ISSN: 2664-9640

Lightweight Road Intelligent Extraction Algorithm For Asymmetric Codec Networks

Peng Xiao, Hao He, Yongfei Li, Dongfang Yang

College of Missile Engineering the Rocket Force University of Engineering, Xian, Shanxi 710025, China

Abstract

In recent years, the application of road network extraction technology in UAV real-time positioning has attracted more and more attention. At present, the accuracy of extracting road network from aerial images using deep learning technology has been greatly improved. However, due to the network model caused by deep depth and large amount of parameters, this kind of algorithm often can not achieve real-time road network extraction. Aiming at the demand of road real-time intelligent extraction, a lightweight algorithm for road intelligent extraction based on asymmetric codec network is proposed. Firstly, the E-Net network with asymmetric codec structure is compressed to remove the redundant parameter layer and reduce the network model storage; Secondly, the convolution is optimized, the conventional convolution is replaced by asymmetric convolution to reduce the parameters , and the hole convolution is introduced to increase the receptive field to reduce the information loss; Fin ally, aiming at the imbalance of road proportion, the cross entropy loss function is improved to solve the tendency of loss function. The improved algorithm is trained on the road data set of deep globe large data set, tested on the deep globe road data set, and compared with the better ASPP -U-net network and D-Link network in extraction accuracy and reasoning speed. The experimental results show that under the condition of ensuring the road extraction accuracy, the extraction speed is increased by 1.72 times compared with ASPP -U-net network and 1.40 times compared with D-Link network. Therefore, it has a good engineering application prospect.

Keywords

Aerial image, road extraction , Lightweight, remote sensing image.

1. Introduction

Visual navigation technology is a new type of navigation technology that has emerged rapidly in recent years. The advantages of low cost, small size, low power consumption and less susceptible to interference make it widely used in various fields[1]-[4].

UAV images contain a variety of ground object information. Among them, road information has the characteristics of wide distribution, strong robustness and high recognition. The research on real-time UAV positioning technology based on road network information has received more and more attention. How to quickly and accurately extract road targets from UAV images has become a hot topic in current research.

Traditional road extraction uses pixel, area block, knowledge model and other methods [5] to segment the image [6]-[12]. These methods determine the location of the road by designing the structural features of the road, such as intersection connections, parallel double edges, road width, texture, etc., and complete the road extraction through the connection algorithm. The above methods have achieved certain results in road extraction tasks, but a large number of threshold parameters need to be established in the design of feature extraction algorithms, and

a large number of parameters need to be manually adjusted, resulting in a lot of resource consumption.

In recent years, with the development of computer vision technology and the excellent performance of convolutional neural networks (Convolutional Neural Networks, CNN) in image semantic segmentation [19]-[23], the use of deep learning technology to achieve intelligent extraction of roads has become more and more popular. much attention. Compared with traditional road extraction methods, the use of deep learning methods to extract roads in aerial images has subversive improvements in accuracy and speed. Reference [16] uses convolutional neural network to extract features of roads in remote sensing images. Due to the low efficiency of sliding windows and too heavy calculation, it is difficult to apply to practical engineering applications. Reference [17] proposes Fully Convolutional Networks (FCN), which converts the fully connected layer at the end of CNN into a convolutional layer, which can accept input images of any size, and reduces significantly when the image is larger However, after the network undergoes multiple downsampling, the resolution of the input feature map is greatly reduced, and the spatial detail information is greatly lost. Reference [18] proposed a U-Net model with a symmetric structure, which uses a classical encoder-decoder structure to reduce the complexity of the model, and fuses information at the same level through skip connections. which effectively alleviates the problem of spatial information loss caused by downsampling. Reference [24] adopts the encoder-decoder network as the main architecture of the segmentation network, and uses atrous spatial pyramid pooling between the encoding layer and the decoding layer of the network to capture more spatial information. The method achieves good results in the task of road extraction. However, due to the deep convolutional network layer, huge network structure, and a large number of network parameters, the reasoning time is too long, and it is difficult to meet the real-time segmentation requirements.

Aiming at the above problems, a lightweight algorithm for road intelligent extraction with asymmetric codec network structure is proposed. First, network pruning is performed on the E-Net [25] network to eliminate redundant parts in the model and reduce the amount of model parameters. Secondly, optimize the convolution block in the network, and replace the conventional convolution with atrous convolution and asymmetric convolution. This method mainly has the following two advantages: 1) Pruning the network can reduce the weight of the network structure , reduce memory usage and improve the real-time processing performance of the network; 2) The convolution module is optimized for processing, introducing a hollow convolution module to increase the receptive field to obtain more spatial information, and introducing asymmetric convolution to reduce the amount of network parameters and make the network model more efficient Lightweight.

