2021 International Conference on Computer Communication and Informatics (*ICCCI* -2021), Jan. 27 – 29, 2021, Coimbatore, INDIA

Attendance Management System

Ravi Kishore Kodali
Department of ECE
National Institute of Technology, Warangal
WARANGAL, INDIA
ravikkodali@gmail.com

Raghu Vamshi Hemadri
Department of ECE
National Institute of Technology, Warangal
WARANGAL, INDIA

Abstract—**Marking of attendance is one of the oppressive tasks in a lecture. Moreover, it takes a lot of time to mark the attendance of students manually. Some of the problems need to be addressed regarding attendance marking are the possibility of a proxy, the analysis of attendance of a student which could include how frequently one is skipping the classes. In recent days, traditional methods of marking attendance are turning out burdensome tasks. A solution for automating the attendance system efficiently can be done by facial recognition and is also a thriving task in face recognition. However, the automation is done using biometric systems in past, but it is not an efficient way of doing as faking a fingerprint is much easier. And also the authenticity and accuracy signifies the methods to be chosen for this task, which many proposed methods lack. This research mainly focuses on the methods of marking attendance and firmly evaluate and analyze intelligent techniques to mark attendance. In this paper, we propose a novel method of marking attendance using facial recognition. The proposed method uses small and accurate deeply supervised network for recognition of faces in a wild classroom scenario. A web application is developed for easy inference to the users. All the analytics is performed on Amazon Elastic Compute Cloud (Amazon EC2) Instance.**

Keywords- MTCNN; EfficientNet; Amazon EC2; Amazon S3

I. INTRODUCTION

Face recognition could be one of the efficient ways of replacing biometric for marking attendance. Making use of facial features of a man or a woman for identification and recognition is novel. Facial features of each person are unique. Being unique and precise, facial features can be an aid for identification and detection of persons in wild which could be accurate and exquisite. Facial recognition can be employed in various fields involving vision directly or indirectly like security and surveillance systems but has not been efficiently used due to noticeable faults and flaws. Most common way of marking attendance has many advantages and disadvantages as well. Any automation mainly aims to reduce the manual effort and the time consumed which are main setbacks for traditional attendance marking system. Focusing on these limitations, improvements are made to use the biometric system to mark attendance. However, biometrics reduces the manual effort but the time consumed is not improved moreover it is sometimes increased. And also, the biometrics is not cheap in cost. No sooner facial recognition is made accurate and efficient by machine learning algorithms, it is employed for attendance marking.

Attendance marking via biometrics is mostly done by scanning the fingerprint and also rarely by iris scan. Rapidly improving research is advancing the current technologies and replacing the traditional methods in many fields. How- ever, most of the classrooms with multiple technologies are being developed biometric has become an integral part of it. Advanced technologies always aim to improve on current technology limitations which in this case are removing hindrances, complex widgets, time taken to complete the process. Most of these limitations can be overcome by using facial recognition for marking attendance. Attendance marking through face recognition is done by using a cam- era, which captures an image of class students. Marking attendance through face recognition mainly involves two stages, first is face detection and second is face identification. Though this method has some limitations like accuracy and resources, this can be used to maintain attendance log easily and the analysis of students' presence in class can be made easier. With improved technology resources and computation are no more limitations. Accuracy can be improved by using proper datasets for training, good environmental conditions for testing and also it can be improved by improving the algorithms used for face recognition. Many face recognition techniques were introduced out of which few use improved concepts and techniques of data mining, statistics and some use typical standard face recognition algorithms. Some methods aim to develop a hardware system which incorporates a camera and a microcontroller whereas some methods use a mobile phone. On researching different face recognition methods for attendance marking and exploring their limitations we proposed a novel method to mark attendance using a web application and amazon web services (AWS). Some of the proposed methods which are notable in attendance marking using face recognition are discussed in the next section under literature review. The third section explains our proposed method and the different AWS services and models it employs. In the last section, the experimental validation of our model is discussed with comparison to different models

II. LITERATURE REVIEW

Research involves analyzing the solutions proposed by others and find out the shortcomings of the proposed systems to put forward a better solution to the problem. To overcome the problems in marking the attendance, the approach used was based on face recognition where the main task is to match the recently taken images with those available in the database.

