IMPROVING SEMANTIC SEGMENTATION PERFORMANCE BY JOINTLY USING HIGH RESOLUTION REMOTE SENSING IMAGE AND NDSM

Ruiqi Yang, Qinlin Dai, Haiyan Cheng, Yue Zhang, Nan Chen, Leiguang Wang*

Southwest Forestry University, Kunming, Yunnan province, China
970825548@qq.com, Daiqinling@126.com,
(603019746, 2486685593,1131689878) @qq.com, wlgbain@126.com
http://www.swfu.edu.cn/

Commission III, WG III/1

KEY WORDS: Semantic Segmentation, Deep Learning, nDSM, ResNet, Resolution Remote Sensing, Augmentation

ABSTRACT:

Semantic segmentation algorithms based on full convolutional neural network have greatly improved segmentation accuracy of high-resolution remote sensing (RS) images. However, the interpretation of RS images from single sensor is still challenging due to the variety and complexity of land objects, the extreme imbalance distributions of land objects on size and numbers. In contrast, multiple sensors can provide complementary information on the land classes, and thus benefit the interpretation. In this context, this research explores the joint use of RGB optical bands and normalized DSM (nDSM) to analyze an urban scene. The method firstly concatenated three channels RGB image and one channel nDSM band into a four-channel image. Thereafter, ResNet-101 network with fine adjustment were utilized as the backbone network to retain multiple feature information by residual blocks. Then the augmented RGB and nDSM images were used to training the network. The established model was evaluated on the Postdam test set. Results show that the proposed method achieves 86.85% on Overall Accuracy (OA), 77.42% Mean Intersection Over Union (MIOU), which is 6.88% and 11.39% higher than the result achieved by single RGB images. Especially, small targets, such as car and tree, are higher. The experimental results show that the simple structure adjustment of ResNet-101 network can achieve good segmentation performance on RS images (especially small targets) after the combination of twice augmented RGB channels and nDSM channels respectively. In addition, with the addition of nDSM, the accuracy of buildings and trees with height information has been improved.

1. INTRODUCTION

Remote sensing (RS) technology utilizes sensors to observe and detect target objects in a long distance. High-resolution RS image is an important window for earth observation (Zheng, 2017). At present, semantic segmentation of high-resolution RS images has become a hot issue in RS image interpretation, and widely used in environmental monitoring (Blaschke et al., 2000), crop cover and type analysis (Yang, 2016), forest tree species analysis (Dechesne, 2017), architectural classification of urban space and land use analysis (Rottensteiner, 2014) etc. However, there are still many complex factors in RS images such as feature diversity of a class of samples, uneven data amount of each class, target space dispersion, variable scale, complex background and shadow etc. which lead to poor segmentation performance and prone to miss segmentation. Due to the characteristics of high-resolution RS images, such as rich shape geometry and texture features, obvious topological relationship of ground object space and huge amount of data (Tang et al., 2013), traditional processing technology cannot make full use of rich details and background information, and results in a phenomenon called "rich data and poor information". This phenomenon makes the segmentation with high precision and high efficiency is still a challenging problem.

With the development of deep learning, semantic segmentation technology has made great progress. Since deep learning methods can automatically extract tailored characteristics for a specific classification task, the processing of RS images over complex scenes has a better choice (Yuan, 2021). The biggest difference between the semantic segmentation method based on convolutional neural network and traditional semantic segmentation method is that the network can automatically learn the features of images, carry out end-to-end classification learning, and greatly improve the accuracy of semantic segmentation. Standard semantic segmentation is the process of classifying each pixel into object classes, and extracting semantic information and image features from a large amount of labeled data by using deep neural network. Pixel-based methods are usually effective in extracting details and edges such as (Zheng, 2022) The quality of image semantic segmentation directly determines the quality of classification or recognition. Therefore, the realization and application of an effective image semantic segmentation algorithm is of important practical significance.

