Secure and Hassle-Free EVM Through Deep Learning Based Face Recognition

Ishani Mondal
Department of Computer Science and Engineering
IIT Kharagpur
Kharagpur, India
ishani340@gmail.com

Sombuddha Chatterjee
IT Department
Tata Consultancy Services, Pvt. Ltd
Kolkata, India
talktosombuddha@gmail.com

***Abstract*—The primary right of voting in the elections is the fundamental yardstick of a democratic citizen. During the modern era, Electronic Voting machine (EVM) has been introduced which has marked a significant change in the conventional voting system in India replacing the ballot papers and boxes which were used earlier. Previously, the ballot papers used to consume a lot of time, due to the malpractices like booth-capturing and ballot-box stuffing, leading to more disputes and delayed results announcement. In this paper, we propose a EVM system which captures the facial image of a voter through a deep CNN based face recognizer, verifies it with the pre-captured images in the database, the result being positive, assumes the voter is a valid one, asks him to cast his vote for a political party. After voting, the facial alignment of the voter gets deleted from the system, ensuring the voter to vote for only once. This article describes the methodology adopted towards building the system and compares the performance of the face-recognizer (99.1%) against a baseline.**

Keywords—Deep Learning, Face Recognition, Electronic Voting Machine, Secure Authentication

I. INTRODUCTION

India is the largest Democratic nation in the world. It comprises of approximately 1000 million voters who cast their vote in the elections. The Indian election commission is an autonomous body of the Constitution, which has successfully hold elections to the Parliament and many other legislative assemblies of the state for a few years in a concise, justified, authenticated and peaceful way. The Commission is popularly known as a Global Gold Standard in Election Management across the World. The Commission is well known as it has successfully embraced, adopted and later implemented the best state-of-the-art technological advances in achieving a global summit as per the process of election has been concerned. Thus the Electronic Voting Machine(EVM) serves as the major tool to record the entry of votes, its processing and finally counting the votes without any manual discrepancy in a secure and credible fashion. Face detection and recognition has been a significant interaction tool used in the security systems, access-control which has gained enough popularity in the last few decades. With the technological advancement, there has been a massive boom of artificial intelligence. Thus many computer vision security techniques such as face recognition system, gesture detection, retina detection have gained some importance as the system is non-intrusive in nature and it has proved to be quite useful to determine one's identity uniquely. Over the years, quite a few number of face detection and recognition techniques have been proposed and that has showed up with promising results. Usually a face recognition module comprises of the following : detecting the face, its alignment detection, extracting out the features, finding out the similarity and its measurement. In comparison to the conventional machine learning methods, the large-scale deep learning approaches have significantly provided much better and scalable performances considering image processing and accuracy. Learning hierarchical representations with a single or a few algorithms is challenging in nature and has mainly beaten records in image recognition techniques, natural language processing, semantic segmentation and several other tasks [3]-[8]. A large variety of deep learning models like Convolutional Neural Network (CNN) , Stacked Autoencoder [18] and Deep Belief Network(DBN) [19,20]. Among all these, CNN provides the most promising results in image recognition and classification tasks as they exploit the local context information using filters of different sizes. CNN is a type of neural network that is popularly used to extract features from the input data by employing various convolution techniques. LeCun initially proposed the central idea of CNN [9] and applied in the task of handwriting recognition. The idea of ImageNet [13] devised by Krizhevsky, Sutskever and Hinton has marked a revolution in computer vision. It is regarded as an influential publication in computer vision and showed that CNNs outperform recognition performances if we compare to handcrafted based methods. In this paper we have proposed a face recognition based on the features extracted by a convolutional neural network from the captured image of a voter. If the captured image's features match with those existing images stored in the database, the result is considered as positive and the voter is asked to cast vote for a political party. Once the vote has been casted, all the facial alignment details pertaining to the voter gets deleted from the system, so that if the same voter comes for voting, the system detects the maliciousness and prevents it instantly. In this paper, the above-stated security problem during elections in India has been dealt properly with face recognition based electronic voting system.