2021 International Conference on Computer Communication and Informatics (ICCCI) | 978-1-7281-5875-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCI50826.2021.9402659

Naveed et al. presented the solution to this in [1], where two databases are available. One of the databases has the student pictures and the other is a database for attendance. When the camera takes a picture of the class, the background of the image and the noise along with it are removed. Consequently, classification of the skin is done and the face which is detected is matched with the face database for marking the attendance in the database for attendance. This method of monitoring everything continuously and fixing the seats to mark the attendance automatically by capturing images of students in a class present in [2] was proposed by Kawagachi.et.al. The system proposed by them constantly observes the attendance and despite the availability of service in the system where videos can be streamed, facial detection and image taking were used for uninterrupted observation. It was also stated by them that, by using different mathematical calculations, they can approximate the seating order followed by the students. The system follows a simple architecture where the number of cameras put into use is two; Out of the two cameras, one is used for sensing purpose and the other for taking the images. The capturing of picture followed by comparing it with the database is done repeatedly for better attendance accuracy and hence, the term "continuous observation" was coined by them.

The process of marking the attendance of the students was ensured by Visar et al. in [3], by using fixed seating. Detection of face in real time on the basis of an already present learning management system was proposed by them. To detect the students' faces and get their attendance marked, they fixed the seat order of the students. This was done for all the classes, by taking a picture shot of the classroom through the camera stationed at the front, they marked the attendance and the camera on the roof was used to take the images of faces and match it with the ones available in the record to mark the students' attendance

A more effective method was utilized by Abhishek Jha et al. for face detection by using statistical techniques PDA and LDA along with the process of captured image matching and stored one to mark the attendance. The process of marking the attendance which, in general is lengthy and liable to error, if compromised will impact the student drastically, was addressed by them. The proposition of a system where it does the image computation in a certain process for match scoring to be done was given by them. While using certain algorithms like, LDA, PCA and color detection can achieve this, the extraction of features like eyes, nose and face outline etc., from the image. The chances of attendance being marked increases with increasing mark score. The image matching and PCA techniques along with calculation of eigen values were used to take the students' attendance. A camera was fixed in the classroom which took a snap shot in the system proposed by Nirmalya et al. [5] and the matching of the face was done by calculation of eigen vectors and eigen values along with the use of PCA algorithm at the backend. The system learned to identify a picture when it repeatedly captures an undistinguished image and adds it to the database.

Llocal binary algorithm (LBA) is a statistical technique which was used by Francisco et al.in [6] along with weighted mask, and is based on local feature based on local or feature based approach to recognize face. Extraction of facial features takes place in this system. Local binary patterns build masks through which the areas to be marked were distinguished for face recognition. To get weighted masks, technique of data mining was applied to the built masks. The most greyish area had greater importance in recognizing face as depicted by weighted masks, and it mainly focuses on the principal features on the face like eyes, nose, cheeks, forehead and lips etc. To address the traditional problem of attendance marking, Stefano Arca et al. in [7], in order to locate facial fiducial points used a different statistical technique called Gabor filters. Special face fiducial points were used and Gabor filters were applied on student picture with color. There could be 4-64 fiducial points however, the system emphasizes on making use of 31 points as the computation is easy with faster operation and less processor power use. In addition to the literature mentioned, other methods to address the problem have also been discussed.

A different solution other than the prior mentioned was given by Muthu Kalyani et al. in [6]. In the put forth system, a mobile device with enhanced 3D modelling was attached to a CCTV camera which was used for facial recognition. After the image was taken, canonical technique and 3D modelling was applied to the taken picture and then it was matched with the image of the student and the stored image. In case there wasn't a match between the images, then the image is stored in the database for strangers, aiding the administration in security enhancement. Another system proposed by Nasser et al. in [8] for attendance of employees in firms argues that the task of marking attendance in large companies is not easy. For this, they suggested a system where after face detection and attendance marking, then SQL database is used to store them. All the employees are initially asked to take a picture. The picture is taken by the camera followed by removing noise in the picture and background, after which the picture is matched with a previously stored picture for attendance marking. The solution to the orthodox problem was given by J.G.Roshan et al. in [9] where existing systems for attendance management were amended or something unattended was pointed out. They pointed out that women/girls have veil covering their faces in Muslim/Islam countries and males have beard, which might change their appearances on everyday basis. The algorithm used by system should be strong such that the presence of veil or different beard styles and length shouldn't hinder the attendance marking procedure.

III. PROPOSED CONCEPTUALMODEL

Our method of attendance management uses Amazon Web Services (AWS) for both storage and analytic purpose. Firstly, images are students under different lightening conditions and different poses are collected. An EfficientNet baseline is trained using Triplet Loss with this dataset. This dataset is stored on Amazon S3 for future training purposes. The trained model is deployed on Amazon Elastic Compute Cloud with a web application developed using Flask.

A. Model architecture

Face recognition is a computer vision task of distinctive and corroborative, of an individual supported by a photo-graph of their face. Recently, deep learning convolutional neural networks have surpassed classical strategies achieving progressive results on normal face recognition datasets with very less computational complexity and inference times. One example of a progressive model is that the EfficientNet model developed by researchers at the Google AI Research. Though the model is difficult to implement, but is resource intensive to coach, it is simply used through the utilization of freely obtainable pre-trained models and third-party open supply libraries. The process of face recognition is divided into 2 major tasks: Face detection and Face identification. For Face detection task Multi-Task Cascade Neural Network (MTCNN) is employed and for face identification, an EfficientNet pre-trained on faces is employed

1) MTCNN for face identification: Multi-task Cascaded Convolutional Neural Networks (MTCNN) is AN formula consisting of three stages, that detects the bounding boxes of faces in a picture at the side of their five purpose Face Landmarks. every stage bit by bit improves the detection results bypassing it's inputs through a CNN, that returns candidate bounding boxes with their scores, followed by non-max suppression. The overall pipeline of the approach is shown in Fig. 1.

Given a picture, it's at the start resized to completely different scales to create an image pyramid, that is that the input of the subsequent three-stage cascaded framework:

Stage 1: we have a tendency to exploit a totally convolutional network, called Proposal Network (P-Net), to get the candidate windows and their bounding box regression vectors in a very similar manner as. Then we have a tendency to use the calculable bounding box regression vectors to calibrate the candidates. After that, we employ non-maximum suppression (NMS) to merge extremely overlapped candidates.

Stage 2: all candidates area unit fed to a different CNN, known as Refine Network (R-Net), that more rejects an outsized variety of false candidates performs standardization with bounding box regression, and NMS candidate merge.

Stage 3: This stage is analogous to the second stage, however during this stage we have a tendency to aim to explain the face in additional details. especially, the network can output 5 facial landmarks' positions.

a) 2) *EfficientNet for face identification:* EfficientNet[9] is one in all the foremost powerful deep neural networks that have achieved superior performance with fewer parameters. EfficientNet has achieved glorious generalization performance on alternative recognition tasks. Their area unit several variants of EfficientNet design i.e., same construct however with a distinct range of layers, namely EfficientNet B(0-7). The EfficientNet B(1-7) are scaled versions of the baseline model EfficientNet B0, using compound scaling method. The EfficientNet B0 is architecture by AutoML algorithm: Neural Architecture Search, by optimizing both accuracy and FLOPS (Floating Point Operations per Second) by using reinforcement learning techniques.

a) Neural Architecture Search and EfficientNet Architecture: The EfficientNet B0 is a baseline architecture of all EfficientNets. The network is built using depth wise convolutions for faster computation of convolutions. This *architecture is*

