

Just a Single Labeling for Structural Knowledge Modeling (Supplementary Materials)

Anonymous CVPR submission

Paper ID 610

1. Sample sets and some detection results

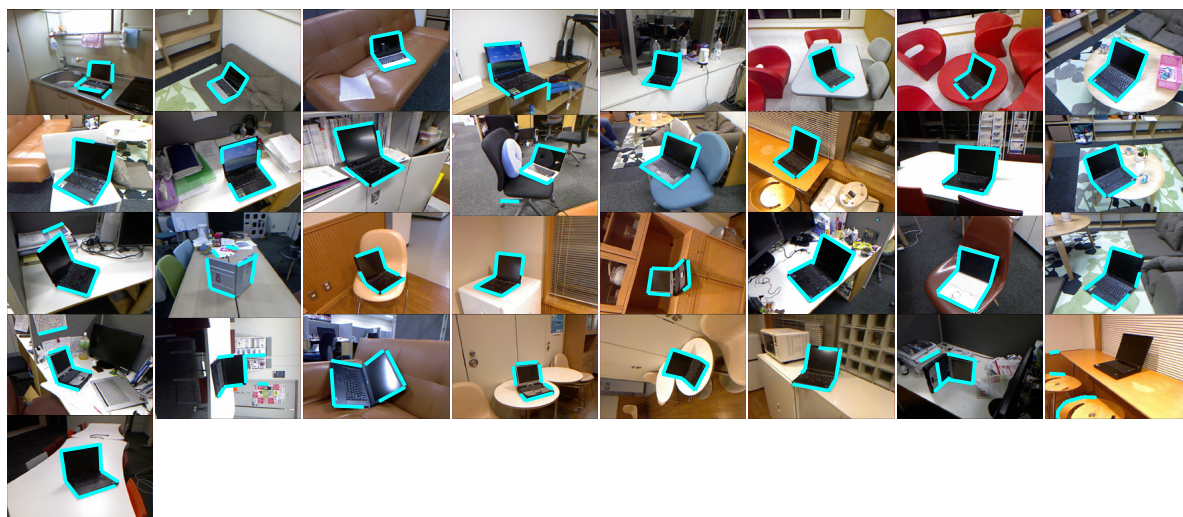
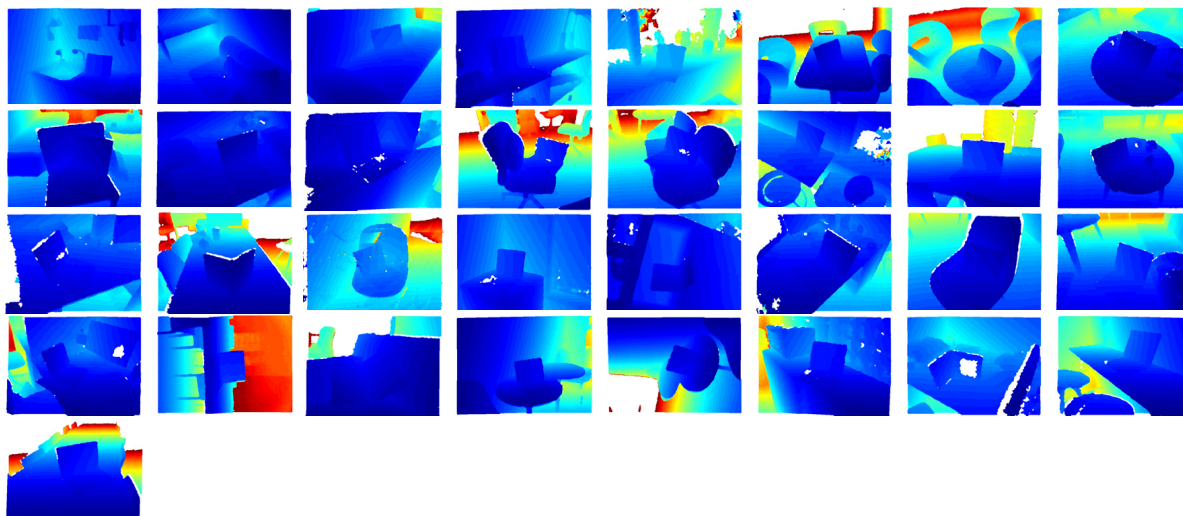


Illustration of the sample set as well as some detection results of the *notebook* category in the cross validation which are only based on the RGB images.



