JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

OmniCity: Omnipotent City Understanding with Multi-level and Multi-view Images

Poster: THU-AM-088

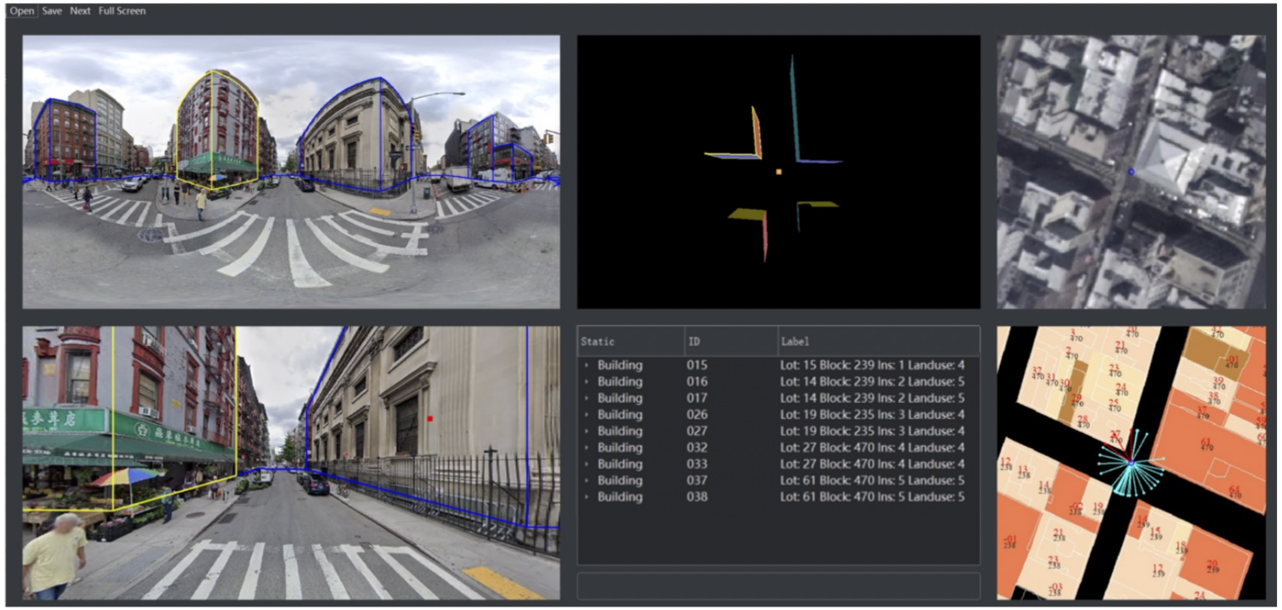
Weijia Li[1] Yawen Lai[2] Linning Xu[3] Yuanbo Xiangli[3] Jinhua Yu[1]
Conghui He[2,4] Gui-Song Xia[5] Dahua Lin[3,4]

[1]Sun Yat-sen University [2]Sensetime Research [3]The Chinese University of Hong Kong
[4]Shanghai AI Laboratory [5]Wuhan University

