Supplementary Materials for: VecRoad: Point-based Iterative Graph Exploration for Road Graphs Extraction

Yong-Qiang Tan Shang-Hua Gao Xuan-Yi Li Ming-Ming Cheng Bo Ren✉
TKLNDST, College of CS, Nankai University
https://mmcheng.net/vecroad/

1. Evaluation Metrics

In all the three road metrics, the pixel metric measures the alignment with ground-truth road pixels, the APLS metric measures the connectivity of the road graph, and the junction metric measures the accuracy of junctions and the connectivity with neighbouring vertices of junctions. We argue that the road alignment (both road centerline and junctions) should be firstly guaranteed, then the connectivity will make sense. In another word, the connectivity without precise alignment is of non-sense.

1.1. Pixel Metric

The most popular evaluation metric in road map extracting area is precision-recall metric [6, 7, 1]. The precision-recall score is defined as

$$\begin{aligned} Precision &= \frac{TP}{TP+FP}, \\ Recall &= \frac{TP}{TP+FN}. \end{aligned} \tag{1}$$

This metric evaluates the pixel-level pairs of predicted and ground-truth road map. The True Positive (TP) denotes that the number of predicted road pixels that are also labeled as a road (True Positives). The $TP+FP$ represents the number of whole predicted road pixels set (True Positives and False Positives), and the $TP+FN$ means the number of the labeled road pixels in ground-truth map. In [6], Mnih *et al*. introduced the relaxed Precision-Recall metric into road extraction, to tolerate the inaccuracy in road pixel-level annotations. The relaxation can be described as a ρ pixel scope for the match of predicted and ground-truth pixels. In particular, the tolerable range ρ is typically assigned with 3 pixels [6, 7], We further calculate the mean F-score to uniformly present the performance. Note that the final goal of the task is to evaluate the performance of the road graph, so we translate the graph representation to road segmentation masks to study the alignment between road graphs and real road. To evaluate the pixel metric on RoadTracer[2] dataset, we transform the graph annotation to road segmentation masks. The road width of both ground-truth and predicted graph is 8 pixels as another tolerance, because the centerline annotation is often subjective and inaccurate.

1.2. Junction Metric

Pixel-level road metric can only describe the alignment between road graph and real roads. To better evaluate road connectivity and topology. Bastani *et al*. [2] proposed a junction-level metric. The metric verifies a predicted junction from the perspective of its coordinate and incident edges. Firstly, same as the relaxation in road segmentation precision-recall metric, Bastani *et al*. [2] use a distance relaxation between the closest pair of a labeled junction v and predicted junction u. If there is a matched pair of (u, v), the $f_v(u)$ is the fraction of incident edges of v that are captured around u, and $f_u(v)$ is the fraction of incident edges of u that appear around v. Otherwise, for each unpaired v, $f_v = 0$, and for each unpaired u, $f_u = 0$. Then, the metric can be defined as a precision-recall format:

$$\begin{aligned} Precision &= \frac{\sum_u f_u}{\sum_u 1}, \\ Recall &= \frac{\sum_v f_v}{\sum_v 1}. \end{aligned} \tag{2}$$

Same as pixel metric, we can take advantage of the mean F-score to represent the performance of junction metric. Although it is vital to find junctions, locating the coordinate of junctions accurately is also necessary. To better evaluate the positioning accuracy of junctions, we set the matching scope radius as 9 meters.

1.3. APLS

The Average Path Length Similarity metric (**APLS**) is introduced from [8]. Having all pairs of corresponding nodes from predicted graph $\hat{G}$ and ground-truth graph G respectively, the APLS metric studies the shortest path length

difference between them:

$$APLS = \frac{1}{N} \sum \left(\frac{2}{\frac{1}{S_{\hat{G} \to G}} + \frac{1}{S_{G \to \hat{G}}}} \right), \quad (3)$$

where

$$S_{\hat{G} \to G} = 1 - \frac{1}{M} \sum \min \left(1, \frac{|L(a,b) - L(\hat{a},\hat{b})|}{L(a,b)} \right) \quad (4)$$

is a shortest path length score mapping from $\hat{G}$ to G. In Equ. (4), M is the number of unique paths in the mapped graph $\hat{G} \to G$. $L(\hat{a},\hat{b})$ and $L(a,b)$ means the length of the path $(\hat{a},\hat{b})$ in $\hat{G}$ and (a,b) in G, respectively. In Equ. (3), N is the number of images belonging to the dataset.

