



Article

An Improved U-Net Network for Sandy Road Extraction from Remote Sensing Imagery

Yunfeng Nie ¹, Kang An ^{1,2,*} , Xingfeng Chen ² , Limin Zhao ², Wantao Liu ³, Xing Wang ^{4,5}, Yihao Yu ¹, Wenyi Luo ², Kewei Li ¹ and Zhaozhong Zhang ¹

- ¹ School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China; 70064@nchu.edu.cn (Y.N.); yuyihaofly@163.com (Y.Y.); 36039@nchu.edu.cn (K.L.); 71033@nchu.edu.cn (Z.Z.)
- ² Aerospace Information Institute, Chinese Academy of Sciences, Beijing 100094, China; chenxf@aircas.ac.cn (X.C.); zhaolm@aircas.ac.cn (L.Z.); lwy2570114470@163.com (W.L.)
- ³ Tianjin Institute of Advanced Technology, Tianjin 300459, China; watoliu@163.com
- ⁴ School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China; wangxing@nju.edu.cn
- ⁵ School of Atmosphere Science, Nanjing University, Nanjing 210023, China
- * Correspondence: akcsat@126.com; Tel.: +86-182-5380-1575

Abstract: The extraction of sandy roads from remote sensing images is important for field ecological patrols and path planning. Extraction studies on sandy roads face limitations because of various factors (e.g., sandy roads may have poor continuity, may be obscured by external objects, and/or have multi-scale and banding characteristics), in addition to the absence of publicly available datasets. Accordingly, in this study, we propose using the remote sensing imagery of a sandy road (RSISR) dataset and design a sandy road extraction model (Parallel Attention Mechanism-Unet, or PAM-Unet) based on Gaofen-2 (GF-2) satellite images. Firstly, the model uses a residual stacking module, which can solve the problem of poor road feature consistency and improve the extraction of fine features. Secondly, we propose a parallel attention module (PAM), which can reduce the occlusion effect of foreign objects on roads during the extraction process and improve feature map reduction. Finally, with this model, the SASPP (Strip Atrous Spatial Pyramid Pooling) structure, which enhances the model's ability to perceive contextual information and capture banding features, is introduced at the end of the encoder. For this study, we conducted experiments on road extraction using the RSISR dataset and the DeepGlobe dataset. The final results show the following: (a) On the RSISR dataset, PAM-Unet achieves an IoU value of 0.762, and its F1 and IoU values are improved by 2.7% and 4.1%, respectively, compared to U-Net. In addition, compared to the models Unet++ and DeepLabv3+, PAM-Unet improves IoU metrics by 3.6% and 5.3%, respectively. (b) On the DeepGlobe dataset, the IoU value of PAM-Unet is 0.658; compared with the original U-Net, the F1 and IoU values are improved by 2.5% and 3.1%, respectively. The experimental results show that PAM-Unet has a positive impact by way of improving the continuity of sandy road extraction and reducing the occlusion of irrelevant features, and it is an accurate, reliable, and effective road extraction method.

Keywords: sandy road extraction; parallel attention mechanism; SASPP structure; GF-2 satellite imagery



Citation: Nie, Y.; An, K.; Chen, X.; Zhao, L.; Liu, W.; Wang, X.; Yu, Y.; Luo, W.; Li, K.; Zhang, Z. An Improved U-Net Network for Sandy Road Extraction from Remote Sensing Imagery. *Remote Sens.* **2023**, *15*, 4899. <https://doi.org/10.3390/rs15204899>

Academic Editor: John Trinder

Received: 31 July 2023

Revised: 6 October 2023

Accepted: 7 October 2023

Published: 10 October 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

Using remote sensing imagery to obtain road information has become a research hotspot. Along with the development of remote sensing technology, the resolution of remote sensing images has been continuously improved [1–3]. Sandy roads represent an indispensable and important part of fragile ecological environments, such as deserts and grasslands [4], and extracting sandy roads in field environments will be beneficial for field ecological inspections, field wind farm inspections, and path planning in complex environments [5,6]. However, due to intricate textural elements, occlusion effects, and low

contrast with surrounding objects, sandy roads are challenging to extract [7]. Deep learning technology is now growing and is widely employed in the feature extraction process for remote sensing images [8–10]. There is an ambition to use deep learning technology to extract sandy roads from remote sensing images, and the technology has some practical application value and potential in this field.

Previous researchers have studied many road extraction techniques, the most common of which are imagery-element-based extraction methods [11,12] and object-oriented road extraction methods [13,14]. Li et al. [15] used a watershed road segmentation algorithm based on threshold labeling, which utilizes the watershed transform to extract road information. Zhou et al. [16] used an object-oriented imagery analysis method to extract roads and achieved some results. As can be shown, the aforementioned techniques work well in certain application settings, but because they are constrained by artificial subjective thresholds, they cannot be generalized for road extraction tasks in multiple scenarios.

In terms of urban road extraction, given the need for navigation and urban planning, as well as the wide application of neural network technology in remote sensing information processing in recent years [17–19], the method for extracting urban roads using neural networks is gradually developing [20–22]. G. Zhou et al. [23] introduced a split depth-wise (DW) Separable Graph Convolutional Network (SGCN), tailored for addressing the challenges encountered in scenarios involving closed tarmac or tree-covered roads. The SGCN proficiently extracts road features and enhances noise interference mitigation throughout the road extraction process. Wang et al. [24] used a convolutional neural network to extract roads from high-resolution remote sensing images, which also optimized the road breaks appearing in the extraction results, and finally obtained complete extraction results for roads. Although deep learning methods have made some progress in road extraction, they still face some challenges. Scholars are now adopting some new approaches with which to improve their accuracy, among which a common strategy is to increase the depth of the network and introduce some feature enhancement modules through which to enhance the extraction ability of the model [25–27]. Jing P et al. [28] enhanced the network's learning ability by deepening the network, using a residual stacking module, and obtained better extraction results. Wu Q et al. [29] proposed a dense global spatial pyramid pooling module, based on Atrous spatial pyramid pooling, to obtain multi-scale information, which enhances the perception and aggregation of contextual information. Qi X et al. [30] proposed the AT replaceable module that incorporates different scale features; it utilizes the rich spatial and semantic information in remote sensing imagery to improve the extraction of roads. Liu et al. [31] introduced an improved asymmetric convolutional block (ACB)-based initiation structure, which extended low-level features in the feature extraction layer to reduce computational effort and obtain better results. It can be seen that the extraction methods based on convolutional neural networks have made great progress in urban road extraction.

The formation of sandy roads is influenced by various factors, including geography, weathering, hydrological processes, and human activities. These factors lead to the development of sandy roads with distinct characteristics. Despite their distinctive attributes, there is a paucity of datasets dedicated to sandy road extraction and relatively limited scholarly research in this domain. The extraction of sandy roads faces major limitations. While the U-Net network is commonly used for image segmentation tasks due to its unique encoding–decoding and skip-connection structure, as well as its advantages in training speed, inference speed, and accuracy, simply applying the original U-Net network is not well-suited to the task of sandy road extraction. The influencing factors are shown in Figure 1. (a) The boundary between vegetation and bare soil is ambiguous. (b) The road has poor continuity. (c) The road is obscured by extraneous features, such as trees and rocks. (d) Sandy roads can be easily mistaken for other features like rivers or canals. (e) The road spans a large scale and has a long-range banded structure [32,33]. Therefore, due to complex structural features interfering with road extraction, convolutional neural networks alone do not achieve the desired results.



Figure 1. Characteristics of sandy roads and challenges in extraction. These images illustrate the difficulties in the extraction process and the characteristics of sandy roads. The first row comprises representative original images from the sandy road dataset, and the second row consists of the resultant images obtained from the extraction of the U-Net network model. The areas outlined in red represent road segments that have not been correctly extracted, due to the following reasons: (a) Roads are confused with bare soil boundaries. (b) Roads have poor continuity and blurred sections in between. (c) Roads are obscured by trees. (d) Streams are incorrectly extracted as roads. (e) Roads contain distinctive banding features.

To address these issues, the following work was conducted in this study:

- (a) This study proposes a sandy road extraction model PAM-Unet based on an improved U-Net [34–37]. To address the issue of poor continuity in sandy roads, PAM-Unet employs stacked residual modules in the encoder section to enhance the model’s feature extraction capability. Meanwhile, at the end of the model encoder, the ASPP module proposed in the DeepLab series of models [38–41] is combined with the stripe pooling module [42] to better perceive the multi-scale features [43] and to adapt to the sandy roads’ long-range banded features. For the occlusion of other targets in the field environment, the parallel attention mechanism (PAM) is proposed and adopted in the feature fusion part of the process to enhance the reducibility of the feature map.
- (b) This study proposes the RSISR dataset, which covers a variety of complex sandy road scenarios including bare soil, grassland, forests, etc. For this dataset, 12,252 data samples were finally obtained. The construction of this dataset provides strong support and a reliable baseline for this study and analysis of sandy roads.
- (c) The PAM-Unet model was tested and analyzed several times on the RSISR dataset and DeepGlobe dataset, which proved that the PAM-Unet model is effective in terms of the extraction of qualitative roads and the improvement of modules. The results showed that the PAM-Unet achieved the ideal extraction results on the sandy road dataset, with an IoU value of 0.762, and obtained a high F1 value and recall, while on the DeepGlobe dataset, the results further demonstrated the positive effects of the model’s modules.

