

A REAL TIME FACE RECOGNITION SYSTEM USING ALEXNET DEEP CONVOLUTIONAL NETWORK TRANSFER LEARNING MODEL

LAWRENCE O. OMOTOSHO¹, IBRAHIM K. OGUNDOYIN¹, JOSHUA O. OYENIYI¹, OLUWASHINA A. OYENIRAN²

¹*Osun State University, Department of Information and Communication Technology, Osogbo, Osun State, Nigeria*

²*Ajayi Crowther University, Department of Computer Science, Oyo, Oyo State, Nigeria*

Abstract. In the field of deep learning, facial recognition belongs to the computer vision category. In various applications such as access control system, security, attendance management etc., it has been widely used for authentication and identification purposes. In deep learning, transfer learning is a method of using a neural network model that is first trained on a problem similar to the problem that is being solved. The most commonly used face recognition methods are mainly based on template matching, geometric features based, algebraic and deep learning method. The advantage of template matching is that it is easy to implement, and the disadvantage is that it is difficult to deal with the pose and scale changes effectively. The most important issue, regardless of the method used in the face recognition system, is dimensionality and computational complexity, especially when operating on large databases. In this paper, we applied a transfer learning model based on AlexNet Deep convolutional network to develop a real time face recognition system that has a good robustness to face pose and illumination, reduce dimensionality, complexity and improved recognition accuracy. The system has a recognition accuracy of 98.95 %.

Keywords: transfer learning, convolutional neural network, deep learning

1. INTRODUCTION

Face recognition involves an automated process of analyzing and identifying facial features of individuals. It belongs to the category of computer vision in the field of deep learning. It has been widely used for authentication and identification purpose in various applications such as access control system, security, attendance management etc. Recent researches in the field of deep learning has shown that the development and usage of convolutional Neural Network (CNN) has greatly improved the accuracy of recognition systems.

The human face is an immensely complex and has dynamic structure with characteristics that can significantly and quickly change in the space of time. Face recognition encompasses a number of behaviors from different areas of human life [1]. Face recognition is a challenge, considering the certain variability in information because of the random variation across different people, including systematic variations from various factors such as lightening conditions and pose. Face pose is a specifically difficult problem because all faces seem similar; specifically, all faces consist of two eyes, mouth, nose, and other features that are in the same location [2].

*Corresponding author, email: thejoshoyeni@gmail.com

Convolutional Neural Networks is a feed-forward network structure that can conveniently extract relevant features from images implicitly with minimal preprocessing. Just like other neural networks, CNN are trained with a version of the back-propagation algorithm and has demonstrated an outstanding performance on image classification tasks [1]. Transfer learning is the process of using models trained on one problem as a starting point on a related problem, the pre-trained models are used directly and integrated as an entirely new model for preprocessing, feature extraction, classification. In deep learning, transfer learning is a method of using a neural network model that is first trained on a problem similar to the problem that is being solved.

Facial features extraction method can be broadly classified into appearance-based (Holistic) and model-based methods. The hybrid method is a combination of these two methods [1]. Convolutional neural networks feature extraction method extracts features by layer-by-layer convolution and then multi-layer nonlinearity. The mapping enables the network to automatically learn from the training samples that have not been specially pre-processed to form feature extractors and classifiers suitable for the recognition task [3].

In order to research the potential application for office door access control using the PCA and artificial neural networks-based eigenfaces technique, an automated face recognition system was developed. [4] The training images can be obtained either offline using advance captured and cropped face images, or online using face detection and recognition training modules on the real frontal face images of the system. The system can recognize faces at a realistic rate of 40 cm to 60 cm distance from the camera with the individual head at rotational angle between -20° to $+20^\circ$. The influences of illumination and pose on the face recognition system was shown in the results obtained.