2. Dataset

2.1. RoadTracer Dataset

The road network dataset constructed by [2] covers the urban center of 40 cities across six countries, with 24 km^2 coverage per image. The aerial images are collected from Google, and the ground-truth is from OpenStreetMap (OSM) project [4]. Each aerial image contains 4096×4096 pixels with the precision of 60 centimeters per pixel. Note that, the road ground-truth from OSM is rather pixel-labeling images, but in the form of graphs. They split RoadTracer dataset into 25 cities for training and 15 cities for testing. It is a great challenge to test the versatility of a road extracting algorithm, since the features (e.g., color and building density in city appearance) are of great variety in different cities. It should be noted that, this dataset with large aerial images is closer to real-world road graph generation, where cases like the mismatch of different patches (in Sec. 4.1) should be taken into account.

3. Experiment

3.1. Ablation Study

To achieve a better performance on road alignment and connectivity, we evaluate the difference between the usages of segmentation cues. We experiment on the model with both segmentation cues and flexible step techniques. Firstly, we only apply segmentation supervision without feeding back any cues to the next move predictor. Secondly, compare to the first experiment, the probability map is reused by concatenating the input of the next move predictor. The model with both segmentation supervision and feature map implicit guidance is the adopted method that we report in the full RP-Net.

As shown in Tab. 1, compare to the method without segmentation cues reuse, the adoption of probability map provides lower P-F1 and J-F1 but higher APLS score, and the employment of feature maps have better performance on all the metrics. We argue that the road alignment (both road centerline and junctions) should be firstly guaranteed, then the connectivity will make sense. As a trade-off, we reuse the feature map of segmentation cues for a better alignment.

Segmentation Cues Reuse	P-F1	J-F1	APLS
None	69.34	59.36	55.81
Probability Map	68.84	58.02	60.23
Feature Map	69.81	59.42	57.28

Table 1. Ablation study on how to make full usage of segmentation cues.

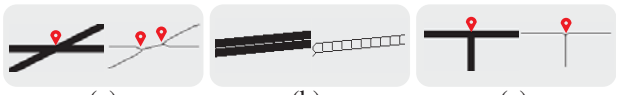
3.2. Runtime

The runtime of our proposed RP-Net is shown in Tab. 2, containing the model which output 4 channels directly (RP-Net-direct) and iteratively (RP-Net). The proposed RP-Net is a real-time model and it is also a trade-off to balance the runtime and prediction performance.

Model	Param	Flops	Runtime	FPS
RP-Net-direct	20.76M	35.38G	10.90ms	91.74
RP-Net	20.76M	63.24G	30.56ms	32.72

Table 2. Ablation study on the runtime given input size of 256×256.

4. Quantitative Analysis



(a) (b) (c)

Figure 1. Skeleton extraction in post-processing changes the geometry and topology of road maps, i.e. (a) one junction splits into two, (b) connected parallel roads become the stepladder-like shape, and (c) shift of junction coordinate.

4.1. Road segmentation to graph annotation

Image Cropping and Splicing. In real-world road graph extraction, the aerial images are often extremely large. To solve the graph extraction task with limited GPU resources, the segmentation-based methods often adopt image patches cropping and splicing [2, 3]. However, the road broken by image boundaries often mismatch due to the different input information (e.g., blue boxes in Fig. 2). Although post-processing techniques may handle some of the cases from purely guessing, a learning-based method is more reliable.

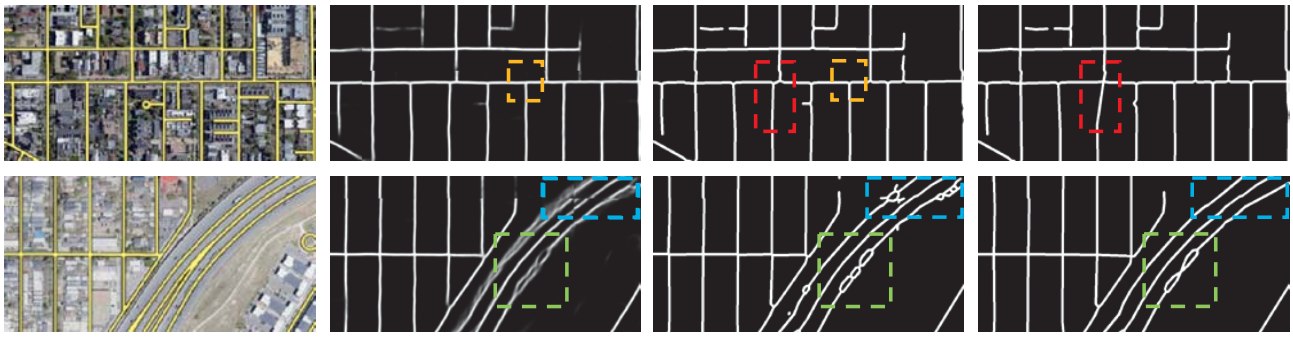


Image & GT | Road Segmentation | Skeleton+RDP | Full Post-processing

Figure 2. Post-processing techniques that adopted by segmentation-based methods to obtain a road graph.

