Classification and Recognition of Vegetable Images Using CNN and Transfer Learning

Final Project
Deep Learning(CS665)

Students
Dipak Dulal,Shusma Kafle,Pooja Rajeev Maheria
Padma Priya Poka,Priyanka Bhauji Wankhede
Professor
Dr.Chengcui Zhang
Department of Computer Science
UAB

April 21, 2022

Abstract

Image classification and recognition using deep learning modules has become the popular in the field of Data Science and Artificial Intelligence. Some years back,binary classification were prevalent but now a days multi-class classification became more popular. Here, we have used deep **Convolution Neural Network(CNN)** and pre-trained transfer learning models: **ResNet152V2,VGG19,InceptionV3** to classify the vegetables images into 15 different vegetable classes. We have tuned different hyper-parameters so as to address over-fitting as well as under-fitting and get better the training,validation and test accuracy.

Key Words: CNN,ResNet152V2,VGG19,InceptionV3

Introduction

Vegetables are the most commonly consumed food all around the world. People are using hundred of species of vegetables but we can get some limited numbers of vegetables are in common use.Vegetables are very important sources of nutrients and minerals for human-beings as well.Though varieties of vegetables are distinguished as per its color size and shape,some vegetables looks same due to similarity in some features. Our project is focused on detecting and recognizing the vegetable image by training 15 thousand vegetable images of 15 different types(class) of labeled images using Convolution Neural Network(CNN) and to get more accurate result we have used Transfer Learning approach also.Specifically, We used CNN model tuning some hyper-parameters and different optimizer and RenNet152V2,InceptionV3 and VGG19 as pre-trained CNN modules.Applying transfer learning techniques, learned features of above DCNN models with very deep network architecture has become pretty effective for our dataset.

Our Model

Convolution Neural Network(CNN)

A Convolutional Neural Network (CNN)[1] is a self-learning network that classifies images similar to how our human brain learns, by observing images of different classes.

CNN basically we start with an input layer and end with an out layer. In the middle we have a lot of filters and processing methods (those are hidden layers) to get the content(feature) of an image. In the hidden layers we will do variety of things like Convolution ((Conv2D), Pooling(Max Pooling and Average Pooling), Activation(ReLU, Sigmoid and Tanh), Fully connected layer, Regularization,Batch normalization and Softmax.

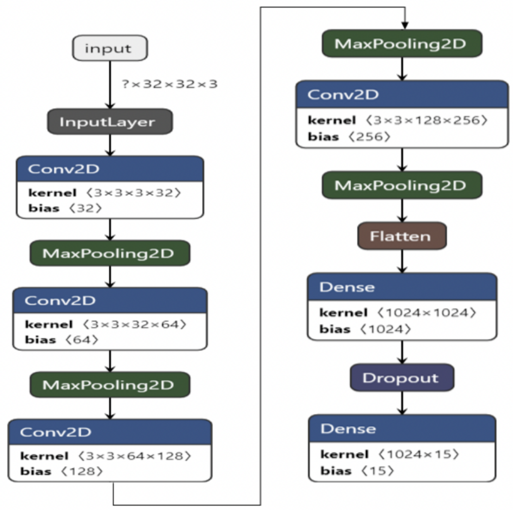

Figure 1: CNN Architecture

In CNN the input layer takes the target image data as input.Then the image is reshaped to an optimal size and forwarded to the convolutional layer. It consists of multiplying a section of the image with a kernel (filter) to produce the exact features of the image as output.Here for this dataset we used RELU and Softmax as activation function.features from the convolutional layer sends to pooling layer(keep only important features from the large image by reducing the parameter(Dropout)),here we used Max Pooling. Features are sent to Fully connected layers(also known as dense layers),connect every connected neuron(features) in the current layer to every connected neuron in the previous layer by applying weight and bias to them.We used some augmentation techniques such as rotation range to extract the important features effectively.