II. RELATED WORK

Face Detection is a relatively new concept. In 2001, Viola and Jones [21] first proposed a cascade Adaboost framework and made the face detection real- time. Off late, the Convolutional Neural Network has been used in computer vision and pattern recognition to achieve the same. Many CNN-based object detection methods which are proposed [22-28].[28] have improved the region proposal based CNN method and proposed the Faster R-CNN frame- work, this

framework introduced the anchors method and made region proposal a CNN classification problem, which could be trained in the whole net during the training stage. The end- to-end trainable Faster R-CNN network was faster and more powerful, which achieved 73% mAP in the VOC2007 dataset with VGG-net. [29] used Faster R-CNN framework making the face detection faster and achieved promising results. Most face recognition methods used aligned faces as the input, it had been shown that adopting alignment in the test stage could have 1% recognition accuracy improvement on the LFW [30] dataset. The usual way for face alignment was predicting facial landmarks from the detected facial patches, such as eyes, nose and mouth. And the geometric transformation between the positions of the predicted facial landmarks and the pre-defined landmarks was applied to the facial patches. The aligned faces with known face identities were then fed into the deep networks and were classified by the last classification layer for training the discriminative feature extractors, the intermediate bottleneck layer was taken as the representation.

A huge collection of face detection, verification and recognition is present. The contribution of [15,16,17]all use a complex system of multiple stages, which employ and merge the output of a deep convolutional network with PCA for dimensionality reduction and an SVM for classification. Zhenyaoet al. [17] employ a deep network to warp faces into a canonical frontal view and then learn CNN that classifies each face as belonging to a known identity. Taigman et al. [16] propose a multi-stage approach that aligns faces to a 3D shape model. They have trained a multi-class network for performing the face recognition task over a large number of entities. They have also taken care of a Siamese network for their experimental purpose where they directly optimize the L1-distance between two face features. They have achieved their best performance on the LFW dataset (97.35%) by using an ensemble of three architectures.

III. METHODOLOGY

Facial recognition is a biometric solution that measures unique characteristics about ones face. Face recognition identifies people from the characteristics of photos and videos. So a face recognizer will identify the facial features and compare the values with those present inside the knowledgebase. Some sort of feature similarity metric aids in computational similarity matching and henceforth the most dominant label is attached as the label of the image. If the value of the similarity metric is lower than a particular value, the classification result will return false,i.e the system will fail to recognize the person. This is how face recognition proceeds and if two persons are same, their feature similarity is too high.

This paper uses a deep CNN model for the purpose of feature extraction from the images. LFW dataset sample has been considered, then finally the dataset has been customized with the images of the valid voters taken into the account of our experiment. The following sections will demonstrate the two-level security authentication provided by the proposed system. First, the entire module of face recognition has been explained such as: a) overview of the deep CNN architecture, b) detection and transformation on input images which ensures that the faces are correctly aligned prior to feeding them into the CNN. c) Use the CNN to extract 200- dimensional representations, or embeddings, of faces from the aligned input images as Euclidean distance directly relates to a measure of face similarity d) finally compute the embeddings to formalize the similarity. The proposed workflow of the entire framework has been shown in the fig. 1.

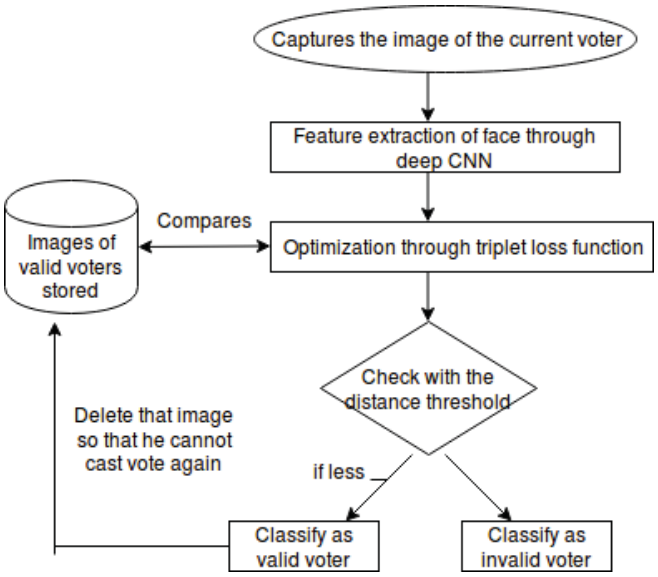

Fig. 1. *Flow chart of the proposed workflow*

A. Overview of CNN architecture for face recognition

The idea of our model is mostly derived from the Inception Architecture[2] of Google. The architecture can be described as : a multi-layer feed-forward layer consisting of 200 intermediate nodes with another L2 weighted regularization layer with convolutional layer in between.