Corresponding depth images in the *notebook* sample set

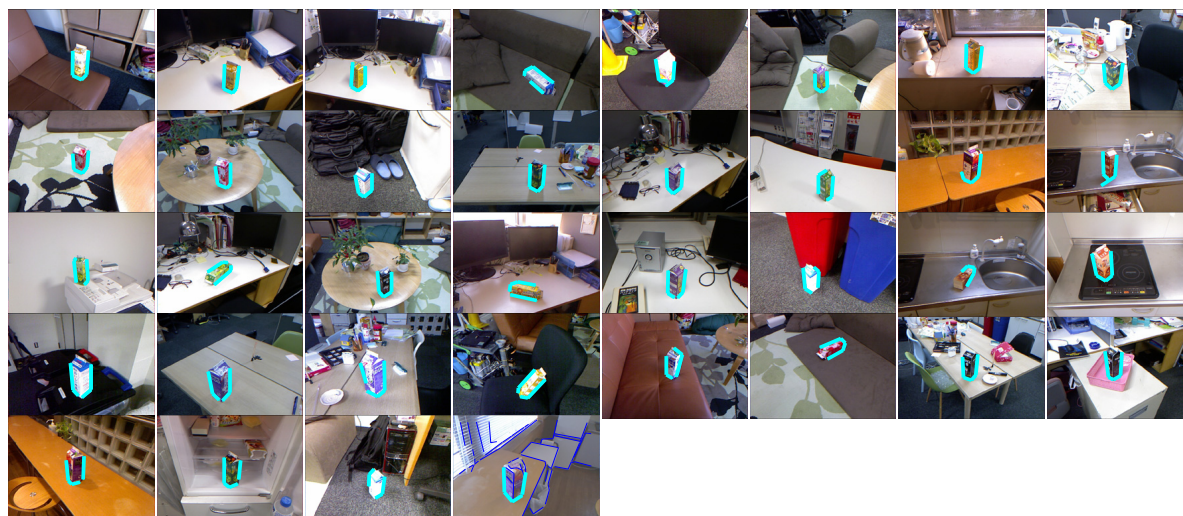
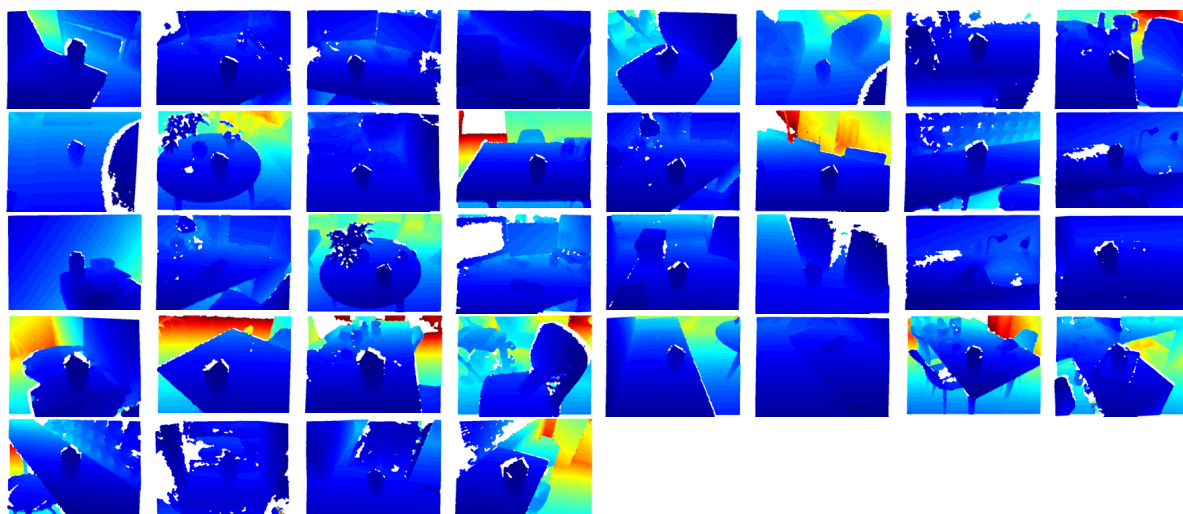


