

Leaf Recognition using Convolution Neural Network - AlexNet

Mihir Samel
Computer Engineering department,
Saraswati College of Engineering,
Kharghar, India

Abhishek Tiwari
Computer Engineering department,
Saraswati College of Engineering,
Kharghar, India

Prof. Suhasini Parvatikar
HOD, Computer engineering
Department,
Saraswati College of Engineering,
kharghar, India

Abstract—In today’s world as everything is equipped with artificial intelligence, we require something to recognize plants too. Recently, Convolution Neural Network is becoming very popular due to its accuracy. AlexNet, an 8-layer convolution neural network is used to perform leaf recognition. First, Data Augmentation is performed, which includes multiple transformations such as rotation, flipping (horizontal or vertical), translation etc. which increases dataset size and also reduces problem of over-fitting. Batch Normalization is then performed to normalize values of each layer which further helps to reduce over-fitting. Dropouts are also used to form numerous different networks which again reduces over-fitting up to a large extent. Network is trained and classification is performed to recognize various plants which can be presented using GUI.

Keywords—CNN, leaf recognition, Batch Normalization, Dropout, Data Augmentation, AlexNet.

I. INTRODUCTION

Trees are the organisms having longest life span on earth and also forms very large part of earth. Lots of people in research as well as botanist requires information about plants. Also, some medical fields use plants as their main form of medicine for e.g. Ayurveda. Main difficulties they face is while identifying various plants as many of the plants may have very high similarities. Leaves are generally recognized through overall shape, pattern of veins, color of leaf, smoothness or roughness of the surface, etc. It is very difficult for a human being to memorize various combinations of above features. Such complex and visually impossible characterization are generally done by machines with very high accuracy and probability of correct results. We are using Deep Learning because it is very effective for categorizing unstructured data or data without any kind of labeling. Deep Learning is subset of artificial intelligence which copies human activities and help to make important decision which are very close to human decisions. Many of today’s applications uses deep learning as it is more accurate and less prone to error as compared to other alternatives. The huge amount of data generated every second needs to be analyzed for proper decision making, this data is generally called as big data. Deep Learning can process this huge amount of data very efficiently and give accurate predictions. Companies realize the power of this technique that can arise from exposing this wealth of information and are becoming increasingly amenable to AI systems for automated support. Deep Learning techniques are hence used today in various fields.

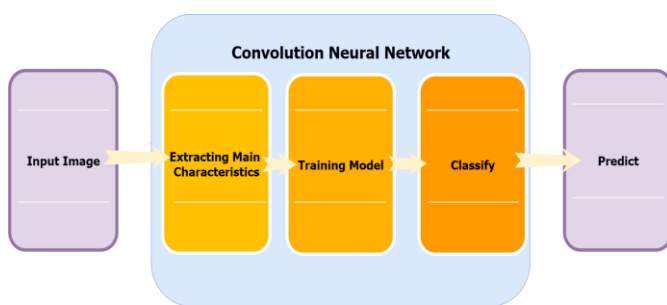


Fig. 1: General flow of our work.

There are various CNN architectures like VGG Net, ResNet, Dense Net, Inception Net, AlexNet and many more. For this particular project we are using AlexNet architecture. It has 5 concentric layers and three layers that are completely connected. The relu is applied after basically each layer of convolution as it is being used as activation function. The dropout is applied to literally drop out some neurons to deal with overfitting problem of the network. AlexNet was a pioneer at CNN and opened a whole new world for researching and development. AlexNet implementation is much easier after such deep learning libraries are released.

