Plant Recognition using Spatial Transformer Network

Azhar Talha Syed, Suresh Merugu, Vijaya Kumar Koppula

***Abstract:** Agriculture is one of the most prominent work sectors in countries like India. However, the majority of farmers are unaware of the modern plant diseases and the methods are to be followed to expect a better yield from their crops. Data science and Machine Learning have made a great progress in recent years for providing a solution to problems like these. Findings: By developing a system which will help the farmers in getting aware about the different species of plants without having a need for definite education would be very helpful to them. Objective: In this paper, we propose an efficient way of recognizing plants using cell phone cameras, as it will be very easy for the farmers and also other people who have their work involving plants, to get information about a plant which will help them in their work. We also provide a performance analysis on our solution and the previous work in this paper. Methods/Statistical Analysis: In Machine Learning terminology this is a multiclass classification problem where the input is an image and the expected output is the class of which the plant in the image belongs to. There are several ways of solving a multi-class classification problem such as using K nearest neighbors, Multiclass Support Vector Machines, Neural Networks, and Convolutional Neural Networks. But for this problem, we also take user convenience into consideration and we suggest the use of Spatial Transformer Network as the classification will still be accurate whilst the image is not properly aligned and has a lot of noise in it.*

***Keywords:** Plant Recognition; Deep Learning; Convolutional Neural Networks; Spatial Transformer Network.*

I. INTRODUCTION

With the advancement of science and technology, the quality of life has improved drastically from before. Especially in the field of agriculture, there has been a lot of improvement. Many new techniques of cultivation have been discovered lately and comparing the yield produced per square feet area from earlier and now, there has been a total average increase in these recent times [10]. But of course, there is always a room for improvement. Educating farmers about the new techniques for cultivation of a crop is important to keep them updated with the new techniques. So we can use the advancements of technologies to meet these goals.

Lately, we have noticed great improvements in computer vision. Algorithms such as Convolutional Neural Networks [1] are one of the main reasons behind this. In this paper, we propose a technique for plant classification, which is the use of Spatial Transformer Networks. The agenda here is to be able to predict the plant species using an image of a leaf of the plant. This is a tough classification problem as the similarities between inter classes are very high and similarities between intra classes are very low. Although a lot of work has been done in Computer Vision and some great results have been noticed, the performance in this particular domain is still not up to the mark and improvements could be necessary.

As the goal here is to build a classifier which can correctly classify the plants for new test data rather than just performing well on the training data, we propose the use of Spatial Transformer Networks. Also as it is not necessary that the picture of the plant is always in alignment, some pictures may contain plants which are not in the center of the image but a little aside from the center or the image could be an inverted image of the usual image in training data. All such problems can be solved using Spatial Transformer Networks, as the dimensions of the image are first adjusted to bring the image close to training data and then the adjusted image is passed through a CNN [2]. Because of this transformation, the classification can still be done accurately although there are problems as discussed.

The structure of the paper is as follows. In Sect. 2. It describes the previous work done related to this problem. In Sect. 3. The methodology used in our solution is provided along with a reasoning for using the solution. In Sect. 4. The results and analysis of our solution are provided. Finally in Sect. 5. A conclusion with a suggested solution is given.

II. RELATED WORK

With a huge increase in the utilization of mobile phones, especially the ones with cameras, new mobile apps such as Pl@ntNet [6] have come into existence where people from all over the world are able to upload pictures of a wide range of species of plants. This will help in training much better classifiers as the chances of overfitting reduce with an increase in the size and variations in the dataset [7].

Jyotismita Chaki and Ranjan Parekh proposed Plant Leaf Recognition using Shape based Features and Neural Network classifiers [8]. Here they suggest the use of two models which are Moments-Invariant (M-I) model and Centroid-Radii (C-R) model for recognizing plant leaves corresponding to three types of plants.

In 2017, Wang-Su Jeon and Sang-Yong Rhee have proposed Plant Leaf Recognition using a Convolutional Neural Network [9].

