3D Convolutional Object Recognition using Volumetric Representations of Depth Data

Ali Caglayan, Ahmet Burak Can
Department of Computer Engineering
Hacettepe University, Ankara, Turkey 06800
{alicaglayan, abc}@cs.hacettepe.edu.tr

This supplementary material provides details of the used benchmark, 3D volumetric representations and experimental evaluations.

1 Washington RGB-D Object Dataset

The Washington RGB-D Object Dataset [1] is a commonly used benchmark with challenging object classes and instances. As explained at the introduction of the main paper, the three reasons that make object recognition a challenging task can be seen in this dataset: (i) The dataset has a diverse intra-class variation as seen in the Figure 1. Instances that belong to "ball" and "coffee mug" object categories can be seen in the figure. (ii) The dataset contains similar instance samples in different object categories as seen in the Figure 2. There are many categories similar in shape (e.g., tomato, potato, ball, orange, peach, apple). This can easily lead to confusion of categories. On the other hand, we use only depth information in our work and thus this creates a more challenging task in our case. (iii) As shown in Figure 3, the environmental illumination (i.e. apple examples) and viewpoint (i.e. pitcher examples) may vary. Also, there are examples with partial object information in different scaling and viewpoints (i.e. food jar and water bottle examples). The images may be noisy (which might be invisible to the human eye). In addition to this, some examples may contain distortions as well but this is very rare (i.e. onion sample). Reflected lights from shiny surfaces such as glass may cause difficulties in recognizing objects, especially in depth images (i.e. calculator sample). Recognition can be difficult in slim sized objects because depth information may not be taken properly in such objects (i.e. toothbrush sample).

2 3D Volumetric Representations

Figure 4 illustrates how the volumetric representations are structured. The binary values in grids are represented in yellow. As the density values increase, the colors become darker. Since the background in the images creates confusion and makes the objects ambiguous, the masked states of the images are represented here for convenience.

3 Additional Evaluations

In this section, we give additional evaluations of the experiments in the main submission. Figure 5 shows the training reports of the first test scenario described in the main paper. Images with backgrounds are used in these illustrations. The columns in this figure represent the results of binary and intensity grids respectively. The rows illustrate the loss function and accuracy results respectively.

Examples of confused object categories are presented in Figure 6. The examples are the results of the second testing scenario, which we have compared with other studies in the literature. The images contain backgrounds. The first column (a) in the figure shows volumetric representations of the samples. In the second column (b), the corresponding RGB images of the misclassified samples are given. In the last column (c), an example is given from the object categories which the related samples are confused with. Here again RGB images are given for convenience (column b and c). Our model only uses depth images, as depicted in the first column (a). When we look at the misclassified examples, it can be seen that these examples are actually very similar to each other in shape. From top to bottom, ball classified as potato, bowl as coffee cup, camera as sponge, food jar as food can, lime as lemon, peach as pear and potato as onion. All these misclassifications are mainly due to shape similarities. On the other hand, samples with shiny surfaces such as cameras may have lack depth information caused by reflections. Because the depth sensors do not properly get reflections from such surfaces.

References

[1] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on.* IEEE, 2011, pp. 1817–1824. ↑1

