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

Shuran Song Samuel P. Lichtenberg Jianxiong Xiao
Princeton University
http://rgbd.cs.princeton.edu

1. Segmetation Result

	wall	floor	cabinet	bed	chair	sofa	table	door	window	bookshelf	picture	counter	blinds	desk	shelves	curtain	dresser	pillow	mirror
RGB NN	37.8	45.0	17.4	21.8	16.9	12.8	18.5	6.1	9.6	9.4	4.6	2.2	2.4	7.3	1.0	4.3	2.2	2.3	6.9
Depth NN	32.1	42.6	2.9	6.4	21.5	4.1	12.5	3.4	5.0	0.8	3.3	1.7	14.8	2.0	15.3	2.0	1.4	1.2	0.9
RGB-D NN	36.4	45.8	15.4	23.3	19.9	11.6	19.3	6.0	7.9	12.8	3.6	5.2	2.2	7.0	1.7	4.4	5.4	3.1	5.6
RGB flow	38.9	47.2	18.8	21.5	17.2	13.4	20.4	6.8	11.0	9.6	6.1	2.6	3.6	7.3	1.2	6.9	2.4	2.6	6.2
Depth flow	33.3	43.8	3.0	6.3	22.3	3.9	12.9	3.8	5.6	0.9	3.8	2.2	32.6	2.0	10.1	3.6	1.8	1.1	1.0
RGB-D flow	37.8	48.3	17.2	23.6	20.8	12.1	20.9	6.8	9.0	13.1	4.4	6.2	2.4	6.8	1.0	7.8	4.8	3.2	6.4
RGBD	43.2	78.6	26.2	42.5	33.2	40.6	34.3	33.2	43.6	23.1	57.2	31.8	42.3	12.1	18.4	59.1	31.4	49.5	24.8

	floormat	clothes	ceiling	books	fridge	tv	paper	towel	shower	box	board	person	nightstand	toilet	sink	lamp	bathtub	bag	mean
RGB NN	0.0	1.2	27.9	4.1	7.0	1.6	1.5	1.9	0.0	0.6	7.4	0.0	1.1	8.9	14.0	0.9	0.6	0.9	8.3
Depth NN	0.0	0.3	9.7	0.6	0.0	0.9	0.0	0.1	0.0	1.0	2.7	0.3	2.6	2.3	1.1	0.7	0.0	0.4	5.3
RGB-D NN	0.0	1.4	35.8	6.1	9.5	0.7	1.4	0.2	0.0	0.6	7.6	0.7	1.7	12.0	15.2	0.9	1.1	0.6	9.0
RGB flow	0.0	1.3	39.1	5.9	7.1	1.4	1.5	2.2	0.0	0.7	10.4	0.0	1.5	12.3	14.8	1.3	0.9	1.1	9.3
Depth flow	0.0	0.6	13.9	0.5	0.0	0.9	0.4	0.3	0.0	0.7	3.5	0.3	1.5	2.6	1.2	0.8	0.0	0.5	6.0
RGB-D flow	0.0	1.6	49.2	8.7	10.1	0.6	1.4	0.2	0.0	0.8	8.6	0.8	1.8	14.9	16.8	1.2	1.1	1.3	10.1
RGBD	5.6	27.0	84.5	35.7	24.2	36.5	26.8	19.2	9.0	11.7	51.4	35.7	25.0	64.1	53.0	44.2	47.0	18.6	36.3

Table 1. **Semantic segmentation.** We evaluate performance for 37 object categories. Here shows accuracy for each category, and the mean accuracy. (wall, floor, cabinet, bed, chair, sofa, table, door, window, bookshelf, picture, counter, blinds, desk, shelves, curtain, dresser, pillow, mirror, floor mat, clothes, ceiling, books, fridge, tv, paper, towel, shower curtain, box, whiteboard, person, nightstand, toilet, sink, lamp, bathtub, bag)

2. Annotation Tool

3D object annotation To perform the 3D object annotations, we instruct the oDesk workers as follows. The user interface of the annotation tool is shown in Figure 2, which is divided into 4 sections. We ask the oDesk workers to label all objects that they can identify. Labeling a single object consists of three main steps: defining the length and width of the objects box, giving the object a name, and defining the height of the objects box.

