Joint Semantic-Geometric Learning for Polygonal Building Segmentation


Weijia Li


Wenqian Zhao


Huaping Zhong


Conghui He


Dahua Lin

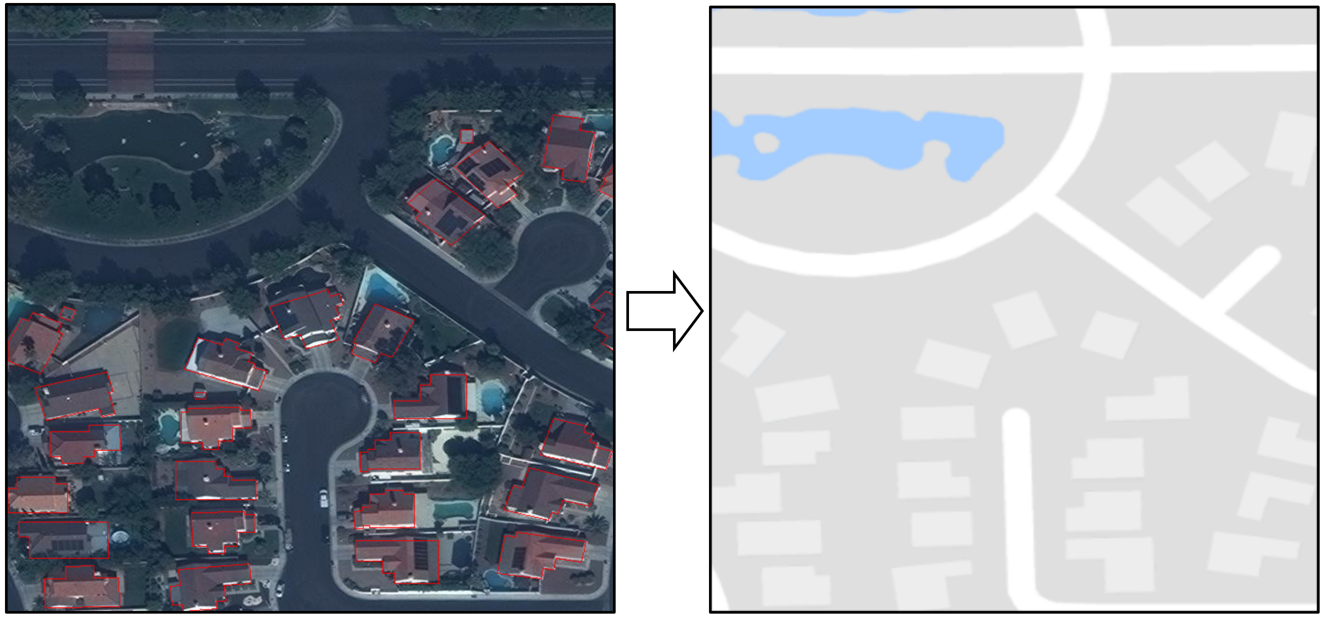
香港中文大學
The Chinese University of Hong Kong

商汤
sensetime

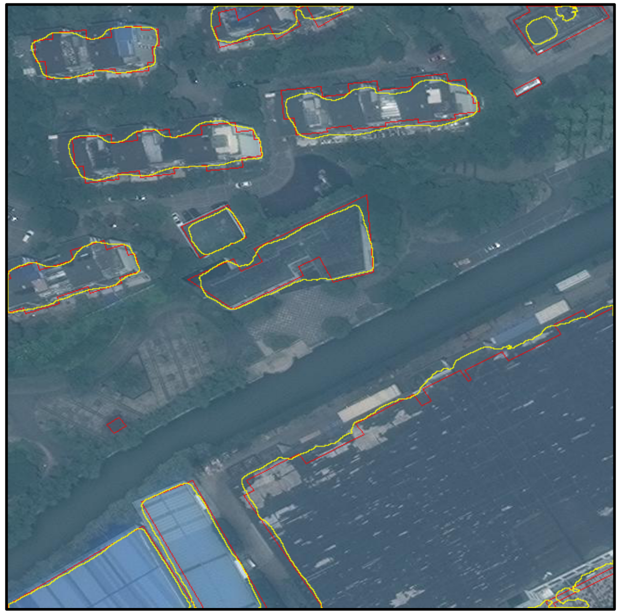
Polygonal Building Segmentation from Remote Sensing Images

- Building footprint extraction from aerial or satellite images
  - An important and popular research topic in both remote sensing and computer vision domains for decades.
  - A fundamental task for urban planning, disaster management, geographical information updating, etc.

- Raster vs. Vector prediction
  - Pixel-wise segmentation models produce building segmentation map in raster format.
  - polygonal building segmentation approaches produce more realistic building polygons that are in the desirable vector format for practical applications.

