

Towards Reliable Evaluation of Algorithms for Road Network Reconstruction from Aerial Images

Supplementary Material

Leonardo Citraro, Mateusz Koziński, and Pascal Fua

Computer Vision Laboratory, École Polytechnique Fédérale de Lausanne (EPFL)
 {firstname.lastname}@epfl.ch

1 Visual examples

We present more examples of road network reconstructions and their scores in Figures tables 1 to 30

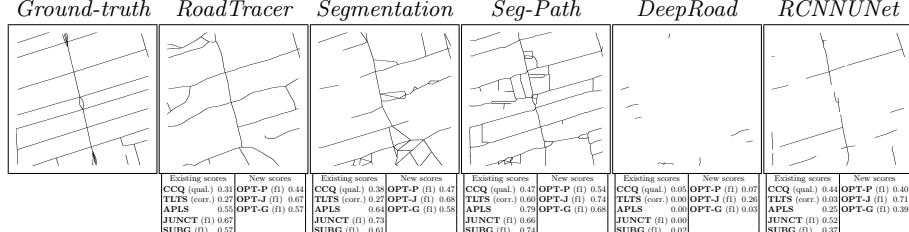


Table 1. Crop of Amsterdam and its reconstructions from aerial images by different methods.

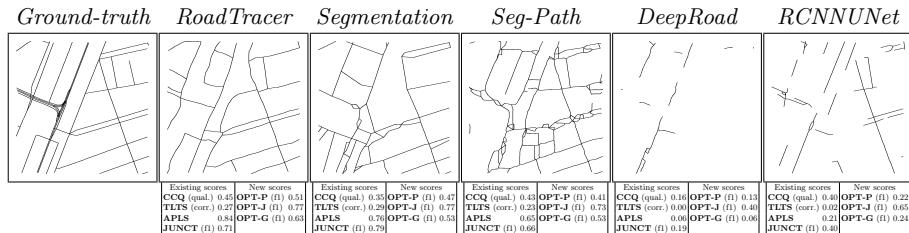


Table 2. Crop of Amsterdam and its reconstructions from aerial images by different methods.

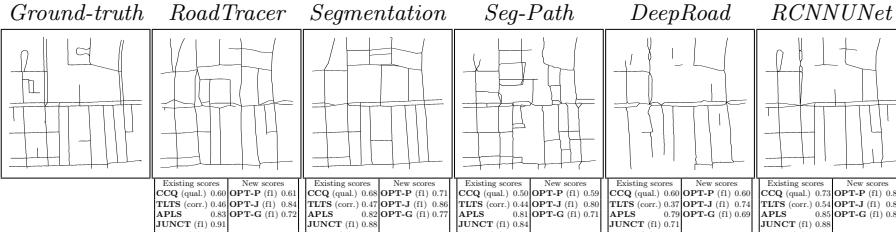


Table 3. Crop of Boston and its reconstructions from aerial images by different methods.

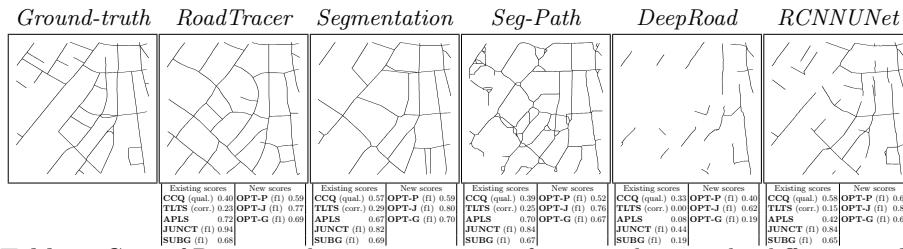


Table 4. Crop of Boston and its reconstructions from aerial images by different methods.

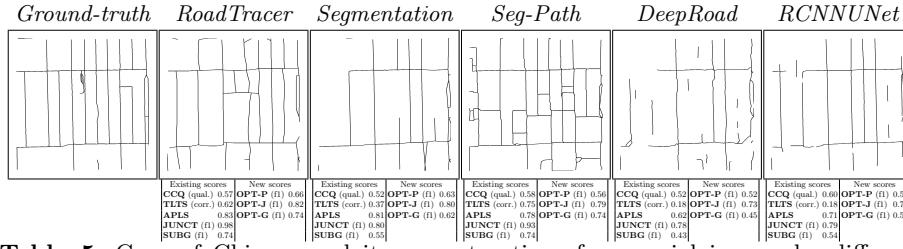


Table 5. Crop of Chicago and its reconstructions from aerial images by different methods.

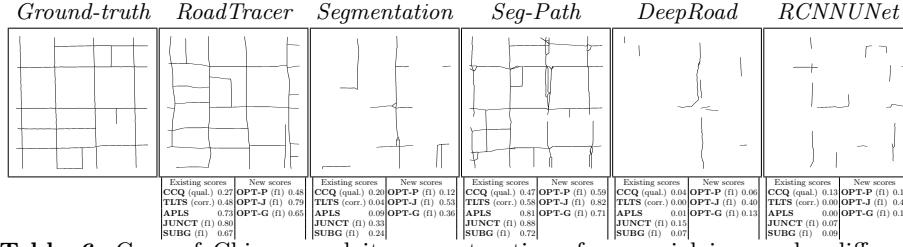


Table 6. Crop of Chicago and its reconstructions from aerial images by different methods.

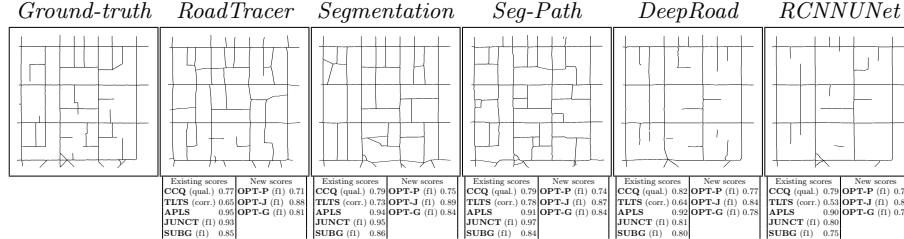


Table 7. Crop of Denver and its reconstructions from aerial images by different methods.

