New Image Classification using Convolutional Neural Network (CNN) by Deep Learning

C.UmaRani L.Ganesan*
Department of CSE, A.C.Govt.College of Engg & Tech, Karaikudi-3, Tamilnadu, India.
** Corresponding author: drcumakkd@gmail.com*

Abstract

Deep learning technologies are becoming the major approaches for natural signal and information processing, like image classification, speech recognition etc, which is inspired by the functioning of human brain. In deep learning networks of artificial neurons analyze large dataset automatically into images, sound, video and text. Convolutional Neural Network(CNN) become very popular for image classification in deep learning. CNN perform better than human subjects on many of the image classification datasets. CNN learns to perform classification tasks directly from images, video, text or sound. It is used for finding patterns in image, recognize and classify the object. It directly learns from the Data using the pattern to classify the images. In this paper, a deep learning convolutional neural network based on keras and tensorflow is deployed using Python for image classification.

CNN adapts two types of classification approach, namely, single class classification and multi class classification. We have experimented both the methods and find their accuracies. Single class classification is experimented with using Apple and Banana images. Multi class classification done by a large number of different images, which contains types of Apple, Banana, Cactus, Avocado, Cherry. In image classification Relu activation function gives higher classification accuracy than other classifier and activation function. By using CNN, given input images are be trained first and using trained images to predict result of new classified images i.e testing. The results obtained are acceptable and comparable with existing approaches reported in literature.

Keywords*: Deep Learning, Convolutional Neural Network(CNN), Tensorflow, Keras, Relu, Opencv, Convolution, Maxpooling.*

1.Introduction

Deep learning is a technology inspired by the functioning of human brain. In deep learning, networks of artificial neurons analyze large dataset to automatically, discover underlying patterns, without human intervention. In deep learning, a system learns to classify images, text and sound. The system is trained with large image datasets and wherein it changes the pixel value of the picture to an internal representation. The classifier can detect patterns on the input image[1]. To increase performance, the application of neural networks to learning tasks contains more than one hidden layers. Deep learning is part of a broader family of machine learning methods, based on learning data representation, as opposed to hard code machine algorithms. One of the most frequently used deep learning method for image classification is the convolutional neural network (CNN). Common problem in image classification using deep learning is low performance because of over fitting. CNN have fewer connection and hyper parameter that make CNN model easy to train and perform slightly better than other models. In this paper, a deep learning convolutional neural network based on keras and tensorflow is deployed using Python for image classification [7].

The conventional methods used for image classification is part and piece of the field of artificial intelligence (AI) formally called as machine learning. The machine learning extracts the important features such as edges, textures etc and a classification module that classify based on the features extracted[2]. 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. This disadvantage is rectified by using the deep learning[6]. Deep learning (DL) is a sub field to the machine learning, capable of learning through its own method of computing. A deep learning model is introduced to persistently break down information with a homogeneous structure like how a human would make determinations, to accomplish this, deep learning

In the area of image recognition and classification, the most successful results were obtained using artificial neural networks. These networks form the basis for most deep learning models. Deep learning is a class of machine learning algorithms that use multiple layers that contain nonlinear processing units . Each level learns to transform its input data into a slightly more abstract and composite representation . Deep neural networks have managed to outperform other machine learning algorithms. They also achieved the first super human pattern recognition in certain domains. Deep neural networks-specifically convolutional neural networks –have been proved to obtain good results in the field of image recognition.It is a network capable of learning unsupervised data that is unstructured (or) unlabeled.

The organization of the paper is as follows, section 1 gives an introduction, section 2 provides brief summary on CNN. Section 3 illustrates the architecture of the proposed CNN and section 4 describes the experimentation and results and section 5 provides conclusion and further scope of our work.

2.Convolutional Neural Network

Machine Learning algorithms like KNN, SVM, logistic regression etc.., learn less while comparing to CNN, because there is no transfer in learning occur and hence less error will occur. People used to create features from images and then feed those features into a classification algorithm like SVM. Some algorithm also used the pixel intensities values of images as a feature vector.

