

Earprint recognition using deep learning technique

Arwa H. Salih Hamdany¹, Aseel Thamar Ebrahim², Ahmed M. Alkababji³

^{1,2}Department of Computer Engineering, Northern Technical University Mosul, Iraq

³Department of Computer Engineering, University of Mosul, Iraq

Article Info

Article history:

Received May 1, 2020

Revised Oct 25, 2020

Accepted Nov 11, 2020

Keywords:

Biometric

Deep learning

Earprint

Recognition

ABSTRACT

Earprint has interestingly been considered for recognition systems. It refers to the shape of ear, where each person has a unique shape of earprint. It is a strong biometric pattern and it can effectively be used for authentications. In this paper, an efficient deep learning (DL) model for earprint recognition is designed. This model is named the deep earprint learning (DEL). It is a deep network that carefully designed for segmented and normalized ear patterns. IIT Delhi ear database (IITDED) version 1.0 has been exploited in this study. The best obtaining accuracy of 94% is recorded for the proposed DEL.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Arwa H. Salih Hamdany,
Department of Computer Engineering
Northern Technical University Mosul, Iraq
Email: arwahamid78@ntu.edu.iq

1. INTRODUCTION

Recognizing people is one of the most important field in security systems. It starts from early stage in humans' life. Basically, individuals were started to be recognized by using their genders, names, ages and nationalities. Then, this matter has been further developed where specific documents have been established for each person in order to provide a clear identity. Examples of these documents are passports and identity documents (IDs). Classical recognition systems that consider ID cards, password and personal identification number (PIN) are not sufficient for reliable identification. Because they can easily be forged, forgotten, misplaced, stolen, or shared [1]. On the other hand, biometric characteristics can electronically and automatically recognize individuals [2]. Generally, biometric characteristics can be classified into physiological biometrics and behavioural biometrics. Physiological biometrics refer to the physiological characteristics within the people's body. Behavioural biometrics points to the behavioural characteristics of people's manner [3]. Physiological characteristics are often more reliable and accurate than the behavioural characteristics as the behavioural of humans may be influenced by the emotional feelings like tension or sickness [3]. Examples of physiological biometrics are iris, fingerprint, face and earprint, and examples of behavioural biometrics are voice, gait and signature [4, 5]. Earprint is a type of physiological biometric. It principally refers to the outer ear shape. It differs between humans, twins and identical twins. Moreover, ear shapes differ between left and right ears [6]. Figure 1 shows the various earprint features.

The aim of this paper is proposing a DL model for earprint recognition. This model is called the deep earprint learning (DEL), this model using Adam optimization to determine the best parameters of convolution and pooling layers to obtain the best error if compare with other training optimization methods

have been examined for the DEL network such as stochastic gradient descent with momentum (SGDM), and root mean square propagation (RMSProp) [7, 8]. The remaining sections are distributed as follows: section 2 provides the literature review of this paper, section 3 describes the DEL method, section 4 discusses the results and section 5 declares the conclusion. A limited number of studies considered the earprint as a type of recognition in the literature. In 2015, automatic recognition systems based on ensemble of local and global earprint features was explored, promising performance was concluded for considering both local and global earprint features [9]. In 2016, a new feature extraction approach was illustrated for the ear geometry recognition. In this approach, both the minimum and maximum ear height lines were employed, then, three ratio-based features were highlighted to enhance the scale of robustness [10]. In the same year, a decision-making of sparse coding-induced was employed with the earprint. It was proved that fusing both residuals and coefficients components can obtain better performances [11]. In 2018 combined different deep convolutional neural network models and analyzed in depth the effect of ear image quality [12]. In the same year, a framework of earprint recognition was described for a light field imaging. A new lenslet light field ear database (LLFEDB) method was illustrated by utilizing the richer spatio-angular features [13]. In 2019, a multi-modal biometric recognition method was explained, where earprint and finger knuckle print (FKP) were used. Techniques of local binary pattern (LBP) and feature level fusion (FLF) were exploited in this study [14]. In the same year, a new approach for a single earprint was proposed. It consists of three phases: providing normalization process, applying a novel Eigenears and utilizing nearest neighbour classifier [15]. In this paper, exploiting a DL technique for earprint recognition is considered. Therefore, a DEL technique is proposed and evaluated.

