Joint Semantic-Geometric Learning for Polygonal Building Segmentation

Weijia Li[1], Wenqian Zhao[1], Huaping Zhong[2], Conghui He[2], Dahua Lin[1]

[1]The Chinese University of Hong Kong [2]SenseTime Group Limited

Polygonal Building segmentation

•A fundamental task for disaster management, urban planning, geographical information updating, etc.

•The pixel-wise segmentation methods in most studies produce building extraction results in raster format.

•Polygonal building segmentation approaches produce more realistic building polygons in the desirable vector format for practical applications.

Limitations of the existing methods

•Relying on a perfect segmentation map to guarantee the quality of vectorization;

•Requiring a complex post-processing procedure;

•Generating inaccurate vertices with a fixed quantity, a wrong sequential order, self-intersections, etc.

Building segmentation from remote sensing images   Vector vs. Raster format

Our proposed approach

•A multi-task segmentation network for joint semantic and geometric learning via three relevant tasks.

•A rule-based vertex generation module to bridge the gap between the image and the graph based network.

•A polygon refinement network to automatically move the polygon vertices into more accurate locations.

•Generating building polygons with a flexible quantity of vertices that are in a proper sequential order.

Multi-task segmentation network

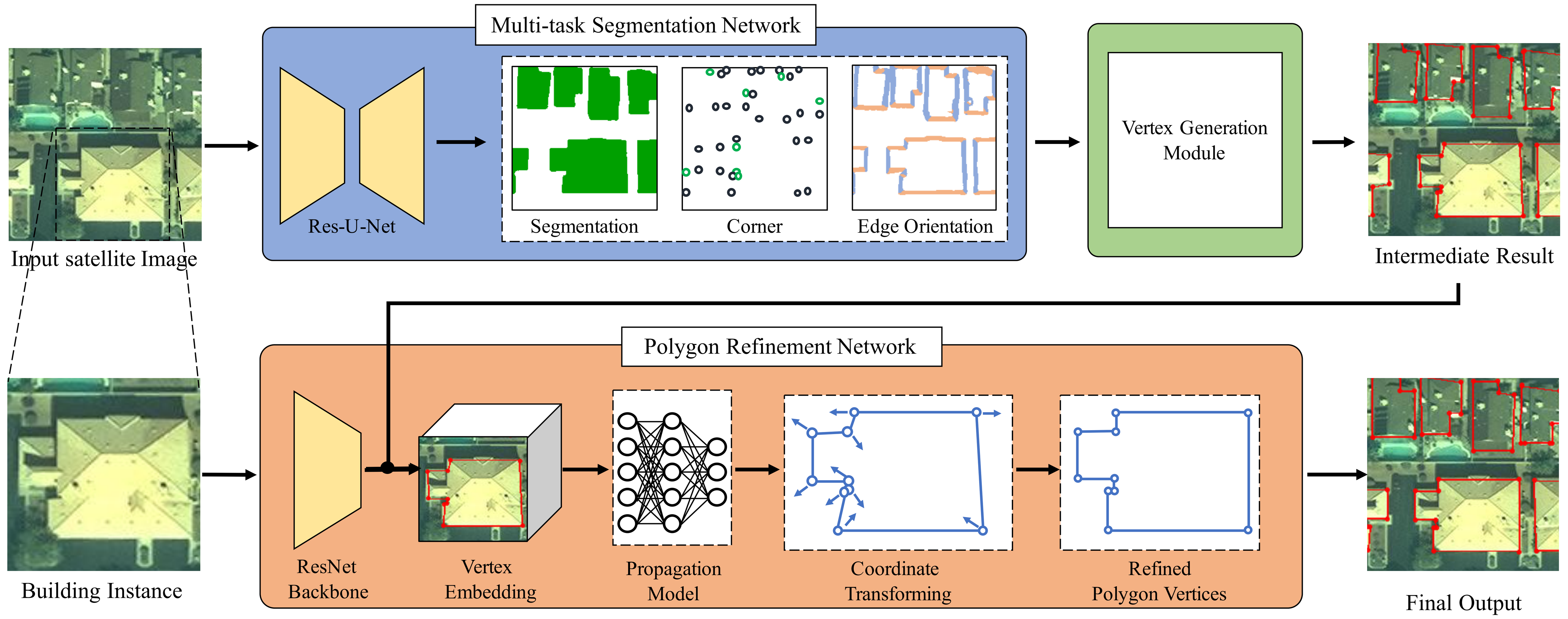
- Building footprint segmentation, multi-class corner prediction, and edge orientation prediction.
- Formulated as pixel-wise classification problems and trained jointly with the cross entropy loss.

Vertex generation module (VGM)

- The initial vertex set is obtained by densely extracting each pixel from the segmentation contour.
- The corner and edge orientation criterions are designed for selecting a set of valid vertices.

Polygon refinement network (PRN)

- GGNN-based model utilizes extra information e.g., the feature of each vertex and their relations.
- Predicting a displacement for each vertex to produce the final result with more accurate vertices.



Experimental result evaluation on building segmentation

- Evaluated on two popular datasets: (1) The CrowdAI mapping challenge dataset; (2) The Vegas dataset of the SpaceNet building dataset.
- For building segmentation results, our method improves the F1-score of current state-of-the-art by 1.5%, 0.4%, and 2.1% under different IoU thresholds.

