

## Segmentation of Remote Sensing Images Based on U-Net Multi-Task Learning

Ni Ruiwen<sup>1</sup>, Mu Ye<sup>1,2,3,4,\*</sup>, Li Ji<sup>1</sup>, Zhang Tong<sup>1</sup>, Luo Tianye<sup>1</sup>, Feng Rui long<sup>1</sup>, Gong He<sup>1,2,3,4</sup>, Hu Tianli<sup>1,2,3,4</sup>, Sun Yu<sup>1,2,3,4</sup>, Guo Ying<sup>1,2,3,4</sup>, Li Shijun<sup>5,6</sup> and Thobela Louis Tyasi<sup>7</sup>

<sup>1</sup>College of Information Technology, Jilin Agricultural University, Changchun, 130118, China

<sup>2</sup>Jilin Province Agricultural Internet of Things Technology Collaborative Innovation Center, Changchun, 130118, China

<sup>3</sup>Jilin Province Intelligent Environmental Engineering Research Center, Changchun, 130118, China

<sup>4</sup>Jilin Province Information Technology and Intelligent Agriculture Engineering Research Center, Changchun, 130118, China

<sup>5</sup>College of Information Technology, Wuzhou University, Wuzhou, 543003, China

<sup>6</sup>Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou, 543003, China

<sup>7</sup>Department of Agricultural Economics and Animal Production, University of Limpopo, Sovenga, Polokwane, 0727, South Africa

\*Corresponding Author: Mu Ye. Email: [muye@jlau.edu.cn](mailto:muye@jlau.edu.cn)

Received: 06 January 2022; Accepted: 23 February 2022

**Abstract:** In order to accurately segment architectural features in high-resolution remote sensing images, a semantic segmentation method based on U-net network multi-task learning is proposed. First, a boundary distance map was generated based on the remote sensing image of the ground truth map of the building. The remote sensing image and its truth map were used as the input in the U-net network, followed by the addition of the building ground prediction layer at the end of the U-net network. Based on the ResNet network, a multi-task network with the boundary distance prediction layer was built. Experiments involving the ISPRS aerial remote sensing image building and feature annotation data set show that compared with the full convolutional network combined with the multi-layer perceptron method, the intersection ratio of VGG16 network, VGG16 + boundary prediction, ResNet50 and the method in this paper were increased by 5.15%, 6.946%, 6.41% and 7.86%. The accuracy of the networks was increased to 94.71%, 95.39%, 95.30% and 96.10% respectively, which resulted in high-precision extraction of building features.

**Keywords:** Multitasking learning; U-net; ResNet; remote sensing image; semantic segmentation

### 1 Introduction

Remote sensing image analysis is a basic and practical research hotspot in remote sensing science. The remote sensing image contains a wealth of features, which can be used in urban planning, agricultural monitoring, ecological services, and geological prospecting. Remote sensing image



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

segmentation is one of the main strategies employed in remote sensing image analysis for effective extraction of semantic information involving various features, which is of great significance to urban decision makers, agricultural growers, and military personnel in national defense [1]. However, with the continuous development of remote sensing technology and earth observation methods, the number of aerial and satellite images has increased sharply [2]. Image resolution has continued to increase along with the proportion of ultra-high resolution remote sensing images. However, the processing frequency and difficulty have gradually increased. Initially, remote sensing image segmentation was completed via visual interpretation by the staff. However, due to the prolonged and expensive features, computer-assisted automated segmentation methods have become the focus of academic and research attention [3], such as IsoData [4], K-Means [5], Maximum Likelihood Method [6], Random Forest [7], Support Vector Machine (SVM) [8–10] and decision tree [11]. However, these methods utilizing spatial and semantic information have lower accuracy and weaker segmentation capabilities, which prevent ultra-high resolution image segmentation. In recent years, deep learning image analysis has made great progress [12–14] in facilitating automated interpretation of high-resolution remote sensing images.

