Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8636973, 8 pages
https://doi.org/10.1155/2022/8636973

Research Article

Extraction of Impervious Surface from High-Resolution Remote Sensing Images Based on a Lightweight Convolutional Neural Network

Lingling Chen,[1] Hongmei Zhang ,[2] and Yuejun Song[3]

[1]*School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China*
[2]*College of Water Conservancy and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China*
[3]*Key Laboratory of Soil Erosion and Prevention of Jiangxi Province, Jiangxi Academy of Water Science and Engineering, Nanchang 330029, China*

Correspondence should be addressed to Hongmei Zhang; 2006992934@nit.edu.cn

Received 28 April 2022; Revised 27 July 2022; Accepted 3 August 2022; Published 28 August 2022

Academic Editor: Zhiguo Qu

Copyright © 2022 Lingling Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to continuous progress of urbanization in China, large area of natural surface has become impervious. Automatic extraction of impervious surface (IS) from high-resolution remote sensing images is important to urban planning and environmental management. Artificial identification of IS is time-consuming and laborious. It is valuable to develop more intelligent recognition patterns. In recent years, semantic segmentation models based on convolutional neural network (CNN) have made great progress in extraction of IS from remote sensing images. However, most existing models focus on improving accuracy and rarely consider computational efficiency. In order to keep balance between computing resource consumption, computing speed, and segmentation accuracy, we propose a lightweight semantic segmentation network model based on CNN, and we named it LWIBNet. LWIBNet uses an efficient encoding-decoding structure as the skeleton and connects the encoding part and the decoding part by the Skip Layer. Moreover, in order to reduce the number of parameters and speed up the calculation, we introduce improved Squeeze-and-Excitation (SE) module, inverted residuals, and depthwise separable convolution to form the Inv-Bottleneck (IB) module and use it as the core to build the LWIBNet model. On the computational complexity, LWIBNet and LWIBNet-TTA have the lowest FLOPs (14.14 G), and SegNet has the second lowest FLOPs, but SegNet is 3.2 times higher than LWIBNet (45.05 G vs 14.14 G). Both the LWIBNet model and classic models are tested and compared on the same data set. The results show that the LWIBNet model achieves a bit higher segmentation accuracy with less computation cost, and its computation speed is faster.

1. Introduction

The rapid urbanization in China has continued for decades, and it changed the pattern of land use and land cover, mainly manifested in the disappearance of natural surface cover and the increase of IS [1]. IS is a typical feature of urban areas and is defined as artificially constructed hard surfaces impermeable to surface water, such as buildings, squares, roads paved with asphalt and concrete, and parking lots [2]. An excessively high IS ratio will not only destroy the surface heat balance and cause a serious heat island effect in the city but also weaken the city's hydrological regulation capacity and cause floods. Therefore, timely grasp of the spatial distribution information of IS is of great significance to urban environmental research [3].

The traditional method of extracting IS information is visual interpretation, but this method has high technical requirements for interpreters, and it has problems such as low production efficiency and high cost. With the development of remote sensing technology and diversification of data, a large number of remote sensing thematic information retrieval technologies have been proposed and widely used (such as spectral mixture analysis, index method, decision tree model method, and regression model method), but most

of the algorithms are only suitable for IS extraction of low-resolution and medium-resolution remote sensing images and are not suitable for high-resolution remote sensing images containing only near-infrared and visible light bands (Xu and Wang 2016). General remote sensing classification methods, such as object-oriented method [4], support vector machine (Okujeni et al. 2015), and random forest (Breiman 2001), are applicable to IS extraction of remote sensing images of all resolutions and can achieve good results, but they also have some shortcomings: Affected by people's subjective consciousness, intelligence and automation are poor, and the extraction efficiency is increasingly unable to meet the massive remote sensing data processing. Obviously, there is an urgent need to explore newer and more intelligent IS extraction methods.

Deep learning (DL) is one of research hotspots in the field of machine learning in recent years, and it has unique characteristics of automatic learning ability and nonlinear complex function and fitting ability [5], and CNN belongs to DL, has excellent performance in image processing, and has been successfully applied to remote sensing visual recognition tasks such as ground object classification, target detection, and semantic segmentation in remote sensing images. Some scholars have applied CNN to IS extraction from high-resolution remote sensing images: He et al. extracted the IS of Chengdu, Sichuan Province, China, from remote sensing images by using CNN and studied the changes of IS from 2009 to 2017 [6]. Fu et al. combined deep convolution neural network with object-based image analysis (OBIA) to accurately extract IS from high-resolution remote sensing [1]. Huang et al. combined CNN with object-oriented segmentation, fuzzy C-means clustering, and improved watershed algorithm to improve the accuracy of automatic recognition of IS [7].