Preview of OmniCity

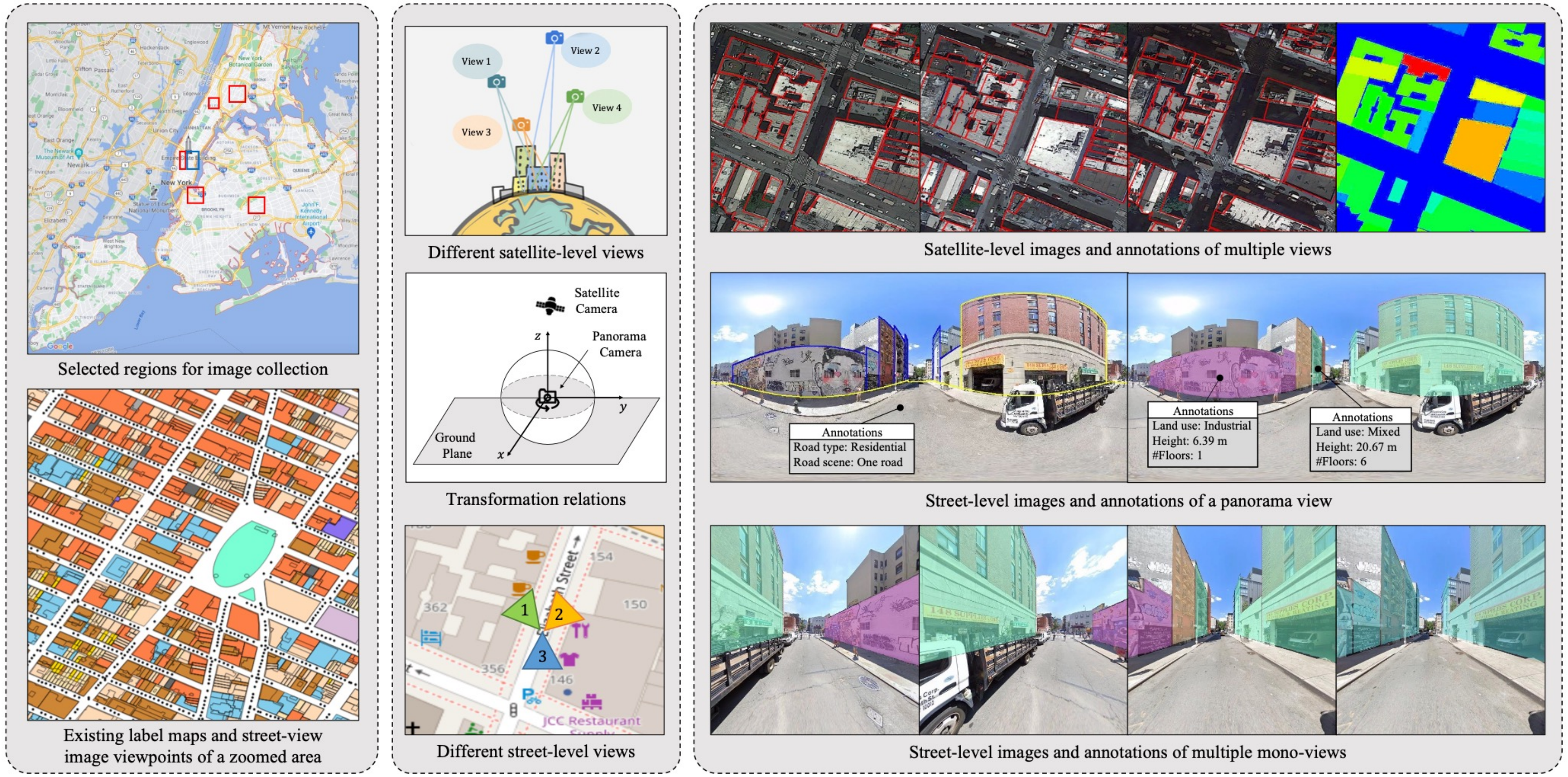
- We construct a dataset that contains multi-view satellite and street-level images, with a larger quantity, richer annotations and more views compared with existing datasets.
- We develop an efficient street-level image annotation pipeline that leverages the existing label maps of satellite view and the transformation relations between different views (satellite-panorama-monoview).

Dataset	#Images	Street	Sate.	Anno.	Attri.	Height
KITTI [14]	15,000	mono	-	semantic	×	×
Cityscapes [10]	25,000	mono	-	semantic	×	×
EuroCity [3]	47,300	mono	-	bbox	×	×
WildPASS [42]	500	multi.	-	semantic	×	×
PASS [41]	400	multi.	-	semantic	×	×
HoliCity [47]	6,300	multi.	-	inst./plane	×	×
SkyScapes [1]	8,820	-	single	semantic	×	×
SpaceNet [38]	60,000	-	multi.	instance	×	×
Christie et al. [9]	11,000	-	single	semantic	×	✓
Li et al. [21]	3,300	-	single	instance	×	✓
TorontoCity [36]	Unknow	multi.	multi.	instance	×	✓
Wojna et al. [39]	49,426	mono	single	image	✓	×
OmniCity	**108,600**	**multi.**	**multi.**	**inst./plane**	✓	✓



Preview of OmniCity

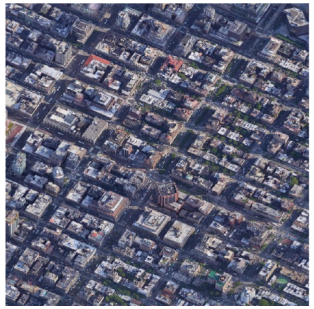
- We conduct a series of benchmark experiments for multiple tasks and data sources, and analyze the limitations of the current benchmarks on OmniCity.
- We provide new problem settings for existing tasks, such as cross-view image matching, synthesis, segmentation, detection, etc., and facilitate new methods and tasks for large-scale city understanding, reconstruction, and simulation.



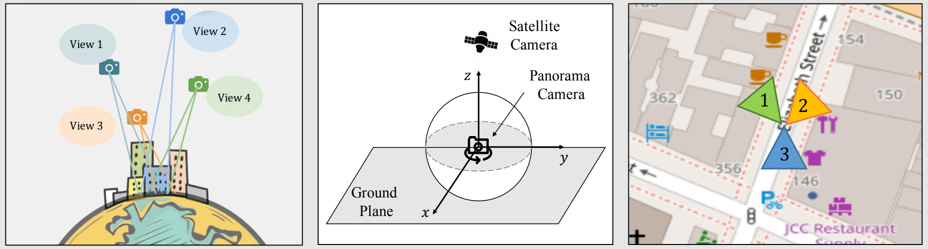
Background and Motivation

Street-level images and datasets
- Image: rich semantic information (e.g., building facade)
- Distribution: sparsely, unevenly, locally distributed
- Annotation: requires extensive human annotation efforts, without fine-grained semantic labels or at only image level


