# CSCI 5622 : Machine Learning
# Project Milestone 2

## Problem space

Our goal in this project is to build a vision guided robotic (VGR) system that performs human-like tasks employing machine learning techniques. A number of autonomous robots have been developed for harvesting tasks [1, 2]. We originally planned to build a fruit harvesting robot. However, since it's difficult to simulate the harvesting process in a robot simulator, we are going to build a robot that picks and places oranges in a robot simulator "Webots". We identified three main sub-problems for this goal. The first subproblem we will address is how to identify objects in real time using an RGB camera and a range-finder (depth camera) in 3D space. The first subproblem we will address is how to improve the segmentation process for creating our dataset and increasing the number of segmented samples. The second subproblem deals with a motion planning approach in recording and acquiring real-time video feed when either the object moves, or the camera moves, or both the object and the camera move. The third subproblem focuses on developing the model from the data so that the object is classified and segmented in real-time. The fourth subproblem deals with a motion planning to reach and grab the target object while avoiding obstacles. And the last subproblem focuses on robotic motion controls for minimal energy consumption and faster task completion.

## Dataset Summary

We planned to use two different machine learning techniques. One is reinforcement learning for robotic path planning and motion control. The other is the instance segmentation for the target localization.

### Reinforcement Learning

Since we employed the Webot robot simulator, we will get training data from the simulator as the training progresses. By putting position sensors and speed sensors at joints of robots, we can get input for the reinforcement learning algorithm. Also employing the supervisor controller of Webots, we can check episode terminating conditions and iterate through the training episodes.

## Instance segmentation

We obtained 17 training datasets from the scene so far. But we planned to get more data. We will set the resolution of the RGB camera to be (480, 320). Thus the number of features for each instance is 480 * 320, or 153600 per instance. Similarly, we are thinking of transforming image samples from the Coco dataset to determine if this data will be useful in training our model to recognize and segment our fruit.

When considering our image data, all information is important, because we do not know in advance where the fruit is located in our images. But since we plan to adapt the feature extraction layer from VGG, ResNet, or others, this feature extraction layer tells us which non-local features are important for instance segmentation.
We want to know which pixels correspond to the target object (orange in this case) or the background. Thus, we have a binary variable for each pixel.


# Planned approach

## Simulation environment (Webots)

Webots robot simulator is chosen for this project because it comes with a wide range of robots and devices, and it is easy to integrate with the ROS robotic operating system, and also it supports a decent 3D renderings. We plan to adapt a “UNIVERSAL ROBOTS” UR5e as a base, which is shown on the right side of the picture. And this robot is equipped with a RGB camera, a range-finder, and “ROBOTIQ” 3-Finger Gripper at its end effector. Two overlays on the left side in the figure above show the camera and the range-finder in the simulator.

## Object localization

Since our robot will be tested in a simulated environment, we think that the training dataset taken in a similar environment is more relevant and will give a more accurate image segmentation model for the test. Thus, we planned to take training datasets in a Webots simulator, and possibly some samples from the Coco dataset.

## Training dataset acquisition

Using the **supervisor controller** of Webots, we can change poses of objects randomly and take pictures from different camera locations and angles. With these images, ground truth segmentation can be manually annotated using VIA annotator. Similarly, as mentioned above, some of the Coco samples may be used.

## Instance segmentation and combining with the range-finder

Image segmentation technique has been applied to object detection problems. [3] We plan to adopt Mask R-CNN or U-Net architecture for instance segmentation. Taking VGG or ResNet as a feature extraction layer, we might be able to save training computations.

Once we have done instance segmentation, combining the instance segmentation with the depth information from a range-finder, we can locate the target object in 3D space. In order to do this, the RGB camera and a range-finder need to be calibrated. We might take this calibration information (frustum or spherical projection) from the official Webots document or calibration data can be obtained from experiments with a checkerboard pattern in the simulator.


# Path planning and motion control

In this project, we were considering adding this as a stretch goal, but decided not to in order to determine how much progress we can make in this project. There are a few different approaches we could implement for this part of the project. One would be employing more conventional approach such as RRT* and PID control, and the other would be a reinforcement learning approach. For the first choice, we can calculate the series of waypoints for a given target location, and with these waypoints we can get joint angles using an inverse kinematic solver such as “ikpy”. For a RL approach, we plan to adapt the actor-critic, DDPG, or others to get the optimal torque values for each joint, which will allow us to obtain the target location with minimal energy consumption.

# References


1. Boaz Arad et al., Development of a sweet pepper harvesting robot,J Field Robotics. 2020;37:1027–1039. DOI: 10.1002/rob.21937
2. Inkyu Sa et al, Peduncle Detection of Sweet Pepper for Autonomous Crop Harvesting - Combined Colour and 3D Information, IEEE Robotics and Automation Letters ( Volume: 2, Issue: 2, April 2017) DOI: 10.1109/LRA.2017.2651952
3. P. Ganesh, K. Volle, T.F. Burks, S.S. Mehta, Deep Orange: Mask R-CNN based Orange Detection and Segmentation, IFAC-PapersOnLine, Volume 52, Issue 30, 2019, Pages 70-75, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2019.12.499.