Image classification can be accomplished by any machine learning algorithms (logistic regression, random forest and SVM). But all the machine learning algorithms require proper features for doing the classification. If you feed the raw image into the classifier, it will fail to classify the images properly and the accuracy of the classifier would be less. Moreover, it is voluminous and time complexity is high.

In CNN architecture, the beginning layers are extracting the low-level features and end level layers extract high-level features from the image. There are so many handcrafted features available(local feature, global feature), but it will take so much time to select the proper features for a solution(image classification) and selecting the proper classification model. CNN handles all these problems. Figure 1. Shows that the diagram for Convolutional Neural Network.

Convolutional layers are named after the convolution operation handled in CNN. In mathematics convolution is an operation on two functions that produces a third function that is the modified (convoluted) version of one of the original functions. The resulting function gives in integral the proposed method which is more suitable for classification.

3. Proposed System

A 100 x 100 image has 10000 pixels and if the first layer has 100 neurons, it would result in 1000000 parameters. Instead of each neuron having weights for the full dimension of the input , a neuron holds weights for the dimension of the kernel input. The kernels slide across the width and height of the input, extract high level features and produce a two dimensional activation map. The stride at which a kernel slides is given as a parameter. The output of a convolutional layer is made by stacking the resulted activation maps, which in turn is used to define the input of the next layer.

Applying a convolutional layer over an image of size 32 X 32 results in an activation map of size 28 X 28 . If apply more convolutional layers, the size will be further reduced , and , as a result the image size is drastically reduced which produces loss of useful information and the vanishing gradient problem. To correct this, padding may be used. Padding increases the size of a input data by filling constants around input data. In most of the cases, this constant is zero so the operation is named zero padding. ”Same” padding means that the output feature map has the same spatial dimensions as the input feature map.

The strides causes a kernel to skip over pixels in an image and not include them in the output. The strides determine how a convolution operation works with a kernel when a larger image and more complex kernel are used. As a kernel is sliding the input, it is using the strides parameter to determine how many positions to skip. ReLUlayer, or Rectified Linear Units layer, applies the activation function max(0, x). It does not reduce the size of the network, but it increases its nonlinear properties.

CNNs can be thought of automatic feature extractors from the image. This algorithm use with pixel vector and a lot of spatial interaction between pixels are lost, a CNN effectively uses adjacent pixel information to effectively down sample the image first by convolution and then uses a prediction layer at the end.

3.1 Pooling Layer

Pooling layers are used on one hand to reduce the spatial dimensions of the representation and to reduce the amount of computation done in the network. The other use of pooling layers is to control over fitting. The most used pooling layer has filters of size 2 x 2 with a stride 2. This effectively reduces the input to a quarter of its original size.

3.2 Fully Connected Layer

Fully connected layers are layers from a regular neural network. Each neuron from a fully connected layer is linked to each output of the previous layer. The operations behind a convolutional layer are the same as in a fully connected layer. Thus, it is possible to convert between the two**.**

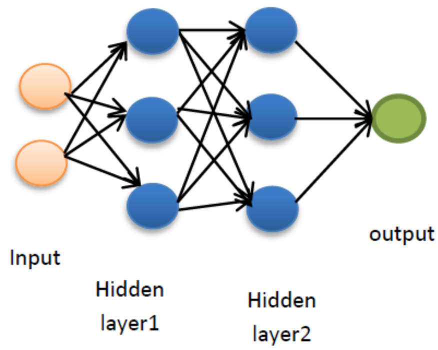


Figure 1. Convolutional Neural Network

In first step image dataset is prepared, there are 3 files in dataset, which contains 1000 images of Apple and banana and avocado, cactus, cherry , where 800 images used for training and 200 images used for testing purpose. In second step, parameters are defined for image classification. In third step, CNN is created with two convolutional layers, then select different combination of activation functions and classifiers for comparison purpose. In next step, the created CNN is filtered to image dataset and Train, Test the system with training and test datasets respectively. Finally, the accuracy for different CNN structures are obtained and these accuracies are compared for performance measurement. The corresponding flow diagram is shown in Figure 2.

