Design of a Face Recognition System based on Convolutional Neural Network (CNN)

Yahia Said
Electrical Engineering Department
College of Engineering
Northern Border University
Arar, Saudi Arabia and
Laboratory of Electronics and
Microelectronics
University of Monastir
Monastir, Tunisia
said.yahia1@gmail.com

Mohammad Barr
Electrical Engineering Department
College of Engineering
Northern Border University
Arar, Saudi Arabia
m.abarr@hotmail.com

Hossam Eddine Ahmed
Electrical Engineering Department
College of Engineering
Northern Border University
Arar, Saudi Arabia and
Faculty of Engineering
Benha University
Benha, Egypt
hossamhalas@gmail.com

Abstract—**Face recognition is an important function of video surveillance systems, enabling verification and identification of people who appear in a scene often captured by a distributed network of cameras. The recognition of people from the faces in images arouses great interest in the scientific community, partly because of the application interests but also because of the challenge that this represents for artificial vision algorithms. They must be able to cope with the great variability of the aspects of the faces themselves as well as the variations of the shooting parameters (pose, lighting, haircut, expression, background, etc.). This paper aims to develop a face recognition application for a biometric system based on Convolutional Neural Networks. It proposes a structure of a Deep Learning model which allows improving the existing state-of-the-art precision and processing time.**

Keywords-face recognition; biometrics; Convolutional Neural Networks (CNNs); artificial intelligence; deep learning

I. INTRODUCTION

Facial recognition consists of automatically identifying one or more people in photos or videos by analyzing and comparing shapes. Typically, face recognition methods extract the facial characteristics of individuals and compare them to a stored database in order to find possible matches. Face recognition has many applications in security, biometrics, robotics, image search by content and surveillance systems, and in image and video indexing systems. For a human being, analyzing and recognizing an immense amount of details about a visual scene is an easy task. However, for a computer this is a very hard task that needs a lot of computation resources and memory.

Deep Learning is based on neural networks and especially Convolutional Neural Networks (CNN). Deep Learning models are neural networks with a deep structure. Deep Learning is inspired by the human brain system and by simulating it, it aims to find a way to solve general learning problems. Deep Learning techniques are a huge success in the field of computer vision. They have been deployed in many applications such as traffic sign detection and identification [1, 2], indoor object detection and recognition [3-5] and many other applications [6, 7]. The recognition of faces is a big challenge and an interesting research subject for different fields: psychology, model identification, computer vision, computer graphics, etc. That's why the relevant literature is vast and diverse.

Authors in [8] presented a long-distance face recognition method, solved by resolving the variation in recognition rate resulting from distance variation. A CNN was used for face recognition and the Euclidean distance was used to measure the similarity. The proposed method achieved excellent performance in various distances, compared to old face recognition methods. Authors in [9] introduced a hybrid system for face recognition by combining a Logistic Regression Classifier (LRC) and a Convolutional Neural Network (CNN). The CNN was trained to localize and identify faces in images, and the LRC was used to classify the features learned by the convolutional network. The reported results on the Yale face dataset [10] showed an improvement in the classification accuracy and a reduction in processing time. Authors in [11] proposed a CNN-based face identification system composed of nine layers: three convolution layers, two pooling layers, two full-connected layers, and one Softmax layer. The proposed CNN was tested on the ORL face [12] and AR face datasets [13]. The obtained results show that the proposed network achieves a higher recognition rate than the traditional machine learning methods and other handcrafted feature methods for face identification. Authors in [14] detailed the implementation of Deep Learning algorithm for face recognition. The proposed algorithm was implemented based on the OpenFace project using FaceNet neural network architecture [15]. The results showed the effectiveness of the incremental learning algorithm for performance improvement. An Active Face Recognition system (AcFR) was proposed in [16]. It deploys a CNN and acts consistently with human behavior in common face recognition scenarios. A pre-trained VGG-Face CNN was used

to extract facial image features, then a nearest-neighbor identity recognition criterion was used for the identification task. The evaluation of the proposed CNN on the CMU PIE face dataset [17] proved that the recognition stage of the AcFR system is more powerful than alternative systems'.

Authors in [18] proposed a new face recognition system using a deep C2D-CNN model decision-level. The proposed method was tested in the case of big differences between the test and the training sets. A novel CNN model was proposed with speed up convergence process and reduced training time, achieving better performance than the state-of-the-art methods. Authors in [19] proposed a deep learning model that improved face identification performance by predicting facial attributes. The proposed model was based on a CNN with two output heads. Experimental results proved that this method was better than the existing face identification and attribute prediction methods. Authors in [20] proposed a Deep CNN with multiple inputs: visible light images and near-infrared images. To generate predictions, the authors fused the information loss strategy and the nearest neighbor algorithm. The experimental results proved that it was very robust against illumination and performed much better than other state-of-the-art models. Authors in [21] presented a survey of different face recognition techniques and methods that claimed to provide effective and accurate face recognition systems. During their investigation of face detection approaches, they implemented a face recognition system using a Deep Neural Network. The performance enhancement found was better than some reported works. An experimental evaluation of the performance of CNN against three well-known image recognition methods, PCA, LBPH, and KNN proved that CNN outperforms them [22]. The experimental results on the ORL dataset [12] demonstrated the effectiveness of CNN-based methods on face recognition. The proposed CNN obtained the best face recognition accuracy of 98.3%. In this paper, a CNN structure for face recognition system is proposed.

II. PROPOSED MODEL FOR FACE RECOGNITION SYSTEM

The trainable Deep Learning system consists of a series of modules, each representing a processing step. Each module is trainable, with adjustable parameters similar to the weights of the linear classifiers. The system is trained from start to finish: in each example, all the parameters of all the modules are adjusted in order to bring the output produced by the system closer to the desired output. The deep classifier comes from the arrangement of these modules in successive layers. In its most common realization, a Deep Learning architecture can be seen as a multilayer network of simple elements, similar to linear classifiers, interconnected by trainable weights. This is called a multilayer neural network. The advantage of deep architectures is their ability to learn to represent the world in a hierarchical manner. As all layers are trainable, there is no need to build a feature extractor by hand. The training will take care of that. In addition, the first layers will extract simple characteristics (presence of contours) that the following layers will combine to form increasingly complex and abstract concepts: assemblies of contours into patterns, patterns into parts of objects, parts of objects in objects, etc. CNNs are designed to automatically extract the characteristics of input images. It is invariant to

slight image distortions and implements the concept of weight sharing allowing considerably reducing the number of network parameters. This weight sharing also allows taking a strong account of the local correlations contained in an image.

A CNN architecture is formed by a stack of independent processing layers:

- The convolution layer (CONV) which processes the received input data.

- The pooling layer (POOL), which allows compressing the information by reducing the size of the intermediate image (often by subsampling).

- The activation layer, often misused as 'ReLU' with reference to the activation function (Rectification Linear Unit).

- The "Fully Connected" (FC) layer, which is a perceptron type layer.

- The classification layer (Softmax) that predicts the class of the input image.

In this work, a CNN is proposed for face identification. The proposed network is composed of two convolution layers, a fully connected layer, and a classification layer. Each convolution layer is followed by an activation layer and a max-pooling layer. Also, we add two regularization techniques after each convolution layer: batch normalization and dropout. After the fully connected layer, we apply the dropout technique to reduce the complexity of calculations and to enhance the performance of the proposed CNN.

The main aim of this work is to identify faces in a biometric system with grayscale images. So, the size of input images for the proposed CNN was fixed to 32x32x1, where the 1 refers to the grayscale space color. For the first convolution layer, the kernel size was fixed to 3x3 and the number of filters was fixed to 16. The number of filters was reduced since the input image is a grayscale image with fewer features to be learned. For the activation layer, we choose the ReLU function as it is the most used in CNNs [23] and has proved efficiency in comparison to other activation functions. The ReLU helps avoiding generating negative values if the input image is damaged and avoiding saturation of the neurons by making the mapping function more flexible and non-linear. For the max-pooling layers, the kernel size was fixed to 2x2 and the stride is 2. In the second convolution layer, the number of filters was doubled to 32 and the same kernel size of the first convolution layer was used. After each block of convolution, activation and max pooling, a batch normalization technique was applied. Batch normalization is a regularization technique used to accelerate the training process by allowing the use of a high-value learning rate while it guarantees the convergence of the network. Also, batch normalization helps initializing the weights of the CNN especially if it is trained from scratch. In addition, batch normalization reduces internal covariant shift and the dependence of gradients on the scale of the parameters or their initial values. Since the input image does not have many features to be learned, adding a dropout technique to avoid the overfitting problem is proposed. The dropout

technique is used to eliminate neurons with weak connections in order to make focus on neurons with strong connections to enhance performance. The dropout reduces the number of connections and the number of neurons resulting in reduced computation complexity. Regularization techniques, batch normalization and dropout are applied only in the training process in order to achieve high performance.

A fully connected layer was added to the network in order to summarize and combine learned features. The dropout technique was applied after the fully connected layer to ovoid overfitting problem. As output layer, the Softmax function was used to compute the probabilities of each class. The sum of the probabilities of all classes must be equal to one. The Softmax is computed as (1):

$$p(y_j|x) = \frac{e^{x^T w_j}}{\sum_{i=1}^{k} e^{x^T w_i}} \quad (1)$$

where x represents the input data, y_j is output of the neural network for class j and w_j, w_i are the weight of the neuron of position i, j. Figure 1 presents the proposed CNN in detail. The proposed neural network has 40 outputs according to the training dataset that contains 40 classes to be identified.
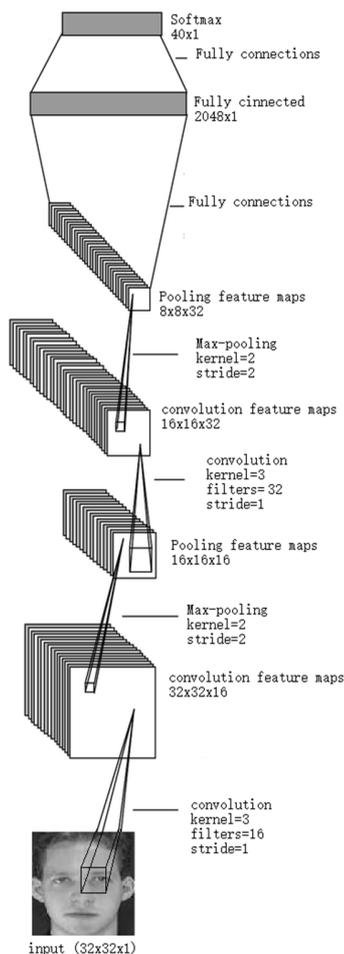

Fig. 1.　　Proposed CNN for face recognition

III.　EXPERIMENTS AND RESULTS

For all experiments, we used a desktop with an Intel i7 CPU and an Nvidia GTX960 GPU. All the algorithms were developed on the Python programming language using the Tensorflow Deep Learning framework and the OpenCV library. To train the proposed CNN for face recognition, we used the ORL dataset [12], built by the AT&T Laboratories of Cambridge University. The faces presented in the dataset were captured between April 1992 and April 1994. The dataset was used in order to build a face recognition project. The dataset contains 40 different faces, and each face is considered as a distinct class. For each class there are 10 images with a size of 92x112 pixels and 256 grey levels per pixel. Figure 2 presents an image from each class of the ORL dataset. The images were organized in 40 different directories where each directory contains 10 images of the same class.
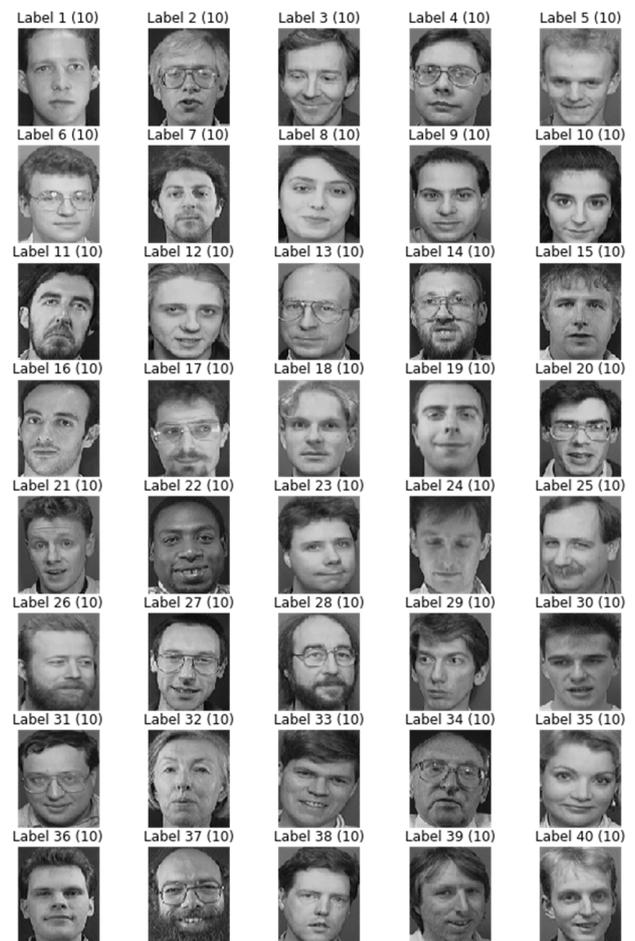

Fig. 2.　　Classes of the ORL dataset [12]. © 2001 AT&T Laboratories Cambridge

The dataset contains 400 .pgm images. All images contain faces in frontal view with an upright or with a slight left-right rotation. Figure 3 shows all the images of the ORL dataset. The dataset was divided into a training set and a testing set. The training was set to 6 images for each class and 4 images were used for the testing set.
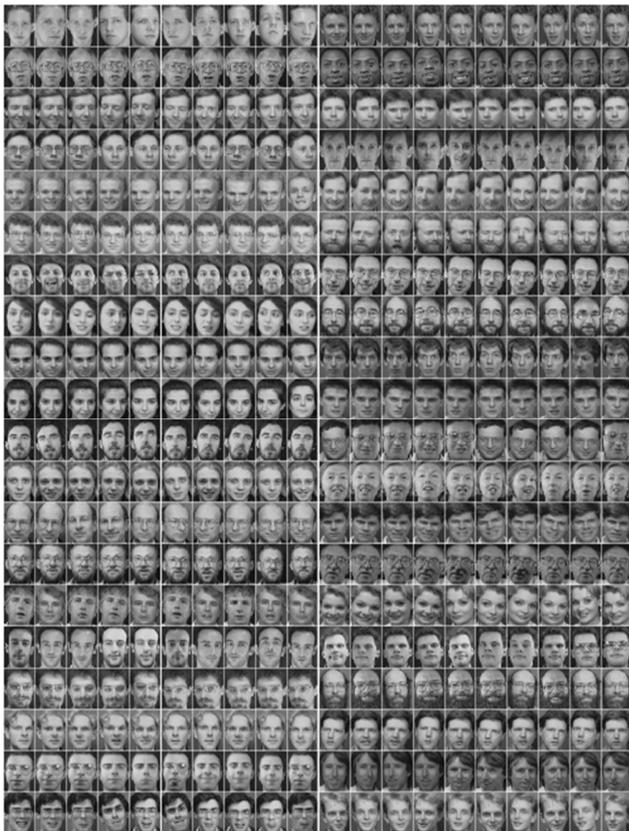
Fig. 3.   ORL dataset images [12]. © 2001 AT&T Laboratories Cambridge

After exploring the data, the Categorical Crossentropy as a loss function was used for the proposed CNN. Generally, this function is used for single label classification tasks, which means that each input image must belong to one output class. It gives an idea about how wrong the prediction of the neural network is. The Categorical Crossentropy can be computed as:

$$L(y, \hat{y}) = -\sum_{j=0}^{m} \sum_{i=0}^{n} (y_{ij} * \log(\widehat{y_{ij}})) \quad (2)$$

where y is the target class and $\hat{y}$ is the predicted class.

