Gender and Age Estimation at Distance in Smart Cities Surveillance: A cascaded Deep Learning-Based Approach

Yahia Said [1,2]

[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

Summary
Smart cities are becoming a reality with a big number of mounted cameras on each street. Those cameras provide a huge amount of data that humans cannot process it. The main goal of those cameras is to detect criminals, detect sick people, and find lost children. In this paper, we propose an intelligent gender and age estimation system for smart city surveillance. The recent improvement of convolutional neural networks for computer vision tasks makes it suitable for the proposed intelligent system. However, the distance separating the camera from the target person can be a hard challenge that must be solved. The idea of this work is to detect faces in real scene images and estimates gender and age by analyzing facial landmarks. To do that, we propose to combine a deep convolutional generative adversarial network, the viola jones algorithm, and a convolutional neural network to generate predictions on gender and age.
The generative adversarial network was used to enhance the resolution of the images, the viola jones algorithm was used to detect and crop faces in the images, and the convolutional neural network was used to generate predictions on the gender and the age of the detected face. The proposed approach was evaluated on the IMDB-WIKI dataset. The obtained results prove the efficiency of the proposed approach.

Key words:
Gender estimation, Age estimation, Smart city surveillance, Deep Learning, Face detection, Generative Adversarial Network.

1. Introduction

Smart cities are considered as our future living style, cameras are everywhere in the streets that collect relevant information about urban living. Those cameras provide a huge amount of visual data which is analyzed by the government and local authorities in order to improve the sustainability and efficiency of urban spaces. This data is very helpful for crime prevention, detect sick people (e.g. Corona infected), and finding lost people. A factor key for most of those functionalities is the gender and age estimation. Knowing the gender and the age of such person make it easier to find it in the crowd.

Generally, the estimation of gender and age is based on face landmark analyses. Most surveillance systems capture images from a long distance. So, the whole body of a person will be presented in the image. The main idea of automatic gender and age estimation is based on face cropping then landmarks analyses. For such conditions, the resolution of the cropped faces will be low and that may lead to false predictions. In this paper, we propose a solution to enhance the resolution of the collected images, the crop faces, and make predictions. In the last few years, the deep learning techniques [1] has been deployed successfully for almost all computer vision applications such as image classification [2], [3], scene recognition [4], indoor object recognition [5], and object detection [6], [7]. Deep learning outperforms the old machine learning techniques which are based on handcrafted features.

Deep learning is based on deep neural networks with tens of hidden layers. Those neural networks can learn by themselves directly from the input data. For computer vision applications, convolutional neural networks (CNN) [8] are the main deep learning model used. CNN is composed of different types of layers, convolution layers, non-linear layers, pooling layers, and fully connected layers. The power of CNN comes from the convolution layer that allows processing of visual data comparatively to the biological system of animals [9]. Based on the mentioned, CNN are perfect for gender and age estimation. Since there is more than one challenge to solve in the gender and age estimation task, we propose to cascade 3 different modules to solve the problem. The first module is used to enhance the resolution of the collected images, the second one is used to crop faces from images and the third one is used to analyses facial landmarks and estimate gender and age.

To enhance the resolution of the captured images, we propose to use the Deep Convolutional Generative Adversarial Network (DCGAN) [25]. Generally, The GAN [10] is composed of 2 neural networks. The first network is the generator which generates new data based on input data. For visual data, taking an image as input, the generator can generate new artificial images that look authentic for humans. The second network is the discriminator that judges the performance of the generator by distinguishing between the input data and the generated data. So, the main goal of the generator is to generate fake data that the discriminator cannot detect that is fake and the main goal of the discriminator is to identify the fake data. The two networks work oppositely.

