PAPER • **OPEN ACCESS**

Face recognition using multitasking cascading convolutional networks

To cite this article: A G Musikhin and S Yu Burenin 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1155** 012057

Face recognition using multitasking cascading convolutional networks

A G Musikhin[1] and S Yu Burenin

MIREA-Russian Technological University, Moscow, 119454, Russia

[1]E-mail: mag.iit.kvt@gmail.com

Abstract. The face recognition has many applications in such areas like intelligent security and access control, biometrics, safeguard, verification, attendance accounting, machine vision, etc. Identification of a personality by the face has a number of advantages over other methods: the person does not need to be contacted physically, which is the most acceptable way for mass applications and no complicated or expensive equipment is required. This article discusses the problem of recognition and identification of a person's face using convolutional neural networks that process frames from a camera in real time or from a recorded video file with the subsequent entry of the identified person into the database. Multitasking Cascade Convolutional Neural Network (MTCNN) has three convolutional networks (P-Net, R-Net, and O-Net) and is capable of outperforming many face detection tests while maintaining real-time performance. The proposed method for human face recognition was developed as a software product, tested and showed the probability of correct recognition in real time 96.02%.

1. Introduction

There is a growing interest in technology face recognition, whose task is to automatically localize faces on video and person identification. The concept of "recognition" can be determined as the assignment of the studied object to one of the known classes. "Identification" is the establishment of identity of the unknown object to the known one basing on the coincidence of recognized features.

The face recognition has many applications in such areas like intelligent security and access control, biometrics, safeguard, verification, attendance accounting, machine vision, etc. Traditional ways of identification such as keys or passwords do not provide high degree of reliability, therefore, the biometric identification systems are being developed dynamically. Identification of a personality by the face has a number of advantages over other methods:

- the person does not need to be contacted physically, which is the most acceptable way for mass applications;
- no complicated or expensive equipment is required.

Various techniques are used to solve the recognition problem, such as based on neural networks, on Karhunen–Loève decomposition, on Haar cascades, on the equal intensity lines, on the algebraic moments, on the elastic (deformable) samples.

The main problem of the most face recognition methods is the dependence of the angle, position, lighting conditions, etc. This problem can be minimized using the neural networks based methods.

For producing an automated system of the person identification and the attendance accounting Chintalapati et al. [1] in 2013 proposed a method based on the Viola Jones method [2] for face detection followed by the equalization histogram for feature extraction and the support vector machine (SVM) classifier for face recognition. Later in 2017 Rathod et al. [3] proposed an automated system of the same technique for face detection and a classifier for face recognition. However, these methods were based on classical machine algorithms. Arsenovic et al. [4] in 2017 proposed a modern FaceTime method using of cascaded convolutional neural network for face detection and convolutional neural network for facial features generation, which were then used for recognition. In this article proposed the structure of multi-task cascaded convolutional networks (MTCNN) [5] for face recognition and Siamese neural network [6] to extract the traits of the faces to create a 128- element vector representing these traits. Most commercial automated systems software use images of faces that should be taken close-up distance and which main condition for the successful work is an open, well-lit face. Image processing is carried out sequentially, which is a rather lengthy process.

2. Materials and methods

This article discusses the problem of recognition and identification of a person's face using convolutional neural networks that process frames from a camera in real time or from a recorded video file with the subsequent entry of the identified person into the database.

One of the main advantages of the proposed solution is robustness of the algorythm against common problems such as occlusion (partially visible / obscured faces), orientation of the face and the influence of room lighting.

The software processes the video file (frames from CCTV cameras) and detects faces using a new method based on the MTCNN [5], recognizes the detected faces using the Siamese neural network [6], which learns to transform facial images into a compact Euclidean space, where the distance corresponds to the measure similarity of faces, and adds the result of face recognition to the database.

Multitasking Cascade Convolutional Neural Network (MTCNN), shown in figure 1, has three convolutional networks (P-Net, R-Net, and O-Net) and is capable of outperforming many face detection tests while maintaining real-time performance. An image is taken and resized to different scales to build a pyramid of images, which is the input to the next three-stage cascading network.