2. Organization of the Text

E-Net network is a real-time semantic segmentation network proposed by Paszke in 2016. It has a simple network structure and can complete pixel-level semantic segmentation of targets for specific tasks. Compared with a large number of floating-point operations in deep neural networks, its smaller parameter amount and faster inference speed meet the requirements of real-time road extraction from aerial images. This paper proposes a light-weight intelligent road extraction algorithm from aerial images by performing model compression and convolution optimization on E - net network.

The E-Net network inherits the main frame of the encoder-decoder (Encoder-Decoder). In order to extract more detailed feature information, most of the network implements the Encoder, and a small part implements the Decoder, forming an asymmetric Encoder-Decoder structure.this structure reduces the convolution operation in the decoder, reduces the amount of parameters, and improves the network inference speed.

ISSN: 2664-9640

In the task of road intelligent extraction, the network is required to perform reasoning and prediction on aerial images quickly and accurately. Due to the single type of semantic segmentation, a deep network layer is not required. At the same time, in order to ensure the accuracy of road extraction, this paper designs the network from the following two aspects to improve the real-time and accuracy of model prediction:

Network pruning, pruning the E-Net model, reducing redundant parameters, further lightweighting the network structure, and improving inference speed;

The convolution module is optimized, and the conventional convolution module is improved to increase the diversity of the convolution module and increase the receptive field to capture rich context information and further improve the segmentation accuracy.

2.1. Model pruning

The E-Net network is a real-time semantic segmentation model. The overall architecture of the network includes seven stages as shown in Figure 1. Except for the initial module and the full convolution module, it is divided into five stages. The first stage consists of a downsampling module. and 4 conventional convolution blocks; the second stage consists of 1 down-sampling module, 4 dilation rate (in order of 2^n($n = 1,2,3,4$)) atrous convolution module, 2 conventional convolution modules, 2 asymmetric convolution blocks; third Stage 2 and Stage 2 have the same structure; the first three stages constitute the encoder part of the network, and the fourth and fifth stages are upsampled to constitute the decoding part of the network.

Name	Type	Output size
initial		16 × 256 × 256
bottleneck 1.0	down sampling	64 × 128 × 128
4×bottleneck 1.x		64 × 128 × 128
bottleneck2.0	down sampling	128 × 64 × 64
bottleneck2. 1		128 × 64 × 64
bottleneck2. 2	dilated 2	128 × 64 × 64
bottleneck2. 3	asymmetric 5	128 × 64 × 64
bottleneck2.4	dilated 4	128 × 64 × 64
bottleneck2. 5		128 × 64 × 64
bottleneck2.6	dilated 8	128 × 64 × 64
bottleneck2.7	asymmetric 5	128 × 64 × 64
bottleneck2. 8	dilated 16	128 × 64 × 64
Repeat section 2, without bottleneck2.0		
bottleneck4.0	upsampling	64 × 128 × 128
bottleneck4.1		64 × 128 × 128
bottleneck4.2		64 × 128 × 128
bottlenecks5.0	upsampling	16 × 256 × 256
bottlenecks5. 1		16 × 256 × 256
fullconv		C × 512 × 512

Figure 1: E-Net network architecture

Road target extraction can be regarded as a two-class semantic segmentation task. In view of the sparse distribution of roads in the image, atrous convolution is introduced in the second stage of the network to increase the receptive field of the image to capture a more complete space. information. In the third stage of the network, since there is no dimensional change in the image, it simply repeats the work of the second stage. For the single-type segmentation task of road extraction, adding the third part will not improve the network too much. On the contrary, it will cause parameter redundancy, generate unnecessary parameters in the encoding process, and reduce the inference speed. Based on this, this paper prunes the third stage of the E-Net

ISSN: 2664-9640

network to remove redundant parameters in the network, lighten the network model, and improve the speed of network inference.

