Communication

LDANet: A Lightweight Dynamic Addition Network for Rural Road Extraction from Remote Sensing Images

Bohua Liu [1]**, Jianli Ding** [1,]*****, Jie Zou** [1]**, Jinjie Wang** [1] **and Shuai Huang** [2]

[1] College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830046, China
[2] College of Geography and Environment, Liaocheng University, Liaocheng 252000, China
***** Correspondence: watarid@xju.edu.cn; Tel.: +86-13579265967

Abstract: Automatic road extraction from remote sensing images has an important impact on road maintenance and land management. While significant deep-learning-based approaches have been developed in recent years, achieving a suitable trade-off between extraction accuracy, inference speed and model size remains a fundamental and challenging issue for real-time road extraction applications, especially for rural roads. For this purpose, we developed a lightweight dynamic addition network (LDANet) to exploit rural road extraction. Specifically, considering the narrow, complex and diverse nature of rural roads, we introduce an improved Asymmetric Convolution Block (ACB)-based Inception structure to extend the low-level features in the feature extraction layer. In the deep feature association module, the depth-wise separable convolution (DSC) is introduced to reduce the computational complexity of the model, and an adaptation-weighted overlay is designed to capture the salient features. Moreover, we utilize a dynamic weighted combined loss, which can better solve the sample imbalance and boosts segmentation accuracy. In addition, we constructed a typical remote sensing dataset of rural roads based on the Deep Globe Land Cover Classification Challenge dataset. Our experiments demonstrate that LDANet performs well in road extraction with fewer model parameters (<1 MB) and that the accuracy and the mean Intersection over Union reach 98.74% and 76.21% on the test dataset, respectively. Therefore, LDANet has potential to rapidly extract and monitor rural roads from remote sensing images.

Keywords: remote sensing image; rural roads; lightweight neural networks; Inception; dynamic weight

1. Introduction

As basic geographic information, rural roads are an important part of the transport system and play a significant role in urban planning, traffic navigation and digital map updating. With the advancement of science and technology, efficient and high-accuracy road extraction must be carried out in geographic mapping. Traditional methods for road area labeling are mostly manual and GPS-based methods, but the former is burdensome and the latter usually loses road details such as width, edges, etc. [1]

Remote sensing images have the advantages of being large-scale, fast-updating, easy to access and rich in information. With increasing resolution, the value and universality of remote sensing images have been greatly expanded. This introduces the possibility and potential of various Earth observation tasks by providing powerful data support [2]. Therefore, using remote sensing images for rapid and efficient road extraction has been a key research topic for many scholars in recent years. There are two main areas of study in remote-sensing-based road network extraction technology. The first is the shallow feature observation method, in which early scholars used the inherent geometric, textural and spectral features of images for road network extraction. However, these features are so simple that the method resulted in low accuracy [3–6], so some researchers combined it with multi-source data fusion, template matching and model orientation application methods to improve accuracy and efficiency. Zhao et al. [7] used the Extended Kalman Filter (EKF) and

Particle Filter (PF) models for road extraction, achieving acceptable performance in moderately and highly noisy backgrounds. Perciano et al. [8] used a two-layer Markov random field (MRF) to analyze multi-source fused data and to improve accuracy. Zang et al. [9] proposed a non-periodic directional structure measure (ADSM) method by introducing the representation of road-like features to enhance its effectiveness. Chinnathev et al. [10] referenced morphological features to develop an automatic road-centerline-extraction field-programmable gate array (FPGA) architecture to meet the demand for real-time road extraction. Although these studies have been effective, there are still problems with low stability and adaptability to complex information, gradient loss and overfitting of results.

The second method is the deep feature mining method, which uses multilayer networks to expand the non-linear mapping of image features and extract them. Hinton et al. [11] introduced a neural network that made the deep learning approach gain widespread attention. Compared with traditional machine learning, deep learning focuses on automatic feature learning from huge datasets with multilayer neuron organization. With the development of artificial intelligence, many scholars have been interested in the intelligent extraction of road networks. Zhong et al. [12] used full convolutional networks (FCNs) for road extraction from remote sensing images and obtained acceptable performance. Varia et al. [13] validated the effectiveness of deep learning methods by using conditional generative adversarial networks (GANs) and FCNs to extract roads from the data of unmanned aerial vehicles (UAVs). Doshi et al. [14] combined the ResNet network and the Inception network to propose a Residual Inception Skip Network and greatly improved the performance of road network extraction. Zhou et al. [15] constructed a D-LinkNet model by improving the extended convolutional layers in LinkNet, which expanded the perceptual field without reducing the resolution of the feature map. Li et al. [16] developed the hybrid convolutional network (HCN) by referring to FCN, Unet and VGG. Boonpook et al. [17] proposed a deep residual deconvolutional network with SegNet, which improves model extraction accuracy by enhancing feature relationships to overcome interference with complex scenes. Lu et al. [18] constructed a multi-scale and multi-task deep learning framework based on Unet, which concerns both road detection and centerline extraction operations, and it outperformed in deep-learning-based road extraction methods. However, it was found that while the higher complexity may achieve better performance, it may also result in greater requirements. Therefore, most of the deep mining models in road extraction are seeking to overcome the problems of training and application.