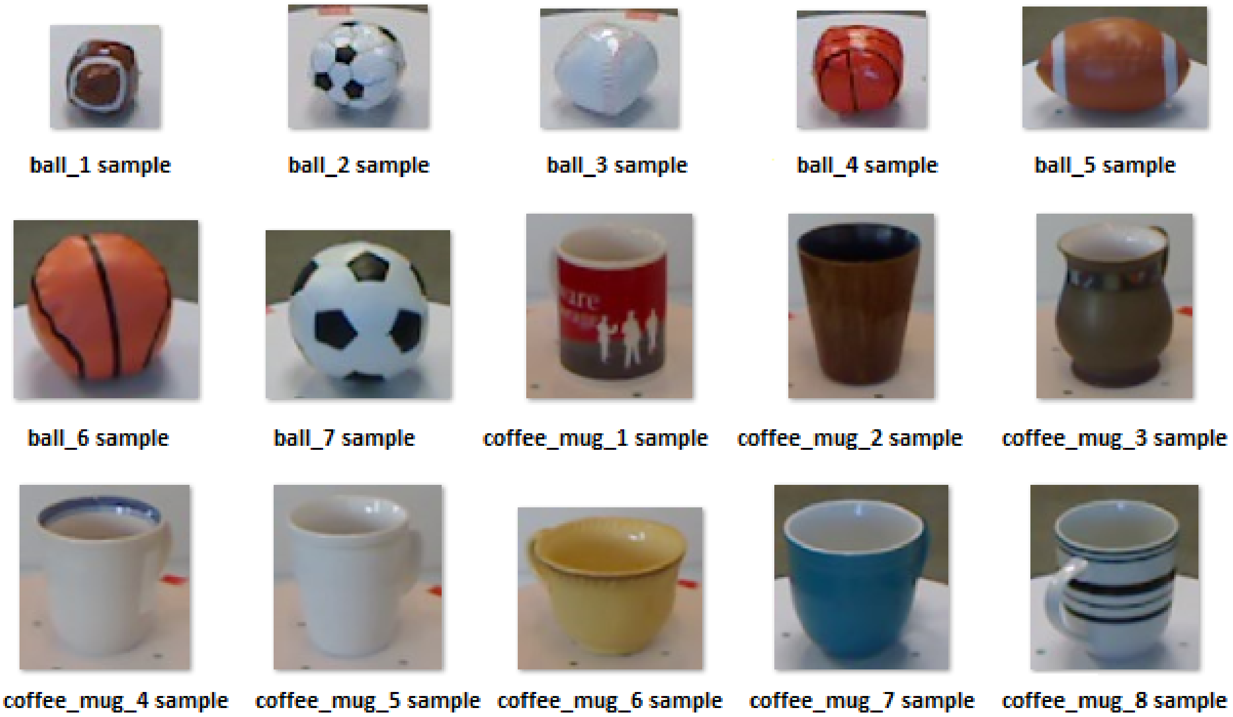

Figure 1. Object examples illustrating the intra-class diversity of the data set.

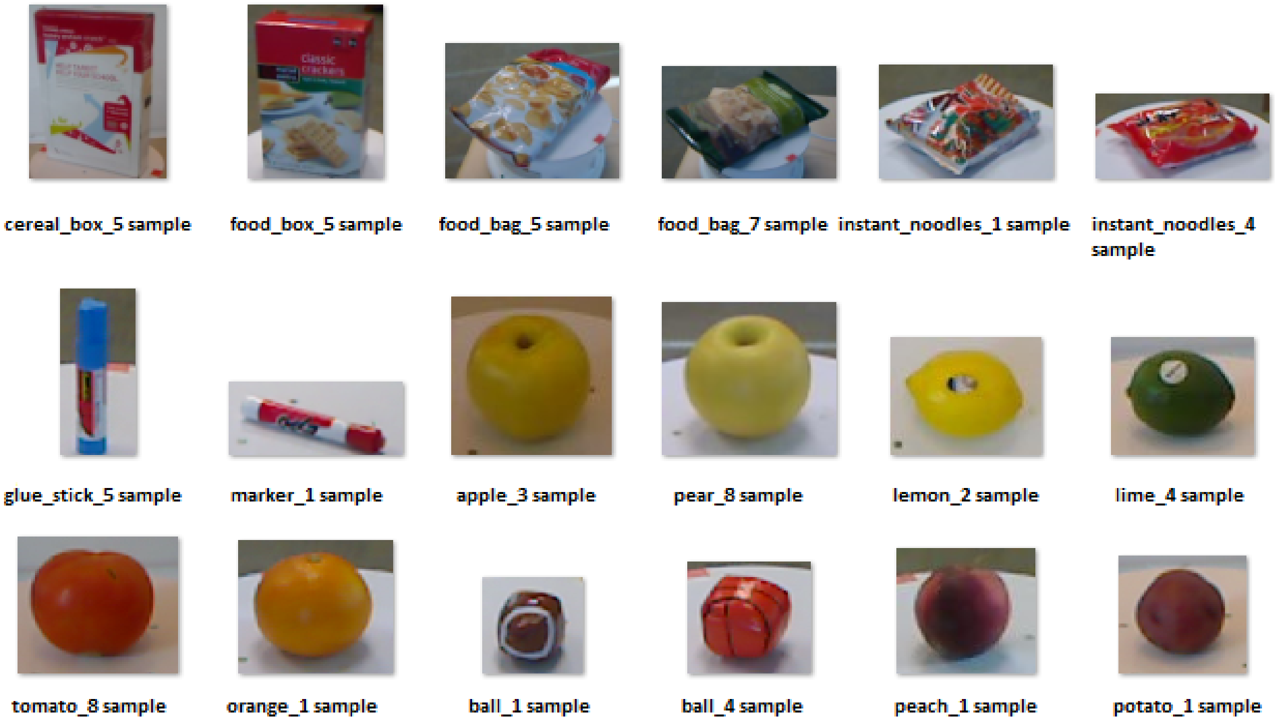

Figure 2. Examples illustrating similar objects in different object classes of the dataset

