SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite

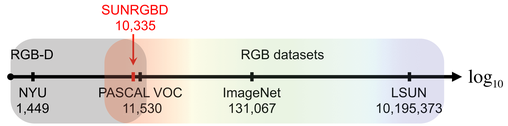
PRINCETON VISION GROUP

Shuran Song Samuel P. Lichtenberg Jianxiong Xiao

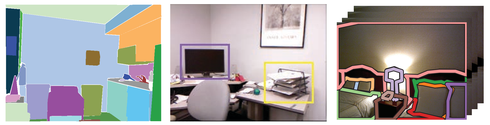
Data & Code:
http://rgbd.cs.princeton.edu

Motivation

RGB-D sensors have also enabled rapid progress for scene understanding. However, the **small dataset size** has become a major bottleneck.



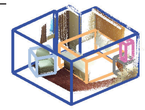
Another problem of existing RGB-D datasets is that most of them are only labeled in **2D**.



NYU depth V2 B3DO SUN3D

We introduce SUN RGB-D, a Pascal-scale RGB-D scene understanding dataset, which has 2D and 3D annotation for both objects and rooms.

	NYU depth V2	SUN RGB-D
Size	1,449	10,335
Sensor Type	1	4
Object Annotation	2D segmentation	2D segmentation 3D Object
Object Pose	None	Yes
Room Annotation	None	Room Layout



SUN RGB-D

References

[NYU] Indoor Segmentation and Support Inference from RGBD Images. N. Silberman, P. Kohli, D. Hoiem and R. Fergus. In ECCV, 2012.
[SUN3D] SUN3D: A database of big spaces reconstructed using SfM and object labels. J. Xiao, A. Owens, and A. Torralba. In ICCV, 2013.
[B3DO] A category-level 3-d object dataset: Putting the kinect to work. A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell. In ICCV Workshop, 2011.

Acknowledgement

This work is supported by gift funds from Intel.

Example Scenes

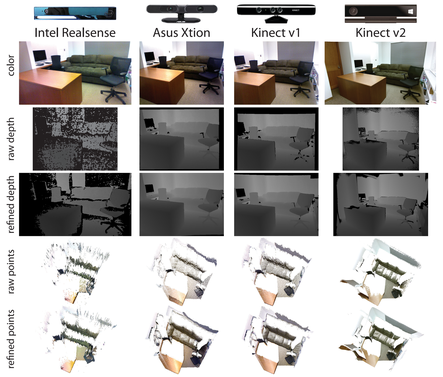
Example Objects

Data Capturing

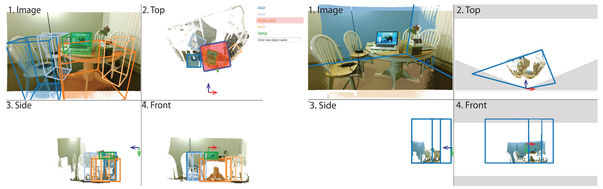
RealSense Kinect v2 and Battery Capturing Setup

RGB-D Sensors

	RealSense	Xtion	Kinect v1	Kinect v2
weight (pound)	0.077	0.5	4	4.5
size (inch)	5.2× 0.25× 0.75	7.1× 1.4× 2	11× 2.3× 2.7	9.8× 2.7× 2.7
power	2.5W USB	2.5W USB	12.96W	115W
depth resolution	628× 468	640× 480	640× 480	512× 424
color resolution	1920× 1080	640× 480	640× 480	1920× 1080

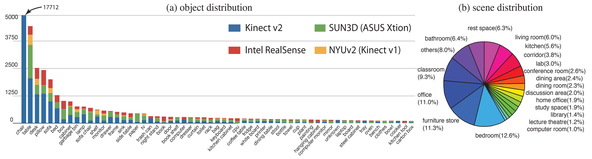


Annotation



Annotation Tool for 3D Object and 3D Room Layout

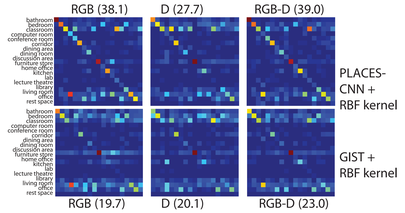
Examples of 2D and 3D Annotation



Statistics of Semantic Annotation

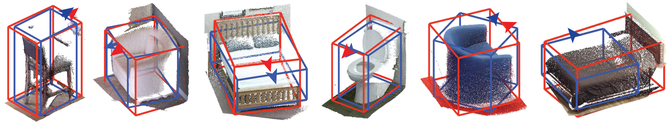
Benchmark Tasks

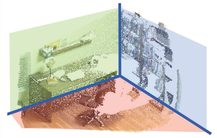
Scene Classification



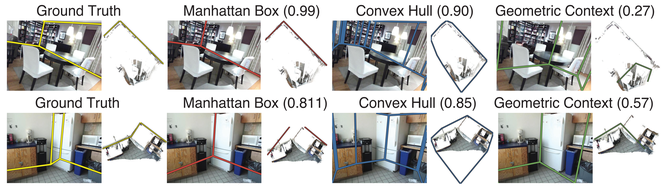
Semantic Segmentation

3D Object Detection and Pose Estimation

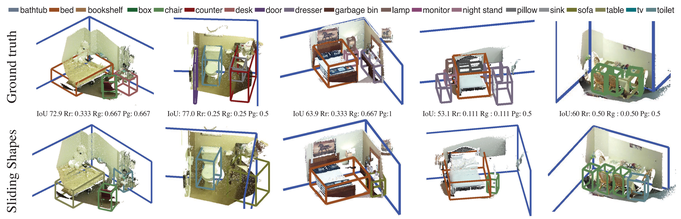


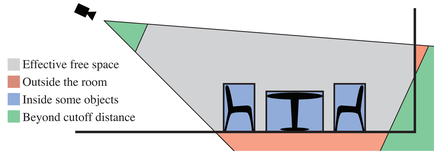


Room Layout Estimation



Total Scene Understanding





Free Space IoU

$$\text{Precision} = \frac{\#\text{ matched pairs with a correct category label}}{\#\text{ prediction boxes}}$$

$$\text{Recall} = \frac{\#\text{ matched pairs with a correct category label}}{\#\text{ ground truth boxes}}$$

Precision Recall for all objects