With the ever-evolving progress of remote sensing technologies, the resolution of RS image is getting higher and higher, and the ground object information is getting richer (Zheng, 2021). With the continuous improvement of semantic segmentation today, there are still many problems to be solved, mainly reflected in the following aspects.

Firstly, the inconsistence between the segmentation result of RS image and semantic information. Due to the rich information of ground objects in high-resolution RS images, general segmentation methods used for another kind of images may show poor performance on RS images. How to make the segmentation results consistent with ground truth of semantic image objects so as to improve the segmentation accuracy and the average image overlap ratio has become an urgent problem to be solved.

* Corresponding author: Leiguang Wang.
Email: wlgbain@126.com

Secondly, the phenomenon of "same object with different spectrum" and "different objects with same spectrum" in high spatial resolution RS images. The high spatial resolution RS image has vivid geometric and attribute details, which makes small targets, texture and shadow of ground objects and other interference factors detectable in images. Meanwhile, the spectral response variation of similar objects or even the same ground objects became obvious with the improvement of spatial resolution (Liu et al., 2011). Therefore, the phenomenon of "same object with different spectrum" and "foreign object with same spectrum" are common in high spatial resolution RS images, which brings great difficulties to the segmentation of relevant ground objects.

Thirdly, the segmentation performance of small targets is poor. In general network model, the basic backbone neural network has several down-sampling processes. Because the size of small targets in the feature map is relatively small, especially only one digit pixel size after down-sampling processing, which results in poor classification performance of the designed classifier on small targets (Nogueira, 2019).

In this context, the main purpose of the study is to establish a deep learning network for semantic segmentation of high-resolution RS images. In this method, three channels RGB images and one channel nDSM images in the Postdam dataset of ISPRSs are superimposed into four channel images. Four-channels images are taken as an input and then put into adjusted ResNet-101 network for training.

2. RELATED WORK

Over the past few decades, researches on RS have emphasized a lot on the application of machine learning, and many deep learning methods have been applied to semantic segmentation of RS images.

Traditional image semantic segmentation techniques mainly include threshold based, edge based, and region based segmentations, and segmentations based on the specific theory. Traditional image segmentation methods are not only difficult to meet the requirements of practical application in real-time scene understanding and image information processing, but also difficult to achieve classification accuracy and segmentation image interpretation efficiency (Liang, 2020). Semantic segmentation based on deep learning can solve the above problems.

In 2015, FCNS (Fully Convolutional Networks) (Long et al., 2015) popularized the original Convolutional Neural Network (CNN) structures. This end-to-end method can process images of arbitrary sizes, which improves processing speed compared with the traditional image block classification method. ResNet (Residual Network) (He et al., 2016) was proposed in 2016. The residual blocks of the network have two structures, "building blocks" and "bottle neck building blocks". Compared with VGGNet and GoogleLeNet, this network, identity mapping and residual mapping are used to transform identity mapping to solve the residual mapping. This method solves the problem that the accuracy decreases with the deepening of the network. SENet (Squeeze-and-Excitation Networks) (Jie et al., 2017) presents a new structural unit called "Squeeze and Excitation" blocks. It adaptively recalibrates channel characteristic responses by modeling interdependencies between the channels. These SE blocks are stacked together to form a SENet. A semantic segmentation method using multi-context paradigm to obtain the optimal patch size is proposed in (Nogueira et al., 2019). This method can capture better ground and context features at the same time, which is of great help in improving the overall classification accuracy and the classification accuracy of small targets (such as vehicles). Resunet-a (Fid, 2020) was proposed for remote sensing image segmentation in 2020. The network consists of a new deep learning architecture, Resunet-a, and a new loss function based on Dice Loss (Dice loss function). Resunet-a uses UNet code structure as the backbone, combines residual joining, empty convolution, pyramid scenario parsing pooling and multi-task reasoning.

