ROAD EXTRACTION FROM SATELLITE IMAGE VIA AUXILIARY ROAD LOCATION PREDICTION

Jingtao Hu, Qi Wang, Xuelong Li*

School of Computer Science and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P.R. China.

ABSTRACT

Road extraction from satellite images is usually interrupted with several disconnected segments so that it does not satisfy the real application. The segmentation-based methods fail to correct separated roads due to the incompleteness information. Therefore, this paper introduces auxiliary Road Location Prediction(RLP), a task leveraging global context information to help road segmentation infer each road segment. The auxiliary task has two branches: horizontal location prediction and vertical location prediction which can predict locations of all the roads. By combining road segmentation and RLP, road extraction performance is effectively improved. As a result, the additional training signals help the primary road segmentation task to aggregate surrounding scene information to reason about its connectivity. The experiments on two public datasets have demonstrated the effectiveness of the proposed method.

***Index Terms*—** Road extraction, road location prediction, global context feature, auxiliary task

1. INTRODUCTION

Road extraction from satellite images has attracted much attention and extensive research in remote sensing. It is used in many fields such as emergency rescue, autonomous driving, city planning, etc[1]. However, it is difficult for the accurate results because of the complex road scene which includes shadows and occlusion caused by trees, vehicles and buildings as shown in Fig. 1, and any other road-like scene.

The problem is regard as a binary segmentation task by post-processing to extract the road map. In order to infer the disconnected segments, it is essential to get a large receptive field. D-linkNet [2] introduced dilated convolution to enlarge the receptive field of features, which obtained a better road segmentation. Combining other information can also improve segmentation performance. [3] improved road connectivity by joint learning of orientation and segmentation. Apart from the segmentation methods, RoadTracer [4] extracted the road, which iteratively generated a graph based on image features and predicted graph. Further, Wei [5] established a multiple starting points tracer that made a great improvement of RoadTracer. However, these methods do not consider the roads diversity and individual road segment information around.

Fig. 1. An example shows the disconnect road due to the shadows and occlusion from trees and buildings.

In this paper, we focus on segmentation-based methods. Although the pix-wise classification can get the contour or edge of a road, the road connectivity is not able to learn from the segmentation. Considering the road connectedness, we propose an auxiliary task to recover the connectivity destroyed by the shadows and occlusion. The road location prediciton task is combined with the road segmentation to connect the broken road. The global context features are used to infer the road disconnect, which get the whole receptive field. Nonetheless, inferring all roads connectivity simultaneously will confuse local connectivity because of the different road categories. Therefore, we define a new road formulation parsing all the road that it can distinguish each road segment. The contributions of the proposed methods are the following:

(1) We propose a multi-task architecture combining road segmentation and road location prediction, which has not been explored together to the best of our knowledge.

(2) We introduce an effective road formulation for road location prediction, which improves the performance of the model in dealing with the shadows and occlusion.

2. METHOD

In this section, we propose a multi-task road extraction method. The details include new road formulation and auxiliary road location prediction, which are described in subsection 2.1 and 2.2 respectively.

*Q. Wang is the corresponding author. This work was supported by the National Key R&D Program of China under Grant 2018YFB1107403, National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, and 61871470.