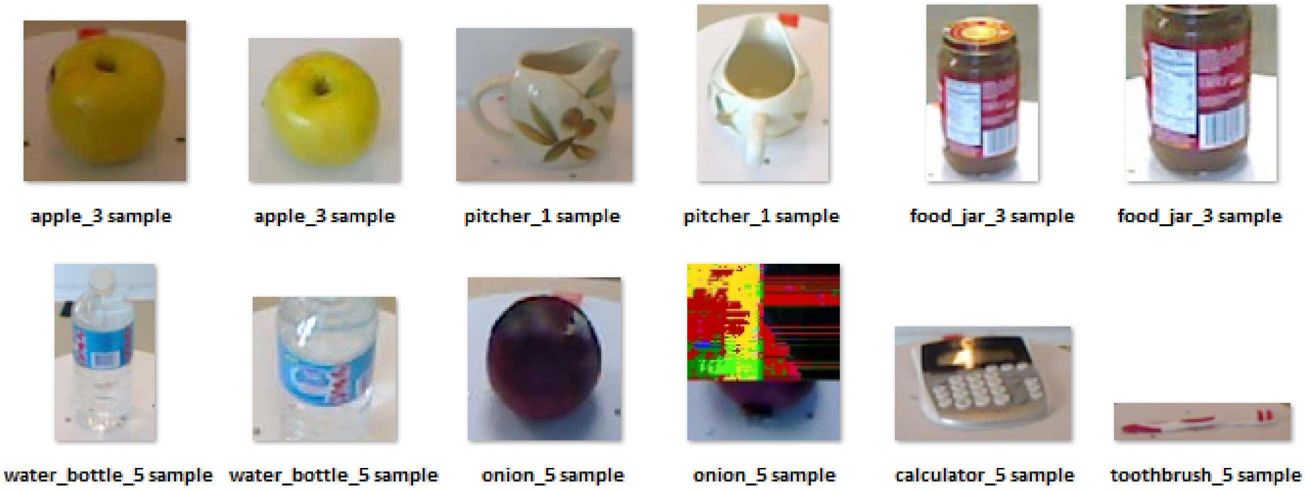
apple_3 sample
apple_3 sample
pitcher_1 sample
pitcher_1 sample
food_jar_3 sample
food_jar_3 sample
water_bottle_5 sample
water_bottle_5 sample
onion_5 sample
onion_5 sample
calculator_5 sample
toothbrush_5 sample

Figure 3. Examples illustrating general challenges in the dataset

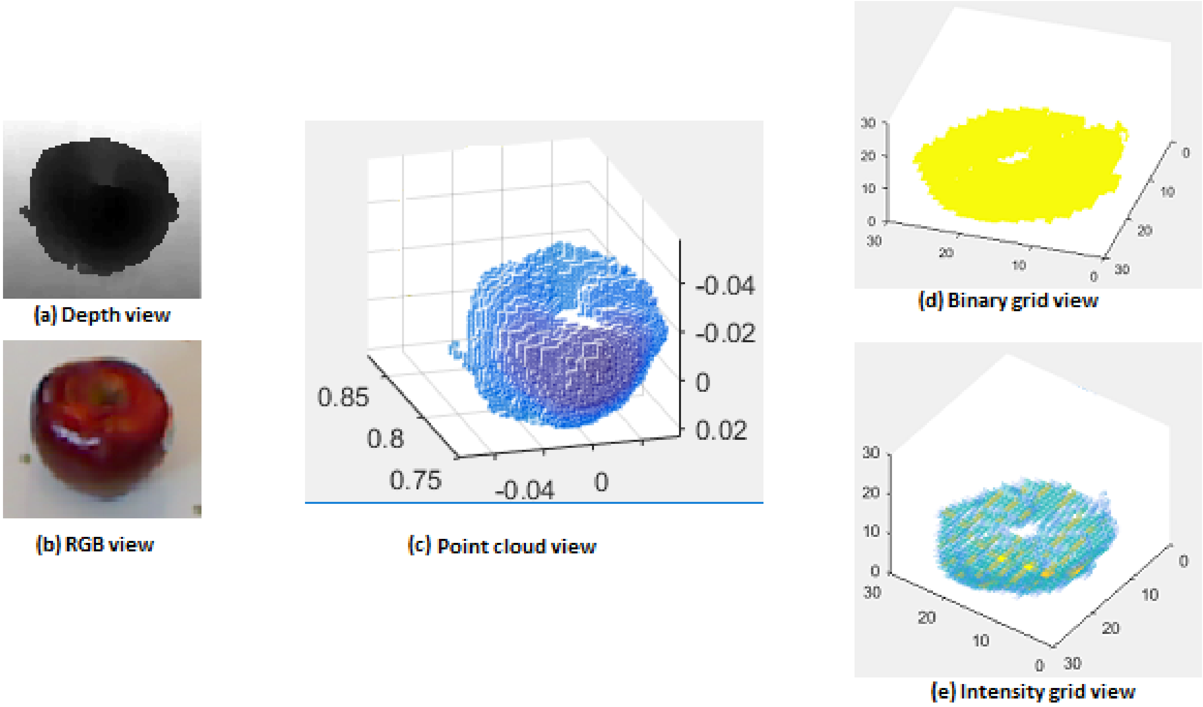
(a) Depth view
(b) RGB view
(c) Point cloud view
(d) Binary grid view
(e) Intensity grid view