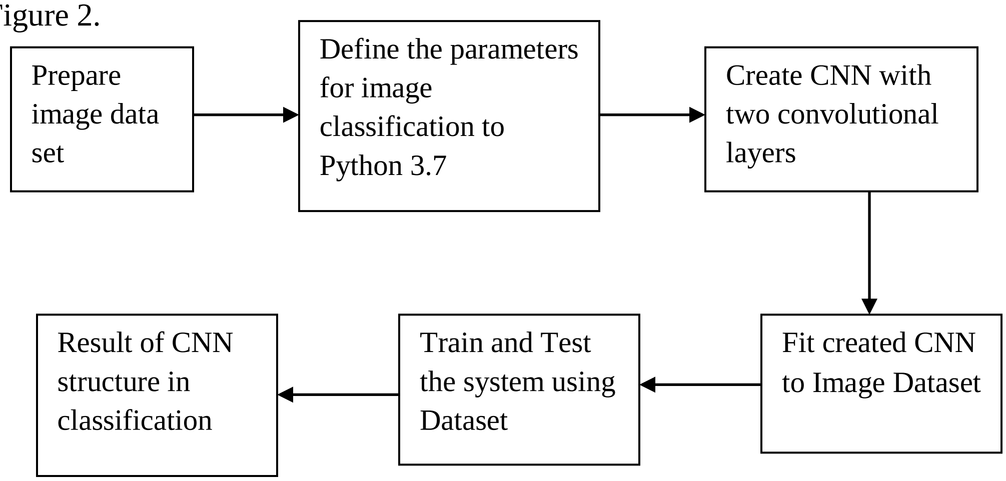


Figure.2. Flow Diagram

Convolutional Neural Network a of feed forward artificial neural network, is inspired by visual cortex. In CNN, the neuron in a layer is only connected to a small region of the layer, instead of all the neurons in a fully connected manner. As a result, CNN handle fewer amounts of weights and also less number of neurons. The CNN model is shown in Figure 3.

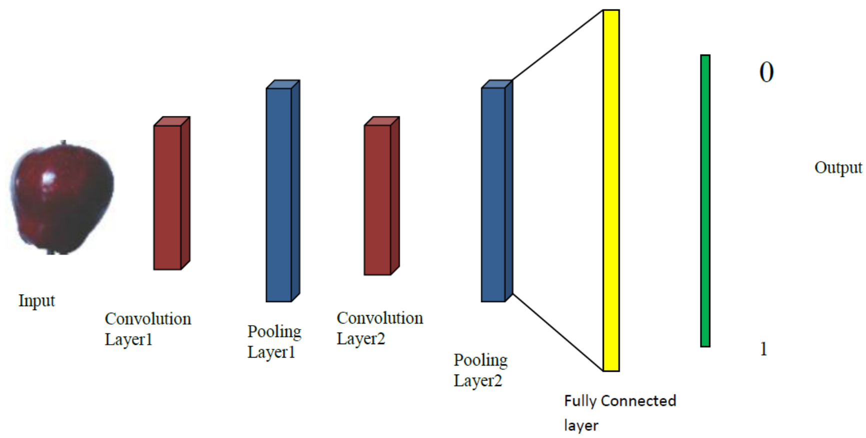


Figure.3. Convolutional Neural Network(CNN) Model

3.3 Architecture Diagram

As previously described the convolutional neural network makes use of convolutional layers, pooling layers, ReLUlayers, fully connected layers and loss layers. In a typical CNN architecture, each convolutional layer is followed by a Rectified Linear Unit (ReLU)layer , then a Pooling layer the none or more convolutional layer and finally one or more fully connected layer. A regular neural network converts the input in a one dimensional array which makes the trained classifier less sensitive to positional changes. The input that used consists of standard RGB images of size 100x100 pixels. The neural network that is used in this work has the structure given in Table 1.