This paper is structured as follows: Section 1 briefly introduces the background of road extraction research, analyzes the difficulties of sandy road extraction, introduces the extraction method of sandy roads, and describes the innovations of this paper. Section 2 describes the proposed PAM-Unet model in detail. First, the overall structure of PAM-Unet is presented, followed by the parallel attention mechanism (PAM) with SASPP structure. Section 3 describes the construction of the sandy road dataset, the application of the DeepGlobe dataset, and the related setup of the experiment. Section 4 presents the results

of comparison and ablation experiments of the model on the dataset to demonstrate the feasibility and effectiveness of the model. The last two sections describe our investigation and summarization.

2. Research Methodology

2.1. Basic U-Net Structure

The U-Net network structure is a highly symmetrical end-to-end segmentation algorithm [44]. It consists of an encoder–decoder framework and hopping connections, which effectively realize the information transfer of low-level and high-level features. The overall structure of the network is shown in Figure 2.

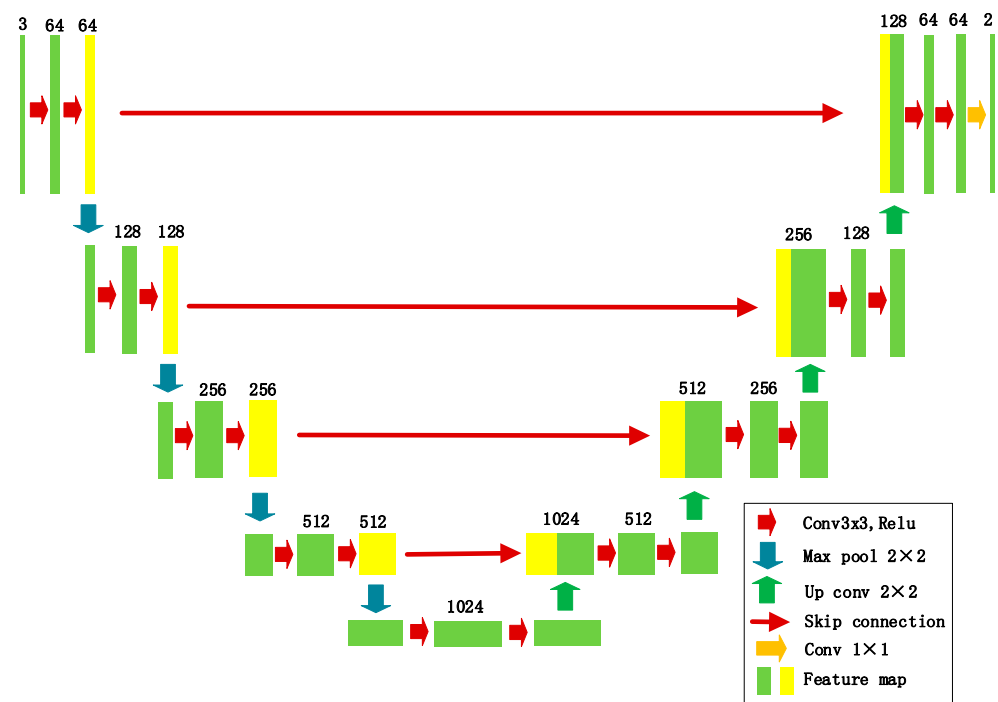


Figure 2. Basic U-Net structure.

In the encoder section, successive downsampling is required to complete the step-by-step extraction of features, which is activated by the Relu activation function after convolution, followed by the pooling operation using the maximum pooling layer with a step size of 2. The number of channels of the feature homography is increased by gradually decreasing the resolution of the feature map, to better obtain the global information. In the decoder section, to recover the features lost in the encoding network part, the decoder part gradually restores the feature maps to the original size, and two convolutional layers are used in the upsampling process. The feature maps are cascaded in each stage of the downsampling process, to recover the low-level semantic information. Finally, the number of channels is adjusted through the convolution to obtain the segmentation result.

2.2. PAM-Unet Structure

The PAM-Unet proposed in this study is an improvement of the U-Net structure, which is shown in Figure 3. In the encoder section, the model employs a stacked residual network to perform the initial extraction of sandy road features. In the feature fusion phase, the model first concatenates the feature maps in high and low dimensions to capture multi-dimensional features. Subsequently, the model incorporates the parallel attention mechanism (PAM) to enhance feature extraction in both channel and spatial dimensions. Finally, in the last stage of the encoder, the model integrates the SASPP module. This operation enhances the receptive field to strengthen contextual connections within the

model, reduces road fragmentation in segmentation results, and enhances the extraction of subtle roads.

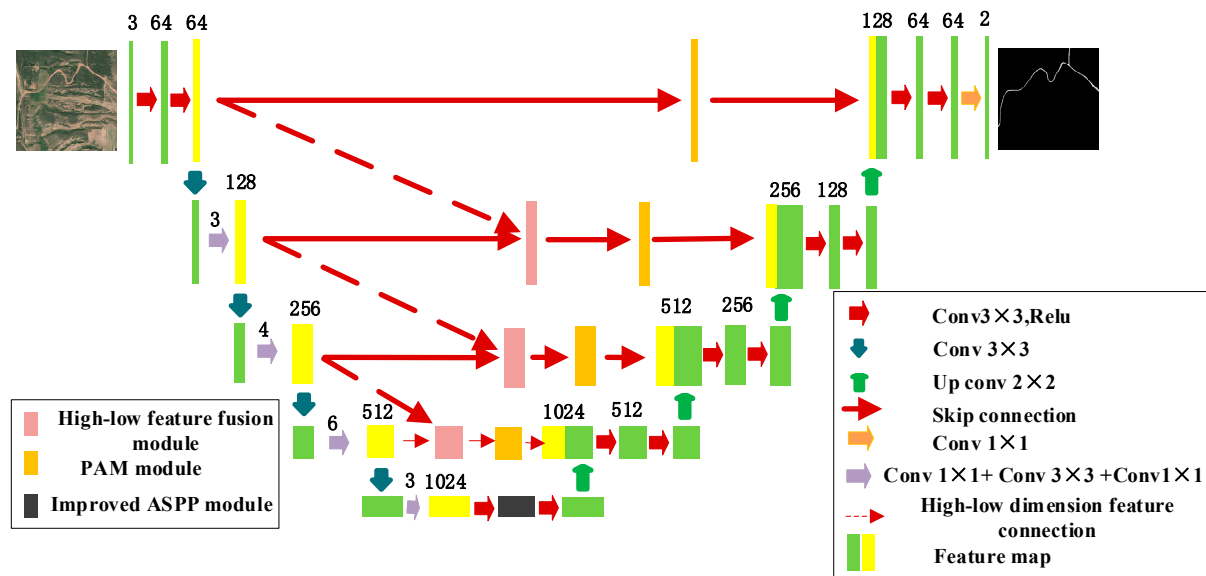


Figure 3. PAM-Unet network.

PAM-Unet is made up of three parts: encoder, feature fusion, and decoder. In the encoder section, we employ the distinctive residual stacking unit, known for its ability to preserve detailed target features effectively. This unit is integrated into the encoder component of the backbone extraction network, ensuring accurate feature extraction. Details of the overall structure of the stacked residual network used in the coding layer of the PAM-Unet model are shown in Table 1. In the feature fusion component of the model, the concatenation of high- and low-dimensional feature maps yields multi-dimensional features. After the fusion of these multi-dimensional features, the parallel attention module, referred to as PAM, is embedded. This module consists of both spatial attention and channel attention components. In these two components, global average pooling and max-pooling operations are separately applied to the input feature maps in the spatial dimension, while global average pooling is applied in the channel dimension. These operations compute weight values for both spatial and channel dimensions, which are used to fuse information with the original feature maps, preserving the integrity of road extraction. Towards the end of the backbone network, the ASPP (Atrous Spatial Pyramid Pooling) module is capable of obtaining multi-scale information through dilated convolutions, contributing to the acquisition of global features. However, due to the specific characteristics of road features, the adoption of square-shaped pooling kernels is insufficient for capturing the stripe-like features of roads. Therefore, the stripe pooling module is integrated into the ASPP structure to enhance the focus on road features and suppress background noise. In the decoder part, it takes the feature maps obtained from the previous two parts, performs feature fusion through dimensional orientation, and reduces them to the size of the original map. The SASPP (Spatial Attention Spatial Pyramid Pooling) structure plays a vital role in acquiring fine road details and improving road fragmentation. Meanwhile, the parallel attention mechanism (PAM) reconstructs road information while reducing the interference of irrelevant features. By integrating these three components, PAM-Unet enhances road extraction accuracy and reduces external noise interference, enabling the precise extraction of sandy roads.

Table 1. The detailed information of encoding layer structure.

