Article

Urban Vegetation Extraction from High-Resolution Remote Sensing Imagery on SD-UNet and Vegetation Spectral Features

Na Lin [1], Hailin Quan [1,*], Jing He [2], Shuangtao Li [1], Maochi Xiao [1], Bin Wang [3], Tao Chen [4], Xiaoai Dai [5], Jianping Pan [1] and Nanjie Li [6]

[1] School of Smart City, Chongqing Jiaotong University, Chongqing 400074, China; linna@cqjtu.edu.cn (N.L.); listao@mails.cqjtu.edu.cn (S.L.); xiaomaochi666@mails.cqjtu.edu.cn (M.X.); panjianping@cqjtu.edu.cn (J.P.)
[2] Chongqing Liangping District Planning and Natural Resources Bureau, Chongqing 405200, China; hejingooooo@163.com
[3] Chongqing Geomatics and Remote Sensing Center, Chongqing 401125, China; arbinwang@email.swu.edu.cn
[4] School of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China; taochen@cug.edu.cn
[5] College of Earth Science, Chengdu University of Technology, Chengdu 610059, China; daixiaoa@cdut.edu.cn
[6] School of Management, Chongqing University of Technology, Chongqing 400054, China; lnj20200078@cqut.edu.cn
* Correspondence: quanhailin@mails.cqjtu.edu.cn; Tel.: +86-1354-7274-070

Abstract: Urban vegetation plays a crucial role in the urban ecological system. Efficient and accurate extraction of urban vegetation information has been a pressing task. Although the development of deep learning brings great advantages for vegetation extraction, there are still problems, such as ultra-fine vegetation omissions, heavy computational burden, and unstable model performance. Therefore, a Separable Dense U-Net (SD-UNet) was proposed by introducing dense connections, separable convolutions, batch normalization layers, and Tanh activation function into U-Net. Furthermore, the Fake sample set (NIR-RG), NDVI sample set (NDVI-RG), and True sample set (RGB) were established to train SD-UNet. The obtained models were validated and applied to four scenes (high-density buildings area, cloud and misty conditions area, park, and suburb) and two administrative divisions. The experimental results show that the Fake sample set can effectively improve the model's vegetation extraction accuracy. The SD-UNet achieves the highest accuracy compared to other methods (U-Net, SegNet, NDVI, RF) on the Fake sample set, whose ACC, IOU, and Recall reached 0.9581, 0.8977, and 0.9577, respectively. It can be concluded that the SD-UNet trained on the Fake sample set not only is beneficial for vegetation extraction but also has better generalization ability and transferability.

Keywords: Gaofen-1 imagery; deep learning; dense connection; separable convolution; SD-UNet; urban vegetation extraction; NIR

check for updates

Citation: Lin, N.; Quan, H.; He, J.; Li, S.; Xiao, M.; Wang, B.; Chen, T.; Dai, X.; Pan, J.; Li, N. Urban Vegetation Extraction from High-Resolution Remote Sensing Imagery on SD-UNet and Vegetation Spectral Features. *Remote Sens.* **2023**, *15*, 4488. https://doi.org/10.3390/rs15184488

Academic Editor: Jon Atli Benediktsson

Received: 27 July 2023
Revised: 7 September 2023
Accepted: 9 September 2023
Published: 12 September 2023

1. Introduction

In urban areas, vegetation holds a crucial position within the urban ecosystem, assuming multifaceted and considerable important roles [1,2], including alleviating environmental pollution [3–5], protecting biodiversity [6,7], regulating the ecosystem [8], reducing the urban heat island effect [9,10], and improving the overall quality of life [11]. The high-precision vegetation information is not only beneficial for decision-makers to exploit the vegetation resources but also helps them to evaluate the urban ecological environment [12], plan urban life [13], and establish sustainable cities [14]. How to efficiently, accurately, and intelligently obtain urban vegetation information has aroused the interest of many researchers.

In recent years, the rapid advancements in space optical remote sensing technology have greatly promoted the Earth Observation System [15] and it has emerged as a potent and impactful tool for acquiring regional and global vegetation information [16–19]. Currently, a variety of optical, hyperspectral, and radar remote sensing data have been widely utilized for vegetation extraction, such as Gaofen and ZY series data [20–22] from China, Sentinel series data from the European Space Agency [23–25], and Landsat series data from the United States [26]. For applying remote sensing technology to urban vegetation research, a critical aspect is to extract ground truth vegetation information from images swiftly and effectively. Traditional methods can be categorized into three kinds: visual interpretation, pixel-based classification [27], and object-oriented classification [28–30]. Manual visual interpretation is usually considered the most accurate extraction method, but there are some disadvantages such as human error, low efficiency, and high cost. The pixel-based classification method was proposed by Roger M. Herold et al. in 1983. It can identify targets based on pixel values, features, and variations. Bey classified land use temporally via the spectral and textural features of Landsat data [31]. Shen proposed a 3D Dabor-wavelet-based method to identify each pixel and finally achieved hyperspectral image classification [32]. Hadi found that maximum likelihood algorithms could achieve higher accuracy by using a combination of original bands and the Ratio Vegetation Index [27]. Although relative research demonstrates that pixel-based classification methods are more efficient than manual visual interpretation, they are limited by "the different objects with the same spectrum" and "the same object with different spectrum" problems. Some researchers also have demonstrated that object-oriented classification methods surpass pixels-based classification methods in accuracy [33–36]. Object-oriented classification methods aim to classify objects by calculating each object's features, such as image spectrum, shape, texture, and size [37]. However, the computational process is too time-consuming and labor-intensive to extract ground truth information swiftly.

With the development of artificial intelligence (AI), deep learning has emerged as a popular approach for extracting remote sensing information, owing to its remarkable precision and efficiency. Deep learning is an important branch of AI, and its development can be traced back to the 1950s. However, deep learning did not experience a breakthrough until the 21st century; many world-renowned universities have invested huge human and financial resources to conduct related research in the field. In recent years, with the proposal of deep learning theory, the Convolutional Neural Network (CNN) has ushered in an era of rapid development. CNN is one of the most successful and important deep learning algorithms whose main network architecture consists of input layer, convolution layer, pooling layer, full connection layer, and output layer [38]. The reasonable architecture ensures powerful learning abilities to automatically learn features from massive samples, making it a main algorithm for academic research such as image classification [39], semantic segmentation [40,41], and land use classification [42,43]. Additionally, a series of semantic segmentation convolutional neural networks have been proposed for providing more options to extract urban vegetation information, including the full convolution network (FCN) [44], U-Net [45], SegNet [46], etc. The study shows that deep learning algorithms extract vegetation information with high accuracy and low cost, whose extraction accuracy has obvious advantages over shallow machine learning algorithms such as Support Vector Machine (SVM) and Random Forest (RF) [47,48], and which can also monitor large-scale vegetation information [49]. However, previous research has primarily focused on visible image data, which possesses limited spectral information for effectively distinguishing vegetation and non-vegetation areas. To enhance the ability of deep learning models to identify vegetation information [50], several scholars have utilized the spectral features of vegetation in the near-infrared band (NIR) to enhance the gap between vegetation and other ground objects. Nezami employed hyperspectral data to construct a training set and determine the most efficient combination of features for tree species classification [51]; Li introduced the red-edge band of GF-6WFV when establishing the training set [52]; Chen evaluated the performance of a proposed neural network by using samples with NIR from Sentinel-2A images [53]; Xu utilized the NIR from Gaofen-2 satellite images to establish a dataset for urban green space classification [54]; Roberto E. Huerta evaluated the potential of two deep learning models for semantic segmentation of urban green spaces using 12 three-band compositions. The bands mainly include vegetation indices (normalized difference vegetation index (NDVI), two-band enhanced vegetation index (EVI2), normalized difference water index (NDWI)) and spectral single bands (near-infrared (NIR), red (R), green (G), blue (B)) obtained from the original WV2 data [55]. The findings indicate that introducing red-edge bands or relevant vegetation indices into the dataset can further enhance the accuracy of vegetation extraction. However, the most suitable sample set may vary depending on the encoder used in different deep learning networks [55]. Therefore, researchers often optimize the encoders of deep learning networks to achieve better accuracy when dealing with irregular urban vegetation boundaries, complex vegetation shapes, uneven vegetation coverage, and scattered vegetation distribution. Liu proposed a multiple architecture method (DeepLabv3plus) for extracting urban green spaces from Gaofen-2 images, employing an Atrous Spatial Pyramid Pooling (ASPP) module within the encoder to capture rich contextual semantic information [47]. Men built a new model known as Concatenated Residual Attention UNet (CRAUNet), which introduced a residual module to preserve more feature information from the original image during the feature extraction process [49]. While these methods demonstrate superior performance in extracting urban green spaces, the price in exchange is that the excessive sampling rate of atrous convolution in the ASPP module leads to the loss of details, and the introduction of the residual network module increases the parameters, resulting in a higher computer burden.

To reduce ultra-fine vegetation omissions, minimize the computational burden, and improve vegetation extraction accuracy, a deep learning network named separable densely UNet (SD-UNet) is proposed, and it utilizes the vegetation spectral features in NIR during network training. The general research process was divided into four steps. Firstly, the acquired Gaofen-1 remote sensing images are preprocessed to ensure the accuracy of subsequent experiments. Secondly, three high-resolution urban vegetation sample sets are established to research the effect of NIR on vegetation extraction results. Each sample set contains 3060 samples of 256 × 256 in size to support the training and verification of the urban vegetation extraction model. Thirdly, SD-UNet, which consists of an encoder and a decoder, is proposed. Specifically, SD-UNet introduces dense connections, separable convolutions, batch normalization layers, and the Tanh activation function. This framework can capture more ultra-fine vegetation features and reduce the computational burden. Finally, the SD-UNet is compared with other methods. The SD-UNet is applied to four different scenes and the best model is applied to two administrative divisions.

The key contributions of this article can be summarized as follows:

1. An optimized convolutional neural network (SD-UNet) was proposed to effectively extract urban vegetation from Gaofen-1 remote sensing images.
2. Three sample sets were established to evaluate the influence of the vegetation spectral features on the model extraction results. The SD-UNet was trained on three sample sets, finally obtaining the best model.
3. The SD-UNet's performance on urban vegetation extraction was compared with U-Net, SegNet, NDVI, and RF. The SD-UNet trained on three sample sets was applied to four scenes and the best model was applied to two administrative divisions to evaluate their generalization ability in vegetation extraction.

The remaining sections of this article are outlined as follows. Section 2 illustrates the proposed network structure, experimental materials, and related works. Section 3 shows the experimental results and provides a comprehensive analysis. Section 4 provides the discussion, and Section 5 offers a conclusion.

2. Materials and Methods

2.1. Study Area and Data Sources

The main urban area of Chongqing is situated in western Chongqing (Figure 1), belonging to the subtropical wet monsoon climate. It is located among the Jinyun, Zhongliang, Tongluo, and Mingyue mountains, at the intersection of the Yangtze River and Jialing River, forming the Yuzhong Peninsula. The geographical scenes are complex: the buildings in the mountainous area are densely arranged and the terrain is undulating; the urban vegetation has serious fragmentation. Therefore, vegetation information extraction is more difficult under cloud and misty conditions.

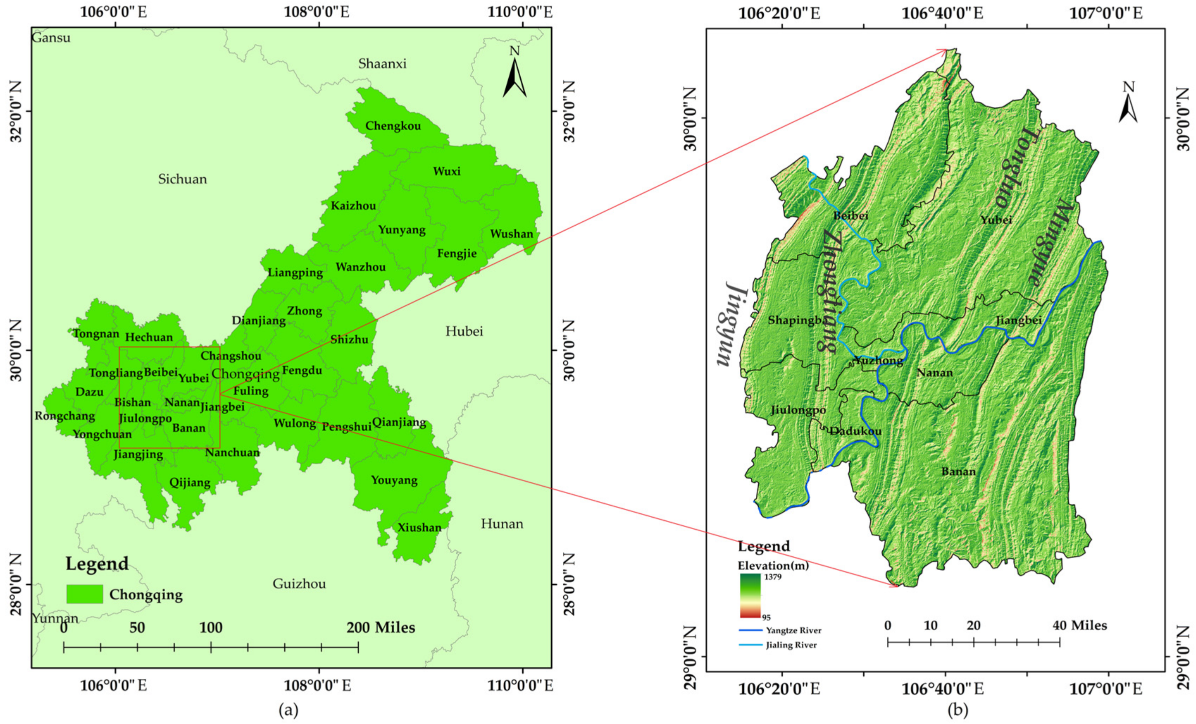


Figure 1. The study area and location (main urban area of Chongqing, China). (**a**) The geographical location of Chongqing; (**b**) Shaded relief map and DEM of the main urban area.