In the first step, the worker considers Section 2, a top view of the room, and begins to label the object by clicking three corners of the rectangle box. The first line segment created is considered to be the "front" of the object, and an orthogonal arrow will appear on that line segment in section 2. If the object has a natural orientation (such as a chair), then we ask the oDesk worker to begin from the correct "front" side; if it does not (such as a round table), then they can begin from any side they choose. Once the top view of the box is defined, worker need to type in the object category name. Each object is given its own label, even if the objects have the same semantic class, so as to discern individual instances. The oDesk workers are selected specifically for high English proficiency (it is extremely difficult to find suitable applicants from within the U.S.) as part of their interview process, so the name labels are far less noisy than they would be if obtained using Amazon Mechanical Turk. Finally, in the third step, the worker considers sections 3 and 4, which are orthogonal side views of the room, in order to adjust the top and bottom of the object's box appropriately. We require that the boxes the oDesk workers drawn be tight. If the object is not fully visible in the image, we ask the workers to try to guess the full size of the object when drawing the box.

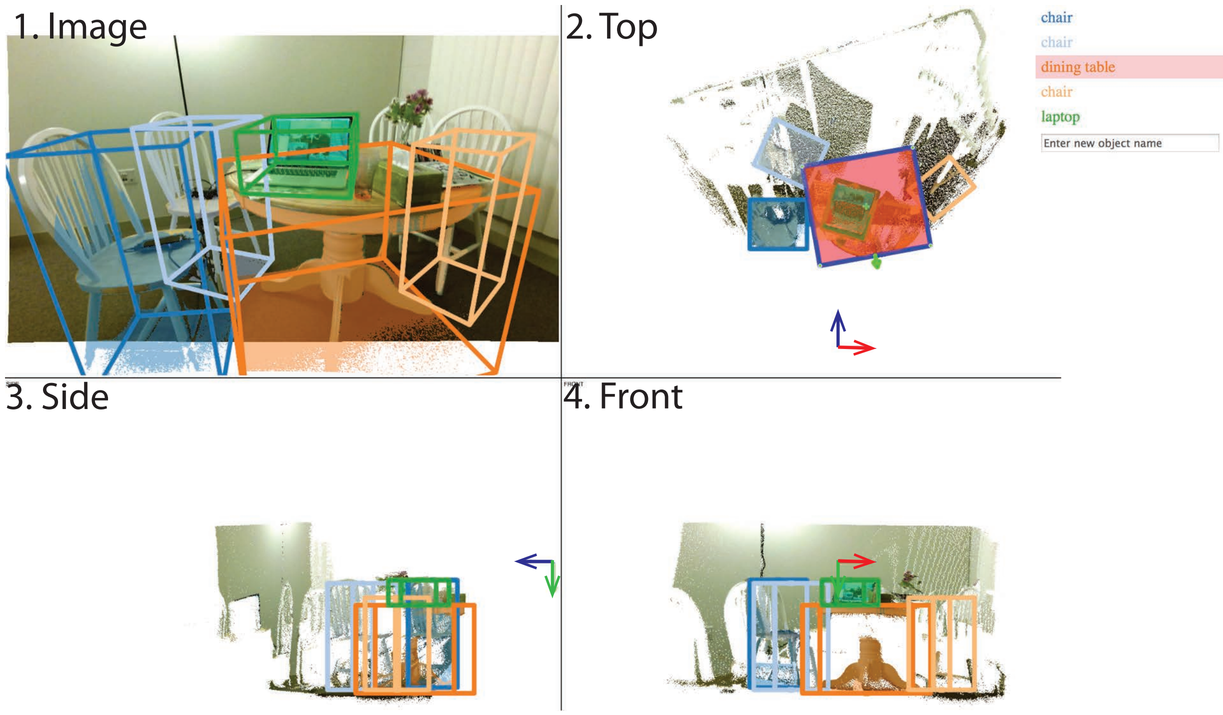

Figure 1. Annotation tool for 3D objects

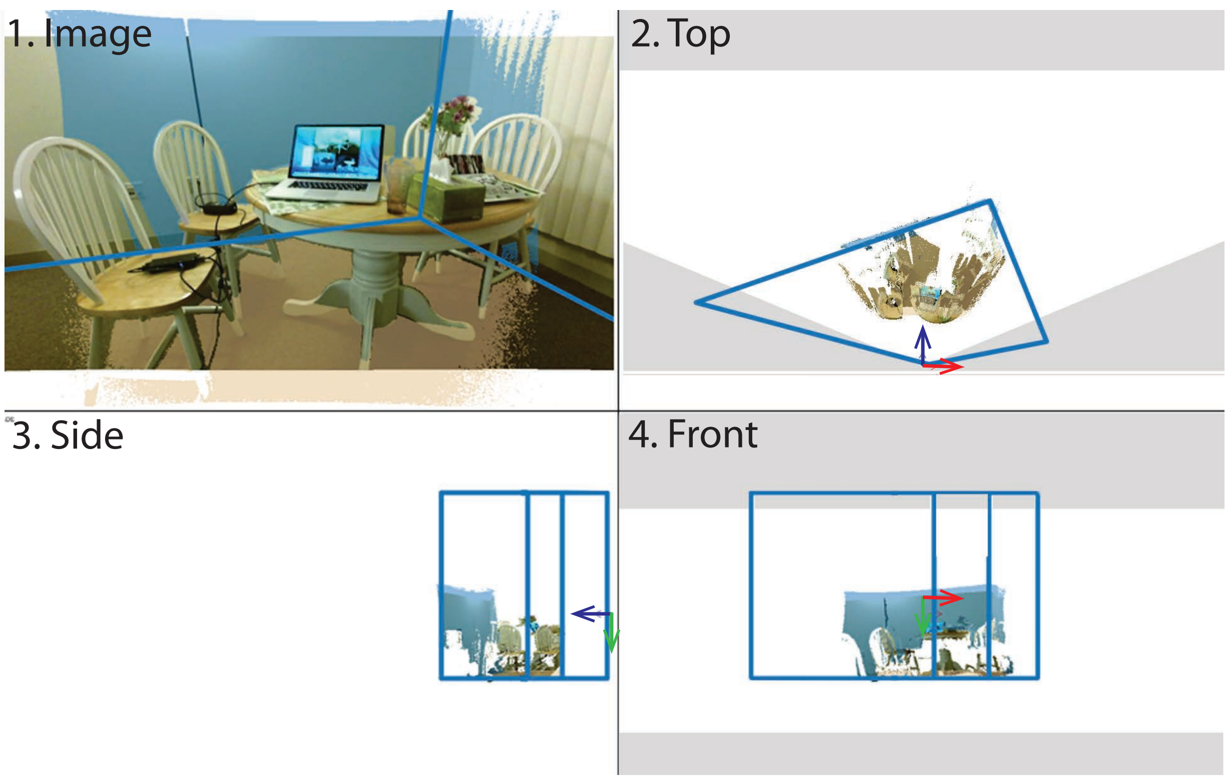

Figure 2. Annotation tool for room layout

After these three steps are completed, the worker has a completed box for the object. The labeling tool colors not only the box but also the points contained in that box onto the RGB image in section 1, which enables the worker to correct any mistakes. This real-time feedback is essential for obtaining high-quality boxes. This process is then repeated for every object in the image, for every image in the dataset.

3D Room Layout Annotation The 3D room layout labeling process is very similar to that of the 3D object annotation. The worker sees an identical interface however, with a gray area indicates the field of view. The room layout labeling process again follows a similar steps.