In [5] three different experiments were conducted to improve PCA performance by decreasing the computational time while same performance. The first experiment was performed to find the best number of images for each person to be used in the training set that gives a highest recognition rate. The analysis is tested using 28 images for each person with 6 images used for training process in the second experiment. The number of eigenvectors decreased in the third experiment, producing less computation time. In terms of accuracy, efficiency is the same with less computational time in the second experiment. This method presented reduce the computation time by 35% compared with the original PCA algorithm particularly with a large database.

In [1] Principal Component Analysis-Back Propagation Neural Network (PCA-BPNN) with Discrete Cosine Transform (DCT) DCT was used to compress the face databases and when combined with PCA, the system recognizes faces easily. Face94 and Grimace databases were used to test the system performance and it gives a recognition rate that is above 90 %.

An open code based on deep-learning method was presented to perform facial recognition [6]. The system is made up of five main steps: face segmentation, facial features detection, face alignment, embedding, and classification. Deep learning approach were used for the fiducial point extraction and embedding and support Vector Machine (SVM) is used for classification task due to its fastness both in training and inference. The system achieved an error rate of 0:12103 for facial features detection, which is pretty close to state of the art algorithms, and 0:05 for face recognition and it is capable of running in real-time.

It was noted in [7] that the application of CNN for face recognition can effectively reduce the requirements for training samples, and the more network layers are learned the more global the features. The problem of large spatial complexity of direct use of face image data was overcome by the method presented in [8], the method employs the use of matrix-like kernels to extract image feature values which increases the nonlinear structure of features, and improves the expression ability of feature vector. Constrained sparse matching method was employed [9] to effectively solve the similarity between the face image sequences and automatically select similar face image sequences, which improves the feature matching accuracy.

The most commonly used face recognition methods are mainly based on template matching, geometric features based, algebraic features based such as principal component analysis (PCA), linear discriminant analysis (LDA) and deep learning method. The advantage of template matching is that it is easy to implement, and the disadvantage is that it is difficult to deal with the pose and scale changes effectively. Regardless of the method used in face recognition system, the most essential concern is dimensionality and computational complexity especially when working on large databases. Regardless of the method used in face recognition system, the most essential concern is dimensionality and computational complexity especially when working on large databases. Thus, in this paper,

we adopt a transfer learning model based on deep convolutional neural network to reduce data dimensionality, computational complexity and enhance recognition accuracy in face recognition system.

2. EXPERIMENTAL SETUP

The method involves acquisition of facial images from individuals, preprocessing of the image to a standard normalization before they flow into couples of convolution using the pretrained AlexNet convolutional neural network transfer learning approach for feature extraction and classification as presented in Figure 1.

