A BOUNDARY AWARE NEURAL NETWORK FOR ROAD EXTRACTION FROM HIGH-RESOLUTION REMOTE SENSING IMAGERY

Haigang Sui [1], Mingting Zhou [1], Mingjun Peng [2, *], Na Xiong [1]

[1] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan, China - haigang_sui@263.net, mintyzhou@whu.edu.cn, naxiong@qq.com
[2] Wuhan Natural Resources and Planning Information Center, Wuhan, China-751119284@qq.com

Commission III, WG III/1

KEY WORDS: Road Extraction, Deep Learning, Semantic Segmentation, Coarse to Fine Learning, Boundary-aware, Boundary Quality

ABSTRACT:

Automatic road extraction from high-resolution remote sensing imagery has various applications like urban planning and automatic navigation. Existing methods for automatic road extraction however, focus on regional accuracy but not on the boundary quality. To address this problem, a Boundary-aware Road extraction Network (BARoadNet) is proposed. BARoadNet is a coarse-to-fine architecture composed of two encoder-to-decoder networks, a Coarse Map Predicting Module (CMPM) and Fine Map Predicting Module (FMPM). The CMPM learns to predict coarse road segmentation maps. The FMPM is used to refine the coarse road maps by learning the difference between the coarse road extraction result and the ground truth. Experiments are conducted on the open Massachusetts Road Dataset. Quantitative and qualitative analysis demonstrate that the proposed BARoadNet can improve the quality and accuracy of road extraction results compared with the state-of-the-art methods.

1. INTRODUCTION

Road network information at high quality is essential for many real-world applications. These include efficient traffic management, urban planning and automatic navigation (Wang et al., 2016). Visual interpretation is still the principle way to update road networks, which is expensive and labour-intensive. Governments and map companies invest hundreds of millions of dollars each year to update road networks; however, there are still thousands of errors reported every day, and error reports are usually handled manually (Bastani et al., 2018). Automatic road extraction based on High-resolution Remote Sensing Images (HRSIs) provides a promising way to update a road network, quickly, precisely, and accurately.

Spectral, geometric, textual and topological properties of road are used to distinguish roads from backgrounds in HRSIs. The roads in HRSIs appear as narrow straight lines composed of a series of homogenous connected areas. Various algorithms have been developed to exploit representative road features to detect road networks automatically. Generally, the existing methods can be sorted into traditional methods and deep learning-based methods (Lu et al., 2019). Traditional methods utilize hand-designed features for road extraction. These can be divided into pixel-based and object-oriented methods. Pixel-based methods extract spectral features and texture features at the pixel level. These features are combined with algorithms such as classification (Grinias et al., 2016; J. Zhang et al., 2018), morphological evolution (Bakhtiari et al., 2017; Sghaier and Lepage, 2015; Zang et al., 2016), and active contours (Miao et al., 2015; Yousif and Ban, 2014; Zhou et al., 2016) to perform road extraction. To represent the road appearances, Lv et al. (2017) introduced a multi-feature sparse model and a sparse constraint regularized mean-shift algorithm was adopted to distinguish roads from the backgrounds. Pan et al. (2018) combined entropy features and spectral features with Digital Surface Models, and proposed an adaptive image matching method to track roads. Jing et al. (2018) generated multi-scale spectral, geometric, and texture features of roads, effectively implementing road extraction on the island. Pixel-based methods extract roads with clear boundaries and simple backgrounds at high quality. However, there is "salt-and-pepper" noise in these pixel-based road extraction results and complex post-processing methods are required for refinement.

Compared with the pixel-based method, the advantage of the object-oriented methods is that it can alleviate the impact of image noises on the road extraction results. Z. Zhang et al. (2018) applied the Fractal Network Evolution Method (FNEA) to segment the image into objects and then used the random forest to classify initial road objects. A series of complex post-processing methods were adopted to obtain a complete road network in Zhang's work. To efficiently extract accurate road targets from HRSIs, Chen et al. (2018) presented a two-stage method combining edge information and region characteristics. To integrate both object and edge features to extract urban road information, Yin et al. (2015) developed a globally optimized method, enhancing the stability when applied to large and complex images. The object-based methods have favourable anti-noise properties. However, it is highly dependent on the segmentation results and can easily be mixed with adjacent spectral alike ground objects. Traditional methods can extract roads at high quality within a small range. However, the manually designed road features are often over-specified and incomplete, which cannot represent the characteristics of roads under different complex conditions. Hence, the generalization ability of traditional methods is usually limited.