Transfer Learning

Transfer Learning is a Machine Learning techniques that helps in improving performance of target learners within the target domain by using the knowledge already gained for other purpose but is relevant to our project.Now a days, this is very popular techniques in deep learning fields such as image detection and classification task. Traditional ML approach is isolated and knowledge is not retained or accumulated from others whereas Transfer Learning is benefited from previously learned tasks and it is more faster and more accurate.There are numerous Pre-trained models, we used here specifically

- ResNet152V2
- InceptionV3
- VGG19

ResNet152V2

RestNet152V2 is the pretrained model with the upgraded version of the previous Resnet model with 152 layers. It has more layers that previous ResNet models.It uses batch normalization before each weight layer which result in better performance compared to the previous.

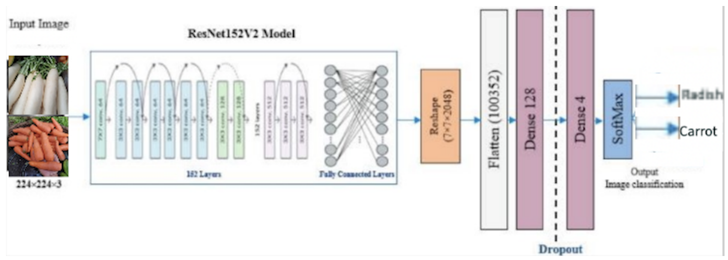

Figure 2: ResNet152V2 Architecture

InceptionV3

[2]It is a modified version of the inception architecture which is more computationally effective with lesser parameter and lower error rate. It consists of 42 layers. Several optimizing techniques were used to make the version more effective which are as follows:

> two 5x5 convolution into two 3x3 will reduce the parameters resulting in effective computational cost.Spatial Factorization into Asymmetric Convolutions Replacing the 3X3 convolution with two 1X3 convolution will increase the computation efficacy by 33%. Effective grid size reduction Max pooling and average pooling were used to reduce the grid size of the feature maps on previous version. In the inception V3 model, to reduce the grid size efficiently, the activation dimension of the network filters is expanded.

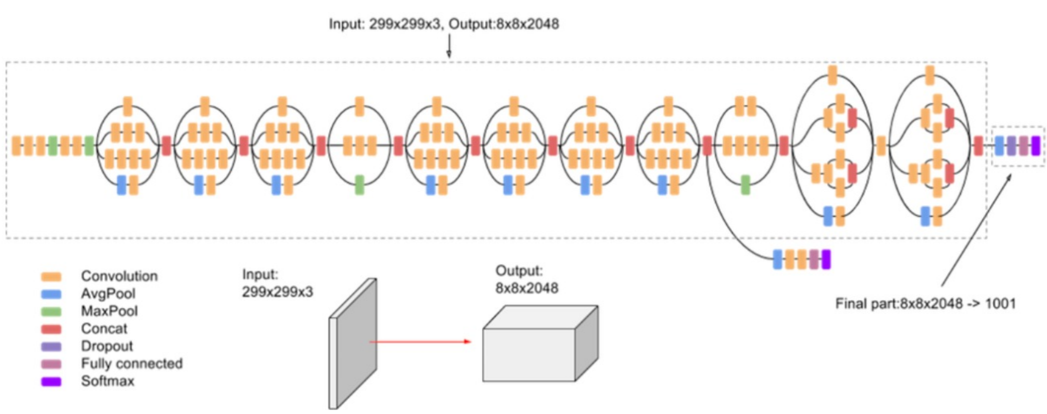

Figure 3: InceptionV3 Architecture[2]

-

VGG19

Vgg19 is the modified version of vgg16 which consists of 19 convolutional layers compare to 16 on the previous version. Those added layers helps us to extract more features from the images. It doesn't necessarily mean the just increasing the convolution layer with better the performance because the weight of the neural networks is updated through the backpropagation so any changes can create the loss of the accuracy of the model.