Satellite-level images and datasets
- Image: limited semantic information (e.g., building roof)
- Distribution: densely and globally distributed
- Annotation: well-aligned with existing label maps

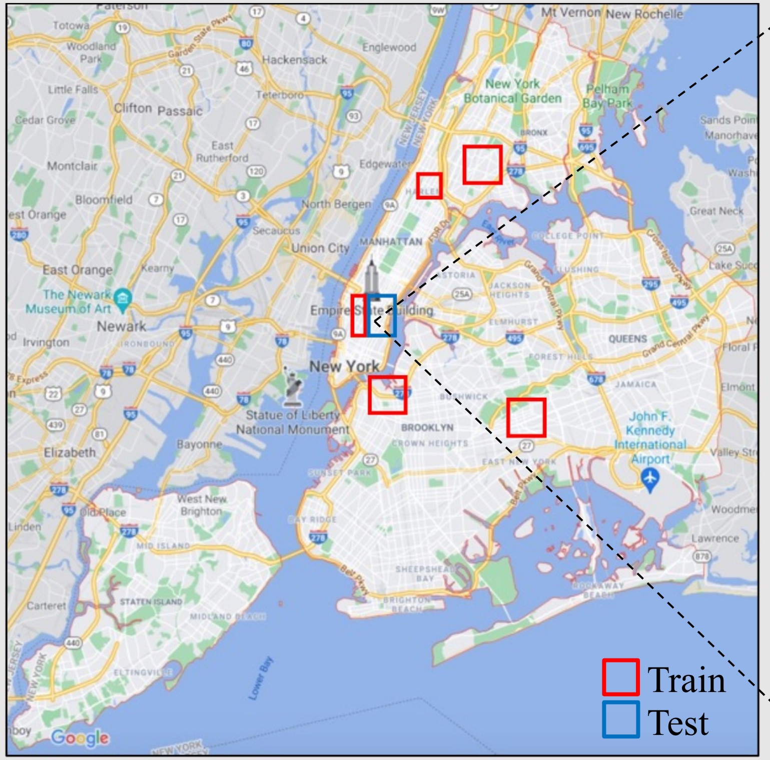


Relation between satellite and street-level images
- Well-aligned at the image level via geographical coordinates
- Complementary characteristics

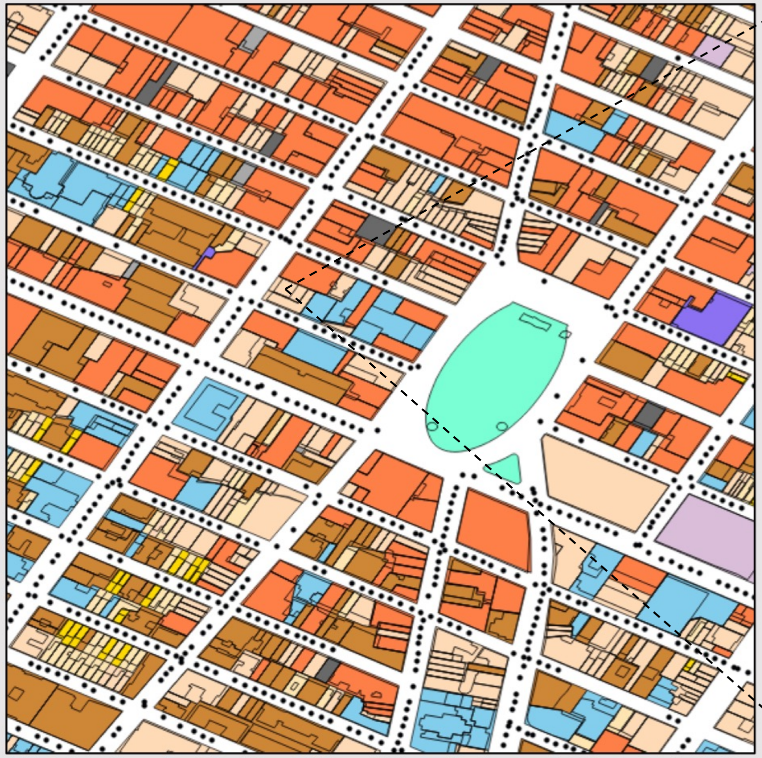


Different satellite-level views　Transformation relations　Different street-level views