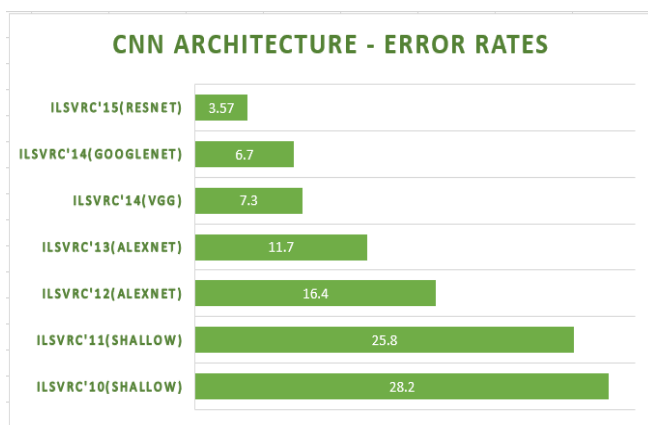


Fig. 2: Architecture error rates

II. LITERATURE SURVEY

A literature survey is nothing but literature which I belonging to some subject and some extra information about it that describes it. It gives ideas about overall information about author, various queries related to respective subject and also various techniques and methods regarding the subject. As such, it is not a research in itself, rather it reports findings of other authors that are relevant to our topic and might help. Chaoyun Zhang, Pan Zhou, "An interactive neural network for recognition of leaves using data augmentation" - data enhancement is adopted to reduce the over-fitting degree of the model. To reduce over-fitting and get better performance, we create surrogate training data with some label-preserving changes. Although this job increases the time for training, the model finds its success with a tremendous load of data with high accuracy.

Christophe Bati and Frank Puppé, "Leaf Identification Using a Deep Convolution Neural Network" - here the "Supervised Pretraining" method, also known as "transfer learning". The basic concept used here is to use large sized data set that is very different from your actual dataset on which your network is going to get pretrain. Filters of fasteners are not randomly initialized, but rather set to multicolored values learning already useful normalization features. Wang-Su Jeon and Song-Yong Rai, "Plant leaf recognition using a convolution neural network" -CNN architecture used here is GoogleNet from Google which has more layers but with less computational complexity and higher accuracy. Inceptions modules used in this architecture uses multiple resolutions in parallel to extract various attributes.

Iofee and Szegady, C2015-Batch Normalization technique makes process of initial weights determination much easier and also allows faster learning. We can skip Dropout in some cases and use batch normalization instead. It gives more accuracy and less training time when applied to network which is a significant difference. Muhammad Rizwan, "Keras using the implementation" <https://engmrk.com/alexnet-implementation-using-keras/> - Extensive Alexnet architecture with multiple convolutions, max-pooling and fully-connected layers. Implementation of each layer using layer library in layer by python. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevski, Ilya Sutskever, Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting" Dropout is generally used to solve problem of overfitting by dropping out some neurons and creating different structures for training. At the time of testing, it suppresses the effect of the average of forecasts for all these thin networks by using a single network with smaller weights. This significantly reduces our problem and gives huge success over other methods.

III. PROCESS OF RECOGNITION

A. Data Augmentation

It is generally better to augment your dataset before training the network. Data Augmentation creates slightly varied copies of each image to train network. This artificially created data helps network to handle predictions of images that varies slightly from data on which network is trained. In our project we are augmenting dataset with variations such as rotating image with some degree, horizontal and vertical flipping and zooming. In our case this helps our network for predicting leaf images taken from various angles and perspectives.

Modern deep learning algorithms, such as the Conventional Neural Network, or CNN, can capture various attributes that are immutable at their location in the image. Nevertheless, enhancements can further aid this transformative approach to learning and help model learning features that vary from top-to-bottom order from left-to-right in variations, light levels in photographs. There are many more adaptations to do. Image data enhancement typically applies only to training datasets, and not to validation or test datasets. Data preparation is different and not suitable to us it does not perform enhancement consistently across whole dataset.

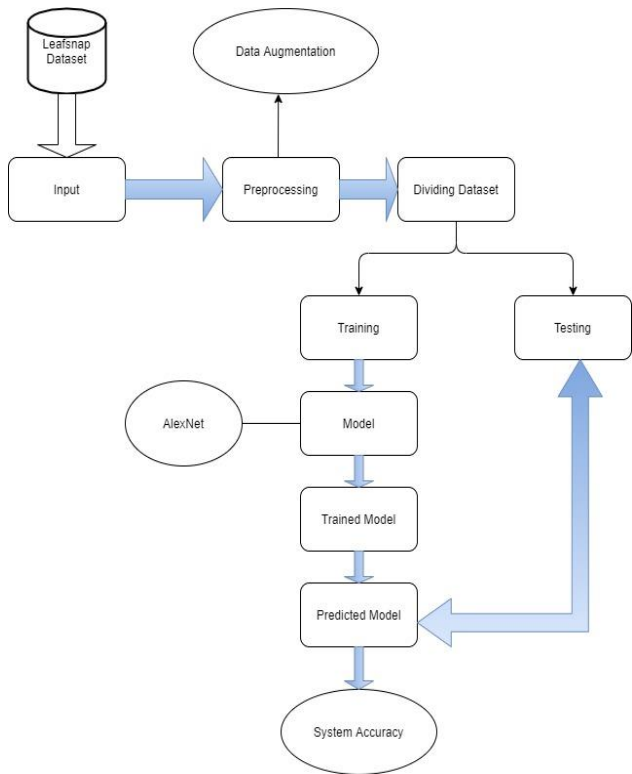


Fig. 3: Process of Recognition

B. Steps in Recognition

In the process of recognition main part is dataset upon which model is going to get trained. In our case dataset is Leafsnap that contains about 185 categories of leaf images with and without segmentation. Preprocessing is done very first step in