Illustration of the sample set as well as some detection results of the *drink box* category in the cross validation which are only based on the RGB images.



Corresponding depth images in the *drink box* sample set

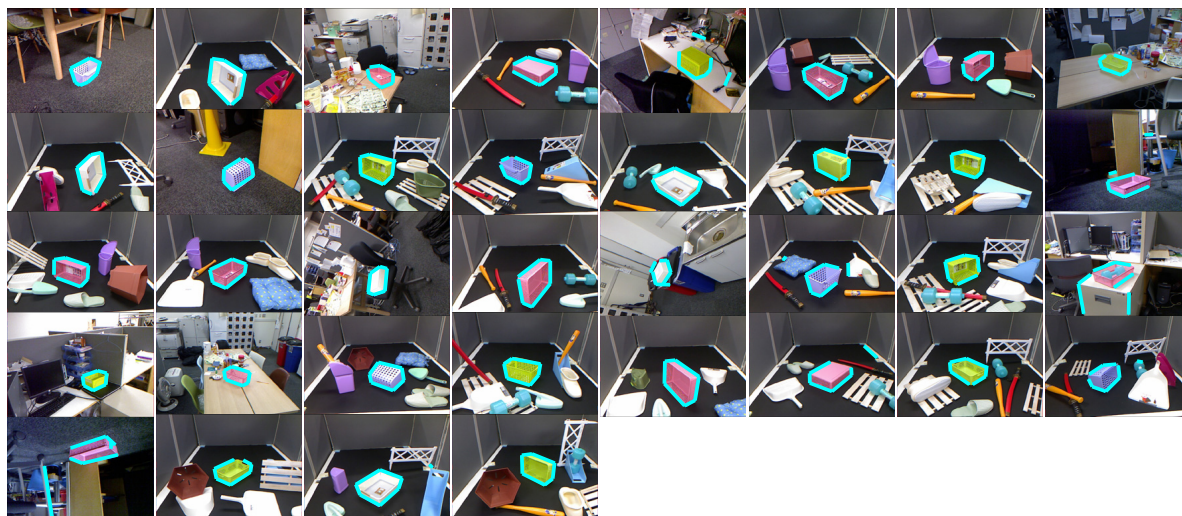
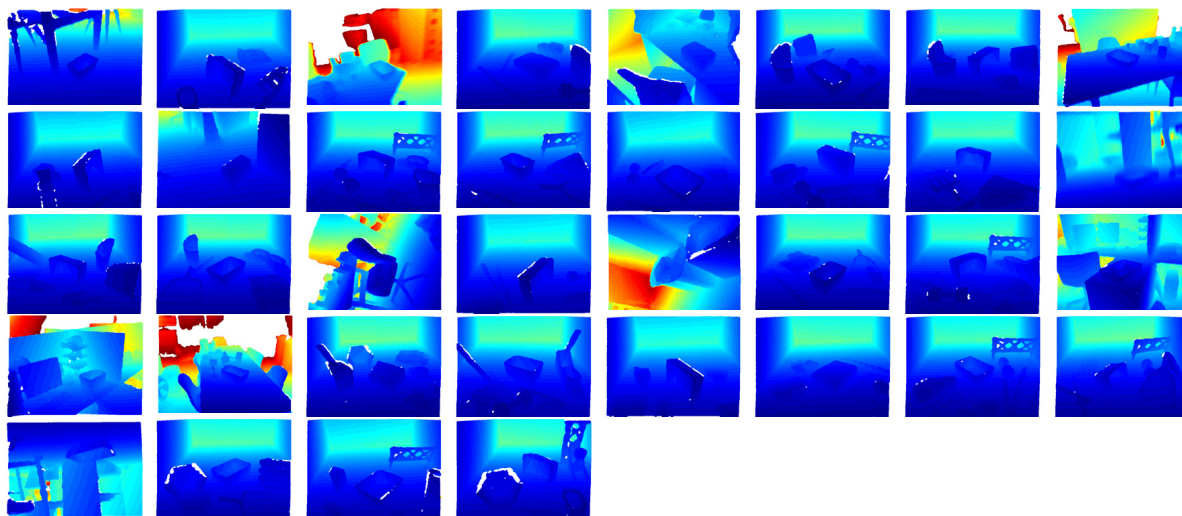
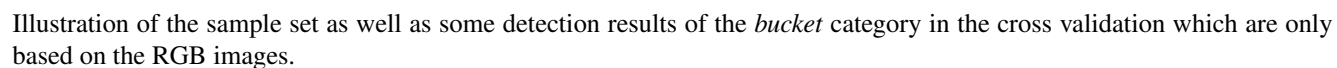
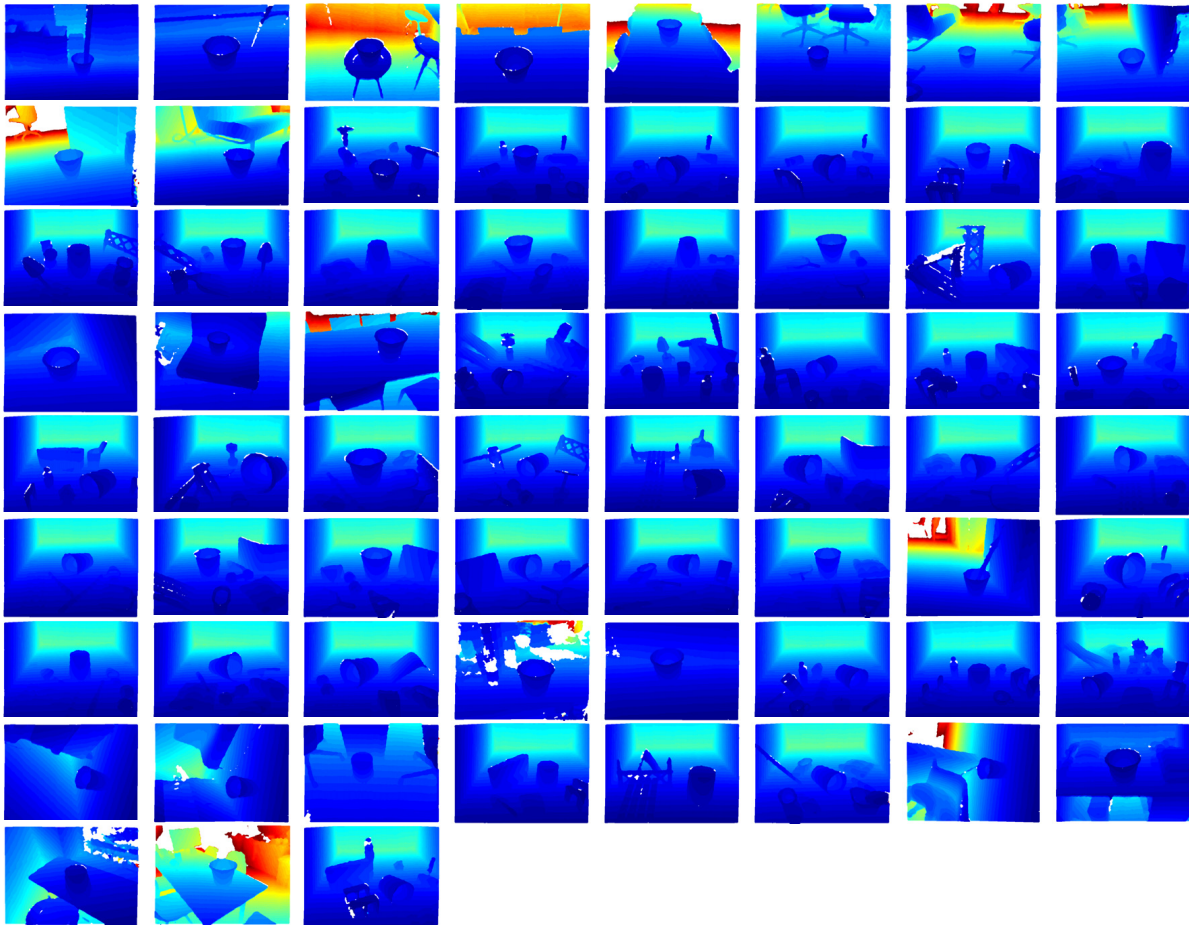


