A MATLAB BASED FACIAL DETECTION SYSTEM USING CONVOLUTIONAL NEURAL NETWORK

Dr.Jesavath Kiran Naik, Lecturer in Electronics and Communication Engineering**,** Government Polytechnic, Adoni**.**

ABSTRACT
Biometrics seems to have assured millions of people with a way to ensure foolproof security. One among such biometrics is the facial detection. It holds a golden pedestal in the realm of successful biometric techniques. This paper delves into the prospect of facial detection using convolution neural networks, performed solely using MATLAB. The proposed CNN can be modified to be made to accept new subjects by training and modifying the last layer of the pre-trained neural network, AlexNet. Viola-Jones algorithm is used as a helping tool for AlexNet. The main aim of the project is to create a fully functional facial detection system using CNNs. Evaluation was done using five subjects from the JAFFE database, producing 95% accuracy within an average time of less than two minutes of training.
KEYWORDS: AlexNet, CNN, Facial detection, JAFFE database, Viola-Jones Algorithm, 95% accuracy.

INTRODUCTION :

A driverless car's navigation system faces Biometrics have certainly proved effective with huge amounts of people relying on retinal recognition, thumbprint recognition, facial recognition[1] and detection, iris detection and recognition, voice authentication, etc. These have proved effective and robust as they cannot be accessed in any way without the explicit presence of the user themselves. However, not all aspects of biometrics such as facial detection and recognition are perfect in all conditions. They face unique issues of their own, such as illumination, pose, scale and imaging parameters. Other issues also include those originating from within the user[2], such as hair, expression, or any medical conditions that might alter any features ever so slightly. difficulty differentiating pedestrians crossing the road from various other vehicles. Filtering, categorizing and tagging billions of videos, gifs and photos uploaded by the users daily on the social media also cannot be handled manually. One way to solve these problems would be through utilization of neural networks. Typical neural networks consist of multi-layer perceptrons. (MLP). A deep neural network is formed through inclusion of various hidden layers to a simple neural network. A CNN[3], [4] itself has several hidden layers within. which CNN thrived was the LeNet-5. CNNs are most often used for pattern recognition[5], image recognition, image detection, object detection[6], etc. LeNet-5 has seven layers, 4 feature extraction layers, 3 MLP layers.

THEORY :

Facial detection is usually done through feature extraction. This is done manually when technologies other than CNN are used. But, in CNN, the feature extraction is done directly by the neural network itself. It is a highly accurate but complex algorithm, primarily having three layers, convolution layer, ReLU (used to overcome the problems arising from back propagation, it is also a step above the sigmoid activation function) and the pooling layer. These are usually followed by the fully connected layer. AlexNet is a pre-trained network as specified before, having 8 layers. The training to testing ratio we implement is 8:2 usually, to ensure prevention of overfitting. The input image size of an RGB image in AlexNet is 227x227.

ISSN: 2278-4632
Vol-14, Issue-7, No.04,July: 2024

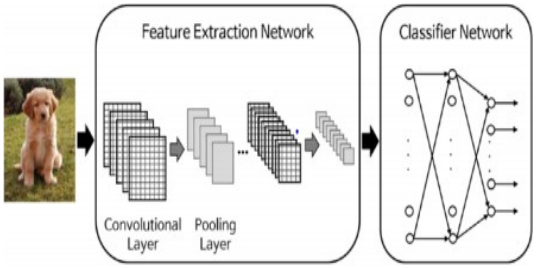


Figure 1: CNN process

Viola Jones Algorithm is used for enabling and aiding AlexNet by using Haar features and identifying them. It was introduced by Paul Viola and Michael Jones in 2001. Feature selection stage outputs are given as inputs to the next step, the feature cascading. Collectively, both these steps form the AdaBoost Learning Algorithm. Then, the classifier at the final stage helps classify the features.

TRANSFER LEARNING :

AlexNet is not inherently capable of recognizing faces on its own. It is an object classifier primarily. However, since it inherently has the capability to recognize patterns, it can be modified quite well to fit our need of detecting faces, by changing the training data set we are using and also modifying the last layer of the neural network, the fully connected layer. Originally it was trained through the ImageNet[7] database to recognize objects up to more than a million while classifying the images into 1000 categories. However, transfer learning is done using the JAFFE database, using the faces of 5 different Japanese models, with 7 basic facial expressions (6 facial expressions + 1 neutral expression), by modifying the fully connected last layer, as stated above. The deep network analyzer is used to modify the layers. The final layer is removed and a layer from the Layer Library, a new classification layer, in this case, is dragged onto the canvas. Eliminating the previous output layer and connecting our new custom layer completes the transfer learning. JAFFE[8] database, however, requires resizing and cropping of the input image to fit its specifications.

