Capstone 2
Fruit Classifier

Sava Dashev

Problem

Customers buy fresh fruits in grocery stores. We want to automatically recognize type of the fruit by the scale – fruit type.

Data source

The source of data is kaggle fruit-recognition data set.

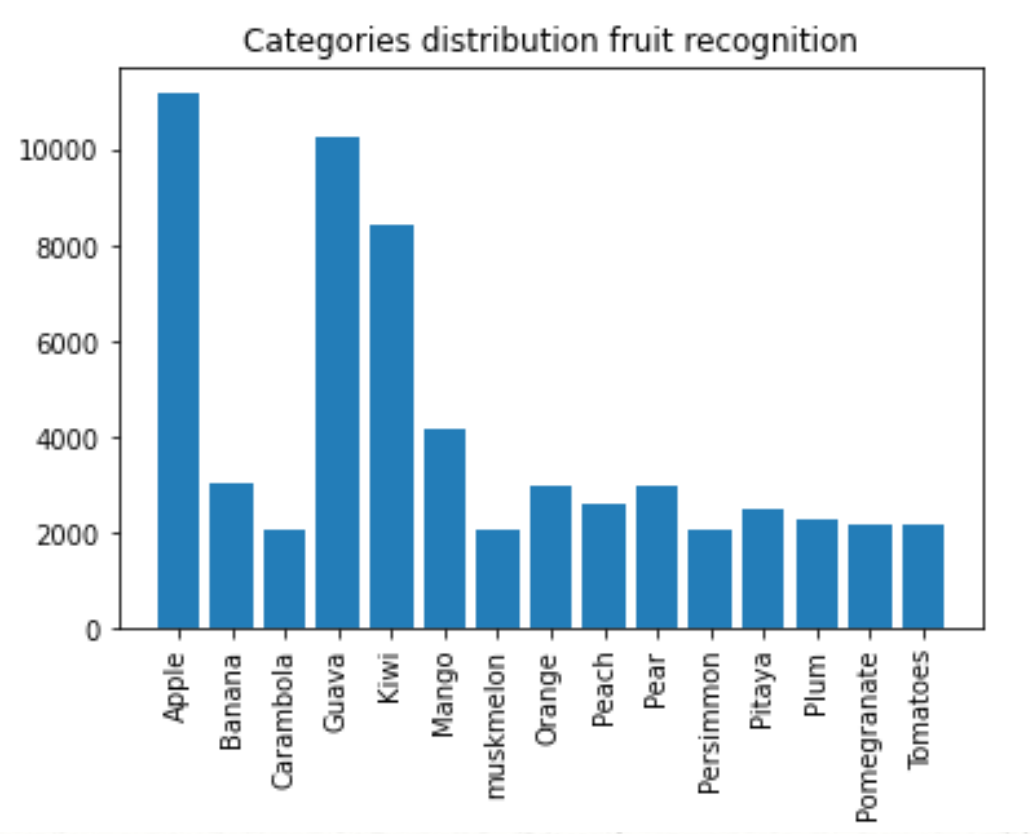
- Size of data is 8 GB.
- Each image is size (200, 200), (320, 258) or (480, 322).
- Only one image is size (320, 240).
- There are 54.

Data - classes

The images represent 15 different fruits.

- 3 classes have more than one sub-folder containing images;
- 9 classes have “resized” or “resized resized” images – 200 by 200 pixels.
-

Data distribution



Data findings

We found 15 type of images.

- Class Apples has the most images.
- Classes Guava and Kiwi has similar, but smaller amount images.
- All other classes have about 2000 to 4000 images each.

Data

We are going to use the images which are the most in the dataset.

We created file with names of images same size.

Further, we mix all names in train test split and saved for machine learning.