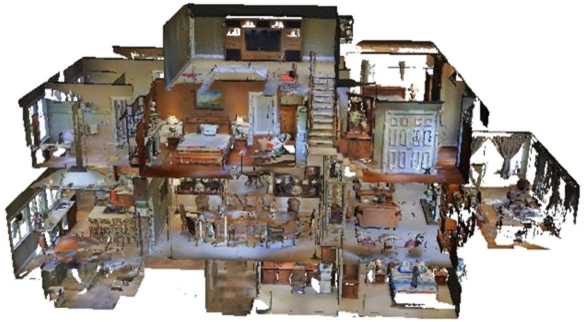
3D Data for Data-Driven Scene Understanding

Thomas Funkhouser

Princeton University*

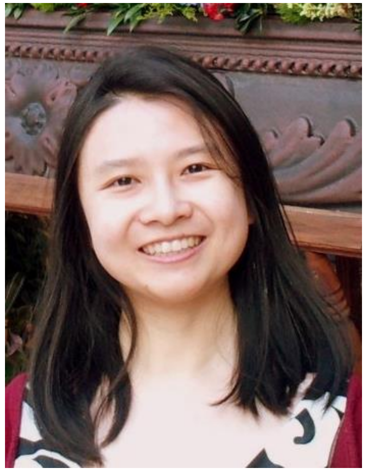


* On Sabbatical at Stanford and Google

Disclaimer: I am talking about the work of these people …

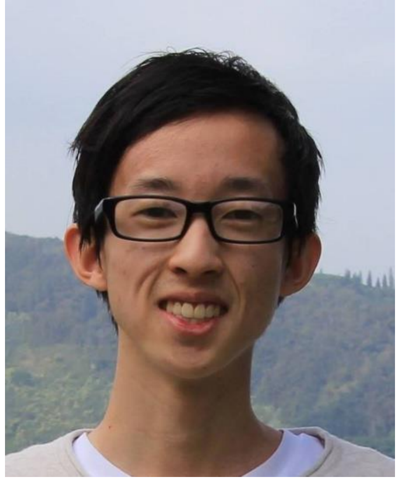
Shuran Song

Manolis Savva



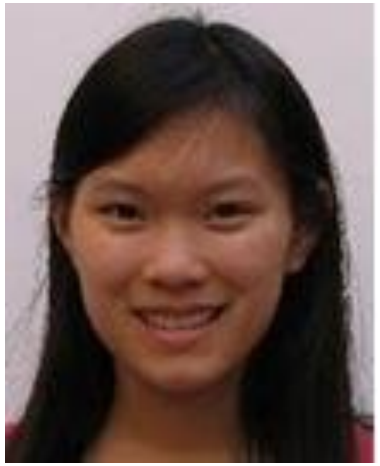
Angel Chang

Yinda Zhang



Andy Zeng

Maciej Halber



Angela Dai

Matthias Niessner

Scene Understanding

Help a computer with cameras to understand indoor environments

- Robotics
- Augmented reality
- Virtual tourism
- Surveillance
- Home remodeling
- Real estate
- Telepresence
- Forensics
- Games
- etc.

Scene Understanding

Help a computer with cameras to understand indoor environments

Input RGB-D Image(s)

Semantic Segmentation

Scene Understanding

Help a computer with cameras to understand indoor environments in 3D

Input RGB-D Image(s)

Semantic Segmentation

+

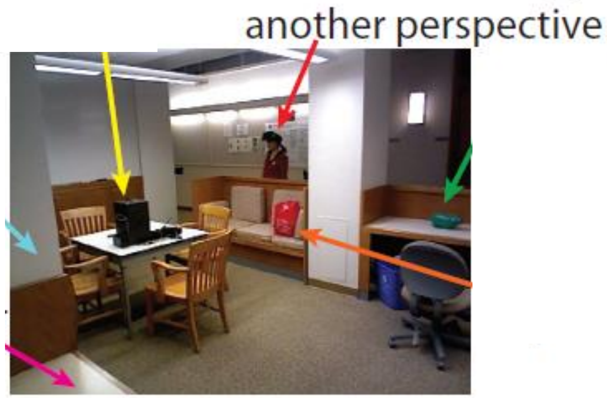
3D Scene Understanding

3D Scene Understanding Research

3D scene understanding research problems:

- Surface reconstruction
- Object detection
- Semantic segmentation
- Scene classification
- Scene completion
- etc.

Semantic Segmentation

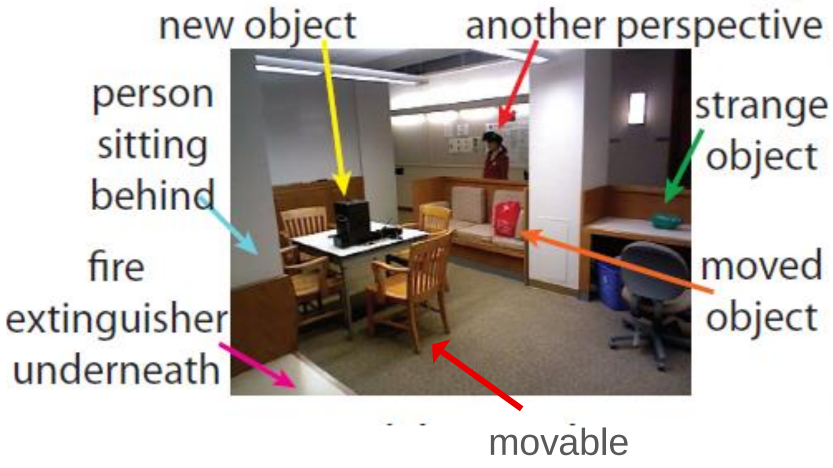


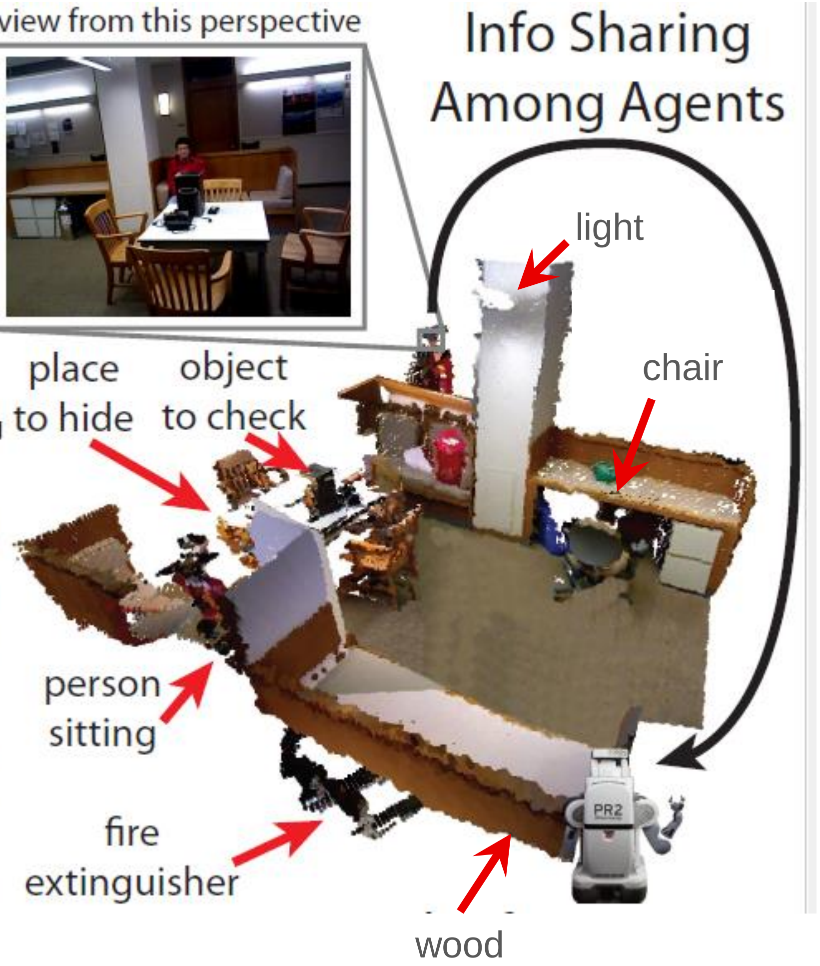
3D Scene Understanding Research

3D scene understanding research problems:

- Surface reconstruction
- Object detection
- Semantic segmentation
- Scene classification
- Scene completion
- Materials, lights, etc.
- Physical properties
- Affordances
- Anomalies
- Changes
- Possibilities
- etc.

Semantic Segmentation





What is the main roadblock for
3D scene understanding research?

What is the main roadblock for
3D scene understanding research?

Data!!!

Outline of This Talk

"New" 3D datasets for indoor scene understanding research:

	Synthetic	RGB-D Image	RGB-D Video
Room			
Multiroom			

Disclaimer: focus on datasets curated by my students and postdocs ☺

Outline of This Talk

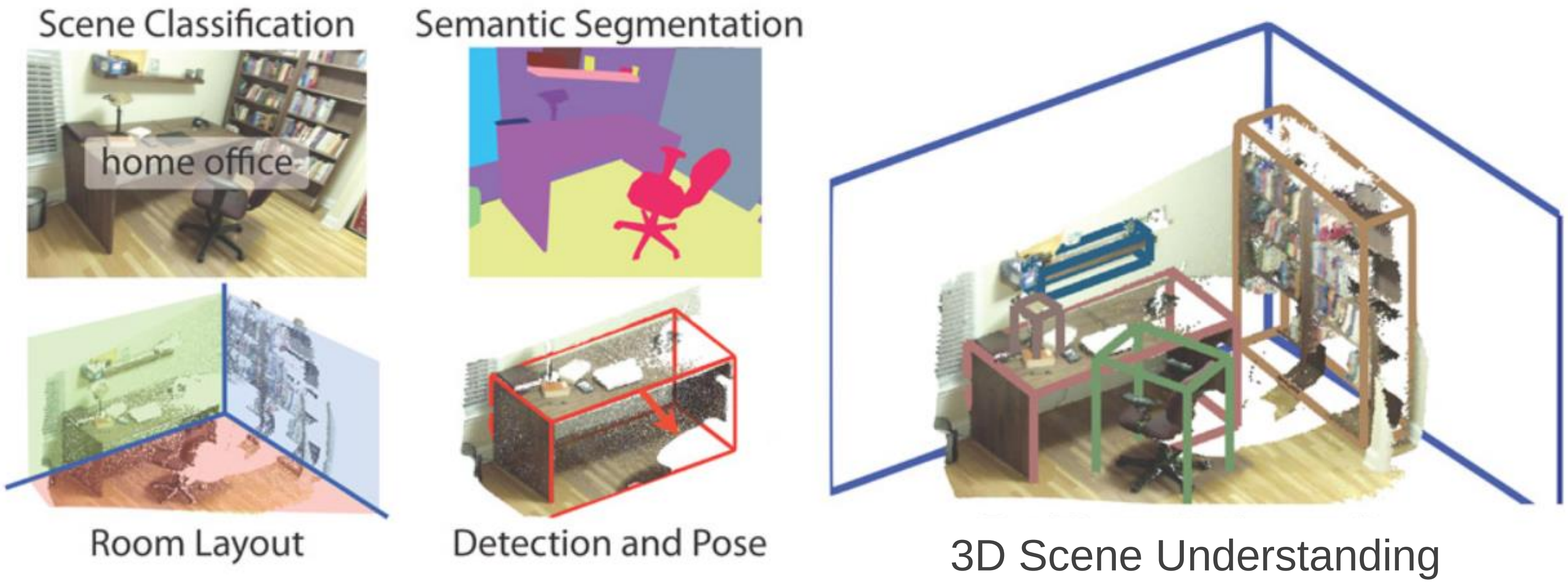
"New" 3D datasets for indoor scene understanding research:

	Synthetic	RGB-D Image	RGB-D Video
Room		SUN RGB-D	
Multiroom			

SUN RGB-D

A RGB-D Scene Understanding Benchmark Suite

- 10,000 RGB-D images
- 147K 2D polygons
- 59K 3D bounding boxes
- Object categories
- Object orientations
- Room categories
- Room layouts



S. Song, S. Lichtenberg, J. Xiao, “SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite,” CVPR 2015

Outline of This Talk

"New" 3D datasets for indoor scene understanding research:

	Synthetic	RGB-D Image	RGB-D Video
Room		SUN RGB-D	
Multiroom			

Outline of This Talk

“New” 3D datasets for indoor scene understanding research:

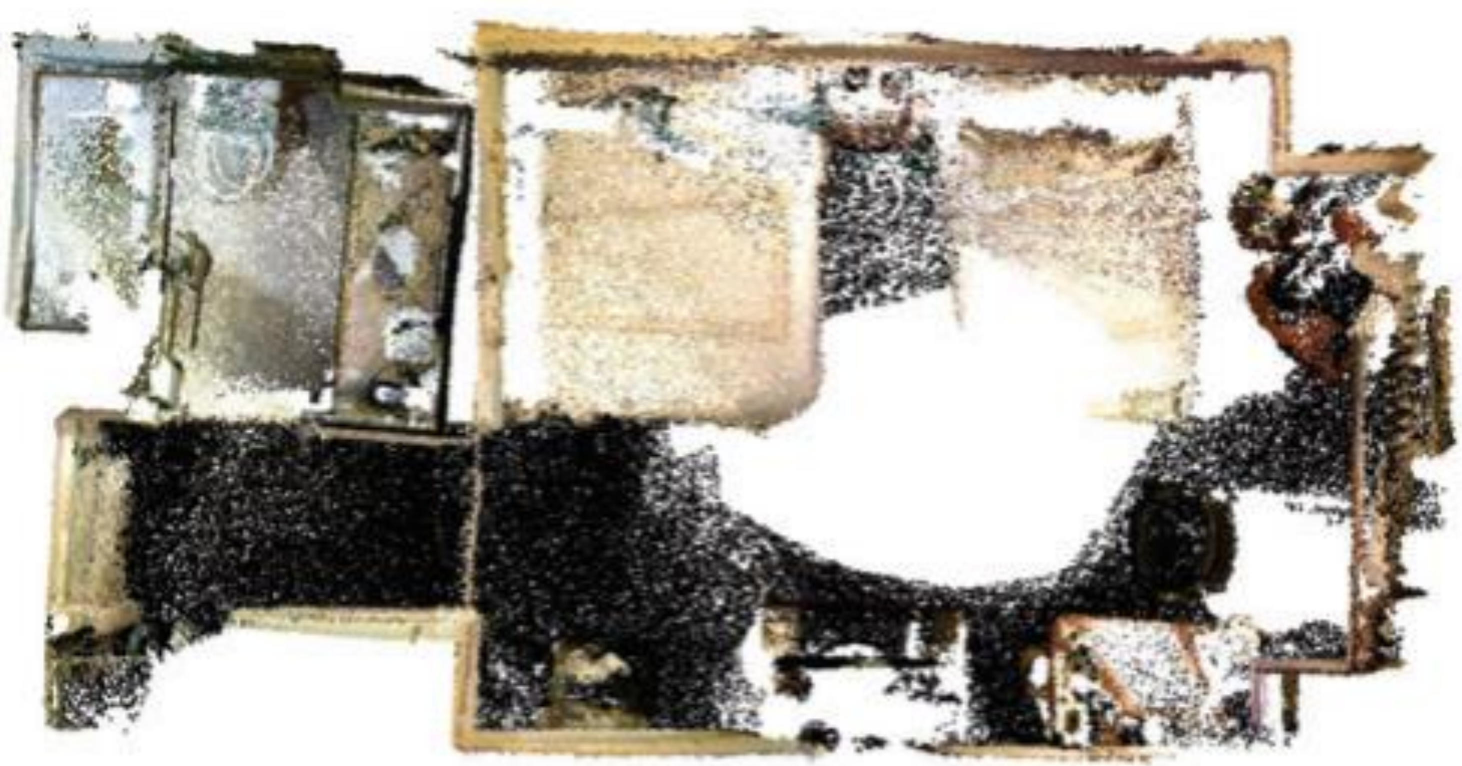
	Synthetic	RGB-D Image	RGB-D Video
Room		SUN RGB-D	
Multiroom			SUN3D

SUN3D

A place-centric 3D database

- 245 spaces
- 415 scans
- Multiple rooms

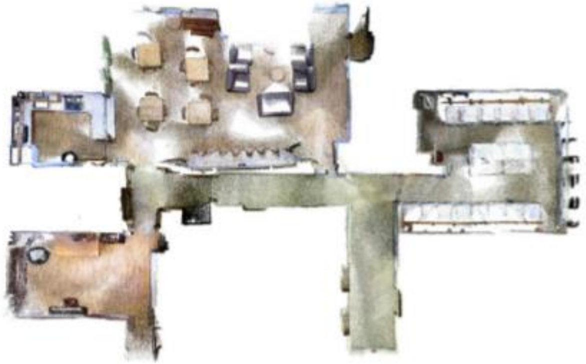
- Originally camera poses distributed for only 8 spaces



J. Xiao, A. Owens, and A. Torralba, “SUN3D: A Database of Big Spaces Reconstructed using SfM and Object Labels,” ICCV 2013

SUN3D

New: camera poses for all 245 spaces (algorithmically reconstructed)

