ENGINEERING IN ADVANCED
RESEARCHSCIENCE AND
TECHNOLOGY

ISSN 2352-8648
Vol.03, Issue.01
June-2022
Pages: -405-415

IMAGE CLASSIFICATION USING ALEXNET

[1]D.DEDEEPYA LAKSHMI, [2]V.HARIKA , [3]R.SUNEETHA, [4]A.V.S.MANIDEEP, [5]B.RAGHAVENDRA, [6] Mr.K.SRINIVASA RAO
[1,2,3,4,5] Student of ECE Dept., Kallam Haranadha Reddy Institute Of Technology,Guntur.
[6]Associate Professor of ECE Dept., Kallam Haranadha ReddyInstitute Of Technology, Guntur.

ABSTRACT : Deep Learning has led the way for demanding and wide-ranging applications in nearly every day life during the previous decade, achieving excellent results on a variety of challenging problems such as facial recognition, motion detection, health informatics, and many more. The performance of a state-of-the-art pre-trained Convolutional Neural Networks (CNNs) model is discussed in this research. Existing models, however, are still behind the acceptable accuracy level, which is required for use in real-world applications, due to the substantial intra-class diversity of images. In this paper, we present a deep learning framework based on the ALEX network for accurately predicting an object in a variety of photos. We use multi-task learning to train our model, and we show that augmenting the classifier's feature embedding with the predicted item in an image can improve object prediction accuracy. Our model performed well ona variety of image datasets. We demonstrate that our model has successfully identified the items by visualising the attention maps of the train model.

Keywords : Deep Learning, AlexNet, Pre Trained Images, Image Classification, Feature Extraction.

INTRODUCTION: Image classification is important in computer vision because it helps to bridge the gap betweencomputer and human perception. For picture classification, several classic classification algorithms have been presented in recent years. Additionally, the training and testing phase comes before several phases like as picture preprocessing, image feature extraction, training a classifier with learnt features,and using a taught classifier to make tactical judgments. Despite the fact that these traditional algorithms contribute significantly to image classification, their performance is far from ideal, possibly due to the poor expression of image semantics due to traditional feature extraction.

The task of identifying an object in an image is known as image classification. Real-world itemsof interest, such as faces and people, provide complex difficulties to represent since they are difficult to model, have a wide range of colour and texture, and have unrestricted backdrops. The detection problem, unlike pattern classification, demands us to distinguish between the object class and the rest of the world. As a result, the class model must account for intra-class variation without jeopardising the object's discriminative strength in cluttered settings.

LITERATURE SURVEY : With an idea in the field of Artificial Neural Networks (ANNs), which intimates the logical structure of the brain by viewing categorization as the foremost active application of localization, creativity and invention bring technology to its pinnacle. Deep learning theory demonstrated better performance with the goal of transforming the future of Artificial Intelligence (AI). Because of the large amount of labelled training data and processing power available, supervised deep learning hasrecently shown considerable promise in computer vision. Deep learning networks, particularly convolutional neural networks, have gained popularity in recent years as a result of the addition of more hidden layers between the input and output layers in neural network architecture.

Simple and complex cells in Neocognitron have been replaced by CNN convolution and sub-sampling layer, and the network's processing capability has been increased by reducing the number of parameters owing to weight sharing in CNN. With the advancement of technology, CNN has demonstrated success in picture categorization challenges. Due to the derivation of many characteristics from LeNet architecture in recent deep CNNs, Convolutional connections and back propagation methods were introduced. Furthermore, CNN rose to prominence in 2012, when CNN, dubbed AlexNet, surpassed numerous previously developed algorithms in the ImageNet Large-ScaleVisual Recognition Challenge (ILSVRC) [22]. Since then, Szegedy introduced a deep convolutional neural network for classification and detection based on a 22-layer architecture on top of ImageNet in 2015. The depth and width of the network were grown while keeping the computational capacity constant, and the quality was optimised using the Hebbian principle and multi-scale processing.
Later, G.Huang presented revolutionary designs such as Resnets and DensNets, which make CNNsdeeper and more powerful by incorporating direct connections between two layers with the same feature-map size.

Proposed Work:

ALEXNET: AlexNet has roughly 650,000 neurons and 60 million parameters, which is a significant increase in network scale over Lenet-5. It won the 2012 ImageNet picture categorization competition by a large margin of 11 percent over the second place finisher. Alexnet is a convolutional neural network created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton incollaboration. Visual was trained to classify 1.2 million high-resolution photos into 1000 differentclasses in the ImageNet Large Scale Recognition Challenge (ILSVRC) 2010.

