Feature level fusion of Face and Iris using Deep Features based on Convolutional Neural Networks

Supreetha Gowda H D
Department of Computer Scinece
Manasagangotri University of Mysore
MYSORE-560 007, India
Email: supreethad3832@gmail.com

Mohammad Imran
125 126, 1st Floor, Brigade Arcade,
Brigade Metropolis, Garudacharpalya
Bengaluru-560048, India
emraangi@gmail.com

Hemantha Kumar G
Department of Computer Scinece
Manasagangotri University of Mysore
MYSORE-560 007, India
Email: ghk2007@gmail.com

***Abstract*—In this work, we have proposed novel deep CNN framework architectures that effectively represent complex image characteristics which performs feature extraction in just two convolution layers and has successfully proved to be an reliable biometric verification system on employment of physiological traits face and iris for our system development. Extensive experiments in configuring the CNN hyper parameters such as number of convolution layers required, filters and its size in each layer, batch size, epochs, iterations and learning rate is a paramount, determining these factors truly depends on the nature of data and its size. Our work has relinquished our novel idea and has obtained 99% of GAR in unimodal biometric verification system itself and definitely the approach has rendered great results when compared with conventional feature extraction and classification techniques.**

***Index Terms*—CNN, Verification, Multimodal biometric, Fusion.**

I. Introduction

Biometric systems are developed for unique identification of humans hence the deep knowledge of subject-technology interface is important [2]. Understanding the stability of biometric modalities which are adopted in deploying the biometric systems in comparatively longer periods of years is definitely a challenging issue. Generalizing the paramount issues in such as, knowing particular trait that best fits an application, understanding which feature extraction algorithm is stable in presence of noise, Classification algorithms, selection of best multibiometric fusion schemes etc., which all greatly matters in developing a system. Analysing on these aspects at basic level distinctiveness is in greater demand.

Some of the applications of artificial intelligence are vehicles speed limit monitoring, banking authorization, self-driving cars, future prediction, space navigation, speech recognition, NLP, image recognition etc., these applications reduces risk of human error and humans are also tiered of computing such things, whereas machines does it confidently on proper training and never feel tiered. Artificial intelligence is a intelligent software which imitates the human behaviour by analysing how human brain thinks, learn and finally decide in solving a problem [1]. Artificial intelligence theoretical concept was introduced in 1956, machine learning (ML) is the subset of artificial intelligence and deep learning is subset of ML. Deep learning works on neural network concepts which is simulated to take decision like human brain.

In ML, the system learns from its ability without being hard programmed and finally predicts. Ex: determining the species with the input features. In ML, the system learns from the data given, once the machine is trained with the training data, it classifies the new input accurately by predicting to which class it belongs on learning. In ML we have two types- supervised and un-supervised learning. In supervised learning, the samples belonging to the classes are predefined with labels and the model is built which consists of generalized templates of all the classes. The test data is passed to the model for claiming its identity. Accuracy is the difference between desired output and actual output. Ex: KNN, decision tree, logistic regression. In unsupervised learning the data does not contain predefined labels, given the data classes are created by making sure these two parameters are preserved high intra class similarity and low inter class similarity. Ex: K means clustering. In reinforcement machine learning the system learns by actions in environment by interacting with a space than being learnt explicitly. The limitations of ML have led to advancement of deep learning area. Some of the major drawbacks of ML are high dimensional, NLP, optimal feature extraction technique depending on data. Deep learning (DL) addresses these issues intelligently by focusing on right features by small guidance from the programmer. DL evolved from ML mimics the human brain. DL ideas are evolved from neural network. In human brain, dendrites provide input to neuron, nucleus performs the function and the output is sent to another neuron from axon output terminal.

DL is based on artificial neural network which generates feature hierarchies by statistical machine learning techniques. DL consists of input layer, hidden layer and output layer. All the nodes in the layers are interconnected to each other. DL consists of thousands of hidden layer, the raw data is provided to the input layer, each node in the layer stores some data they come across, then the data is sent to the next layer nodes which learns abstract information of data [4].

