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

Yahia Said[1,2], Mohammad Barr[1] and Hossam Eddine Ahmed[1,3]

[1]Electrical Engineering Department, College of Engineering,
Northern Border University, Arar, Saudi Arabia
[2]Laboratory of Electronics and Microelectronics (LR99ES30),
Faculty of Sciences of Monastir, University of Monastir, Tunisia
[3]Benha Faculty of Engineering, Benha University, Egypt

Summary
Face recognition is an important function of video surveillance systems to enable the verification and identification of people who appear in a scene captured by a distributed network of cameras. The recognition of people from the faces in images arouses great interest in the scientific community; this is 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's about proposing a structure of a Deep Learning model which makes it possible to improve the precisions existing in the state of the art and the processing time by images.
Keywords:
Face Recognition, Biometrics, Convolutional Neural Networks, Artificial Intelligence, Deep Learning.

1. Introduction

In all areas of security and access control, we use passwords or keys that consist of numbers or letters. But, in recent times with the advancement of technology these passwords have become easily falsifiable and passable. This is why researchers from different fields have focused their works on keys and passwords that are impossible to forge, safe and above all effective. Biometrics are required areas that need high level of security and control. Face recognition is an adapted technology for the existing biometric systems.

Facial recognition consists of identifying one or more people automatically 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, as well as in image and video indexing systems.

Face recognition remains the most acceptable technique since it is similar to visual interaction used by humans. Face identification is more advantageous in comparison to other applications. It is a method without intrusion, that does not require any action from the individuals observed from a distance. Second, the data collection sensors are cheap (an ordinary camera) unlike other applications such as the fingerprint where the individuals must be close to the sensor and must cooperate for data acquisition. In addition, the acquisition equipment necessary for fingerprint are very expensive. Although some say that face recognition is a relatively unreliable biometric, the acquired signal is subject to much greater variations than other characteristics, such as the variation of lighting, the change of position facial, the presence or absence of glasses and others. But in recent years several image processing techniques have emerged, such as face detection, lighting standardization, and so on. Not to mention the considerable development of digital camera technologies, which neglects the effect of these problems.

The face recognition task is a high-level visual task for a human being. In effect, humans can detect and identify faces with a simple galance, but building an artificial system for face identification is a serious challenge. Variable conditions can make the challenge harder. For face identification there is two main challenges that must be solved: inter-class and intra-class variation. Inter-class variation is less challenging because of the shape similarity between individuals. On the other hand, the intra-class variation is much harder. It can be due to many factors. For each different image, individual face can generate a wide variety. This great diversity complicates the analysis of face images. Also, variations in the appearance of face images pose great problems to the identification.

For a human being, analyzing and recognizing an immense amount of details about the 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. Recently, a new technique named Deep Learning brings intelligence to computers. Deep Learning is based on neural networks and especially Convolutional Neural

Manuscript received March 5, 2025
Manuscript revised March 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.3.23

Networks (CNN) that made great success on computer vision field. It is an important deep learning model for image and video processing. Deep Learning models are neural networks with a deep structure. It was inspired by the simulation of the human brain system that aims to find a way to solve general learning problems. Deep Learning techniques are a huge success in the field of computer vision. Mostly known for their classification performance, these techniques can also be used in image enhancement, object detection, tracking, multi-camera analysis, and 3D reconstruction. It has been deployed for many applications such as traffic signs detection and identification [13], [14], indoor object detection and recognition [15], [16], [17] and many other applications.

In this paper, we propose a Convolutional Neural Network structure for face recognition system. The remainder of the paper is organized as follows. In Section 2, related works on face detection and recognition are presented. The proposed structure for the face recognition system is described in section 3. In Section 4, experiments and results are detailed. Finally, Section 5 concludes the paper.

2. Related Works

The recognition of faces is a big challenge and interesting research subject from different fields: psychology, identification of models, computer vision, computer graphics ... That's why the literature is so vast and diverse. We present below, some related works on implementations of face recognition structures and the obtained results.

Moon et al. [1] presented a long-distance face recognition method. That was solved by resolving the variation in recognition rate resulting from distance variation. They fixed the issue arising from distance change through the method of standardizing the size of face images and the use of face images by distance as training data for the CNN. The CNN was used for face recognition and the Euclidean distance was used to measure the similarity. The proposed method was evaluated through the tests and it achieves excellent performance in various distances, compared to old face recognition methods.