ARCHITECTURE OF ALEXNET AND TRAINING :

Transfer learning is followed by the training of the neural network. The training is done in no more than a minute. AlexNet network is split into two pipelines. In the training phase, as in any neural network, each epoch increases the accuracy of the algorithm. This is followed by inputting the data to the first layer, the convolution layer. The convolution layer has filters that act upon the input matrix. Several filters exist in this layer, and form several convolved matrices (feature maps). AlexNet has more filters per layer, as opposed to LeNet, with all the convolution layers being stacked in the former architecture. It has 11x11, 5x5 and 3x3 convolutions. Dropout and data augmentation are the two techniques that are also employed. Each of these filters are used for identification of a unique feature and formed on their own during the training process. The end result is the convolved matrix, this is what is further processed in the next layer. The next layer is a ReLU layer. In AlexNet multiple ReLU activations take place. This data, then are processed in the ReLU layer, wherein, the negative values are removed, and a scale of values ranging from 1 to 10 is applied to each data matrix. A result of 1 implies a certain event is the least probable; while a result of 10 indicates it is most probable. These, in turn, when passed onto thepooling layer, max pooling is performed and it finally moves on to the fully connected layer.

ISSN: 2278-4632
Vol-14, Issue-7, No.04,July: 2024

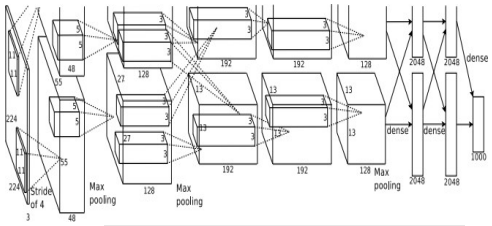


Figure 2: AlexNet Architecture

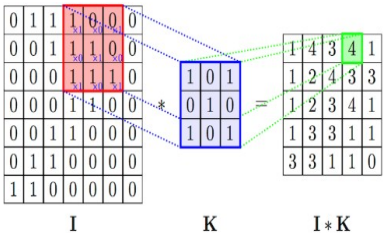


Figure 3: Convolution operation being performed in the first layer, the convolution layer[9]

Edge detection[10] is used for identifying the contours, patterns, edges and curves of the faces. This is done in the first layer itself, the convolution layer.

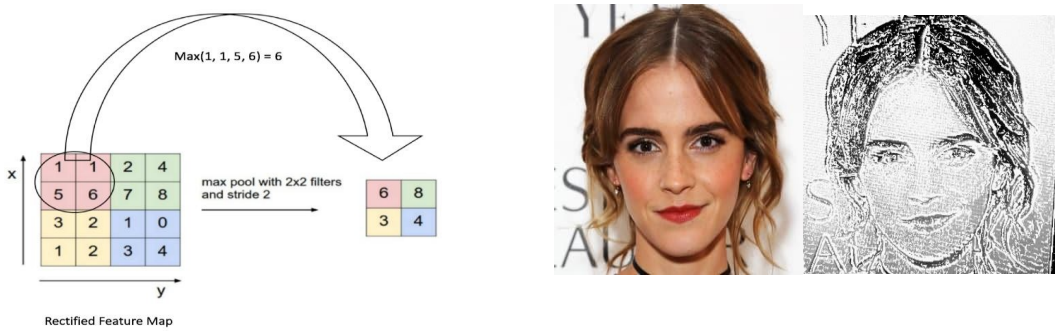


Figure 5: Pooling layer – Max pooling[12]

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:04	31.25%	1.9906	0.0010
20	20	00:01:39	100.00%	0.0026	0.0010

As depicted in the table above, the total time elapsed while training is 1 minute and 39 seconds, with an accuracy of 100 percent by the end of 20 epochs. Each epoch has an increasing accuracy of 31.25%, with a loss of 1.9906. Learning rate (alpha) is 0.0010.

