ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2021.0367

COMBINING DEEP RESIDUAL NEURAL NETWORK FEATURES WITH SUPERVISED MACHINE LEARNING ALGORITHMS FOR REAL-TIME FACE RECOGNITION-BASED INTELLIGENT SYSTEMS

M. Lakshaga Jyothi and R.S. Shanmugasundaram
Department of Computer Science and Engineering, Vinayaka Mission's Kirupananda Variyar Engineering College, India

Abstract

Face Recognition is the domain of technology in Computer Vision that deals with the process of identifying faces of known and unknown persons-based on facial patterns. Despite all the recent researches these years on Face Recognition technology, the development of real-time face recognition has always been a challenging task. This kind of technology has its applications widespread like security, medical diagnosis, educational sectors, etc. The advancements in High-end processors and High-Definition cameras led to the design of Face recognition systems that use offline or real-time input datasets. In this paper, our main aim is to focus on real-time video feeds taken from a framed classroom environment to identify the students or the faculty members and tag their names, com- paring them with the already stored face databases. Attendance marking is a daily routine that follows calling the name or passing the attendance books, which is very time-consuming, and they tend to begin proxy at-tendances. This study proposes a attendance marking system using face recognition and Deep Learning techniques on a Raspberry Pi board. The proposed system delivers an approach to make real-time face recognition-based attendance systems by extracting deep facial features using deep residual network (ResNet)-based CNNs. Then that deep facial feature dataset is combined with Machine Learning Algorithm such as SVM to perform face detection and recognize the faces. The maximum recognition accuracy of 80% is obtained using the planned system on the real- time face images provided and will be further improved.

Keywords:

Face Recognition, Deep Learning, Deep Convolutional Neural Network, Face Tagging, Classroom Attendance, Support Vector Machines

1. INTRODUCTION

Over the recent advancements, evolution and the role of Artificial Intelligence and Machine Learning has taken its path in combining the Internet of Things and Artificial Intelligence leading to the rise of Embedded Intelligence. Face recognition system implementations with the help of deep learning techniques provides remarkable performances both in time and accuracy [1].

Although there are various achievements in Face Recognition technology, implementing such systems in memory-constrained devices and its end-to-end application development is still a challenging task. In Face Recognition technology, the initial step is detection of faces and then face recognition, where the vital stage is to detect human faces. Face recognition and detection are the necessary components for an attendance marking system. Attendance management plays a trivial role in the academic performances of staff members and the students in the colleges and the universities. Acceptance of biometric systems corresponds to an extent how the people accept these biometric systems [2].

Frequent absences from the classes can increase the risk of dropping out of the classrooms. Checking the attendance helps improve the students' attendance and the staff members in the classroom environments.

To overcome the difficulties faced through traditional attendance management systems, various distinguished technologies are explored. First, in fingerprint-based technology, to collect the fingerprint databases to monitor their attendances. Second, RFID cards for attendance management system to record classroom attendance is quite reliable. Iris-based attendance management where the camera scans and collects the iris dataset to form an iris database against which the attendance is marked. Later on, Face-based attendances systems have started to get the attention of most of the researchers [3].

Despite its recent contributions by the researchers in developing attendance management systems-based on face recognition using Machine Learning and Deep Learning approaches, it is pretty chal lenging to create an end-to-end face-based attendance system deployment is still a difficult task. Therefore, more attempts are made to develop hardware systems for accelerating the performances [4].

Since, Deep Learning algorithms are a representative structure of hundreds of parameters, enabling Deep Learning for IoT devices is still a challenging task. After most exploration [5], few techniques were discussed to enable Deep Learning for IoT-based system devices with ARM cores that tends to handle low-end video processing tasks in computer vision applications to support by edge computational devices like Raspberry Pi, Jetson Nano boards, etc. to run Deep Learning techniques in IoT devices via Direct Model processing on Mobile/Edge devices, Acceleration of running Deep Learning paradigms in IoT devices. In this research, the former approach is taken into consideration for the experimentation.

This paper tries to implement the face recognition-based system for marking attendances-based on pre-trained neural network models.

This paper is structured as follows. Section 1 and 2. Discusses the background knowledge and the recent research on the face-based attendance systems implementation. Section 3 describes the proposed methodology and its technical aspects have been described in section 4. The experimental implementation, data collection, experimental setup, overview of the proposed system, and outcomes are then summarized in Section 4, Section 5 tends to give the limitations of the proposed system, finally, Section 6 conclude the work along with future plans.