The Gaofen-1 satellite, launched on 31 March 2018, takes images with a panchromatic band and four multispectral bands, offering spatial resolution of 2 m and 8 m, respectively. The study data are Gaofen-1 remote sensing images of Chongqing city, obtained from the Application Center of the Chinese High-Resolution Earth Observation System. The imaging time was in September 2021 and the cloud cover was less than 5%. Images were preprocessed by radiometric calibration, atmospheric correction, orthorectification, and image fusion. Finally, a multispectral image with 2 m spatial resolution was obtained, and the digital number (DN) quantization depth is 16 bit, ranging from 0 to 65,535.

2.2. Experimental Process

The research workflow is shown in Figure 2. The research process was divided into four parts, including preprocessing the Gaofen-1 images, establishing the vegetation sample sets, training the models, and evaluating the model's generalization ability. For remote sensing image interpretation, preprocessing is a very important step. The goal of preprocessing is to eliminate or reduce the influences (such as noise, uneven brightness, and geometric distortions), ensuring the reliability and accuracy of subsequent vegetation extraction results. Therefore, in the first part, Gaofen-1 images are preprocessed by radiometric calibration, atmospheric correction, orthorectification, and image fusion. Finally, a multispectral image with 2 m spatial resolution is obtained. In the second part, three urban vegetation sample sets are established by combining different bands, namely the Fake sample set (NIR-RG), NDVI sample set (NDVI-RG), and True sample set (RGB). In the third part, SD-UNet is proposed, and the specific structure and content are shown in Section 2.5.1. Additionally, for evaluating the advantages of SD-UNet and the impacts of different vegetation spectral features on the model's vegetation extraction accuracy, this part trains different networks using different sample sets, resulting in nine models. In the last part, nine models are validated. The validation results are compared and analyzed. For evaluating SD-UNet's performance in practical application scenes, the SD-UNet is applied to four different scenes.

To make the experimental results more convincing, the vegetation of two administrative divisions is extracted by using the best model.

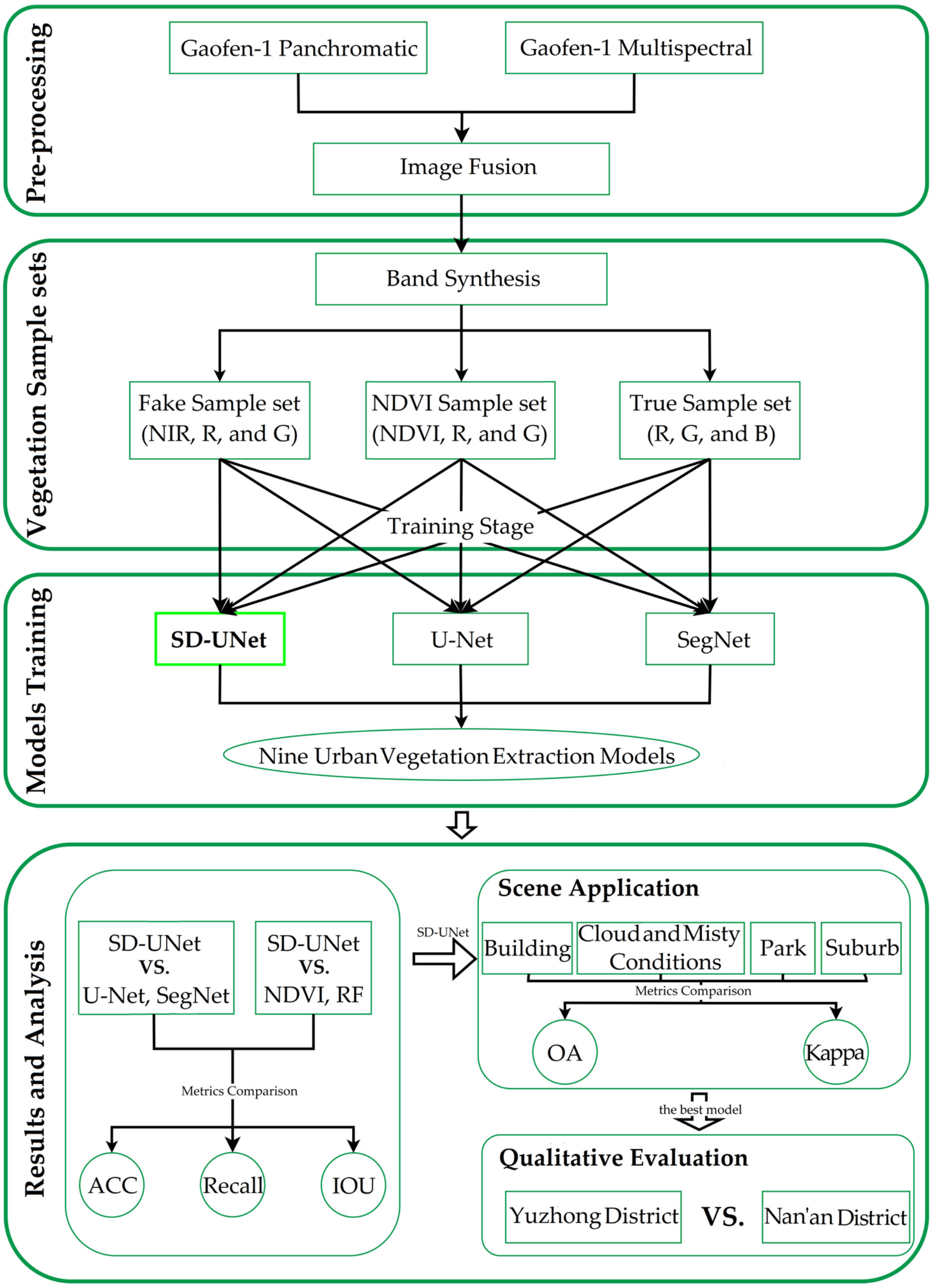


Figure 2. Research flowchart.

2.3. Sample Sets

Deep learning requires massive samples to automatically learn features and extract targets. Three urban vegetation sample sets were employed to train and validate the deep learning model. The following briefly describes the sample set.

The urban vegetation sample sets are composed of images and labels. The labels consist of vegetation and background areas, where white areas represent urban vegetation and black areas represent non-vegetation. All labels are established by manual visual interpretation. The process is mainly realized through three steps. Firstly, the vector boundaries of vegetation in the study area were delineated on Gaofen-1 images using ArcGIS 10.2 software, with reference to high-resolution Google images. Secondly, these vector boundaries were corrected through field surveys (Figure 3). Notably, not all labels were field surveyed, only the uncertain ground truth types in Gaofen-1 remote sensing images were physically verified. Finally, the results are stored in the form of raster images as labels. The labels (Figure 4a) in the sample set are binary images, where 1 represents urban vegetation and 0 represents background. Since the sample size is fixed, both images and labels were clipped to a size of 256 × 256.

Figure 3. Field survey images. (**a**) Vegetation on both sides of the road; (**b**) vegetation under the viaduct; (**c**) vegetation in the park; (**d**) vegetation on the plaza.

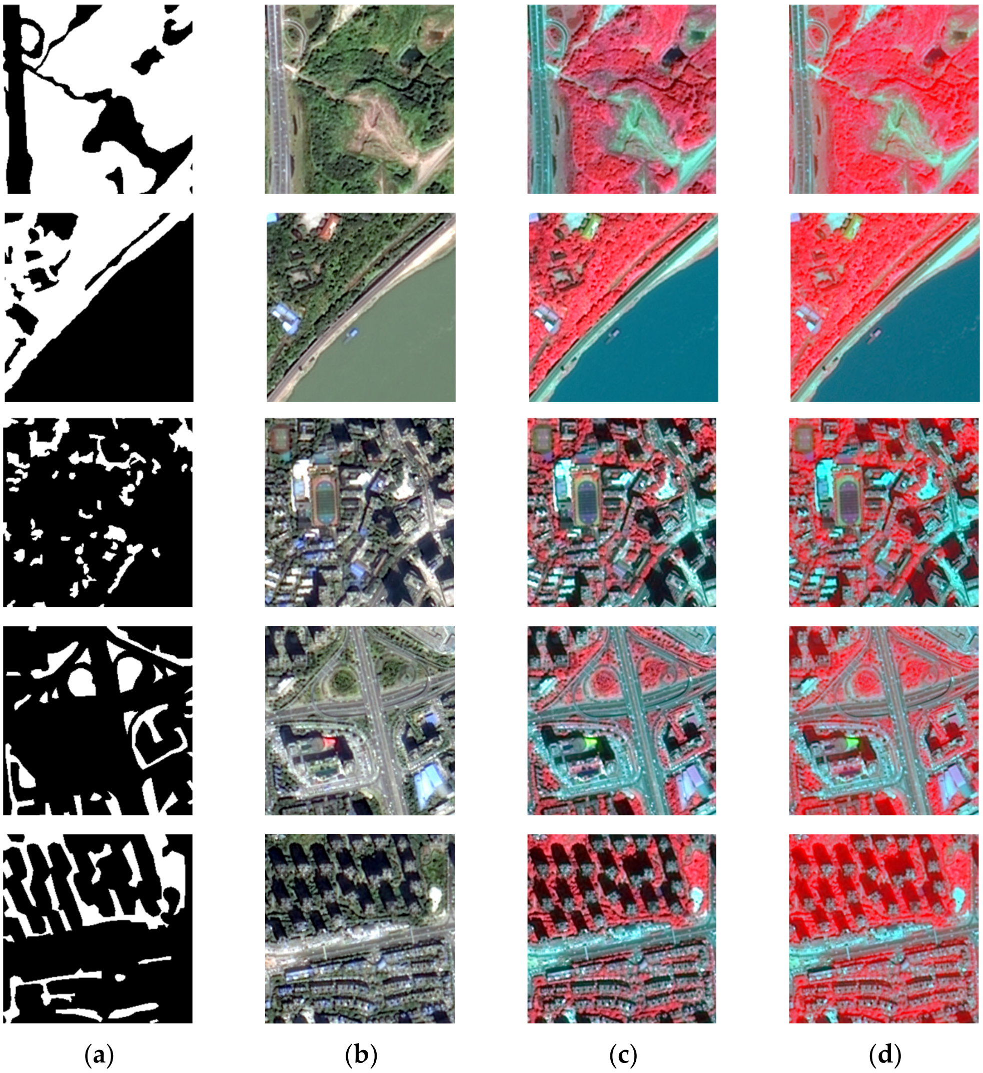


Figure 4. Sample sets. (**a**) Labels; (**b**) True sample set; (**c**) Fake sample set; (**d**) NDVI sample set. The white and black pixels in (**a**) represent the vegetation and background, respectively. And the red areas in (**c**,**d**) represent the vegetation.

Three sample sets are used to train and validate SD-UNet, as well as to calculate vegetation extraction accuracy. These sample sets include the standard false-color image vegetation sample set (Fake sample set, Figure 4c), the normalized difference vegetation index sample set (NDVI sample set, Figure 4d), and the true-color sample set (True sample set, Figure 4b). The Fake sample set was synthesized by three bands: NIR, R, and G. Vegetation is particularly prominent in standard false-color images due to its higher reflectance and distinct spectral features in the near-infrared region [56]. The NDVI sample set was synthesized by NDVI, R, and G bands. NDVI is widely applied as a key parameter to distinguish vegetation and non-vegetation areas [57,58]. Its value equals the ratio of the difference between NIR and R to the sum of NIR and R. Generally, NDVI values for vegetation are greater than 0, while most of the values for non-vegetation are less than 0 (with some exceptions, such as barren land of rock, sand, snow, and soil [59–61]). Thus, the NDVI values were determined to range from 0.355 to 0.854 through numerical statistics. Finally, the True sample set with R, G, and B bands was used as the research control group for model training and scene application. Considering that training data must contain as many different samples as possible to avoid over-fitting the generated model, data augmentation techniques such as geometric transformations (flipping, rotation, and random cropping), adding noise, and fuzzy transformations were applied. The final sample sets consisted of 2448 training samples and 612 validation samples, with a ratio of 8:2.

2.4. Scene Data

To verify the transferability and generalization ability of SD-UNet, four 1000 × 1000 pixel images with typical mountainous urban scene characteristics were used to extract vegetation and analyze the results. Notably, all scene images did not participate in the sample set construction. In addition, to make the verification results more convincing, the vegetation of Yuzhong District and Nan’an District was extracted. In order to eliminate the influence of the atmosphere condition and sunlight, four images were preprocessed via radiometric calibration, atmospheric correction, orthorectification, and image fusion.

As shown in Figure 5: (a) represents a scene image with cloud and misty conditions. Chongqing City, a famous “fog city”, the biggest interference factors are clouds and misty when extracting vegetation from remote sensing images. Hence, extracting vegetation in this scene can evaluate the transferability of SD-UNet. (b) represents a scene image located in the urban center where the vegetation and buildings are mixed. Thus, the biggest challenge is how to avoid interference from buildings in the process of extracting vegetation. Extracting vegetation in this scene can evaluate SD-UNet’s performance in accurately extracting fine vegetation within building gaps. (c) represents a suburban scene located at the edge of the city, where residential areas, factories, cultivated land, and roads are intertwined, and the vegetation is distributed randomly. Extracting vegetation in this scene can evaluate SD-UNet’s performance in a complex background. (d) represents park vegetation in the city where vegetation is concentrated and covers a large area, but is affected by interference from the lake. Extracting vegetation in this scene can evaluate SD-UNet’s performance in large-scale areas.

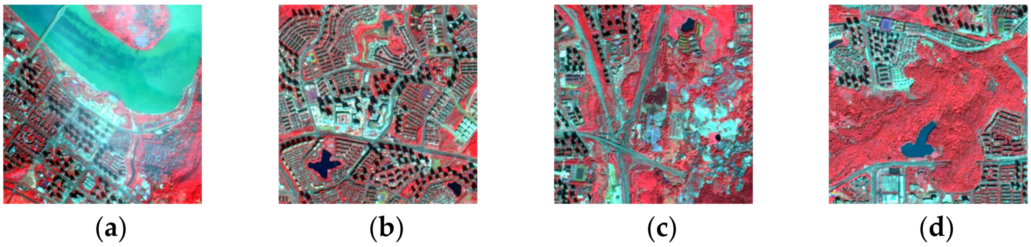


Figure 5. Four scene images. (**a**) Cloud and misty conditions scene; (**b**) high-density building scene; (**c**) suburban scene; (**d**) park scene. The red areas of the four scene images represent the vegetation.