For meeting the requirements of practical applications, many lightweight networks have been proposed, with the primary intention of lightening the network in size and speed, while maintaining as much accuracy as possible. MobileNets [19–21] replaces the standard convolutions with depth-wise separable convolutions (DSC), which effectively reduces the number of parameters by decomposing the convolution operations. ShuffleNets [22,23] came up with a split–shuffle structure to speed up the model training process and enhance inter-channel correlation. ENet [24] was proposed to solve the problem of the large number of floating point operations in the network by compressing the channel, but with the resulting loss of spatial information, the accuracy does not perform well. In recent years, multiple networks have been proposed that employ optimized convolution, which can balance the network's number of parameters, inference speed and segmentation accuracy. LiteSeg [25] applies the Atrous Spatial Pyramid Pooling module to MobileNetV2 and has shown strong performance. ERFNet [26] uses residual connections and factorized convolutions to remain efficient while retaining remarkable accuracy. EDANet [27] proposes a bottleneck structure that combines asymmetric convolution with depth-wise convolution. ESPNet [28] and ESPNetv2 [29] introduce aurous convolution into their networks because they can obtain more features over larger areas using the same parameters. ELANet [30] uses an attention mechanism to strengthen different levels of information. MADNet [31] proposes a dense lightweight network by combining the residual multiscale module with an attention mechanism with the dual residual path block, which effectively reduces the model parameters and complexity. Despite these achievements, we believe that the balance

between the accuracy and efficiency of these design strategies for the extraction of single vulnerable targets still need to be improved. Designing a reasonable network structure that achieves greater accuracy with few computational resources by making full use of remote sensing image features is important to facilitate the application of deep learning to road extraction.

To achieve a lightweight, efficient and high-accuracy real-world application of rural road extraction, we designed a lightweight dynamic addition network (LDANet) based on the encoder–decoder framework. Firstly, we constructed an improved ACB-based Inception structure to extend the low-level features in the feature extraction layer. Then, we developed a deep feature association module by introducing DSC and an adaptation-weighted overlay to reduce the computational complexity. Finally, we built a typical rural road dataset to evaluate the performance and utilize a dynamically weighted combined loss function to solve the sample imbalance. The main contributions of this article are as follows:

1. A lightweight rural road extraction model is proposed and shows significant performance on two datasets, enhancing the applicability of remote sensing techniques.
2. We extended shallow features using ACB-based Inception and designed a lightweight deep correlation module by referring to DSC and an adaptation-weighted overlay.
3. We designed a dynamic hybrid loss function to improve the accuracy of unbalanced samples.

2. Data

In this paper, we evaluate the performance of the LDANet on two datasets: (1) the typical rural road dataset and (2) the Massachusetts roads dataset [32].

2.1. The Typical Rural Roads Dataset

This dataset is constructed based on the DeepGlobe Land Cover Classification Challenge dataset [33], which is a publicly available dataset of high-resolution remote sensing images that can be obtained from the website http://Deepglobe.org/challenge.html (accessed on 10 March 2022). The dataset's spatial resolution is 50 cm/pixel and includes 1000 images, each 1024×1024 in size. The training set, validation set and test set were divided from the typical rural road dataset in a ratio of 75%, 15% and 10%, respectively, and the types of roads contained in the dataset include unpaved roads, paved roads and dirt roads. In the experiment, we used the data augmentation method which includes geometric transformations (scaling and rotation) and pixel transformations (noise addition and gamma transform) [34] to expand the training set and the validation set to a size of 7500 and 1500 images, respectively, with image sizes of 256×256. In the test set, 100 images of size 1024×1024 were used. Sample images are shown in Figure 1.