Where they used Flavia dataset to classify 32 plant species and used a Convolutional Neural Network with a structure similar to GoogleNet for classification.

III. . METHODOLOGY

In this section of the paper, we will discuss the Spatial Transformer Network that we used for Plant Recognition and

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

the tweaks that were used to boost the performance of classification. For training,
we have used publicly available 50,000 images of 1000 plants and 10,000 images for testing.

3.1. Spatial Transformer Network

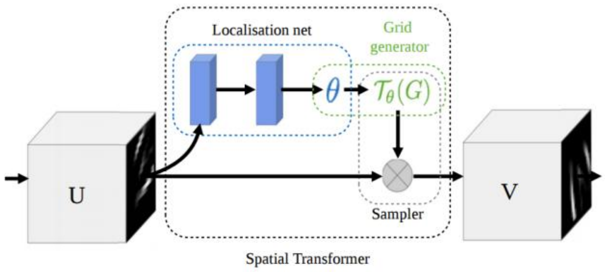

Fig. 1 General Architecture of Spatial Transformer Networks

In a general Convolutional Neural Network, there are no transformations performed on the input image before sending it to the Convolution layers. Whereas in an STN the dimensions of the input image are adjusted such that the image is oriented in a similar fashion as the images in training data, for a better prediction. In addition to passing the input image through a CNN, the image is first passed through the Spiral Transformer.

The Spiral Transformer is split into three parts, which are:

- Localization Network, this network takes the image as input and outputs a parameter , which is later used in transformation. For doing this prediction either a fully connected layer can be used or another Convolution Network can be used.
- Grid Generator, using predicted by the Localization Network, a sampling kernel is generated.
- Sampler, to transform the input image, sampling kernel is applied centered at specific location in the input image.

3.2. Spatial Transformer Network for Plant Recognition

Fig. 2 Classification of maple tree using Spiral Transformer Networks

To implement Spatial Transformer Network we have used Fully connected Neural Network for Spiral Transformation and a Convolutional Neural Network with a structure similar to AlexNet were used.

I. Structure and training methodology of Fully connected Neural Network for Spiral Transformation

- The Network contains two hidden layers with 1024 neurons in each layer with ReLU as the activation function
- The output layer has 227 neurons representing each pixel values of the transformed image
- To overcome overfitting Dropout [3] the two hidden layers were used during the training with a dropout rate of 0.25
- Xavier Initialization [4] for initializing weights and biases of the network
- Batch Normalization after the first hidden layer
- Adam optimizer with 0.001 learning rate

(II). After the transformation, we will have an image with dimensions (227x227x3). As the image is already transformed, it will be easier to make a prediction. However, Plant classification is still a very tough job hence we decided to use a very strong Convolutional Neural Network such as ZFNet [5]. The Architecture of ZFNet is as following.

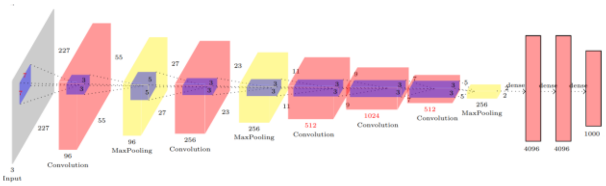

Fig. 3 The architecture of CNN for plant classification

. Convolution Layer: Input - a tensor of dimensions (227 x 227 x 1); number of filters - 96; kernel dimensions - (7 x 7 x 1); Padding - 0; stride - 4;

b. Max Pooling Layer: Input - a tensor of dimensions (54 x 54 x 96); Pool size - (2 x 2 x 1); strides - (2 x 2 x 1);

c. Convolution Layer: Input - a tensor of dimensions (27 x 27 x 96); number of filters-256; kernel dimensions - (5 x 5 x 96); Padding - 0; stride - 1;

d. Max Pooling Layer: Input - a tensor of dimensions (23 x 23 x 256); Pool size - (3 x 3 x 1); strides - (2 x 2 x 1);