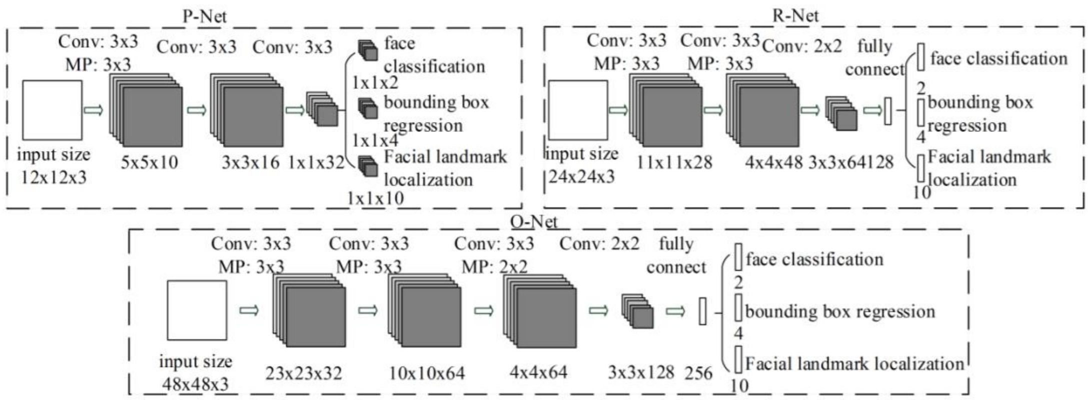

Figure 1. MTCNN architecture.

Three stage of the MTCNN:

- The Proposal Network (P-Net) This first stage is FCN (Fully Convolutional Network). The difference between CNN (Convolutional Neural Network) and FCN is that a fully convolutional network does not use the dense layer as part of the architecture. P-Net is used to obtain candidate windows and their bounding box regression vectors.

- The Refine Network (R-Net) All candidates from P-Net go to R-Net. This network is a CNN and not an FCN like the previous one, since there is a dense layer at the last stage of the network architecture. R-Net outputs the decision of whether the image is a face or not, in the form of a 4-element vector, which is the bounding box for the face, and a 10-element vector for localizing the facial landmarks.
- The Output Network (O-Net) This step is similar to R-Net, but this output network aims to describe the face in more detail and output the positions of the five facial landmarks for the eyes, nose, and mouth.

The MTCNN's job is to deduce three things: classification (face/non-face), bounding box coordinates, and facial landmarks location. The Siamese network shown in figure 2 is a type of a neural network architecture that learns to differentiate input data. That is, it allows you to learn to understand which images are similar and which are not. Siamese networks are made up of two identical neural networks, each with the same exact weights. Each network takes one of two images as input. The outputs of the last layers of each network are then sent to a function that determines if the images contain the same IDs. This is done by calculating the distance between the two outputs. The Siamese network is used to extract quality facial features and create a 128-element vector to represent those features.

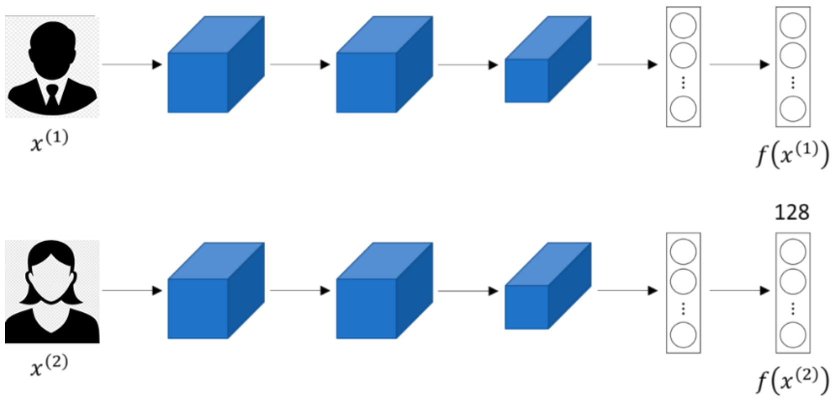

Figure 2. Siamese network.

The operation of the software includes a classifier training process and an identification process. When trained (figure 3), the input video is passed through the MTCNN, which gives the coordinates of a bounding box around the face. With the help of coordinates, the face is cut out of the frame and transferred to the change, including flipping, blurring, adding noise, horizontal displacement, vertical displacement, increasing and decreasing the brightness of the image. The resized frames are sized to 160 by 160 pixels and are saved to the folder corresponding to the person ID. After creating the database, it is necessary to train the Siamese network on it. The Siamese network model is trained until the classification accuracy is 99%. After training, the data file is saved and used later for identification tasks.