Table 1: Structure of neural network

Layer type	Dimensions	Output
Convolutional	5 X 5X 32	32
Maxpooling	2 X 2 – Stride : 2	--
Convolutional	3 X3X 96	96
Maxpooling	2 X 2 – Stride : 2	--
Convolutional	3 X 3X 128	123
Maxpooling	2 X 2 – Stride : 2	--
Convolutional	3 X 3X 128	128
Maxpooling	2 X 2 – Stride : 2	--
Fully connected	3 X 3X 128	1024
Fully connected	1024	192
Softmax	192	15

A visual representation of the neural network used is given in Figure 4,

The first layer (Convolution#1)is a convolutional layer which applies 32, 5 x 5 filters. On this layer we apply max pooling with a filter of shape 2 x 2 with stride 2 which specifies that the pooled regions do not overlap (Max-Pool #1). This also reduces the width and height to 50 % each.

The second convolutional (Convolution #2) layer applies 96, 5 x 5 filters which outputs 32 activation maps. We apply on this layer the same kind of max pooling(Max-Pool #2) as on the first layer, shape 2 x 2 and stride 2.

The third convolutional (Convolution#3) layer applies 128, 5x5 filters. Following is another maxpool layer (Max-Pool#3) of shape 2x2 and stride 2.

The fourth convolutional (Convolution #4) layer applies 128, 5 x 5 filters after which we apply a final max pool layer (Max-Pool #4).

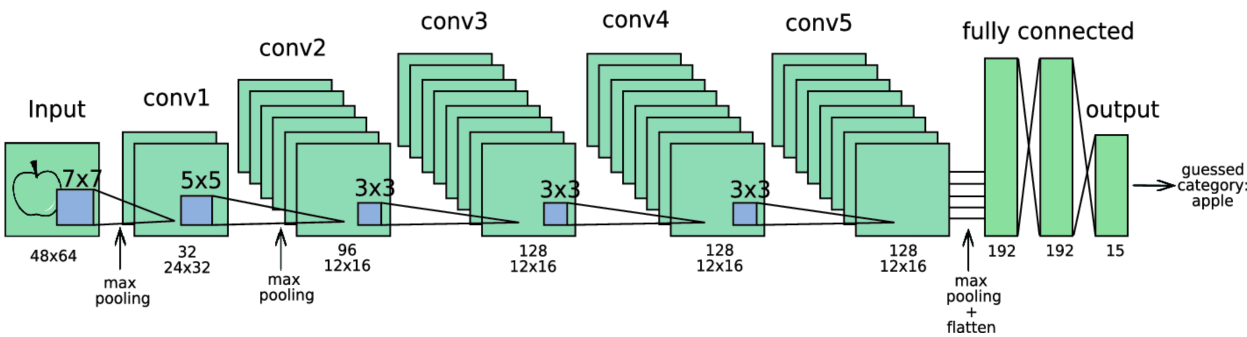


Figure.4. Visual Representation of Neural Network

The following three ways, based on which the images are classified in CNN.

- Size of image
- Color of image
- Dimension of image

4.Experiments and Results

4.1 Image Collection

The input image dataset is split in 2 parts: training set and testing set. This is a binary file that contains protocol buffers with a feature map. In this map it is possible to store information such as the image height, width, depth and even the raw image. Using these files it create queues in order to feed the data to the neural network. By calling the method shufflebatchprovide randomized input to the network. The way used this method was providing it example tensors, for images and labels. It returned tensors of shape batch size x image dimensions and batch size x labels. This helps greatly lower the chance of using the same batch multiple times for training, which in turn improves the quality of the network.

4.2 Conversion

For multiple scenarios in which the neural network was trained using different levels of data augmentation and preprocessing:

- Convert the input RGB images into gray scale
- Keep the input images in the RGB color space itself.
- Convert the input RGB images in to the HSV color space

4.3. Training Images

For each of the above scenario the neural network is trained over 75000 iterations with batches of 60 images selected at random from the training set. Every 50 steps, the accuracy is calculated using cross-validation. For testing, the trained network is executed on the test set. The results for each case are presented in Table 2.