However, most of the mainstream CNN models currently applied to IS extraction and even to the whole remote sensing field are complex in structure and have a large number of parameters. Although they have excellent performance in accuracy, the limited hardware resources and computing power will lead to difficulties in model training, time-consuming, and even impossible to implement, which will hinder the realization of more research and industrial applications.

To handle the above problems and achieve rapid and high-precision segmentation, we proposed a lightweight semantic segmentation network model based on CNN (LWIBNet), aiming at obtaining IS extraction graphs without loss of segmentation accuracy with less consumption of computational resources and time.

2. Dataset

The dataset used in this paper is divided into training set, verification set, and test set. The training set and verification set are taken from Nanchang City, Jiangxi Province, China, while the test set is from Chongqing City, China. All remote sensing images are collected from LocaspaViewer, with a spatial resolution of 0.262 m and three bands of RGB. The dataset was obtained by ArcGIS visual interpretation and strict registration with the original image position. Due to the large size of the high-resolution image (which would lead to memory overflow), the sliding window cutting method with a repetition rate of 0.1 was adopted to cut the images of the training set and the verification set into 256×256 subgraphs to improve the training efficiency. In addition, in order to avoid overfitting caused by too few images in the training set, we enhanced the training set by horizontal flip, vertical flip, and diagonal flip and generated 82880 training images (the size of each image is 256×256), 20220 verification images (the size of each image is 256×256), and 2 test images (the size of each image is 3091×1451). Finally, we shuffled the training set and the verification set to make the distribution of samples more reasonable.

3. Methodology

3.1. Construction of LWIBNet Model. The structure of the LWIBNet model is shown in Figure 1, and its basic structure is an encoding-decoding structure with excellent performance in DL semantic segmentation [8]. The role of the encoding part is to learn the semantic information of the input image and extract features that IS, while the role of the decoding part is to enhance the feature extraction results of the encoding part and restore the spatial information and resolution of the feature map (FP). Tables 1 and 2 show the detailed information of encoding-decoding structure.

Generally, feature extraction of the network model requires appropriate receptive field to obtain sufficient rich context information. The traditional method is to combine convolution layer with pooling layer and then use this combination extensively. Although it can capture more translation invariance features, too many convolution layers greatly increase the computation in the learning process, and a large number of pooling operations will also cause the loss of internal data structure and spatial hierarchical information, which will lead to the inability to reconstruct some detailed information during decoding [9]. Therefore, in order to reduce computation and retain more information, we built the Inv-Bottleneck (IB) module and used it as the main model, supplemented by standard convolution +upsampling to build the model.

As shown in Figure 1 and Tables 1 and 2, the coding part is divided into five modules, which are composed of eleven IB modules, two standard convolutions, and one maximum pool and will output five feature maps with different scale information to the decoding part. The decoding part is also divided into five modules, which are composed of four IB modules, four standard convolutions, and five upsampling layers. The decoding part will fuse the five feature maps output by the encoding part in turn to form thicker feature information. The encoding-decoding connection mode adopts Skip Layer (E-FP1 connects D-FP4, E-FP2 connects D-FP3, E-FP3 connects D-FP2, E-FP4 connects D-FP1, input E-FP5 to D-FP5), which makes low-level features splice with high-level features, making full use of shallow local detail information and deep global information. At the end of the network is a softmax classifier, which can summarize the pixels with the same semantics according to