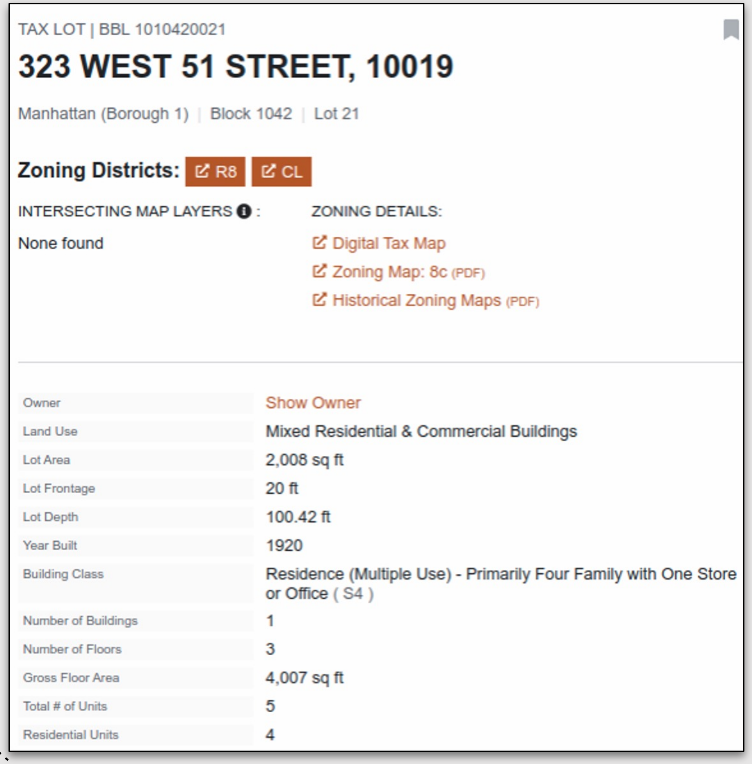
Dataset: collection of images and existing annotations



Regions for collecting the training and test images of multiple views



Existing label maps and street-view image viewpoints of a zoomed area

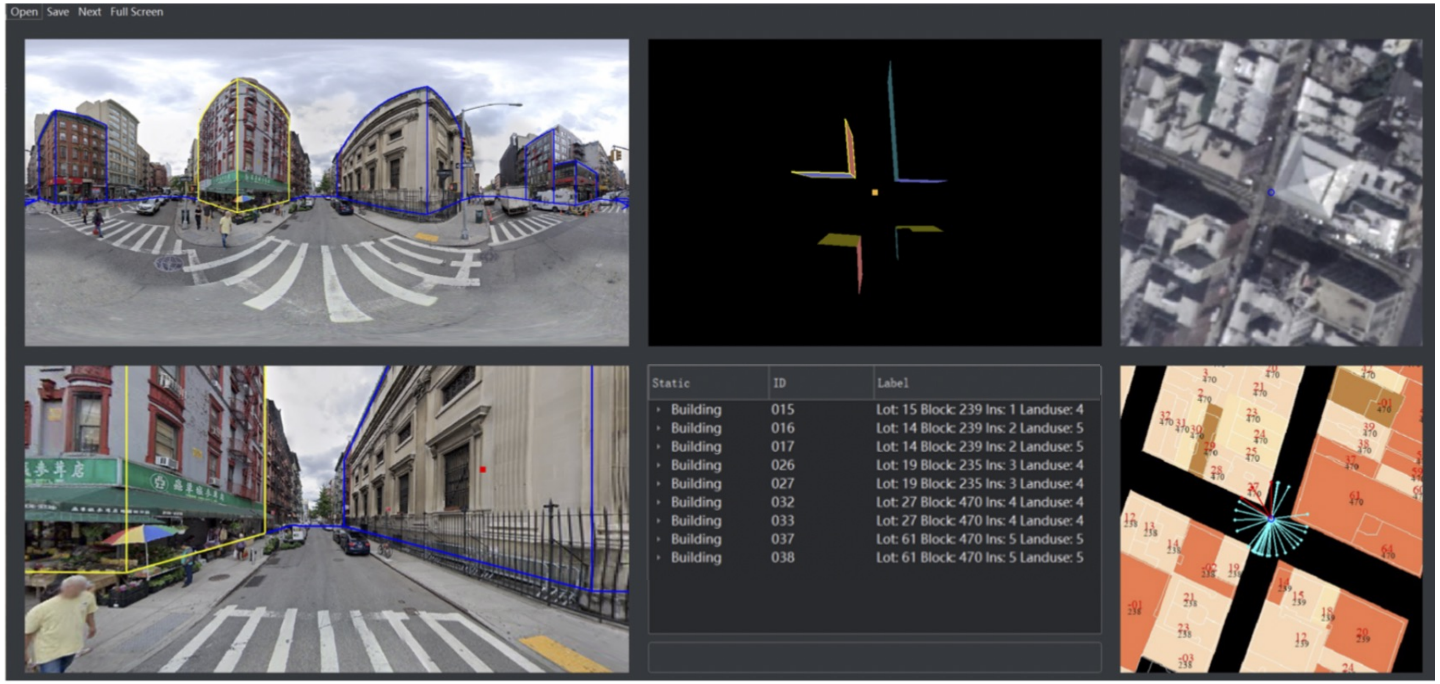


Fine-grained attributes of each building (Usage, #floors, Year built, etc.)