2.5. Methods

2.5.1. SD-UNet Model

U-Net, originally introduced in 2015 by Ronneberger et al. [45], is an improved network based on a fully convolutional network (FCN), named after its U-shaped architecture. Compared to other semantic segmentation networks, U-Net possesses significant advantages, particularly its symmetric encoder–decoder structure and skip connections. The encoder mainly reduces the spatial resolution of the input image through down-sampling and gradually extracts high-level semantic features, while the decoder mainly restores the spatial resolution of the feature maps through up-sampling. Skip connections enable the decoder to directly obtain feature information from different levels of the encoder and integrate these features with its own extracted features, thereby improving the capability of image semantic segmentation. To address U-Net's overfitting and accuracy fluctuations during training with small samples, SD-UNet (Figure 6) was proposed to enhance the model's generalization ability and transferability. The letter "S" in SD-UNet indicates two meanings: one refers to the separable convolutions, and the other refers to "Septal" which represents the alternating arrangement of ordinary convolution and separable convolution. The letter "D" represents the dense connections.

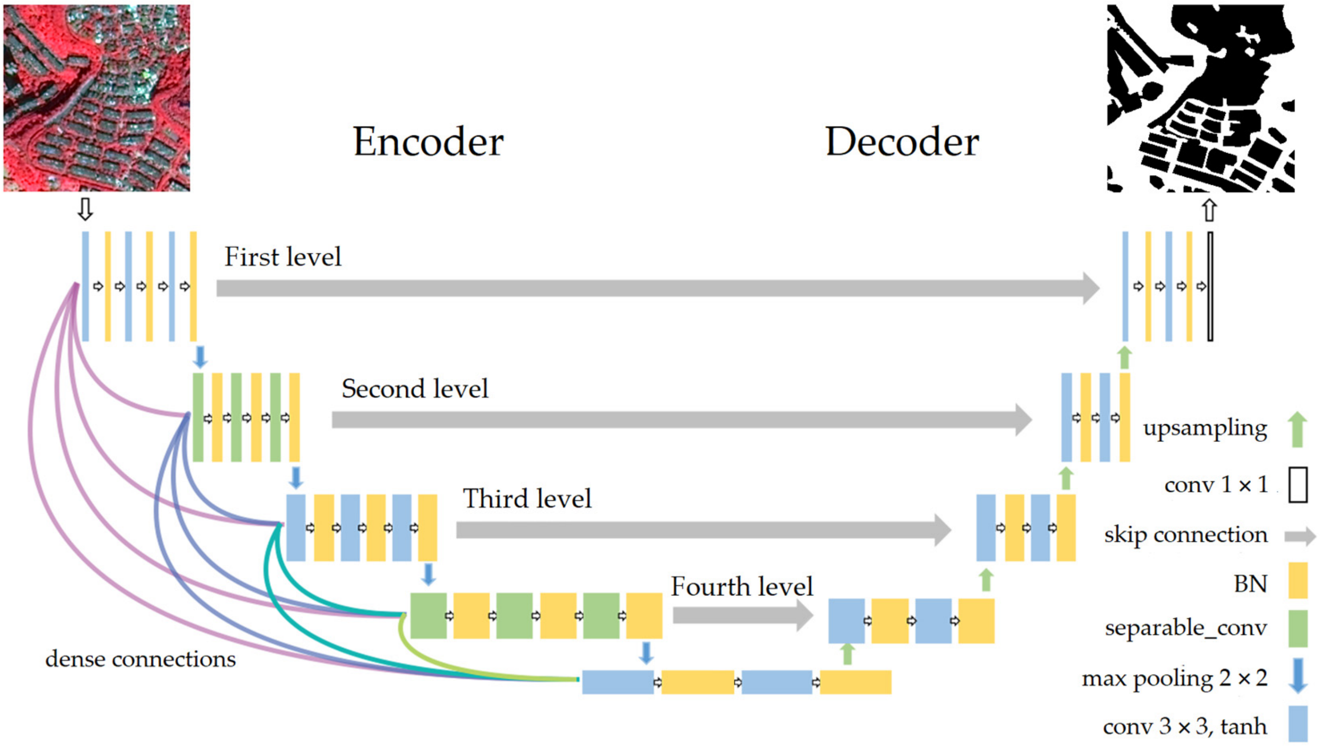


Figure 6. The structure of the SD-UNet.

The design of SD-UNet mainly focuses on optimizing the encoder by increasing the network depth and introducing dense connections to enhance vegetation extraction accuracy. For better fitting vegetation features and expanding the receptive field of the network, the number of convolutions was increased at each level of the encoder. This enhancement enables the network to capture global semantic contextual information more effectively. At the same time, to better connect the feature information between the encoders, the idea of dense connections was introduced. Specifically, a dense connection layer was added at the input of each level to enhance its relationship with all the previous levels. The dense connection can enhance the feature transfer and alleviate gradient disappearance. By introducing dense connections, the utilization of feature maps at each scale can be maximized, and low-level vegetation features can be retained as much as possible, thereby reducing the loss of fine vegetation during the convolution process.

However, the aforementioned two methods also increased network parameters, so that the computational burden was increased significantly. To address this issue, separable convolutions were introduced [62]. Compared to conventional convolutions, separable convolutions have fewer parameters, which streamline the network backbone. However, some details may be overlooked by the overuse of separable convolutions. Through research, the separable convolutions finally were positioned in the second and fourth levels of the network, which not only increases the network depth but also reduces computational parameters, while preserving SD-UNet's ability to identify and extract fine vegetation. Furthermore, a batch normalization (BN) layer [63] was introduced after each convolution layer. BN plays a crucial role in alleviating the issues of gradient disappearance and exploding, while also accelerating the convergence speed of the network. It can not only stabilize the network's training process by normalizing the inputs and outputs at each convolution operation but it can also effectively avoid overfitting and improve the network's overall performance. Particularly in scenes with limited samples, overfitting is prone to occur in traditional U-Net architectures, but SD-UNet can reduce the dependence on the data and enhance the convergence and fitting ability for urban vegetation extraction. Finally, for a better convergent and fitting network, the Relu activation function [64] was replaced with Tanh.

2.5.2. Experimental Environment

The TensorFlow framework was used to build SegNet, U-Net, SD-UNet, and train network models. The hardware configurations mainly included an Inter Core i9-10900F CPU and an NVIDIA GeForce RTX2060 GPU with 12-GB frame buffer in the experiment. The image and label sizes in the sample set were set to 256 × 256 pixels, the initial learning rate was set to 0.001, and the epoch was set to 200. Considering the limitation of GPU memory size, the batch size was set to 4. The learning rate determined the step size of parameter update during each iteration to decrease the model gradient until the best. Batch size represents the sample number in one iteration. Epoch represents the number of complete training samples.

2.5.3. Assessment Measures

A single image element is typically used to calculate accuracy evaluation metrics. Four metrics were selected to evaluate the vegetation extraction results, namely accuracy (ACC), intersection over union (IOU), Recall, and loss which is calculated by the cross-entropy loss function. ACC is the proportion of the correctly identified subjects among all ground subjects in model extraction results, IOU is the proportion of the correctly identified vegetation in all actual and identified vegetation, and Recall is the proportion of all actual vegetation that is correctly identified as vegetation.

Three metrics are calculated as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$\mathrm{IOU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2}$$

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

where TP indicates the number of vegetation regions correctly classified as belonging to vegetation (i.e., true positives), FN indicates the number of vegetation regions incorrectly classified as belonging to non-vegetation (i.e., false negatives), FP indicates the number of non-vegetation regions incorrectly classified as belonging to vegetation (i.e., false positives), and TN indicates the number of non-vegetation regions correctly classified as belonging to non-vegetation (i.e., true negatives).

3. Results and Analysis

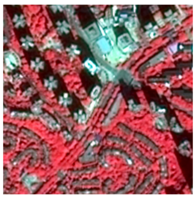
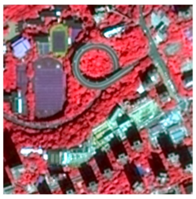
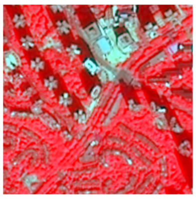
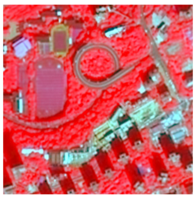
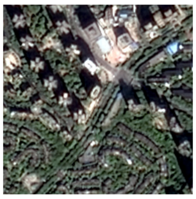
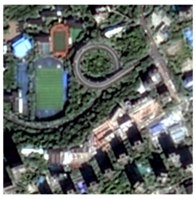
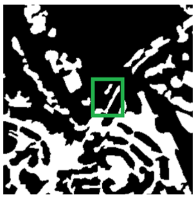
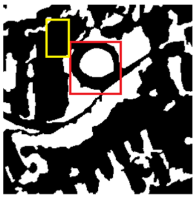
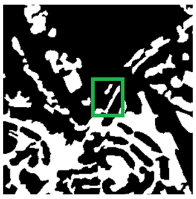
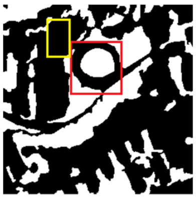
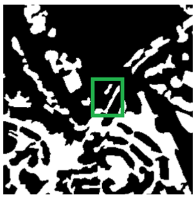
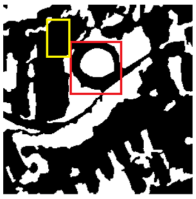
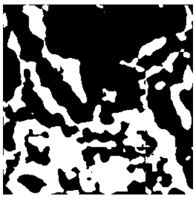
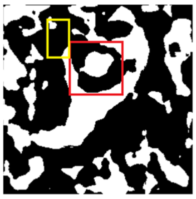
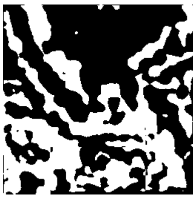
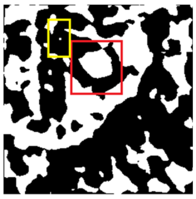
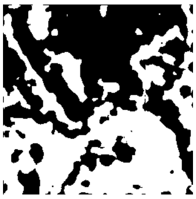
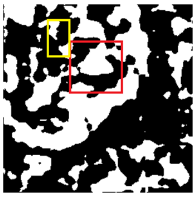
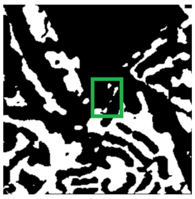
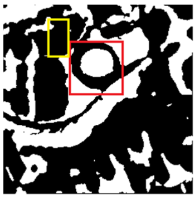
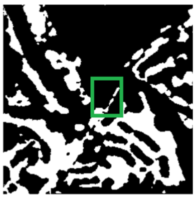
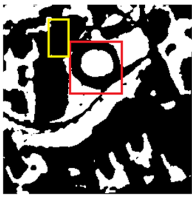
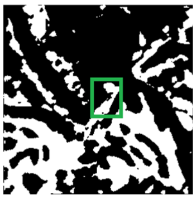
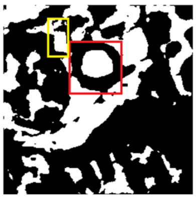
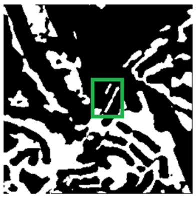
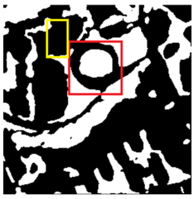
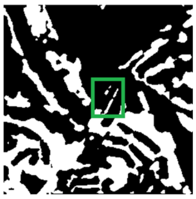
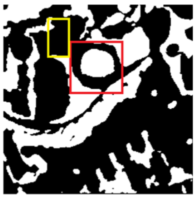
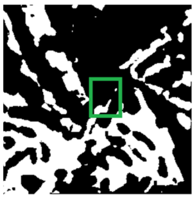
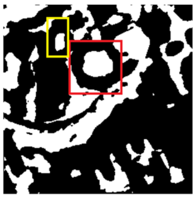
To evaluate the advantages of SD-UNet, SD-UNet was compared with SegNet, U-Net, NDVI, and RF.

3.1. Results

3.1.1. SD-UNet Results

As shown in Figure 2, nine models were obtained. In this section, the nine models are validated. The vegetation extraction results are shown in Table 1, where white areas represent vegetation and black represent non-vegetation.

Table 1. Vegetation extraction results from different deep learning methods.

Methods	Fake Samples		NDVI Samples		True Samples	
	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
Images						
Labels						
SegNet						
U-Net						
SD-UNet						

Note: The red areas of the images represent the vegetation. The colorful boxes in Block 1 and Block 2 represent the marked areas, the white pixels represent the vegetation, and the black pixels represent the background.

From the vegetation extraction results of all models, it can be seen that the SegNet performed the worst in urban vegetation extraction. The vegetation boundaries are difficult to accurately fit the labels, and the misclassification and omissions are pretty clear. Especially in block 2, there was a loss of detail in the circular road region (red box), and the extracted results were obviously different from the ground truth delineated by the labels.

In contrast, the SD-UNet and U-Net showed significantly better results in extracting urban vegetation information, with most of the extraction results matching labels. However, there were still minor differences in the extraction results for fine vegetation. In block 1, the SD-UNet showed an apparent ability to identify and extract finer strip vegetation along the sides of roads (green box). In contrast, the U-Net model performed slightly less effectively in this aspect. Taking a different perspective to compare the results obtained by models trained on three sample sets, as shown in block 2 (yellow box), the model trained on the True sample set showed obvious misclassification such as the playground, whose artificial turf was misclassified as vegetation. In contrast, the models trained on the Fake sample set and NDVI sample set successfully avoided such errors. Finally, the vegetation extraction results of the same network trained on the Fake sample set and the NDVI sample set were compared, as shown in the green box. It can be seen that the network trained on the Fake sample set performed better. In conclusion, the SD-UNet trained on the Fake sample set is the best model.

