CSCI 5622 Machine Learning - Vision Guided Picker

Young Il Joe - Cameron Mattson - Venkata Janakirama Pradyumna Gudluru

University of Colorado, Boulder

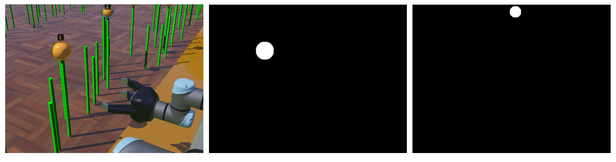
Problem Statement

Our goal in this project it to build a vision guided robotic (VGR) system that performs human-like tasks employing machine learning techniques. We build a robot that picks and places oranges in a robot simulator “Webots”. We identified four main sub-problems for this goal. The first sub-problem is how to collect training data and annotate them for segmentation tasks. The second sub-problem deals with instance segmentation and analysis of an RGB camera image. In the third sub-problem we will address how to localize objects using a instance segmentation map and a range-finder (depth camera) in 3D space. The fourth sub-problem deals with a motion planning to reach and grab the target object while avoiding obstacles and its execution.

Approach

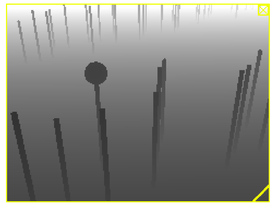
1. Instance segmentation data acquisition and annotations
   - Taken in a simulated environment Webots
   - Annotated using VGG Image Annotator
2. Orange localization
   - Using an instance segmentation map and a depth image
3. Path Planning
   - Hindsight Experience Replay was attempted
   - Using RRT was implemented by MoveIt!
4. Motion control
   - FollowJointTrajectory ROS action server

Data



The training dataset consisted of 200 samples. Each of these samples contained a segmentation mask, even if the image did not contain an orange. Data acquisition was conducted in episodes. For each episode, oranges and empty green sticks were spawned at random locations, and then 20 images were taken. Similarly, the test dataset consisted of 40 image-mask pairs.

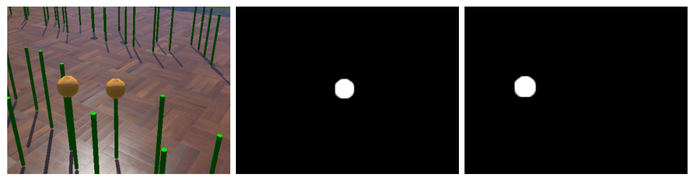
The training data as annotated using VGG Image Annotator and one example of annotations, two circles for two orange instances, are shown in the left figure above along with two binary mask images for these two instances in the middle and right figures.



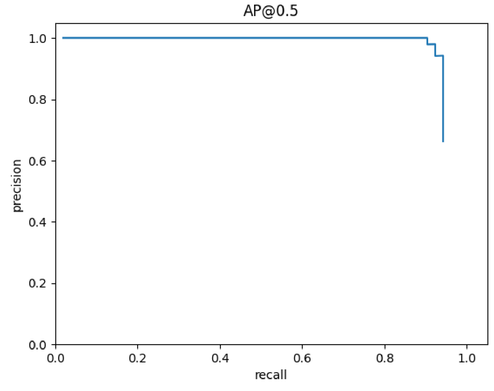
Results

Even with a small training dataset, a Mask R-CNN model [1] is successfully trained on this dataset and gives accurate enough 3D location estimations with which a robotic arm can perform pick-and-place operations for an orange. Similarly, we successfully train the U-net model to segment oranges using the small training set [2]. However, we did not use this model to locate oranges in the simulation environment. The modified U-net architecture was trained on images from the simulation to create segmenation masks on the test dataset with an Intersection Over Union (IoU) of 0.74

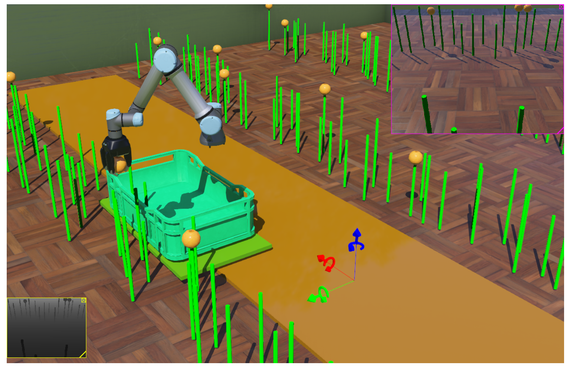
Fig. 3 shows predictions on one of test images which also has two instances of oranges. A trained model gives two predictions with greater than 0.998 of confidence scores.



To quantify the performance of the instance segmentation model, a precision-recall curve, which is commonly used performance metric for instance segmentation tasks, is calculated at the IoU threshold of 0.5.



With each cycle we first detect an orange of interest, move the XY stage for the second detection, detect the orange again, localize the orange, and then plan and execute the pick-and-place operation. Repeating these cycles, oranges are placed on one side of the field by the robotic arm. Fig. 5 shows the pick-and-place operation just after an orange is picked up by the end-effector.



Discussion

We successfully built an autonomous orange picker by combining image segmentation algorithms, range-finder, RRT motion planning, and ROS "FollowJointTrajectoryAction" action server for UR5e. This demonstrates that image segmentation techniques can be applied to vision guided robots to perform autonomous operations. Since our ultimate goal is picking oranges, a few false negatives using image segmentation will not severely hinder the performance of the system, since we gather many pictures of images using the video feed. As long as the oranges can be detected at a certain angle and position, the robot can approach the fruit and pick it up. To measure the overall performance of an autonomous orange picker, appropriate performance measures or benchmarks need to be devised.

Using both the U-net and clustering algorithms was not a viable alternative to instance segmentation algorithms, since pixels of adjacent object instances may not be separable using current clustering techniques. However, using this approach for disjoint segmentation tasks could be used a novel approach.

We see many possible improvements over the current implementation.

1. Combined motion planning for the XY stage and the the robotic arm
2. A camera placement near the end-effector. It might make a more reliable pick-and-place operation.
3. A motion control from the reinforcement learning technique (possibly employing HER)
4. Object localization using a machine learning technique on a subset of point cloud
5. Object grasping planning using a reinforcement learning technique

Takeaway

1. During completion of this project, we learned that there are many sub-tasks of vision guided robots to which various machine learning techniques can be applied including localization, pose estimations, motion control, grasping.
2. In image segmentation tasks and path planning tasks, we learned that the algorithms are dependent on the environment, as well as the hardware running and following the commands of the algorithms. In a smaller team, a broad knowledge base may be required. However, this can come at a cost in terms of both computational complexity and time. These algorithms have to be fault tolerant in the real-world and at the very least, these algorithms must use reasonable hyper-parameters.
3. From a robotics perspective or computer vision perspective this idea enhances the usage of autonomous robots to identify and pluck different vegetables over large farms reducing the human efforts and errors.

References

[1] Kaiming He et al. *Mask R-CNN*. 2017. DOI: 10.48550/ARXIV.1703.06870. URL: https://arxiv.org/abs/1703.06870.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.