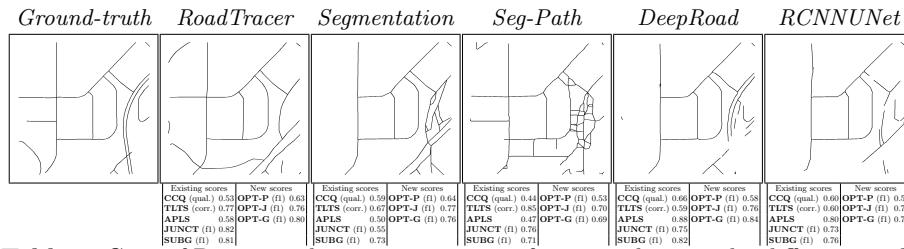


Table 8. Crop of Denver and its reconstructions from aerial images by different methods.

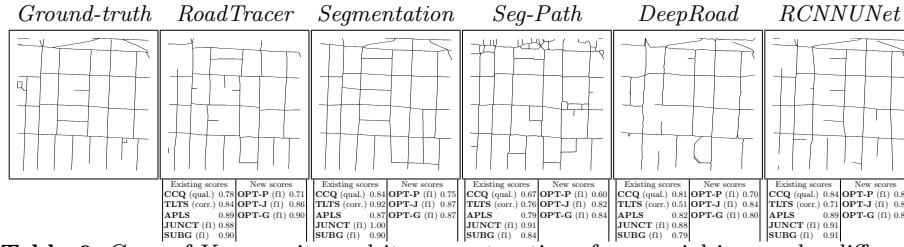


Table 9. Crop of Kansas city and its reconstructions from aerial images by different methods.

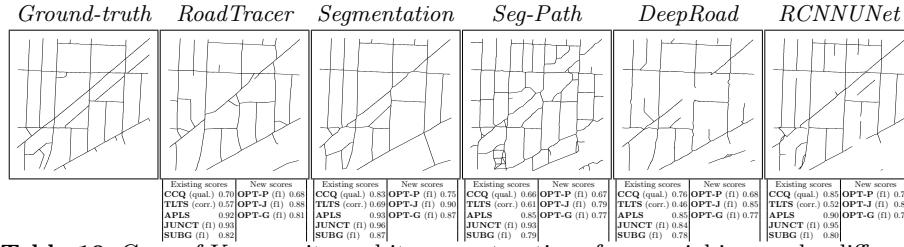


Table 10. Crop of Kansas city and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.51 TIFS (corr.) 0.40 APLS 0.80 JUNCT (fl) 0.81 SUBG (fl) 0.70	New scores OPT-P (fl) 0.65 OPT-J (fl) 0.82 OPT-G (fl) 0.86	Existing scores CCQ (qual.) 0.54 TIFS (corr.) 0.68 APLS 0.86 JUNCT (fl) 0.85 SUBG (fl) 0.85	Existing scores CCQ (qual.) 0.76 TIFS (corr.) 0.88 APLS 0.90 JUNCT (fl) 0.91 SUBG (fl) 0.86	Existing scores CCQ (qual.) 0.62 TIFS (corr.) 0.74 APLS 0.81 JUNCT (fl) 0.90 SUBG (fl) 0.80	Existing scores CCQ (qual.) 0.29 TIFS (corr.) 0.60 APLS 0.81 JUNCT (fl) 0.47 SUBG (fl) 0.41

Table 11. Crop of Los Angeles and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.49 TIFS (corr.) 0.56 APLS 0.72 JUNCT (fl) 0.81 SUBG (fl) 0.72	New scores OPT-P (fl) 0.60 OPT-J (fl) 0.82 OPT-G (fl) 0.72	Existing scores CCQ (qual.) 0.57 TIFS (corr.) 0.65 APLS 0.82 JUNCT (fl) 0.94 SUBG (fl) 0.77	Existing scores CCQ (qual.) 0.41 TIFS (corr.) 0.73 APLS 0.76 JUNCT (fl) 0.92 SUBG (fl) 0.73	Existing scores CCQ (qual.) 0.10 TIFS (corr.) 0.08 APLS 0.14 JUNCT (fl) 0.21 SUBG (fl) 0.13	Existing scores CCQ (qual.) 0.70 TIFS (corr.) 0.65 APLS 0.82 JUNCT (fl) 0.70 SUBG (fl) 0.74

Table 12. Crop of Los Angeles and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.38 TIFS (corr.) 0.45 APLS 0.69 JUNCT (fl) 0.91 SUBG (fl) 0.62	New scores OPT-P (fl) 0.60 OPT-J (fl) 0.75 OPT-G (fl) 0.68	Existing scores CCQ (qual.) 0.43 TIFS (corr.) 0.57 APLS 0.72 JUNCT (fl) 0.85 SUBG (fl) 0.63	Existing scores CCQ (qual.) 0.45 TIFS (corr.) 0.61 APLS 0.73 JUNCT (fl) 0.80 SUBG (fl) 0.74	Existing scores CCQ (qual.) 0.08 TIFS (corr.) 0.19 APLS 0.07 JUNCT (fl) 0.13 SUBG (fl) 0.08	Existing scores CCQ (qual.) 0.54 TIFS (corr.) 0.65 APLS 0.72 JUNCT (fl) 0.82 SUBG (fl) 0.70

Table 13. Crop of Montreal and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.48 TIFS (corr.) 0.50 APLS 0.77 JUNCT (fl) 0.91 SUBG (fl) 0.81	New scores OPT-P (fl) 0.64 OPT-J (fl) 0.84 OPT-G (fl) 0.77	Existing scores CCQ (qual.) 0.42 TIFS (corr.) 0.58 APLS 0.77 JUNCT (fl) 0.80 SUBG (fl) 0.78	Existing scores CCQ (qual.) 0.51 TIFS (corr.) 0.69 APLS 0.85 JUNCT (fl) 0.86 SUBG (fl) 0.79	Existing scores CCQ (qual.) 0.28 TIFS (corr.) 0.44 APLS 0.34 JUNCT (fl) 0.24 SUBG (fl) 0.21	Existing scores CCQ (qual.) 0.62 TIFS (corr.) 0.60 APLS 0.68 JUNCT (fl) 0.88 SUBG (fl) 0.84

