applied sciences

MDPI

Article

Advancing Rural Building Extraction via Diverse Dataset Construction and Model Innovation with Attention and Context Learning

Mingyang Yu *, Fangliang Zhou, Haiqing Xu and Shuai Xu

School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 250101, China; 2021165123@stu.sdjzu.edu.cn (F.Z.); 2021160105@stu.sdjzu.edu.cn (H.X.); 2021165102@stu.sdjzu.edu.cn (S.X.)
* Correspondence: ymy@sdjzu.edu.cn

Abstract: Rural building automatic extraction technology is of great significance for rural planning and disaster assessment; however, existing methods face the dilemma of scarce sample data and large regional differences in rural buildings. To solve this problem, this study constructed an image dataset of typical Chinese rural buildings, including nine typical geographical regions, such as the Northeast and North China Plains. Additionally, an improved remote sensing image rural building extraction network called AGSC-Net was designed. Based on an encoder–decoder structure, the model integrates multiple attention gate (AG) modules and a context collaboration network (CC-Net). The AG modules realize focused expression of building-related features through feature selection. The CC-Net module models the global dependency between different building instances, providing complementary localization and scale information to the decoder. By embedding AG and CC-Net modules between the encoder and decoder, the model can capture multiscale semantic information on building features. Experiments show that, compared with other models, AGSC-Net achieved the best quantitative metrics on two rural building datasets, verifying the accuracy of the extraction results. This study provides an effective example for automatic extraction in complex rural scenes and lays the foundation for related monitoring and planning applications.

Keywords: rural building extraction; deep learning; diverse dataset; attention mechanisms; context collaboration networks; remote sensing

check for updates

Citation: Yu, M.; Zhou, F.; Xu, H.; Xu, S. Advancing Rural Building Extraction via Diverse Dataset Construction and Model Innovation with Attention and Context Learning. *Appl. Sci.* **2023**, *13*, 13149. https://doi.org/10.3390/app132413149

Academic Editor: Igal M. Shohet

Received: 3 November 2023
Revised: 22 November 2023
Accepted: 8 December 2023
Published: 11 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Accurate extraction of urban buildings is significant for urban planning, disaster assessment, building area estimation, 3D urban modeling, and so on [1–4]. Compared with urban buildings, rural buildings have their own characteristics: rural buildings have smaller scales, lower floors, scattered distributions in villages, and higher dispersion degrees; second, building materials, designs and construction times vary greatly, leading to larger internal differences [5]. Therefore, extracting rural buildings is more challenging, and related research is relatively scarce. In addition, the high cloud cover in rural areas increases the extraction difficulty [6]. Accurately identifying rural building roof types is of great importance for rural revitalization, environmental planning, energy assessment and disaster management [7–10]. Most previous rural building identification studies have focused on some local areas without considering the sparse distribution and differential features of buildings across regions, resulting in limited generalization capabilities [11]. Therefore, it is imperative to develop rural building extraction methods applicable to large areas.

Traditional manual measurement methods conduct field surveys on buildings, which are inefficient and costly. The complex rural topography increases the workload, making it difficult to meet modern requirements [12]. In recent years, the spatial resolution of high-resolution remote sensing images has significantly improved to submeter levels, providing richer texture and contour details [13]. Therefore, an increasing number of scholars use

high-resolution remote sensing images to extract building information [14]. Traditional simple classifiers mostly rely on spectral and texture features to extract buildings from remote sensing images. Faced with increasingly complex high-resolution remote sensing images with abundant information and details, traditional extraction methods are no longer applicable [15]. Object-oriented and machine learning methods are widely used to identify features, such as buildings, in complex images [16]. Chen and colleagues used classifiers such as AdaBoost, Random Forest and support vector machine (SVM) to identify buildings from remote sensing images by segmenting images and removing shadows and vegetation. The newly introduced features of the edge regularity index (ERI) and shadow line index (SLI) significantly improved the accuracy [17]. Related studies also adopt the morphological building index (MBI) to detect potential building areas and use an SVM for training and self-correction to remove falsely detected buildings [18]. However, object-oriented methods rely on image segmentation results. Machine learning methods such as SVMs require large sample sets, and their simple structures cannot acquire in-depth information on building details [19], leading to poor extraction effects on rural buildings with large internal differences.

Deep learning technologies based on big data and computer technology have been widely used in image recognition fields and are particularly suitable for analyzing high-resolution remote sensing images. Typical convolutional neural networks (CNNs), such as FCNs, UNet, and VGG16, can autonomously learn complex spectral and texture features to achieve automatic extraction of buildings, roads and other targets [20]. Many scholars have improved classic semantic segmentation network frameworks to adapt to the feature extraction of different land cover types. Yu et al. [21] integrated attention mechanism modules in the building extraction model AGs-UNet. By optimizing feature selection and focusing on small-scale building feature expression, the quality of building contour extraction in high-resolution remote sensing images has been significantly improved. Transfer learning is an effective method to overcome overfitting in neural network training with small data volumes. Smail et al. [22] adopted the idea of transfer learning, combining the UNet basic architecture with ResNet101 and ResNet152 weights. By applying the feature parameters trained on large datasets to new classification tasks, the efficiency and accuracy of small data classification problems improved, reducing data requirements and improving the building extraction capabilities of basic models on remote sensing images. Gong et al. [23] designed context collaboration networks using self-attention mechanisms to capture long-range dependency between buildings, providing complementary semantic and positional information. The above studies show that introducing attention and other feature learning modules can effectively reduce category confusion in semantic segmentation networks and enhance the recognition capability of complex building types.

Although improved deep learning methods have achieved good performance in urban building recognition, their applications in rural scenes are still limited [24]. Difficulties exist in recognizing complex and diverse rural building types. Additionally, the high cloud cover in rural areas affects satellite image quality [6]. To solve this problem, low-altitude remote sensing technologies have advantages, such as low cost and high-quality data acquisition, which can effectively improve the accuracy of rural building recognition. For example, Zhou et al. [25] used UAV oblique photography to construct rural building datasets and generate digital surface models to realize building contour extraction and area measurement. Deng et al. [26] designed a fully convolutional network based on attention mechanisms to extract multiscale features of rural buildings. However, the above studies mainly focus on local regions and have limited generalization ability.

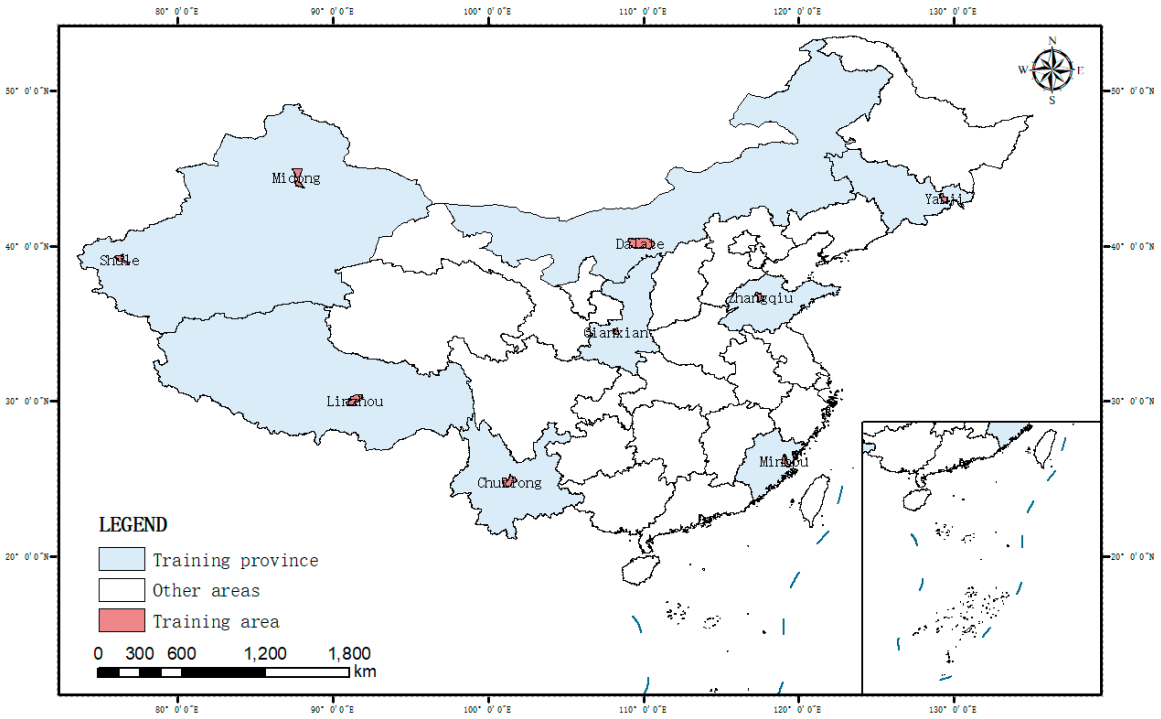
In summary, to solve the problems of insufficient rural building samples and large regional differences, this study constructs a dataset of building images from typical rural areas across China to train deep learning models and improve their generalization ability. Considering the limitations of existing methods, which tend to confuse buildings with objects such as vegetation and roads that have similar spectral features when extracting complex architectural features of rural buildings [27], this paper proposes an improved

remote sensing rural building extraction network called AGSC-Net. Experiments show that the network can effectively extract semantic information from different building types and significantly improve the identification between buildings and backgrounds. The main sections of this paper are arranged as follows. Section 2 introduces the study area and dataset construction and processing. Section 3 presents the main experimental methods, including the improvements and implementation of the AGSC-Net model. Section 4 shows the results of rural building recognition experiments and provides an analysis. Section 5 discusses the effects of different feature combinations on the training results. Finally, Section 6 summarizes the study.

