Machine Learning Engineer Nanodegree
Capstone Project Proposal
Classifying Flowers Data Set
Reddy Pavan Kumar
january 23rd, 2019

Proposal:-
Classifying Flowers Data Set

Domain Background:-
In modern times, people have sought ways to cultivate, buy, wear, or otherwise be around flowers and blooming plants, partly because of their agreeable appearance and smell. Around the world, people use flowers to mark important events in their lives:

- For new births or christenings
- As a corsage or boutonniere worn at social functions or for holidays
- As tokens of love or esteem
- For wedding flowers for the bridal party, and for decorations for the hall
- As brightening decorations within the home
- As a gift of remembrance for *bon voyage* parties, welcome-home parties, and "thinking of you" gifts
- For funeral flowers and expressions of sympathy for the grieving
- For worship. In Hindu culture adherents commonly bring flowers as a gift to temples

People therefore grow flowers around their homes, dedicate parts of their living space to flower gardens, pick wildflowers, or buy commercially-grown flowers from florists.

They are so many types of flowers around the world for that I have taken 5 flowers to classify

1. Daisy:- Bellis perennis is a common European species of daisy, of the Asteraceae family, often considered the archetypal species of that name. Many related plants also share the name "daisy", so to distinguish this species from other daisies it is sometimes qualified as common daisy, lawn daisy or English daisy.(Wikipedia)

2. Dandelion:- Taraxacum is a large genus of flowering plants in the family Asteraceae, which consists of species commonly known as dandelions. They are native to Eurasia and North America, but the two commonplace species worldwide, T. officinale and T. erythrospermum, were introduced from Europe and now propagate as wildflowers.(Wikipedia)

3. Rose:-A rose is a woody perennial flowering plant of the genus Rosa, in the family Rosaceae, or the flower it bears. There are over three hundred species and thousands of cultivars. They form a group of plants that can be erect shrubs, climbing or trailing with stems that are often armed with sharp prickles.(Wikipedia)

4. Sunflower:-Helianthus annuus, the common sunflower, is a large annual forb of the genus Helianthus grown as a crop for its edible oil and edible fruits. This sunflower species is also used as wild bird food, as livestock forage, in some industrial applications, and as an ornamental in domestic gardens.(Wikipedia)

5. Tulip:-Tulips form a genus of spring-blooming perennial herbaceous bulbiferous geophytes. The flowers are usually large, showy and brightly coloured, generally red, pink, yellow, or white. They often have a different coloured blotch at the base of the tepals, internally.(Wikipedia)

Symbolism:-

Many flowers have important symbolic meanings in Western culture. The practice of assigning meanings to flowers is known as floriography. Some of the more common examples include:

- Red roses are given as a symbol of love, beauty, and passion.
- Daisies are a symbol of innocence.

Reference links:- https://en.wikipedia.org/wiki/Flower

Essential Uses of Flowers in Our Daily Life:-

- Flowers Make The Best Gifts.
- Flowers Make Decorations Heavenly.
- Flowers Make Meals More Appetizing.
- Flowers Can Convey Your Deepest Emotions.
- Flowers Can Release Stress And Anxiety.
- Flowers Make You Look More Beautiful.

Reference links:-
https://manipalblog.com/6-essential-uses-of-flowers-in-our-daily-life
https://www.quora.com/What-are-the-uses-of-flowers

personal motivation:-

- I choose this project because I am really inserted in learning **Convolutional Neural Networks .** I know project some what challenging but still I am taking this project to learn **Convolutional** neural networks.
- I think in this project I can learn how to use the keras_pretrained models like to load them.
- In this project I can learn how to handle the image data and pre-processes datasets for the training the data.
- In this project I can also learn ,how to use of kaggle kernals.

Problem Statement:-

To understand the difference between types of flowers. The aim of this project is to predict the type of flower it belongs to by visualizing image.

In the project, I am going to use various Convolutional Neural Networks to predict the types of images and compare their performance and finally declare my final model.

Datasets and Inputs:-

The dataset that I am working is downloaded from https://www.kaggle.com/alxmamaev/flowers-recognition. This dataset contains 4242 images of flowers.

content :-

The pictures are divided into five classes: Daisy, tulip, rose, sunflower, dandelion. For each class there are about 800 photos. Photos are not high resolution, about 320x240 pixels. Photos are not reduced to a single size, they have different proportions

Citation:-

The data collection is based on the data flicr, google images, yandex images.

Dataset for classifying different flowers. Main categories have been taken:-

https://yandex.com/images

https://www.flickr.com/photos/tags/data

The data is open – sourced and can be download for education purpose with no citation.

Solution Statement:-

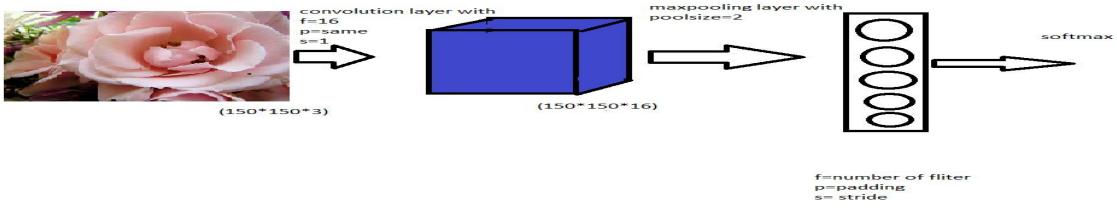
The classifier is a Convolutional Neural Network,which contain softmax activation function and I want to use momentum gradient descent optimizer. I explore the data set with opencv and matplotlib.pyplot libraries.in this project the following parameters can be tuned to optimize the classifier:

❖ Hyper parameters :-

1. learning rate alpha
2. Number of layers
3. Mini-batch size

Benchmark Model:-

In this flower classification, I want to use random assignment to set the worst score benchmark. With one convolution layer and one maxpooling layer and one dense with 5 units that classifies the data with minimum 20% accuracy.



.

Evaluation Metrics:-

I want to use accuracy as evaluation metric for flowers classification. Accuracy is a common metric for categorical classifiers

Accuracy = (images correctly classified) / (all images)

During development, a validation set was used to evaluate the model. I want to use a small set of training images as my validation images.

For validation I want to use "categorical_crossentropy" as loss metric for CNN, optimizer as "momentum gradient descent" and also metrics as "accuracy"

Project Design:-

The project is composed of different steps as follows:-

pre-processing:-

- First task is to read the dataset and perform visualizations on it to get some insights about the data.
- I want use the cv2 library for the reading the dataset and resize the (150*150*3) seens it is RGB image
- If any corrupted images are present in the dataset I skip the reading step for that kind of images
- After Data Exploration, I want to split the train data into training and validation sets and normalize the data to make it suitable for .Convolutional neural networks.
- Since I have small data set so I want use data augmentation so that my model may not overfit the data.

First step in training:-

First, I want to choose a Benchmark model which will at least gives testing accuracy around 20 % accuracy.

second step in training:-

- I want to apply RESNET50 Convolutional Neural Networks with some modification and use pre-trained weights on the data
- I want to apply VGG 16 Convolutional Neural Networks with some modification and use pre-trained weights on the data.
- Then I want to tune the hyper parameter like learning rate,number of hidden units,number of epochs etc.

After that I announce the best model from above two.

I want to train my model in kaggle which provides both cpu and gpu for training, RAM and Disk to store the data.

Finally, I will declare the model with highest accuracy on both training and testing data sets as my model.

Sample images:-


Daisy


Dandelion


Rose


sunflower


Tulip