Table 14. Crop of Montreal and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.62 TFLS (corr.) 0.82 APLS 0.90 JUNCT (I) 0.88 SUBG (II) 0.85	New scores OPT-P (I) 0.52 OPT-J (I) 0.77 OPT-G (II) 0.84	Existing scores CCQ (qual.) 0.66 TFLS (corr.) 0.51 APLS 0.89 JUNCT (I) 0.83 SUBG (II) 0.77	Existing scores CCQ (qual.) 0.51 TFLS (corr.) 0.61 APLS 0.87 JUNCT (I) 0.68 SUBG (II) 0.61	Existing scores CCQ (qual.) 0.63 TFLS (corr.) 0.84 APLS 0.93 JUNCT (I) 0.77 SUBG (II) 0.82	Existing scores CCQ (qual.) 0.48 TFLS (corr.) 0.10 APLS 0.69 JUNCT (I) 0.38 SUBG (II) 0.36
Existing scores CCQ (qual.) 0.62 TFLS (corr.) 0.82 APLS 0.90 JUNCT (I) 0.88 SUBG (II) 0.85	New scores OPT-P (I) 0.52 OPT-J (I) 0.77 OPT-G (II) 0.84	Existing scores CCQ (qual.) 0.66 TFLS (corr.) 0.51 APLS 0.89 JUNCT (I) 0.83 SUBG (II) 0.77	Existing scores CCQ (qual.) 0.51 TFLS (corr.) 0.61 APLS 0.87 JUNCT (I) 0.68 SUBG (II) 0.61	Existing scores CCQ (qual.) 0.64 TFLS (corr.) 0.78 APLS 0.92 JUNCT (I) 0.77 SUBG (II) 0.82	Existing scores CCQ (qual.) 0.48 TFLS (corr.) 0.57 APLS 0.84 JUNCT (I) 0.77 SUBG (II) 0.94

Table 15. Crop of New York and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.60 TFLS (corr.) 0.82 APLS 0.90 JUNCT (I) 0.90 SUBG (II) 0.90	New scores OPT-P (I) 0.77 OPT-J (I) 0.90 OPT-G (II) 0.88	Existing scores CCQ (qual.) 0.61 TFLS (corr.) 0.50 APLS 0.85 JUNCT (I) 0.57 SUBG (II) 0.72	Existing scores CCQ (qual.) 0.56 TFLS (corr.) 0.73 APLS 0.86 JUNCT (I) 0.73 SUBG (II) 0.73	Existing scores CCQ (qual.) 0.44 TFLS (corr.) 0.60 APLS 0.85 JUNCT (I) 0.79 SUBG (II) 0.79	Existing scores CCQ (qual.) 0.60 TFLS (corr.) 0.79 APLS 0.80 JUNCT (I) 0.90 SUBG (II) 0.90
Existing scores CCQ (qual.) 0.60 TFLS (corr.) 0.82 APLS 0.90 JUNCT (I) 0.90 SUBG (II) 0.90	New scores OPT-P (I) 0.77 OPT-J (I) 0.90 OPT-G (II) 0.88	Existing scores CCQ (qual.) 0.61 TFLS (corr.) 0.50 APLS 0.85 JUNCT (I) 0.57 SUBG (II) 0.72	Existing scores CCQ (qual.) 0.56 TFLS (corr.) 0.73 APLS 0.86 JUNCT (I) 0.73 SUBG (II) 0.73	Existing scores CCQ (qual.) 0.44 TFLS (corr.) 0.60 APLS 0.85 JUNCT (I) 0.79 SUBG (II) 0.79	Existing scores CCQ (qual.) 0.60 TFLS (corr.) 0.80 APLS 0.80 JUNCT (I) 0.75 SUBG (II) 0.74

Table 16. Crop of New York and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.20 TFLS (corr.) 0.79 APLS 0.81 JUNCT (I) 0.95 SUBG (II) 0.77	New scores OPT-P (I) 0.57 OPT-J (I) 0.77 OPT-G (II) 0.75	Existing scores CCQ (qual.) 0.57 TFLS (corr.) 0.61 APLS 0.80 JUNCT (I) 0.82 SUBG (II) 0.84	Existing scores CCQ (qual.) 0.66 TFLS (corr.) 0.61 APLS 0.82 JUNCT (I) 0.86 SUBG (II) 0.56	Existing scores CCQ (qual.) 0.49 TFLS (corr.) 0.60 APLS 0.65 JUNCT (I) 0.69 SUBG (II) 0.56	Existing scores CCQ (qual.) 0.40 TFLS (corr.) 0.60 APLS 0.62 JUNCT (I) 0.64 SUBG (II) 0.49
Existing scores CCQ (qual.) 0.20 TFLS (corr.) 0.79 APLS 0.81 JUNCT (I) 0.95 SUBG (II) 0.77	New scores OPT-P (I) 0.57 OPT-J (I) 0.77 OPT-G (II) 0.75	Existing scores CCQ (qual.) 0.57 TFLS (corr.) 0.61 APLS 0.80 JUNCT (I) 0.82 SUBG (II) 0.84	Existing scores CCQ (qual.) 0.66 TFLS (corr.) 0.61 APLS 0.82 JUNCT (I) 0.86 SUBG (II) 0.56	Existing scores CCQ (qual.) 0.49 TFLS (corr.) 0.60 APLS 0.65 JUNCT (I) 0.69 SUBG (II) 0.56	Existing scores CCQ (qual.) 0.50 TFLS (corr.) 0.62 APLS 0.62 JUNCT (I) 0.80 SUBG (II) 0.75

Table 17. Crop of Paris and its reconstructions from aerial images by different methods.

