2021 ICCV OCTOBER 11-17 VIRTUAL

3D Building Reconstruction from Monocular Remote Sensing Images






Weijia Li[1,2] Lingxuan Meng[2,3] Jinwang Wang[2,4] Conghui He[2] Gui-Song Xia[4] Dahua Lin[1,5]

[1]CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong, [2]SenseTime Research
[3]University of Electronic Science and Technology of China, [4]Wuhan University, [5]Shanghai AI Laboratory

Background

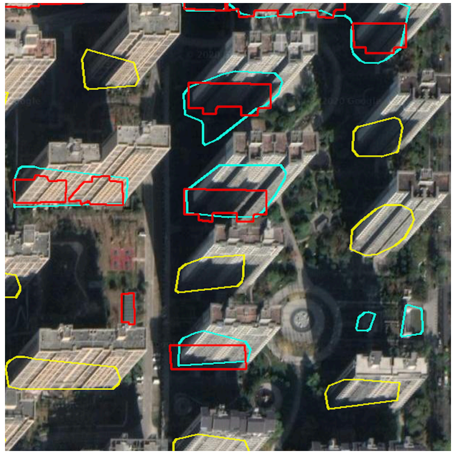
- An important research problem and an economic solution to large-scale city modeling, compared with using LiDAR data or multi-view imagery.
- Challenges of monocular 3D building reconstruction:
  - The partial invisibility of building footprints and facades
  - The serious shadow effect
  - The extreme variance of building height in large-scale areas

3D Building Reconstruction from Monocular Remote Sensing Images

Related Work

Low-rise buildings

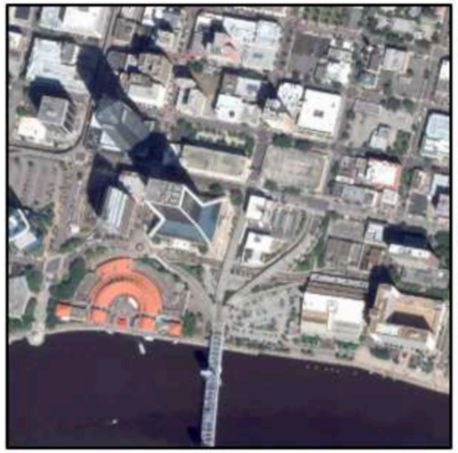
High-rise buildings



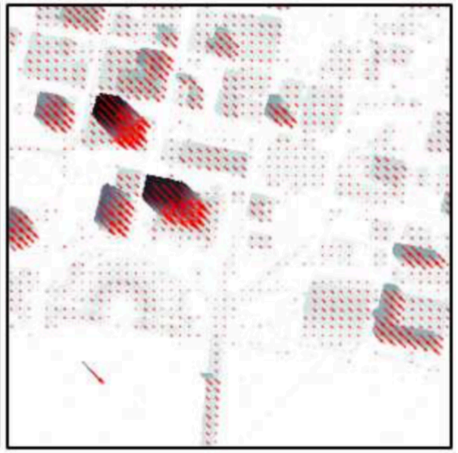
Building footprint extraction

- Regarding different parts of a building instance as a unified entirety.
- Achieving satisfying results for low-rise buildings.
- Producing poor segmentation boundaries when extracting high-rise buildings from oblique images.

