Journal of Artificial Intelligence and Data Science (JAIDA) Vol. 1, No. 1, pp. 28-34, 2021
Research Article
https://jaida.ikcu.edu.tr/

A Face Authentication System Using Landmark Detection

Velican ERCAN, M. Erdal ÖZBEK*

İzmir Kâtip Çelebi University, Faculty of Engineering and Architecture,
Department of Electrical and Electronics Engineering, İzmir, Turkey

Abstract

Biometric data is the key for many security applications. Authentication relies on the individual's measurable biometric properties collected as features. In this study, a face authentication system is built to be used in opening the entrance door accessing to the apartments and housing estates. The proposed system consists of three stages. In the first stage, landmarks on the face are captured using a deep neural network. Then six selected features from the landmarks are extracted and traditional machine learning algorithms are used to authenticate users. In the last stage, a user interface is built. Face recognition tests achieved an accuracy rate of 89.79%.
Keywords: Biometrics, face authentication, facial recognition, landmark detection.

1. Introduction

The physiological or behavioral characteristics of an individual is referred as biometrics [1]. They are preferred in security systems or applications, due to their advantages over traditional systems in authentication. They are unique and inseparable for an individual, thus they cannot be lost nor forgotten. Therefore, it is the effective authentication method that allows authenticating the user directly, rather than her/his smartphone, smart card, or any secret she/he knows. Face is one of the physiological characteristics which is now frequently used in smartphone unlock screens [2].

The first facial recognition system was designed in the 1960s under the leadership of Woody Bledsoe. The system, which could identify forty people per hour at that time [3], can now recognize people in times measured in milliseconds. A facial recognition system is a technology capable of matching a human face from a digital image or a video frame against a database of faces, which works by pinpointing and measuring facial features from a given image or video [4]. In the early 1970s, facial recognition was considered as a two-dimensional pattern recognition problem [5, 6]. A recent review of the technology discussing the future development direction can be found in [7].

Basically, there are three steps in the face recognition process. These are face detection, extraction of attributes, and recognition. Face detection is defined as the process of extracting face/faces from images or frames. It is difficult to determine the detection methods precisely because most of the algorithms are designed to work with other methods, e.g., object detection, in order to improve the performance computed based on accuracy and/or speed [8]. Generally, detection methods can be divided into two categories: knowledge-based and image-based [9]. Knowledge-based methods are rule-based methods. These methods use information about face features, skin colors or template matching. Some rules are easy to guess, e.g., a face generally has two symmetrical eyes, and the eyes are darker than the cheeks. The image-based method uses a training-learning approach to make comparisons between face and non-face images. These kinds of methods have to be trained with large number of face and non-face images to increase the accuracy of this system. There are various face detection algorithms in the literature using Eigenface, AdaBoost, Neural Networks, and Support Vector Machines (SVMs) [10-13].

Recognition consists of preprocessing the face image, vectorizing the image matrices, creating a database of images, and finally classifying the image. Because face recognition is a pattern recognition problem, a learning method should be utilized to compare faces. The most common methods for two-dimensional recognition are Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Gabor wavelet, and spectral feature analysis [14, 15]. The two-dimensional methods generate features for each individual and then classify them. It is also important to generate abstract feature vectors in order to reduce the size of the input images. Lighting is another important aspect where image enhancement techniques are used to suppress the bad lighting conditions by a logarithm transformation and normalization [9]. The illumination difference between the left and right sides of the face is eliminated by using the mirror of the average illuminated part [16]. Promising results using deep learning approaches on different conditions such as lower and upper face occlusions, misalignment, varying head posture angles have been reported [17-19].

The aim of this study is to strengthen the security in the entrance of the apartments with a biometric system. Biometrics are basically measurable biological traces of a person. By considering the biometric data as the basis

*Corresponding author e-mail address: merdal.ozbek@ikcu.edu.tr

Received: 16/07/2021; Accepted: 25/08/2021

of the face recognition system, landmarks are extracted using deep neural networks. Authentication is based on the extracted features from the landmarks where machine learning algorithms are used to establish identification. A simple user interface is also generated for ease of common usage.

The remainder of the paper is organized as follows: Section 2 explains the system architecture and implementation of the proposed face recognition system including the landmark extraction and user interface. Section 3 reports the accuracy performance of the system. Section 4 summarizes and concludes the paper.

2. Methodology

This section presents the information about the proposed system, determining landmarks, extraction of features from landmarks, and the implementation of the graphical user interface (GUI).

