remote sensing

Article

Automatic Extraction of Urban Impervious Surface Based on SAH-Unet

Ruichun Chang [1,2,3], Dong Hou [1,2,*], Zhe Chen [1,2,3,4,5] and Ling Chen [1,2]

[1] College of Mathematics and Physics, Chengdu University of Technology, Chengdu 610059, China
[2] Digital Hu Line Research Institute, Chengdu University of Technology, Chengdu 610059, China
[3] Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu 610059, China
[4] International Research Centre of Big Data for Sustainable Development Goals (CBAS), Beijing 100094, China
[5] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
* Correspondence: 2021021053@stu.cdut.edu.cn

Abstract: Increases in the area of impervious surfaces have occurred with urbanization. Such surfaces are an important indicator of urban expansion and the natural environment. The automatic extraction of impervious surface data can provide useful information for urban and regional management and planning and can contribute to the realization of the United Nations Sustainable Development Goal 11—Sustainable Cities and Communities. This paper uses Google Earth Engine (GEE) high-resolution remote sensing images and OpenStreetMap (OSM) data for Chengdu, a typical city in China, to establish an impervious surface dataset for deep learning. To improve the extraction accuracy, the Small Attention Hybrid Unet (SAH-Unet) model is proposed. It is based on the Unet architecture but with attention modules and a multi-scale feature fusion mechanism. Finally, depthwise-separable convolutions are used to reduce the number of model parameters. The results show that, compared with other classical semantic segmentation networks, the SAH-Unet network has superior precision and accuracy. The final scores on the test set were as follows: Accuracy = 0.9159, MIOU = 0.8467, F-score = 0.9117, Recall = 0.9199, Precision = 0.9042. This study provides support for urban sustainable development by improving the extraction of impervious surface information from remote sensing images.

Keywords: impervious surface; remote sensing; deep learning; SAH-Unet; Google Earth images

check for updates

Citation: Chang, R.; Hou, D.; Chen, Z.; Chen, L. Automatic Extraction of Urban Impervious Surface Based on SAH-Unet. *Remote Sens.* **2023**, *15*, 1042. https://doi.org/10.3390/rs15041042

Academic Editor: Yuji Murayama

Received: 27 December 2022
Revised: 1 February 2023
Accepted: 9 February 2023
Published: 14 February 2023

1. Introduction

In recent years, with the acceleration of urbanization processes around the world, the area occupied by cities and towns has expanded. This is causing a series of environmental problems, such as reductions in and degradation of natural habitats, losses in biodiversity, land subsidence and water pollution [1]. According to the United Nations, the urbanization of developing countries has been particularly prominent in the past decade. Their data show that urban expansion is most obvious in Asia, especially China and India [2]. Rapid urbanization has resulted in increases in the area of impervious surfaces. The United States Geological Survey (USGS) defines impervious surfaces as hard areas that do not allow water to seep into the ground [3]. Specifically, they refer to any natural or man-made substance that can hinder water infiltration and, thus, affect the flood runoff, material precipitation and pollution profile. They include building roofs covered with waterproof materials, parking lots and sidewalks. In general, with ongoing global urbanization, the expansion of urban impervious surfaces is having very important impacts on the ecological balance, hydrological conditions and environment of urban areas [4]. To monitor and evaluate sustainable urban development, the United Nations proposed 17 sustainable development goals (SDGs) in 2015. Among them, SDG11 refers to sustainable cities and communities; specifically, strategies to make cities and other human settlements inclusive, safe, resilient and sustainable. The accurate quantification of impervious surfaces is an important planning tool for urban land use development. Careful planning can mitigate the adverse effects of urban heat islands, water quality degradation and natural habitat loss caused by increases in impervious surface area [5]. The automatic extraction of accurate real-time data on impervious surfaces is very important for urban planning and environmental and resource management [6,7]. Research on the automatic extraction of impervious surface data is of great significance to urban ecological construction and the monitoring of urban dynamics and for achieving sustainable development in urban and rural areas.

In the early days, impervious surfaces were mainly studied by manual survey and mapping. Although such methods are highly accurate, they are costly and have poor real-time performance. Compared with traditional surveying and mapping methods, satellite-based remote sensing technology is lower cost, very practical and provides wider coverage. Concurrent with its rapid development, remote sensing technology has become widely used to obtain data on impervious surfaces and has become an important research method in sustainable urban development. The traditional method of extracting impervious water surface data from remote sensing images is to analyse differences in the reflected spectral characteristics of different ground objects through spectral analysis and mixed pixel decomposition. However, data resolution and spectral interference from different ground objects limit the accuracy of impervious surface extraction. For instance, Deyong et al. applied a classification and regression tree (CART) to Landsat and night light data to effectively extract data on impervious water surfaces [8]. Yu et al. proposed the joint use of multi-source remote sensing data, including multispectral images, high-spatial-resolution images and airborne LIDAR data, to extract impervious surfaces [9]. They made full use of visible light, near-infrared radiation, thermal infrared radiation, elevation and other features extracted from the multi-source remote sensing data to achieve a more accurate understanding of urban impervious surfaces. In general, in early research, the extraction of impervious surfaces was mostly based on simple machine learning algorithms. Nevertheless, the features and algorithms must be adjusted for use in different scenarios, applications or geographical areas [10]. These presented many problems, such as a low utilization rate of underlying features, extreme dependence on manual work, poor automation of the extraction process and poor overall accuracy.

In recent years, deep learning has become a major focus of machine learning. It is characterized by its unique automatic feature learning ability and strong ability to represent and fit nonlinear functions. It can generate abstract high-level representations, attributes or features by processing and integrating low-level features [9]. Due to its great advantages over traditional machine learning algorithms, deep learning and related methods have been successfully applied to various computer vision tasks, such as image classification, instance segmentation and target detection. Convolutional neural networks (CNNs) have been gradually applied to remote sensing image processing because they can automatically mine the relevant context representation of images and deeply learn the abstract image features [11]. A fully convolution neural network (FCN) extends image-level classification to the pixel level, greatly promoting the development of semantic segmentation networks [12]. The Unet model based on an encoder–decoder architecture combines the characteristics of deconvolution and jump networks. Many studies have applied it to remote sensing image research and achieved good results [13]. The Feature Pyramid Network (FPN) is a feature pyramid model that combines multi-level features to solve multi-scale problems. It fuses high- and low-level features to increase the expression ability of low-level features and improve network performance. This allows targets of different scales to be allocated to different layers for prediction, following a strategy of “divide and conquer” [14]. The DeepLabv3 network architecture adds a module for multi-scale object segmentation and uses serial and parallel hole convolution modules. It uses a variety of different hole rates to obtain multi-scale content information, which improves the performance of multi-scale object instance segmentation [15,16]. LinkNet links an encoder and decoder to maintain the accuracy of a network model while reducing the number of parameters on a large scale [17]. Through context aggregation based on different regions, the Pyramid Scene Parsing Network (PSPNet) allows the network model to make full use of context information and improve the network’s performance under scenarios with different resolutions [18]. The DeepLabv3+ architecture adds a new decoding module to the DeepLabv3 architecture to reconstruct object boundaries more accurately for image segmentation [19]. The Pixel Aggregation Network (PAN) architecture adds a bottom-up pyramid based on FPN to transfer the underlying features. This allows the model to combine semantic and positioning information to improve performance [20]. The multi-scale attention network (MAnet) introduces a Position-wise Attention Block (PAB) and Multi-scale Fusion Attention Block (MFAB) to capture the channel dependencies between any feature maps by multi-scale semantic feature fusion, providing advancements in medical image segmentation [21]. Compared with classic machine-learning methods, deep-learning methods have better performance in image segmentation [22]. Several network models have been used to extract impervious surfaces; Bowen et al. used a depth convolution neural network to extract data on impervious surfaces from Gaofen 2 satellite remote sensing images of Wuhan city [23]. The efficiency and accuracy of the deep-learning methods were better than those of traditional machine-learning algorithms such as random forest and support vector machine. Parekh et al. used a Unet series to extract data on impervious surfaces from Landsat 8 remote sensing images and achieved good results [3]. Based on the local attention mechanism model in a densely-connected FCN, Pang Bo et al. extracted data on impervious surfaces from GF-2 remote sensing images of Tianjin [24]. Their method had better integrity than other methods in extracting details of impervious surfaces from remote sensing images. In addition, the research of Furkan et al. shows that, even if the sample annotation precision is less than 100%, using a depth neural network classifier to classify remote sensing images can still obtain superior classification results [25]. Even though previous studies have demonstrated satisfactory performance in impervious surface data extraction based on DL networks, some limitations remain that need to be tackled [26]. For instance, as the network hierarchy deepens, small details such as impervious surfaces and edges will be lost. In addition, due to incomplete imaging, such models may commission or omission certain ground objects. To retain detailed information and extract more accurate impervious surface data, further exploration of network models is required. This must ensure their ability to extract multi-scale image features and be suitable for the extraction of impervious surface data from high-resolution remote sensing images [27,28].

To sum up, using a deep-learning method to extract impervious surface information can overcome the main shortcoming of traditional methods—the requirement for a large amount of prior knowledge. Its end-to-end learning method can optimise the model parameters, reduce the dependence on prior knowledge and human intervention, and provide more accurate extractions on impervious surfaces. The present study produced an impervious water surface dataset for deep learning, which is based on high-resolution remote sensing images of Chengdu, a typical Chinese city. The data are analysed using a proposed model—the Small Attention Hybrid Unet (SAH-Unet). Compared with other classical semantic segmentation networks, SAH-Unet demonstrates better performance in extracting impervious surface data. This study proposes a new method for the automatic extraction of impervious surface information from high-resolution remote sensing images. The method provides support for monitoring the sustainable development of cities.