Figure 4. 3D volumetric representations of an apple sample. (a) The input of our method is a depth image (b) Corresponding RGB image for visualization purpose. Our model only uses depth images. (c) Related point cloud data of the input depth map (d) Volumetric binary grid (e) Volumetric intensity grid

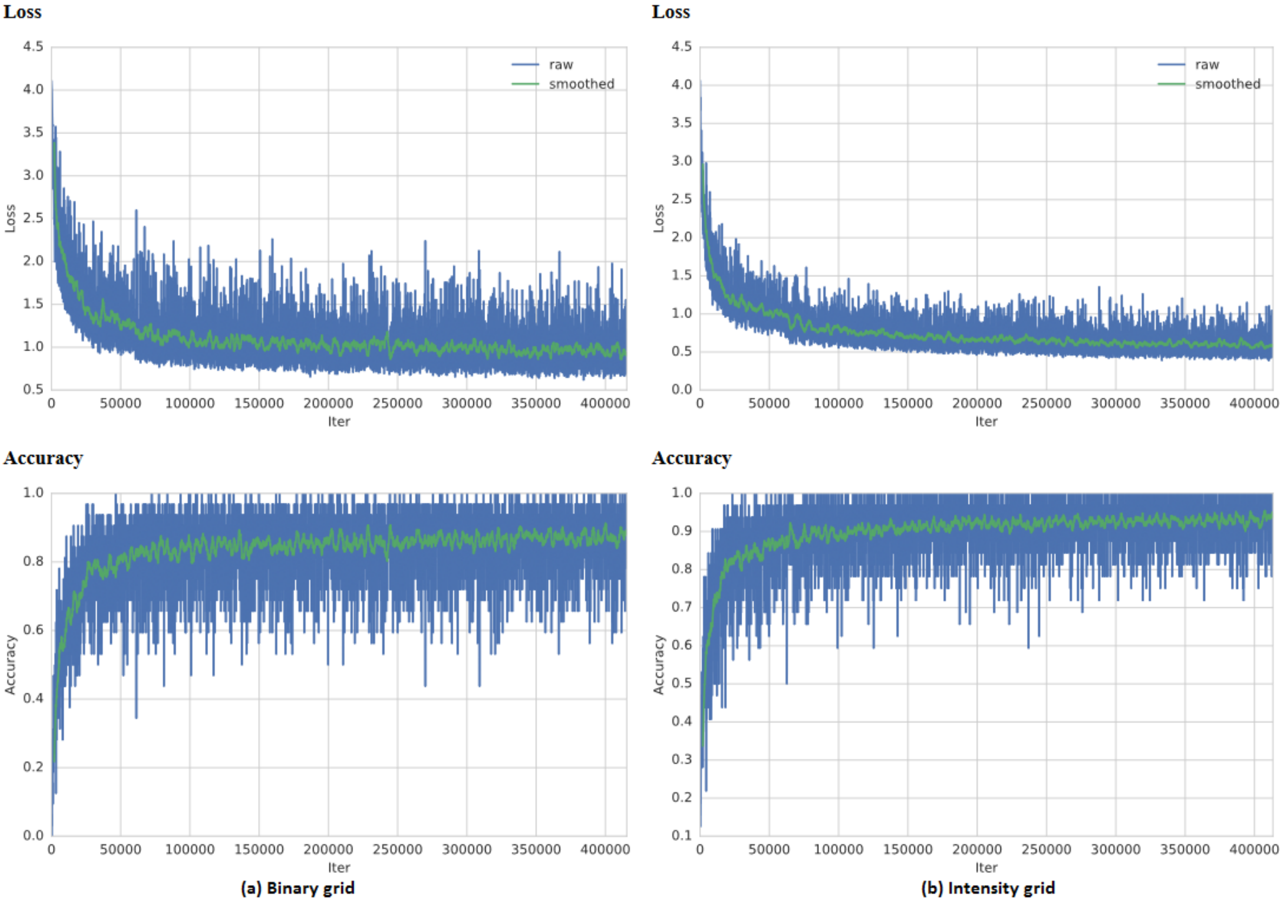