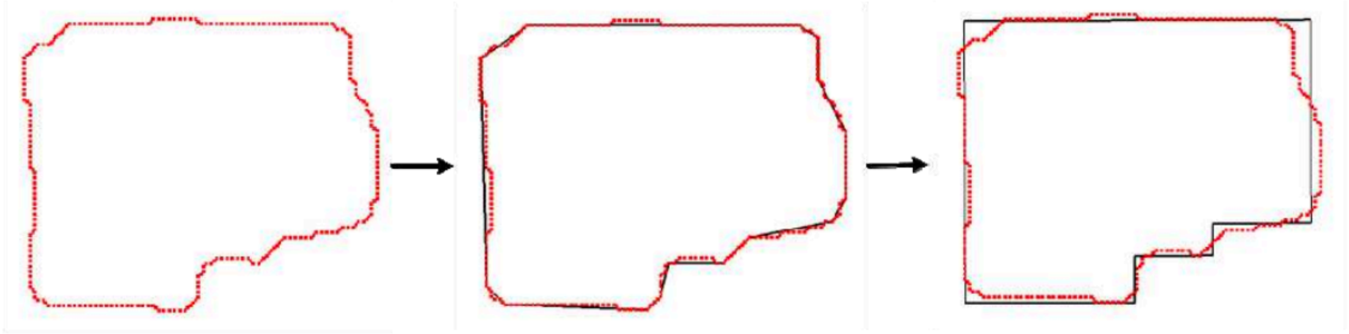


Building footprint extraction from aerial or satellite images



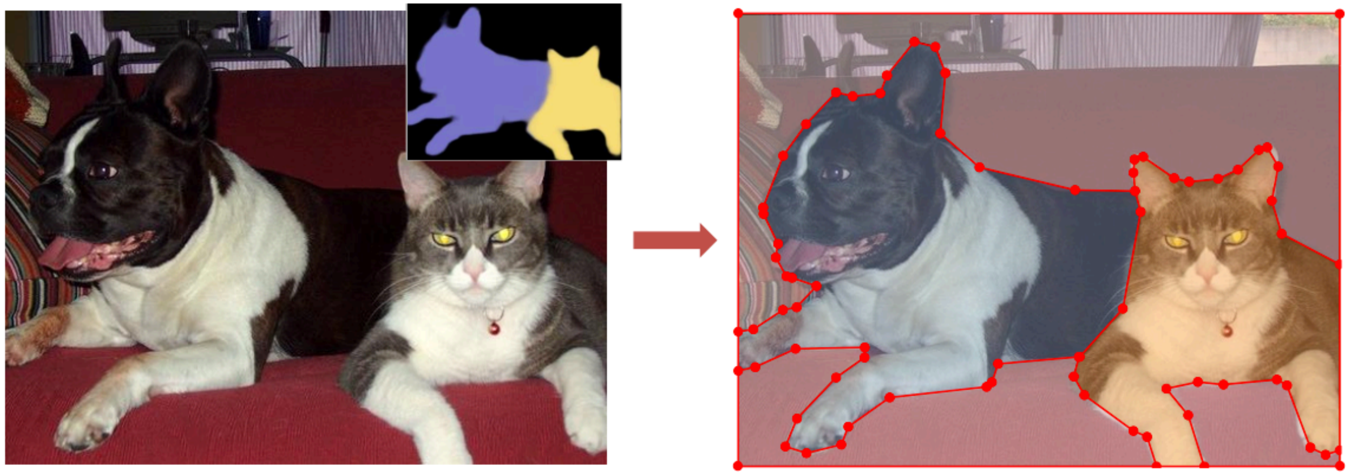
Raster vs. vector format

Related Work: Post-processing based Polygonization



(a) (b) (c)

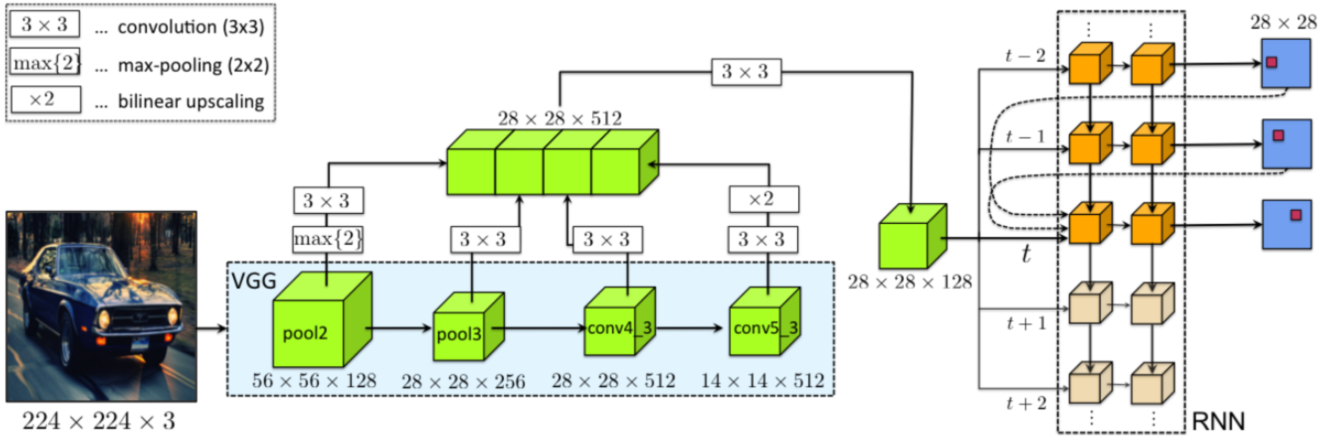
- Zhao et al. (CVPRW 2018) proposed a multi-step boundary regularization method to simplify the building segmentation results predicted from Mask-RCNN.



