Indonesian Journal of Electrical Engineering and Computer Science
Vol. 22, No. 3, June 2021, pp. 1548~1555
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v22.i3.pp1548-1555

Pedestrian age estimation based on deep learning

Nawal Younis Abdullah, Mohammed Talal Ghazal, Najwan Waisi
Department of Computer Engineering Technology, Northern Technical University, Mosul, Iraq

Article Info

Article history:

Received Jan 24, 2021
Revised Mar 23, 2021
Accepted Mar 29, 2021

Keywords:

Convolutional neural networks
Deep learning
Pedestrian
ResNet-50
VGG-Face

ABSTRACT

The large-scale distribution of camera networks in the traffic area resulted in the increasing popularity of video surveillance systems. As pedestrian detection and tracking are the critical monitoring targets in traffic surveillance, many studies focus on pedestrian detection algorithms across cameras. This paper addressed the effect of using the age estimation based on deep convolution neural network (CNN) as a convenience for pedestrian monitoring who is crossing at intersections. Two popular deep convolutional neural networks (DCNNs) pre-trained models have been used in this work, which have recently achieved the best performance in facial features extraction tasks: VGG-Face and ResNet-50. We combined these two models to increase the efficiency of the proposed system. We ran our experiments to evaluate the system based on the VGGFace2 dataset consisting of 3.31 million face images. From the experimental results, we observed a gap in the detection performances between those age groups: children from (00-10) years and elderly with 55 years and more. Moreover, it noted that the proposed pedestrian age estimation model performance is high, also a good result can be obtained by using the model for new purpose.

Corresponding Author:

Mohammed Talal Ghazal
Department of Computer Engineering Technology
Northern Technical University
Mosul, Iraq
Email: mohammed.ghazal@ntu.edu.iq

1. INTRODUCTION

The World Health Organization (WHO) reported the statistics for the number of road traffic deaths in 2018, the rate of death reached 1.35 million in 2016 with an average rate of 27.5 deaths per 100,000 population as shown in Figure 1. The report recommended the need to improve the roads infrastructure, which represents the main cause of the fatal injury in road traffic collisions [1]. Pedestrian monitoring systems provide quantitative insights into the pedestrian traffic state, although they are easy to install and good under normal conditions, but they are not 100% accurate [2]. This work focuses on age estimation for specific age groups, since elderly people fall during across the street is among the most damaging event and the most common occurrence, beside the children between those two phases, ages 4 through 10. The majority of pedestrian injuries happen in mid-block crosswalks, where young children try to cross a street without checking for traffic, or at intersections [3].

The proposed system is a part of the computer vision field, which plays a significant role in face recognition applications [4]. Briefly, the computer vision algorithms task is to extract the facial features from images and compare them with datasets of face profiles [5]. Recently deep learning plays a major role as a computer vision tool [6], [7]. In deep learning technology, a convolutional neural network (CNN) is a deep neural network architecture class that is mostly used to analyze the visual imagery [8], [9]. A deep CNN's mathematical construction consists of several types of layers: convolutional layers, pooling, and a connected

layer. The convolution and pooling layers' function are to perform feature extraction, while the fully connected layer feds with the result of these processes to get the final classification decision [10], [11].

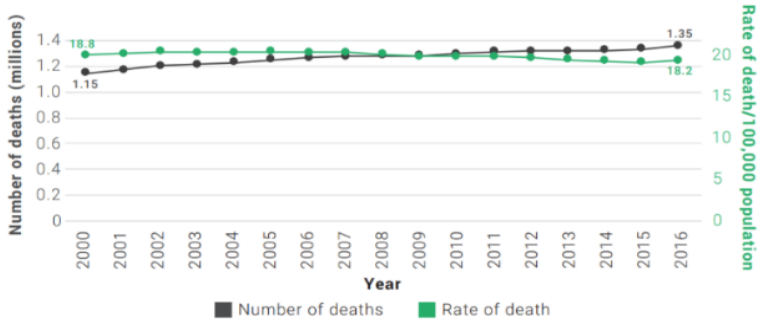

Figure 1. Death rate from road traffic for the period (2000–2016) per 100,000 population