The use of deep learning-related models and methods to extract image features is highly accurate and image segmentation has become a leading trend in image processing. TONG [15] used pseudo-labels based on inversion resulting in supervision and fine-tuning of the pre-trained model. The fine-tuned convolutional neural network combines block-by-block classification and hierarchical segmentation to facilitate hybrid classification and construction of a large-scale land cover data set, namely GID (<http://captain.whu.edu.cn/GID/>). Zhao et al. [16] used a multi-scale convolutional neural network to extract land cover information from a variety of public data sets. This learning mechanism can be combined with additional classifiers such as support vector machines and random forests. The overall accuracy is as high as 91.12%, indicating that the multi-scale convolutional neural network has strong practicability in object-based image classification. Dang et al. [17] used the AlexNet model for the classification of forest land, cultivated land, water areas, and houses on 1,875 map spots obtained from the geographical census. The accuracy of the classification of houses and cultivated land was 99%. Insufficient training samples for woodland resulted in poor classification accuracies of 43.59% and 62.73%, respectively. Li et al. [18] used the “CCF Satellite Image AI Classification and Recognition Competition” dataset to classify vegetation, roads, buildings, waters and other land types in a selected area of southern China using the U-Net model, with a final training accuracy of 94%. Shi et al. [19] obtained good results using the transfer learning CNN to classify the land use scenes of the experimental area map blocks based on satellite imaging. Sun et al. [20] migrated the VGG11 network trained on the Carvana dataset to the coding structure of the U-Net network for building extraction. They found that the pre-trained model rapidly converged to its stable value. In view of the difficulty of fuzzy class and detail loss in semantic segmentation, Badrinarayanan et al. [21] used Empty convolution to obtain an encoder without down-sampling and restored the label to full resolution during training or similar to previous studies [22,23] involving features that combine multiple resolutions. Kemker et al. [24] and Chen et al. [25] improved the decoder structure by designing symmetric transposed convolutional layers and skip connections using probability models and filters [26] or fusion of unsupervised segmentation [27]. The results of semantic segmentation are post-processed. Gul et al. [28] proposed that optimal cooperative spectrum sensing based on butterfly optimization algorithm. Kwon et al. [29] proposed data traffic reduction with compressed sensing in an Artificial Intelligence of Things (AIoT) system. Islam et al. [30] reported land cover

classification and its impact on land surface temperature in Peshawar using remote sensing, Jiang et al. [31] proposed a crowdsourcing price game model for crowd sensing, Cheng et al. [32] proposed crowdsensing based on compressed sensing of orthogonal matching pursuit algorithm image recovery. Zhang et al. [33] reported a robust 3-D medical watermarking based on wavelet transform for data protection, Sun et al. [34] reported robust reversible audio watermarking scheme for telemedicine and privacy protection. Although the foregoing research yielded robust results in building feature extraction, two serious challenges remained to be addressed [14]: ① The building feature segmentation method via post-processing steps is too complex and the integration between modules is difficult; ② The method of extracting different features via multiple different networks and combining these features is hindered by complex networks, increased need for hardware equipment and a long learning curve.

In brief, the U-net network based on ResNet can extract clear boundaries and segment accurate target objects [25]. A deep network based on multi-task learning for training a variety of different tasks on one network subject obviates the need for building different networks for multiple tasks [27]. Therefore, this study builds a semantic segmentation network based on ResNet under the U-net network framework. In order to further improve the accuracy of building feature extraction, a multi-task learning strategy is used to add a boundary distance prediction layer to the network to extract the complete building feature boundary for high-precision extraction of building features, while avoiding waste of computing resources.

## 2 Model Establishment

### 2.1 Multitasking Network

The multi-task network proposed here uses multi-task learning, which not only enhances the segmentation of architectural features during objective loss function, but also introduces the boundary information to improve the final segmentation outcomes.

The multi-task network not only enhances the segmentation of the semantic information of the building, but also extracts the boundary information of the building features during the early training process. Based on the true value map of the building features, it is convenient to extract the edge, shape and other related geometric information of the boundary of the building. In this paper, the distance between the pixels of the building feature to the boundary is used as the training data for the network to generate geometric attributes. The advantages of this training data for the network are as follows. The boundary distance map can be quickly produced from the existing building ground truth map via distance transformation. The loss function designed by the boundary distance map (such as mean square error or negative logarithm) facilitates the calculation and learning of the boundary position information of each pixel in the image by the network and implicitly capture its geometric properties.