Skeletonization. To obtain a road graph from road segmentation, a skeletonization technique [5] is adopted using morphlogical thinning. As shown in Fig. 1, skeleton extraction in post-processing changes the geometry and topology of road maps. In Fig. 1(a), one junction splits into two, which is against the original representation of road segmentation. In Fig. 1(b), as a common situation in road segmentation, two parallel roads may overlap each other, after morphlogical thinning, connected parallel roads become the stepladder-like shape. In Fig. 1(c), shift of junction coordinate (also shown in orange boxes in Fig. 2) is caused also by the pixel-wise corrosion.

Hard-coded rules. There are many post-processing techniques like nearby junction mergence, small circle detecting and short branch clipping. To solve case shown in Fig. 1(a), a nearby junction mergence is applied but not suit for all situations (e.g., red boxes in Fig. 2). Without learning strategy, the two cases is hard to distinguish. Another technique in post-processing is small circle detecting, which is proposed in [5]. Small circles is often caused by segmentation blur (green boxes in Fig. 2) or small holes (in Fig. 1(b)). However, the introduced noise by segmentation is usually hard to be solved by the artificially defined threshold. The short branch clipping is proposed to remove short road segments, and further remove bad junctions. After all, through step-by-step post-processing, the obtained graph is better than the skeletoned one but still noisy and intermittent.

4.2. Iterative Graph Exploration

The short-sight iterative exploration is a continuous decision-making process, where every step affects the next starting position. So such an exploration framework should be carefully designed, because all the training strategies are teaching the neural network how to trace the road centerline. Global information is ultilized to guide the exploration, otherwise once the tracer runs into non-road area, the exploration will lead to bad alignment or extra road branch. Without post-processing like skeletonization and the consideration of image boundaries, the obtained graph will be of a good connectivity.

5. Visual comparison

We mainly compare the performance with methods that directly generating road graphs. The visulization results of segmentation-based methods are also given as a support. As shown in Fig. 3 4 5 6, segmentation-based methods [5, 2] have a good road alignment but is struggled by blurs and post-processing issues mentioned in Sec. 4.1. The iterative exploration methods [2] tackles the post-processing problem to achieve a better connectivity, but the alignment is not ensured so the connectivity is also limited. As a combination of both road alignment and graph connectivity, our proposed RP-Net have a better visual performance.

References

[1] R. Alshehhi and P. R. Marpu. Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. *ISPRS J. Photogramm. Rem. S.*, 126:245–260, 2017. 1

[2] 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 *CVPR*, 2018. 1, 2, 3, 4, 5, 6, 7

[3] A. Batra, S. Singh, G. Pang, S. Basu, C. Jawahar, and M. Paluri. Improved road connectivity by joint learning of orientation and segmentation. In *CVPR*, pages 10385–10393, 2019. 2, 4, 5, 6, 7

[4] M. Haklay and P. Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4):12–18, 2008. 2

[5] G. Máttyus, W. Luo, and R. Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *ICCV*, 2017. 3, 4, 5, 6, 7

[6] V. Mnih and G. E. Hinton. Learning to detect roads in high-resolution aerial images. In *ECCV*, pages 210–223. Springer, 2010. 1

[7] S. Saito, T. Yamashita, and Y. Aoki. Multiple object extraction from aerial imagery with convolutional neural networks. *Electronic Imaging*, 2016(10):1–9, 2016. 1