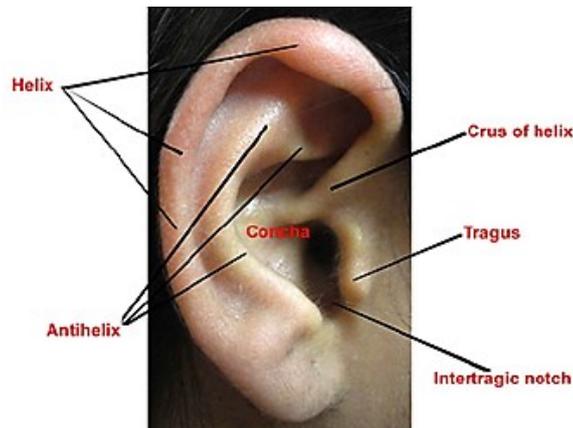


Figure 1. Various earprint features

2. RESEARCH METHOD

In this work, we are proposing the DEL network. It is a DL technique and a type of convolutional neural network (CNN). It is designed to accept earprint patterns. Firstly, the DEL network can be trained with various earprint patterns that are acquired from different persons. Then, the DEL network is tested for new earprint patterns that have not been seen before. The training stage will be resulted by producing useful values (weights). These values can be stored in order to be used later in the testing stage. Figure 2 shows the general DEL framework structure. DEL network consists of multi-layers. These are: the earprint input, convolution, rectified linear unit (ReLU) layer, pooling, fully connected (FC), softmax and classification layers. The architecture of the DEL network is given in Figure 3. The input layer is adapted for the earprint patterns. It accepts grayscale two-dimensional (2D) images. Thus, each earprint image E involves only one channel.

Regarding the convolution layer, the input image E is transformed into group of feature maps. The feature maps are convolved input image by a kernel weights matrix. The following equation represents the convolution process of the convolution layer [16]:

$$c_{v,u,z^{la}} = \beta z^{la} + \sum_{j=-ke_h^{la}}^{ke_h^{la}} \sum_{i=-ke_w^{la}}^{ke_w^{la}} \sum_{z^{la-1}=1}^{z^{la-1}} W_{j+ke_h^{la}, i+ke_w^{la}, z^{la-1}}^{z^{la}} c_{v+j, u+i, z^{la-1}} \quad (1)$$

where: $c_{v,u,z^{la}}$ is convolution layer output, (v, u) is the coordinate of a specific pixel, ke_h and ke_w are the kernels of height and width, respectively, βz^{la} is a bias, z^{la} is the channel of a specific layer, $W_{j,i,z^{la-1}}^{z^{la}}$ is a

kernel parameter, and la and $la-1$ are current and previous layers, respectively. ReLU process can be described by the following equation [17]:

$$q_{v,u,z^{la}} = \max(0, c_{v,u,z^{la}}) \tag{2}$$

where q_{u,v,c^l} is a ReLU output? Subsequently, the pooling layer further decreases the sizes of previous channels. The following equation represents the pooling computation [18]:

$$o_{x^{la},y^{la},z} = \underset{0 \leq x < p_h, 0 \leq y < p_w}{\text{ope}} q_{x^{la} \times p_h + x, y^{la} \times p_w + y, z} \tag{3}$$

where $o_{x^{la},y^{la},z}$ is a pooling output, $0 \leq x^{la} < p_h^{la}$, p_h^{la} is a pooled channel height, $0 \leq x^{la} < p_w^{la}$, p_w^{la} is a pooled channel width, $0 \leq z < p_z^{la} = p_z^{la-1}$, p_z^{la} is a pooled channel matrix, ope is the maximum operation, p_h is a sub-channel height and p_w is a sub-channel width. Hence, FC layer can match between the pooling neurons and required recognizing subjects. The following equation demonstrates the FC operation [19]:

$$F_r = \sum_{x=1}^{n_1^{la-1}} \sum_{y=1}^{n_2^{la-1}} \sum_{z=1}^{n_3^{la-1}} W_{x,y,z,d,r}^{la} (\mathbf{O}_z)_{x,y}, \forall 1 \leq r \leq m^l \tag{4}$$

where F_r is a FC output, n_1^{la-1} is the prior channel height of, n_2^{la-1} is the prior channel width, n_3^{la-1} is the number of prior channels, $W_{x,y,z,d,r}^{la}$ is a connection weight between FC and pooling, \mathbf{O}_z is the vector/vectors of pooling layer outputs, and n^{la} is the number of required classes. Now, Softmax transfer function can be computed as follows [20, 21]:

$$b_r = \frac{\exp(F_r)}{\sum_{su=1}^{n^{la-1}} \exp(F_{su})} \tag{5}$$

where b_r is a softmax output? Hereafter, the classification layer is required. The following equation can be considered [2]:

$$D_r = \begin{cases} 1 & \text{if } b_r = \max \\ 0 & \text{otherwise} \end{cases}, \quad r = 1, 2, 3, \dots, m \tag{6}$$

where D_r is a classification output, \max is obviously a maximum b_r value and m is the number of classes.

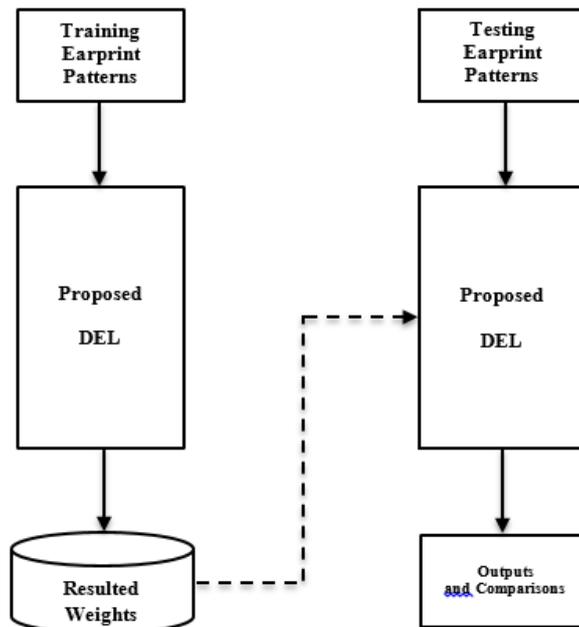


Figure 2. General DEL framework structure

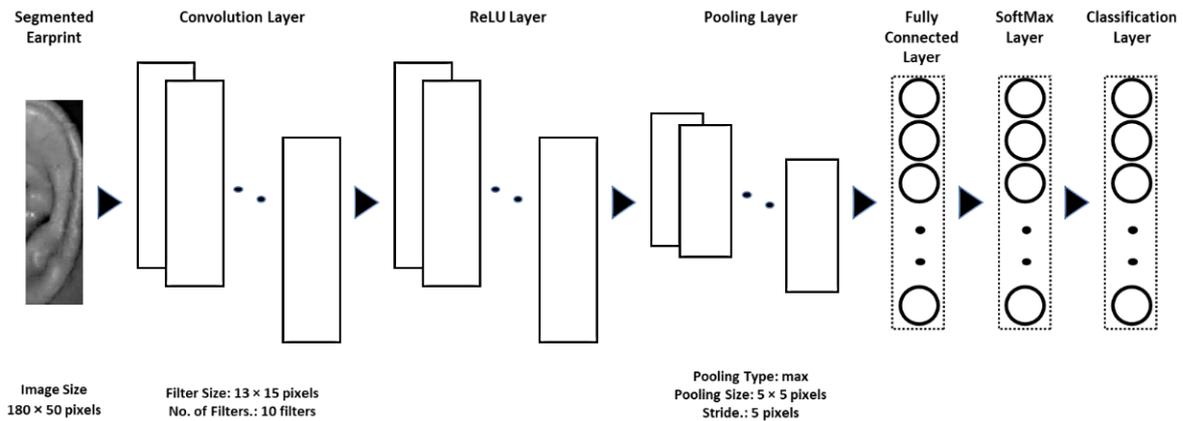


Figure 3. The architecture of the proposed DEL network

3. RESULTS AND ANALYSIS

First of all, earprint database are available for the IITDED version 1.0. It consists of touchless earprint images. It was collected from students and staff from the IIT Delhi in India. It was acquired between October 2006 to June 2007. Earprint images were captured using indoor environment and simple imaging setup. 221 persons in the age between 14 to 58 years were participated with multiple image samples (at least three earprints). All earprint images are of resolution 272×204 pixels and they are of type jpeg format. Furthermore, segmented earprint images are also available within the same database, each with a size of 50×180 pixels [22, 23]. The segmented earprint images of the IITDED version 1.0 has been employed in this paper but for input image size of 180×50 pixels. Two third number of earprint samples has been used in the training stage. Whereas, 100 evaluated case has been used in the testing stage. The training parameters have been set as: Adaptive moment estimation (Adam) optimizer, maximum epochs equal 50, initial learn rate equal 0.0001, decay rate of gradient moving average of 0.9, decay rate of squared gradient moving average of 0.999 and denominator offset of 10^{-8} .

