Single Sample Face Recognition Using Convolutional Neural Networks for Automated Attendance Systems

Foteini P. Filippidou and George A. Papakostas
Human-Machines Interaction Laboratory (HUMAIN-Lab)
Department of Computer Science, International Hellenic University
Kavala, Greece
{fnfilip, gpapak}@cs.ihu.gr, gpapak@teiemt.gr

***Abstract*— Convolutional Neural Networks (CNNs) have been developed as powerful models for image recognition problems requiring large-scale labeled training data. However, estimating millions parameters of deep CNNs requires a huge amount of labeled samples, restricting CNNs being applied to problems with limited training data. To address this problem, a two-phase method combining data augmentation and CNN transfer learning i.e., fine-tuning pre-trained CNN models are studied herein. In this paper, we focus on the case of a single sample face recognition problem, intending to develop a real-time visual-based presence application. In this context, five well-known pre-trained CNNs were evaluated. The experimental results prove that DenseNet121 is the best model for dealing with practice problems (up to 99% top-1 accuracy) is the best and most robust model for dealing with the single sample per person problem, which are related to using deep CNNs on a small dataset and specifically to single sample per person face recognition task.**

Keywords— Single Sample per Person Face Recognition (SSPP FR), computer vision, convolutional neural network (CNN), student attendance systems

I. Introduction

Students' attendance is a vital issue in checking their regular performance at all educational levels. A student is only eligible to take part in exams if he/she has sufficient attendance in the course. In a traditional system, students have to sign an attendance sheet during the lesson or the teacher calls their names to check and record their attendance. The traditional presence methods are time-consuming, spend educational time, are not completely safe as a teacher may make a mistake or the attendance form may be lost or damaged or even one student could mark the attendance of another absentee student especially in a large classroom. Besides in some special circumstances, it may be necessary to implement online courses as has recently happened due to COVID-19 pandemic, so student's attendance verification based on face recognition could be a quick, easy and safe method.

Face recognition (FR), as one of the most commonly used biometric methods, must be robust and efficient in many variations such as changes in face pose and lighting conditions.

Conventional ways of defining an individual rely on a broad and representative collection of data that is not available in most real-life circumstances with a very limited or even a single sample per person (SSPP) available. According to available data, the FR methods are distinguished in MSPP (multi-sample per person) category when we have multiple samples per person and SSPP (single sample per person) when we have only one, usually frontal sample per person. SSPP face recognition is an extremely difficult task as the lack of sufficient sample information the large differences in lighting, expression, etc. impede the generalization.

The SSPP face recognition is of great research interest due to its application in many realistic scenarios such as in the case of airport passengers identification, identification at an entrance, classroom attendance, credit card verification etc [1].

Usually, an educational organization has only one frontal image per student, taken in a surveillance environment while the identification photos are taken in an environment without restrictions. Our contribution is to build an automated attendance system based on a SSPP face recognition method, that can be applied at a lecture, lesson, workshop or exams, where students are in person or distance. This work studies for the first time the ability of known CNNs to learn by a single sample per class for the needs of implementing an automated students attendance system.

The rest structure of the paper is as follows: Section II introduces the related work and the description of our system. Section III presents the settings of the conducted experimental study along with the corresponding simulation results. Finally, Section IV concludes the paper and indicates the future work.

II. Materials and Methods

A. Automated Attendance Systems

As Smartphones are very popular among teachers and students, they were used in a classroom monitoring system [2]. In another system [3], while the teacher enters the application and selects the current lesson the whole list of students will appear on his monitor and confirms the presence by calling. As soon as the instructor posts the presentations, all absences appear on the screen and once confirmed, the instructor submits this file. By completing the database entry, students can check their participation by entering the application.

Some systems based on special equipment in combination with other attendance methods. Radio Frequency Identification (RFID) is a technology that uses radio waves to transfer data through a reader that is attached to an object-electronic tag to detect and monitor the object. Some RFID tags can be read several meters away. Under the proposed system [4] all classes must have a computer with a connected RFID reader that can read the student's RFID card and also a webcam to take pictures of them. Computers are connected to a server and they send data to be stored in a database. The camera is intended to prevent a student from scanning another card. The RFID reader reads the student's card, while the Web

This work was supported by the MPhil program "Advanced Technologies in Informatics and Computers", hosted by the Department of Computer Science, International Hellenic University, Greece.