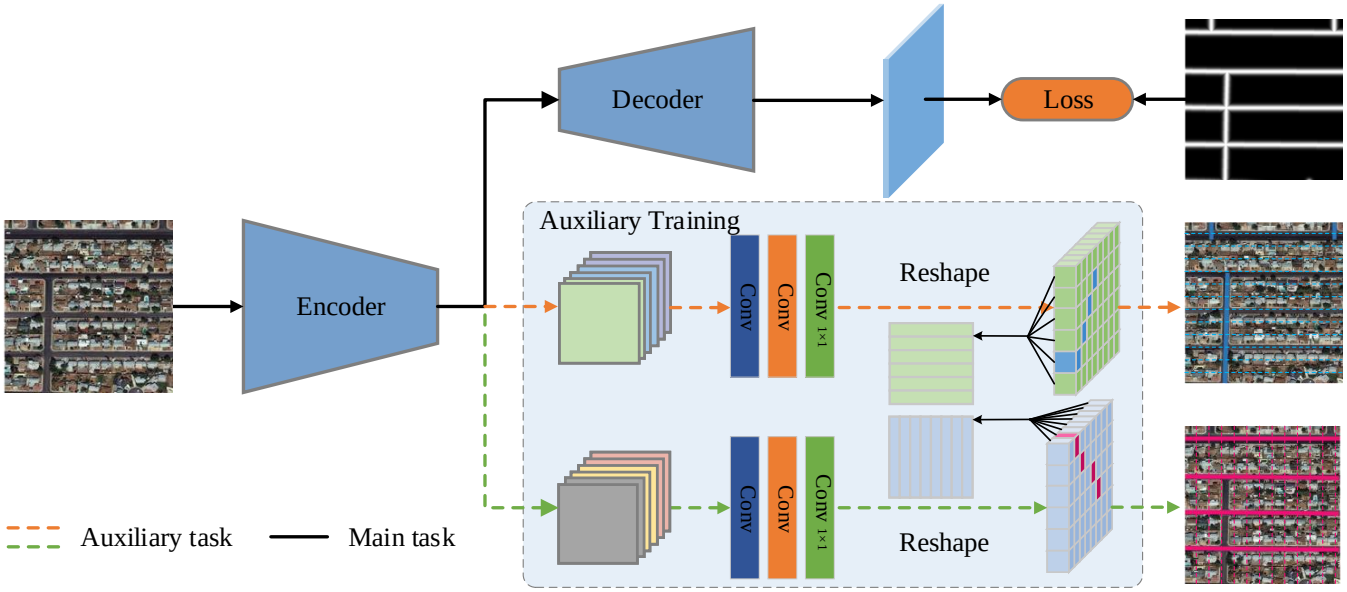
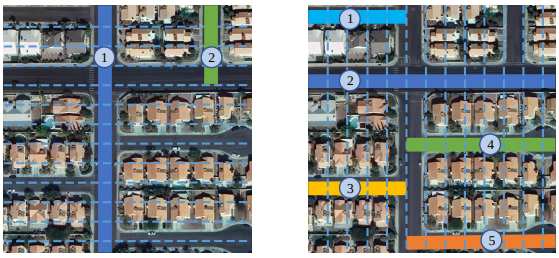


Fig. 2. Illustration of the proposed multi-task framework for road extraction.

2.1. Road Formulation

As mentioned in the introduction section, road extraction performance is greatly affected by shadows and occlusion, resulting in many disconnected segments. And it is challenging to solve the above problems by using the method based on segmentation. To address the issues, we define a new formulation for road extraction to locate each road segment.



(a) row anchor (b) column anchor

Fig. 3. Row anchors and column anchors for RLP.

Inspired by [6], we formulate the road extraction to road location prediction based on several line anchors to learn the global context information. The road location is to assist road segmentation, which mainly aims to complete the lost features around the road. As the road has a complex topological structure, we divide line anchors into row anchors and column anchors to cover all road segments, which is used to solve the prediction confusion caused by parallel road segments and line anchors shown in Fig. 3.

Firstly, we need to classify road segments to determine the uniqueness of road segments in each line anchor because road segments will confuse every prediction of line anchors, as shown in Fig. 3. The roads can be easily classified by the direction. However, there may be more than one road segment with the same orientation in a dense road scene. Therefore, we need to sort the roads to make each prediction unique. Suppose that the set of road segments belonging to the same direction is $S_\theta = \{s_\theta^1, s_\theta^2, \ldots, s_\theta^n\}$ in which θ is the road orientation, n is the total number of the roads with θ. We can calculate road types number $N_{rt} = N \times n$ from the road direction and its roads amount, in which N is the number of road orientations. The road segments with different orientations can be extracted from the road ground truth, and do not require any extra annotation effort. According to the above definition, we can get the unique location index of road segments in row anchor or column anchor. Also, selecting the grid number is essential. The positions of the road segments in the row anchor is in the vertical block interval, the same as the column anchor is in the horizontal block. The griding number of row anchor and column number is equal to W.

2.2. Auxiliary Road Location Prediction

In the complicated road situation, the results of road segmentation are usually noisy due to the shadows and occlusion. We introduce the road location prediction as an auxiliary task, which will enhance the feature representation ability of the intermediate features and therefore optimize the output of road segmentation.