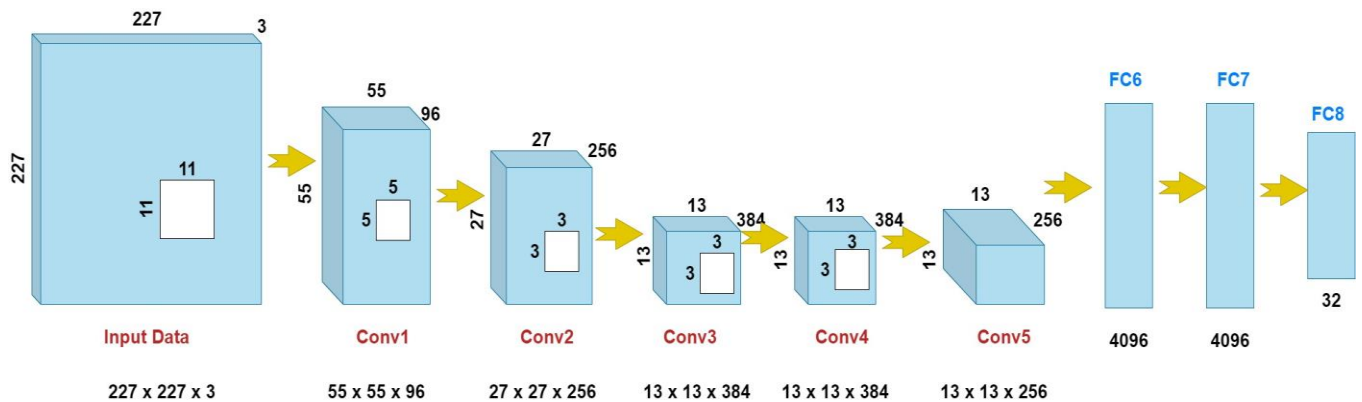


Fig. 4: AlexNet Architecture

discussed in previous section. In next step complete dataset is divided in training and testing datasets in some proportion for validation and testing of data trained on training dataset. For training data, we are using AlexNet model consisting of 8 layers out which five are convolution layers and three are fully connected layers. Further in process after model is trained predictions are done on testing dataset as we have discussed earlier. Once predictions are done system accuracy can be determined by comparing desired results with observed results.

IV. CNN ARCHITECTURE FOR RECOGNITION

A. Why CNN ?

The ImageNet initiative is competition for object classification. When it comes to training a fully connected network large numbers of parameters comes into picture. For example, an image of 64 x 64 x 3 needs 12288 parameters in first layer and increases by every layer. Also, for an image with high resolution huge numbers of parameters will have to be trained. This problem mainly arises due to each neuron in fully connected network is connected to all neurons in next layer. But convolution neural networks this is not the case, perhaps each neuron in previous layer is connected to nearby neurons in the next layer. This property of CNN makes it special as it reduces numbers of parameters to be train and also resolves overfitting issues up to an extent.

B. AlexNet

AlexNet is a CNN architecture consisting of five convolution layers and three fully connected layers. This architecture converts image of any dimensions into size of $224 \times 224 \times 3$. First layer is convolution layer consisting of 96 filters each of size $11 \times 11 \times 3$. These filters are also called as feature maps which are used to capture features of the input image. First layer is followed by max-pooling layer which is used to reduce feature map dimensions by calculating largest or maximum value in the patch of feature map. Second layer is also a convolution layer with 256 filters having size $5 \times 5 \times 3$ and followed again by a max-pooling layer.

Third, Fourth and Fifth layers are also convolution layers 384, 384 and 256 feature maps respectively and each of size $3 \times 3 \times 3$. These three consecutive convolution layers are followed by a single max pooling layer. Next three layers are fully connected layers and lastly an output layer. Sixth layer is of size 9216 neurons and followed by next two layers with 4096 neurons each.

All above layers except output layer have ReLU as an activation function. It is default activation function used in many networks as it is very easy to implement using simple maximum function and it is capable of outputting a true zero value. It outputs the input directly if it is positive else just outputs a zero. Output layer on the other hand consist Softmax activation function which is a logistic regression that normalizes input values to array of probability distribution that sums up to 1.