VGG-Face and ResNet-50 are top-rated deep convolutional neural network (DCNN) models; we propose to use these two models to improve the estimation of pedestrian age. Using these architectures for feature extraction from images facilitates traditional methods' difficulty because it is needed to select which features are significant in each given image [12], [13]. We don't require to use feature descriptors: scale-invariant feature transform (SIFT), speeded up robust features (SURF), BRIEF**,** histogram-oriented gradient (HOG) and Support vector machine (SVM) for further recognition or classification tasks [14]. Machine learning is also not free of difficulties; overfitting problems represent one of the most difficult problems when using small datasets. This problem results from the enormous numbers of deep network parameters since they have multiple layers and thousands of nodes. Most of the studies on age estimation have occurred recently, so we find that most of the databases built for this purpose are small in size. To solve the overfitting problem, we built our proposed pedestrian age estimation system based on a deep CNN model, which trained on a huge database [15]. The main contribution of this paper is to achieve the best performance for human age estimation through the combination of VGG-Face and ResNet-50 models.

Many related studies have been published on using video data to investigate pedestrians crossing behaviours. An investigation of pedestrian's location violations in [16], a study on prediction of pedestrians red-light crossing intentions based on the appearance characteristics: gender, age and head direction in [17], a SORT tracking model in [18], and a study based on region-based convolutional neural network (R-CNN) object detection model has been proposed in [19]. Although these previous studies were based on machine learning models, they were not efficient in detecting the relationships in the series of data for future predictions. Our work in pedestrian detection is based on age estimation, a study based on deep expect of visible age from a single image, which is performed using CNN architecture presented in [20]. The problem of this method is approached as a classification problem with 101 age classes. In section 2 of this work, we described the proposed model. The experimental model training and results have been proposed in section 3. Finally, the conclusions are reviewed in section 4.

2. THE PROPOSED MODEL

Our proposed method for pedestrian age estimation focused on training age-dependent face representation to enhance system performance, and it is based on two popular pre-trained models (ResNet-50 and VGG-Face) to exploit age information for improving the detection system. At first, the system detects faces from the input images and enclosing them by a bounding box based on Haar-cascade method [21]. Haar-cascade is a machine learning algorithm for object detection. In this method the cascade function is trained by a lot of positive and negative images, where the object that want to be detected is exist in the positive images, while the negative images are those where it is not. This method has been used to detect faces in images. Then, an age estimator module has been used to predicate the age of pedestrians automatically. Figure 2 shows a diagram of the model architecture.

2.1. The pre-trained deep CNNs models

Both ResNet-50 convolution network model suggested in [22], and VGG-Face convolution network model, suggested in [23], [24] have been used in this work to achieve some of the best performance in an age estimation task. ResNet-50 is a deep CNNs based on residual neuronal network architecture. This network is

distinguished by adding a shortcut connection between blocks, which switches the convolution network to a residual neural network version. The network consists of 34 residual layers for training and 41,192,951 parameters; it has been adapted to age estimation to be used in this study. The ResNet-50 model structure is shown in Figure 3.

The VGG-Face network is another deep CNNs based on VGG-16 network architecture, it consists of 11 layers, 8 blocks of convolutional layers, and three fully-connected layers with rectified linear unit (ReLU) activations, the network configuration is shown in Table 1. VGG-Face is applied as a facial feature extractor for age estimation purposes from the face images.