ISSN: 2278-4632
Vol-14, Issue-7, No.04,July: 2024

SAMPLES OF DATABASE :

The database used for transfer learning is the JAFFE database. We use 5 different Japanese models, with 7 basic facial expressions of each (6 facial expressions + 1 neutral expression), and modify the fully connected last layer, as stated above.

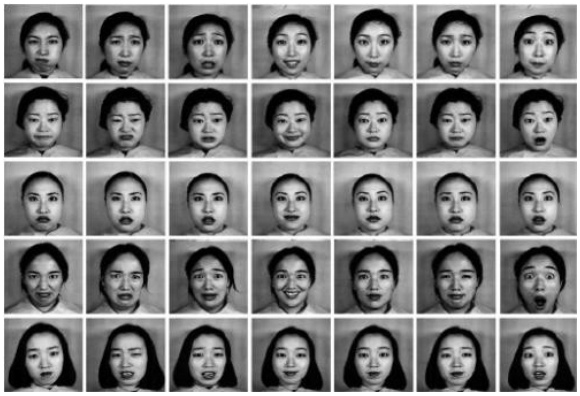
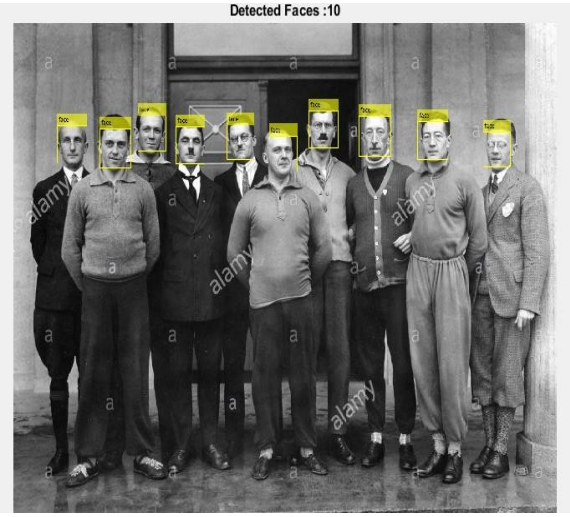


Figure 6: Samples of JAFFE database

EXPERIMENTAL RESULTS AND CONCLUSION

The proposed experiment is run on a 2.20GHz Intel(R) Xeon(R) CPU with Windows 10 Pro system. It is supplemented with Nvidia Quadro K600 GPU, Tesla K20c.

After the training process is finished, the input image is to be given, and the system proceeds to detect the image.



In the output image, ten people are standing, and ten faces have been detected. Less than a minute is required for the whole process to occur. The methodology and time taken also show that with exactly 20 epochs, and less than two minutes of training and detection time, a record number of faces can be detected. However, the more the number of faces available for training, the more the accuracy of the network increases.In conclusion, the proposed experimental system can detect faces with an accuracy of 95% in under an average of two minutes of time. However, this proposed methodology is not fully suitable for a very varied range of facial expressions, poses and illumination. For detection of a wide range of poses, etc, very deep neural networks are required. [13], [14]

REFERENCES :

[1] Mankar, Vijay & G Bhele, Sujata. (2012). A Review Paper on Face Recognition Techniques. International Journal of Advanced Research in Computer Engineering & Technology. 1. 339-346.

[2] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural computation, 1989.Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network".

[3] Deshpande, Adit. "The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part 3)".

[4] A MATLAB based Convolutional Neural Network Approach for Facial Recognition System (2014)
Syafeeza A.R, JBPR.

[5] Carlos E. Perez. "A Pattern Language for Deep Learning".

[6] Habibi, Aghdam, Hamed (2017-05-30). Guide to convolutional neural networks: a practical application to traffic-sign detection and classification. Heravi, Elnaz Jahani. Cham, Switzerland. ISBN 9783319575490.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.

[8] http://www.kasrl.org/jaffe.html JAFFE DATABASE OFFICIAL URL

[9] https://ujjwalkarn.me/2016/08/11/intuitive-explanation- convnets/

[10] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In ECCV, 2014.

[11] https://www.biography.com/people/emma-watson-20660247

[12] http://perso.mines-paristech.fr/fabien.moutarde/ES_MachineLearning/TP_conv Nets/convnet-notebook.html

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[15] MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence
Phil Kim, ISBN-13: 978-1-4842-2844-9 BOOK.