Manuscript received September 5, 2020
Manuscript revised September 20, 2020

DOI: 10.22937/IJCSNS.2020.20.09.2

In this work, the DCGAN will be used to generate images with different resolutions based on the originally captured image. In the training process, the generated images are considered as data augmentation and for the testing process, they are used as extra input data for better predictions. The second module proposed for this task is the viola-jones algorithm [11] which used to crop faces from the images. In this work, we used an improved version of the viola-jones algorithm [12] for better detection and faster processing to meet real-time conditions. The viola-jones algorithm was designed for face detection in natural images in real-time. The last module in the gender and age estimation application is the convolutional neural network used to generate predictions. For this purpose, we design a custom CNN model based on lightweight Blocks proposed by the mobileNet model [13]. In the mobileNet model, the original convolution layers were replaced by depthwise separable convolution blocks. Each depthwise separable convolution block is composed of a depthwise convolution and pointwise convolution. The depthwise convolution does the same functionality of the original convolution layer but it does not affect the number of channels, which means the number of channels at the input equal to the number of channels at the output. The pointwise convolution is a convolution layer with a 1x1 kernel size.

The combination of the depthwise convolution and the pointwise convolution results in a normal convolution layer but they prove that separating the functionalities of the convolution results in a lighter model and faster processing. The main contribution of this work is to build a gender and age estimation system for smart cities based on analyzing images from surveillance cameras. The distance separating the cameras and the person is considered a hard challenge since faces appear too small and with low resolution.

In this paper, we propose to use the DCGAN to enhance the resolution of the images then the viola-jones algorithm was deployed to crop faces and a custom-made CNN was used to generate prediction by analyzing facial landmarks. To train and evaluate the proposed approach, we propose to use the IMDB-WIKI dataset [30] which is a dataset a publicly available dataset of face images collected from the IMD and the Wikipedia web sites. The dataset provides gender and age labels for the training data. The evaluation of the proposed approach proves its efficiency with an accuracy of gender classification of 97.3% and an age estimation accuracy of 91.2%. The achieved results outperform the state-of-the-art approaches.

The remainder of the paper is organized as a fellow. Related works are detailed and evaluated in section 2. Section 3 is reserved for describing the proposed approach. Experiment and results are discussed in section 4. In section 5, the paper was concluded and potential future work is proposed.

2. Related works

In the last years, automatic gender and age estimation systems become an important research topic. Many works have been proposed to enhance the performance of these systems in both accuracy and speed. Since the first use of automatic gender and age estimation, it was based on calculating the ratios between facial geometric features [14]. The geometry features can be deployed to distinguish between adult and a baby but it was unable to separate close ages. The active appearance model [15] was proposed to get better results. However, all the proposed methods were not suitable for real-word conditions such as illumination, pose variation, expressions, and occlusions. The deep learning has boosted the state-of-the-art performance.

Duan et al. [16] propose a combination between a convolutional neural network and an extreme learning machine for age and gender classification. The convolutional neural network was used without its fully connected layers for feature extraction and the extreme learning machine was used for age and gender classification. The proposed approach achieves a maximum accuracy of 88%. The proposed approach was not suitable for natural low-resolution images and works only on pre-cropped faces.

A convolutional neural network was proposed in [17] based on the residual network [18] for age group and gender estimation. The proposed convolutional neural network was pre-trained on the image Net data set then fine-tuned using the IMDB-WIKI 101 dataset [30]. for age estimation, the proposed approach achieves an accuracy of 61.78% and for gender classification, it achieves an accuracy of 90.87%. the proposed approach struggles to make predictions on low-resolution images.

Bahat et al. [19] propose a convolutional neural network for gender classification. The proposed CNN is composed of 6 convolution layers, each is followed by a nonlinear layer and max-pooling layer, and 2 fully connected layers. CNN was trained and evaluated on the IMDB-WIKI dataset [30]. The proposed approach achieves a maximum accuracy of 40% and it was very slow in terms of processing time.

In [20], the authors propose to deploy the VGG-Face [21] for gender classification. The VGG-Face was designed for face identification. The transfer learning technique was applied to the VGG-Faces to make it suitable for gender classification by replacing the IDs classifier by a binary classifier. The proposed network was pre-trained on the Celebrity faces dataset then it was fine-tuned using a subset of 200 faces images of the Celebrity faces dataset. 160 images were used as a training set and 40 images were used as a testing set. The evaluation of the proposed approach proves its efficiency with an accuracy of 95% but a very slow processing time.

