YOLOv5 Guided 3D Point Cloud Segmentation

Edward Sun,[1] Benjamin Wen,[2] Edward Sun,[1] Benjamin Wen,[2] Aditya Patankar,[3] Nilanjan Chakraborty[3]
[1]Torrey Pines High School, 3710 Del Mar Heights Road, San Diego, CA 92130 [2]Aragon High School, 900 Alameda de las Pulgas, San Mateo, CA 94402 [3]Department of Mechanical Engineering, Stony Brook University, Stony Brook, NY 11790

Introduction

Assistive robots have become popular in recent years to aid the disabled in performing daily tasks. This study designed an end-to-end implementation of a You Only Look Once (YOLOv5) guided point cloud segmentation that can efficiently localize objects from a 3D point cloud data collected from a Intel D-415 RGBD camera to improve the speed and accuracy of assistive robotic grasps. The setup is depicted in **Fig. 4.**

Methods

The camera interface was built on the **ROS**. We used the Yale-CMU-Berkeley objects dataset to train a YOLOv5 model. **YOLOv5** was implemented in Python and used to label and localize each object with a bounding box in 2D. The localized 2D object was transformed to 3D real-world coordinates via a perspective projection from 2D to 3D. To finalize the image to world frame transformation of the segmented point cloud data, we applied reference frame transformation from the local camera coordinate system to the global coordinate system. Post processing and normal vector removal were performed to clean the noisy point cloud data. Finally, a density-based spatial clustering algorithm (**DBSCAN**) was applied to cluster the point cloud data.