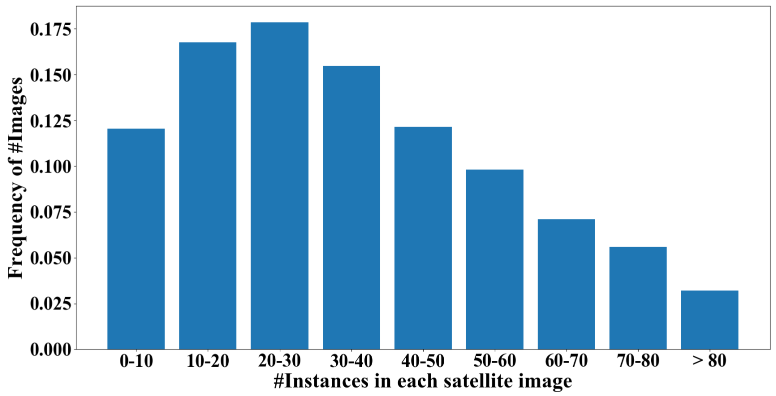
- Images: google street-view panorama images and the corresponding google earth images
- Annotations: OpenStreetMap (footprint and height) and PLUTO (land use, #floors, year built, …)

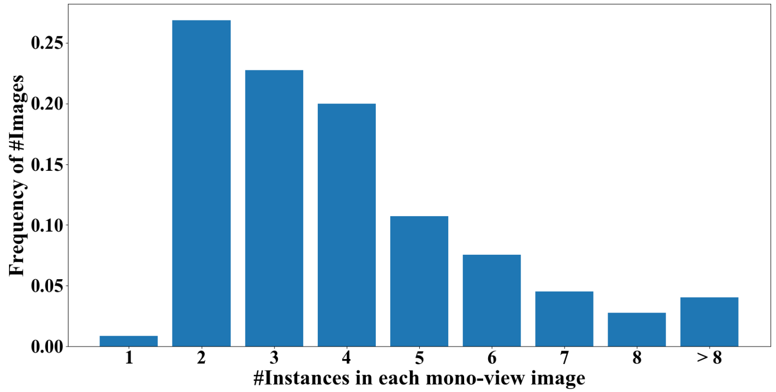
Dataset: street-level image annotation

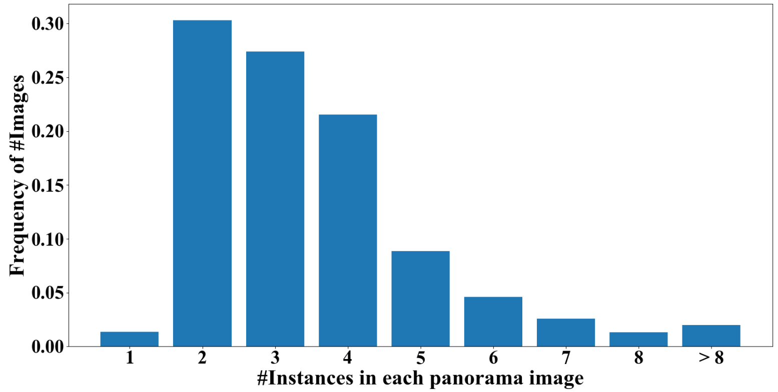
- **Image selection:** select the panorama images that are essential to be annotated according to building coverage, occlusion extent, etc;
- **Segmentation annotation:** adjust the floor/top line to fit the bottom/roof of each building, and add the boundary split line considering both auxiliary information (in the bottom-right window) and building appearance (e.g. texture discrepancy, doors, etc.);
- **Attribute assignment:** assign attributes (instance ID, block-lot id and land use type) for each building plane;
- **Quality assessment:** check the annotation quality and remove the unqualified images.

