

ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes

Angela Dai¹, Angel X. Chang^{1,2}, Manolis Savva^{1,2}, Maciej Halber², Thomas Funkhouser², Matthias Nießner^{1,3}

¹Stanford University ²Princeton University ³Technical University of Munich

IEEE 2017 Conference on Computer Vision and Pattern Recognition



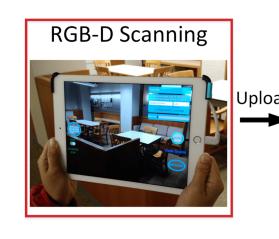
Introduction

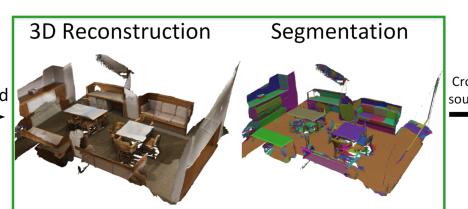
One of the main challenges of applying deep learning models is the availability of large, labeled datasets. Unfortunately, current densely reconstructed RGB-D scene datasets cover only a small range of scene views and have limited semantic annotations. To address this, we introduce ScanNet, an RGB-D video dataset of 1500 scans annotated with 3D camera poses, surface reconstructions, semantic labels, and CAD model placements

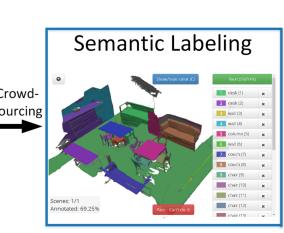
http://www.scan-net.org

Framework

Our RGB-D scanning and semantic annotation framework was designed to enable scalable, crowdsourced data acquisition.







Crowdsourced Semantic Labeling



Data Management Interface

CSV Column visibility Default order Show 25 • entries					Search:						
nowing 51 to 75 of 3	1,524 entries	createdAt	updatedAt (frames	Previous name	1 2 3	4 5	61 Next progress tags	vertices	% annotated	actions
7		2016-10- 04T02:56:10	2016-11- 12T20:03:25	745	mit-d400 mit-d400-3	Office	checked	annotation	120105	70	Annotate Annotatio Details Files Download
	No.	2016-10- 04T02:55:44	2016-11- 12T20:03:56	714	mit-d400 mit-d400-2	Office	checked	annotation	110014	64	Annotatio Details Files Download
		2016-10- 04T02:55:16	2016-11- 12T20:03:10	635	mit-d400 mit-d400-1	Office	checked	annotation	118006	64	Annotatio Details Files Download
BA		2016-10- 04T02:54:11	2016-11- 12T20:04:02	768	mit-d458 mit-d458-3	Office	checked	annotation	176438	50	Annotate Annotatio Details Files

Related work

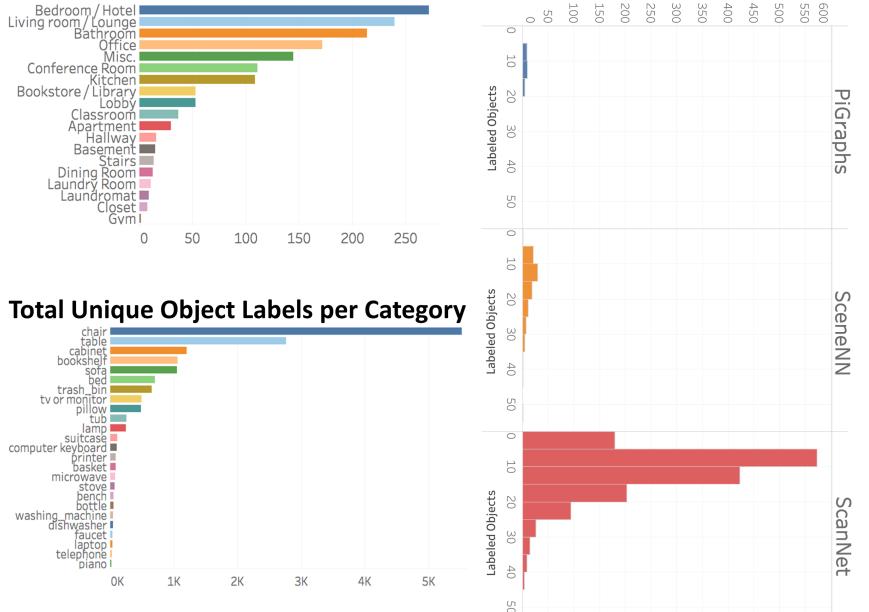
There is a lack of large-scale, semantically annotated dense RGB-D reconstructions of real-world scenes.

ScanNet (ours)	1513 scans 2.5M frames	1513 scans	dense 3D, crowd-sourced MTurk labels also proj. to 2D frames
SceneNN [32]	100 scans	100 scans	dense 3D, by the authors [60]
PiGraphs [71]	26 scans	26 scans	dense 3D, by the authors [71]
BuildingParser [3]	265 rooms	265 rooms	CloudCompare [24]
SUN RGB-D [75]	10k frames	10k frames	2D polygons + bounding boxes
SUN 3D [92]	415 scans	8 scans	2D polygons
TUM [81]	47 scans	none	-
NYU v2 [58]	464 scans	1449 frames	2D LabelMe-style [69]
Dataset	Size	Labels	Annotation Tool

Data and Statistics

Scan Environment per Category

Labeled Object Distribution Comparison

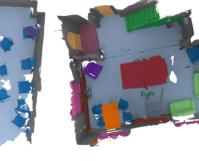


Example Semantically Annotated Scenes Bathroom Conference Room Lounge Living Room



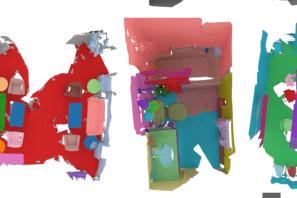


















Tasks and Benchmarks

Semantic Voxel Labeling

Voxel Predictions	Ground Truth	Voxel Predictions	Ground Truth	
				T
				Si Si Si Si Si Si
			100	

curtain counter door window shower curtain refrigerator picture cabinet otherfurniture

3D Object Classification

	Synth	etic Test Sets	Real Test Sets		
Training Set	ShapeNet	ShapeNet Partial	SceneNN	ScanNet	
ShapeNet	92.5	37.6	68.2	39.5	
ShapeNet Partial	88.5	92.1	72.7	45.7	
SceneNN	19.9	27.7	69.8	48.2	
NYU	26.2	26.6	72.7	53.2	
ScanNet	21.4	31.0	78.8	74.9	
ScanNet +ShapeNet Par.	79.7	89.8	81.2	76.6	

Check out our website for data and code!