A full convolution decoder network is built for the road location prediction. Suppose the number of row anchors is L as well as the column anchor. According to the different anchors, the auxiliary task includes two branches as shown in Fig. 1. In the row anchor branch, the final goal of the auxiliary network is to output the $\mathbf{Y}_{row}$,

where $\mathbf{Y}_{row} = \{y^1_{row}, y^2_{row}, \ldots, y^W_{row}, y^{W+1}_{row}\}$, in which $y^W_{row} \in \mathbb{R}^{L \times N_{rt}}$, containing the correct position for each road segment in that channel. So does the column anchor branch which outputs $\mathbf{Y}_{col} = \{y^1_{col}, y^2_{col}, \ldots, y^W_{col}, y^{W+1}_{col}\}$ in which $y^W_{col} \in \mathbb{R}^{L \times N_{rt}}$. Assume F is the intermediate feature map (global context features) and g^{row} is the classifier used for determining the road location on the i-th road type, j-the row anchor. Then the location prediction for road segments can be written as:

$$\mathbf{Y}^{i,j}_{row} = g^{i,j}_{row}(F), \text{ s.t. } i \in [1, N_{rt}], j \in [1, L], \quad (1)$$

$$\mathbf{Y}^{i,j}_{col} = g^{i,j}_{col}(F), \text{ s.t. } i \in [1, N_{rt}], j \in [1, L], \quad (2)$$

where $\mathbf{Y}^{i,j}_{row}$ and $\mathbf{Y}^{i,j}_{col}$ represents the probability of selecting $(W+1)$ gridding blocks for the i-th road type, j-th anchor. Suppose $\mathbf{T}_{row}$ and $\mathbf{T}_{col}$ is the truth label of correct road locations. Then, we show the optimization corresponding to:

$$L_{row} = L_{FL}(\mathbf{Y}_{row}, \mathbf{T}_{row}), \quad (3)$$

$$L_{col} = L_{FL}(\mathbf{Y}_{col}, \mathbf{T}_{col}), \quad (4)$$

where L_{FL} is the Focal Loss [7]. From Eq. (1) and Eq. (2), we can see that RLP predicts the probability distribution of all locations on each row anchor or column anchor based on global context features. Consequently, the maximum probabilities represent the correct predicted locations.

2.3. Main Task Prediction

The road segmentation decoder is built on the features F to perform the main task of ground truth G. As the same as [3], we use the soft IoU loss [8] to train the main task:

$$L_{seg} = L_{soft-IoU}(f_{seg}(F), G). \quad (5)$$

The final learning objective function utilizes three losses:

$$Loss = L_{seg} + \lambda_{row} L_{row} + \lambda_{col} L_{col}, \quad (6)$$

where λ_{row} and λ_{col} are the parameters and we empirically set $\lambda_{row} = \lambda_{col} = 4$ in this paper.

3. EXPERIMENTS

3.1. Datasets

To verify the effectiveness of the proposed method, the relative experiments are trained and tested on the two public data sets SpaceNet [9] and DeepGlobe [10]. We follow the datasets splits of [3]. SpaceNet dataset has 2780 high resolution images which have 2213 splits images for training and 567 for testing. We also split the DeepGlobe dataset into 4696 images for training and 1530 for testing.

3.2. Implement Details

Preprocessing: Before training, we should generate the ground truth of road locations. The process is as follows: Firstly, the orientation gound-truth is extracted from the binary mask to get the orientation of each road segment. Then, we can get the vertical coordinates(row anchor) or the horizontal coordinates(column anchor) of road segments in the same orientation. Finally, the road location ground truth $\mathbf{T}_{row} \in \mathbb{R}^{L \times N_{rt}}$ and $\mathbf{T}_{col} \in \mathbb{R}^{L \times N_{rt}}$, we set $L = 27$ as well as $N_{rt} = 36$ that means $N = 9$ and $n = 4$.
Training: The Stochastic Gradient Decent(SGD) optimizer is used to train the network with moment $= 0.0009$, learning rate $= 0.01$ and weight decay $= 0.0005$. Random 256x256 crops of images is used to accelerate the training speed and reduce the GPU memory. To enhance the generalization ability of the model, random horizontal flip and random 0°/90°/180°/360°rotation is used as the augmentation.
Inference: Auxiliary task is not activated in the inference. We evaluate the segment performance by IoU_{road} and F1-score, and the road connectivity by Average Path Length Similarity (APLS) [9] respectively.