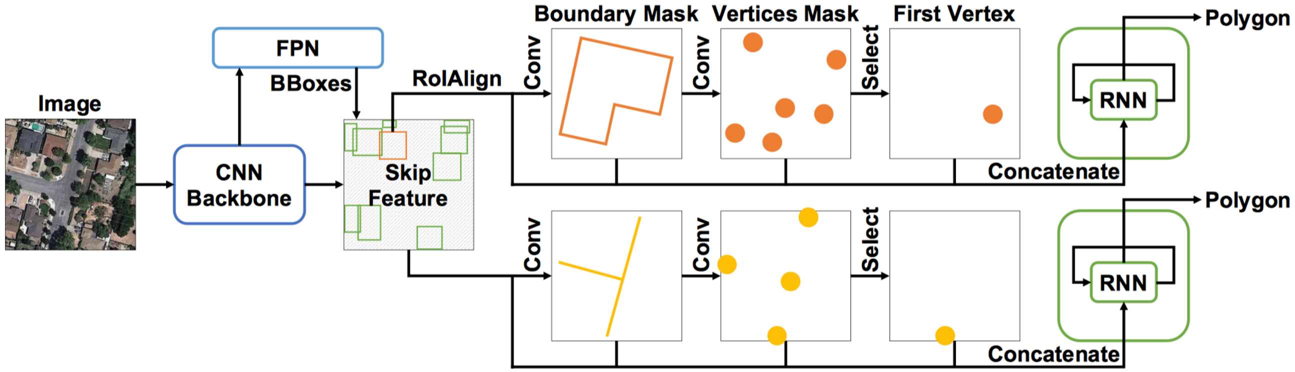
- Li et al. (CVPR 2020) proposed a polygonal partition refinement method for vectorizing the output probability maps of a U-Net based model.

These post-processing based methods not only require a complex processing procedure, but also a perfect segmentation map to ensure the quality of the polygonization results.

Related Work: Deep Neural Network based Vertex Prediction



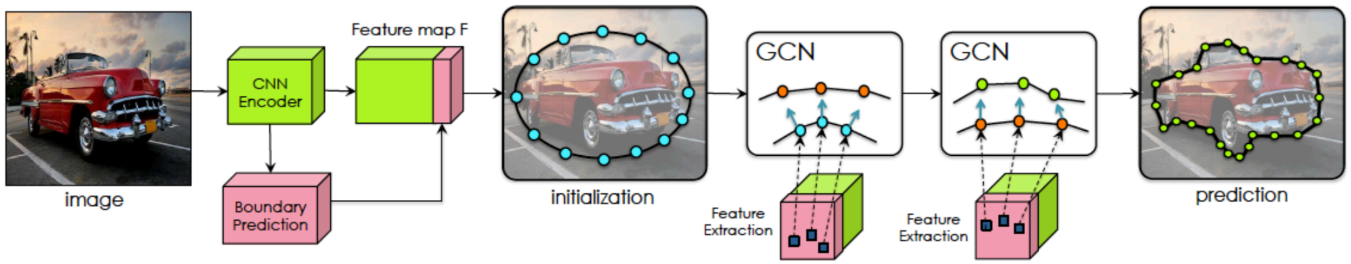
- Castrejon et al. (CVPR 2017) proposed a polygonal annotation method (**Polygon-RNN**) that directly predicts a polygon vertex at each time step using a CNN-RNN architecture.
- Acuna et al. (CVPR 2018) proposed **Polygon-RNN++** that further improved Polygon-RNN to achieve better performance.