Suppose  $Q$  represents the set of pixels at the boundary of the building and  $C$  represents the set of pixels belonging to the building, for each pixel  $p$  in the image, the cutoff distance  $D(p)$  is:

$$D(p) = \delta \min \left[ \min_{q \in Q} d(p, q), R \right],$$

$$\delta_p = \begin{cases} +1 & (p \in C) \\ -1 & (p \notin C) \end{cases} \quad (1)$$

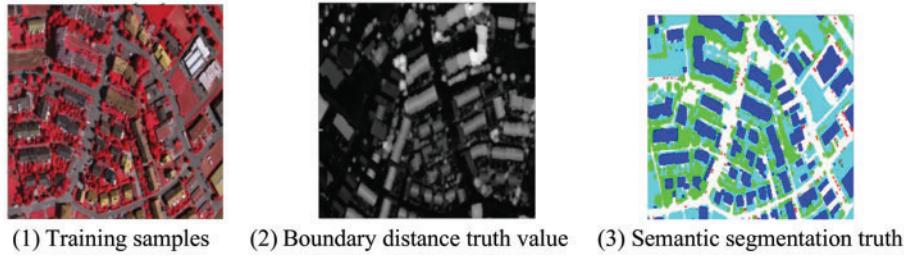
In the equation above,  $d(p, q)$  is the Euclidean distance between pixel  $p$  and  $q$ ;  $R$  is the truncation threshold; the symbol  $\delta p$  denotes the weighting of pixel  $p$ , indicating whether the pixel was inside or outside the building.

The continuous distance value is quantized uniformly for training. The boundary distance graph is encoded as k-dimensional binary vector  $\mathbf{B}(p)$  using one-HOT, i.e.,

$$D(P) = \sum_{k=1}^K r_k b_k(p), \sum_{k=1}^K b_k(p) = 1 \quad (2)$$

In the Eq. (2) above,  $r_k$  is the distance of the corresponding  $k$ . The binary pixel distance graph obtained by  $k$  represents the boundary distance graph of each pixel in the KTH boundary distance.

At this point, the data for the training multitask network has been generated. Fig. 1 illustrates the training sample images and the corresponding semantic segmentation and boundary distance truth values. Among them, pairs of similar images are used to test the robustness of the network and the effectiveness of ground object segmentation for small-scale buildings. The third graph represents the distance between the building features and the boundary. The larger the distance, the less the pixel belongs to the boundary pixel, and the smaller the value the more likely the value is to be the boundary pixel. Thus, the network trained by the boundary distance truth graph retains the boundary information of the building features to the maximum.

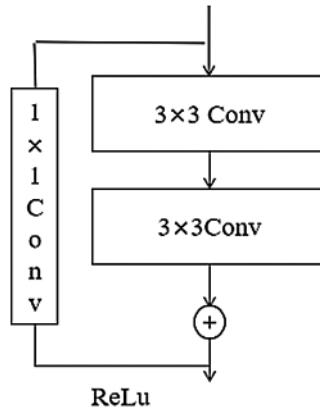


**Figure 1:** Visualization of training data

## 2.2 Multitasking Network Structure

The multi-task network architecture proposed in this paper is constructed based on U-net. U-net is a network structure with complete symmetry between convolutional encoding and decoding. It can capture features at different levels and integrates them through feature superposition. Features of different levels or receptive fields of different sizes show different sensitivities to target objects of different sizes. However, the U-net network has a simple structure. Although the location of building objects can be accurately detected when it is used to extract building objects from remote sensing images, the results often reveal a few round spots with different sizes. Most building objects cannot be detected and substantial boundary information is lost. Therefore, the multi-task network presented in this paper uses ResNet as the basic framework to reconstruct U-net network.

The residual block of the network is composed of two  $3 \times 3$  convolutional layers and a shortcut layer that completes feature dimension matching, The ResNet structure used is shown in Fig. 2:



**Figure 2:** Residual block structure

It includes two components: contraction network and expansion network. The contraction network is similar to the original contraction structure of U-net; however, the output results of each layer are first activated via batch standardization (BN) and modified linear unit (ReLU) activation function. The upsampling structure consists of a residual block and a  $2 \times 2$  maximum pooling layer. In the downsampling process, the image size is reduced 2-fold compared with the original image, and the extracted features are increased by 2-fold. The expansion network is similar to the original expansion structure of U-net. The upsampling component is composed of a residual block and a single  $2 \times 2$  upsampling, which is the same as the compression network. The output results of each layer need to be standardized in batches and activated using the activation function. Finally, a  $1 \times 1$  convolution is performed to output the corresponding results of the feature mapping.