In [2], Khalajzadeh et al. have 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. To train the proposed hybrid system, a training algorithm was developed based the gradient descent algorithm. The evaluation of the proposed architecture was performed after training the network to recognize faces completed on the Yale face dataset [20]. The reported results show an improvement in the classification accuracy and the processing time was reduced.

Yan et al. [3] proposed a face identification system based on Convolution Neural Network (CNN). The network was composed from 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 [21] and AR face datasets [22]. The obtained results show that the proposed network achieves a higher recognition rate than that of the traditional machine learning methods and other handcrafted features methods for face identification.

In [4], Lufan et al. detailed the implementation of Deep Learning algorithm for face recognition. The proposed algorithm was implemented based on the OpenFace project using FaceNet neural network architectures [23]. They proposed an intelligent model training method using an incremental SVM algorithm named S-DDL (self-detection, decision, and learning). The reported results show the effectiveness of the incremental learning algorithm for performance improvement.

An active face identification system AcFR is proposed in [6]. It deploys a Convolutional Neural Network and acts consistently with human behaviors 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 proposed recognition system achieved higher accuracy on images acquired at angles more than those saved in memory. The evaluation of the proposed CNN on the CMU PIE face dataset [24] proved that the recognition stage of the AcFR system is more powerful than alternative systems.

Jing et al. [8] proposed a new face recognition system using a deep C2D-CNN model decision-level. The proposed method was tested in case of having a big difference between the test and the training set to solve face recognition task. A novel CNN model was proposed and they speed up convergence process and reduced the training time by using a normalization layer in the network. Also, a new activation function was used to adapt the activation function for the normalized data. Finally, probabilistic max-pooling was applied to preserve the feature representations in order to maximize extent while maintaining feature invariance. The proposed method solves recognition accuracy degradation because of the difference between test and training datasets. The reported results show that the proposed method achieves better performance compared to the state-of-the-art methods.

In [9], Taherkhani et al. proposed a deep learning model that improve face identification performance by predicting facial attributes. The proposed model is based on a Convolutional Neural Network (CNN) with two output heads; the first head predicts facial attributes while the second head was used for face identification in images.

Experimental results proved that this method was better than the existing face identification and attribute prediction methods.

To overcome challenge in face recognition applications, Guo et al. [10] proposed a Deep Convolutional Neural Network with multiple inputs; a 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 prove that the proposed network model is very robust against illumination and perform much better than other state-of-the-art models that deals with the same challenge of illumination changes.

Arya et al. [7] presented a survey of different face recognition techniques and methods that claimed to provide an effective and accurate face recognition systems. During an investigation of the face detection approaches in this work, the authors have implemented face recognition system using Deep neural network. As reported, the performance enhancement found was better than some works reported in the literature.

An experimental evaluation presented in [5] of the performance of Convolutional Neural Network (CNN) against three well-known image recognition methods such as PCA, LBPH, and KNN proved that CNN outperforms the state of the art methods. The experimental results on the ORL dataset [21] demonstrated the effectiveness of methods based on CNN for face recognition. The proposed CNN have obtained the best face recognition accuracy of 98.3 %.

3. 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 so as 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.

Convolutional Neural Networks (CNNs) designed to automatically extract the characteristics of input images. It is invariant to slight image distortions and implements the concept of weight sharing allowing to considerably reduce the number of network parameters. This weight sharing also makes it possible to take 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 (ReLU), often misused as 'ReLU' with reference to the activation function (Rectification Linear unit).

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

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

In this work, we propose a Convolutional Neural Network (CNN) 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 technique after each convolution layer, we apply batch normalization and dropout techniques. After the fully connected layer, we apply the dropout technique to reduce the complexity of calculation and to enhance the performance of the proposed convolutional neural network.

The main task in this work, is to identify faces in a biometric system. Images in the biometric system are grayscale. So, as input images for the proposed CNN, we fix the size of images 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 and there are few features to be learned. For the activation layer, we choose the ReLU function as it was the most used in the convolutional neural network since the breakthrough of using a CNN for image classification [18] and proving the efficiency of the ReLU function in comparison to other activation functions. The ReLU helps to avoid generating negative values if the input image is damaged and to avoid the 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 and guaranteeing the convergence of the network. Also, batch normalization helps to initialize the weights of the Convolutional Neural Network especially if it was 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 do not have many features to be learned, we propose to add a dropout technique to avoid the overfitting problem. The dropout technique is used to eliminate neurons with weak connections in order to make focus on neurons with strong connections to enhance performance. Figure 1 present an example of dropout technique application on a multi-layers neural network. The dropout reduces the number of connections and the number of neurons resulting in a reduce of computation complexity.