3. METHODS

In this paper, three-channel RGB images, labels and their corresponding one-channel nDSM are augmented twice in the image preprocessing process. The augmented images are divided into training set and verification set. When the image is read, the three-channel RGB image and one-channel nDSM image are stacked. The 4-channel images are input into adjusted ResNet-101 network, and then output TIF format images compared with the test set for evaluation.

3.1 Augmentation

In this paper, 5 methods including random image clipping, gaussian blur, special affine transform enhancement (called Rotation), noise enhancement and color enhancement, were used randomly in the first augmentation at the same time. The second augmented images were obtained by horizontal, vertical and mirror inversion of the first augmented images. Each image was expanded into three images. The secondary augmented images are simultaneously put into the network. The three-channel labels and single-channel nDSM corresponding to high resolution RS images were expanded to the same number. The augmentation process is shown (in Fig. 1), where (a) is the original image, (b) and (c) are obtained after the first augmentation (clip, rotation and noise). (d), (e), (f) were obtained from (b) after the image were flipped horizontally, vertically and mirroring. (g), (h), (i) were obtained from (c) after the image were flipped horizontally, vertically and mirroring.

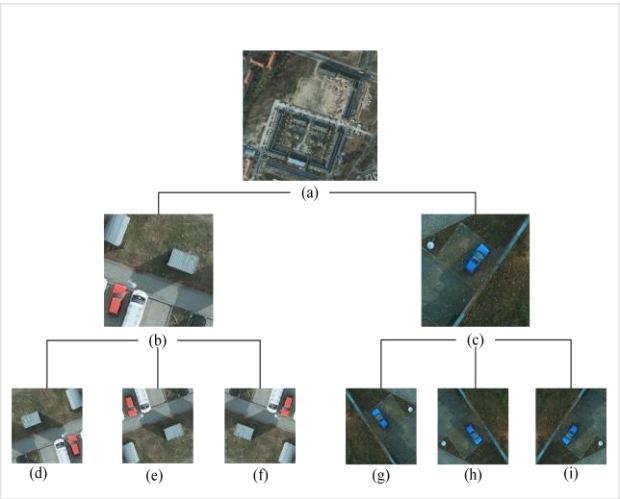

Figure 1. The augmentation processes.

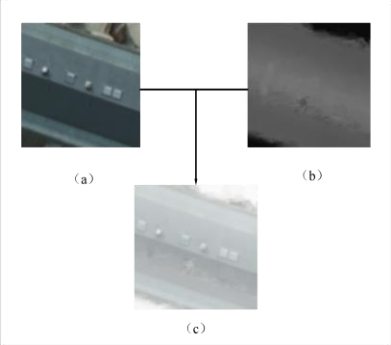

Figure 2. RGB is combined with nDSM.

