TopDiG: Class-agnostic Topological Directional Graph Extraction from Remote Sensing Images – Supplementary Material

Bingnan Yang[1] , Mi Zhang[1†] , Zhan Zhang[2] , Zhili Zhang[1] , Xiangyun Hu[1]

[1] School of Remote Sensing and Information Engineering, Wuhan University, China

[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China

[†] Corresponding author { mizhang@whu.edu.cn }

The supplementary material contains: 1) Additional quantitative comparisons between TopDiG and other approaches; 2) Comparison of attentive maps produced by different methods; 3) Additional visual comparisons between TopDiG and other approaches.

1. Additional Quantitative Comparisons

To further evaluate TopDiG, we provide additional quantitative comparisons with classic or recent relevant approaches. For segmentation-based methods, we evaluate 12 pure semantic segmentation models on both polygon-shape and line-shape targets. We also evaluate a classic building extraction model named Frame field [5] on polygon-shape targets. In terms of contour-based approaches, two influential workflows called Curve-GCN [8] and Deep Snake [10] are evaluated on polygon-shape targets. For graph generation methods, we select Enhanced-iCurb [13] since it focuses on line-shape targets.

1.1. Compare with Segmentation-based Method

We compare TopDiG with a few of segmentation-based methods on **Inria** and **Massachusetts**. In terms of **Inria** (Table 1), TopDiG reports score of approximately 85% $\mathbf{mIoU}^{mask}$ with respect to pixel-wise metrics. It surpasses all those segmentation-based methods with at least 1% $\mathbf{mIoU}^{topo}$ and 3% **APLS** regarding topology-wise metrics. For **Massachusetts** (Figure 2), TogDiG outperforms achieves highest $\mathbf{mIoU}^{topo}$ and **APLS** with scores of 71% and 60%. Visual examples in Figure 2 and Figure 3 clearly show that segmentation-based methods require post-processing to obtain topology from masks and suffer from low quality topological graphs.

1.2. Compare with Contour-based Method

Quantitative comparisons are conducted between TopDiG and two classic contour-based approaches,namely Deep Snake and Curve-GCN, on **Inria** dataset. As shown in Table 1, TopDiG notably surpasses these two methods on both pixel-wise and topology-wise metrics with at least 6% $\mathbf{mIoU}^{mask}$, 4% $\mathbf{mIoU}^{topo}$ and 15% **APLS**. The main drawback of contour-based methods is the unavoidable contour initialization procedure which obstructs their applications on targets with complicated topological structures (see image with red cross in Figure 2).

1.3. Compare with Graph Generation Method

Table 2 presents comparison between TopDiG and graph generation approach Enhanced-iCurb. It reports that TopDiG achieves superiority over its competitor with 13% $\mathbf{mIoU}^{topo}$ and 22% **APLS**. In Figure 3, visual instances illustrate that the iterative prediction and imitation learning strategies employed in Enhanced-iCurb pull the extracted roads to the road boundaries instead of central areas. By contrast, TopDiG concentrates on centerlines of roads and extracts reliable topological graphs.

2. Attentive Maps of Different Methods

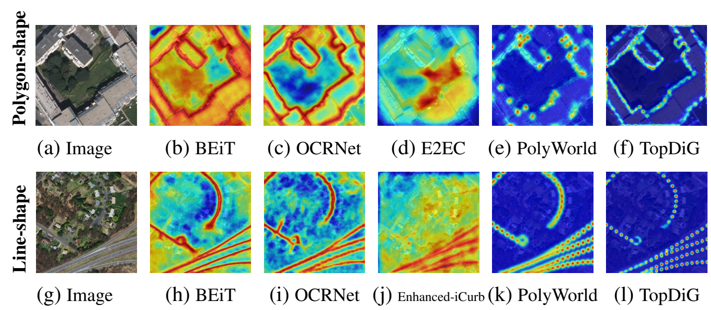

Figure 1. **Visual comparison of attentive maps** for a few approaches on polygon-shape and line-shape targets. TopDiG obtains compact perception of topological components.