Unlike traditional methods, road extraction methods based on deep learning show stronger generalization ability across regions and areas. Deep learning-based methods extract multi-scale semantic features efficiently and automatically. Fully convolutional networks (FCNs) (Long et al., 2015) are the most commonly used road extraction architecture, but well-annotated samples are required to train these deep learning models. To ease deep network training, Z. Zhang et al. (2018) proposed Deep Residual UNet, that combines the advantages of UNet with residual blocks. To extract roads with various widths, Gao et al. (2018) proposed a multi-feature pyramid network (MFPN).

These studies focused on the road region extraction using deep learning methods. In order to further improve road extraction accuracy, some researches (Cheng et al., 2017; Liu et al., 2018; Lu et al., 2019; Yang et al., 2019) constructed cascaded neural networks to perform multi-task learning to extract road surfaces, centrelines, and edges simultaneously. These multiple tasks are mutually restrained to achieve the purpose of further improving the accuracy of road extraction, but still have limitations.

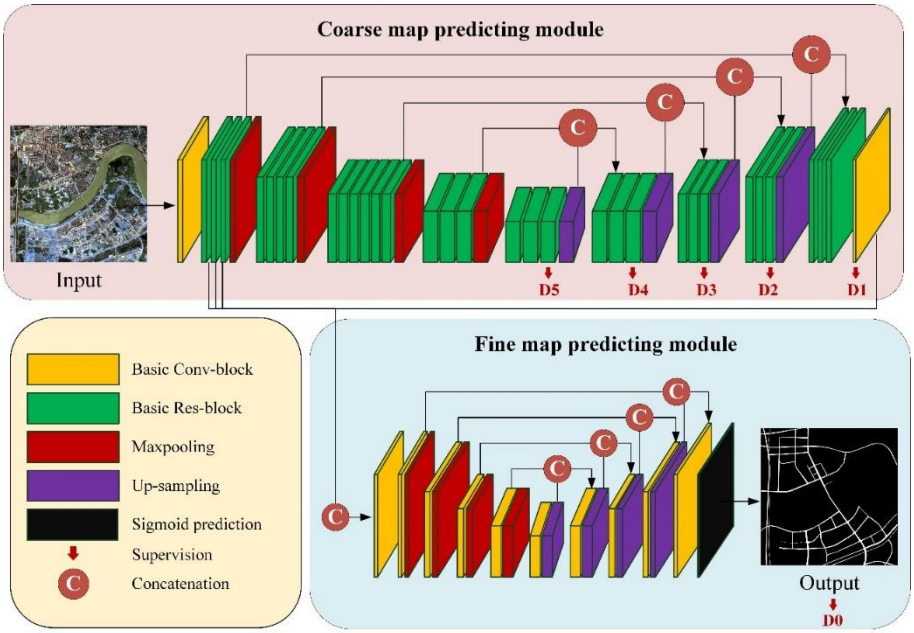


Figure 1 Architecture of our proposed BARoadNet

These methods have greatly improved the accuracy and generalization ability of current road extraction methods; however, since these methods are mostly encoder-decoder structures, the boundary accuracy of the road extraction results will decrease during the down sampling process. In the encoder part, with the network going deeper, the number of feature maps increases and the spatial resolution decreases. Although in the decoder part, the spatial resolution of feature maps is gradually recovered to the same as the input through up sampling, the edge details are lost. Roads are artificial objects with clear boundaries, focusing on the boundary accuracy improves the quality of road network extraction.

In this paper, a boundary-aware road extraction neural network (BARoadNet) is proposed to address the boundary-accuracy problem. BARoadNet was designed as a coarse-to-fine architecture to improve the accuracy of road boundaries. BARoadNet includes a Coarse Map Predicting Module (CMPM) and a Fine Map Predicting Module (FMPM). The CMPM transforms the input optical image into a road probability map, called "coarse road map". The FMPM instead, is focused on the road boundary accuracy; it receives the shallow features of the CMPM and the predicted coarse road map as the input, and learns the residuals between the coarse road map and the ground truth, thus optimizing the boundary of roads. Experiments on the Massachusetts Road Dataset demonstrate that the proposed BARoadNet outperforms the state-of-the-art methods quantitatively and qualitatively.

2. METHODOLOGY

BARoadNet solves the problems of inaccurate road boundaries. The proposed architecture is composed of two modules, a Coarse Map Predicting Module (CMPM) and a Fine Map Predicting Module (FMPM), as shown in Figure 1.

In Figure 1, the red box and the blue box are the architecture of the Coarse Map Predicting Module (CMPM) and the Fine Map Predicting Module (FMPM). The CMPM is a UNet (Ronneberger et al., 2015) like encoder-decoder architecture that learns to predict coarse road segmentation maps. The FMPM learns the difference between the coarse road extraction result and the ground truth, refining the road boundary to match the true boundary on a remote sensing image closely. Taking an optical image as the input, there are six side outputs, i.e. D0, D1, D2, D3, D4, and D5 in Figure 1, produced by our proposed BARoadNet. The six side outputs are deeply supervised by the ground truth to facilitate training. The implementation details of CMPM and FMPM will be described in detail in Sections 2.1 - 2.2. The loss function will be introduced in Section 2.3.

