ARC2017 RGB-D Dataset for Object Detection and Segmentation

Ryosuke Araki[1], Takayoshi Yamashita[2] and Hironobu Fujiyoshi[3]

I. ARC2017 RGB-D DATASET

In July 2017, Amazon Robotics LLC held a competition called the Amazon Robotics Challenge (ARC) in Nagoya City, Japan, to compare different logistics automation technologies. Contestants had to complete two tasks — a *Stow* task, where robots were required to stow 40 different types of items from a tote into a storage system, and a *Pick* task, where they had to pick items from the storage system and place them in the cardboard boxes. In the final round, these two tasks had to be performed consecutively. The items included various rigid, non-rigid, and translucent objects.

We competed as part of the Team MC^2 and created our own dataset for the ARC2017 tasks. This dataset comprises images of the 40 items used in the ARC 2017 competition that were in a red tote. For every scene, the dataset includes an RGB image, a depth map image, and correctly labeled bounding-box and segmentation data. Image samples are shown in Figs. 1 (a) and (b). This dataset is available on the website of our research group [4].

A. RGB Images and Depth Images

The dataset includes 1280×960 pixel RGB images and depth images corresponding to the same scenes. Using the set of 40 ARC2017 items and 20 APC (Amazon Picking Challenge) 2015 items, we created 1,510 scene images, as shown in TABLE I. Because the ARC2017 required recognition of the unknown-class objects, we prepared the APC2015 items instead.

All images in the dataset were captured using MELFA-3D vision sensor made by Mitsubishi Electric Corporation. It incorporates a miniature projector and uses coded structured light to obtain a depth point cloud.

B. Bounding Boxes

The RGB images are annotated with bounding box information. The annotation files are provided separately for each image and consist of text data indicating the box coordinates and the items' IDs. The annotations are only found in the regions of the image that can be confirmed when the annotator is looking at the image, so in the cases where occlusion occurs, the boxes only surround the visible parts, as shown in Fig. 1 (c).

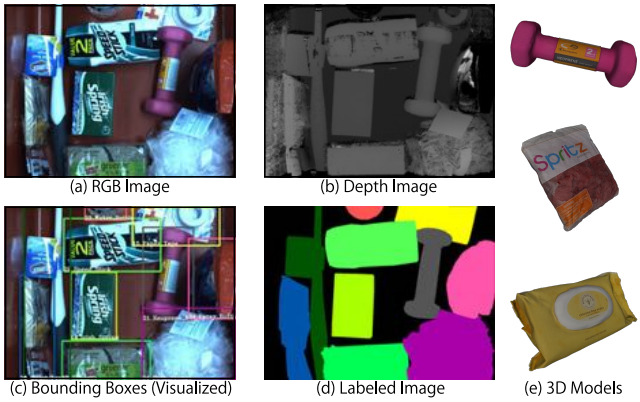

Fig. 1. Dataset contents for a single scene and examples of 3D models

TABLE I
SCENE DETAILS

		Single item	Multiple items	Total
Train	ARC'17	410	800	1,210
Test	Known-ARC'17		200	200
	Unknown-APC'15		100	100
Total		410	1,100	1,510

C. Segmentation Labeled Image

The dataset includes RGB labeled images colored on a per-pixel basis in each item region. A different color is used for each item, and the background is painted black, as shown in Fig. 1 (d). Since the items are painted without anti-aliasing, they can be used for training and testing of semantic segmentation.

D. 3D Models

We created 3D models of all 40 items. Fig. 1 (e) shows some examples of the 3D models. These models were created using a 3D MFP machine made by Ortery Technologies, Inc. The models were saved in OBJ file format together with material data in MTL format and texture data in JPG format. By applying scale information to these models, it is possible to test recognition and gripping of items in a simulator.

II. CONCLUSION

We have released the ARC2017 dataset and will show the benchmark results of the state of the art for object detection and segmentation by deep learning-based approach such as Faster-RCNN, Single Shot Multibox Detector, and SegNet.

[1]Ryosuke Araki is with the Department of Computer Science, Graduate School of Engineering, Chubu University, Aichi, Japan ryorsk@mprg.cs.chubu.ac.jp

[2]Takayoshi Yamasita is with Faculty of Computer Science, College of Engineering, Chubu University, Aichi, Japan yamashita@cs.chubu.ac.jp

[3]Hironobu Fujiyoshi is with Faculty of Robotics Science and Technology, College of Engineering, Chubu University, Aichi, Japan hf@cs.chubu.ac.jp

[4]http://mprg.jp/en/research/arc_dataset_2017_e