FaceTime – Deep Learning Based Face Recognition Attendance System

Marko Arsenovic, Srdjan Sladojevic, Andras Anderla, Darko Stefanovic
University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
arsenovic@uns.ac.rs, sladojevic@uns.ac.rs, andras@uns.ac.rs, darkoste@uns.ac.rs

***Abstract*— In the interest of recent accomplishments in the development of deep convolutional neural networks (CNNs) for face detection and recognition tasks, a new deep learning based face recognition attendance system is proposed in this paper. The entire process of developing a face recognition model is described in detail. This model is composed of several essential steps developed using today's most advanced techniques: CNN cascade for face detection and CNN for generating face embeddings. The primary goal of this research was the practical employment of these state-of-the-art deep learning approaches for face recognition tasks. Due to the fact that CNNs achieve the best results for larger datasets, which is not the case in production environment, the main challenge was applying these methods on smaller datasets. A new approach for image augmentation for face recognition tasks is proposed. The overall accuracy was 95.02% on a small dataset of the original face images of employees in the real-time environment. The proposed face recognition model could be integrated in another system with or without some minor alternations as a supporting or a main component for monitoring purposes.**

Keywords—face recognition; deep learning, attendance system;

I. Introduction

One necessary component of every business system is recording employees' work hours and activities, despite the capacity of the system. This process could be time consuming if it is managed manually. As a result of a rapid growth in information technologies, automatic solutions have become a standard option for these types of business processes.

There are now plenty of systems which differ in many aspects: core technology they are based on, way of use, cost, reliability, security and etc. Many of those depend on employees having to carry specific identification devices. One of the common types of the attendance systems is Radio Frequency Identification (RFID) where employees have to carry appropriate RFID cards. There are also location based attendance tracking systems. The location of an employee can be determined via Global Positioning System (GPS). The presence is determined by calculating the proximity between an employee's and the company's location. Both of the above mentioned types of the attendance systems have weaknesses. Employees could forget the RFID card or the location device, or someone else could check instead of them. This could also be a potential security issue. Therefore, there are systems that exclude the usage of external devices for attendance purposes by exploiting the individual attributes: fingerprints, iris, voice, face and etc. These types of systems are heavily based on computer vision and machine learning algorithms. Recent advances in these areas, especially in deep learning, provide possibilities to use these methods searching for practical solutions. These solutions could be more flexible and could reduce human errors.

The method proposed in this paper provides solution for face recognition tasks combining various modern approaches and state-of-the-art crafts in deep learning.

The rest of the paper is organized as follows: Section II presents the related work, Section III presents the methodology, Section IV presents the results and discussion, and finally, Section V holds the conclusion.

II. Related Work

As a result of the active progress in software technologies, there are now many different types of computerized monitoring and attendance systems applied in companies. These systems mostly differ in the core technology they use. The authors' previous work [1] introduced one solution for developing the RFID based type of attendance system. Employees' entrances and exits records are gathered using cards and RFID reader devices which send data via GPRS to the remote server, where it is, then, stored in the database. This data could be accessed by a web application for authenticated users. Similar RFID based systems are proposed by Sharma et al. in [2]. Sultana et al. in [3] proposed a location based attendance tracking system using an Android for extracting the GPS data. Rao et al. in [4] presented an attendance system using biometrics authentications. They used a common minutiae and a pattern based matching for fingerprint verification in order to accurately distinguish the identity of the people whose attendance was logging. Soewito et al. in [5] used a smartphone. They integrated both the location and individual attributes in order to accurately track attendances. Their system uses fingerprint or voice recognition. Within the application, the user sends the GPS coordinates, date and time along with a fingerprint or voice to the server. Minutiae and texture feature matching algorithms are applied for fingerprint recognition. A voice recognition algorithm uses spectrogram or voiceprint which is converted from the electronic signal to a voice that matches the template voices stored in the database. The tested system achieved the accuracy of 95%. Kadry et al. in the paper [6] presented the attendance wireless system that is based on iris recognition.

978-1-5386-3855-2/17/$31.00 ©2017 IEEE

The system uses an eye scan sensor and Daugman's algorithm for the iris recognition.

Patil et al. in [7] applied face recognition for classroom attendance. They used Eigenface for the recognition, but the overall accuracy of the system was not mentioned in the paper. Similar approach using Eigenface for the face recognition based attendance tracking system was proposed in [8]. They achieved overall recognition accuracy of 85% for unveil faces. Tharanga et al. in [9] used Principle Component Analysis (PCA) method for the face recognition for their attendance system, achieving the accuracy of 68%.

