially with constrained datasets, and the model's complexity complicates result interpretation. Moreover, Res-Unet may exhibit sensitivity to hyperparameter optimization and could transmit mistakes across its encoder–decoder architecture.

The technique developed in this study can be used for other domains, including UAV imagery for landslide crack segmentation and the detection and segmentation of landslides utilizing satellite images.

4. Conclusions and Further Work

This study has demonstrated the significant contribution of building extraction in landslide-affected dispersed mountainous areas to sustainable development, particularly through enhanced disaster risk management, promoting sustainable land use, protecting the environment, and supporting socio-economic development. The proposed Res-Unet model, applied to the example of the Fanjiaping landslide in the Three Gorges Reservoir area, has shown superior performance in extracting buildings from UAV data in landslide-affected regions. Compared to traditional models like DeepLabV3 and PSP-Net, Res-Unet achieved higher F1-score (0.9892) and IOU score (0.9785) values, along with better convergence speed and feature extraction capability. These results highlight the effectiveness of Res-Unet in handling complex rural environments, where accurate building identification is critical for disaster risk assessment.

The potential application of the Res-Unet model to several types of remote sensing data presents promising pathways for further exploration. For instance, the segmentation of landslide cracks from UAV images and the detection and segmentation of landslides using satellite imagery could significantly improve the model's applicability and practicality. Future research should investigate the model's adaptability to various geographical regions and its capacity to generalize across numerous environmental situations. Furthermore, integrating additional different data sources, such as temporal UAV imaging or multi-sensor fusion, could enhance the model's robustness and accuracy in dynamic and developing environments.

Author Contributions: Conceptualization, C.T. and T.C.; methodology, C.T.; software, J.L.; validation, C.T., T.C. and X.D.; formal analysis, C.T.; investigation, H.W.; resources, X.D.; data curation, J.L.; writing—original draft preparation, C.T.; writing—review and editing, C.T.; visualization, C.T.; supervision, J.M.; project administration, J.M.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC) (42177147), the NSFC_Yangtze River Water Science Joint Fund Project (U2340230), China Yangtze Power Co., Ltd. (Z532302036), the Key Research and Development Project of the Hubei Provincial Technology Innovation Plan (2023BCB117), the Foundation for Innovative Research Groups of Hubei Province of China (2024AFA015), and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (CUG2642022006). We very much appreciate their support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of this study or in the collection, analysis, or interpretation of data.

References

1. Si, R.; Wen, J.; Yin, Z. Emergency events database (EM-DAT) and its applications. *Sci. Technol. Rev.* **2007**, *25*, 60–67.
2. Pereira, S.; Garcia, R.A.; Zêzere, J.L.; Oliveira, S.C.; Silva, M. Landslide quantitative risk analysis of buildings at the municipal scale based on a rainfall triggering scenario. *Geomat. Nat. Hazards Risk* **2017**, *8*, 624–648. [CrossRef]
3. Qiu, H.; Nie, W.; Zhou, L.; Wei, Y.; Wang, J. Regional Emigration—China's New Approach to Geo-Disaster Mitigation. *J. Earth Sci.* **2024**, *35*, 1786–1788. [CrossRef]
4. Li, Q.; Zhou, Z.; Huang, D.; Xiao, D.; Yan, L.; Huang, Y.; Zhang, Y. Extraction of Complex Buildings in the Karst Mountains Based on U-Net Deep Learning Model. In Proceedings of the 2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS), Zhoushan, China, 22–24 April 2022; pp. 127–138.