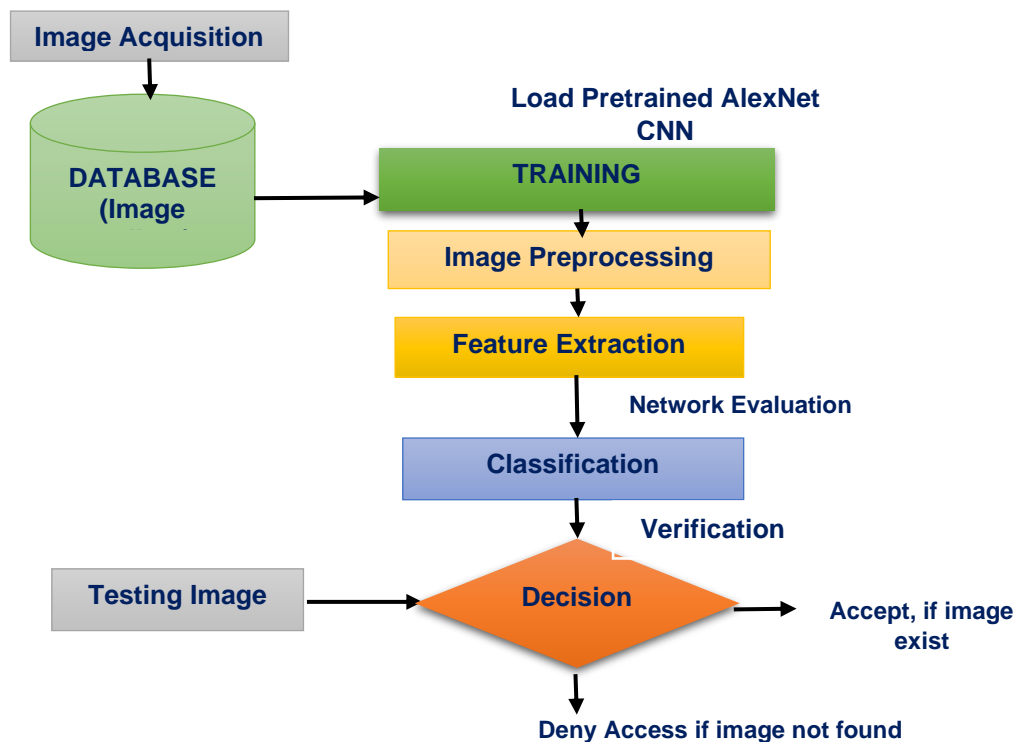


Fig. 1. The designed framework for face recognition system based on AlexNet transfer learning.

1. Image Acquisition: Image acquisition is the very first step. At this an electro optical camera (an inbuilt webcam) was used to capture the facial image of subjects. The system creates the face object detector using cascade object detector, after which a point tracker and webcam object was created (using HP WEBCAM). The system then creates a video player objects to capture 50 images of an individual subject using a loop function in a 45 seconds. 50 individual were captured, the images are then stored into a designated folder for each subject in the database. The total number of images used as dataset is 2500.

2. Image Pre-processing: At the preprocessing stage, region of interest (ROI) was extracted from the acquired images and normalization was done. The purpose of the pre-processing is to reduce or eliminate some of the variations in the images due to illumination and improve the visual quality. To maintain the uniformity in the database, all the images are resized to 277×277 pixels by convolutional neural networks (CNNs).

3. Load Pre-trained Network and the Train Network: This phase is also called the modeling phase. At this stage, the content of the Image Dataset were split into 80 % by 20 %, such that 80 % of the image in the repository were used as training dataset and the remaining 20 % were used as testing dataset. The system loads the pre-trained network (AlexNet CNN) based on transfer learning model and trained the network using the initial learning rate 0.00001 and other parameters shown in Table 1.

4. Training: At the training stage, the (n) number of training dataset samples are divided into small batches (b), then the model coefficients are updated using model error (equation (1)):

$$T = \frac{n}{b} \quad (1)$$

where T represent the total n is number of iterations per training epoch and b represent the batches. The weights (w) of CNN is optimized using error function defined in equation (2):

$$E_t[f(w)] = \frac{1}{b} \sum_i^{tb} = (t-1)b + 1 f(w, x_0) \quad (2)$$

where X_0 is the sample of training data and W represent the weight. At each iteration the weights are updated by rule mini batch gradient descent update rule with learning rate given in equation (3):

$$w^{t+1} = w^t - \mu \nabla_w E_t[f(w)] \quad (3)$$

Table 1. Parameter for the training.

Parameter	Specification
Learning rate	0.00001
Mini batch size	64
Epoch	20
Filter size	1
Iteration per epoch	12
Maximum iteration	240

5. *Feature Extraction*: Features extraction is the method of extracting a feature set from the data acquired. It involves finding a set of vectors that efficiently represent an observation's information content while reducing the dimensionality. The last dense sigmoid/logistic layer was trained by the CNN network to extract features from the CNN model. Convolutional neural networks are based on profoundly supervised learning. The layer is also called hidden feature extractor which describes the internal connectivity of the image region.

6. *Classification*: Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model predict the value of one or more outcomes. Outcomes are labels that can be applied to a dataset. CNNs classifier which is a deep convolutional neural network was used for classification.

7. *Matching*: Matching is a method where the feature extracted from the image called user template are compared with the template of the image stored in the database. It helps us to verify the authenticity of the person.