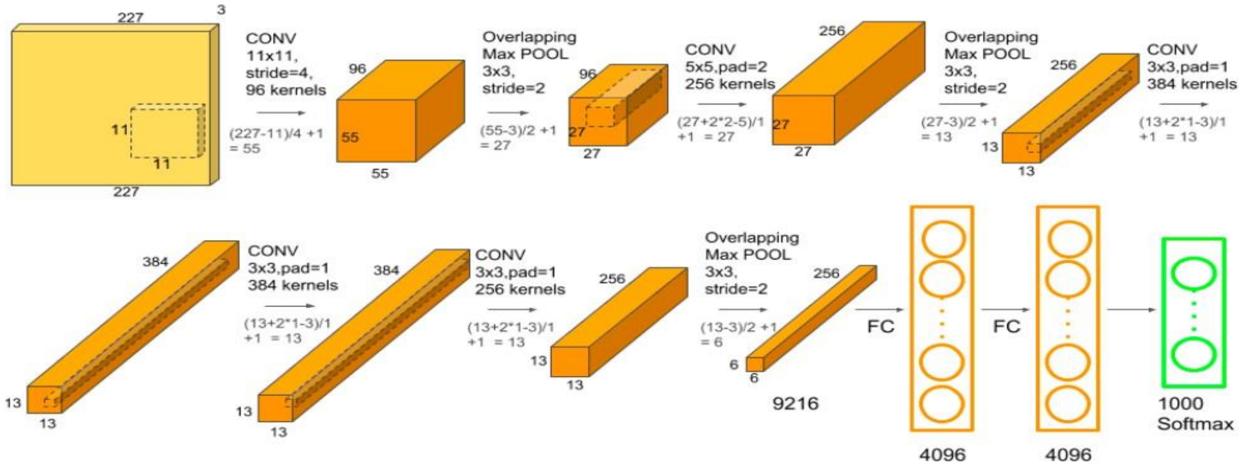


Figure 1: The architecture of Alexnet model

Alexnet is a 14 layer convolution neural network. The Alexnet contains one input layer, five convolution layers, three fully connected layers, three max-pooling layers and two droplet layers.Each layer having its own learnable parameters (Kernals and weights).Two droplet layersare used to overcome the overfitting problem of the network.

A convolutional window of size 11 ×11 is utilised in the first layer. We need to utilise a large kernel to capture the object because the input size is enormous. In the next layers, the convolutional window shape is gradually lowered to 5 ×5 and 3 × 3, but the number of filters israised in parallel. Max-pooling layers are implemented after the two initial and last convolutional layers, with a pooling window of size 3 ×3 and a stride of 2 steps. As a result, theoutput size is cut in half across these pooling layers.

BLOCK DIAGRAM :

In this article, we are going to discover the architecture of this network, as well as its implementation on Matlab. We further apply this network to the problem of classifying images.

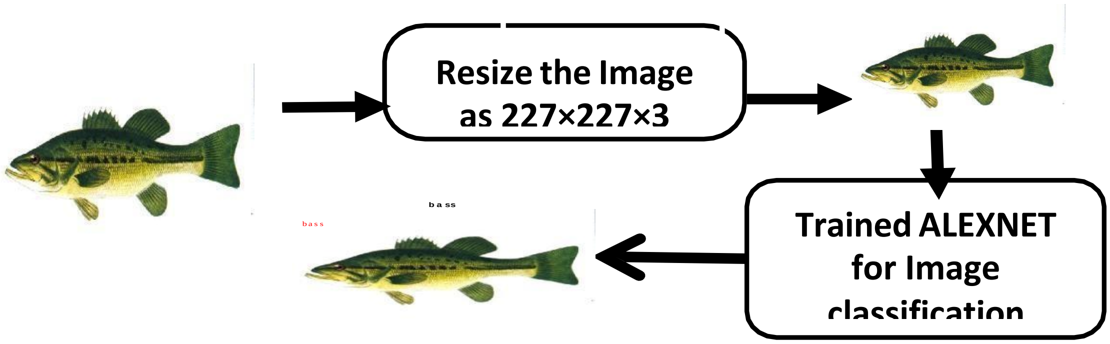

Resize the Image as 227×227×3
bass
bass
Trained ALEXNET for Image classification

IMAGE CLASSIFICATION USING ALEXNET

The primary goal of our study is to train ALEXNETs so that they can automatically identify objects in images in multiple databases. Using multiple databases, train the ALEXNETs such thatthey can automatically recognise the age and gender of a human image. To train the network, we gathered diverse databases from various websites. This database is one of Alexnet's most important resources.