978-1-7281-8084-7/20/$31.00 ©2020 IEEE

camera takes his / her photo at the same time and is sent to the server for verification and identification. According to [5], routers are installed throughout the university combined with specific rooms. At constant intervals, the client application communicates with the server and sends the information of the nearest access point (router) thereby eliminating the presence of a student in the particular room that is closest to the router. If a student is present throughout the lesson, the name of the class will be permanently recorded in his/her file and this is the proof of his attendance. A simple software with a graphical interface based on smartphone technology too, was developed in [6]. There must be a WiFi router in each room where students' smartphones are connected and their unique MAC addresses are automatically recognized when entering the room. After connecting the device to the LAN, the fingerprint of the user is requested. Using multi-sample image processing techniques in a few minutes the system identifies the number of students present in the classroom and if it does not match the number of students monitored by the biometric method, the teacher verifies the results. An FR based system [7] developed to verify the identified students by RFID. The system can identify any student whose photo with its name and identification number has been recorded.

A lot of systems based on FR attempt to provide an automated way to monitor and record students' attendance. Principle Component Analysis (PCA) [8], [9], [10], [11], Linear Discriminate Analysis (LDA) [8] and LBP [9] based on feature extraction, were used to identify students. A facial recognition system [12], based on mathematical methods, seeks to extract relevant information from an image, to encode and compare them with another image stored in a database. Eigen Face Detection, Fisherfaces and Haar cascade methods used in [13] to record students' attendance. A Field Programmable Gate Array (FPGA) system [14] is an FR based application used for various aspects in a classroom. A camera was used to monitor students' attendance and engagement. A carbon dioxide sensor records the CO2 concentration associated with decision making and the general concentration capacity. A microphone measures the level of noise generated by the school environment so that the system could detect harmful sounds inside or outside the classroom and report them. Finally, a brightness sensor is used to ensure the proper lighting. A camera fixed at a specific distance within a classroom capture video of the students' frontal images and converted into frames. The CNN model detects faces and compares them with the persons stored in the database [15].

Some developed systems include a software application where the teacher enters and simply records presences without using a paper. Smartphone technology used to locate the student by connecting and recognizing his/her mobile phone. Finally, some systems used face recognition in combination with some other identification methods and equipment. In these cases, the school should have a lot of different student's images. Knowing that the school usually has only one frontal photo of each student, we decided to create a system that applies the SSPP FR approach. Specifically, the use of convolutional neural networks in conjunction with transfer learning and image augmentation was chosen.

B. *Single Sample Face Recognition*

Various SSPP face recognition methods have been proposed in the past that can generally be divided into five categories [1]:

I. *Generic databases based:* a separate database is used to learn the parameters which cannot be otherwise learnt from the single available training face image. These parameters are further used to extract image features.

II. *Virtual sample generation:* pseudo-face images are built from a single face image using a variety of methods.

III. *Feature-based:* discriminative features are extracted from all or some local patches of the single image.

IV. *Hybrid:* correlates some of the above methods

V. *Others*: include methods that do not belong to one of the above categories

In [16], [17] a generic dataset used to prepare certain feature extractors. In [16] the authors combine the traditional eigenface with the artificial neural network (ANN) method to learn the discriminant vectors from generic sets. The SSPP FR framework proposed in [17], combine transfer learning and sample expansion in feature space using generic dataset. By training a deep convolutional neural network (DCNN) on a multi-sample dataset we get a well-trained model. The generic and gallery dataset used as input to the pre-trained model. To expand the single sample face features by the intra-class variance of the generic dataset we use the k-class feature transfer (KCFT) algorithm. Then the expanded features used to train the last layer of softmax classifier and finally, the trained model used for testing the target set.