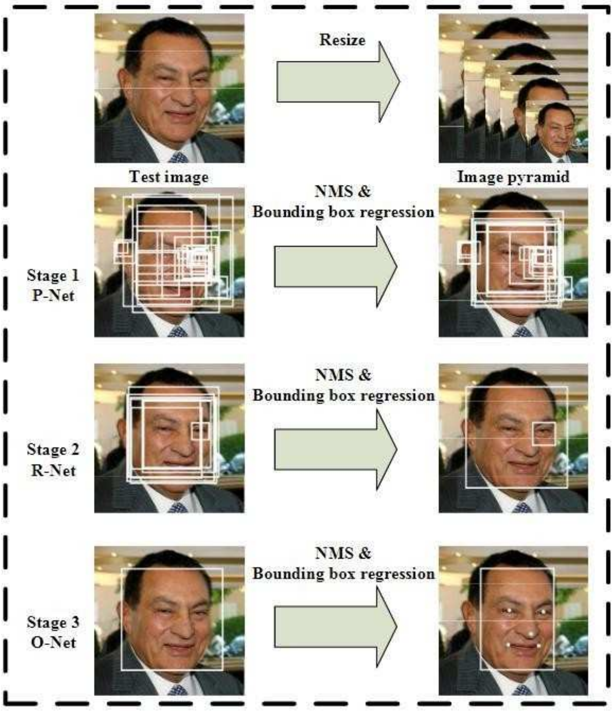

Fig. 1. Overall pipeline of the MTCNN approach

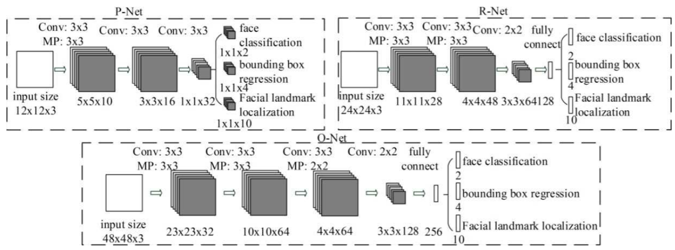

Fig. 2. Architectures of P-net, R-net and O-net

developed in a similar approach to MNasNet paper i.e, to use Neural Architecture Search (NAS), a reinforcement learning based algorithm. NAS aims to optimize

$$ACC(m) * [FLOPS(m)/T]w \qquad (1)$$

where ***ACC*(m)** is the accuracy of the model, ***FLOPS*(*m*)** is the accuracy of floating-point operations per second(FLOPS), ***T*** is the targeted FLOPS, ***w*** is manually chosen to be 0.07, to achieve higher accuracies. The optimal architecture given by NAS is shown in Fig. 3. The MBConv is an inverted bottleneck block with convolution layers(kernel size specified), shown in Fig. 4. These blocks have convolution layers along with squeeze and excitation blocks(SE Blocks). The SE block, which has an output shape of (1x1x1 Channels), adds weightage to each channel layer, rather than weighing each channel equally like in ordinary convolution.Moreover, EfficientNets employ Swish activation, which is found to perform better than most used activation function ReLu. The Swish activation function is given

$$Swish(x) = x * Sigmoid(x) = x * \frac{1}{1+e^{-x}} \qquad (2)$$

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Fig. 3. Bottle neck structure of MBConv3 and MBConv6

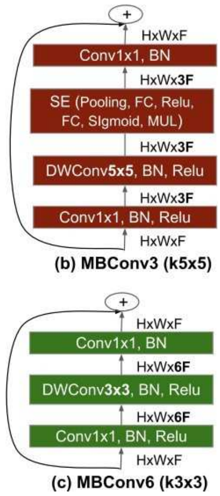

Fig. 4. Architecture of EfficientNet B0

A. Amazon S3

Amazon Simple Storage Service (Amazon S3) is an object service. This service offers data availability and network security, thereby enhancing its performance on a large amount of data for overall industry-leading scalability. Customers of different sizes and industries can use it to store and protect their data from various sources such as websites, mobile applications, enterprise applications and IoT devices. This data can also include backup and restore data, archived data. This enables its use in big data analytics. Amazon S3 provides uncomplicated-manageable features to organize customers data and configure finely-tuned access controls to meet their organizational, business and compliance requirements. Amazon S3 has a simple web services interface to store and retrieve the data. This feature provides developer access to similar highly reliable quick data storage infrastructure used by Amazon to run its global network. The service aims toassist developers.

