Towards a Practical Real-Time Applications of Face verification

Mohamed E. Elhamahmy[1] Mohamed.M.A. Elgazzar[2] Abdel-Hamid M. Emara[3,4]

[1]Dr., Information Technology, mezzat1967@yahoo.com, Cairo, Egypt.
[2]Asso.Prof., m_elgazzar1961@yahoo.com,Higher Institute of Computer Science and Information Systems, Fifth Settlement, New Cairo, Egypt.
[3]Department of Computers and Systems Engineering, Faculty of Engineering, Al-Azhar University, Cairo, Egypt.
[4]Department of Computer Science, Taibah University, Medinah, Kingdom of Saudi Arabia, abdemara@yahoo.com.

Corresponding author: [2]Asso.Prof., Mohamed Elgazzar, m_elgazzar1961@yahoo.com

Abstract— In the last few years, the deep learning of face recognition has achieved good and satisfied results. Deep learning techniques used a deep convolution neural network (CNN) for face recognition is extremely complex because of the dataset size and the requirement of a high performance computing power for training as well as testing the data set. However, the learning outcomes transferred between different parties facilitated to have a large scale training dataset. Once learning outcomes transferred many applications may be implemented based on the pre-trained weights of the deep learning model. The face verification is one of the applications of the face recognition. Face verification application is used to verify whether the input image is related to the authorized candidate. In this paper, the transfer learning concept is maintained to build a real-time face verification application. It is aimed to make use of the already trained data set weights in a real-time application. The application is built in Keras with Tensorflow back end using LFW dataset with an open deep face recognition model. The one-shot training is applied using opencv with a traditional web cam. The result was promising, as the accuracy was 99% for comparing one image against 100 pre-saved images in real-time. It augmented the transfer learning concept in utilizing the deep learning results.

Keywords-Deep learning; face verification; one shot learning; real-time verification; transfer learning.

Introduction

Differentiation between face verification and face recognition is frequently confounded. Face recognition is a general theme that incorporates both face verification and face identification. Face recognition technique is used for proofing whether the input face image is from a set of an authorized group of individuals. However, face verification technique is used for comparing a face image to another one in order to proof the authorization claims of someone. So that, face verification technique may be considered as one-to-one mapping. on the other hand, face recognition technique may be considered as one-to-many mapping. As the face recognition is a general theme, so the fulfilled consequences of deep learning inspired the more inquires about in face verification. Since deep learning includes various dimensions of portrayal and different layers of non-direct preparing units (or neurons), it has appeared to be fitting to depict them as deep. By differentiate; all non-deep learning methodologies can be qualified of shallow learning. Also, it happens that the absolute first neural network architectures, which had not the learning limit of the present deep neural networks, had just several hidden layers, all the more regularly only one. Convolutional Neural System (CNN) is a deep learning architecture. In neural systems, Convolutional neural system (ConvNets or CNNs) is one of the fundamental classifications to do pictures acknowledgment, pictures groupings. CNN picture classification takes an information picture, process it and group it under specific classifications. Computer sees an info picture as cluster of pixels and it relies upon the picture goals. Actually, deep learning CNN models to prepare and test, each info picture will go it through a progression of convolution layers. Convolution is the primary layer to extricate highlights from an info picture. Convolution saves the connection between pixels by learning picture highlights utilizing little squares of information. It is a numerical activity that takes two data sources, for example, picture network and a channel. The first proposed frameworks of CNN architectures neglected to perform well on complex issues because of the absence of vast preparing information and computational power. Since the coming of GPGPUs and their utilization in machine learning, the field of CNN has experienced a renaissance stage [1]. A few distributions have set up progressively productive approaches to prepare convolutional neural systems utilizing GPU registering. For example, GoogleNet [2] and VGGNet [3], have appeared enhance its execution. Oxford visual geometry a mass reported its profound face recognition architecture VGGNet. It can perceive several pictures simply applying transfer learning. Transfer learning is an examination issue in machine learning that centers around putting away learning picked up while taking care of one issue and applying it to an alternate yet related issue. In this way, learning outcomes transferred between different parties. Furthermore, one doesn't need to have a large scale training dataset once learning outcomes transferred.