2. RELATED WORK

Face recognition systems to process videos in real-time is still a challenging task. To solve computer vision-based tasks, remarkable success is achieved via deep learning in facial feature extraction by overcoming the impact of various peripheral factors and the trustworthiness of face recognition technology. In recent years, many researchers have discussed multiple approaches towards face-based attendance marking systems in the classroom environment resulting in intelligent system implementations using various deep learning techniques to promote intelligent classroom experiences.

As proposed in [6], various android-based Deep learning models were explored to take classroom attendances. Similarly, using SOTA techniques, an attendance management system was developed [7] to attain an accuracy of 98.67% through Labelled Faces in the Wild (LFW) dataset. Furthermore, to deal with storage efficiencies, a neural network-based attendance system via a web-based portal along with XAMPP web servers was explored. Finally, student attendances were marked using FaceNet and MTCNN. In addition, CNN models were examined [8] to recognize students in the frontal, side, and down-faced face angles reaching 81.25%,73%, and 43%, accuracies respectively.

More variants of neural network models have evolved, such as Convolutional Neural Network (CNN) model, Recurrent Neural Network (RNN) model, etc. The deep CNNs can be best to find the image patterns [9], Siamese Network. Smart security and smart educational sectors can take control of various administrative activities and easily automate multiple tasks such as Attendance marking systems in classrooms and exam halls [10].

In recent advancements, deep learning-based face recognition systems are equipped with novel methods to train facial features and extract patterns. Being the sub field of machine learning, deep learning enables the automatic task of accomplishing responsibilities that were expensive for humans in the past. The robust nature of any facial recognition system is-based on the factors considered while proposing the algorithm and the datasets, such as illumination expression, gender occlusion, etc. A model named Deep CNNs has been developed to achieve remarkable accuracies on very complex visual pattern recognition tasks to match or exceed the performances of the human perception level. Pioneer researchers have proposed various novel methods in solving computer vision tasks namely ImageNet, QuocNe, FaceNet [11], Inceptionv3 (GoogleNet), ResNet [12] etc.

3. METHODOLOGY AND EXPERIMENTS

In our research work, this architecture does not demand handcrafted features. Face detection is done through CNN techniques to extract facial features by itself automatically. Face-based attendance marking system has been explored in this research. This proposed system comprises mainly two sub-modules. i. Face detection and ii. Face Recognition. To detect the faces in given video or the image or the real-time input, in this research work has deployed SSD (Single Shot Multibox Detector) [13] framework, which is developed on the base on ResNet-10 neural network model. Inorder to work on real-time computer vision techniques related to deep learning system implementation, a popular library named OpenCV (Open-Source Computer Vision) [14] along with its deep learning module called deep neural network (dnn) module is focused on the facial recognition pipeline. This OpenCV module supports Deep Learning frameworks such as Tensorflow, Caffe, PyTorch, etc. Using the pre-trained neural network models released along with OpenCV can achieve face detection with remarkable accuracy. OpenCV face detector is-based on the proposed architecture as discussed below.

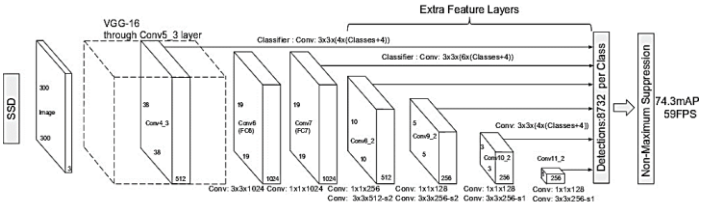

Fig.1. ResNet Architecture

3.1 SSD NETWORK ARCHITECTURE

SSD's Single Shot MultiBox detector uses the technique of extracting feature maps through the convolution layers and then applying those convolutional filters for face detection. They usually use VGG16 as a base network to remove the features. Then it tries to detect the faces using Conv4 3 layer. Finally, it makes predictions on each location via the grid detector technique. SSD models widely use grid techniques to detect multiple objects in different locations. The face image is split into numerous grid cells (m*n), each cell associated with a simple object detection model. Each detector is specialized in detecting and classifying the objects falling in their cell. SSD then uses feature maps from different layers of the base network, i.e., VGG-16. The primary focus of SSD is to use the base network, i.e., VGG-16, to take the input image and detect faces of different scales, locations, and different shapes. This study uses SSD 300*300, which takes input for the network architecture with 300*300 image pixel size.