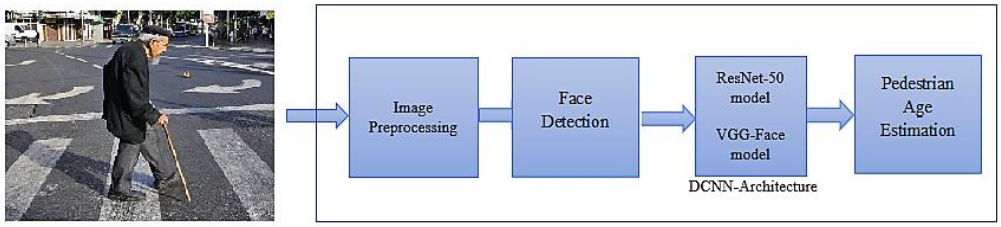

Figure 2. The pedestrian age estimation model architecture

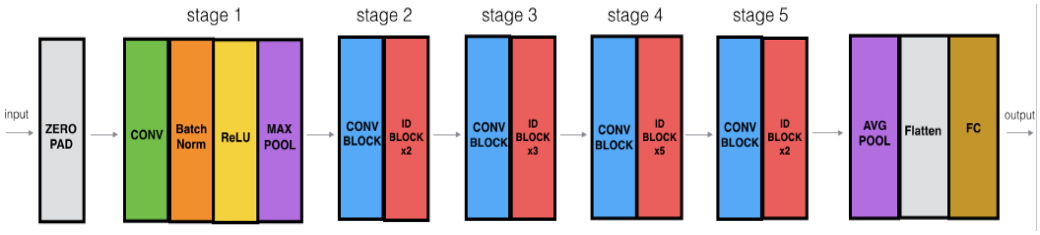

Figure 3. ResNet-50 model structure

Table 1. VGG-Face layers structure [22]

Layer type name	Support	Filt dim	Num filt	Stride	Pad	Layer type name	Support	Filt dim	Num filt	Stride	Pad
0 input	-	-	-	-	-	19 relu relu 4_1	1	-	-	1	0
1 conv conv1_1	3	3	64	1	1	20 conv conv4_2	3	512	512	1	1
2 relu relu1_1	1	-	-	1	0	21 relu relu 4_2	1	-	-	1	0
3 conv conv1_2	3	64	64	1	1	22 conv conv4_3	3	512	512	1	1
4 relu relu 1_2	1	-	-	1	0	23 relu relu4_3	1	-	-	1	0
5 mpool pool 1	2	-	-	2	0	24 mpool pool4	2	-	-	2	0
6 conv conv2_1	3	64	128	1	1	25 conv conv5_1	3	512	512	1	1
7 relu relu 2_1	1	-	-	1	0	26 relu relu 5_1	1	-	-	1	0
8 conv conv2_2	3	128	128	1	1	27 conv conv5_2	3	512	512	1	1
9 relu relu2_2	1	-	-	1	0	28 relu relu5_2	1	-	-	1	0
10 mpool pool2	2	-	-	2	0	30 relu relu5_3	1	-	-	1	0
11 conv conv3_2	3	128	256	1	1	31 mpool pool5	2	-	-	2	0
12 relu relu3_1	1	-	-	1	0	32 conv fc6	7	512	4096	1	0
13 conv conv3_2	3	256	256	1	1	33 relu relu6	1	-	-	1	0
14 relu relu3_2	1	-	-	1	0	34 conv fc7	1	4096	4096	1	0
15 conv conv3_3	3	256	256	1	1	35 relu relu7	1	-	-	1	0
16 relu relu3_3	1	-	-	1	0	36 conv fc8	1	4096	2622	1	0
17 mpool pool3	2	-	-	2	0	37 softmx prob	1	-	-	1	0
18 conv conv4_1	3	256	512	1	1						

2.2. Assemble learning for age estimation

The proposed age estimator is built based on fine-tuning the pre-trained models ResNet-50 and VGG-Face to leveraging an ensemble approach since they are achieving greater discrimination rates when compare them with other similar models. A vector of weights is obtained from the second to the last layer of the convolutional model, which can be used as feature vectors. The weights feature vectors considered the latent representation for the input image which each model learned. In the proposed method, we combined these latent representations by concatenating the feature vectors to form an overall feature vector inputted into logistic regression models for the final age prediction.

2.3. Database

VGGFace2 dataset used in this work, provides a dataset of annotations of 3.31 million face images belong to 9131 subjects and about 362.6 images for each subject. VGGFace2 has been developed for the advanced study of face recognition systems, it has a large variety and available labels for age and pose. The VGGFace2 dataset's annotations are made on faces samples taken from Google Image [25]. Although the black and white face images are robust in face recognition but they are excluded, The VGGFace2 dataset focused on facial and image variation due to color processing as shown in Figure 4. Five age classes have been included in this study {(00-10), (11-20), (21-35), (36-54), (55-90)}.