A. Challenges of face verification problem

Deep learning tasks usually expect to be fed multiple instances of a custom class to learn (e.g. lots of pictures of someone). This makes face recognition task satisfactory because training should be handled with limited number of instances – mostly one shot of a person exists. Moreover, adding new classes should not require reproducing the model. Various new thoughts were consolidated including utilizing different CNNs, a Bayesian learning system to prepare a measurement, perform multiple tasks learning over face verification [4]. The goal is to achieve the best performance on LFW "Labeled Faces in the Wild" dataset. Facebook researchers named their proposed face recognition model DeepFace [5]. Trained on a data set of some 4 million photos "from a popular social network," the DeepFace model uses 3D modeling techniques and artificial neural networks to recognize similarities between two images of the same person. There is a need for representing the images as fingerprints values in order to differentiate between them. The fingerprints computed by the autoencoder are vectors of real numbers and those vectors can be compared using cosine distance [6]. In face verification application, it is needed to verify a person's face image giving a few images. Though most machine learning based article arrangement calculations require preparing on hundreds or thousands of tests/pictures and huge datasets, one-shot learning expects to take in data about item classes from one, or just a couple, preparing tests/pictures. Also, the images taken for the person's face to be verified is needed to be represented in a vector. Opencv reads and elaborates images as numeric structures, or 2-dimensional vector. Accordingly, there are varieties of techniques to select from and integrate between them in order to build a real time application for face verification.

B. Problem Statement

How to utilize the good results achieved by deep learning architecture models without need to retrain the huge dataset in order to produce a real-time application? Should we need to know the construction of the connected neural networks when using the pre-trained weights? The problem is how to apply the transfer learning as a concept in order to simply build a wide range of applications. So that, learning outcomes transfers between different parties.

I. RELATED WORKS

In last few years, there are many researches on the face recognition field. Google researchers proposed their face verification model named FaceNet [7]. FaceNet achieved nearly 100-percent accuracy on a popular face recognition dataset named "Labeled Faces in the Wild (LFW)". DeepFace model employs a nine-layer deep neural net, and was trained on four million images uploaded by Facebook users [8]. Oxford visual geometry group (VGG) announced its deep face recognition architecture (VGG-Face) [4]. VGG-Face model is deeper than Facebook's DeepFace, it has 22 layers and 37 deep units. These mentioned models need high computing power as well as a long time for training their proposed models. However, long preparing time because of the high computational complexity keeps PC vision calculations from being practically utilized in versatile applications [9]. OpenCV (and other related APIs) have made it simpler for application engineers to utilize PC vision. They are all around reported and energetic open source extends that continue developing, and they are being adjusted to new processing advancements [9, 10, 11]. This paper aimed to share the way for creating the next generation of facial recognition systems using machine learning. Another challenge is how to extract features from normal pictures. Convolutional Auto-Encoder (CAE) as an unsupervised learning model was proposed for extracting various leveled highlights from normal pictures [12]. CAE is like a conventional auto-encoder aside from it utilizes convolutional layers for the concealed layers in the system. Moreover, it is conceivable to utilize the weights learned by a (stacked) CAE to introduce the weights in a convolutional neural network for classification tasks.

II. THE PROPOSED APPROACH

Machine learning researchers, especially face recognition ones, would like to share outcomes. They might spend a lot of time to construct a neural networks structure, and train the model. The models take a long time for training. They might also run learning process on highly cost hardware such as GPUs and parallelized systems. However, one can run the same model in seconds if he has the pre-constructed network structure and pre-trained weights. So, this paper focused on the usability of the pre-trained models. In order to make use of already trained models of face recognition, as they specified a good results lately. The application of face verification is different from those of face recognition. Face verification applications focus on test a face image against another for similarity. However, face recognition applications interested in building a model with certain algorithm and prepare datasets for training using that algorithm. On other hand, the face verification applications are based on the results of the face recognition models. The real-time application of face recognition relies on the one-shot training. One-shot training in face recognition mean that it is needed just one image related to the person and no more just one image for recognition. So that, it may use a camera such as mobile camera or laptop camera or either a web cam for the one-shot based applications. One case study for applying the concept of learning transfer is the Institute Gate Access. There is a need for application of face verification is to access an institute gate by students. The gate is dedicated for students only. So, there is a need for an application to verify students before allowing them to pass through the gate. Figure (1) shows the modules of the required application. The student should pass through a camera that takes a picture of his face. Then the image should pass to the face verification module. The face verification module will reply with "yes" for similarity or "no" for no similarity. It has the images of the students pre-saved. The VGG-Face model is pre-trained weights model related to face recognition. So that, the similarity is just a distance calculation between two vectors represent two images. One of these images is the entered image come from the "camera module" and the other related to one of the pre-saved images related to students that allowed passing through the gate. It should test one-to-many images, the one related to that came from the camera module against the images that pre-saved related to the allowed students.

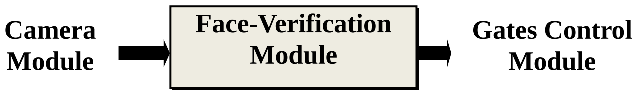


Figure (1): The proposed main modules of the required application for face verifications

The result of "similarity-test" is passed to the "Gates Control" module. If the result was "Yes" it should allow the student to pass through, and if it was "No" it should not allow him to go through.