Encoder	Type	Input	Output	Kernel	Stride	Padding
Encoder0	Conv	$512 \times 512 \times 3$	$256 \times 256 \times 64$	7	2	3
	max pooling	$256 \times 256 \times 64$	$128 \times 128 \times 64$	3	2	1
Encoder1	$3 \times \text{res_block}$	$128 \times 128 \times 64$	$128 \times 128 \times 128$	3	1	1
Encoder2	Conv	$128 \times 128 \times 128$	$64 \times 64 \times 256$	1	2	0
	$4 \times \text{res_block}$	$64 \times 64 \times 256$	$64 \times 64 \times 256$	3	1	1
Encoder3	Conv	$64 \times 64 \times 256$	$32 \times 32 \times 512$	1	2	0
	$6 \times \text{res_block}$	$32 \times 32 \times 512$	$32 \times 32 \times 512$	3	1	1
Encoder4	Conv	$32 \times 32 \times 512$	$16 \times 16 \times 1024$	1	2	0
	$3 \times \text{res_block}$	$16 \times 16 \times 1024$	$16 \times 16 \times 1024$	3	1	1

2.3. Parallel Attention Mechanism (PAM)

The downsampling network, comprised of stacked residual units, excels at finely extracting features. In the feature fusion phase, at each stage of the downsampling process, a unique feature map is generated. These feature maps possess distinct characteristics, and their diversity is enhanced by overlaying them [45,46]. After the feature maps are superimposed, the parallel attention mechanism (PAM) is added to the superimposed feature maps to enhance the attention to the target. This module obtains the global features by compressing the channels and enhancing the spatial information, which, in turn, improves the overall accuracy of the model.

The model adopts the feature fusion and attention mechanism enhancement module in Figure 4. The image is fused with high- and low-dimensional features to obtain a feature map of size $X \in R^{C \times H \times W}$ (C , H , and W representing the number of channels, length value, and width value of the feature map, respectively), which is fed into two parallel modules, where the spatial information extraction part is mainly for the enhancement of spatial dimensional information, while the channel side compresses the channel to extract the features and makes the feature map better for cross-channel interaction through the adaptive convolution kernel. In the spatial information extraction part, keeping the spatial dimension unchanged, two feature maps F_{max}^s and F_{avg}^s are obtained by max-pooling and average pooling of the feature map $X \in R^{C \times H \times W}$. Then, the dimension of the feature map is converted from $X \in R^{C \times H \times W}$ to $X \in R^{1 \times H \times W}$. Subsequently, the feature maps are concatenated, and the weights are obtained through the sigmoid activation function. The weights are multiplied with the input feature map to obtain the result of the spatial feature extraction part. The spatial attention calculation formula is shown in Equation (1):

$$M_F^S(F) = \sigma(f([AvgPool(F); Max(F)])) = \sigma\left(f\left(\left[F_{max}^s; F_{avg}^s\right]\right)\right) \quad (1)$$

In Equation (1), the sigmoid activation function is denoted by σ while the convolution process is denoted by f , F represents the feature map, and S represents spatial attention.

In the channel information compression and extraction part, the input feature map size is also $X \in R^{C \times H \times W}$. Firstly, the feature map is subjected to the flat pooling operation in the channel direction to obtain the feature map F_{avg}^c , and the size of the feature map is converted to $X \in R^{C \times 1 \times 1}$. Then, the dynamic convolution operation is carried out. Unlike ordinary channel attention, this module does not perform a dimensionality reduction operation but uses a 1D convolution of size 1 for fast implementation. This operation effectively captures cross-channel interaction information. Subsequently, the output feature map is passed through the sigmoid function to obtain the weights, and the weights are multiplied by the original feature map to obtain the result. The channel characteristic compression calculation formula is shown in Equation (2):

$$M_F^C(F) = \sigma(f([AvgPool(F)])) = \sigma\left(f\left(F_{avg}^c\right)\right) \quad (2)$$

In Equation (2), the sigmoid activation function is denoted by σ while the convolution process is denoted by f , $F1$ represents the feature map, and c represents channel attention. Finally, the feature maps, which have integrated information from both aspects, undergo an additional operation to obtain the enhanced feature map.

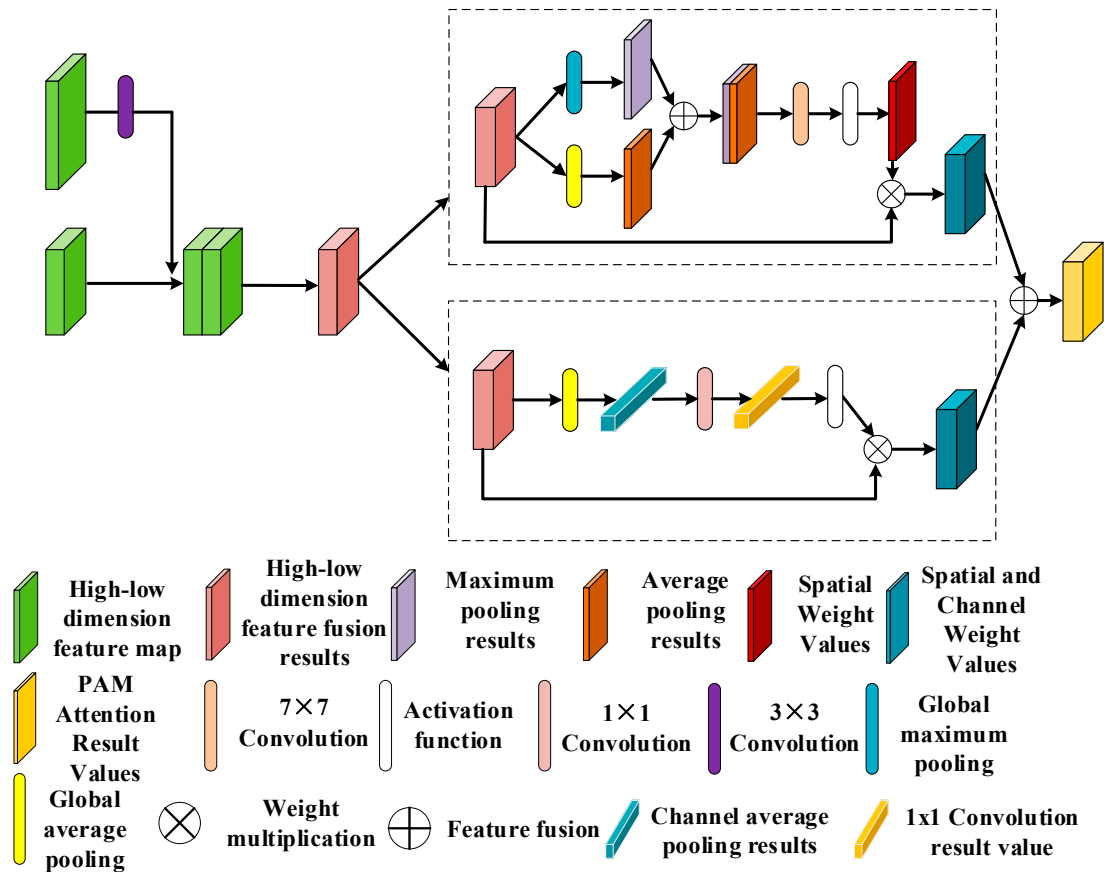


Figure 4. Parallel attention mechanism.

2.4. Improved ASPP Module

In 2015, Kaiming He et al. [47] first proposed the ASPP structure, which consists of multiple null convolutions with different expansion rates and a global average pooling module. Increasing the sensory field of the entire feature map without loss of resolution is achieved by superimposing null convolutions with different dilation rates. Therefore, we can focus on the multi-scale features of the target and later fuse the multi-scale feature maps. The original ASPP structure is shown in Figure 5.

Roads exhibit long-range banded features, and employing a large square pooling window would result in the inclusion of noise from irrelevant regions. Furthermore, due to the extensive span of road targets, long-range convolutions can weaken contextual connections between feature maps. The stripe pooling structure, which maintains a relatively narrow kernel shape in the spatial dimension, is capable of capturing long-distance relationships within isolated regions. The ASPP (Atrous Spatial Pyramid Pooling) structure enhances the receptive field, enabling the capture of multi-scale information regarding the overall target [48,49]. Effectively combining these two modules allows for the synthesis of advantages from both aspects. Therefore, in the overall model, this structure is introduced into ASPP to expand the sensing field while establishing long-distance relationships to extract the multi-scale features of the road.

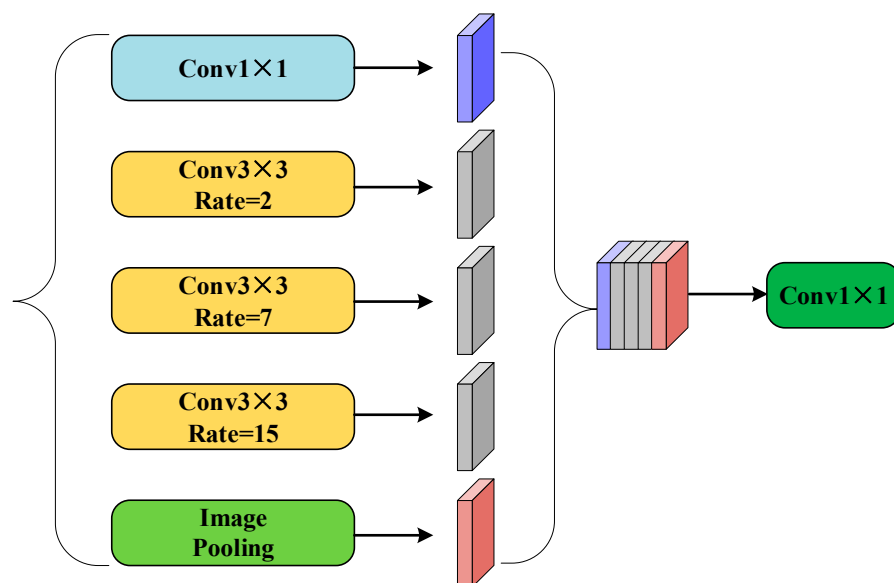


Figure 5. ASPP network architecture module.