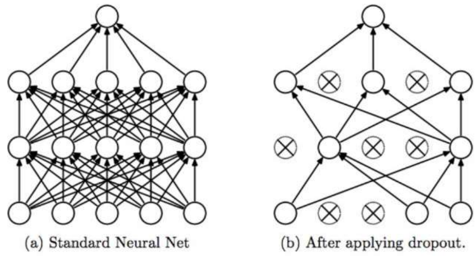

Fig. 1 Dropout technique applied to a neural network

Both 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 leaned features and to combine them. A 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).

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

Where x is 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 2 present the proposed Convolutional Neural Network with details. The proposed neural network has 40 output according to the training dataset that contains 40 class to be identified.

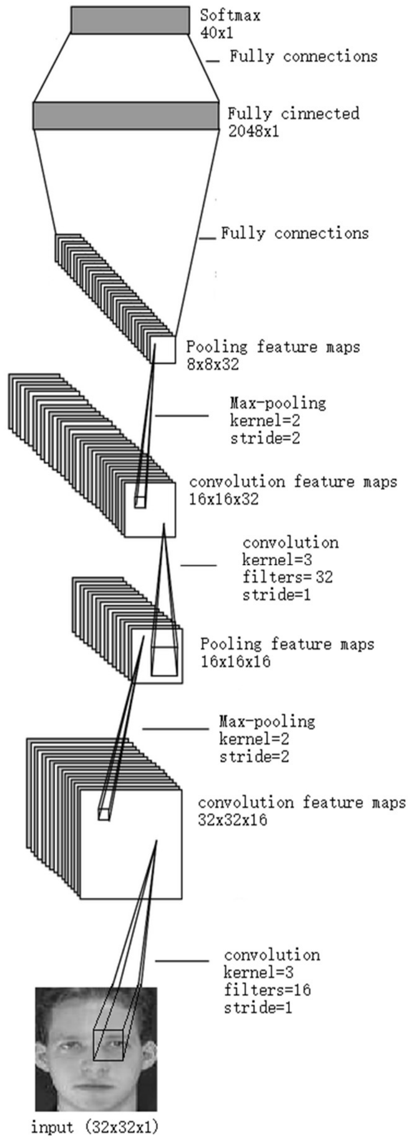

Fig. 2 Proposed CNN for face recognition

4. Experiments and results

For all experiments, we used a desktop with an Intel i7 CPU with 32 GB of Ram 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 Convolutional Neural Network for face recognition, we propose the use of the ORL dataset [21]. The ORL dataset is builded by the AT&T Laboratories of Cambridge university. The faces presented in dataset were captured of researchers of the AT&T Laboratories between April 1992 and April 1994. The dataset was used in order to build a face recognition project. The dataset contains 40 different faces each face is considered as distinct class. For each class there is 10 images with a size of 92 x 112 pixels and 256 grey levels per pixel. Figure 3 present 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. 3 Classes of the ORL dataset

The dataset contains 400 images with the .pgm extension. All images contain faces in frontal view with an upright or with a slight left-right rotation. Figure 4 present all the images of the ORL dataset. The dataset was divided into training set and testing set. We set the training set to six images for each class and four images for the testing set. After exploring the data, we propose to use the Categorical Crossentropy as a loss function for the proposed Convolutional Neural Network. Generally, this function is used for single label classification tasks means each input image must belong to one output class. It gives an idea about how wrong the prediction of the neural network. The Categorical Crossentropy can be computed as (2).