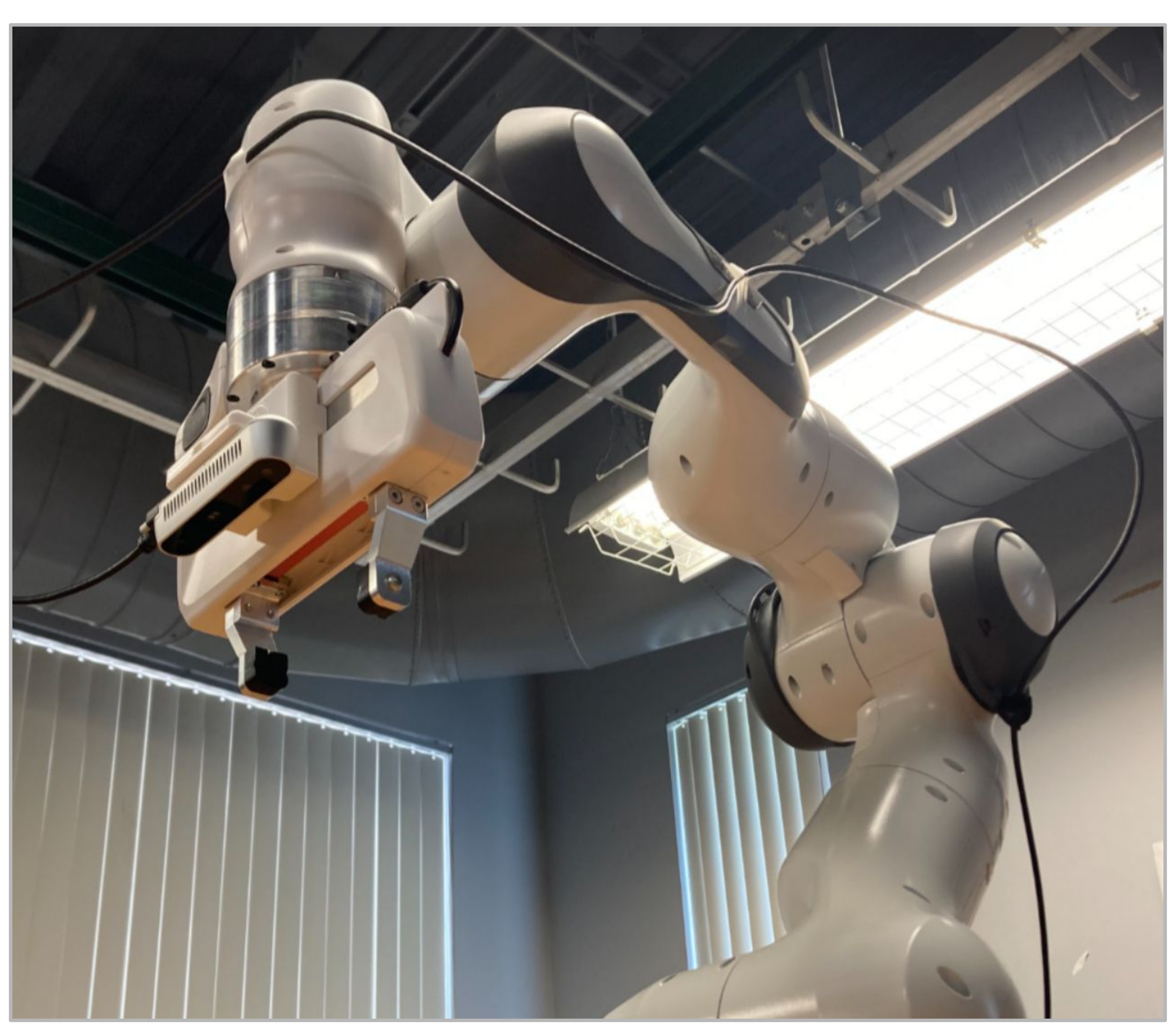

Fig. 4 Hardware Setup (Intel D-415 on a Robotic Arm)

Results

The results show the acceptable accuracy of the segmentation of the various objects **(Fig. 3)**. Although only capturing data from one perspective, the results show that the position of the object can be accurate estimated by a bounding box (**Fig. 1).** Furthermore, the from the coordinate frame transformations, the results show that the pose of the object can also be estimated as seen in **Fig. 2** where the arrows represent the transformation of the arm's base and camera lens. These results demonstrate that this pipeline segments and reasonably estimates of object's position in the world.