Due to the rapid progress in deep learning, the accuracy of face recognition is drastically improved by the usage of deep CNNs. Schroff et al. in [10] presented the revolutionary system – FaceNet which depends on the Deep Neural Network (DNN) for the face recognition task. The proposed method achieved astonishing results on the Labeled Faces in the Wild (LFW) dataset, 99.63% accuracy.

Motivated by these results, the authors of this paper decided to use an alternated version of this approach as part of a model for the deep learning based face recognition attendance system.

III. Materials and Methods

The whole method of developing the deep learning based attendance system is explained in detail in this section. The developing procedure is divided into several important stages, including obtaining the training dataset and augmentation, preparing images and training DNNs and last but not least, integration into the existing system in order to test the proposed method.

A. Dataset preparation and augmentation

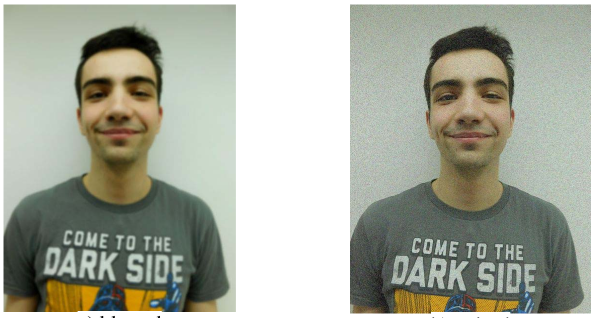
The system proposed in this paper was tested in an IT company where the authors' previous work [1] was integrated. Five employees volunteered in this research. The dataset included the photographs of them. Also, this dataset was only used for training the DNN. The employees took several different positions while being photographed. In order to make this approach applicable for production usage, it is of great importance to capture a small number of photographs of every employee at the site, Fig. 1.

a) front b) from above

c) 3/4 view right d) 3/4 view left

Fig. 1. An employee's original photographs taken in several positions.

Due to the fact that it is possible to achieve high accuracy by using the DNN on larger datasets, the augmentation process was applied on the original images. The authors of this paper proposed a novel approach of face augmentation for the purpose of extending the dataset which could lead to achieving higher accuracy on smaller datasets of the original images. The augmentation process was split into two stages. The first stage included common image augmentation techniques: noising and blurring the images on different levels, Fig. 2.



a) blurred b) noised

Fig. 2. Noised and blurred images.

The reason for using these techniques lies in the fact that the employees would be monitored at the entrance of the company by a standard IP camera. Poor network traffic or some other technical problems could potentially bring noise to the data. Using these augmented images in the dataset could adopt DNN for partially noised data. For this part of augmentation a Python script was written using OpenCV [11] interface to automatically generate new augmented images out of the original ones.

The second stage included a new approach of augmenting the images for deep learning face recognition tasks. This stage used the Dlib [12], machine learning toolkit for marking the location of a person's nose, eyes, chin and mouth on the image. Knowing the actual positions of these parts of a face on the image, the Python script automatically adds random accessories: mustaches, glasses, etc. and creates new images for training dataset, Fig. 3.

Fig. 3. Examples of generated images with some extra accessories used in training DNN.

By applying this method, the dataset is enlarged with newly generated images with a goal to reduce overfitting during the DNN's learning and to improve the accuracy.

B. Developing Face Recognition Model

This model includes several important steps: face detection, image preprocessing – finding face landmarks and face positioning, generating face embeddings and classification, Fig. 4.

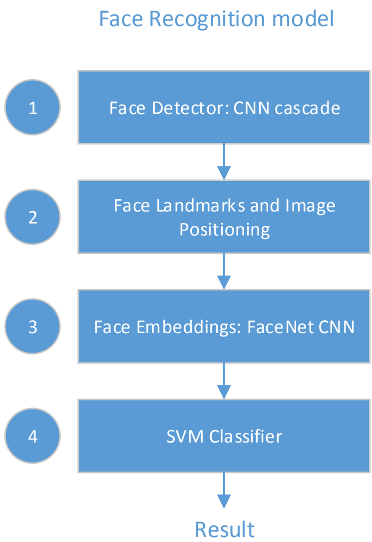


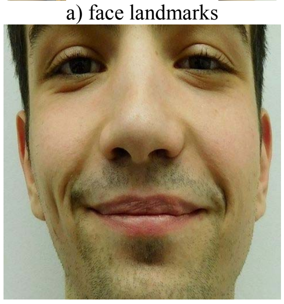
Fig. 4. Face Recognition Model.

The first step of the face recognition process is face detection. Face detection presents the well-studied field in the computer vision domain. As a result of decades of research, nowadays there are numerous machine learning algorithms applicable for this task. In recent years, CNNs achieved advanced results in image classification [14] and object detection [15].