3.2 RESNET NETWORK ARCHITECTURE (RESIDUAL NETWORKS)

The Microsoft Research team introduced this network, and Convolutional Neural Networks are the primary forms of the deep learning paradigms, which are the provable approach to image classification and recognition. Variation of ConvNet is ResNet for Residual Network that was introduced in 2015 by Kaiming He.

In ResNet architecture, as shown above (Fig.1.), which appears similar to the CNN model, the first layer tries to learn the edges, the second layer tries to find the textures, and then the third layer tries to detect the objects. Unlike the traditional CNN model, the maximum threshold for depth as in the later stages of adding more layers tends to degrade the performances due to the vanishing gradient problem. Hence, to overcome such issue training deep neural networks,-based on which the detailed system architecture (Fig.2.) of experimentation on amalgam of Residual networks along with SVM classifier for real-time face recognition system for marking attendance systems.-based on which student attendances can be experimented in the future for multi-view face recognition in the classroom environment installing a camera in the classrooms using this model as a background knowledge with multiple number of participants and finally receiving suggestions or feed backs from the student users for further improvisation.

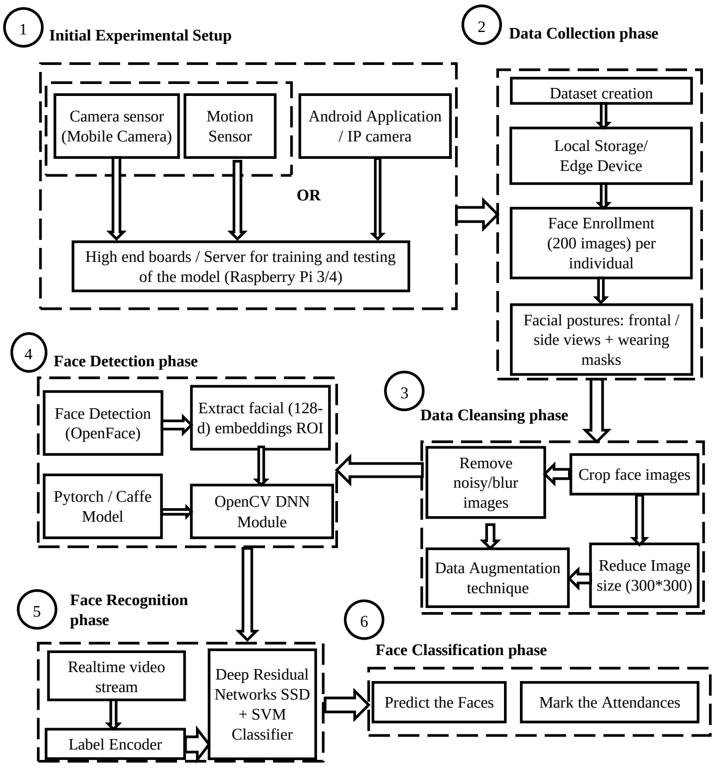

Fig.2. Proposed system architecture

The ResNet architecture proved to be efficient and robust in recognizing visual and non-visual tasks. The depth of the network model plays a vital role in the model performance, but much more deeper networks tend to pose the difficulty in training the network, resulting in more parameters. ResNet consisting of Residual blocks comes into the picture as shown below (Fig.3.). The ResNet framework simplifies the process of training such deeper networks and thereby increasing the model performance

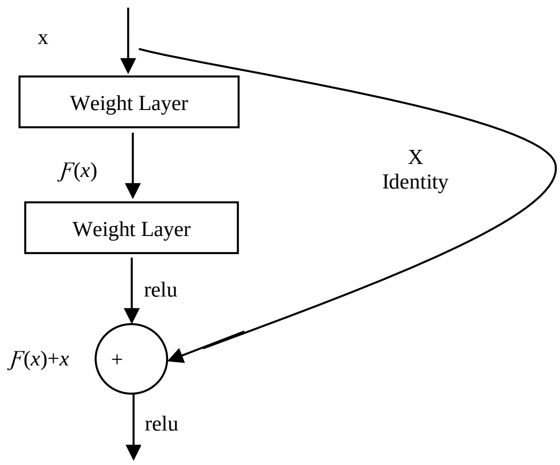

Fig.3. Residual building blocks (ResNet Architecture)

3.3 FACE DETECTION

The initial step in the Face recognition pipeline is Face detection, which deals with identifying and confirming any human faces available in the video frame or the real-time data inputs. First and foremost, the algorithm has to analyse the pictures and point out all the human faces in the picture that are in different constraints like lighting conditions, poses, occluded faces, etc.-based on the selected faces, the features are extracted from those faces like nose, eye features, lips, etc. Then, at last, the features are matched with the person's name known. This is what a human brain does with any time-lapse instantly. This face detection takes the BGR image as an input and draws bounding rectangles or boxes around the detected faces. Those bounding boxes with the most robust confidence are chosen for the next phase of Face recognition. In this study, SSD framework-based Face detection model called OpenFace is used. This framework is-based on the reduced ResNet-10 model.

