

Object Detection within a Robotic Application with YOLOv2

Chia-Wen Tsai

amysai17@hotmail.com

Felizitas Kunz

felizitas.kunz@online.de

Christoph Caprano

Caprano@in.tum.de

Oskar Haller

o.haller@outlook.com

February 4, 2018

Project Proposal

<https://arxiv.org/abs/1612.08242>

1 Introduction

We want to teach a Robot learn how to play the child game memory/pairs. Therefore, it should be able to detect playing cards, their position on a table as well as the images that they contain. For object detection YOLOv2-Real-Time-Object detection is used for transfer learning with datasets containing images of our playing cards.

1.1 Related Works

Applying deep learning for object detection is a broad field of research. Similarly, various firms and researchers try to apply vision techniques in robotic applications.

One example that falls into that field is the Amazon Picking Challenge(APC) (<https://arxiv.org/pdf/1609.09475.pdf>). Thereby, most of the teams apply multiple camera views to prevent arising issues such as clutter, self-occlusions, and bad reflections. To enable communication between embedded devices ROS has proven to be suitable in many applications (MLA Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." 2009, <http://www.willowgarage.com/papers/ros-open-source-robot-operating-system>).

For the object detection part, we plan to do transfer learning with YOLO/Darknet, which should result in real-time detection. (YOLO9000: Better, Faster, Stronger, (Redmon, Joseph and Farhadi, Ali)

2 Dataset

For the object recognition, we plan to take images of our playing cards, and label them manually. For transfer learning we can use any dataset that provides sufficient amount of data. A promising dataset which is suited for object detection and is available publicly is the PASCAL VOC Dataset (<http://host.robots.ox.ac.uk/pascal/VOC/>). Facing the playing cards from different angles might be a challenge, because a change in the angle reshapes the object from a camera perspective. Furthermore, reflection issues can cause noise in the input camera stream.

3 Methodology

To achieve our goal, "fast object detection within a robotic use case", we use a stream of images as input. The stream comes from a webcam mounted on the robotic arm. A RaspberryPi hands the stream to a computer where the images are processed.

To enable real-time detection of the different playing cards as well as getting position data relative to the camera, we aim at implementing transfer learning with the Darknet/YoloV2 Architecture on our Dataset.

Resource wise we plan to train our network on the Google Cloud Platform. For processing the images we use a Laptop. The Raspberry Pi takes in the camera stream and sends it further to the Laptop. Last but not least, Motor

commands are sent from the Laptop via an Arduino Uno to the Robot (compare Figure 1).

Figure 1: Braccio Robotic Arm

4 Outcome

First priority is that the robot can classify the images on the playing cards. With the goal to be able to correctly identify pairs.

For a first stage the Robot shall be able to detect the cards at predefined positions by taking a single picture of each card.

As a next step single cards shall be recognizable anywhere on the table, possibly in real-time.

As optional outcome the robot should be able to pick up a flipped faced down card, turn it around and build stacks of pairs.