3.1.2. NDVI Results

The NDVI is a widely used traditional method in the realm of vegetation monitoring. Its principle is to segment vegetation and non-vegetation by setting a threshold. Through numerical statistics, the range of NDVI value was determined to be between 0.355 and 0.854. In the previous section, the best model was obtained. Comparing the best model with the NDVI, it can be seen that there is a clear difference in the vegetation extraction results of the two methods, as shown in Figure 7. (a) is the images, (b) is the fake samples; (c) is the labels, (d) is the extraction results of NDVI, and (e) is the extraction results of the best model. It can be clearly seen that the edges of the vegetation extracted by NDVI are not smooth enough and appear obviously jagged. The overall results did not align well with the labels, and many non-vegetation areas were misclassified as vegetation. For example, when comparing the vegetation extraction results on the sides of narrow roads, the obvious differences between the NDVI extraction results and labels can be seen just from the human eye. In contrast, the vegetation results extracted by the best model were closer to the labels. In general, the vegetation extraction results only relying on NDVI were unsatisfactory, particularly the phenomenon of “same object with different spectra” and “different objects with same spectra” in remote sensing images had an apparent impact on NDVI. Therefore, to improve classification accuracy and enhance the model’s discriminative capability between vegetation and non-vegetation, it is necessary to fuse multiple bands of information [65].

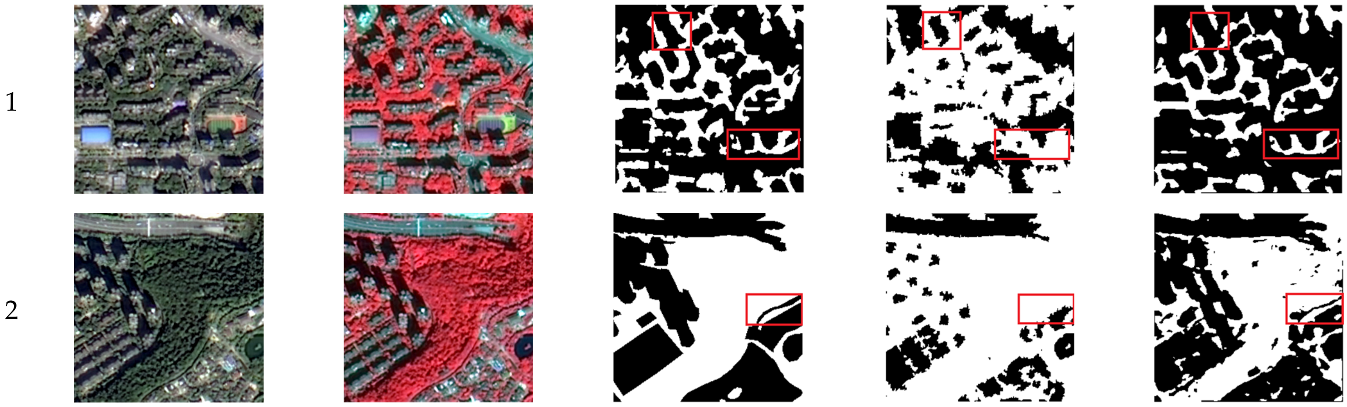


Figure 7. *Cont.*

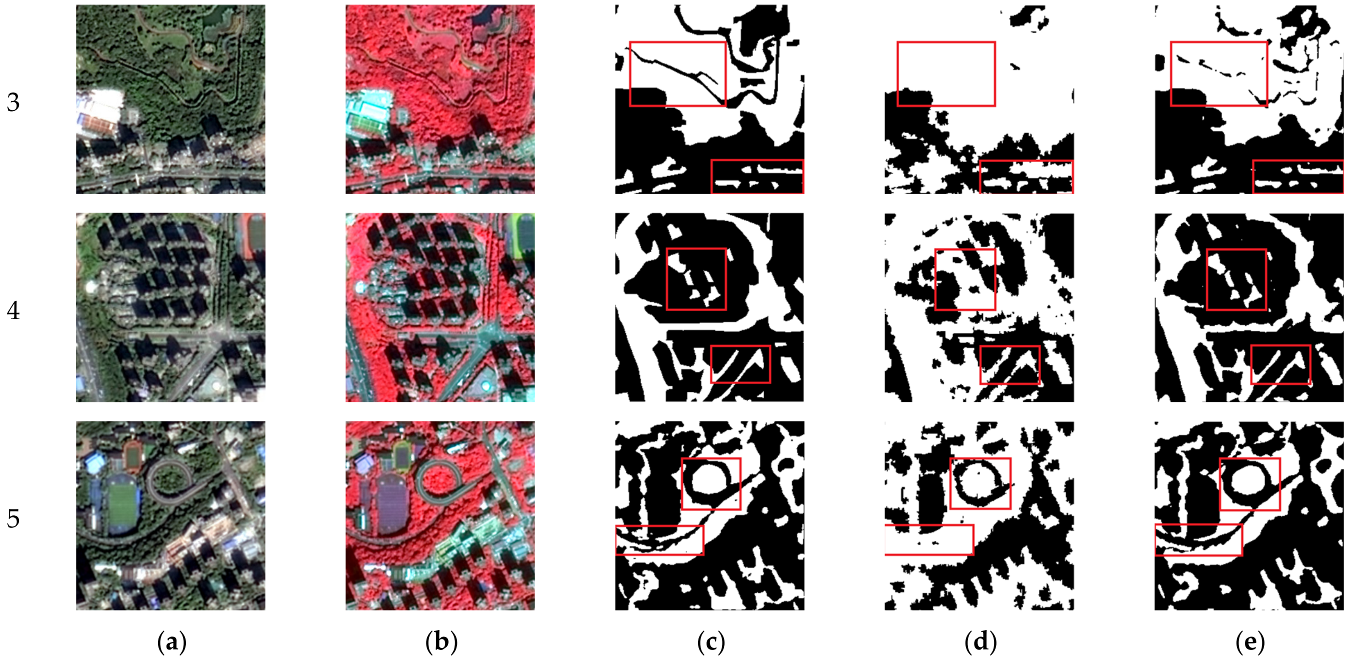


Figure 7. The urban vegetation extraction result of NDVI and the best model. (**a**) Images; (**b**) Fake samples; (**c**) labels; (**d**) the extraction result of NDVI; (**e**) the extraction results of the best model. The red areas in (**b**) represent the vegetation. The red boxes in (**c**–**e**) represent the marked areas, the white pixels represent the vegetation, and the black pixels represent the background. 1–5 represent five images in the sample set.

3.1.3. Random Forest Results

Random Forest is a relatively representative machine learning method to extract targets or specific information from remote sensing images. It has been extensively applied in various fields, including target detection, land cover classification, vegetation extraction, etc. The principle of RF is to achieve target extraction by classifying or regressing pixels in an image. The best model was compared with RF, and it was found that the vegetation extraction results of the best model were significantly better than RF. To provide reliable evaluation results, the same images were used, as shown in Figure 8. (a) shows the images, (b) shows the fake samples; (c) shows the labels, (d) shows the extraction results of RF, and (e) shows the extraction results of the best model. In Figure 8d, it can be clearly seen that the extraction results not only exhibited a “salt and pepper” phenomenon but also showed numerous omissions with noticeable “foggy” vegetation boundaries. Specifically, as shown in the red box of the fourth row, omissions on narrow roads and buildings could be observed. In contrast, the vegetation results extracted by the best model showed more integrity, and the omission phenomenon was rarely produced in the same images. In addition to the above-mentioned problems, the vegetation results extracted by NDVI and RF exhibited opposite misclassification characteristics. The former misclassified more vegetation as non-vegetation, while the latter misclassified more non-vegetation as vegetation.

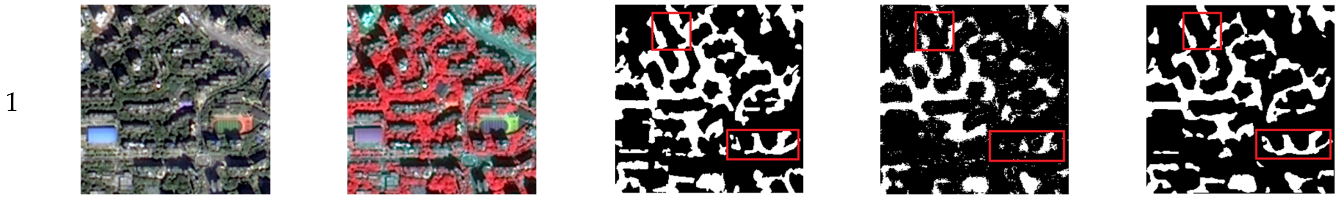


Figure 8. *Cont.*

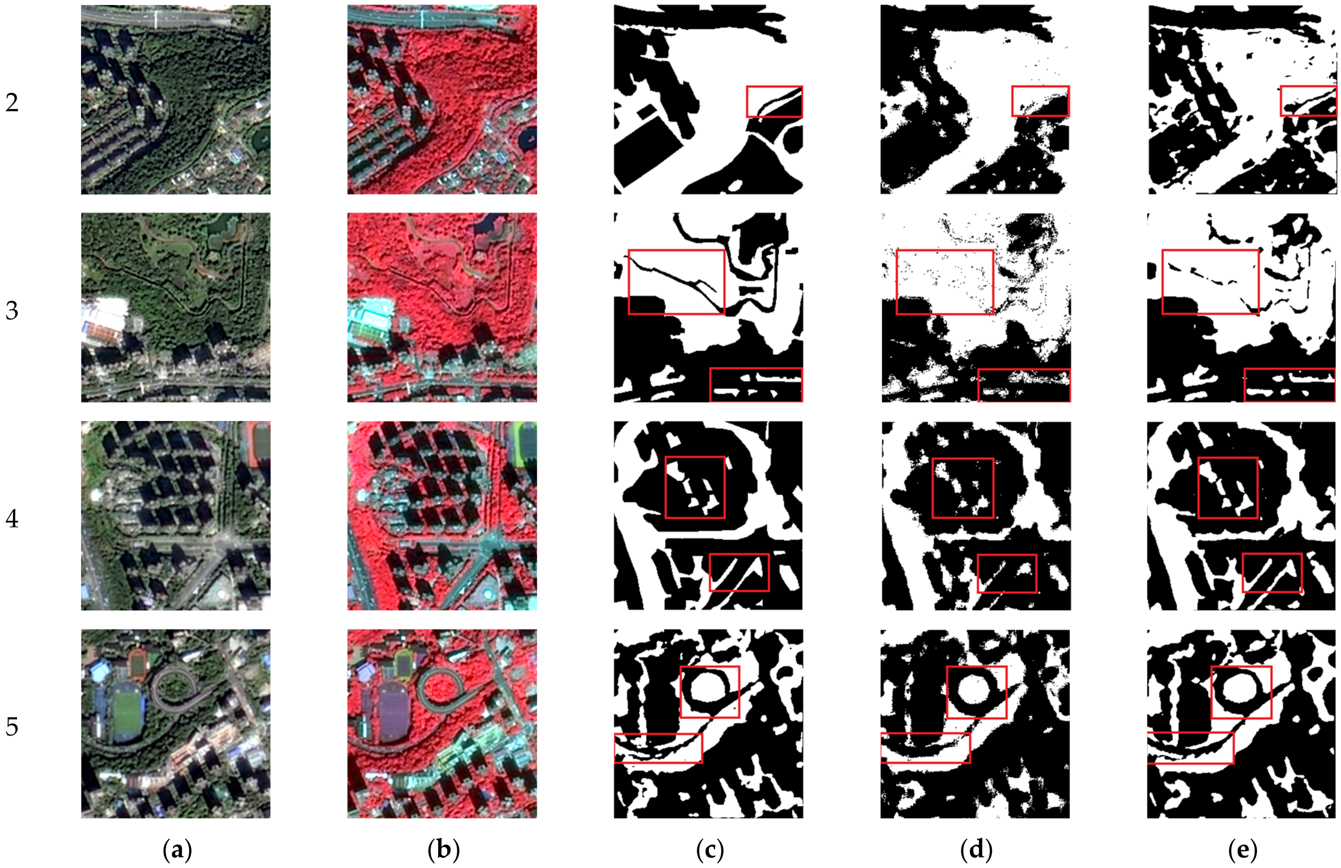
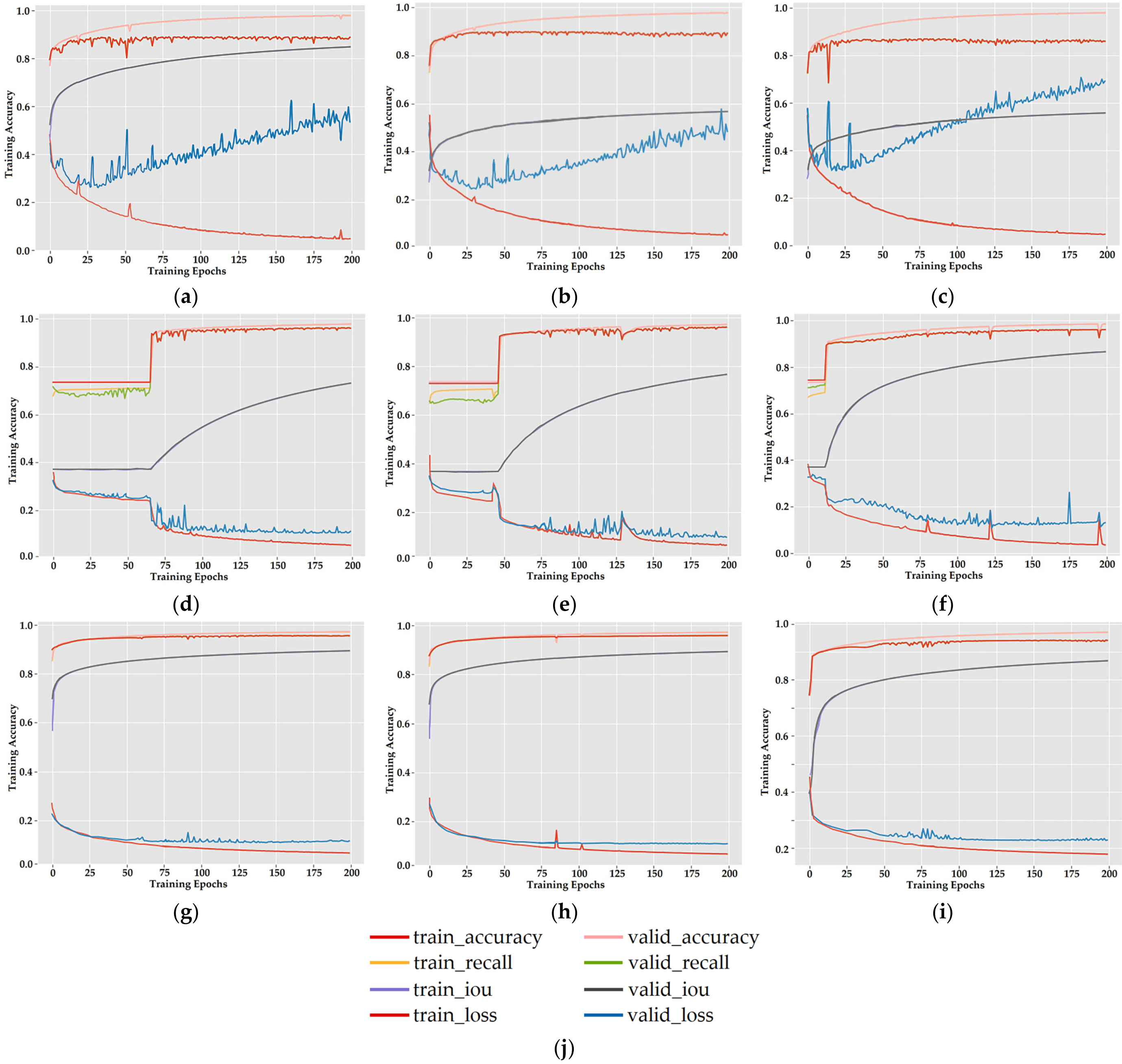