2. Study Area and Data

2.1. Introduction of Study Area

Buildings are closely related to their geographical environment, and regionality is one of the inherent attributes of architecture [28]. When building homes, people often use abundant local materials and technical traditions to construct buildings adapted to the local climate; thus, different regions have formed unique regional architectural cultures [29]. With the industrialization and standardization of the construction industry, the regional features of urban buildings have gradually weakened, while the regionality of rural buildings is still pronounced in many areas, resulting in high heterogeneity among rural buildings [30]. Considering the vast territory and complex diverse landforms in China, as well as the dilemma of insufficient samples and large differences in rural building extraction, this study selected nine typical geographical regions, including the Northeast Plain (Yanji), North China Plain (Zhangqiu), and Loess Plateau (Qian County), to construct a rural building image dataset to cover the main geomorphological types, climate characteristics and regional architectural styles in China with strong typicality and representativeness [31]. Figure 1 presents an overview of the study area.



(**a**)

Figure 1. *Cont.*



(**b**) (**c**)

Figure 1. Study area. (**a**) Province of the study area; (**b**) Remote Sensing Images of Yuan Village, Qian County, Shaanxi Province; (**c**) Yuan Village Building Label (The green pixels of the map represent the marking of the building contour).

2.2. Research Data

Building extraction based on deep learning is a computer vision semantic segmentation task, that is, dividing real objects into buildings and non-buildings [32]. We obtained Google remote sensing images of some rural areas in the above research areas in 2020 with a resolution of 0.29 m from Google Earth and used the open-source LabelMe software version 5.2.1 based on PyQt for image annotation (buildings in red, non-buildings in black) to generate annotated images corresponding to the original images [33]. To adapt to the training process of deep learning, each village orthophoto was uniformly cropped into 256 × 256 pixels to generate cropped images. Data augmentation methods can effectively improve the generalization capability of models and solve overfitting problems [34]. Therefore, this study adopts data augmentation methods to expand rural building sample data. The main methods include applying clockwise rotation of 90 degrees, 180 degrees, 270 degrees, horizontal flipping and vertical flipping to the cropped images obtained in the previous step to generate augmented images after data enhancement. The enhanced dataset contains a total of 13,428 images and annotations, of which 10,742 were randomly selected for model training and 2686 for testing. Figure 2 shows a set of original images, annotated images and augmented images after data enhancement.

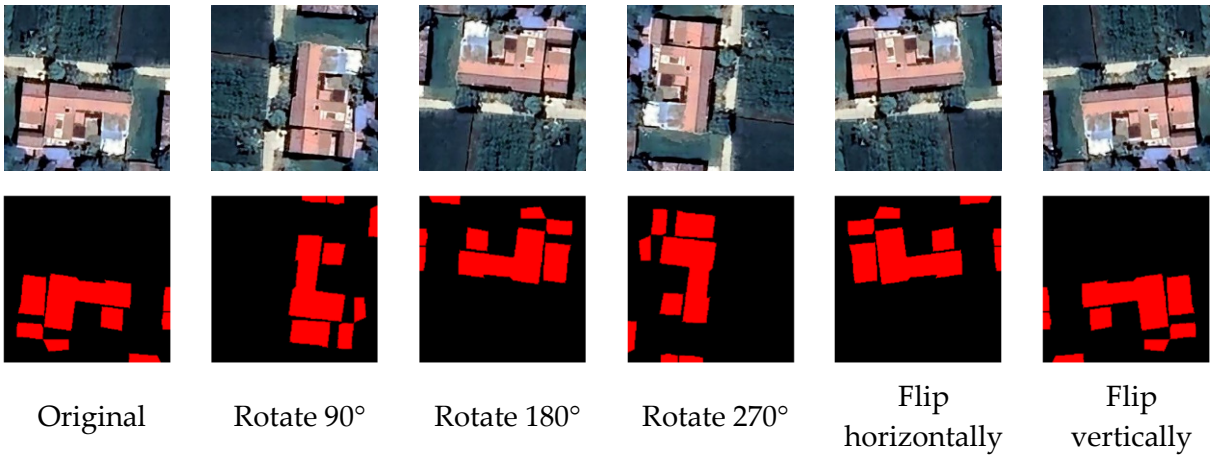


Figure 2. An example of data augmentation by rotating and flipping the Building Image Dataset of Typical Rural Areas in China.

3. Methodology

As presented in Figure 3, as an attention control mechanism-based remote sensing image building extraction network, the main structure of AGSC-Net is as follows. An

encoder–decoder structure based on the U-Net model is adopted to achieve end-to-end building extraction. The encoder gradually obtains semantically richer feature representations through convolution and pooling, while the decoder gradually restores spatial resolution to output prediction results consistent with the input image size [35]. This structure ensures the end-to-end trainability of the entire model. The encoder consists of 4 convolutional blocks, each containing 2 convolutional layers, 2 normalization layers and 2 ReLU activation function layers. The max-pooling layers in the model can extract the maximum values from local regions of the feature maps and reorganize them into new feature maps. The converter consists of 4 AG and Context Collaboration Network (CC-Net) modules from the "skip connections". Traditional FCNs take the global context of the entire input image but cannot distinguish features that are more important for building extraction [36]. Attention modules are introduced between the encoder and decoder to explicitly model building features. The AG module in AGSC-Net can reweight each feature map to enhance the feature response related to buildings and suppress irrelevant feature responses, thereby improving the representation capability for buildings [37].

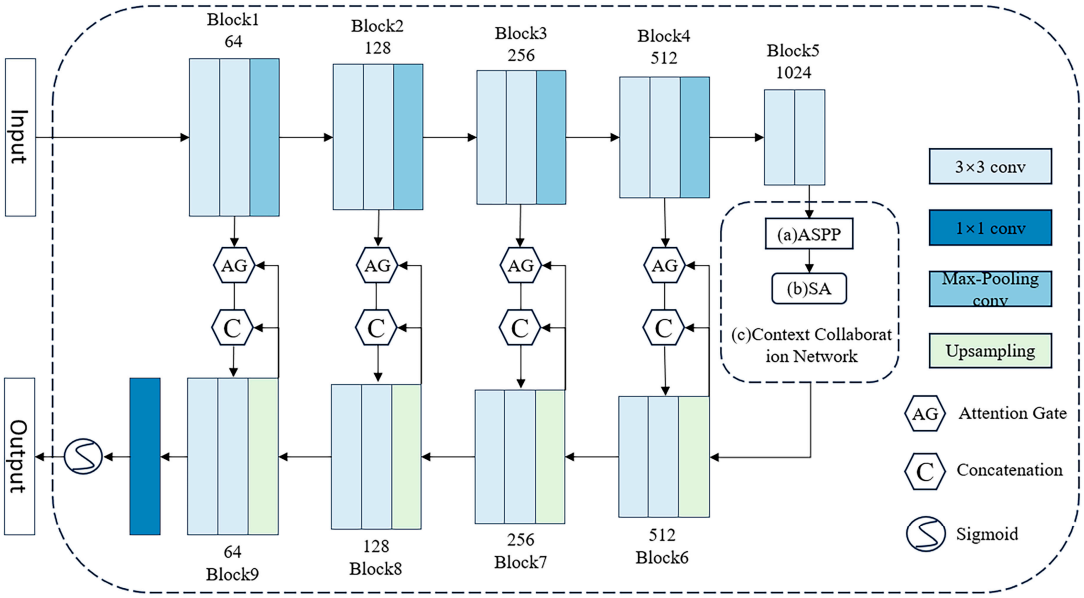


Figure 3. Framework of the proposed AGSC-Net. (a) Atrous Spatial Pyramid Pooling. (b) Self-Attention. (c) The proposed Context Collaboration Network (CC-Net).

The jump connections directly pass the features of each encoder layer to the decoder. This connection allows the decoder end to obtain image features at different levels, which helps restore detailed information and mitigate the vanishing gradient problem in deep networks, making the model easier to optimize [38]. To obtain multiscale building features, AGSC-Net not only uses image pyramid pooling modules but also designs context collaboration networks to learn long-range dependencies between different building instances for more accurate localization. By constructing the feature converter with AG and CC-Net modules, the encoder features are converted before being passed to the decoder, which further enhances the representation capability of building features at the decoder end. Fully convolutional networks are utilized to achieve end-to-end training and prediction. Compared with the segmented processing flow in traditional methods, this end-to-end approach makes feature extraction more tailored to the final goal of building extraction and facilitates model training and deployment [39]. In summary, AGSC-Net integrates the advantages of encoder–decoder, attention mechanisms, multiscale context extraction and other modules to form an efficient network structure for remote sensing image building extraction. This structure retains the stable and efficient basic framework of UNet and significantly improves the expression capability of rural building features through attention mechanisms and multiscale fusion. The specific parameters of AGSC-Net are shown in Table 1.

Table 1. Detailed blocks of the proposed AGSC-Net.