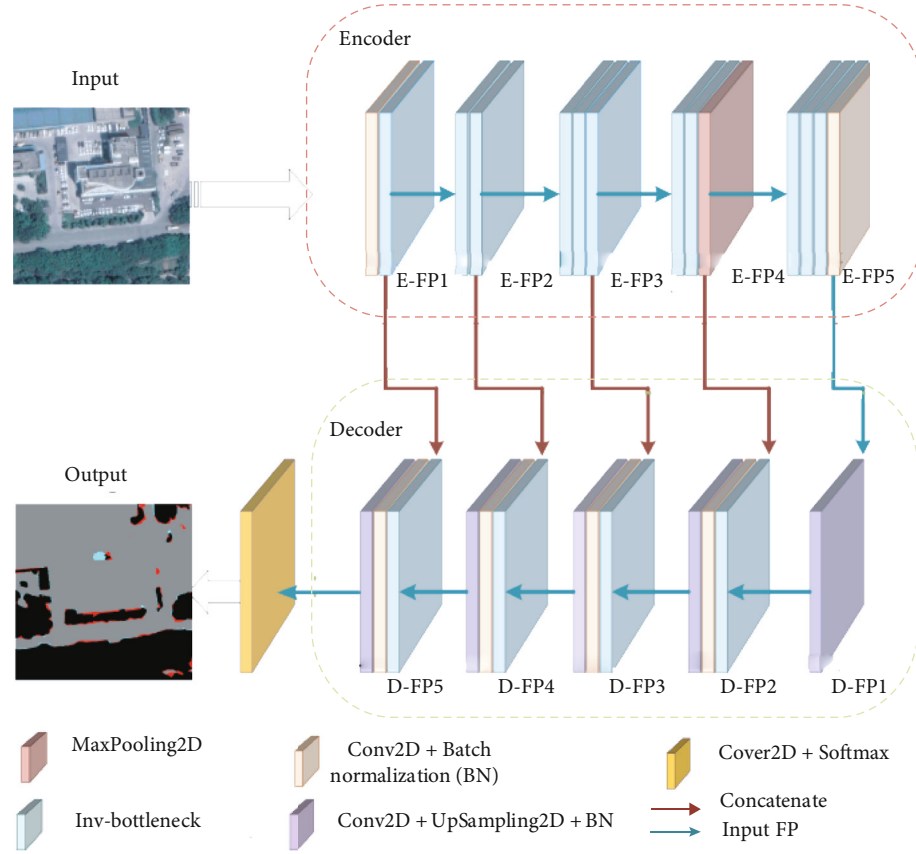


Figure 1: The network structure of the LWIBNet model. E-FP and D-FP, respectively, represent the feature maps of the encoding and decoding parts, each with 5 layers.

Table 1: The specific structure and information of the encoding part.

FP	Input	Operator	Up	Output	SE	Activation	Stride	Contact
E-FP1	$256^2 \times 3$	Conv+BN, 3×3	—	$128^2 \times 16$	—	Hard_swish	2	D-FP4
	$128^2 \times 16$	IB, 3×3	16	$128^2 \times 16$	—		1	
E-FP2	$128^2 \times 16$	IB, 3×3	72	$64^2 \times 24$	—		2	D-FP3
	$64^2 \times 24$	IB, 3×3	88	$64^2 \times 24$	—		1	
E-FP3	$64^2 \times 234$	IB, 5×5	96	$32^2 \times 40$	√	Hard_swish	2	D-FP2
	$32^2 \times 40$	IB, 5×5	240	$32^2 \times 40$	√	Hard_swish	1	
	$32^2 \times 40$	IB, 5×5	240	$32^2 \times 40$	√	Hard_swish	1	
E-FP4	$32^2 \times 40$	IB, 5×5	120	$32^2 \times 48$	√	Hard_swish	1	D-FP1
	$32^2 \times 48$	IB, 5×5	144	$32^2 \times 48$	√	Hard_swish	1	
	$32^2 \times 48$	Maxpooling, 2×2	—	$16^2 \times 48$	—	—	—	
E-FP5	$16^2 \times 48$	IB, 5×5	288	$8^2 \times 96$	√	Hard_swish	1	(To)D-FP5
	$8^2 \times 96$	IB, 5×5	576	$8^2 \times 96$	√	Hard_swish	1	
	$8^2 \times 96$	IB, 5×5	576	$8^2 \times 96$	√	Hard_swish	1	
	$8^2 \times 96$	Conv+BN, 1×1	—	$8^2 \times 96$	—	Hard_swish	2	

*Note that columns 1 to 9, respectively, represent the name of the feature map, the size of the input feature map, the size of the output feature map, the current operation, the number of channels changed, whether improved Squeeze-and-Excitation (SE) is used, the type of activation function, the step size, and the connection method.

Table 2: The specific structure and information of the decoding part.

Contact	FP	Input	Operator	Up	Output	SE	Activation	Stride
E-FP4	D-FP1	$8^2 \times 96$	Conv, 2×2 + UpSamp, 2×2	—	$16^2 \times 48$	—	ReLU	1
		$16^2 \times 48$	IB, 3×3	144	$16^2 \times 48$	√	Hard_swish	1
E-FP3	D-FP2	$16^2 \times 48$	Conv+BN, 3×3	—	$16^2 \times 48$	—	ReLU	1
		$16^2 \times 48$	Conv, 2×2 + UpSamp, 2×2	—	$32^2 \times 40$	—	ReLU	1
		$32^2 \times 40$	IB, 3×3	240	$32^2 \times 40$	√	Hard_swish	1
E-FP2	D-FP3	$32^2 \times 40$	Conv+BN, 3×3	—	$32^2 \times 40$	—	ReLU	1
		$32^2 \times 40$	Conv, 2×2 + UpSamp, 2×2	—	$64^2 \times 24$	—	ReLU	1
		$64^2 \times 24$	IB, 3×3	88	$64^2 \times 24$	—	ReLU6	1
E-FP1	D-FP4	$64^2 \times 24$	Conv+BN, 3×3	—	$64^2 \times 24$	—	ReLU	1
		$64^2 \times 24$	Conv, 2×2 + UpSamp, 2×2	—	$128^2 \times 16$	—	ReLU	1
		$128^2 \times 16$	IB, 3×3	16	$128^2 \times 16$	—	ReLU6	1
E-FP5(To)	D-FP5	$128^2 \times 16$	Conv+BN, 3×3	—	$128^2 \times 16$	—	ReLU	1
		$128^2 \times 16$	Conv, 3×3 + UpSamp, 2×2	—	$256^2 \times 2$	—	ReLU	1

the training results and output an IS information map with the same size as the input image.