Lee et al. [22] propose a combination of CNN based on the residual learning model [18] for gender and age estimation. The main idea was to use a network for gender classification, a network for female age estimation, and a network for male age estimation. Each network is composed of 50 layers with residual connections. The proposed approach was trained and evaluated on the IMDB-WIKI dataset [30]. An accuracy of 88.5% was achieved for gender classification and an accuracy of 52.2% was achieved for age estimation. The proposed approach struggle with small faces and does not achieve real-time processing.

A combination of wavelet pre-processing and Convolutional Neural Networks was proposed in [23]. The input images were transformed into the YCbCr color space. The Y component was resized to 32x32 using the discrete wavelet transform. Then it was fed to the AlexNet [24] convolutional neural network model for gender classification. The proposed approach allows to process low-resolution images and accelerate the processing time. An accuracy of 98.84% was achieved on the FERET dataset. The proposed approach works only on pre-cropped faces and cannot be applied to natural scene images.

As mentioned above, many works have been proposed to solve the problem of gender and age estimation but none of them take in account the real word conditions such geometric transformation, pose variation, a different point of view, occlusion, etc. In this work, we propose a solution to handle those conditions and considering the distance separating the acquisition camera from the target person. More details will be provided in the next section.

3. Proposed Approach

In this section, the proposed approach was detailed. The proposed approach is composed of 3 modules. The first module is based on a DCGAN for image resolution enhancement, the second module is used for face detection and cropping based on the viola-jones algorithm and the third module is a CNN model for gender and age prediction.

In a smart city, all surveillance cameras are mounted in a high position to cover larger areas. For gender and age estimation, facial landmarks must be analyzed. The main problem in this task that faces do not occupy more than 5% of the image size. So cropped faces will have low resolution and this will lead to false predictions. To solve this problem, we propose to use a GAN to enhance the resolution of the captured images.

In this work, we adopted the Deep Convolutional Generative Adversarial Network (DCGAN) for image generation and resolution enhancement. The DCGAN is a class of unsupervised learning model designed for high-resolution image generation. Taking an image as input, the DCGAN can generate similar images with better resolution and more discriminative features. The DCGAN generates realistic images that the human observatory cannot detect that they are fake. Figure 1 illustrates the typical architecture of the DCGAN.

The DCGAN is based on a Convolutional Neural Network where all pooling layers are replaced with fractional-strided convolution layers in the generator and with strided convolution layers in the discriminator. As proved in [25], using strided convolution allow to learn spatial upsampling without losing important features. Also, all fully connected layers are removed for deeper architecture. In the generator, the ReLU was used as an activation function except in the output layer that uses the Tanh activation function. In the discriminator, the Leaky-ReLU was used as an activation function. The Batch normalization was applied in both generator and discriminator. The DCGAN can take a random noise as input and generate images as output that looks real and the discriminator cannot detect that are fake. In this work, the DCGAN was trained to generate natural images similar to the images used as an input with higher resolution.

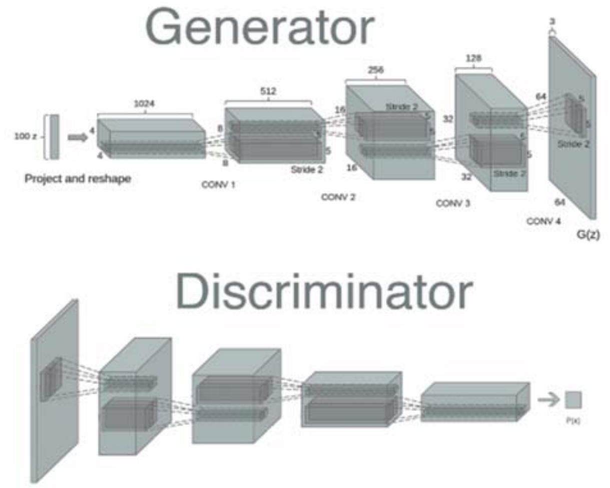

Fig. 1 Architecture of the DCGAN

In the training process, the DCGAN was considered as a data augmentation technique where the generated images were fed to further processing as new data but with the same labels. In the testing process, the generated images were used for over-sample testing to get better prediction. The DCGAN was tested on many datasets including faces datasets. So, it was easy to fine-tune it on the IMDB-WIKI dataset. The DCGAN was designed to generate training images that contain the same faces with higher resolution or new faces that have the same gender and the same age range. In effect, this allow to get more training data based on a small dataset. The testing images must contain the same faces with higher resolution.