Block	Type	Filter	Channel Size	Output Size
Block 1–4	Conv1	(3,3)	64	256 × 256
	Maxpool1	(2,2)	64	128 × 128
	Conv2	(3,3)	128	128 × 128
	Maxpool2	(2,2)	128	64 × 64
	Conv3	(3,3)	256	64 × 64
	Maxpool3	(2,2)	256	32 × 32
	Conv4	(3,3)	512	32 × 32
	Maxpool4	(2,2)	512	16 × 16
Block 5	PPM		1024	16 × 16
Block 6–9	Up_conv4	Up-(3,3) Conv-(2,2)	512	32 × 32
	AG4		512	32 × 32
	Up4		512	32 × 32
	Up_conv3	Up-(3,3) Conv-(2,2)	256	64 × 64
	AG3		256	64 × 64
	Up3		256	64 × 64
	Up_conv2	Up-(3,3) Conv-(2,2)	128	128 × 128
	AG2		512	128 × 128
	Up2		512	128 × 128
	Up_conv1	Up-(3,3) Conv-(2,2)	64	256 × 256
	AG 1		64	256 × 256
	Up1		64	256 × 256
Block 10	Conv_1×1	(1,1)	1	256 × 256
	Sigmoid		1	256 × 256

Conv: convolution; Maxpool: the maximum pooling; Up: up-sampling, AG: attention gates; Up_conv: up-sampling and convolution.

3.1. AG Module

For the building extraction task from high-resolution remote sensing images, the attention gate (AG), as a spatial attention mechanism, enables the model to selectively focus on different spatial regions of the image to accurately extract building contours in remote sensing images, which can significantly improve segmentation model performance [40]. AG contains two branches for different feature processing. These two branches are fused at the pixel level, which can better integrate the multidimensional information of the input image. AG uses two different activation functions: ReLU for feature normalization and Sigmoid for generating attention coefficients. The combination of these two activation functions introduces nonlinearity to AG. Second, AG contains three sets of trainable parameters that control the gate coefficients W_g, attention coefficients W_x, and output adjustment parameters, making AG more adaptable to specific segmentation tasks. Additionally, its behavior can be controlled by adjusting the parameters. AG is an efficient, controllable and scalable spatial attention mechanism that is well-suited for fine-grained building segmentation tasks from high-resolution remote sensing images. It improves the model's ability to identify key target areas by dual-branch feature fusion, adjustable parameters, and suppressing irrelevant regions. Integrating AG into standard convolutional networks can significantly improve the extraction accuracy of small and densely distributed buildings [41]. The AG module structure is illustrated in Figure 4.

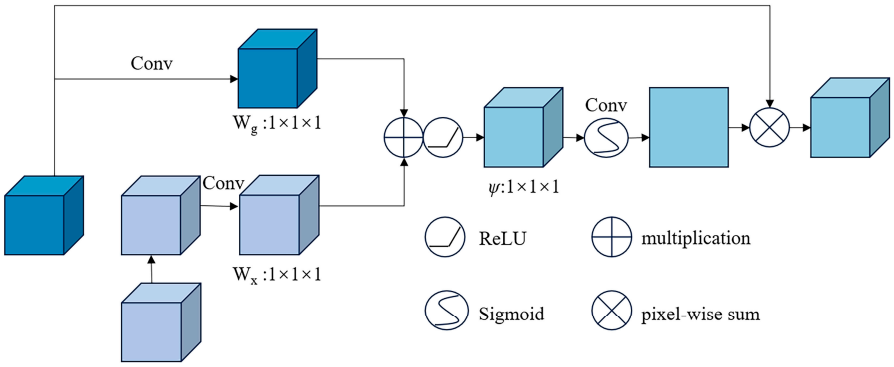


Figure 4. Structure of an AG module.

3.2. Context Collaboration Network (CC-Net)

Atrous convolution can effectively expand the receptive field by using dilated convolutional kernels [42], thereby capturing multiscale information. Based on this, ASPP uses multiple parallel dilated convolutional branches with different dilation rates to extract feature maps and then fuse multiscale features. This multibranch structure can efficiently encode multiscale contexts to improve segmentation performance [43]. Self-attention mechanisms compute correlations between different positions in a sequence to achieve self-attention and context modeling. It can effectively capture long-range dependencies and provide global contextual clues [44]. The structure of our context collaboration network (CC-Net) is illustrated in Figure 5.

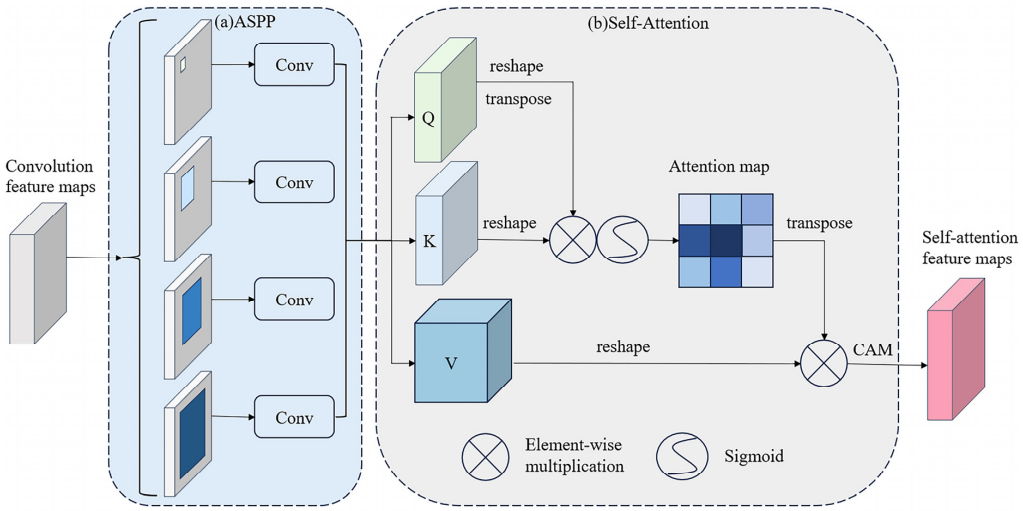


Figure 5. Structure of the CC-Net module. (**a**) Atrous Spatial Pyramid Pooling (ASPP). (**b**) Self-Attention.

CC-Net organically integrates ASPP and Self-Attention. ASPP obtains multiscale feature expressions, and self-attention encodes global contexts and models positional relationships, complementing each other to form a context-collaborative structure. In the decoder, the self-attention features of the encoder output are inputted through skip connections, passing rich global context and localization information. In addition, CC-Net uses channel attention to focus on relevant semantic features. This structure considers both localization information and multiscale features, which can accurately segment buildings in remote sensing images. Through end-to-end training, the multiscale features of ASPP, the global context modeling and positioning information of self-attention, and the collaborative connection between them enable CC-Net to form an efficient building segmentation network with strong capabilities in complex scene understanding, achieving excellent localization and multiscale capabilities.

The convolutional feature maps are generated as the output from the Block5 section of the encoder, and these maps serve as the input for the atrous spatial pyramid pooling. The ASPP module contains 4 parallel branches using dilated convolutions with different dilation rates (1, 6, 12, 18) to obtain feature information at different scales. The 4 groups of feature maps after 1×1 convolution are concatenated to obtain feature expressions sensitive to multiscale features. The SA module contains query, key, and value 1×1 convolutions to generate query, key, and value feature maps. The attention map is generated by calculating the correlations between the query and key feature maps. Then, the value feature map is weighted by the attention map to obtain a new feature map fused with global context information. Finally, the CAM unit reweights the feature channels. The multiscale feature maps provided by ASPP are input into the SA module to fuse global context information. The output of SA will contain rich localization and multiscale features. By extracting multiscale features through ASPP and providing global context and localization information through SA, the combination and connection of the two in the encoder–decoder structure enable CC-Net to consider both localization information and integrate multiscale features, achieving excellent effects in segmenting buildings in complex rural scenes.

4. Experimental Datasets and Evaluation

4.1. Datasets

4.1.1. Building Image Dataset of Typical Rural Areas in China

To solve the problems of insufficient sample data and large regional differences in rural buildings, this study constructed an open-source dataset of typical rural buildings in China for training rural building automatic extraction models based on deep learning. Considering the vast territory and complex diverse landforms in China, rural buildings in different regions have significant differences. This dataset selected sample images from nine geographical regions, including the Northeast Plain, North China Plain, and Yungui Plateau, representing the main geomorphological types, climate characteristics and regional architectural styles in China, ensuring the comprehensiveness and representativeness of the samples [31]. All samples of this dataset come from 2020 Google remote sensing images with a resolution of 0.29 m, clearly reflecting the details of rural buildings. Based on the manual drawing of fine annotations containing building outlines and categories, the diversity of samples is enriched through rotation, flipping and other augmentation means, which can effectively enhance the robustness of models and prevent overfitting problems [34], making up for the overall shortage of rural building samples. This dataset contains 13,428 images and annotations, of which 10,742 were randomly selected for model training and 2686 for testing. Figure 6 shows an original image and its corresponding label, where black represents the background and red represents the building.

4.1.2. China Rural Building UAV Imagery Dataset

This dataset uses UAV images obtained by DJI Inspire2 with a spatial resolution of up to 0.1 m, clearly capturing building boundary features and providing delicate visual information to meet the requirements of deep learning for high-quality image samples. It covers sample images from provinces, including Shaanxi, Jiangsu and Sichuan, with extensive geographical distributions. The images cover common rural building types in China, including masonry, stone and brick, with strong representativeness [45]. The dataset contains 6060 original images and annotated samples, 4848 of which were selected for model training and 1212 for testing. An example of an input image and its corresponding label are presented in Figure 7. The dataset has been made public, and the data service system website is https://www.scidb.cn/en/detail?dataSetId=807518619180204032&version=V1 (accessed on 8 May 2023).