Table.2: Results of training the neural network on the fruits dataset.

Scenario	Accuracy on training set	Accuracy on test set
Grayscale	99.82%	92.65%
RGB	99.82%	94.43%
HSV	99.80%	94.40%
HSV + Grayscale	99.78%	94.74%
HSV + Grayscale +hue / saturation change + flips	99.58%	95.23%

4.4.CNN Accuracy

It is also important to notice that training the grayscale images only yielded the best results on the train set but very weak results on the test set. To investigate this problem and discover that many of images containing apples are incorrectly classified on the test set. In order to further investigate the issue to run a round of training and testing on just the apple classes of images. The results are similar, with high accuracy on the training data, but low accuracy on the test data. We attribute this to over fitting, because the gray scale images lose too many features, the network does not learn properly how to classify the images.

In order to determine the best network configuration for classifying the images in our dataset, multiple configurations are taken, used the train set to train them and then calculated their accuracy on the test and training set. Results of training different network configurations on the fruits dataset is shown in Table 3.

Table 3. : Results of training different network configurations on the fruits dataset.

<table>
<tr><th>S.No.</th><th colspan="3">Configuration</th><th>Accuracy on training set</th><th>Accuracy on test set</th></tr>
<tr><td rowspan="6">1</td><td>Convolutional</td><td>5X5</td><td>16</td><td rowspan="6">99.58%</td><td rowspan="6">95.23%</td></tr>
<tr><td>Convolutional</td><td>5X5</td><td>32</td></tr>
<tr><td>Convolutional</td><td>5X5</td><td>64</td></tr>
<tr><td>Convolutional</td><td>5X5</td><td>128</td></tr>
<tr><td>Fully Connected</td><td>---</td><td>1024</td></tr>
<tr><td>Fully Connected</td><td>---</td><td>256</td></tr>
<tr><td>2</td><td>Convolutional</td><td>5X5</td><td>8</td><td>99.68%</td><td>95.02%</td></tr>
</table>

	Convolutional	5X5	32		
	Convolutional	5X5	64		
	Convolutional	5X5	128		
	Fully Connected	---	1024		
	Fully Connected	---	256		
3	Convolutional	5X5	32	99.24%	94.06%
	Convolutional	5X5	32		
	Convolutional	5X5	64		
	Convolutional	5X5	128		
	Fully Connected	---	1024		
	Fully Connected	---	256		

4.5. Classify incorrect Images

The evolution of accuracy during training It can be seen that the training rapidly improves in the first 1000 iterations (accuracy becomes greater than 90%) and then it is very slowly improved in the next 74000 iterations.

Some of the incorrectly classified images are given in Figure 5.

Apple Golden 2 Apple Golden 3 Braeburn(Apple) Peach

Apple Golden 3 Granny Smith(Apple) Apple Red 2 Apple Red Yellow
96.54% 95.22% 97.71% 97.85%

Figure. 5: Classification and misclassification images

4.6 Image Classification

The 2nd, 4th, and 5th convolutional layers bits are related just to the part maps in the previous layer which dwell on the same Graphics Processing Unit (GPU) said in the figure. The kernels of the 3rd convolutional layer are associated with all kernel maps in the 2nd layer. The neurons in the fully connected layers are associated with all neurons in the past layer.

The 3rd, 4th, and 5th convolutional layers are associated with each other with no interceding pooling or standardization layers. The 3rd convolutional layer has 384 parts of size 3 × 3 × 256 associated with the (standardized, pooled) yields of the 2nd convolutional layer. The fourth convolutional layer has 384 kernels of size 3 × 3 × 192 and the fifth convolutional layer has 256 kernels of size 3 × 3 × 192.