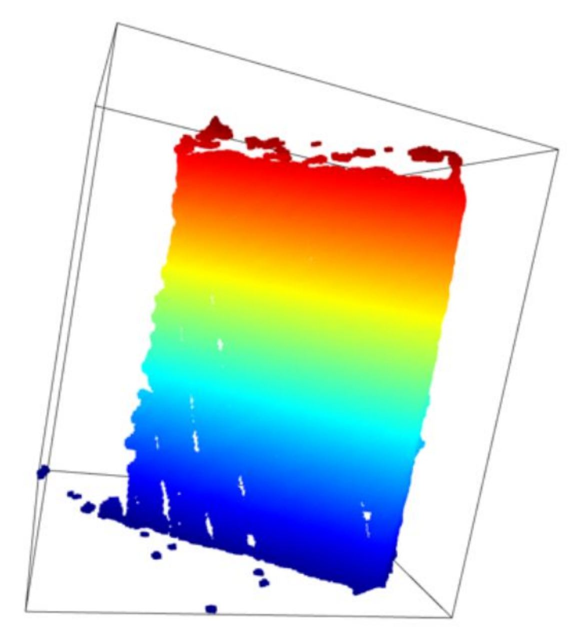

Fig. 1 Bounded and Segmented Cereal Box

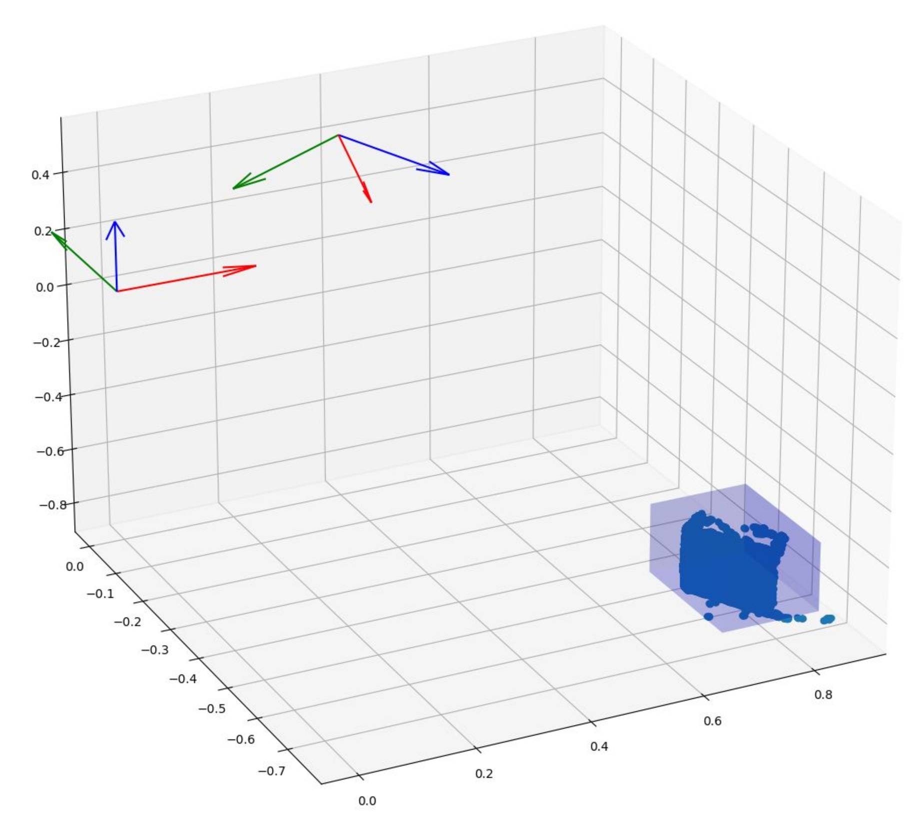

Fig. 2 Pose of the Cereal Box Transformed from Camera

This procedure is lightweight and accurate compared to other methods of 3D object segmentation. Further studies may seek to infer the full shape of the object and implement robotic grasping calculations based on the more accurate model.Taking advantage of the reduced complexity of this segmentation pipeline, this system will be better suited for wheelchair-mounted robot grasping systems that support individuals with reduced mobility.