AlexNet convolutional neural network is a deep feed forward neural network, which is based on the LeNet neural network. Compared with the LeNet neural network, the AlexNet neural network has a deeper network layer and more convolution kernel parameter. The layers in AlexNet CNN are:

- *Convolutional Layer*: The convolution kernel is an $n \times n$ matrix that moves on the map in a set step size. A convolution operation is performed on the map at the corresponding location each time it moves in steps to eventually create a new map. Extracting features from the input image is the primary objective of this layer. It consists of a series of learning filters (kernels) used to identify image patterns.

- *Pooling layer*: The layer of sampling is also called the pooling layer. Its primary role is to reduce the scale of data. By using certain functions to summarize sub-regions, such as taking the average or maximum value and making learned features more robust, it reduces the size of the feature maps by making changes in scale and orientation more invariant.

- *Flat layer*: After multiple convolutions and pooling operations, the work of the flat layer is to transform the data output into a one-dimensional vector output in a matrix form to provide calculations for the fully connected layer.

- *Fully connected layer*: Each neuron of the fully connected layer is connected to each neuron of the upper layer, and the final calculation result is obtained by the selected activation function. FC layer takes all neurons in the previous layer (be it fully connected, pooling, or convolutional) and connects it to every single neuron it has. Adding a fully-connected layer is also a way of learning non-linear combinations of these features.

- *Output layer*: The output layer acts as the last layer and is used to calculate the probability response.

3. RESULT AND DISCUSSION

The simulation was done on Matlab 2018a. Figure 2 show the graphical user interface (GUI) of the system for image acquisition. A sample of the face database for each subject captured is shown in Figure 3. Figure 4 shows the forward and back propagation during the training, forward propagation denotes the step where the input data are transformed into output through CNN layers, back propagation propagate the errors from a layer to the previous one and compute the derivate of the error with respect to weight and biases. Figure 5 show that the system has at an epoch of 1, the training iteration value is 1 with a mini batch accuracy of 6.25 % and mini batch loss of 3.5139. At an epoch of 5, the training iteration value is 50 with a mini batch accuracy of 73.44 % and mini batch loss of 0.9185. At an epoch 9 of the training iteration value is 100 with a mini batch accuracy of 89.06 % and mini batch loss of 0.3638.

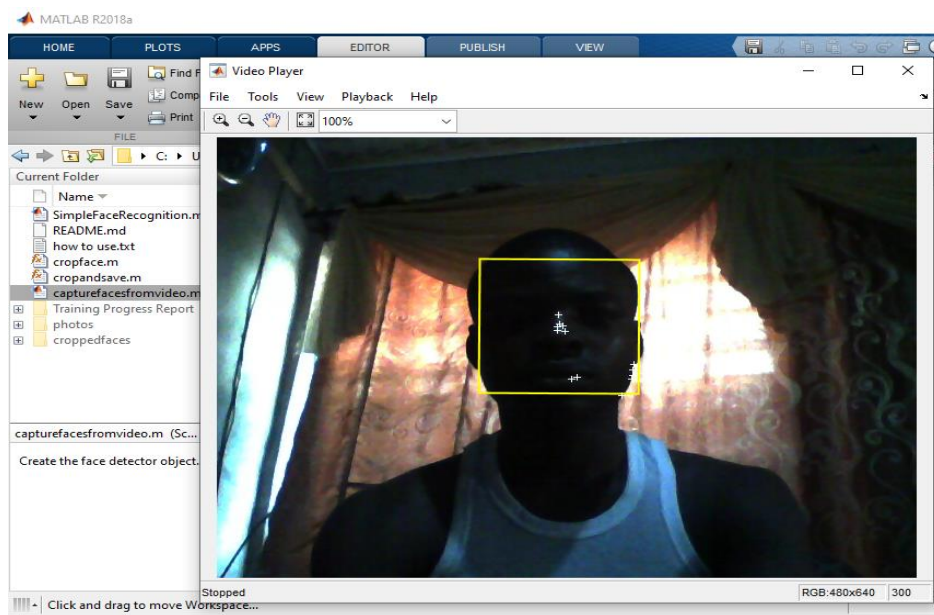


Fig. 2. Image acquisition.