In addition, we found that traditional dropout can degrade the performance of the small base model; so, we dropped the function of dropout and add batch normalization (BN) after each convolution operation to reduce the risk of overfitting.

It should be noted that IB module is the key to reduce the number of parameters and the amount of calculation and can help the model more accurately identify the IS in the complex features. Figure 2 shows the structure diagram of IB module and its components. IB module is composed of inverted residuals, depthwise separable convolution, and improved SE module.

Inverted residual structure can increase the ability of feature expression and solve the gradient disappearance problem caused by the increase of network depth during training. Its specific structure is shown in Figure 2(b). We take the inverse residual structure as the backbone of IB module and introduce the *h*-swish activation function [10], which has less computation but can improve the model accuracy.

Depth separable convolution is a key tool of the efficient neural network model, which splits a standard convolution into two independent operations: Depthwise convolution (DW) and pointwise convolution (PW), which can effectively reduce the amount of computation and cost while achieving similar (or slightly better) performance as standard convolution [11]. We use 5*5 DW and 3*3 DW in IB module to make the model more robust.

SE module is a mechanism that enables the model to calibrate features. It can make the effective weight larger and the invalid or ineffective weight smaller, but it will increase the total number of parameters and the total calculation amount of the model. Therefore, we use the improved SE (first, the result after exclusion 1 is changed to 1/4 of the original one, which reduces the parameter amount. Secondly, the sigmoid function is replaced by *h*-swish, which reduces the calculation amount.). The improved SE is shown in Figure 2(c), which is applied to the last layer of IB module.

3.2. Model Training and IS Extraction. In the training stage, the loss function adopts the cross entropy function, the optimizer abandons Adam and adopts a better Nadam optimization algorithm, the classifier adopts the softmax activation function, the batch size is set to 36, and the initial learning rate is set to $1e^{-4}$. When three epoch pass and the loss of the verification set do not drop, the learning rate is halved. Finally, set the loss of the verification set for 10 consecutive rounds and stop training if there is no decline. In the prediction stage (IS extraction), there are two prediction methods: one is ordinary remote sensing large image prediction (firstly, overlapping cutting, then image prediction, and finally, splicing prediction results by ignoring edges). The other is the prediction method combined with postprocessing (test time augmentation, TTA) on the basis of the former (firstly, the input test images are enhanced, then the results of different ways of data enhancement are predicted, and finally, all the prediction results are averaged). LWIBNet combined with TTA (LWIBNet-TTA) can improve the accuracy, and the comparison of its results will be exhibited in Section 3.

4. Results

To verify the effectiveness of the proposed model, we compare the extraction results of LWIBNet with those of classical models (U-Net, SegNet, and Deeplabv3+ (based on ResNet50)) and all models adopt the same experimental setup. The experimental hardware and software environment of this paper is shown in Table 3.

The qualitative segmentation results of all models are shown in Figures 3 and 4. Among them, Figure 3 is a comparison diagram after the difference between the tag map