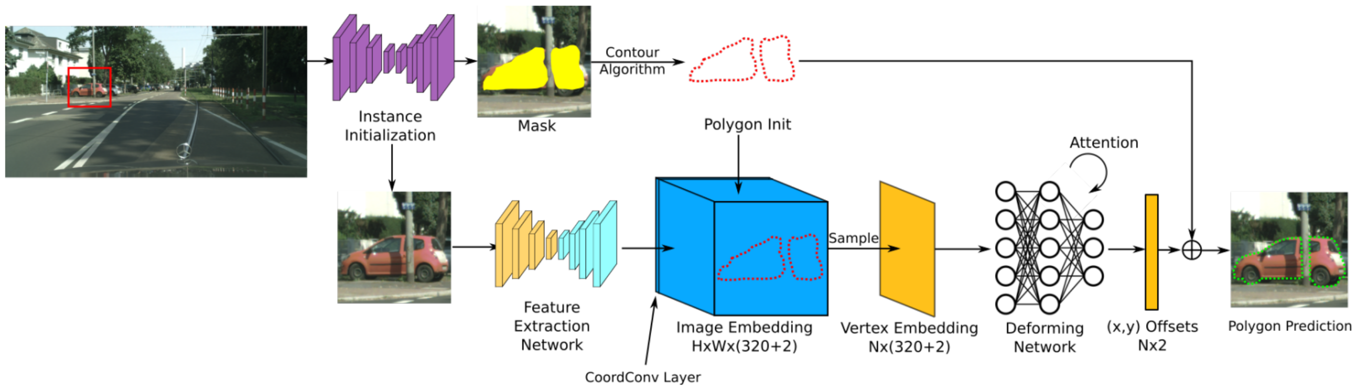
- Li et al. (ICCV 2019) proposed **PolyMapper** that extended Polygon-RNN for automatic building segmentation task.

These RNN-based methods usually achieve desirable prediction results for buildings with simple shapes. However, the sequential manner of the recurrent model limits its capability of correctly predicting vertices for complex buildings, producing vertices with wrong sequential order and self-intersections.

Related Work: Deep Neural Network based Vertex Prediction



- Ling et al. (CVPR 2019) proposed **Curve-GCN** that represents an object as a graph with a fixed number of vertex, and predicts an offset for each vertex simultaneously in a regression manner.



- Liang et al. (CVPR 2019) proposed **PolyTransform** that selects the initial vertices uniformly from a segmentation mask contour via a given distance and predicts their offsets simultaneously.

These methods usually generate over redundant vertices for buildings with simple shape and insufficient vertices for buildings with complex contour.

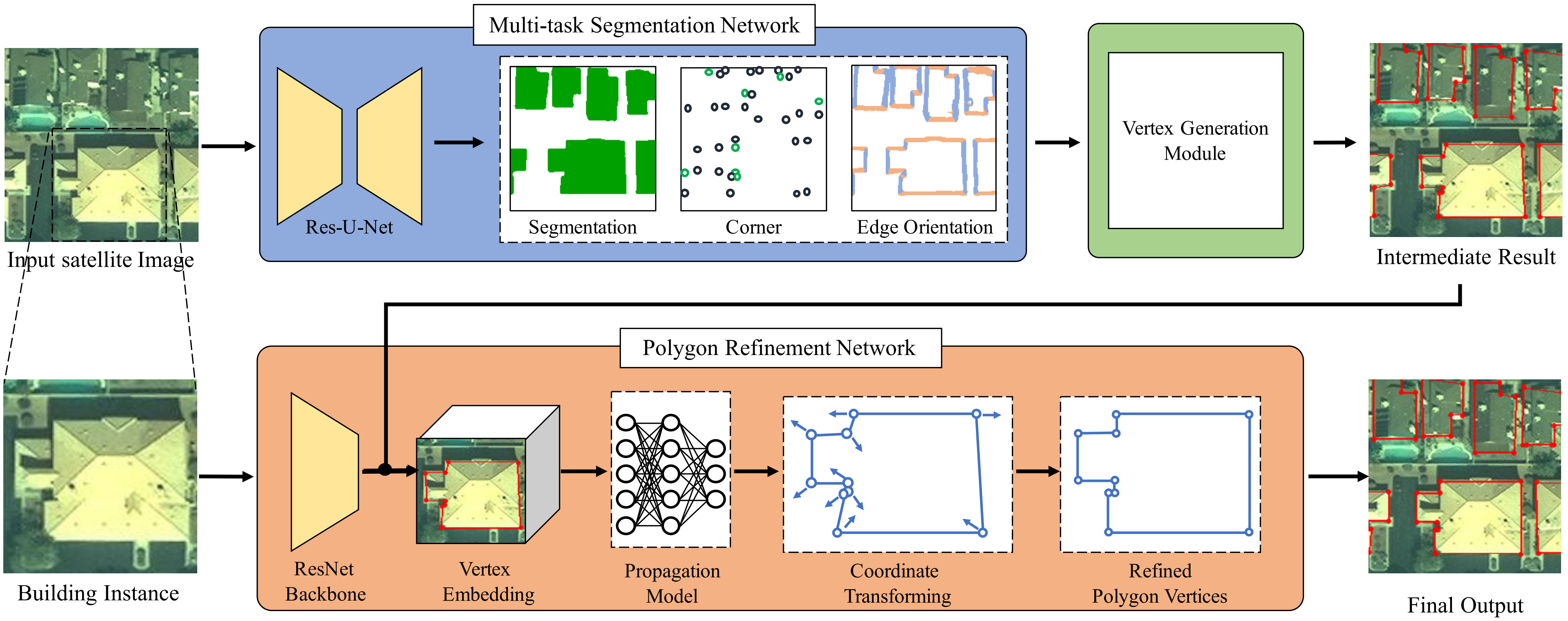
Our Work

In summary, state-of-the-art polygonal segmentation methods suffer from several limitations, e.g., (1) relying on a perfect segmentation map to guarantee the vectorization quality; (2) requiring a complex post-processing procedure; (3) generating inaccurate vertices with a fixed quantity, a wrong sequential order, self-intersections etc.