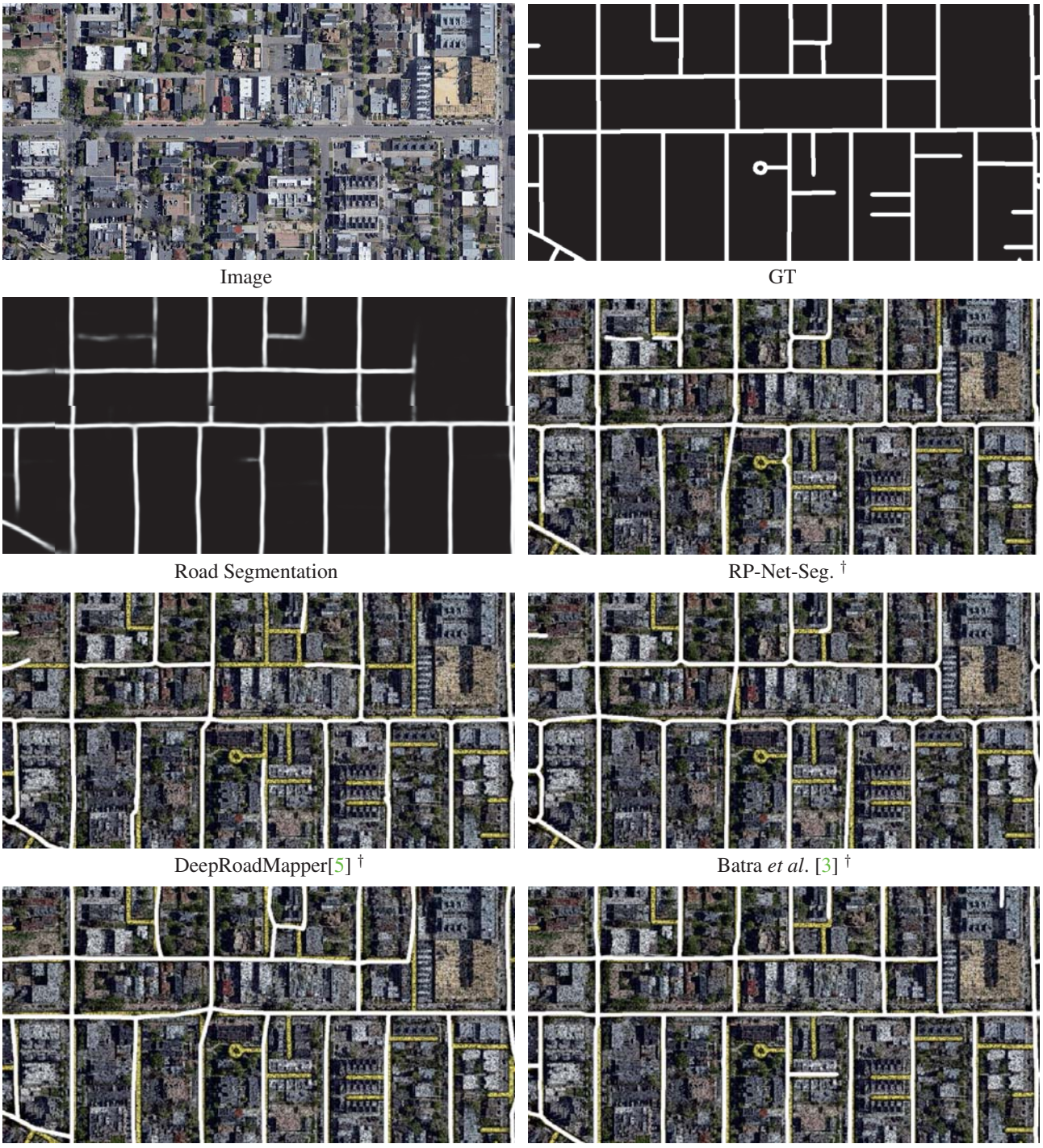


Figure 3. Qualitative comparison of various methods. The '†' denotes that the method needs post-processing.

[8] A. Van Etten, D. Lindenbaum, and T. M. Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018. 1