2.2. The Massachusetts Roads Dataset

This dataset was downloaded from https://www.cs.toronto.edu/~vmnih/data/ (accessed on 15 December 2021). The dataset's spatial resolution is 1.2 m/pixel and the image size is 1500×1500. The Massachusetts roads dataset covers both urban and suburban areas. In this paper, we split the images from the original dataset into multiple 256×256 images and expanded the training set and validation set to 6000 and 1200 images, respectively, for the experiment. In the test set, we used 100 images of size 1500×1500. Sample images are shown in Figure 2.
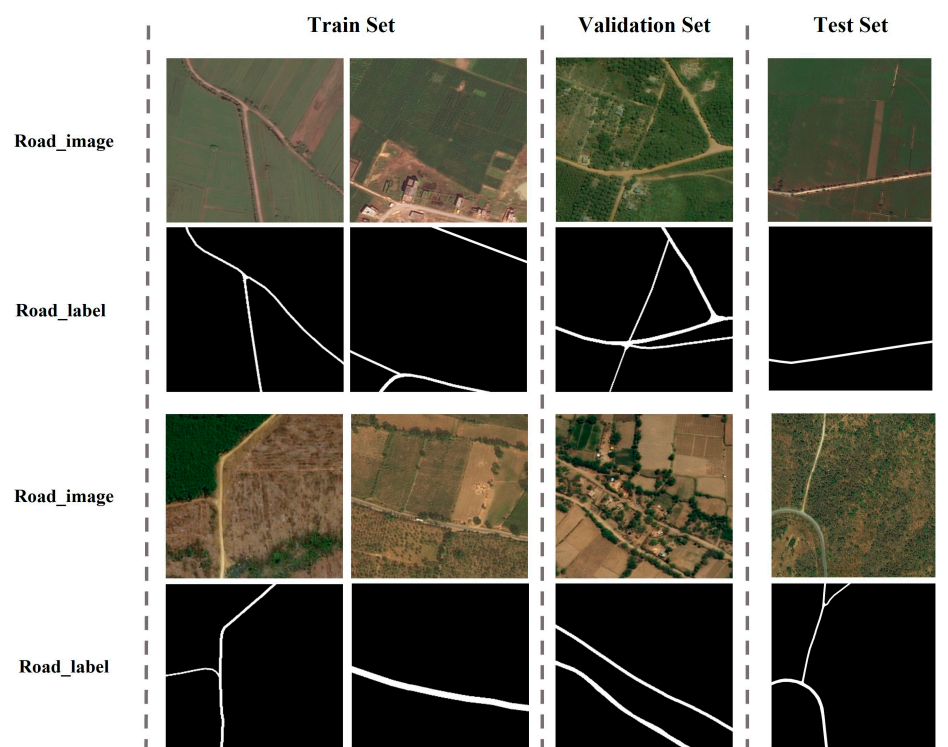
Figure 1. Sample images of the training set, validation set and test set of the rural road dataset.
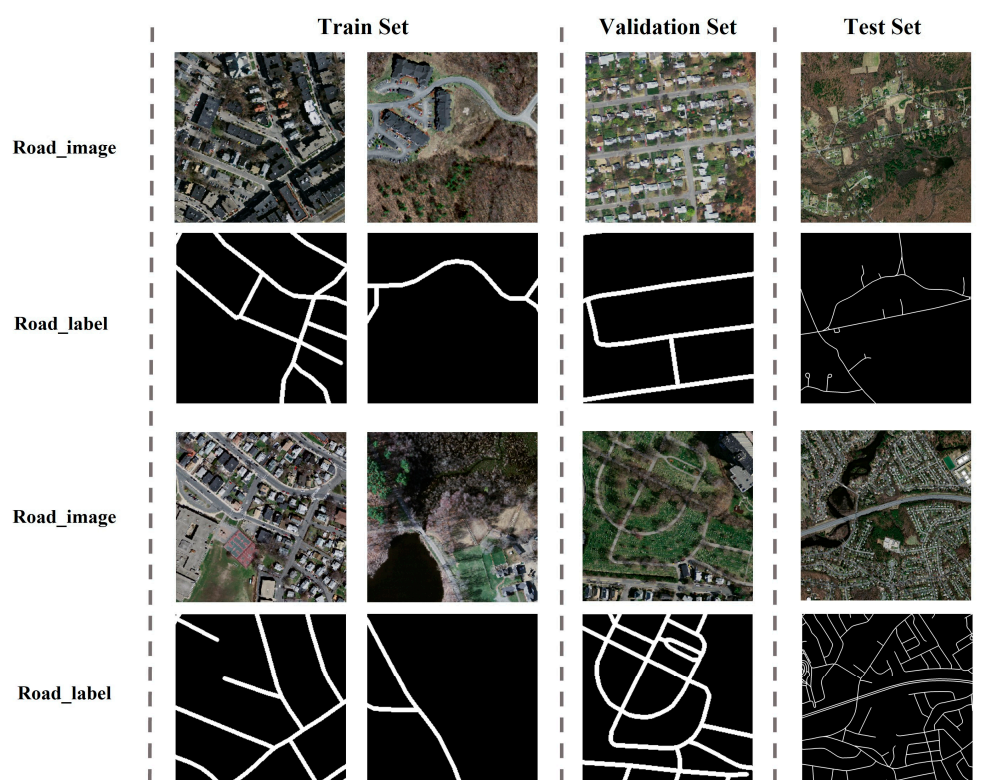

Figure 2. Massachusetts roads dataset training set, validation set and test set sample images.

3. Methodology

In this work, we aimed to construct a lightweight dynamic addition network (LDANet) for rural road extraction. LDANet consists of two modules: the feature expansion module based on the improved Inception framework and the deep feature association module

based on the DSC and dynamic feature superposition. The model framework is shown in Figure 3.
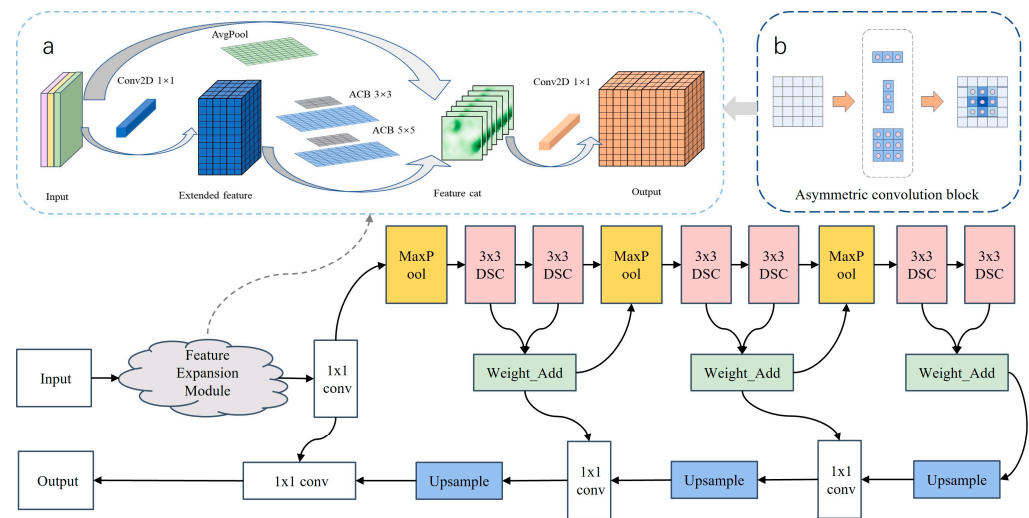

Figure 3. Structure of the lightweight dynamic addition network. (**a**) Feature expansion module. (**b**) Asymmetric convolution block (ACB), depth-wise separable convolution (DSC).

In the feature expansion module, the model obtains global and multi-scale features from the input image by averaging pooling and convolving at different scales, and uses 1×1 convolution for the fusion and dimensionality enhancement of feature layers. The deep feature association module is divided into two stages: the encoder and the decoder. In the encoder stage, six convolutional layers are constructed to extract the deep features of the expanded feature layers, where every two layers are pooled once to compress the image size, and the two feature layers are dynamically weighted and superimposed. In the decoder stage, the image size of the deep feature layers is expanded by upsampling, then the expanded layers are fused with the upper superimposed layers, and finally, a result map the same size as the input data is produced.