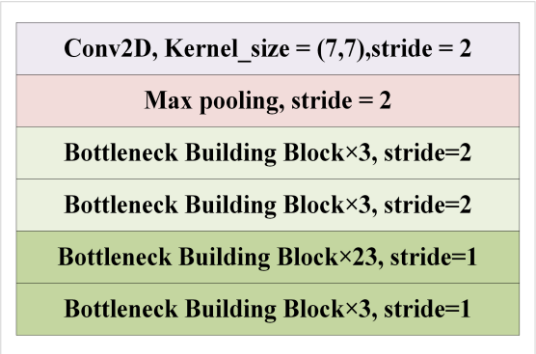

Figure 3. Adjusted ResNET-101

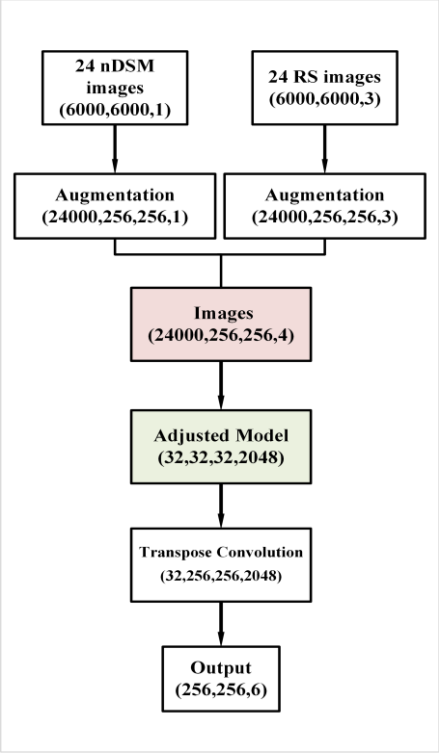

Figure 4. The overall architecture.

3.2 The Channel Stacking

In this paper, RGB and nDSM of RS images are read by Tifffle function, whose array forms are (H, W, 3) and (H, W) respectively. At the same time, they are stacked with channel numbers. Then the array output of RS images is (H, W, 4). As shown (in Fig. 2),where (a) is the 3-channel RGB image, (b) is the 1-channel nDSM image, (c) is the image visualization combining RGB image and nDSM image.

3.3 ResNet-101

The characteristic of this network is to use a kind of connection called "short connection", which effectively solves the problem of gradient explosion and gradient disappearance caused by the deepening of the network. When using this network to extract features, this method changed the stride, retained the feature graph to a greater extent, and reduced the loss of small target information.

3.4 Model Structure

The previous three sections describe the preprocessing methods and channel merging. This section will introduce the architecture of overall approach in detail.

The model takes ResNet-101 network as the backbone, and the last two layers of the model, i.e., GlobalAvgPool2D and Flatten layers are discarded. The method uses ResNet-101 as main framework with five convolution layers as shown (in Fig. 3). The first convolution layer is the convolution with a kernel size of 7×7 and the Max Pooling with a kernel size of 3×3 and the stride of 2. The second layer and the third layer consist of three “bottleneck” building blocks with stride of 2. The fourth convolution layer and fifth convolution layer consist of 23 and 3 “bottleneck” building blocks respectively, the stride both is 1. Compared with the original network, the stride size of last two layers changes from 2 to 1, and output shape of the feature map changes from (8, 8, 2048) to (32, 32, 2048).

The whole architecture is presented (in Fig. 4). The first step of the method is augmentation. RS images are augmented twice, the first augmented method is clipping, random Gaussian blur, random special affine Transform enhancement, random noise enhancement and color enhancement, and the second augmentation mode is horizontal, vertical and mirror flipped. The nDSM images also are augmented twice. The first augmentation is clipping. The second augmentation is the same as RS image. The second step is to concatenate three-channel RGB images and one-channel nDSM images. Thirdly, the image is restored to its original size by transpose convolution, and the convolution kernel size of transpose convolution is set as 64×64, the stride set to 8, and the initialization is carried out by bilinear interpolation. Finally, the convolution layer of 1×1 is used. The dataset has 6 categories (including background), so the output shape is (256, 256, 6).

4. EXPERIMENTAL RESULTS AND DISSCUSSION

In this section, experimental settings will be introduced more specifically. Section 4.1 shows the dataset used in this experiment. Section 4.2 describes the implementation details of the experiment. Evaluation functions are provided in Section 4.3. Finally, the experimental results are analyzed in section 4.4.

4.1 Dataset

ISPRS 2D Semantic Labeling Contest Potsdam dataset (https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx) is a high-resolution aerial image dataset. This dataset has 38 patches of the same size (6000 ×

6000 pixels) and a spatial resolution of 0.5 meter. Each patch of the dataset was extracted from orthophoto images, with a total of 24 RS images, and corresponding semantic labels were performed on them. RS images files are composed of different channels, including IRRG (3 channels, IR-R-G), RGB (3 channels, R-G-B), (1 channel, nDSM) and (one channel, DSM). In this experiment, RGB images are combined with nDSM images. Dataset labels are divided into six categories (Background, Building, Impervious surfaces, Tree, Low Vegetation, and Car respectively).