2.1. Landmark detection

The semantic meanings of facial contours and facial components are delivered via facial landmarks positioned around those contours and components, as well as the subject's face [20]. Facial landmark detection has attracted a lot of interest in the field of computer vision as a crucial step for a variety of face-related applications [21]. It can be basically stated as predicting the coordinates of the center of eyes, nose tip, eyebrows, and mouth for a given face. The most important step of the facial recognition system is the determination of the key points on the face.

Improvements in landmark detection have been made recently thanks to the use of convolutional neural networks (CNNs). Because of its fault tolerance and self-learning capabilities, this type of feedforward neural network with shared weights and local connections has been widely used in the field of image processing. Therefore, in this study, landmark detection is applied by deep CNN models.

In literature, an efficient CNN model approach for facial key points detection has been offered [22]. In this approach, they augmented the image data with horizontally flipping, then vertically stacked. The image is pre-processed with the pixels by normalizing in the range of [0,1], and then the training targets are zero centered to take the values in the range of [-1,1]. They have created fifteen different so-called NaimishNet models. Each of the fifteen different models has been filtered out the non-missing values. They have shown that the predicted locations are very close to the actual locations.

The facial key points detection model of this study is based on the NaimishNet model. The dataset used in the training of the landmark model belongs to Dr. Yoshua Bengio [23]. The dataset is augmented in order to artificially increase the number of training samples. However, it is found that in NaimishNet model, insufficient light affects the performance as shown in Figure 1.

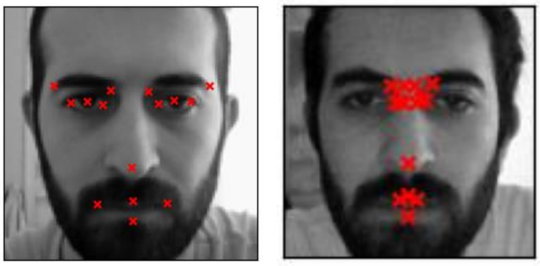

Figure 1. *Images with normal (left) and low light (right) conditions with detected key points.*

Therefore, we propose a 45-layer CNN model instead of 25-layer, and train models with artificially manipulated brightness values in order to increase accuracy in any poor light conditions while augmenting data. The proposed CNN architecture is shown in Figure 2.

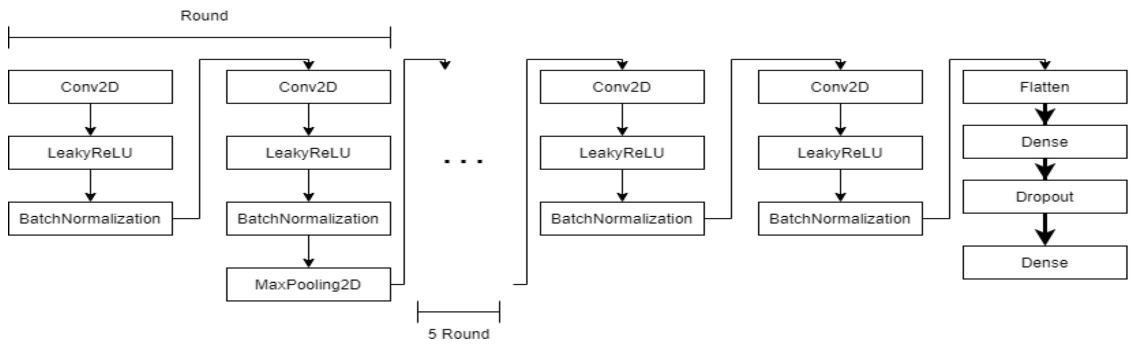

Figure 2. *CNN architecture.*

The proposed architecture is different from the NaimishNet architecture. This model includes additional LeakyReLU and BatchNormalization layers after each of the Conv2D layers. The number of total model parameters is 7,264,318. In order to avoid overfitting, the number of epochs for the training of the neural network was determined as 100, where the validation and training error values were started to increase.

2.2. Feature extraction

The landmark model returns an array of coordinates which are then converted to distances including the golden ratio [24], which were also selected in previous studies [25, 26]. The features are extracted based on the key points by proportioning the calculated distances to each other. In this study, two different approaches are considered for feature extraction, with twenty features and six features, respectively. The first approach with twenty features was found too complex for standard machine learning algorithms. Especially if the number of data (image dataset in this study) is not sufficient, the algorithm may not be able to learn and may memorize the data what is known as "overfitting" [27]. In cases of overfitting, very high accuracy values may be observed when training, however very low scores will most likely occur when testing. Besides, similar features are highly correlated. For example, the eye center distance and mouth midpoints are reflected in five features, while nose to eye points are reflected in three features. Therefore, it affects three or five features for each unit pixel estimation that will be found faulty. This may end up with inconsistent results, leading the trained model to make inaccurate predictions.