Figure 6. Image and label selected from Building Image Dataset of Typical Rural Areas in China.

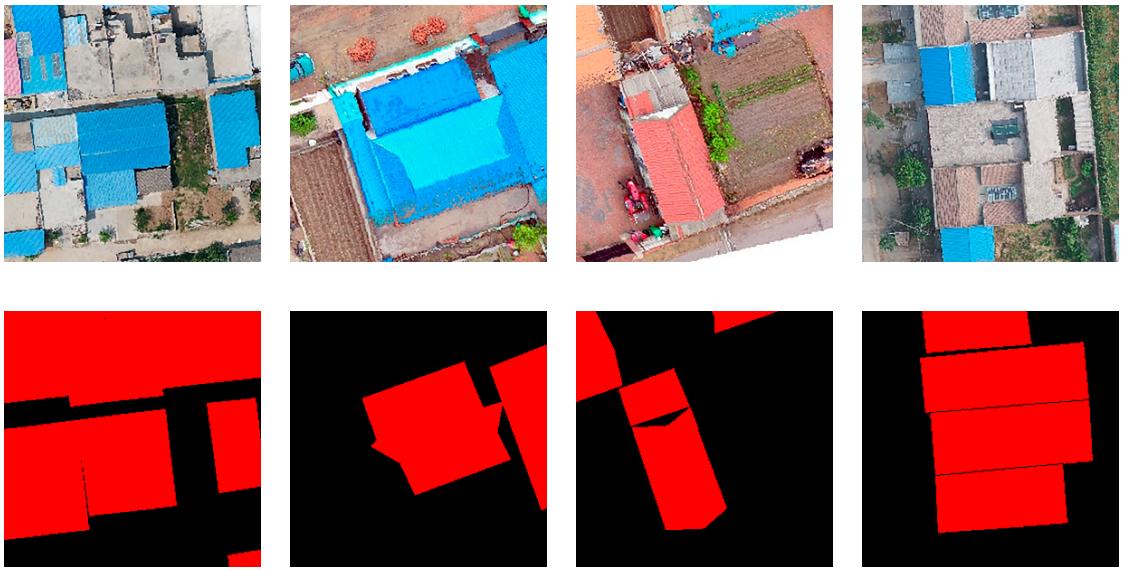


Figure 7. Image and label selected from China Rural Building UAV Imagery Dataset.

4.1.3. Varied Drone Dataset (VDD)

The Varied Drone Dataset (VDD) comprises 400 high-resolution aerial photographic images captured utilizing DJI MAVIC AIR II unmanned aerial vehicles from 23 geographical locations within Nanjing, China. It incorporates pixel-level semantic annotations across seven categories, labeled utilizing LabelMe software. In contrast to other datasets focused solely on urban environments, VDD incorporates diverse rural and natural landscapes, encompassing agricultural lands, villages, forests, hills, and bodies of water. With its large-scale annotated imagery spanning urban, industrial, rural, and natural areas captured from varying angles and conditions, VDD furnishes rich semantic details. VDD is freely available on the website https://vddvdd.com/download/ (accessed on 13 November 2023) [46]. We extracted the portion of the VDD dataset belonging to rural buildings to enrich our China Rural Building Unmanned Aerial Vehicle Imagery Dataset to verify the extraction potential for rural buildings within the unmanned aerial vehicle image dataset utilizing the model. Figure 8 illustrates an example of input images and their corresponding labels in the VDD dataset.

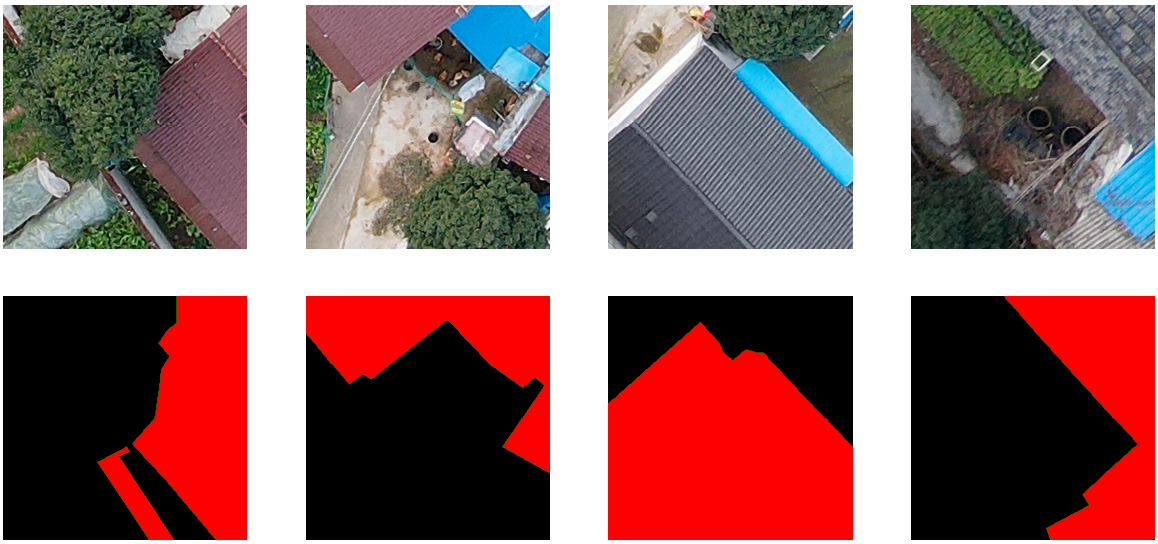


Figure 8. Image and label selected from Varied Drone Dataset.

4.2. Evaluation Metrics

In this study, five metrics, including overall accuracy (OA), precision, recall, F1 score and intersection over union (IoU), were adopted to evaluate the building extraction effects of different models. Among them, OA refers to the ratio of the number of correctly classified pixels to the total number of test pixels: precision reflects the ratio of true positives in all predicted positives, recall represents the ratio of true positives correctly identified, F1 score is the harmonic average of precision and recall, and IoU calculates the ratio of the intersection and union between the predicted and true targets, reflecting the segmentation accuracy. The definitions of the above metrics are as follows:

$$\text{Overall Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

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

where TP, TN, FP and FN represent the number of pixels of true positives, true negatives, false positives and false negatives, respectively. These metrics reflect the building extraction performance of models from different perspectives. OA focuses on overall classification results, precision and recall focus on identification capabilities, F1 score combines precision and recall, and IoU directly examines segmentation accuracy. This study will use the above metrics to evaluate the building extraction results of different models and determine the pros and cons of models based on quantitative results.

4.3. Model Comparisons

To evaluate our rural building extraction method, we selected the typical encoder–decoder architecture extraction model SegNet, AGs-UNet with a structure similar to our building extraction model, C^3Net, and CSA-UNet, also applied to rural building extraction for comparative experiments. A brief introduction of these models is as follows:

(1) SegNet is an image semantic segmentation network based on the encoder–decoder architecture proposed by Badrinarayanan et al. in 2015 [47]. Compared with FCNs, the difference is that the decoder end reuses the max-pooling indices of the encoder pooling layers to help restore segmentation results with rich details [48]. SegNet shows good real-time processing capabilities and semantic segmentation accuracy and is especially suitable for accurate pixel-level building extraction, so it is selected as the benchmark model.

(2) AGs-UNet is an improved version of UNet with multiple AG modules embedded to achieve enhanced representation of building-related features [21]. Its attention mechanism can suppress irrelevant features, thereby improving the model's ability to identify key building areas [49]. AGs-UNet has verified the improvement of extraction performance on multiple high-resolution urban remote sensing image datasets and is selected to evaluate the effect of the attention module on this task.

(3) C^3Net adopts an encoder–decoder structure. The decoder integrates context-aware and edge residual refinement modules to achieve a balance between extraction accuracy and integrity [23]. This model has been proven to make progress on multiple open datasets and is considered to provide comprehensive comparative extraction capabilities for this task.

(4) CSA-UNet is based on the encoder–decoder structure, using channel spatial attention mechanisms to focus on key building features [50]. Compared with other datasets, its specifically constructed rural building sample dataset using UAV images is more consistent with the research objectives of this study. Considering the impact of dataset differences on model generalization capabilities, CSA-UNet is selected to evaluate the adaptability of models under new sample distributions.

In summary, these models have different focuses on building extraction tasks and can serve as comparative benchmarks to analyze the strengths and weaknesses of the AGSC-Net model in the application of extracting buildings in typical rural areas of China, evaluating the extraction capabilities of the model. To evaluate the extraction capabilities of the proposed model, all networks are tested on two datasets: the Building Image Dataset of Typical Rural Areas in China and the complementary China Rural Building UAV Imagery Dataset. The experimental results from these datasets are utilized to verify the extraction potential of the AGSC-Net model. The empirical results of these comparative experiments are detailed in the subsequent sections.