Figure 8. The urban vegetation extraction result of RF and the best model. (**a**) Images; (**b**) Fake samples; (**c**) labels; (**d**) the extraction result of RF; (**e**) the extraction results of the best model. The red areas in (**b**) represent the vegetation. The red boxes in (**c**–**e**) represent the marked areas, the white pixels represent the vegetation, and the black pixels represent the background. 1–5 represent five images in the sample set.

3.2. Analysis

3.2.1. SD-UNet vs. U-Net, SegNet

SD-UNet, U-Net, and SegNet were trained on three urban vegetation sample sets with the same hyperparameter settings. The accuracy curves show the fitting process of the network macroscopically, as shown in Figure 9. It can be seen that each model was trained for 200 epochs, and the accuracy curve trends of models obtained by training the same network with different sample sets are similar. (a–c) show the accuracy curve during the training SegNet. The three models showed noticeable fluctuations during the fitting process, and all the loss curves showed a phenomenon from a downward trend to an upward trend. (d–f) show the accuracy curve during the training U-Net. It can be seen that the accuracy of the three models remained stable during the initial few epochs. Subsequently, the loss curve sharply decreased and the other accuracy curves rapidly rose. Eventually, all the curves gradually reached a stable fitting state. (g–i) show the accuracy curve during the training SD-UNet. Compared to the previous accuracy curves, the accuracy curve of SD-UNet appeared smoother and more uniform changes. Meanwhile, SD-UNet achieved an accuracy of over 0.9 in the early few epochs, followed by a gradual increase.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**) (**h**) (**i**) (**j**)

Figure 9. The accuracy curve of nine models. (**a**) The accuracy curve of SegNet trained on the Fake sample set; (**b**) the accuracy curve of SegNet trained on the NDVI sample set; (**c**) the accuracy curve of SegNet trained on the True sample set; (**d**) the accuracy curve of U-Net trained on the Fake sample set; (**e**) the accuracy curve of U-Net trained on the NDVI sample set; (**f**) the accuracy curve of U-Net trained on the True sample set; (**g**) the accuracy curve of SD-UNet trained on the Fake sample set; (**h**) the accuracy curve of SD-UNet trained on the NDVI sample set; (**i**) the accuracy curve of SD-UNet trained on the True sample set; (**j**) legend of all accuracy curves.

Nine models extracted the urban vegetation from the validation images and calculated the accuracy based on labels. The final results are shown in Table 2. Overall, SegNet showed the lowest extraction accuracy, followed by U-Net, while SD-UNet achieved the highest extraction accuracy. In the same network, the model trained on the Fake sample set showed higher extraction accuracy than other models trained on the NDVI sample set and True sample set. In conclusion, the SD-UNet trained on the Fake sample set (best model) achieved the highest results: ACC (0.9581), IOU (0.8977), and Recall (0.9577).

Table 2. Accuracy comparison of nine models.

Method	Fake Sample Set			NDVI Sample Set			True Sample Set		
	ACC	IOU	Recall	ACC	IOU	Recall	ACC	IOU	Recall
SegNet	0.8902	0.7962	0.8895	0.8910	0.5659	0.8910	0.8673	0.5544	0.8568
U-Net	0.9432	0.7703	0.9419	0.9429	0.7658	0.9422	0.9182	0.8305	0.9163
SD-UNet	0.9581	0.8977	0.9577	0.9577	0.8942	0.9572	0.9447	0.8740	0.9442

3.2.2. SD-UNet vs. RF, NDVI

To further intuitively evaluate the best model, it was compared with traditional methods in accuracy, as shown in Table 3. The ACC of the best model achieved 0.9581, which was 0.0678 higher than the RF and 0.1269 higher than the NDVI, indicating that it showed better accuracy. The IOU of the best model achieved 0.8977, which was 0.2424 higher than the RF and 0.2981 higher than the NDVI, indicating that its extraction results had a higher overlap with the labels. The Recall of the best model achieved 0.9577, which was 0.0843 higher than the RF and 0.1453 higher than the NDVI, indicating that it had a stronger ability to correctly identify targets.

Table 3. Accuracy comparison of the best model and traditional methods.

Method	ACC	IOU	Recall
NDVI	0.8312	0.5996	0.8124
RF	0.8903	0.6553	0.8734
The best model	0.9581	0.8977	0.9577

3.3. Scene Application

To further research the generalization ability of the SD-UNet in practical application scenes, four images with dimensions of 1000 × 1000 pixels were selected. The four images represent different urban scenes, respectively, namely the high-density buildings scene, cloud and misty conditions scene, park scene, and suburban scene, as shown in Figure 10. It can be seen that the extraction results of the SD-UNet trained on different sample sets are significantly different. (1) is a scene with high-density buildings. In this scene, the vegetation extraction results of SD-UNet trained on the True sample set generally showed interference from buildings, and many buildings were misclassified as vegetation (red box). In contrast, the SD-UNet trained on the Fake sample set and the NDVI sample set had better performance. (2) is a scene with cloud and misty conditions. In this scene, the differences between the extraction results were very obvious. Especially for the vegetation along the river banks (red box), the extraction results of the SD-UNet trained on the Fake sample set were obviously better than SD-UNet trained on the other two sample sets. This is because vegetation has high reflectivity and water has strong absorption in the near-infrared spectrum, so adding the NIR to the sample set not only improves the SD-UNet's anti-interference to water but also enhances the differences between vegetation and background, thereby reducing misclassifications. (3) is the park scene, which has the highest vegetation extraction accuracy among the four images. The main drawback of the vegetation results extracted by SD-UNet trained on the True sample set was the blurred vegetation boundaries (red box) and poor integrity when extracting large-area vegetation (yellow box). The vegetation extraction results of the SD-UNet trained on the Fake sample set and NDVI sample set were similar, but the difference was the omission of fine vegetation (green box). The vegetation extraction result of SD-UNet trained on the NDVI sample set had a higher omission rate. (4) is a suburban scene, which is the most difficult area from which to extract vegetation. In this scene, the vegetation is too fragmented. It can be seen that the extraction results of the SD-UNet trained on the True sample set had many misclassifications (red box), while the extraction results of the SD-UNet trained on the NDVI sample set had many omissions (yellow box). In contrast, the extraction results of the SD-UNet trained on the Fake sample set were closer to the labels. Moreover, the OA and kappa of the three models were shown to more intuitively compare their performance, as shown in Table 4. The SD-UNet trained on the True sample set performed the worst, the SD-UNet trained on the NDVI sample set outperformed it, and the SD-UNet trained on the Fake sample set was the best model.