(1) ReLU function as activation function. It is expressed as follows:

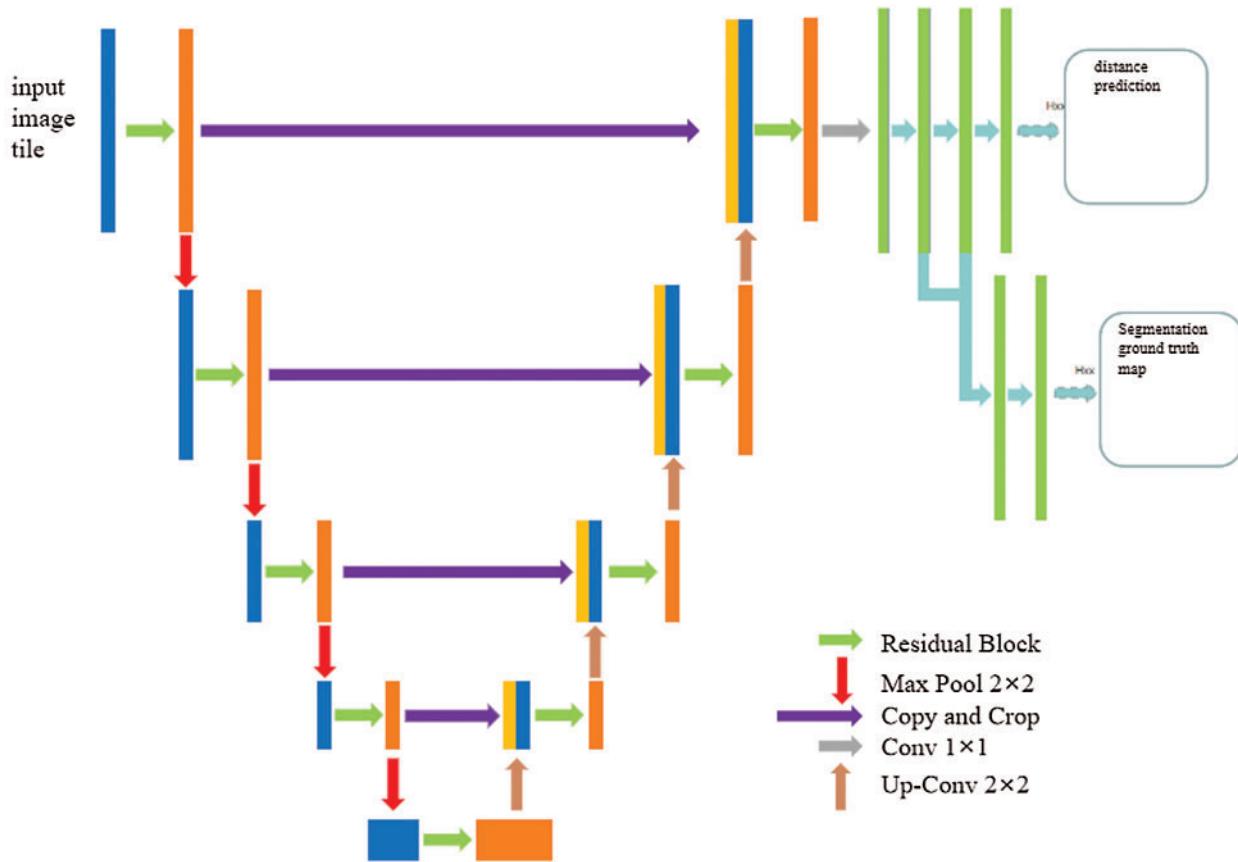
$$\text{ReLU}(x) = \max(0, x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (3)$$

(2) BN layer. From the perspective of the activation function, ReLU resolves the gradient saturation problem to a large extent. However, in order to prevent data from falling into the saturation zone during the training process, i.e., the phenomenon of gradient dispersion and slow network convergence, the BN layer is introduced into the model.

The final network structure is shown in Fig. 3 below.

