Capstone Project Proposal — seif tarek elshamy
Machine Learning Nano-degree — 12th of october 2018

I. Domain Background

Being updated with the latest technology such as deep learning and computer vision is one of the main reasons i wanted to do the project i will state next, these technologies specially deep learning became one of the most essentials in many fields (medical, self driving cars) and much more.
So in this project we have about 55k images in the data sets for 81 kinds of fruits (i will illustrate below the details of the data sets) our goal is to identify a new image(fruit) tested by the model to tell us what fruit is it. That can be made for a more advanced project that can be able to identify more than fruits or maybe identify anything at all.

II.Problem Statement

Fruits have many types even if you travel alot and get exposed to new fruits as there is some fruits grows up in different territories.

In the project we will train a CNN model to classify fruits into certain categories using keras.

III. Datasets and Inputs

In this project, I'll be using a data set from kaggle. it contains 55244. images of different kinds Fruits.

Total number of images: 55244.

Training set size: 41322 images (one fruit per image).

Test set size: 13877 images (one fruit per image).

Multi-fruits set size: 45 images (more than one fruit (or fruit class) per image)

Number of classes: 81 (fruits).

https://www.kaggle.com/moltean/fruits/home

License

MIT License

Copyright (c) 2017-2018 Mihai Oltean, Horea Muresan

Preview of the dataset

Input:

We will use Tenso Flow as backend, specifically, Keras CNN

IV.Solution Statement

The solution of this problem is to build a model that can classify the new images into one of the 81 classes accurately.
Im planning to make image classification using CNN using keras

i will use inception, xcepetion or VGG architectures, i will decide when i try on the project myself.

V.Benchmark model

I will benchmark my model against this research paper of the datasets owner (https://www.researchgate.net/publication/321475443_Fruit_recognition_from_images_using_deep_learning) this research paper of Mihai Oltean which achieved approximatly 96.2% of accuracy

VI. Evaluation Metrics

I'll be evaluating my model using accuracy score test to check
the accuracy i achieved.

VII. Project Design

Firstly, i will be importing the datasets (showed in the link above), then i will start preprocessing the images in the dataset will turn the data set into numpy array, will be also shuffeling the data(if needed) to avoid overfitting or i will manually do a validation set instead. Also will be rescaling of the images.
I will be creating a cnn model using keras(didnt decide which architecture yet)
And finally i will test the model to see what accuracy i achieved.
Just to re-ensure the model is working i will test the model with external images to see if it does good.