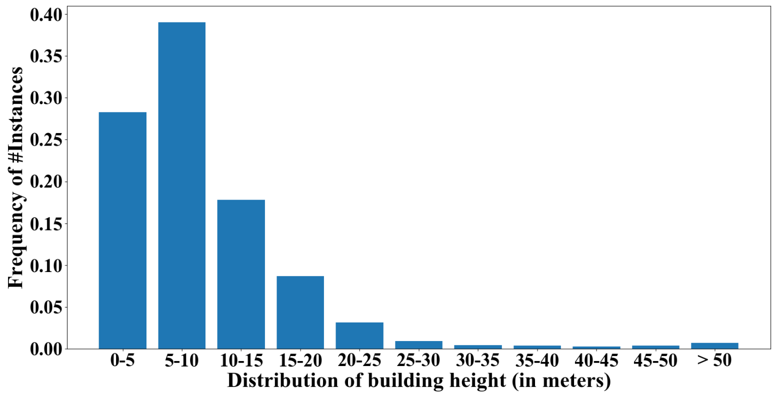


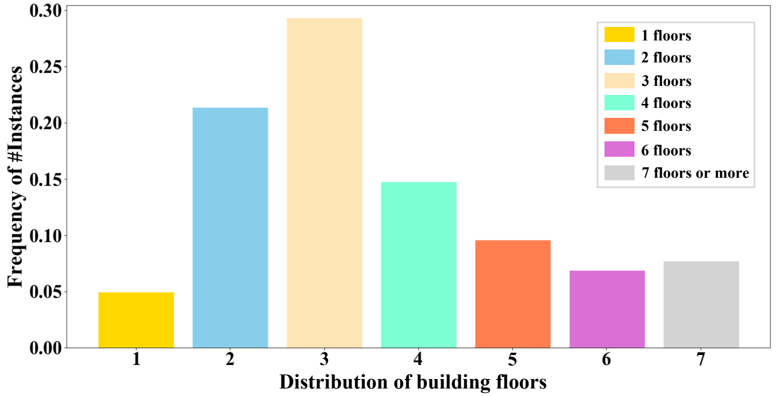
Dataset: statistics

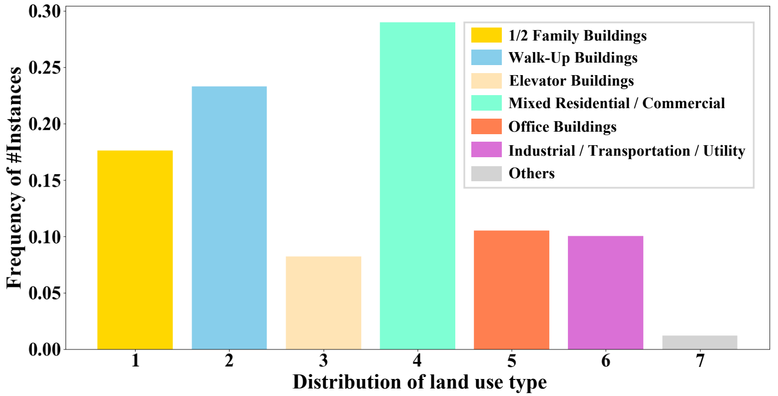






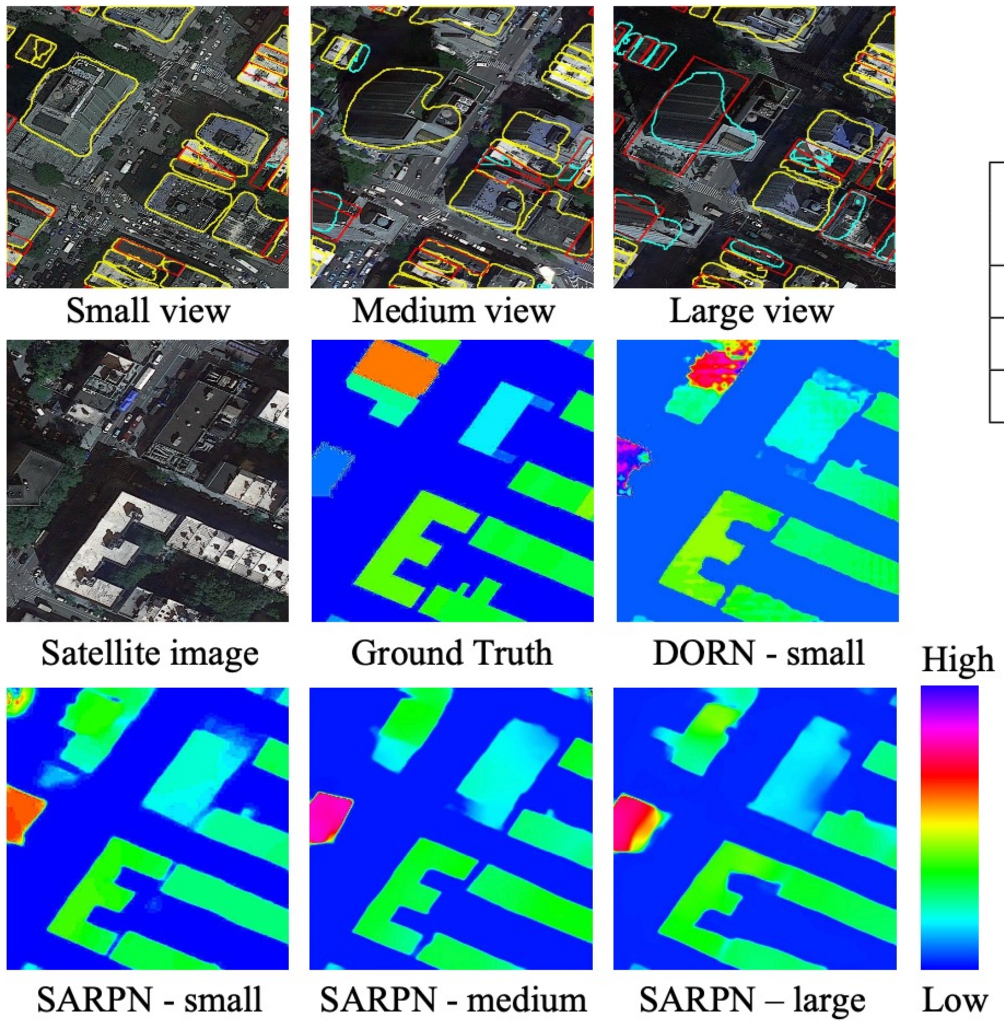






- 75K satellite images in three types of view angles and 33K street-level panorama and mono-view images
- The initial building attributes of PLUTO are merged into seven categories in total
- The numbers of building instances have a great discrepancy between different categories

