This project is taken from kaggle competition:

https://www.kaggle.com/c/invasive-species-monitoring/

Domain Background

Concentration: Computer Vision

Tangles of kudzu overwhelm trees in Georgia while cane toads threaten habitats in over a dozen countries worldwide. These are just two invasive species of many which can have damaging effects on the environment, the economy, and even human health. Despite widespread impact, efforts to track the location and spread of invasive species are so costly that they're difficult to undertake at scale.

Currently, ecosystem and plant distribution monitoring depends on expert knowledge. Trained scientists visit designated areas and take note of the species inhabiting them. Using such a highly qualified workforce is expensive, time inefficient, and insufficient since humans cannot cover large areas when sampling.

This problem can be solved using Image classification using neural nets. This is one of the applications of computer vision Domain. This is one of the problems we should consider solving now as its important for the maintenance of ecosystem and it's a burning issue.

This particular project is picked up from kaggle challenge. More Details of which can be found here.

https://www.kaggle.com/c/invasive-species-monitoring

Problem Statement

The problem at hand is to analyse the sample images and label them correctly as efficiently as possible, whether a particular sample has dangerous levels of these two invasive species or not. We can use Image classification to train the neural network model with the already

available training data to label the new unseen samples. we can use the accuracy metric of the neural networks to define the accuracy.

Datasets and Inputs

The data set contains pictures taken in a Brazilian national forest. In some of the pictures there is *Hydrangea*, a beautiful invasive species original of Asia. Based on the training pictures and the labels provided, the participant should predict the presence of the invasive species in the testing set of pictures.

.Each data sample used here is an image which needs to be analysed for excessive growth of invasive plant species. This data set is obtained from the kaggle competition of invasive plant species challenge.

Data Files Description

- train.7z - the training set (contains 2295 images).
- train_labels.csv - the correct labels for the training set.
- test.7z - the testing set (contains 1531 images), ready to be labelled by your algorithm.

Train.7z contains 2295 image samples, where train_label.csv contains the label for that image(image number and label pairs).

Solution Statement

we can use deep neural networks for classifying the images. Given the data set is a set of images it would be better to use convolutional neural networks to achieve better accuracy. We can use gradient descent for training the neural networks.

Benchmark Model

We can use a fully connected layer with one hidden layer as a bench mark model. It serves as a basic model for classification. And we can then try and improve the accuracy using convolutions.

Evaluation Metric

We can use the basic accuracy as the evaluation metric. Meaning simply the number of samples it got correctly out of all the samples in test data, calculated in percentages.

Project Design

The approach would be to split the train data into training data and validation data first. And then train the benchmark model on train data by testing it on the validation data simultaneously and find the accuracy of that on the test data. Then train convolutional neural network model on the training data by testing it on validation data. The training data can be split into many batches and fed to the model. And this can be repeated for many number of epochs and also being careful not to over fit the model. we can use gradient descent to minimise the loss. The number of hidden layers and hidden layer's size and size of the convolution filter and number of epochs used can determine the accuracy. They can be sufficiently tuned to achieve greater accuracies. Finally, when a the validation accuracy is satisfactory this model can be tested with the testing data for the final accuracy.

This particular project is picked up from kaggle challenge. Details of which can be found here.

https://www.kaggle.com/c/invasive-species-monitoring