One epoch is completed when the entire dataset is passed forward and backward only once through the network. There is no ideal answer in fixing the right number of epochs as it directly depends on nature of data, when the number of epochs increases, weight gets updated in the network leading from underfitting to optimal fit curve. The entire dataset is divided into batches, as we cannot pass the entire dataset all at once to the network. Batch size is the number of training samples present in a batch. Iteration is the total number of batches needed in completing an epoch. Learning rate is the hyper parameter where we tune the neural network by adjusting weights, learning rate has to be optimal else if the learning rate is high then training may not converge.

II. Literature Review

A. S. Al-Waisy et.al [5] proposed a multimodal biometric identification system adopting face, left and right irises traits. Employing the deep learning system followed by softmax classifier that gives the probability distribution on all the classes. Traditional machine learning approaches such as LBP, LPQ should undergo preprocessing transformations etc, whereas DCNNs plays a reliable substitute that performs both feature extraction and classification. Veeru Talreja et.al [6] developed multibiometric system using the feature level fusion adopting iris and facial traits. The deep features are extracted from face and iris traits by CNN, then these features are further fused with fully connected layer or bilinear layer.

Raghavendra et.al [8] proposed a novel work for periocular verification based on smatphone device, features were extracted by maximum response filter, then the extracted texture features are classified using deep neural networks by deeply coupled autoencoders and obtained genuine match rate of 92%. Iris segmentation is done in one of the four ways based on implementation. The first method is boundary based method where the pupil is identified, firstly eyelid and limbic areas are separated from iris. Ex: Hough transform, Daughmans intergo differential operator. In the second method, pixel based segmentation is done by identifying iris boundary using color, texture and illumination in differentiating between iris pixel and neighborhood pixel. Third method is based on circle fitting based methods. The weakness identified by these three methods are less accuracy in segmentation, detection error, when calculated gradient is affected where we fail to identify iris boundary. To address these drawbacks learning based via deep CNN was used by Arsalan [**?**] proposing two stage iris segmentation, in first stage noise removal, edge detection by canny, contrast enhancement was done. In second stage deep CNN with 21*21 pixels of input used to fit true iris boundary.

III. Methods and Materials

Neural network: Generally neural network models works as follows, let $(x_1, x_2 \ldots x_n)$ be the input data, the input is multiplied with corresponding weights that are randomly assigned in beginning $(w_1, w_2 \ldots w_n)$, the processing element performs the summation $\sum x_i w_i$, then the transfer function F(s) is generated, the activation function sigmoid can be used, then the neuron gets fired if it is above the threshold else it does not. Finally a network is framed by connection of multiple neurons for information processing, if the actual output is not equal to desired output, then the weights are updated again and the process is repeated until actual output is equal to desired output. The updating of weights in achieving the desired results is called as back propagation [3].

Convolution Neural network(CNN): CNN is a feed forward artificial neural network, generally contains four layers convolution, ReLU (rectified linear unit), pooling and fully connected layer. CNN solves the case effectively when there are certain deformations in images. CNN compares images by patches roughly by filters at the same position and looks into similarity. Once the filter patches are chosen, it is applied on the entire image and if it matches then the test image is classified correctly. In convolution layer the filter is moved on all possible position on the image, then each pixel of the image is multiplied with corresponding filter pixel, then we add them and finally division is done by counting total number of pixels in the filter patch, this process is done on all positions of image and compare how well that filter matches that area. Likewise the convolution operation is done similarly by all chosen feature patches on all possible directions of image and we get filtered output. In ReLU layer, the negative values are removed from filtered images, where zeros are replaced which avoid summing of zero values.

ReLU layer activates if the value is above the threshold. $f(x) = \begin{Bmatrix} 0 & if & x < 0 \\ x & if & x \geq 0 \end{Bmatrix}$.