2.1 Coarse Map Predicting Module

The CMPM (red branch in Figure 1) is a UNet like encoder-decoder architecture. The UNet architecture was originally proposed to perform segmentation on biomedical images. UNet captures context information at multiple scales via contracting (encoder) and expanding (decoder) paths. It can be trained with relatively small amounts of data. Such a structure propagates low-level but high-resolution details to the high-level semantic features that can optimize the training process. To facilitate training and forbid the degradation problem, in our CMPM, the encoder path in the original UNet was replaced by an architecture similar to ResNet-34 (He et al., 2016). Our decoder is almost symmetrical to the encoder. The difference is that the repeated times of residual convolutional block are less than those in the corresponding encoder.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-3-2020-67-2020 | © Authors 2020. CC BY 4.0 License.

2.2 Fine Map Predicting Module

The Coarse Map Predicting Module (CMPM) outputs "coarse" road maps; "coarse" refers to extracted results with blurry and noisy boundaries. There is already published research addressing coarse-to-fine learning and refinement of coarse segmented outputs. Peng et al. (2017) proposed Residual Refinement Module based on local context for boundary refinement. Based on this work, Islam et al. (2017) iteratively use the Residual Refinement Module at different scales to optimize the predicted coarse map by learning the residuals between the coarse maps and the ground truth. Wang et al. (2017) applied dilated convolutions with different kernel sizes and dilation rates on coarse maps to capture multi-scale context features and concatenate these features to obtain a refined map. The modules in these applications however, are shallow and thus do not capture high-level information for refinement. Qin et al. (2019) employed a residual encoder-decoder architecture that does capture high level information for refinement.

Our Fine Map Predicting Module (FMPM) is inspired by Qin's work (Qin et al., 2019)--Residual Refinement Module (RRM). However, the original RRM receives only the output coarse map as the input, without feature sharing between the coarse branch and the refine branch. In this way, the error in the coarse map will be passed to the refined map. Although in RRM, the coarse map is optimized by learning the differences between the predicted coarse map and the ground truth, the refinement of the coarse map does not refer to the actual land cover pattern, since there is no participation of the original information provided by the high-resolution remote sensing image. Hence, our FMPM receives the coarse output together with the shallow feature maps of the encoder path of the CMPM as the input, to get a refined road map. The FMPM employs the encoder-to-decoder architecture. The architecture of our FMPM is similar to but simpler than our CMPM. The difference is that the encoder part of FMPM uses plain convolutional blocks for feature extraction, rather than residual convolutional blocks.

2.3 Loss formulation

Our BARoadNet aims to distinguish two classes (i.e., road and background), falling into a binary semantic segmentation problem. Binary Cross Entropy (BCE) is the most commonly used loss function in road extraction tasks, in which the pixel-wise difference is calculated between the predicted result and the true value for backward propagation. BCE is adopted as the basic loss function in our BARoadNet. It is defined as:

$$l_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\overline{y_i}) + (1-y_i)(1 - (\log(\overline{y_i}))) \quad (1)$$

where y_i is the predicted probability of being road of pixel i, while $\overline{y_i}$ is the ground truth label and $\overline{y_i} \in \{0,1\}$.

Road networks are highly structured, and their pixels exhibit strong dependencies on spatial relationships. However, BCE evaluates the difference between the prediction map and the ground truth in the pixel level only, causing blurry road segmentation results. It is necessary to add the structural similarity evaluation measure to the loss function to maintain the road shape and continuity. The Structural Similarity Index Metric (SSIM) (Wang et al., 2004) was originally designed for image quality assessment. It calculates the structural information of an image, and a higher SSIM means cleaner results. Hence, SSIM is integrated to our training loss to learn and extract structural information from the road maps.

Roads are structured both locally and globally. Locally, roads are long shape targets; globally, roads are topologically connected. It is more appropriate to use SSIM loss to evaluate the local structural similarity of roads than the global evaluation since roads are unevenly distributed in an image. The global SSIM may overwhelm the effective local information and may reduce the efficiency of a similarity estimation (He et al., 2019). Thus, the local SSIM loss is calculated by a square sliding window. The window size is set to 11, the same as (He et al., 2019). Given that the predicted map and the ground truth is cropped by the sliding window into $x = \{x_j : j = 1,2,\dots,N\}$ and $y = \{y_j : j = 1,2,\dots,N\}$, the local SSIM loss of x and y is defined as:

$$l_{SSIM} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{({\mu_x}^2 + {\mu_y}^2 + C_1)({\sigma_x}^2 + {\sigma_y}^2 + C_2)} \quad (2)$$

where μ_x, μ_y and σ_x, σ_y are the mean and standard deviations of x and y, and σ_{xy} is their covariance. $C_1 = 0.01^2$ and $C_2 = 0.03^2$ are used to avoid dividing by zero. The overall SSIM loss of the completely predicted map is the mean of local SSIM of all the cropped patches.