Therefore, 24 images were used as training set and validation set. The 24 RS images augmented 24,000 images were randomly divided into train set and validation set. When there were 15 images for training, one image was used for validating. There are 22,400 training images and 1,600 verification images.

Since the Benchmark Challenge ended in the summer of 2018, all reference data for all benchmarks are available for download, so 14 images without corresponding semantic labels were served as test set. 4000 images randomly cropped from 14 images were put into the pre-trained model for prediction. 4,000 TIF format images were generated and compared with the reference labels provided in the official benchmark.

4.2 Implementation Details

In this experiment, training equipment of deep learning network is 8-core 16-thread Intel I9-9900K CPU. NVIDA RTX3090 Graphic card, 24G Memory with CUDA11.2.

The software environment is 64-bit Microsoft Windows10, operating system and the development platform is Anaconda-5.2.0. The built-in Python version is 3.8.8. The deep learning software framework is TensorFlow2.5.0. Adam optimizer was used with a 1×10^{-3} learning rate. A total of 100 training epochs with a batch size of 32.

4.3 Model Evaluation Function

In order to comprehensively evaluate the performance of the proposed model, Overall Accuracy (OA), Precision, Recall, F1, and Mean Intersection Over Union (MIOU) were used to evaluate experimental results. The above evaluation indexes are often used in previous papers and compared with the recognized evaluation indexes of semantic word segmentation. The calculation formula of each evaluation index is as follows:

$$OA = \frac{TP + TN}{P + N} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

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

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{FN + FP + TP} \tag{5}$$

Where P is the number of positive samples; N is the number of negative samples; TP is the number of positive samples; FP is the number of positive samples predicted falsely; TN is the number of negative samples predicted correctly FN is the number of negative samples predicted falsely.

4.4 Evaluation and Discussion

This part provides a comparison of three methods. The first method is the segmentation result without using our proposed method and simply using ResNet-101 model without adding nDSM (Table1 None and Table2 None). The second method is the segmentation result without nDSM, but our method was used (Table1 Ours and Table2 Ours). The third method is the segmentation result added the one-channel nDSM and ours was used. Experimental results show that the third method has the best results (Table1 Ours+nDSM and Table2 Ours+nDSM).

Global Model Analysis: OA, Precision, Recall, F1 and MIOU were applied to evaluate our model on the Potsdam test set (in Table1). The segmentation result of the first method that without nDSM and ours is not used were the lowest for the five evaluations. The first method reached the OA and MIOU to 79.97% and 66.03%. The second method reached the OA and MIOU to 86.18% and 75.95%. The third method added one-channel nDSM based on the second method, and it can be seen that the five evaluation indicators of the third method have been improved. The third method reached the OA and MIOU to 86.85% and 77.42%. Compared with the "None" method, our method with nDSM improved OA and MIOU by 6.88% and 11.39%.

Method	OA	Precision	Recall	F1	MIOU
None	79.97	78.41	79.80	79.19	66.03
Ours	86.18	85.45	86.02	85.97	75.95
Ours+nDSM	**86.85**	**86.07**	**86.99**	**86.99**	**77.42**

Table 1. Evaluation results of the three methods.

Class	None		Ours		nDSM + Ours	
	F1	MIOU	F1	MIOU	F1	MIOU
Building	88.43	79.27	94.52	89.61	**94.66**	**89.87**
Road	83.28	71.35	89.74	81.39	**89.81**	**81.51**
Tree	69.57	53.33	76.70	62.20	**79.28**	**65.68**
Vegetation	74.93	59.91	80.70	67.64	**81.68**	**69.02**
Car	79.93	66.29	88.21	78.90	**89.50**	**81.00**

