Machine Learning Engineer Nanodegree

Capstone Proposal

by Mayur Nehete

Proposal

Domain Background

Convolution Deep Learning has a good impact on the image recognition capabilities.Even simple models are able to make highly accurate predictions on datasets like the MNIST database of handwritten images.

To recognise image on the real world condition is much more difficult. As to differentiate plants seedlings. It is kaggle Dataset provided by Aarhus University Department of Engineering Signal Processing Group. The classification of the seedling is very important because so it can differentiate as Hazardous or Non Hazardous.

The images of the Dataset contains leaves of the seedlings with is sometimes very small and the Images shapes ranges from (49,49) to (3000,3000)px. So it is require to recognise patterns in the image . So the Machine Learning is not ideal for this problem. Convolutonal Net should be use to recognise patterns in the image

Problem Statement

The goal of this project, however, is to train a deep learning model that it able to recognize an plants seedlings. The model should be able to recognize plants seedlings with a degree of accuracy.

Datasets and Inputs

This project will use dataset of Plant Seedling Classification dataset. a dataset containing images of approximately 960 unique plants belonging to 12 species at several growth stages. The dataset have 4750 training images and 794 test images.
The Datastet is unbalanced lowest number of training example of the class is 221 and highest is 654 so the lowest number of 221 examples of each category to be used for train the model.
.

Labels	Count
Black-grass	263
Charlock	390
Cleavers	287
Common Chickweed	611
Common wheat	221
Fat Hen	475
Loose Silky-bent	654
Maize	221
Scentless Mayweed	516
Shepherds Purse	231
Small-flowered Cranesbill	496
Sugar beet	385

The image size of images vary in the Dataset. Image size ranges from (49, 49) to (3991, 3457) so all the images in the Datasets needs to resize.

Dataset Used: https://www.kaggle.com/c/plant-seedlings-classification#description

Solution Statement

As deep learning techniques have been very effective in image classification over the years, in this project, transfer learning will be used to train a convolutional neural network to classify images of fish to their respective classes. Transfer learning refers to the process of using the weights from pre-trained networks on large dataset. Fortunately many such networks such as RESNET, Xception, pretrained on [imagenet challenge](http://www.image-net.org/challenges/LSVRC/) is available for use publicly.

Benchmark Model

By using SVM on the features extracted from xception model got a score of 0.52 on the Validation set This represents an ambitious but more attainable goal. Model accuracy will be judged based on predictions made on the Validation set..

Evaluation Metrics

Submissions are evaluated on MeanFScore, which at Kaggle is actually a micro-averaged F1-score.
MeanFScore= 2 (Precision * Recall) / (Precision + Recall)

Project Design

Programming language:- Python 3.6
Libraries:- Keras, Scikit-learn
Workflow:-
1. There are 221 example class and 654 of other class to get a bias result the lowest 221 examples of the each class is used to train the model because by data augmentation if the dataset size increase the RAM on the kaggle burns.
2. The Image is resize to (299,299) because to retain information quality in the images ,images is preprocess by preprocessing of the exception model then features are extrated from pretrained exception model.
3. Trainin a Dense Deep neural network on the Training set that has 12 output neurons on the last layer to get predictions.

References

1. http://cs231n.github.io/transfer-learning/ "CS231, Andrej Karpathy's overview on Transfer Learning".