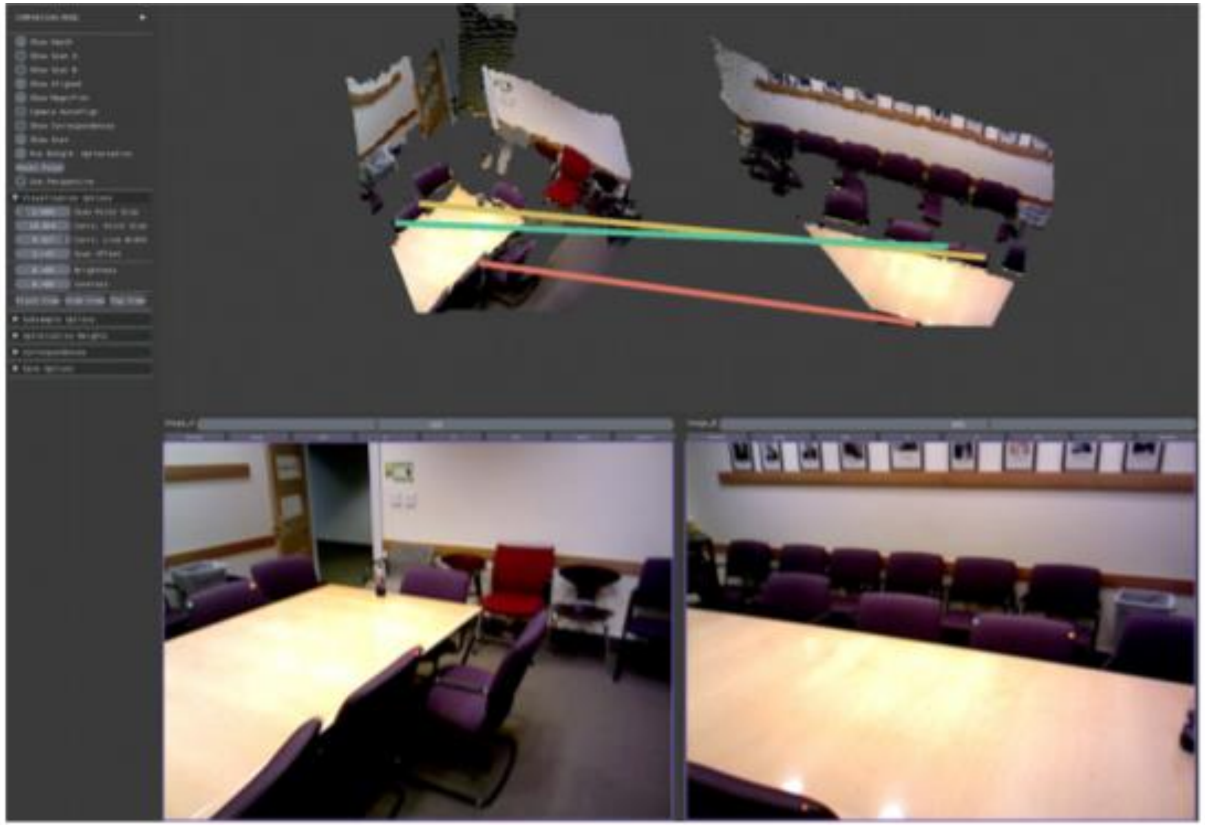


M. Halber and T. Funkhouser, “Fine-to-Coarse Registration of RGB-D Scans,” CVPR 2017

SUN3D

New: “ground truth” point correspondences and camera poses for 25 spaces

- 10,401 manually-specified point correspondences
- Surface reconstructions without visible errors



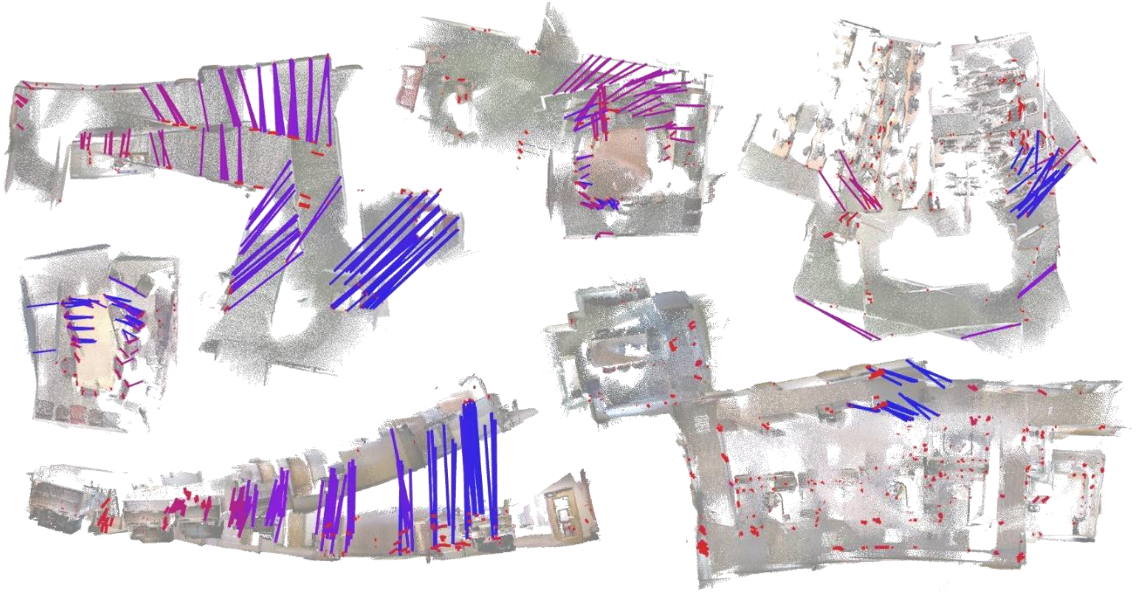
Point correspondence interface

M. Halber and T. Funkhouser, “Fine-to-Coarse Registration of RGB-D Scans,” CVPR 2017

What Can Be Done with SUN3D?

What Can Be Done with SUN3D?

1) Benchmark SLAM algorithms



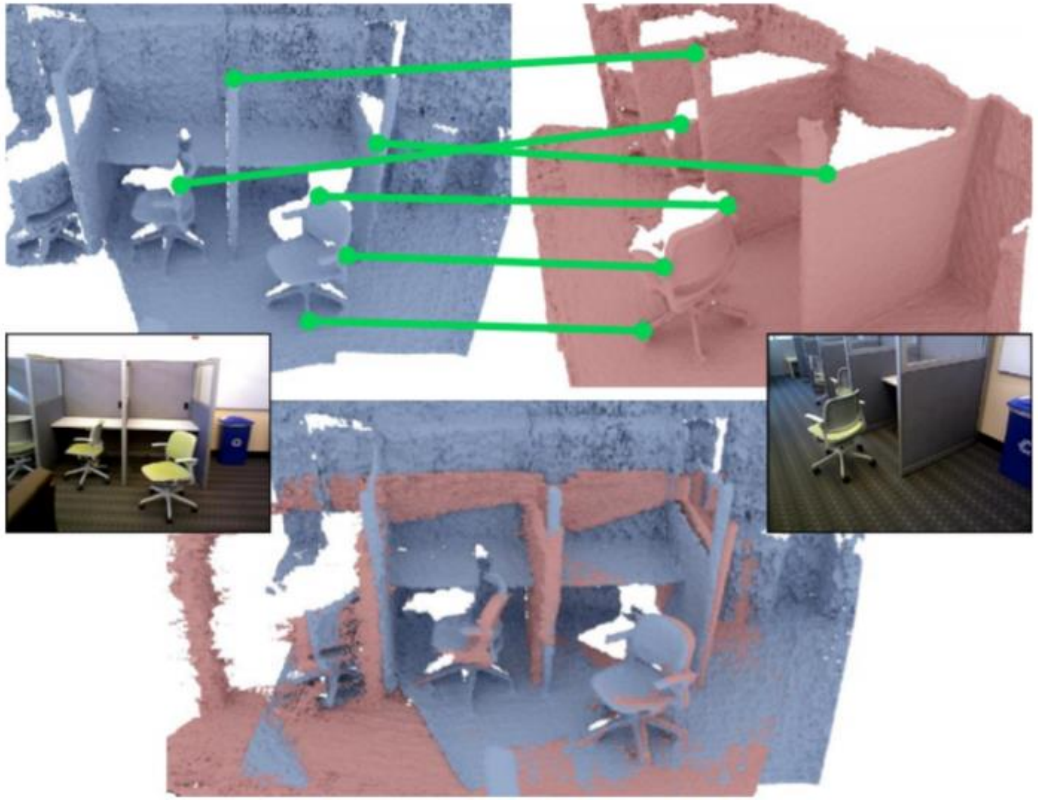
Sequence Name	FtC	SUN3D	RR	Elastic Fusion	Kintinuous
brown_bm_1	0.08345	0.25424	1.60400	1.90877	1.15671
brown_bm_4	0.10545	2.00690	4.12032	0.64936	1.78738
brown_cogsci_1	0.07161	0.89468	1.52869	0.75887	0.55985
brown_cs2	0.06346	0.21408	3.55556	0.89136	0.47414
brown_cs3	0.10796	1.90186	5.90101	2.90157	1.58114
hv_c11_2	0.06471	0.40341	0.27989	0.18390	0.15577
hv_c3_1	0.06541	0.09465	0.41692	0.30158	0.31309
hv_c5_1	0.07766	0.26991	0.11158	0.29152	0.28333
hv_c6_1	0.07524	0.62119	0.26693	0.27570	0.30313
hv_c8_3	0.08656	0.45715	0.24724	0.38132	0.28994
home_at_scan1_2013_jan_1	0.04063	0.21196	0.07570	1.18692	1.23930
home_bksh_oct_30_2012	0.05871	0.15002	1.23549	1.47723	0.58745
home_md_scan9	0.06063	0.16358	1.04740	1.29805	0.54559
nips_4	0.05109	0.15168	0.06181	0.45188	0.40953
scan1	0.06788	0.52143	1.91663	1.98147	1.46379
scan3	0.05042	0.07849	0.06207	0.13804	0.13694
maryland_hotel1	0.06140	0.30138	0.05156	0.65117	0.25950
maryland_hotel3	0.05794	0.20083	0.05260	0.15046	0.11797
d507_2	0.13874	0.32074	0.08354	0.57447	0.52683
ted_lab_2	0.04699	0.11556	0.05600	0.61538	0.59755
76-417b	0.04852	0.09020	0.04724	0.70408	0.68069
76-1studyroom2	0.05347	0.17491	0.12469	0.55497	0.27545
dorm_next_sj	0.08861	0.21222	0.23403	0.19009	0.12923
lab_hj	0.09000	0.67366	0.10347	0.47529	0.16703
sc_athena	0.09680	0.13690	1.41592	1.40803	0.23490

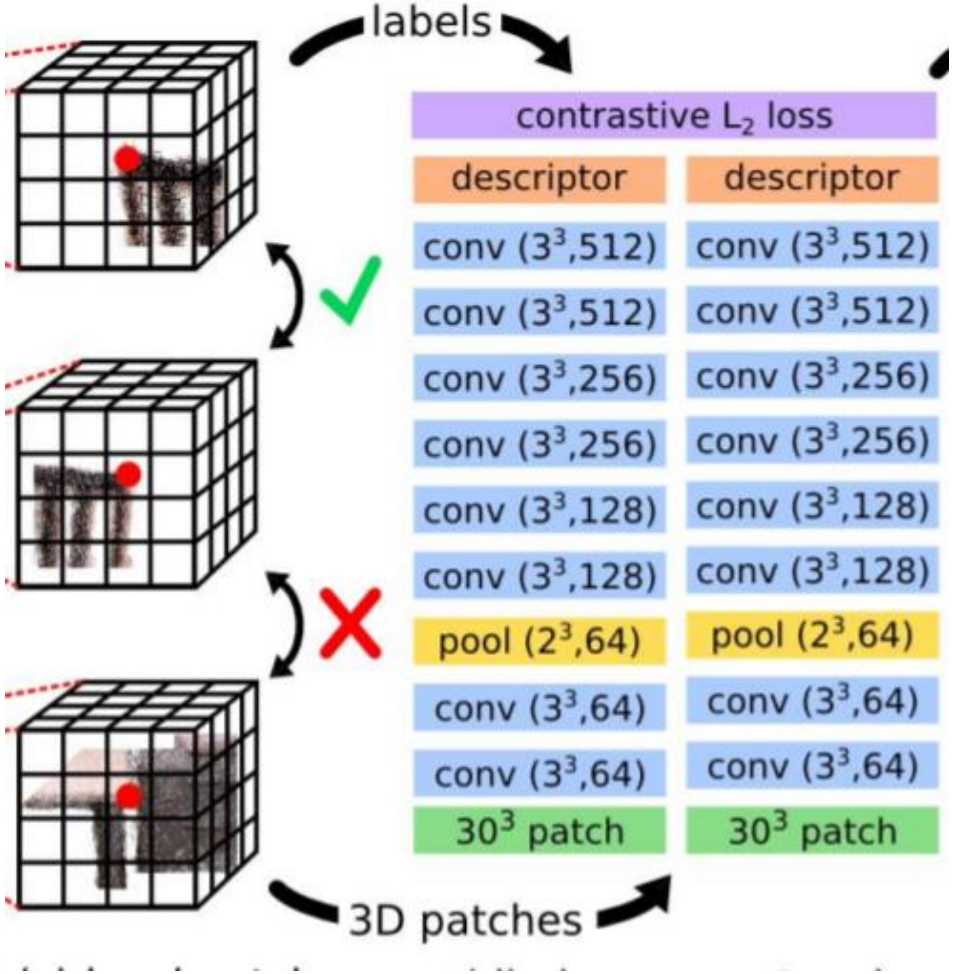
RMSE of ground truth correspondences (in meters)

M. Halber and T. Funkhouser, “Fine-to-Coarse Registration of RGB-D Scans,” CVPR 2017

What Can Be Done with SUN3D?

2) Learn 3D shape descriptors





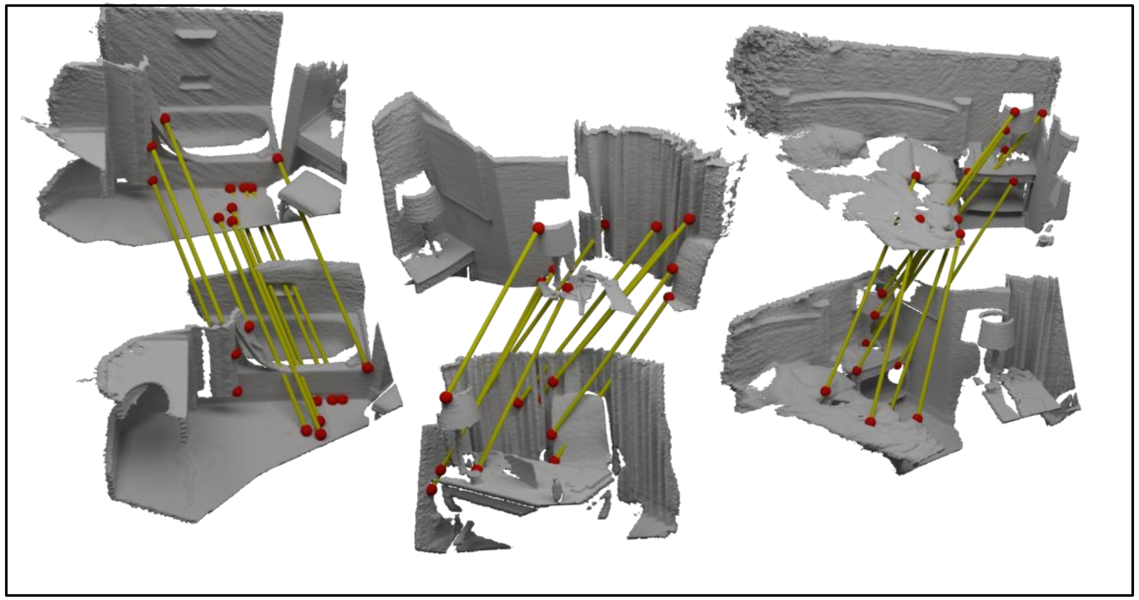
A. Zeng, S. Song, M. Niessner, M. Fisher, J. Xiao, and T. Funkhouser, "3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions," CVPR 2017

What Can Be Done with SUN3D?

2) Learn 3D shape descriptors – descriptors learned from scenes of SUN3D and three other datasets outperform other descriptors

Method	Error
Johnson *et al.* (Spin-Images) [18]	83.7
Rusu *et al.* (FPFH) [27]	61.3
2D ConvNet on Depth	38.5
Ours (3DMatch)	**28.5**

Match classification error at 95% recall



Method	Recall (%)	Precision (%)
Rusu *et al.* [27] + RANSAC	44.2	30.7
Johnson *et al.* [18] + RANSAC	51.8	31.6
Ours + RANSAC	**60.1**	**36.0**

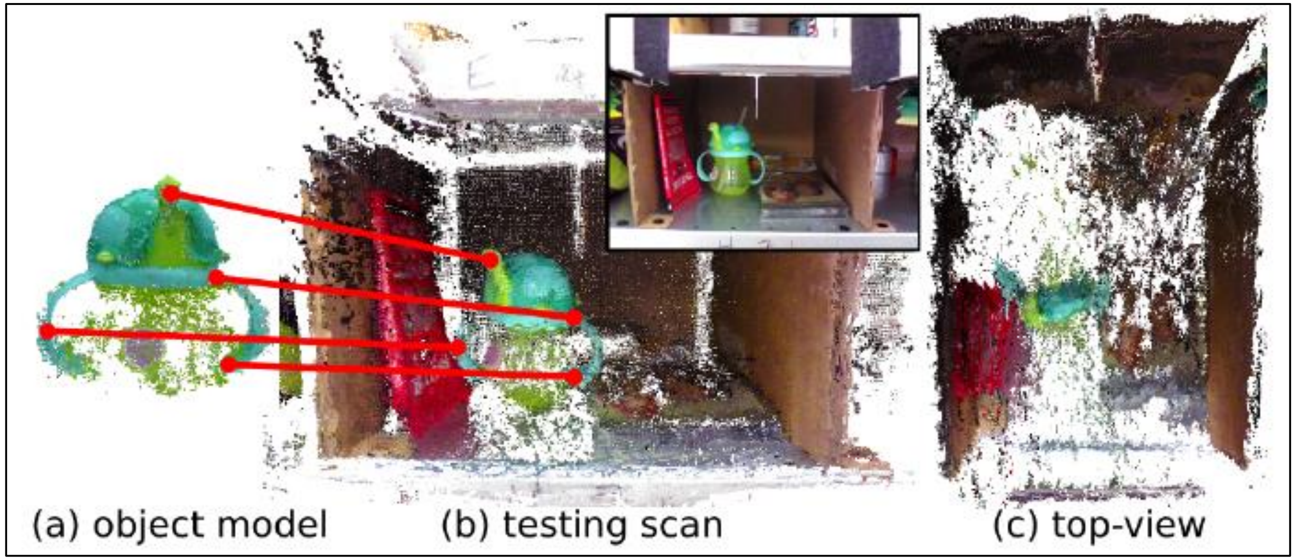
Fragment Alignment Success Rate

What Can Be Done with SUN3D?