Table 2. Compare the F1 and MIOU of each class on the Potsdam test set

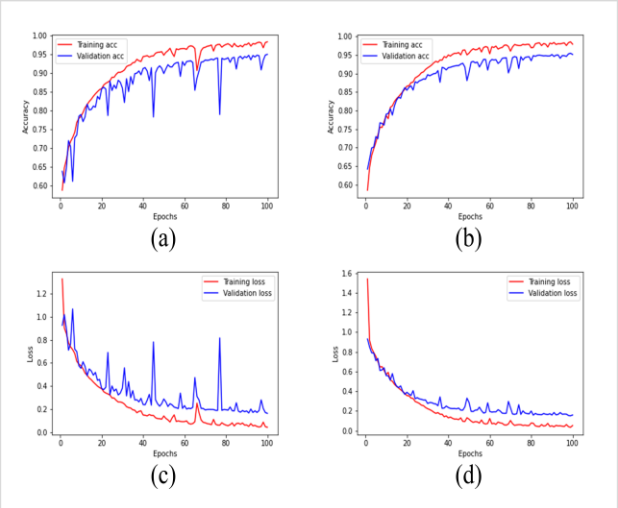

Figure 5. Accuracy and loss of training set, validation set.

The output shows competitive performance in all classes (in Table 2). It can be seen that F1 and MIOU of each category have been improved in our method compared with method 1. The segmentation performance of buildings is the best, while tree is the lowest. Compared with method "None", segmentation results (MIOU) of building, road, tree, vegetation and car are improved 10.6%, 10.16%, 12.35%, 9.11%, 14.29%, respectively. Especially, small targets like car and tree increased more than the other classes in F1 and MIOU

Convergence Analysis: The convergence of our method is analysed in this part (in Fig. 5). The (a) and (c) are visualizations of training validation accuracy and loss without nDSM, while (b) and (d) are visualizations of training validation accuracy and loss with nDSM.

As can be see from the result, both methods converge gradually, and the accuracy and loss are more stable after nDSM is added. In our method, the training accuracy can reach 98.57% and the verification accuracy can reach 95.48%. The lowest training loss can reach 0.0348 and verification loss can reach 0.1516.

Comparison of Experiments on Potsdam Datasets: This part compares the segmentation results in different methods (in Table 3)The method used nDSM(Wenkai Zhang et al., 2018) also. At the same time, the segmentation results of similar ResNet-101 networks (Wang Y et al., 2019) in the same Postdam data set. So we compare the class segmentation results of the two methods. (in Table 4).

In general (in Table 3), our method achieved good results and improved efficiently only through augmentation and model adjusted under the same Postdam dataset and the same nDSM segmentation. The OA and F1 of the proposed method are both higher than the segmentation results of the fusion proposed by the second method. The third method (Wang Y et al., 2019) uses Atrous Spatial Pyramid Pooling (ASPP) in combination with ResNet networks and Superpixel-CRF, while our proposed approach achieves similar results by only slightly tweaking the ResNET-101 network.

Method	OA	F1	MIOU
Sherrah J et al., 2016	87.8	81.2	--
Wenkai Zhang et al., 2018	79.21	79	--
Wang Y et al., 2019	--	88.80	--
Ours+nDSM	86.85	86.99	77.42

Table 3. Compare the F1 and MIOU of each class on the Potsdam test set

We can see that (in Table4.) the buildings, trees, vegetation and cars are better than the segmentation results of (Wenkai Zhang et al., 2018). The result of car segmentation shows that the proposed method is 15.5% higher than the first method. Our proposed method also performs better on buildings with high levels of information.
When compared with the method proposed by (Wang Y et al., 2019), it is found that the accuracy of small-target vehicles can reach 89.5% even without adding ASPP to increase the complexity of the network and without using Superpixel-CRF for prediction.