3.1. Feature Expansion Module

Based on previous research [35], we have found that road texture, color and geometric features have a positive effect on road extraction. Rural roads are narrow, complex and diverse, so enriching shallow information is essential and important. To fully reflect the performance of shallow features, we built a feature expansion module based on the Inception network module (Figure 3a). First, the input layer is up-dimensioned using a 1×1 convolution kernel, which not only reduces the convolution parameters but also integrates features across channels to ensure the validity of the features while simplifying the model. Then, we use the asymmetric convolution block (ACB) [36] to expand the feature layers. The ACB (Figure 3b) contains three branches—a 3×1 horizontal kernel, a 1×3 vertical kernel and a 3×3 global kernel—and each branch extracts the layer features separately and then fuses the results. Compared with double-layer convolution, ACB not only reduces the model parameters but also improves the saliency of node features while ensuring feature globalization. As roads have certain geospatial correlations, expanding the perceptual field can further enhance the road feature information, so we set the convolutional kernel size to 3×3 versus 5×5. Finally, the features are up-dimensioned and output by 1×1 convolution.

3.2. Deep Feature Association Module

As shown in Figure 4b, we use a multi-layer convolutional network for deep feature extraction and to enhance feature association with dynamic weighted superposition. This module is based on the Unet [37] network framework (Figure 4a). In the encoder stage, we

expand the feature space dimension by 3×3 convolution to enhance the deep information and then compress the feature layer size by MaxPool to reduce the model's computational complexity and to improve the computational efficiency. In the decoder stage, we recover and output the image by jump connection and upsampling. In this module, we introduce an adaptation-weighted overlay layer. In the encoder stage, the layer has an overlap and dynamically adjusts the weights of the convolutional output feature layer, which increases the expressiveness of salient features and improves the efficiency of feature extraction. In the decoder stage, dynamically weighted overlay layers and 1×1 convolution realize the jump connection in the network. Compared with the complex jump connection in Unet++ [38] and Unet+++ [39], this method can not only strengthen the association of salient features but can also better preserve the structure of the feature space and achieve model simplification based on ensuring the accuracy of extraction. As shown in Figure 4b, this module takes fewer feature dimensions than Unet because, on the one hand, the road is a two-class model, and too rich a feature space will cause greater calculation pressure, and on the other hand, this module uses DSC instead of traditional convolution calculation to reduce model parameters. Because one convolution kernel is responsible for one channel in DSC, and one channel is only convolved by one convolution kernel, the number of features mapped is the same as the number of channels in the input layer [40]. So, we added an intermediate feature layer to the DSC, first by up-dimensioning the input features by 1×1 convolution, then using DSC to generate the intermediate feature layer, and finally using the 1×1 convolution output; the convolution process is shown in Figure 5.
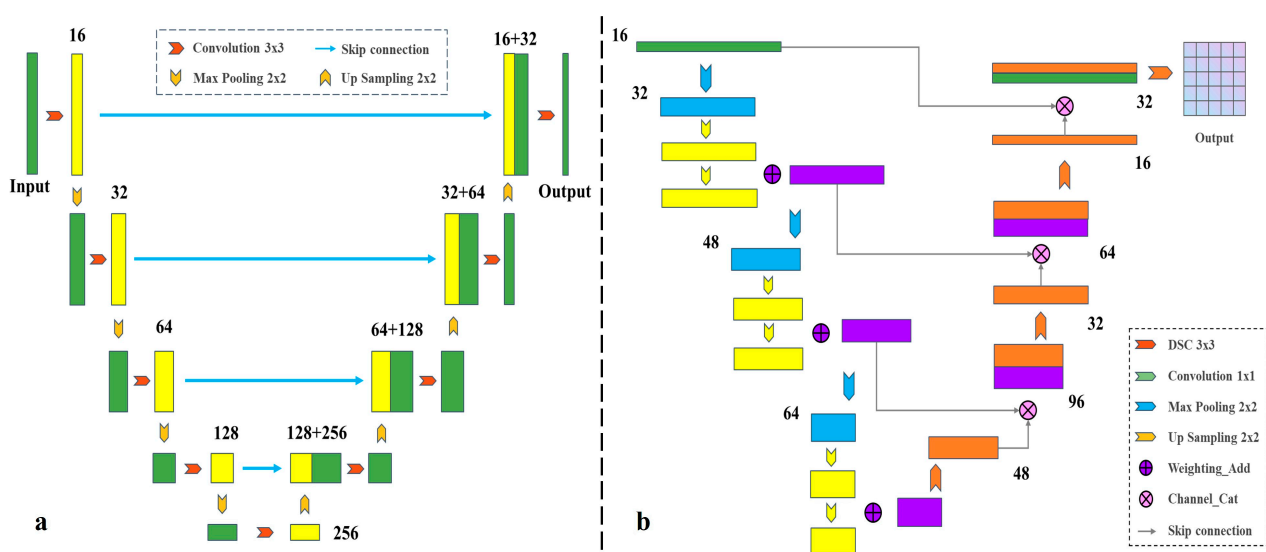