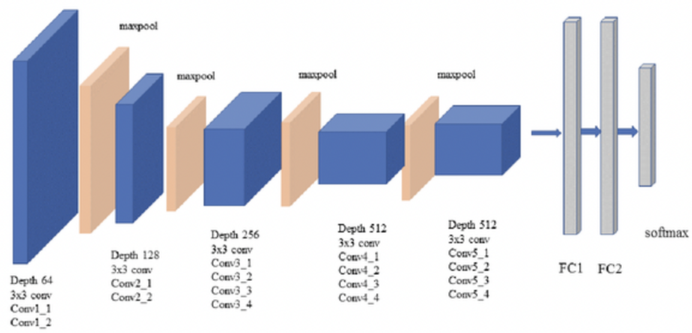

Figure 4: VGG19 Architecture [6]

Data Processing and Visualization

Data Structure

We have downloaded the data from Kaggle data repository [4][5]. Our images data consists of 15000 labeled training images of 15 categories:bean, bitter gourd, bottle gourd, brinjal, broccoli, cabbage, capsicum, carrot, cauliflower, cucumber, papaya, potato, pumpkin, radish and tomato.Each catetory consist of 1000 images for training purpose. There are 3000 images(200 in each category) for validation and 300 images for test purpose.

Figure 5: Some Random Images of Vegetables from Training Data Set

Data Generation

For the efficacy, robustness and time friendlier, we used different data generation technique. First, we rescaled the data values into target values from 0 to 1.. we adjusted the target size as

(224,224). We opened the shuffle, activated the horizontal shift by 0.15.Similarly, we set the rotation range as 30, zoom range = 0.15,width shift range 0.2,shear range, 0.15 and fill mode, ”nearest”.

Evaluation

Training,Testing and Validation Accuracy Comparison

S.N	Model	Optimizer	Train_accurancy (%)	Val_accurancy (%)	Test_accurancy(%)	Epochs	Time (Minutes)
1	CNN	RMSprop	98.71	92.3	92.4	20	49.44
2	CNN	Adam	98.44	96.23	96.6	20	181.08
3	CNN	SGD	98.92	95.17	95.83	20	96.52
4	ResNet152v2	RMSprop	99.93	99.57	99.53	6	63.22
5	InceptionV3	RMSprop	99.93	99.57	99.57	11	30.65
6	VGG19	RMSprop	99.91	99.61	99.57	6	62.1