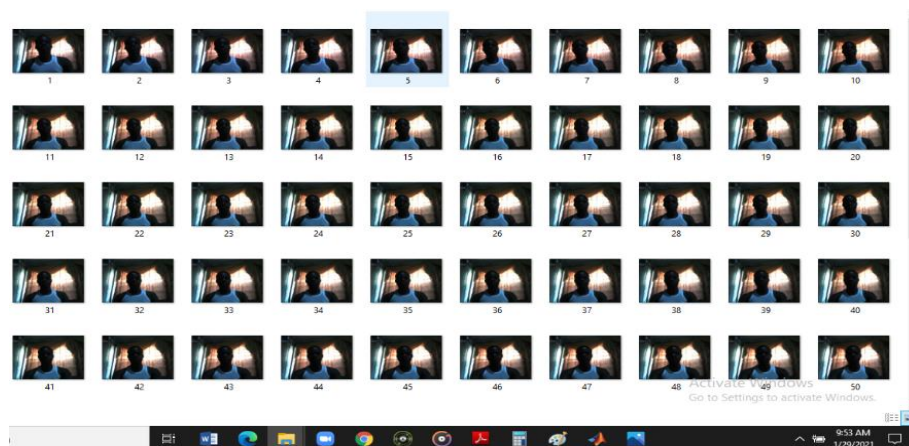


Fig. 3. Sample of facial database for individual.

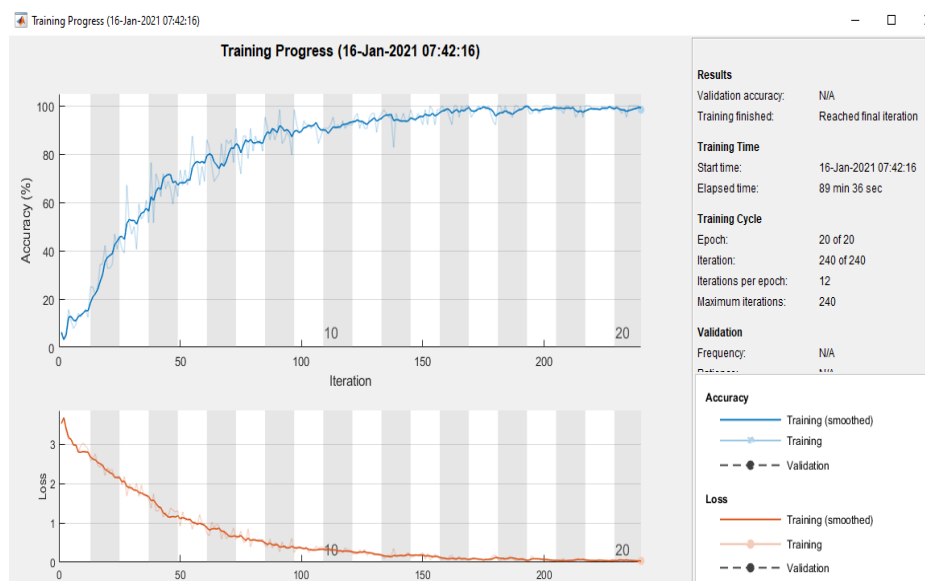


Fig. 4. Training process.