The SASPP structure consists of six convolutional blocks parallel to each other, and the overall structure is shown in Figure 6. The number of channels in the first and last convolutional blocks remains unchanged, preserving the original information of the feature map. Meanwhile, the middle four convolutional blocks have 256 output channels each to enhance feature extraction capabilities. Since different dilation rates help the model to perceive features at different scales, in order to ensure the effectiveness of the combination with the bar pool, we slightly adjusted the value of r in the intermediate convolution block of SASPP in such a way that the adjustment balances the relationship between global context and local details. Finally, a cascade operation is used to fuse the features to obtain the performance of the image at different scales of the dilatation rate, and lastly, the cascaded feature maps are produced through 1×1 convolution to generate the feature maps.

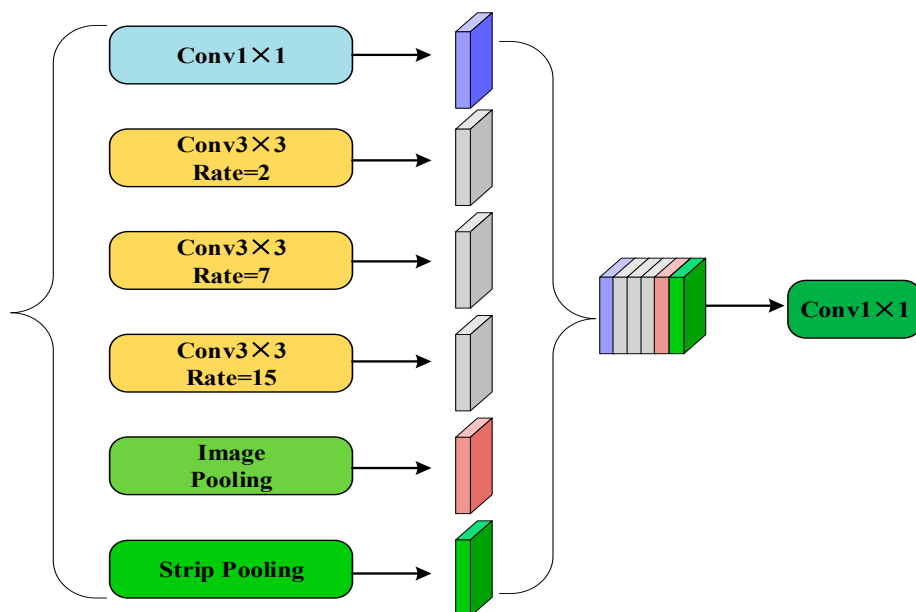


Figure 6. SASPP network architecture.

3. Dataset and Experimental Setup

3.1. Sandy Road Dataset Construction

For the experiments, high-resolution imagery from the GF-2 remote sensing satellite was utilized to create the RSISR dataset. Sandy road images with diverse and complex backgrounds spanning the timeframe of 2018 to 2022 were carefully selected. A series of processes were applied to the downloaded remote sensing images, including radiometric calibration [50], ortho-correction [51], atmospheric correction [52], image alignment and fusion [53], and so on. Then, the fused remote sensing imagerys were labeled with roads to obtain the true value of the roads, the label pixel value of the road samples was set to 1, and the background pixel value was set to 0. The imagerys were cropped with a step size of 512 to obtain 12,252 datasets, which were randomly divided into 10,415 training sets, 612 validation sets, and 1225 test sets. Some of the imagerys and labels in the training and test datasets are shown in Figures 7 and 8.

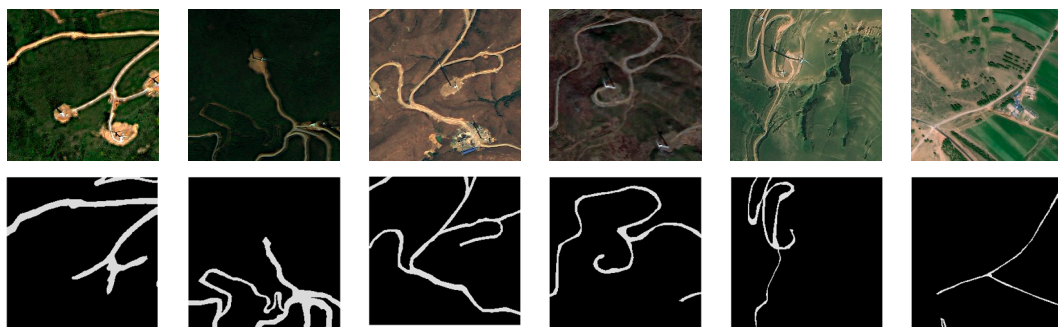


Figure 7. Some imagerys and labels from the training set.



Figure 8. Some imagerys and labels from the test set.

3.2. DeepGlobe Dataset

To validate the effectiveness of the improved model and test the generalization of the model, its performance on different datasets was evaluated. The experiment used the DeepGlobe dataset for further training and testing in addition to the sandy road dataset. The DeepGlobe dataset [54] is a widely used and highly regarded remote sensing imagery road dataset, which is specifically for road extraction and road network analysis studies. The dataset consists of 6226 aerial images, each of which has a size of 1024×1024 and a resolution of 0.5 m. Before using the DeepGlobe dataset, the dataset was pre-processed by cropping the imagery in step 512; based on the division ratio, this created 10,174 training sets, 598 validation sets, and 1196 test sets.

For the sample imagerys and labels, the DeepGlobe dataset demonstrated accurate labeling information for roads. Examples of imagerys and labels are shown in Figure 9. These imagerys were further used for training and testing to validate the model's ability to extract roads from real scenes.

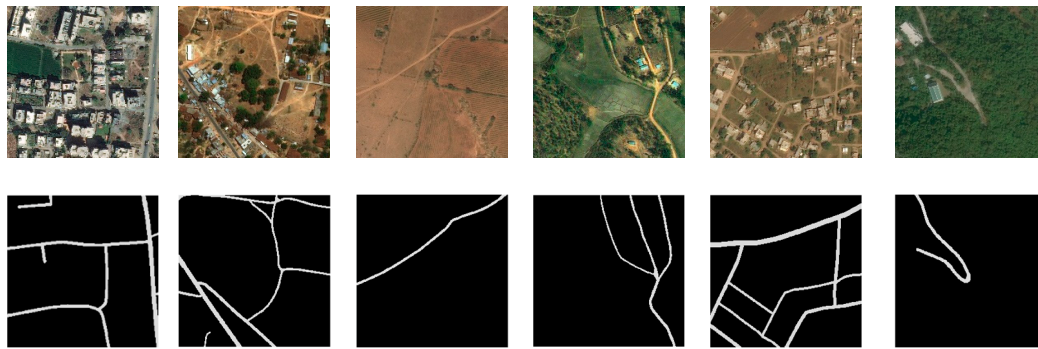


Figure 9. DeepGlobe dataset imagery and labeling.

3.3. Experimental Setup

To validate the feasibility and superior performance of PAM-Unet in sandy road extraction, in this study, we conducted comparative experiments using the same set of training and testing samples with other relevant models. We used the same set of training samples and test samples for comparison experiments, where the experimental computer operating system was Windows 10 Professional, the CPU configuration was Intel(R) Xeon(R) Silver 4214, the memory capacity of the workstation was 256G, the graphics card model was NVIDIA Quadro RTX 5000, the video memory was 16G, the CUDA version was 11.1, and the PyTorch version was 1.7.0. The experimental parameters are shown in Table 2.

Table 2. Experimental parameters.

Name of the Parameter	Parameter Value
learning rate	0.001
optimizer	Adam
loss function	binary cross entropy
batch size	8
epochs	150
weight_decay	0.0005
momentum	0.9
gamma	0.98

Since road extraction is a binary classification problem, a binary cross-entropy loss function was chosen, and the mathematical expression for this loss function is shown in Equation (3), where $p(x)$ is the distribution of the true value of the road, $q(x)$ is the distribution of the predicted value of the road, and i represents the pixel position.

$$H(p, q) = \sum_{i=1}^n -p(x_i) \log(q(x_i)) \quad (3)$$

3.4. Evaluation Indicators

In the field of semantic segmentation, extracting sandy roads from remote sensing imagery is a binary classification task, where roads are positive samples and backgrounds are negative samples. In the prediction process, all the prediction results can be classified into four categories: True Positive (TP) denotes the number of pixels correctly classified as roads, True Negative (TN) denotes the number of pixels correctly classified as background, False Positive (FP) indicates the number of background pixels that are misclassified as roads, and False Negative (FN) indicates the number of road pixels that are misclassified as background. To evaluate the performance of the model, the following metrics were used in this experiment: precision, recall, F1 score, and average intersection–union ratio (IoU). Their formulas are shown in (4)–(7).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (7)$$

In terms of evaluation metrics, the accuracy ratio measures the proportion of all correctly extracted road pixels among all predicted road pixels, while the recall ratio indicates the proportion of all correctly extracted road pixels among the actual labeled road pixels. The F1 score is the reconciled average of the accuracy and recall ratios, which is used to comprehensively assess the accuracy and completeness of the road extraction. The intersection and concatenation ratio, on the other hand, is a metric that comprehensively assesses the overall quality of road extraction by dividing the number of intersecting pixels by the number of concatenated pixels, thus making the assessment more comprehensive.