Local SSIM loss evaluates the structural difference between the predicted map and the ground truth locally. Further, to evaluate the performance of our model globally, Intersection over Union (IoU) is introduced as another loss function in our model. IoU used to be an accuracy evaluation metric for measuring the similarity of the predicted map and the ground truth, but it has been also used as a loss function to evaluate the difference between a predicted map and the ground truth globally. l_{IoU} is defines as:

$$l_{IoU} = 1 - \frac{\sum_{i=1}^{N} y_i\overline{y_i}}{\sum_{i=1}^{N}(y_i + \overline{y_i} - y_i\overline{y_i})} \quad (3)$$

where y_i is the predicted probability of being roads, $\overline{y_i}$ is the ground truth label of the pixel i.

The BCE loss learns the difference between the predicted map and the ground truth in the pixel level, and the SSIM loss evaluates the difference locally while the IoU loss measures the performance of the training model globally. To obtain high quality road segmentation maps with clear boundaries, these three losses are integrated to train the network, and the overall loss is calculated as:

$$l = l_{BCE} + l_{SSIM} + l_{IoU} \quad (4)$$

Since the BARoadNet is deeply supervised with six side outputs, each loss function is a weighted summation of all the side outputs. The final loss is defined as:

$$L = \sum_{i=1}^{6} l_i \quad (5)$$

where l_i is the overall loss of the ith side output calculated by Equation (4).

3. EXPERIMENTS

Extensive experiments and analysis are presented in this section. There are two subsections. The first subsection demonstrates experimental setups including the test datasets, the training

details, the evaluation metrics, and the algorithms for comparative evaluation. The second subsection presents the experimental results along with a qualitative and quantitative analysis of the tested methods on the Massachusetts Road Dataset.

3.1 Experimental setups

The public Massachusetts Road Dataset is used to evaluate the performance of the proposed BARoadNet. The raw Massachusetts Road Dataset was seamlessly cropped into image tiles with the size of 500×500 pixels and there are 9972 training samples, 126 validating samples and 441 testing samples after image cropping. All the experiments were conducted on a server with one NVIDIA Tesla K80 GPU accelerator, with 12 GB GPU memory. Limited by the size of the GPU memory, if we input the input image directly into the network at the size of 500×500 pixels, the GPU will be full of memory when the batch size is set to two. However, when the batch size is only two, the training process for the model was not very stable. Therefore, in order to increase the number of images that the model can receive per iteration, the input image was resized to 256×256 pixels and fed into the network when training, thus the batch size was increased to 10. We empirically find that the network delivers higher performance, if we resize the input images to 256×256 pixels rather than random cropping. We trained the network until the validation loss converged. For the Massachusetts Road Dataset, it takes about three days to converge, occurring after 4k iterations.

We applied four standard metrics widely used for evaluating road detection performance, i.e. precision, recall, quality and F1-score. Precision measures the percentage of correctly classified road pixels among all predicted road pixels while recall measures the percentage of correctly classified road pixels among all actual road pixels. Quality and F1-score are two comprehensive metrics that combine precision and recall. The reason for using both quality and F1-score evaluation indicators is that, we could directly compare our extraction results on the Massachusetts Road Dataset with the accuracy report of this dataset in published papers, while some articles use quality and the others use F1-score as the evaluation indicator. All of the four evaluation metrics were calculated pixel-to-pixel. Equation (6)-(9) are the definitions for these evaluation metrics.

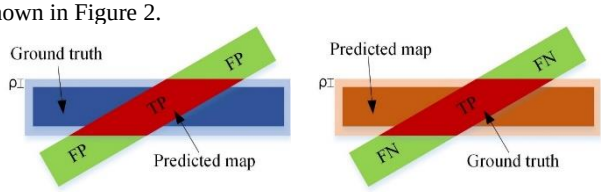
$$Precision = \frac{TP}{TP + FP} \tag{6}$$

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

$$Quality = \frac{Precision * Recall}{Precision + Recall - Precision * Recall} \tag{8}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

where *TP*, *FN* and *FP* are true positive, false negative, and false positive, respectively. True positive is the number of road pixels correctly identified; the false negative is the number of road pixels wrongly identified as non- road pixels; false positive is the number of non-road pixels identified as road pixels. Relaxed evaluation was adopted to assess the accuracy of our method, to compare the results with previous work. Compared with hard evaluation, relaxed evaluation allows an offset of ρ distance between the extraction result and the reference (Lu et al., 2019). The calculation methods of relaxed precision and recall score are shown in Figure 2.