Oblique image



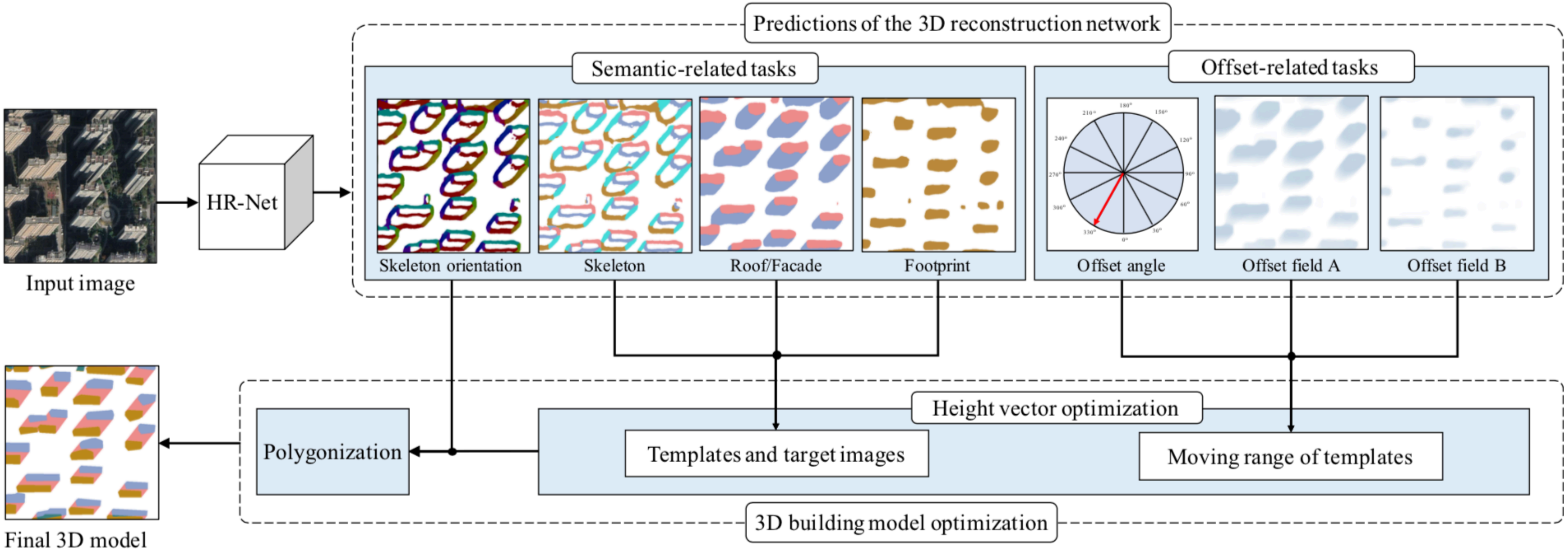
Flow vector (height)



Building height estimation and 3D reconstruction

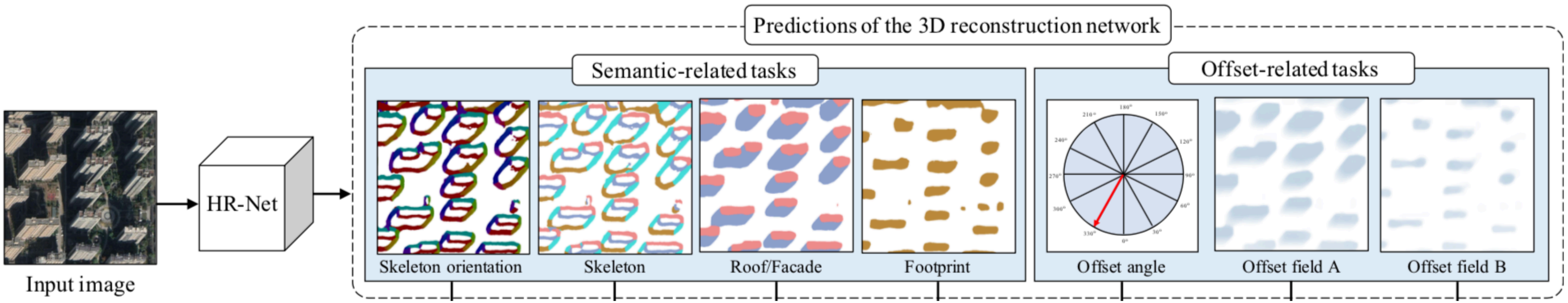
- Most existing methods focus on height estimation from near-nadir images.
- A recent study proposed a method for building height estimation from oblique images.
- It only focus on the single-task height estimation instead of 3D reconstruction.

Methods



Predictions of the 3D reconstruction network
Semantic-related tasks
Offset-related tasks
Input image
HR-Net
Skeleton orientation
Skeleton
Roof/Facade
Footprint
Offset angle
Offset field A
Offset field B
Height vector optimization
Templates and target images
Moving range of templates
Polygonization
Final 3D model
3D building model optimization