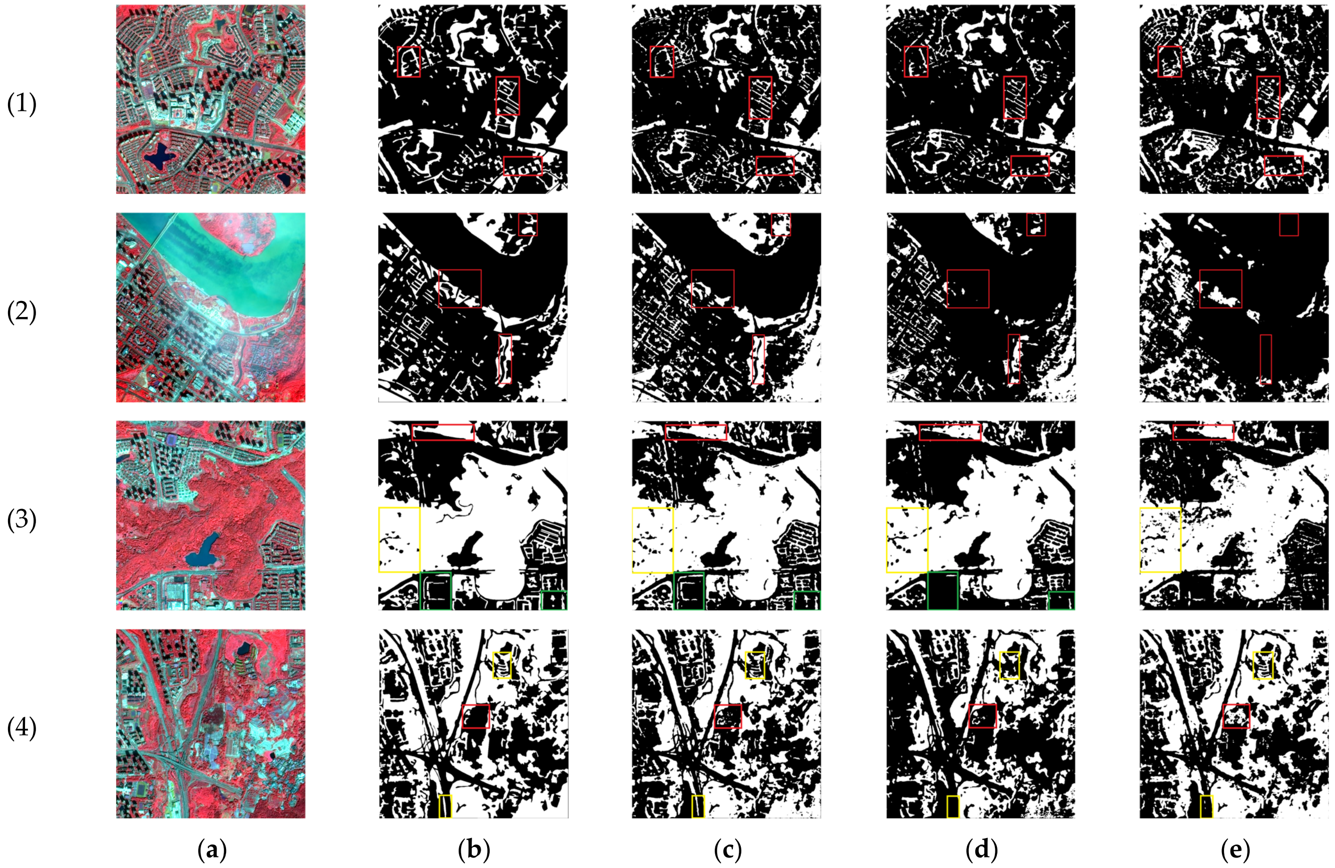


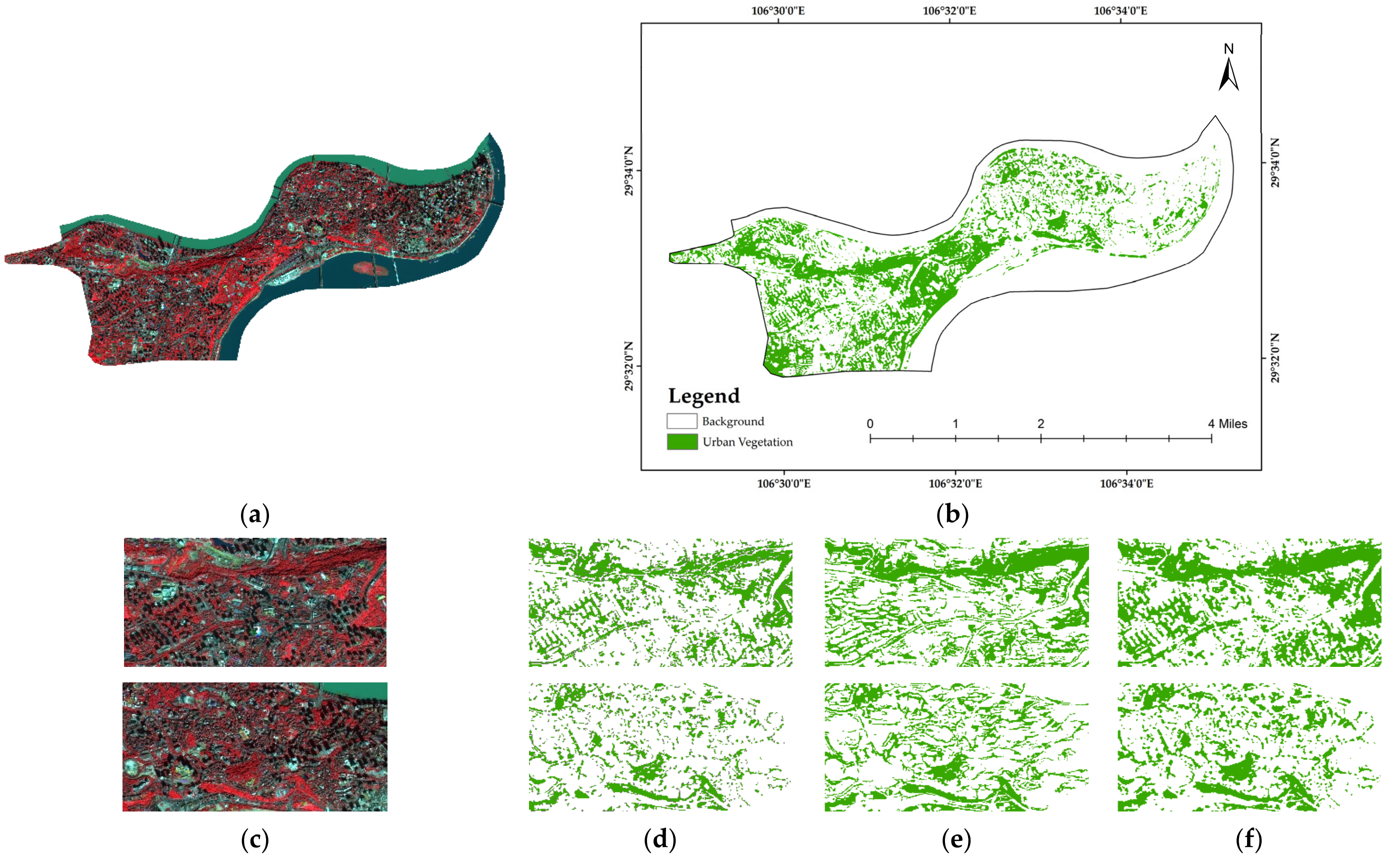
Figure 10. Vegetation extraction result of SD-UNet in four scenes. (1) High-density building scene; (2) cloud and misty conditions scene; (3) park scene; (4) suburban scene; (**a**) four scene images with standard false color; (**b**) labels; (**c**) the extraction result of SD-UNet trained on the Fake sample set; (**d**) the extraction result of SD-UNet trained on the NDVI sample set; (**e**) the extraction result of SD-UNet trained on the True sample set. The red areas in (**a**) represent the vegetation. The colorful boxes in (**b**–**e**) represent the marked areas, the white pixels represent the vegetation, and the black pixels represent the background.

Table 4. Accuracy comparison of SD-UNet trained on different sample sets in four scene images.

Scenes	Fake Sample Set				NDVI Sample Set				True Sample Set			
	Clouds and Misty	Building	Suburb	Park	Clouds and Misty	Building	Suburb	Park	Clouds and Misty	Building	Suburb	Park
OA	0.9256	0.9301	0.9214	0.9392	0.9085	0.9210	0.8605	0.9286	0.8238	0.8821	0.8817	0.8823
KAPPA	0.8631	0.8750	0.8832	0.8893	0.8055	0.8574	0.8102	0.8565	0.6589	0.7635	0.8325	0.7645

3.4. Qualitative Evaluation of Administrative Divisions

To further study the SD-UNet's applicability to regions of different scales and complexities, the qualitative analysis is carried out in this section by using the best model from Section 3.3. For more contrast, Yuzhong District and Nan'an District were selected for vegetation extraction. Both districts belong to the main urban area of Chongqing. Yuzhong District is located in the center of the main urban area, covering an area of 23.24 km^2. Nan'an District is adjacent to Yuzhong District, covering an area of 262.43 km^2. Both districts are characterized by high temperatures, limited sunshine, a long lengthy rainy season, high humidity, and the frequent presence of clouds and mist. From the image selection perspective, two districts contain the four scenes. The Yuzhong District's geography is particularly complex, with dense and scattered buildings, undulating terrain, and extremely fragmented vegetation. In contrast, the distribution of vegetation in Nan'an District is relatively good. Although the scene near the residential buildings is complicated, the vegetation in most areas is relatively complete, and the terrain is relatively flat. From the validation perspective, the urban vegetation extraction results of two districts can be evaluated visually by using the available ground truth. Due to time and human resource limitations, large-scale ground truths are lacking to evaluate accuracy, the accuracy metrics cannot be calculated. The ground truth is obtained by manual visual interpretation, so it has a certain human error. In addition, to better demonstrate the advantages of the best model, the vegetation extraction results of the best model and deep learning (DL) framework [66] are compared. The DL framework enables large-scale and general urban vegetation extraction. Its vegetation extraction results can be downloaded from https://doi.org/10.57760/sciencedb.07049, accessed on 10 September 2023. The visualization comparisons in the two districts are shown in Figures 11 and 12.



(**a**) (**b**)

(**c**) (**d**) (**e**) (**f**)

Figure 11. The vegetation extraction results of Yuzhong District. (**a**) The image of Yuzhong District with standard false color; (**b**) the vegetation coverage map of Yuzhong District; (**c**) some example images of Yuzhong District; (**d**) ground truth; (**e**) the vegetation extraction results of DL framework; (**f**) the vegetation extraction results of the best model. The red areas in (**a**,**c**) represent the vegetation. The green pixels of (**b**) and (**d**–**f**) represent the vegetation, and the white pixels represent the background.

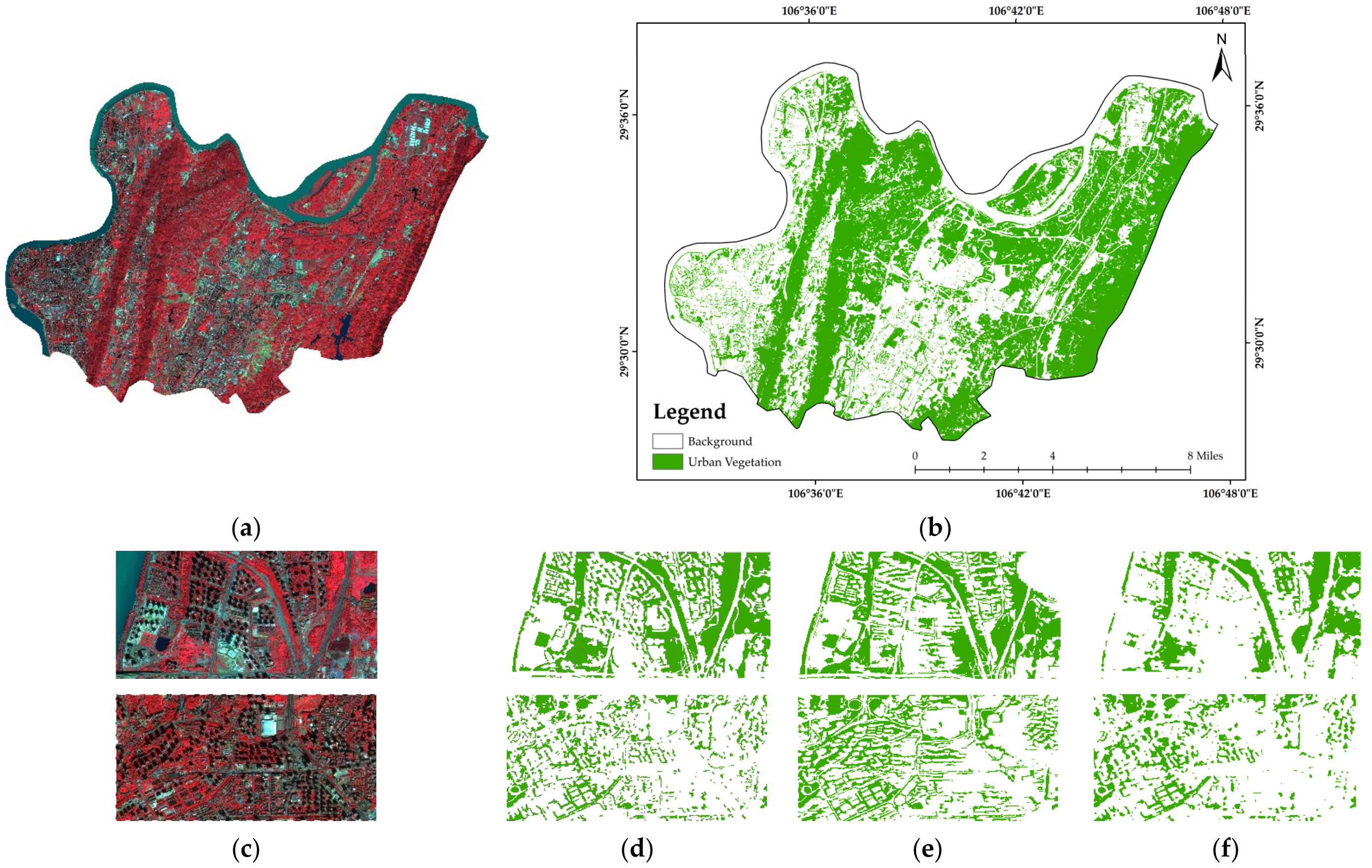


Figure 12. The vegetation extraction results of Nan'an District. (**a**) The image of Nan'an District with standard false color; (**b**) the vegetation coverage map of Nan'an District; (**c**) some example images of Nan'an District; (**d**) ground truth; (**e**) the vegetation extraction results of DL framework; (**f**) the vegetation extraction results of the best model. The red areas in (**a**,**c**) represent the vegetation. The green pixels of (**b**) and (**d**–**f**) represent the vegetation, and the white pixels represent the background.

From the overview images of Yuzhong Districts, it can be seen that the vegetation extraction results were relatively good overall. Specifically, the best model performed well in the extraction of ultra-fine vegetation alongside roadsides, riverbanks, and building gaps, although they were complex and broken in morphology. In contrast, the vegetation extraction results of the DL framework in these aspects were slightly insufficient. This may be because the DL framework is more suitable for large-scale and general vegetation extraction, so it has some limitations in ultra-fine vegetation. Similarly, the vegetation extraction results of the best model also have some shortcomings. For example, the urban vegetation extraction results of the best model are mainly influenced by the building shadows. As shown in Figure 11f, the urban vegetation features were blocked by building shadows, resulting in a relatively poor extraction effect. However, according to Figure 11e, at the same location in the image, it can be seen that the building shadows were extracted as vegetation and the vegetation was misclassified as non-vegetation. This may be caused by the different angles of the images taken, and Shi's article also validated this. Additionally, the vegetation extraction results were also affected by the vegetation sparseness. When an area is covered by sparser vegetation, it is more easily misclassified due to its similar appearance to other land types. From the overview images of Nan'an Districts, it can be seen that the vegetation extraction results were generally average. Compared to the Yuzhong District, its results are slightly inferior, especially the vegetation near dense buildings. There may be three reasons for this. Firstly, the area of Nan'an District is relatively large and the vegetation distribution is uneven, so some details may be ignored when the model captures the overall vegetation distribution. Secondly, the increase in the diversity of ground features can make it difficult for the model to accurately distinguish vegetation and non-vegetation. Finally, it can be seen from Figure 12c,f that the Nan'an District's vegetation extraction results are also affected by building shadows and vegetation sparseness. However, the vegetation extraction results of the best model also have many advantages. For example, the best model can extract large-area vegetation more completely without lake interference. When the buildings are arranged regularly, the vegetation between the building gaps can also be well extracted. In summary, despite the possible deficiencies, the best model still has a strong generalization ability.

4. Discussion

Currently, the combination of high-resolution remote sensing images and deep learning methods is widely used in semantic segmentation [67] and change detection [68]. High-resolution remote sensing images have many advantages, including providing rich geospatial information, enhancing classification accuracy, alleviating the problem of "the different objects with the same spectrum" and "the same object with different spectrum", and supporting fine-grained classification. Especially in complex scenes, high-resolution images can provide more detailed geospatial features and contextual information for finer feature classification and detection. However, traditional methods selecting features artificially are not suitable for massive data and complex feature details. In contrast, deep learning networks have become the preferred choice. Deep learning networks exhibit stronger learning capabilities and higher levels of automation, enabling automatic extraction and analysis of various features from high-resolution images, such as vegetation, buildings, and roads. In addition, deep learning models have better generalization capabilities on new scenes. However, existing networks still have some problems in extracting vegetation, such as blurred vegetation edges, misclassification, and omissions of ultra-fine vegetation. To solve these problems, SD-UNet was proposed to improve the accuracy of urban vegetation extraction. Considering the spectral features of vegetation in the NIR and NDVI, two different spectral sample sets were additionally established. In this section, the following three aspects are discussed. Firstly, the sample sets with NIR and NDVI are compared with the True sample set, and the impact of NIR, sample sets quality, and quantity on the model's vegetation extraction accuracy are discussed. Secondly, the advantages of SD-UNet's structure are discussed in detail. Finally, SD-UNet's generalization ability in four scenes and two administrative divisions is discussed.

4.1. Sample Set Evaluation

The NIR band makes the vegetation extraction model perform better, which is similar to Huerta's research results using ResNet34 to extract urban vegetation information [55]. Moreover, Ayhan's experiments also confirm this [69]. In the comparison of deep learning methods, the CNN model trained on the Vasiliko dataset (RGB and NIR bands) exhibited the best vegetation extraction results, followed by the DeeplabV3+ trained on the Vasiliko dataset (NDVI-GB). The CNN model and DeeplabV3+ trained on the Vasiliko dataset (RGB only) performed the worst. However, in previous deep learning studies, about half of the researchers only used the R, G, and B band information from remote sensing images to establish sample sets. Few researchers utilized the electromagnetic radiation features of vegetation in the NIR. Similarly, there were also fewer comparisons between the sample set with NIR and the True sample set (RGB) in the vegetation extraction results. In order to study the effect of NIR on SD-UNet's vegetation extraction results, a Fake sample set, an NDVI sample set, and a True sample set were established. Training three deep networks on three sample sets, nine urban vegetation extraction models were obtained. Then, the validation images were used to extract urban vegetation, and the extraction results and accuracy of nine models were obtained. Through the previous analysis, it has been proven that the models trained on the Fake sample set and NDVI sample set can improve the urban vegetation extraction accuracy. This is because the spectral feature sample set with NIR and NDVI can provide a more comprehensive feature representation, which can enhance models' identification and extraction capabilities for vegetation [70,71]. Moreover, the vegetation extraction results of the model trained on the Fake sample set were slightly superior to the model trained on the NDVI sample set under the same network and research area. This is because the reflectance difference between the NIR and R band was lower than normal, making NDVI unable to reach the threshold value for identifying vegetation. As a result, omissions were produced in the process of vegetation extraction. In general, the sample sets with different spectral features have many potential possibilities for the application. In the future, more detailed vegetation spectral features can be utilized to train deep learning networks, thereby achieving the classification of different vegetation types and monitoring vegetation changes (seasonal changes, degradation, etc.).