4. Experimental Results and Analysis

This section discusses the ablation experiment for building the network, which was experimented with and analyzed on the homemade sandy road dataset RSISR as well as on the public dataset DeepGlobe.

4.1. Road Extraction Results and Experiments on RSISR Dataset

The model trained on the sandy road dataset was used for the prediction and analysis of the results, where the extraction target was the sandy roads in the imagery. Comparison experiments were conducted using four semantic segmentation models: PAM-Unet, U-Net, Deeplabv3+, and Unet++. Figure 10 shows the results of the comparison of the current experiments, and the results of the comparison of the test set are shown in the following order: the test set original imagery, the groundtruth, the U-Net result map, the Deeplabv3+ result map, the Unet++ result map, the D-LinkNet result map, and the PAM-Unet result map. The circled portion in the figure represents a clearly distinguishable region for assessing the quality of sandy road extraction results. For the U-Net model, the performance of the prediction results shows that the extraction is relatively poor and the continuity of the sandy roads is poorly maintained. For the Deeplabv3+ model, there is a relative improvement in road continuity, but there are still more road breaks. For the Unet++ model, the extraction of certain major roads is ignored. D-LinkNet does not extract well in the face of finer road features and produces many broken patches. For the PAM-Unet model, the synthesis of the extraction is optimal concerning the other network models, and the extraction results obtained are relatively complete. The manifestations of this are mainly in two areas: Firstly, when dealing with roads obscured by complex backgrounds, PAM-Unet extracts complete sandy roads relatively well, enhances the integrity of roadway extraction, relatively effectively avoids shading by trees, and reduces the extraction errors in the extraction result maps. This improvement is mainly attributed to the enhanced ability of the parallel attention mechanism to extract road information, thus improving the overall performance of road extraction. Secondly, the present model also shows better results in extracting slender sandy roads. The improved ASPP structure enhances the connection between contexts and improves the ability to capture subtle features of the target, which leads to better continuity of the extracted sandy roads, and thus these improvements enhance the performance of the model to a certain extent.

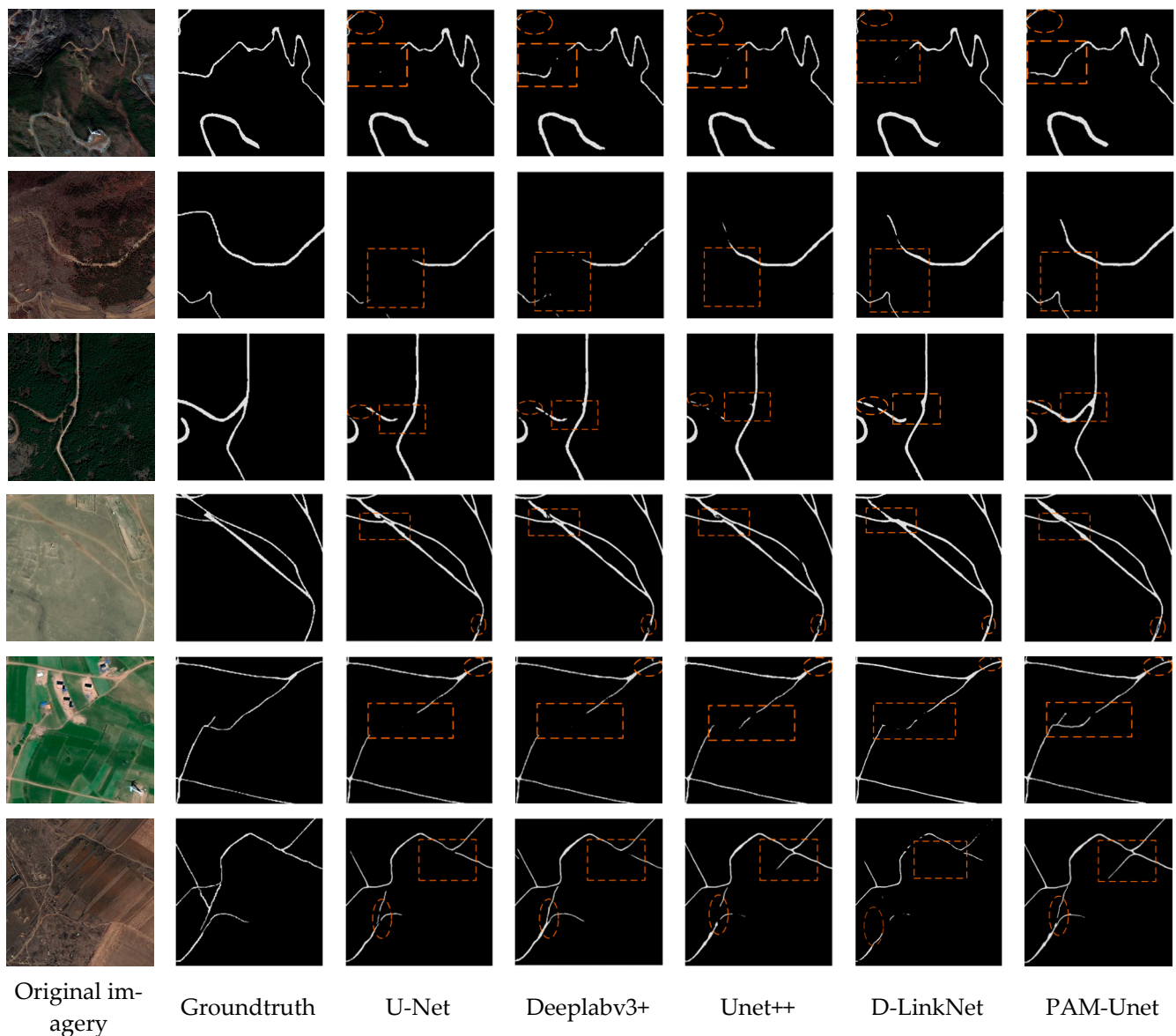


Figure 10. Comparison of model extraction results on the RSISR dataset.

Firstly, through comparative experiments, the PAM-Unet model was tested several times on the test set with other semantic segmentation network models. Table 3 demonstrates the metric values obtained from the evaluation tests for IoU, precision, recall, and F1 score. The metrics' comparison values with the related models show that the proposed PAM-Unet model produces improvements of 0.036, 0.053, and 0.041 in the IoU metric and 0.024, 0.035, and 0.027 in the F1 score metric. Since the parallel attention mechanism with the SASPP module extracts more detailed features and is more complex, the time to infer an image is 70 s, which is longer than that of the other models, but we slightly increase the running time while obtaining relatively higher accuracy. When comparing the data, it is clear that the PAM-Unet model has higher extraction accuracy for sandy roads and outperforms the other semantic segmentation models in the comparison experiments in terms of extraction results.

Table 3. Statistics of extraction results of PAM-Unet and comparison models on RSISR dataset.

Method	IoU	Precision	Recall	F1 Score	Running Time
Unet++	0.726	0.832	0.851	0.841	70 s
Deeplabv3+	0.709	0.814	0.847	0.830	68 s
Unet	0.721	0.831	0.844	0.838	66 s
D-LinkNet	0.721	0.822	0.854	0.838	67 s
PAM-Unet	0.762	0.863	0.868	0.865	72 s

Secondly, to verify the effectiveness of the model improvement, we conducted ablation tests of each part of the PAM-Unet model on the test set, and the results are shown in Table 4. PA-Unet incorporates the parallel attention mechanism (PAM) in the feature fusion part on the constructed Unet network, AS-Unet incorporates the SASPP structure at the end of the encoder in the Unet network, and PAM-Unet incorporates both of the above-improved structures. In Table 2, it can be seen that PAM-Unet performs optimally in terms of IoU indicator values, indicating that with the simultaneous incorporation of the parallel attention mechanism and the improvement of the SASPP structure, the parallel attention mechanism has been enhanced in terms of focusing on the target, while the SASPP focuses on the multi-scale features and establishes a long-distance linkage. Furthermore, the two modules are used jointly, thereby enhancing the overall performance of the model for sandy road extraction. In terms of testing time, the model's testing time is slightly longer than that of PA-Unet and AS-Unet due to the reduction in the number of parameters resulting from the simultaneous removal of either of the two modules, but an overall improvement in accuracy is obtained.

Table 4. Statistics of PAM-Unet and ablation model extraction results on RSISR dataset.