Figure 4. Module structure of (**a**) Unet, and (**b**) deep feature association module.
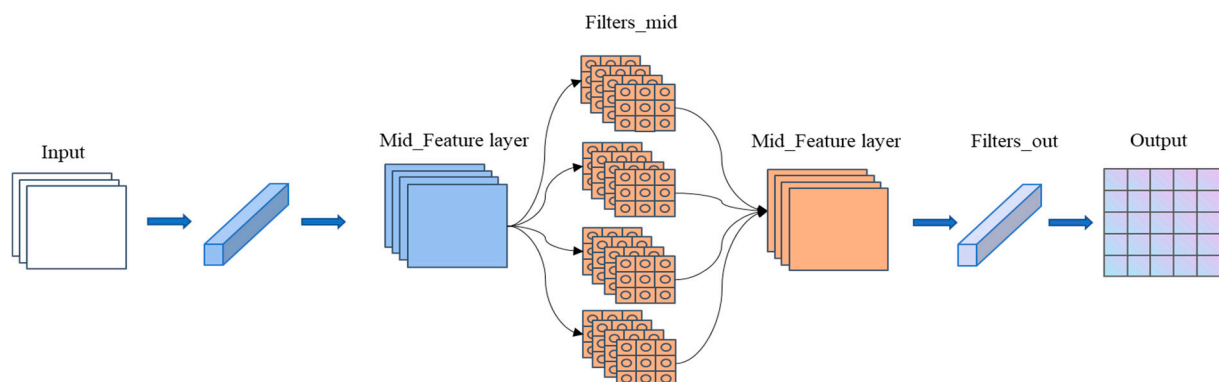

Figure 5. Depth-wise separable convolution (DSC) process.

3.3. Loss Function

Loss functions are often used to measure the extent to which a model's predictions differ from the actual data, and selecting the correct loss function can guide model training in the right direction.

Cross-Entropy Loss is a loss function that evaluates the difference between the probability distribution and the true distribution of the current training set to judge the training effect on the model. The model in this paper is dichotomous, so Binary Cross-Entropy Loss (BCE Loss) was chosen as the loss function.

$$\text{BCE Loss} = -\frac{1}{N} \sum_i [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \tag{1}$$

where y_i is the sample value, the positive class is 1 and the negative class is 0. p_i is the predicted value, taking values within (0, 1).

Roads behave narrowly in the dataset, covering a small area, and can suffer from severe sample imbalance during training. Dice Loss solves the problem of having too small a proportion of foregrounds by measuring the overlap of two samples, and it has acceptable performance in binary classification problems.

$$\text{Dice Loss} = 1 - \sum_1^N \frac{2y_i p_i + \varepsilon}{y_i + p_i + \varepsilon} \tag{2}$$

where y_i is the sample value, the positive class is 1 and the negative class is 0. p_i is the predicted value, taking values within (0, 1).

Dice Loss has acceptable performance for scenarios with a severe imbalance between positive and negative samples, but the loss tends to be unstable when training small targets, which leads to drastic gradient changes. Therefore, this paper proposes a Combined Weighted Loss (CWL), which sets the weights according to the ratio of BCE Loss to Dice Loss, giving higher weights to functions with larger loss values and lower weights to functions with smaller loss values, to increase the proportion of high-value loss functions while using low-value loss functions to maintain the stability of the model, thereby accelerating the convergence of the model and improving its accuracy.

$$\text{CWL} = W_{CB} \cdot \text{BCE Loss} + W_{CD} \cdot \text{Dice Loss} \tag{3}$$

$$W_{CB} = \frac{|\text{BCE Loss}_n|}{|\text{BCE Loss}_n| + |\text{Dice Loss}_n|} \tag{4}$$

$$W_{CD} = 1 - W_{CB} \tag{5}$$

where W_{CB} is the proportion of BCE Loss, and W_{CD} is the proportion of Dice Loss.

4. Experimental Study

4.1. Model Evaluation Criteria

In this paper, four precision metrics, Precision, Recall, F1 Score and IoU (Intersection over Union), were chosen to evaluate the road extraction results, and their expressions are as follows.

Precision is the ratio of the area of the real road area in the resulting image to the area of the road in the labeled image.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Recall is the proportion of correctly identified roads to the total roads in the tagged image.

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

The F1 Score, a statistical measure of the accuracy of a dichotomous model, is often used to determine the overall performance of a dichotomous model.

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

IoU is the degree of overlap between the pre-target measurement results and the target label.

$$\text{IoU} = \frac{\text{Result(Road)} \cap \text{Label(Road)}}{\text{Result(Road)} \cup \text{Label(Road)}} \tag{9}$$

where TP (true positive) represents a positive label and a positive prediction, FP (false positive) represents a negative label and a positive prediction and FN (false negative) represents a positive label and a negative prediction.

4.2. Loss Function Selection

In this section, we use BCE Loss, Dice Loss, BCE Loss + Dice Loss and CWL as model loss functions and test them on the rural road dataset. The experimental results were evaluated using both Precision and IoU, and the results are shown in Table 1. From the comparison of the single-loss-function applications, it can be seen that BCE Loss performs better in terms of extraction accuracy compared to Dice Loss, but Dice Loss has better performance in terms of IoU, which is because Dice Loss focuses more on the overlapping area of image samples, while BCE Loss focuses more on the global performance. The combined loss results show that the Precision and IoU of BCE Loss + Dice Loss and CWL are higher than that of using a single loss function, indicating that the combination of loss functions can effectively improve the performance of the model. The results show that the Precision and IoU of CWL are the highest, which shows that the dynamic weighting method can give full play to the advantages of the combination of loss functions, so that the model in this paper can fully utilize the advantages of Dice Loss in the secondary classification while ensuring stability and improving the applicability of the model to situations of serious imbalance between positive and negative samples. Therefore, CWL was selected as the loss function for road extraction in this paper.

Table 1. Effect of different loss functions on road extraction results.