3.4 FACE RECOGNITION

Face recognition is a pipeline used to authenticate a person-based on the facial features present in the stored template. The face image template includes eyes, nose, mouth, ears, chin, lips, and cheek captured using the camera or similar sensor technology.

Pseudo-code for the Proposed System:

Step 1: Collect the participants frontal and side viewed face images from the onboard camera / Mobile camera that acts as an IP camera

Step 2: Apply ResNet-based Neural Network model (Detection of Faces)

Step 3: Execute pre-processing techniques for instance scaling, rotating, cropping, etc

Step 4: Resize the images to suit network architecture 300*300

Step 5: Extract the 128-d features from detected faces

Step 6: Train all the images using OpenFace

Step 7: Generate Label Encodings

Step 8: Perform classification using Support Vector Machines (Recognition of faces)

Step 9: Put Attendance-based on recognized faces

Step 10: Take live video data for face recognition

Step 11: Extract the features from detected faces

Step 12: Compare Results

3.5 SETTING

All the experiments are conducted with the platform configuration of Raspbian OS, Debian version installed in the Raspberry Pi 3 Model B+ prototype board with Python 3.7 configuration as shown below (Fig.4.). Since the Pi camera has reduced resolution, in this work, a Mobile Camera is used in the laboratory-based classroom environment at the Centre for Research and Development (CRD) lab of the institution. The mobile camera has been used for higher resolution images, connecting the Mobile camera to the Raspberry Pi using Python Programming language. A typical Android application named 'IP camera has been used for this study. As mentioned earlier, before training, the raw images that were collected have to undergo pre-processing. All the datasets were scale transformed to an input size of 300*300 for ResNet architecture. The pretrained CNN system with SVM classifier is then evaluated on the metric to test the model accuracy using recognition accuracy,

Fig.4. Raspberry Pi board used for experimentation

which is the disclosure of predictable labels that are correct and the unknown labels which are new to the data sources.

3.6 DATASET DESCRIPTION

For any deep learning algorithms approached to develop face recognition systems, face images are the system's heart. Dataset creation is the primary challenging task in our study. Maintaining the quality of the data and the quantity remains a difficult task. For the face recognition pipeline, initially, dataset images of the faculties have been taken with a sample of 200 images per faculty with a total of around images for seven faculties with happy, sad, normal, or casual mode with several side views (left, right, top or bottom). Then using Data augmentation techniques, a few more augmented images have been added to the face datasets reaching more 100 images to improve the model's performance in real-time video inputs using three different system configurations.

4. EXPERIMENTAL RESULTS

This section tries to present the experimental results found in the face recognition pipeline using the Deep Residual Networks and SVM classifier. First, the image facial features from the given dataset for training were created-based on our own image dataset. The images were initially of 1024*768 pixels. But the neural network architecture demands 300*300 size input size. So the idea is then re-scaled to 300*300 pixels.

Fig.5. Recognizing without a mask as in [15]

Fig.6. Recognizing wearing a mask as in [15]

Face recognition system had been experimented with [15] in the previous research work as the reproduced results-based on the research work on SCADA architecture [16], results of achieving accuracy 80-90% in the real-time testing inferences are shown (Fig.5-Fig.8).