The features are reduced to six in order to avoid overfitting and to satisfy that the landmark model predicts with the higher accuracy. For example, the deviation in the pupil is more than that in the eyebrow ends. These ratios given in Table 1 are also expected to be less than one, eliminating the necessity of normalizing the data.

Table 1. *Extracted features.*

Feature name	Feature
f1	Eyebrow Mid to Nose / Eyebrow Left Out to Mouth Left
f2	Nose to Mouth Right / Eyebrow Distance
f3	Mouth Width / Eyebrow Mid to Nose
f4	Eyebrow Distance / Eyebrow Right In to Mouth Right
f5	Nose to Mouth Left / Nose tip to Eyebrow Mid
f6	Right Eyebrow / Mouth Width

2.3. Proposed face recognition system

After the extraction of six features from the ratios of the landmarks, the collected biometric data is used to perform the recognition step. The traditional machine learning algorithms are available to be used but SVM is found to have the best performance.

The basis of the SVM algorithm used for recognition in this study is based on finding the optimal hyperplane of the separable two classes with the maximum margin [28]. The goal of classifying is to find a line that divides the input space into distinct areas. Assuming the independent and identically distributed input-output training data pair (x, y) follows an unknown probability distribution,

$$(x_1, y_1), \dots, (x_n, y_n) \in R^N, y_i = \{-1, +1\} \tag{1}$$

If the training data can be separated by a hyperplane, the margin is the shortest distance between a sample and the decision hyperplane. A kernel function is used to translate the input data to a high-dimensional feature space, and then a linear classifier is created in this space to solve the following optimization problem

$$min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \tag{2}$$

where w is the weight vector, and b is a scalar (bias), subject to,

$$y_i(\varphi(x_i)w + b) \geq 1 - \xi_i, \quad \xi \geq 0, \; i = 1, \dots, n, \quad C > 0 \tag{3}$$

where φ is a function mapping the input data to feature space, ξ_i is slack variable and C is the trade-off parameter between the error and margin. Then, the discriminant function can be used to predict the label of unknown data using a kernel function $k(x_i, x)$ by

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b) \tag{4}$$

where α_i is the Lagrange multiplier.

As a summary, we propose a system consisting of three essential stages, including deep CNN for landmark extraction, six features derived from the landmarks, and machine learning algorithms, in particular SVM, for recognition and authentication. Figure 3 illustrates the proposed system architecture.

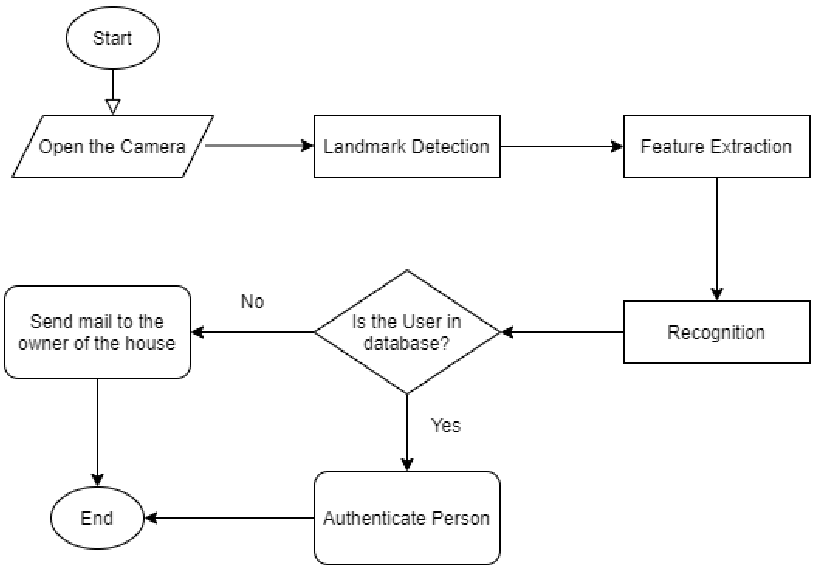

Figure 3. *Proposed system architecture.*

2.4. Graphical user interface

A graphical user interface is built to merge all the processes for simplicity and ease of use. For this purpose, PyQt5, a set of Python bindings is considered. It consists of over thirty-five extension modules [29] that allow Python to be used as an alternative to C++ for application development on all supported platforms, including iOS and Android. The interface consists of three main pages. The first page allows the user to create a dataset. The camera is deployed in the middle captures images and a new folder is created with a hundred cropped gray-level face photos that are taken. In the second page, the model can be trained by selecting a machine learning algorithm with the users added to the database, later to be used in the authentication. The classification accuracy, confusion matrix, and classification report can be observed in the lower right corner. If the model is trained as desired without overfitting or underfitting, the model can be saved as shown in Figure 4.