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

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

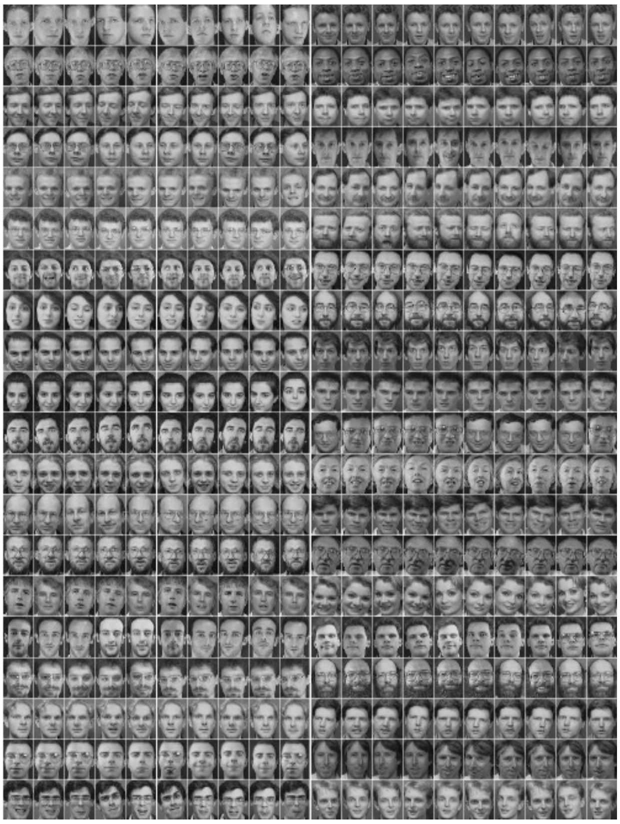

Fig. 4 ORL dataset images

To optimize the proposed loss function, we propose the use of the Adam gradient descent algorithm. The Adam algorithm take advantage of the momentum acceleration and the adaptive gradient descent algorithm methodology for weights update. Also, the Adam updates the learning rate automatically in order to achieve better performance. Weights update using the Adam algorithm can be computed as (3).

$$\widehat{m_t} = \frac{m_t}{(1-\beta_1^t)}$$
$$\widehat{v_t} = \frac{v_t}{(1-\beta_2^t)}$$
$$w_{t+1} = w_t - \frac{\vartheta}{\sqrt{\widehat{v_t}}+\varepsilon}\widehat{m_t} \qquad (3)$$

Where $\beta_1 = 0.9$,$\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. w_{t+1} is the updated weights and w_t is the old weights. $\widehat{m_t}$ is the bias corrected first moment and $\hat{v}_t$ 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 in each epoch 400 iteration. The curves of loss optimization and of the accuracy are presented in figure 5. 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% and a validation accuracy of 98.7%, and an inference speed 231 frames per second.

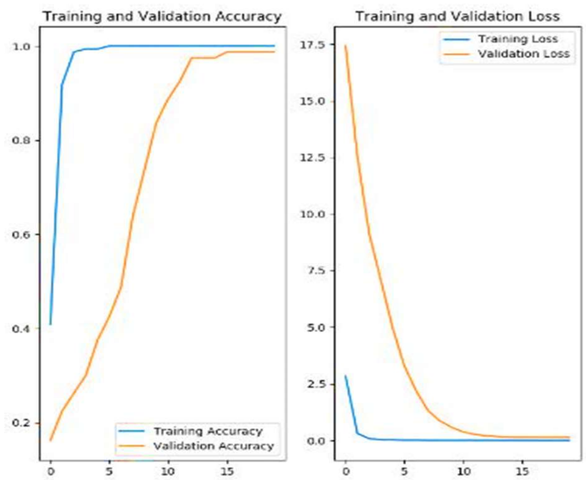

Fig. 5 Curves of the accuracy and the loss function optimization

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 state of the art methods. Table 1 present a comparison between state of the art methods for face identification in terms of accuracy. Based on the reported results, the proposed method achieves state of the art accuracy performance.

Table 1. Comparison of face recognition methods on ORL database

Approach	*Accuracy (%)*
Kamencay et al. [5]	98,3
Eigenface [11]	97,5
ICA [11]	97,75
2DPCA [12]	98,3
Our's	98.75

6. Conclusion

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

Acknowledgments

The authors wish to acknowledge the approval and the support of this research study by the grant N° ENG-2018-3-9-F-7880 from the Deanship of the Scientific Research in Northern Border University, Arar, KSA.