ARCHITECTURE OF ALEXNET:

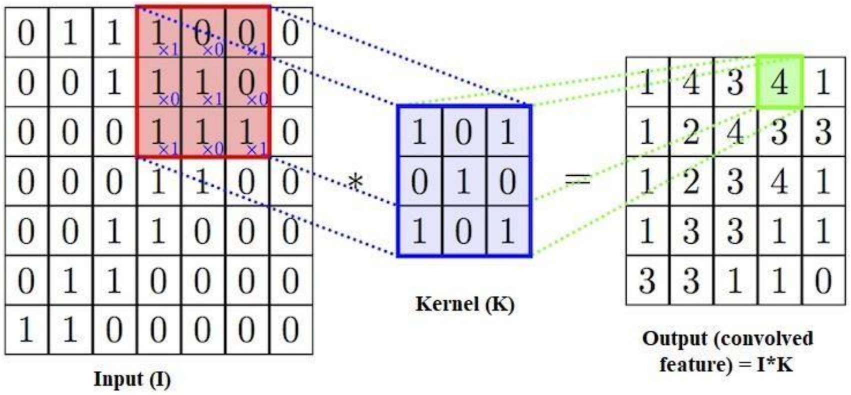
Layer	Filters	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227×227×3	-
Conv 1	96	11×11	4	-	55×55×96	ReLU
Max Pool 1	-	3×3	2	-	27×27×96	-
Conv2	256	5×5	1	2	27×27×256	ReLU
Max Pool2	-	3×3	2	-	13×13×256	-
Conv 3	384	3×3	1	1	13×13×384	ReLU
Conv 4	384	3×3	1	1	13×13×384	ReLU
Conv 5	256	3×3	1	1	13×13×256	ReLU
Max Pool 3	-	3×3	2	-	6×6×256	-
Dropout 1	Rate = 0.5	-	-	-	6×6×256	-
Fully connected 1	-	-	-	-	4096	ReLU
Dropout 2	Rate = 0.5	-	-	-	4096	-
Fully connected 2		-	-	-	4096	ReLU
Fully connected 3		-	-	-	1000	Softmax

Input layer :

It is the First layer of the ALEXNET.This layer act as a barrier between databaseandnetwork.It accepts the RGB image whose size is 227×227×3 and given as aninput to the network. This layer didn't accept the gray scale images and other sizeof images.This layer didn't maintain anylearnable parameters and activation function.

CONVOLUTION LAYER :

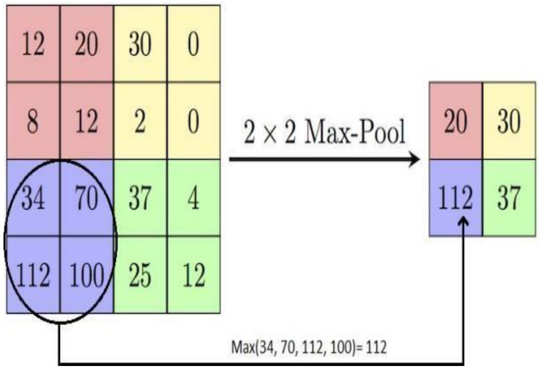
This layer performs convolution operation on the image.Convolution operation isused to detect the feature map of the image.To obtain accurate feature map of theimage, we need to use number of convolution layers in the network.The operationof convolution is as follows.



Feature map size = [[(Input size − filter size + 2∗padding size)/stride] + 1

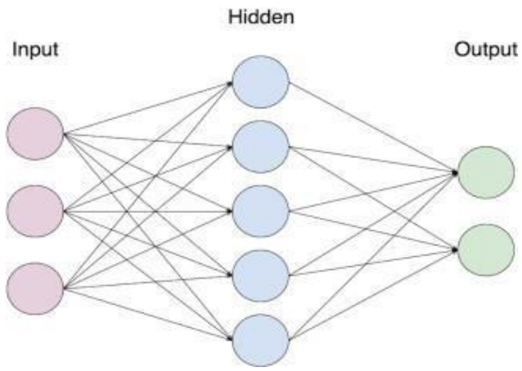
MAX POOLING LAYER :

This layer performs max operation on the image.Max operation is used to reducethe size of the feature map of the image.In Max-pooling layer, we use empty kernel To obtain required size feature map of the image, we need to use number of max pooling layers in the network.