To determine the best DEL network parameters, many experiments were executed and evaluated. Table1 shows various DEL network experiments with Adam optimization to determine the best parameters of convolution and pooling layers. In this table, convolution layer and pooling layer parameters are evaluated by changing a single parameter and fixing the values of all the remaining parameters. The DEL network performance can simply be evaluated by the obtained accuracy. It can be observed that the best convolution layer parameters are recorded for: filter size of 13×13 , number of filters equal 10 and padding of 0. Furthermore, it can be seen that the best pooling layer parameters are reported for: pooling type of maximum, filter size of 5×5 , stride of 10 and padding of 0. This is because the DEL accuracy after using these parameters has benchmarked its highest value of 94%. Tuning any parameter value for slightly more or less than the recorded parameter would decrease the accuracy value.

The training curves of the DEL network by using the best obtaining parameters are given in Figure 4. Moreover, different training optimization methods have been examined for the DEL network as given in Table 2. In this table, different training optimizations have been evaluated. These are: stochastic gradient descent with momentum (SGDM), root mean square propagation (RMSProp) and Adam. Obviously, Adam optimization has attained best accuracy of 94% compared with the SGDM and RMSProp as they attained 71% and 93%, respectively as see in Figure 4. It is worth mentioning that the performance of increasing the DEL architecture by adding more than one convolution, ReLU and pooling layers are also investigated. That is, by using two sequential convolutions, ReLU, pooling, FC, Softmax and classification layers the accuracy decreased to 82%. Also, by using convolution₁, ReLU₁, convolution₂, ReLU₂, pooling, FC, Softmax and classification layers the accuracy decreased to 83%. According to these results the reported accuracies are too far from the best attained accuracy. Thus, it is not worth to increase the complexity of the DEL network architecture. The DEL network has been compared with state-of-the-art methods as shown in Table 3. From Table 3, it can clearly be seen that our proposed method has achieved the best accuracy of 94% over state-of-the-art methods after applying their proposed architectures to our data. That is, the proposed CNN architecture in [24] achieved 63% and the novel deep finger texture learning (DFRL) architecture in [25] achieved 72%.

Table 1 Various DEL network experiments with Adam optimization to determine the best parameters of convolution and pooling layers

Convolution layer			Pooling layer				Accuracy (%)
Filter size	No. of filters	Padding	Type Maximum (Max) or Average (Ave)	Size	Stride	Padding	
15×15	10	0	Max	5×5	5	0	86
13×13	10	0	Max	5×5	5	0	94
11×11	10	0	Max	5×5	5	0	85
13×13	8	0	Max	5×5	5	0	89
13×13	12	0	Max	5×5	5	0	86
13×13	10	0	Max	3×3	3	0	87
13×13	10	0	Max	7×7	7	0	84
13×13	10	0	Ave	5×5	5	0	89
13×13	10	1	Max	5×5	5	0	83
13×13	10	2	Max	5×5	5	0	86
13×13	10	0	Max	5×5	5	1	85
13×13	10	0	Max	5×5	5	2	87