Benchmark results of satellite-level tasks



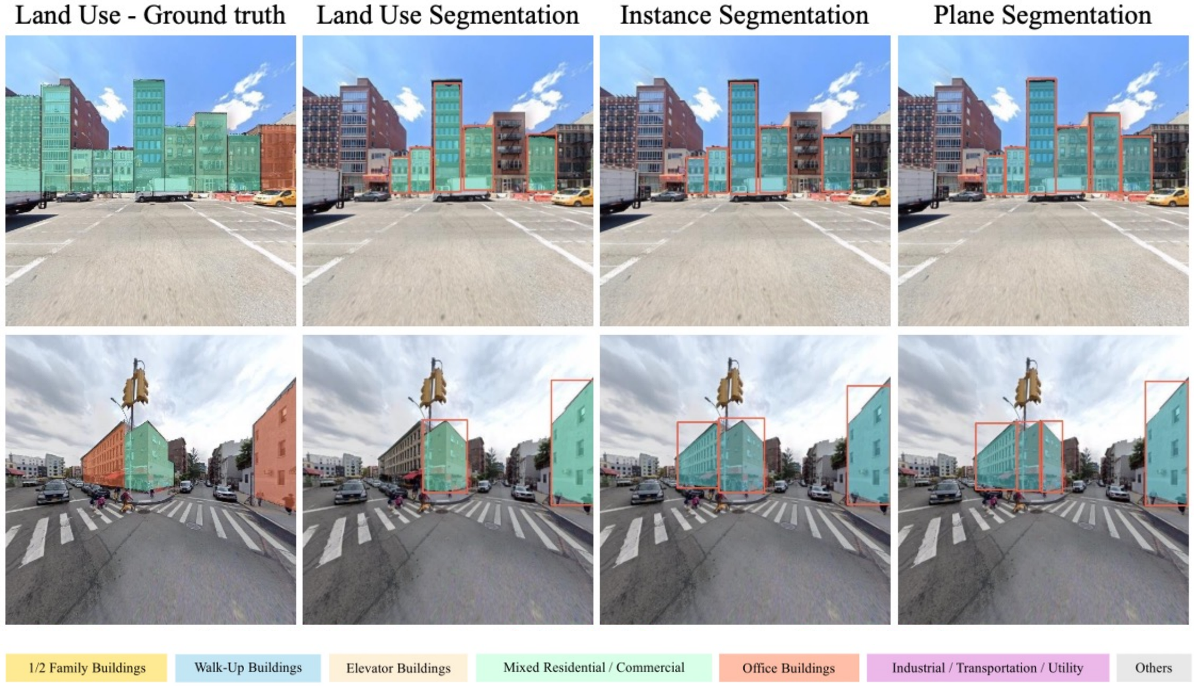
Quantitative results of instance segmentation for satellite images with different view angles (V1/V2/V3: Small/Medium/Large)

View	Metrics of various thresholds						threshold = 0.5		
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	P	R	F1
V1	**29.7**	**66.0**	**23.5**	**15.9**	**33.9**	**36.7**	**76.9**	**66.3**	**71.2**
V2	23.7	56.6	16.1	11.5	27.2	30.3	73.9	55.0	63.1
V3	18.9	51.4	9.6	9.1	21.5	25.3	70.7	51.7	59.7

Quantitative results of height estimation for satellite images with different view angles

View	SARPN [8]			DORN [13]		
	MAE	MSE	RMSE	MAE	MSE	RMSE
V1	16.18	870.34	29.50	12.71	670.52	25.89
V2	**13.75**	**694.17**	**26.35**	**12.24**	**628.06**	**25.06**
V3	15.32	823.01	28.69	13.40	730.67	27.03

Benchmark results of street-level tasks



Task	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Landuse Seg.	23.9	32.1	26.7	0.3	10.6	27.5
Instance Seg.	68.3	88.8	73.8	3.2	33.3	76.1
Plane Seg.	65.1	87.4	71.0	5.0	40.7	73.8