In order to obtain multi-scale features, the convolution decoding component was designed in series with corresponding modules of the convolution coding structure. Each module in the convolutional decoding structure includes an input corresponding to convolutional coding and a lower module to ensure retention of high-frequency information by the convolutional decoding component. At the end of the network, two convolution layers are added, which are respectively used to predict the distance Hdist from each pixel in the image to the boundary of building features and the distance Hseg used to predict the segmentation result of building features, based on the distance prediction convolution layer. The two convolutional layers are accompanied by the corresponding SoftMax layer to complete different prediction tasks, so that the multi-task network can fully utilize the semantic and geometric attributes in the feature mapping of convolution decoding. Therefore, ResNet network can be used as the feature extractor to address gradient loss caused by the increased number of convolutional layers and extract effective image features in the convolutional coding component. The

serial connection in the convolutional decoding structure can learn the features of multiple scales and different network layers, which can increase the robustness of the network and improve the accuracy of building segmentation. Finally, a multi-task prediction structure is added to enable extraction of semantic and geometric attributes of the target object by the network.



**Figure 3:** Model structure

### 3 Experiment

#### 3.1 Data Set and Data Set Amplification

In this paper, experiments were carried out on large-scale ISPRS Vaihingen [35] aerial remote sensing image building object annotation data set. The sample images in this dataset include RGB images with a spatial resolution of 0.3 m after orthography correction. The size of each image is 5 000 pixels  $\times$  5 000 pixels, covering an area of 1 500  $\times$  1 500 m<sup>2</sup>. The data set annotates only two semantic classes including architectural and non-architectural features, and the training sets contain complete annotation truth values.

The goal of data augmentation is to generate new sample instances. In case of fewer training samples, data augmentation is very useful for improving the robustness of the network. For remote sensing images, many data augmentation methods are available including color dithering, random cropping, horizontal/vertical flipping, shifting, rotation/reflection, noise, cutting, and switching frequency bands. Since most remote sensing images are orthophotos, the changes are mainly reflected

in the direction and scale. However, the images in the data set used in this article exhibit the same spatial resolution without large-scale changes. Therefore, only three common augmentation methods are used: horizontal/vertical flipping, rotation, and random cropping. An image block with a size of 224 pixels  $\times$  224 pixels is randomly extracted from the original image and flipped horizontally and vertically, and rotated at different angles. After data augmentation, the original data set can be expanded 14-fold. It should be noted that only the original data training set is augmented, and the validation set is no longer augmented.

### 3.2 Experimental Results and Analysis

In model training, the batch size was set to 32, while the momentum was set to 0.9, and the learning rate is 0.001. Training is regularized by weight decay and dropout regularization of two dense layers (dropout ratio is set to 0.5). The experiment uses Keras as the developmental framework, and the models are trained for 40 iterations, with an average of 25 h for each network model. The following figure is a graphic representation of the function of the loss rate of the training set and the validation set of the model. In order to verify the performance of the proposed semantic method segmentation, we define an intersection-union ratio (IoU), suggesting that the predicted result and truth graph are derived from the intersection of buildings divided by their union.

$$(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (4)$$

In the formula: A denotes the building predicted by different methods and B is the building in the truth graph.

In this study, the superiority of the method has been verified by deepening the coding and decoding layers of the semantic segmentation network, and adding the boundary prediction layer based on cascaded multi-task learning to build the U-net network. Towards this end, the remote sensing images of 5 cities were selected and the method was compared with the FCN. Combining the MLP method (FCN + MLP) [17], the U-net network was based on VGG16 (VGG16) [24]. The experimental results of THE U-net network based on VGG16 (VGG16 + boundary prediction) and the U-net network built based on ResNet50 (ResNet50) [25] were compared and analyzed. At the same time, we added only ordinary remote sensing images and real values of remote sensing images to the model to train the model, and added ordinary remote sensing images and the real value of boundary distance to conduct two ablation experiments. The experimental results are presented in [Tab. 1](#).