	Precision	IoU
BCE Loss	0.9862	0.7322
Dice Loss	0.9850	0.7561
BCE Loss + Dice Loss	0.9866	0.7605
CWL	0.9874	0.7621

4.3. Results and Discussion

To estimate the performance of our model, we compared the model with five other models: Unet, Unet++, Unet+++, MACUnet [41] and MobileNet. Unet++, Unet+++ and MACUnet are all Unet-based networks, with Unet++ and Unet+++ refining the jump layer and enhancing global feature linkage, and MACUnet improving upon Unet with the ACB module and multi-scale jumps to enhance network feature acquisition. MobileNet was used as a benchmark for the comparison of lightweight models. All experiments were implemented on an NVIDIA Tesla P100 GPU with Adam as the optimizer and a learning rate of 0.001. LDANet uses CWL as the loss function and the rest of the models use a Cross-Entropy Loss function.

4.3.1. Results of the Typical Rural Roads Dataset

In this section, we compare LDANet with five other image segmentation models on the rural road dataset. Table 2 shows the results of four accuracy evaluation metrics, Precision, Recall, F1 Score and IoU, for the six models on the rural road test set. Since we aimed to

build fast and efficient lightweight models for rural road extraction, it was necessary to evaluate the model complexity, so we conducted statistics on the training time and model parameters for each of the six models mentioned above to measure the model performance.

Table 2. Comparison of six models evaluated on the rural roads dataset.

	Precision	Recall	F1 Score	IoU	Parameters (M)	Train Time/Epoch (S)
Unet	0.9754	0.9728	0.9741	0.7482	9.85	580
Unet++	0.9831	0.9822	0.9826	0.7593	11.80	1318
Unet+++	0.9881	0.9875	0.9878	0.7644	6.75	1530
MACUnet	0.9840	0.9817	0.9829	0.7617	5.15	725
MobileNet	0.9683	0.9632	0.9657	0.7431	0.17	178
LDANet	0.9874	0.9870	0.9872	0.7621	0.20	183

From Table 2, Unet++ has the highest Precision and IoU, scoring 0.9881 and 0.7644, respectively. This is because it deepens the network context and can fully learn the characteristics of the target. However, the complex network structure degradation increases its operation difficulty, making it not perform well in terms of model parameters and training time. In terms of training efficiency, MobileNet uses DSC to replace the traditional convolution process, which reduces the cost of convolution and makes it perform best in terms of model parameters and training time. However, due to image restoration only occurring through the upsampling method, the global feature association of the target is insufficient in the modeling process, resulting in low extraction accuracy.

In terms of model accuracy, LDANET has a Precision of 0.9874 and an IoU of 0.7621, which are 1.98% and 2.13% lower than Une+++, respectively, but LDANET greatly improves training efficiency with a model parameter count of 0.20 M and a training speed of 183 s/epoch, thus having 3375% fewer parameters and being 836% faster than Unet+++. Meanwhile, LDANET performed comparably to MobileNet in terms of training efficiency, but with 1.91% and 1.9% improvements in accuracy and IoU, respectively.

The results show that compared with Unet++, LDANet is relatively lightweight because it introduces ACB and DSC at the cost of slightly reducing model accuracy, which greatly enhances the applicability of the model. Compared with MobileNet, although LDANet's calculation parameters have increased, it can significantly improve the accuracy of the model. This is because LDANet uses the encoder and decoder structure to strengthen the feature association and to significantly improve the expression of dominant features by introducing dynamically weighted overlay layers.

Figure 6 shows the extraction results of the six models on the rural road test set. From the figure, it can be seen that all six models have some defects in the road extraction effects, but compared with other models, Unet+++ and LDANet have an obvious improvement in terms of road extraction. In summary, the use of LDANet for rural road extraction has a high application value.
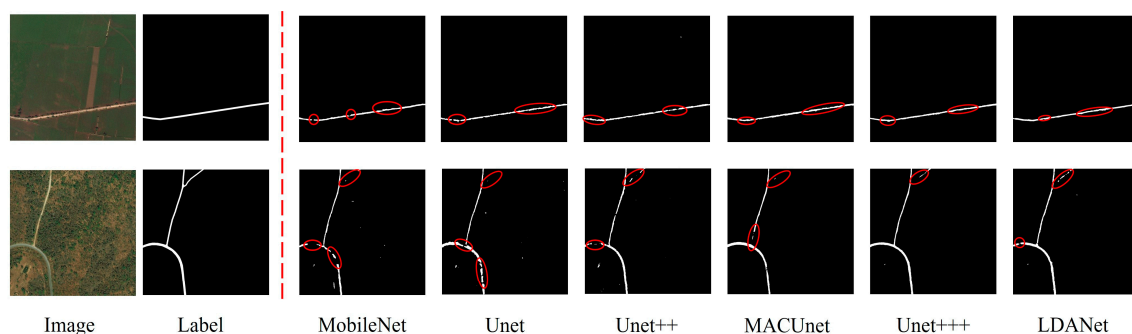

Image | Label | MobileNet | Unet | Unet++ | MACUnet | Unet+++ | LDANet

Figure 6. Extraction results of six models on the rural roads dataset.

4.3.2. Results of the Massachusetts Roads Dataset

To further validate the superiority and robustness of our LDANet, we also compared the LDANet with other methods on the Massachusetts roads dataset, using Precision, Recall, F1 Score and IoU metrics for evaluation.

The comparison results are shown in Table 3. The table shows that LDANet also performs well on the Massachusetts roads dataset, with Precision, Recall, F1 Score and IoU reaching 0.9755, 0.9707, 0.9731 and 0.6834, respectively. Figure 7 shows the extraction results of LDANet on the Massachusetts roads dataset compared to the ground truth map comparison, and it can be seen that our method achieves adequate visual results of road extraction.

Table 3. Comparison of the six models evaluated on the Massachusetts roads dataset.