Since different variations can not be properly represented by a single image, the widening of the available data set is very often applied, creating virtual images from the frontal-view face. A bi-directional neural network, inspired by the neocortex functional model, is proposed in [18]. After the person and pose information is separated, the virtual samples are synthesized and a neural network is trained. In another work [19] a deep neural network designed as a nonlinear image information processing model for separating pose and person information and use it to generate virtual images. These virtual images are used to train a neural network classifier. In [20], an easy sample expanding method has been proposed which creates variations in expression, disguise and mixed, combining traditional and deep learning methods. For the expansion of the training dataset, the best constructed intraclass variation set is selected. Applying transfer learning a well-trained DCNN is fine-tuned with the expanding samples to be used for the experiments. Considering that facial variations can be shared among individuals, a new method of sample expansion is proposed in [21]. Then a well-trained DCNN model is fine-tuned by these expanding samples. Instead of using a generic dataset, in [22] simple transformations to the unique sample creating a large number of face augmentation was applied. Highly discriminative deep features extracted by applying the VGG-face net and subsequently the linear discriminant analysis (LDA).

To extract robust features, stable to the intraclass variances so that the same person should have the same (ideally) or similar features is another method used in FR. According to the proposed method in [23], a deep neural network is trained using the image gallery, the probe images and their labels, for feature extraction. The activation function of the encoders' previous layer, is applied to both clean and corrupted data, and the outputs serve as the input data to train the supervised auto-encoder in the next layer. After network's training the output of the highest level is used as the features to represent the image. Assuming that the occlusion variations of one subject can be approximated by a sparse linear combination of other subjects the authors in [24] proposed an extension of the Fisher Discrimination Dictionary Learning (FDDL) model, based on deep features.

C. Convolutional Neural Networks

The architecture of a typical CNN contains different processing levels [25].

I. <u>Input Layer</u>: accepts raw images, and promotes them to further levels to extract features.

II. <u>Convolution Layer</u>: At this level, several filters are applied to images to extract features from them. These features are used to calculate the matches in the testing phase

III. <u>ReLU (Rectified-Linear Unit)</u>: replaces the negative numbers of the Convolution Layer outputs with zero, which helps in faster and more efficient training.

IV. <u>Pooling Concentration</u>: The exported characteristics are sent to the Pooling level. This level takes large images and reduces them to keep important information.

V. <u>Fully Connected Layer</u>: The Final Layer which receives high-level filtering images and translates them into tagged categories.

VI. <u>Softmax Layer</u>: This level calculates the decimal probabilities in each class. These decimal places are in between 0 and 1.

The first four stages are called feature extraction stages and the last two are the classification stages

In this paper, we evaluate, MobileNetV2 [26], ResNet50V2 [27], DenseNet121[28], InceptionV3 [29], and VGG16 [30] CNNs. The aim of our work is to study and evaluate established CNN based object classifiers of general interest without considering specialized for face recognition (e.g Face Net [31], Gaussian Face [32] and Deep Face [33]) models.

MobileNetV2 is a new architecture of neural networks that are specifically adapted for mobile and resource-constrained environments, significantly reducing the number of tasks and the memory requirement while maintaining the same accuracy. Its basic building block is a bottleneck depth-separable convolution with inverted residuals. [26].

ResNet50V2 levels in neural networks are not limited to sequential layout only but form a graph and RESNET is an architecture that has the advantage of this flexibility. A Residual block consists of two or three consecutive convolutional layers and a separate identity (repeater) connector that connects the input of the initial level to the output of the final level. Each block has two paths. One is known also to other neurons, and the second is the skip connection or as it is called *identity shortcut connection*. The network can access several blocks thanks to the skip access path. Resnet is popular not only for its high precision accuracy but also for its simplicity and versatility [27].

To ensure the maximum flow of information, all levels are connected directly to each other, so each level receives additional inputs from all previous levels and passes its feature maps to all subsequent layers. Because of this dense connection, these networks are called Dense Convolutional Network (DenseNet). In addition to better performance, a great advantage of DenseNets is an improved flow of information and gradients across the network, that makes easy their train. It is also be observed that the dense connections have a regularizing effect, which reduces overfitting when the training set is small. DenseNets can be good feature extractors for various computer vision tasks based on convolutional features [28].