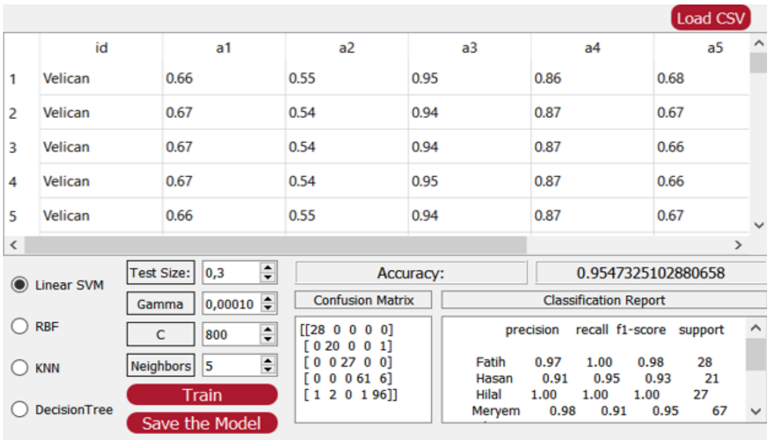

Figure 4. *Training page.*

The third page is designed for authentication as an automatic door lock. As shown in Figure 5, there is a live camera view on the left, and there is a keypad and a menu on the right, where the trained machine learning algorithm can be selected. If the user is known and recognized as a valid user, the text “Welcome User” will be displayed. If the user is not in the database and is accepted as an intruder, the text “Access Denied” will appear in the text box and the photo of the person trying to log in is sent to the e-mail address of the host, with the time and date information included.

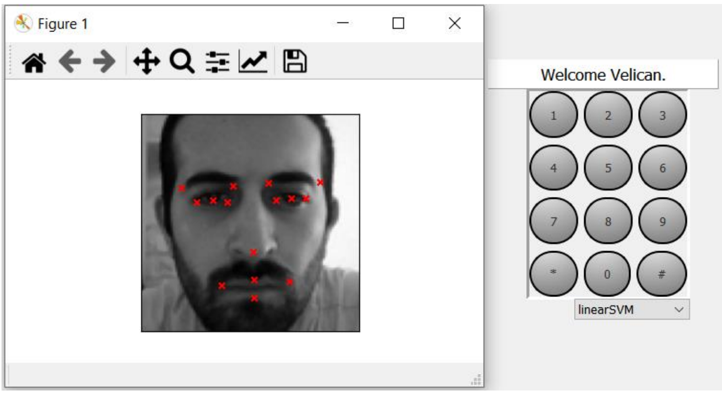

Figure 5. *Authentication page.*

3. Results

The deep CNN experiments were made with NVIDIA Tesla K80 Graphics Processing Unit on the Google Colaboratory cloud system. The deep CNN model used in landmark detection was trained by monitoring the mean absolute error and accuracy values. Loss functions are used to determine how well an algorithm matches the data it is trained on. The difference between the anticipated and actual values is used to calculate the loss. The loss function will generate a very big number if the projected values are distant from the actual values. The basic goal of a learning model is to change the weight vector values using different optimization approaches, such as backpropagation in neural networks, to minimize the loss function's value with regard to the model's parameters. At the end of the 100 epoch, the best loss values in the training set were 2.779 while 1.1952 in the validation set, respectively. Figure 6 shows the training and validation loss variations, and it is seen that 100 epochs were enough for the models to converge.

