Fruit Recognition using Deep Learning

Rounak Raj Surana
Computer Science & Information Management
Asian Institute of Technology
Pathumthani, Thailand 12120
st119979@ait.asia

***Abstract* - Automation in food processing comes into play to increase productivity, quality and profitable growth of the country. Traditional fruit classification methods have often relied on manual operations based on visual ability and such methods are tedious, time consuming and inconsistent. External shape appearance is the main source for fruit classification. In recent years, image processing, machine learning, deep learning techniques have been found increasingly useful in the fruit industry, especially for applications in quality inspection and color, size, shape sorting and many more. In this document we are gonna see classification of varied variety of fruits using one of the deep learning concept CNN and transfer learning with VGG16 and CNN architecture.**

Index Terms – Deep learning, CNN, Transfer Learning, VGG16 .

I. INTRODUCTION

Your Classification is a systematic arrangement of groups and categories based on the features so formed from the images. Image classification came into picture to fill the gap between computer vision and human vision by training computer with data. The image classification is achieved by differentiation the image into the prescribed category based on the content of vision.

The demand of fruit classification has increased due to various varieties of fruits coming up in the market and that to in many large quantities which again have to be distributed to various retail shops. Doing it manual would be a very tedious repetitive job to classify tons of fruits in a short time. Hence, a vision based fruit classification which could also assists in packaging.Here, we explore fruit recognition/classification using deep learning. As machine learning consists of feature extraction module that extracts the important features such as edges, textures etc and a classification module that classify based on the features extracted. The main limitation of machine learning is, while separating, it can only extract certain set of features on images and unable to extract differentiating features from the training set of data. Deep learning (DL) is a sub field to machine learning, capable of learning through its own method of computing and exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation and for pattern analysis and classification. In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. The working and the results obtained are described further.

II. RELATED WORK

In this section we briefly review the architectures and concepts used for fruit classification are: Convolutional Neural Network (CNN), Transfer learning and Visual Geometry Group (VGG) Net.

A. Convolutional neural networks

Convolutional neural networks. Sounds like a weird combination of biology and math with a little CS sprinkled in, but these networks have been some of the most influential innovations in the field of computer vision.

Convolutional Neural Networks have a different architecture than regular Neural Networks. Regular Neural Networks transform an input by putting it through a series of hidden layers. Every layer is made up of a set of neurons, where each layer is fully connected to all neurons in the layer before. Finally, there is a last fully-connected layer—the output layer—that represent the predictions.

CNN are a bit different. First of all, the layers are organised in 3 dimensions: width, height and depth. Further, the neurons in one layer do not connect to all the neurons in the next layer but only to a small region of it. Lastly, the final output will be reduced to a single vector of probability scores, organized along the depth dimension.

CNNs have two components:

The Hidden layers/Feature extraction part In this part, the network will perform a series of convolutions and pooling operations during which the features are detected and next the Classification part Here, the fully connected layers will serve as a classifier on top of these extracted features.

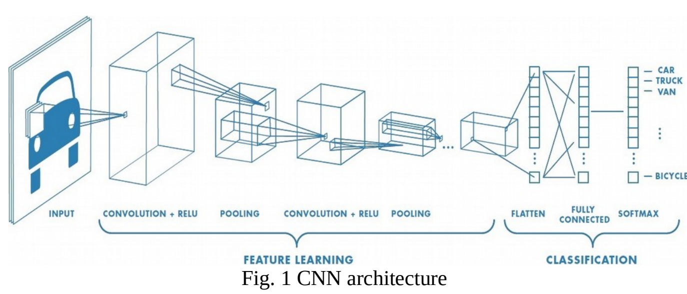


Fig. 1 CNN architecture
CNN architecture with showing feature extraction and classification components

B. Transfer Learning

Conventional machine learning and deep learning algorithms, so far, have been traditionally designed to work in isolation. These algorithms are trained to solve specific tasks. The models have to be rebuilt from scratch once the feature-space distribution changes. Transfer learning is the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related ones. Traditional learning is isolated and occurs purely based on specific tasks, datasets and training separate isolated models on them. No knowledge is retained which can be transferred from one model to another. In transfer learning, you can leverage knowledge (features, weights etc) from previously trained models for training newer models and even tackle problems like having less data for the newer task.