4.4. Experimental Settings

The model training and testing in this study are based on the PyTorch deep learning framework and use open-source image processing modules, including TorchVision, Skimage and Matplotlib. To accelerate the calculation, the computing device used in this study is the NVIDIA GeForce GTX 3090 graphics processing unit, with a memory capacity of 24 GB, and CUDA 11.0 has been installed for GPU acceleration. Considering the constraint of GPU memory resources, all image samples are randomly cropped to 256×256 pixels to meet the memory requirements for model training and cross-validation in each round. In the hyper-parameter configuration of the model, we performed multiple comparative experiments to determine the optimal settings for each model. The Adam optimization algorithm [51] is used for model training, with an initial learning rate set to 0.0001. To suppress overfitting, L2 regularization is introduced in all convolutional layers, and the weight decay coefficient [52] is set to 0.0001. Additionally, considering the limitation of the GPU memory of the computing device, the training of each model takes batches of 18 images as input and iterates 200 epochs on each dataset. To ensure the fairness of comparative experiments between different models, all comparison models are trained under the same software and hardware environment. The training accuracy and loss profiles [53] of the proposed AGSC-Net architecture on both datasets are presented in Figure 9. As shown, the model accuracy consistently improves over training iterations and eventually converges to approximately 0.94 with minor fluctuations. Concurrently, the loss monotonically decreases and reaches a plateau, indicating effective optimization of the AGSC-Net model. Unless otherwise specified, all comparative models were trained

under the same hardware and software configuration detailed previously to enable fair benchmarking against the proposed AGSC-Net.

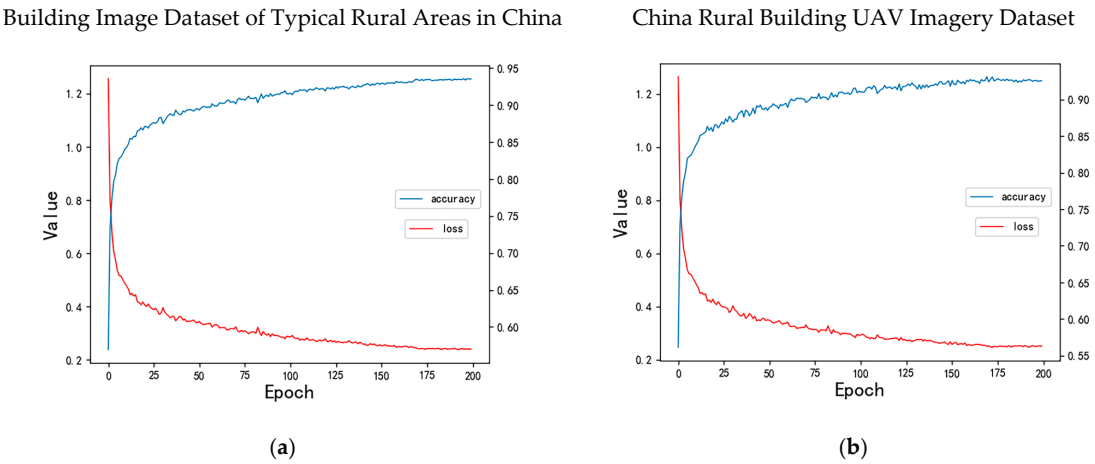


Figure 9. Variation in training accuracy and loss value of AGSC-Net.

4.5. Results

To validate the superiority of our self-constructed datasets for rural building extraction model training, we will train and test the proposed model on the self-constructed Building Image Dataset of Typical Rural Areas in China and the open China Rural Building UAV Image Dataset respectively. Test data from four geographic regions with distinct characteristics were selected from each dataset to assess the proposed model's building extraction capabilities. Representative sample images were chosen from each image set for qualitative analysis. The proposed rural building extraction network AGSC-Net was compared to models such as SegNet, AGs-UNet, C^3Net and CSA-UNet. As shown in Figures 10 and 11, the original remote sensing images and corresponding ground truths occupy the first two rows, while rows 3–7 exhibit the building extraction outputs of the different models. The green and black pixels in the extraction results denote regions classified by the models as buildings and non-buildings, respectively.

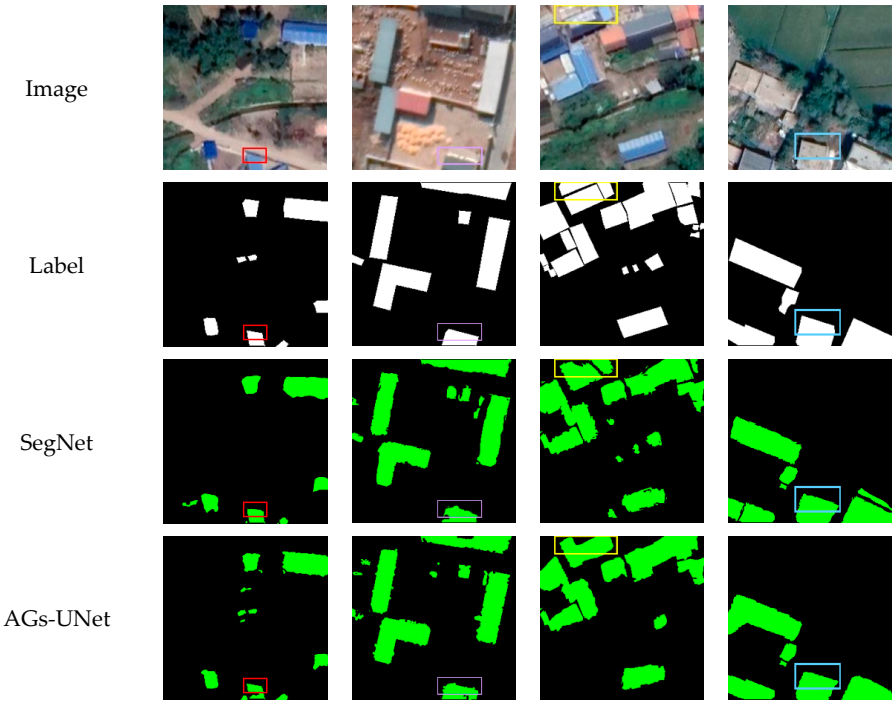


Figure 10. *Cont.*

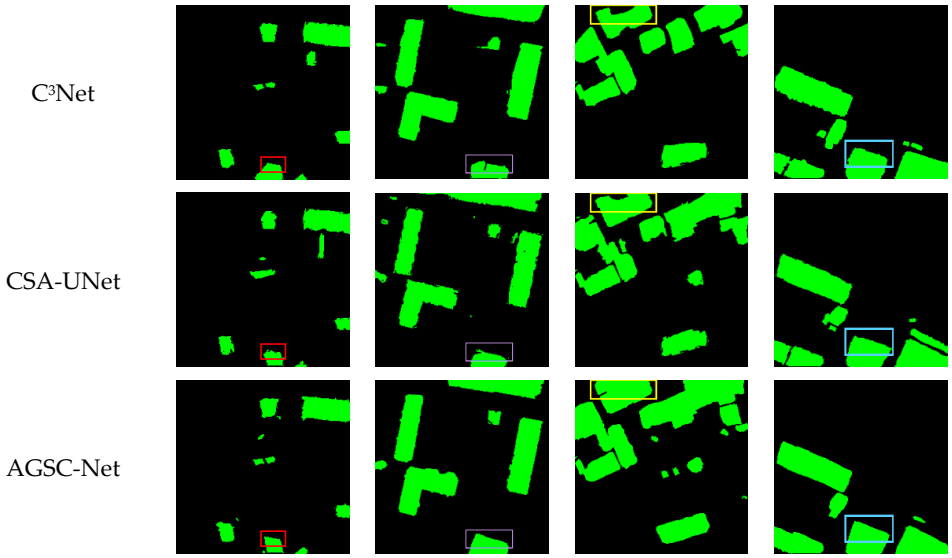


Figure 10. Comparison of extraction results of each model building in Building Image Dataset of Typical Rural Areas in China. The first two rows are aerial images and ground truth, respectively. Rows 3–7 are building extraction results of SegNet, AGs-UNet, C^{3}Net, CSA-UNet, and our proposed AGSC-Net, respectively. The green and black pixels of the maps represent the predictions of true positive and true negative, respectively. The red, purple, yellow, and blue rectangles represent the parts of the extraction results that need to be focused on in the cases of sparse building, moderately dense building, dense building, and dense vegetation, respectively.