2. Materials and Method

2.1. Data Collection and Processing

The data used in the study were images of Chengdu, China, obtained from GEE and OSM. A corresponding impervious surface binary map was obtained by ENVI processing. Finally, a dataset for model training was constructed through clipping and image enhancement. The following subsections describe these processes in more detail.

2.1.1. Study Area

Chengdu is located in Sichuan Province at 30°05′–31°26′N, 102°54′–104°53′E, at the western edge of Sichuan Basin and the hinterland of Chengdu Plain (Figure 1). To the east are the Longquan Mountains and Pengzhong Mountains (mountain areas within the basin). The central region is the Chengdu Plain, which has a dense river network and fertile land. To the west are the

Qionglai Mountains, which have a great range of elevations and varied topography. Chengdu has a monsoon-influenced humid subtropical climate (Köppen Cwa). It is mostly warm with high relative humidity all year and an annual mean temperature of 16.27 °C (61.3 °F) [29]. It generally features flat terrain, mainly of plains, with a crisscrossing river network. The average elevation in the urban area is 500 m [30]. The density of the city's river network is as high as 1.22 km/km^2, the densest in Sichuan Province [31]. In the past two decades, the city's population has surged and urbanization has continued to expand. Chengdu has gradually become an important central city and economic, cultural, financial and transportation centre of Southwest China [32]. It is an ideal area for studying cities.

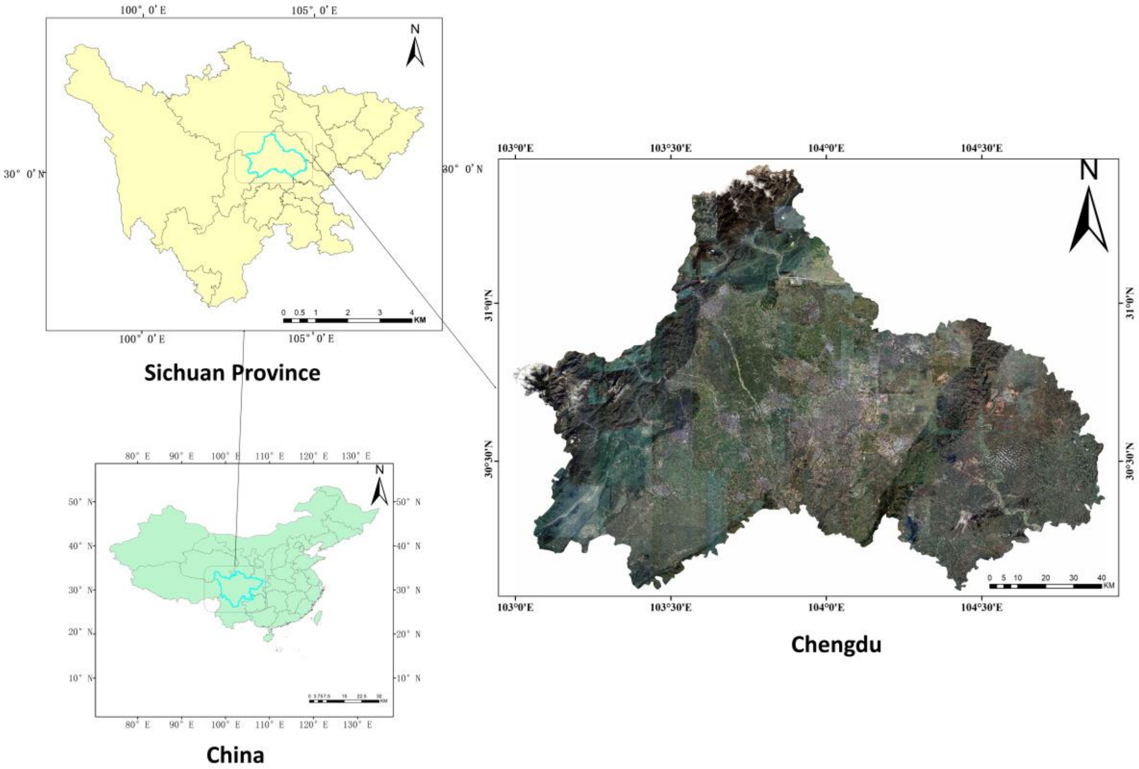


Figure 1. Geographical location and remote sensing image of the study area.

2.1.2. Data Sources

The Google Earth Engine (GEE) is an open-source remote sensing image platform that combines satellite, aerial, 3D and Street View images (Data SIO, NOAA, US Navy, NGA, GEBCO; image Landsat/Copernicus; image IBCAO, etc.) [33]. The data used in the study were high-spatial-resolution (2.15 m) satellite images of Chengdu obtained from the GEE. Due to the characteristics of the integrated images on the GEE platform, the image acquisition dates were inevitably inconsistent. The main source was the Maxar satellite, with an acquisition date range of 2015–2022. As there was a huge number of data, it was difficult to process data for the whole research field at the same time. On the premise of ensuring image quality, three images of the main urban area of Chengdu were selected as training samples and one was selected as a test sample. The size of these images was 4352 × 4352 pixels, and their selection positions are shown in Figure 2.

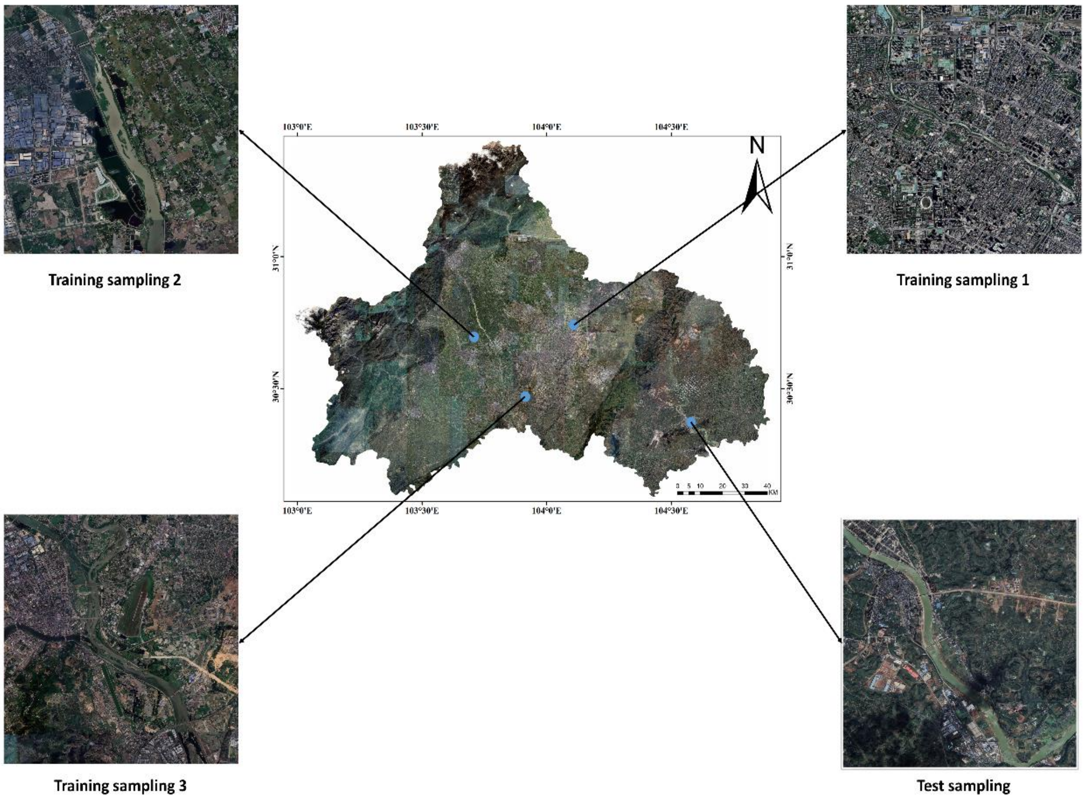


Figure 2. Training and test sample selection location and remote sensing image.

OpenStreetMap (OSM) is a free, open-source geographic database that is updated and maintained by a community of volunteers via open collaboration [34]. Volunteers manually drew various features, such as roads, bridges, pavements and buildings, and rendered this information on a map. Since the model needs to adjust the corresponding weights based on data feedback, it is crucial to obtain accurate sample data. To ensure the accuracy of impervious surface labels as much as possible, we used OSM data to help identify impervious surface [35,36]. The overall OSM data of Chengdu is shown in Figure 3 and the correspondence between the OSM data and high-resolution remote sensing images are shown in Figure 4 (taking the test samples as an example).