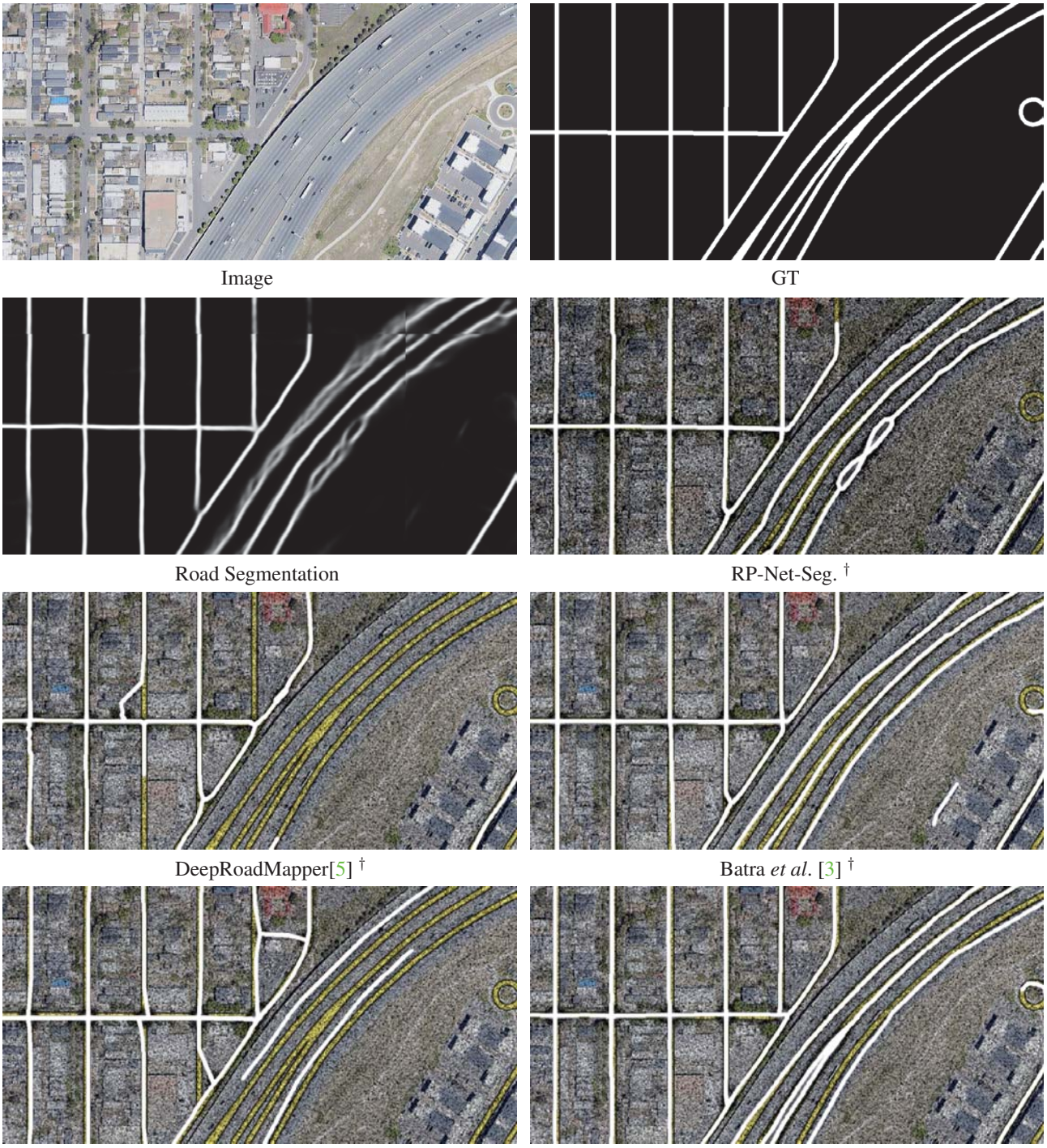


Image GT

Road Segmentation RP-Net-Seg. †

DeepRoadMapper[5] † Batra *et al*. [3] †

RoadTracer[2] RP-Net

Figure 4. Qualitative comparison of various methods. The '†' denotes that the method needs post-processing.

Image

GT

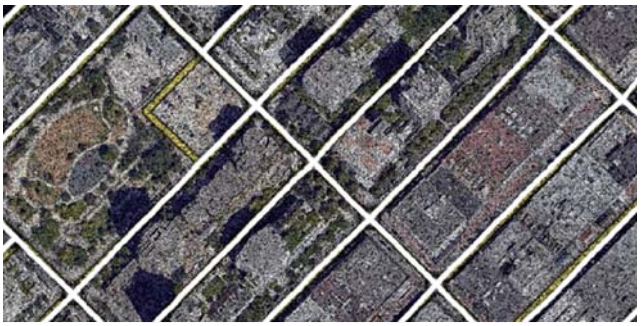
Road Segmentation

RP-Net-Seg. †

DeepRoadMapper[5] †

Batra *et al*. [3] †

RoadTracer[2]

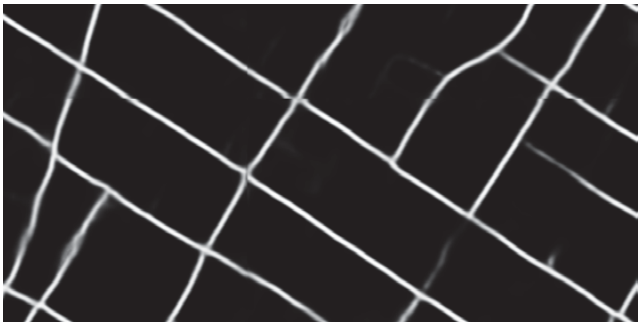


RP-Net

Figure 5. Qualitative comparison of various methods. The '†' denotes that the method needs post-processing.

Image

GT



Road Segmentation

RP-Net-Seg. †

DeepRoadMapper[5] †

Batra *et al*. [3] †

RoadTracer[2]

RP-Net

Figure 6. Qualitative comparison of various methods. The '†' denotes that the method needs post-processing.