Method	IoU	Precision	Recall	F1 Score	Running Time
Unet	0.721	0.831	0.844	0.838	66 s
PA-Unet	0.746	0.851	0.858	0.854	67 s
AS-Unet	0.733	0.843	0.849	0.846	69 s
PAM-Unet	0.762	0.863	0.868	0.865	72 s

4.2. Road Extraction Results and Experiments on DeepGlobe Dataset

To verify the effectiveness of each module in the model improvement, the PAM-Unet model was further tested on the DeepGlobe dataset. Compared to the sandy road dataset, there are fewer rural, suburban, and bare land samples and more urban samples. The experimental results are shown in Figure 11; from left to right, they are the original imageries of the test set, the groundtruth, the original U-Net result map, the PA-Unet result map, the AS-Unet result map, and the PAM-Unet result map, respectively. The circled parts of the figure are the areas where the contrast is more pronounced. From the experimental images, we surmised that PA-Unet maintains the continuity of the overall road extraction better, but ignores some detailed information; and AS-Unet is excellent for the global extraction of the road, but the continuity of the road extraction fails to meet the ideal requirements. The PAM-Unet model synthesizes the extraction effect better, which is mainly reflected in the following aspects: Firstly, PAM-Unet effectively maintains the overall shape of the road, extracts the continuity of the road better, and recognizes the road more completely in the case of shade of trees and buildings, as well as in the confusion of grass and urban buildings. Secondly, for small features, its better extraction ability compared to other models improves the consistency of road features, the model has a stronger focus for some small roads and inconspicuous roads, and PAM-Unet is better in terms of the comprehensive extraction of the road.

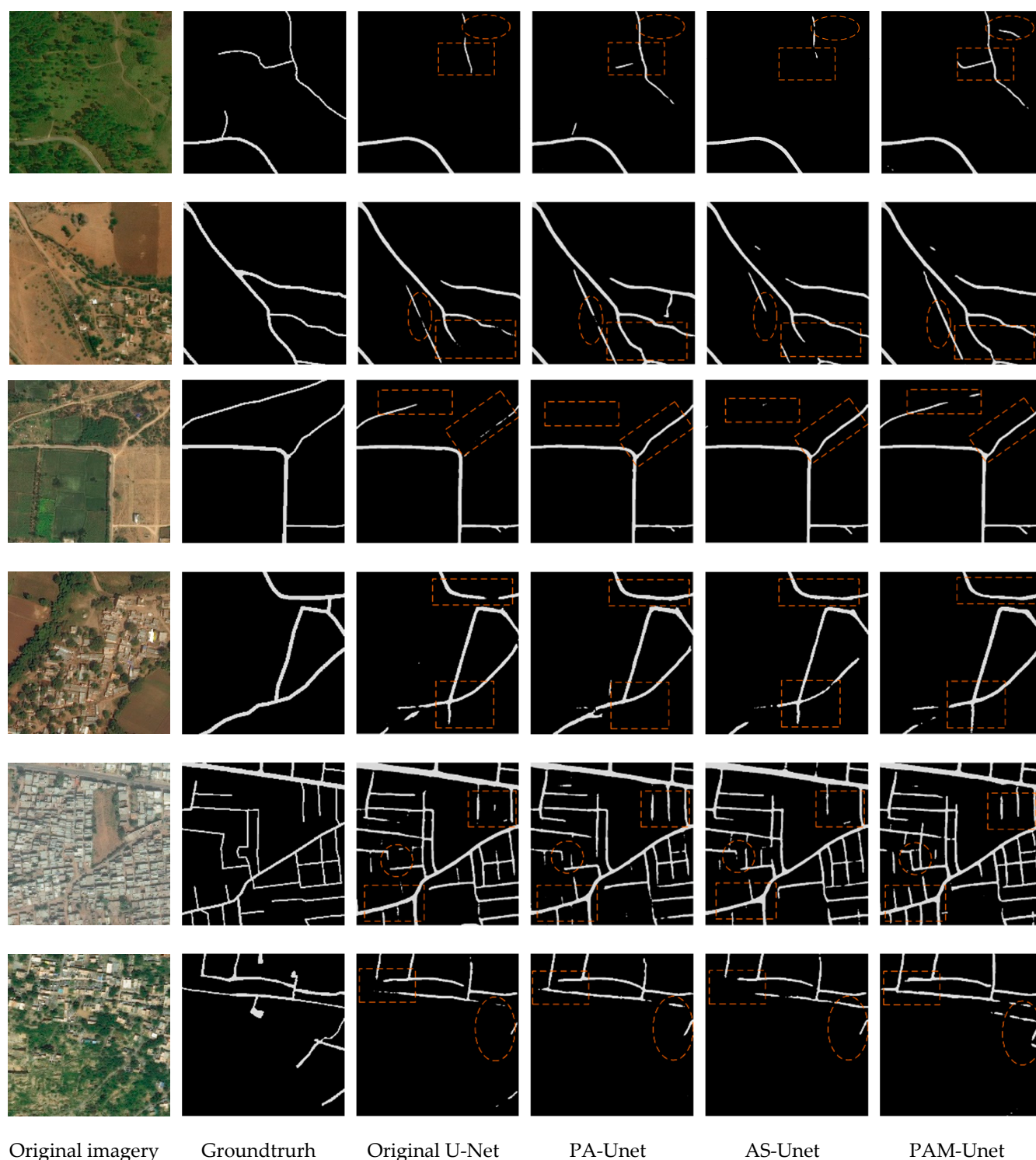


Figure 11. Comparison of model extraction results on the DeepGlobe dataset.

Table 5 shows the relevant metrics obtained with several models when compared on the DeepGlobe dataset. From the metrics, it can be seen that PAM-Unet improves by 0.031 on the IoU metric and 0.025 on the F1 metric compared to the original U-Net. The model also improves the relevant evaluation metrics compared to PA-Unet and AS-Unet. Taken together, the improved model improves in accuracy relative to the original model, and the evaluation metrics reflect the effectiveness of the improved modules, with several improved modules working with each other to obtain excellent metric values.

Table 5. Statistics of PAM-Unet and ablation model extraction results on the DeepGlobe dataset.

Method	IoU	Precision	Recall	F1 Score	Running Time
U-Net	0.627	0.758	0.781	0.769	64 s
PA-Unet	0.651	0.789	0.789	0.789	65 s
AS-Unet	0.644	0.780	0.787	0.783	67 s
PAM-Unet	0.658	0.799	0.789	0.794	70 s

In comparative experiments, the PAM-Unet model was tested several times on the test set for its performance compared to other semantic segmentation network models. Table 6 demonstrates the metric values obtained from the evaluation tests in IoU, precision, recall, and F1 score. The results on the DeepGlobe test set were 0.658, 0.799, 0.789, and 0.794, respectively. The IoU value was improved by 3.1% compared to the original model U-Net, which was a significant improvement compared to networks such as Deeplabv3+ and Unet++. This was achieved because our model incorporates the parallel attention mechanism (PAM) and SASPP module. These enhance the focus on the target and the perception of the extent and shape of sandy roads, respectively, thereby improving the extraction of small details of road targets. Together, they enhance the model's generalizability and improve its image segmentation performance.

Table 6. Statistics of extraction results of PAM-Unet and comparison models on DeepGlobe dataset.

Method	IoU	Precision	Recall	F1 Score	Running Time
Unet++	0.643	0.791	0.774	0.782	68 s
Deeplabv3+	0.636	0.770	0.786	0.778	66 s
U-Net	0.627	0.758	0.781	0.769	64 s
SegNet	0.629	0.765	0.779	0.771	65 s
PAM-Unet	0.658	0.799	0.789	0.794	70 s

5. Discussion

The results from the ablation experiments highlight the effectiveness of the introduced parallel attention mechanism (PAM) in enhancing road continuity, consistency, and completeness. The SASPP module, on the other hand, is better adapted to the banding characteristics of roads as well as increasing the sensory field during the extraction process. The use of stacked residual networks, on the other hand, ensures the accuracy of the whole extraction process and lays the foundation for the extraction of sandy roads. In this study, stacked residual networks were used as the basis and these two feature enhancement modules were introduced into the model to construct the PAM-Unet network model, with which better experimental results were achieved. In this regard, we believe that PAM-Unet is valuable for field ecological patrols and path planning. In the comparison experiments, PAM-Unet was tested against other semantic segmentation models using the RSISR dataset. The IoU value obtained with PAM-Unet was 0.762, which was higher than the other semantic segmentation models (evaluation values shown in Table 3). Figure 10 demonstrates a comparison of the effect of this model against those of the other models on the test set, from which it can be seen that the extraction of PAM-Unet for occluded roads and subtle road features is enhanced compared to the extraction of other models. This is thanks to the dual application of the parallel attention mechanism (PAM) and the SASPP structure, which yields relatively good results. To verify the effectiveness of the improvement, an ablation test of each module was carried out on the DeepGlobe dataset. It can be seen in Table 5 that the optimal result was obtained after adding both the SASPP structure and the parallel attention mechanism, with an IoU value of 0.658. Meanwhile, the IoU value of adding only the parallel attention mechanism was 0.651, and the IoU value of adding only the SASPP structure was 0.644. This proves the optimization of the results obtained by adding the two modules at the same time and justifies the interplay between the two modules. Meanwhile, from Figure 11, it can be seen that the improved model can extract the roads more

completely and outperforms the other models in terms of comprehensive performance. In order to further validate, compare, and substantiate the performance differences with the improved method, we conducted a statistical hypothesis test using Student's *t*-test on multiple sets of experimental values obtained during the experiments with PAM-Unet and U-Net. We selected IoU and F1 score as the key performance indicators. The calculation formula for the *t*-test is shown in Equation (8), aimed at ascertaining whether there was a statistically significant difference in performance between the improved model and the original model.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (8)$$

in which $\bar{X}_1$ and $\bar{X}_2$ are the means of the two sample groups, S_1 and S_2 are the standard deviations of the two sample groups, and n_1 and n_2 are the sample sizes of the two sample groups.

