Improved Road Connectivity by Joint Learning of Orientation and Segmentation - Supplementary Material

Anil Batra [*1]   Suriya Singh [* †2]   Guan Pang[3]   Saikat Basu[3]   C.V. Jawahar[1]   Manohar Paluri[3]

[1]IIIT Hyderabad   [2]MILA / Polytechnique Montréal   [3]Facebook

Abstract

The supplementary material contains: (1) Architecture details, (2) Dataset Details, (3) Additional Implementation Details, (4) Junction Learning, an additional related task for multi-task learning, (5) Additional ablation study, (6) Additional qualitative results for DeepRoadMapper [5], Topology Loss [6], LinkNet34 [2], Stacked Multi-branch (Ours) (7) Qualitative results of Orientation prediction overlay on the image, and (8) Qualitative result for Connectivity refinement.

1. Architecture Details

We use encoder-decoder structured model with two intermediate stacks of multi-branch module to predict orientation and segmentation at different scales. We perform downsampling in the model using 2×2 `max pool` layer. We use three basic Resnet [4] blocks in the shared encoder and multi-branch module. We add `BatchNorm` layer after each convolution layer. The shared encoder reduces the input resolution to $H/4, W/4$ using the strided convolution and max-pool layer, which is fed to the multi-branch module. The shared encoder and final decoder has 64, 64, 64, 128, 64, 32 channels and each multi-branch module has 128. The final decoder block uses bottleneck deconvolution similar to LinkNet [2]. There are 29.02 Million parameters in the joint model.

2. Dataset Details

We perform our experiments on the two challenging road datasets Spacenet [8] and DeepGlobe[3] to assess the significance of orientation learning task. In the current scope we utilize only 3-Band RGB image of both datasets to detect the curvilinear road structures.

---

*Equal Contribution

†Work partially done as Research Fellow at IIIT Hyderabad

Spacenet [8]: This dataset provides imagery from four different cities: `Paris`, `Las Vegas`, `Shanghai`, and `Khartoum`. The imagery is available at ground resolution of 30cm/pixel and pixel resolution of 1300×1300 in GeoTiff format (16-bit). Dataset consists of different road types (`Motorway`, `Primary`, `Secondary`, `Tertiary`, `Residential`, `Unclassified`, `Cart Tracks`) from the four cities, having diverse road widths and visual appearance. Road annotation is provided in the form of line-string, representing the center line of roads. Each image may have multiple line-strings and each line-string consists of pixel coordinates {X Y} depicting road centerline points in the 2D image plane, assuming top-left corner as the origin.

The public dataset consists of 2567 images with road vector data as labels. We split the dataset into 2000 images for training and 567 for testing. When we split the imagery, each city equally contributes to train (80% per city) and validation set (20% per city). To augment the training dataset we create crops of 650×650 with overlapping region of 215 pixels, thus providing ∼32K images. For validation we use the crops of same size without overlap. For fair comparison with other datasets, we also convert 16-bit imagery to 8-bit using MinMax scaling followed by adaptive histogram equalization on each channel independently.

DeepGlobe [3]: It includes GeoTiff imagery from three different areas: `Thailand`, `Indonesia`, and `India`. The ground resolution of 3-Band RGB image is 50cm/pixel and pixel resolution is 1024×1024. Pixel level annotation is provided for `road` and `background` classes. It contains 4696 images for training phase and 1530 for validation. We augment it by creating crops of size 512×512 with overlapping region of 256 pixels, yielding ∼42K images for training phase.

We evaluate and report the road connectivity metrics on full resolution images at inference time for each dataset.

3. Addition Implementation Details

An identical training procedure for all the experiments is implemented in PyTorch [7] framework and the parameters of each model are initialized with He *et al.* [4]. Segmentation outputs from the joint learning module is converted into network graph with each linear road segment representing an edge. We perform simple graph processing to remove the duplicate edges and small hanging road segments. The proposed graph is converted into line strings and used it to evaluate the connectivity metric APLS [8] of road network.

4. Junction Learning

We perform experiments to show that the improvement in road connectivity is due to the proposed task in contrast to the multi-task learning [1]. We develop another related task of classifying the junctions and road tracing points, described as *Junction Learning*. We convert the road mask into a graph. If node of the graph has a degree more than 3 then we consider the node of a graph as road junction, otherwise node is assumed to be a tracing point of road. The junctions and tracing points are generated using a distance transform with Gaussian kernel as shown in Figure.1. We use pixel wise loss to learn junctions with three classes: `background`, `tracing points`, `junctions`.
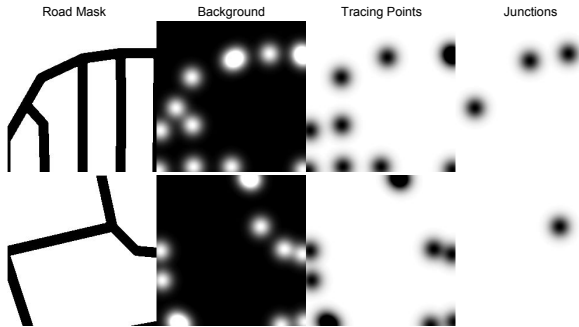

Figure 1: Ground truth for *Junction Learning*

5. Ablation Studies

Quantization Size	Spacenet		DeepGlobe	
	road IoU$^\alpha$	APLS	road IoU$^\alpha$	APLS
5°	63.35	63.12	66.62	72.39
10°	63.75	**63.65**	**67.21**	**73.12**
20°	**63.80**	63.01	67.02	72.75

Table 1: Effect of different quantizations on orientation angles in the proposed stacked multi-branch module. **road IoU$^\alpha$**: accurate pixel based intersection over union. **APLS**: average path length similarity on the extracted graph from road segmentation.

Quantization of Orientation Angles: We perform ablation study on the different quantization levels for orientation angles and report the results in Table.1. The results shows that quantization of $10°$ is good choice for better road connectivity and we use it in all the comparison methods.

Road Width of Spacenet Mask: We convert the road line strings of Spacenet dataset [8] using distance transform with Gaussian kernel along the center line of roads. This provide a choice to choose threshold corresponding to different road widths. We perform experiments with different thresholds (as shown in Figure.2) using LinkNet34 [2] model and choose the threshold of 0.76 in all the experiments. The threshold of 0.76 correspond to road width of 6-7 meters.
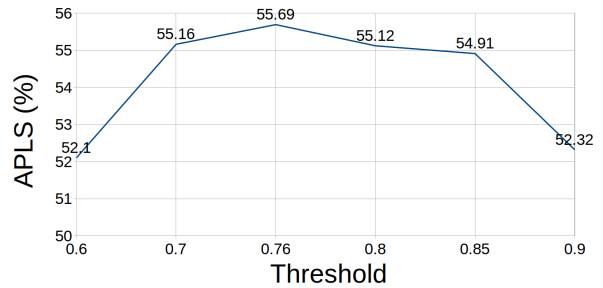

Figure 2: Effect of different road widths for Spacenet [8] Road masks using LinkNet34 [2] model on the connectivity metric APLS.