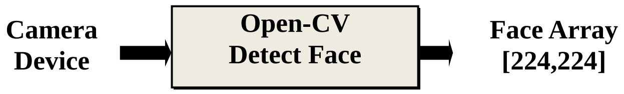


Figure (2): The proposed "Camera" module components

As shown in Figure (2), the used camera is a traditional one, it is any traditional web cam such as that built-in the available laptop. So, there no special requirements of the image quality. However, the image should be taken in sufficient light. OpenCV is used for extracting the features from the captured image. It produces a two-dimensional array [224,224] representing the image vector.

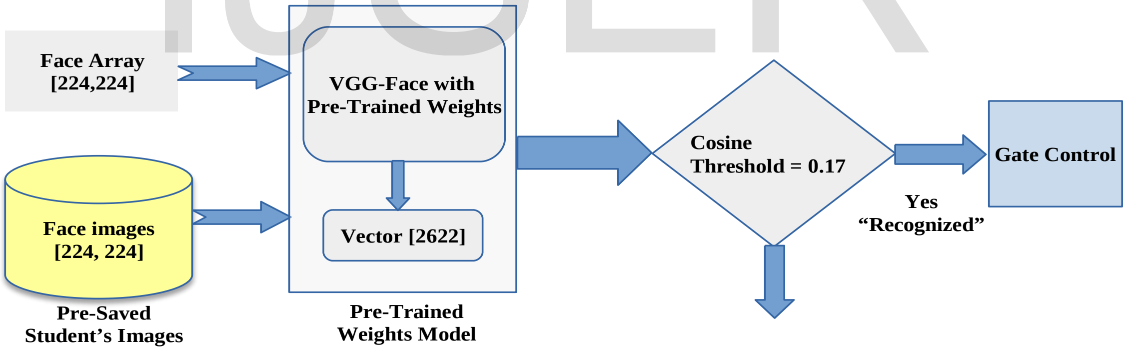


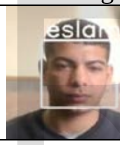
Figure (3): The proposed "Face Verification" module components.

Oxford visual geometry group (VGG) has announced its deep face recognition architecture [1]. It can be recognize hundreds of images just applying transfer learning concept. Basically, in this paper, the transfer learning will be applied and make use of pre-trained weights of VGG Face model. What's more, we will consume the model as auto-encoder to represent images as vectors. As shown on the above Figure (3), the face array [224, 224] will be passed to the pre-trained weights model. The VGG pre-trained model will represent the image as a vector [2622]. The images should be taken off the pre-saved student's images one-by-one in a loop until the last saved image. Each image should be passed to the pre-trained model and represented as a vector [2622] as well. The Cosine algorithm for calculating the distance is used to calculate the distance between the two vectors. One of them is regarded the taken image of the student against the other one regarded an image taken from the pre-saved images database. The threshold value is considered 0.17. It is experimented that it is the best value for verification accuracy. The similarity result, only, should be passed to the next module "Gate Control" if the calculated Cosine distance is less than or equal to 0.17.

III Results:

The implementation of the proposed work is maintained by a laptop system type x64-based pc. os name: Microsoft windows 10 home edition – version: 10.0.17134 build 17134. Processor Intel(r) core(tm) i5-7300hq CPU is 2.50 GHz, 2501 MHz, 4 core(s), 4 logical processor(s). Installed physical memory (ram) is 8.00 GB. Total physical memory is 7.87 GB. Available physical memory: 1.64 GB. Total virtual memory: 19.4 GB. Available virtual memory: 3.08 GB. Page file space: 11.5 GB. As shown in table (1), a sample of 4 students is trying to access the gate. Student is declined as the cosine distance of his face image is greater than the threshold value. However the others are passed as their images have the cosine distance less than the threshold. It is cleared that there are faces are much similar with respect to the human sense. Such as the person named "Shehab", and the person named "Shaban" at the last column. However the cosine distance of the vectors represented their images were greater than the threshold value (0.17). In spite of the low quality images, in addition to they are shouted once by the web cam, the proposed model succeeded to verify them accurately. The main target of the proposed application is to present to which extent it can be utilized the "transfer learning" concept. The importance of sharing the pre-trained data sets for the research community in order to maximize the usability of these data. The average time required for calculating the vector related to the input image is about 0.428 seconds. The average time required for calculating the cosine difference between the image-vector related to the input image against that pre-saved is about 0.003 seconds. The proposed model is evaluated for face verification of one input image against 100 pre-saved images. The accuracy was 99%, as the proposed model verified all input face-images except one image. However, the experiment is repeated for the person who did not verified. The threshold value is adopted at 0.17 for cosine similarity, then he accurately verified.