2) Learn 3D shape descriptors – descriptors learned from scenes of SUN3D and three other datasets outperform other descriptors

Method	Error
Johnson *et al.* (Spin-Images) [18]	83.7
Rusu *et al.* (FPFH) [27]	61.3
2D ConvNet on Depth	38.5
Ours (3DMatch)	**28.5**

Match classification error at 95% recall



Method	Recall (%)	Precision (%)
Rusu *et al.* [27] + RANSAC	44.2	30.7
Johnson *et al.* [18] + RANSAC	51.8	31.6
Ours + RANSAC	**60.1**	**36.0**

Fragment Alignment Success Rate

Useful for detecting object poses of small objects

What Can Be Done with SUN3D?

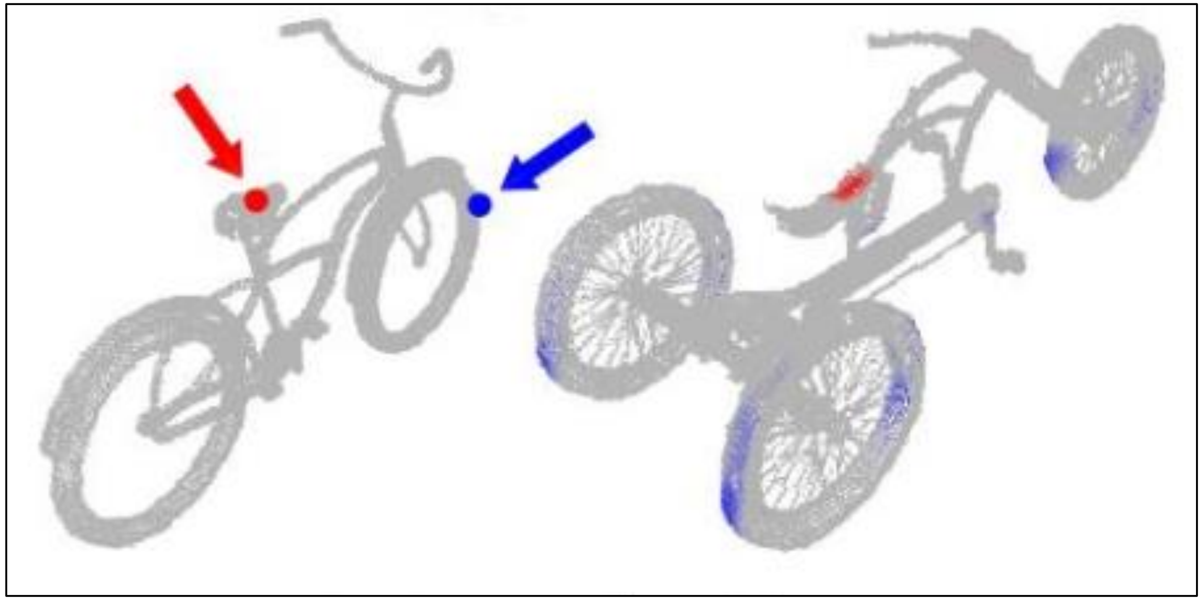
2) Learn 3D shape descriptors – descriptors learned from scenes of SUN3D and three other datasets outperform other descriptors

Method	Error
Johnson *et al.* (Spin-Images) [18]	83.7
Rusu *et al.* (FPFH) [27]	61.3
2D ConvNet on Depth	38.5
Ours (3DMatch)	**28.5**

Match classification error at 95% recall

Method	Recall (%)	Precision (%)
Rusu *et al.* [27] + RANSAC	44.2	30.7
Johnson *et al.* [18] + RANSAC	51.8	31.6
Ours + RANSAC	**60.1**	**36.0**

Fragment Alignment Success Rate



Useful for detecting surface matches in CG models

Outline of This Talk

"New" 3D datasets for indoor scene understanding research:

	Synthetic	RGB-D Image	RGB-D Video
Room		SUN RGB-D	
Multiroom			SUN3D

Outline of This Talk

New 3D datasets for indoor scene understanding research:

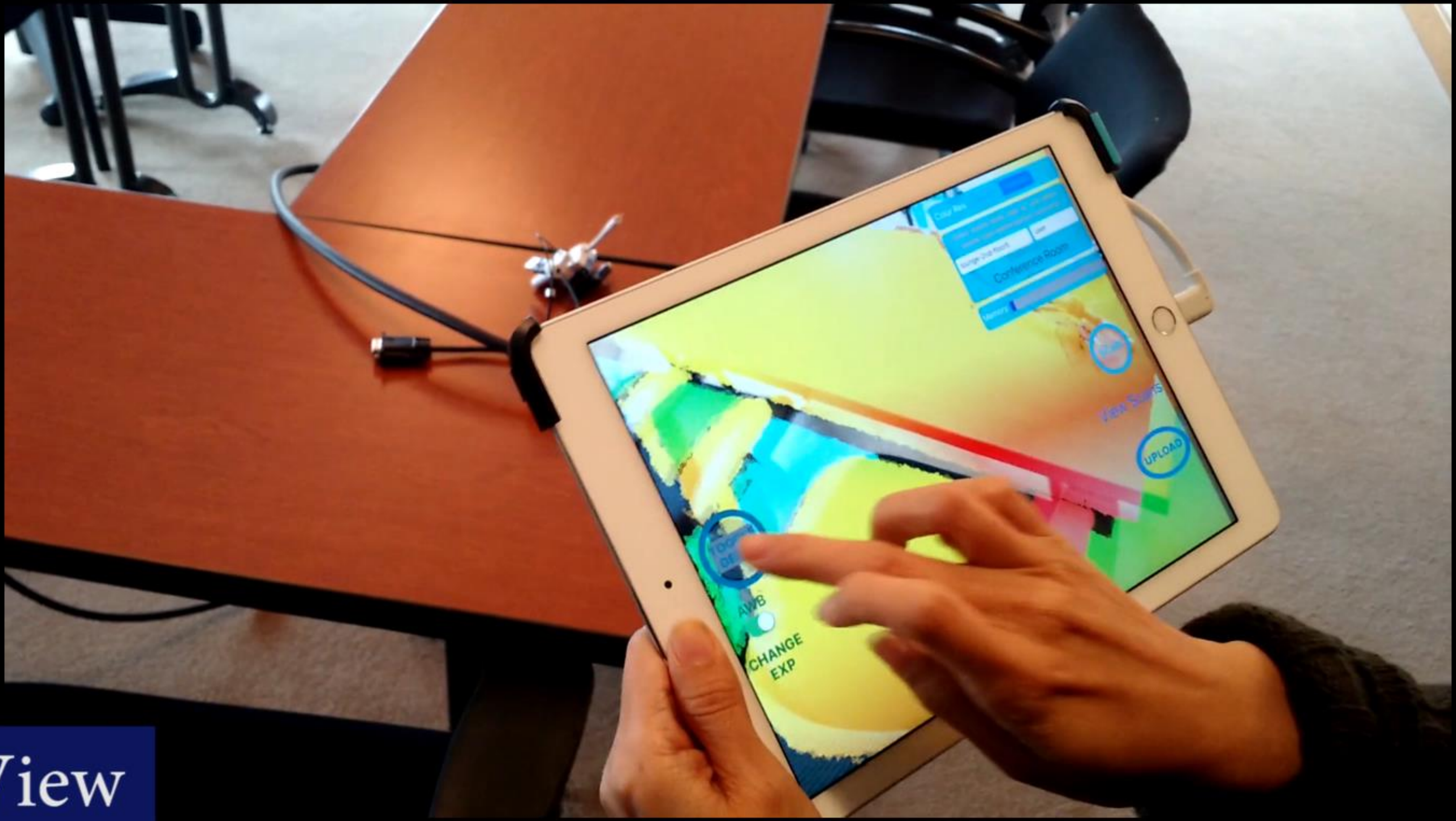
	Synthetic	RGB-D Image	RGB-D Video
Room		SUN RGB-D	**ScanNet**
Multiroom			SUN3D

ScanNet

3D reconstructions and annotations of rooms scanned with RGB-D video

A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

RGB-D Scanning with Commodity Sensors



Depth View

ScanNet

3D reconstructions and annotations of rooms scanned with RGB-D video

- Raw RGB-D video
- Surface reconstructions
- Labeled objects
- CAD model placements

Surface reconstructions

Labeled objects

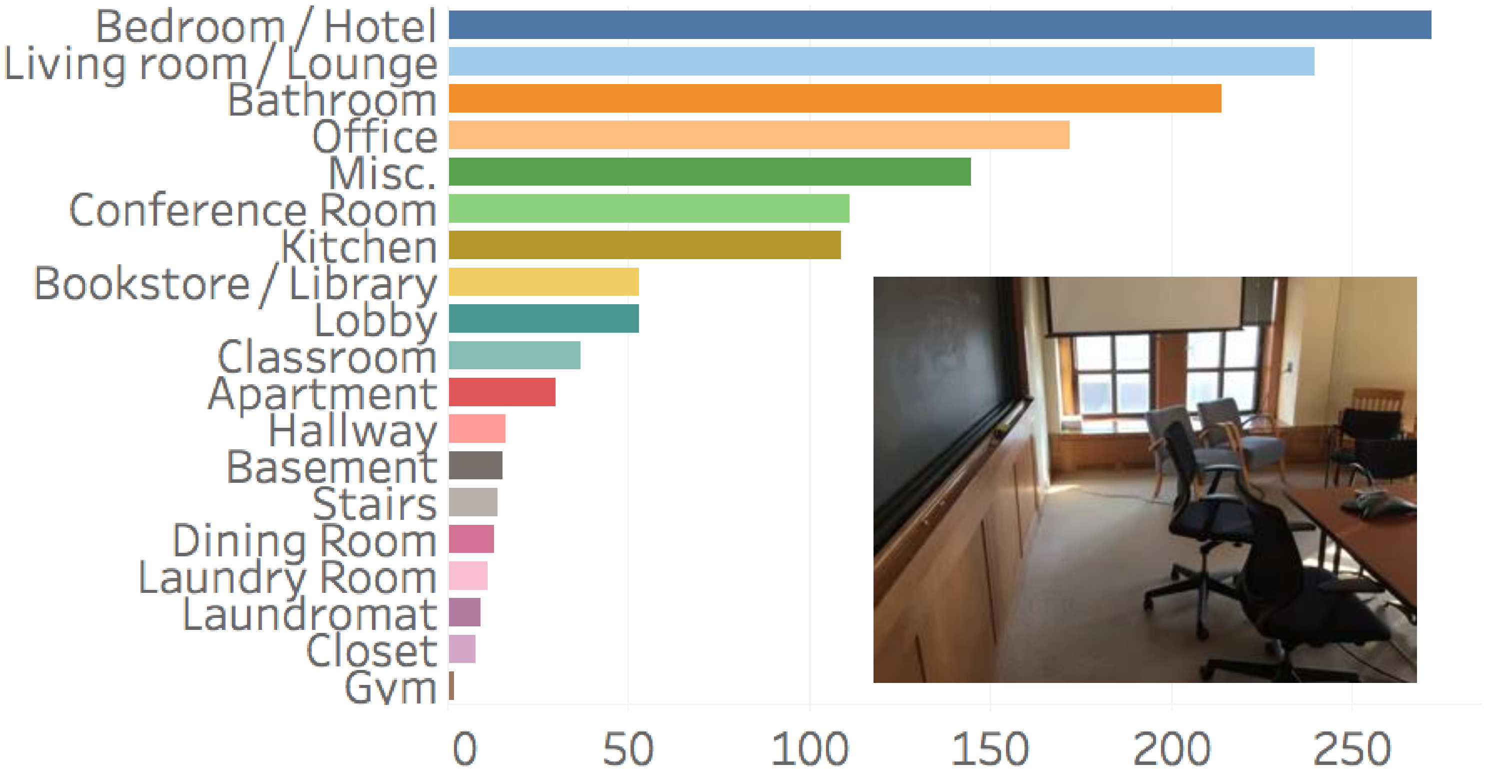
CAD model placements

A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

ScanNet

3D reconstructions and annotations of rooms scanned with RGB-D video

- 1500 scans
- 700 rooms
- 2.5M frames
- 78K sq meters

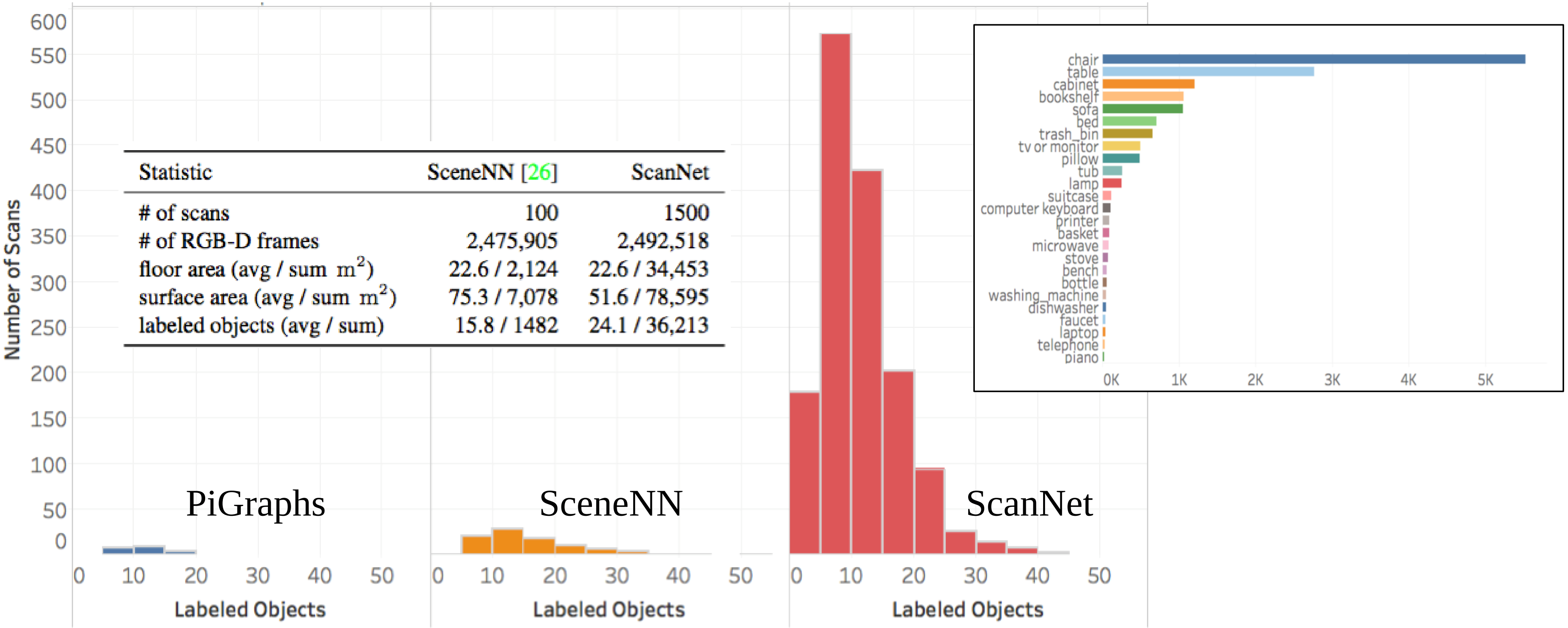


A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

ScanNet

3D reconstructions and annotations of rooms scanned with RGB-D video

- 36K object annotations



Statistic	SceneNN [26]	ScanNet
# of scans	100	1500
# of RGB-D frames	2,475,905	2,492,518
floor area (avg / sum m^2)	22.6 / 2,124	22.6 / 34,453
surface area (avg / sum m^2)	75.3 / 7,078	51.6 / 78,595
labeled objects (avg / sum)	15.8 / 1482	24.1 / 36,213

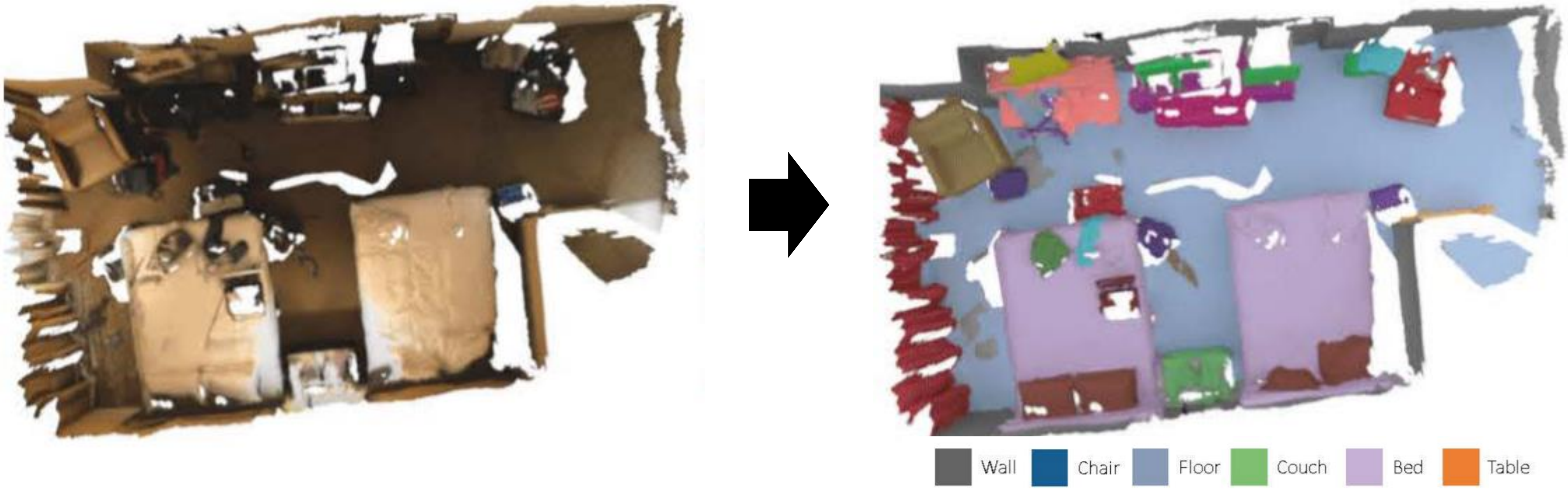
A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

What Can Be Done with ScanNet?

What Can Be Done With ScanNet?