Firstly, the worker draws the boundary of the room, starting and ending at the camera position, in the top-down view of section 2. This time, the worker can connect line segments to draw an arbitrary polygon–not just a rectangle–to express the room structure. The worker is presented with a white area that covers all points in the room, and a gray border(outside the field of view). Any line segments used to help draw the room layout box, but that do not actually represent walls of the room, should lie only in the gray boundary area.

There is an additional complicating factor: oftentimes rooms will have hallways or other passageways leading out of the field of view of the camera. To handle this case, we ask workers to draw both the walls defining this passageway into the gray boundary area, and connect them there with a line segment, outside the white area considered to represent the room layout.

Then, the worker adjusts the height of the room layout box. The bottom of the box should align with the floor, and the top with the ceiling. If the floor is not clearly visible in the image, we ask the workers to make the best guess they can. If the ceiling is not visible in the image, we ask the workers to draw the ceiling up into a gray boundary area at the top of section 4. The process is repeated for each image in the dataset.

3. Dataset

In this section, we show more statistics of our SUN RGB-D dataset. Figure 3 shows the distributions of object 3D location for various categories. Figure 4 shows the distribution of object orientation. In Figure 5, we show the distribution of object 3D sizes. In Figure 6 and Figure 7, we show some examples for 2D and 3D annotations.