(a) Relaxed precision (b) Relaxed recall

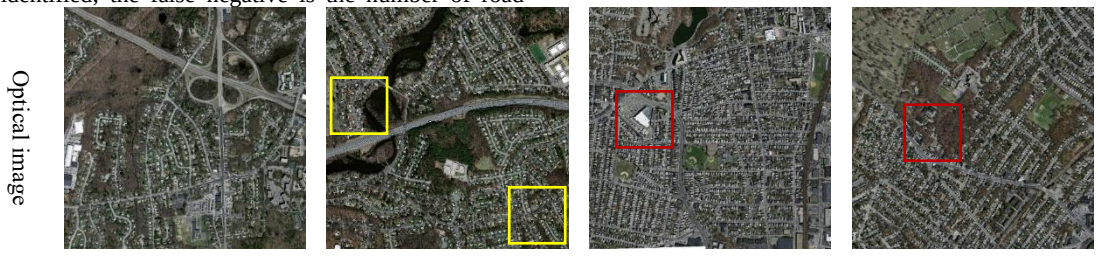
Figure 2 Illustration of the method for relaxed evaluation

In Figure 2, subgraphs (a), (b), and (c) show the detailed calculation of relaxed precision, relaxed recall, and relaxed quality, respectively. As shown in Figure 2- (a), the relax precision is the ratio of the predicted map, which lies within the buffered ground truth. Figure 2- (b) indicates that the relax recall is the ratio of the ground map, which lies within the buffered predicted map. Figure 2- (c) demonstrates that relax quality is the ratio of the overlap area to the union area of the buffered predicted map and the buffered ground truth.

Since the Massachusetts Road Dataset is a public dataset, many algorithms have been tested on it and accuracy reports have been given. Therefore, we compared the accuracy of our method with methods tested on the Massachusetts Road Dataset published in the recent two years, to demonstrate the accuracy of our method. The comparative methods on the Massachusetts Road Dataset were ASPP-UNet-SSIM (He et al., 2019), RDRCNN (Gao et al., 2019), JointNet (Zhang and Wang, 2019), GL-DenseUNet (Xin et al., 2019), and the Improved GAN (Zhang et al., 2019).

3.2 Experiments on Massachusetts Road Dataset

The experiment was conducted on the Massachusetts Road Dataset, published by Mnih (Mnih, 2013). Much work on road network extraction has been done with this dataset and published in open access journals. We directly cite the accuracy reports in their published works for quantitative comparison. We have not attempted to implement these complex algorithms for further visual comparison, as important details of their methods are missing, so our implementation would produce unsatisfactory and unfair results. Figure 2 shows the visual results of BARoadNet on the Massachusetts Road Dataset. There are three rows and four columns. The columns are images and results for four representative samples in the test dataset. The first to the third rows are the optical image, the ground truth image, and the predicted map of the proposed BARoadNet.