6. Qualitative Comparisons

We present the qualitative comparison with state-of-art segmentation based methods DeepRoad Mapper [5], Topology Loss [6], LinkNet34 [2] in Figures.3, 4, 5, 6, 7, 8.

7. Qualitative Results of Orientation

We present the qualitative results for orientation and segmentation prediction in Figures.9, 10, 11, 12 from our approach.

8. Qualitative Results of Connectivity Refinement

We present the qualitative results of connectivity refinement over orientation and segmentation learning in Figures.13, 14 via LinkNet34 [2] as joint learning module.

References

[1] R. Caruna. Multitask learning: A knowledge-based source of inductive bias. In *ICML*, 1993. 2

[2] A. Chaurasia and E. Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *VCIP*, 2017. 1, 2, 4, 5, 6, 7, 8, 9, 14, 15

[3] 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 *CVPRW*, 2018. 1

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1, 2

[5] G. Máttyus, W. Luo, and R. Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *ICCV*, 2017. 1, 2, 4, 5, 6, 7, 8, 9

[6] A. Mosinska, P. Márquez-Neila, M. Kozinski, and P. Fua. Beyond the pixel-wise loss for topology-aware delineation. In *CVPR*, 2018. 1, 2, 4, 5, 6, 7, 8, 9

[7] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPSW*, 2017. 2

[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, 2
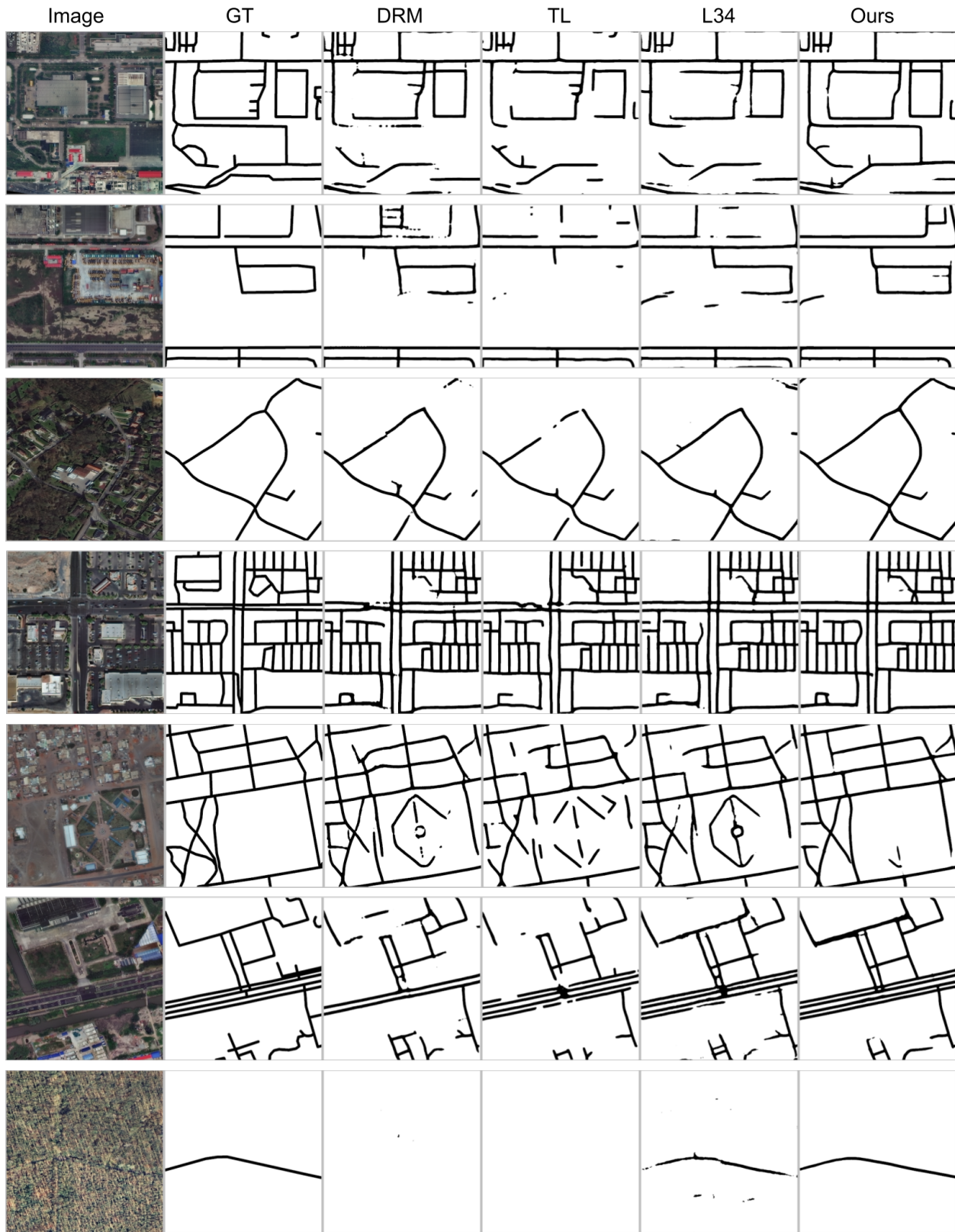