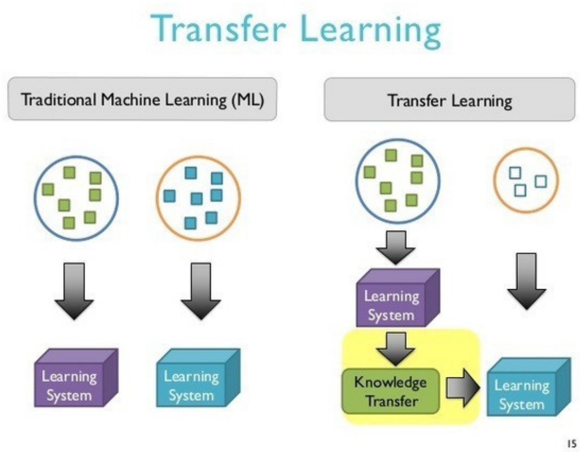


Fig. 2 Traditional learning vs Transfer learning

For many complex problems the training time and the amount of data required for such deep learning systems are much more than that of traditional ML systems.

These pre-trained networks/models form the basis of transfer learning in the context of deep learning, or what I like to call 'deep transfer learning'.
One of the fundamental requirements for transfer learning is the presence of models that perform well on source tasks which gets into picture the pre-trained models. Pre-trained models are usually shared in the form of the millions of parameters/weights the model achieved while being trained to a stable state. Pre-trained models are available for everyone to use through different means. The famous deep learning Python library, keras, provides an interface to download some popular models. You can also access pre-trained models from the web since most of them have been open-sourced. As such of i have used only VGG16 for the classification using pre-trained model.

C. VGGNet

VGGNet consists of 16 convolutional layers and is very appealing because of its very uniform architecture. The most important is that there are many other models built on top of VGGNet or based on the 3x3 convolutions, but lots of filters. The most preferred choice in the community for extracting features from images. The weight configuration of the VGGNet is publicly available and has been used in many other applications and challenges as a baseline feature extractor. However, VGGNet consists of 138 million or more parameters making it to be hard to handle.

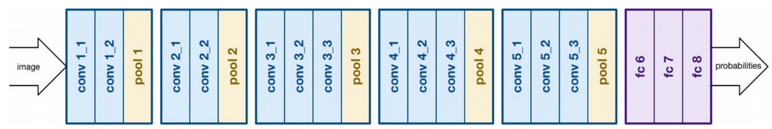


Fig.3 VGG16 Architecture

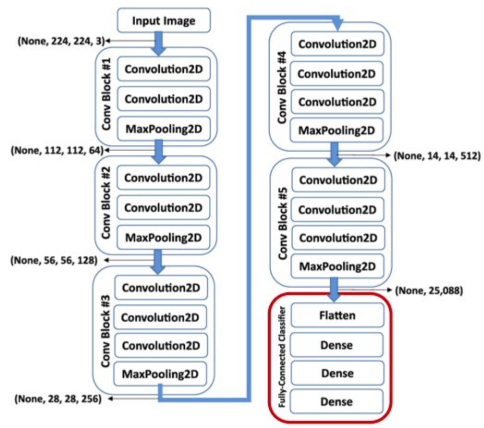


Fig. 4 Semantic Representation of VGG16

III. Approach

In this section we just formulate the problem of fruit classification using CNN and pre-trained network VGG16 and then we are gonna see the training methodology of our proposed network.

A. Problem Definition:

To give a basic idea of what my project deals with is fruit detection and classification but more of fruit classification that is to recognize the given fruit lies in which type of fruit class as they have many classes and each class is unique in it own way, few are similar too as such of citric fruits (lemon, orange,etc). Classifying them is the key role of my project using CNN architecture and using the concept of transfer learning as seen earlier with the help of VGG16 architecture.

IV. Experiments

I conducted experiments to evaluate the proposed model of fruit detection and classification using CNN architecture and transfer learning method on the fruit images given. Further we get to the datasets used implementation details and results obtained from the models trained.

A. Implementation Details

In this section, we show the implementaion details of our approach using CNN, VGG16 and implementation for detection using Faster R-CNN.

Convolutional Neural Network

Dataset used for training consists of a 14 {Apple, Avocado, Banana, Cherry, Guava, Kiwi, Lemon, Lychee, Mango, Mulberry, papaya, Pineapple, Pomegranate, Strawberry} classes of fruits with total of 3813 images with 20% used for validation of the dataset that is dividing it we get 3051 images for training of the model and rest 762 for validation of the model. Here, the model used is CNN with 4 convolutional layers, 4 max-pooling layers for feature extraction and then we have 1 flatten layer and 2 fully connected layers for classification of the fruits based on the class of the fruit so given. So, after training the dataset training accuracy seems to be 98% as the resolution of the pictures where quite high. Whereas the validation accuracy stays back at 41% and the loss to be 4.021 by the end of 100 epochs.