As the model is getting trained with a triplet loss function, an embedding f(x)of image x such that the relative distance measured between different pairs of samples get maximized whereas that between similar pairs of samples gets minimized. The purpose of the loss function is to bring close the anchor record or the current image x^a_i and a positive image x^p_i and the negative record x^n_i (some other person) will be separated by some considerable large distance and the margin will be α.

$$L = \sum_{i=1}^{m} ||f(x_i^a) - f(x_i^p)||_2^2 - f(x_i^a) - f(x_i^n)||_2^2 + \alpha]_+$$

where [z]+ means max(z,0) and m is the number of triplets in the training set.

The model was not trained end-to-end from the scratch as it is quite expensive in terms of time and effort, thus, we have used the pre-trained models.

B. Facial Recognition

A subpart of the entire LFW dataset has been used to demonstrate the usage of the model. The dataset consists of 100 images of 10 people. Later, we have tested the system with the pre-captured images of the valid voters. The results and analysis has been shown in the following section. The face-detection, transformation and cropping has been done on the images.

The 200-dimensional embedding vectors are then calculated by feeding the aligned and scaled images into the pre-trained network. While calculating the distance between the images of same person, the distance must be relatively less than that between two different persons. That is the main utility of using a triplet loss function.

The same computation has been done using the custom dataset of the pre-captured images from the voters who are willing to cast their vote in the election. Thus, the test image i.e. the person currently willing to cast his vote is then compared with the database images, if the comparison score is below the optimal threshold then the voter is considered as valid and is allowed to undergo through the next authentication level i.e. the fingerprint recognition system.

IV. RESULTS AND ANALYSIS

The face recognition based electronic voting system has been evaluated on a standard dataset and a custom dataset prepared for the purpose of testing real-time.

A. Dataset Visualization

The dataset has been embedded into 2D clusters. Thus, t-SNE is applied to 200-dimensional embeddings. Distance between positive and negative pairs has been shown in the Fig. 2.

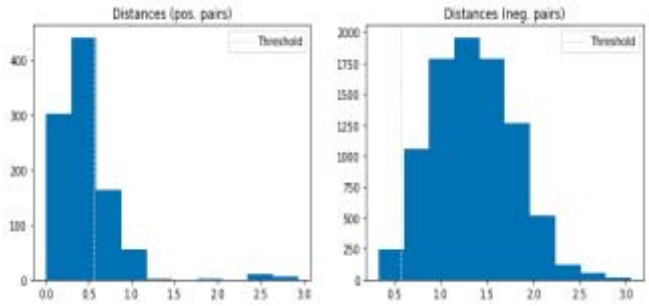

Fig. 2. *Distance between positive and negative pairs of the images*

B. Distance Threshold and accuracy measurement

This is very important to find the optimal value of the distance Threshold, T. Therefore, to solve this purpose, the face recognition model has been tested and verified over a large distribution of distance threshold measures. The threshold measure decides whether the embedding vectors will be classified as either positive (same person) or negative(different person), it has been shown in Fig .3.

The face verification accuracy at T = 0.56 is 95.7%.

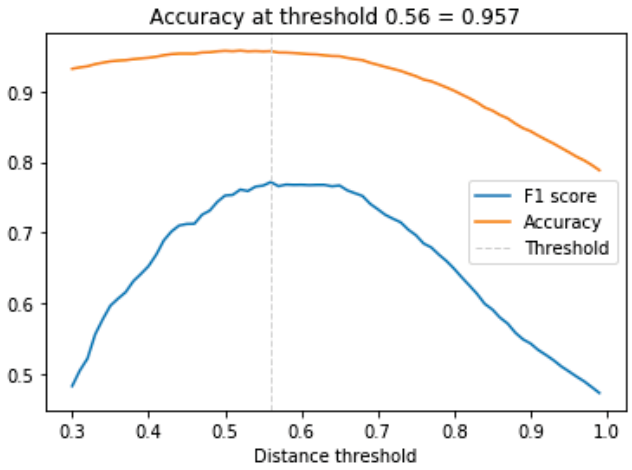

Fig. 3. *Measurement index vs Distance Threshold*

C. . Face Recognition

Now as we have an estimate of the distance threshold τ, it is easy to calculate face recognition as it is the distances between an input embedding vector and all embedding vectors in a database. The performance is quite robust in nature as it efficiently assigns the positive level to the entry in the database whose relative distance is the smallest. As a result, it achieves the highest probability.