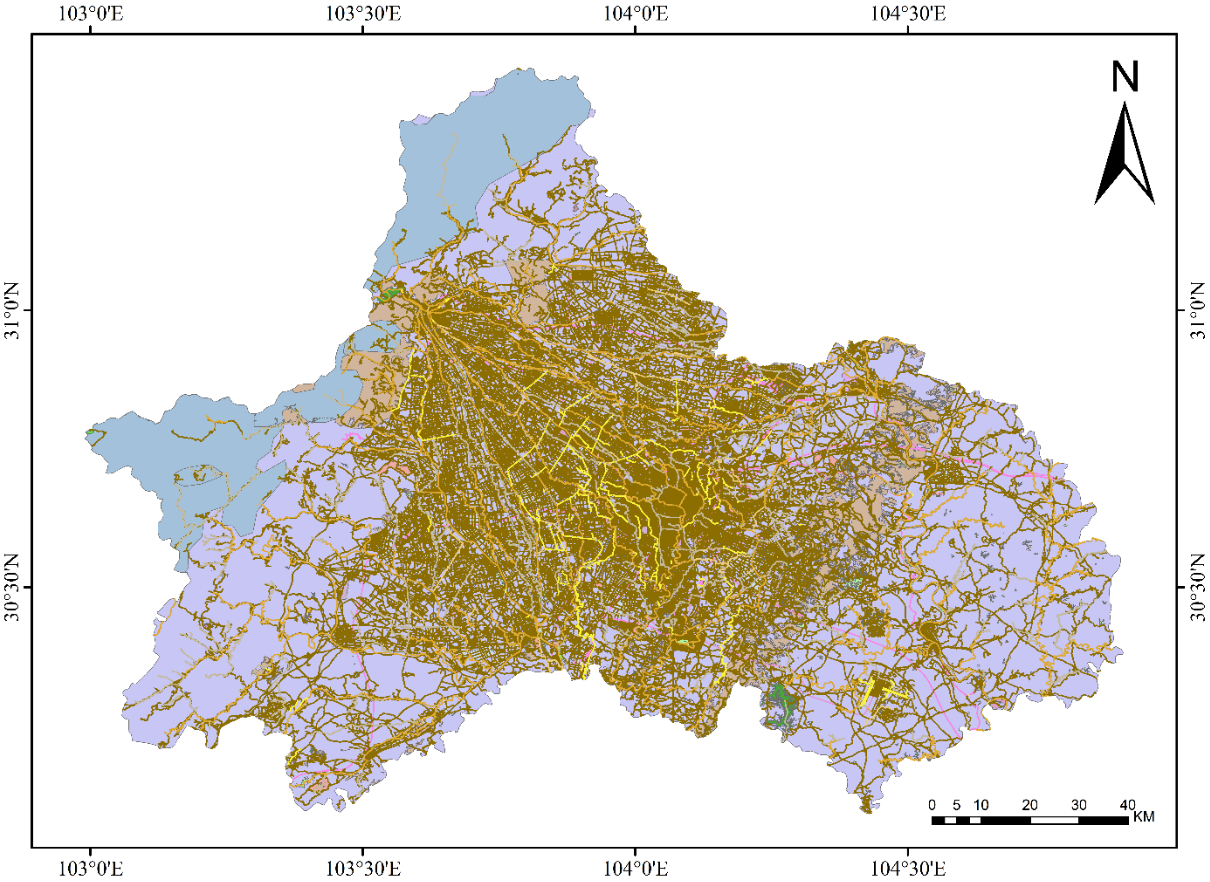


Figure 3. Overview of Chengdu OSM data.

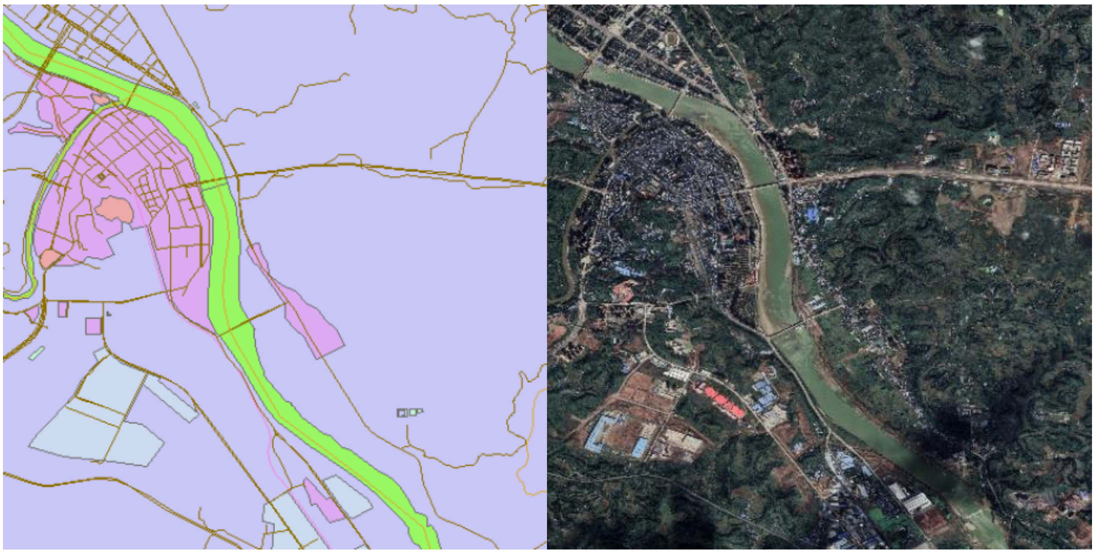


Figure 4. Example mapping of the OSM and high-resolution image.

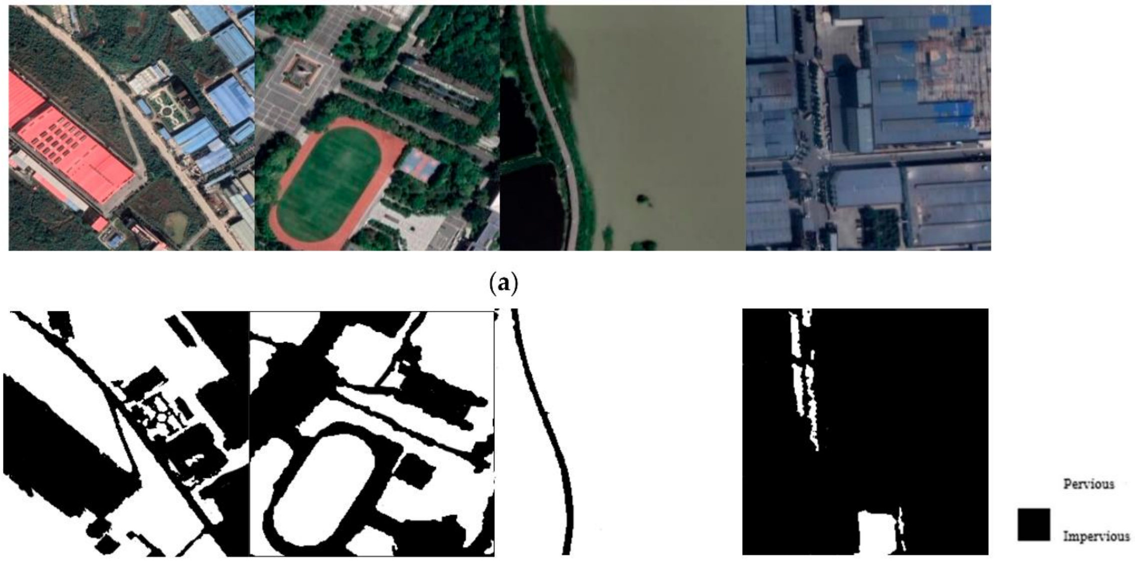
2.1.3. Dataset Construction

An important problem in extracting impervious surfaces using deep-learning methods is data scarcity. There are many types of urban land cover, such as buildings, roads, bridges, forests, farmland, bare land, water and other surface objects, which have complex spatial structures. Accurately labelling as many different types of permeable and impervious surfaces as possible is key to the stability of model training and increasing the model's generalisability.

In this paper, we obtained high-resolution remote sensing images from the GEE to make a binary map of impervious surfaces. To ensure the accuracy of labels as much as possible, we used OSM data as mask data corrected by visual interpretation and used ENVI to mark the impervious surface areas in the image in strict accordance with their original positions in the image. Supervised classification was used for the initial classification of the dataset to ensure that the separation between samples was >1.9. After the initial classification, Majority/Minority Analysis, clump classes and sieve classes operations were carried out to eliminate data processing errors and obtain a better binary map of impermeable surfaces.

High-resolution remote sensing images have extremely rich spatial, colour and texture features. The different levels of feature information provide a basis for the extraction of impervious surfaces [37]. However, directly inputting the images into the network model will cause a huge computational burden and lead to memory overflow. To retain as much image information as possible and reduce computational pressure, the images were cut into blocks before being input into the network. This allowed the detection of some small and meaningful texture and contour features related to impervious surfaces that were contained in a small number of adjacent pixels. When there are limited data, data augmentation can improve the quality and quantity of existing data, expand the learning range during model training, and enhance data robustness. This is done by adding noise to better train the network model and enhance its generalisability. Due to the angle and temporal image of remote sensing image sensors [38], the image quality can be improved through geometric changes and other data enhancement methods. This can make the image features more obvious, thus enhancing the recognition of impervious surfaces. In this paper, we used noise and Gaussian blur to enhance the learning ability of the model and improve its robustness. Multi-angle rotation was adopted to resolve the inconsistency between the remote sensing image distribution and reality.

Based on the above analysis, the unified picture block size used in this paper was 256 × 256. After image enhancement, the images with poor quality (where the features were not obvious) and unbalanced sample distribution were filtered out to form 7605 images of 256 × 256 pixels. There were 5760 used in the training set, 1440 in the verification set and 405 in the test set. An example of the final sample is shown in Figure 5. The marked impervious surfaces mainly include buildings and roads, while the other surfaces include vegetation, water bodies and bare land.