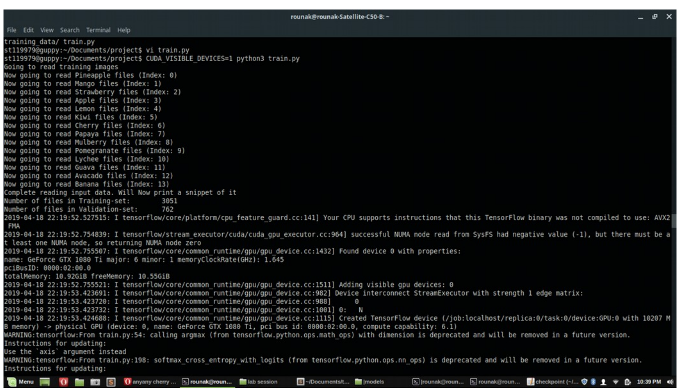


Fig. 5 Screenshot of the training using CNN

Accuracy level after testing of the data goes to 98%.

VGG16

As we have understood what transfer learning is in the earlier chapters of the documentation and as also understood that VGG16 is a pre-trained model which could be used for learning. So, for my project I did use pretrained model of imagenet which consists of weights of around 58889256 images and many different classes as we have seen the limitations of VGG16 that the layers are quite small but the data processing takes a longer time which made me decrease my iterations from 1000 to 100 and try running the training data, testing data and validation data on only 10 different classes as my system with i3 processor and intel graphics card did hand after half of the interations which did waste most of the time. So, better would be to use a GPU based good processing system to run the keras program for transfer learning.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0

Total params: 14,714,688
Trainable params: 14,714,688
Non-trainable params: 0

Fig. 6 Screenshot of the VGG16 architecture after training imagenet dataset from the weights so downloaded

Faster R-CNN

I tried implementing faster R-CNN for my project as of to detect the fruit and tehn apply cnn for classification but as the data did require annotaion and annotaions on 3000 images would take a loads of time which made me divide the data to 750 and with only 5 classes {Apple, banana, mango, lychee, strawberry}and did try to train the data but it took me 2 long days to run 20 epochs with batch size of 32 rather leading to no output. That is after all the wait the model which was created after training using Faster RCNN was nil it couldn’t detect any object. Which made me decrease my project topic to only classification. Might be a GPU could have helped for training for more longer epochs.

YOLO

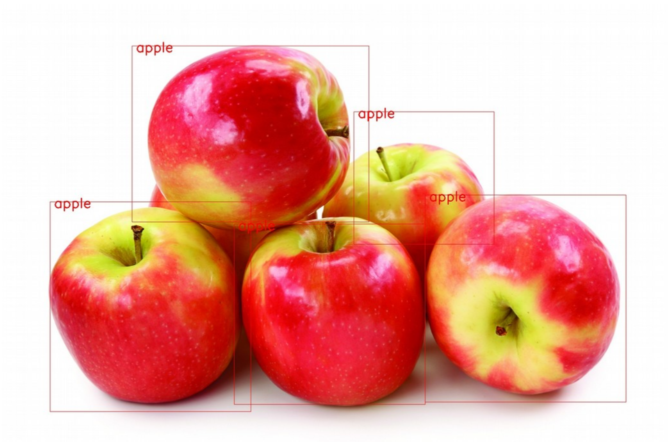
You Only Look Once was my hope to continue with for detection as it was the fastest way possible for training in a very less time but as the annotaions which I did choose where not for real time which would rather make me loose my accuracy. So, I couldn’t even use YOLO for my fruit detection.

V. Conclusion

CNN gave a better result for classification compared to any other architecture on which my system could help me work on and it could classify most of the fruit classes with a higher accuracy rate.

VI. Future Work

Future work would be rather than not only classification but as well as detection for the fruit in a image given so an example of image detection and classification is shown below.



References

SIMPLE-INTRODUCTION-TO-CONVOLUTIONAL-NEURAL-NETWORKS

IMAGE CLASSIFICATION USING DEEP LEARNING

A STUDY ON IMAGE PROCESSING METHODS FOR FRUIT CLASSIFICATION

FRUIT CLASSIFICATION SYSTEM USING COMPUTER VISION: A REVIEW