<i>Ground-truth</i>	<i>RoadTracer</i>	<i>Segmentation</i>	<i>Seg-Path</i>	<i>DeepRoad</i>	<i>RCNNUNet</i>
Existing scores CCQ (qual.) 0.20 TFLS (corr.) 0.79 APLS 0.75 JUNCT (I) 0.82 SUBG (II) 0.76	New scores OPT-P (I) 0.50 OPT-J (I) 0.79 OPT-G (II) 0.73	Existing scores CCQ (qual.) 0.22 TFLS (corr.) 0.59 APLS 0.33 JUNCT (I) 0.57 SUBG (II) 0.36	Existing scores CCQ (qual.) 0.44 TFLS (corr.) 0.60 APLS 0.49 JUNCT (I) 0.69 SUBG (II) 0.61	Existing scores CCQ (qual.) 0.45 TFLS (corr.) 0.60 APLS 0.45 JUNCT (I) 0.69 SUBG (II) 0.56	Existing scores CCQ (qual.) 0.40 TFLS (corr.) 0.60 APLS 0.48 JUNCT (I) 0.66 SUBG (II) 0.60
Existing scores CCQ (qual.) 0.20 TFLS (corr.) 0.79 APLS 0.75 JUNCT (I) 0.82 SUBG (II) 0.76	New scores OPT-P (I) 0.50 OPT-J (I) 0.79 OPT-G (II) 0.73	Existing scores CCQ (qual.) 0.22 TFLS (corr.) 0.59 APLS 0.33 JUNCT (I) 0.57 SUBG (II) 0.36	Existing scores CCQ (qual.) 0.44 TFLS (corr.) 0.60 APLS 0.49 JUNCT (I) 0.69 SUBG (II) 0.61	Existing scores CCQ (qual.) 0.45 TFLS (corr.) 0.60 APLS 0.45 JUNCT (I) 0.69 SUBG (II) 0.56	Existing scores CCQ (qual.) 0.51 TFLS (corr.) 0.70 APLS 0.59 JUNCT (I) 0.51 SUBG (II) 0.56

Table 18. Crop of Paris and its reconstructions from aerial images by different methods.

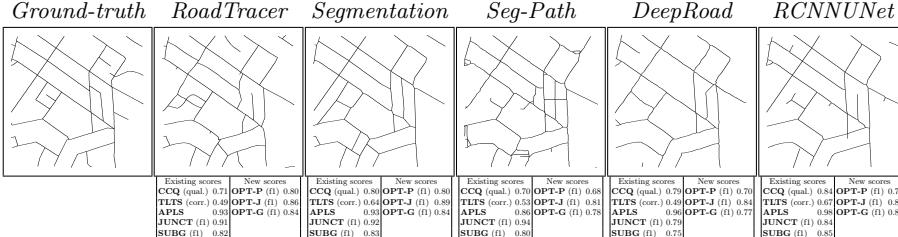


Table 19. Crop of Pittsburgh and its reconstructions from aerial images by different methods.

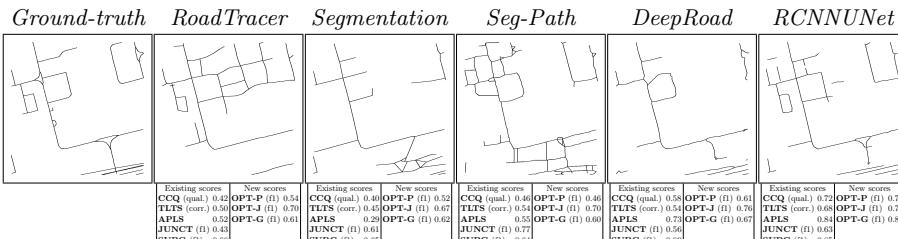


Table 20. Crop of Pittsburgh and its reconstructions from aerial images by different methods.

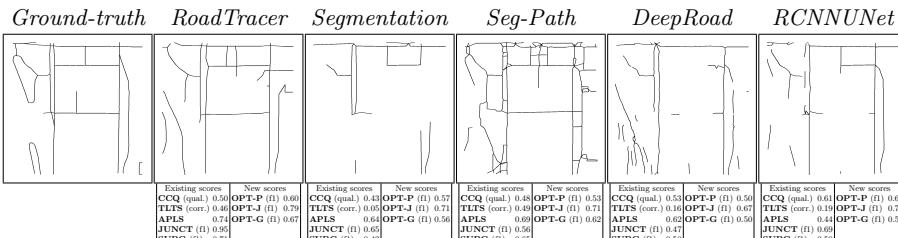


Table 21. Crop of Saltlakecity and its reconstructions from aerial images by different methods.

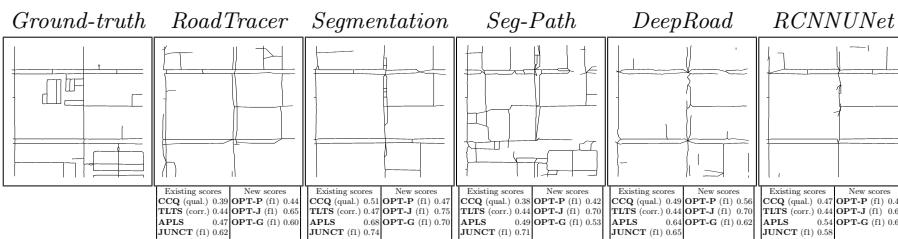


Table 22. Crop of Saltlakecity and its reconstructions from aerial images by different methods.

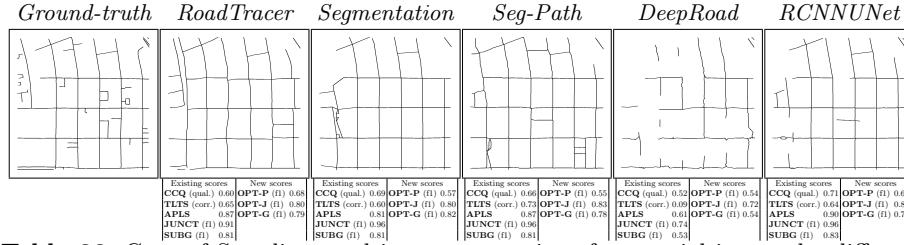


Table 23. Crop of San diego and its reconstructions from aerial images by different methods.