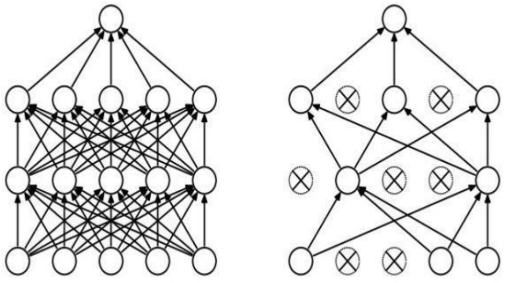
FULLY CONNECTED LAYER :

Fully connected layer is used to flattened (e×1) the feature map of the image. Inthis layer, first the feature map is converted into neurons. This can be achieved by resize the feature map [m×n×p] into [e×1] size, where e = m*n*p. This [e×1] size neurons act as an input to this layer. Next,these input neurons are fully connected with the output neurons with the weights.The schematic representation and operation of fully connected layer is shown below:



DROPOUT LAYER :

Dropout layer performs dropout technique on neurons which is used to improve theefficiency of the network.It is mainly used at input of the fully connected layers.In fully connected layers, some adjacent neurons having same values. Dueto this, it carries same information, takes huge time and more operations are required to obtain the output of fully connected layer. In this situation, the drop out layer removes some neurons which are having same values at the input .



(a) Standard Neural Net (b) After applying dropout.

ACTIVATION FUNCTIONS :

The activation function is a node that is put at the end of convolution layer or fully connectedlayer.It helps to decide if the neuron would fire or not.ALEXNET uses two types of activation functions and they are RELU (Rectifed Linear Unit).

NETWORK TRAINING :

In this project, we use ALEXNET for image classification.We use number of databases fortrain the alexnet .Flickr-Faces-HQ and NUAA Imposter, CVonline, ImageNet and USC-SIPI imagedatasets are used to train the alexnet for image classification.Totally, 52348 images are used for image classification in which 41,878 images for training.In the same manner, the remaining images are used for testing the classifcation network. The alexnet train with 0.0002 learning rate,‘piecewise’ learning rate schedule, 0.2 learn rate drop factor, 32 mini batch size, 45 maximum iterations.

IMPORTANCE OF TRAINING THE NETWORK :

The efficiency of the network output is also depends upon the training of network.In training, the network learns how to recognize or classify the object.While train the network with huge images,it stores the [m×1] size feature map of the images.Based on the feature map of the images, it classifyor recognize the image at testing time

NETWORK TESTING :

After completion of training, we test the network with sample images.When ever one image was passed into trained networks then its identifies the object in an image.

The experimental results are shown in below

Input Image

Resized Image

Output Image

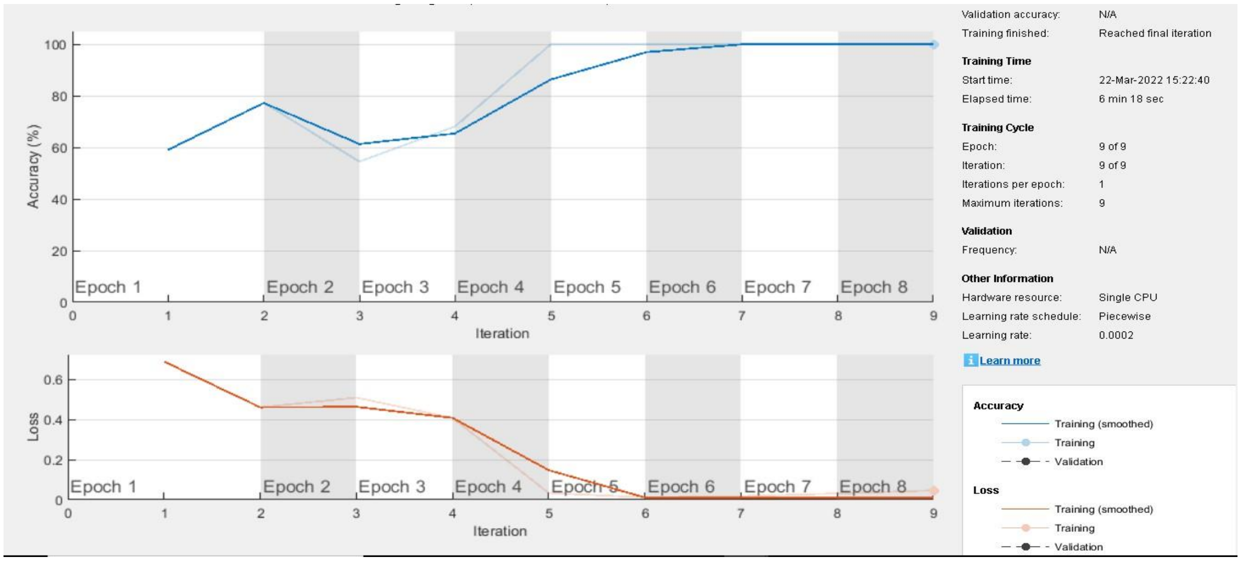
RESULTS :