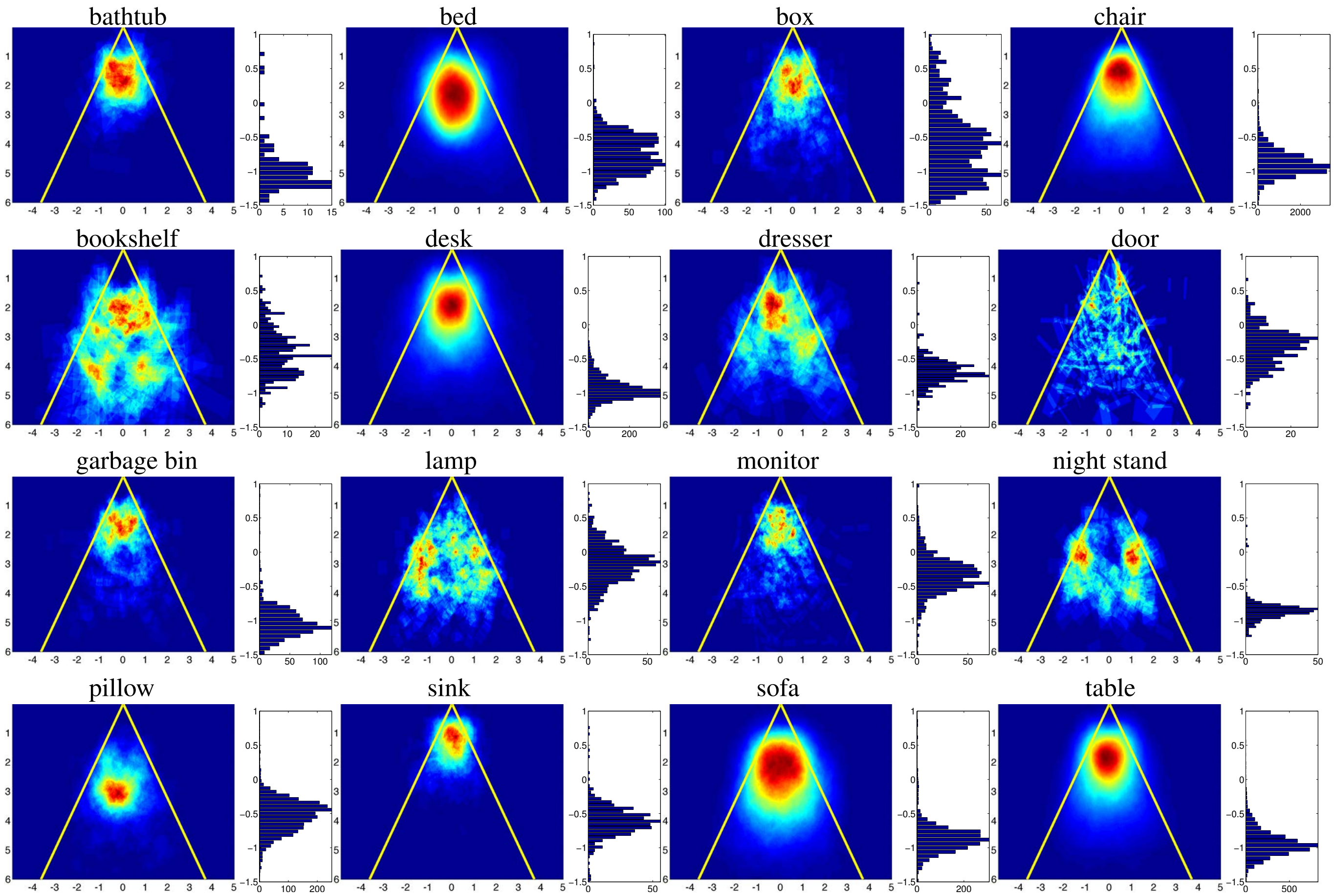

Figure 3. **3D location priors for selected object classes.** For each object category we show their 3D location distribution from top view (Left) and their center's height location distribution (right), in meters. We draw the Kinect v1's FOV (in yellow) for reference.