Table 1. Results of combining road local prediction(RLP) for road segmentation. It shows that the improvement is due to the auxiliary task.

Method	SpaceNet		DeepGlobe	
	IoU_{road}	APLS	IoU_{road}	APLS
LinkNet34	60.33	55.69	62.75	65.33
LinkNet34+RLP	**62.02**	**56.24**	**63.93**	**67.14**
D-LinkNet34	61.08	55.09	62.82	64.60
D-LinkNet34+RLP	**61.74**	**56.03**	**63.84**	**66.54**

Table 2. Comparisons of road segmentation and road connectivity on SpaceNet dataset.

Method	SpaceNet				
	Precision	Recall	F1-score	IoU_{road}	APLS
LinkNet34	61.30	61.45	61.39	60.33	55.69
D-LinkNet34	62.35	62.44	62.39	61.08	55.09
MatAN	49.84	50.16	50.01	52.86	46.44
DeepRoadMapper	60.61	60.80	60.71	59.99	54.25
Ours	**62.39**	**62.50**	**62.44**	**62.02**	**56.24**

Table 3. Comparisons of road segmentation and road connectivity on DeepGlobe dataset.

Method	Deepglobe				
	Precision	Recall	F1-score	IoU_{road}	APLS
LinkNet34	78.34	78.85	78.59	62.75	65.33
D-LinkNet34	79.03	79.37	79.20	62.82	64.60
MatAN	57.59	56.96	40.13	46.88	47.15
DeepRoadMapper	79.82	80.31	80.07	62.58	65.56
Ours	**80.63**	**80.98**	**80.81**	**63.93**	**67.14**

3.3. Results

Road Location Prediction: We choose two architectures LinkNet34 [11] and D-LinkNet34 [2] to study the performance of road location prediction. In order to verify the effectiveness of the auxiliary task, we assemble the auxiliary task to two architectures. In Table 1, the results show that our proposed auxiliary task for road connectivity is generalized to different architectures. Incorporating the road location prediction as an auxiliary task improves the IoU_{road} for both networks by 1.69% and 0.66% for SpaceNet, respectively. For DeepGoble, it also has 1.18% and 1.02% increase. This suggests that RLP improves the intermediate feature generalization to better road extraction.

Comparisons with Other Methods: We compare the proposed methods with other segmentation-based methods [2] [8] [11] [12]. In Table 2 and Table 3, we can see that our method achieves the competitive quantitative segmentation results. The road connectivity also has a better improvement with more than 0.55 % and 1.58% on two datasets respectively.

Tackle the Shadows and Occlusion: As shown in Fig. 4, we observe that the disconnected regions have been corrected. It means that our method effectively handles the shadows and occlusion problems.

(a) Satellite Images (b) LinkNet34 (c) LinkNet34+RLP (d) Ground Truth

Fig. 4. Visualization of the results on the SpaceNet dataset

4. CONCLUSION

In this paper, we have proposed a road formulation for road location prediction and achieve notable accuracy. The effectiveness of the auxiliary road location prediction task is adequately confirmed with both qualitative and quantitative experiments. With the multi-task architecture, our method could achieve a competitive result compared with the above methods in five metrics. In the future, we will explore the other auxiliary task and improve the way of generating road locations to further refine the output of road segmentation.

5. REFERENCES

[1] C. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 3322–3336, June 2017.

[2] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 182–186.

[3] A. Batra, S. Singh, G. Pang, S. Basu, C.V. Jawahar, and M. Paluri, "Improved road connectivity by joint learning of orientation and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10385–10393.

[4] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt, "Roadtracer: Automatic extraction of road networks from aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4720–4728.

[5] Y. Wei, K. Zhang, and S. Ji, "Road network extraction from satellite images using cnn based segmentation and tracing," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 3923–3926.

[6] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," in *Proceedings of the European Conference on Computer Vision*, 2020.

[7] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.

[8] G. Mattyus, W. Luo, and R. Urtasun, "Deeproadmapper: Extracting road topology from aerial images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3438–3446.

[9] A. Van Etten, D. Lindenbaum, and T. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232*, 2018.

[10] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 182–186.

[11] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *Proceedings of the IEEE Visual Communications and Image Processing*, 2017, pp. 1–4.

[12] G. Mattyus and R. Urtasun, "Matching adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8024–8032.