We visually present attentive maps in different methods to illustrate the distinction between TopDiG and other approaches. As shown in Figure 1, segmentation-based methods BEiT and OCRNet can obtain coarse semantic attention on topological components such as boundaries and centerlines but they require elaborate post-processing to achieve topological graphs. Contour-based method E2EC adopts semantic segmentation model as backbone and neglects geometric textures of polygon boundaries. Graph generation approach Enhanced-iCurb fails to obtain compact attention on essential target topology. Another graph generation method PolyWorld suffers from insufficient geometir-

Category	Method		Backbone	Pixel-wise Metrics			Topology-wise Metrics			
				PA^{mask}↑	$F1^{mask}$↑	$mIoU^{mask}$↑	PA^{topo}↑	$F1^{topo}$↑	$mIoU^{topo}$↑	APLS↑
Segmentation-based	FCN [9]	[CVPR2015]	ResNet-101	0.92	0.81	0.79	0.90	0.48	0.60	0.30
	CCNet [6]	[ICCV2019]	ResNet-101	0.92	0.81	0.79	0.90	0.46	0.60	0.27
	DANet [4]	[CVPR2019]	ResNet-101	0.92	0.80	0.79	0.90	0.47	0.60	0.29
	GCNet [2]	[ICCV2019]	ResNet-101	0.92	0.79	0.78	0.90	0.46	0.60	0.27
	EncNet [15]	[CVPR2018]	ResNet-101	0.92	0.80	0.79	0.90	0.46	0.59	0.29
	OCRNet [14]	[ECCV2020]	HRNet-V2	0.92	0.81	0.79	0.89	0.47	0.60	0.30
	PSPNet [16]	[CVPR2017]	ResNet-101	0.92	0.80	0.79	0.90	0.47	0.60	0.28
	UperNet [11]	[ECCV2018]	ResNet-101	0.93	0.82	0.80	0.90	0.50	0.62	0.31
	SegFormer [12]	[Neurips2021]	MIT-B5	0.93	0.82	0.81	0.90	0.50	0.62	0.33
	MaskFormer [3]	[Neurips2021]	ResNet-101	0.93	0.83	0.81	0.90	0.52	0.62	0.34
	MemoryNetV2 [7]	[TPAMI2022]	Swin-transformer	0.92	0.80	0.78	0.89	0.44	0.59	0.26
	BEiT [1]	[arXiv2021]	BEiT-L	0.95	0.88	0.86	0.92	0.60	0.67	0.45
	Frame Field [5]	[CVPR2021]	HRNet-V2	0.92	0.85	0.77	0.92	0.68	0.59	0.37
Contour-based	Curve-GCN [8]	[CVPR2019]	ResNet-50	0.87	0.84	0.75	0.93	0.62	0.55	0.31
	Deep Snake [10]	[CVPR2020]	DLA	0.93	0.86	0.79	0.93	0.73	0.64	0.33
Ours	TopDiG		TCND	0.95 (+0)	0.91 (+0.03)	0.85 (-0.01)	0.94 (+0.01)	0.78 (+0.05)	0.68 (+0.01)	0.48 (+0.03)

Table 1. **Quantitative comparisons on polygon-shape targets**. We evaluate the pixel-wise and topology-wise metrics on **Inria**. TopDiG achieves competitive scores on pixel-wise metrics and outperforms all other approaches on topology-wise metrics. Red and Blue represent the top-2 scores. We use ↑ and ↑ to indicate the increases crossing all datasets.

Category	Method		Backbone	Pixel-wise Metrics			Topology-wise Metrics			
				PA^{mask}	$F1^{mask}$	$mIoU^{mask}$	PA^{topo}↑	$F1^{topo}$↑	$mIoU^{topo}$↑	APLS↑
Segmentation-based	FCN [9]	[CVPR2015]	ResNet-101	0.96	0.37	0.59	0.93	0.54	0.65	0.12
	CCNet [6]	[ICCV2019]	ResNet-101	0.96	0.11	0.51	0.92	0.21	0.52	0.05
	DANet [4]	[CVPR2019]	ResNet-101	0.96	0.17	0.53	0.92	0.29	0.54	0.06
	GCNet [2]	[ICCV2019]	ResNet-101	0.96	0.11	0.51	0.92	0.20	0.51	0.05
	EncNet [15]	[CVPR2018]	ResNet-101	0.96	0.12	0.51	0.92	0.22	0.52	0.05
	OCRNet [14]	[ECCV2020]	HRNet-V2	0.96	0.33	0.58	0.92	0.45	0.61	0.11
	PSPNet [16]	[CVPR2017]	ResNet-101	0.96	0.08	0.50	0.92	0.16	0.50	0.04
	UperNet [11]	[ECCV2018]	ResNet-101	0.96	0.38	0.60	0.92	0.50	0.63	0.14
	SegFormer [12]	[Neurips2021]	MIT-B5	0.96	0.36	0.59	0.93	0.49	0.63	0.10
	MaskFormer [3]	[Neurips2021]	ResNet-101	0.88	0.36	0.57	0.80	0.37	0.51	0.56
	MemoryNetV2 [7]	[TPAMI2022]	Swin-transformer	0.96	0.34	0.58	0.92	0.43	0.59	0.12
	BEiT [1]	[arXiv2021]	BEiT-L	0.96	0.54	0.66	0.92	0.65	0.70	0.57
Graph generation	Enhanced-iCurb [13]	[LRA2021]	FPN	-	-	-	0.89	0.68	0.58	0.38
Ours	TopDiG		TCND	-	-	-	0.95(+0.02)	0.80 (+0.12)	0.71 (+0.01)	0.60 (+0.03)