Method	F1 Scores		
	Wenkai Zhang et al., 2018	Wang Y et al., 2019	Ours+nDSM
Building	83.00	95.9	94.66
Road	90.00	90.20	89.81
Tree	77.00	84.30	79.28
Vegetation	72.00	83.90	81.68
Car	74.00	89.60	89.50

Table 4. Comparison of F1 scores on Potsdam test set.

Compared with Three Methods: The results (in Fig. 6) show the comparison of prediction results of different strategies on test set, and the prediction performance of ours in the study is better. The segmentation only using ResNet-101 without nDSM lost more tree information, and the classification boundary of cars were blurred (in Fig. 6 None). Compared with the segmentation results without using our method and simply using ResNet-101 model without adding nDSM, small targets (cars) predicted by our method have clearer boundaries and less predicted tree loss information (in Fig. 6 Ours). When we add nDSM and use our method, it can be seen that the segmentation results of building and tree with height information are better than the other two methods. There is less misclassification phenomenon (in Fig. 6 Ours+nDSM).

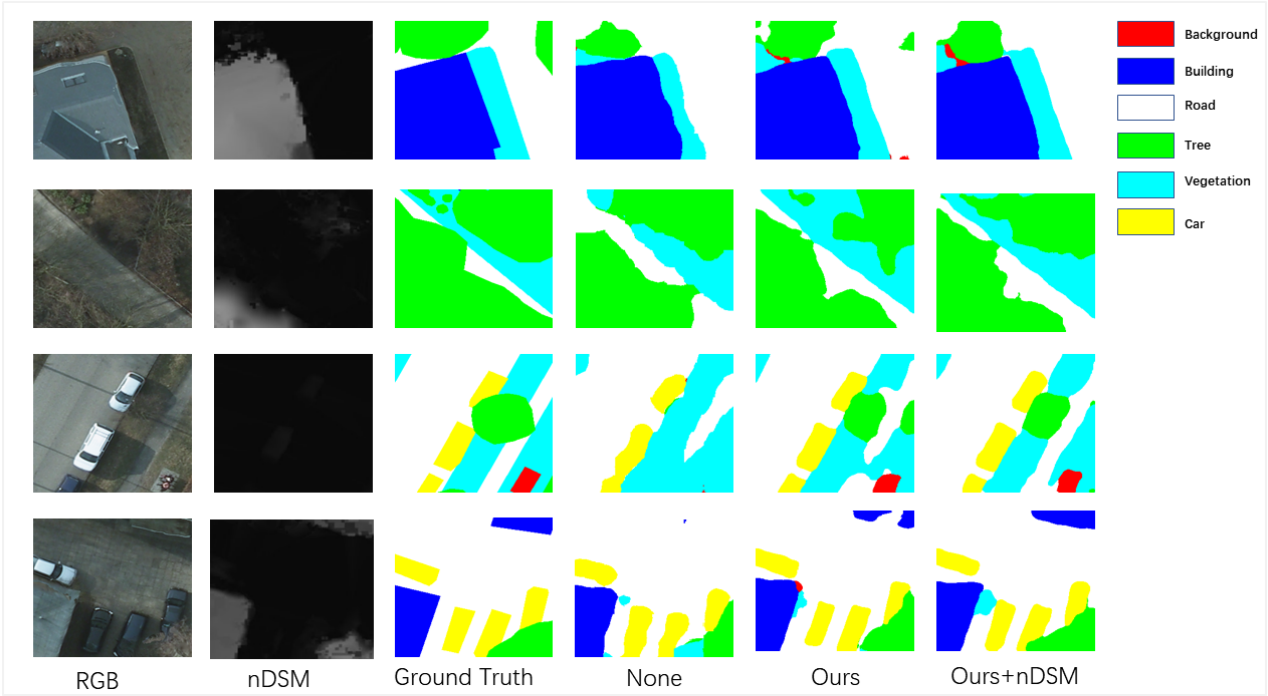

Figure 6. Comparison of different methods

5. CONCLUSION

In the current study, three-channels RGB image and one-channel nDSM image in Postdam data set provided by ISPRS are used for semantic segmentation. The method in this paper is adjusted to ResNet-101 network. The OA and MIOU can be improved quickly by using a simple method.