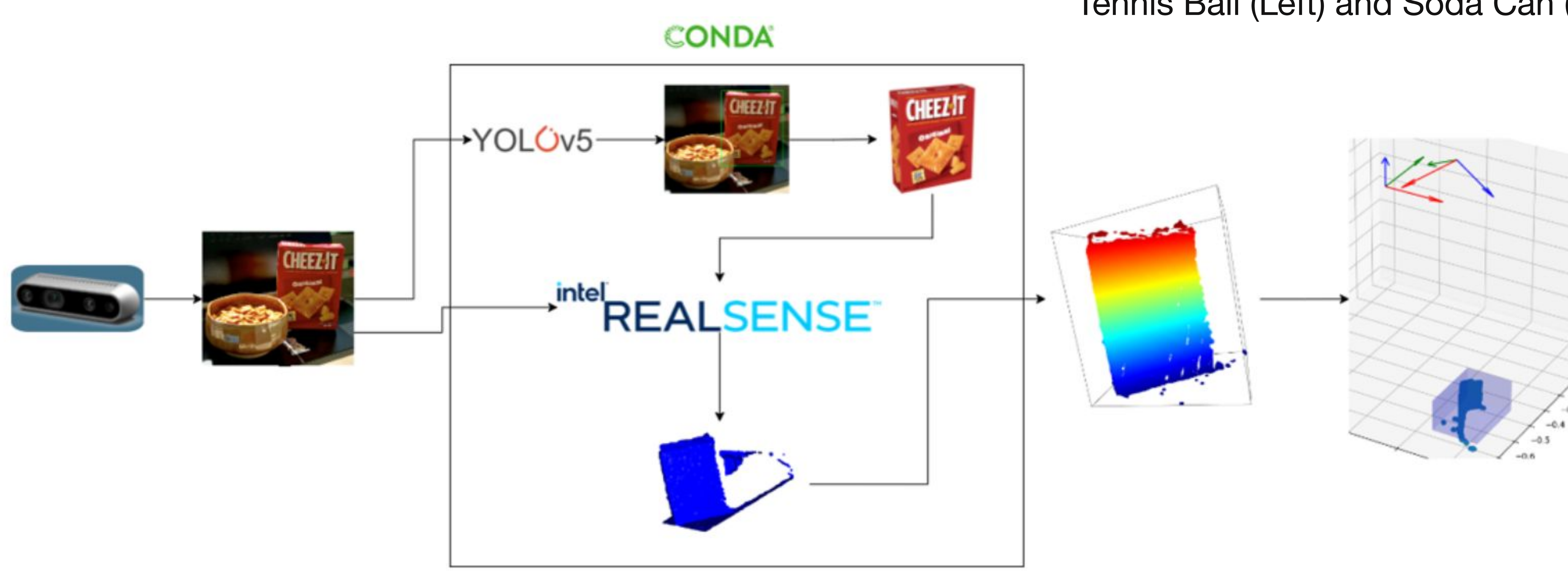

Fig. 3 Segmentation Pipeline from Capture to Results

Conclusion

Ailments like muscular dystrophy, cerebral palsy, arthritis, or severe spinal cord injuries can cause serious mobility impairments on one or many limbs, restricting an individual's ability to perform fundamental tasks. Living with such mobility impairments cause individuals to become highly dependent on the care from others. However, these limitations may be overcome with the usage of a robotic arm that assists with daily tasks. Previous studies have shown that caregiving time can be reduced by 41% with the aid of a Kinova Jaco robotic arm. As dexterous manipulation is dependent upon proper point cloud segmentation objects, speed and accuracy must be considered in perception of objects to be grasped. This end-to-end framework explores avenues of lightweight yet accurate segmentation methods.

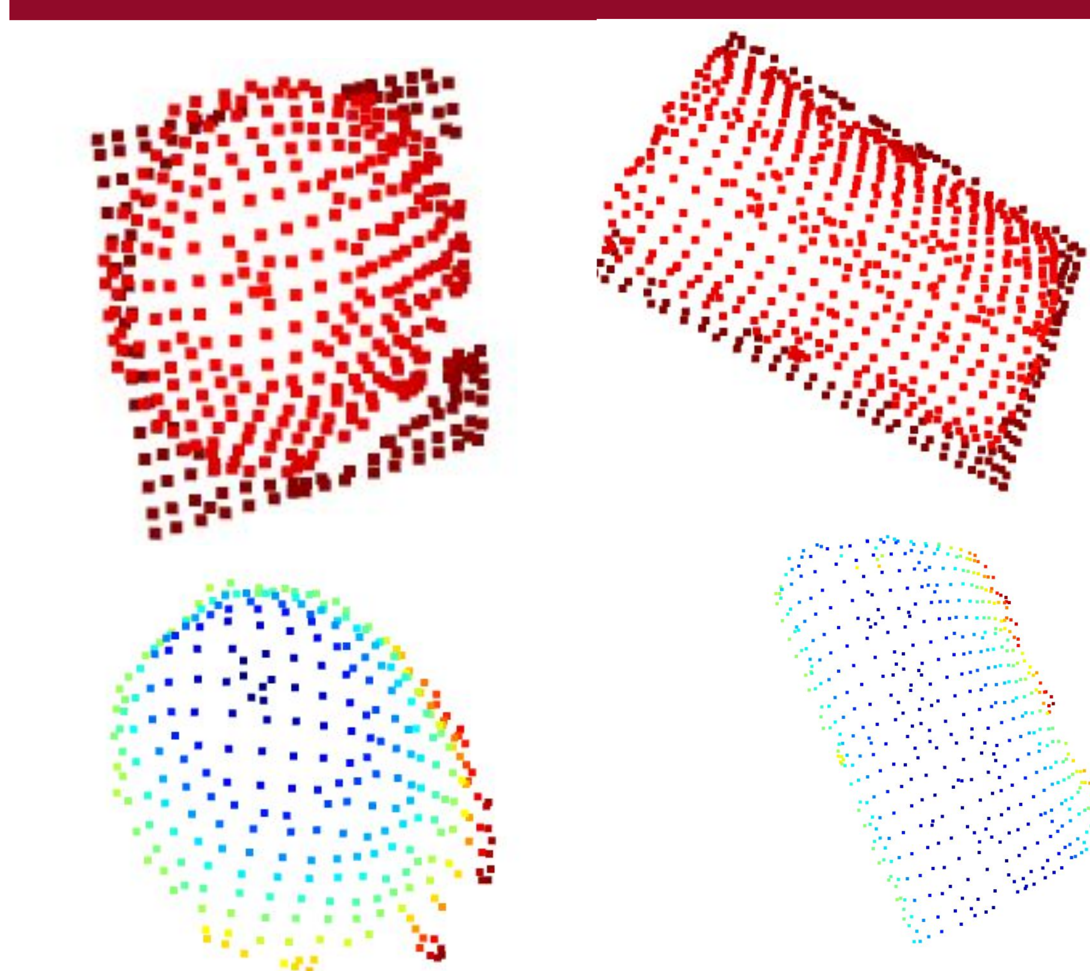

Fig. 3 Other Objects Segmented: Tennis Ball (Left) and Soda Can (Right)