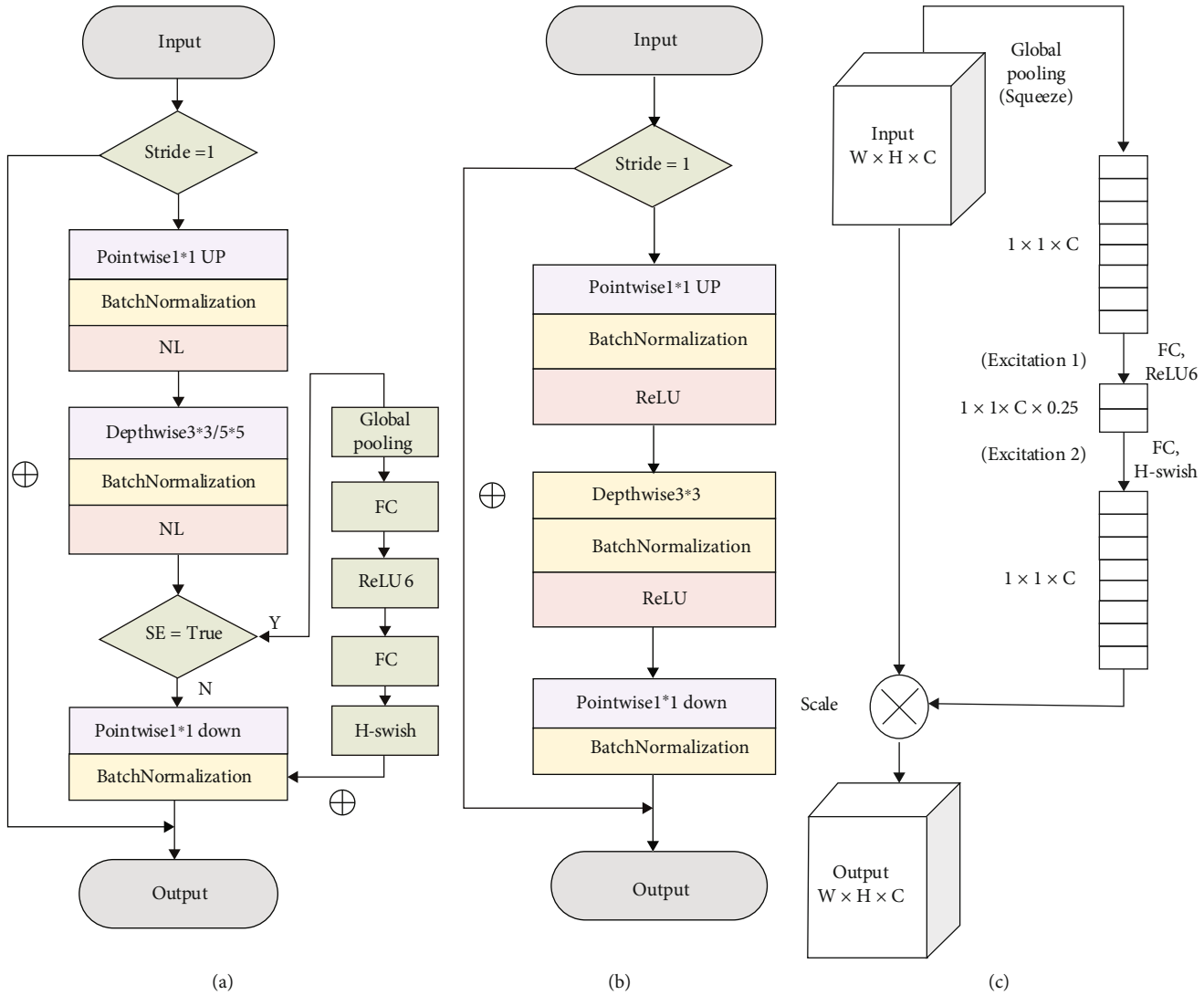


FIGURE 2: The structure of the IB module and its components. Up and down represent dimension increase and dimension decrease, respectively.

of the whole remote sensing image and the prediction map obtained by each model, which can clearly see the misclassified pixels. Figure 4 is a detailed comparison diagram of the misclassified areas in Figure 3, which can more specifically identify correctly classified and wrongly classified ground objects. Overall analysis of Figure 3 shows that the SegNet model has a large number of blue pixels and only a small number of red pixels, which indicates that its extraction ability is uneven and its effect is the worst. The number of red pixels in U-Net, Deeplabv3+, and LWIBNet is similar, while the number of blue pixels in LWIBNet is slightly less. LWIBNet-TTA has the least number of red and blue pixels, which indicates that its extraction effect is the best.