(**a**)

(**b**)

Figure 5. Example of a remote sensing impervious surface map. (**a**) Sub-map image generated after original image clipping. (**b**) Impervious surface binary map corresponding to the sub-map image.

2.2. Methodology

This section introduces the proposed SAH-Unet model proposed for the extraction of impermeable surfaces, and describes its training process and evaluation index.

2.2.1. SAH-Unet

The model that we propose here builds upon and extends the Unet architecture. As shown in Figure 6, the Unet architecture consists of an encoder–decoder structure that results in a U-shape [39]. The encoder part (corresponding to the left half of Figure 6) applies max-pooling and convolution, which halves the image size and doubles the number of feature maps, respectively. The encoders are subsequently followed by the same number of decoders (corresponding to the right half of Figure 6), which consist of three parts: a bilinear upsampling operation to double the feature map size, a concatenation of the resulting feature maps with the previous encoder's output via skip-connections, and, lastly, convolution to half the number of feature maps. Unet can use multiple scales of input to generate its output, which is conducive to the extraction of impervious surfaces.

The Small Attention Hybrid Unet (SAH-Unet) makes three modifications to Unet. Firstly, we add the CBAM attention mechanism to the encoder part. Secondly, we introduce the multi-scale feature fusion (MFF) mechanism to the encoder part. Thirdly, we transform the regular convolutional operations to depthwise-separable convolutions (DSC).

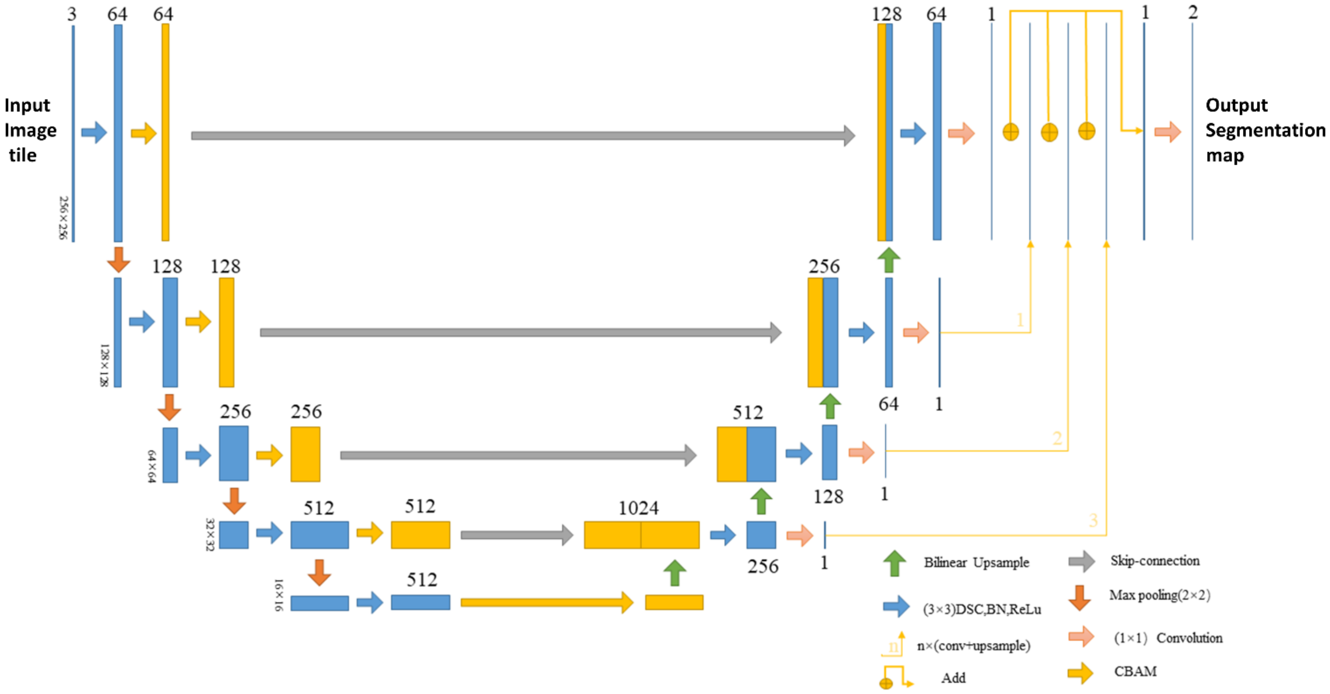


Figure 6. SAH-Unet model architecture. Each box corresponds to a multi-channel feature map. The number of channels is denoted at the top of the box. The size is provided at the lower left edge of the box. The arrows denote the different operations.

2.2.2. CBAM Attention Mechanism

In the field of deep learning, the introduction of an attention mechanism means that the network model does not need to process huge amounts of input information, some of which may be redundant, to the same standard, allowing the network to focus on specific parts of the input [40,41]. In the detection of impervious surfaces, due to the existence of small, complex and overlapping samples, the spatial perception ability of the model is also very important. As it is lightweight and generalisable, a convolutional block attention module (CBAM) can be integrated into any CNN architecture seamlessly with negligible overhead. Such modules have demonstrated usefulness in feature extraction [42]. In CBAM, the attention mechanism is applied first across the channels of the image and subsequently to the spatial dimension, as shown in Figure 7. The channel attention module focuses on deciding what is meaningful information. It uses two parallel max-pooling (Maxpool) layers and an average-pooling (Avgpool) layer. It then passes through a shared multi-layer perceptron (MLP) module and, finally, adds the two output results and uses a sigmoid activation function to obtain the weights of each channel use. The channel attention is computed as follows:

$$M_c(F) = \sigma(MLP(Avgpool(F)) + MLP(Maxpool(F))) \tag{1}$$

where F denotes the input feature and σ denotes the sigmoid function.

The spatial attention module focuses on the informative parts. It obtains two feature maps through the Maxpool and Avgpool layers, then concatenates the feature maps and, finally, obtains the weights of each pixel in the entire image through 7 × 7 convolution and the sigmoid activation function. In short, it is computed as:

$$M_s(F') = \sigma\left(f^{7\times7}([Avgpool(F'); Maxpool(F')])\right) \tag{2}$$

where F' denotes the input characteristics of the through channel attention module and $f^{7\times7}$ represents a convolution operation with a filter size of 7 × 7.

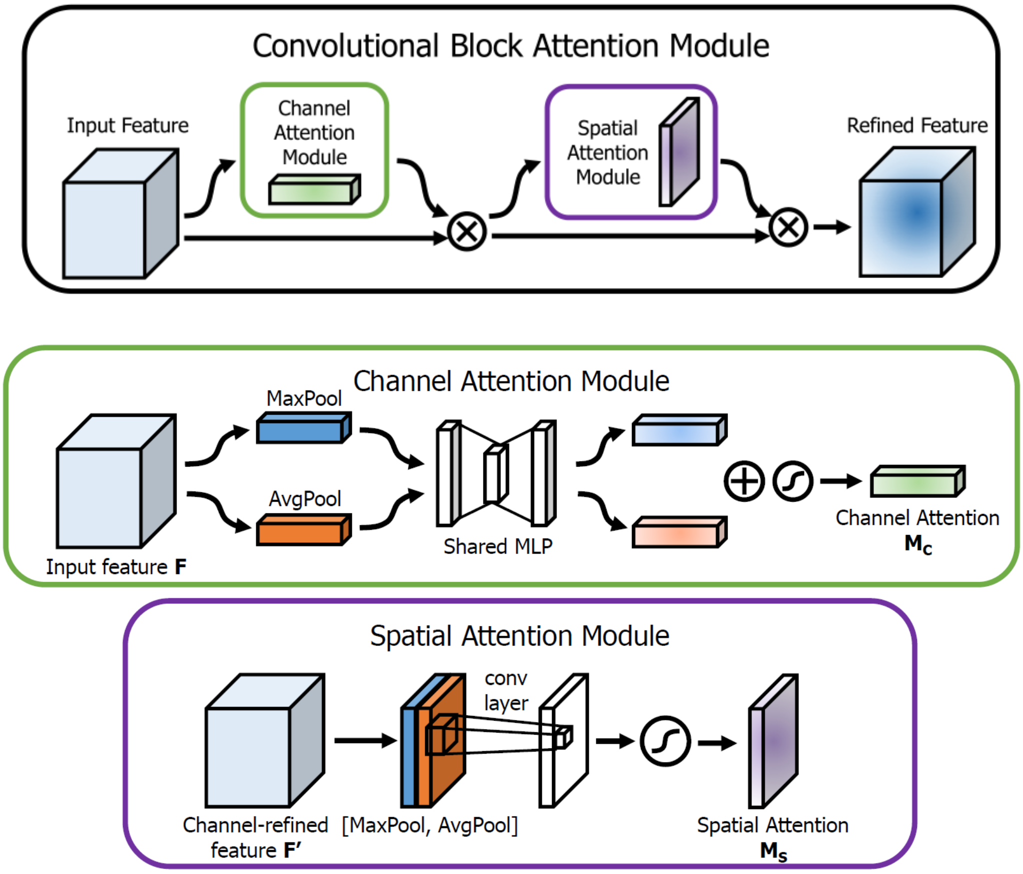


Figure 7. CBAM schematic diagram. The module has two sequential sub-modules: channel and spatial. the channel sub-module utilises both max-pooling outputs and average-pooling outputs with a shared network; the spatial sub-module utilises similar two outputs that are pooled along the channel axis and forward them.