Figure 6: Accuracy Table

Comparison on Confusion Matrices

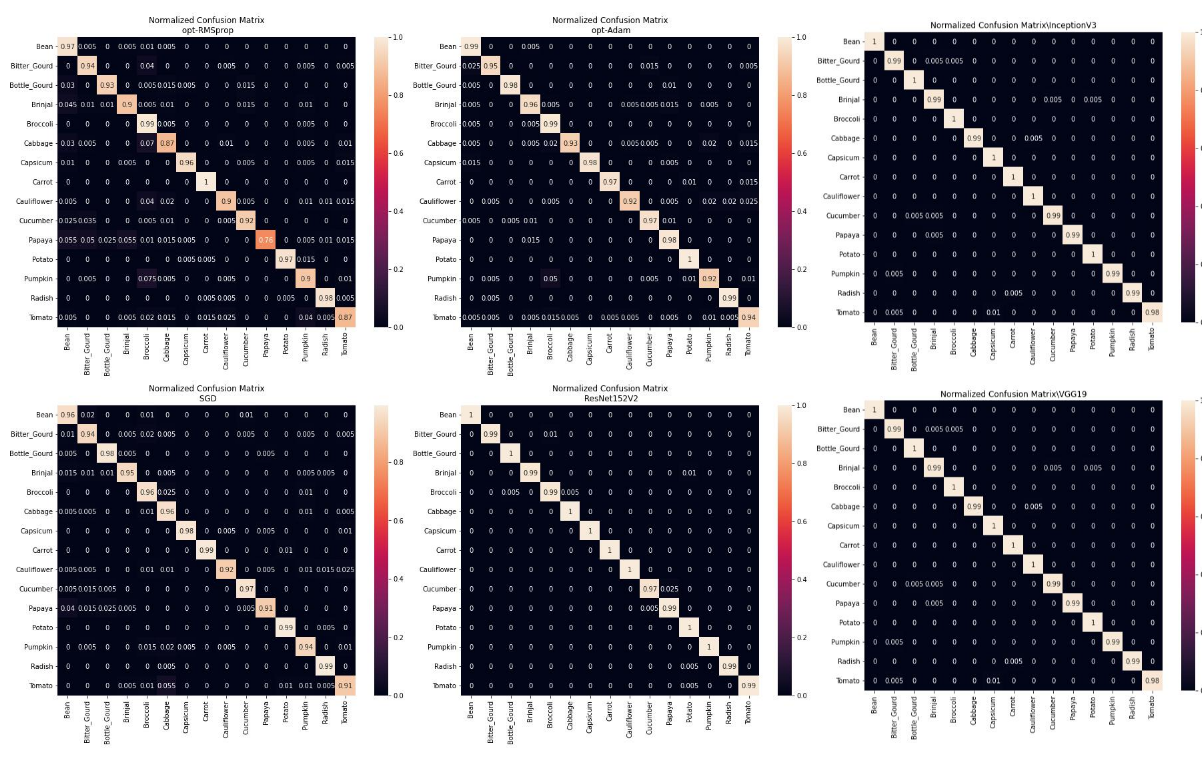

Figure 7: Confusion Matrices

Accuracy and Loss Propagation

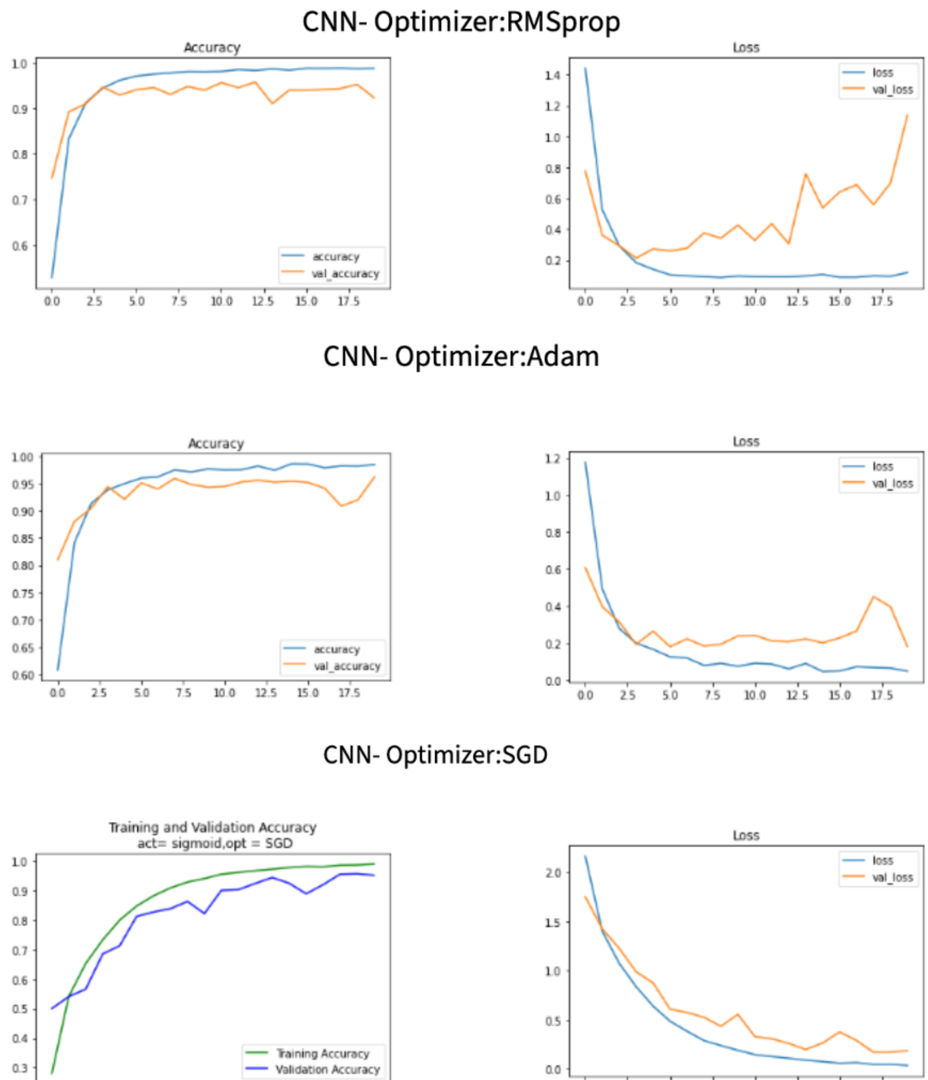

Figure 8: Loss and Accuracy Values Propagation(CNN)