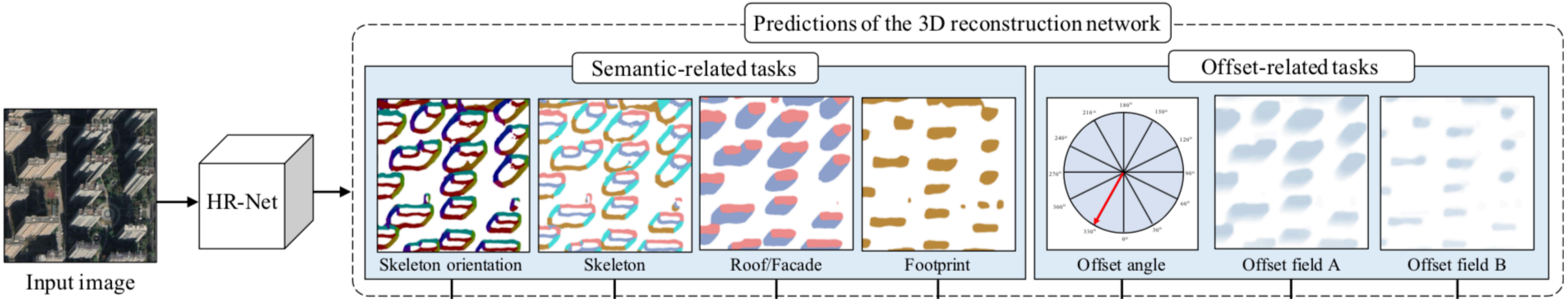
Methods: Task definition of MTBR-Net



- **Roof and Facade** have complete contours on the monocular remote sensing images.
- **Footprint** is partially invisible but has the same shape as the corresponding roof.
- **Building Skeleton** includes 4 semantic edges of a 3D building model.
- **Skeleton Orientation** is designed for converting the raster map into vector 3D model.

- **Offset Field A** is defined as 2 pixel-wise values (x, y) located in roof and facade regions.
- **Offset Field B** is defined as 2 pixel-wise values (x, y) located in footprint regions.
- **Offset Angle** is an absolute value for each image.

Methods: Training of MTBR-Net

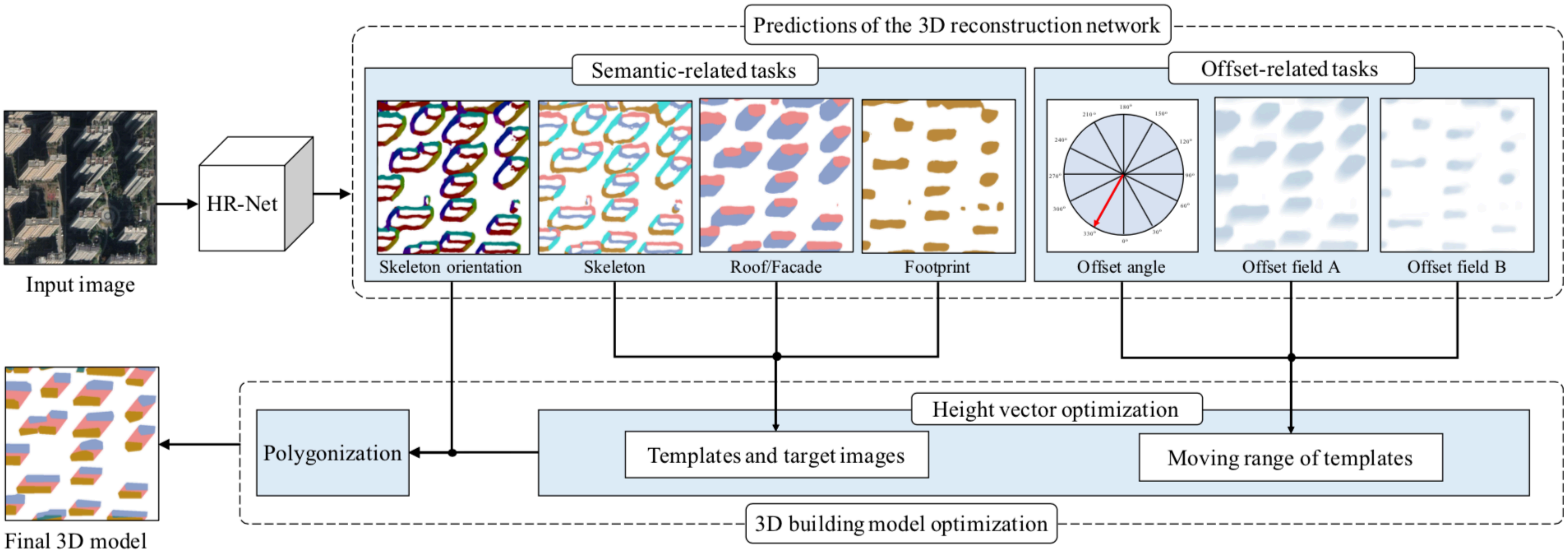


- **The four semantic-related tasks** are all formulated as pixel-wise classification problems and trained with the cross entropy loss.
- **The two offset field prediction tasks** are defined as pixel-wise regression tasks, of which the loss is calculated by the endpoint error (EPE).
- **The image-wise offset angle prediction task** is trained with the cross entropy loss.
- The total loss of the seven tasks is the weighted sum of each task-specific loss.