The K-nearest neighbor classification when performed with the help of Euclidean distance metric labels the input data with the help of highest K scoring entries present inside the database. Again a linear kernel SVM classifier is trained with the database entries and can be made to perform well on the unseen data, such as to identify new entries.

While the model has been trained on non-parametric classifier like K-nearest neighbor, it gives 97% accuracy while the same has been done with the help of Support Vector Machine, it provides 99.1 % accuracy.

If the system can correctly identify the image of the voter, or in other words, if any of the aligned image of the voter is present inside the database, then the Electronic voting machine will allow the voter to cast his vote, after the voting has been done correctly, all the aligned images of that particular identity will be automatically detected from the database. As a result of this, the next time the voter comes to cast vote, he will not be recognized, thus the system will prevent him from voting more than once.

The classic results of the face recognition system has been tested both on LFW dataset and custom dataset. The Table. 1. shows the comparison of results on two different datasets.

TABLE I: Dataset Analysis

System	Accuracy(%)
LFW dataset	**99.1**
Custom dataset	98

V. EVM AS A SIMULATOR

The proposed method has been illustrated in a simulation and the outline has been described below. As a naive user who is willing to cast his vote in general elections in India, has to undergo through the following set of steps like:

*1)*Register his name before casting vote, or in other words, he should register himself by image capturing.This is the first step to recognize himself as a valid voter.

*2)*While casting vote, the voter again needs to capture his image, the electronic voting machine(EVM) takes some time to capture the image, verifies his identity through the above explained procedure.

*3)*If the face verification result turns out to be positive, or in other words, if the voter is a pre-registered candidate, then he will be asked to cast the vote for a political party, otherwise the result will show that his face cannot be identified and he will be informed that he is not allowed to cast vote.

*4)*If he is allowed to cast vote for the first time, all kinds of aligned pre-stored face images will be deleted from the database, so that he will not be able to vote more than once.

VI. CONCLUSIONS

Face recognition involves a large number of challenges when it comes to deal with the visual analysis. Recognizing face of human beings is quite a challenging task as those can exhibit a huge variation of characteristics of age, expression, nose, distance between two eyes etc. Thus, this development or approach plays a significant role in security applications such as legal documents identification, identification of terrorists in the public places like railway station, ports, shopping malls. Despite the fact that a large number of techniques have proved to work superiorly in terms of detection and recognition of human faces, still it remains challenging to develop a computationally efficient algorithm to match the human face with those present in the large database. Thus face recognition can be referred to as a superior computer vision task. A good number of machine learning models like SVM, random forest and other powerful classifiers like artificial neural networks have given some promising results to achieve a reasonable performance on this task. In this paper, a secure and hassle- free face recognition based electronic voting machine has been proposed which is intended to solve the tamparability and security issues faced during the elections in India. The proposed method has been used in real-time, runs perfectly on the standard benchmark dataset as well as the custom dataset prepared by us. The performance has been better compared to the traditional approaches just because of the deeper architecture of the convolutional neural networks. In the last two decades, a large number of government-owned companies have conducted the elections in India by allowing the voters to cast their vote through digital system of Electronic voting machine. The EVM devices are quite simple in design, easy to use, and reliable as well. But during these days, there have been reports of its irregular usage. But from the point of business and customer satisfaction, these digital devices have gained much popularity and our method promises to work as a standalone authentication system.

VII. FUTURE WORK

In the proposed system, we have experimented the system through one level of authentication biometric system. A face recognizer determines the characteristics of a person's image by capturing the same through a video camera. The entire structure has been analyzed including the proximity of the eyes, nose, jaws and mouth. Their relative distance serves as the unique feature to identify a person. Those measurements are kept inside a database and used to compare a user's identity when he stands before the camera. In near future, we would like to incorporate fingerprint recognition system as second level of authentication system to eradicate any form of security threats to general elections in India.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their valuable feedback.