2.2. Convolution Optimization Processing

Roads contain rich semantic information. Attributes such as straightness, intersection, and continuity contribute to the accuracy of classification. At the same time, features such as symmetrical double edges, texture, and color also have a certain impact on improving the accuracy of detail segmentation. The network will reduce the resolution of the feature map during the downsampling process, resulting in the loss of edge information. In order to improve the road extraction accuracy and reduce the loss of edge information, this paper proposes a convolution optimization processing scheme. The specific improvement method is shown in Figure2. Optimize the four conventional convolution modules in the first stage of the network, replace the second conventional convolution module with a dilated convolution module with an expansion rate of 2, and replace the third conventional convolution module with 5×1, 1×5 asymmetric convolution, the fourth regular convolution module is replaced by a dilated convolution module with a dilation rate of 4.

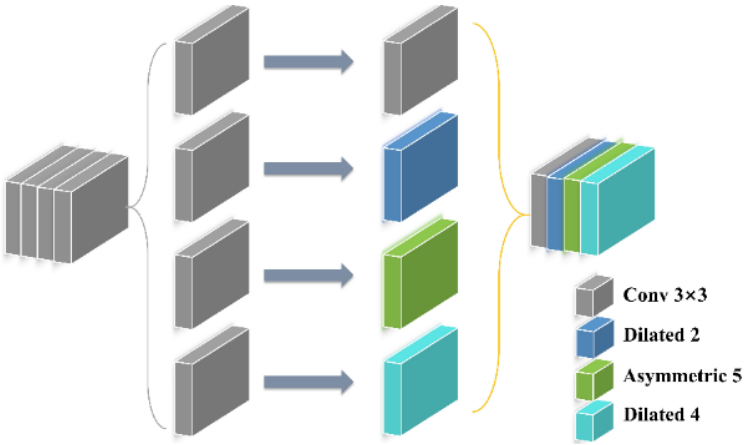


Figure 2: convolution optimization

The hole convolution increases the receptive field without reducing the spatial resolution, can extract more feature information, reduce the loss of spatial information, and retain more detailed information. At the same time, the multi-scale hole convolution can capture more fully the context information of the network; asymmetric convolution can reduce the amount of network operation and improve the operation efficiency. At the same time, the diversity of convolution in network coding can reduce the risk of overfitting.

2.3. Overall network architecture

The overall network architecture designed in this paper is shown in Figure 3. The encoder part has been downsampled twice, which effectively reduces the resolution loss caused by network downsampling and retains more complete spatial information; the diversity of convolution modules It also makes the network have stronger generalization performance, and the network can extract relatively complete road information in the coding part. The simpler decoding part significantly reduces the network parameters, while the shallower network structure greatly reduces the amount of network parameters, which increases the network's timeliness performance and engineering application value.

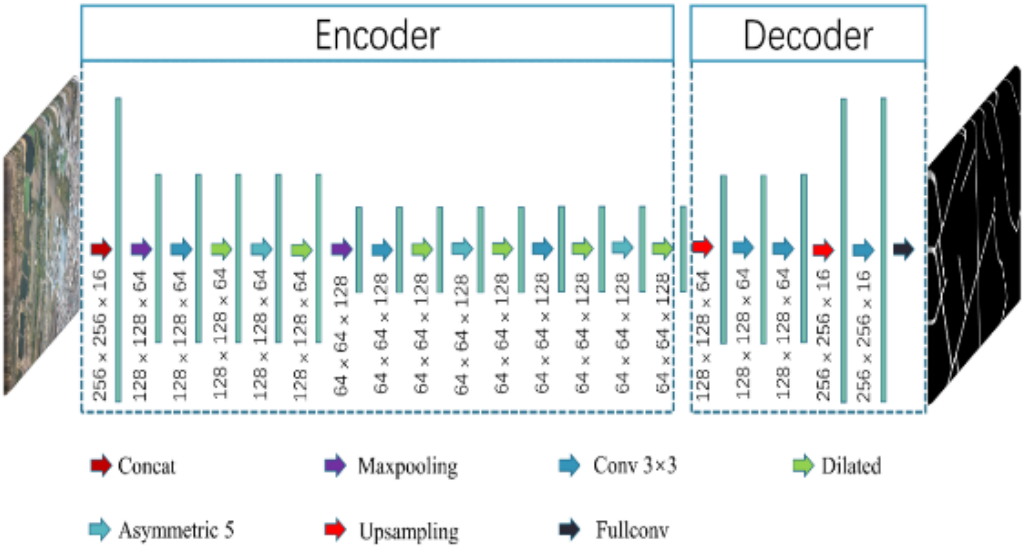


Figure 3: asymmetric codec network structure for road extraction task

2.4. Loss function improvements

Road extraction is a pixel-level semantic segmentation task. In image classification and semantic segmentation tasks, the optimization network usually uses the cross entropy loss function:

$$L(y,p) = -y\log(p) - (1-y)\log(1-p) \tag{1}$$

where p is the predicted probability; y is the true value of road pixels. There are only two types of positive samples and negative samples in the binary classification task. y Take 0 to represent a negative example, and y take 1 to represent a positive example, so there are:

$$L(y,p) = \begin{cases} -\log(p) & y=1 \\ -\log(1-p) & y=0 \end{cases} \tag{2}$$

In the road extraction task, the predicted image only contains the background and the foreground (road), so the road extraction can be regarded as a two-category semantic segmentation task, so the constraint function in the road extraction task usually selects the two-category cross entropy loss function .

$$L = -\frac{1}{N}\left(\sum_{y_i=1}^{m}\log(p) + \sum_{y_i=0}^{n}\log(1-p)\right) \tag{3}$$

where: m is the number of road pixels, is the number n of background pixels, N is the total number of pixels in the sample, and $m+n=N$. $y_i=1$ Represents the i th pixel as the road, and $y_i=0$ represents the i th pixel as the background. In the road extraction task, the proportion of the road part in the image is extremely small, and the distribution of positive and negative samples is uneven (as shown in Figure 4). In Equation (3), each gradient return gives the same attention to the road and the background. When the positive and negative samples are unbalanced, the distribution of the loss function will be more inclined to the background, resulting in a decrease in model training efficiency.

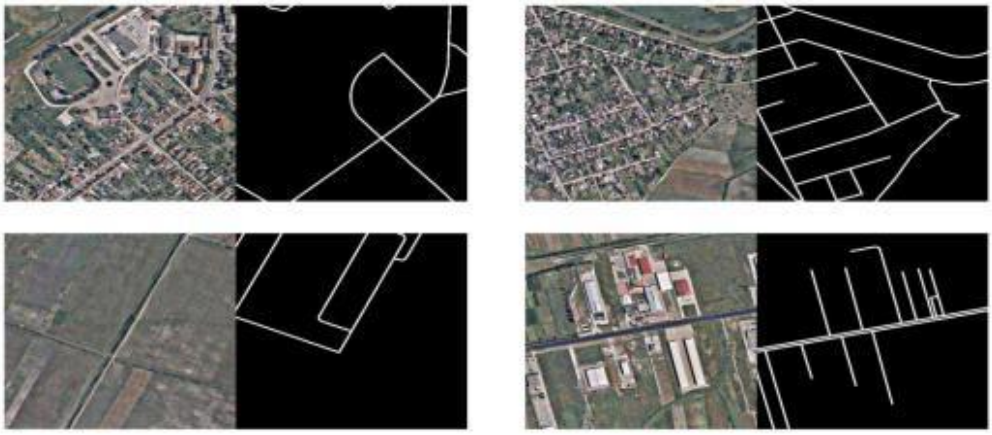


Figure 4: distribution of positive and negative samples

Therefore, giving the same attention to positive and negative samples does not meet the requirements of road extraction tasks. Aiming at the problem of the tendency of loss function caused by sample imbalance, this paper redesigns the loss function suitable for the road extraction task:

$$L = -\frac{1}{N}\left(\sum_{y_i=1}^{m} \alpha \log(p) + \sum_{y_i=0}^{n} (1-\alpha)\log(1-p)\right) \tag{4}$$

In formula (4), the scale factor α represents the weight given to the road pixels, which $1-\alpha$ represents the weight given to the background pixels. Since the proportion of positive and negative samples is seriously unbalanced, it is very important to reasonably design the scale factor α to constrain model training. Therefore, considering the propensity of the loss function, the scale coefficient is designed as:

$$\alpha = \frac{n}{m}(1-\alpha) \tag{5}$$

According to the actual task needs, this paper sets the α value to 0.8, which effectively improves the weight ratio of road pixels and solves the problem of function tendency caused by the small proportion of roads.

3. Experiments

3.1. Datasets

Test sample data using Deep The road dataset in the large globe dataset includes urban, rural, suburban, seaside, tropical rainforest and other scenes. All image sizes are 1024 × 1024. The article selects 4064 images as the training sets, 1360 image as a validation set, and 325 test images.