TRAINING REPORT OF ALEXNET :

The alexnet reaches to significant accuracy after train with several images.

Epoch :

Training the neural network with all the training data for one cycle. In an epoch, we use all of the data exactly once. A forward pass and a backward pass together are counted as one pass: An epoch is made up of one or more batches, where we use a part of the dataset to train the neural network.



Validation accuracy: N/A
Training finished: Reached final iteration

Training Time
Start time: 22-Mar-2022 15:22:40
Elapsed time: 6 min 18 sec

Training Cycle
Epoch: 9 of 9
Iteration: 9 of 9
Iterations per epoch: 1
Maximum iterations: 9

Validation
Frequency: N/A

Other Information
Hardware resource: Single CPU
Learning rate schedule: Piecewise
Learning rate: 0.0002

CONCLUSION:

This project is used to identify the object to one particular image. For this we used the alexnet and trained the network with number of datasets. Flickr Faces-HQ, NUAA Imposter, CVonline, ImageNet and USC-SIPI image datasets are used to train the alexnet for image classification. Totally, 52348 images are used for image classification in which 41,878 images for training. In the same manner, the remaining images are used for testing the classifcation network. The alexnet trainwith 0.0002 learning rate, 'piecewise' learning rate schedule, 0.2 learn rate drop factor, 32 mini batch size, 45 maximum iterations. After completion of training, we tested the network with sample images. When ever one particular input image was passed into trained network then its identifies an object to that image. The accuracy of this project is 97% for image classification.

REFERENCES:

1. Arora, S., Bhaskara, A., Ge, R., & Ma, T. (2014). Provable bounds for learning some deep representations. In Proceedings of the 31th International Conference Machine Learning (pp. 584– 592). N.p.: International Machine Learning Society.
2. Ba, J., & Frey, B. (2013). Adaptive dropout for training deep neural networks. In C.
3. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), Advances in neural information processing systems, 26 (pp. 3084–3092). Red Hook, NY: Curran.

4. Ba, J., Mnih, V., & Kavukcuoglu, K. (2015). Multiple object recognition with visual at- tention. InProceedings of the 3rd International Conference on Learning Representations (pp. 1–10). N.p.: Computational and Biological Learning Society.
5. Chen, W., Wilson, J. T., Tyree, S., Weinberger, K. Q., & Chen, Y. (2015). Compressing neural networks with the hashing trick. ArXiv preprint:1504.04788v1.
6. Cheng, Y., Felix, X. Y., Feris, R. S., Kumar, S., Choudhary, A., & Chang, S. (2015). Fast neural networks with circulant projections. ArXiv preprint:1502.03436.
7. Bottou, L. (1998). Online learning and stochastic approximations. On-Line Learning in Neural Networks, 17(9), 142–177.
8. Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In Proceedings of the International Conference on Computational Statistics (pp. 177–186). Berlin: Physica-VerlagHeidelberg.
9. Boureau, Y., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In Proceedings of the 27th International Confer- ence on Machine Learning (pp. 111–118). N.p.: International Machine Learning Society.
10. Bromley, J., Bentz, J. W., Bottou, L., Guyon, I., LeCun, Y., Moore, C., Shah,
11. R. (1993). Signature verification using a “Siamese” time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7(4), 669– 688.