3D Semantic Voxel Labeling

- Task: predict the semantic category of every visible voxel

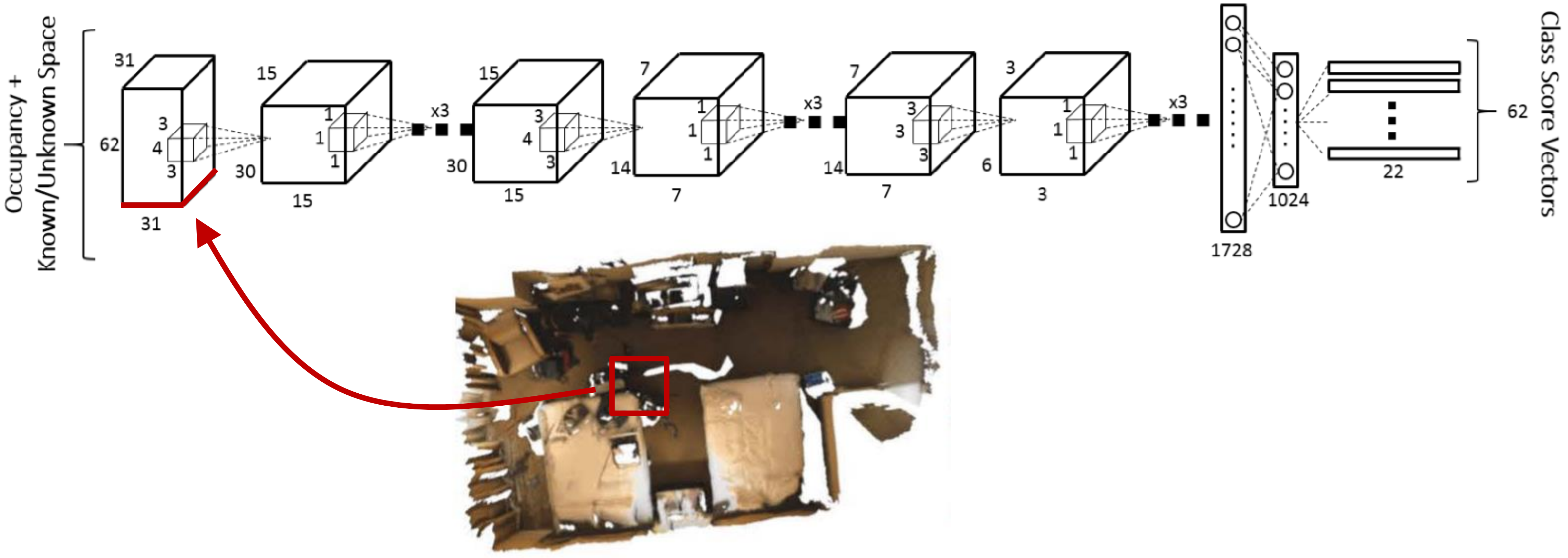


A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, “ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes,” CVPR 2017.

What Can Be Done With ScanNet?

3D Semantic Voxel Labeling

- Method: 3D ConvNet for Sliding Windows

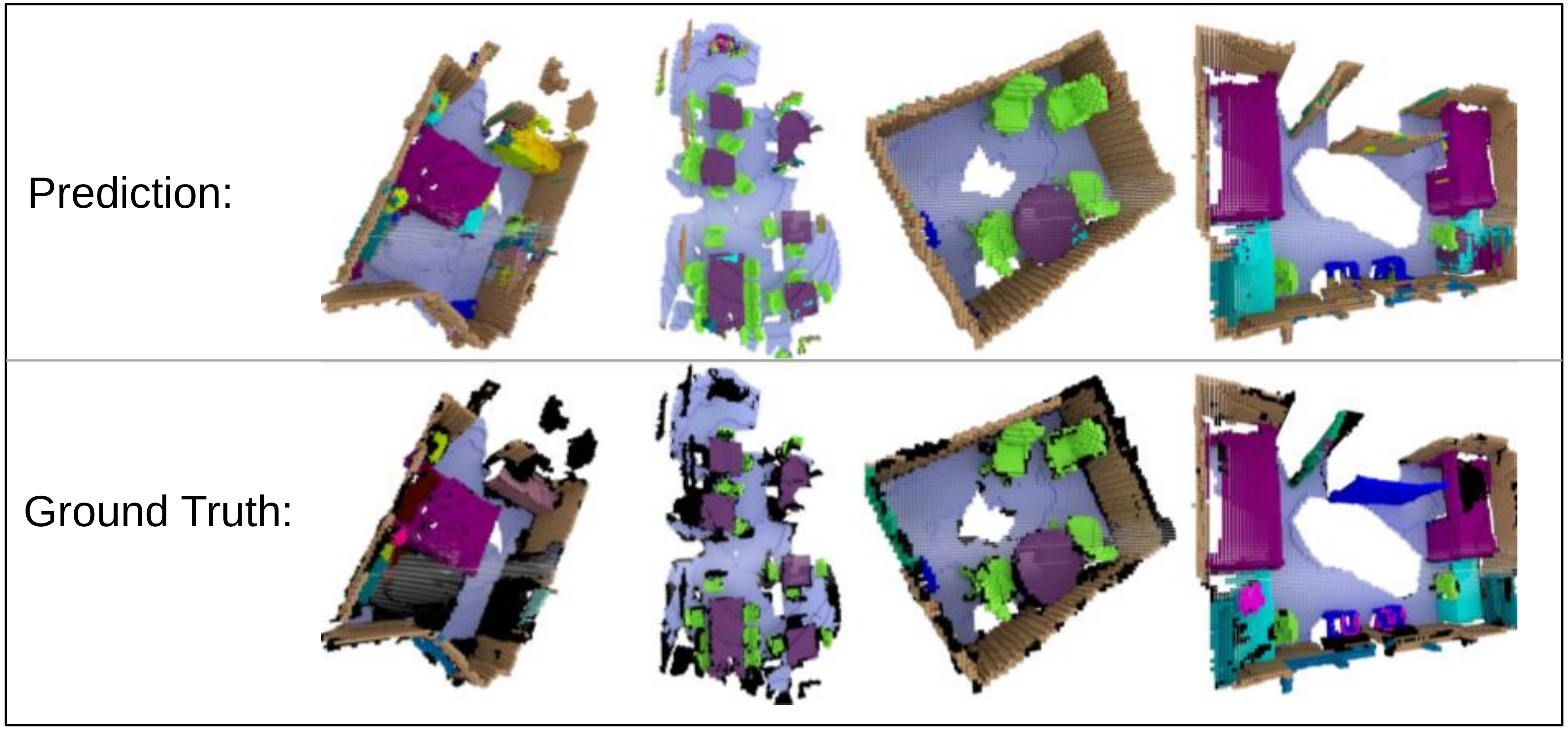


A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

What Can Be Done With ScanNet?

3D Semantic Voxel Labeling

- Results: 73% accuracy overall on 20 classes



Class	% of Test Scenes	Accuracy
Floor	35.7%	90.3%
Wall	38.8%	70.1%
Chair	3.8%	69.3%
Sofa	2.5%	75.7%
Table	3.3%	68.4%
Door	2.2%	48.9%
Cabinet	2.4%	49.8%
Bed	2.0%	62.4%
Desk	1.7%	36.8%
Toilet	0.2%	69.9%
Sink	0.2%	39.4%
Window	0.4%	20.1%
Picture	0.2%	3.4%
Bookshelf	1.6%	64.6%
Curtain	0.7%	7.0%
Shower Curtain	0.04%	46.8%
Counter	0.6%	32.1%
Refrigerator	0.3%	66.4%
Bathtub	0.2%	74.3%
OtherFurniture	2.9%	19.5%
Total	-	73.0%

Voxel classification accuracy on ScanNet test set

A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

What Can Be Done With ScanNet?

3D Semantic Voxel Labeling

- Results: pretraining on ScanNet helps prediction for NYUv2

	floor	wall	chair	table	window	bed	sofa	tv	objs.	furn.	ceil.	avg.
Hermans et al. [31]	91.5	71.8	41.9	27.7	46.1	68.4	28.5	**38.4**	8.6	37.1	**83.4**	49.4
SemanticFusion [54]*	92.6	**86.0**	58.4	34.0	60.5	61.7	47.3	33.9	**59.1**	63.7	43.4	58.2
SceneNet [28]	96.2	85.3	61.0	43.8	30.0	72.5	62.8	19.4	50.0	60.4	74.1	59.6
Ours (ScanNet + NYU)	**99.0**	55.8	**67.6**	**50.9**	**63.1**	**81.4**	**67.2**	35.8	34.6	**65.6**	46.2	**60.7**

Dense pixel classification accuracy on NYUv2

A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Niessner, "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes," CVPR 2017.

Outline of This Talk

New 3D datasets for indoor scene understanding research:

	Synthetic	RGB-D Image	RGB-D Video
Room	**SUNCG**	SUN RGB-D	ScanNet
Multiroom	**SUNCG**		SUN3D

SUNCG

Computer graphics models of houses

- 46K houses
- 50K floors
- 400K rooms

S. Song, F. Yu, A. Zeng, A.X. Chang, M. Savva, T. Funkhouser, "Semantic Scene Completion from a Single Image," CVPR 2017.

SUNCG

Computer graphics models of houses

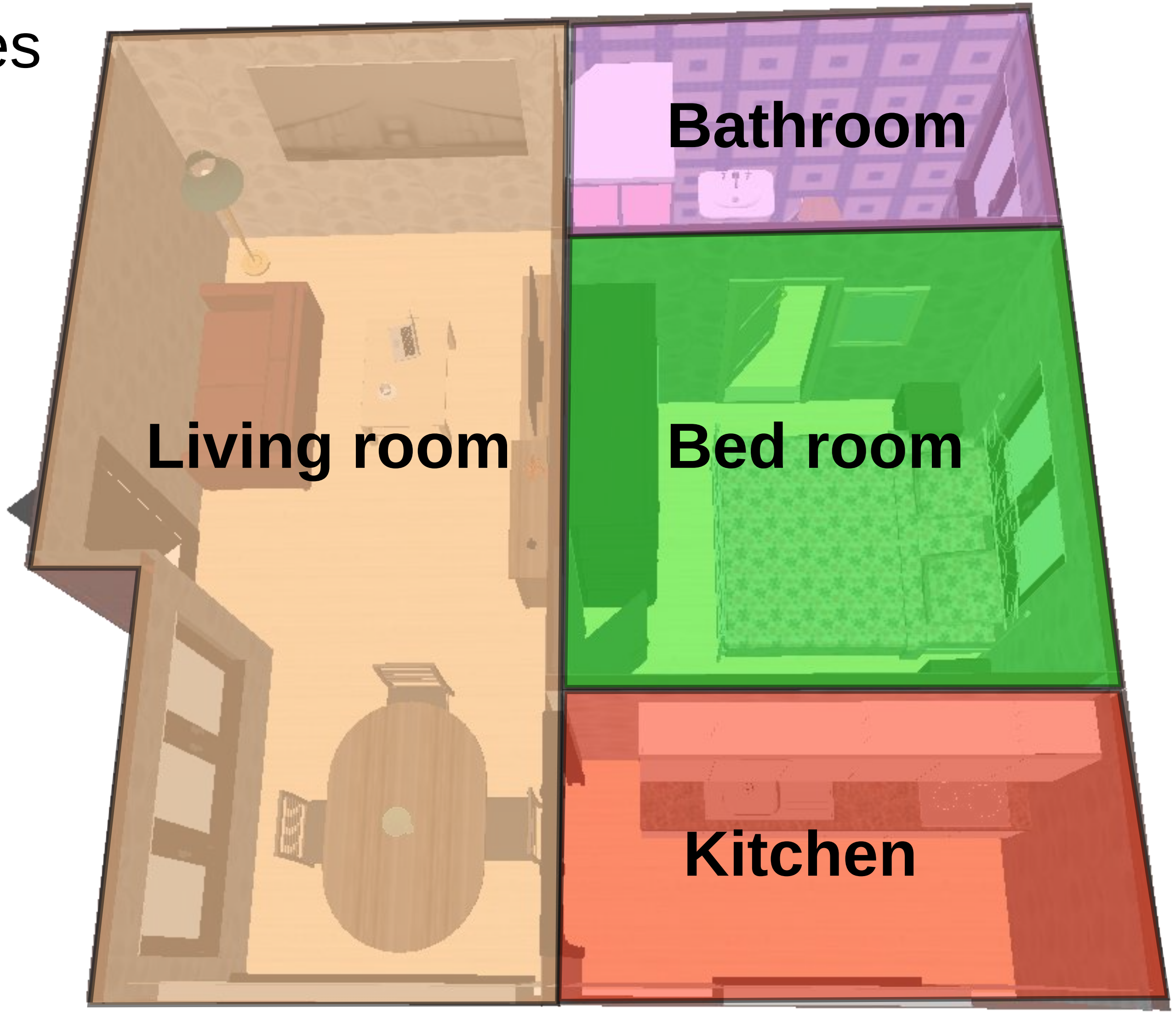
- Furnished (5.6M object instances)

S. Song, F. Yu, A. Zeng, A.X. Chang, M. Savva, T. Funkhouser, "Semantic Scene Completion from a Single Image," CVPR 2017.

SUNCG

Computer graphics models of houses

- Furnished (5.6M object instances)
- Labeled room categories

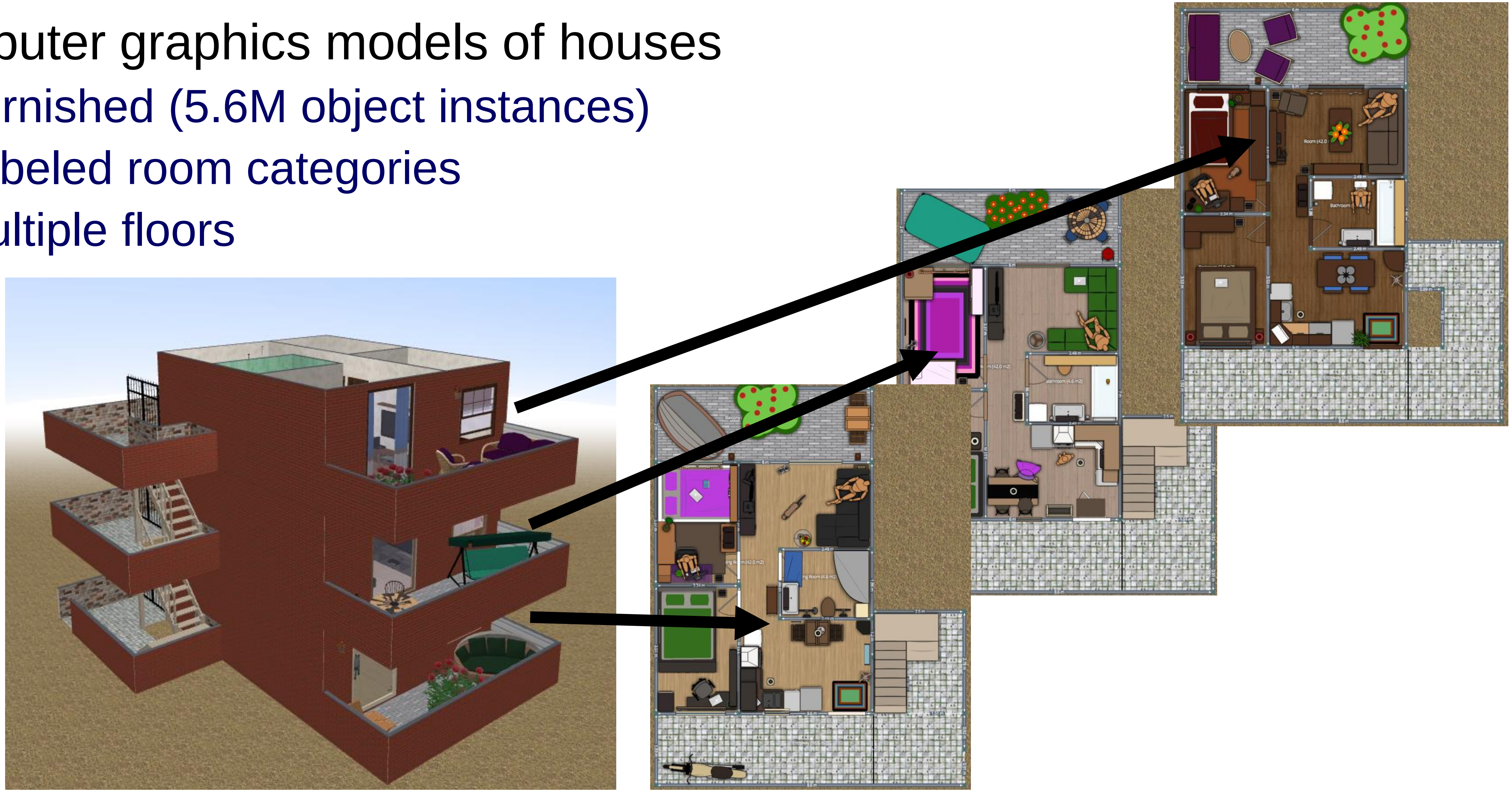


S. Song, F. Yu, A. Zeng, A.X. Chang, M. Savva, T. Funkhouser, "Semantic Scene Completion from a Single Image," CVPR 2017.

SUNCG

Computer graphics models of houses

- Furnished (5.6M object instances)
- Labeled room categories
- Multiple floors



S. Song, F. Yu, A. Zeng, A.X. Chang, M. Savva, T. Funkhouser, “Semantic Scene Completion from a Single Image,” CVPR 2017.