Training a network model with good performance usually requires a large amount of data as support, and the amount of existing data is difficult to meet the needs of network training. In this paper, the training set is flipped horizontally, vertically, and horizontally and vertically (as shown in Figure 5), and the data volume is expanded to 4 times the original data volume to expand the data volume and improve the data diversity.

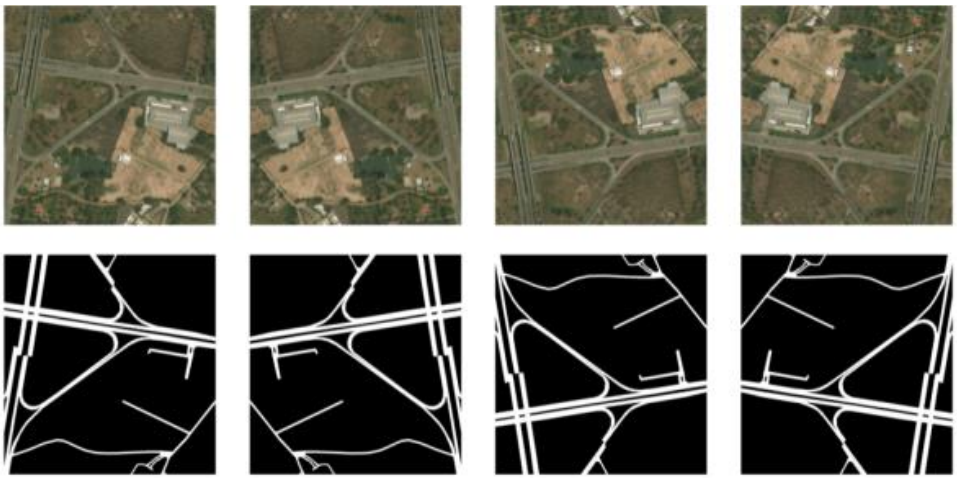


Figure 5: flipping of data

3.2. Evaluation indicators

Road extraction can be regarded as a two-class semantic segmentation problem, and the prediction results can be divided into two categories: road and non-road. For the two-category problem, according to the combination of the true value category and the predicted value category of the sample, the sample data can be divided into true positive (TP), false positive (FP), true negative (TN), and false negative (FN) Four categories. This paper uses evaluation indicators commonly used in the field of semantic segmentation: precision, recall, and comprehensive evaluation indicator F1-Score (F1) to evaluate the performance of the model, which are defined as equations (6)-(8) respectively.

$$precision = \frac{TP}{TP + FP} \tag{6}$$

$$recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} = \frac{2TP}{2TP + FN + FP} \tag{8}$$

Among them, the precision rate and recall rate correspond to the correctness and completeness in the classical road extraction [26]-[27] task, respectively. Therefore, using the above indicators can effectively quantify the road extraction effect.

In order to verify the timeliness performance of the method in this paper, three common indicators for measuring the size of the network model are used to evaluate the timeliness of the network: the amount of computation, the amount of parameters, and the size of the model.

4. Experimental Results and Discussion

4.1. Training details

The hardware environment tested in this paper is Ubuntu18.04 system, CPU E5-2603, memory 8G, GeForce GTX 1080 graphics card, GPU acceleration library is CUDA 10.1, based on Python3.6 programming language, pytorch1.6.0 deep learning open source library for algorithm implementation. At the same time, reasonable configuration of network hyperparameters is very important for model training. In this paper, Adam optimization algorithm is used, which can iteratively update neural network weights based on training data. According to the convergence of the model, set the initial learning rate α=0.001, β1=0.9, β2=0.999, ε=0.00000001, limited by computing resources, the batch size is set to 2.

ISSN: 2664-9640

4.2. Test Results and Analysis

In order to demonstrate the effectiveness of the algorithm proposed in this paper, a comparative experiment is conducted with ASPP-U-Net [24] and D-L ink [28], which have better road extraction accuracy. U-Net has a typical encoder and decoder network architecture, a completely symmetrical U-shaped structure and a skip connection method, which makes it achieve good results in medical image segmentation tasks. In the encoder stage of the network, a total of 4 downsampling times, in order to capture high-level semantic information; in order to reduce the loss of spatial information caused by downsampling, skip connections are used to capture more underlying information in the upsampling process, but this method can only recover some shallow information. Based on the basic architecture of the U-Net network, ASPP-U-Net increases the receptive field of feature points by introducing atrous convolution and multi-scale perception modules. The D - Link network adopts ResNet34 with strong feature extraction ability as the encoding layer of the network in the encoding stage, which significantly improves the feature extraction ability of the model and performs well in road extraction tasks.