The first two fully connected layers have 4096 neurons each. The local response are used for normalization in the normalization layer. There are two normalization layers present in the

AlexNet architecture. The Deep Neural Network with ReLU Nonlinearity can train very fast than with the identical of the function tanh units. The ReLU considers quicker and more compelling training by mapping the negative esteems to zero and keeping up positive esteems. Signifying by the movement of a neuron figured by applying kernel i at position (x, y) and after that applying the ReLU nonlinearity. Result for multiclass classification(Apple,Avocato) is shown in Figure 6. Result for single class classification is shown in Figure.7.

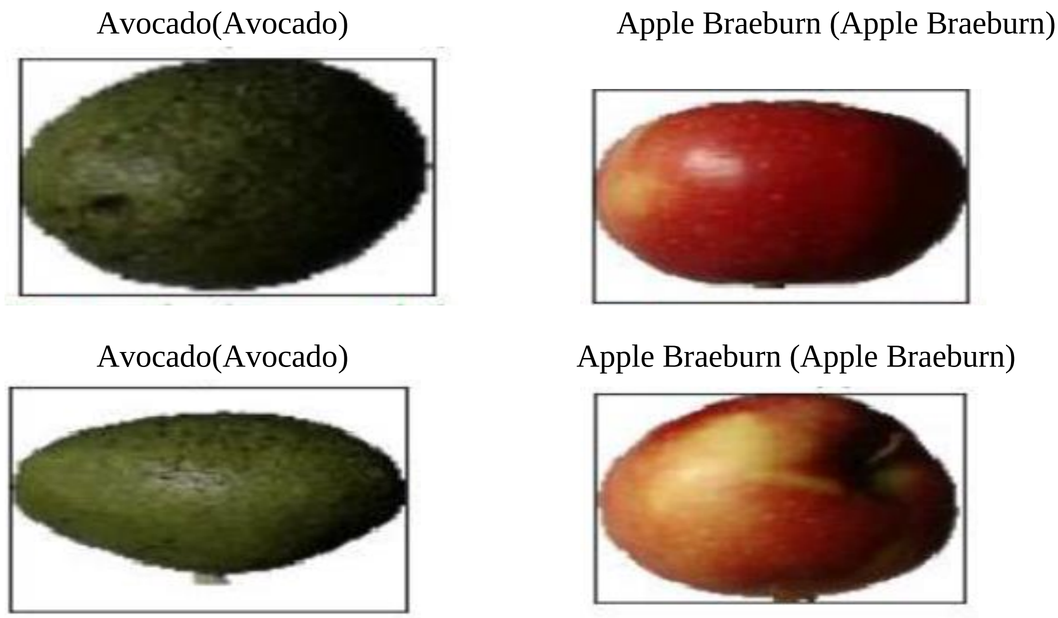


Figure.6 : Result for multiclass classification(Apple, Avocato)

pred: Apple Red Yellow 1,true: 0_100 pred: Apple Red Delicious,true:108_100 pred: Apple Red Yellow 1,true:149_100

Figure.7: Result for single class classification

4.6.1. Collecting Dataset:

In first step to collect Apple, banana, Cactus, Avacato, Cherry images.

a. Training images

To train the images using convolutional neural network(CNN

In CNN using two method 1)Convolution 2)Pooling.

b. Testing images

To test the images using test data set that is used to know the performance of the images.

c. Validate new image

Apply the collection of new images for classification purpose.


19_100jpg 21_100jpg 32_100jpg 23_100jpg 28_100jpg 29_100jpg


75_100jpg 77_100jpg 79_100jpg 157_100jpg 165_100jpg 167_100jpg


106_100jpg 108_100jpg 118_100jpg 106_100jpg 190_100jpg 192_100jpg

Figure.8: Cherry images **Figure.9: Apple images**


4_100-Copy.jpg 4_100.jpg 5_100-Copy.jpg 28_100-.jpg 30_100.jpg 34_100.jpg


21_100.jpg 24_100.jpg 27_100.jpg 107_100.jpg 112_100.jpg 119_100.jpg

84_100.jpg 87_100.jpg 91_100.jpg 163_100.jpg 166_100.jpg 176_100.jpg

Figure.10: Banana images **Figure.11: Cactus images**

Pred: Apple Gravy Smith,true:0_100 Pred: Apple Red Delicious,true:108_100 Pred: Apple Red Yellow,true:149_100