In addition, the quality and quantity of the sample set is also an important influencing factor for vegetation extraction results. Training different quality sample sets will lead to the SD-UNet's accuracy difference. Deep learning is a big data-driven approach, so sample sets are particularly important for the extraction accuracy of deep learning models. The vegetation sample set established by manual visual interpretation inevitably has certain human errors, so it is impossible to completely guarantee that the ground truths fit the labels. However, the labels were delineated on the Gaofen-1 remote sensing images by using ArcGIS and referring to the high-precision Google images. This step ensures the quality of the sample set to a certain extent. High-quality samples can provide accurate label and feature information, while an adequate number of samples can help the model learn and generalize more effectively. For better learning the statistical characteristics and patterns of the data, numerous samples are required to train deep learning networks. Therefore, the sample data were augmented by geometric transformation, adding noise, and fuzzy transformations, and 2448 training samples and 612 validation samples were finally obtained.

4.2. SD-UNet's Superiority in the Structure

The optimized SD-UNet is beneficial for urban vegetation extraction. It is an optimized version of U-Net, with several enhancements. Firstly, the SD-UNet introduces additional convolutions at each level of the U-Net's encoder, capturing deeper semantic information and enhancing the accuracy, continuity, and integrity of vegetation extraction results. Secondly, dense connections are introduced at the input side of each encoder layer. By directly connecting all layers, multiscale feature reuse can be strengthened and the information flow between layers can be maximized [72,73]. Therefore, the SD-UNet can preserve vegetation details from low levels as much as possible and reduce the omissions in ultra-fine vegetation. Thirdly, the separable convolutions are reasonably arranged. Although adding traditional convolutional layers can capture deeper semantic information, a large number of parameters are generated [74]. To further reduce the computation cost of the networks, some traditional convolutions are replaced by separable convolutions [75,76]. Fourthly, a BN layer is introduced after each convolutional layer. The BN layer can improve the network training speed [77], prevent over-fitting [78], and enhance the stability of the model [79]. Therefore, adding BN can not only save SD-UNet's training time and resources but also improve the accuracy and stability of the urban vegetation extraction model. Finally, the activation function Relu is replaced with Tanh. Experimental findings indicated that, compared to the Relu activation function, the Tanh activation function could reduce loss in SD-UNet and accelerate convergence.

4.3. SD-UNet's Applicability

To verify the generalization ability of SD-UNet in practical applications, four 1000 × 1000 pixels scene images that do not participate in the construction of urban vegetation sample sets were selected to extract vegetation. Four images represent different urban scenes: high-density buildings, cloud and misty conditions, park, and suburban scenes. To more intuitively compare the urban vegetation extraction results, overall accuracy (OA) and Kappa were used to quantify the accuracy. The experimental results show that the best model achieved the highest vegetation extraction accuracy in the same scene. In other words, the sample set with NIR can avoid interference from other ground objects. In addition, the results and analysis in the preceding sections also indicate that the vegetation extraction accuracy was the highest in the park scene and the lowest in the suburban scene and the results' contrast was the most obvious in the cloud and misty conditions scene. In the park, the vegetation is dense and has a high reflection in NIR, so the extensive vegetation areas can be more completely and accurately extracted. In contrast, the relationship among ground subjects in a suburban scene is highly complex, and the spectral difference between vegetation and non-vegetation is small, so the precision of vegetation extraction is the lowest. At the same time, the suburban environment varies from one city to another. The suburban scene only represents partial suburban features and lacks strict representativeness compared with the other three scenes. Therefore, the extraction results and conclusions about the suburban scene only have a certain reference value. In the cloudy and misty conditions scene, visible light images are significantly impacted. The cloud and mist can blur vegetation details and reduce the contrast between vegetation and non-vegetation. However, NIR with higher vegetation reflectance can penetrate clouds, so images with NIR can provide more vegetation features. In this case, SD-UNet trained on the Fake sample set can learn these vegetation features to distinguish vegetation and non-vegetation in the cloud and misty condition scene.

While the previous evaluations and comparisons have proven the best model's advantages and validity in vegetation extraction, its potential utility is not very convincing due to the smaller scene images compared to the study area. Therefore, the vegetation in Yuzhong District and Nan'an District was extracted to further display the best model's performance. Since the lack of large-scale and complete urban vegetation ground truth, a visual comparison was made. The experimental results showed that the vegetation extraction results in the Yuzhong District were good, while the vegetation extraction results in the Nan'an District were generally average. Considering the difference in complexity and area between the two regions, the following suggestion is made: To efficiently and accurately obtain urban vegetation in a large area (with complex areas and general areas), the vegetation area can be divided into several small areas according to the complexity. The best model is used to extract vegetation in complex regions, and other methods (such as the DL framework [66]) are used to extract vegetation in general regions. However, even though the best model has proven to be practicable for complex vegetation extraction, it still has to be pointed out that there may be a small number of missing labels in the sample set, especially for the vegetation covered by building shadows. As analyzed above, building shadows easily affect the model's vegetation extraction performance. This also affects the establishment of labels. Fortunately, despite the possible deficiencies, the best model can still learn different vegetation features from a large number of accurate labels due to its strong generalization ability.

5. Conclusions

To effectively solve the problem of misclassification, ultra-fine vegetation omissions, and heavy computational burden, a new convolutional neural network SD-UNet is proposed to extract urban vegetation from Gaofen-1 remote sensing imagery. At the same time, three sample sets are established to evaluate the impact of the NIR on the model's vegetation extraction accuracy, namely the Fake sample set, the NDVI sample set, and the True sample set. SD-UNet is also compared with U-Net, SegNet, NDVI, and RF on the sample sets created in Section 2.3. To ensure the fairness and reliability of the experiment, the same training data and parameters are used. Additionally, three metrics are selected to intuitively evaluate the vegetation extraction performance of the model, including ACC, IOU, and Recall. The experimental results show that the Fake sample set can effectively improve the accuracy of urban vegetation extraction results. The SD-UNet can achieve the highest performance; the ACC, IOU, and Recall of the SD-UNet trained on the Fake sample set reached 0.9581, 0.8977, and 0.9577, respectively. Based on the accuracy evaluation results, it can be concluded that the SD-UNet trained on the Fake sample set is the best model to extract urban vegetation. Finally, the SD-UNet trained on three sample sets are applied to four scenes to evaluate its generalization ability and transferability. To make the result more convincing, the best model is applied to two administrative divisions to extract vegetation. The result shows that the best model is suitable for small-scale vegetation extraction. It also shows that the research in this article can provide decision-making support for the sustainable development of the urban environment.

Although the SD-UNet can achieve satisfactory extraction results, limitations still exist. Firstly, the SD-UNet still produces omissions for ultra-fine vegetation among other ground subjects. The reason is probably that the ultra-fine vegetation information is continuously lost in the pooling process of down-sampling with the increasing network depth, so it cannot be correctly judged and identified. Secondly, the vegetation covered by building shadows also cannot be accurately identified and extracted. These problems need to be further improved in future research.

Author Contributions: Conceptualization, N.L. (Na Lin); Investigation, S.L. and M.X.; Methodology, N.L. (Na Lin) and H.Q.; Project administration, N.L. (Na Lin); Resources, J.H. and B.W.; Software, J.H. and M.X.; Supervision, N.L. (Na Lin); Validation, S.L. and M.X.; Writing—original draft, N.L. (Na Lin) and H.Q.; Writing—review and editing, B.W., T.C., X.D., J.P. and N.L. (Nanjie Li) All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Foundation of Chongqing (Grant Number: CSTB2023NSCQ-MSX0781); the National Key R&D Project of China (Grant Number: 2021YFB2600603); the Key R&D Program of Ningxia Autonomous Region (Grant Number: 2022CMG02014); the Chongqing Graduate Training Base Construction Project (Grant Number: JDLH-PYJD2019004); and the China Postdoctoral Science Foundation (Grant Number: 2021M693920).

Data Availability Statement: The data that support the findings of this study are available from the Chongqing Data and Applications Center of the Chinese High-Resolution Earth Observation System. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from http://www.map023.cn/ (accessed on 6 March 2021) with permission. The link to download the urban vegetation sample sets in this article can be found at https://github.com/FadedTree/Chongqing-urban-area (accessed on 24 July 2023).

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

References

1. Abdollahi, A.; Pradhan, B. Urban Vegetation Mapping from Aerial Imagery Using Explainable AI (XAI). *Sensors* **2021**, *21*, 4738. [CrossRef] [PubMed]
2. Guo, J.H.; Xu, Q.S.; Zeng, Y.; Liu, Z.H.; Zhu, X.X. Nationwide urban tree canopy mapping and coverage assessment in Brazil from high-resolution remote sensing images using deep learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *198*, 1–15. [CrossRef]
3. Xing, Y.; Brimblecombe, P. Role of vegetation in deposition and dispersion of air pollution in urban parks. *Atmos. Environ.* **2019**, *201*, 73–83. [CrossRef]
4. Zheng, T.; Jia, Y.P.; Zhang, S.J.; Li, X.B.; Wu, Y.; Wu, C.L.; He, H.D.; Peng, Z.R. Impacts of vegetation on particle concentrations in roadside environments. *Environ. Pollut.* **2021**, *282*, 117067. [CrossRef]
5. Lee, E.S.; Ranasinghe, D.R.; Ahangar, F.E.; Amini, S.; Mara, S.; Choi, W.; Paulson, S.; Zhu, Y. Field evaluation of vegetation and noise barriers for mitigation of near-freeway air pollution under variable wind conditions. *Atmos. Environ.* **2018**, *175*, 92–99. [CrossRef]
6. Threlfall, C.G.; Mata, L.; Mackie, J.A.; Hahs, A.K.; Stork, N.E.; Williams, N.S.G.; Livesley, S.J. Increasing biodiversity in urban green spaces through simple vegetation interventions. *J. Appl. Ecol.* **2017**, *54*, 1874–1883. [CrossRef]
7. Paiva, P.F.P.R.; Ruivo, M.D.P.; da Silva, O.M.; Maciel, M.D.M.; Braga, T.G.M.; de Andrade, M.M.N.; dos Santos, P.C.; da Rocha, E.S.; de Freitas, T.P.M.; Leite, T.V.D.; et al. Deforestation in protect areas in the Amazon: A threat to biodiversity. *Biodivers. Conserv.* **2020**, *29*, 19–38. [CrossRef]
8. Liu, G.; Shao, Q.; Fan, J.; Huang, H.; Liu, J.; He, J. Assessment of Restoration Degree and Restoration Potential of Key Ecosystem-Regulating Services in the Three-River Headwaters Region Based on Vegetation Coverage. *Remote Sens.* **2023**, *15*, 523. [CrossRef]
9. Yan, C.H.; Guo, Q.P.; Li, H.Y.; Li, L.J.; Qiu, G.Y. Quantifying the cooling effect of urban vegetation by mobile traverse method: A local-scale urban heat island study in a subtropical megacity. *Build. Environ.* **2019**, *169*, 106541. [CrossRef]
10. Susca, T.; Gaffin, S.R.; Dell'Osso, G.R. Positive effects of vegetation: Urban heat island and green roofs. *Environ. Pollut.* **2021**, *159*, 2119–2126. [CrossRef]