Figure 4. A sample of face image from VGGFace2 dataset

2.4. Preprocessing dataset

The environment of this work is based on TensorFlow, which represent the core open-source library to help researcher develop and train ML models. About 30k images of VGGFace2 have been used. Before loading input images, the working environment need some preparation such as installing the required libraries, then start to load the dataset. A preprocessing data, is required at this stage including cleansing the dataset from low quality images which will confuse the training model, A relabeling age group images to five categories {(00-10), (11-20), (21-35), (36-54), (55-90)}, splitting dataset to 80% train, 10% test and 10% validation and finally, set the batch size. The age estimation model is trained with three pre-trained-weight, and two deep learning algorithms (VGG-Face and ResNet50) as it illustrated in the experimental work.

3. EXPERIMENTAL WORK AND RESULTS

The proposed system's experimental work was designed to train an automatic pedestrian age estimator using face images dataset VGGFace2 and analyze the effect of using the two deep learning models, ResNet-50 and VGG-Face. To evaluate the system performance and efficiency, many experiments have been performed (Experiment 1); the ResNet-50 model is loaded with the pre-trained weight of "ImageNet", to use it as a base model. The ImageNet large scale visual recognition challenge (ILSVRC) is a large visual database containing more than 14 million images included with a total of 20,000 categories, which is used for object category classification and detection [26]-[29]. The result from ResNet-50 presents that both train and test performance are less than 50% as shown in Figure 5, this is due to the ImageNet is not specified for face images and contains various image categories. Figure 5 illustrate the accuracy of ResNet-50, the curve of accuracy on the training data (acc) while (val_acc) is the curve of accuracy on the validation data.

(Experiment 2); The VGGFace2 dataset is divided into the training set, validation set, and test set, to train both VGG-Face and ResNet-50 pre-trained models as the baseline, which resulted from a higher performance of more than 90%. Moreover, The ResNet-50 resulted in a higher accuracy performance than VGG-Face, as shown in Figure 6. Finally, (3), to increase the system efficiency and enhance the performance we combined the two DCNN models in this experiment, the system accuracy increased both in train (0.9689) and test set (0.8801). The system trained using 150 epochs. Age prediction accuracy becomes (0.8678) on the validation set at epoch 50. We use the Adadelta optimization and sparse categorical cross-entropy, the model implemented using TensorFlow library. The system performance for experiments 3 illustrated in Figure 7.

In this study, the class data is imbalanced. Therefore, a confusion matrix with normalization has been used. The best category prediction is on the age of 00-10 years old, and the worst prediction is on 55-90 years old. This is because of face structure; the children face from 00-10 years old can easily

distinguish from other age groups. Therefore, the system classified this age group as the best. The confusion matrix is shown in Figure 8. The proposed system's test performance is (88.01%) and can be increased if the pedestrian face image taken in the close shot will be the high-quality resolute image, some of the test result images are shown in Figure 9.

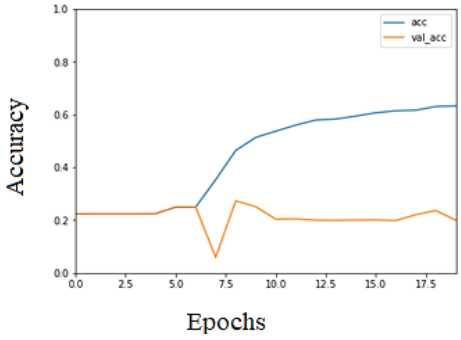

Figure 5. The accuracy of ResNet-50 loaded with ImageNet weights for the proposed system