After enhancing the resolution of the input images, faces must be detected and cropped for gender and age estimation. For this task, an improved version of the viola-jones algorithm was used. The viola-jones algorithm was designed for object detection and spatially or face detection in real-time. It was developed since 2001 and still until now a powerful framework for face detection in real-time. The viola-jones is consisting of 5 stages, gray scaling, haar-like descriptor [27], Integral Image generation [28], AdaBoost classifier [29], and cascaded classifiers. As a first step, the viola-jones algorithm converts input images to gray scale in order to optimize the processing time and complexity since there is no difference between detecting a color face and grayscale face. For the Haar-like descriptor, the viola-jones algorithm defines 3 important features, the edge features, the line features, and the four-sided features. The defined features allow to understand the image and to detect important face features like the mouth, nose, and eyebrows. The Haar-like features are presented in Figure 2.

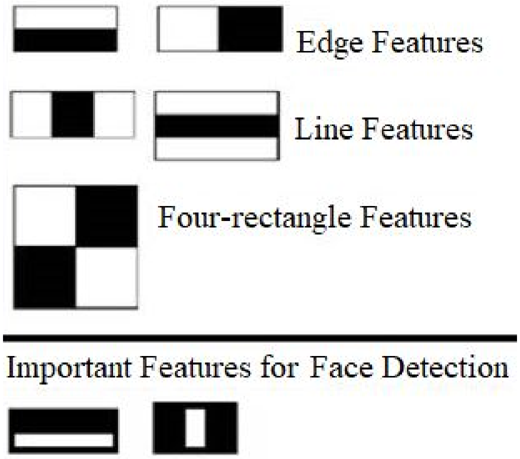

Fig. 2 Haar-like features from the viola-jones algorithm

Since calculating features is computationally intensive, the integral image was deployed to accelerate the processing time. In the viola-jones algorithm using the Haar-like descriptor are rectangles, so using the Integral Image make it easier to find important features in the images and by subtracting two rectangles of the integral image it is easy to find the difference between 2 features. So, even if the input image is high-resolution images, using an integral image will reduce the computation effort of calculating features. To classify the extracted features, the AdaBoost classifier was adapted. Many weak AdaBoost classifiers were trained each on particular features. All those weak classifiers were combined at the cascading stage to form a strong classifier. The cascading stage was used to speed up the processing time and to boost accuracy. The workflow of the viola-jones algorithm is illustrated in Figure 3.

Fig. 3 Workflow of the Viola-Jones algorithm

The Viola-Jones was used for face detection and cropping. The detected faces were sent to a convolutional neural network to generate predictions on gender and age. The proposed CNN was based on the lightweight locks proposed by the MobileNet model. In the MobileNet model, they propose to replace the regular convolution layers with a depthwise separable convolution block. Each block is composed of a depthwise convolution layer, non-linear layer, batch normalization layer, and pointwise convolution layer. The main idea was to separate the functionalities of the convolution layer which filters and combine features to make more reliable features. At the depthwise separable convolution block, the depthwise convolution filter the features then the pointwise convolution combines the channels. The depthwise separable convolution blocks ensure the same functionalities of the regular convolution layer but it is computed faster and uses fewer computation resources. An illustration of the depthwise separable convolution block is presented in Figure 4.