Figure 5. The training report of learning on volumetric binary and intensity grids.

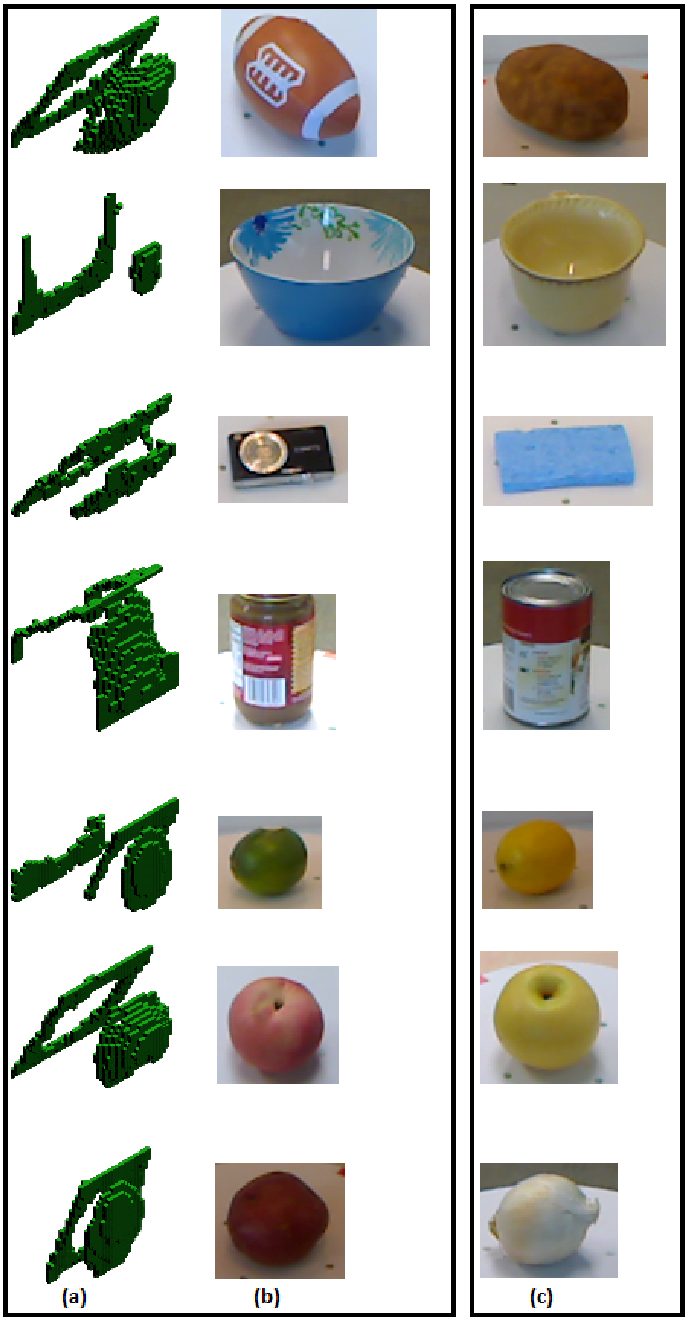

Figure 6. Misclassification examples. (a) Volumetric representations of confused examples. (b) Corresponding RGB view of the misclassified examples. RGB images are given for illustration purposes. (c) Sample RGB images from the predicted object categories.