As shown in [Tab. 1](#), the multi-task network discussed in this paper has the following advantages. The U-net network built with deeper coding and decoding layers yields better segmentation of building features. As shown in [Tab. 1](#), the FCN + MLP method uses a simple 4-layer convolutional coding layer to build FCN, followed by the use of MLP to combine the feature maps of different layers resulting in the final building feature prediction. Although the MLP combines feature maps from different layers, due to its shallow coding and decoding layers, it is impossible to fully extract the changing features of the building, resulting in poor feature extraction. VGG16 and ResNet50 networks were utilized to verify the importance of the depth of the encoder and the decoding layers in the construction of the U-net network. The network weights of the coding layer in the newly constructed U-net network were initialized using the network weights pre-trained by VGG16 and ResNet50 networks on ImageNet [28]. The network weights of the decoding layer were initialized via Gaussian distribution. [Tab. 1](#) presents the experimental results of U-net networks constructed by different networks: Compared with FCN + MLP method, the IoU mean values of VGG16 network [24], VGG16 + boundary prediction, ResNet50 network and the proposed method were improved by 5.15, 6.94, 6.41 and 7. At 86 hundred

points, Acc average points increased to 94.71%, 95.39%, 95.30% and 96.10%, indicating that it is difficult to extract the deep abstract features of remote sensing image for building object segmentation task using the FCN + MLP method. Compared with VGG16 network, the mean values of IoU and Acc in ResNet50 network increased by 1.26% and 0.59%, respectively.

**Table 1:** Experimental results of different models

Model	Accuracy	IOU	Accuracy of training with unbounded distance to ground truth	Accuracy of ground-truth images without remote sensing
FCN+MLP	94.42	64.47	87.34	86.23
VGG-16	94.71	69.82	88.23	89.36
VGG-16+boundary prediction	95.93	71.61	90.56	92.65
ResNet50	95.30	71.08	90.34	93.34
Our model	96.10	72.52	93.23	90.17

In order to verify the advantages of the multi-task network with boundary distance prediction proposed in this paper, we added the boundary distance prediction layer to the U-net network based on VGG16 and ResNet50. Thus, the boundary distance prediction layer Hseg was not only used to segment the result prediction layer, but also the distance prediction layer Hdist was added. A large number of experiments reported the highest segmentation accuracy of remote sensing image building features using the proposed method to create the boundary distance map and generate training data, when truncation distance  $R = 20$  in Formula (1) and interval number  $K = 10$  in Formula (2). As shown in Tab. 1, the mean values of IoU and Acc of VGG16 + boundary prediction method are 1.79% and 0.62% higher than those of VGG16 segmentation, respectively. The mean values of IoU and Acc in this method are 72.53% and 96.10%, respectively. The segmentation accuracy of building features in remote sensing image is the highest. Therefore, the joint boundary layer distance prediction of multitasking network enhances the classification accuracy, as the boundary layer distance prediction via U-net network in the training process was combined with the main body of the multitasking network layer (layer encoding and decoding) to construct geometric features and predict segmentation layer-restricted construction features of the boundary information. Thus, a higher precision of semantic segmentation can be obtained.

In order to further verify the effectiveness of the method in this paper, the segmentation results of different remote sensing image building features are presented in Tab. 2.

As can be seen from Tab. 2, among the results of building features extracted from five remote sensing images by different methods, only the segmentation of FCN + MLP square show an obvious “circular spot”, while the results obtained via other methods are very close to the truth value. As shown in Tab. 2, compared with FCN + MLP method, IoU and Acc values of the other four methods are greatly improved, resulting in enormous differences in the results of segmentation. In addition, the size of the 5 images listed in Tab. 2 is 500 pixels  $\times$  500 pixels, which is only 1/10,000 of the remote sensing image of one scene, suggesting poor visual effects. However, after careful observation of the segmentation results of images 1 and 2, the method presented in this paper can be used to segment

small-scale building features more accurately than the other four methods. As shown in figure 4 in Tab. 2, due to the small spacing between the building objects, it is easy to merge different degrees resulting in the appearance of rough edges during segmentation, while the phenomenon of rough edges in the segmentation results of ResNet50 is reduced. As shown in the figure, the segmentation results of the VGG16 + boundary prediction method and the proposed method are closer to the truth value, with accurate and distinct boundaries.

**Table 2:** Segmentation results of different remote sensing image buildings

Image number	Remote sensing image	FCN+MLP	VGG16	VGG16+ boundary prediction	ResNet 50	Method of this article	Truth value	True value of boundary distance
1								
2								
3								
4								
5								

Nonetheless, the VGG16 + boundary prediction method and the method in this paper are very close to the true value of the boundary distance. The complete boundary of the building ground can be identified from the five sets of images. Therefore, the features of remote sensing images extracted using the multi-task network with the boundary prediction layer are better than the single-task network under the same framework. In addition, the multi-task network with the boundary prediction layer facilitates the extraction of the boundary of the building features using the U-net network, resulting in additional geometric information associated with the building features for the prediction layer of the segmentation result. The method described in this paper is significantly better than the method used for VGG16 + boundary prediction, which demonstrates that the U-net network with deeper coding and decoding layers is more effective in extracting the details of remote sensing image building features.