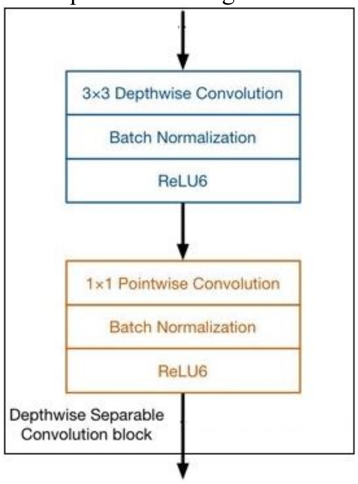

Fig. 4 Depthwise separable convolution block

Age estimation can be determined based on analyzing different facial landmarks. All facial landmark features are calculated and summed up to make age estimation. For this purpose, we propose to use a linear regression layer to make age predictions. Also, gender is used as additional features for age estimation since the facial landmark of a male may be different from the age of a female. The proposed CNN takes as input the grayscale image generated by the viola-jones algorithm. The grayscale image can be processed 3 times faster than the color image.

The architecture of the proposed CNN for gender and age estimation is presented in Figure 5.

The proposed approach is composed of 3 main blocks. The first block is a DCGAN used for resolution enhancement. The second block is the viola jones algorithm used for face detection and cropping. The third block is a CNN model used for gender classification and age estimation. The workflow of the proposed approach is illustrated in figure 6.

4. Experiment and Results

The experimental environment used in this work is based on a desktop with a Linux operating system occupied with an Intel i7 CPU and an Nvidia GTX960 GPU and 32 G of RAM. The CNN model and the DCGAN was built based on the TensorFlow deep learning framework. The viola-jones algorithm was based on the open CV library. Besides, the open CV was used for image manipulation and processing.

To train and evaluate the proposed approach, we propose to the IMDB-WIKI dataset [30]. It is a publicly available dataset that contains 523,051 images collected from the IMDB and Wikipedia sites. The dataset contains faces of 20,284 celebrities captured under real conditions. The images were collected in real life, movies, and festivals. All the images were annotated automatically by defining the age based on the date of birth and the date when the image was captured extracted from the IMDB or the Wikipedia sites. Also, the faces locations were defined and the gender of each face was assigned. The dataset was divided into 3 sets. Training set which is 60% of the dataset, a validation set which is 10% and a testing set which is 30% of the dataset.

Each lock of the proposed approach was trained separately since there is no way to train the system by single forward but at the same time where the output of each block was used as input of the next block. First, the DCGAN was trained to generate high-resolution images where all the generated images are stored in a temporary folder. Second, the viola-jones algorithm was trained to detect faces and crop the detected faces. Finally, the cropped faces were used to train the CNN model to predict gender and age.

For the evaluation of the proposed method, we used 2 metrics for age estimation and 1 metric for gender classification. For age estimation, the Mean Absolute Error (MAE) was reported which is the difference between the predicted age and the correct age, and the accuracy was reported as another evaluation metric. For gender classification, the accuracy was reported as an evaluation metric.

The proposed approach achieves an MAE of 2.6 years. To the best of our knowledge, all old methods except the DEX [30] do not achieve an MAE bellow to 3 years. For more convenience, the achieved accuracy was reported. Table 1 summarizes the achieved accuracy of the best state-of-the-art methods for age estimation and the proposed approach where all the mentioned methods where trained and tested on the IMD-WIKI dataset. As reported in table 1 the proposed approach achieves an accuracy for the exact age estimation of 72.2%. The achieved accuracy outperforms the best existing method with a big margin.

TABLE I. COMPARISON AGAINST STATE-OF-THE-ART APPROACHES FOR AGE ESTIMATION

Method	*Exact age (%)*	*1-off (%)*
DEX [30]	64	96.6
CNN [31]	50.7	84.7
Residual [22]	52.2	92
Pre-RoR-82+SD [17]	61.7	92.15
Ours	72.3	97.8

For the gender classification, the accuracy was reported as an evaluation metric for a fair comparison against state-of-the-art methods. Table 2 present the reported accuracy of the proposed approach achieved on the IMDB-WIKI dataset and the best existing methods. The proposed approach achieves an accuracy of 92.7% which outperforms state-of-the-art methods.

The proposed approach was designed to achieve high performance in terms of accuracy and processing time. A processing time of 20 FPS was achieved as processing time. None of the existing methods has reported its processing time. So, there is no possibility to compare our approach to existing works.