SUNCG

Computer graphics models of houses

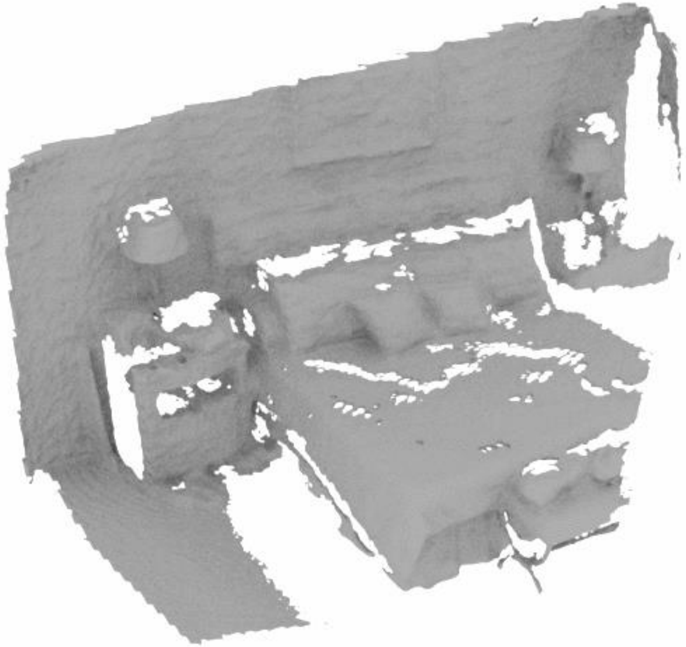
- Furnished (5.6M object instances)
- Labeled room categories
- Multiple floors
- Materials
- Textures
- Lights

S. Song, F. Yu, A. Zeng, A.X. Chang, M. Savva, T. Funkhouser, “Semantic Scene Completion from a Single Image,” CVPR 2017.

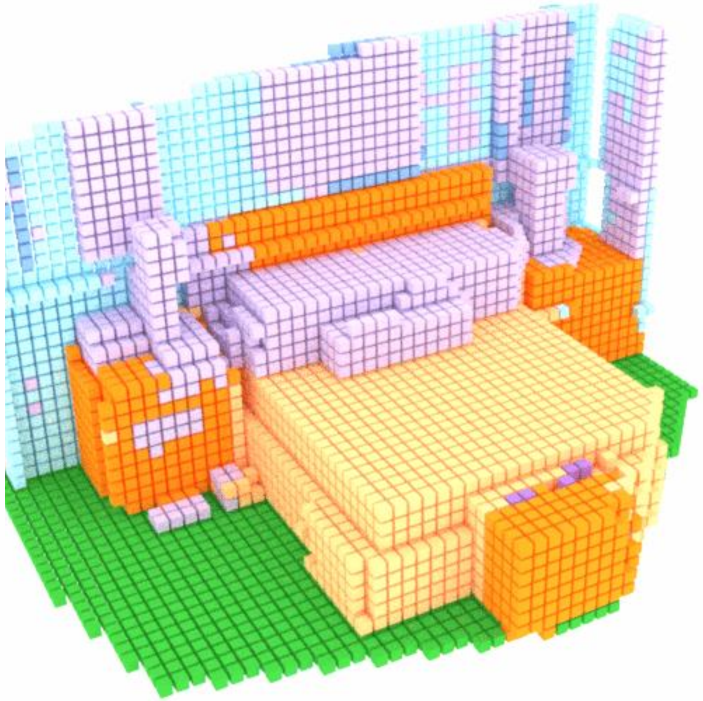
What Can Be Done with SUNCG?

What Can Be Done With SUNCG?

1) Semantic Scene Completion (label ALL voxels, not just visible ones)



Input: Single view depth map



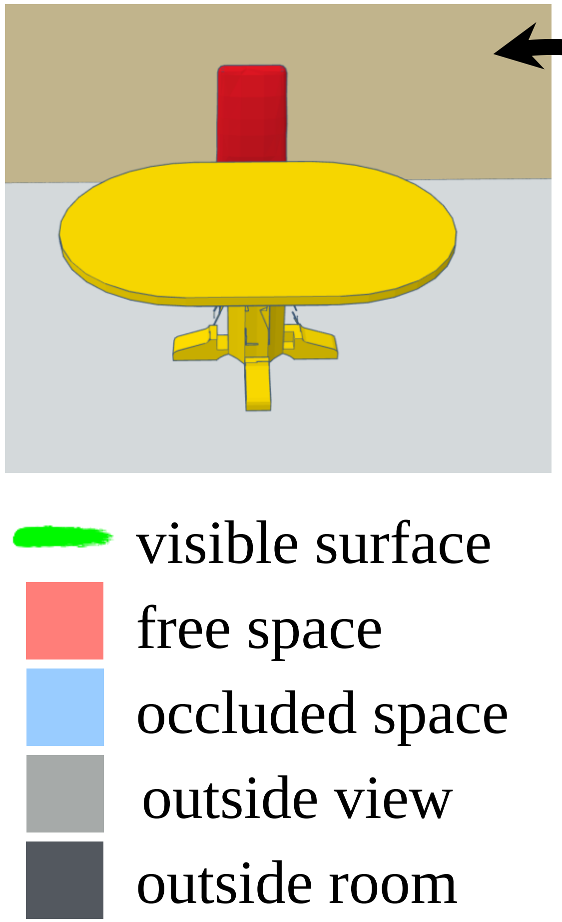
Output: Semantic scene completion

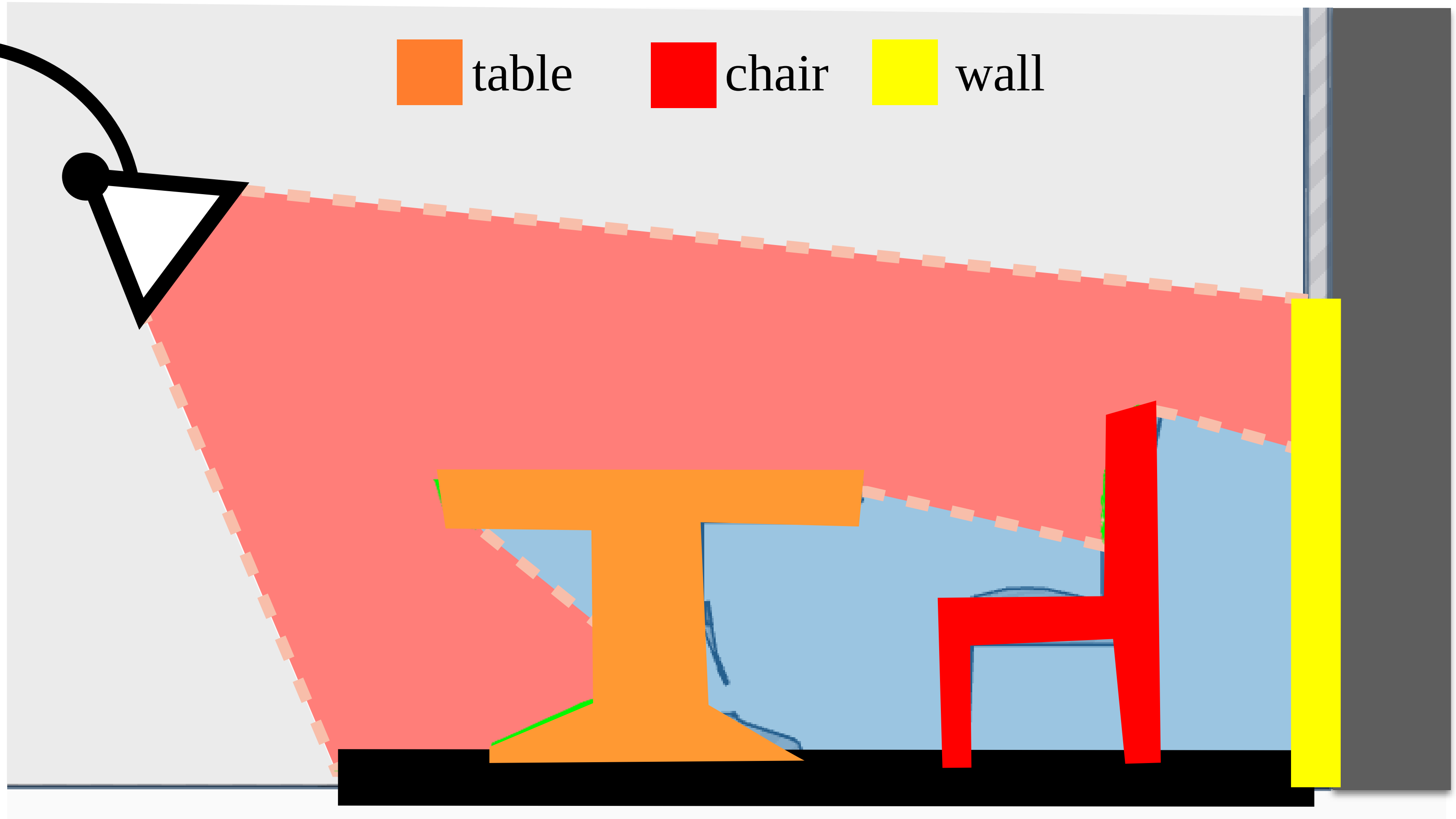
floor wall window chair bed
sofa table tvs furn. objects

S. Song, F. Yu, A. Zeng, A.X. Chang, M. Savva, and T. Funkhouser, "Semantic Scene Completion from a Single Image," CVPR 2017.

What Can Be Done With SUNCG?

1) Semantic Scene Completion (label ALL voxels, not just visible ones)

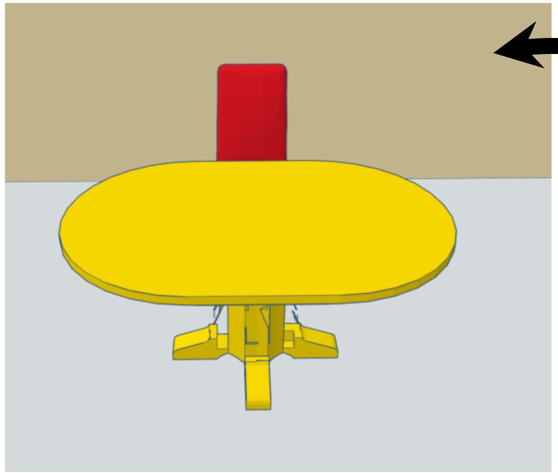


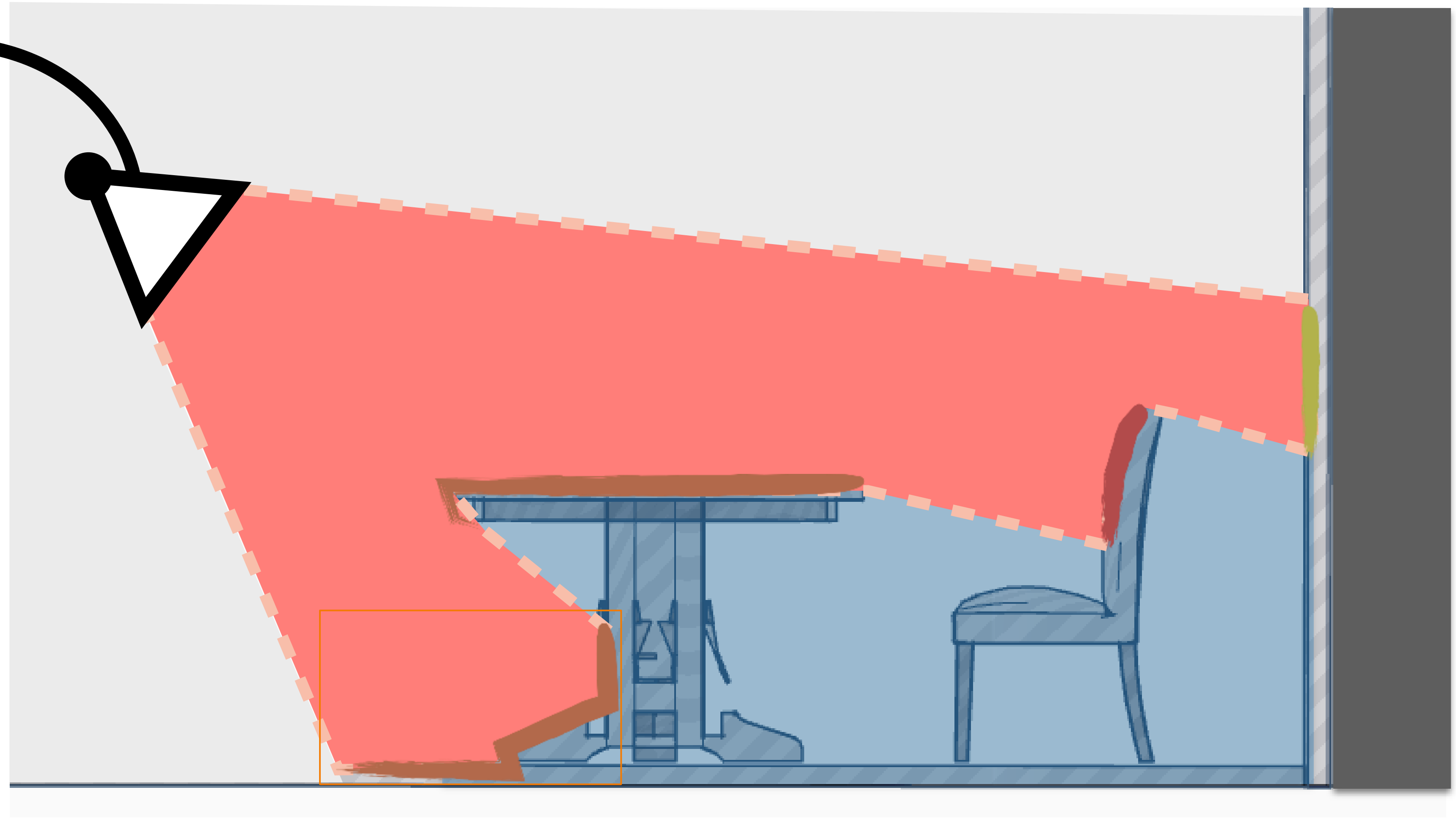


3D Scene

Semantic Scene Completion

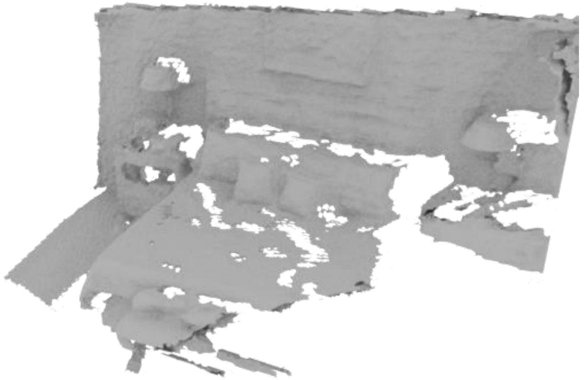
1) Semantic Scene Completion (label ALL voxels, not just visible ones)



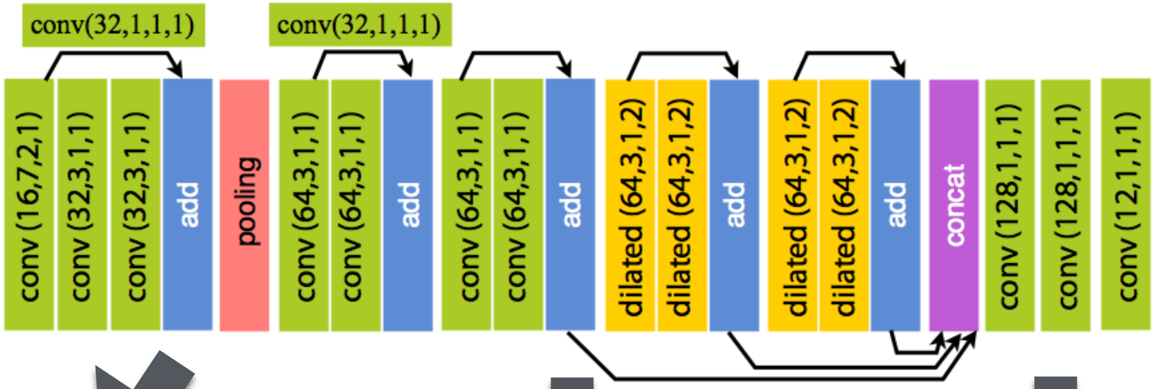


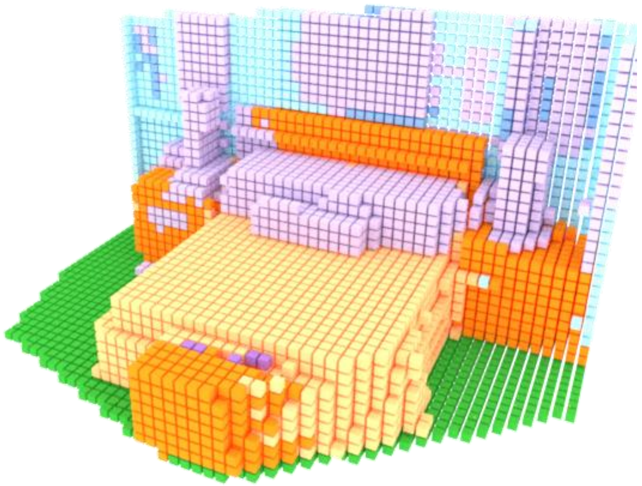
3D Scene

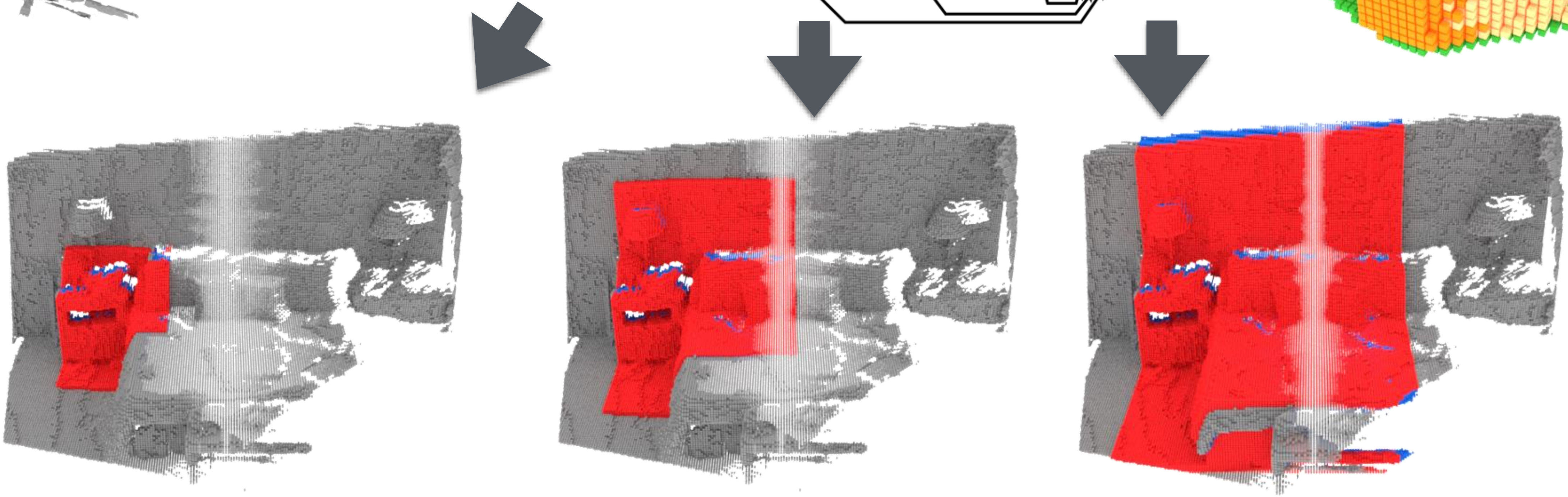
Semantic Scene Completion : “SSCNet”