In SAH-Unet, CBAM is placed after the convolution calculation of each encoder to enlarge important features and suppress relatively unimportant ones at the image scale. It is worth noting that the input of each encoder is still based on the features of the previous encoder after convolution and maximum pooling down-sampling, rather than on the feature information with the attention mechanism applied. This way, the original image features are preserved until the last encoder. In the subsequent network, the features processed by the attention mechanism are connected with the feature information sampled on the decoder through skip-connection.

2.2.3. Multi-Scale Feature Fusion Mechanism

The original Unet only predicts the last layer from top to bottom. The shallow-layer features can provide more accurate location information, while multiple up-sampling and down-sampling network operations will introduce errors into the positioning information of the deep network, resulting in easy loss of detail in small coverage areas, and irregular shape coefficients for impervious surfaces [43]. The SAH-Unet model adds a branch path to the decoder (corresponding to the part of the decoder marked by the yellow arrow in Figure 6). The specific implementation is divided into three parts: (1) 1 × 1 convolution is conducted to obtain a feature map with 1 channel; (2) convolution and upsampling of this map are done a corresponding number of times to obtain a feature map with the same size and channel number as the final output; (3) we add the outputs from each encoder to obtain the final output of the model. By predicting the classification results at different scales in the upsampling step, the model can use multi-scale information in the backpropagation and weight update process and apply the feature map of each layer to the prediction. This enhances the sensitivity of the network to impervious surface details, improves the ability to extract the impervious surfaces of small targets, and thus improves the accuracy of impervious surface data extraction.

2.2.4. Depthwise-Separable Convolutions

Finally, we use depthwise-separable convolutions in our model to reduce the number of parameters. It consists of two parts: depthwise convolution and pointwise convolution [44,45]. Each channel conducts a convolution operation to obtain the corresponding output, and then passes a 1 × 1 convolution kernel to obtain the final output, as shown in Figure 8. Its advantage is that it can greatly reduce the number of parameters and amount of computation when extracting multi-scale features of impervious surfaces. Assume that the size of the input feature map is $D_G \times D_G \times M$. In depthwise convolution, a channel can only be convolutioned by one convolution kernel; thus, assume that the size of the filter is $D_K \times D_K \times 1$. Then, the calculation amount after depthwise convolution is $M \times {D_G}^2 \times {D_K}^2$. Because depthwise convolution only calculates a single channel, the information between each channel is not exchanged. The introduction of pointwise convolution completes the further fusion of feature channels. Because its convolution kernel size is 1×1, assume that the size of the output feature map is $D_G \times D_G \times N$. Thus, the filter is $1 \times 1 \times M$ and the number is N. Then, the calculation amount after pointwise convolution is $N \times M \times {D_G}^2$ and that after regular convolution is $N \times {D_G}^2 \times {D_K}^2 \times M$. The formula for calculating the ratio of the depthwise-separable and regular convolution calculation amounts can be obtained as follows:

$$\frac{\text{Depthwise–separable Conv}}{\text{Regular Convolutional}} = \frac{M \times {D_G}^2 \left({D_K}^2 + N\right)}{N \times {D_G}^2 \times {D_K}^2 \times M} = \frac{{D_K}^2 + N}{{D_K}^2 \times N} \quad (3)$$

In SAH-Unet, we substitute all convolutions of the original Unet model with depthwise-separable convolutions, but in CBAM we still apply regular convolutions. Table 1 compares the models' parameters. It can be seen that the number of parameters of the proposed SAH-Unet model is greatly reduced after improving the original Unet, and the increase in parameters compared with Unet with DSC is also within an acceptable range. The final network model parameters are about 4 M.

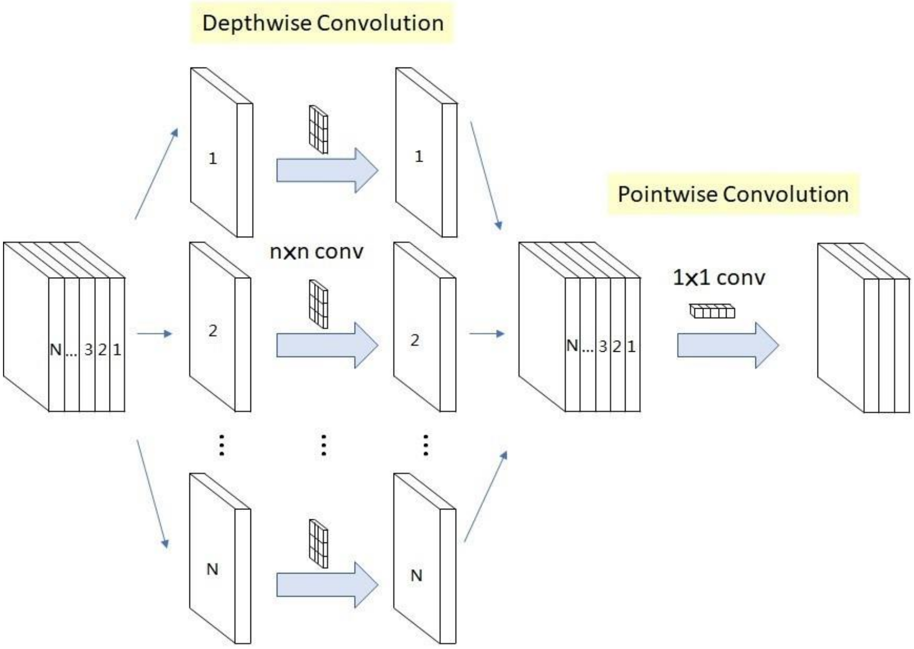


Figure 8. Depthwise-separable convolutions schematic diagram. It consists of two parts: depthwise convolution and pointwise convolution. Depthwise convolution uses only one convolution core for each channel. Pointwise convolution performs 1 × 1 convolution on all channels.

Table 1. Comparison Table of Model Parameters.

Model	Parameters
Unet	17,272,577
Unet with CBAM	17,428,781
Unet with DSC	3,955,185
SAH-Unet	4,121,398

2.2.5. Model Training and Evaluation

To better evaluate the performance of each model, we pre-trained them on the ImageNet dataset [46], and then used pre-trained weights to extract impervious surfaces. In model training, the epochs were pre-set to 200 and an early stop standard was used whereby training stopped when the verification loss did not improve for five consecutive epochs [47]. The experiment shows that all the models were satisfied within 200 epochs. When the gradient descent optimisation algorithm is used to optimise the objective function, the learning rate should be reduced appropriately so that the global minimum of the loss function can be approached as much as possible. The cosine function has the characteristics of slowly decreasing and then accelerating and then decelerating with the increase in the value of the independent variable. To avoid falling into local optimisation, we used cosine annealing to optimise the learning rate. Its formula is:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right) \quad (4)$$

where η_t denotes current learning rate, η_{max} denotes maximum learning rate, η_{min} denotes minimum learning rate, T_{cur} denotes the current number of iterations and T_{max} denotes the maximum number of iterations.

The initial learning rate was set to 0.001; we used the Adam optimiser with default values [48]. The loss function used was the Dice Loss function between the impervious surface label of the predicted image and the actual label [49,50]. This is a region-related loss function that provides an ideal effect in the training of positive and negative sample imbalance. Its formula is:

$$\text{DiceLoss} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i} \quad (5)$$

where y_i represents the label value of pixel i, $\hat{y}_i$ represents the predicted value of pixel i, and N represents the total number of pixels.

2.2.6. Model Evaluation

To compare the impervious surface data extraction accuracy of each model, we used the Accuracy, MIOU, F-score, Recall and Precision indicators [51,52]. A true positive (TP) is when an impervious surface is consistent with the model's identification, a false positive (FP) is when a permeable surface is misidentified by the model as impervious, a false negative (FN) is when an impervious surface is misidentified by the model as permeable, and a true negative (TN) occurs when a permeable surface is correctly identified by the model [53,54]. These values are achieved by the Confusion Matrix defined in Table 2. The metrics are calculated as follows:

Table 2. Explanation of the confusion matrix.

Actual Labels	**Predicted Labels**	
	Impervious	**Permeable**
Impervious	TP	FN
Permeable	FP	TN

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (8)$$

$$\text{F-score} = \frac{2\times recall \times precision}{recall+precision} \quad (9)$$

$$\text{MIOU} = \frac{TP}{FP+FN+TP} \quad (10)$$

3. Results

The experiments were conducted in Python language based on the Pytorch 1.12 framework. The experiments were done on a single Tesla P100 GPU with 12.5 GB of RAM. To further demonstrate the performance of the proposed SAH-Unet in impervious surface extraction, the classical semantic segmentation network architectures Unet, DeepLabv3+, PAN, FPN, LinkNet, PSPNet and PAN were trained under the same environment for comparison.

3.1. Training and Validation Results

The final training results of each model on the validation and test sets are shown in Table 3. The results are the averages of five experiments.

Table 3. Comparison table of model training and test results.