Due to its runtime performance, for this step, a state-of-the-art CNN cascade is used for a face detection task, introduced by Haoxiang Li et al in [16]. The cascade consists of 6 CNNs, 3 CNNs for binary classification (face and non-face) and 3 CNNs for bounding box calibration. A Torch [17], machine learning framework is used for developing this face detector used as the first step of face recognition model.

Due to the problem of turning the face in different directions, which could seem different to the machine, the second step deals with the positioning of the face. A human face has 68 specific points – face landmarks. The primary goal of this step is to detect the face landmarks and to position the image by applying an affine transformation in order to centralize these landmarks as much as possible without distorting the image. A Python script was used to automatically detect the face landmarks based on the algorithm proposed in [18] and to position the face based on them, Fig. 5.

a) face landmarks



b) positioned face image based on face landmarks on the original image

Fig. 5. Face landmarks and positioning.

The third step presents the embedding process using the proposed system in [10] – FaceNet, as mentioned in the Section II. This method uses deep CNN for learning mapping from face images to Euclidean space where distances match to the face similarity measurements. This results in generating 128-bytes embeddings per face. Training of the network consists of triplets: the face image of a target person, the test face image of the target person and the face image of another person. OpenFace library [19] with pre-trained FaceNet network was used for training this deep CNN.

The final step of developing the face recognition model for tracking employees' attendance consists of training the classifier based on the previously generated embedding from employees' dataset by the deep CNN. Due to the fact that this system is based on smaller dataset, linear Support Vector Machine (SVM) was applied for this classification task.

C. Integration with the Existing System

For testing purposes, the developed face recognition model was integrated as an independent Face Recognition API in the existing RFID based employee attendance system which was the authors' previous work. The system consists of the RFID reader device, remote server along with the database and web application for administration and monitoring purposes. An IP camera was set at the entrance of the company where the reader device was placed. In order to validate the accuracy of the model, 5 employees who took part in this research continued to register with RFID card as usual. A face Recognition API was gathering video frames from the web camera, while cascade CNN face detector ran continuously as a background thread which was fed by video frames. If face was detected, then the image was preprocessed and passed to the deep CNN to generate 128-byte embedding. A SVM classifier determines the employee's identity and stores required data to the database: employee's identity, accuracy percentage, image, date and time. The primary reason of storing the image and accuracy is only for further research purposes and analyzing, while date and time is needed to compare results with the RFID reader device to validate face recognition model accuracy.

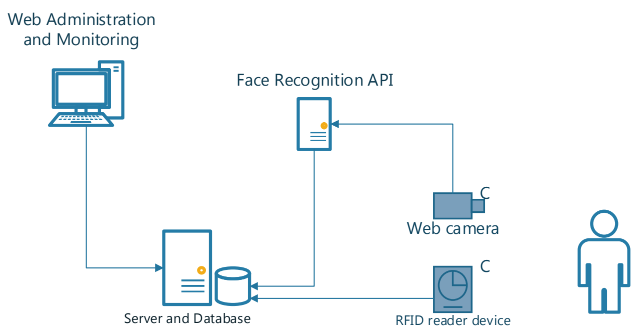


Fig. 6. Face Recognition API and current RFID-based system.

IV. RESULTS AND DISCUSSION

During the period of 3 months, the integrated face recognition system was actively recording every entrance and exit of the targeted employees. After this period, all the collected data was cross validated against the data gathered by the RFID card from the database. The prediction results are presented in Table I.

TABLE I. CONFUSION MATRIX

Classes					
Empl 1	***Empl 2***	***Empl 3***	***Empl 4***	***Empl 5***	***Predictions***
230	8	0	6	1	***Empl 1***
4	269	0	3	1	***Empl 2***
0	0	301	0	3	***Empl 3***
8	4	9	138	0	***Empl 4***
2	5	6	1	227	***Empl 5***

Accuracy per class is presented in Fig. 7. The overall accuracy of the system is 95.02%.

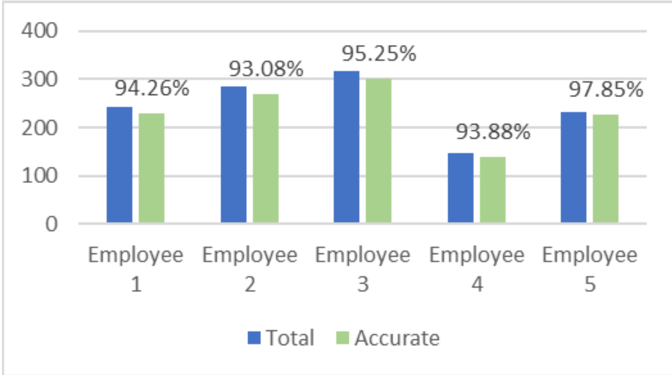


Fig. 7. Accuracy per class.