Multi-scale aggregation







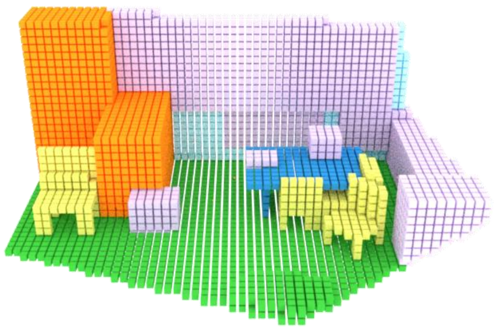
Receptive field: 0.98 m

Receptive field:1.62 m

Receptive field: 2.26 m

Semantic Scene Completion : “SSCNet”

1) Semantic Scene Completion results



Ground Truth

SSCNet

method	training	prec.	recall	IoU
Zheng *et al.* [36]	NYU	60.1	46.7	34.6
Firman *et al.* [3]	NYU	66.5	69.7	50.8
SSCNet completion	NYU	66.3	96.9	64.8
SSCNet joint	NYU	75.0	92.3	70.3
SSCNet joint	SUNCG+NYU	**75.0**	**96.0**	**73.0**

Comparison to previous algorithms for volumetric completion

	scene completion			semantic scene completion											
method (train)	prec.	recall	IoU	ceil.	floor	wall	win.	chair	bed	sofa	table	tvs	furn.	objs.	avg.
Lin *et al.* (NYU) [17]	58.5	49.9	36.4	0	11.7	13.3	**14.1**	9.4	29	24	6.0	7.0	16.2	1.1	12.0
Geiger and Wang (NYU) [4]	65.7	58	44.4	10.2	62.5	19.1	5.8	8.5	40.6	27.7	7.0	6.0	22.6	5.9	19.6
SSCNet (NYU)	57.0	**94.5**	55.1	15.1	**94.7**	24.4	0	12.6	32.1	35	13	7.8	27.1	10.1	24.7
SSCNet (SUNCG)	55.6	91.9	53.2	5.8	81.8	19.6	5.4	12.9	34.4	26	13.6	6.1	9.4	7.4	20.2
SSCNet (SUNCG+NYU)	**59.3**	92.9	**56.6**	**15.1**	94.6	**24.7**	10.8	**17.3**	**53.2**	**45.9**	**15.9**	**13.9**	**31.1**	**12.6**	**30.5**

Comparison to previous algorithms for 3D model fitting

What Else Can Be Done with SUN3D?

What Can Be Done With SUNCG?

2) Learn from synthetic images

- Rendered 400K synthetic images with Metropolis Light Transport in Mitsuba

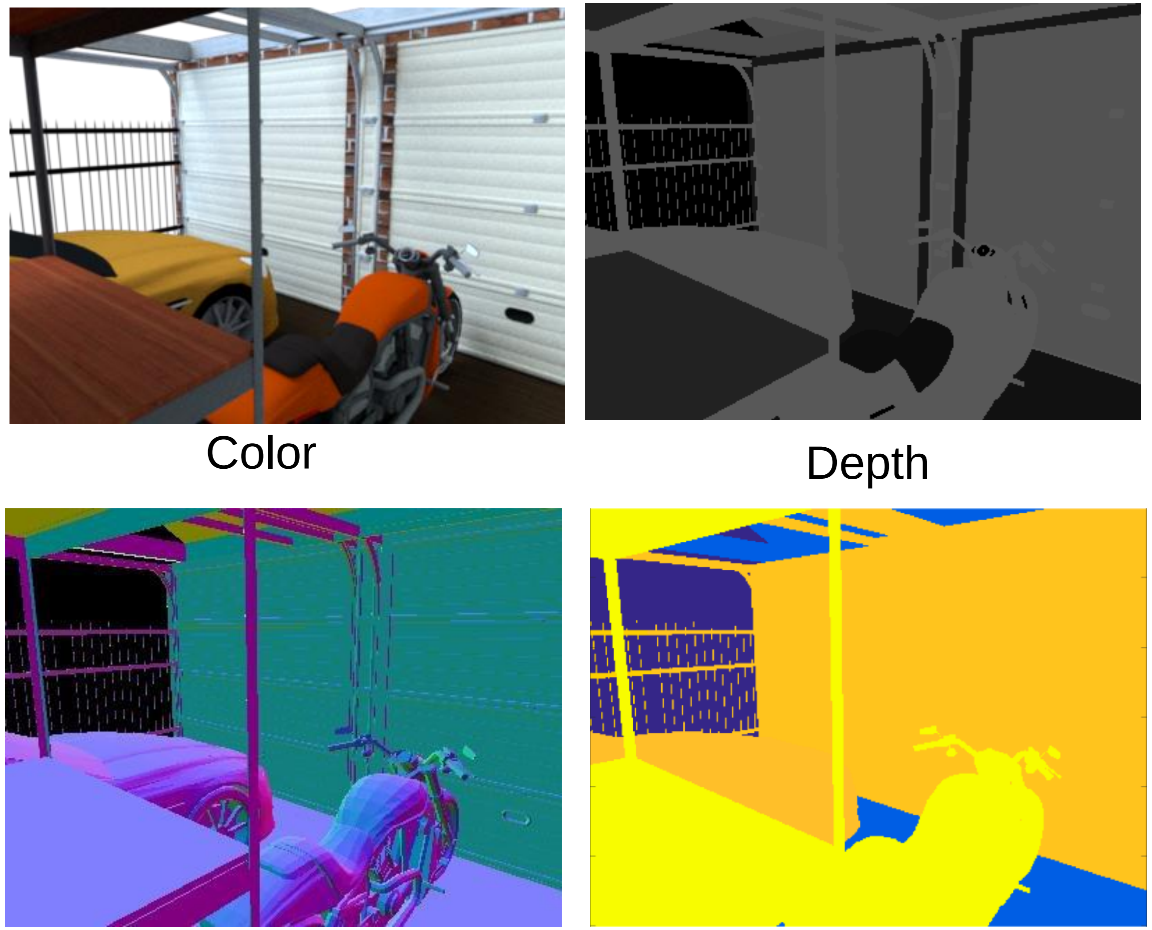
Y. Zhang, S. Song, E. Yumer, M. Savva, J. Lee, H. Jin, T. Funkhouser "Physically-Based Rendering for Indoor Scene Understanding Using CNNs," CVPR 2017.

What Can Be Done With SUNCG?

2) Learn from synthetic images

- Rendered 400K synthetic images with Metropolis Light Transport in Mitsuba

- All images annotated with depths, normals, boundaries, segmentations, labels, etc.



Color Depth

Normal Segmentation

Y. Zhang, S. Song, E. Yumer, M. Savva, J. Lee, H. Jin, T. Funkhouser "Physically-Based Rendering for Indoor Scene Understanding Using CNNs," CVPR 2017.

What Can Be Done With SUNCG?

2) Learn from synthetic images

- Rendered 400K synthetic images with Metropolis Light Transport in Mitsuba
- All images annotated with depths, normals, boundaries, segmentations, labels, etc.
- Experiments show that pre-training on these images improves performance on 3 scene understanding tasks
… and better rendering helps more

Pre-Train	Finetune	Selection	Mean (°) ↓	Median(°) ↓
Eigen *et al.* [8]			22.2	15.3
NYUv2			27.30	21.12
MLT Object	-	-	48.78	47.49
MLT-OL	-	No	49.33	42.30
MLT-IL/OL	-	No	29.33	22.62
MLT-IL/OL	-	Yes	28.59	22.61
OPENGL-DL	-	Yes	36.89	31.97
OPENGL-IL	-	Yes	35.93	30.91
OPENGL-IL	NYUv2	Yes	23.65	15.71
MLT-IL/OL	NYUv2	Yes	**22.06**	**14.78**

Normal Estimation Errors (degrees)

Pre-train	Finetune	OSD↑	OIS↑	AP↑	R50↑
NYUv2[28]	-	0.713	0.725	0.711	0.267
OPENGL-IL	-	0.523	0.555	0.511	0.504
MLT-IL/OL	-	0.604	0.621	0.587	0.749
OPENGL-IL	NYUv2	0.716	0.729	0.715	**0.893**
MLT-IL/OL	NYUv2	**0.725**	**0.736**	**0.720**	0.887

Boundary Estimation Accuracy

Input	Pre-train	Mean IoU
HHA	ImageNet	4.1
	ImageNet+OpenGL	4.3
RGB	Long *et al.* [16]	31.6
	Yu *et al.* [29]	31.7
	ImageNet + OPENGL-DL	32.8
	ImageNet + OPENGL-IL	32.9
	ImageNet + MLT-IL/OL	**33.2**

Semantic Segmentation Accuracy (%)

Y. Zhang, S. Song, E. Yumer, M. Savva, J. Lee, H. Jin, T. Funkhouser "Physically-Based Rendering for Indoor Scene Understanding Using CNNs," CVPR 2017.

Outline of This Talk

New 3D datasets for indoor scene understanding research:

	Synthetic	RGB-D Image	RGB-D Video
Room	SUNCG	SUN RGB-D	ScanNet
Multiroom	SUNCG		SUN3D

Outline of This Talk

New 3D datasets for indoor scene understanding research:

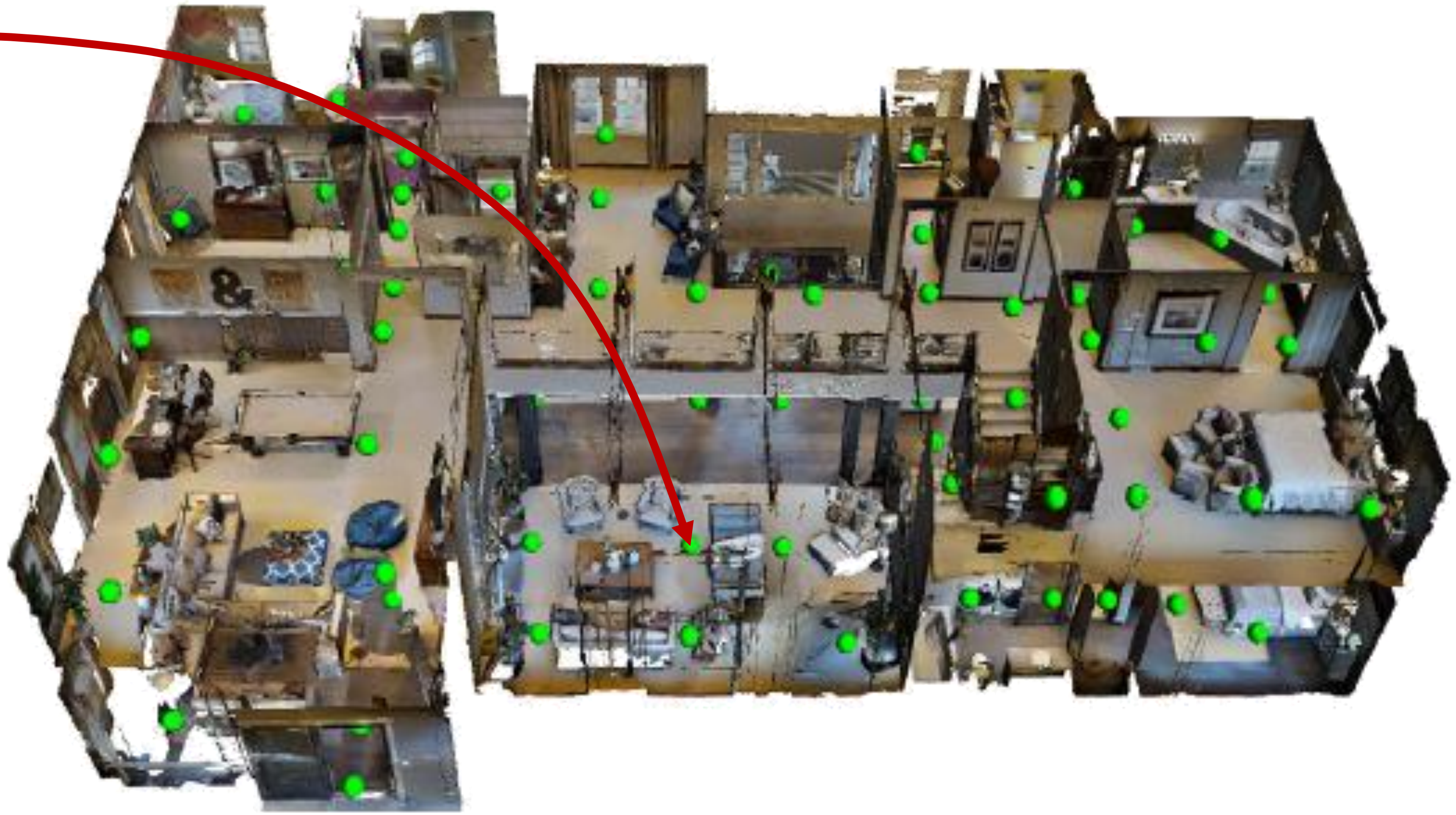
	Synthetic	RGB-D Image	RGB-D Video
Room	SUNCG	SUN RGB-D	ScanNet
Multiroom	SUNCG	Matterport3D	SUN3D

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

Matterport Camera

RGB-D Panorama



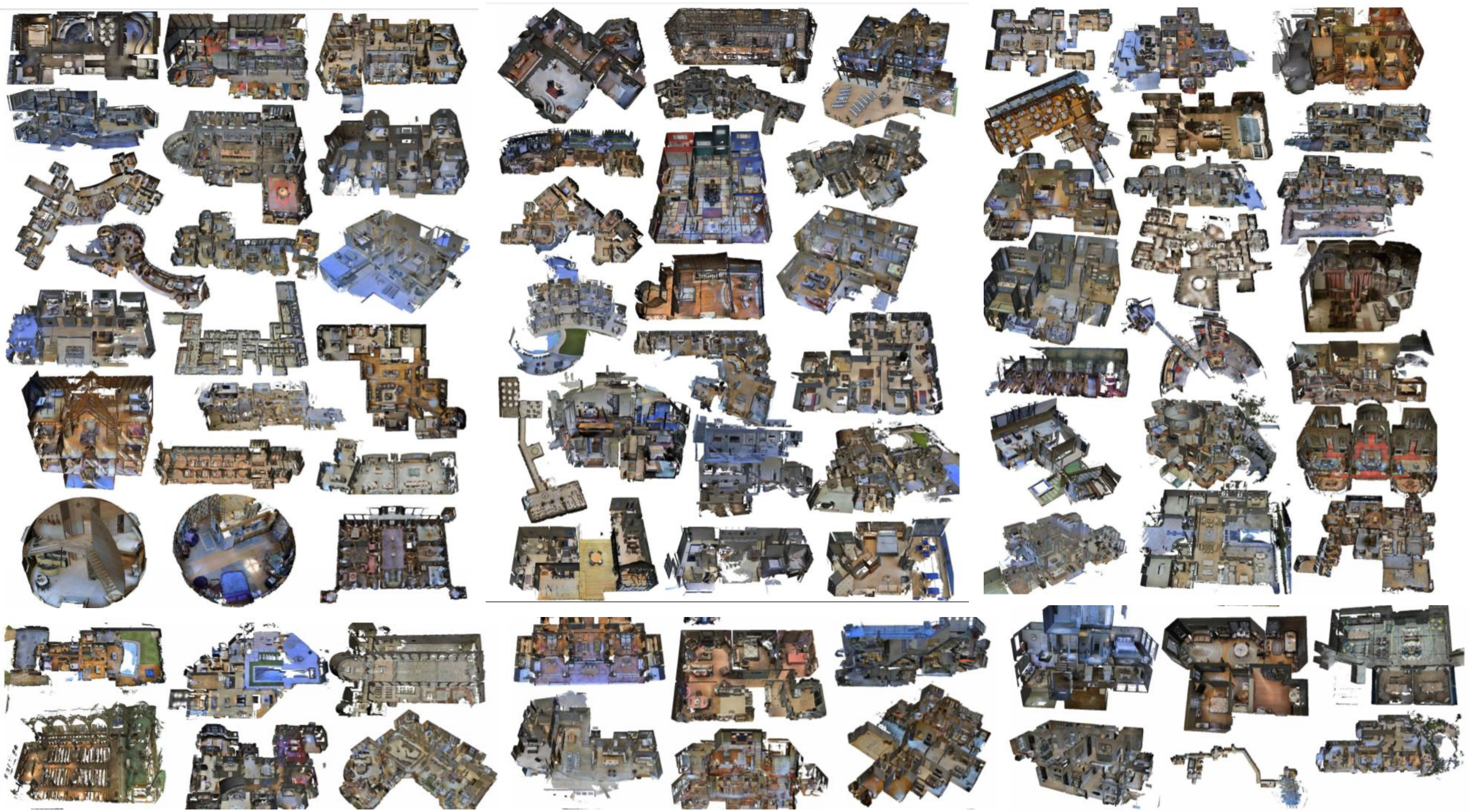
3D Textured Mesh Reconstruction

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matteport3D: Learning from RGB-D Data in Indoor Environments," 3DV 2017

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

- 90 Buildings
- 231 Floors
- 1K Rooms
- 11K Panoramas
- 194K Images
- 46K m^2



A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matteport3D: Learning from RGB-D Data in Indoor Environments," 3DV 2017