Figure 3: Qualitative Comparisons with state-of-the-art methods - **DRM**: DeepRoad Mapper[5], **TL**: Topology Loss [6], **L34**: LinkNet34 [2]
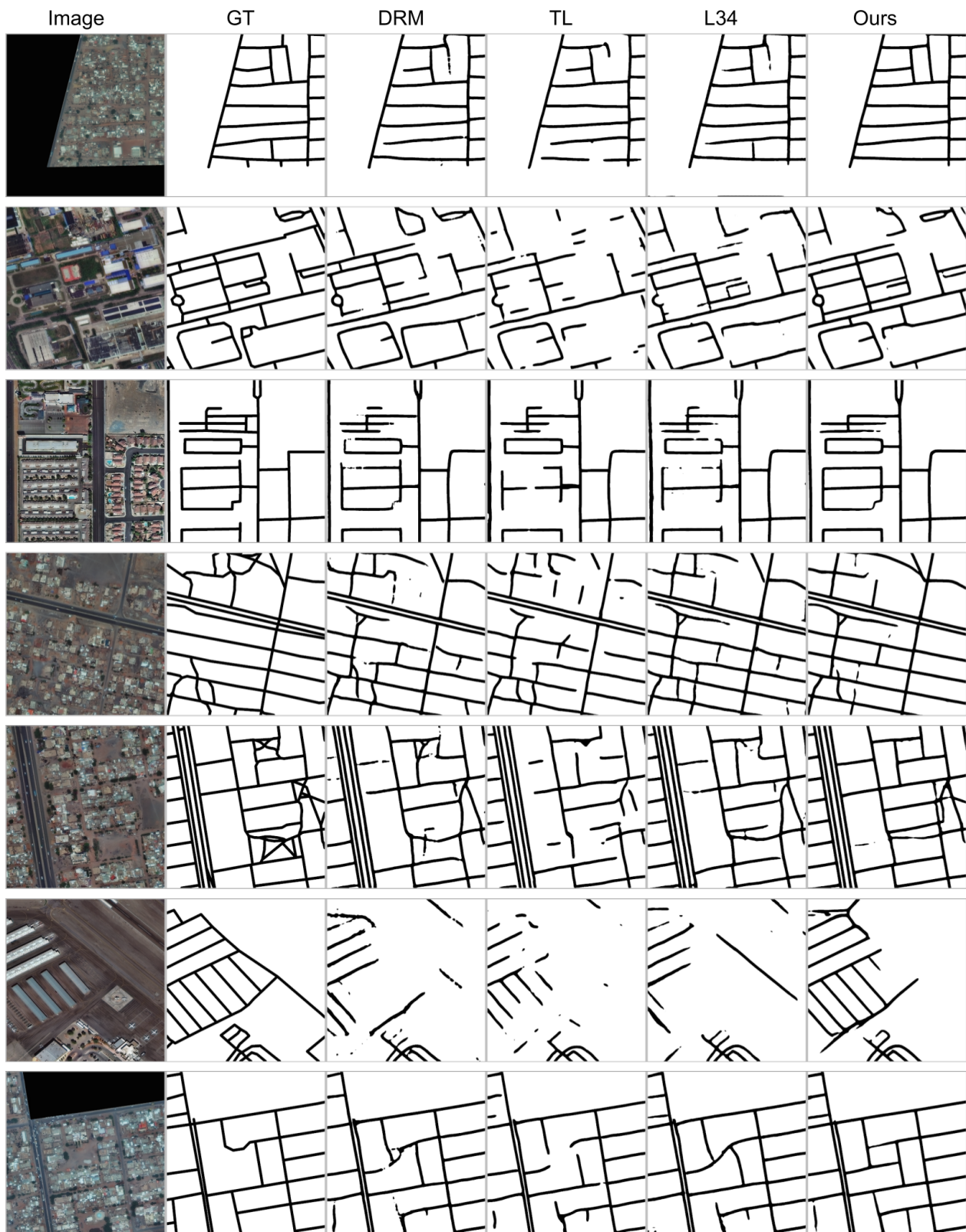
Figure 4: Qualitative Comparisons with state-of-the-art methods - **DRM**: DeepRoad Mapper[5], **TL**: Topology Loss [6], **L34**: LinkNet34 [2]
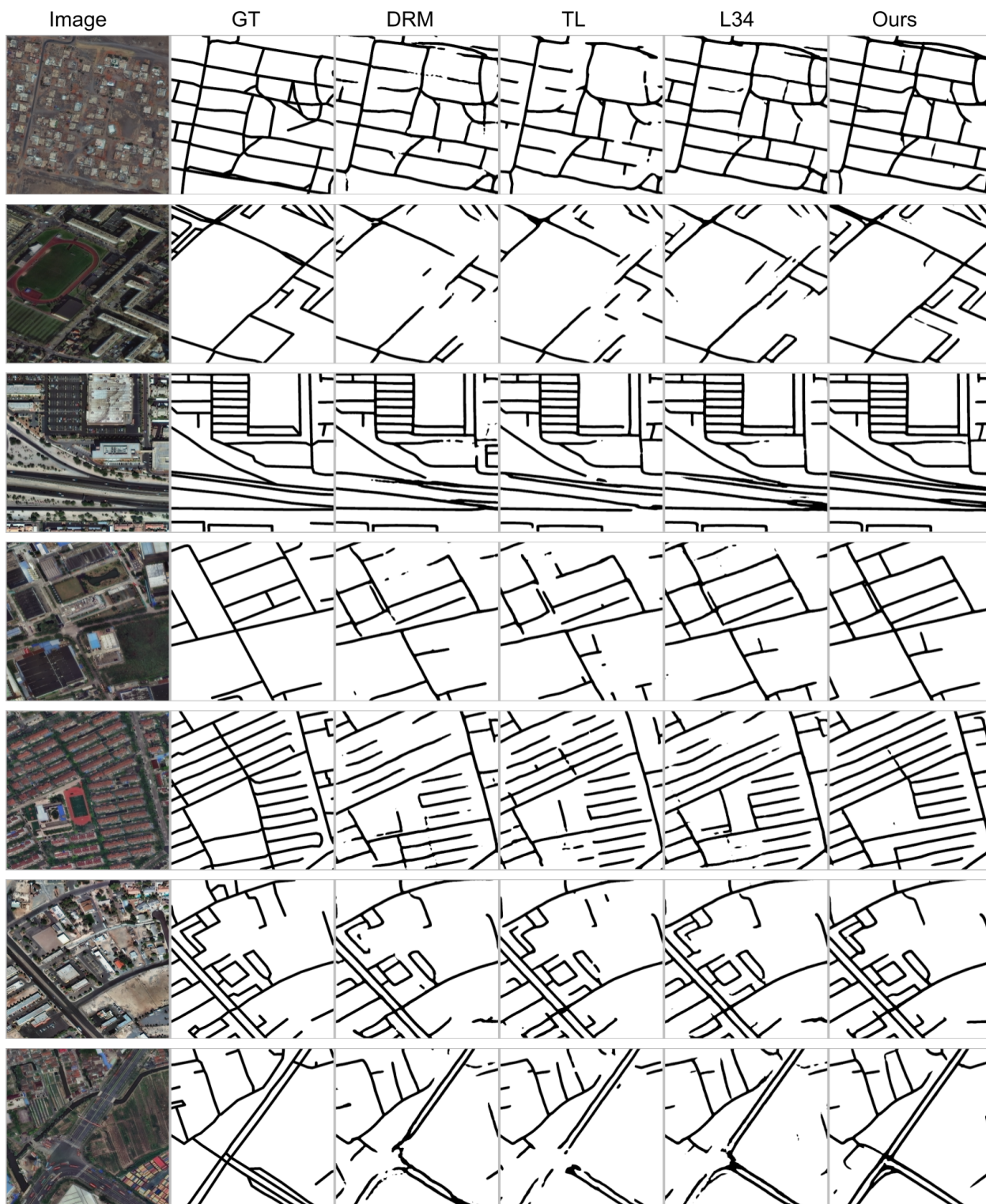
Figure 5: Qualitative Comparisons with state-of-the-art methods - **DRM**: DeepRoad Mapper[5], **TL**: Topology Loss [6], **L34**: LinkNet34 [2]
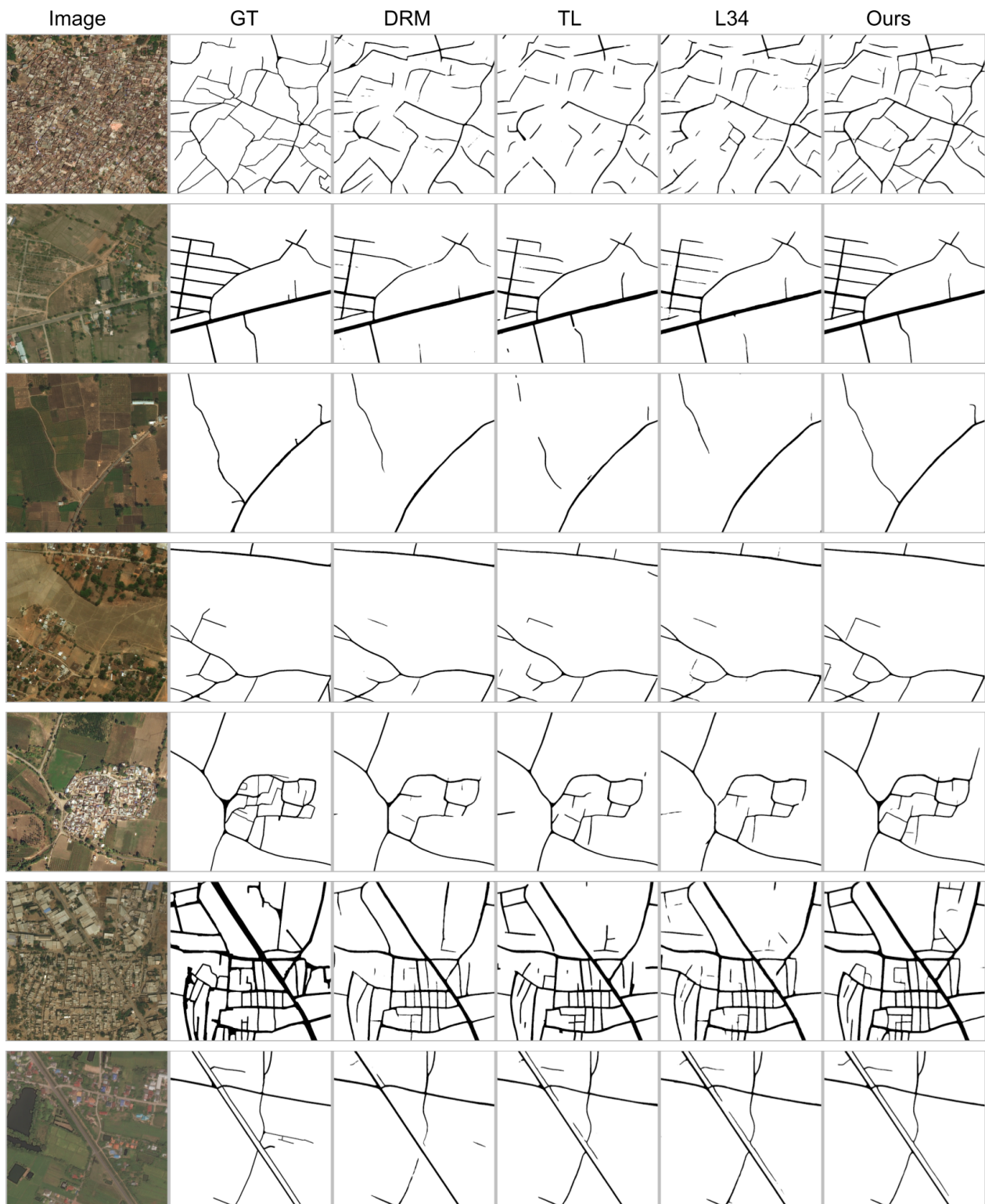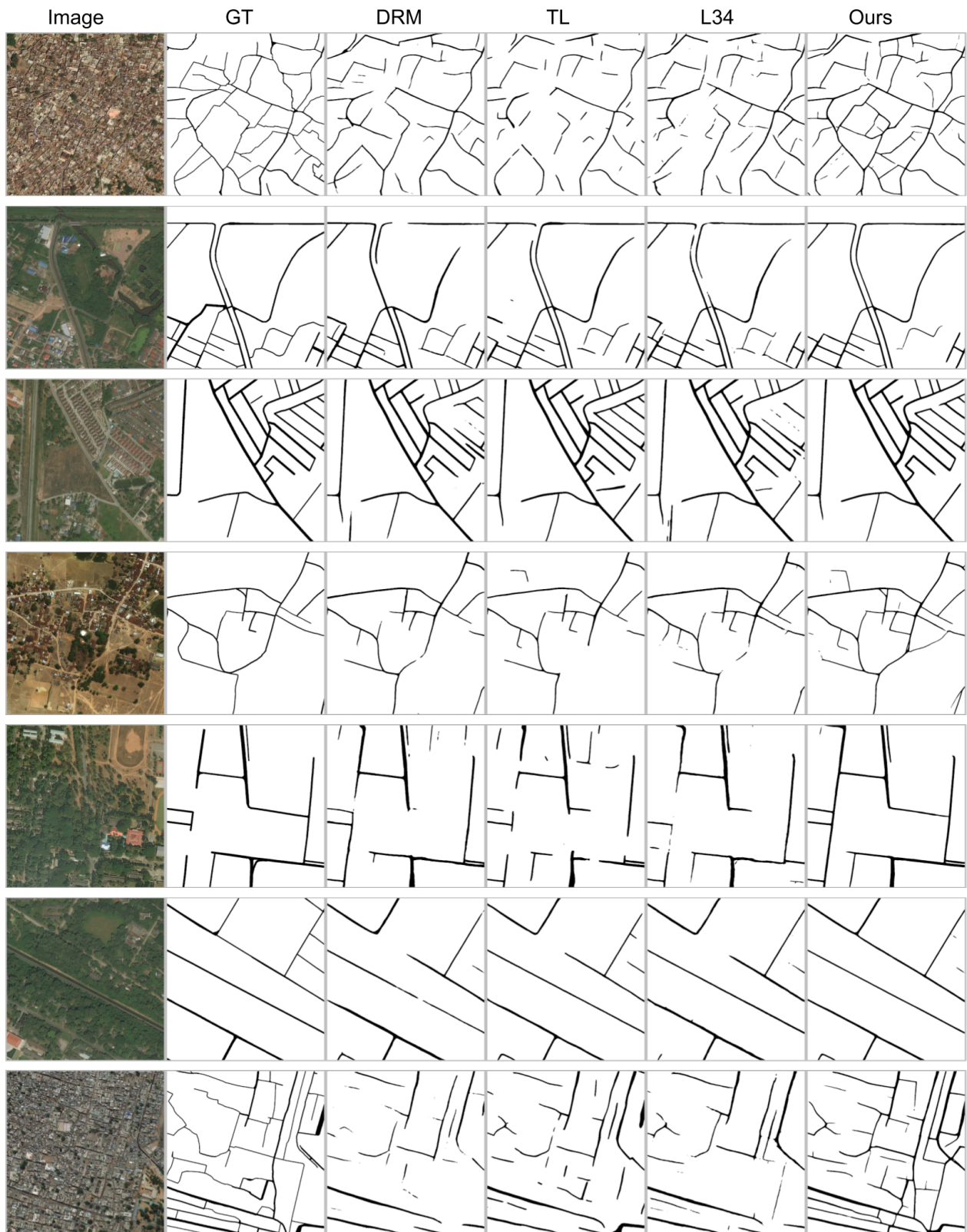
Figure 6: Qualitative Comparisons with state-of-the-art methods - **DRM**: DeepRoad Mapper[5], **TL**: Topology Loss [6], **L34**: LinkNet34 [2]