In pooling layer, the image stack obtained after ReLU layer, image is shrinked by picking a window size and stride optimally, usually kept small size. The window is moved through out the filtered image and finally the maximum value in the window is retained, which reduces the image size. This process is repeated for all the images obtained after the ReLU layer process. Again convolution, ReLU, pooling is done further in obtaining the smaller resized image.

In fully connected layer, each and every neuron is connected to every other neuron in previous layer. The convolution and pooling layer extract the higher level of features; actual classification is done in fully connected layer.

IV. Experimental results

We have proposed a novel deep CNN architecture in modeling a robust and reliable biometric verification system adopting the promising traits face (ORL dataset) and iris (CASIA dataset), considering 40 users. Considering the standard benchmark threshold values for FAR 0.01%, 0.1%, 1%, we have performed extensive experimentation in

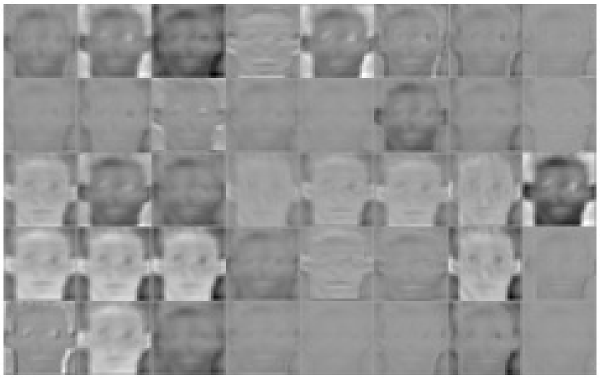

Fig. 1. Visualization of CNN's second layer : Convolution

analyzing the optimal curve fitting for the data, for which we have worked in finding out the right number of batch size, epochs to be fixed for this particular data. Though generalization of fixing up of batch size, epochs, learning rate is not possible and it directly depends on nature of data, which could be done by iterative method in fixing up the curve optimally rather being under fit or over fit, by gradient descent method we usually fix up the curve by weight update.

The proposed novel deep learning architecture for the adopted face and iris data is as follows; the original image size has been re-sized to $60 \times 60 \times 1$, after the rigorous experimentation we have fixed the number of convolution layers to two, once the convolution process is done by the suitable filters ReLU and Max pooling operations are carried out. The output of first convolution layer is the input for the next convolution layer. Initial Learn Rate is $1e-3$, maximum epochs is 10, Mini Batch Size is 10.

The novel deep CNN arhitecture of two layers is as follows: In the first CNN layer, we have considered 20 output filters with 3×3 kernel size and padding is set to 1 because $(k-1)/2$ is padding, where k is kernel size. Once the convolution process is done ReLU activation function is set which retains the positive part of its arguments, finally Max pooling is performed with 2×2 size and stride 2. In the second CNN layer, we have considered 40 output kernels with 3×3 size, as we have already discussed that the input for second layer is obtained from first layer, hence 40 kernels. Then ReLU followed by Max pooling is done. Finally the fully connected layer which contains the connectivity of all the neurons in the previous layers, here the case is 40 fully connected layers, as we know the total number of classes is also 40 considered for experimentation. Softmax function performs the categorical probability distribution which analysis and predicts the multi class classification problem in identifying to which class this particular sample belongs.