Table 2. Different examined training optimization methods for the DEL network

Training optimization method	Accuracy (%)
SGDM	71
RMSProp	93
Adam	94

Table 3. Comparisons with different state-of-the-art methods

References	DL Network	Accuracy
[22]	CNN	63%
[23]	DFTL	72%
Suggested method	DEL	94%

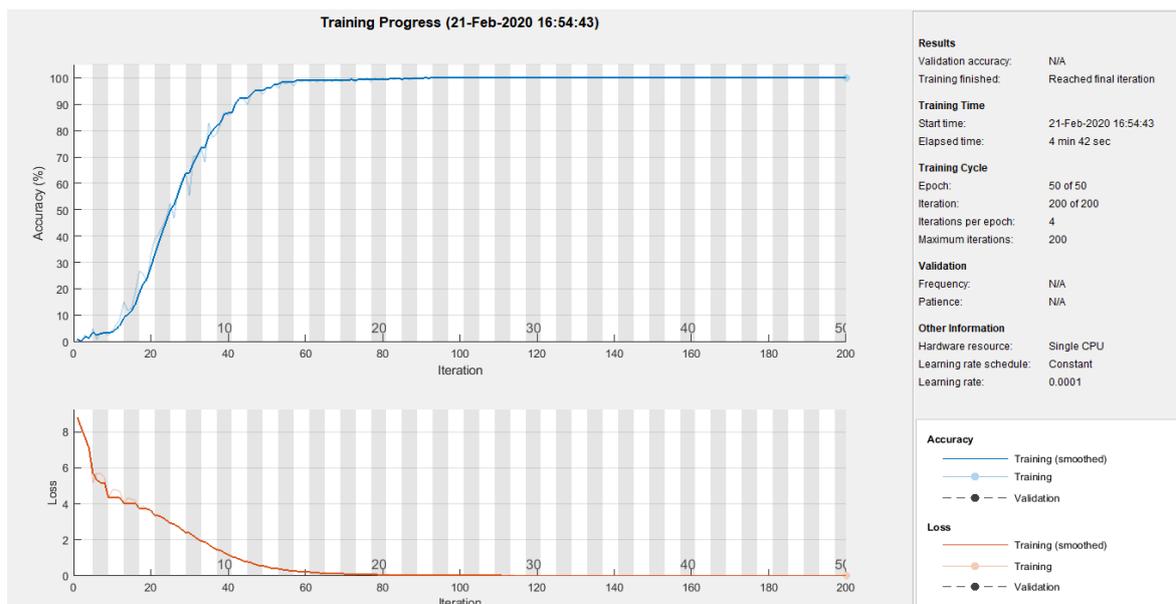


Figure 4. Training curves of the proposed DEL network

4. CONCLUSION

In this study, we proposed an efficient DEL network model. This network has the ability to recognize different earprint patterns. The suggested method was investigated and evaluated for different DL parameters. It reported best recognition accuracy of 94% and this can be considered as a promising performance. Also, the DEL outperformed other state-of-the-art networks.

ACKNOWLEDGEMENTS

"Portions of the work tested on the IITD Ear Database version 1.0".