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

$$\mathcal{L}_{field} = \frac{1}{N}\sum_{i=1}^{N} ||\vec{O}_i^{pred} - \vec{O}_i^{gt}||_2$$

Methods: Optimization of the 3D building model



- **Height vector optimization:** a prior knowledge based template matching method for optimizing the height estimation results.
- **Polygonization:** a skeleton orientation based method for converting the raster results into vector 3D building model with valid shapes.

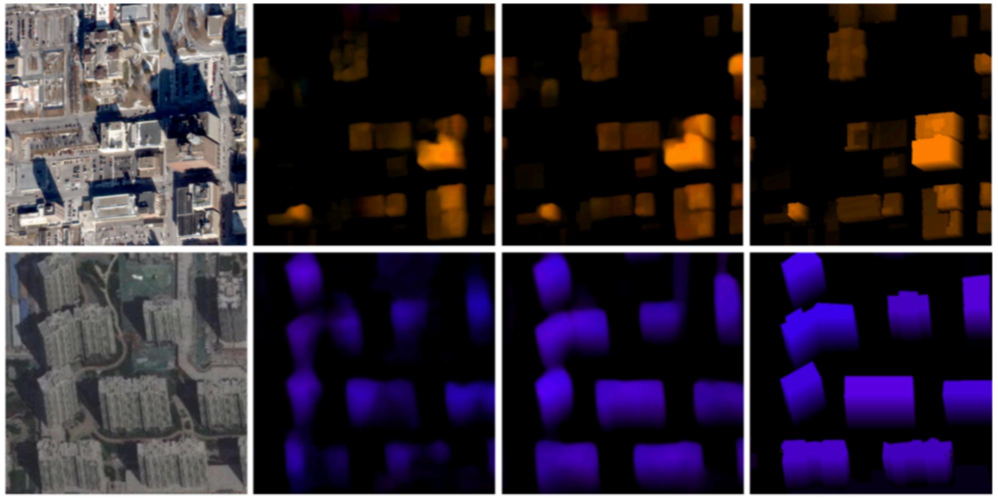
Results: Height estimation performance

Table 1. Comparison of building height estimation on our proposed dataset. We report the EPE of the roof instances within different height range and the average EPE of all instances. Our method reduces the EPE of high-rise buildings by 5 to 24 pixels compared with [7].

Dataset	Method	EPE of different height range (in pixels)											Average EPE
		0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	>100	
In-domain	Christie et al. [7]	6.22	5.26	7.04	9.01	10.94	12.52	14.89	19.47	24.50	73.07	50.41	6.19
	Ours	**4.92**	**4.24**	**6.02**	**5.91**	**6.87**	**7.82**	**8.39**	**12.45**	**20.75**	**61.41**	**26.69**	**4.88**
Out-domain	Christie et al. [7]	7.99	**9.83**	9.81	10.41	13.31	16.11	19.41	24.13	21.27	26.17	75.21	12.31
	Ours	**6.63**	9.96	**8.33**	**8.56**	**9.32**	**9.45**	**12.55**	**15.75**	**10.76**	**11.82**	**52.52**	**9.59**

Table 2. Comparison of building height estimation on DFC19 dataset, in terms of the MAE and RMSE of actual height (in meters), EPE of offset vector (in pixels), and angle error (in degrees).

Method	Actual Height		Offset EPE		Angle Error
	MAE	RMSE	Roof	Facade	
Kunwar [16]	8.33	19.65	-	-	-
Zheng et al. [34]	8.72	19.32	-	-	-
Christie et al. [7]	7.73	16.87	5.44	7.11	15.09
Ours	**4.75**	**9.57**	**4.67**	**5.35**	**8.40**