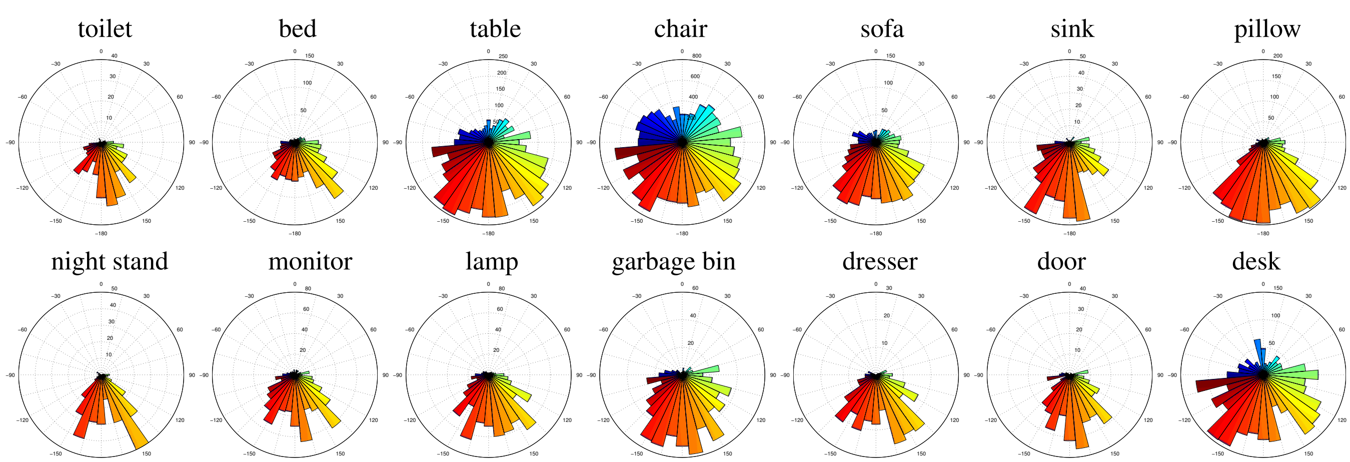

Figure 4. **3D orientation priors for selected object classes.**

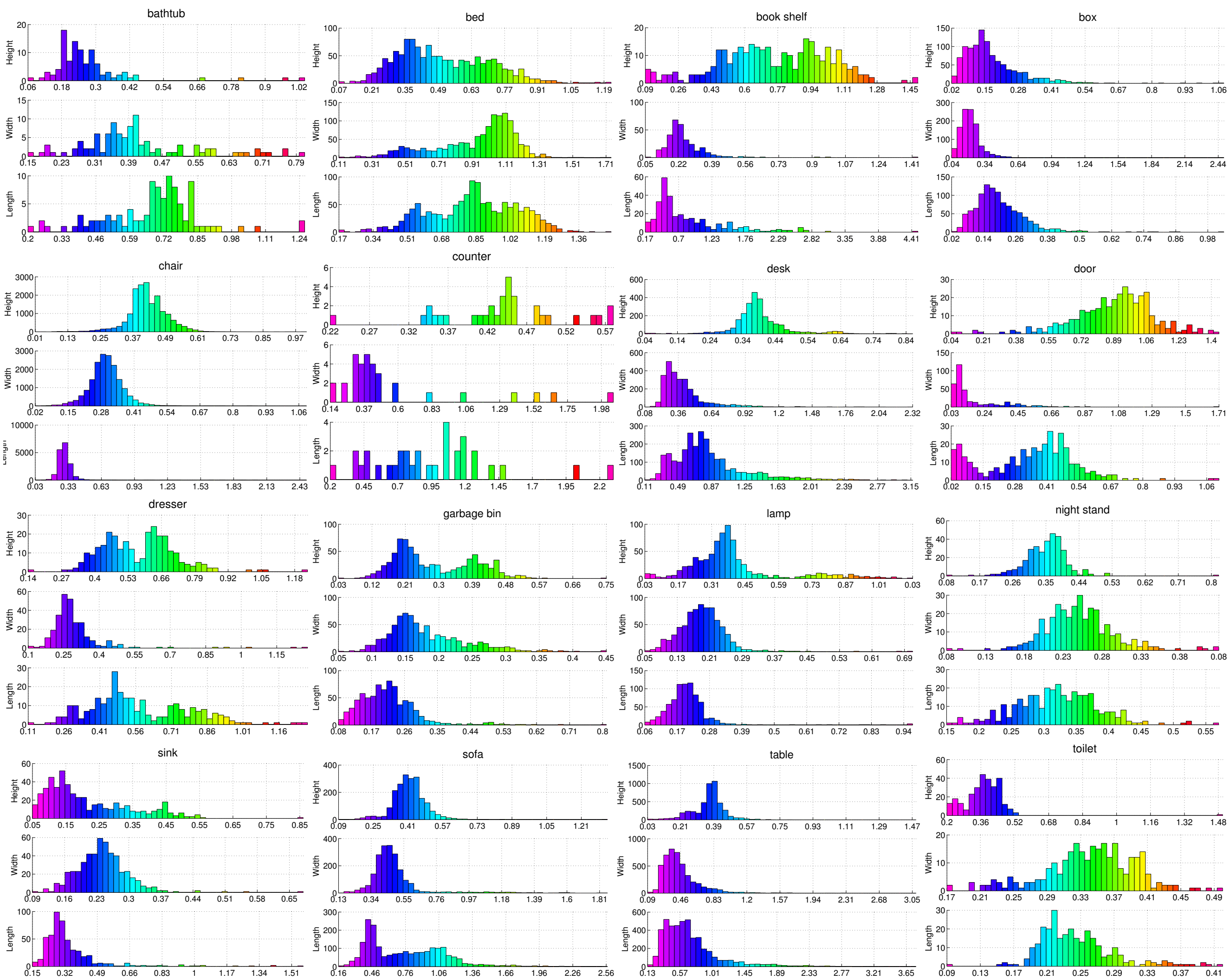

Figure 5. **Object 3D size distribution.**

Figure 6. **Examples of 2D Segmentation Annotation in our SUN RGB-D dataset**

Figure 7. **Examples of 2D Segmentation Annotation in our SUN RGB-D dataset**