In this work, we propose a novel polygonal building segmentation approach to tackle the above limitations and make the following contributions:

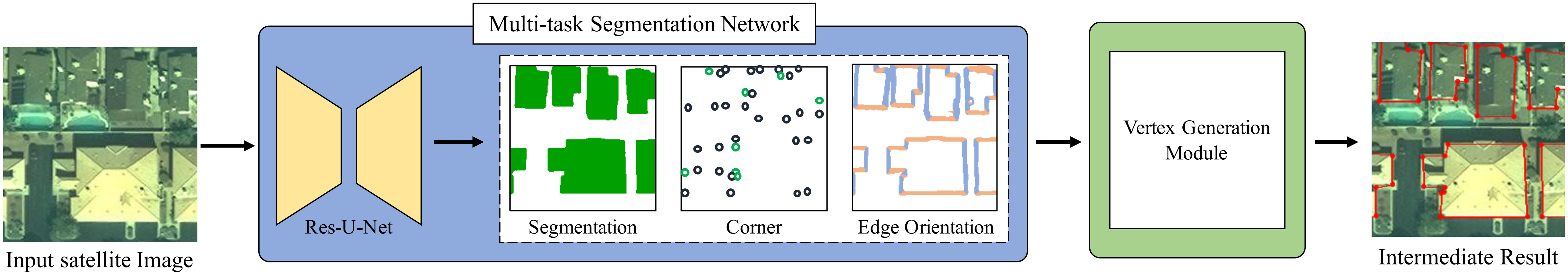
1. A multi-task segmentation network for joint semantic and geometric learning via three tasks, i.e., pixel-wise building segmentation, multi-class corner prediction, and edge orientation prediction.
2. A simple but effective vertex generation module for transforming the segmentation contour into valid polygon vertices.
3. A polygon refinement network that automatically moves the polygon vertices into more accurate locations.

Methods: Framework Overview



1. Multi-task segmentation network: building segmentation, multi-class corner prediction, and edge orientation prediction.
2. Vertex generation module: converting the former three types of outputs into a set of valid polygon vertices.
3. Polygon refinement network: predicts a displacement for each vertex and produces the final building polygons.

Methods: Multi-task Segmentation Network



Representation of multi-class corners

- Convex corner: the interior angle of a vertex is smaller than 180° (denoted by black circles).
- Concave corner: the interior angle of a vertex is larger than 180° (denoted by green circles).
- Background: pixels that are not on the polygon vertices.

Representation of edge orientations

- Pixels on the building edges: obtained via discretizing the orientation angle of the edge into a class.
- Pixels at the building corners: randomly assigned with the one of its neighbor pixel of an edge.
- Pixels not on any edges or corners: defined as zero.

Multi-task Network Training

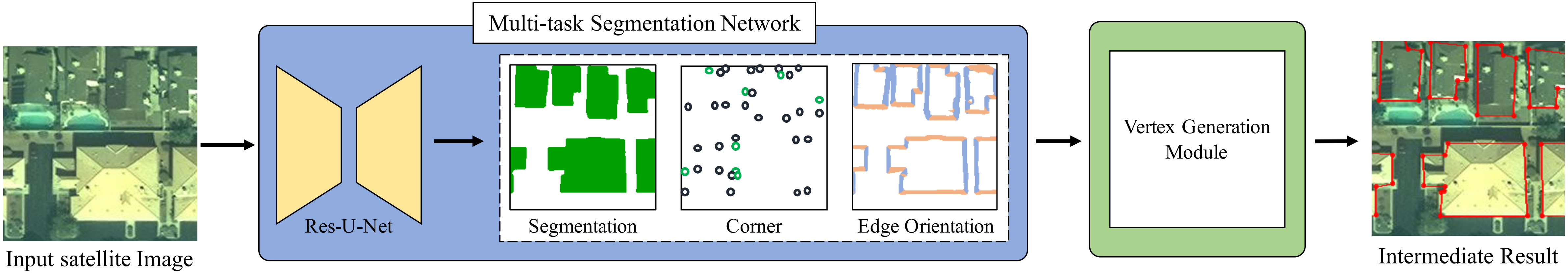
- Based on Res-U-Net, which achieved promising performance in many building segmentation challenges and studies.
- The three tasks are all formulated as pixel-wise classification issues and trained jointly with the cross entropy loss:

$$L = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \times log(p(y_{i,c}))$$