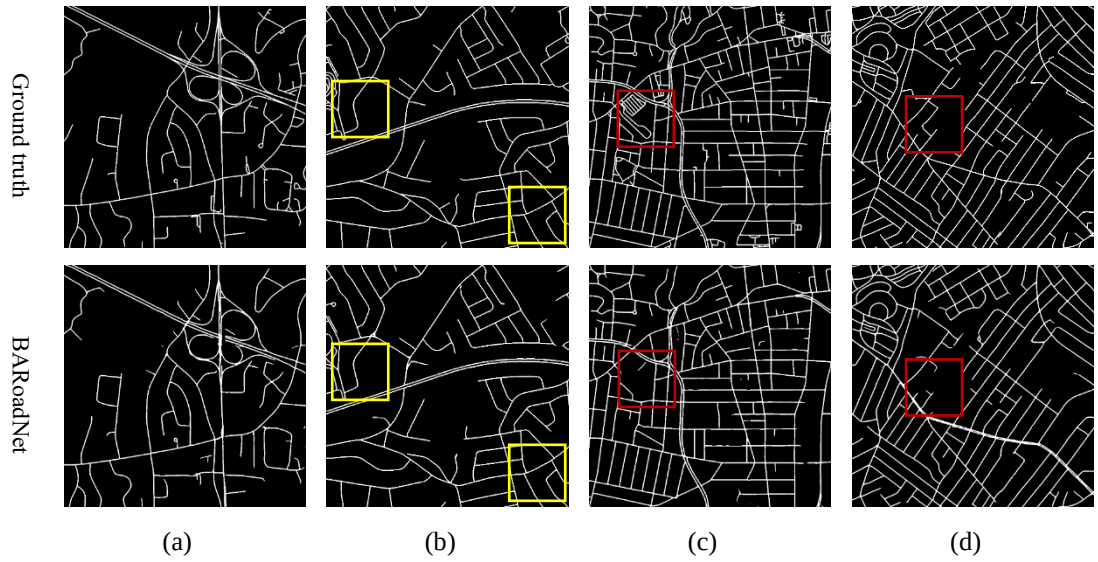


Figure 2 Road extraction results on the Massachusetts Road Dataset

As can be seen in Figure 2, the biggest difficulty in the Massachusetts Road Dataset is that roads on the image are covered by trees. Our algorithm is not affected by occlusions, and can extract continuous, smooth road boundaries, such as the area marked by the yellow box seen in Figure 2(b); but, there are differences between our results and the ground truth in some details, such as the area marked by the red box in Figure 2(c)-(d). Figure 3 shows a detailed view of the area marked in Figure 2.

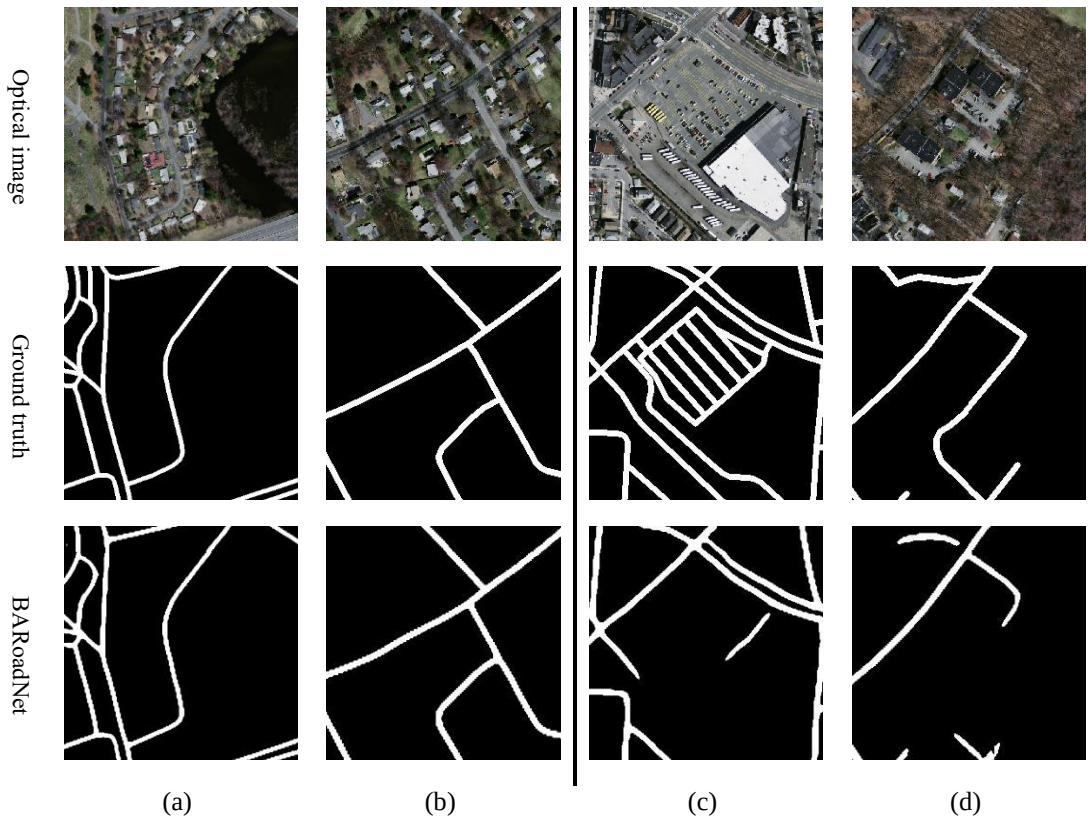


Figure 3 Close ups of red/yellow boxes marked in Figure 2

In Figure 3, subfigure (a) and (b) are detailed extraction results of typical occluded roads, and (c) and (d) are magnifications of areas where the prediction result is different from the ground truth. It can be seen from (a) and (b) that in the occluded area, the results extracted by our algorithm are consistent with the real road boundary on the image, indicating that the BARoadNet proposed in this paper can effectively improve the edge accuracy, and can eliminate the impact of occlusion on road extraction. Subfigures (c) and (d) are two failed examples. In subgraph (c), our method cannot extract detailed roads as the ground truth, because it is in a parking lot. In the ground truth, the aisle left between parking spaces is labelled as road due to semantic thinking. However, the spectral differences between these walkways and parking spaces are minimal, so our method identifies them as background. Subfigure (d) is similar to subfigure (c). In subfigure (d), the roads missed by our method constitute a path connecting two houses and spectrally similar to the surroundings. For roads in cases (c) and (d), it is not an

exception for our method to fail. Without manual intervention, roads with no representative shape features, no topological connection to the existing road network, and no obvious spectral differences from the background, cannot be extracted by automatic algorithms at high quality.

