

A Data Science Approach to Object Classification

Vera Degtiareva and Ryan Z. Nie
eLab, Physics Department, Boston University, 02215

(Dated: May 2, 2022)

We propose an object recognition sorting device that will use machine learning to recognize fruits. We will use an APDS9960 RGB sensor to collect RGB color data of a fruit and preprocess the data on Python. We will then implement several machine learning algorithms (i.e Deep Neural Network (DNN), K-Nearest Neighbors (KNN), etc.) to predict a probability that the object is of a certain type of fruit. Neural network training and other machine learning algorithms will be done on Python and deployed onto the Arduino.

I. INTRODUCTION

Object recognition is the basis for modernizing the assembly line in all sectors of production. Developing tools for computer-based decision making ensures that the factories of the future are as automated as possible and only require human control at key points in the production. We propose a simple device that can categorize objects that fall into one of the predetermined categories based on their color characteristics. In our case, we will be training the device to differentiate between three types of fruit, but it can easily be trained to differentiate between different colored objects on an assembly line.

We are planning to use various machine learning algorithms such as neural networks and K-Nearest Neighbors to differentiate between objects. The ultimate goal of our project is to determine the most accurate algorithm for classifying fruits.

The device we propose will use only an inexpensive RGB sensor and an Arduino powered by 3V, which means that implementing it on a larger scale will require few monetary and energy resources.

II. HARDWARE

The full schematic of our project can be seen in Figure 1. We will be using the APDS9960 sensor, which can recognize RGB values. Arduino Uno will be used for this project, but other models can be used as well.

III. SOFTWARE

To acquire the data, we will be using the `Arduino_APDS9960` package [1] which will help us extract the RGB data from the sensor. The processing of all data will be done using popular Python data science libraries (numpy, pandas, sklearn).

One of the algorithms we implemented was a deep neural network. Tensorflow [2] was used to create the DNN architecture, and 10-fold cross validation is used to find

the optimal hyperparameters. Our model yielded an accuracy of 98% on test data for recognizing apples, oranges, and lemons and 88% on validation data.

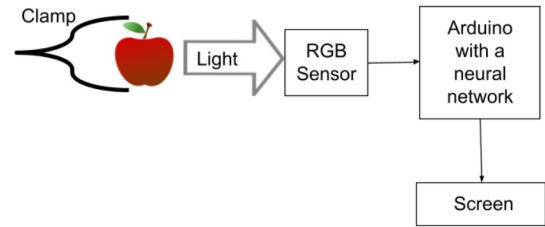


FIG. 1. Schematic of the project. The light is received by the RGB sensor which transmits the R, G and B values to the Arduino. After the data has been processed the results are displayed on the computer screen.

Another algorithm we will be testing is K-Nearest Neighbors. We will use the Arduino implementation of KNN [3]. The implemented model showed accuracy of 87%.

IV. LIST OF PARTS REQUIRED

- APDS9960 sensor
- Arduino Mega
- Clamp
- Breadboard and wires
- Fruits (Apples, Oranges, Lemons)

ACKNOWLEDGMENTS

We wish to acknowledge the support of the elab staff: Larry Sulak, Yaokun Situ and Jon Barlow.

[1] https://www.arduino.cc/reference/en/libraries/arduino_apds9960/.

[2] <https://keras.io/>.

[3] https://www.arduino.cc/reference/en/libraries/arduino_knn/.