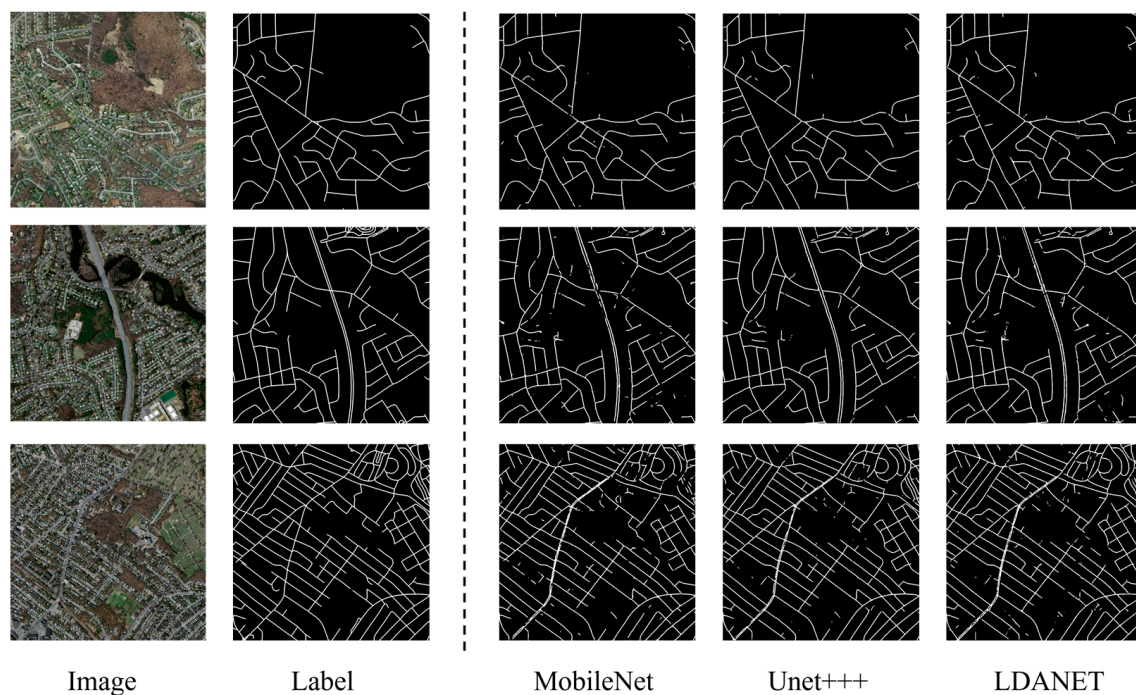|	Precision (%)	Recall	F1 Score	IoU	Parameters (M)	Train Time/Epoch (S)
Unet	0.9612	0.9584	0.9598	0.6513	9.85	480
Unet++	0.9710	0.9667	0.9688	0.6769	11.80	1130
Unet+++	0.9768	0.9716	0.9742	0.6957	6.75	1280
MACUnet	0.9721	0.9683	0.9702	0.6774	5.15	605
MobileNet	0.9533	0.9412	0.9472	0.6455	0.17	152
LDANet	0.9755	0.9707	0.9731	0.6834	0.20	163

| Image | Label | MobileNet | Unet+++ | LDANET |

Figure 7. LDANet extraction results on the Massachusetts roads dataset.

4.3.3. Discussion

Deep learning methods to achieve road extraction from remote sensing images are important to promote the application of remote sensing images and the development of cities. In this work, a lightweight model was constructed based on the Unet network structure. The model enhances the shallow feature information by introducing a feature expansion module and uses dynamic weighted superposition to improve the feature representation. Compared with Boonpook and Lu et al., this method can significantly reduce the modeling parameters by using DSC, making the model lighter and faster. Compared with MobileNets and ShuffleNets, this model can obtain high-accuracy road

extraction results by upsampling and hopping connections of the Unet network structure to fully learn contextual features.

5. Conclusions

We have proposed a lightweight extraction model based on rural road data. The model is composed of a feature expansion module and a deep feature association module. In addition, we used a dynamically weighted loss function according to the small proportion of rural roads. Compared with complex methods which are expensive to calculate, our method focuses on enriching shallow features and strengthening the correlation of deep salient features, which can effectively balance reliability and speed. In the typical rural roads dataset, our model's accuracy was 0.9874, and its IoU was 0.7621. In the Massachusetts roads dataset, our model also performed well. Our model has the characteristics of a small number of parameters and a fast training speed, which can greatly reduce the requirements for hardware while still ensuring extraction accuracy in practical applications. Therefore, it is of great significance to promote the portable and rapid application of remote sensing technology.

Future work will involve optimization strategies based on combined model applications to achieve multi-objective learning applications by enhancing global information interactivity based on constrained model complexity.