V. DATASET

The Dataset we have used is leafnap dataset containing about 185 classes of leaf images. This dataset is divided into two major parts lab color images and segmented images. Further both the major categories are divided into lab images taken in lab with proper artificial lighting conditions and field images taken in daylight conditions. We have mainly used color images for the process of recognition.

VI. RESULT

The AlexNet model described previously was tested and the system is able to recognize categories of leaf with on an average 95 % accuracy.

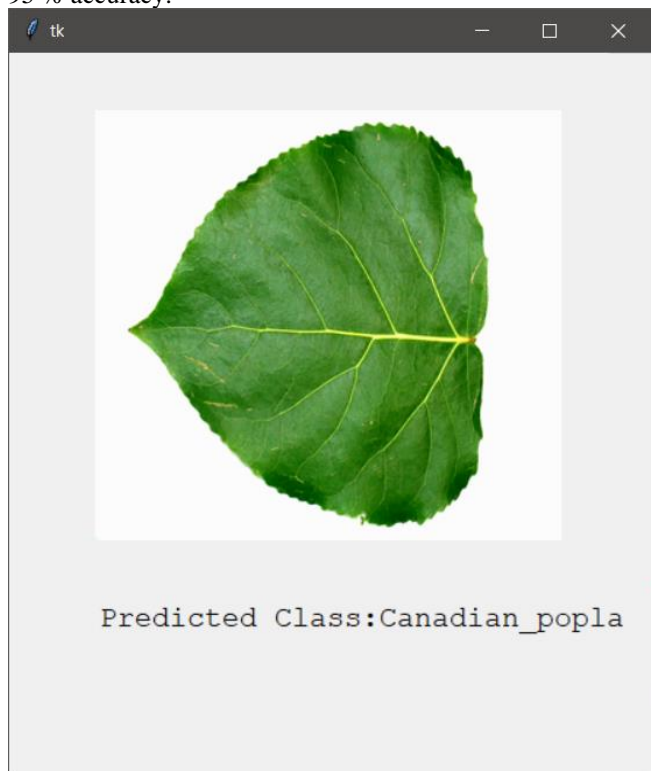


Fig. 5: Leaf predicted as Canadian popla

```
[INFO] loading network...
[INFO] classifying image...
abies_nordmanniana: 71.77%
acer_campestre: 19.12%
acer_palmatum: 99.95%
ailanthus_altissima: 91.90%
betula_nigra: 79.84%
liriodendron_tulipifera: 19.47%
pinus_cembra: 11.05%
```

Fig. 6. Matching percentage of categories

The image above shows different categories with their respective matching percentage with the image of interest.

VII. CONCLUSION

The accuracy factor of the system proposed is been meticulous throughout since the filtering of the layers for the parameters have been very precisely operated and the system efficiency can be increased by implementing and overlapping of other methods and enhance the research and development and squelch the errors extensively. AlexNet has the edge over other methods adopted for the implementation of this system, however it is not presumptuous. The improvident parameters can be kept vacant and it is copious to have this method to work upon a variety of resonating fields and development. Input image goes from various filters in architecture and distinguishing features from each image are extracted and recorded. These features are further used for detecting various types of plants using their leaves. Finally, the system will able to identify plant from given image within few seconds.

VIII. REFERENCES

- [1] D. G. Lowe, "Distinctive image features from scale invariant key points," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004. <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- [2] Wang-Su Jeon and Sang-Yong Rhee, "Plant Leaf Recognition Using a Convolution Neural Network" <http://dx.doi.org/10.5391/IJFIS.2017.17.1.26>, ISSN 2093-744X
- [3] Image preprocessing for CNN", Available at " <https://towardsdatascience.com/image-pre-processing-c1aec0be3edf>"
- [4] Ioffe, S, and Szegedy, C 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift., Proceedings of the 32nd International Conference on Machine Learning, Lille, France, pp.448-456.
- [5] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever Ruslan Salakhutdinov," Dropout: A Simple Way to Prevent Neural Networks from Overfitting".
- [6] "Loading Datasets from various directories into keras network for various uses in CNN" Available at : " <https://machinelearningmastery.com/how-to-load-large-datasets-from-directories-for-deep-learning-with-keras/>"
- [7] Leafsnap dataset for leaf recognition, Available at <http://leafsnap.com/dataset/>