The InceptionV3 is the third iteration of the inception architecture, first developed for the GoogLeNet model. Note that the traditional 7×7 convolution has been factorized into three 3×3 convolutions [29]. The key aspect of the Inception architecture is to figure out how an ideal local sparse structure in a convolutional vision network can be approached and covered by easily available dense components. Assuming translation invariance means that the network is constructed by convolutional blocks. One of the main benefits of its architecture is that we can significantly increase the number of units at each stage without causing an uncontrollable increase in computational complexity [34].

The VGG16 model was developed by the Visual Graphics Group (VGG) at Oxford University. Examining the depth, an important aspect of ConvNet architecture design, other parameters of the architecture were corrected, and the network depth was steadily increased by adding more convolutional layers, which is possible due to the use of very small assembly filters (3×3) in all. As a result, much more precise ConvNet architectures have been achieved, which provide excellent performance on various image recognition data sets. The image in this CNN goes through a stack of convolutional layers, where filters with a very small reception field: 3×3 (which is the smallest size to record the notion left/right, up/down, centre) [30] are used.

D. Proposed Automated Attendance System

Our proposed system is an automatic presentation system in a classroom or workshop that uses face recognition to control the student's attendance. The main core of our system is the model responsible to recognize the students by using facial information. For this purpose, a thorough study of the behaviour of well-known pre-trained convolutional neural networks for face recognition having only a single frontal photo per person is needed to decide the most suitable for our system. Moreover, the educational institutions have only one frontal photo for each student and thus a data augmentation process is applied for increasing the data used to train the deep learning models. The student's attendance is taken by capturing images of the entire classroom using a webcam. When the image is captured, it is sent to the processing unit for the classification process where firstly the faces are detected by MTCNN face detector algorithm. The detected faces are cropped and then the recognition process is carried out by a pre-trained CNN which counters the percentage the face to belong to each class and the maximum of the above percentages is valued. If the maximum percentage is above a threshold, multiple times in different photos to reduce the chance of misidentification, the attendance is marked for that particular student. During the lesson, the teacher can check or unchecked the presence of a student. When the recognition process is completed, the attendance report is uploaded to the web-page that is designed to serve this purpose and stored in the database. The web-page can be accessed by the authenticated persons only.

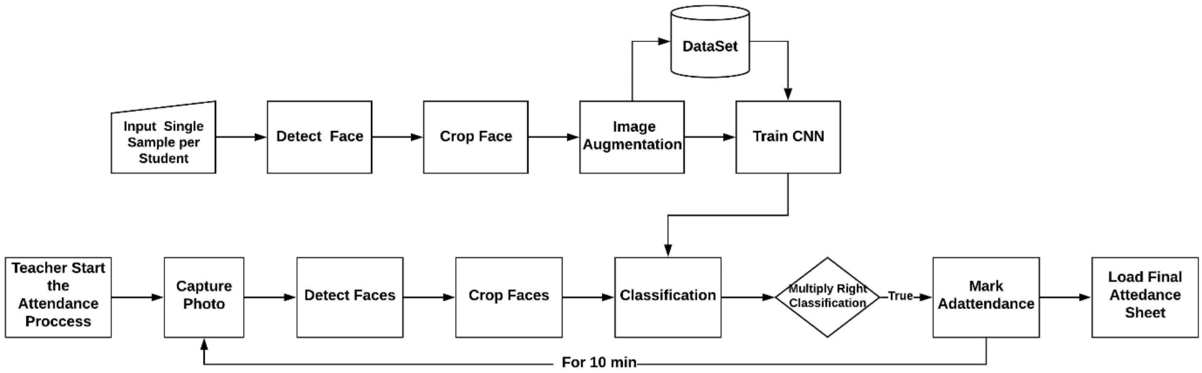

Fig. 1. The proposed SSPP FR based attendance system

Schematic presentation of the proposed methodology is given in Figure 1.