e. Convolution Layer: Input - a tensor of dimensions (11 x 11 x 256); number of filters - 512; kernel dimensions - (3 x 3 x 256); Padding - 0; stride - 1;

f. Convolution Layer: Input - a tensor of dimensions (9 x 9 x 512); number of filters - 1024; kernel dimensions - (3 x 3 x 512); Padding - 0; stride - 1;

g. Convolution Layer: Input - a tensor of dimensions (7 x 7 x 1024); number of filters - 512; kernel dimensions - (3 x 3 x 1024); Padding - 0; stride - 1;

h. Max Pooling Layer: Input - a tensor of dimensions (5 x 5 x 512); Pool size - (2 x 2 x 1); strides - (2 x 2 x 1);

i. Fully Connected Layer: Input - a vector of length 1024; output neurons - 4096;

j. Fully Connected Layer: Input - a vector of length 4096; output neurons - 4096;

k. Softmax Layer: input - a vector of length 4096; output neurons - 1000;

Training methodologies

- Batch normalization after first and second max pooling layer
- Adam optimizer with learning rate 0.001
- Loss function categorical cross entropy

The network totally has approximately 29 million parameters. For the whole system i.e. the Spiral Transformer and the Convolutional Neural Network, the system has approximately 30 million parameters.

IV. RESULTS

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fig. 4 Some predicted classes by the system

After training both of the Neural Networks we achieved an accuracy of 99.12% on test dataset for classifying 1000 plant species on the Pl@ntNet dataset. However, after training we collected 40 leaf images of different plant species and fed them as input to our system and 37 of them were classified correctly although these had a lot of noise and were not properly oriented. Some of the successful classifications made by our system are shown in Fig 4

V. CONCLUSION

In the real world, it is not necessary that all the images are taken in the same orientation, by the same camera or in the same conditions as the training data especially in the case of Plant Recognition by images is provided by a usual human being. As Spatial Transformer Networks are specially designed to handle such situations, we noticed such fruitful results in our experiment.

So In this paper, the use of Spatial Transformer Networks was proposed as an improvement over the current existing methodologies for Plant Recognition, along with the implementation of the proposed system.

REFERENCES

1. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton in "ImageNet Classification with Deep Convolutional Neural Networks" Advances in Neural Information Processing Systems 25 (NIPS 2012).
   Max Jaderberg, Karen Simonyan, Andrew Zisserman, Koray Kavukcuoglu in "Spatial Transformer Networks" [Online]. Available: https://arxiv.org/pdf/1506.02025.pdf.
2. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov in "Dropout: A Simple Way to Prevent Neural Networks from Overfitting" The Journal of Machine Learning Research, 2014, pp 1929-1958.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun in "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification".
4. Matthew D. Zeiler, Rob Fergus, 2013, "Visualizing and Understanding Convolutional Networks".
5. Hervé Goëau, Pierre Bonnet, Alexis Joly, Julien Barbe, Souheil Selmi, Vera Bakic, Vera Bakic, Jennifer Carré, Daniel Barthelemy, Nozha Boujemaa. "Pl@ntNet Mobile App" .
6. Michael Nielsen, in is book Neural Networks and Deep learning - Chapter 3 - Overfitting and Regularization. [Online].Available: http://neuralnetworksanddeeplearning.com/chap3.html#other_techniques_for_regularization. Jyotismita Chaki and Ranjan Parekh. "Plant Leaf Recognition using Shape based Features and Neural Network classifiers" in International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011.
7. Wang-Su Jeon and Sang-Yong Rhee. "Plant Leaf Recognition Using a Convolution Neural Network " in International Journal of Fuzzy Logic and Intelligent Systems Vol. 17, No. 1, March 2017, pp. 26-34.
8. Jonathan P. Caulkins (2010) in his research work "Estimated Cost of Production for Legalized Cannabis" at RAND, Drug Policy Research Center

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication