Interactive Object Segmentation in Mobile Robots

Averill, Charles
charles.averill@utdallas.edu

Xu, Zesheng
zesheng.xu@utdallas.edu

Xiang, Yu
yu.xiang@utdallas.edu

1. Introduction

As a human interacts with their environment by manipulating objects, they learn about their surroundings. This provides important context about the objects and their surroundings, allowing the human to make better-informed decisions to accomplish tasks. This process is known as Interactive Perception. Robots can utilize vision and physical interaction to learn through Interactive Perception as well.

Computer Vision (CV) is a group of machine learning tasks that are heavily researched due to their abundant applications in industry. Common CV solutions use Deep Learning, a set of statistical algorithms that build a model based on the attributes of input data. These attributes are used to predict unknown information, such as the kinds of objects in any given image. These models function well, however they rely on massive amounts of input data, typically tens to hundreds of thousands of images.

This issue is the leading cause for Interactive Perception research. By manipulating and interacting with its environment, a robot can generate its own training data and learn to solve common CV tasks. This process of realtime learning is also known as ”online learning”.

2. Research Goals

Interactive Perception is a relatively new field of CV, and therefore requires researchers to not only discover new techniques of learning from interaction, but also to replicate and tune foundational works in the field. We will cover Object Segmentation, a task in which an agent draws a boundary mask around each object in an input image (Fig. 1, ”Instance Label” columns). We seek to implement and optimize existing Object Segmentation solutions in the novel context of a mobile robot.

Although Object Segmentation has been demonstrated with Convolutional Neural Networks, it has been shown that learning these segmentation masks during interaction with the scene of the input images can significantly improve performance with smaller amounts of training data. We seek to:

- Design a simulation allowing a mobile agent to push objects in an artificial tabletop scene (Fig. 1, ”RGB” columns)
- Implement existing Object Segmentation online solutions for the simulated agent
- Optimize these existing methods for both simulated and physical agents

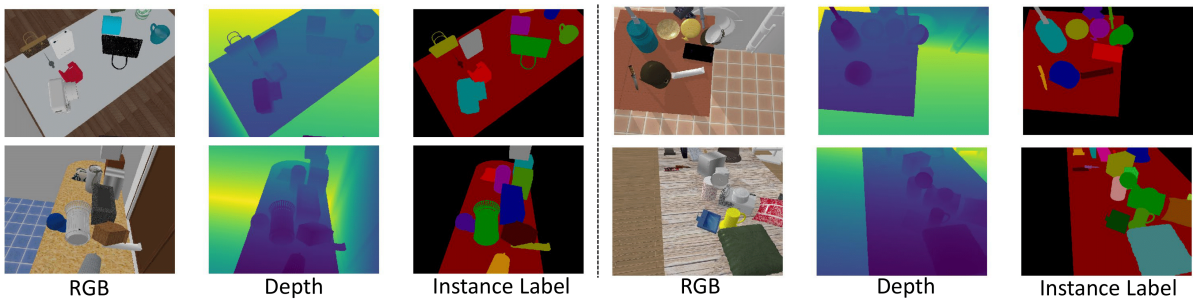

Figure 1. Object Segmentation masks for simulated tabletop images

3. Purpose

Interactive Perception is a necessary alternative to common CV solutions because of its accessibility. Although large datasets of images of cluttered desks and countertops exist, there are limitless specific applications of CV in which large datasets or refined algorithms do not exist. Interactive Perception is the solution to designing models without the need for large datasets.

4. References

- **Object Segmentation masks for simulated tabletop images**

  ”Example RGB-D images and the corresponding instance labels from the Tabletop Object Dataset”

  Reprinted from “Learning RGB-D Feature Embeddings for Unseen Object Instance Segmentation” by Xiang et al, 2020, with permission

  https://yuxng.github.io/xiang_corl20.pdf

- **Learning Instance Segmentation by Interaction** - Pathak et al.

  https://arxiv.org/pdf/1806.08354.pdf

- **Self-Supervised Object-in-Gripper Segmentation from Robotic Motions** - Boerdijk et al.

  https://elib.dlr.de/139332/1/wout_boerdijk20.pdf