To optimize the proposed loss function, the Adam gradient descent algorithm was utilized. The Adam algorithm takes advantage of the momentum acceleration and the adaptive gradient descent algorithm methodology for weights update. Also, it updates the learning rate automatically in order to achieve better performance. Weights update by the Adam algorithm can be computed as:

$$\widehat{m_t} = \frac{m_t}{(1-\beta_1^t)}$$

$$\widehat{v_t} = \frac{v_t}{(1-\beta_2^t)}$$

$$w_{t+1} = w_t - \frac{\vartheta}{\sqrt{\widehat{v_t}}+\varepsilon} \widehat{m_t} \quad (3)$$

where $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. w_{t+1} presents the updated weights and w_t the old weights, while $\widehat{m_t}$ is the bias corrected first moment and $\widehat{v_t}$ is the the bias corrected second moment.

After defining the loss function and the optimization algorithm, the proposed CNN was trained on the ORL dataset for 20 epochs, with 400 iterations in each epoch. The loss optimization and the accuracy curves are presented in Figure 4. The loss function reaches a minimum value of 0.12 on the testing set (validation). The proposed CNN achieves a training accuracy of 99.78%, a validation accuracy of 98.7%, and an inference speed of 231 frames per second.
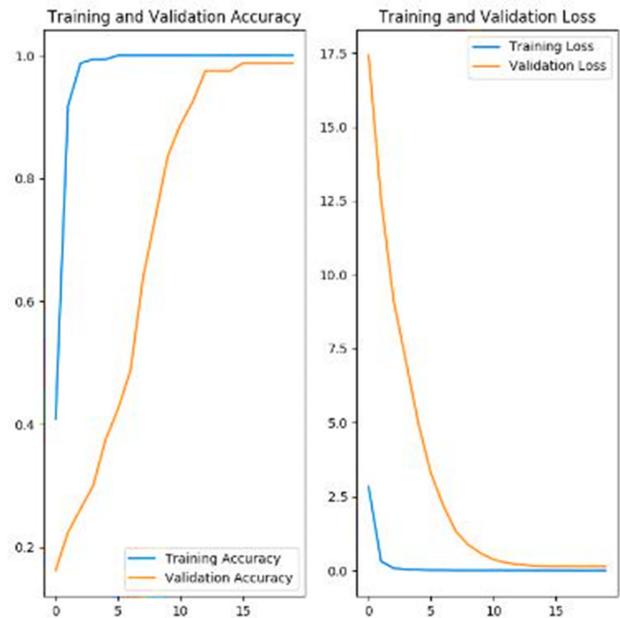

Fig. 4.   Curves of the accuracy and the loss function optimization

The obtained results show the efficiency of the proposed CNN for face identification in biometric systems. In order to prove the performance of the proposed method, we compare it against the state-of-the-art methods. Table I presents a comparison between the state-of-the-art methods for face identification with the proposed one in terms of accuracy. We can see that the proposed method achieves state-of-the-art accuracy performance.

TABLE I.   FACE RECOGNITION METHODS COMPARISON ON [12]

Method	Accuracy (%)
[22]	98,3
Eigenface [24]	97,5
ICA [24]	97,75
[25]	98,3
Proposed	98.75

IV.   CONCLUSION

Face identification is one of the most important tasks for applications such as video surveillance. In this work, a face identification application based on Convolutional Neural Networks is proposed. The proposed CNN achieves high accuracy performance. As future work, the proposed network must be optimized for embedded implementation.