Table (1): The cosine distance value for a sample of 4 students

	Images taken by a web cam at the gate			
				
Pre-Saved Images	Declined	Passed	Passed	Passed
	0.3314738869667053	**0.08947175741195679**	0.4153124690055847	0.2977548837661743
	0.17165738344192505	0.4647138714790344	**0.09387362003326416**	0.28989458084106445
	0.40253347158432007	0.38822972774505615	0.2891366481781006	**0.15313202142715454**

IV. Conclusion and Future Work

Transfer learning concept may resulted in a wide applications that utilize the good results satisfied by deep learning. One of these applications is the face verification problem. In this paper, the transfer learning is applied and make use of pre-trained weights of VGG Face model. The VGG face model is model as auto-encoder to represent images as vectors. The integration between OpenCV, VGG face model, gate control system and the traditional web cam proved the concept of transfer learning. The threshold value is adopted more than once just to improve the face verification accuracy. As shown above the threshold satisfied good results at the value (0.17) for calculating the cosine similarity. The time required for verification process is accepted for 10, 20, and 100 persons. It was no more than 2 seconds starting from taking the face image until the verification results come out. The real-time application based on the the pre-trained weights of the deep learning model has an accepted results and can be relied on. In future work, it is planned to make use of another pre-trained model on another application.

REFERENCES

[1] Ashwin Bhandare, Maithili Bhide, Pranav Gokhale, and Rohan Chandavarkar. 2016. Applications of convo-lutional neural networks.International Journal of Computer Science and Information Technologies, pages2206–2215.

[2] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[3] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556(2014).

[4] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. Proceedings of the British Machine Vision 1(3), 6 (2015)F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A uni-fied embedding for face recognition and clustering. In Proc. CVPR, 2015.

[5] K. Grm, V. Struc, A. Artiges, M. Caron, H. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations", IET Biometrics, vol. 7, no. 1, pp. 81-89, 2017.

[6] L. Xiong, J. Karlekar, J. Zhao, J. Feng, S. Pranata, and S. Shen. A good practice towards top performance of face recognition: Transferred deep feature fusion. arXiv preprint arXiv:1704.00438, 2017.

[7] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognitionand clustering. InProc. CVPR, 2015

[8] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface:Closing the gap to human-level performance in face verifica-tion. InIEEE Conf. on CVPR, 2014

[9] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov,"Real-time computer vision with OpenCV,"Communi-cations of the ACM, vol. 55, no. 6, pp. 61–69, 2012.

[10] A. Mohammadi, S. Bhattacharjee, S. Marcel, "Deeply vulnerable: a study of the robustness of face recognition to presentation attacks", IET Biometrics, vol. 7, no. 1, pp. 15-26, 2018.

[11] A. RoyChowdry, T. Lin, S. Maji, E. Learned-Miller, "One-to-many face recognition with bilinear CNNs", WACV, 2016.

[12] J. Masci, U. Meier, D. Cireşanet al., "Stacked convolutional auto-encoders for hierarchical feature extraction," inArtificial Neural Net-works and Machine Learning–ICANN 2011. Springer, 2011, pp. 52–59.

AUTHOR'S PROFILE

FIRST AUTHOR: Dr. Eng. Mohamed Ezzat Elhamahmy is a cyber-technology expert. He is from Egypt. He has received his Ph. D. in computer networks security from Faculty of Computers and information, Cairo University, 2011. He has received his B.Sc. in Computer Engineering from MTC, Cairo in 1989. He received his M. Sc. in Computer Engineering, Faculty of Engineering, Al-Azhar University, Cairo in 2001. His past research was in building text to speech synthesis system in Arabic. His current research interests are Network Security and Machine Learning.

SECOND AUTHOR: He received the B.Sc. degree in Electronic and Communications from MTC, Cairo, Egypt, in 1984 and M.Sc. degree in Computer Networks from Faculty of Engineering, Cairo University, Egypt, in 1991 and the Ph.D. degree in Computer Networks at Faculty of Engineering, from Cairo University, Egypt, in 1995. He has experience of 33 years which includes both academic and research. He is currently an Associate Professor in Computer Science Department at Higher institute of Computer Science and information systems fifth Settlement, New Cairo, Egypt. His research interests Computer Networks, Network Security and Machine Learning. He published several research papers.

THIRD AUTHOR: Abdel-Hamid Emara received the B. S., M. S., and Ph.D. degree in computers engineering from Al-Azhar University in 1992, 2000, 2006, respectively. He works in Computers and Systems Engineering Department at Faculty of Engineering, Al-Azhar University, Cairo, Egypt. He has experience of 12 years which includes both academic and research. He is currently an Assistant Professor in Computer Science Department at the Taibah University, Al Madinah Al Monawarah, KSA. His research interests educational data mining, Machine Learning, Arabic text mining, and intelligent, and adaptive systems. He published several research papers.