References

1. Chen, Z.; Wang, C.; Li, J.; Fan, W.; Du, J.; Zhong, B. Adaboost-like End-to-End Multiple Lightweight U-Nets for Road Extraction from Optical Remote Sensing Images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102341. [CrossRef]
2. Jiang, X.; Li, Y.; Jiang, T.; Xie, J.; Wu, Y.; Cai, Q.; Jiang, J.; Xu, J.; Zhang, H. RoadFormer: Pyramidal Deformable Vision Transformers for Road Network Extraction with Remote Sensing Images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *113*, 102987. [CrossRef]
3. Li, C.; Zeng, Q.; Fang, J.; Wu, N.; Wu, K. Road Extraction in Rural Areas from High Resolution Remote Sensing Image Using a Improved Full Convolution Network. *Natl. Remote Sens. Bull.* **2021**, *25*, 1978–1988. [CrossRef]
4. Herumurti, D.; Uchimura, K.; Koutaki, G.; Uemura, T. Urban Road Extraction Based on Hough Transform and Region Growing. In Proceedings of the FCV 2013—19th Korea-Japan Joint Workshop on Frontiers of Computer Vision, Incheon, Republic of Korea, 30 January–1 February 2013.
5. Shi, W.; Miao, Z.; Debayle, J. An Integrated Method for Urban Main-Road Centerline Extraction from Optical Remotely Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3359–3372. [CrossRef]
6. Lian, R.; Wang, W.; Mustafa, N.; Huang, L. Road Extraction Methods in High-Resolution Remote Sensing Images: A Comprehensive Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5489–5507. [CrossRef]
7. Zhao, J.Q.; Yang, J.; Li, P.X.; Lu, J.M. Semi-Automatic Road Extraction from SAR Images Using EKF and PF. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences—ISPRS Archives, Kona, HI, USA, 21–23 July 2015; Volume 40.
8. Perciano, T.; Tupin, F.; Hirata, R.; Cesar, R.M. A Two-Level Markov Random Field for Road Network Extraction and Its Application with Optical, SAR, and Multitemporal Data. *Int. J. Remote Sens.* **2016**, *37*, 3584–3610. [CrossRef]
9. Zang, Y.; Wang, C.; Cao, L.; Yu, Y.; Li, J. Road Network Extraction via Aperiodic Directional Structure Measurement. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3322–3335. [CrossRef]
10. Sujatha, C.; Selvathi, D. FPGA Implementation of Road Network Extraction Using Morphological Operator. *Image Anal. Stereol.* **2016**, *35*, 93–103. [CrossRef]
11. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [CrossRef]

12. Zhong, Z.; Li, J.; Cui, W.; Jiang, H. Fully Convolutional Networks for Building and Road Extraction: Preliminary Results. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; Volume 2016.

13. Varia, N.; Dokania, A.; Senthilnath, J. DeepExt: A Convolution Neural Network for Road Extraction Using RGB Images Captured by UAV. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI, Bangalore, India, 18–21 November 2018.

14. Doshi, J. Residual Inception Skip Network for Binary Segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; Volume 2018.

15. Zhou, L.; Zhang, C.; Wu, M. D-Linknet: Linknet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; Volume 2018.

16. Li, Y.; Guo, L.; Rao, J.; Xu, L.; Jin, S. Road Segmentation Based on Hybrid Convolutional Network for High-Resolution Visible Remote Sensing Image. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 613–617.

17. Boonpook, W.; Tan, Y.; Bai, B.; Xu, B. Road Extraction from UAV Images Using a Deep ResDCLnet Architecture. *Can. J. Remote Sens.* **2021**, *47*, 450–464. [CrossRef]

18. Lu, X.; Zhong, Y.; Zheng, Z.; Liu, Y.; Zhao, J.; Ma, A.; Yang, J. Multi-scale and multi-task deep learning framework for automatic road extraction. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9362–9377. [CrossRef]

19. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.

20. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

21. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; Volume 2019.

22. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. *Shufflenet V2: Practical Guidelines for Efficient Cnn Architecture Design*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11218 LNCS.

23. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

24. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *arXiv* **2016**, arXiv:1606.02147.

25. Emara, T.; Abd El Munim, H.E.; Abbas, H.M. Liteseg: A novel lightweight convnet for semantic segmentation. In Proceedings of the 2019 Digital Image Computing: Techniques and Applications (DICTA), Perth, WA, Australia, 2–4 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7.

26. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 263–272. [CrossRef]

27. 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.

28. Mehta, S.; Rastegari, M.; Caspi, A.; Shapiro, L.; Hajishirzi, H. *ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11214 LNCS.

29. Mehta, S.; Rastegari, M.; Shapiro, L.; Hajishirzi, H. ESPNetv2: A Light-Weight, Power Efficient, and General Purpose Convolutional Neural Network. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; Volume 2019.

30. Yi, Q.; Dai, G.; Shi, M.; Huang, Z.; Luo, A. ELANet: Effective Lightweight Attention-Guided Network for Real-Time Semantic Segmentation. *Neural Process. Lett.* **2023**, *55*, 1–18. [CrossRef]

31. Lan, R.; Sun, L.; Liu, Z.; Lu, H.; Pang, C.; Luo, X. MADNet: A fast and lightweight network for single-image super resolution. *IEEE Trans. Cybern.* **2020**, *51*, 1443–1453. [CrossRef]

32. Mnih, V. *Machine Learning for Aerial Image Labeling*; University of Toronto: Toronto, ON, Canada, 2013.

33. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raska, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; Volume 2018.

34. Ran, S.; Ding, J.; Liu, B.; Ge, X.; Ma, G. Multi-U-Net: Residual module under multisensory field and attention mechanism based optimized U-Net for VHR image semantic segmentation. *Sensors* **2021**, *21*, 1794. [CrossRef] [PubMed]

35. Wang, W.; Yang, N.; Zhang, Y.; Wang, F.; Cao, T.; Eklund, P. A Review of Road Extraction from Remote Sensing Images. *J. Traffic Transp. Eng. Engl. Ed.* **2016**, *3*, 271–282. [CrossRef]

36. DIng, X.; Guo, Y.; DIng, G.; Han, J. ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; Volume 2019.

37. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

38. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef] [PubMed]

39. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, 4–8 May 2020; Volume 2020.

40. Ren, Y.; Zhang, X.; Ma, Y.; Yang, Q.; Wang, C.; Liu, H.; Qi, Q. Full Convolutional Neural Network Based on Multi-Scale Feature Fusion for the Class Imbalance Remote Sensing Image Classification. *Remote Sens.* **2020**, *12*, 3547. [CrossRef]

41. Li, R.; Duan, C.; Zheng, S.; Zhang, C.; Atkinson, P.M. MACU-Net for Semantic Segmentation of Fine-Resolution Remotely Sensed Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8007205. [CrossRef]