REFERENCES

[1] R. Ayachi, M. Afif, Y. Said, M. Atri, "Traffic signs detection for real-world application of an advanced driving assisting system using deep learning", Neural Processing Letters, Vol. 51, pp. 837-851, 2020

[2] R. Ayachi, Y. E. Said, M. Atri, "To perform road signs recognition for autonomous vehicles using cascaded deep learning pipeline", Artificial Intelligence Advances, Vol. 1, No. 1, pp. 1-58, 2019

[3] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, M. Atri, "Indoor image recognition and classification via deep convolutional neural network", 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, Maghreb, Tunisia, December 18-20, 2018

[4] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, M. Atri, "An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation", Neural Processing Letters, Vol. 51, pp. 1-15, 2020

[5] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, M. Atri, "Indoor object classification for autonomous navigation assistance based on deep CNN model", IEEE International Symposium on Measurements & Networking, Catania, Italy, July 8-10, 2019

[6] D. Virmani, P. Girdhar, P. Jain, P. Bamdev, "FDREnet: Face detection and recognition pipeline", Engineering, Technology & Applied Science Research, Vol. 9, No. 2, pp. 3933-3938, 2019

[7] U. Khan, K. Khan, F. Hasssan, A. Siddiqui, M. Afaq, "Towards achieving machine comprehension using deep learning on non-GPU machines", Engineering, Technology & Applied Science Research, Vol. 9, No. 4, pp. 4423-4427, 2019

[8] H. M. Moon, C. H. Seo, S. B. Pan, "A face recognition system based on convolution neural network using multiple distance face", Soft Computing, Vol. 21, pp. 4995-5002, 2017

[9] H. Khalajzadeh, M. Manthouri, M. Teshnehlab, "Face recognition using convolutional neural network and simple logistic classifier", Advances in Intelligent Systems and Computing, Vol. 223, pp. 197-207, 2014

[10] Yale Face Database, available at: http://vision.ucsd.edu/content/yale-face-database

[11] K. Yan, S. Huang, Y. Song, W. Liu, N. Fan, "Face recognition based on convolution neural network", 36th Chinese Control Conference, Dalian, China, July 26-28, 2017

[12] AT&T Database of Faces: ORL face database, available at: http://cam-orl.co.uk/facedatabase.html

[13] A. Martinez, R. Benavente, The AR face database, technical report series, CVC, 1998

[14] L. Li, Z. Jun, J. Fei, S. Li, "An incremental face recognition system based on deep learning", Fifteenth IAPR International Conference on Machine Vision Applications, Nagoya, Japan, May 8-12, 2017

[15] F. Schroff, D. Kalenichenko, J. Philbin, "Facenet: A unified embedding for face recognition and clustering", 28th IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, June 7-12, 2015

[16] M. Nakada, H. Wang, D. Terzopoulos, "AcFR: Active face recognition using convolutional neural networks", IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, July 21-26, 2017

[17] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, "Multi-pie", Image and Vision Computing, Vol. 28, No. 5, pp. 807-813, 2010

[18] J. Li, T. Qiu, C. Wen, K. Xie, F. Q. Wen, "Robust face recognition using the deep C2D-CNN model based on decision-level fusion", Sensors, Vol. 18, pp. 1-27, 2018

[19] F. Taherkhani, N. M. Nasrabadi, J. Dawson. "A deep face identification network enhanced by facial attributes prediction", IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, June 18-23, 2018

[20] K. Guo, S. Wu, Y. Xu, "Face recognition using both visible light image and near-infrared image and a deep network", CAAI Transactions on Intelligence Technology, Vol. 2, No. 1, pp. 39-47, 2017

[21] A. Shraddha, A. Agrawal, "Face recognition with partial face recognition and convolutional neural network", International Journal of Advanced Research in Computer Engineering & Technology, Vol. 7, No. 1, pp. 91-94, 2018

[22] P. Kamencay, M. Benco, T. Mizdos, R. Radil, "A new method for face recognition using convolutional neural network", Digital Image Processing and Computer Graphics, Vol. 15, No. 4, pp. 663-672, 2017

[23] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, Vol. 25, No. 2, pp. 1097-1105, 2012

[24] M. H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods", 5th International Conference on Automatic Face Gesture Recognition, Washington, USA, May 21-21, 2002

[25] J. Yang, D. Zhang, A. F. Frangi, J. Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition", IEEE Transactionson Pattern Analysis and Machine Intelligence, Vol. 26, No. 1, pp. 131-137, 2004