Figure 7: Qualitative Comparisons with state-of-the-art methods - **DRM**: DeepRoad Mapper[5], **TL**: Topology Loss [6], **L34**: LinkNet34 [2]

Figure 8: Qualitative Comparisons with state-of-the-art methods - **DRM**: DeepRoad Mapper[5], **TL**: Topology Loss [6], **L34**: LinkNet34 [2]
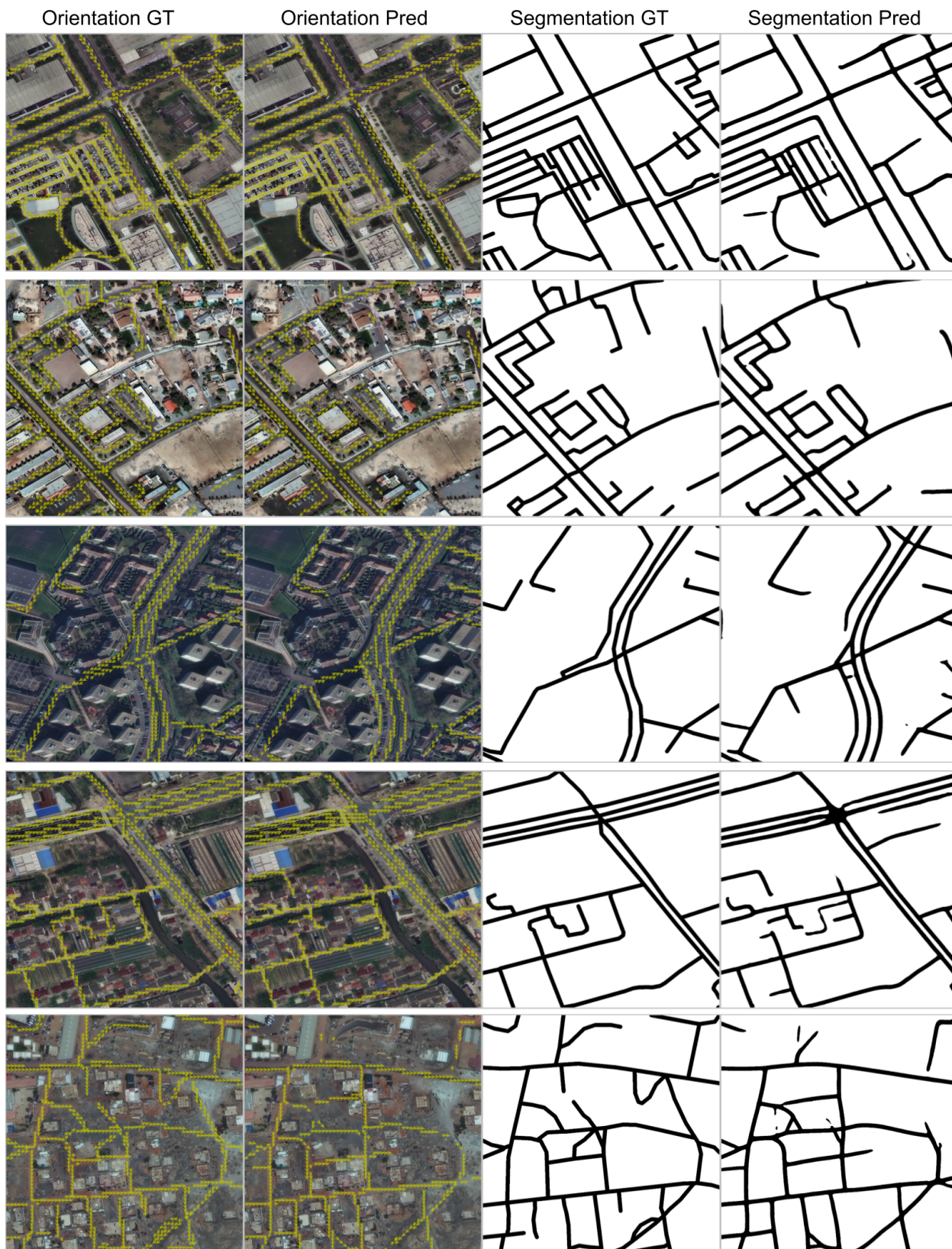
Orientation GT	Orientation Pred	Segmentation GT	Segmentation Pred