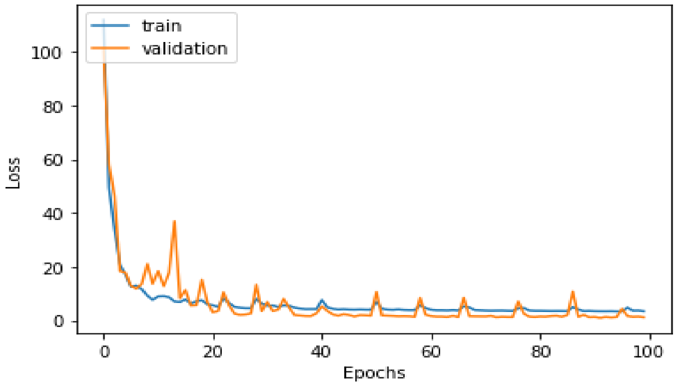

Figure 6. *Loss for training and validation.*

In the tests, the key points from wide angles were detected with high accuracy even with the moving head and obtained consistent results. This model, which is quite ahead of other models, can make consistent predictions in changing conditions. The importance of consistency of results is vital because the landmarks that do not change in the same pose would be helpful in the face recognition part. The test results of the trained landmark model are shown in Figure 7.

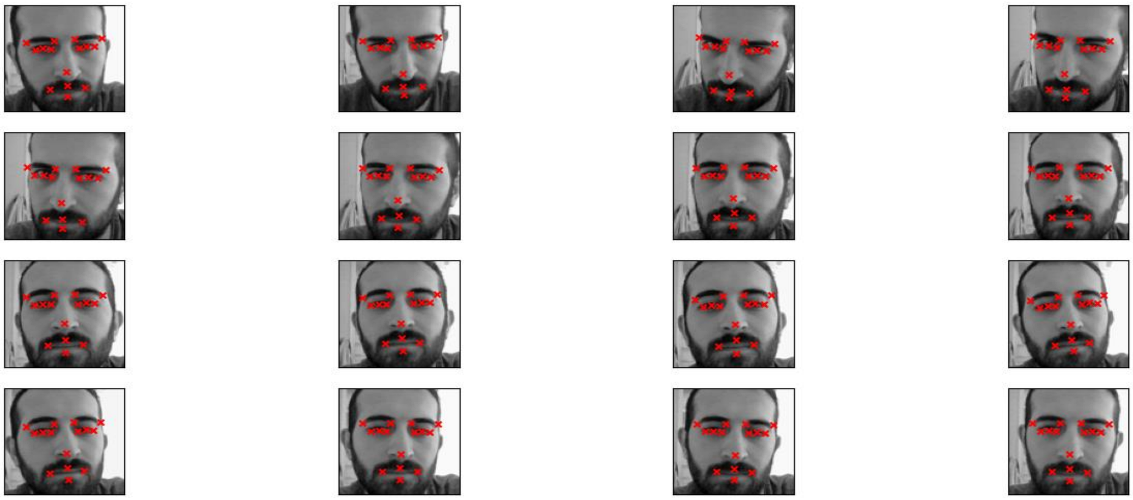

Figure 7. *Landmark model test results.*

The proposed system is developed and tested in Python Software, on an eight-core Intel i5 2.50 GHz computer on a database consisting of images of five people and forty-nine test photos with a resolution of 200x200 pixels. All the test photos were taken at a distance of fifty-five cm from the camera and in the same pose. Five of the results from forty-nine test photos independent of the training data are wrongly classified. Therefore, the recognition rate is obtained as 89.79% with a linear SVM kernel.

4. Conclusions

Artificial intelligence, machine learning, and deep learning started to change our lives day by day and shape our future. This new ecosystem touches every aspect of our lives, disrupts our habits, and makes our work easier. Facial recognition systems are becoming more important as they become a part of biometric systems, where a physical need such as a key is ended, time is saved and security is increased by providing instant information to the host.

The main purpose of this work is to build and implement a facial recognition system to be used in automatic door unlocking. Face detection and recognition procedures were used to determine the landmarks of the face by the deep learning model. The biometric data was obtained by using the distances and ratios between the landmarks, and the authentication was carried out by the recognition process performed by the machine learning model. As facial expression, imaging condition, and poor illumination can affect the proposed system, augmentation of data to combat these conditions were proposed. The realization of the recognition was processed in a very short time and the rapid provision of security constitute the advantages of the system. On the other hand, misrecognition of the face due to different pose angles was found as the major disadvantage. The solution was to take an image at the same distance and at the same angle. In this way, the system has achieved an accuracy of 89.79%.

In the future this study might be extended to be used in banks, in ATM's, for identification of users, or in factories for workers' entry pass as of today most of the working places still use signature or card based systems. Moreover, this system can be set out to be used in classrooms for attendance control with crowd analysis.

Declaration of Interest

The authors declare that there is no conflict of interest.