We use relaxed evaluation to assess the accuracy of 441 test data in the Massachusetts Road Dataset. The buffer width ρ was set to three, the same to (Mnih, 2013). The quantitative evaluation results are shown in Table 1. There are six columns and eight rows. The rows are the methods we compared, with the last row showing the quantitative results of our proposed BARoadNet. The columns are values of evaluation metrics including precision, recall, quality and F1- score for each method. The best results are marked in bold and the secondary ones are underlined.

Method	Relaxed evaluation metrics (%), ρ=3			
	Precision	Recall	Quality	F1-score
ASPP-UNet-SSIM	87.10	80.50	-	<u>83.50</u>
RDRCNN	81.82	70.47	-	80.31
JointNet	85.36	71.90	<u>64.00</u>	-
GL-DenseUNet	78.48	70.09	-	73.98
Improved GAN	<u>93.00</u>	<u>82.00</u>	-	-
BARoadNet	**94.50**	**86.80**	**82.74**	**90.29**

Table 1 Quantitative evaluation on the Massachusetts Road Dataset

Table 1 indicates that the precision, recall, quality and F1-score of BARoadNet are higher than those of ASPP-UNet-SSIM (He et al., 2019), RDRCNN (Gao et al., 2019), JointNet (Zhang and Wang, 2019), GL-DenseUNet (Xin et al., 2019), and the Improved GAN (Zhang et al., 2019). As can be seen from the table, the Improved GAN ranks second in precision and recall. Our method has a significant improvement of 4.8% on recall compared to the Improved GAN, indicating that completeness of roads extracted by our method is higher than the Improved GAN. ASPP-UNet-SSIM uses the SSIM as a structural loss function and ranks the second on F1-score. Our method shows improvements of 7% on precision and 6% on recall than ASPP-UNet-SSIM. This indicates that our coarse-to-fine learning strategy delivers more accurate road boundaries, and improves the robustness for occlusion, as compared to using structured loss only.

4. CONCLUSION

In this paper, a boundary-aware road extraction neural network (BARoadNet) is proposed to solve the problems of inaccurate road boundaries. The proposed architecture is composed of two modules, the Coarse Map Predicting Module (CMPM) and the Fine Map Predicting Module (FMPM). The CMPM learns to predict coarse road segmentation maps, while the FMPM learns the difference between the coarse road extraction result and the ground truth, refining the road boundary closer to the true boundary on the remote sensing image. Experiments on the public Massachusetts Road Dataset show that the proposed BARoadNet can improve the quality and accuracy of road extraction results compared with the state-of-the-art methods.

ACKNOWLEDGEMENTS

This research was funded by the National Natural Science Foundation of China under Grant No.41771457, the National Key Research and Development Program of China (No.2016YFB0502603).

REFERENCES

Bakhtiari, H.R.R., Abdollahi, A., Rezaeian, H., 2017. Semi automatic road extraction from digital images. Egypt. J. Remote Sens. Sp. Sci. 20, 117–123. https://doi.org/10.1016/j.ejrs.2017.03.001

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

Chen, L., Zhu, Q., Xie, X., Hu, H., Zeng, H., 2018. Road Extraction from VHR Remote-Sensing Imagery via Object Segmentation Constrained by Gabor Features. ISPRS Int. J. Geo-Information 7, 362.

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Trans. Geosci. Remote Sens. 55, 3322–3337.

Gao, L., Song, W., Dai, J., Chen, Y., 2019. Road Extraction from High-Resolution Remote Sensing Imagery Using Refined Deep Residual Convolutional Neural Network. Remote Sens. 11, 552.

Gao, X., Sun, X., Zhang, Y., Yan, M., Xu, G., Sun, H., Jiao, J., Fu, K., 2018. An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. IEEE Access 6, 39401–39414.