Figure 9: Qualitative results of Orientation and Segmentation prediction of **Ours** method. Orientation GT and Prediction are visualized as overlay on the image.
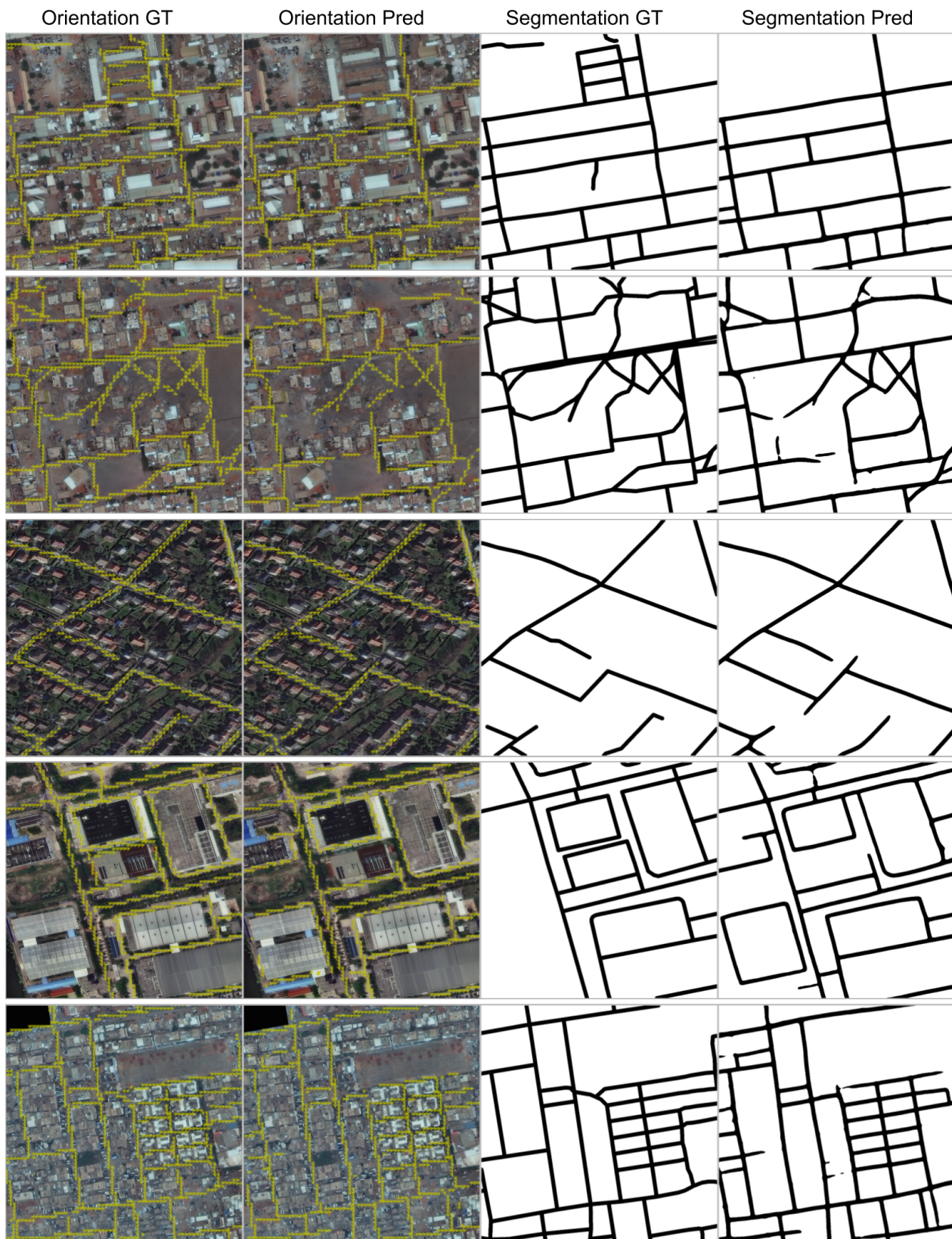
Orientation GT	Orientation Pred	Segmentation GT	Segmentation Pred

Figure 10: Qualitative results of Orientation and Segmentation prediction of **Ours** method. Orientation GT and Prediction are visualized as overlay on the image.
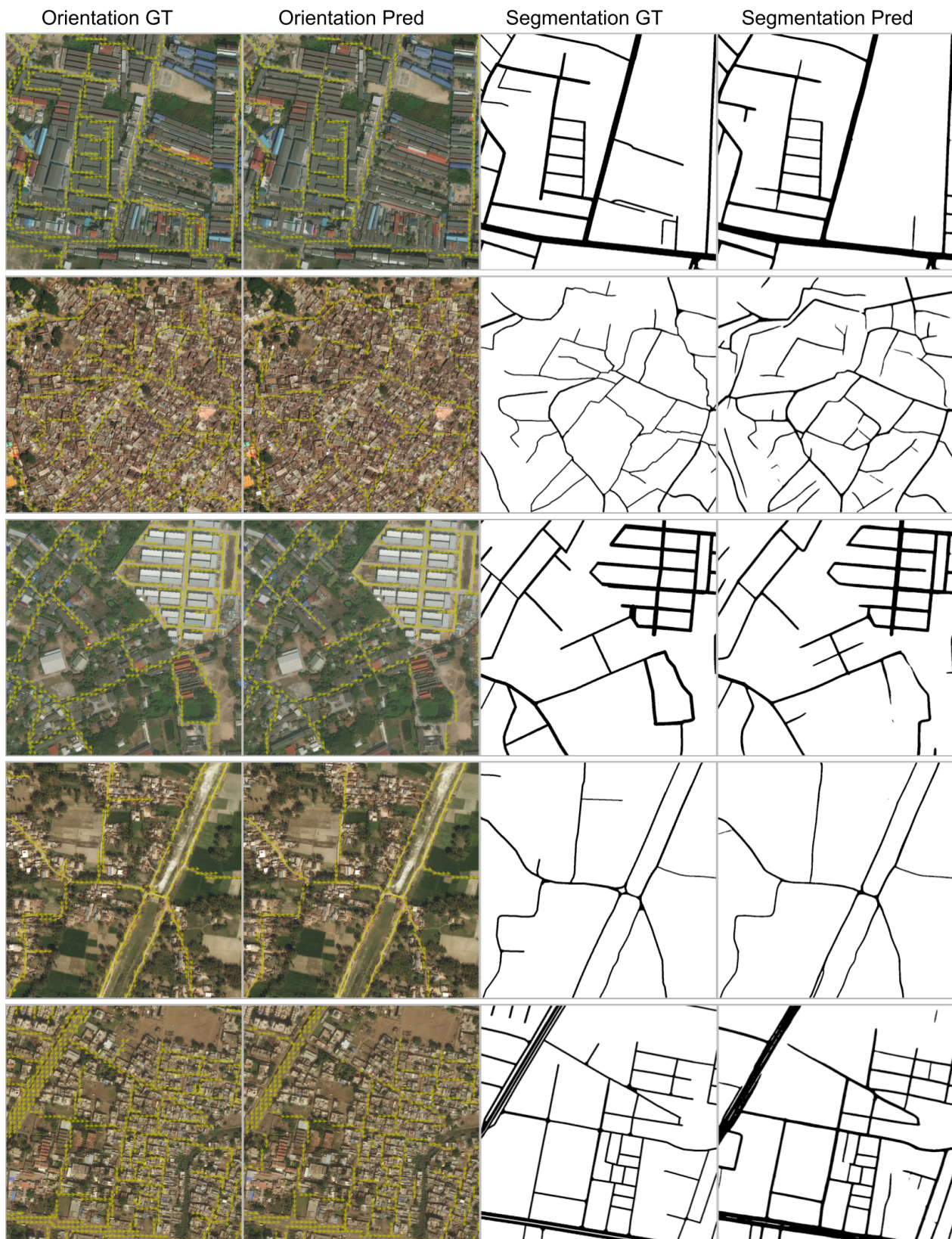
Orientation GT	Orientation Pred	Segmentation GT	Segmentation Pred