References

[1] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," IEEE Transactions on Information Forensics and Security, vol. 1, no. 2, pp. 125-143, June 2006.
[2] R. Spolaor, Q. Li, M. Monaro, M. Conti, L. Gamberini, and G. Sartori, "Biometric authentication methods on smartphones: A survey," PsychNology Journal, vol. 14, no. 2-3, pp. 87-98, 2016.
[3] S. B. Thorat, S. K. Nayak, and J. P. Dandale, "Facial recognition technology: An analysis with scope in India," arXiv preprint, arXiv:1005.4263, 2010.
[4] C. A. Hansen, Face Recognition, Institute for Computer Science University of Tromso, Norway, 2009.
[5] A. K. Jain and S. Z. Li, Handbook of Face Recognition, vol. 1, New York: Springer, 2011.
[6] Wikipedia, The Free Encyclopedia, [online] Available at: https://en.wikipedia.org/wiki/Facial_recognition_system, [Accessed: June 21, 2021].
[7] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," IEEE Access, vol. 8, pp. 139110-139120, 2020.
[8] C. Li, R. Wang, J. Li, and L. Fei, "Face detection based on YOLOv3," Recent Trends in Intelligent Computing, Communication and Devices, vol. 1006, pp. 277-284, 2020.
[9] C. Gürel, Development of a Face Recognition System, Master of Science Thesis, Atilim University, 2011.
[10] M. Turk, M and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
[11] R. E. Schapire, Explaining Adaboost, In: Empirical Inference, Springer, Berlin, Heidelberg, pp. 37-52, 2013.
[12] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, 1998.
[13] W. S. Noble, "What is a support vector machine?," Nature Biotechnology, vol. 24, no. 12, pp. 1565-1567, 2006.
[14] C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," IEEE Transactions on Neural Networks, vol. 14, pp. 919-928, 2003.
[15] A. H. Boualleg, C. Bencheriet, and H. Tebbikh, "Automatic face recognition using neural network-PCA," In Proc. 2nd Information and Communication Technologies (ICTTA), Damascus, Syria, 2006, pp. 1920-1925.
[16] Y. Song, Y. Kim, U. Chang, and H. B. Kwon, "Face recognition robust to left-right shadows facial symmetry," Pattern Recognition, vol. 39, pp. 1542-1545, 2006.
[17] P. S. Prasad, R. Pathak, V. K. Gunjan, and H. V. R. Rao, "Deep learning based representation for face recognition," ICCCE 2019, Lecture Notes in Electrical Engineering, vol. 570, Springer, Singapore, 2020, pp. 419-424.
[18] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," In X. Xie, M. W. Jones, and G. K. L. Tam, editors, Proceedings of the British Machine Vision Conference (BMVC), 2015, pp. 41.1-41.12.
[19] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2884-2896, 2018.
[20] W. J. Baddar, J. Son, D. H. Kim, S. T. Kim, and Y. M. Ro, "A deep facial landmarks detection with facial contour and facial components constraint," In IEEE International Conference on Image Processing (ICIP), 2016, pp. 3209-3213.
[21] M. Köstinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," In IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 2144-2151.
[22] N. Agarwal, A. Krohn-Grimberghe, and R. Vyas, "Facial key points detection using deep convolutional neural network-NaimishNet," arXiv preprint, arXiv:1710.00977, 2017.
[23] Y. Bengio, "Facial keypoints detection," [Online] Available at: https://www.kaggle.com/c/facial-keypoints-detection, [Accessed: August 21, 2021].
[24] E. P. Prokopakis, I. M. Vlastos, V. A. Picavet, T. G. Nolst, R. Thomas, C. Cingi, and P. W. Hellings, "The golden ratio in facial symmetry," Rhinology, vol. 51, no. 1, pp. 18-21, 2013.
[25] P. S. Gaikwad, V. B. Kulkarni, "Face recognition using golden ratio for door access control system," In: Merchant S. N., Warhade K., Adhikari D. (eds), Advances in Signal and Data Processing, Lecture Notes in Electrical Engineering, vol. 703. Springer, Singapore, 2021, pp. 209-231.
[26] R. Feng and B. Prabhakaran, "A novel method for post-surgery face recognition using sum of facial parts recognition," IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 1082-1089.
[27] T. Dietterich, "Overfitting and undercomputing in machine learning," ACM Computing Surveys, vol. 27, no. 3, pp. 326-327, 1995.
[28] N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.
[29] Python bindings for the Qt cross platform application toolkit, [online] https://pypi.org/project/PyQt5/, [Accessed: June 27, 2021].