Capstone Proposal

Machine Learning Engineer Nanodegree

Plant Seedlings Classification

Domain Background

The ability to efficiently differentiate between a weed and crop seedling would allow crop growers and farmers to improve their crop yields and maintain the health of their corner of the environment. In addition, it would aid amateur and hobbyist crop growers in identifying any weeds growing among their crops. This capstone project is based on the Plant Seedlings Classification challenge on Kaggle [1].

Problem Statement

The goal of this project is to correctly identify the species of a weed or crop seedling given its image. Given multiple plant images to be classified, the goal would be to maximize the micro-averaged F1-score produced by the classifier's predictions; this would depend on the number of true positives, false positives, and false negatives reported by the classifier [2]. The classifier should be able to work with any seedling that belongs to one of 12 specific plant species.

Datasets and Inputs

The dataset to be used in this project was provided by the Aarhus University Signal Processing Group in collaboration with the University of Southern Denmark [3]. This data comprises 4750 plant images split among 12 different plant species; the colour images themselves vary in dimensions and may require normalization before they can be used in developing the classifier.

Since this data is directly related to this project's problem statement, it can be used to both train the classifier and test it for accuracy. To make the most out of this dataset and reduce overfitting in general, the data will be split into separate training and testing sets.

Solution Statement

A Convolutional Neural Network would be the tool of choice for this project; since the challenge is to classify images of plants, a CNN is well suited for detecting basic features within certain areas of an image and then building on them to learn more complex features, allowing it to accurately identify high-level elements within an image [4]. For this particular project, a relatively simple CNN that consists of 3 convolutional layers interleaved with 3 pooling layers will form the core of the initial, exploratory model. This network would accept a normalized image as an input and would have its outputs passed into a global average pooling layer for dimensionality reduction, followed by a final dense layer that comprises 12 nodes, one for each of the plant species to be predicted.

Benchmark Model

A basic benchmark model would be a classifier that randomly assigns a plant species to a given image. Such a model would provide an intuitive baseline against which a more complex model could be evaluated; by calculating the micro-averaged F1-score produced by this random classifier, it can be easily compared against the same score produced by other models.

Evaluation Metrics

As mentioned in the Problem Statement and Benchmark Model sections, the micro-averaged F1-score will be used to evaluate the accuracy of the benchmark model and any classifiers:

$$F1_{micro} = \frac{2Precision_{micro}Recall_{micro}}{Precision_{micro} + Recall_{micro}}$$

This metric is based on the precision and recall produced by a classifier's test, which are in turn based on the number of true positives, false positives, and false negatives (for each class *k*) produced by the test:

$$Precision_{micro} = \frac{\sum_{k \in C} TP_k}{\sum_{k \in C} TP_k + FP_k}$$

$$Recall_{micro} = \frac{\sum_{k \in C} TP_k}{\sum_{k \in C} TP_k + FN_k}$$

Project Design

To begin, each image will be resized to 224 x 224 pixels, converted to an array of size 224 x 224 x 3 (accounting for the 3 RGB channels), and rescaled by having each pixel in each channel divided by 255. This normalization would allow all images to be fed into the same CNN, which accepts images of fixed size, and would scale each colour channel such that it ranges from 0 to 1 rather than 0 to 255.

After the images have been normalized, a random 80% of images will be designated the training set, while the remaining 20% will be designated the testing set. This would give the classifier a chance to predict data that it had never "seen" (i.e. trained on) before; the results of which would be used to determine the accuracy of the classifier according to the established evaluation metric.

With the images normalized and the training and testing sets prepared, the training data will be fed into a CNN of (initially) 3 convolutional layers interleaved with 3 pooling layers, as outlined in the solution statement. Each convolutional layer will consist of a 2 x 2 kernel, 16 filters, and a ReLU activation function, while each pooling layer will be represented by a max pooling function with a 2 x 2 filter.

This narrowing and deepening network is intended to extract only the relevant features of an image while keeping the total number of parameters manageable. The output of the final pooling layer will be fed into a global average pooling layer to further condense the network into a 1D array of nodes; this will connect to a final dense layer that comprises 12 nodes, one for each of the plant species to be predicted.

A summary of such a model may be visualized as follows:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 224, 224, 16)      208
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 112, 112, 16)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 112, 112, 32)      2080
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 56, 56, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 56, 56, 64)        8256
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 28, 28, 64)        0
_________________________________________________________________
global_average_pooling2d_1 ( (None, 64)                0
_________________________________________________________________
dense_1 (Dense)              (None, 12)                780
=================================================================
Total params: 11,324.0
Trainable params: 11,324.0
Non-trainable params: 0.0
```

To start off, this model will be trained on the training set for 5 epochs. From that point on, the number of epochs, layers, and their specific parameters will be tuned with the intent of producing a reasonably accurate plant species classifier.

References

[1] https://www.kaggle.com/c/plant-seedlings-classification

[2] https://www.kaggle.com/wiki/MeanFScore

[3] https://vision.eng.au.dk/plant-seedlings-dataset/

[4] http://deeplearning.net/tutorial/lenet.html