Grinias, I., Panagiotakis, C., Tziritas, G., 2016. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. ISPRS J. Photogramm. Remote Sens. 122, 145–166. https://doi.org/10.1016/j.isprsjprs.2016.10.010

He, H., Yang, D., Wang, S.S., Wang, S.S., Li, Y., 2019. Road Extraction by Using Atrous Spatial Pyramid Pooling Integrated Encoder-Decoder Network and Structural Similarity Loss. Remote Sens. 11, 1015. https://doi.org/10.3390/rs11091015

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Islam, M.A., Kalash, M., Rochan, M., Bruce, N.D.B., Wang, Y., 2017. Salient Object Detection using a Context-Aware Refinement Network., in: BMVC.

Jing, R., Gong, Z., Zhu, W., Guan, H., Zhao, W., 2018. Island road centerline extraction based on a multiscale united feature. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 3940–3953.

Liu, Yahui, Yao, J., Lu, X., Xia, M., Wang, X., Liu, Yuan, 2018. RoadNet: Learning to Comprehensively Analyze Road Networks in Complex Urban Scenes From High-Resolution Remotely Sensed Images. IEEE Trans. Geosci. Remote Sens. 57, 2043–2056.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Lu, X., Zhong, Y., Zheng, Z., Liu, Y., Zhao, J., Ma, A., Yang, J., 2019. Multi-Scale and Multi-Task Deep Learning Framework for Automatic Road Extraction. IEEE Trans. Geosci. Remote Sens. 57, 9362–9377.

Lv, Z., Jia, Y., Zhang, Q., Chen, Y., 2017. An adaptive multifeature sparsity-based model for semiautomatic road extraction from high-resolution satellite images in urban areas. IEEE Geosci. Remote Sens. Lett. 14, 1238–1242.

Miao, Z., Shi, W., Gamba, P., Li, Z., 2015. An object-based method for road network extraction in VHR satellite images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4853–4862.

Mnih, V., 2013. Machine learning for aerial image labeling. Citeseer.

Pan, H., Jia, Y., Lv, Z., 2018. An adaptive multifeature method for semiautomatic road extraction from high-resolution stereo mapping satellite images. IEEE Geosci. Remote Sens. Lett. 16, 201–205.

Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large Kernel Matters--Improve Semantic Segmentation by Global Convolutional Network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4353–4361.

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M., 2019. BASNet: Boundary-Aware Salient Object Detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7479–7489.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241.

Sghaier, M.O., Lepage, R., 2015. Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 1946–1958.

Wang, J., Qin, Q., Gao, Z., Zhao, J., Ye, X., 2016. A new approach to urban road extraction using high-resolution aerial image. ISPRS Int. J. Geo-Information 5, 114. https://doi.org/10.3390/ijgi5070114

Wang, T., Borji, A., Zhang, L., Zhang, P., Lu, H., 2017. A stagewise refinement model for detecting salient objects in images, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 4019–4028.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., others, 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. image Process. 13, 600–612.

Xin, J., Zhang, X., Zhang, Z., Fang, W., 2019. Road Extraction of High-Resolution Remote Sensing Images Derived from DenseUNet. Remote Sens. 11, 2499.

Yang, X., Li, X., Ye, Y., Lau, R.Y.K., Zhang, X., Huang, X., 2019. Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net. IEEE Trans. Geosci. Remote Sens.

Yin, D., Du, S., Wang, S., Guo, Z., 2015. A direction-guided ant colony optimization method for extraction of urban road information from very-high-resolution images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4785–4794.

Yousif, O., Ban, Y., 2014. Improving SAR-based urban change detection by combining MAP-MRF classifier and nonlocal means similarity weights. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 4288–4300.

Zang, Y., Wang, C., Cao, L., Yu, Y., Li, J., 2016. Road network extraction via aperiodic directional structure measurement. IEEE Trans. Geosci. Remote Sens. 54, 3322–3335.

Zhang, J., Wang, Y., Zhao, W., 2018. An improved probabilistic relaxation method for matching multi-scale road networks. Int. J. Digit. Earth 11, 635–655. https://doi.org/10.1080/17538947.2017.1341557

Zhang, X., Han, X., Li, C., Tang, X., Zhou, H., Jiao, L., 2019. Aerial Image Road Extraction Based on an Improved Generative Adversarial Network. Remote Sens. 11, 930.

Zhang, Z., Wang, Y., 2019. JointNet: A Common Neural Network for Road and Building Extraction. Remote Sens. 11, 696.

Zhang, Z., Zhang, X., Sun, Y., Zhang, P., 2018. Road Centerline Extraction from Very-High-Resolution Aerial Image and LiDAR Data Based on Road Connectivity. Remote Sens. 10, 1284.

Zhou, H., Kong, H., Wei, L., Creighton, D., Nahavandi, S., 2016. On detecting road regions in a single UAV image. IEEE Trans. Intell. Transp. Syst. 18, 1713–1722.