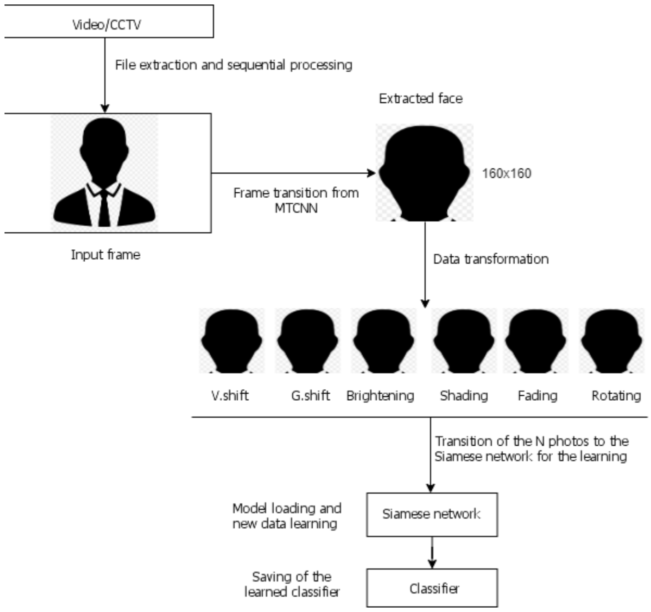

Figure 3. The learning process scheme.

During identification (figure 4) the video frame passes through the MTCNN, which returns a twodimensional array containing the coordinates of the bounding boxes of all detected faces. Then the faces are extracted from the frame and, after normalizing the size, are processed for recognition using a trained classifier. Confidence scores corresponding to all classes are recorded, the decision on whether a person belongs to one of the classes is made according to the maximum credibility score when the set threshold is exceeded and added to the list of recognized faces. If the set threshold is not exceeded, a decision is made about the presence of an unknown person. The list containing the names of the recognized persons is then transferred to the attendance accounting module. This module creates an attendance file and adds the names of people from the database to it. Next, the module creates a new column in the file with the current date and adds a mark and time of the visit of the recognized person.

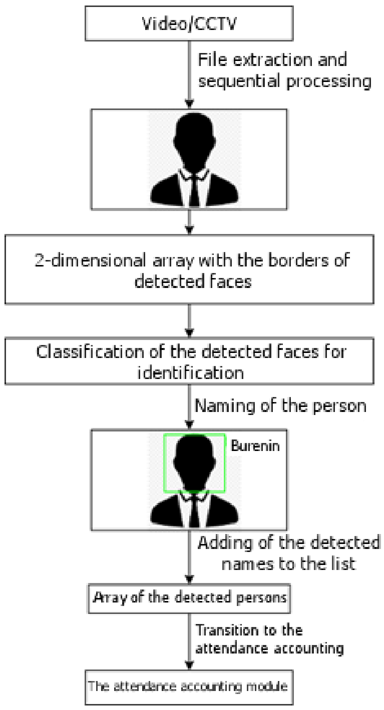

Figure 4. The identification and the attendance registration process scheme.

3. Results and discussion

The proposed method for human face recognition was developed as a software product, tested on a dataset of 10,000 images of 20 people, that is, 500 images per person, and showed the probability of correct recognition in real time 96.02% with the probability of false recognition less than 10%. This result exceeds indicator of other automated attendance recording systems used in practice (table 1).

Table 1. Comparison with other methods.

Number	Method	Accuracy
1	Chintalapati et al. [1]	95.00%
2	FaceTime [4]	95.02%
3	Proposed method	96.02%

4. Conclusion

The proposed system is sensitive to the distance to the faces in the video, as well as to video with low resolution. Further work may include improved face detection and recognition independent of video resolution by preprocessing with an image resolution enhancer.

References

[1] Chintalapati S and Raghunadh M V 2013 Automated attendance management system based on face recognition algorithms *IEEE International Conference on Computational Intelligence and Computing Research.* Enathi, India 1-5

[2] Paul V and Jones M J 2004 Robust real-time face detection *International Journal of Computer Vision* **57(2)** 137-54

[3] Rathod, Hemantkumar, Yudhisthir Ware, Snehal Sane, Suresh Raulo, Vishal Pakhare and Imdad A Rizvi 2017 Automated attendance system using machine learning approach *IEEE Explore*

[4] Arsenovic M, Sladojevic S, Anderla A and Stefanovic D 2017 FaceTime – Deep learning based face recognition attendance system *Conference: – IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY) At* Subotica, Serbia

[5] Zhang Kaipeng, Zhanpeng Zhang, Zhifeng Li and Yu Qiao 2016 *Joint face detection and alignment using multitask cascaded convolutional networks* [Electronic resource] 1499-503 Link : https://arxiv.org/ftp/arxiv/papers/1604/1604.02878.pdf

[6] Koch G, Zemel R and Salakhutdinov R 2015 Siamese Neural Networks for One-shot Image Recognition *Proceedings of the32nd International Conference on Machine Learning* Lille France (JMLR: W&CP **37**) p 3-6