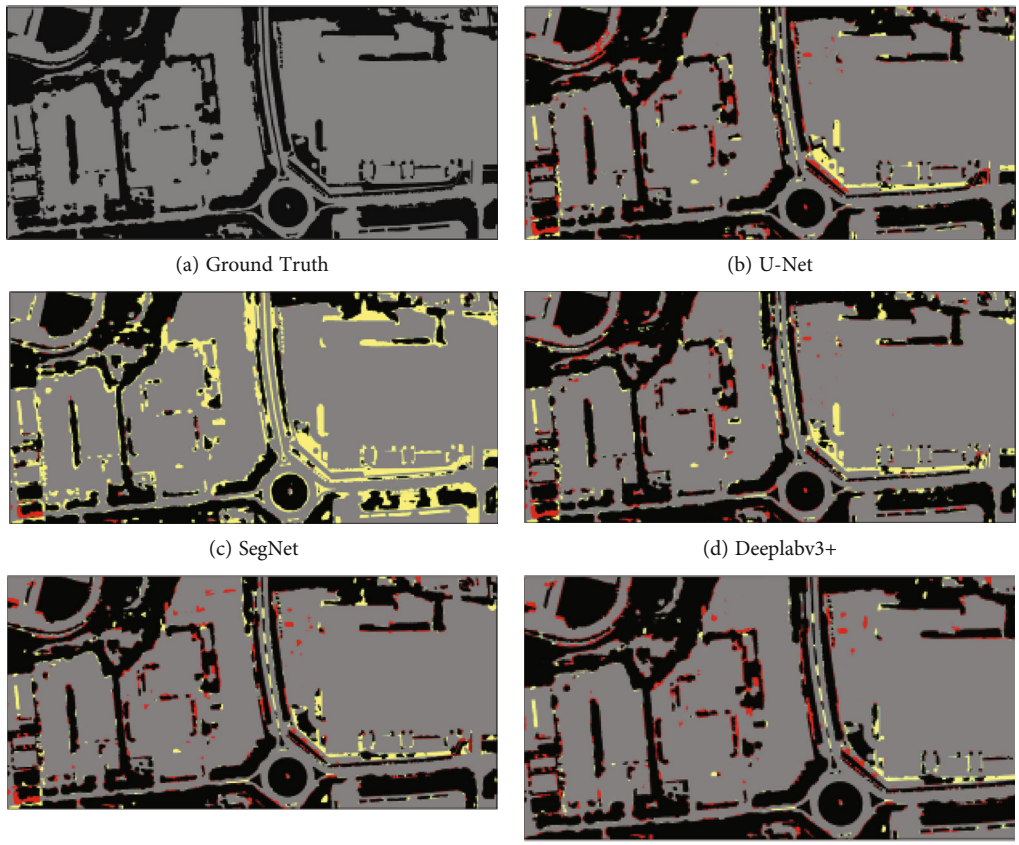
From the detailed analysis of Figure 4, it can be seen that each model has some problems: the concrete bare land covered by shadow is easily classified as permeable to water, and the bare grassland with bare and yellow is easily classified as impermeable to water, insensitive to black and dark IS objects, and easily confused and misclassified. According to these extraction difficulties, it can be seen that SegNet is the most serious, LWIBNet-TTA is lighter, and LWIBNet is similar to other models.

LWIBNet-TTA means adding TTA operation based on the LWIBNet model. The black, gray, red, and blue pixels represent the predictions of "Pervious Surface (PS)," "Impervious Surface (IS)," "IS misclassified as PS," and "PS misclassified as IS", respectively.

The quantitative results of each model are shown in Table 4. The indexes in the table are producer accuracy (Prd Acc), user accuracy (User Acc), overall accuracy (OA), F1 score, mean intersection over union (MloU), kappa coefficient, parameter quantity (Param), floating point operations (FLOPs), training time (TT), and segmentation time (ST). The first six indexes consider the basic performance indexes

Table 3: The experimental hardware and software environment.

Hardware or software	Version
CPU	Intel(R) Core i7-10875H
GPU	NVIDIA GeForce RTX 2070
Operating system	Windows10
Development tools	Tensorflow-Gpu 2.3.0, python3.8, and CUDA10.1



(a) Ground Truth (b) U-Net

(c) SegNet (d) Deeplabv3+

(e) LWIBNet (f) LWIBNet-TTA

Figure 3: IS extraction diagram of label dimension.

of the semantic segmentation model, while the last four indexes consider the complexity of the model itself, computational complexity, and time consumption.

On the basic performance index, LWIBNet-TTA has the best performance. Although SegNet has high drawing accuracy, its other indexes are very low; so, its overall performance is the worst. The overall performance of U-Net, Deeplabv3+, and LWIBNet has little difference, but LWIBNet still has 0.4-0.5 percentage points advantages in User Acc, MloU, and Kappa. In terms of parameters, the parameters of the LWIBNet model are about 11 times (1.07 vs 11.55, 11.85) lower than those of SegNet and Deeplabv3+ and even 29 times (1.07 vs 31.06) lower than those of U-Net. On the computational complexity, LWIBNet and LWIBNet-TTA have the lowest FLOPs (14.14G), and SegNet has the second lowest FLOPs, but SegNet is 3.2 times higher than LWIBNet (45.05 G vs. 14.14 G). In terms of segmentation time, each step of LWIBNet-TTA (processing a 256*256 size image) takes the most time, each step of U-Net takes the second most time, and each step of LWIBNet takes the least time.

5. Discussion

5.1. The Validity of IB Module. Inverted residuals, improved SE module, and depthwise separable convolution are commonly used efficient structures for lightweight models [10, 11]. However, whether the IB module composed of them can perform well is a problem. To verify the effectiveness of this IB module in reducing the amount of calculation and parameters and maintaining certain accuracy, we keep the main structure of LWIBNet unchanged, replace IB module with traditional convolution+pooling module, and carry out comparative experiments. The experiment is divided into three groups: A1 and A2 groups which do not use IB