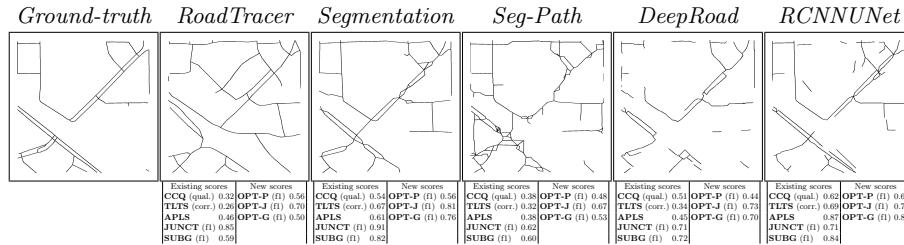


Table 24. Crop of San diego and its reconstructions from aerial images by different methods.

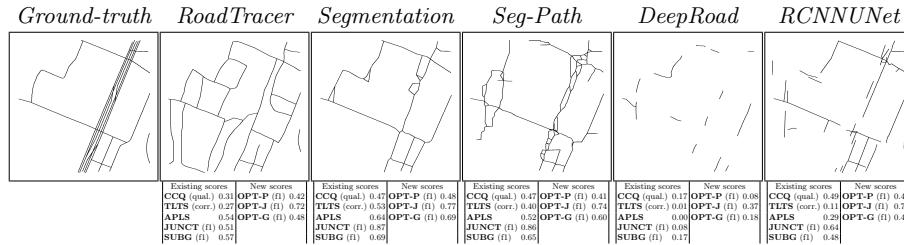


Table 25. Crop of Tokyo and its reconstructions from aerial images by different methods.

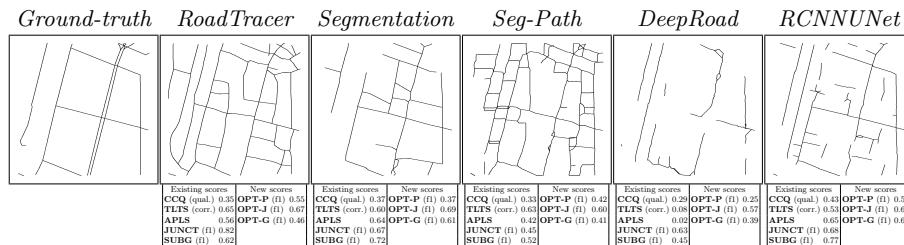


Table 26. Crop of Tokyo and its reconstructions from aerial images by different methods.

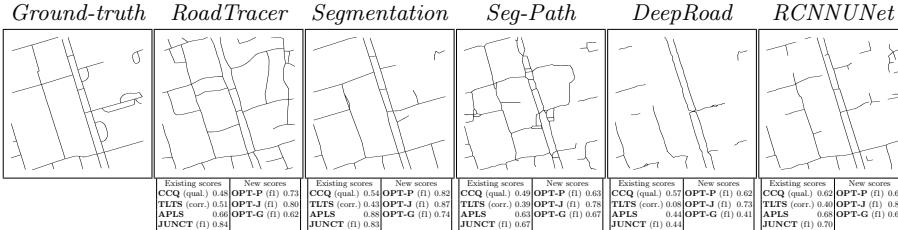


Table 27. Crop of Toronto and its reconstructions from aerial images by different methods.

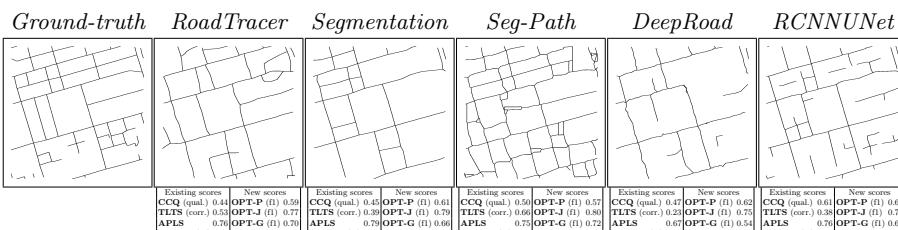


Table 28. Crop of Toronto and its reconstructions from aerial images by different methods.

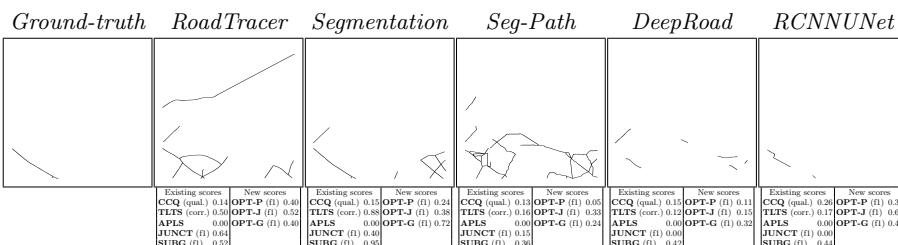


Table 29. Crop of Vancouver and its reconstructions from aerial images by different methods.

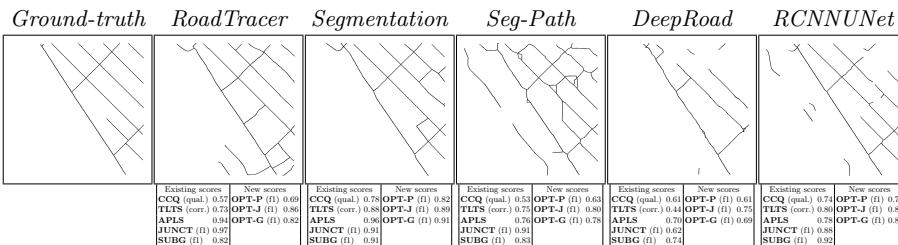


Table 30. Crop of Vancouver and its reconstructions from aerial images by different methods.

2 Pairwise comparison of reconstruction methods

In Figure 8 of the paper we presented example pairwise comparison of two road network reconstruction methods in terms of the existing and the new metrics, broken down across different cities of the RoadTracer test set. In Figure 1, we provide the remaining pairwise comparison of different methods on this dataset, and we present the corresponding plots for the DeepGlobe dataset in Figure 2. DeepGlobe is not divided into cities, so we only present the aggregated scores.