III. Experimental Study

A. Study Settings

It is known that the educational institutions collect only one frontal photo during student's registration, which means one sample per student in the source domain (i.e. SSPP problem) and on the other hand a CNN needs a huge amount of sample for training. For the experiments, we collect one frontal face image per person, with random size and resolution. The collected persons where both men and women and at variable ages. Firstly we detect and crop the faces to isolate them from the background. Then in the cropped photos, we apply data augmentation techniques to create a dataset with several photos to fine-tune five pre-trained CNN models. Specifically we evaluate DenseNet121, MobileNetV2, InceptionV3, ResNet50V2 and VGG16 models. The augmented data generated by applying different illumination conditions, Horizontal and Vertical Shift while keeping the image dimensions the same, Random Rotation clockwise by a given number of degrees, Random Brightness by either randomly darkening or brightening images, Random Zoom Augmentation, ZCA whitening and Shearing is also used to transform the orientation of the image. In each experiment from the total number of images, we use 10% of them for testing, and the remaining 90% for training and validation by applying a validation_split=0.3.

In deep learning, a model is trained with a large amount of data and learns the weights and biases during training. A pre-trained model has already been trained using a big dataset and its weights are transferred to other network models for testing [25]. We fine-tune five models pre-trained with Imagenet [35] a database of 1,461,406 images and 1000 classes. First, we removed the "head" of the pre-trained CNNs, that includes the final set of fully connected levels where the class label predictions are returned. We then replace these fully connected layers with a new set. From there, we freeze all layers below the head except batch layers so their weights cannot be updated (ensuring that any previous robust features learned by the CNNs are not destroyed) and then compile the new network. Fine-tuning allows us to use pre-trained networks to identify classes that were not initially trained on, achieving great accuracy. As we know, the final set of layers (i.e., the "head") includes fully connected layers with our softmax classifier. Then we train and validate each CNN model, save the weights and perform face recognition tests by using the stored weights.

The source code is written in Python using OpenCV (Open Source Computer Vision) and Keras library, running on x-64, Intel ® Core™ i5-85250U processor with NVIDIA GeForce card. We use the MTCNN package for face detection, while the Adam optimizer with a learning rate of 0.045 is applied to train the models for 10 epochs.

B. Results

To evaluate the performance of the CNN models concerning the number of augmented images, we measured the Top-1 accuracy during each experiment. The models are evaluated for 25 students per classroom, which is a common number of students in a classroom. We conduct five experiments, using totally for each person, 200, 400, 600, 800 and 1000 augmented images respectively. From these images, 10% were used for testing. The results are shown in Table I.

TABLE I. Top-1 Accuracy (%) Per Number Of Images

Number of images	*DenseNet121*	*MobileNetV2*	*InceptionV3*	*ResNet50V2*	*VGG16*
200	99.2	93.4	74.2	95.8	96.0
400	99.8	93.4	75.8	97.0	95.9
600	100.0	93.9	77.8	97.8	96.9
800	100.0	94.6	77.9	98.4	97.5
1000	99.8	95.1	76.0	98.4	96.9
Average	99.8	94.1	76.3	97.5	96.6

The results of Table I, are also illustrated in Figure 1, where we can see the change of Top-1 concerning the available images for training, validation and test. In Table I, we see the total number of augmented samples.

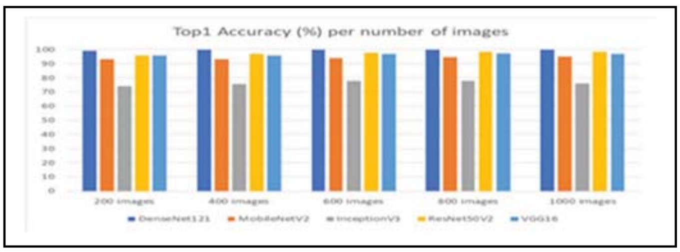

Fig. 2. Top-1 Accuracy concerning the number of image samples

According to the values in Table I and Figure 1, we observe a small improvement in the recognition accuracy of all models with the increase of the total number of samples for each of the 25 students under study. Besides, according to Table I, it is revealed that DenseNet121 model presents the highest average accuracy (99.76%) among the five models while InceptionV3 has the lowest (76.34%). Moreover, the former model shows lower dependency on the training dataset, compared to the other models which seem to be more influenced by the data size.