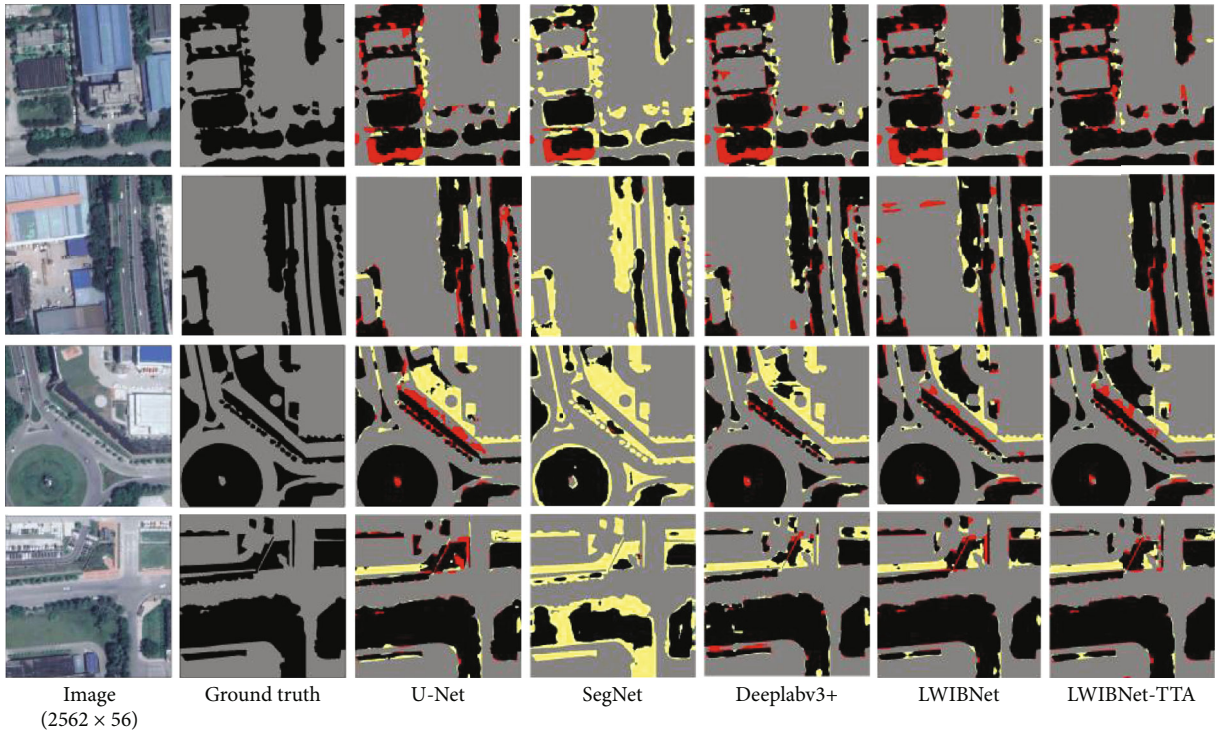


Figure 4: Comparison of the details of areas with severe misclassification.

Table 4: Comparison of the quantitative evaluation indices.

Model	Prd Acc	User Acc	OA	F1 score	MIoU	Kappa	Param/M	FLOPs/G	TT/ms	ST/ms
U-Net	0.965	0.957	0.947	0.961	0.886	0.932	31.06	109.3	637	20
SegNet	**0.994**	0.853	0.881	0.918	0.743	0.848	11.55	45.05	233	8
Deeplabv3+	0.970	0.946	0.943	0.958	0.877	0.927	11.85	52.57	277	9
LWIBNet	0.964	0.961	0.949	0.962	0.891	0.936	**1.07**	**14.14**	**139**	**5**
LWIBNet-TTA	0.962	**0.974**	**0.957**	**0.968**	**0.908**	**0.946**	**1.07**	**14.14**	**139**	25

*The indices include the producer accuracy (Prd Acc), user accuracy (User Acc), overall accuracy (OA), F1 score, mean intersection over union (MIoU), kappa coefficient, parameter quantity (Param), floating-point operations (FLOPs), training time (TT), and segmentation time (ST) (time consumption for processing a 256 × 256 size image). The optimal value of each index is presented in bold.

Table 5: Experimental results with and without IB module.

Group	OA	MIoU	Kappa	Param/M	FLOPs/G
A1	0.924	0.840	0.903	93.00	59.05
A2	0.902	0.791	0.876	**0.970**	**11.28**
B	**0.949**	**0.891**	**0.936**	1.070	14.14

*The optimal value of each index is presented in bold.

module (traditional convolution+pooling module is used) and B group which uses IB module. Among them, the channel number change of FP in group A1 follows U-Net and SegNet, and the channel number change of FP in group A2 follows the IB module. Their experimental results are shown in Table 5, it can be seen that group A1 has a large number of parameters, and its performance is worse than that of group B, while group A2 has the least amount of parameters and calculation, but its performance is also the worst. Group B using IB module has less parameters and the best performance. The comparison results prove the efficiency and applicability of using IB module as the core module to extract impervious surface from high-resolution remote sensing images.