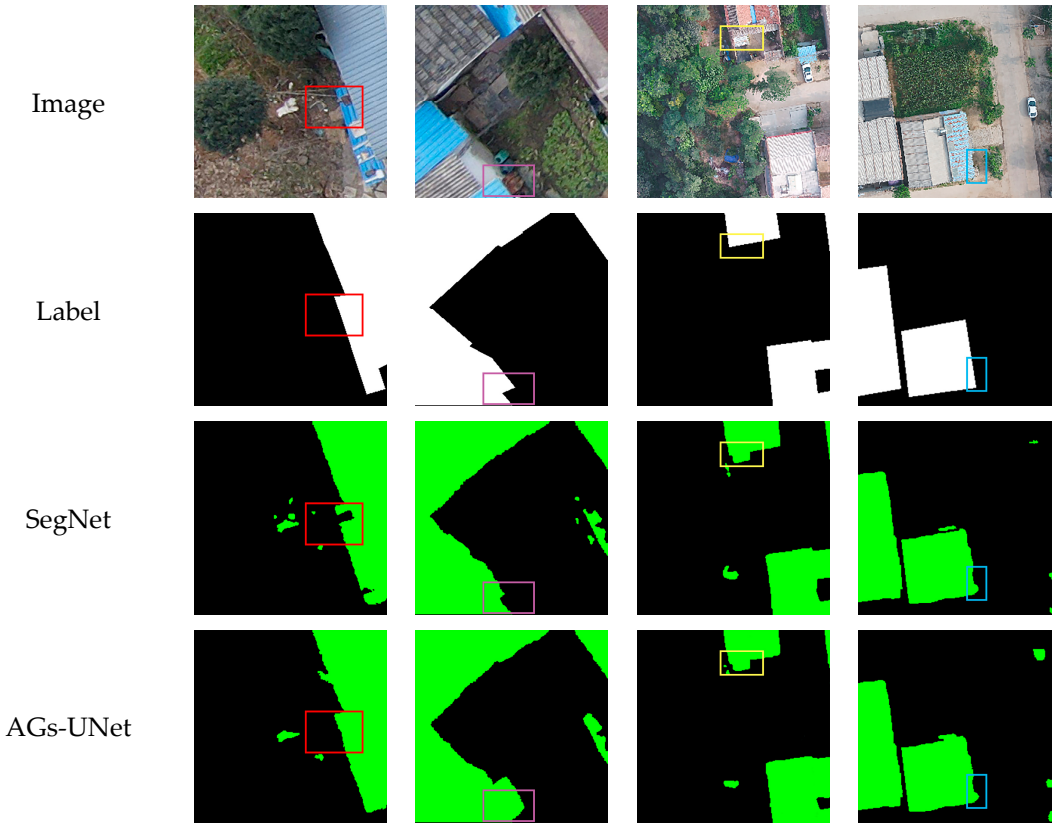


Figure 11. *Cont.*

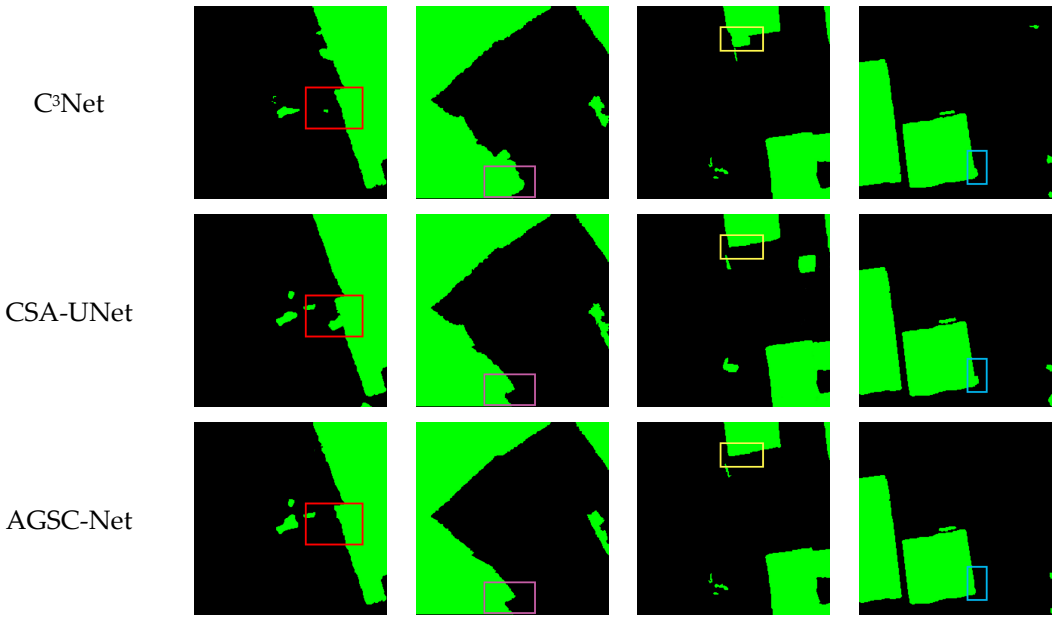


Figure 11. Comparison of extraction results of each model building in China Rural Building UAV Imagery Dataset. The first two rows are aerial images and ground truth, respectively. Rows 3–7 are building extraction results of SegNet, AGs-UNet, C^3Net, CSA-UNet, and our proposed AGSC-Net, respectively. The green and black pixels of the maps represent the predictions of true positive and true negative, respectively. The red, purple, yellow, and blue rectangles represent the parts of the extraction results that need to be focused on in the cases of sparse building, moderately dense building, dense building, and dense vegetation, respectively.

In Figure 10, the red and purple rectangles in the first two columns contain small buildings. The extraction results of SegNet and AGs-UNet display varying degrees of internal structural deficiencies. C^3Net exhibits more complete internal information but insufficiently defined edges. CSA-UNet has sharper yet less smooth and natural edges. By contrast, AGSC-Net accurately and seamlessly extracts small building edges, contours and internal structures. The yellow rectangle in the third column encapsulates a densely populated large-scale building area. All models misdetect or omit some regions, but AGSC-Net has fewer missed detections, uniquely and accurately extracting even the buildings inside the yellow box. The blue rectangle area in the fourth column suffers from vegetation interference. AGs-UNet and C^3Net display localized contour deficiencies. SegNet and CSA-UNet also contain some misdetections. Comparatively, AGSC-Net obtained more precise extraction results. Overall, the AGSC-Net model demonstrates superior performance across diverse complex scenarios, validating the efficacy of its designed modules and network architecture.

Figure 11 exhibits the extraction outputs on the UAV image dataset. With higher resolution, this dataset harbors more complex spectral and texture traits alongside profound detail. Moreover, with a smaller volume, it poses greater difficulty for extraction. When processing the sharper building contours, SegNet and AGs-UNet's outlines show noticeable deficiencies and misclassifications. The contours from C^3Net and CSA-UNet appear more complete yet risk false positives with higher predictive values due to vegetation. Accounting for improvements conferred by the AG module and Contextual Correlation Network (CC-Net) module, AGSC-Net surpasses the other models, extracting building edges more accurately with fewer internal gaps.

From the quantitative evaluation, compared with other models, AGSC-Net achieved the best performance in all metrics. From the quantitative results of the first dataset, as illustrated in Table 2, it can be seen that AGSC-Net outperforms the other models in all metrics. Specifically, compared with other models, AGSC-Net improved the OA metric

by 0.3–3 percentage points and the IoU metric by 1–2.8 percentage points. This indicates that AGSC-Net can identify building pixels more accurately and provide finer building segmentation results. On the second dataset, as delineated in Table 3, AGSC-Net improved the OA metric by 0.2–0.5 percentage points and the IoU metric by 0.6–1 percentage points compared to other models, maintaining significant performance advantages. Although C^3Net has the highest precision P on the first dataset, AGSC-Net has a higher F1 score by balancing precision and recall, evaluating the overall effect of the model more comprehensively. In addition, the IoU metric combines accuracy and recall metrics, and the IoU metric of AGSC-Net is significantly higher than that of the other models, indirectly verifying that its extraction results are more accurate and complete.

Table 2. Quantitative accuracy comparison of different methods on the Building Image Dataset of Typical Rural Areas in China.

Model	OA	P	R	F1	IoU
SegNet	0.898	0.843	0.859	0.851	0.741
AGs-UNet	0.898	0.836	0.853	0.844	0.730
C^3Net	0.896	0.830	0.864	0.847	0.734
CSA-UNet	0.900	0.846	0.863	0.855	0.746
AGSC-Net	0.901	0.847	0.863	0.855	0.747

Table 3. Quantitative accuracy comparison of different methods on the China Rural Building UAV Imagery Dataset.

Model	OA	P	R	F1	IoU
SegNet	0.881	0.828	0.829	0.828	0.707
AGs-UNet	0.878	0.811	0.827	0.819	0.694
C^3Net	0.856	0.800	0.832	0.816	0.689
CSA-UNet	0.884	0.824	0.846	0.835	0.717
AGSC-Net	0.886	0.837	0.833	0.835	0.717

AGSC-Net has stronger adaptability to different types of rural buildings. It can be seen from the comparison of the two datasets that the metrics of all models on the second dataset are generally slightly lower than those on the first dataset. This is because the second dataset covers samples from the main terrain types in China, including masonry, stone, brick and other building types, but has fewer data samples than the first dataset, and the building styles are more complex, increasing the extraction difficulty. However, the performance decline in the AGSC-Net model is smaller than that of the other models, proving its stronger adaptability to different types of rural buildings. Specifically, although the P, R, and F1 metrics of all models decreased slightly on the second dataset compared to the first dataset, the decline in AGSC-Net is the smallest, and the R metric even increased slightly, indicating that its extraction results are more comprehensive and complete. At the same time, the IoU metric of AGSC-Net only dropped from 0.717 to 0.747, a decrease of less than 0.5%, significantly better than the 1–2% decrease in other models. Compared with other models using similar improvements, e.g., attention mechanisms such as CSA-UNet, the AGSC-Net model goes further by simultaneously utilizing the AG and context collaboration network (CC-Net), improving the representation capability and modeling capability of building features, thus achieving optimal performance. In summary, by achieving the best extraction results on two representative rural building datasets, the AGSC-Net model has proven its stronger generalization adaptability to extracting different building types in complex rural scenes.

5. Ablation Experiments

The following three ablation experiments are set up: (a) only containing the U-shaped convolutional neural network (UNet) architecture, (b) an improved network (UNet + AG)

with AG modules based on UNet, and (c) the complete model (AGSC-Net) with both AG and context collaboration (CC-Net) modules. Under the same computer software and hardware environment and hyperparameter settings, the above three models were trained for 200 epochs. After training, we selected the same images from the validation set of the Building Image Dataset of Typical Rural Areas in China for model effect verification to ensure fair comparison under the same conditions for each experiment.

Figure 12 displays the visualization results of the three models on test sets of three different scenarios. The first row is a scene with sparse small building targets. The UNet model has significant building omission problems, while the UNet + AG model has some commission errors regarding small, scattered buildings. The second row is a complex rural building scene. The UNet model has many omissions of building edges, while the building edges extracted by the AGSC-Net model are smoother, indicating that the AG module effectively enhances the ability to identify building edges. The third row shows a scene with considerable vegetation interference. In this case, the detection effects of UNet and UNet + AG are poor, resulting in many misjudgments. The AGSC-Net model correctly excludes vegetation interference and accurately extracts buildings. The results prove that AGSC-Net can effectively suppress background interference and accurately extract buildings through the multilevel attention mechanism of the AG and CC-Net modules and the extracted multidimensional features.