Table 2. **Quantitative comparisons on line-shape targets**. We evaluate the pixel-wise and topology-wise metrics on **Masschusetts**. TopDiG obtains better topology quality than all other methods. Red and Blue represent the top-2 scores. We use ↑ to indicate the increases crossing all datasets.

c textures when tacking relatively complicated topological structures. By contrast, TopDiG concentrates on topological components and perceives compact texture features.

3. Additional Visual Comparisons

We provide examples visually comparing TopDiG with segmentation-based, contour-based and graph generation approaches. For polygon-shape targets (Figure 2), rectangles in 1st column illustrate that TopDiG can precisely delineate concave building boundaries and images in 4th column show its ability of resisting against shadows. Furthermore, as demonstrated in the 5th column, TopDiG can also obtain interior detailed outlines of a circular building.

In terms of line-shape targets (Figure 3), segmentation-based methods suffer from severe unconsciousness, omission and jaggies (red rectangles) in obtained masks. Roads extracted by Enhanced-iCurb tend to move towards boundary areas (green rectangle in 3rd column) and can hardly solve accumulated prediction errors (green rectangle in 2nd column). As for PolyWorld, purple rectangles in 1st and 2nd columns release the omitted and redundancy connections. In contrast with these methods, TopDiG achieves reliability in aforementioned scenarios.

References

[1] Hangbo Bao, Li Dong, and Furu Wei. Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*, 2021. 2

[2] Yue Cao, Jiarui Xu, Stephen Lin, Fangyun Wei, and Han Hu. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019. 2

[3] Bowen Cheng, Alex Schwing, and Alexander Kirillov. Per-pixel classification is not all you need for semantic segmentation. *Advances in Neural Information Processing Systems*, 34:17864–17875, 2021. 2

[4] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019. 2

[5] Nicolas Girard, Dmitriy Smirnov, Justin Solomon, and Yuliya Tarabalka. Polygonal building segmentation by frame field learning. In *CVPR 2021-IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1, 2

[6] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross

Figure 2. **Visual comparisons on the polygon-shape targets.** These images come from the Inria dataset. Top - bottom: BEiT, OCRNet, E2EC, PolyWorld and TopDiG. Green line: segmentation contours of buildings; Red line: simplified polygons using the DouglasPeucker algorithm; Yellow dots: detected/sampled nodes; Cyan arrow lines: directional connections between node pairs; Red cross: no predicted building; Orange rectangles: concave building outlines.