Illustration of the sample set as well as some detection results of the *basket* category in the cross validation which are only based on the RGB images.



Corresponding depth images in the *basket* sample set

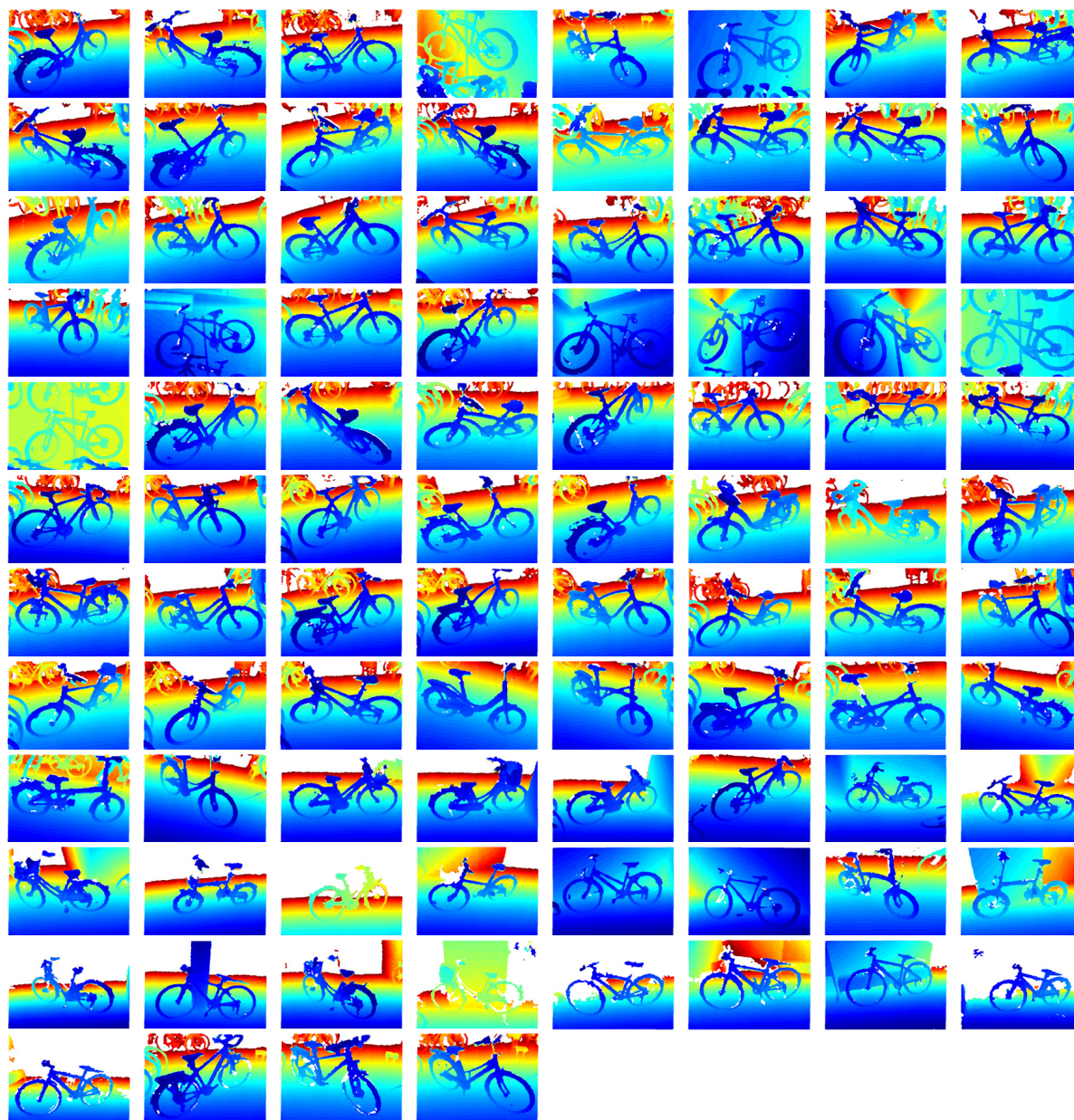




Corresponding depth images in the *bucket* sample set