When designing the network architecture, this paper fully considers the relatively single characteristics of the road segmentation task, and only downsamples the input image twice, which preserves the resolution of the feature map to a large extent and reduces the loss of spatial information. At the same time, the shallower encoding layers and smaller decoding layers provide the network model with good timing performance.

4.3. Remote Sensing Image Data Test Results and Analysis

In order to demonstrate the feature extraction ability of the model, the test data in the Deep Globe road dataset is used for testing. The test analysis results are shown in Table 1. The accuracy, recall and F1 scores of the test results are 73.50 % and 99.56 %. By observing Figure 6, the overall performance of road extraction is good, and the ground truth value can be predicted almost completely.

Table 1: Comparison of road extraction results in re mote images

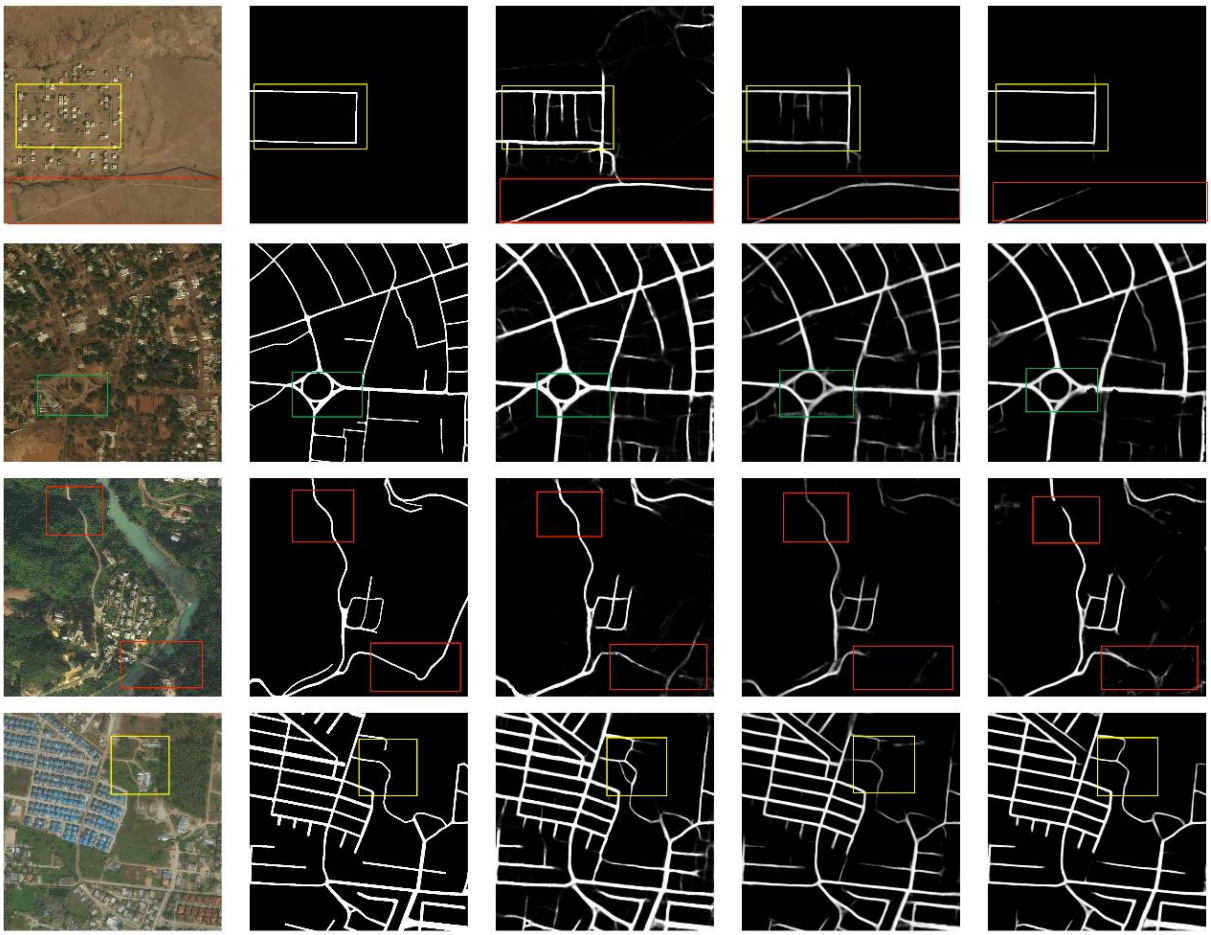
Method	Recall	Precision	F1-score
ASPP-U-Net	0.8836	0.7400	0.7857
D -Link	0.9976	0.7523	0.8366
Ours	0.9956	0.7350	0.8131