Pred: Apple Red Yellow,true:23_100 Pred: Apple Red Yellow,true:apple_clipart

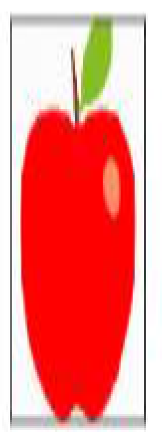
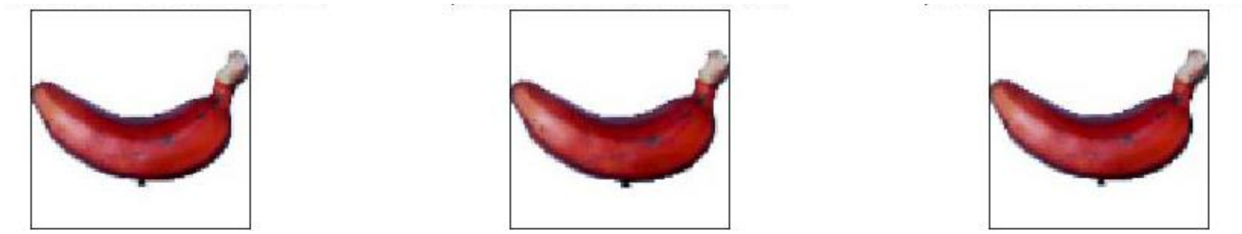
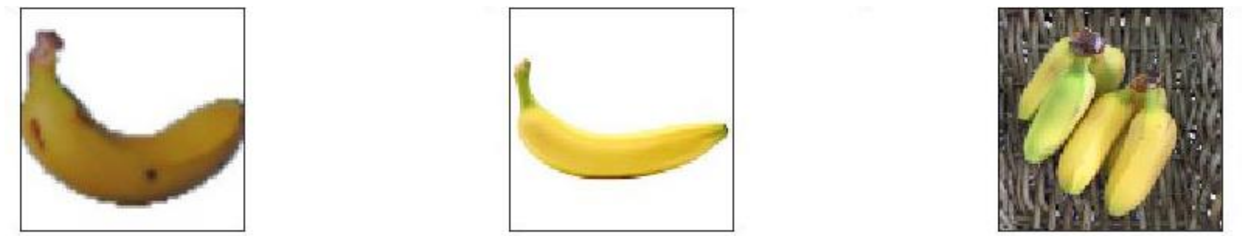


Figure.12. Result for Apple image Classification



Pred:Banana Red,true:Banana Red(2) Pred:Banana Red,true:Banana Red Pred:Banana Red,true:Banana Red3



Pred:Banana,true:Banana Pred:Banana,true:Banana 1 Pred:BananaLadyFingertrue:Banana_lady_finger

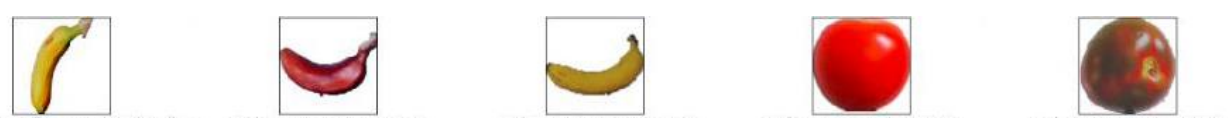
Figure.13. Result for Banana image classification

Pred:AppleGolden3,true:AppleGolden3 Pred:Apple Braebum,true:AppleBrae Pred:AppleCrimsonSnow,true:Applecrimsnow
Pred:Apple Golden1,true:AppleGolden1 Pred:Apple Golden2,true:Apple Golden2