Table-I shows the verification results of iris and face unimodal biometric system. For all the standard benchmark threshold values, we have obtained significant results on which

Fig. 2. Visualization of CNN's second layer : Rectified Linear Unit

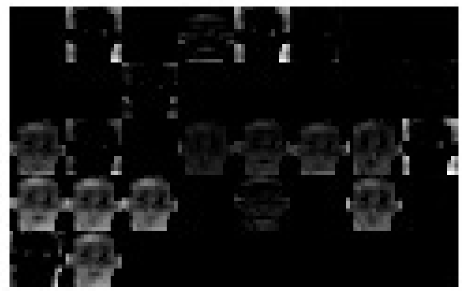

Fig. 3. Visualization of CNN's second layer : Max pooling

TABLE I
RESULTS OF UNIMODAL BIOMETRIC VERIFICATION SYSTEM

FAR%	Iris(GAR%)	Face(GAR%)
0.01	98	98.5
0.1	98.5	99
1	99	99

TABLE II
RESULTS OF MULTIMODAL BIOMETRIC VERIFICATION SYSTEM

FAR%	Multimodal feature level fusion of face and iris (GAR%)
0.01	99
0.1	99
1	100

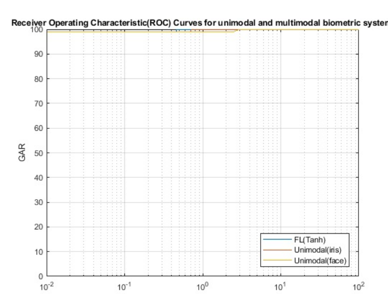

Fig. 4. ROC Curve of Iris and Face - unimodal and multimodal biometric system

one can completely rely on the proposed system for tighter security systems.

Table-II shows the verification results of multimodal biometric systems implemented using feature level fusion scheme. We have obtained 100% GAR on 1% FAR, the results depicts the confidence of the novel proposed deep CNN architecture and the fine tuning of its hyper parameters in obtaining remarkable result.

V. Conclusion

The proposed unimodal biometric verification system is performing well and of course it serves the tighter security needs, feature level fusion under multimodal biometric system has obtained 100% accuracy. Now, its up to the system developers to rely on multimodal biometric system deployment, as we can see from our results unimodal verification system itself is yielding prominent results. When deep CNN was not that active, previous research works in generalizing the dynamic selection of feature extraction, feature selection and reduction, classification techniques was very tedious.

On comparing the conventional methods of feature extraction and classification techniques with deep CNN architecture, current research on deep CNN models prove significant and promising in the entire machine learning approach problems and of course non deterministic problems would also be addressed successfully in future.

Our future work would address the issues in generalizing the deep learning CNN parameters such as selection of optimal output filters, number of epochs, batch size, learning rate for physiological and behavioural biometric modalities.

References

[1] Van Gerven M, Bohte S. Editorial: Artificial Neural Networks as Models of Neural Information Processing. Frontiers in Computational Neuroscience. 2017;11:114. doi:10.3389/fncom.2017.00114.

[2] H. D. S. Gowda, G. H. Kumar and M. Imran, "Robust multimodal biometric verification system based on face and fingerprint," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 243-247

[3] Werbos, P.J., Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, https://books.google.co.in/books?id=z81XmgEACAAJ, 1975, Harvard University.

[4] Deng, Li Yu, Dong. (2013). Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing. 7. 10.1561/2000000039.

[5] A. S. Al-Waisy, R. Qahwaji, S. Ipson and S. Al-Fahdawi, "A multimodal biometrie system for personal identification based on deep learning approaches," 2017 Seventh International Conference on Emerging Security Technologies (EST), Canterbury, 2017, pp. 163-168.

[6] Veeru Talreja, Matthew C. Valenti, Nasser M. Nasrabadi , Multibiometric secure system based on deep learning. 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (2017): 298-302.

[7] Raghavendra, Ramachandra and Christoph Busch. Learning deeply coupled autoencoders for smartphone based robust periocular verification. 2016 IEEE International Conference on Image Processing (ICIP) (2016): 325-329.

[8] Arsalan, Muhammad Hong, Hyung Naqvi, R.A. Beom Lee, Min Cheol Kim, Min Seop Kim, Dong Sik Kim, Chan Kang, Ryoung , Park. (2017). Deep Learning-Based Iris Segmentation for Iris Recognition in Visible Light Environment. Symmetry Deep Learning-Based Biometric Technologies.