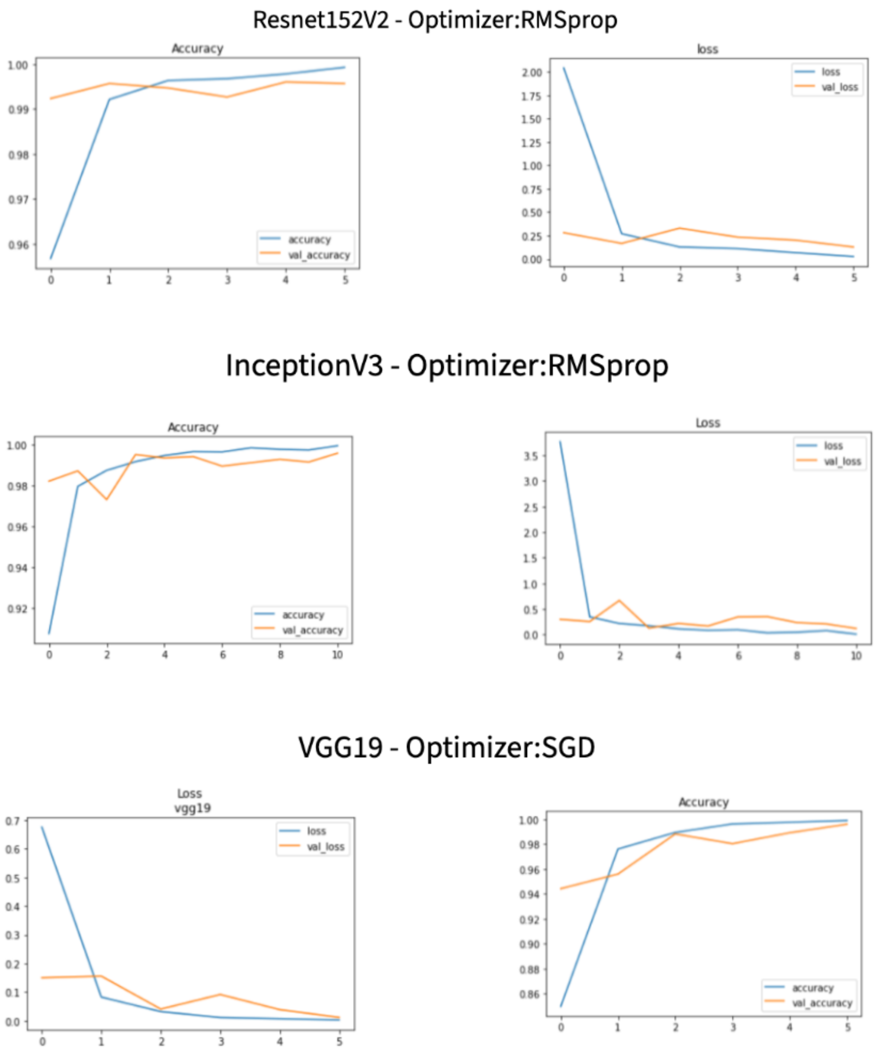

Figure 9: Loss and Accuracy Values Propagation(TF)