REFERENCES

[1] Florian Schroff, Dmitry Kalenichenko, James Philbin.: FaceNet: A Unified Embedding for Face Recognition and Clustering.Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015. Publisher, Boston, MA (2015)

[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabi-novich: Going Deeper with Convolutions. Computer Vision and Pattern Recognition.

[3] M. Liang and X. Hu, Recurrent convolutional neural network for object recognition, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 33673375, 2015.

[4] P. Pinheiro and R. Collobert, Recurrent convolutional neural networks for scene labeling, Proc. 31st Int. Conf., vol. 32, no. June, pp. 8290, 2014.

[5] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang, DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection, in Proceedings of the IEEE Computer Society Conference on

Computer Vision and Pattern Recognition, vol. 0712, pp. 39823991, June 2015

[6] M. A. K. Mohamed, A. El-Sayed Yarub, and A. Estaitia, Automated Edge Detection Using Convolutional Neural Network, Int. J. Adv. Comput. Sci. Appl., vol. 4, no. 10, pp. 1117, 2013

[7] M. A. K. Mohamed, A. El-Sayed Yarub, and A. Estaitia, Automated Edge Detection Using Convolutional Neural Network, Int. J. Adv. Comput. Sci. Appl., vol. 4, no. 10, pp. 1117, 2013

[8] Dan Cire an, Deep Neural Networks for Pattern Recognition.

[9] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural Language Processing (Almost) from Scratch, J. Mach. Learn. Res., vol. 12, pp. 24932537, 2011.

[10] R. Collobert and J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, Proc. 25th Int. Conf. Mach. Learn., pp. 160167, 2008.

[11] E. Shelhamer, J. Long, and T. Darrell, Fully Convolutional Networks for Semantic Segmentation, IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640651, 2017.

[12] Y. LeCun, Backpropagation Applied to Handwritten Zip Code Recognition, Neural Comput., vol. 1, no. 4, pp. 541551, Dec. 1989

[13] A. Krizhevsky, I. Sutskever, and H. E. Geoffrey, ImageNet Classification with Deep Convolutional Neural Networks, Adv. Neural Inf. Process. Syst. 25, pp. 19, 2012.

[14] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In S. Thrun, L. Saul, and B. Schlkopf, editors, NIPS, pages 4148. MIT Press, 2004. 2

[15] Y. Sun, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. CoRR , abs/1406.4773, 2014. 1, 2, 3

[16] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In IEEE Conf. on CVPR , 2014. 1, 2, 5, 7, 8, 9

[17] Z. Zhu, P. Luo, X. Wang, and X. Tang. Recover canonical view faces in the wild with deep neural networks. CoRR , abs/1404.3543, 2014. 2

[18] R. Xia, J. Deng, B. Schuller, and Y. Liu, Modeling gender information for emotion recognition using Denoising autoencoder, in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings, pp. 990994, 2014.

[19] G. E. Hinton, S. Osindero, and Y. W. Teh, A fast learning algorithm for deep belief nets, Neural Comput., vol. 18, no. 7, pp. 15271554, 2006.

[20] Y. Bengio, Learning Deep Architectures for AI, vol. 2, no. 1, 2009.

[21] P. Viola and M. Jones. Robust real-time face detection. In Proc. Eighth IEEE Int. Conf. Computer Vision. ICCV 2001 , volume 2, page 747, 2001.

[22] J. Dai, Y. Li, K. He, and J. Sun. R-fcn: Object detection via region-based fully convolutional networks. arXiv , 2016.

[23] R. Girshick. Fast r-CNN. In Proc. IEEE Int. Conf.Computer Vision (ICCV), pages 14401448, Dec. 2015.

[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-based convolutional networks for accurate object detection and segmentation.IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(1):142158, Jan. 2016

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed,C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector.arXiv, 2015.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pages 779788, June 2016.

[27] J. Redmon and A. Farhadi. Yolo9000: Better, faster,stronger.arXiv, 2016.

[28] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-CNN: Towards real-time object detection with region proposal networks.IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99):1, 2016.

[29] H. Jiang and E. Learned-Miller. Face detection with the faster r-cnn.arXiv, 2016

[30] G. B. Huang, M. Ramesh, T. Berg, and E. Learned Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.