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

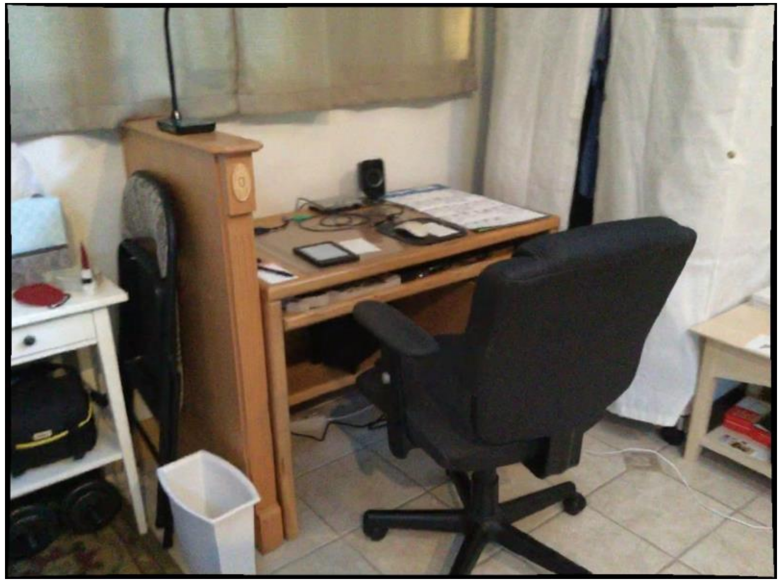
- Entire buildings
- Mostly personal living spaces
- Comprehensive coverage

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

- Calibrated panoramas
- Stationary cameras
- 1280x1024 images
- HDR color

SUN3D



ScanNet

Matterport3D

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

- Calibrated panoramas
- Stationary cameras
- 1280x1024 images
- HDR color

SUN3D

ScanNet

Matterport3D

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

- Calibrated panoramas
- Stationary cameras
- 1280x1024 images
- HDR color

SUN3D

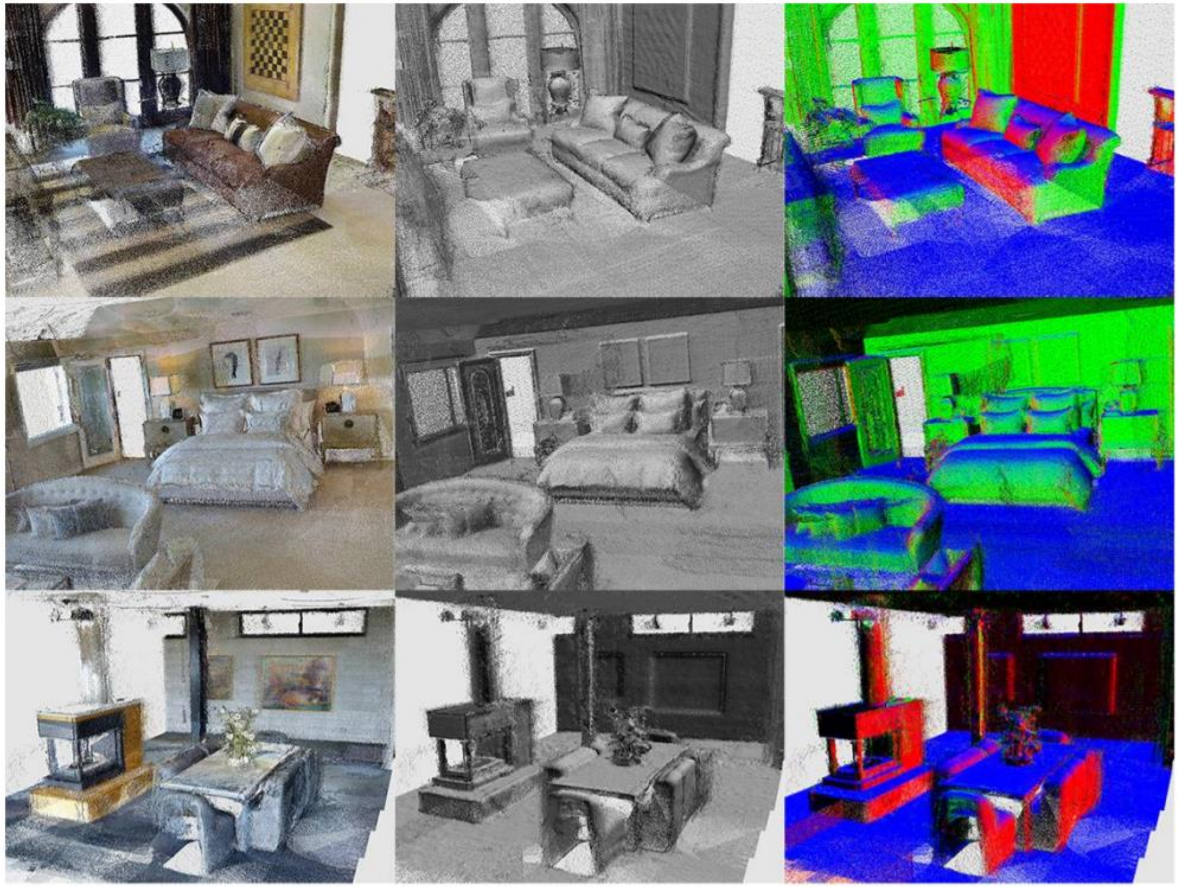
ScanNet

Matterport3D

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

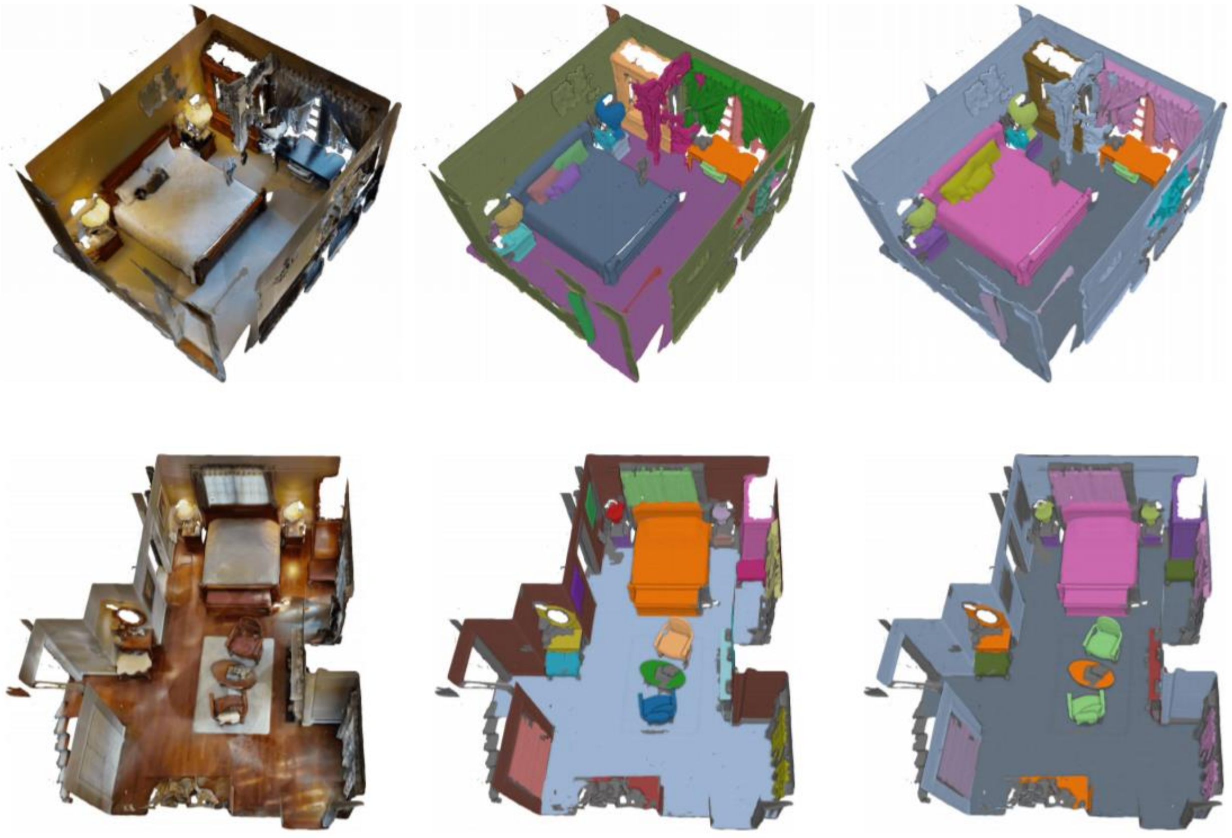
- Evenly-spaced view sampling (panorama are ~2.25m apart)
- Precise global alignment
- Textured mesh reconstruction



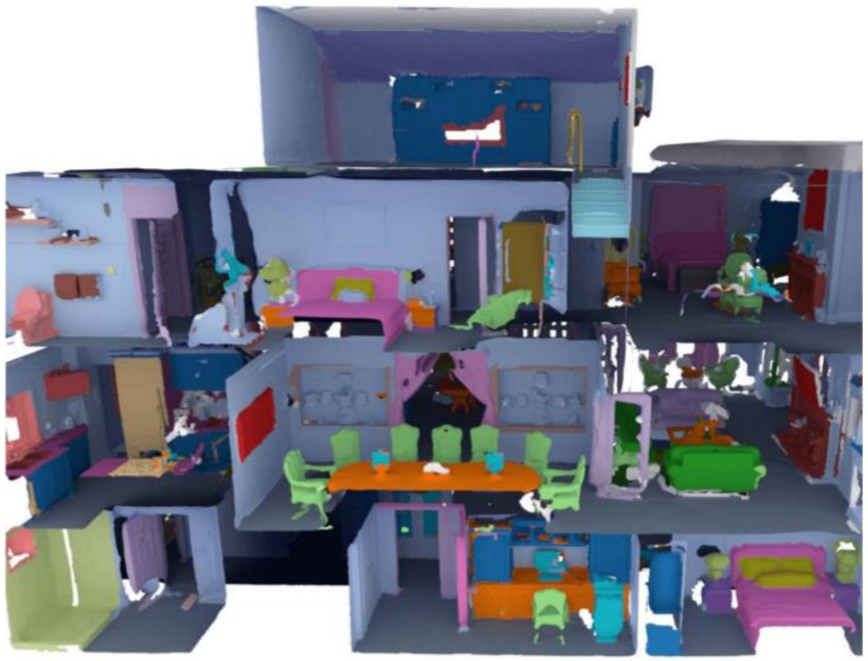
Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

- 50K object segmentations and labels



Textured Mesh Segmentation Labels

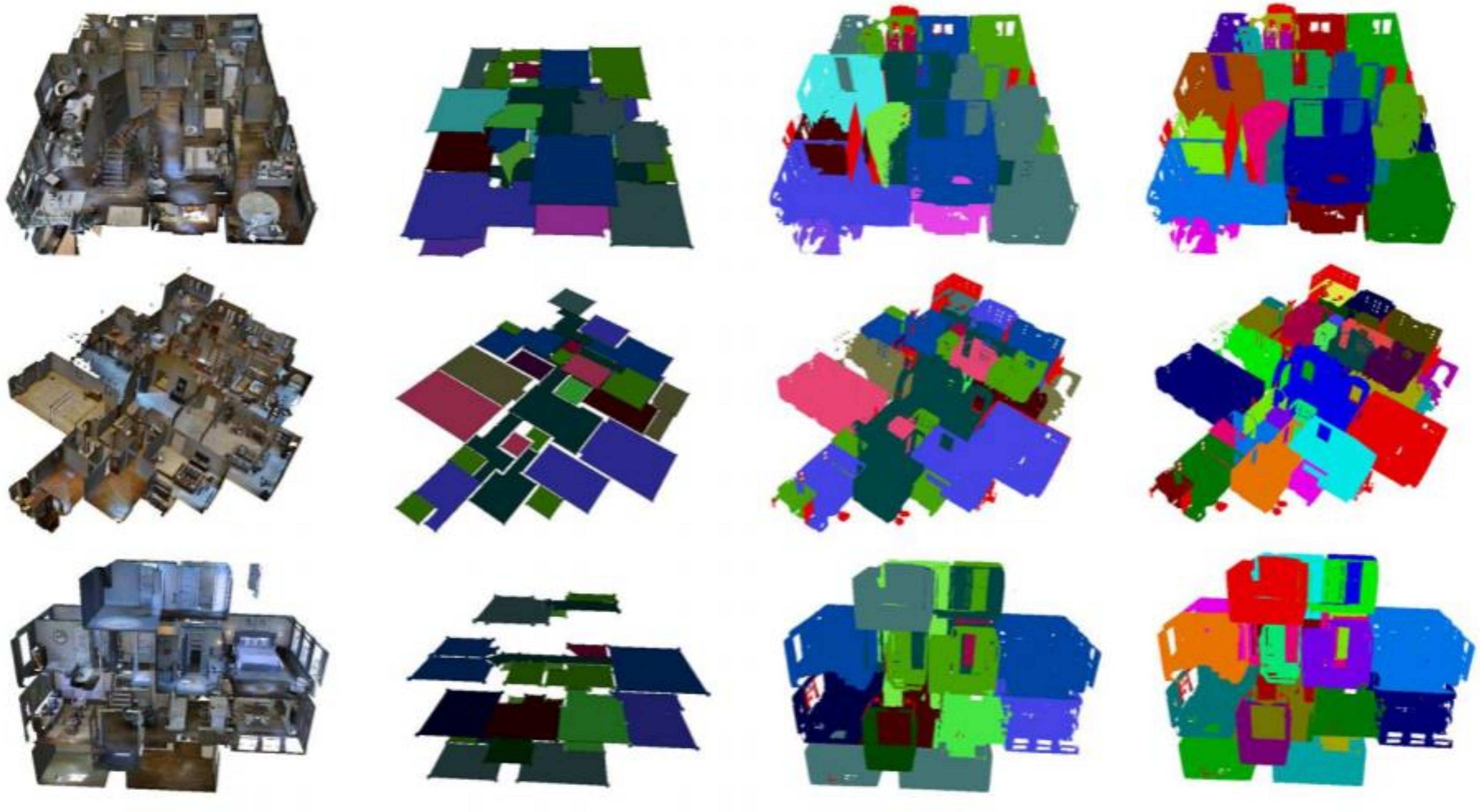


Labels

Matterport3D

Annotated 3D reconstructions of large spaces scanned with RGB-D panoramas

- 2K region segmentations and labels

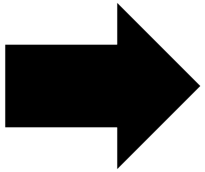


What Can Be Done with Matterport3D?

What Can Be Done With Matterport3D?

View classification

- Given an arbitrary RGB image, predict what type of room contains the camera



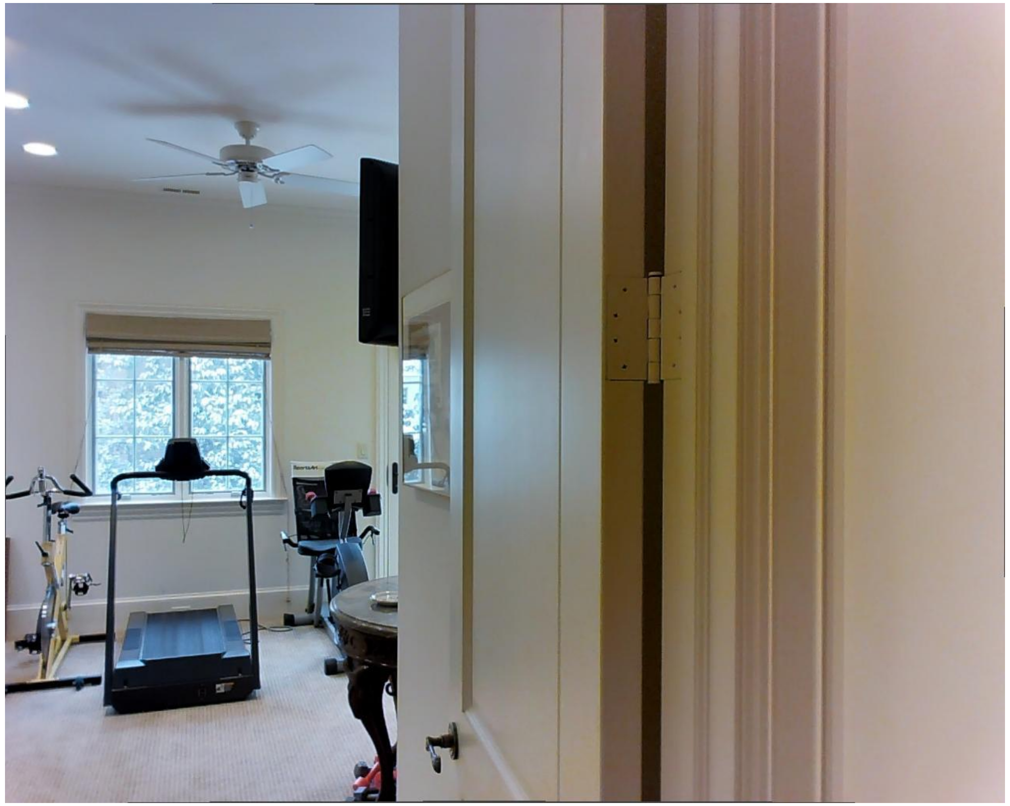
Living Room

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Given an arbitrary RGB image, predict what type of room contains the camera



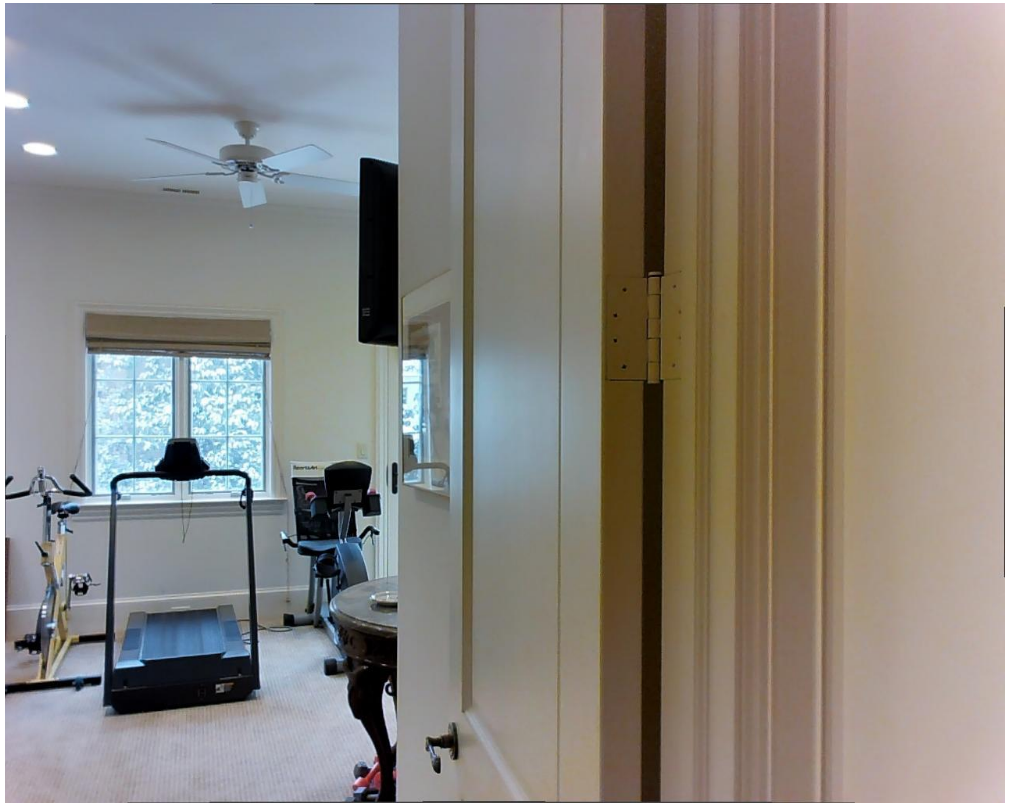
?