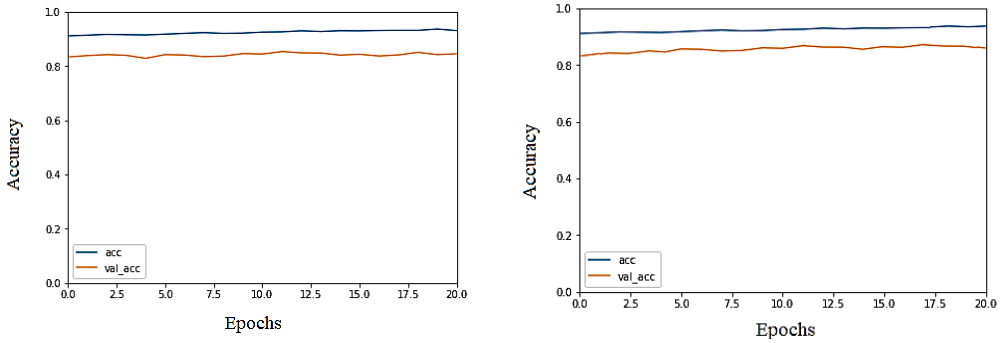

Figure 6. The accuracy of system training and testing for both VGG-Face and ResNet-50 models with VGGFace2 dataset: (a) VGG-Face trained with VGGFace2 dataset accuracy. The best train accuracy: 0.9092. The best test accuracy: 0.8603 and (b) ResNet-50 trained with VGGFace2 dataset accuracy. The best train ac curacy: 0.9392. The best test accuracy: 0.8751