#### 4 Conclusion

In order to achieve high-precision segmentation of building features in remote sensing images, this study proposes a multi-task learning U-net network based on ResNet50. The network mainly improves the semantic segmentation of architectural features in remote sensing images using a deeper ResNet network to build a U-net network, and a cascaded multi-task learning protocol to combine the constructed U-net network with the geometric boundary information of the building features providing input to FCN for effective semantic segmentation. The experimental results show that the method can increase the mean IoU of the semantic segmentation of remote sensing image features to 72.53%. The average value of Acc increased to 96.10%, which is partially accurate and timely for the actual remote sensing image segmentation of building features. In practice, the segmentation accuracy of remote sensing image building features in Xinxiang High-tech Zone reached 86.93%. However, the depth of the network reported in this paper is still limited, and its boundary distance uses simple Euclidean distance. Therefore, we plan to use ResNet101 and ResNet200 networks to continue to deepen the coding and decoding layers of the U-net network, and use Markov distance to generate boundary distance prediction maps to improve the semantic segmentation accuracy of remote sensing images.

**Funding Statement:** This research was supported by National Key Research and Development program [2018YFF0213606-03 (Mu, Y., Hu, T. L., Gong, H., Li, S. J. and Sun, Y. H.) <http://www.most.gov.cn>], the Jilin Province Science and Technology Development Plan focusing on research and development projects [20200402006NC (Mu, Y., Hu, T. L., Gong, H. and Li, S. J.) <http://kjt.jl.gov.cn>], the science and technology support project for key industries in southern Xinjiang [2018DB001 (Gong, H., and Li, S. J.) <http://kjj.xjbt.gov.cn>], and the key technology R & D project of Changchun Science and Technology Bureau of Jilin Province [21ZGN29 (Mu, Y., Bao, H. P., Wang X. B.) <http://kjj.changchun.gov.cn>].

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### References

- [1] B. Zhang, C. Wang and Y. Shen, “Fully connected conditional ran-dom fields for high resolution remote sensing land use/land cover classification with convolutional neural networks,” *Remote Sensing*, vol. 10, no. 12, pp. 1889–1903, 2018.
- [2] G. Heng, X. Xie, J. Han and L. Guo, “Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. 1, pp. 3735–3756, 2020.
- [3] C. Yao, Y. Zhang and H. Liu, “Application of convolutional neural network in classification of high resolution agricultural remote sensing images. ISPRS-international archives of the photogrammetry,” *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W7*, vol. 1, no. 1, pp. 989–992, 2017.
- [4] W. Li, R. Dong and H. Fu, “Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks,” *Remote Sensing*, vol. 11, no. 1, pp. 11–31, 2019.
- [5] Z. Lv, T. Liu, J. A. Benediktsson and H. Du, “Novel land cover change detection method based on k-means clustering and adaptive majority voting using bitemporal remote sensing images,” *IEEE Access*, vol. 7, no. 1, pp. 34425–34437, 2019.