Set	Model	Mean Accuracy Score	Mean Loss
Training set	LinkNet	0.9153	0.0979
	DeepLabv3+	0.9237	0.0831
	PAN	0.9293	0.0745
	Unet	0.9352	0.0706
	MAnet	0.9337	0.0690
	PSPNet	0.9089	0.0947
	FPN	0.9262	0.0807
	SAH-Unet	0.9432	0.0640
Validation set	LinkNet	0.8543	0.1511
	DeepLabv3+	0.8714	0.1327
	PAN	0.8656	0.1400
	Unet	0.8863	0.1206
	MAnet	0.8830	0.1209
	PSPNet	0.8375	0.1673
	FPN	0.8682	0.1367
	SAH-Unet	0.8873	0.1169

Table 3 shows that the proposed SAH-Unet architecture achieves the best performance after training, with the lowest training loss of 0.0640. It also has the lowest loss and highest accuracy of the verification set: 0.1206 and 0.8863, respectively.

3.2. Metric Results

The final metric results of each model on the test set are shown in Table 4. The results are the averages of five experiments.

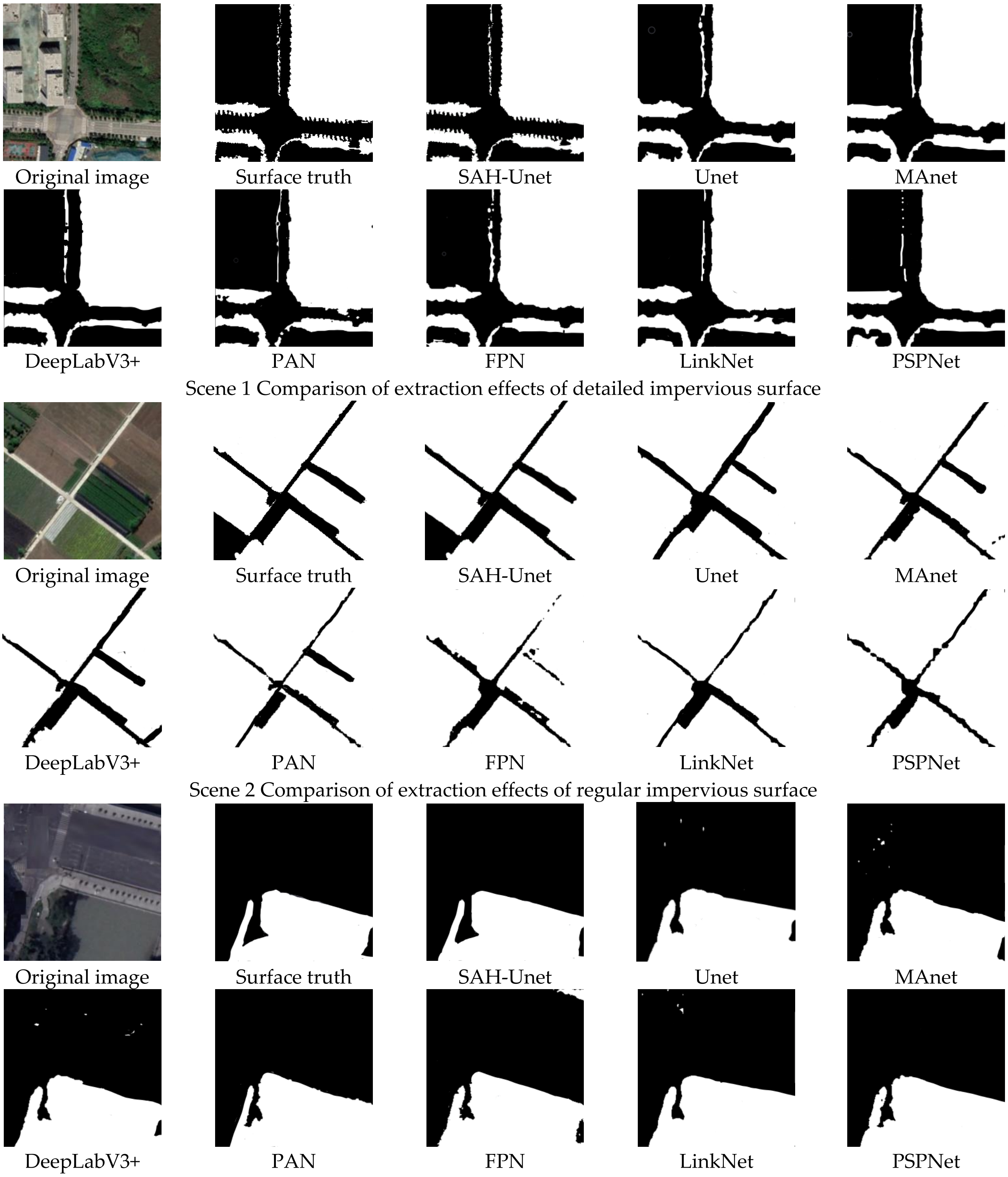
Table 4. Comparison table of model metric results.

Model	Accuracy	MIOU	F-Score	Recall	Precision
LinkNet	0.8759	0.7836	0.8726	0.8866	0.8596
DeepLabv3+	0.8974	0.8166	0.8944	0.9096	0.8805
PAN	0.8907	0.8062	0.8875	0.9003	0.8756
Unet	0.9078	0.8339	0.9041	0.9149	0.8943
MAnet	0.9052	0.8299	0.9020	0.9150	0.8902
PSPNet	0.8747	0.7799	0.8712	0.8821	0.8609
FPN	0.8885	0.8019	0.8859	0.9040	0.8693
SAH-Unet	0.9159	0.8467	0.9117	0.9199	0.9042

Table 4 compares the Precision of each model on the test set. The results show that SAH-Unet has advantages in terms of five evaluation indicators—Accuracy, MIOU, F-score, Recall and Precision—as well as having the highest extraction precision of all models. Compared with other semantic segmentation networks, the Unet network also achieved relatively good results, indicating that its unique jump-connection architecture can improve the accuracy of extracting impermeable water surface data from high-resolution remote sensing images, making it suitable for this task.

3.3. Visualization Results

To verify the actual effect of extracting impervious surfaces with SAH-Unet, the extraction results were visually evaluated. As a large range of extraction results cannot clearly reflect the differences between various methods, three typical areas in the test sample were selected for detailed comparison. A detailed comparison of the impervious surface extraction results of the SAH-Unet, LinkNet, DeepLabV3+, PAN, Unet, MAnet, PSPNet and FPN networks is shown in Figure 9.



Scene 1 Comparison of extraction effects of detailed impervious surface

Scene 2 Comparison of extraction effects of regular impervious surface

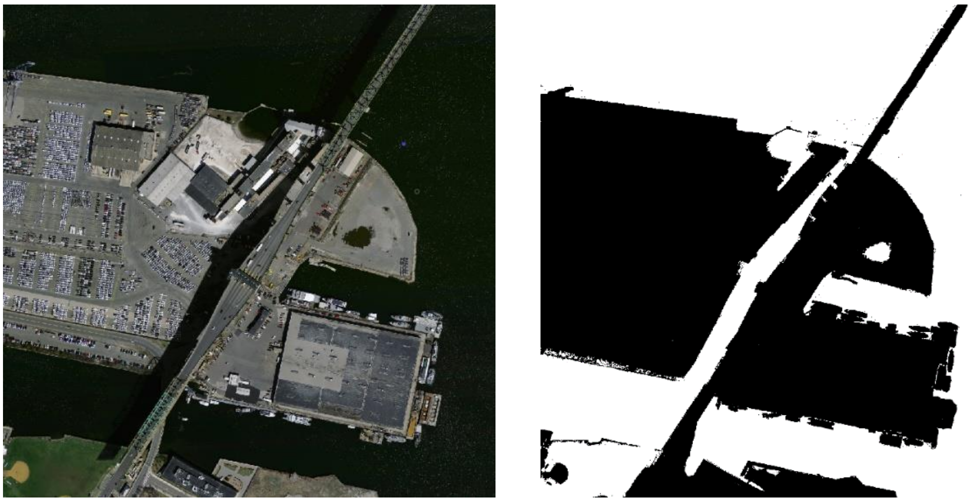
Scene 3 Comparison of extraction effects of special shaped impervious surface

Figure 9. Results of impervious surface extraction by different methods.

From the visualisation of the impervious surface extraction results shown in Figure 9, it can be seen that the extraction results of SAH-Unet have the highest rate of coincidence with the true values. The contour is relatively complete, the boundary is clear and the extraction effect is the best. In scene 1, due to shooting restrictions, vegetation blocks the scene, resulting in extremely irregular contours. Unlike SAH-Unet, the other models' extractions of impermeable surfaces in scene 1 have commissions or omissions. This indicates that the proposed model is also good at extracting complex and irregular contours. In scene 2, due to the size limitation of the image, the shed in the lower-left corner is only partly visible, which increases the difficulty of recognition. As a result, all the models except SAH-Unet produce wrongly divided results. In addition, the proposed model also shows better performance in extracting ground objects with clear and fixed contours, such as roads. The difficulty in scene 3 lies in the extraction of unobvious details on the right side and the interference of noise. Through comprehensive comparison, we can see that SAH-Unet achieves the best extraction effect, with no incorrect division due to noises.

3.4. Generalization Results

To verify the generalisation performance and scalability of the proposed model, aerial images of Boston, Massachusetts, USA that were not used in training were selected for model verification. The results of impervious surface extraction are shown in Figure 10.



(**a**) High-resolution images (**b**) SAH-Unet extraction results

Figure 10. Test results based on high-resolution images from different times and regions.

It can be seen from Figure 10 that SAH-Unet can still accurately extract impervious surfaces from images taken at different times and in different regions. The distribution of impervious surfaces in this area is clear, and the road and building structures are obvious.