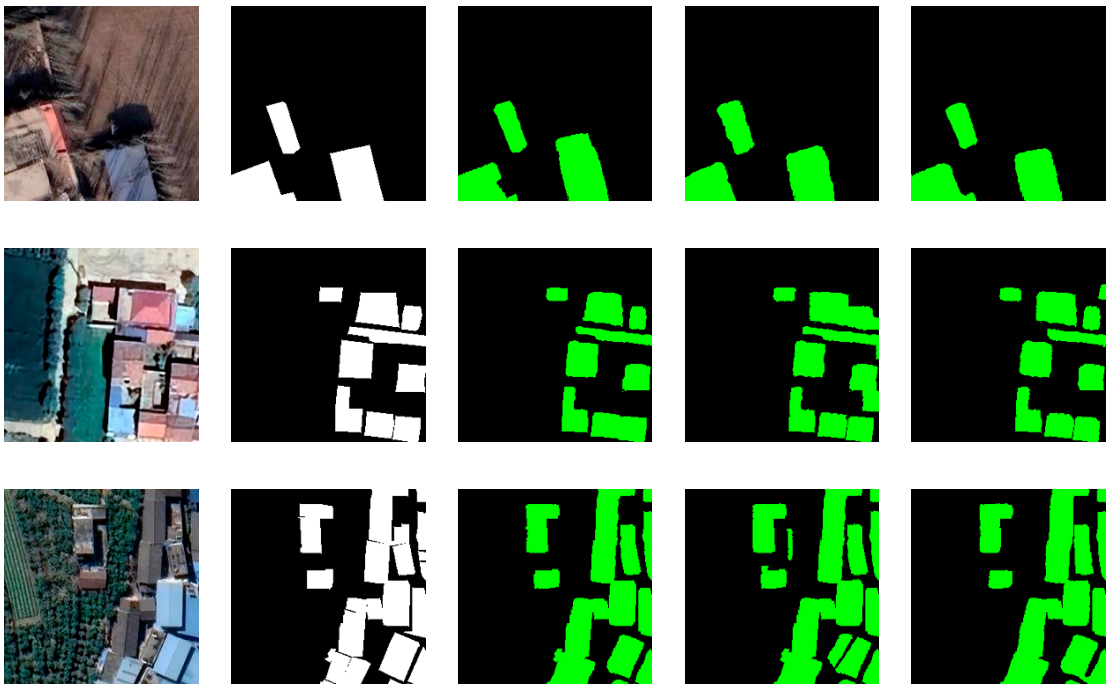


Figure 12. Comparison of ablation experimental results. The first two columns are aerial images and ground truth, respectively. Columns 3–5 are building extraction results. The green and black pixels of the maps represent the predictions of true positive and true negative, respectively.

This study also verified the positive effects of the AG and CC-Net modules on building extraction through quantitative analysis, providing references for constructing more accurate building extraction models. As indicated in Table 4, compared with the baseline UNet model, the UNet + AG model with AG modules improved the detection precision (P) by 0.17 percentage points, mainly because the AG module enhanced the model's ability to identify building edges. Compared with the UNet + AG model, the complete AGSC-Net model improved the detection precision (P) by 0.64 percentage points, and recall and other metrics were further improved. Especially in complex scenes with rich background

interference, AGSC-Net significantly reduced commission errors. This proves that, under the synergistic effect of the AG and CC modules, the multilevel attention mechanism can effectively improve the extraction accuracy of buildings by enhancing building-related features and eliminating background interference.

Table 4. Accuracy statistics of ablation experiment in Building Image Dataset of Typical Rural Areas in China.

Model	OA	P	R	F1	IoU
(a) UNet	0.8969	0.8425	0.8567	0.8495	0.7384
(b) UNet + AG	0.8973	0.8442	0.8565	0.8503	0.7396
(c) AGSC-Net	0.9012	0.8506	0.8613	0.8559	0.7481

6. Conclusions

To address the problems of insufficient samples and large regional differences faced by existing rural building extraction techniques, this study constructed a dataset of building images from typical rural areas across China to provide sample support for model training. Meanwhile, an improved rural building extraction network called AGSC-Net was proposed by integrating attention mechanisms and multiscale contextual feature extraction modules, which significantly enhanced the model's ability to recognize and represent buildings in complex rural scenes. Experiments showed that, compared with other models, AGSC-Net achieved the optimal metrics on two representative rural building datasets, validating the accuracy of its extraction effects. Ablation experiments also proved that the introduction of AG and CC-Net modules can improve the model's robustness in complex environments. The innovations of this study are as follows: (1) constructing a dataset of rural building images from typical areas in China to compensate for the lack of samples; and (2) designing the AGSC-Net model that integrates attention mechanisms and multiscale modeling to enhance feature expression capabilities and validate the model's generalization ability on multiple datasets. This study provides new ideas for developing rural building automated extraction techniques by constructing datasets and designing improved models. However, this study has some limitations: (1) the dataset covers a wide area but has fewer examples for some regions, and (2) the model's extraction effects for buildings with complex textures need to be improved. Future studies could continue to expand the scale and diversity of the dataset and explore methods to introduce texture features to further enhance model adaptability.

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

Funding: This research was funded by the China National Key R and D Program during the 13th Five-year Plan Period, grant number 2019YFD1100800, and the National Natural Science Foundation of China, grant number 41801308.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: China Rural Building UAV Imagery Dataset available at https://www.scidb.cn/en/detail?dataSetId=807518619180204032&version=V1 (accessed on 8 May 2023). Varied Drone Dataset available at https://vddvdd.com/download/ (accessed on 13 November 2023). The source code is released at https://github.com/Yiyizzzzz/AGSCNet (accessed on 13 November 2023). The rest of the data and code of this study can be obtained from the corresponding authors according to requirements.

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

References

1. Zhang, J.; Fukuda, T.; Yabuki, N. Development of a City-Scale Approach for Façade Color Measurement with Building Functional Classification Using Deep Learning and Street View Images. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 551. [CrossRef]
2. Rao, A.; Jung, J.; Silva, V.; Molinario, G.; Yun, S.-H. Earthquake Building Damage Detection Based on Synthetic-Aperture-Radar Imagery and Machine Learning. *Nat. Hazards Earth Syst. Sci.* **2023**, *23*, 789–807. [CrossRef]
3. Le, Q.H.; Shin, H.; Kwon, N.; Ho, J.; Ahn, Y. Deep Learning Based Urban Building Coverage Ratio Estimation Focusing on Rapid Urbanization Areas. *Appl. Sci.* **2022**, *12*, 11428. [CrossRef]
4. Zhao, Q.; Zhou, L.; Lv, G. A 3D Modeling Method for Buildings Based on LiDAR Point Cloud and DLG. *Comput. Environ. Urban Syst.* **2023**, *102*, 101974. [CrossRef]
5. Wei, R.; Fan, B.; Wang, Y.; Yang, R. A Query-Based Network for Rural Homestead Extraction from VHR Remote Sensing Images. *Sensors* **2023**, *23*, 3643. [CrossRef] [PubMed]
6. Wang, C.; Sun, W.; Wu, H.; Zhao, C.; Teng, G.; Yang, Y.; Du, P. A Low-Altitude Remote Sensing Inspection Method on Rural Living Environments Based on a Modified YOLOv5s-ViT. *Remote Sens.* **2022**, *14*, 4784. [CrossRef]
7. Hine, J.; Sasidharan, M.; Torbaghan, M.E.; Burrow, M.; Usman, K. *Evidence of the Impact of Rural Road Investment on Poverty Reduction and Economic Development*; Institute of Development Studies: Brighton, UK, 2019.
8. Cogato, A.; Cei, L.; Marinello, F.; Pezzuolo, A. The Role of Buildings in Rural Areas: Trends, Challenges, and Innovations for Sustainable Development. *Agronomy* **2023**, *13*, 1961. [CrossRef]
9. Zhu, L.; Wang, B.; Sun, Y. Multi-Objective Optimization for Energy Consumption, Daylighting and Thermal Comfort Performance of Rural Tourism Buildings in North China. *Build. Environ.* **2020**, *176*, 106841. [CrossRef]
10. Torvi, D.A. Fire Protection in Agricultural Facilities: A Review of Research, Resources and Practices. *J. Fire Prot. Eng.* **2003**, *13*, 185–215. [CrossRef]
11. Jia, Y.; Zhang, X.; Xiang, R.; Ge, Y. Super-Resolution Rural Road Extraction from Sentinel-2 Imagery Using a Spatial Relationship-Informed Network. *Remote Sens.* **2023**, *15*, 4193. [CrossRef]
12. Benarchid, O.; Raissouni, N.; Adib, S.E.; Abbous, A.; Azyat, A.; Achhab, N.B.; Lahraoua, M.; Chahboun, A. Building Extraction Using Object-Based Classification and Shadow Information in Very High Resolution Multispectral Images, a Case Study: Tetuan, Morocco. *Can. J. Image Process. Comput. Vis.* **2013**, *4*, 1–8.
13. Xing, H.; Zhu, L.; Chen, B.; Zhang, L.; Hou, D.; Fang, W. A Novel Change Detection Method Using Remotely Sensed Image Time Series Value and Shape Based Dynamic Time Warping. *Geocarto Int.* **2022**, *37*, 9607–9624. [CrossRef]
14. Hou, D.; Wang, S.; Tian, X.; Xing, H. PCLUDA: A Pseudo-Label Consistency Learning- Based Unsupervised Domain Adaptation Method for Cross-Domain Optical Remote Sensing Image Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5600314. [CrossRef]
15. Wang, S.; Hou, D.; Xing, H. A Self-Supervised-Driven Open-Set Unsupervised Domain Adaptation Method for Optical Remote Sensing Image Scene Classification and Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5605515. [CrossRef]
16. Muhammed, E.; El-Shazly, A.; Morsy, S. Building Rooftop Extraction Using Machine Learning Algorithms for Solar Photovoltaic Potential Estimation. *Sustainability* **2023**, *15*, 11004. [CrossRef]
17. Xing, H.; Niu, J.; Feng, Y.; Hou, D.; Wang, Y.; Wang, Z. A Coastal Wetlands Mapping Approach of Yellow River Delta with a Hierarchical Classification and Optimal Feature Selection Framework. *CATENA* **2023**, *223*, 106897. [CrossRef]
18. Avudaiammal, R.; Elaveni, P.; Selvan, S.; Rajangam, V. Extraction of Buildings in Urban Area for Surface Area Assessment from Satellite Imagery Based on Morphological Building Index Using SVM Classifier. *J. Indian Soc. Remote Sens.* **2020**, *48*, 1325–1344. [CrossRef]
19. Wei, S.; Zhang, T.; Ji, S.; Luo, M.; Gong, J. BuildMapper: A Fully Learnable Framework for Vectorized Building Contour Extraction. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 87–104. [CrossRef]
20. Xia, L.; Mi, S.; Zhang, J.; Luo, J.; Shen, Z.; Cheng, Y. Dual-Stream Feature Extraction Network Based on CNN and Transformer for Building Extraction. *Remote Sens.* **2023**, *15*, 2689. [CrossRef]
21. Yu, M.; Chen, X.; Zhang, W.; Liu, Y. AGs-UNet: Building Extraction Model for High Resolution Remote Sensing Images Based on Attention Gates U Network. *Sensors* **2022**, *22*, 2932. [CrossRef]
22. Ait El Asri, S.; Negabi, I.; El Adib, S.; Raissouni, N. Enhancing Building Extraction from Remote Sensing Images through UNet and Transfer Learning. *Int. J. Comput. Appl.* **2023**, *45*, 413–419. [CrossRef]
23. Gong, M.; Liu, T.; Zhang, M.; Zhang, Q.; Lu, D.; Zheng, H.; Jiang, F. Context–Content Collaborative Network for Building Extraction from High-Resolution Imagery. *Knowl.-Based Syst.* **2023**, *263*, 110283. [CrossRef]
24. 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]
25. Zhou, J.; Liu, Y.; Nie, G.; Cheng, H.; Yang, X.; Chen, X.; Gross, L. Building Extraction and Floor Area Estimation at the Village Level in Rural China Via a Comprehensive Method Integrating UAV Photogrammetry and the Novel EDSANet. *Remote Sens.* **2022**, *14*, 5175. [CrossRef]