Figure 11: Qualitative results of Orientation and Segmentation prediction of **Ours** method. Orientation GT and Prediction are visualized as overlay on the image.
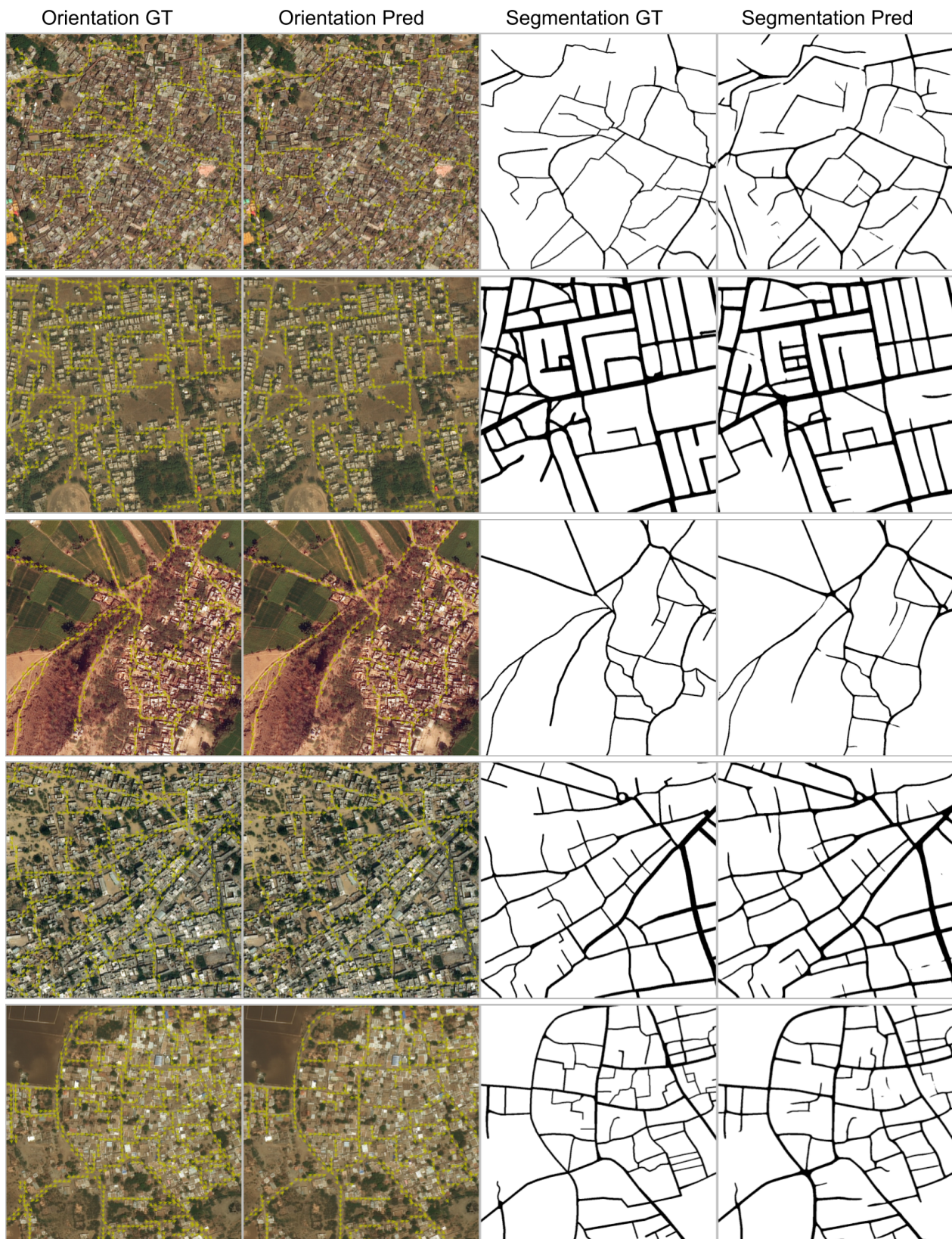
Orientation GT	Orientation Pred	Segmentation GT	Segmentation Pred