In the existing road dataset, there are some small roads that are intentionally not marked, as shown in the yellow rectangle box in Figure 6, both ASPP-U-Net and D-Link ignore the small road parts to varying degrees, while this paper The algorithm can effectively extract this part, which proves that the algorithm in this paper has strong feature extraction ability. The part framed by the green rectangle is a roundabout, and the part where it connects with the road has holes. Both the algorithm in this paper and the two comparison algorithms can clearly identify the empty parts. The part outlined in red is the road area with insignificant road features or occluded by trees. Both the ASPP-U-Net model and the D - Link model cannot effectively identify the road in the road prediction.

Both comparison methods use 4 downsampling operations in the encoding phase of the network, which greatly reduces the feature resolution of the image, resulting in more and more blurred spatial information of the input road feature map. Although the two network architectures are in the encoding layer A multi-scale feature fusion module is introduced in the middle of the decoding layer to learn road information of different spatial scales, but it still cannot fully recover the road feature spatial information.

The algorithm in this paper only performs two downsampling operations on the input road feature image, effectively retaining the road space information, so that the model can extract

more detailed road information during the training process; The optimization process is carried out to increase the feature receptive field, retain more detailed spatial information, and enhance the feature extraction ability of the network.



(a) Image (b) Ground truth (c) Ours (d) ASPP-U-Net (e) D-Link

Figure 6: visual comparison of road extraction results from remote sensing images

4.4. Lightweight and timeliness analysis

Under the premise of ensuring the accuracy of road extraction, the algorithm in this paper focuses on realizing the time performance of the network. A 1024 × 1024 × 3 image was randomly generated in the experiment as a test sample, and the time-effectiveness of the method in this paper was proved by comparing the model calculation amount, parameter amount, and aerial image prediction time. The specific indicators are shown in Table 2. The computing power required for the forward propagation of the algorithm in this paper accounts for 1.069% of the computing power required by the ASPP - U - Net network model, and 11.176 % of the computing power required by the D - Link network model; the model size only accounts for the ASPP-U-Net network model. -Net network model size is 0.6%, accounting for 1.34% of D - Link network model size; model parameters account for 7.436 % of ASPP-U-Net network model parameters, accounting for 10.382 of D-Link network model parameters %.

Table 2: Comparison of timeliness of road extraction results by different methods

Name	Type	Model size (#params)	Model size(MB)	GFLOPs (forward pass)
ASPP-U-Net	CNN	75044865.0	300.2MB	1428.8840
D-Link	CNN	33613121.0	134.0MB	136.6727

Ours	CNN	348975.0	1.8MB	15.2817

In order to fully demonstrate the timeliness performance of this algorithm, 325 1024×1024 remote sensing images are selected for testing. The inference time of the ASPP-U-Net model is 119.01 s, the D - Link model is 94.61 s, and the algorithm of this paper is 69.28 s.

In summary, the algorithm in this paper not only ensures the accuracy of road extraction, but also significantly reduces the amount of network parameters, reduces the complexity of the model, and realizes the lightweight of the network. Therefore, it has good engineering application value.

5. Conclusion

Pixel-level semantic segmentation of target tasks using deep learning methods is widely used in various scenarios. It is aimed at the practical problems of deep network model layers and many parameters. In this paper, a large-encoder-small-decoder real-time segmentation network with diverse convolutional modules is designed. Experiments show that the algorithm in this paper significantly reduces the amount of network parameters while ensuring the accuracy of road extraction. It can achieve accurate and real-time extraction of road targets in UAV images, so it has good engineering application value.

Deep learning is a data-driven algorithm. Usually, the size of the data determines the generalization ability of the model, but it is difficult to obtain massive data, and training a network model often takes a lot of time. How to use road features the combination of prior knowledge and the method in this paper to further improve the accuracy of road extraction is the next key research direction.