Experimental results show that the proposed method achieves good results in five evaluation functions by simple adjustment. Compared with non-NDSM, the boundary with nDSM is clearer. In the end, compared with the original method, the accuracy of trees and small targets (cars) prone to misclassification is greatly improved.

ACKNOWLEDGEMENTS

Thanks to ISPRS for providing the Postdam dataset of airborne data images of urban classification, which is free and publicly available. For more information, see the ISPRS Test Project on Urban Classification and 3D Building Reconstruction-2D Semantic Labeling Contest official website.

This work was supported in part by the National Natural Science Foundation of China under Grants 32160369, 31860182, 41961053, and 41771375, in part by the Key Development and Promotion Project of Yunnan Province under Grant 202102AE090051; in part by the Fund of Reserve Talents for Young and Middle-Aged Academic and Technological Leaders of Yunnan Province under Grant 2018HB026.
(Corresponding author: Leiguang Wang.)

REFERENCES

Zheng C, Zhang Y, Wang L. Semantic segmentation of remote sensing imagery using an object-based Markov random field model with auxiliary label fields [J]. IEEE Transactions on geoscience and remote sensing, 2017, 55(5): 3015-3028.

Blaschke T , Lang S , Lorup E , et al. Object-Oriented Image Processing in an Integrated GIS/Remote Sensing Environment and Perspectives for Environmental Applications. 2000.

Yang S, Chen Q, Yuan X, et al. Adaptive coherency matrix estimation for polarimetric SAR imagery based on local heterogeneity coefficients[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(11): 6732-6745.

Dechesne C, Mallet C, Le Bris A, et al. Semantic Segmentation of Forest Stands of Pure Species as a Global Optimization Problem[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, 4: 141.

Rottensteiner F, Sohn G, Gerke M, et al. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction[J]. ISPRS journal of photogrammetry and remote sensing, 2014, 93: 256-271.

Yuqi Tang. Object-oriented multi-feature change detection for high resolution image cities. Wuhan University, 2013.

Yuan X, Shi J, Gu L. A review of deep learning methods for semantic segmentation of remote sensing imagery[J]. Expert Systems with Applications, 2021, 169: 114417.

Zheng C, Chen Y, Shao J, et al. An MRF-Based Multigranularity Edge-Preservation Optimization for Semantic Segmentation of Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, PP (99):1-5.

Zheng C, Zhang Y, Wang L. Multigranularity Multiclass-Layer Markov Random Field Model for Semantic Segmentation of

Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, PP (99):1-20.

Jianhua Liu. Research on adaptive segmentation method of high spatial resolution remote sensing image [D]. Fuzhou University,2011.

Nogueira K, Dalla Mura M, Chanussot J, et al. Dynamic multicontext segmentation of remote sensing images based on convolutional networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(10): 7503-7520.

Xinyu Liang, Chen Luo, Jichuan Quan, et al. Research progress of image semantic segmentation based on deep learning [J]. Computer Engineering and Applications, 2020.

Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

Nogueira K, Dalla Mura M, Chanussot J, et al. Dynamic multicontext segmentation of remote sensing images based on convolutional networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(10): 7503-7520.

Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 94-114.

Sherrah J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery[J]. arXiv preprint arXiv:1606.02585, 2016.

Zhang W, Huang H, Schmitz M, et al. Effective fusion of multi-modal remote sensing data in a fully convolutional network for semantic labeling[J]. Remote Sensing, 2018, 10(1): 52.

Wang Y, Liang B, Ding M, et al. Dense semantic labeling with atrous spatial pyramid pooling and decoder for high-resolution remote sensing imagery[J]. Remote Sensing, 2019, 11(1): 20.

https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-potsdam/