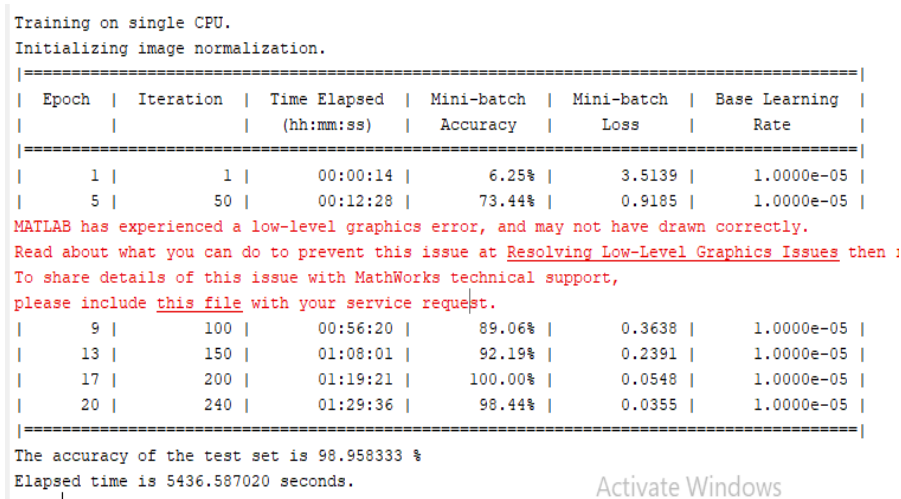


Fig. 5. Result of the test set.

At an epoch of 13, the training iteration value is 150 with a mini batch accuracy of 92.19 % and mini batch loss of 0.2391. At an epoch 17 of the training iteration value is 200 with a mini batch accuracy of 100 % and mini batch loss of 0.0548. At an epoch of 20, the training iteration value is 240 with a mini batch accuracy of 98.44 % and mini batch loss of 0.0355. There is a significant reduction in the batch loss as training goes through epoch and iteration increases. The recognition accuracy of the test set is 98.95 % as shown in Figure 5.

The result shows that AlexNet Deep convolutional network improved the recognition accuracy and thus reduce complexity in the system.

4. CONCLUSION

Face recognition algorithm based on transfer learning model using AlexNet Convolutional neural network has better robustness to light, pose and complex background. The use of AlexNet Convolutional neural network for feature extraction, training, and classification and testing greatly reduced complexity and dimensionality in face recognition system, thus, recognition accuracy was improved. The system is quite simple to implement for identity management, attendance, and security system. However, future work should consider the learning effect of optimization algorithm, data augmentation on dataset that consist of n a larger number of subjects.

REFERENCES

- [1] Barnouti, N.H., Mahmood, S.S., Matti, W.E., Face recognition: A literature review, *International Journal of Applied Information Systems, Foundation of Computer Science FCS*, vol. 11, no. 4, 2016, p. 21-25.
- [2] Jain, A.K., Li, S.Z., *Handbook of face recognition*, Springer, New York, 2011.
- [3] Zhiming, X., Junjie, L., Hui, S., High performance computing and computational intelligence conference, *Journal of Physics, Conference Series*, ser. 1395 012006, 2019, p. 1-6.
- [4] Ibrahim, R., Zin, Z.M., Study of automated face recognition system for office door access control application, *Communication Software and Networks (ICCSN), IEEE 3rd International Conference on Communication Software and Networks*, 2011, p. 132-136.
- [5] Abdullah, M., Wazzan, M., Bo-Saeed, S., Optimizing face recognition using PCA, *International Journal of Artificial Intelligence & Applications (IJAIA)*, vol. 3, no. 2, 2012, p. 23-31.
- [6] Passos, W.L., Quintanilha, I.M., Araujo, G.M., Real-time deep-learning-based system for facial recognition, *Simposio Brasileiro de Telecomunicacoes e Processamento de Sinais (SBrT)*, vol. 37, 2018, p. 16-19.
- [7] Chen, Y., Wang, L., Convolutional neural network for face recognition, *Journal of Northeast Normal University (Natural Science Edition)*, vol. 48, no. 2, 2016, p. 70-76.
- [8] Zang, H., Face recognition method based on feature fusion of quasi-matrix neural kernel, *Computer Engineering and Applications*, vol. 54, no. 2, 2018, p. 116-122.
- [9] Zhang, J., Ni, S., Cao, Z., Face recognition via deep learning and constraint sparse representation, *Transactions of Beijing Institute of Technology*, vol. 39, no. 3, 2019, p. 255-261.