5.2. Limitations of This Article. Although LWIBNet has achieved good performance and efficiency, it still has limitations. The ground objects in high-resolution remote sensing images are complex and detailed [12], and the spectral similarity between different objects and shadows of tall buildings or trees limits the correct extraction of IS. There is a problem that the ground objects covered by shadows are easy to be confused (it is difficult to distinguish by pixels alone). In future research, we will add more spectral information of bands, combine multisource remote sensing data from different sensors, and integrate multidisciplinary knowledge to further improve the automatic extraction ability of IS. Based on the constructed spectral homogeneity and

spectral heterogeneity indexes, the strategy of "coarse estimation + precise determination" is adopted, and an optimal segmentation result after multilevel optimization is gradually obtained.

6. Conclusion

In this paper, a lightweight semantic segmentation network model based on CNN is proposed, which is used to automatically extract impervious surface from high-resolution remote sensing images. In order to strike a balance between computing resource consumption, computing speed, and segmentation accuracy, the model takes an efficient encoding-decoding structure as the skeleton and IB module (composed of inverted residuals, improved SE module, and depthwise separable convolution) as the flesh and blood.

At the same time, the qualitative and quantitative results show that the parameters and computation of the LWIBNet model in this paper are far less than those of the classical semantic segmentation model, which is in line with the characteristics of lightweight network model. The LWIBNet model is comparable to the classical models (U-Net, SegNet, and Deeplabv3+) in segmentation accuracy, which preliminarily achieves the balance of computing resources, computing speed, and segmentation accuracy. It can quickly extract IS and can achieve similar or even slightly higher extraction accuracy with less computational resources than the classical model. In addition, the LWIBNet-TTA method can obtain higher extraction accuracy without considering the time-consuming segmentation.

In the future research, we will add more spectral information, combine multisource remote sensing data from different sensors, and integrate multidisciplinary knowledge to further improve the automatic extraction ability of IS.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Acknowledgments

This work was supported by the General Project of Key R&D Plan of Jiangxi Provincial Department of Science and Technology (No. 20192BBGL70054), the Water Conservancy Science and Technology Project of Jiangxi Province (No. 201821ZDKT18, No. 201922ZDKT08, No. 202123YBKT16, No. 202124ZDKT21, and No. 202124ZDKT25), and the General Project of Science and Technology Plan of Jiangxi Provincial Department of Education (No. GJJ170982).

References

[1] Y. Fu, K. Liu, Z. Shen, J. Deng, and K. Wang, "Mapping impervious surfaces in town–rural transition belts using china's GF-2 imagery and object-based deep cnns," *Remote Sensing*, vol. 11, no. 3, p. 280, 2019.

[2] C. L. Arnold Jr. and C. J. Gibbons, "Impervious surface coverage: the emergence of a key environmental indicator," *Journal of the American Planning Association*, vol. 62, no. 2, pp. 243–258, 1996.

[3] X. Cheng, R. Luo, G. Shi, L. Xia, and Z. Shen, "Automated detection of impervious surfaces using night-time light and landsat images based on an iterative classification framework," *Remote Sensing Letters*, vol. 11, no. 5, pp. 465–474, 2020.

[4] X. Hu and Q. Weng, "Impervious surface area extraction from IKONOS imagery using an object-based fuzzy method," *Geocarto International*, vol. 26, no. 1, pp. 3–20, 2011.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] Y. He, X. Zhang, J. Shi et al., "Change of impervious surface of Chengdu City, China," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2894–2897, Waikoloa, HI, USA, 2020.

[7] F. Huang, Y. Yu, and T. Feng, "Automatic extraction of impervious surfaces from high resolution remote sensing images based on deep learning," *Journal of Visual Communication and Image Representation*, vol. 58, no. 58, pp. 453–461, 2019.

[8] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany*, pp. 801–818, Munich, Germany, 2018.

[9] J. Lin, W. Jing, H. Song, and G. Chen, "ESFNet: efficient network for building extraction from high-resolution aerial images," *IEEE Access*, vol. 7, pp. 54285–54294, 2019.

[10] A. Howard, M. Sandler, G. Chu et al., "Searching for mobilenetv3," in *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, Seoul, Korea (South), 2019.

[11] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: an extremely efficient convolutional neural network for mobile devices," in *In Proceedings ofthe IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6848–6856, Salt Lake USA, 2018.

[12] Y. Li, L. Xu, J. Rao, L. Guo, Z. Yan, and S. Jin, "A Y-net deep learning method for road segmentation using high-resolution visible remote sensing images," *Remote Sensing Letters*, vol. 10, no. 4, pp. 381–390, 2019.