- [6] B. Rimal, S. Rijal and R. Kunwar, "Comparing support vector machines and maximum likelihood classifiers for mapping of urbanization," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 1, pp. 71–79, 2019.
- [7] M. A. Huijuan, X. Gao and X. T. Gao, "Random forest classification of landsat 8 imagery for the complex terrain area based on the combination of spectral, topographic and texture information," *Journal of Geo-Information Science*, vol. 21, no. 3, pp. 359–371, 2019.
- [8] R. Khatami, G. Mountrekis and S. V. Stehman, "A Meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sensing of Environment*, vol. 177, no. 1, pp. 89–100, 2016.
- [9] X. Niu and Y. Ban, "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *International Journal of Remote Sensing*, vol. 34, no. 1–2, pp. 1–26, 2013.
- [10] X. Niu and Y. Ban, "A novel contextual classification algorithm for multitemporal polarimetric SAR data," *IEEE Geoscience & Remote Sensing Letters*, vol. 11, no. 3, pp. 681–685, 2014.
- [11] C. Yang, G. F. Wu and K. Ding, "Improving land Use/Land cover classification by integrating pixel unmixing and decision tree methods," *Remote Sensing*, vol. 9, no. 12, pp. 1222, 2017.
- [12] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097–1105, 2012.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, vol. 1, no. 1, pp. 1–12, 2014.
- [14] C. D. Storie and C. J. Henry, "Deep learning neural networks for land use land cover mapping," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Valencia, Spain, vol. 1, no. 1, pp. 3445–3448, 2018.
- [15] X. Y. Tong, G. S. Xia and Q. Lu, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sensing of Environment*, vol. 237, no. 1, pp. 111322, 2018.
- [16] W. Zhao and S. Du, "Learning multiscale and deep representations for classifying remotely sensed imagery," *Isprs Journal of Photogrammetry & Remote Sensing*, vol. 113, no. 1, pp. 115–165, 2016.
- [17] Y. Dang, J. X. Zhang and K. Z. Deng, "Remote based on deep learning AlexNet study on classification and evaluation of sensing image land cove," *Geo-Information Science*, vol. 19, no. 11, pp. 1530–1537, 2017.
- [18] L. Li, J. Liang and M. Weng, "A Multiple-feature reuse network to extract buildings from remote sensing imagery," *Remote Sensing*, vol. 10, no. 9, pp. 1350–1368, 2018.
- [19] W. Z. Shi and J. Q. Liu, "Building extraction from high-resolution re-motely sensed imagery based on neighborhood total variation and potential histogram function," *Journal of Computer Applications*, vol. 37, no. 6, pp. 1787–1792, 2017.
- [20] X. Sun, X. Lin and S. Shen, "High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field," *ISPRS International Journal of Geo - Information*, vol. 6, no. 8, pp. 245–271, 2017.
- [21] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [22] K. Chen, K. Fu and M. Yan, "Semantic segmentation of aerial images with shuffling convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 173–177, 2018.
- [23] M. Zhang, X. Hu and L. Zhao, "Learning dual multi-scale manifold ranking for semantic segmentation of high-resolution images," *Remote Sensing*, vol. 9, no. 9, pp. 500, 2017.
- [24] R. Kemker, C. Salvaggio and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, no. 1, pp. 60–77, 2018.
- [25] G. Chen, X. Zhang and Q. Wang, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1633–1644, 2018.

- [26] Y. Y. Xu, L. Wu and Z. Xie, "Building extraction in very high resolution remote sensing imagery using deep learning and guided filters," *Remote Sensing*, vol. 10, no. 1, pp. 144, 2018.
- [27] W. Zhao, S. Du and Q. Wang, "Contextually guided very-high-resolution imagery classification with semantic segments," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 132, no. 1, pp. 48–60, 2017.
- [28] N. Gul, S. Ahmed, A. Elahi, S. M. Kim and J. Kim, "Optimal cooperative spectrum sensing based on butterfly optimization algorithm," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 369–387, 2022.
- [29] H. Kwon, S. Hong, M. Kang and J. Seo, "Data traffic reduction with compressed sensing in an aiot system," *Computers, Materials & Continua*, vol. 70, no. 1, pp. 1769–1780, 2022.
- [30] S. U. Islam, S. Jan, A. Waheed, G. Mehmood, M. Zareei *et al.*, "Land-cover classification and its impact on peshawar's land surface temperature using remote sensing," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 4123–4145, 2022.
- [31] W. Jiang, X. Liu, D. Shi, J. Chen, Y. Sun *et al.*, "Research on crowdsourcing price game model in crowd sensing," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 1769–1784, 2021.
- [32] C. Cheng and D. Lin, "Based on compressed sensing of orthogonal matching pursuit algorithm image recovery," *Journal of Internet of Things*, vol. 2, no. 1, pp. 37–45, 2020.
- [33] X. R. Zhang, W. F. Zhang, W. Sun, X. M. Sun and S. K. Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science & Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.
- [34] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.
- [35] Y. Xuan, S. S. Li, Z. C. Chen, J. Chanussot, X. P. Jia *et al.*, "An attention-fused network for semantic segmentation of very-high-resolution remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, no. 1, pp. 238–262, 2021.