- The total loss of the three tasks can be summarized as:

$$L_{total} = \lambda_1 L_{seg} + \lambda_2 L_{corner} + \lambda_3 L_{orient}$$

Methods: Vertex Generation Module



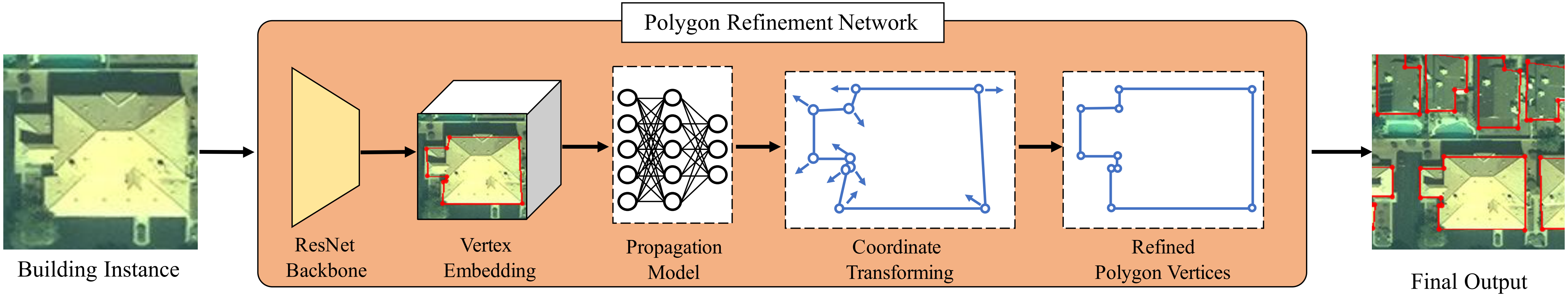
Initial vertex set: For each predicted building instance, the building segmentation map is converted to a mask contour with a width of one pixel. Each pixel on the mask contour is extracted via dense sampling in an anticlockwise order, constituting the initial vertex set.

Corner criterion: Vertices with a corner prediction probability smaller than T_{cor} are removed from the initial vertex candidates. Each group of adjacent vertices are further converted into one valid vertex, i.e., the local maximum of the corner prediction probability.

Edge orientation criterion: We calculate the absolute difference of the orientation angle between two neighboring vertices for each initial vertex candidate. The vertices with an absolute difference greater than T_{ori} are selected as the valid vertices.

Output valid vertex set: The valid vertices selected by corner and edge orientation criterions are combined into a union vertex set. Each group of adjacent vertices are further merged into one vertex, constituting the final output of the vertex generation module.

Methods: Polygon Refinement Network



Backbone: The ResNet backbone is served as an encoder for extracting features from the input image that is cropped by single building instance, producing feature map that will be further used for vertex embedding.

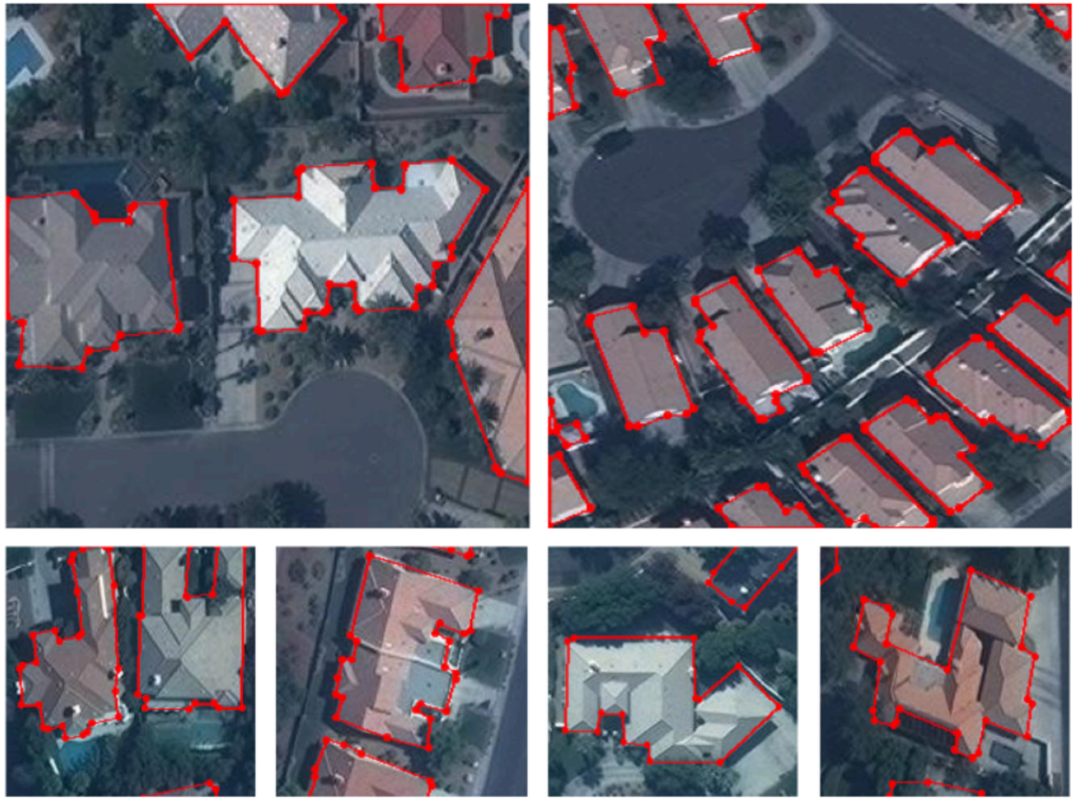
Vertex Embedding: The valid vertices obtained from VGM are transformed accordingly for vertex embedding on the final feature map of the backbone (the red points on the cube). Each vertex is assigned with the features extracted from the channel direction.