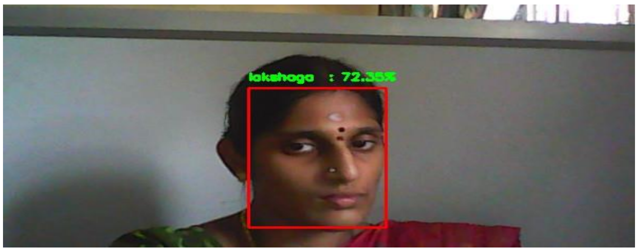

Fig.7. Recognizing the face without a mask using mobile camera

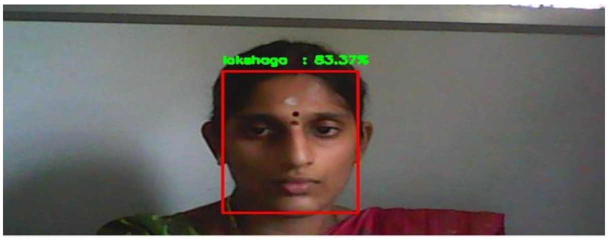

Fig.8. Recognizing the face without a mask using mobile camera

Table.1. Tabulation of the system configurations vs. FPS rate configurations

System configuration	Python Version	Camera resolution	Elapsed time	FPS
Mobile camera	Python 3.7	1920*1080	574.75	0.64
Laptop camera	Python 3.8	1920*1080	118.46	1505.94
Webcam	Python 3.7	640*480	121.39	0.63

The experimental dataset was collected using Mobile camera and the inferences of the model in achieving a decent accuracy has been tested in different environments as shown in the below tabulation (Table.1.). The graph on tabulation clearly shows that the system no matter what the language configuration is such as Python 3.7 or Python 3.8, The model can show inferences of face recognition processing images in the real-time video in the FPS rate of around 0.6ms by testing using web-camera (Logitech C270) with the resolution of 640*480, mobile camera with the resolution of 1920*1080 and finally Laptop Builtin Camera (Lenono V110) on Windows 10 environment.

5. LIMITATIONS

5.1 MEMORY CONSTRAINTS

When dealing with video processing tasks like face recognition systems, high-resolution images cost expensive resources and high-end cameras. In this study, HD mobile cameras were used for data collection. Processing these high-resolution images in a matter of seconds through Deep Learning-based algorithm implementations. Meanwhile, in the future, image transfer via the internet should be addressed using some proper protocols.

5.2 IMAGE QUALITY AND SIZE

Digital images are usually captured in different dimensions-based on the camera considered for the work. When the image size is reduced, it will affect the recognizing performance of the face recognition systems. In this experimental study, frontal face images from a particular distance of the mobile camera is carried out. Much longer distances are not yet explored that can be approached by using image enhancement techniques or incremental learning transforming from high quality to low-quality images. In the future study, longer distance images will be considered for data collection yet maintaining the stable quality in the image taken to recognize the person.

5.3 ANGLE FIXATION

To build a generalized face recognition system that can easily recognize a person from all angles. Since it is possible to create proxy faces to fool the system by appearing spontaneously, with beards, masks, spectacles, etc. Hence in this study, faces were captured with causal faces appearing happy, sad/neutral, and normal faces on causal talks, focusing from various angles such as left, right, top or bottom.

5.4 PROXY TIMINGS

The main purpose of this work is to automatically recognize the faces of the individuals and thereby simultaneously avoid time fraud in the collegiate lecture hours. Furthermore, it will serve as a considerable advantage to automatically monitor on an hourly basis without any demanded workforce to monitor their check-in and check-out time without relying on identity cards or fingerprint scanners to swipe for marking the attendances.

6. CONCLUSION AND FUTURE WORK

In this paper, pretrained CNN architectures are experimented with for handling face recognition systems. Initially, this work applied the ResNet-10 architecture model to extract facial features, and then the Support Vector Machine (SVM) is finally used for classification of face images. The developed and implemented face recognition system is experimented on improved recognition metrics based on the exploration of using Deep Learning for IoT devices, OpenFace is used with the dnn module of the OpenCV library supported in Python-based programming language. The mobile camera has been used for dataset creation and experimentation to address the problem of processing low-resolution images. This paper applies ResNet-10 + SVM to mark attendance with decent satisfactory results. In the future, they can be improved for much more robust results. This research tries to propose a systematic method contributed to image dataset preparation and image annotation, capturing the faces of the participants in the frontal view sitting in a laboratory class-based environment. This dataset focuses on frontal face images and side view images of faces covering factors like lighting, rotations, occlusions like wearing masks-based on the current pandemic scenarios. In the future, this work will try to propose a public access student classroom dataset contributing to face detection and recognition systems to train and test on future developments. It is needed to ensure that the proposed dataset has its limitations and advantages on storage processing and quality of the images. Only South Indian faces of limited participants have been considered for variability in the individuals in this work. This research is planned to further progress on the contribution of images taken from classroom environments to support our proposed work.