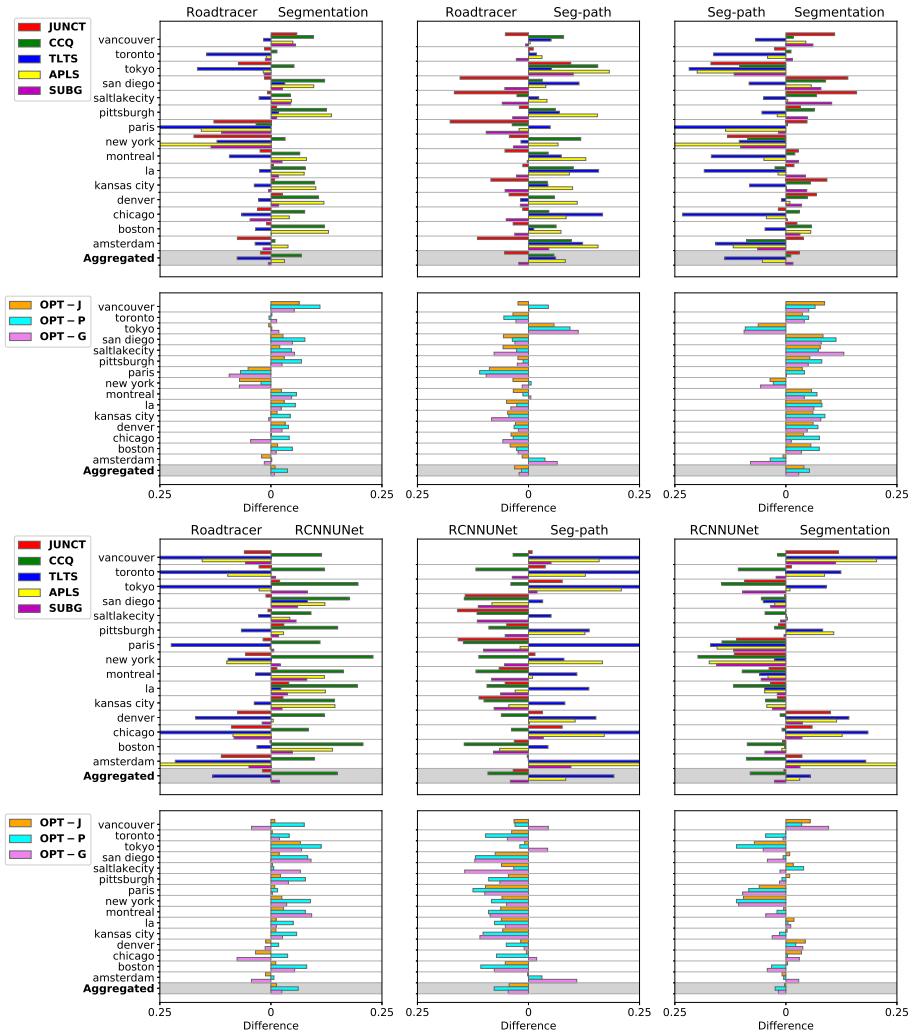


Fig. 1. Differences of scores computed for all cities from the RoadTracer test set. Bars extending to the right express preference for the first method, ones extending to the left indicate the second method scored higher.

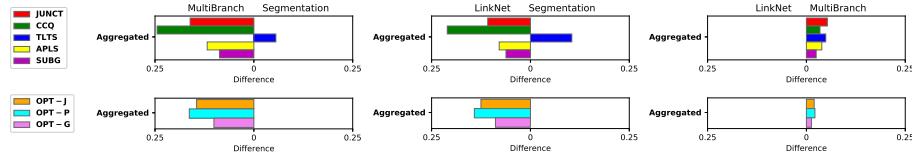


Fig. 2. Differences of scores computed for DeepGlobe test set. Bars extending to the right express preference for the first method, ones extending to the left indicate the second method scored higher.

3 Correlation of the scores on DeepGlobe

In Figure 8 of the main paper we illustrated the correlations of the scores on the RoadTracer dataset. We illustrate the correlations for the DeepGlobe dataset in Figure 3.

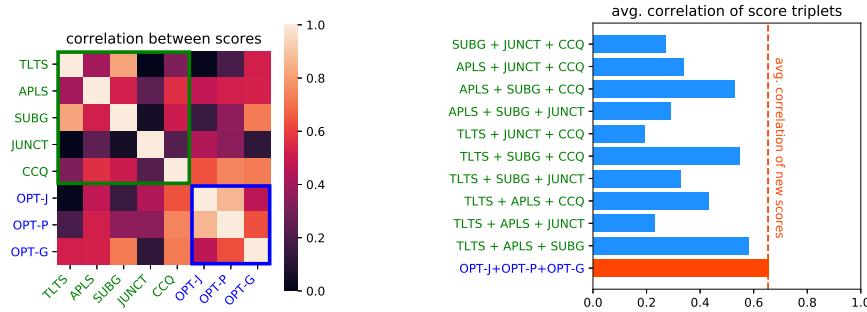


Fig. 3. Correlations between the existing and the new scores on the DeepGlobe dataset. *Left:* A matrix of correlations of the scores computed for the maps reconstructed by different methods on the DeepGlobe dataset. The correlation coefficients of the old scores are outlined in green, the correlation coefficients of the new scores in blue. *Right:* The average correlation of all possible score triplets (blue bars) against the average correlation of the three new scores (dashed red line).