The model was trained based on a small number of images per employee and using the proposed method of augmentation. This led to the enlargement of the initial dataset and the improvement of the overall accuracy. By analyzing the images stored in the database during the acquisition period, it could be seen that the light conditions influenced the recognition process. Most of the images predicted incorrectly were exposed to the daylight while the door was open. This could potentially be corrected by applying gradient transformation on the images. A small number of images affected by noise of the unknown cause were predicted correctly. The overall accuracy could be improved by applying on time interval automatic re-training of the embedding deep CNN together with the newly gathered images predicted by the model with the high accuracy rate.

V. CONCLUSION

Nowadays, various attendance and monitoring tools are used in practice in industry. Regardless the fact that these solutions are mostly automatic, they are still prone to errors.

In this paper, a new deep learning based face recognition attendance system is proposed. The entire procedure of developing a face recognition component by combining state-of-the-art methods and advances in deep learning is described. It is determined that with the smaller number of face images along with the proposed method of augmentation high accuracy can be achieved, 95.02% in overall.

These results are enabling further research for the purpose of obtaining even higher accuracy on smaller datasets, which is crucial for making this solution production-ready. The future work could involve exploring new augmentation processes and exploiting newly gathered images in runtime for automatic retraining of the embedding CNN. One of the unexplored areas of this research is the analysis of additional solutions for classifying face embedding vectors. Developing a specialized classifying solution for this task could potentially lead to achieving higher accuracy on a smaller dataset. This deep learning based solution does not depend on GPU in runtime. Thus, it could be applicable in many other systems as a main or a side component that could run on a cheaper and low-capacity hardware, even as a general-purpose Internet of things (IoT) device.

REFERENCES

[1] Andric, Milan, et al. "Web application as a support system for records of working time, monitoring business processes and activities of company employees."

[2] Sharma, Saumya, S. L. Shimi, and S. Chatterji. "Radio frequency identification (RFID) based employee monitoring system (EMS)." International Journal of Current Engineering and Technology 4.5 (2014): 3441-3444.

[3] Sultana, Shermin, Asma Enayet, and Ishrat Jahan Mouri. "A SMART, LOCATION BASED TIME AND ATTENDANCE TRACKING SYSTEM USING ANDROID APPLICATION." International Journal of Computer Science, Engineering and Information Technology (IJCSEIT) 5.1 (2015).

[4] Rao, Seema, and K. J. Satoa. "An attendance monitoring system using biometrics authentication." International Journal of Advanced Research in Computer Science and Software Engineering 3.4 (2013).

[5] Soewito, Benfano, et al. "Smart mobile attendance system using voice recognition and fingerprint on smartphone." Intelligent Technology and Its Applications (ISITIA), 2016 International Seminar on. IEEE, 2016.

[6] Kadry, Seifedine, and Mohamad Smaili. "Wireless attendance management system based on iris recognition." Scientific Research and Essays 5.12 (2013): 1428-1435.

[7] Patil, Ajinkya, and Mrudang Shukla. "Implementation Of Classroom Attendance System Based On Face Recognition In Class." International Journal of Advances in Engineering & Technology 7.3 (2014): 974.

[8] Balcoh, Naveed Khan, et al. "Algorithm for efficient attendance management: Face recognition based approach." IJCSI International Journal of Computer Science Issues 9.4 (2012): 146-150.

[9] Tharanga, JG Roshan, et al. "SMART ATTENDANCE USING REAL TIME FACE RECOGNITION (SMART-FR)." Department of Electronic and Computer Engineering, Sri Lanka Institute of Information Technology (SLIIT), Malabe, Sri Lanka.

[10] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015

[11] Bradski, Gary, and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.", 2008.

[12] King, Davis E. "Dlib-ml: A machine learning toolkit." Journal of Machine Learning Research 10.Jul (2009): 1755-1758.

[13] Li, Haoxiang, et al. "A convolutional neural network cascade for face detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[14] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115.3 (2015): 211-252.

[15] Everingham, Mark, et al. "The pascal visual object classes (voc) challenge." International journal of computer vision 88.2 (2010): 303-338.

[16] Li, Haoxiang, et al. "A convolutional neural network cascade for face detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[17] Collobert, Ronan. "Torch." Workshop on Machine Learning Open Source Software, NIPS. Vol. 113. 2008.

[18] Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[19] Amos, Brandon, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. OpenFace: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.