Moreover, this paper will further investigate this work leading to its suitability for its usage in the classroom environment based on the results, it is clearly defined how the method achieves 70-80% accuracy to recognize the faces. As a demand, this face recognition system will be limited to indoor usage with good lighting conditions. In the future, illumination variation will be addressed to improve the model accuracy and generalization. When enhanced in the future, this study would initiate researchers to use this opportunity tailored towards classroom-based intelligent systems preventing bogus attendances. This attendance system has been implemented in the desktop environment as well as tested in the edge computational device like Raspberry Pi 3 Model B + board. The system was tested in the laboratory environment for its accurate face recognition. Even though their were false positives and negatives while recognizing the faces, the model seemed to perform well in the frontal facial posture as well as in the side views for recognition of the individual. Considering occlusion problem via face masks, the model fails to recognize the faces in the side positions wearing the masks. This issue has to be rectified in the future development model. This model after final deployment can be transformed to a desktop-based application software or a Raspberry Pi-based edge software that runs automatically on the Raspberry Pi boards without manual initiation. The above results shows that the system reaches a maximum accuracy of 80-90% on frontal faces, 70-80% on side view angles respectively based on the final tabulation, it can be concluded that the system works better in the Edge device using webcamera or the Mobile camera for testing. In future work, student attendances can be experimented on multi-view face recognition in the classroom environment.

REFERENCES

[1] Y. Sun, D. Liang, X. Wang, and X. Tang,"Face Recognition with very Deep Neural Networks", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-11, 2015.

[2] A. Jain, Ross and Nandakumar, "*Introduction to Biometrics*", Springer, 2011.

[3] S. Gupta, A. Agrawal, K. Gopalakrishnan and P. Narayanan, "Deep Learning with Limited Numerical Precision", *Proceedings of International Conference on Machine Learning*, pp. 1737-1746, 2015.

[4] X. Qi and C. Liu, "Enabling Deep Learning on IoT Edge: Approaches and Evaluation", *Proceedings of IEEE/ACM Symposium on Edge Computing*, pp. 367-372, 2018.

[5] O.M. Parkhi, A. Vedaldi and A. Zisserman, "Deep Face Recognition", *Proceedings of International Conference on Machine Learning*, pp. 1-12, 2015.

[6] P. Sarkar, D. Ranjan, G.R.S. Mishra, and Subhramanyam, "Automatic Attendance System using Deep Learning

Framework", *Proceedings of International Conference on Machine Intelligence and Signal Analysis*, pp. 335-346, 2019.

[7] T. Sutabri, A.K. Pamungkur and R.E. Saragih, "Automatic Attendance System for University Student using Face Recognition Based on Deep Learning", *International Journal of Machine Learning and Computing*, Vol. 9, No. 5, pp. 1-13, 2019.

[8] Y. Lecun, Y. Bengio and G. Hinton, "Deep Learning", *Nature*, Vol. 521, No. 7553, pp. 436-444, 2015.

[9] D. Lizzie and J. Harirajkumar, "Contactless Attendance System using Siamese Neural Network based Face Recognition", *Materials Today: Proceedings*, Vol. 45, No. 2, pp. 1-7, 2020.

[10] V.K. Polamarasetty, M.R. Reddem, D. Ravi and M.S. Madala, "Attendance System based on Face Recognition", *Work*, Vol. 5, No. 4, pp. 1-19, 2018.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, 2015, "Rethinking the Inception Architecture for Computer Vision", *Proceedings of International Conference on Machine Intelligence and Signal Analysis*, pp. 44-56, 2015.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu and A.C. Berg, "SSD: Single Shot Multibox Detector", *Proceedings of International Conference on Machine Learning*, pp. 899-906, 2016.

[13] Open-Source Computer Vision Library, Available at https://opencv.org/, Accessed at 2021.

[14] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A General-Purpose Face Recognition Library with Mobile Applications", Technical Report, CMU School of Computer Science, pp. 1-145, 2016.

[15] M. Lakshaga Jyothi and R.S. Shanmugasundaram, "Design and Implementation of Intelligent Classroom Framework through Light-Weight Neural Networks based on Multimodal Sensor Data Fusion Approach", *Intelligence Artificielle*, Vol. 35, No. 4, pp. 291-300, 2021.

[16] L.S. Huang, J.Y. Su and T.L. Pao, "A Context Aware Smart Classroom Architecture for Smart Campuses", *Applied Sciences*, Vol. 9, No. 9, pp. 1837-1837, 2019.