Propagation model: The GGNN-based Propagation model is able to utilize the extra information such as the feature of each node (vertex) and their relations, which is designed to learn the relative displacement between a predicted valid vertex and its nearest GT vertex.

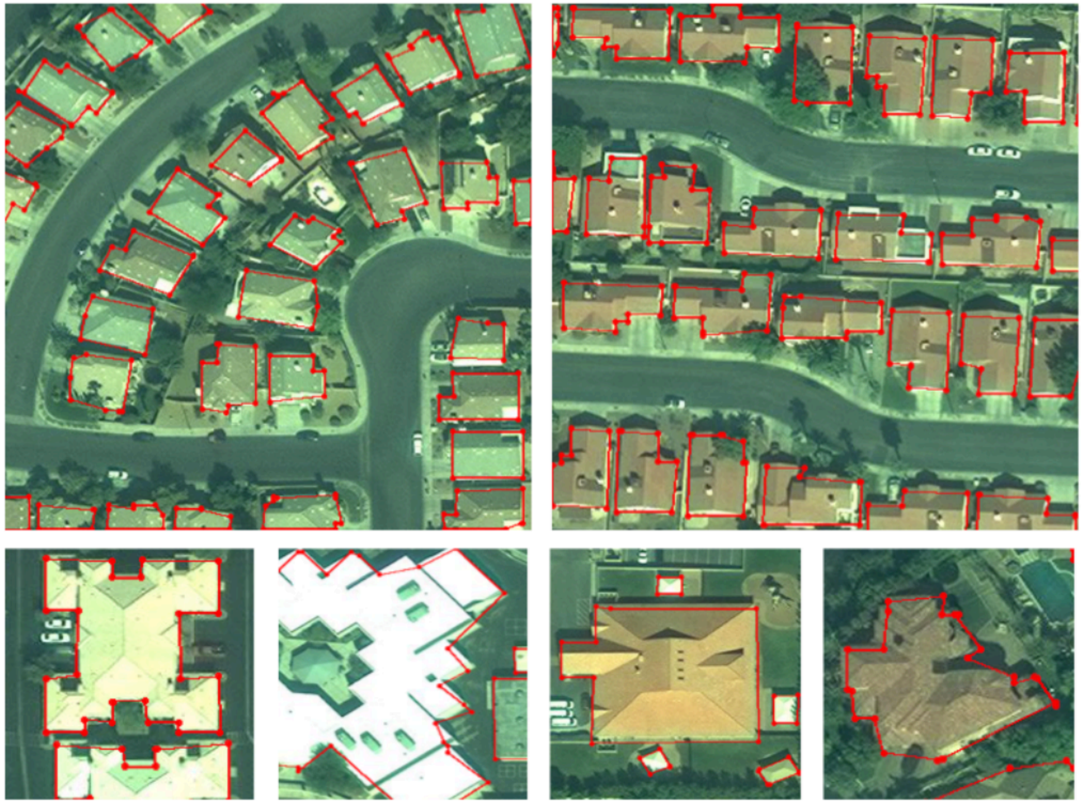
Coordinate transforming: The predicted displacement classes are converted to the displacement coordinates, and added to the corresponding vertex coordinates of VGM to obtain the final building polygons.

Results: Qualitative Results on Two Datasets

- Our method is evaluated on two popular building datasets: (1) CrowdAI mapping challenge dataset, including around 3 million buildings. (2) Vegas dataset of SpaceNet building dataset, including around 108,000 buildings.
- Our method produces vectorized outputs with accurate vertices and edges, even for buildings with complex shapes.



CrowdAI mapping challenge dataset (CrowdAI)



Vegas dataset of SpaceNet building dataset (Vegas)

Results: Comparison with state-of-the-art

Quantitative comparison on CrowdAI dataset.

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**

Quantitative comparison on Vegas dataset.

Method	U-Net	Mask-RCNN	Zhao et al.	Ours
F1-score	88.5	88.1	87.9	**89.4**

Comparison of our method and ASIP on vertex prediction scores

	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**

- 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.
- For vertex prediction results, our method achieves the F1-score gain of 6.64% and 6.82% compared with ASIP.