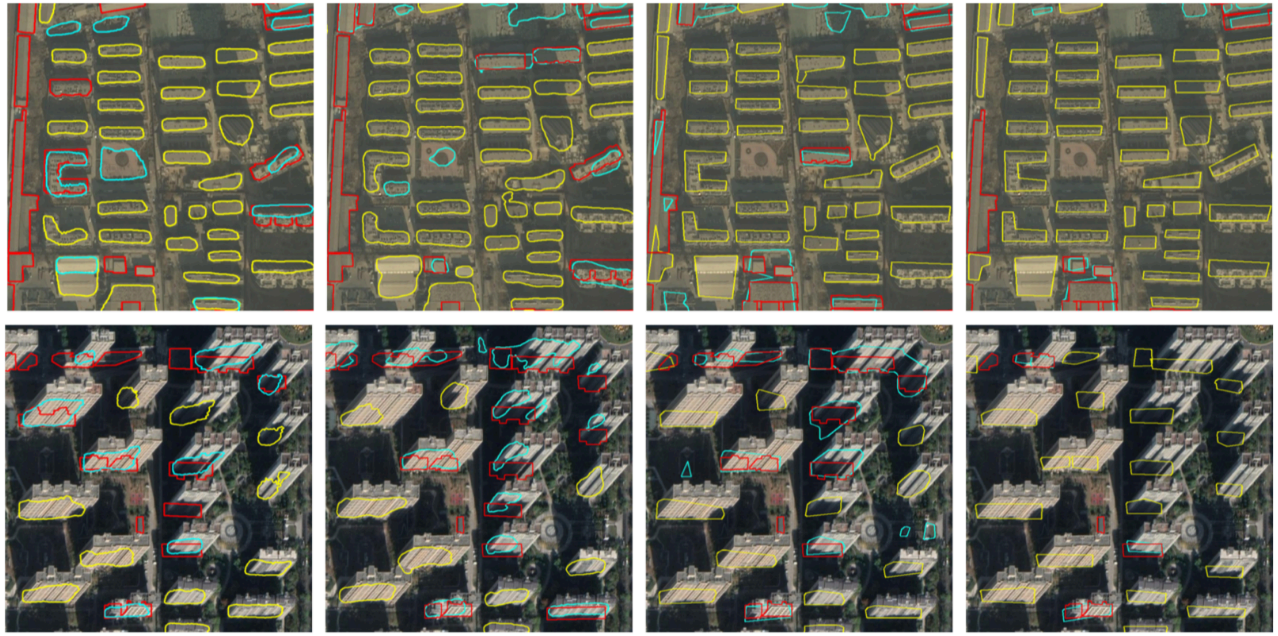


Input image Christie et al. [7] Ours Ground truth

The proposed dataset will be released on https://liweijia.github.io/projects/building_3d/.

Results: Building segmentation performance

Method	In-domain dataset (Roof)			In-domain dataset (Footprint)			Out-domain dataset (Roof)			Out-domain dataset (Footprint)		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Cascade Mask R-CNN [4]	66.68	67.06	66.87	61.27	61.48	61.37	48.39	48.74	48.56	40.73	39.31	40.00
Mask R-CNN [13]	67.98	69.35	68.66	63.43	63.85	63.64	59.65	52.09	55.62	50.30	41.29	45.35
PANet [21]	68.38	67.98	68.18	64.03	61.91	62.95	62.11	50.46	55.68	52.54	41.03	46.08
HR-Net [30]	68.78	66.09	67.41	64.19	64.29	64.24	55.76	46.62	50.78	41.95	35.06	38.20
Li et al. [18]	71.76	69.25	70.48	65.71	66.37	66.04	60.44	56.40	58.35	49.69	45.77	47.65
Ours (w/o optimization)	**72.72**	**71.37**	**72.04**	66.85	**68.05**	67.44	**65.20**	**57.97**	**61.37**	54.34	46.37	50.04
Ours (w/ optimization)	**72.72**	**71.37**	**72.04**	**69.47**	67.71	**68.58**	**65.20**	**57.97**	**61.37**	**56.45**	**48.17**	**51.98**



Mask R-CNN [13] HR-Net [30] Li et al. [18] Ours

- We compare the segmentation performance of our approach with the current SOTA and 4 other competitive segmentation methods.
- Our method improves the F1-score of SOTA by 2.5% and 4.3% for footprint extraction, indicating the effectiveness of warping the predicted roof instances to footprints using the offset vectors.

Conclusions

- We have presented a novel 3D building reconstruction method that produces vector 3D building model with accurate roof, facade, footprint, and height,
- Qualitative and quantitative evaluations demonstrate the significant performance gain over current state-of-the-art.
- To the best of our knowledge, this is the first work that produces vector 3D building model reconstruction results from monocular remote sensing images using deep neural networks.
- We believe that this paper provides effective solutions for 3D building reconstruction in large-scale and complex application scenes.

Thank you!