It is noteworthy that the shadow of the highway is not correctly identified as an impervious surface. On the other hand, small shadows of cars and buildings are correctly recognised. It is speculated that the ground feature model containing shadows in the image block was sufficiently studied during training, and the recognition ability of the large image models is insufficient due to the limitations of image size. In fact, regardless of the overall image, for some shadows in the image block, it is also difficult for humans to accurately identify the impervious surfaces. In general, the extraction result of SAH-Unet is in good agreement with the real values, and it also extracted some small-scale details well, indicating that the model has a certain generalisability in impervious surface data extraction.

3.5. Ablation Study

It can be seen from Table 5 that, compared with Unet, the introduction of CBAM and multi-scale feature fusion (MFF) can further improve the accuracy, while the introduction of DSC causes accuracy loss while reducing the number of model parameters. SAH-Unet significantly improves the modelling accuracy while greatly reducing the parameters. Compared with Unet, the indexes of the extraction results of SAH-Unet are improved by 0.89%, 1.53%, 0.84%, 0.55% and 1.11%, respectively, which shows that the introduction of attention modules and feature fusion can achieve better model performance, even with a reduced number of parameters. In general, SAH-Unet is feasible for the extraction of impervious surface data and has excellent accuracy. The total accuracy of impervious surface data extraction based on high-resolution images is 0.9159, and the MIOU, F-score, Recall and Precision are 0.8467, 0.9117, 0.9199 and 0.9042, respectively.

Table 5. Comparison table of the test accuracy of different Unet models.

Model	Accuracy	MIOU	F-Score	Recall	Precision
Unet	0.9078	0.8339	0.9041	0.9149	0.8943
Unet with CBAM	0.9096	0.8386	0.9095	0.9164	0.9003
Unet with MFF	0.9105	0.8391	0.9102	0.9166	0.9009
Unet with DSC	0.9021	0.8320	0.9019	0.9100	0.8895
SAH-Unet	0.9159	0.8467	0.9117	0.9199	0.9042

Scene 3 in Figure 9 is used as an example to visually analyse the Unet-series models (Figure 11). It can be seen from the figure that Unet CBAM does not suffer from noise misclassification, which indicates that the introduction of an attention mechanism makes the model more accurate in identifying impervious surfaces and more capable of dealing with noise. However, there is still some room for improvement in extracting certain details of the impervious surface. Unet with MFF has a strong ability to extract the impervious surface details, which indicates that the introduction of a multi-scale feature fusion mechanism increases the ability to extract network details. However, it is affected by some other factors that result in misclassification. SAH-Unet combines the advantages of both and shows the best performance.

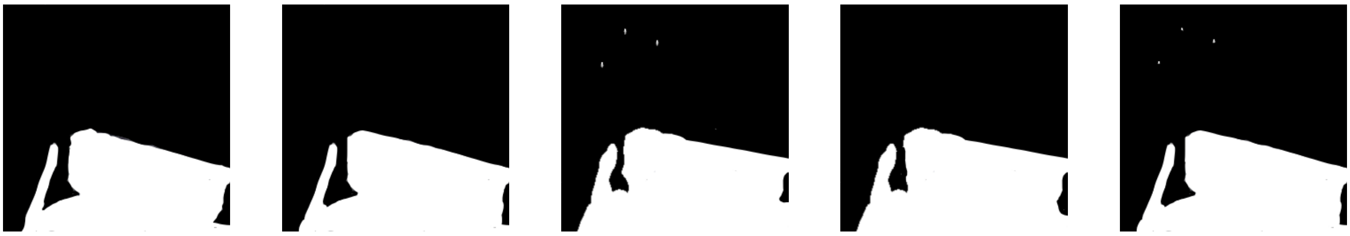

(**a**) Surface truth (**b**) SAH-Unet (**c**) Unet (**d**) Unet with CBAM (**e**) Unet with MFF

Figure 11. Impervious Surface Data Extraction Results of Different Unet Architectures.

4. Discussion

This paper proposes the SAH-Unet network model for extracting impervious surface data from high-resolution remote sensing images. The experimental results show that the network structure setting is effective. Table 5 and Figure 11 show that the introduction of CBAM helps the model to extract impermeable surface information more accurately, while the introduction of MFF enhances its ability to extract impermeable surface details. Tables 1 and 5 show that the introduction of depth separable convolution greatly reduces the number of model parameters while maintaining model performance. In the experiment testing the generalisation ability of the proposed model with high-resolution remote sensing images of Chengdu, good results in the extraction of impervious surface information were achieved.

Historically, impervious surface modeling is based on statistical indices computed to accentuate impervious surfaces in satellite imagery; the use of deep-learning methods to extract the impermeable surface is still a frontier topic [55]. SAH-Unet achieves the best results in terms of target edges and details and has certain advantages over the LinkNet, DeepLabV3+, PAN, Unet, MAnet, PSPNet and FPN frameworks. Tables 3 and 4 show the precision results of each model on the training, validation and test set: the total extraction precision of SAH-Unet on the test set was 0.9159, while the MIOU, F-score, Recall and Precision were 0.8467, 0.9117, 0.9199 and 0.9042, respectively, the best of all models. Figure 9 shows the visualisation results of impervious surface extraction. Figure 10 uses high-resolution remote sensing images of different time images and regions to test the generalisation ability of SAH-Unet. SAH-Unet also has great advantages over other models. In view of the difficulties of detail and shadow extraction with impervious surface data, this method also has some improvements.

It is worth mentioning that both buildings and vegetation will produce shadows, and when large shadows are used as input features, due to the incomplete information contained in the image, it greatly increases the difficulty of model recognition, resulting in the phenomenon of misclassification. In addition, due to the small amount of bare surface leakage in the urban area, it may lead to insufficient recognition by the model.

In general, there is still much room for improvement in the extraction of impervious surface information from high-resolution remote sensing images. With the continuous development of network architectures and remote sensing technology, further progress will be made in the extraction of data on urban impervious surface via deep learning.

5. Conclusions

The real-time extraction of impervious surface data is of great significance to urban sustainable development. This paper used GEE high-resolution remote sensing images and OSM data for Chengdu, a typical city in China, to make an impervious surface dataset via pre-processing and image enhancement. This was applied to network model training based on deep learning. In addition, this paper improved upon Unet by proposing the SAH-Unet architecture. It introduced an attention mechanism and feature prediction combined with multi-scale fusion, and, finally, used depthwise-separable convolution instead of traditional convolution to reduce the model parameters. The experimental results show that, compared with other classical models, the proposed model can extract impervious surface data better, while it achieved higher accuracy.

As future research, we will introduce other attention mechanisms and test their effects on the model. Additionally, later research will strive to build a higher quality dataset and further explore the network architecture.

Author Contributions: D.H.: methodology, validation, conceptualization, data curation, writing—original draft, and software. R.C.: writing—review and editing, funding acquisition, supervision, and project administration. Z.C.: investigation, formal analysis, resources, and visualization. L.C.: supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant numbers XDA19090121, XDA19030000, and XDA19090000), the Key Research and Development Program of Guangxi (GuikeAB22035060), the Key Research and Development Program of the Sichuan Provincial Science and Technology Department (grant number 2022YFS0486), Chengdu University of Technology Postgraduate Innovative Cultivation Program: Tunnel Geothermal Disaster Susceptibility Evaluation in Sichuan-Tibet Railway Based on Deep Learning (CDUT2022BJCX015).

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank data providers including the United States Geological Survey (USGS), Google, pytorch, and the OpenStreetMap (OSM) community for provision of resources which made this study possible. The authors are grateful for helpful comments from many researchers and colleagues.

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

Abbreviations

Abbreviation	**Definition**
Avgpool	average-pooling
CART	classification and regression tree
CBAM	convolutional block attention module
CNNs	Convolutional neural networks
DSC	depthwise-separable convolutions
FCN	fully convolution neural network
FN	false negative
FP	false positive
FPN	Feature Pyramid Network
GEE	Google Earth Engine
MAnet	multi-scale attention network
Maxpool	max-pooling
MFAB	Multi-scale Fusion Attention Block
MFF	multi-scale feature fusion
MLP	multi-layer perceptron
OSM	OpenStreetMap
PAB	Position-wise Attention Block
PAN	Pixel Aggregation Network
PSPNet	Pyramid Scene Parsing Network
SAH-Unet	Small Attention Hybrid Unet
SDGs	sustainable development goals
TN	true negative
TP	true positive
USGS	United States Geological Survey