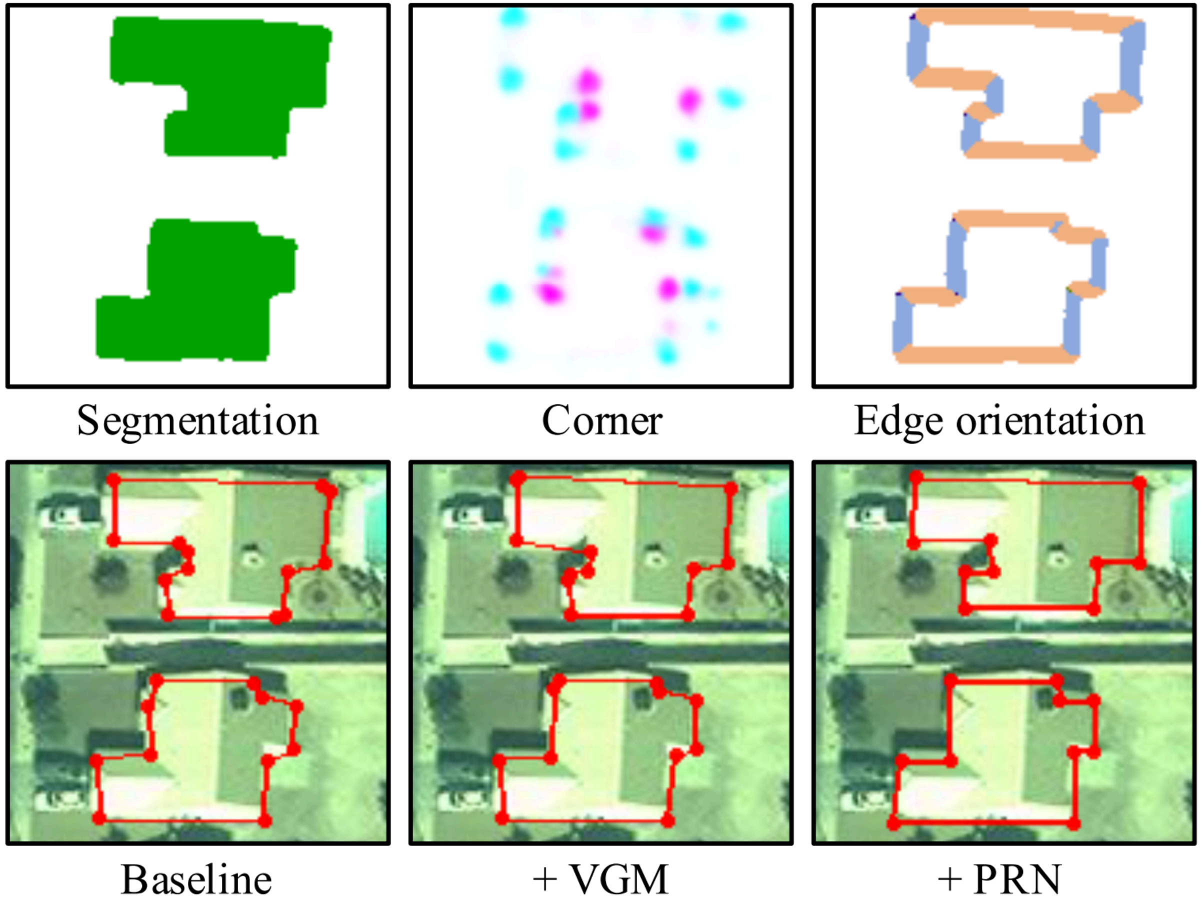
Results: Comparison with state-of-the-art

Qualitative comparison of the predictions of ASIP and our method. The building polygons predicted by our method have more accurate vertices in terms of locations, quantities, and angles.

Results: Failure Case Analysis

Three typical examples of failure cases of our method. Our method has difficulties in producing accurate polygons for buildings that are seriously sheltered by trees (left), buildings with multiple extremely short edges (middle), and high-rise buildings with serious parallax effect (right), which should be explored and solved in our future work.

Results: Ablation Study



Results of ablation study on vertex prediction scores

	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**

- The vertex generation module produces much better vertex prediction results compared with the Baseline, via effectively utilizing the corner and edge orientation predictions to filter out invalid vertices and remain the valid vertices with accurate quantity.
- The polygon refinement network further improves the vertex prediction F1-scores by adjusting the vertices to more accurate locations.

Summary

- In this work, we have presented a novel building segmentation approach that is capable of producing vector building polygons from remote sensing images.
- Qualitative and quantitative evaluations on two popular building segmentation datasets demonstrate that our proposed approach achieves significant improvements over state-of-the-art methods. The effect of each component of our approach is also verified in the ablation study.
- We believe that this paper motivates novel ideas for predicting vectorized object representations and provides effective solutions for practical applications in Geographic Information Systems.
- In our future work, we would like to explore novel methods for more complex application scenarios, such as producing the vectorized roof and footprint polygons for highrise and dense buildings.

Thank you!