5. Dutta, D.; Serker, N. Urban building inventory development using very high-resolution remote sensing data for urban risk analysis. *Int. J. Geoinf.* **2005**, *1*, 109–116.
6. Matsuoka, M.; Mito, S.; Midorikawa, S.; Miura, H.; Quiroz, L.G.; Maruyama, Y.; Estrada, M. Development of building inventory data and earthquake damage estimation in Lima, Peru for future earthquakes. *J. Disaster Res.* **2014**, *9*, 1032–1041. [CrossRef]
7. Yamazaki, F.; Matsuoka, M. Remote sensing technologies in post-disaster damage assessment. *J. Earthq. Tsunami* **2007**, *1*, 193–210. [CrossRef]
8. Moselhi, O.; Bardareh, H.; Zhu, Z. Automated data acquisition in construction with remote sensing technologies. *Appl. Sci.* **2020**, *10*, 2846. [CrossRef]
9. Fua, P.; Hanson, A.J. An optimization framework for feature extraction. *Mach. Vis. Appl.* **1991**, *4*, 59–87. [CrossRef]
10. Li, Q.; Mou, L.; Sun, Y.; Hua, Y.; Shi, Y.; Zhu, X.X. A Review of Building Extraction from Remote Sensing Imagery: Geometrical Structures and Semantic Attributes. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–15. [CrossRef]
11. Wang, J.; Yang, X.; Qin, X.; Ye, X.; Qin, Q. An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 487–491. [CrossRef]
12. Gu, L.; Cao, Q.; Ren, R. Building extraction method based on the spectral index for high-resolution remote sensing images over urban areas. *J. Appl. Remote Sens.* **2018**, *12*, 045501. [CrossRef]
13. Luo, L.; Li, P.; Yan, X. Deep learning-based building extraction from remote sensing images: A comprehensive review. *Energies* **2021**, *14*, 7982. [CrossRef]
14. Guo, Z.; Chen, Q.; Wu, G.; Xu, Y.; Shibasaki, R.; Shao, X. Village building identification based on ensemble convolutional neural networks. *Sensors* **2017**, *17*, 2487. [CrossRef] [PubMed]
15. Zhang, L.; Wu, J.; Fan, Y.; Gao, H.; Shao, Y. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN. *Sensors* **2020**, *20*, 1465. [CrossRef]
16. Abdollahi, A.; Pradhan, B.; Alamri, A.M. An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images. *Geocarto Int.* **2022**, *37*, 3355–3370. [CrossRef]
17. Alsabhan, W.; Alotaiby, T. Automatic building extraction on satellite images using Unet and ResNet50. *Comput. Intell. Neurosci.* **2022**, *2022*, 5008854. [CrossRef]
18. Hui, J.; Du, M.; Ye, X.; Qin, Q.; Sui, J. Effective building extraction from high-resolution remote sensing images with multitask driven deep neural network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 786–790. [CrossRef]
19. Wang, H.; Miao, F. Building extraction from remote sensing images using deep residual U-Net. *Eur. J. Remote Sens.* **2022**, *55*, 71–85. [CrossRef]
20. Qiu, W.; Gu, L.; Gao, F.; Jiang, T. Building extraction from very high-resolution remote sensing images using refine-UNet. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
21. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699.
22. Raghavan, R.; Verma, D.C.; Pandey, D.; Anand, R.; Pandey, B.K.; Singh, H. Optimized building extraction from high-resolution satellite imagery using deep learning. *Multimed. Tools Appl.* **2022**, *81*, 42309–42323. [CrossRef]
23. Nurkarim, W.; Wijayanto, A.W. Building footprint extraction and counting on very high-resolution satellite imagery using object detection deep learning framework. *Earth Sci. Inform.* **2023**, *16*, 515–532. [CrossRef]
24. Li, X.; Zhou, G.; Zhou, L.; Lv, X.; Li, X.; He, X.; Tian, Z. A New Technique for Urban and Rural Settlement Boundary Extraction Based on Spectral–Topographic–Radar Polarization Features and Its Application in Xining, China. *Remote Sens.* **2024**, *16*, 1091. [CrossRef]
25. Chen, T.-H.K.; Pandey, B.; Seto, K.C. Detecting subpixel human settlements in mountains using deep learning: A case of the Hindu Kush Himalaya 1990–2020. *Remote Sens. Environ.* **2023**, *294*, 113625. [CrossRef]
26. Ma, J.; Lei, D.; Ren, Z.; Tan, C.; Xia, D.; Guo, H. Automated machine learning-based landslide susceptibility mapping for the three gorges reservoir area, China. *Math. Geosci.* **2024**, *56*, 975–1010. [CrossRef]
27. Lei, D.; Ma, J.; Zhang, G.; Wang, Y.; Deng, X.; Liu, J. Bayesian ensemble learning and Shapley additive explanations for fast estimation of slope stability with a physics-informed database. *Nat. Hazards* **2024**, 1–30. [CrossRef]
28. Yin, Y.; Huang, B.; Wang, W.; Wei, Y.; Ma, X.; Ma, F.; Zhao, C. Reservoir-induced landslides and risk control in Three Gorges Project on Yangtze River, China. *J. Rock Mech. Geotech. Eng.* **2016**, *8*, 577–595. [CrossRef]
29. Gao, D.; Li, K.; Cai, Y.; Wen, T. Landslide Displacement Prediction Based on Time Series and PSO-BP Model in Three Georges Reservoir, China. *J. Earth Sci.* **2024**, *35*, 1079–1082. [CrossRef]
30. Wang, Y.; Bai, Z.; Lin, Y.; Li, Y.; Shen, W. Sentinel-1 Quasi-PS InSAR for identification and monitoring of landslide deformation. In Proceedings of the IET International Radar Conference (IET IRC 2020), Online, 4–6 November 2020; pp. 946–949.
31. Deng, M.; Huang, X.; Yi, Q.; Liu, Y.; Yi, W.; Huang, H. Fifteen-year professional monitoring and deformation mechanism analysis of a large ancient landslide in the Three Gorges Reservoir Area, China. *Bull. Eng. Geol. Environ.* **2023**, *82*, 243. [CrossRef]
32. Ren, Z.; Ma, J.; Liu, J.; Deng, X.; Zhang, G.; Guo, H. Enhancing deep learning-based landslide detection from open satellite imagery via multisource data fusion of spectral, textural, and topographical features: A case study of old landslide detection in the Three Gorges Reservoir Area (TGRA). *Geocarto Int.* **2024**, *39*, 2421224. [CrossRef]
33. Wang, Y.; Deng, N.; Xin, B.; Wang, W.; Xing, W.; Lu, S. A novel three-dimensional surface reconstruction method for the complex fabrics based on the MVS. *Opt. Laser Technol.* **2020**, *131*, 106415. [CrossRef]