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matteport3D: Learning from RGB-D Data in Indoor Environments," 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Given an arbitrary RGB image, predict what type of room contains the camera



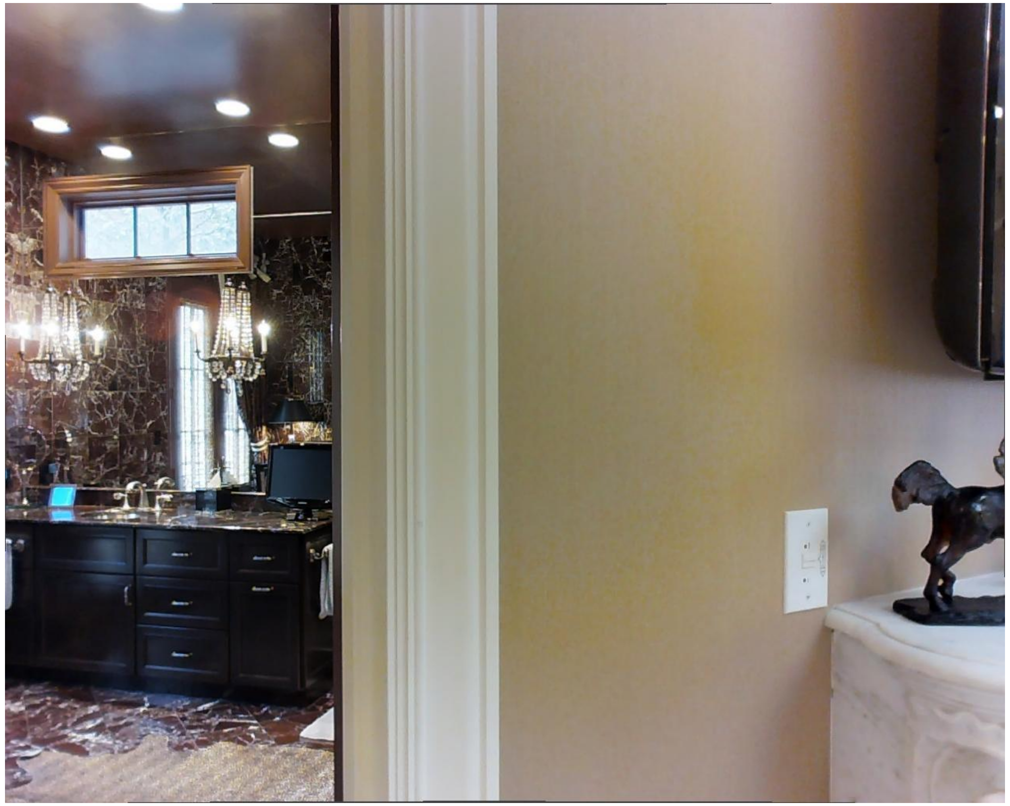
Hallway

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Given an arbitrary RGB image, predict what type of room contains the camera



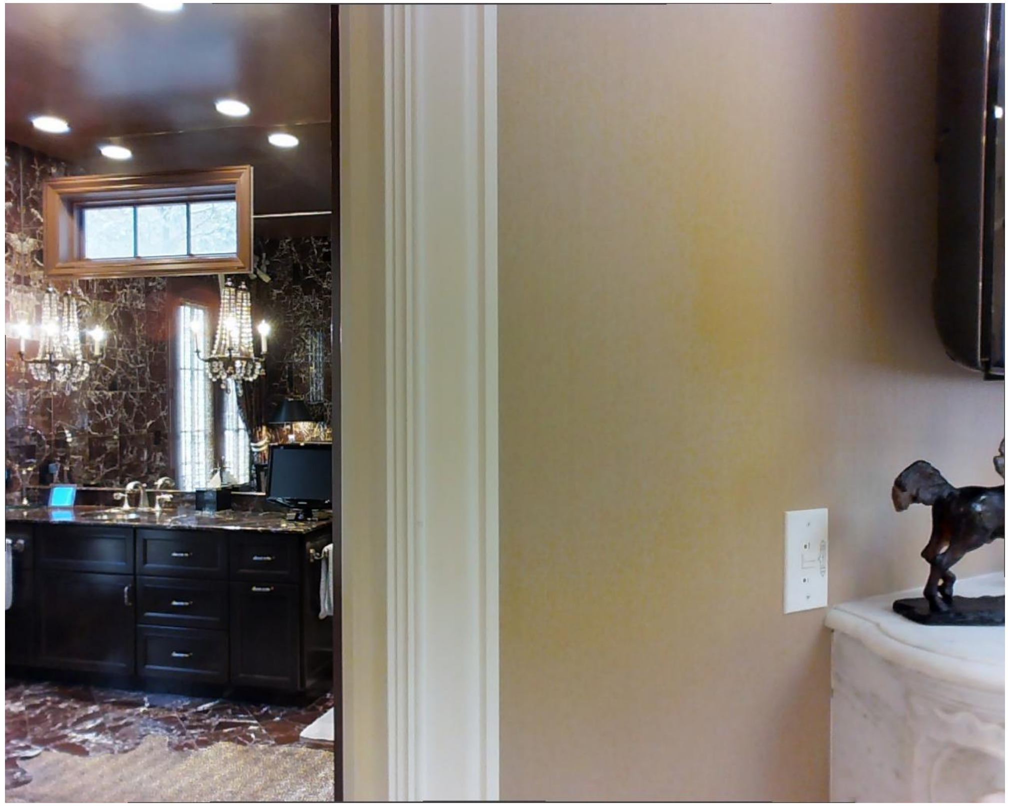
?

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Given an arbitrary RGB image, predict what type of room contains the camera



Bedroom

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Given an arbitrary RGB image, predict what type of room contains the camera

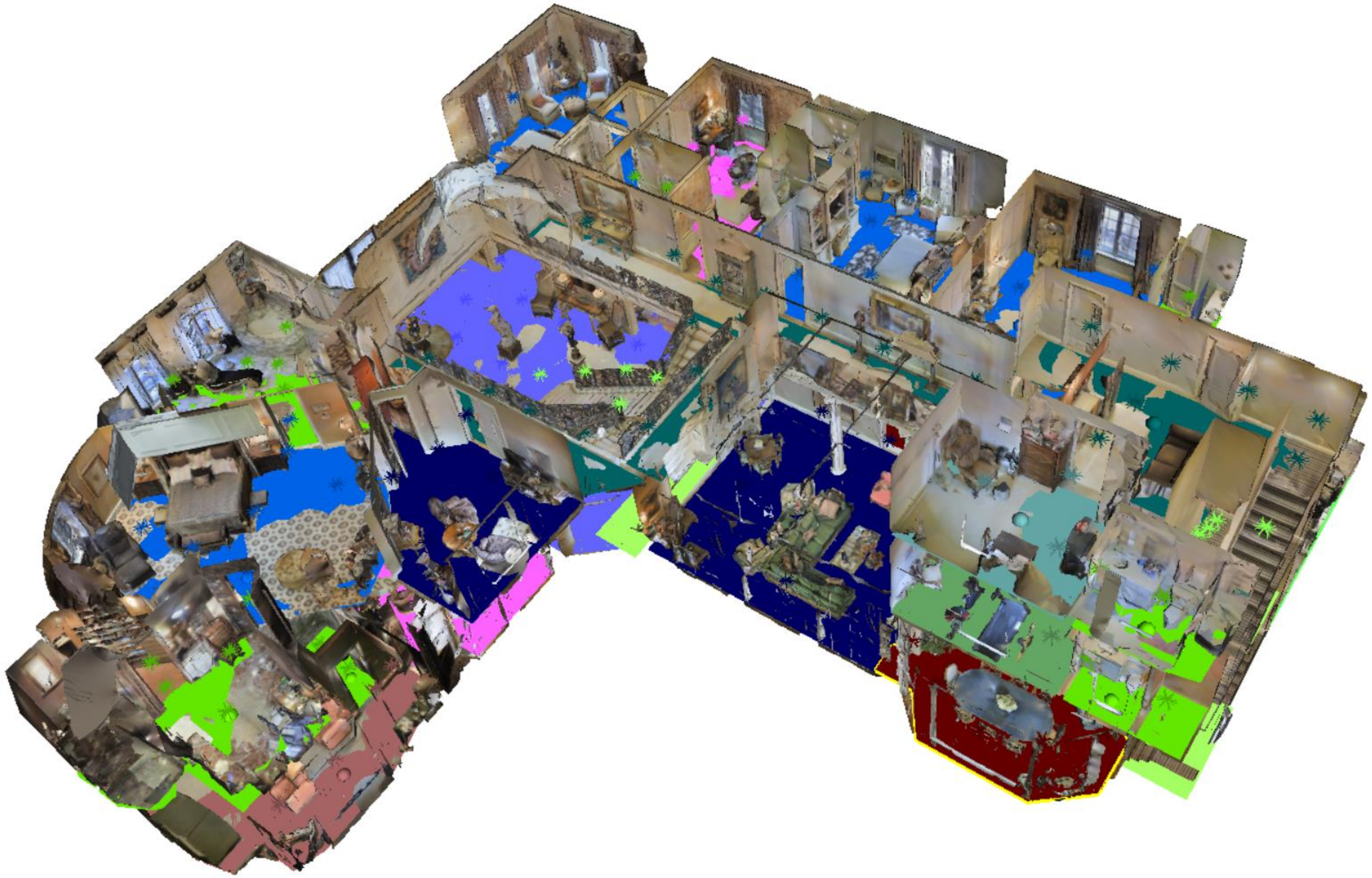
?

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Can use the region annotations to classify views for training and testing

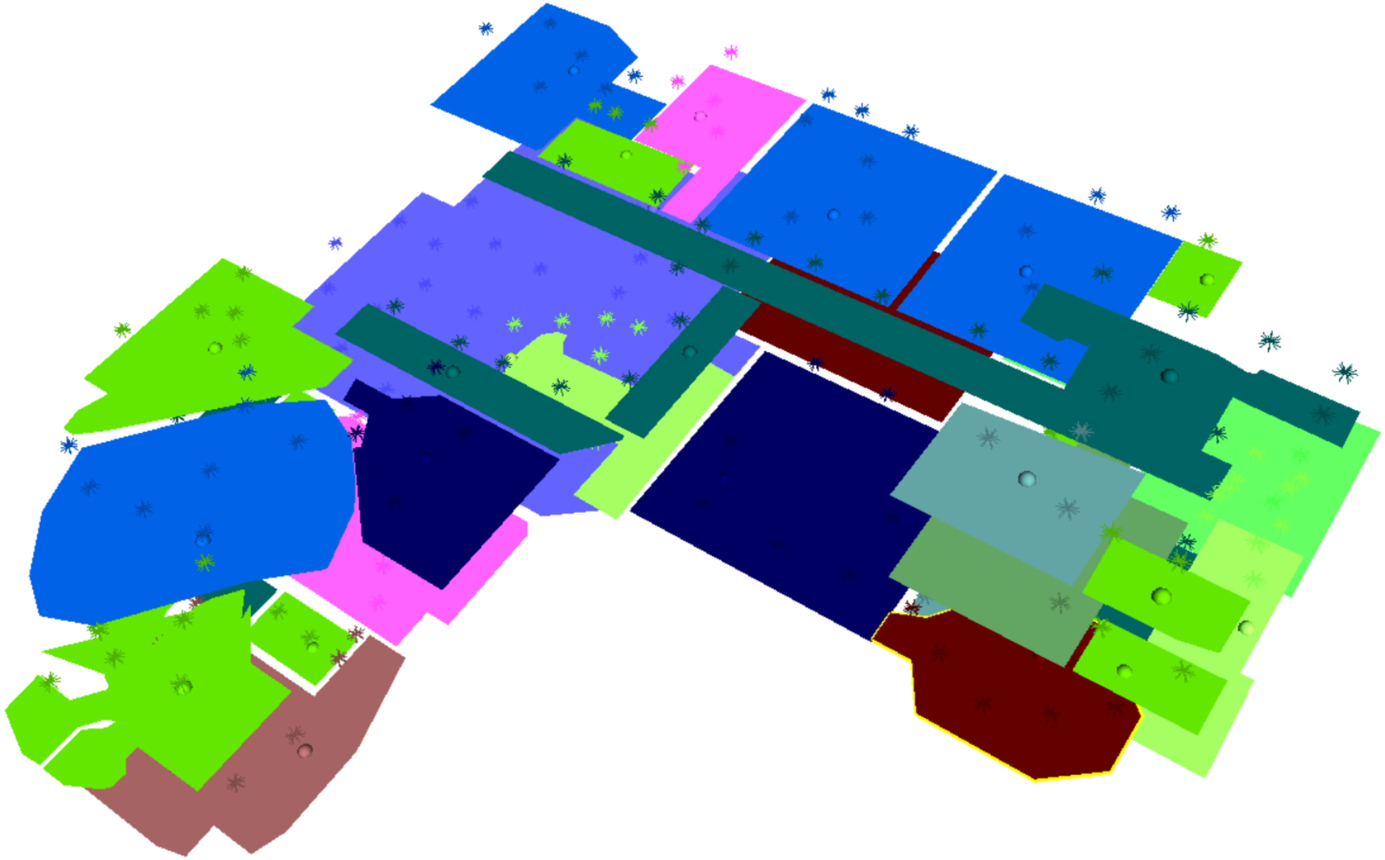


A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Can use the region annotations to classify views for training and testing



A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matteport3D: Learning from RGB-D Data in Indoor Environments," 3DV 2017

What Can Be Done With Matterport3D?

View classification

- Results for ResNet-50

class	office	lounge	familyroom	entryway	dining room	living room	stairs	kitchen	porch	bathroom	bedroom	hallway
single	20.3	21.7	16.7	1.8	20.4	27.6	49.5	52.1	57.4	44.0	43.7	44.7
pano	26.5	15.4	11.4	3.1	27.7	34.0	60.6	55.6	62.7	65.4	62.9	66.6

Classification accuracies (%)

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, “Matteport3D: Learning from RGB-D Data in Indoor Environments,” 3DV 2017

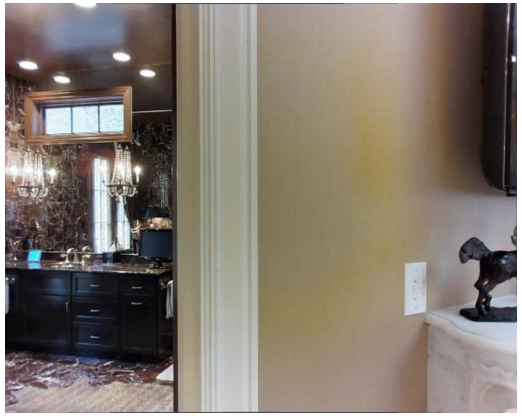
What Can Be Done With Matterport3D?

View classification

- Results for ResNet-50

class	office	lounge	familyroom	entryway	dining room	living room	stairs	kitchen	porch	bathroom	bedroom	hallway
single	20.3	21.7	16.7	1.8	20.4	27.6	49.5	52.1	57.4	44.0	43.7	44.7
pano	26.5	15.4	11.4	3.1	27.7	34.0	60.6	55.6	62.7	65.4	62.9	66.6

Classification accuracies (%)



Single image

Panoramic image

A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matteport3D: Learning from RGB-D Data in Indoor Environments," 3DV 2017

Summary and Conclusion

3D datasets are just now becoming available – they provide new opportunities for research in 3D scene understanding

	Synthetic	RGB-D Image	RGB-D Video
Room	SUNCG	SUN RGB-D	ScanNet
Multiroom	SUNCG	Matterport3D	SUN3D

I think each of these datasets is the largest and most richly-annotated of its kind

Future Work

More data:

- Internet-scale 3D scanning?

Richer annotations:

- Lighting, materials, physical properties, etc.

Multimedia data associations:

- Images, CAD models, floorplans, etc.

Real-time scene understanding tasks:

- Real-time scene parsing
- Robot navigation

Acknowledgments

Students and postdocs:

- Angel Chang, Maciej Halber, Jerry Liu, Manolis Savva, Shuran Song, Fisher Yu, Yinda Zhang, Andy Zeng

Collaborators:

- Angela Dai, Matthias Niessner, Ersin Yumer, Matt Fisher, Jianxiong Xiao, Kyle Simek, Craig Reynolds, Matt Bell

Data:

- Matterport, Planner5D

Funding:

- NSF, Facebook, Intel, Google, Adobe, Pixar

Thank You!

Dataset Webpages

- SUN3D – http://sun3d.cs.princeton.edu
- SUN RGB-D – http://rgbd.cs.princeton.edu
- SUNCG – http://suncg.cs.princeton.edu
- ScanNet – http://www.scan-net.org
- Matterport3D – http://github.com/niessner/Matterport

- ShapeNet – http://shapenet.org
- ModelNet – http://modelnet.cs.princeton.edu
- LSUN – http://lsun.cs.princeton.edu