REFERENCES

- [1] Abeer A., *et al.*, "Biometric Face Recognition Using Multilinear Projection and Artificial Intelligence," PhD thesis, School of Engineering, Newcastle University, UK, 2013.
- [2] R. R. Al-Nima, "Signal Processing and Machine Learning Techniques for Human Verification Based on Finger Textures," PhD thesis, School of Engineering, Newcastle University, UK, 2017.
- [3] V. Matyas, *et al.*, "Biometric authentication systems," in *verfugbar uber: http://grover. Informatik. Uni-augsburg. De/lit/MMSeminar/Privacy/riha00biometric. Pdf*. Citeseer, 2000.
- [4] M. Abdullah, *et al.*, "Cross-spectral Iris Matching for Surveillance Applications," Springer, *Surveillance in Action Technologies for Civilian, Military and Cyber Surveillance*, pp. 105-125, 2017.
- [5] S. Minaee, *et al.*, "An experimental study of deep convolutional features for iris recognition," *IEEE signal processing in medicine and biology symposium (SPMB)*, pp 1-6, 2016.
- [6] Nitin Kaushal, *et al.*, "Human earprints: a review," *Journal of biometrics and biostatistics*, vol. 2, no. 5, 2011.
- [7] R. R. Al-Nima, "Human authentication with earprint for secure telephone system," *Iraqi Journal of Computers, Communications, Control and Systems Engineering IJCCE*, vol. 12, no. 2, pp.47-56, 2012.
- [8] M. Oravec, *et al.*, "Mobile ear recognition application," in *International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2016.
- [9] Morales, *et al.*, "Earprint recognition based on an ensemble of global and local features," *International Carnahan Conference on Security Technology (ICCST)*, IEEE, 2015.
- [10] Omara, *et al.*, "A novel geometric feature extraction method for ear recognition," Elsevier, *Expert Systems with Applications*, vol. 65, pp. 127-135, 2016.
- [11] G. Mawloud, *et al.*, "Sparse coding joint decision rule for ear print recognition," *Optical Engineering*, vol. 55, no. 9, 2016
- [12] F. I. Eyiokur, D. Yaman and H. K. Ekenel, "Domain adaptation for ear recognition using deep convolutional neural networks," in *IET Biometrics*, vol. 7, no. 3, pp. 199-206, 2018.
- [13] A. Sepas-Moghaddam, F. Pereira and P. L. Correia, "Ear recognition in a light field imaging framework: a new perspective," in *IET Biometrics*, vol. 7, no. 3, pp. 224-231, 2018.
- [14] A.M. Kumar, *et al.*, "Local Binary Pattern based Multimodal Biometric Recognition using Ear and FKP with Feature Level Fusion," *IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, IEEE, 2019.
- [15] N. Kumar, "A Novel Three Phase Approach for Single Sample Ear Recognition," In *International Conference on Computing and Information Technology*, Springer, Cham, 2019.
- [16] Simo-Serra, *et al.*, "Learning to simplify: fully convolutional networks for rough sketch cleanup," *ACM Trans. on Graphics (TOG)*, vol. 4, no. 35, p. pp. 121, 2016.
- [17] S. Chris, *et al.*, "Fingerphoto recognition with smartphone cameras," BIOSIG-Proceedings of the *International Conference of Biometrics Special Interest Group (BIOSIG)*, IEEE, 2012.
- [18] Wu, *et al.*, "Introduction to convolutional neural networks," *National Key Lab for Novel Softw. Technol., Nanjing University*, 2017.
- [19] Stutz, *et al.*, "Neural Codes for Image Retrieval," *Neural Codes for Image Retrieval*, pp. 584-599, 2014.
- [20] Kevin Jarrett, *et al.*, "What is the best multi-stage architecture for object recognition?" In *Computer Vision, International Conference on*, Kyoto, Japan, September 2009, pp. 2146-2153.
- [21] D. Stutz, "Neural codes for image retrieval," *Proceeding of the Computer Vision-ECCV*, Zurich, Switzerland, 2014.
- [22] "IIT Delhi Touchless Palmprint Database version 1.0" [Online]. Available: http://webold.iitd.ac.in/~biometrics/Database_Ear.html
- [23] Kumar, *et al.*, "Automated Ear Identification using Ear Imaging," *Pattern Recognition*, vol. 45, no. 3, pp. 956-968, 2012.
- [24] George, *et al.*, "Real-time eye gaze direction classification using convolutional neural network," In: *International conference on signal processing and communications (SPCOM)*, 2016.
- [25] R. R. Omar, T. Han, S. A. M. Al-Sumaidae and T. Chen, "Deep finger texture learning for verifying people," in *IET Biometrics*, vol. 8, no. 1, pp. 40-48, 1 2019.