TABLE II. COMPARISON OF ACCURACY OF THE PROPOSED APPROACH AGAINST STATE-OF-THE-ART FOR GENDER CLASSIFICATION

Method	*Accuracy (%)*
VGG-Face [20]	95
CNN [31]	86.8
Residual [22]	88.5
Pre-RoR-82+SD [17]	90.87
Ours	97.2

The reported results prove the efficiency of the proposed approach. The use of the DCGAN has improved the accuracy by generating high-resolution images for further processing. The Viola-Jones algorithm was very important to detect faces and accelerating the processing time. The use of grayscale images as input for CNN accelerate the processing time and reduces the computation effort and complexity. In addition, the use of the lightweight was perfect to get a balance between processing speed and accuracy. The CNN model was pre-trained on the ImageNet dataset and this help to accelerate the training process.

The use of the DCGAN and the Viola-Jones with CNN have improved the performance. CNN can achieve good accuracy only of the input image contains a face with high resolution and cannot be applied to smart city surveillance. Images from surveillance cameras are captured from a long distance to the target person. So, enhancing the resolution and face detection is an important task for age and gender estimation in smart city surveillance.

The proposed approach was tested using images collected from the internet. Figure 7 presents an example of tested image. The proposed approach shows a high generalization power when tested on new images that do not exist in the training and testing dataset.

Fig. 7 Demo of the proposed approach for gender and age estimation

5. Conclusion

Gender and age estimation has been considered vital information for many purposes in smart city surveillance. In this work, we propose a gender and age estimation system based on the combination of a DCGAN, the viola-jones algorithm, and a CNN model. The DCGAN was used to enhance the resolution of the captured images, the Viola-Jones algorithm was used for face detection and cropping, and the CNN model based on a lightweight block of the MobileNet model was used for gender and age estimation.

The experimental results have proved the efficiency of the proposed approach. In terms of accuracy, the proposed approach achieved state-of-the-art performance while achieving a processing time of 20 FPS. The proposed approach is suitable for smart city surveillance and it can work on low-resolution images and achieves high performance. As future work, the proposed approach will be implemented on a real surveillance system.

Acknowledgments

The author wishes to acknowledge the help of Mr. Ayachi Riadh from Laboratory of Electronics and Microelectronics at University of Monastir for assistance with implementing the proposed model.

References

[1] Voulodimos, Athanasios, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. "Deep learning for computer vision: A brief review." Computational intelligence and neuroscience 2018 (2018).

[2] Riadh Ayachi, Yahia ElFahem Said, and Mohamed Atri. "To Perform Road Signs Recognition for Autonomous Vehicles Using Cascaded Deep Learning Pipeline." Artificial Intelligence Advances 1.1 (2019): 1-58.

[3] Aftf, Mouna, Riadh Ayachi, Yahia Said, Edwige Pissaloux, and Mohamed Atri. "Indoor object c1assification for autonomous navigation assistance based on deep CNN model." In 2019 IEEE International Symposium on Measurements & Networking (M&N), pp. 1-4. IEEE, 2019.

[4] Afif, Mouna, Riadh Ayachi, Yahia Said, and Mohamed Atri. "Deep Learning Based Application for Indoor Scene Recognition." Neural Processing Letters (2020): 1-11.

[5] Afif, Mouna, Riadh Ayachi, Yahia Said, Edwige Pissaloux, and Mohamed Atri. "Indoor image recognition and classification via deep convolutional neural network." In International conference on the Sciences of Electronics, Technologies of Information and Telecommunications, pp. 364-371. Springer, Cham, 2018.

[6] Ayachi, Riadh, Mouna Afif, Yahia Said, and Mohamed Atri. "Traffic signs detection for real-world application of an advanced driving assisting system using deep learning." Neural Processing Letters 51, No. 1 (2020): 837-851.

[7] Afif, Mouna, Riadh Ayachi, Yahia Said, Edwige Pissaloux, and Mohamed Atri. "An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation." Neural Processing Letters (2020): 1-15.

[8] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." In 2017 International Conference on Engineering and Technology (ICET), pp. 1-6. IEEE, 2017.

[9] Lindsay, Grace. "Convolutional neural networks as a model of the visual system: past, present, and future." Journal of Cognitive Neuroscience (2020): 1-15.