34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
35. Shafiq, M.; Gu, Z. Deep residual learning for image recognition: A survey. *Appl. Sci.* **2022**, *12*, 8972. [CrossRef]
36. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Proceedings, Part III, Munich, Germany, 5–9 October 2015; pp. 234–241.
37. Zhou, Z.; Wang, F.; Xi, W.; Chen, H.; Gao, P.; He, C. Joint multi-frame detection and segmentation for multi-cell tracking. In Proceedings of the Image and Graphics: 10th International Conference, ICIG 2019, Proceedings, Part II, Beijing, China, 23–25 August 2019; pp. 435–446.
38. Chen, W.; Liu, B.; Peng, S.; Sun, J.; Qiao, X. S3D-UNet: Separable 3D U-Net for brain tumor segmentation. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Revised Selected Papers, Part II, Granada, Spain, 16 September 2018; pp. 358–368.
39. Singh, N.J.; Nongmeikapam, K. Semantic segmentation of satellite images using deep-unet. *Arab. J. Sci. Eng.* **2023**, *48*, 1193–1205. [CrossRef]
40. Abderrahim, N.Y.Q.; Abderrahim, S.; Rida, A. Road segmentation using u-net architecture. In Proceedings of the 2020 IEEE International Conference of Moroccan Geomatics (Morgeo), Casablanca, Morocco, 11–13 May 2020; pp. 1–4.
41. Karki, S.; Kulkarni, S. Ship detection and segmentation using unet. In Proceedings of the 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 19–20 February 2021; pp. 1–7.
42. Cao, K.; Zhang, X. An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sens.* **2020**, *12*, 1128. [CrossRef]
43. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]
44. Buslaev, A.; Seferbekov, S.; Iglovikov, V.; Shvets, A. Fully convolutional network for automatic road extraction from satellite imagery. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 207–210.
45. Zhang, H.; Hong, X.; Zhou, S.; Wang, Q. Infrared image segmentation for photovoltaic panels based on Res-UNet. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Xi'an, China, 8–11 November 2019; pp. 611–622.
46. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 27–29 October 2020; pp. 1–7.
47. Wang, H.; Liu, J.; Zeng, S.; Xiao, K.; Yang, D.; Yao, G.; Yang, R. A novel landslide identification method for multi-scale and complex background region based on multi-model fusion: YOLO+ U-Net. *Landslides* **2024**, *21*, 901–917. [CrossRef]
48. Hou, H.; Chen, M.; Tie, Y.; Li, W. A universal landslide detection method in optical remote sensing images based on improved YOLOX. *Remote Sens.* **2022**, *14*, 4939. [CrossRef]
49. Nava, L.; Bhuyan, K.; Meena, S.R.; Monserrat, O.; Catani, F. Rapid mapping of landslides on SAR data by attention U-Net. *Remote Sens.* **2022**, *14*, 1449. [CrossRef]
50. Shi, W.; Zhang, M.; Ke, H.; Fang, X.; Zhan, Z.; Chen, S. Landslide recognition by deep convolutional neural network and change detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4654–4672. [CrossRef]
51. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
52. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
53. Deng, S.; Wu, S.; Bian, A.; Zhang, J.; Di, B.; Nienkötter, A.; Deng, T.; Feng, T. Scattered mountainous area building extraction from an open satellite imagery dataset. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
54. Li, C.; Fu, L.; Zhu, Q.; Zhu, J.; Fang, Z.; Xie, Y.; Guo, Y.; Gong, Y. Attention enhanced u-net for building extraction from farmland based on google and worldview-2 remote sensing images. *Remote Sens.* **2021**, *13*, 4411. [CrossRef]
55. Wang, Y.; Li, S.; Teng, F.; Lin, Y.; Wang, M.; Cai, H. Improved mask R-CNN for rural building roof type recognition from uav high-resolution images: A case study in hunan province, China. *Remote Sens.* **2022**, *14*, 265. [CrossRef]
56. Wang, X.; Liang, K.; Sui, L.; Zhong, M.; Zhu, J. Rural buildings extraction based on deep learning model with dilated convolution and pyramid representation. *Bull. Surv. Mapp.* **2022**, 61–65.