26. Deng, W.; Shi, Q.; Li, J. Attention-Gate-Based Encoder–Decoder Network for Automatical Building Extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2611–2620. [CrossRef]
27. Deng, R.; Guo, Z.; Chen, Q.; Sun, X.; Chen, Q.; Wang, H.; Liu, X. A Dual Spatial-Graph Refinement Network for Building Extraction from Aerial Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 23903570. [CrossRef]
28. Li, H.; Qiu, P.; Wu, T. The Regional Disparity of Per-Capita CO_2 Emissions in China's Building Sector: An Analysis of Macroeconomic Drivers and Policy Implications. *Energy Build.* **2021**, *244*, 111011. [CrossRef]
29. Thi, H.L.V.; Nguyen, T.Q. Adaptive Reuse of Local Buildings in Sapa, Vietnam for Cultural Tourism Development towards Sustainability. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *878*, 012032. [CrossRef]
30. Zhang, T.; Chen, X.; Zhang, F.; Yang, Z.; Wang, Y.; Li, Y.; Wei, L. A Case Study of Refined Building Climate Zoning under Complicated Terrain Conditions in China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8530. [CrossRef]
31. Mao, T.; Li, Q. Research on the Relationship between the Formation of Local Construction Culture and Geographical Environment Based on Adaptability Analysis. *J. King Saud Univ. Sci.* **2023**, *35*, 102387. [CrossRef]
32. Liu, Z.; Shi, Q.; Ou, J. LCS: A Collaborative Optimization Framework of Vector Extraction and Semantic Segmentation for Building Extraction. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5632615. [CrossRef]
33. Dai, G.; Hu, L.; Fan, J. DA-ActNN-YOLOV5: Hybrid YOLO v5 Model with Data Augmentation and Activation of Compression Mechanism for Potato Disease Identification. *Comput. Intell. Neurosci.* **2022**, *2022*, e6114061. [CrossRef]
34. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [CrossRef]
35. 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*, 6002905. [CrossRef]
36. Yu, M.; Zhang, W.; Chen, X.; Liu, Y.; Niu, J. An End-to-End Atrous Spatial Pyramid Pooling and Skip-Connections Generative Adversarial Segmentation Network for Building Extraction from High-Resolution Aerial Images. *Appl. Sci.* **2022**, *12*, 5151. [CrossRef]
37. Liu, W.; Liu, H.; Liu, C.; Kong, J.; Zhang, C. AGDF-Net: Attention-Gated and Direction-Field-Optimized Building Instance Extraction Network. *Sensors* **2023**, *23*, 6349. [CrossRef]
38. Li, P.; Sun, Z.; Duan, G.; Wang, D.; Meng, Q.; Sun, Y. DMU-Net: A Dual-Stream Multi-Scale U-Net Network Using Multi-Dimensional Spatial Information for Urban Building Extraction. *Sensors* **2023**, *23*, 1991. [CrossRef]
39. Guo, H.; Su, X.; Wu, C.; Du, B.; Zhang, L. Decoupling Semantic and Edge Representations for Building Footprint Extraction from Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5613116. [CrossRef]
40. Wang, Z.; Xu, N.; Wang, B.; Liu, Y.; Zhang, S. Urban Building Extraction from High-Resolution Remote Sensing Imagery Based on Multi-Scale Recurrent Conditional Generative Adversarial Network. *GIScience Remote Sens.* **2022**, *59*, 861–884. [CrossRef]
41. Zhou, Y.; Chen, Z.; Wang, B.; Li, S.; Liu, H.; Xu, D.; Ma, C. BOMSC-Net: Boundary Optimization and Multi-Scale Context Awareness Based Building Extraction from High-Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5618617. [CrossRef]
42. Qiu, Y.; Wu, F.; Yin, J.; Liu, C.; Gong, X.; Wang, A. MSL-Net: An Efficient Network for Building Extraction from Aerial Imagery. *Remote Sens.* **2022**, *14*, 3914. [CrossRef]
43. Chan, S.; Wang, Y.; Lei, Y.; Cheng, X.; Chen, Z.; Wu, W. Asymmetric Cascade Fusion Network for Building Extraction. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 23709453. [CrossRef]
44. Liang, S.; Hua, Z.; Li, J. Enhanced Self-Attention Network for Remote Sensing Building Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4900–4915. [CrossRef]
45. Liu, Y.; Yang, X.; Li, J.; Cheng, H.; Zhou, J.; Fan, X.; Zhang, H.; Li, X.; Qi, W.; Li, Z.; et al. A dataset of building sampling and labeling of UAV images in rural China. *Sci. Data Bank* **2022**, *7*, 182–194. [CrossRef]
46. Cai, W.; Jin, K.; Hou, J.; Guo, C.; Wu, L.; Yang, W. VDD: Varied Drone Dataset for Semantic Segmentation. *arXiv* **2023**, arXiv:2305.13608.
47. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
48. 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]
49. Tian, Q.; Zhao, Y.; Li, Y.; Chen, J.; Chen, X.; Qin, K. Multiscale Building Extraction with Refined Attention Pyramid Networks. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 21408934. [CrossRef]
50. Shi, X.; Huang, H.; Pu, C.; Yang, Y.; Xue, J. CSA-UNet: Channel-Spatial Attention-Based Encoder–Decoder Network for Rural Blue-Roofed Building Extraction From UAV Imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6514405. [CrossRef]
51. Liu, Y.; Gross, L.; Li, Z.; Li, X.; Fan, X.; Qi, W. Automatic Building Extraction on High-Resolution Remote Sensing Imagery Using Deep Convolutional Encoder-Decoder with Spatial Pyramid Pooling. *IEEE Access* **2019**, *7*, 128774–128786. [CrossRef]

52. Dixit, M.; Chaurasia, K.; Mishra, V.K. Automatic Building Extraction from High-Resolution Satellite Images Using Deep Learning Techniques. In Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences, Kurukshetra, India, 1–3 May 2020; Dave, M., Garg, R., Dua, M., Hussien, J., Eds.; Algorithms for Intelligent Systems. Springer: Singapore, 2021; pp. 773–783, ISBN 9789811575327.
53. Zhang, X.; Akber, M.Z.; Zheng, W. Predicting the Slump of Industrially Produced Concrete Using Machine Learning: A Multiclass Classification Approach. *J. Build. Eng.* **2022**, *58*, 104997. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.