In this analysis, we focused on the *t*-statistic values and corresponding *p*-values for IoU and F1 score. For IoU, the *t*-statistic value was 2.158, and the *p*-value was 0.043. Based on a significance level of 0.05, the *p*-value was less than the significance level, indicating that we can reject the null hypothesis. This demonstrates that our improved algorithm shows a statistically significant difference in IoU compared to the original model. Regarding F1 score, the *t*-statistic value was 2.605, and the *p*-value was 0.017. Similarly, based on a significance level of 0.05, the *p*-value was less than the significance level, leading to the rejection of the null hypothesis. In the case of F1 score, we also found a statistically significant difference between the two algorithms. Considering the results above, for both the statistical IoU and F1 score values, the PAM-Unet model exhibits a significant difference compared to the Unet model across multiple statistical tests. The results indicate that PAM-Unet outperformed the original model in multiple statistical tests. The hypothesis-testing statistical experiments further enhance the credibility of our improvements, demonstrating the improved model's performance advantages and stronger generalizability across multiple applications. We hope that our improved model design will inspire future researchers to advance the work on sandy road extraction, ultimately contributing to more precise sandy road extraction in the future.

In summary, compared to traditional manual road extraction methods, sandy road extraction through semantic segmentation offers the advantage of accurately identifying sandy roads in complex environments. This capability proves valuable for tasks like vehicle navigation, path planning, and ecological patrols in challenging landscapes, making it highly adaptable. However, our model suffers from some limitations when facing unfriendly geographic environments as well as bad weather in real-world applications, and it remains a challenge to continue to improve the accuracy of model recognition. Future research should focus more on cross-data source fusion and real-time semantic segmentation, and investigate how to integrate different data sources to improve the robustness of road extraction. We also propose developing efficient semantic segmentation models that can run in real-time environments. Furthermore, researchers may investigate semi-supervised and unsupervised learning methods to reduce the reliance on large amounts of labeled data. Finally, there is research scope to combine semantic segmentation with other perceptual modalities (e.g., object detection and speech recognition) to improve the comprehensiveness of scene understanding.

6. Conclusions

The main aim of this study was to address the problem of extraction of sandy roads. We firstly established a sandy road extraction dataset RSISR and secondly improved on the U-Net network by proposing the PAM-Unet. Compared to the baseline U-Net network, our model yields significantly improved prediction results. Notably, PAM-Unet excels in enhancing the extraction of fine sandy road features, improving road continuity, and reducing the interference from extraneous features. The achieved results align with our expectations.

From the results, it can be seen that PAM-Unet improves the completeness and continuity of the road extraction, which makes the sandy road extraction result clearer and more holistic. Comparison experiments, on the other hand, verify that the model outperforms other semantic segmentation models, and ablation experiments demonstrate the usefulness of the individual improvement modules. In summary, the challenges in sandy road extraction can be effectively dealt with by reasonably designing the network structure and feature fusion method, which effectively improves the accuracy and robustness of road extraction from remote sensing images. However, some limitations of the model still exist, which relate to the size and diversity of the dataset and the rationalization of the combination of network model structures. Future research can explore these issues to further improve the performance and generalizability of road extraction, and it is hoped that the research results in this paper will provide useful references and insights for further research and applications in the field of sandy road extraction from remote sensing imageries.

Author Contributions: Y.N., X.C. and K.A. designed the project; X.C. and L.Z. provided the image and experimental resources; K.A., Y.Y. and W.L. (Wenyi Luo) performed the dataset-related production; K.A. wrote the paper and performed the related experiments; X.W., W.L. (Wantao Liu) and K.A. revised the manuscript; K.L. and Z.Z. supervised the experiment. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the National Natural Science Foundation of China (Grant No. 41101426), Natural Science Foundation of Jiangxi Province (Grant No. 20202BABL202040), National Key Research and Development Program (Grant No. 2019YFE0126600) and 03 Special Project and 5G Project of the Science and Technology Department of Jiangxi Province (Grant No. 20212ABC03A03). The authors would like to thank the handling editor and anonymous reviewers for their valuable comments.

Data Availability Statement: Not applicable.

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

References

1. Li, S.F.; Liao, C.; Ding, Y.L.; Hu, H.; Jia, Y.; Chen, M.; Xu, B.; Ge, X.M.; Liu, T.Y.; Wu, D. Cascaded residual attention enhanced road extraction from remote sensing images. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 9. [\[CrossRef\]](#)
2. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sens.* **2020**, *12*, 1444. [\[CrossRef\]](#)
3. Chen, Z.Y.; Deng, L.A.; Luo, Y.H.; Li, D.L.; Junior, J.M.; Goncalves, W.N.; Nurunnabi, A.A.M.; Li, J.; Wang, C.; Li, D. Road extraction in remote sensing data: A survey. *Int. J. Appl. Earth Obs.* **2022**, *112*, 102833. [\[CrossRef\]](#)
4. Zhao, Y.Y.; Zhang, Y.; Yuan, M.T.; Yang, M.; Deng, J.Y. Estimation of initiation thresholds and soil loss from gully erosion on unpaved roads on China's Loess Plateau. *Earth Surf. Proc. Land* **2021**, *46*, 1713–1724. [\[CrossRef\]](#)
5. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **2020**, *37*, 362–386. [\[CrossRef\]](#)
6. Kiran, B.R.; Sobh, I.; Talpaert, V.; Mannion, P.; Sallab, A.A.A.; Yogamani, S.; Pérez, P. Deep reinforcement learning for autonomous driving: A survey. *IEEE Trans. Intell. Transp.* **2021**, *23*, 4909–4926. [\[CrossRef\]](#)
7. Li, C.K.; Zeng, Q.G.; Fang, J.; Wu, N.; Wu, K.H. Road extraction in rural areas from high resolution remote sensing image using an improved Full Convolution Network. *Nat. Remote Sens. Bull.* **2021**, *25*, 1978–1988. [\[CrossRef\]](#)
8. Shamsolmoali, P.; Zareapoor, M.; Zhou, H.Y.; Wang, R.L.; Yang, J. Road segmentation for remote sensing images using adversarial spatial pyramid networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4673–4688. [\[CrossRef\]](#)
9. Wei, Y.; Ji, S. Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5602312. [\[CrossRef\]](#)
10. Hou, Y.W.; Liu, Z.Y.; Zhang, T.; Li, Y.J. C-UNet: Complement UNet for remote sensing road extraction. *Sensors* **2021**, *21*, 2153. [\[CrossRef\]](#)
11. Zhu, Q.Q.; Zhang, Y.N.; Wang, L.Z.; Zhong, Y.F.; Guan, Q.F.; Lu, X.Y.; Zhang, L.P.; Li, D. A global context-aware and batch-independent network for road extraction from VHR satellite imagery. *ISPRS J. Photogramm.* **2021**, *175*, 353–365. [\[CrossRef\]](#)
12. Lin, S.F.; Yao, X.; Liu, X.L.; Wang, S.H.; Chen, H.M.; Ding, L.; Zhang, J.; Chen, G.H.; Mei, Q. MS-AGAN: Road Extraction via Multi-Scale Information Fusion and Asymmetric Generative Adversarial Networks from High-Resolution Remote Sensing Images under Complex Backgrounds. *Remote Sens.* **2023**, *15*, 3367. [\[CrossRef\]](#)
13. Zhang, Y.C.; Yang, H.L.; Xin, Z.B.; Lu, L.L. Extraction of Small Watershed Terraces in the Hilly Areas of Loess Plateau through UAV Images with Object-oriented Approach. *J. Soil Water Conserv.* **2023**, *37*, 139–146.