4 Hyper-parameters used in the experiments

In table 31, we present the parameters that we used for computing the scores.

		name	description	value
existing	CCQ	N/A	Positive pixel distance. The score is based on a relaxation of the notion of true positives from perfect coincidence of predicted and ground truth positives to a threshold on the distance between the closest predicted/ground truth positives. For example, for computation of <i>Completeness</i> (precision), a positive pixel in the ground truth is considered a true positive if there is a predicted positive pixel within this threshold distance. For computation of <i>Correctness</i> (recall), a pixel predicted as positive is considered a true positive if there is a ground truth positive pixel within this distance.	8 px
	TLTS	N/A	Maximum distance of corresponding path end points. For a path end point randomly selected in one graph, if there is no point in the other graph within this distance from the end point, the end point is considered absent from the other graph, and the path is considered infeasible. (It is also considered infeasible if the end points exist, but are not connected.)	25 px
	APLS	N/A	Maximum relative length difference for classifying a path as correct. The path is too-long or too-short if the difference is larger.	5 %
	JUNCT	N/A	Maximum distance of corresponding junctions. For a junction selected in one graph, if there does not exist a junction in the other graph within this distance, the junction in the first graph is considered not matched. (We used the original implementation [?] with the default parameters.)	110 px
	SUBG	N/A	Maximum distance between matching control points. If two control points from two different graphs lie within this distance, they are considered matched and are counted as true positives.	25 px
		N/A	Maximum distance between corresponding starting points. For a starting point in one graph, if there is no point in the other graph within this distance, all the control points inserted to the first graph are counted as unmatched.	25 px
		N/A	The maximum distance travelled when exploring the neighborhood of the starting point and inserting control points.	300 px
		N/A	Distance between two consecutive inserted control points.	10 px
	OPT-P	N/A	Maximum distance between a path node and its position in the other graph. If there is no graph edge within that distance, the path node remains unmatched.	25 px
		N/A	Maximum distance between a path edge and the corresponding trajectory in the graph. If possible, two consecutive nodes in a path are matched to connected points in the graph. Points are considered connected, if the trajectory that joins them is within a certain distance from the edge that joins the path nodes.	25 px
ours	OPT-J	d^{\max}	Maximum distance of corresponding junctions. For a junction in one graph, if there does not exist a junction within this distance in the other graph, the junction in the first graph is considered not matched.	25 px
		α	Coefficient that balances the contribution of the distance and the difference of node degree in the matching cost.	100
	OPT-G	N/A	Maximum distance between matching control points. If a pair of control points from two different graphs lies within this distance, they are considered matched and contribute to the count of true positives.	25 px
		N/A	Maximum distance between corresponding starting points. For a starting point in one graph, if there is no point in the other graph within this distance, all the control points inserted to the first graph are counted as unmatched.	25 px
		N/A	The maximum distance travelled when exploring the neighborhood of the starting point and inserting control points.	300 px
		N/A	Distance at which consecutive control points are inserted to the graph.	10 px

Table 31. Parameters used for computing the scores.

5 Supplementary material for the description of the proposed scores

In this section we put additional information about our scores that we found too detailed for the main body of the paper.

5.1 Details and comments on the new path-based score OPT-P

To complement the description of the path-based score, we present the associated algorithm 1. It consists in iterating sampling a path from one graph, matching it to the other graph, and computing the contribution to the score.

Algorithm 1: Computing our path recall R_P , a part of our path-based score **OPT-P** from section 3.1.

```

input :
    A predicted graph  $G^{\text{est}}$ 
    A ground truth graph  $G^{\text{gt}}$ 
output:
    path recall  $R_P$ 
 $\Pi \leftarrow \emptyset$  // the set of paths
while some edges are left both in  $G^{\text{est}}$  and in  $G^{\text{gt}}$  do
     $\pi \leftarrow \text{SamplePath}(G^{\text{gt}})$ 
     $\Pi \leftarrow \Pi \cup \pi$ 
    // see section 3.1 of the paper for the definitions of  $\mathcal{S}$  and  $\Psi$ 
     $\mathcal{S}, \Psi \leftarrow \text{MatchPathToGraph}(\pi, G^{\text{est}})$ 
    compute  $P_\pi \leftarrow \sum_{s \in \mathcal{S}} \frac{\sum_{s \in \mathcal{S}} l(s)^2}{l(\pi)^2}$  and store it //  $l(\cdot)$  is path or segment
    length
    remove the edges of  $\pi$  from  $G^{\text{gt}}$ 
    remove the edges of each  $\psi \in \Psi$  from  $G^{\text{est}}$ 
end
 $R_P \leftarrow \frac{1}{|\Pi|} \sum_{\pi \in \Pi} P_\pi$ 
```

Path sampling. Path sampling consists in picking a road end point as a path starting point and randomly exploring the graph depth-first until another end point is found. If the network has one, or no end points, a random loop is returned, unbuckled to form a path. For completeness, we present the details of this procedure in algorithm SamplePath.

Path matching. We formulate the problem of matching a path sampled from the ‘source’ network to the ‘target’ network as energy minimization and solve it using dynamic programming. A trajectory of an individual path $\pi \in \Pi$ is represented as a sequence of points $\pi = (p_i)_{0 \leq i < n}$ evenly spaced in the source

Procedure SamplePath.

The decomposition of a graph into paths, from section 3.1. It is based on a standard, randomized depth-first graph traversal. We outline it here for the reader's convenience.

```

input : A graph  $G$ 
           with vertices  $V$ 
           and edges  $\mathcal{E}$ 
output: A path  $\pi$ 

// each node and edge can be unseen, onPath, or offPath
mark all  $v \in V$  as unseen ;
mark all edges  $e \in \mathcal{E}$  as unseen ;
if a singly-connected node exists in  $G$  then
     $v \leftarrow$  random singly-connected vertex ;
else
     $v \leftarrow$  random vertex ;
end
mark  $v$  as onPath ;
 $C \leftarrow (v)$ ;                                // a stack representing the current path
loop  $\leftarrow C$ ;                                // returned if no open-ended path is found
while  $C$  is not empty do
     $n \leftarrow$  LastElement( $C$ );
    if  $v$  has 1 edge onPath and no other edges then
        return  $C$ ;                                // found a path
    else
        if  $v$  has unseen edges then
             $u \leftarrow$  randomly selected neighbor of  $v$  along an unseen edge;
            if  $u$  is onPath then          // loop found; don't descend to  $u$ 
                loop  $\leftarrow C$ ;
                mark the edge  $(v, u)$  offPath ;
            else if  $u$  is offPath then          // do not descend to  $u$ 
                mark the edge  $(v, u)$  offPath ;
            else if  $u$  is unseen then          // descend to  $u$ 
                mark the edge  $(v, u)$  onPath ;
                mark  $u$  onPath ;
                push  $u$  on top of  $C$  ;
            end
        else                                // backtrack
            //  $C[-2]$  is the 2nd from last element of  $C$ 
            mark the edge  $(C[-2], v)$  offPath ;
            mark  $v$  as offPath ;
            pop  $v$  from  $C$  ;
        end
    end
end
return loop ; // no open-ended path found, return an unbuckled loop