Keeping constant the number of images, we successively increase the number of recognized classes. Specifically, we used a total of 1000 augmented images for each student and measured the Top-1 accuracy of each CNN model, using 25, 50, 75 and 100 students per classroom. This experiment aims to study how models behave as the number of classes increases. The measurements are shown in Table II

TABLE II. TOP1 ACCURACY (%) PER NUMBER OF CLASSES

Number of classes	***DenseNet121***	***MobileNetV2***	***InceptionV3***	***ResNet50V2***	***VGG16***
25	99.8	95.1	76.0	98.4	96.9
50	99.9	86.2	56.0	94.7	86.8
75	99.6	87.1	42.4	93.3	81.7
100	99.3	85.1	33.5	90.9	73.5
Average	99.6	88.4	41.6	94.3	86.5

With the values of Table II, we draw the graph in Figure 2, where we can see the change of Top-1 accuracy with the number of classes, having a total of 1000 images for each student.

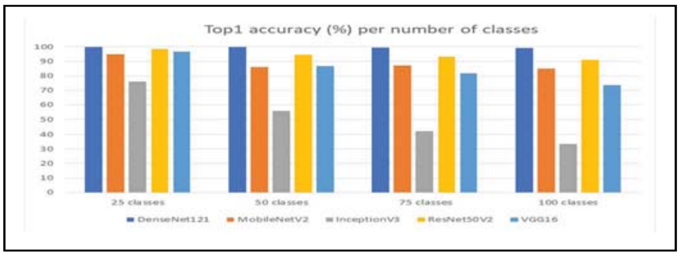

Fig. 3. Top1 Accuracy concerning the number of classes

According to Table II and Graph in Figure 2, we observe that maintaining the same number of samples, as the classes increase, the accuracy of classification decreases. However, the DenseNet121 model shows a more robust behaviour since its accuracy does not affect significantly with the increasing number of classes studying the average accuracy of each model, we conclude that DenseNet121 has the highest accuracy (99.65%) while InceptionV3 presents the lowest (41.58%).

Table III summarizes the average performance for each model in the entire experimental study. The values in Table III show that in both experiments DenseNet121 presents the best performance, while InceptionV3 the worse. We also notice that DenseNet121shows a small decrease in its performance in the second experiment while InceptionV3 shows a very big drop.

TABLE III. AVERAGE TOP-1 ACCURACY (%) FOR BOTH EXPERIMENTS

Number of classes	***DenseNet121***	***MobileNetV2***	***InceptionV3***	***ResNet50V2***	***VGG16***
Average (Table I)	99.8	94.1	76.3	97.5	96.6
Average (Table II)	99.6	88.4	41.6	94.3	86.5

To confirm the correctness of the results of the Top-1 accuracy measurements, the AUC (Area Under Curve) measurement for each experiment is also calculated. More precisely, in TABLE IV the AUC for each model is presented with gradually increasing the number of available photos. Furthermore, TABLE V depicts the AUC measurements in conjunctin with the number of classes for each examined model.

TABLE IV. AUC PER NUMBER OF IMAGES

Number of images	***DenseNet121***	***MobileNetV2***	***InceptionV3***	***ResNet50V2***	***VGG16***
200	0.9999	0.9983	0.9783	0.9987	0.9994
400	0.9999	0.9967	0.9778	0.9994	0.9995
600	1.0000	0.9979	0.9821	0.9999	0.9997
800	1.0000	0.9988	0.9788	0.9999	0.9995
1000	1.0000	0.9990	0.9776	0.9993	0.9997
Average	0.9999	0.9981	0.9789	0.9994	0.9996

TABLE V. AUC PER NUMBER OF CLASSES

Number of classes	***DenseNet121***	***MobileNetV2***	***InceptionV3***	***ResNet50V2***	***VGG16***
25	0.9999	9990	0.9776	0.9993	0.9997
50	0.9999	9950	0.9496	0.9988	0.9956
75	0.9999	9973	0.9292	0.9987	0.9935
100	0.9999	9959	0.9109	0.9983	0.9880
Average	0.9999	9968	0.9418	0.9980	0.9942

Examining the measurements in the Table IV and Table V we find out that the DenseNet model presents the best and more robust performance in both experiments. We also note that, as the number of classes increases, the performance of the models decreases, except for the DenseNet, which performance maintains the highest. Finally, we ascertain that Inception model presents the worst performance as we found out with the Top-1 measurements. In general, the AUC measurement confirms the conclusions we reached with the Top-1 measurements.