[10] Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). Generative Adversarial Networks. Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.

[11] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1, pp. I-I. IEEE, 2001.

[12] Huang, Jing, Yunyi Shang, and Hai Chen. "Improved Viola-Jones face detection algorithm based on HoloLens." EURASIP Journal on Image and Video Processing 2019, no. 1 (2019): 1-11.

[13] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

[14] Y. Kwon and N. Lobo, ‘‘Age classification from facial images,’’ in Proc. CVPR, 1994, pp. 762–767.

[15] A. Lanitis, C. Draganova, and C. Christodoulou, ‘‘Comparing different classifiers for automatic age estimation,’’ IEEE Trans. Syst., Man, B, Cybern., vol. 34, no. 1, pp. 621–628, Feb. 2004.

[16] Duan, Mingxing, Kenli Li, Canqun Yang, and Keqin Li. "A hybrid deep learning CNN–ELM for age and gender classification." Neurocomputing 275 (2018): 448-461.

[17] Zhang, Ke, Ce Gao, Liru Guo, Miao Sun, Xingfang Yuan, Tony X. Han, Zhenbing Zhao, and Baogang Li. "Age group and gender estimation in the wild with deep RoR architecture." IEEE Access 5 (2017): 22492-22503.

[18] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[19] Bhat, Salma Fayaz, and Taniya Ashraf Dar. "Gender Prediction from Images Using Deep Learning Techniques." In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1-6. IEEE, 2019.

[20] Dhomne, Amit, Ranjit Kumar, and Vijay Bhan. "Gender recognition through face using deep learning." Procedia Computer Science 132 (2018): 2-10.

[21] Cao, Qiong, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. "Vggface2: A dataset for recognising faces across pose and age." In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 67-74. IEEE, 2018.

[22] Lee, Seok Hee, Sepidehsadat Hosseini, Hyuk Jin Kwon, Jaewon Moon, Hyung Il Koo, and Nam Ik Cho. "Age and gender estimation using deep residual learning network." In 2018 International Workshop on Advanced Image Technology (IWAIT), pp. 1-3. IEEE, 2018.

[23] Aslam, Aasma, Khizar Hayat, Arif Iqbal Umar, Bahman Zohuri, Payman Zarkesh-Ha, David Modissette, Sahib Zar Khan, and Babar Hussian. "Wavelet-based convolutional neural networks for gender classification." Journal of Electronic Imaging 28, no. 1 (2019): 013012.

[24] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105. 2012.

[25] Yu, Yang, Zhiqiang Gong, Ping Zhong, and Jiaxin Shan. "Unsupervised representation learning with deep convolutional neural network for remote sensing images." In International Conference on Image and Graphics, pp. 97-108. Springer, Cham, 2017.

[26] Ayachi, Riadh, Mouna Afif, Yahia Said, and Mohamed Atri. "Strided convolution instead of max pooling for memory efficiency of convolutional neural networks." In International conference on the Sciences of Electronics, Technologies of Information and Telecommunications, pp. 234-243. Springer, Cham, 2018.

[27] Mita, Takeshi, Toshimitsu Kaneko, and Osamu Hori. "Joint haar-like features for face detection." In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, pp. 1619-1626. IEEE, 2005.

[28] Afif, Mouna, Yahia Said, and Mohamed Atri. "Efficient implementation of integrall image algorithm on NVIDIA CUDA." In 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), pp. 1-5. IEEE, 2018.

[29] Schapire, Robert E. "Explaining adaboost." In Empirical inference, pp. 37-52. Springer, Berlin, Heidelberg, 2013.

[30] Rothe, Rasmus, Radu Timofte, and Luc Van Gool. "Deep expectation of real and apparent age from a single image without facial landmarks." International Journal of Computer Vision 126, no. 2-4 (2018): 144-157.

[31] Gil Levi and Tal Hassner, “Age and Gender Classification using Convolutional Neural Networks”, IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), at the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, June 2015.

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 Embedded Vision, Image and Video Processing, SoC Design, Pattern Recognition, Artificial Intelligence, Deep Learning and Convolutional Neural Networks.