```

network. A similar trajectory of connected points $\tau = (t_i)_{0 \leq i < n}$ is sought in the target network. The similarity of the trajectories is measured in terms of the matching cost

$$c(\pi, \tau) = \sum_{0 \leq i < n} d^2(p_i, t_i), \quad (1)$$

where d is the Euclidean distance. We anticipate situations where the trajectory is only partly represented in the target network, in which case some p_i will remain unmatched, and some pairs p_{i-1}, p_i might be matched to points that are disconnected.

The search for the most similar trajectory is performed by means of dynamic programming. To that end, we define the set of candidate positions for t_i , denoted T_i , to contain one candidate per each edge of the target network within a distance d^{\max} from p_i , and an element denoted \emptyset to model the special case when the point p_i remains unmatched. For $0 \leq i < n$ and $t \in T_i$, we define the cost of matching the path point p_i to position t in the graph

$$c_{it} = \begin{cases} d^2(p_i, t) & t \neq \emptyset \\ c^{\max} & t = \emptyset \end{cases}, \quad (2)$$

where c^{\max} is a high value that penalizes leaving p_i unmatched. In order to penalize disconnected trajectories we define for $0 < i < n$ and $t \in T_{i-1}$, $t' \in T_i$ the pairwise cost $c_{itt'} = c^{\max}$ if t and t' are not connected by a path that runs within a distance d^{\max} from the edge (p_{i-1}, π_i) . We set $c_{itt'} = 0$ otherwise. We find the trajectory

$$\arg \min_{(t_i)_{0 \leq i < n}} \sum_{0 \leq i < n} c_{it_i} + \sum_{0 < i < n} c_{it_{i-1}t_i} \quad (3)$$

using the Viterbi algorithm.

After the path has been matched to the graph and its corresponding trajectory $\tau = (t_i)_{0 < i < n}$ has been identified, we find its connected sub-trajectories $\psi = (t_i)_{k_\psi \leq i < k'_\psi}$, such that all $t_{i-1}, t_i \in \psi$ are connected, but for $\psi \neq \psi'$ any $t \in \psi$, $t' \in \psi'$ remain disconnected. We denote the sets of these connected sub-trajectories by Ψ . Each of the $\psi \in \Psi$ corresponds to a sub-path of π denoted as $s_\psi = (p_i)_{k_\psi \leq i < k'_\psi}$. We denote the set of such sub-paths by \mathcal{S} . The $s \in \mathcal{S}$ are used to compute the path precision and recall as indicated in section 3.1 of the paper.

Score interpretation. The path precision and recall can loosely be interpreted as probabilities that a path randomly selected in one network is present (and connected) in the other network, if the distribution governing this selection is defined by our path sampling algorithm. Choosing a different decomposition algorithm might result in a different probability, and ideally we would like to use a distribution that is in some sense ‘fair’, for example, one that assigns equal probabilities to all paths in a network. However, the number of paths in our

networks is very large and sampling from such a distribution would be computationally infeasible. We therefore chose to keep our simple sampling scheme and experimentally demonstrated its good behaviour.

5.2 Additional details of the new graph-based score **OPT-G**

The main difference between the existing subgraph-based score and ours are described in section 2.3 and section 3.3 respectively. Our score relies on iteratively sampling a starting point in one of the graphs, finding its closest point in the other graph, cropping out subgraphs containing graph locations accessible by travelling a predefined distance in the graphs away from the starting points, and comparing the resulting subgraphs. The cropping of a graph around a starting point is performed by a breadth-first graph traversal. Traversal of each branch terminates whenever a predefined distance has been travelled. In addition, only the nodes that lead away from the starting point are considered. Control points are inserted in equal intervals during the traversal. The comparison of the subgraphs is based on one-to-one matching of the control points. This is performed by the Hungarian algorithm with the cost of matching two control points equal to the Euclidean distance between them. Only points within a predefined distance are matched. Calculation of the score is based on the number of matched and unmatched control points. We define subgraph-based precision as

$$P_G = \frac{\text{TP}_G}{\text{PP}_G} \quad (4)$$

and subgraph-based recall as

$$R_G = \frac{\text{TP}_G}{\text{AP}_G}, \quad (5)$$

where TP_G is the total number of matched control points, PP_G is the number of control points in the predicted graph and AP_G is the number of control points in the ground truth graph.

6 Stability of the metrics

Since the computation of **OPT-P** and **OPT-G** involves executing random node and path sampling, we need to make sure that this does not cause variations of the score that could influence the evaluation. In table 32 we report mean and standard deviation over a ten-fold computation of the metrics for a single image of the RoadTracer dataset. The standard deviation for **OPT-P** amounts to roughly 2% of the value of the score and is smaller than that of **TLTS** and slightly larger than **APLS**. The standard deviation for **OPT-G** represents around 2% of the score and is smaller than that of **SUBG**. Note that the reported numbers have been computed for a single test image, and the score aggregated over many test images will in general have a much lower variance.

	existing scores				new scores			
	CCQ	TLTS	APLS	JUNCT	SUBG	OPT-P	OPT-J	OPT-G
	qual.	corr.		f1	f1	f1	f1	f1
Mean	0.345	0.191	0.561	0.690	0.583	0.394	0.706	0.530
SD	0	0.0126	0.0049	0	0.0092	0.0085	0	0.0080

Table 32. Standard deviation of the existing and the new metric on the road network of Amsterdam reconstructed by the *Segmentation* method.