B. Amazon EC2

Amazon Elastic Compute Cloud (EC2) is a vital part of Amazon's cloud-computing platform and Amazon Web Services (AWS). So, users can use virtual computers to run their applications. EC2 enhances the scalable deployment of applications through a web service in which the user can boot an Amazon Machine Image (AMI) to configure the virtual machine (called "Instance" by Amazon), consisting of the desired software. User is free to create, launch and terminate server-instances as required under different circumstances. User can access the platform by renting the active server on seconds basis (so the term "Elastic" is used). Using EC2 users can have control over the geographical location of instances that allows for latency optimization and high redundancy levels. Amazon Elastic Compute Cloud (Amazon EC2) supports scalable computing capacity in the Amazon Web Services (AWS) cloud. Amazon EC2 eliminates the need to invest in hardware upfront for quick development and deployment of applications. Amazon EC2 can be used to launch a wide range of virtual server customer needs by configuring security and managing storage. Amazon EC2 enables the users to enhance changes in requirements or variations in popularity, reducing the need to forecast traffic.

A. The Attendance Management Web Application

The trained model is deployed on AWS EC2 t2.micro instance. The t2.micro instance has 1 CPU with 1 GB RAM. Faster inference can be guaranteed with higher computational devices. A web application is developed for easy access to the users and is available anywhere over the internet. The web application is created using Flask for easy testing and usage of the attendance management system. Fig.5 shows the response of the web application after testing. This web app also has a feature of the attendance sheet for all the days which is automatically generated.

Fig. 5. The web application output of the testing inference

IV. RESULTS

During testing, the MTCNN is used to extract faces and the trained model is used to find the embeddings of a face in test image. The embeddings are compared with embeddings of all the images in database by euclidean distance. The image having the least euclidean distance is selected, and the accuracy is reported as top 1% accuracy. The top 1% accuracy and inference time is compared and presented in Table 1. The image shown in fig 6. is used for testing, the resolution of testing image is 4068x3456. The EfficientNet B4 is found best for deployment with good accuracy. The EfficientNet B4 and ResNet-50 has almost equal number of parameters and FLOPS but EfficientNet B4 has an accuracy of 94% which is almost 8% more than that of ResNet-50.

Fig. 6. Captured image for testing

The EfficientNet B0 however has very less parameters of 5.3 million and has inference time of only 0.008 seconds for a batch of 16 images on 8 GB NVIDIA GTX 1080 GPU. The accuracy of EfficientNet B0 is however not more than 79%, but is found to perform equally accurate to that of ResNet-50 when the network is scaled with the parameters w = 1.8, d = 1.4, r = 380 and dropout rate of 0.4 resulting in the architecture of EfficientNet B4. The EfficientNet B4 has an inference time of only 0.02 seconds whereas ResNet-50 has 0.05 seconds of inference time on NVIDIA GTX 1080 GPU. After an extensive comparison and analysis, it can be concluded that Attendance Management System proposed with EfficientNet B4 is accurate and fast. Moreover, it is trained using Triplet Loss, so whenever new students are added to class no retraining of network is required. Because we learn to map features of same class closer in the latent space and of different class farther. So, this proposed model works even for new faces, without any additional cost of retraining. This model being trained on wide range of faces captured under different lightening conditions and real world environment conditions, this model works for class room environments with higher accuracy.

REFERENCES

[1] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks", IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016.

[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)

[3] T. Ahonen, A. Hadid, M. Pietikainen, "Face description with local binary patterns: Application to face recognition", Pattern Analysis and Machine Intelligence IEEE Transactions on, vol. 28, no. 12, pp. 2037-2041, 2006.

[4] Nagi J, Ducatelle F, Di Caro GA et al (2011) Max-pooling convolutional neural networks for vision-based hand gesture recognition. In: 2011 IEEE international conference on signal and image processing applications (ICSIPA2011), pp 342–347

[5] Aitkenheada MJ, Mcdonald AJS (2003) A neural network face recognition system. Eng Appl Artif Intell 16(3):167–176

[6] Ahonen T, Pietikainen M, Hadid A (2004) Face recognition based on the appearance of local regions. In: 17th international conference on pattern recognition, vol 3, pp 153–156

[7] Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by inde- pendent component analysis. IEEE Trans Neural Netw 13(6):1450–1464

[8] 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), Boston, MA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682

[9] Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." arXiv preprint arXiv:1905.11946(2019)