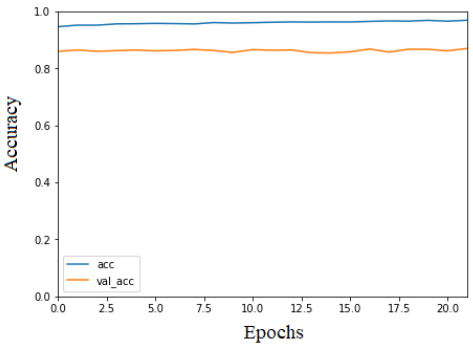

Figure 7. The overall system performance by combining the two DCNN models

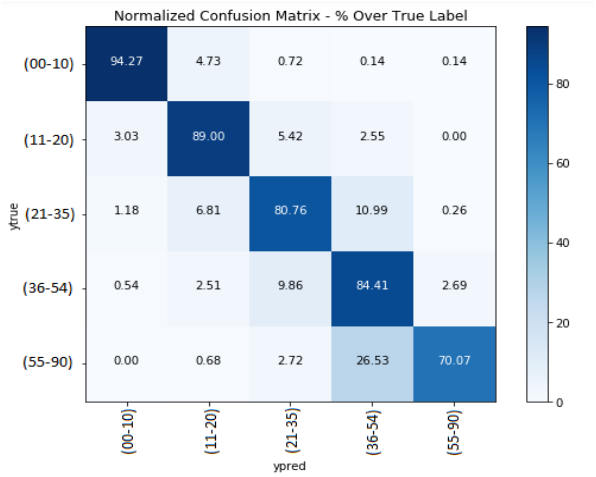

Figure 8. Confusion matrix

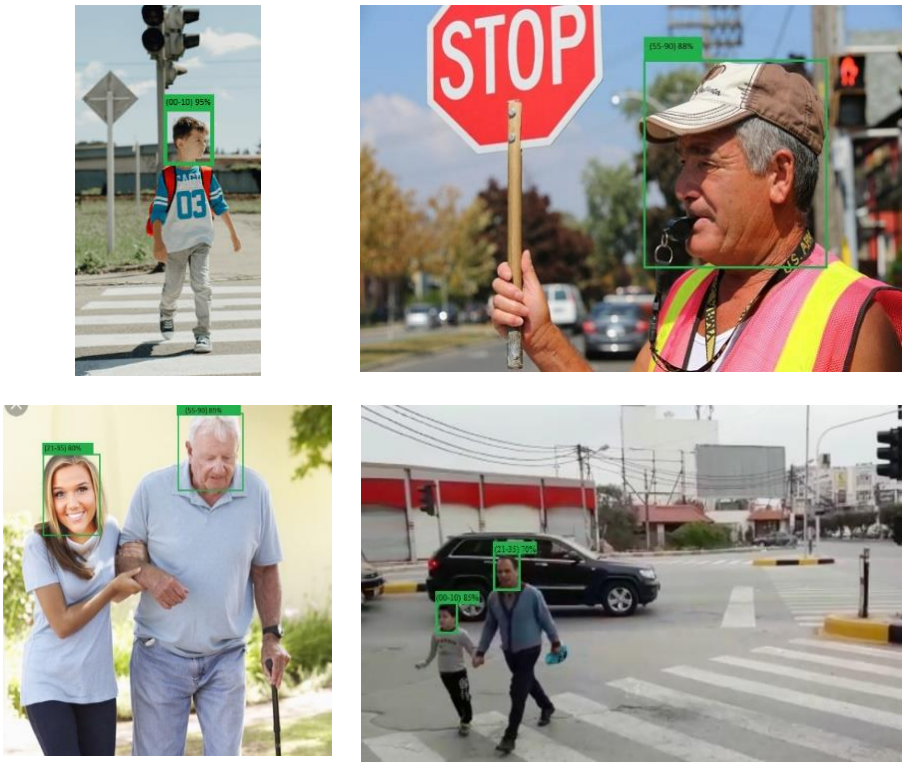

Figure 9. Some of test result images

4. CONCLUSION

This paper implies a combination of two deep convolutional neural network DCNN model's architecture to estimate the pedestrians' age from the provided images of the surveillance camera system. The research contributes to reducing accidents resulting from crossing the street, in addition to alerting vehicle drivers to the presence of pedestrians on the road. The experimental work shows the efficiency and performance of the proposed method. It is also observed that there are some difficulties to estimate the elderly age group, besides the poor image resolution problem. We will enhance the proposed model in the future to improve the efficiency of other pedestrian face attributes.

ACKNOWLEDGEMENTS

This research is funded by cooperation between deep learning team in Northern Technical University NTU in Iraq. Website: For Northern Technical University https://www.ntu.edu.iq

REFERENCES

[1] World Health Organization, "Global status report on road safety," World Health Organization, 2018.
[2] D. Duives, W. Daamen, and S. Hoogendoorn, "Monitoring the Number of Pedestrians in an Area: The Applicability of Counting Systems for Density State Estimation," *Journal of Advanced Transportation*, vol. 2018, pp. 1-14, 2018, doi: 10.1155/2018/7328074.
[3] D. C. Schwebel, A. L. Davis, and E. E. O'Neal, "Child pedestrian injury: A review of behavioral risks and preventive strategies," *American journal of lifestyle medicine*, vol. 6, no. 4, pp. 292-302, 2012, doi: 10.1177/0885066611404876.
[4] P. Patel and A. Thakkar, "The upsurge of deep learning for computer vision applications," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 538-548, 2020, doi: 10.11591/ijece.v10i1.pp538-548.
[5] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 1015-1027, 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.
[6] D. A. Jasm, M. M. Murtadha, and A. T. H. Alrawi, "Deep image mining for convolution neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, pp. 347-352, 2020, doi: 10.11591/ijeecs.v20.i1.pp347-352.