Method	AP	AP_{50}	AP_{75}	AR	AR_{50}	AR_{75}	$F1$	$F1_{50}$	$F1_{75}$
Mask-RCNN (He et al. 2017)	41.9	67.5	48.8	47.6	70.8	55.5	44.6	69.1	51.9
PANet (Liu et al. 2018)	50.7	73.9	62.6	54.4	74.5	65.2	52.5	74.2	63.9
PolyMapper (Li et al. 2019)	55.7	86.0	65.1	62.1	88.6	71.4	58.7	87.3	68.1
FrameField (Girard et al. 2020)	50.5	76.6	59.3	55.3	78.1	64.0	52.8	77.3	61.6
ASIP (Li et al. 2020)	65.8	87.6	73.4	**78.7**	**94.3**	**86.1**	71.7	90.8	79.2
Ours	**73.8**	**92.0**	**81.9**	72.6	90.5	80.7	**73.2**	**91.2**	**81.3**

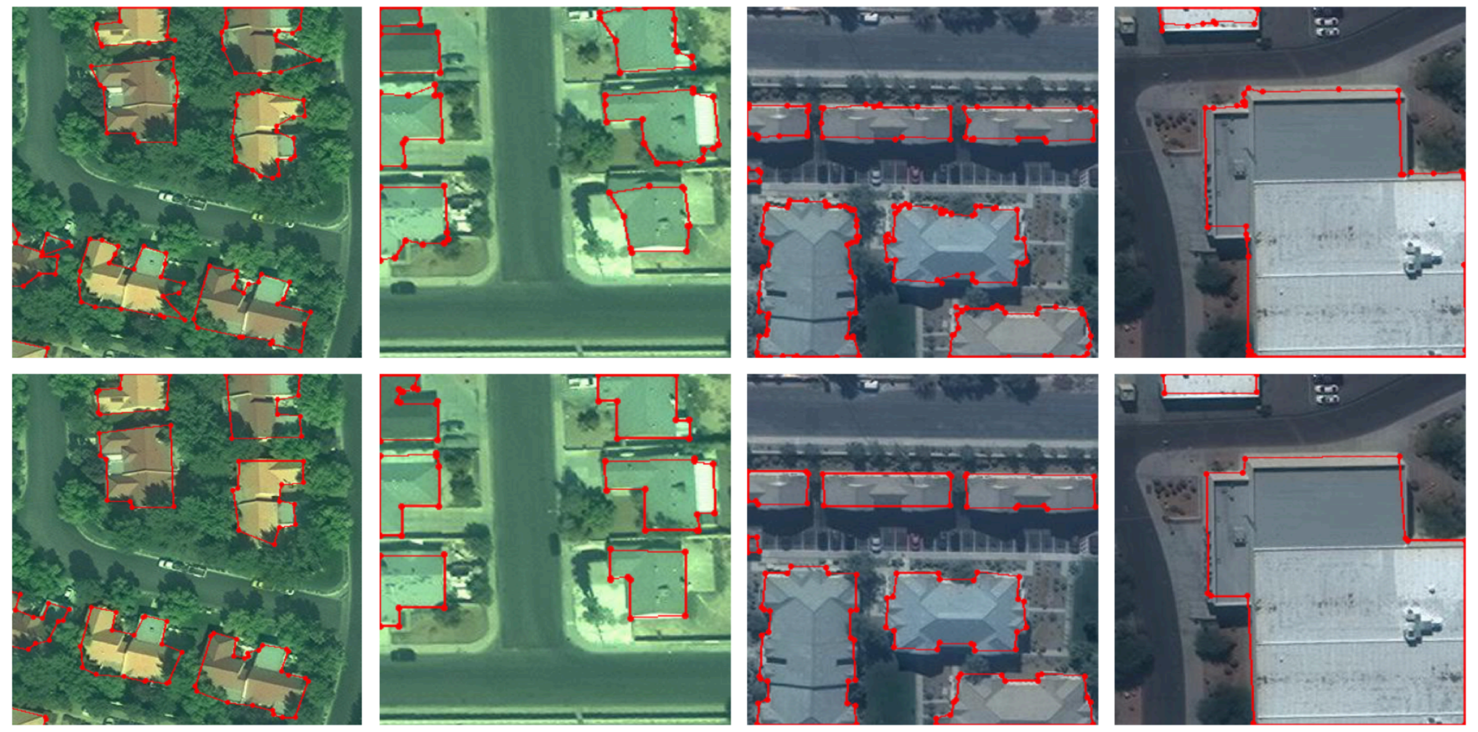
Results of vertex prediction and ablation study

- For vertex prediction results, our method achieves the F1-score gain of 6.64% and 6.82% compared with ASIP.
- Our method produces more accurate polygon vertices in terms of locations, quantities, angles, etc.
- The VGM produces better F1-score compared with the Baseline via filtering out the invalid vertices.
- The PRN further improves the vertex F1-scores by adjusting the vertices to more accurate locations.

	P_{3px}	R_{3px}	$F1_{3px}$	P_{5px}	R_{5px}	$F1_{5px}$
ASIP	51.13	**73.55**	60.32	69.25	**89.27**	78.00
Ours	**64.25**	69.90	**66.96**	**83.81**	85.85	**84.82**

	P_{3px}	R_{3px}	$F1_{3px}$	P_{5px}	R_{5px}	$F1_{5px}$
Baseline	51.67	50.24	50.94	76.66	74.31	75.47
+ VGM	56.71	52.53	54.54	81.54	75.22	78.25
+ PRN	**69.50**	**61.70**	**65.37**	**86.54**	**76.56**	**81.24**

Qualitative comparison of ASIP (top) and our method (bottom)



An example of our prediction results at different stages

Baseline   + VGM   + PRN