Quantitative results on street-level mono-view images

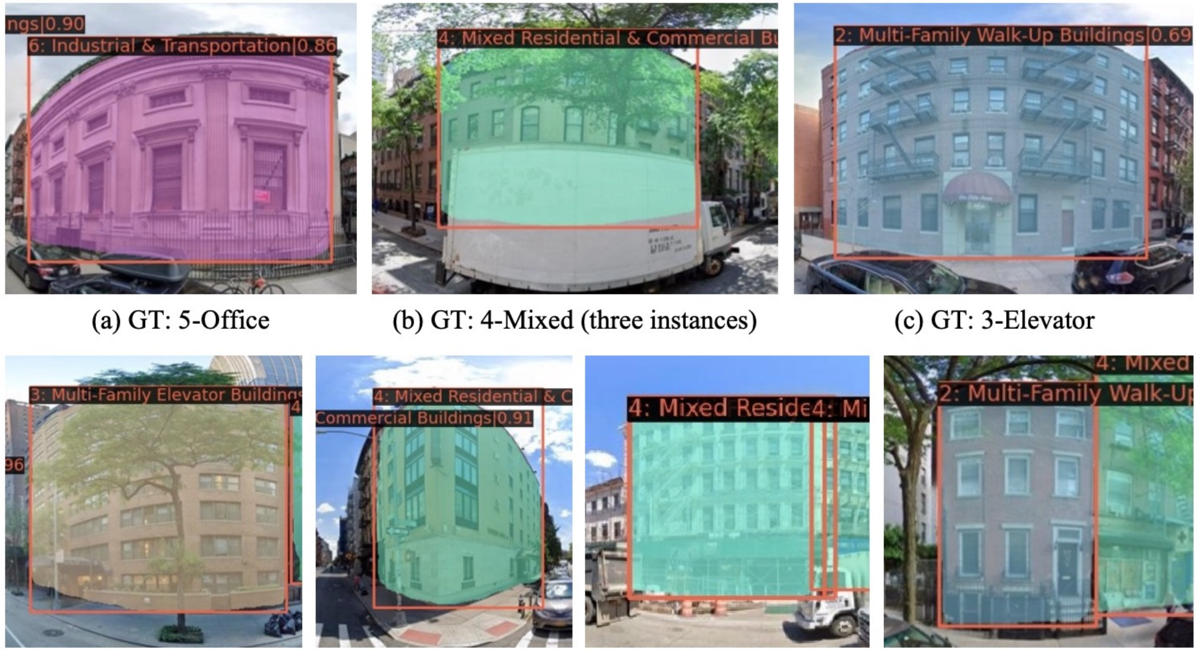
Task	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Landuse Seg.	26.0	34.7	28.5	0.3	12.0	30.4
Instance Seg.	66.7	86.5	72.5	1.7	40.2	74.1

Quantitative results on street-level panorama images

More results will be updated on OmniCity homepage: https://city-super.github.io/omnicity/

Results analysis and discussion

- Existing methods target at general instance segmentation tasks for common datasets (such as COCO and CityScapes), without considering the special properties of panorama images.
- Existing methods have difficulties in recognizing the building instances with a small area and the categories with a small number of building instances, with serious confusions between different categories.
- New instance segmentation methods should be designed for solving the limitations via considering the characteristics of panorama images, building instances, fine-grained categories, etc.



(a) GT: 5-Office (b) GT: 4-Mixed (three instances) (c) GT: 3-Elevator

(d) GT: 5-Office (e) GT: 5-Office (f) GT: 7-Others (g) GT: 1-Family

Typical failure cases of the current benchmark methods

Method	Overall Metrics					
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Mask R-CNN [15]	26.0	34.7	28.5	**0.3**	12.0	30.4
MS R-CNN [19]	27.1	**35.8**	29.8	0.1	**12.4**	31.5
Cascade [5]	25.9	33.8	28.3	0.2	11.4	30.5
CARAFE [35]	25.9	34.5	28.5	0.1	11.9	30.2
HTC [6]	**27.2**	35.7	**29.9**	**0.3**	**12.4**	**32.0**

Method	Metrics of each category						
	C1	C2	C3	C4	C5	C6	C7
Mask R-CNN [15]	19.6	37.5	25.8	39.2	36.9	22.2	0.8
MS R-CNN [19]	**22.5**	**39.1**	26.2	**40.8**	38.0	21.7	**1.2**
Cascade [5]	20	38.3	25	38.5	36.7	22.1	0.3
CARAFE [35]	19.6	37.3	24.9	39.9	37.2	21.5	0.8
HTC [6]	20.8	38.7	**27.2**	39.9	**38.4**	**24.5**	**1.2**

Conclusions and future work

- In this paper, we have proposed OmniCity, a new dataset for omnipotent city understanding from over 100K satellite and street-level images of multiple views, of which the annotations are generated from both existing label maps and our proposed annotation pipeline.

- We provide benchmark experimental results for multiple tasks and data sources based on state-of-the-art methods and analyze their limitations.

- We believe that OmniCity will not only promote new algorithms and application scenarios for existing tasks, but facilitate novel tasks for 3D city reconstruction and simulation.

- In our future work, we plan to enrich OmniCity with more properties of buildings and other geographical object types, extend it to more cities of different countries, and develop new methods for object detection, instance segmentation, and 3D reconstruction from cross-view images.

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thank you!

Project homepage:
https://city-super.github.io/omnicity/