Random Test for Image Prediction from Test Data

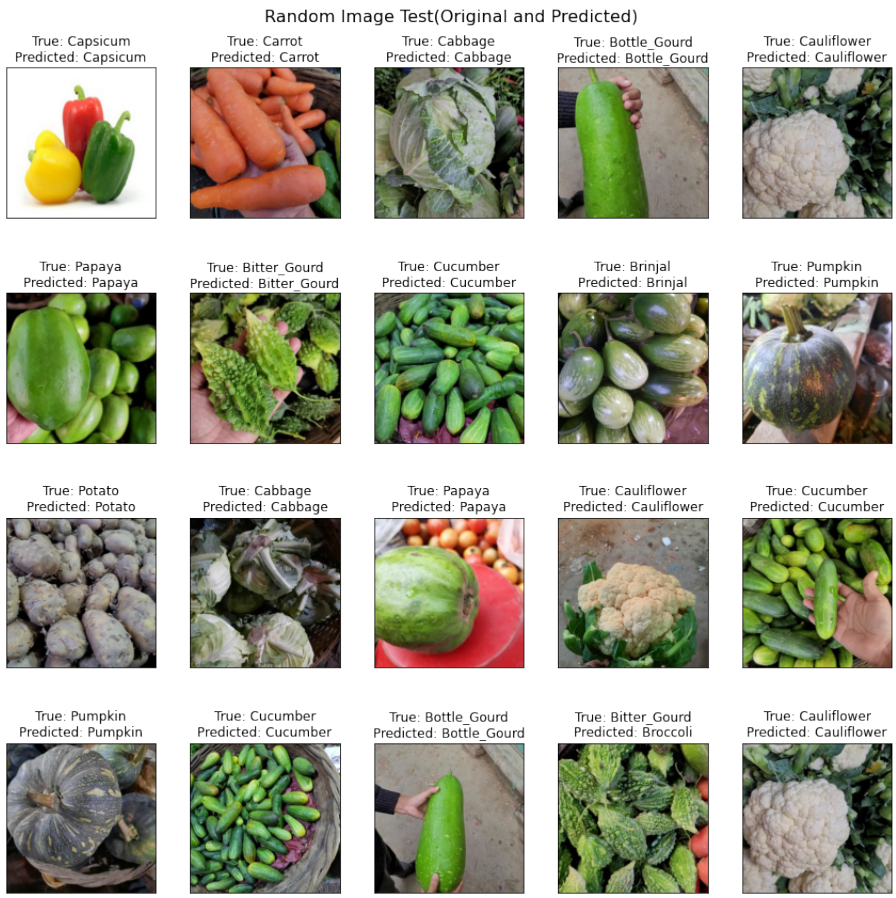

Figure 10: Original and Predicted Label for Random Test Images(CNN:RMSProp)

Figure 11: Original and Predicted Label for Random Test Images(INceptionV3)

Conclusion

In conclusion, this project on image classification using deep learning via TensorFlow framework is quite effective on comparison of different CNN models with some transfer learning techniques. On the basis of the result we have obtained, with fine tuning hyper-parameters, augmentation and other data generation techniques on our vegetable image dataset,the model using InceptionV3 gives the most robust and effective and highly accurate result.Testing randomly 20 images from test dataset, we found at least one error prediction on some random selection using CNN model(without Pretrained Model), whereas it is quite hard to find error prediction while using previously trained transfer learning models .

Future Work

We successfully executed our model to get high accuracy using the labeled fine images of equal sizes with high accuracy. In future we can test the raw data of different sizes and expanding the size of class i.e.increasing the categories of vegetables also.In addition, we can include vegetables, fruits and cereals images together and apply our models to train large data set with big size of class and explore for the high accuracy.

References

[1] Ian Goodfellow,Yoshua Benhio and Aaron Courvile, Deep Learning

[2] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna,Rethinking the Inception Architecture for Computer Vision

[3] Lecture Slides of Deep Learning CS665,Prof. Dr. Chengcui Jhang

[4] Kaggle Data
https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset

[5] M. I. Ahmed, S. Mahmud Mamun and A. U. Zaman Asif, "DCNN-Based Vegetable Image Classification Using Transfer Learning: A Comparative Study," 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), 2021, pp. 235-243, doi: 10.1109/ICCCSP52374.2021.9465499.

[6] https://blog.techcraft.org/vgg-19-convolutional-neural-network/

[7] Wikipedia