Pred:AppleGrannySmith,true:AppleGranny Pred:Applepinklady,true:ApplepinkladyPred:Applepreddelicious,true:Appledelicious
Pred:AppleRed1,true:AppleRed Pred:AppleRed3,true:AppleRed3

Pred:AppleRedYellow1,true:Appleyellow1 Pred:AppleRedYellow2,true:Appleyellow2 Pred:AppleRed2,true:Applered2
Pred:Avocada ripe,true:AvocadoRipe Pred:Avocada,true:Avocado

Pred:BananaLadyFinger,true:BananaLady Pred:BananaRed,true:bananared Pred:Banana,true:Banana Pred:Tomoto cherryRed,true:Tomotocherry Pred:Tomoto Maroon,true:TomotoMaroon



Pred:TomotoYellow,true:Tomotoyellow Pred:Tomoto1,true:Tomoto1 Pred:Tomoto2,true:Tomoto2 Pred:Tomoto3,true:Tomoto3
Pred:Tomoto4,true:Tomoto4

Figure.14. Result for Multi class classification

5.Conclusions

A new and complex database of images with fruits are used for classification into single class and multiclass. Numerical experiments by using TensorFlow library in order to classify the images according to their content are done. Convolutional Neural Networks (CNN) is used for image classification. The data sets used for both training and testing purpose using CNN. It provides the accuracy rate 98%. The highlighting feature of our experimentation revealed that the accuracy in classification is higher when used deep learning approach than conventional existing approaches reported in literature. The difficult discriminating features are seen in face recognition when images of twins are experimented. Our further scope is to use the same set up for classification in such trivial applications.

References

1. *Bargoti, S., and Underwood, J. "Deep fruit detection in orchards" in 2017 IEEE International Conference on Robotics and Automation (ICRA) ,pp. 3626–3633(May 2017).*
2. *Chan, T. F., and Vese, L. A. "Active contours without edges", IEEE Transactions on Image Processing 10, 2 266–277, (Feb 2001).*
3. *Cheng,H.,Damerow, & etal, "Early yield prediction using image analysis of apple fruit and tree canopy features with neural network", Journal of Imaging 3, 1 (2017).*
4. *Cires D. C., Giusti, A., & etal "Deep neural networks segment neuronal membranes in electron microscopy images" in Proceedings of the 25th International Conference on NeuralInformationProcessingSystems-Volume2 ,NIPS'12, Curran Associates Inc., pp. 2843–2851, (USA, 2012).*
5. *Cires¸an,D.C.,Meier,U,& etal "High performance convolutional neural networks for image classification", in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Two ,IJCAI'11, AAAI Press, pp. 1237–1242,(2011).*
6. *Sa, I., Ge, Z., Dayoub, F, & etal, " Deepfruits: A fruit detection system using deep neural networks" Sensors 16, 8 (2016).*
7. *Schmidhuber, J. Deep learning in neural networks: An overview. CoRR abs/1404.7828 (2014).*
8. *Song, Y., Glasbey, C, & etal "Automatic fruit recognition and counting from multiple images", Biosystems Engineering 118 ,203 – 215(2014).*
9. *Springenberg, J. T., Dosovitskiy, A., & etal "Striving for simplicity: The all convolutional network ",CoRR abs/1412.6806 (2014).*
10. *Srivastava, R. K., Greff, K., & etal, "Training very deep networks.", CoRR abs/1507.06228 (2015).*
11. *Szegedy, C., Liu, W., . & etal," Going deeper with convolutions" , CoRR abs/1409.4842 (2014). TensorFlow. Tensorflow. [Online; accessed 05.05.2018].*
12. *Xiong, J., Liu, Z. & etal,. "Green grape detection and picking-point calculation in a night-time natural environment using a charge-coupled device (ccd) vision sensor with artificial illumination ", Sensors 18, 4 (2018).*
13. *Zawbaa, H., Abbass, M., & etal , "Automatic fruit image recognition system based on shape and color features ", Communications in Computer and Information Science 488 (11 2014).*