References

1. Elmqvist, T.; Andersson, E.; Frantzeskaki, N.; McPhearson, T.; Olsson, P.; Gaffney, O.; Takeuchi, K.; Folke, C. Sustainability and resilience for transformation in the urban century. *Nat. Sustain.* **2019**, *2*, 267–273. [CrossRef]
2. United Nations Department of Economic and Social Affairs (UN DESA). Commission on Population and Development, Fifty-Sixth Session. 2023. Available online: https://www.un.org/development/desa/pd/events/CPD56 (accessed on 1 September 2022).
3. Parekh, J.R.; Poortinga, A.; Bhandari, B.; Mayer, T.; Saah, D.; Chishtie, F. Automatic detection of impervious surfaces from remotely sensed data using deep learning. *Remote Sens.* **2021**, *13*, 3166. [CrossRef]
4. Mohajerani, A.; Bakaric, J.; Jeffrey-Bailey, T. The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *J. Environ. Manag.* **2017**, *197*, 522–538. [CrossRef] [PubMed]
5. Shrestha, B.; Ahmad, S.; Stephen, H. Fusion of Sentinel-1 and Sentinel-2 data in mapping the impervious surfaces at city scale. *Environ. Monit. Assess.* **2021**, *193*, 556. [CrossRef] [PubMed]
6. United Nations Department of Economic and Social Affairs (UN DESA). Sustainable Development Goals Report 2017. Available online: https://www.un.org/en/desa/sustainable-development-goals-report-2017 (accessed on 1 November 2022).
7. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015.
8. Hu, D.; Chen, S.; Qiao, K.; Cao, S. Integrating CART algorithm and multi-source remote sensing data to estimate sub-pixel impervious surface coverage: A case study from Beijing Municipality, China. *Chin. Geogr. Sci.* **2017**, *27*, 614–625. [CrossRef]
9. Yang, L.; Huang, C.; Homer, C.G.; Wylie, B.K.; Coan, M.J. An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Can. J. Remote Sens.* **2003**, *29*, 230–240. [CrossRef]
10. Coseo, P.; Larsen, L. Accurate characterization of land cover in urban environments: Determining the importance of including obscured impervious surfaces in urban heat island models. *Atmosphere* **2019**, *10*, 347. [CrossRef]
11. Bau, D.; Zhu, J.-Y.; Strobelt, H.; Lapedriza, A.; Zhou, B.; Torralba, A. Understanding the role of individual units in a deep neural network. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 30071–30078. [CrossRef]
12. Zhao, J.; Mao, X.; Chen, L. Learning deep features to recognise speech emotion using merged deep CNN. *IET Signal Process.* **2018**, *12*, 713–721. [CrossRef]
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

15. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
16. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
17. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
18. Chaurasia, A.; CulurcielloLinknet, E. Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017.
19. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
20. Li, H.; Xiong, P.; An, J.; Wang, L. Pyramid attention network for semantic segmentation. *arXiv* **2018**, arXiv:1805.10180.
21. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface water mapping by deep learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4909–4918. [CrossRef]
22. Khankeshizadeh, E.; Mohammadzadeh, A.; Moghimi, A.; Mohsenifar, A. FCD-R2U-net: Forest change detection in bi-temporal satellite images using the recurrent residual-based U-net. *Earth Sci. Inform.* **2022**, *15*, 2335–2347. [CrossRef]
23. Cai, B.; Wang, S.; Wang, L.; Shao, Z. Extraction of urban impervious surface from high-resolution remote sensing imagery based on deep learning. *J. Geo-Inf. Sci.* **2019**, *21*, 1420–1429.
24. Pang, B.; Huang, Z.; Lu, Y. Mapping of Impervious Surface Extraction of High Resolution Remote Sensing Imagery Based on Improved Fully Convolutional Neural Network. *Remote Sens. Inf.* **2020**, *35*, 47–55.
25. Sun, Z.; Zhao, X.; Wu, M.; Wang, C. Extracting urban impervious surface from worldView-2 and airborne LiDAR data using 3D convolutional neural networks. *J. Indian Soc. Remote Sens.* **2019**, *47*, 401–412. [CrossRef]
26. Fu, Y.; Liu, K.; Shen, Z.; Deng, J.; Gan, M.; Liu, X.; Lu, D.; Wang, K. Mapping impervious surfaces in town–rural transition belts using China's GF-2 imagery and object-based deep CNNs. *Remote Sens.* **2019**, *11*, 280. [CrossRef]
27. Zhang, H.; Wan, L.; Wang, T.; Lin, Y.; Lin, H.; Zheng, Z. Impervious surface estimation from optical and polarimetric SAR data using small-patched deep convolutional networks: A comparative study. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2374–2387. [CrossRef]
28. McGlinchy, J.; Muller, B.; Johnson, B.; Joseph, M.; Diaz, J. Fully Convolutional Neural Network for Impervious Surface Segmentation in Mixed Urban Environment. *Photogramm. Eng. Remote Sens.* **2021**, *87*, 117–123. [CrossRef]
29. Jia, J.; Liang, X.; Ma, G. Political hierarchy and regional economic development: Evidence from a spatial discontinuity in China. *J. Public Econ.* **2021**, *194*, 104352. [CrossRef]
30. Global Times. Another Turkish Consulate General Approved to be Set Up in Chengdu. 2021. Available online: https://www.globaltimes.cn/page/202107/1228068.shtml (accessed on 1 September 2022).
31. Guo, S.; Deng, X.; Ran, J.; Ding, X. Spatial and Temporal Patterns of Ecological Connectivity in the Ethnic Areas, Sichuan Province, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12941. [CrossRef] [PubMed]
32. Figueira, A.R. Rupturas e continuidades no padrão organizacional e decisório do Ministério das Relações Exteriores. *Rev. Bras. Polít. Int.* **2010**, *53*, 05–22. [CrossRef]
33. Hamama, I.; Yamamoto, M.-Y.; ElGabry, M.N.; Medhat, N.I.; Elbehiri, H.S.; Othman, A.S.; Abdelazim, M.; Lethy, A.; El-Hady, S.M.; Hussein, H. Investigation of near-surface chemical explosions effects using seismo-acoustic and synthetic aperture radar analyses. *J. Acoust. Soc. Am.* **2022**, *151*, 1575–1592. [CrossRef] [PubMed]
34. Wiki, O. Slippy Map Tilenames. Available online: https://wiki.openstreetmap.org/wiki/Slippy_map_tilenames (accessed on 1 September 2022).
35. Google Earth. Google. Retrieved January 1. 2021. Available online: https://en.wikipedia.org/wiki/Google_Earth (accessed on 1 September 2022).
36. "Openstreetmap-Website/Config/Locales at Master". Archived from the Original on 28 February 2017. Retrieved 30 September 2019. Available online: https://github.com/openstreetmap/openstreetmap-website/tree/master/config/locales (accessed on 1 September 2022).
37. "OpenStreetMapStatistics". OpenStreetMap. OpenStreetMapFoundation. Archived from the Original on 13 August 2021. Retrieved 18 October 2022. Available online: https://planet.openstreetmap.org/statistics/data_stats.html (accessed on 1 September 2022).
38. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice loss for data-imbalanced NLP tasks. *arXiv* **2019**, arXiv:1911.02855.
39. Hutchinson, M.; Samsi, S.; Arcand, W.; Bestor, D.; Bergeron, B.; Byun, C.; Houle, M.; Hubbell, M.; Jones, M.; Kepner, J. Accuracy and performance comparison of video action recognition approaches. In Proceedings of the 2020 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 22–24 September 2020; pp. 1–8.
40. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
41. Wang, S.-H.; Fernandes, S.L.; Zhu, Z.; Zhang, Y.-D. AVNC: Attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sens. J.* **2021**, *22*, 17431–17438. [CrossRef]

42. Chen, L.; Sun, Q.; Wang, F. Attention-adaptive and deformable convolutional modules for dynamic scene deblurring. *Inf. Sci.* **2021**, *546*, 368–377. [CrossRef]
43. Canayaz, M. C+ EffxNet: A novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet. *Chaos Solitons Fractals* **2021**, *151*, 111310. [CrossRef]
44. Du, Y.; Song, W.; He, Q.; Huang, D.; Liotta, A.; Su, C. Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection. *Inf. Fusion* **2019**, *49*, 89–99. [CrossRef]
45. Guo, C.; Fan, B.; Zhang, Q.; Xiang, S.; Pan, C. Augfpn: Improving multi-scale feature learning for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12595–12604.
46. Frickenstein, A.; Rohit Vemparala, M.; Unger, C.; Ayar, F.; Stechele, W. DSC: Dense-sparse convolution for vectorized inference of convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
47. Civalek, Ö. Buckling analysis of composite panels and shells with different material properties by discrete singular convolution (DSC) method. *Compos. Struct.* **2017**, *161*, 93–110. [CrossRef]
48. Recht, B.; Roelofs, R.; Schmidt, L.; Shankar, V. Do imagenet classifiers generalize to imagenet? In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 5389–5400.
49. Cazenave, T.; Sentuc, J.; Videau, M. Cosine Annealing, Mixnet and Swish Activation for Computer Go. In *Advances in Computer Games: 17th International Conference, ACG 2021, Virtual Event, 23–25 November 2021*; Revised Selected Papers; Springer International Publishing: Cham, Switzerland, 2022; pp. 53–60.
50. Misra, D. Mish: A self regularized non-monotonic neural activation function. *arXiv* **2019**, arXiv:1908.086814.
51. Reddi, S.J.; Kale, S.; Kumar, S. On the convergence of adam and beyond. *arXiv* **2019**, arXiv:1904.09237.
52. Yuan, Y.; Xie, J.; Chen, X.; Wang, J. Segfix: Model-agnostic boundary refinement for segmentation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 489–506.
53. Nash, W.; Drummond, T.; Birbilis, N. Quantity beats quality for semantic segmentation of corrosion in images. *arXiv* **2018**, arXiv:1807.03138.
54. Chang, Y.-T.; Wang, Q.; Hung, W.-C.; Piramuthu, R.; Tsai, Y.-H.; Yang, M.-H. Weakly-supervised semantic segmentation via sub-category exploration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8991–9000.
55. Shah, D.K. *Impervious Surface Probability Distribution Mapping of Kathmandu Valley*; University of Salzburg: Salzburg, Austria, 2021.

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.