References

[1] C HENG JH, GAO G, KU XS, et al. Review of road network extraction from SAR images[J]. Journal of Image and Graphics, 2013(01):11-23.

[2] Cheng G, Zhu F, Xiang S, et al. Road Centerline Extraction via Semisupervised Segmentation and Multidirection Nonmaximum Suppression[J]. IEEE Geoscience & Remote Sensing Letters, 2017, 13(4):545-549.

[3] Gongjian W, Runsheng W, et al. Automatic Extraction of Main Roads from Aerial Remote Sensing Images[J]. Journal of Software, 2000,11(7):957-964.

[4] ZHANG YH, HE J, KAN X, et al. Summary of road extraction methods for remote sensing images[J]. Computer Engineering and Applications, 2018, 54(13): 1-10.

[5] CAO YG, WAGN ZP, YANG L. Advances in method on road extraction from high resolution remote sensing images[J]. Remote Sensing Technology and Application, 2017, 32 (1): 20-26.

[6] Sheng Q, Zhu F, Chen S, et al. Automatic Road Extraction from Remote Sensing Images Based on Fuzzy Connectedness[C]// Fifth International Conference on Geo-information Technologies for Natural Disaster Management. IEEE, 2014.

[7] Shi W, Zhu C. The line segment match method for extracting road network from high-resolution satellite images[J]. IEEE Transactions on Geoscience & Remote Sensing, 2002, 40(2):511-514.

[8] Christophe E, Inglada J. Robust Road Extraction for High Resolution Satellite Images[C]// IEEE International Conference on Image Processing. IEEE, 2007.

[9] Zhang R, Zhang JX, Hai-Tao LI. Semi-automatic Extraction of Ribbon Roads from High Resolution Remotely Sensed Imagery Based on Angular Texture Signature and Profile Match[J]. Journal of Remote Sensing, 2008.

[10] ZUO J, LI Y J. Road extraction from high resolution remote sensing image based on texture and shape feature[J]. Surveying and Mapping, 2013(03):111-113.
[11] LUO QZ, YIN Q, KUANG D B. Research on extracting road based on its spectral feature and shape feature[J]. 2007(3).
[12] LIN Xiangguo, ZHANG Jixian, LI Haitao, YANG Jinghui. Semi-automatic Extraction of Ribbon Road from High Resolution Remotely Sensed Imagery by a T-shaped Template Matching. GEOMATICS AND INFORMATION SCIENCE OF WUHAN UNIVERS, 2009, 34(3):293- 296.
[13] Jiuxiang, Hu, Razdan, et al. Road Network Extraction and Intersection Detection From Aerial Images by Tracking Road Footprints. [J]. IEEE Transactions on Geoscience & Remote Sensing, 2007.
[14] Gaetano R, Zerubia J, Scarpa G, et al. Morphological road segmentation in urban areas from high resolution satellite images[C]// International Conference on Digital Signal Processing. IEEE, 2011.
[15] Zhang J, Lin X, Liu Z, et al. Semi-automatic road tracking by template matching and distance transformation in urban areas[J]. International journal of remote sensing, 2011, 32(23): 8331-8347.
[16] Saito S, Yamashita T, Aoki Y. Multiple object extraction from aerial imagery with convolutional neural networks[J]. Electronic Imaging, 2016, 2016(10): 1-9.
[17] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.
[18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
[19] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// IEEE Computer Society. IEEE Computer Society, 2013.
[20] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
[21] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep C onvolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017:1-1.
[22] Noh H, Hong S, Han B. Learning Deconvolution Network for Semantic Segmentation[J]. IEEE, 2016.
[23] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[J]. IEEE Computer Society, 2014.
[24] H EH, WANG SC, YANG DF, et al. An road extraction method for remote sensing image based on Encoder-Decoder network[J]. Acta Geodaetica et Cartographica Sinica, 2019, 048(003):330-338.
[25] Paszke A, Chaurasia A, Kim S, et al. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation[J]. 2016.
[26] Wegner J, Montoyazegarra J, Schindler K. A Higher-Order CRF Model for Road Network Extraction[C]// IEEE. IEEE, 2013.
[27] Zhang Z, Liu Q, Wang Y. Road Extraction by Deep Residual U-Net[J]. IEEE Geoscience and Remote Sensing Letters, 2017, PP(99):1-5.
[28] Zhou L, Zhang C, Ming W. D- LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018.