[7] H. Kim, S. Lee, and H. Jung, "Human activity recognition by using convolutional neural network," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6, pp. 5270-76, 2019, doi: 10.11591/ijece.v9i6.pp5270-5276.
[8] B. Kim, N. Yuvaraj, K. Sri Preethaa, R. Santhosh, and A. Sabari, "Enhanced pedestrian detection using optimized deep convolution neural network for smart building surveillance," *Soft Computing*, vol. 24, no. 2, pp. 1-12, 2020, doi: 10.1007/s00500-020-04999-1.
[9] F. H. K. Zaman, J. Johari, and A. I. M. Yassin, "Learning Face Similarities for Face Verification using Hybrid Convolutional Neural Networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 3, p. 1333-1342, 2019, doi: 10.11591/ijeecs.v16.i3.pp1333-1342.
[10] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, no. 4, pp. 611-629, 2018, doi: 10.1007/s13244-018-0639-9.
[11] Z. Kadim, M. A. Zulkifley, and N. Hamzah, "Deep-learning based single object tracker for night surveillance," *International Journal of Electrical & Computer Engineering*, vol. 10, no. 4, pp. 33576-87, 2020, doi: 10.11591/ijece.v10i4.pp3576-3587.
[12] N. O'Mahony *et al.*, "Deep learning vs. traditional computer vision," in *Science and Information Conference*, 2019, pp. 128-144.
[13] M. H. A. Hijazi, S. K. T. Hwa, A. Bade, R. Yaakob, and M. S. Jeffree, "Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images," *International Journal of Artificial Intelligence*, vol. 8, no. 4, pp. 429-435, 2019, doi: 10.11591/ijai.v8.i4.pp429-435.
[14] A. Dhomne, R. Kumar, and V. Bhan, "Gender recognition through face using deep learning," *Procedia Computer Science*, vol. 132, pp. 2-10, 2018, doi: 10.1016/j.procs.2018.05.053.
[15] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, "Deep convolutional neural network for age estimation based on VGG-face model," *arXiv preprint arXiv:1709.01664*, 2017.
[16] M. H. Zaki and T. Sayed, "Automated analysis of pedestrians' nonconforming behavior and data collection at an urban crossing," *Transportation research record*, vol. 2443, no. 1, pp. 123-133, 2014, doi: 10.3141/2443-14.
[17] D. Ka, D. Lee, S. Kim, and H. Yeo, "Study on the framework of intersection pedestrian collision warning system considering pedestrian characteristics," *Transportation research record*, vol. 2673, no. 5, pp. 747-758, 2019, doi: 10.1177/0361198119838519.
[18] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," *2017 IEEE international conference on image processing (ICIP)*, 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.
[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.
[20] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 10-15, doi: 10.1109/ICCVW.2015.41.
[21] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, doi: 10.1109/CVPR.2001.990517.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[23] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," Visual Geometry Group Department of Engineering Science University of Oxford, 2015.
[24] A. A. Moustafa, A. Elnakib, and N. F. Areed, "Optimization of deep learning features for age-invariant face recognition," *International Journal of Electrical & Computer Engineering*, vol. 10, no. 2, pp. 1833-41, 2020, doi: 10.11591/ijece.v10i2.pp1833-1841.
[25] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 67-74, doi: 10.1109/FG.2018.00020.
[26] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211-252, 2015.
[27] H. Sofian, J. T. C. Ming, S. Muhammad, and N. M. Noor, "Calcification detection using convolutional neural network architectures in intravascular ultrasound images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 3, pp. 1313-1321, 2020, doi: 10.11591/ijeecs.v17.i3.pp1313-1321.
[28] R. F. Rahmat, O. S. Sitompul, S. Purnamawati**,** and R. Budiarto, "Advertisement billboard detection and geotagging system with inductive transfer learning in deep convolutional neural network," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 17, no. 5, pp. 2659-2666, 2019, doi: 10.12928/telkomnika.v17i5.11276.
[29] M. Ghazal, N. Waisi, and N. Abdullah, "The detection of handguns from live-video in real-time based on deep learning," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 18, no. 6, pp. 3026-3032, 2020, doi: 10.12928/telkomnika.v18i6.16174.

BIOGRAPHIES OF AUTHORS

Nawal Younis Abdullah obtained her M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2014. Her M.Sc. thesis entitled: "FPGA Based Video Scene Boundaries Detection Using Enhanced Sobel Filter" and her current research interests include the deep learning CNNs models and object detection.

Mohammed Talal Ghazal obtained his M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2016. His M.Sc. thesis entitled: "Wheelchair Robot Control Using EOG signals". His research interests include the design of face recognition algorithms, deep learning CNNs models and object detection.

Najwan Zuhair Waisi obtained her M.Sc. degree from Computer Science Department, Mousl University, Mosul, Iraq in 2014. Her M.Sc. thesis entitled: "Design and Implementation of Client Honeypot" and her current research interests include the deep learning CNNs models and object detection.