



Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences



RnD Topic Introduction

Multi-Input Model for RoboCup Dataset based on key points

August 18, 2023

Gokul Krishna, Chenchani

Advisors

Deebul Nair

Introduction

- Currently in the RoboCup scenario, YOLOv8 model is used for object classification.
- During every perception done by the Intel Realsense D435 RGB-D camera mounted on the YouBot, the model gives the list of all objects identified along with the bounding boxes for the input image passed to it.
- The objects list from the model is checked against the target label of the object to be manipulated given by the planner.

Introduction

The perception pipeline for the current approach is shown in fig. 1

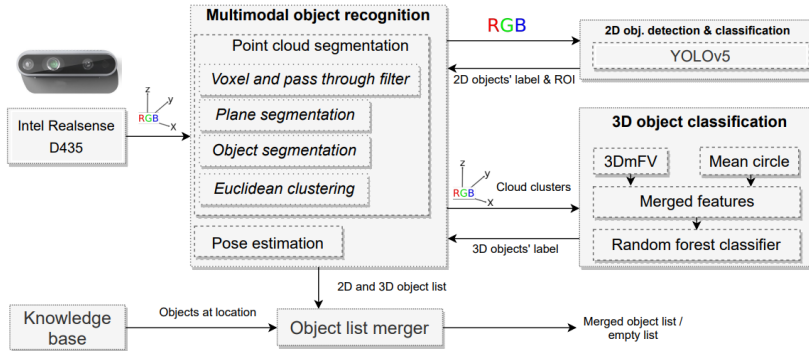


Figure 1: Object Perception Pipeline

Introduction

The simplified perception pipeline for the current approach is shown in fig. 2

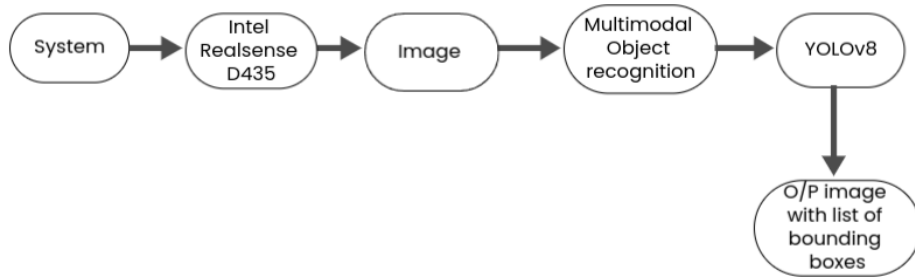


Figure 2: Simplified Object Perception Pipeline

Problem Statement

- The configured KUKA arm manipulator can manipulate only one object at a time.
- Also, Due to the possibility of the use of DBC (Direct Base Controller) in the YouBot to manipulate, the pipeline is configured to perceive every time before manipulating an object.
- Detecting additional classes from the input image through the model also consumes more time and efficiency of the system.
- YOLO based detection often fails on arbitrary surfaces.

Proposed Approach

- Since the system knows which object to be manipulated from the planner even before perceiving, we can give a multi-input to the model in the form of image and an encoded label or text of the target object.
- Few deep learning approaches for this multi-input process can be applied along with few fusion methods to fuse the image and the label data as a single entity to the model.
- The expected output from this model is the detection of only the single object along with its bounding box coordinates or key point values for the KUKA arm to manipulate.

Proposed Approach

An abstract pipeline for the proposed approach is shown in fig. 3

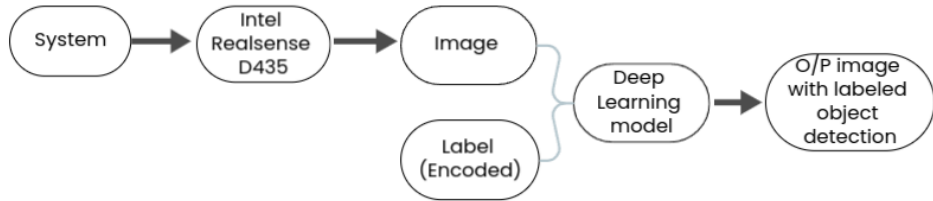


Figure 3: Object Perception Pipeline for Proposed Approach

Research and Development Questions

Research Questions:

- Which encoding deep learning approach can be used for comparative evaluation?
- Which fusion methods can be implemented for comparative evaluation?
- Comparative run-time analysis with current YOLO model used in the YouBot for performance and test on arbitrary surfaces.

Research and Development Questions

Development Questions:

- Deploy the final study and implementation on the real YouBot hardware and test the performance and efficiency compared with YOLO model.

Possible Approaches

Below are few possible researched methods which can be tested for the proposed approach:

- One-hot encoding
- Ordinal encoding
- Transformers using key, value and query.

THANK YOU