Illustration of the sample set as well as some detection results of the *bicycle* category in the cross validation which are only based on the RGB images.



Depth images in the *bicycle* sample set. There depth images are not sorted to correspond to RGB images in the figure above.

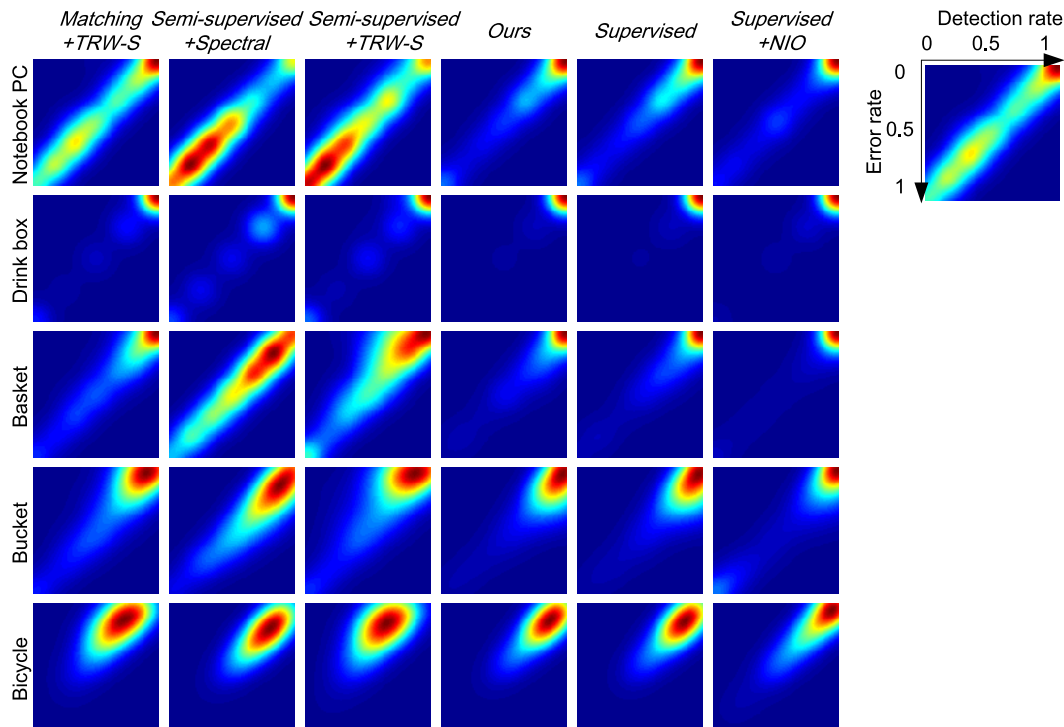
Errors in object detection:



Illustration of some obvious errors in object detection.