IV. CONCLUSION AND FUTURE WORK

According to the previous study, it is concluded that the known Convolutional Neural Networks are able to be applied for a single sample per student face recognition based attendance systems. This high performance was achieved by augmenting the frontal images towards increasing the training data. Furthermore, the DenseNet121 presents the best and most robust performance among the studied CNNs. Our future research plans include the deployment of the DesnseNet121 model on the attendance system and evaluate it in real classroom conditions.

REFERENCES

[1] "Single Sample Face Recognition in the Last Decade: A Survey | wizdom.ai - intelligence for everyone.," *wizdom.ai*. https://www.wizdom.ai/publication/10.1142/S0218001419560093/title/single_sample_face_recognition_in_the_last_decade_a_survey (accessed Apr. 03, 2020).

[2] M. M. Islam, M. K. Hasan, M. M. Billah, and M. M. Uddin, "Development of smartphone-based student attendance system," 2018, vol. 2018-January, pp. 230–233, doi: 10.1109/R10-HTC.2017.8288945.

[3] A. Vamshi Krishna and K. Ganapathy, "Mobile application for uploading marks and student attendance management system," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6 Special Issue, pp. 379–383, 2019, doi: 10.35940/ijeat.F1080.0886S19.

[4] N. Saparkhojayev and S. Guvercin, "Attendance Control System based on RFID-technology," *International Journal of Computer Science Issues*, vol. 9, May 2012.

[5] U. Gupta, R. Khurana, and P. Singh, "Making organizations pervasive using existing infrastructure: Case study of a university," 2014, vol. 9-11-October-2014, doi: 10.1145/2662117.2662119.

[6] D. Vishal, H. Saliq Afaque, D. Saketh Vishnu, and T. K. Ramesh, "A mobile integrated classroom," 2018, pp. 125–130, doi: 10.1109/MITE.2017.00028.

[7] M. S. Akbar, P. Sarker, A. T. Mansoor, A. M. Al Ashray, and J. Uddin, "Face Recognition and RFID Verified Attendance System," 2019, pp. 168–172, doi: 10.1109/iCCECOME.2018.8658705.

[8] A. Jha and M. Hooda, "Class Room Attendance System Using Facial Recognition System," Accessed: Mar. 22, 2020. [Online]. Available: https://www.academia.edu/28754712/Class_Room_Attendance_System_Using_Facial_Recognition_System.

[9] N. D. Veer and B. F. Momin, "An automated attendance system using video surveillance camera," 2017, pp. 1731–1735, doi: 10.1109/RTEICT.2016.7808130.

[10] S. Poornima, N. Sripriya, B. Vijayalakshmi, and P. Vishnupriya, "Attendance monitoring system using facial recognition with audio output and gender classification," presented at the International Conference on Computer, Communication, and Signal Processing: Special Focus on IoT, ICCCSP 2017, 2017, doi: 10.1109/ICCCSP.2017.7944103.

[11] S. Rinesh, B. Hemanth Kumar Reddy, Y. Jagadesh, V. Parthipan, and S. Magesh Kumar, "Face recognition system for attendance marking," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 5822–5826, 2019, doi: 10.35940/ijrte.C5275.098319.

[12] J. R. De Toledo Quadros, F. Paschoal, and F. P. Laercio Brito Goncalves, "Facial recognition system for automatic presence control in a classroom," presented at the Iberian Conference on Information Systems and Technologies, CISTI, 2017, doi: 10.23919/CISTI.2017.7975712.

[13] C. B. Yuvaraj, M. Srikanth, V. Santhosh Kumar, Y. V. Srinivasa Murthy, and S. G. Koolagudi, "An approach to maintain attendance using image processing techniques," 2018, vol. 2018-January, pp. 1–3, doi: 10.1109/IC3.2017.8284353.

[14] M. Hanna *et al.*, "Smart fpga-based system for enhancing educational programs," 2019, pp. 122–125, doi: 10.1109/NILES.2019.8909344.