14. Sun, Y.X.; Zhou, L.P.; Pi, Y.Z. Object-oriented classification based on high resolution UCE aerial images. *Survey World* **2022**, *1*, 55–58.
15. Li, Y.; Han, X.; Ma, H.; Deng, L.; Sun, Y. Road image segmentation based on threshold watershed algorithm. *J. Nonlinear Convex A* **2019**, *20*, 1453–1463.
16. Zhou, A.X.; Yu, L.; Feng, J.; Zhang, X.Y. Road Information Extraction from High-Resolution Remote Sensing Image Based on Object-Oriented Image Analysis Method. *Geomat. Spat. Inf. Technol.* **2017**, *40*, 1–4.
17. Wu, Z.H.; Pan, S.R.; Chen, F.W.; Long, G.D.; Zhang, C.Q.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [\[CrossRef\]](#)
18. Goswami, A.; Sharma, D.; Mathuku, H.; Gangadharan, S.M.P.; Yadav, C.S.; Sahu, S.K.; Pradhan, M.K.; Singh, J.; Imran, H. Change detection in remote sensing image data comparing algebraic and machine learning methods. *Electronics* **2022**, *11*, 431. [\[CrossRef\]](#)
19. Wang, G.H.; Li, B.; Zhang, T.; Zhang, S.B. A network combining a transformer and a convolutional neural network for remote sensing image change detection. *Remote Sens.* **2022**, *14*, 2228. [\[CrossRef\]](#)
20. Feng, J.N.; Jiang, Q.; Tseng, C.H.; Jin, X.; Liu, L.; Zhou, W.; Yao, S.W. A deep multitask convolutional neural network for remote sensing image super-resolution and colorization. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5407915. [\[CrossRef\]](#)
21. Cheng, X.J.; He, X.H.; Qiao, M.J.; Li, P.L.; Hu, S.K.; Chang, P.; Tian, Z.H. Enhanced contextual representation with deep neural networks for land cover classification based on remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102706. [\[CrossRef\]](#)
22. Abdollahi, A.; Pradhan, B.; Shukla, N. Road extraction from high-resolution orthophoto images using convolutional neural network. *J. Indian Soc. Remote* **2021**, *49*, 569–583. [\[CrossRef\]](#)
23. Zhou, G.D.; Chen, W.T.; Gui, Q.S.; Li, X.J.; Wang, L.Z. Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5614115. [\[CrossRef\]](#)
24. Wang, B.; Chen, Z.L.; Wu, L.; Xie, P.; Fan, D.L.; Fu, B.L. Road extraction of high-resolution satellite remote sensing images in U-Net network with consideration of connectivity. *Remote Sens. Bull.* **2020**, *24*, 1488–1499. [\[CrossRef\]](#)
25. Xie, T.Y.; Liang, X.W.; Xu, S. Research on Urban Road Extraction Method Based on Improved U-Net Network. *Comput. Digit. Eng.* **2023**, *51*, 650–656.
26. Zhang, K.; Chen, Z.J.; Qiao, D.; Zhang, Y. Real-Time Image Detection via Remote Sensing Based on Receptive Field and Feature Enhancement. *Laser Optoelectron. Prog.* **2023**, *60*, 0228001-1–0228001-10.
27. Li, M.; Hsu, W.; Xie, X.D.; Cong, J.; Gao, W. SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. *IEEE Trans. Med. Imaging* **2020**, *39*, 2289–2301. [\[CrossRef\]](#)
28. Jing, P.; Yu, H.Y.; Hua, Z.H.; Xie, S.F.; Song, C.Y. Road Crack Detection Using Deep Neural Network Based on Attention Mechanism and Residual Structure. *IEEE Access* **2022**, *11*, 919–929. [\[CrossRef\]](#)
29. Wu, Q.Q.; Luo, F.; Wu, P.H.; Wang, B.; Yang, H.; Wu, Y.L. Automatic road extraction from high-resolution remote sensing images using a method based on densely connected spatial feature-enhanced pyramid. *IEEE J. Sel. Top Appl. Earth Obs. Remote Sens.* **2020**, *14*, 3–17. [\[CrossRef\]](#)
30. Qi, X.Q.; Li, K.Q.; Liu, P.K.; Zhou, X.G.; Sun, M.Y. Deep attention and multi-scale networks for accurate remote sensing image segmentation. *IEEE Access* **2020**, *8*, 146627–146639. [\[CrossRef\]](#)
31. Liu, B.; Ding, J.L.; Zou, J.; Wang, J.J.; Huang, S.A. LDANet: A Lightweight Dynamic Addition Network for Rural Road Extraction from Remote Sensing Images. *Remote Sens.* **2023**, *15*, 1829. [\[CrossRef\]](#)
32. Luo, Y.B.; Chen, J.X.; Shi, Z.; Li, J.Z.; Liu, W.W. Mechanical characteristics of primary support of large span loess highway tunnel: A case study in Shaanxi Province, Loess Plateau, NW China primary. *Tunn. Undergr. Space Technol.* **2020**, *104*, 103532. [\[CrossRef\]](#)
33. Diaz-Gonzalez, F.A.; Vuelvas, J.; Correa, C.A.; Vallejo, V.E.; Patino, D. Machine learning and remote sensing techniques applied to estimate soil indicators—Review. *Ecol. Indic.* **2022**, *135*, 108517. [\[CrossRef\]](#)
34. Liu, W.Q.; Wang, C.; Zang, Y.; Hu, Q.; Yu, S.S.; Lai, B.Q. A survey on information extraction technology based on remote sensing big data. *Big Data Res.* **2022**, *8*, 28–57.
35. Kong, J.Y.; Zhang, H.S. Improved U-Net network and its application of road extraction in remote sensing image. *Chin. Space Sci. Techn.* **2022**, *42*, 105.
36. Xu, L.L.; Liu, Y.J.; Yang, P.; Chen, H.; Zhang, H.Y.; Wang, D.; Zhang, X. HA U-Net: Improved model for building extraction from high resolution remote sensing imagery. *IEEE Access* **2021**, *9*, 101972–101984. [\[CrossRef\]](#)
37. Ma, Y. Research Review of Image Semantic Segmentation Method in High-Resolution Remote Sensing Image Interpretation. *J. Front. Comput. Sci. Technol.* **2023**, *1*, 4200153.
38. He, C.; Liu, Y.L.; Wang, D.C.; Liu, S.F.; Yu, L.J.; Ren, Y.H. Automatic Extraction of Bare Soil Land from High-Resolution Remote Sensing Images Based on Semantic Segmentation with Deep Learning. *Remote Sens.* **2023**, *15*, 1646. [\[CrossRef\]](#)
39. Wang, J.X.; Feng, Z.X.; Jiang, Y.; Yang, S.Y.; Meng, H.X. Orientation Attention Network for semantic segmentation of remote sensing images. *Knowl. Based Syst.* **2023**, *267*, 110415. [\[CrossRef\]](#)
40. Mahmud, M.N.; Azim, M.H.; Hisham, M.; Osman, M.K.; Ismail, A.P.; Ahmad, F.; Ahmad, K.A.; Ibrahim, A.; Rabiani, A.H. Altitude Analysis of Road Segmentation from UAV Images with DeepLab V3+. In Proceedings of the 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 21–22 October 2022; pp. 219–223.

41. Yan, Y.T.; Gao, Y.; Shao, L.W.; Yu, L.Q.; Zeng, W.T. Cultivated land recognition from remote sensing images based on improved deeplabv3 model. In Proceedings of the 2022 China Automation Congress (CAC), Xiamen, China, 25–27 November 2022; pp. 2535–2540.
42. Hou, Q.B.; Zhang, L.; Cheng, M.M.; Feng, J.S. Strip pooling: Rethinking spatial pooling for scene parsing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4003–4012.
43. Wulamu, A.; Shi, Z.X.; Zhang, D.Z.; He, Z.Y. Multiscale road extraction in remote sensing images. *Comput. Intell. Neurosci.* **2019**, *2019*, 2373798. [[CrossRef](#)]
44. 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.
45. Xiao, B.; Yang, Z.Y.; Qiu, X.M.; Xiao, J.J.; Wang, G.Y.; Zeng, W.B.; Li, W.S.; Nian, Y.J.; Chen, W. PAM-DenseNet: A deep convolutional neural network for computer-aided COVID-19 diagnosis. *IEEE Trans. Cybern.* **2021**, *52*, 12163–12174. [[CrossRef](#)] [[PubMed](#)]
46. Ye, H.R.; Zhou, R.; Wang, J.H.; Huang, Z.L. FMAM-Net: Fusion Multi-Scale Attention Mechanism Network for Building Segmentation in Remote Sensing Images. *IEEE Access* **2022**, *10*, 134241–134251. [[CrossRef](#)]
47. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
48. Liu, R.R.; Tao, F.; Liu, X.T.; Na, J.M.; Leng, H.J.; Wu, J.J.; Zhou, T. RAANet: A residual ASPP with attention framework for semantic segmentation of high-resolution remote sensing images. *Remote Sens.* **2022**, *14*, 3109. [[CrossRef](#)]
49. Li, Y.Z.; Cheng, Z.Y.; Wang, C.J.; Zhao, J.L.; Huang, L.S. RCCT-ASPPNet: Dual-Encoder Remote Image Segmentation Based on Transformer and ASPP. *Remote Sens.* **2023**, *15*, 379. [[CrossRef](#)]
50. Shin, T.; Jeong, S.; Ko, J. Development of a Radiometric Calibration Method for Multispectral Images of Croplands Obtained with a Remote-Controlled Aerial System. *Remote Sens.* **2023**, *15*, 1408. [[CrossRef](#)]
51. Wang, C.S.; Wang, L.H.; Zhang, J.; Ye, A.W.; Zhong, G.X.; Wang, Y.Q.; Cui, H.X.; Li, Q.Q. Remote Sensing Video Production and Traffic Information Extraction Based on Urban Skyline. *Geomat. Inf. Sci. Wuhan Univ.* **2023**, *48*, 1490–1498.
52. Kong, Z.; Yang, H.T.; Zheng, F.J.; Li, Y.; Qi, J.; Zhu, Q.Y.; Yang, Z.L. Research advances in atmospheric correction of hyperspectral remote sensing images. *Remote Sens. Nat. Resour.* **2022**, *34*, 1–10.
53. Xu, H.; Yuan, J.; Ma, J. MURF: Mutually Reinforcing Multi-modal Image Registration and Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12148–12166. [[CrossRef](#)]
54. Lu, X.Y.; Zhong, Y.F.; Zheng, Z.; Chen, D.Y. GRE and Beyond: A Global Road Extraction Dataset. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 3035–3038.

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.