References

[1] Mo. Hae-Min, C. H. Seo, and S. B. Pan, “A face recognition system based on convolution neural network using multiple distance face”, Soft computing, Vol. 21, No. 17, pp. 4995-5002, 2017.
[2] K. Hurieh, M. Mansouri and M. Teshnehlab, “Face recognition using convolutional neural network and simple logistic classifier”, Soft Computing in Industrial Applications, Springer, Cham, pp. 197-207, 2014.
[3] Y. Kewen, et al., “Face recognition based on convolution neural network”, 36th IEEE Chinese Control Conference (CCC), 2017.
[4] L. Lufan, et al., “An incremental face recognition system based on deep learning”, Fifteenth IEEE IAPR International Conference on Machine Vision Applications (MVA), 2017.
[5] K. Patrik, et al., “A new method for face recognition using convolutional neural network”, Digital Image Processing and Computer Graphics, Vol. 15, No. 4, 2017.
[6] N. Masaki, H. Wang and D. Terzopoulos, “AcFR: active face recognition using convolutional neural networks”, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017.
[7] A. Shraddha and 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, 2018.
[8] J. Li, T. Qiu, C. Wen, K. Xie and F. Q. Wen, “Robust face recognition using the deep C2D-CNN model based on decision-level fusion”, Sensors, Vol. 18, No. 7, 2018.
[9] F. Taherkhani, N. M. Nasrabadi and J. Dawson. “A Deep Face Identification Network Enhanced by Facial Attributes Prediction”, arXiv preprint arXiv:1805.00324 (2018).
[10] G. Kai, S. Wu, and 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.
[11] M. H. Yang, “Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods”, Fgr, Vol. 2, p. 215, 2002.
[12] J. Yang, D. Zhang, A. F. Frangi, et a1., “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.
[13] R. Ayachi, M. Afif, Y. Said, and M. Atri, "Traffic Signs Detection for Real-World Application of an Advanced Driving Assisting System Using Deep Learning", Neural Processing Letters, pp. 1-15, 2019.
[14] R. Ayachi, Y. Said, and M. Atri, "To Perform Road Signs Recognition for Autonomous Vehicles Using Cascaded Deep Learning Pipeline", Artificial Intelligence Advances.
[15] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, and M. Atri, "Indoor image recognition and classification via deep convolutional neural network", International conference on the Sciences of Electronics, Technologies of Information and Telecommunications, Springer, Cham, pp. 364-371, 2018.
[16] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, and M. Atri, "An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation", Neural Processing Letters, pp 1-15, 2020.
[17] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, and M. Atri, "Indoor Object C1assification for Autonomous Navigation Assistance Based on Deep CNN Model", IEEE International Symposium on Measurements & Networking (M&N), pp. 1-4, 2019.
[18] K. Alex, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, pp. 1097-1105, 2012.
[19] I. Sergey, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167,2015.
[20] Yale face dataset. http://vision.ucsd.edu/content/yale-face-database
[21] AT&T Database of Faces: ORL face database, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[22] A. M. Martinez and R. Benavente, The AR Face Database. CVC Technical Report #24, June 1998. http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html
[23] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering", Proceedings of the IEEE conference on computer vision and pattern recognition, pp.815-823, 2015.
[24] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie", Image and Vision Computing, Vol. 28, No. 5, pp.807-813, 2010.

Yahia Said received the B.Sc., M.Sc., and Ph.D. degrees in Electronics from the Faculty of Science of Monastir, Monastir University, Tunisia in 2008, 2010, and 2016 respectively. He worked as an Assistant Professor at Faculty of Science of Monastir, Monastir University, Tunisia. Now, he is an Assistant Professor at Electrical Engineering Department, College of Engineering, Northern Border University, Arar, Saudi Arabia. His research interests are Pattern Recognition, Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Image and Video Processing, Embedded Vision and SoC Design.

Mohammad Barr received the M.Sc., and Ph.D. degrees in Electronic Engineering from Faculty of Technology, De Montfort University-Leicester, United Kingdom, in 2013 and 2018, respectively. He is an Assistant Professor at Electrical Engineering Department, College of Engineering, Northern Border University, Arar, Saudi Arabia. His research interests are Image and Video Processing, Digital Signal Processing and Embedded Systems.

Hossam Eddine Ahmed received the B.SC and M.Sc. degrees in Electronics and Communications from Faculty of Engineering-Benha University, Egypt, in 1996 and 2004, respectively and the Ph.D. degree in Electronics and Communications from Faculty of Engineering, Ain Shams University, Egypt, in 2010. He worked as an Assistant Professor at Faculty of Engineering-Benha University, Egypt. Now, He is an Assistant Professor at Electrical Engineering Department, College of Engineering, Northern Border University, Arar, Saudi Arabia. His research interests are Digital Signal Processing, Frequency Hopping Spread Spectrum, Embedded Systems and Digital Communications.