11. Olsen, J.R.; Nicholls, N.; Mitchell, R. Are urban landscapes associated with reported life satisfaction and inequalities in life satisfaction at the city level? A cross-sectional study of 66 European cities. *Soc. Sci. Med.* **2019**, *226*, 263–274. [CrossRef] [PubMed]
12. Lehmann, I.; Mathey, J.; Rossler, S.; Brauer, A.; Goldberg, V. Urban vegetation structure types as a methodological approach for identifying ecosystem services—Application to the analysis of micro-climatic effects. *Ecol. Indic.* **2014**, *42*, 58–72. [CrossRef]
13. Zhou, W.; Cao, F.; Wang, G. Effects of Spatial Pattern of Forest Vegetation on Urban Cooling in a Compact Megacity. *Forests* **2019**, *10*, 282. [CrossRef]
14. Du, J.Q.; Fu, Q.; Fang, S.F.; Wu, J.H.; He, P.; Quan, Z.J. Effects of rapid urbanization on vegetation cover in the metropolises of China over the last four decades. *Ecol. Indic.* **2019**, *107*, 105458. [CrossRef]
15. Rast, M.; Painter, T.H. Earth Observation Imaging Spectroscopy for Terrestrial Systems: An Overview of Its History, Techniques, and Applications of Its Missions. *Surv. Geophys.* **2019**, *40*, 303–331. [CrossRef]
16. Zhu, X.L.; Liu, D.S. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 222–231. [CrossRef]
17. Yang, L.; Jia, K.; Liang, S.; Wei, X.; Yao, Y.; Zhang, X. A Robust Algorithm for Estimating Surface Fractional Vegetation Cover from Landsat Data. *Remote Sens.* **2017**, *9*, 857. [CrossRef]
18. Chen, X.; Yang, Y.; Zhang, D.; Li, X.; Gao, Y.; Zhang, L.; Wang, D.; Wang, J.; Wang, J.; Huang, J. Response Mechanism of Leaf Area Index and Main Nutrient Content in Mangrove Supported by Hyperspectral Data. *Forests* **2023**, *14*, 754. [CrossRef]
19. Zhang, C.; Liu, Y.; Tie, N. Forest Land Resource Information Acquisition with Sentinel-2 Image Utilizing Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Trees and Multi-Layer Perceptron. *Forests* **2023**, *14*, 254. [CrossRef]
20. Mao, X.; Deng, Y.; Zhu, L.; Yao, Y. Hierarchical Geographic Object-Based Vegetation Type Extraction Based on Multi-Source Remote Sensing Data. *Forests* **2020**, *11*, 1271. [CrossRef]
21. Tang, Z.; Sun, Y.; Wan, G.; Zhang, K.; Shi, H.; Zhao, Y.; Chen, S.; Zhang, X. Winter Wheat Lodging Area Extraction Using Deep Learning with Gaofen-2 Satellite Imagery. *Remote Sens.* **2022**, *14*, 4887. [CrossRef]
22. Zhan, Z.Q.; Zhang, X.M.; Liu, Y.; Sun, X.; Pang, C.; Zhao, C.B. Vegetation Land Use/Land Cover Extraction From High-Resolution Satellite Images Based on Adaptive Context Inference. *IEEE Access* **2020**, *8*, 21036–21051. [CrossRef]
23. Abdollahi, A.; Liu, Y.X.; Pradhan, B.; Huete, A.; Dikshit, A.; Tran, N.N. Short-time-series grassland mapping using Sentinel-2 imagery and deep learning-based architecture. *Egypt. J. Remote Sens. Space Sci.* **2022**, *25*, 673–685. [CrossRef]
24. Cheng, X.M.; Liu, W.D.; Zhou, J.H.; Wang, Z.Z.; Zhang, S.Q.; Liao, S.X. Extraction of Mountain Grasslands in Yunnan, China, from Sentinel-2 Data during the Optimal Phenological Period Using Feature Optimization. *Agronomy* **2022**, *12*, 1948. [CrossRef]
25. Adagbasa, E.G.; Adelabu, S.A.; Okello, T.W. Application of deep learning with stratified K-fold for vegetation species discrimination in a protected mountainous region using Sentinel-2 image. *Geocarto Int.* **2019**, *37*, 142–162. [CrossRef]
26. Dai, Y.; Feng, L.; Hou, X.; Tang, J. An automatic classification algorithm for submerged aquatic vegetation in shallow lakes using Landsat imagery. *Remote Sens. Environ.* **2021**, *260*, 112459. [CrossRef]
27. Hadi, H.A.; Danoedoro, P. Comparing several pixel-based classification methods for vegetation structural composition mapping using Sentinel 2A imagery in Salatiga area, Central Java. In Proceedings of the Seventh Geoinformation Science Symposium, Yogyakarta, Indonesia, 25–28 October 2021.
28. Meng, X.; Shang, N.; Zhang, X.; Li, C.; Zhao, K.; Qiu, X.; Weeks, E. Photogrammetric UAV Mapping of Terrain under Dense Coastal Vegetation: An Object-Oriented Classification Ensemble Algorithm for Classification and Terrain Correction. *Remote Sens.* **2017**, *9*, 1187. [CrossRef]
29. Shen, Y.; Zhang, J.; Yang, L.; Zhou, X.; Li, H.; Zhou, X. A Novel Operational Rice Mapping Method Based on Multi-Source Satellite Images and Object-Oriented Classification. *Agronomy* **2022**, *12*, 3010. [CrossRef]
30. Zhao, F.; Wu, X.; Wang, S. Object-oriented Vegetation Classification Method based on UAV and Satellite Image Fusion. *Procedia Comput. Sci.* **2020**, *174*, 609–615. [CrossRef]
31. Bey, A.; Jetimane, J.; Lisboa, S.N.; Ribeiro, N.; Sitoe, A.; Meyfroidt, P. Mapping smallholder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique. *Remote Sens. Environ.* **2020**, *239*, 111611. [CrossRef]
32. Shen, L.L.; Jia, S. Three-Dimensional Gabor Wavelets for Pixel-Based Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 5039–5046. [CrossRef]
33. Sun, Z.P.; Shen, W.M.; Wei, B.; Liu, X.M.; Su, W.; Zhang, C.; Yang, J.Y. Object-oriented land cover classification using HJ-1 remote sensing imagery. *Sci. China Earth Sci.* **2010**, *53*, 34–44. [CrossRef]
34. Xu, Q.; Jin, M.T.; Guo, P. A High-Precision Crop Classification Method Based on Time-Series UAV Images. *Agriculture* **2023**, *13*, 97. [CrossRef]
35. Rizayeva, A.; Nita, M.D.; Radeloff, V.C. Large-area, 1964 land cover classifications of Corona spy satellite imagery for the Caucasus Mountains. *Remote Sens. Environ.* **2023**, *284*, 113343. [CrossRef]
36. Saba, F.; Zoej, M.J.V.; Mokhtarzade, M. Optimization of Multiresolution Segmentation for Object-Oriented Road Detection from High-Resolution Images. *Can. J. Remote Sens.* **2016**, *42*, 75–84. [CrossRef]
37. Xu, W.C.; Lan, Y.B.; Li, Y.H.; Luo, Y.F.; He, Z.Y. Classification method of cultivated land based on UAV visible light remote sensing. *Int. J. Agric. Biol. Eng.* **2019**, *12*, 103–109. [CrossRef]
38. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.

39. Lee, H.; Kwon, H. Going Deeper with Contextual CNN for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef]
40. Zheng, C.; Hu, C.; Chen, Y.C.; Li, J.Y. A Self-Learning-Update CNN Model for Semantic Segmentation of Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6004105. [CrossRef]
41. Zhang, H.; Zheng, X.C.; Zheng, N.S.; Shi, W.Z. Building extraction from high spatial resolution imagery based on MAEU-CNN. *J. Geo-Inf. Sci.* **2022**, *24*, 1189–1203.
42. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. Introducing Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. In Proceedings of the IEEE international geoscience and remote sensing symposium, Valencia, Spain, 22–27 July 2018; pp. 204–207.
43. Pan, S.Y.; Guan, H.Y.; Chen, Y.T.; Yu, Y.T.; Gonçalves, W.N.; Marcato, J.; Li, J. Land-cover classification of multispectral LiDAR data using CNN with optimized hyper-parameters. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 241–254. [CrossRef]
44. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
45. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI, Munich, Germany, 5–9 October 2015; pp. 234–241.
46. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
47. Liu, W.Y.; Yue, A.Z.; Shi, W.H.; Ji, J.; Deng, R. An automatic extraction architecture of urban green space based on DeepLabv3plus semantic segmentation model. In Proceedings of the International Conference on Image, Vision and Computing (ICIVC), Xiamen, China, 5–7 July 2020; pp. 311–315.
48. Lu, T.; Wan, L.; Wang, L. Fine crop classification in high resolution remote sensing based on deep learning. *Front. Environ. Sci.* **2022**, *10*, 991173. [CrossRef]
49. Men, G.; He, G.; Wang, G. Concatenated Residual Attention UNet for Semantic Segmentation of Urban Green Space. *Forests* **2021**, *12*, 1441. [CrossRef]
50. Zhou, X.; Zhou, W.; Li, F.; Shao, Z.; Fu, X. Vegetation Type Classification Based on 3D Convolutional Neural Network Model: A Case Study of Baishuijiang National Nature Reserve. *Forests* **2022**, *13*, 906. [CrossRef]
51. Nezami, S.; Khoramshahi, E.; Nevalainen, O.; Pölönen, I.; Honkavaara, E. Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 1070. [CrossRef]
52. Li, Q.J.; Liu, J.; Mi, X.F.; Yang, J.; Yu, T. Object-oriented crop classification for GF-6 WFV remote sensing images based on Convolutional Neural Network. *Natl. Remote Sens. Bull.* **2021**, *25*, 549–558. [CrossRef]
53. Chen, Y.; Weng, Q.; Tang, L.; Liu, Q.; Zhang, X.; Bilal, M. Automatic mapping of urban green spaces using a geospatial neural network. *GISci. Remote Sens.* **2021**, *58*, 624–642. [CrossRef]
54. Xu, Z.; Zhou, Y.; Wang, S.X.; Wang, L.T.; Wang, Z.Q. U-Net for urban green space classification in Gaofen-2 remote sensing images. *J. Image Graph.* **2021**, *26*, 700–713.
55. Huerta, R.E.; Yépez, F.D.; Lozano-García, D.F.; Cobián, V.H.G.; Fierro, A.L.F.; Gómez, H.D.; González, R.A.C.; Vargas-Martínez, A. Mapping Urban Green Spaces at the Metropolitan Level Using Very High Resolution Satellite Imagery and Deep Learning Techniques for Semantic Segmentation. *Remote Sens.* **2021**, *13*, 2031. [CrossRef]
56. Xie, D.J.; Lv, C.L.; Zu, M.; Cheng, H.F. Research progress of bionic materials simulating vegetation visible-near infrared reflectance spectra. *Spectrosc. Spectr. Anal.* **2021**, *41*, 1032–1038.
57. Yu, J.J.; Ji, S.X.; Li, X.L. Automatic extraction method of crop leaves from complex background based on multi/hyperspectral imaging. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 240–249.
58. Liang, S.H.; Lv, C.B.; Wang, G.J.; Feng, Y.Q.; Wu, Q.B.; Wan, L.; Tong, Y.Q. Vegetation phenology and its variations in the Tibetan Plateau, China. *Int. J. Remote Sens.* **2018**, *40*, 3323–3343. [CrossRef]
59. Sharma, M.; Bangotra, P.; Gautam, A.S.; Gautam, S. Sensitivity of normalized difference vegetation index (NDVI) to land surface temperature, soil moisture and precipitation over district Gautam Buddh Nagar, UP, India. *Stoch. Environ. Res. Risk Assess* **2022**, *36*, 1779–1789. [CrossRef] [PubMed]
60. Jin, X.M.; Guo, R.H.; Zhang, Q.; Zhou, Y.X.; Zhang, D.R.; Yang, Z. Response of vegetation pattern to different landform and water-table depth in Hailiutu River basin, Northwestern China. *Environ. Earth Sci.* **2013**, *71*, 4889–4898. [CrossRef]
61. Escobar-Flores, J.G.; Lopez-Sanchez, C.A.; Sandoval, S.; Marquez-Linares, M.A.; Wehenkel, C. Predicting Pinus monophylla forest cover in the Baja California Desert by remote sensing. *PeerJ.* **2018**, *6*, e4603. [CrossRef] [PubMed]
62. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
63. Wang, J.; Li, S.; An, Z.; Jiang, X.; Qian, W.; Ji, S. Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines. *Neurocomputing* **2018**, *329*, 53–65. [CrossRef]
64. Maas, A.L.; Hannun, A.Y.; NG, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30 th International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013; pp. 456–462.
65. Yao, J.; Wu, W.B.; Kang, T.J. Extraction method of urban vegetation information based on TM image. *Sci. Surv. Mapp.* **2010**, *35*, 113–115.

66. Shi, Q.; Liu, M.; Marinoni, A.; and Liu, X. UGS-1m: Fine-grained urban green space mapping of 31 major cities in China based on the deep learning framework. *Earth Syst. Sci. Data* **2023**, *15*, 555–577. [CrossRef]
67. Fu, J.; Yi, X.; Wang, G.; Mo, L.; Wu, P.; Kapula, K.E. Research on Ground Object Classification Method of High Resolution Remote-Sensing Images Based on Improved DeeplabV3+. *Sensors* **2022**, *22*, 7477. [CrossRef]
68. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [CrossRef]
69. Ayhan, B.; Kwan, C.; Budavari, B.; Kwan, L.; Lu, Y.; Perez, D.; Li, J.; Skarlatos, D.; Vlachos, M. Vegetation Detection Using Deep Learning and Conventional Methods. *Remote Sens.* **2020**, *12*, 2502. [CrossRef]
70. Sasidhar, T.T.; Sreelakshmi, K.; Vyshnav, M.T.; Sowmya, V.; Soman, K.P. Land Cover Satellite Image Classification Using NDVI and SimpleCNN. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019.
71. Unnikrishnan, A.; Sowmya, V.; Soman, K.P.; Buyya, R.; Sherly, K.K. Deep AlexNet with Reduced Number of Trainable Parameters for Satellite Image Classification. *Procedia Comput. Sci.* **2018**, *143*, 931–938. [CrossRef]
72. Cui, B.; Chen, X.; Lu, Y. Semantic Segmentation of Remote Sensing Images Using Transfer Learning and Deep Convolutional Neural Network With Dense Connection. *IEEE Access.* **2020**, *8*, 116744–116755. [CrossRef]
73. Tian, T.; Li, L.; Chen, W.; Zhou, H. SEMSDNet: A Multiscale Dense Network With Attention for Remote Sensing Scene Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5501–5514. [CrossRef]
74. Liu, R.; Jiang, D.; Zhang, L.; Zhang, Z. Deep Depthwise Separable Convolutional Network for Change Detection in Optical Aerial Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1109–1118. [CrossRef]
75. Song, Z.S.; Zhang, Z.T.; Yang, S.Q.; Ding, D.Y.; Ning, J.F. Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. *Comput. Electron. Agric.* **2020**, *179*, 105812. [CrossRef]
76. Zhang, T.W.; Zhang, X.L.; Shi, J.; Wei, S.J. HyperLi-Net: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 123–153. [CrossRef]
77. Shi, C.J.; Zhou, Y.T.; Qiu, B.; Guo, D.J.; Li, M.C. CloudU-Net: A Deep Convolutional Neural Network Architecture for Daytime and Nighttime Cloud Images' Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1688–1692. [CrossRef]
78. Zhao, J.L.; Hu, L.; Dong, Y.Y.; Huang, L.S.; Weng, S.Z.; Zhang, D.Y. A combination method of stacked autoencoder and 3D deep residual network for hyperspectral image classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102459. [CrossRef]
79. Zhong, Z.L.; Li, J.; Luo, Z.M.; Chapman, M. Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 847–858. [CrossRef]

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