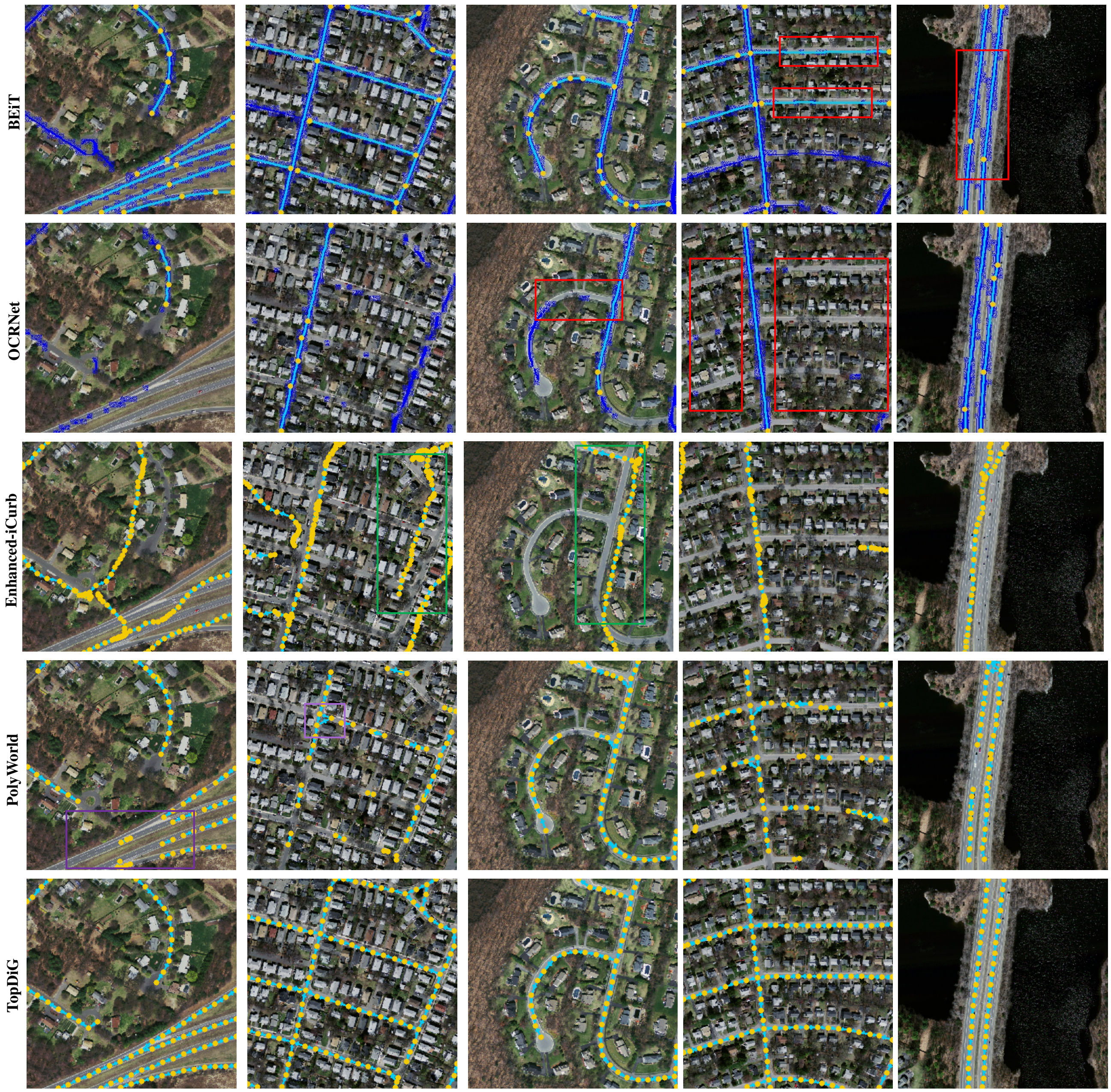

Figure 3. **Visual comparisons between TopDiG and other approaches on the line-shape targets.** These images come from the **Massachusetts** dataset. Top - bottom: BEiT, OCRNet, Enhanced-iCurb, PolyWorld and TopDiG. Blue masks: segmentation masks of roads; Yellow dots: detected/sampled nodes; Cyan arrow/straight lines: directional/non-directional connections between node pairs; Red rectangles: omitted or jagged roads masks; Green rectangles: typical errors of Enhanced-iCurb; Purple rectangles: omitted and redundancy connections produced by PolyWorld. The centerlines of BEiT and OCRNet are obtained from masks by applying the DouglasPeucker algorithm.

attention for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 603–612, 2019. 2

[7] Zhenchao Jin, Dongdong Yu, Zehuan Yuan, and Lequan Yu. Mcibi++: Soft mining contextual information beyond image for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2

[8] Huan Ling, Jun Gao, Amlan Kar, Wenzheng Chen, and Sanja Fidler. Fast interactive object annotation with curve-gcn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5257–5266, 2019. 1, 2

[9] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2

[10] Sida Peng, Wen Jiang, Huaijin Pi, Xiuli Li, Hujun Bao, and Xiaowei Zhou. Deep snake for real-time instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8533–8542, 2020. 1, 2

[11] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European conference on computer vision (ECCV)*, pages 418–434, 2018. 2

[12] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 2

[13] Zhenhua Xu, Yuxiang Sun, and Ming Liu. Topo-boundary: A benchmark dataset on topological road-boundary detection using aerial images for autonomous driving. *IEEE Robotics and Automation Letters*, 6(4):7248–7255, 2021. 1, 2

[14] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. In *European conference on computer vision*, pages 173–190. Springer, 2020. 2

[15] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context encoding for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7151–7160, 2018. 2

[16] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 2