Figure 12: Qualitative results of Orientation and Segmentation prediction of **Ours** method. Orientation GT and Prediction are visualized as overlay on the image.
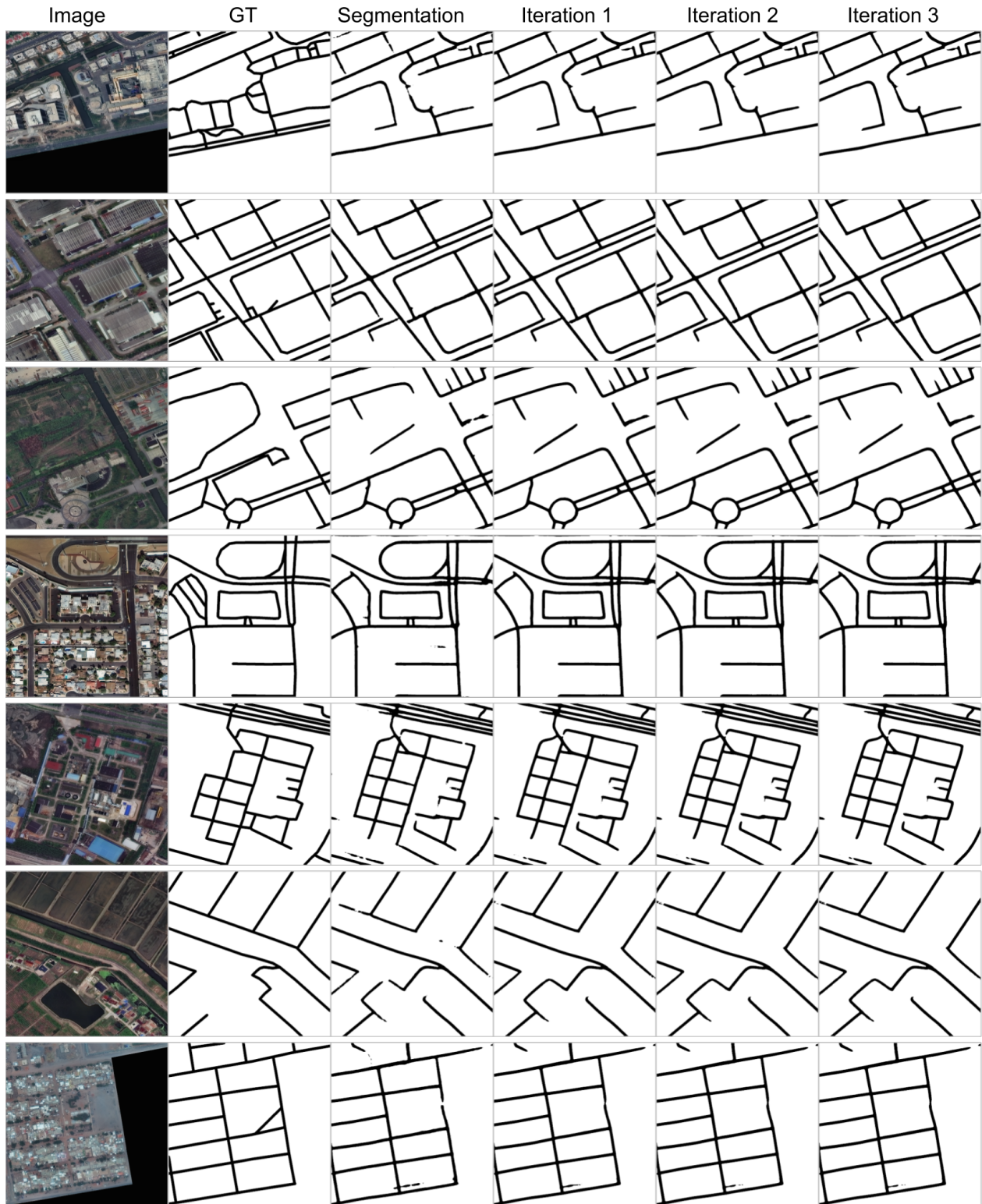
Figure 13: Qualitative results of Connectivity refinement over segmentation and orientation learning with LinkNet34 [2] as joint learning module.
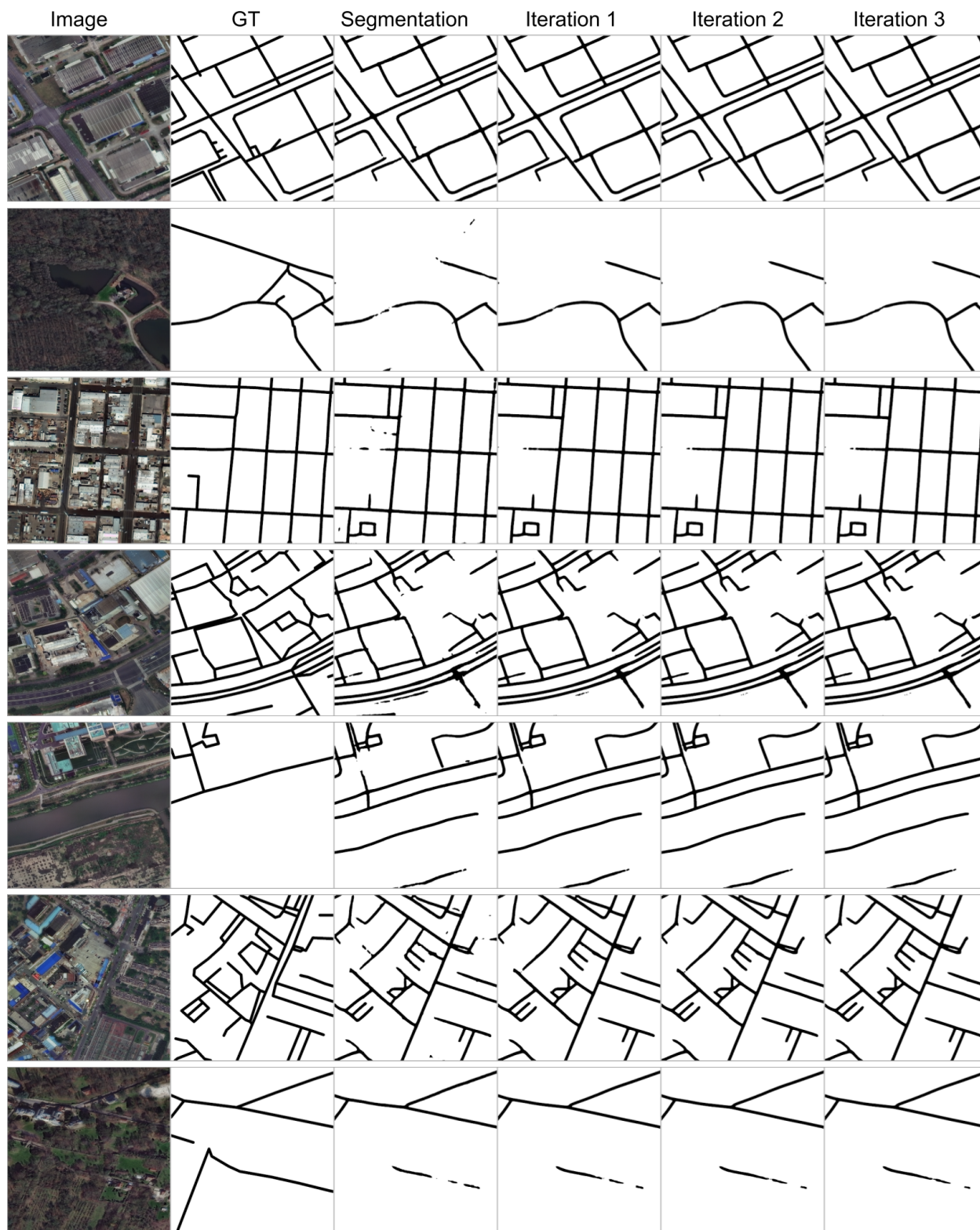
Figure 14: Qualitative results of Connectivity refinement over segmentation and orientation learning with LinkNet34 [2] as joint learning module.