[15] "C11230283S19.pdf." Accessed: Mar. 20, 2020. [Online]. Available: https://www.ijeat.org/wp-content/uploads/papers/v8i3S/C11230283S19.pdf.

[16] F. Hafiz, A. A. Shafie, and Y. M. Mustafah, "Face Recognition From Single Sample Per Person by Learning of Generic Discriminant Vectors," *Procedia Engineering*, vol. 41, pp. 465–472, Jan. 2012, doi: 10.1016/j.proeng.2012.07.199.

[17] R. Min, S. Xu, and Z. Cui, "Single-sample face recognition based on feature expansion," *IEEE Access*, vol. 7, pp. 45219–45229, 2019, doi: 10.1109/ACCESS.2019.2909039.

[18] F. Abdolali and S. A. Seyyedsalehi, "Improving face recognition from a single image per person via virtual images produced by a bidirectional network," 2012, vol. 32, pp. 108–116, doi: 10.1016/j.sbspro.2012.01.019.

[19] T. Zhuo, "Face recognition from a single image per person using deep architecture neural networks," *Cluster Computing*, vol. 19, no. 1, pp. 73–77, 2016, doi: 10.1007/s10586-015-0513-1.

[20] J. Zeng, X. Zhao, J. Gan, C. Mai, Y. Zhai, and F. Wang, "Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition," *Computational Intelligence and Neuroscience*, vol. 2018, 2018, doi: 10.1155/2018/3803627.

[21] J. Zeng, X. Zhao, C. Qin, and Z. Lin, "Single sample per person face recognition based on deep convolutional neural network," 2018, vol. 2018-January, pp. 1647–1651, doi: 10.1109/CompComm.2017.8322819.

[22] V. Cuculo, A. D'amelio, G. Grossi, R. Lanzarotti, and J. Lin, "Robust single-sample face recognition by sparsity-driven sub-dictionary learning using deep features," *Sensors (Switzerland)*, vol. 19, no. 1, 2019, doi: 10.3390/s19010146.

[23] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single Sample Face Recognition via Learning Deep Supervised Autoencoders," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2108–2118, Oct. 2015, doi: 10.1109/TIFS.2015.2446438.

[24] H. Ouanan, M. Ouanan, and B. Aksasse, "Non-linear dictionary representation of deep features for face recognition from a single sample per person," *Procedia Computer Science*, vol. 127, pp. 114–122, 2018, doi: 10.1016/j.procs.2018.01.105.

[25] S. T. Krishna and H. K. Kalluri, "Deep Learning and Transfer Learning Approaches for Image Classification," vol. 7, no. 5, p. 7, 2019.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *arXiv:1801.04381 [cs]*, Mar. 2019, Accessed: May 27, 2020. [Online]. Available: http://arxiv.org/abs/1801.04381.

[27] I. Vasilev, *Advanced Deep Learning with Python: Design and implement advanced next-generation AI solutions using TensorFlow and PyTorch*. Packt Publishing Ltd, 2019.

[28] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *arXiv:1608.06993 [cs]*, Jan. 2018, Accessed: May 30, 2020. [Online]. Available: http://arxiv.org/abs/1608.06993.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016, pp. 2818–2826, Accessed: Jun. 09, 2020. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html.

[30] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Apr. 2015, Accessed: May 30, 2020. [Online]. Available: http://arxiv.org/abs/1409.1556.

[31] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, Jun. 2015, doi: 10.1109/CVPR.2015.7298682.

[32] C. Lu and X. Tang, "Surpassing Human-Level Face Verification Performance on LFW with GaussianFace," *arXiv:1404.3840 [cs, stat]*, Dec. 2014, Accessed: Jul. 15, 2020. [Online]. Available: http://arxiv.org/abs/1404.3840.

[33] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 1701–1708, doi: 10.1109/CVPR.2014.220.

[34] C. Szegedy *et al.*, "Going Deeper with Convolutions," *arXiv:1409.4842 [cs]*, Sep. 2014, Accessed: Jun. 11, 2020. [Online]. Available: http://arxiv.org/abs/1409.4842.

[35] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.