Sometimes, some nodes in the model may not be able to be compatible with most other nodes in matching for object detection. We set a matching choice “none” for these nodes, but this matching choice cannot prevent all the errors, and some significant matching errors still exist. These errors are shown in both the figure above and the object detection results in video frames in the supplementary demo.

2. Bias problem of conventional semi-supervised (unsupervised) methods



Distribution of detection and error rates of different models for each categories. Some models are greatly biased in the semi-supervised learning process and have low detection rates and high error rates. In our experiments, we train N models for a category of N objects by taking different objects as the initial labeling. Except the initial labeled one, $2/3$ and $1/3$ of the objects are randomly collected for training and testing, so the each model has $(N/3)$ detections of objects for evaluation, and there are $(N^2/3)$ detections in total to create the distribution figure for the category. $N = 33, 36, 36, 67, 92$ for the *notebook PC*, *drink box*, *basket*, *bucket*, and *bicycle* categories.

3. Current RGBD datasets

Various RGBD datasets has been built in recent years [1, 2, 3, 4, 5, 6, 7, 8] (reviewed by Allison Janoch [3]). Lai *et al.* built a large RGBD object dataset containing 300 objects belonging to 50 categories [1], as well as an RGBD scenes dataset, which is of great value for object recognition. Koppula *et al.*, Silberman *et al.*, and Browatzki *et al.* also constructed RGBD image sets of indoor environments [2, 4, 5]. A dataset for the solution of the perception challenge [6] consists of 35 objects for training. The UBC robot vision survey dataset [7] contains four categories, and the 3D table top dataset [8] covers three categories without significant variation in viewpoint. Janoch *et al.* also built a large dataset [3].

However, according to our scenario of learning from casually captured scenes, there are three requirements. 1) Target objects in RGBD images should not be hand-cropped or aligned. 2) Objects belonging to the same category should have different textures, rotations, and scales. 3) Each category needs to contain enough samples for training. Therefore, we build a new dataset containing approximately 900. Five large categories—*notebook PC*, *drink box*, *basket*, *bucket*, and *bicycle*—are used, each being contained in 33, 36, 36, 67, and 92 scenes, respectively.

References

- [1] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox, “A large-scale hierarchical multi-view rgb-d object dataset”, *In Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1817–1824, 2011. 10
- [2] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena, “Semantic labeling of 3d point clouds for indoor scenes”, *In NIPS*, pp. 244–252, 2011. 10
- [3] Allison Janoch, “The berkeley 3d object dataset. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-85.html>”, Master’s thesis, EECS Department, University of California, Berkeley, May 2012. 10
- [4] Nathan Silberman and Rob Fergus, “Indoor scenen segmentation using a structrued light sensor”, *In IEEE International Conference on Computer Vision Workshop (ICCV Workshops)*, pp. 601–608, 2011. 10
- [5] B. Browatzki, J. Fischer, G. Birgit, H. Bulthoff, and C. Wallraven, “Going into depth: Evaluating 2d and 3d cues for object classification on a new, large-scale object dataset”, *In IEEE International Conference on Computer Vision Workshop (ICCV Workshops)*, pp. 1189–1195, 2011. 10
- [6] *Solution in perception challenge*. <http://opencv.willow-garage.com/wiki/SolutionsInPerceptionChallenge>. 10
- [7] *UBC Robot Vision Survey*. <http://www.cs.ubc.ca/labs/lci/vrs/index.html>. 10
- [8] M. Sun, G. Bradski, B.-X. Xu, and S. Savarese, “Depth-encoded hough voting for joint object detection and shape recovery”, *In The 11th European Conference on Computer Vision (ECCV 2010)*, 2010. 10