RESEARCH REVIEW JOURNALS
Volume-04
Issue-10
October-2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[Peer Reviewed Journal]

Bunk Alert System Based on Convolutional Neural-Network in Real-Time

[1]Nidhi Chauhan, [2]Soni Yadav, [3]Prof. Sneha Sankhe

[1&2]UG Student, Theem College of Engineering, I.T Department, University of Mumbai, Mumbai, Maharashtra (India)
[3]Professor, Theem College of Engineering, I.T Department, University of Mumbai, Mumbai, Maharashtra (India)

ARTICLE DETAILS

Article History
Published Online: 14 Oct 2019

Keywords
Bunk Alert, ConvNet, Face Detection, HOG, MLP, Max-Margin Object Detection, OpenCV, SVM, Triplet Loss.

***Corresponding Author**
Email: nidhic826[at]gmail.com

ABSTRACT

Nowadays, students find it easy to bunk lectures in college unnecessarily. As a result, to stop and to decrease them we have decided to build a system which will detect their faces. The system is composed of five main steps: face segmentation, facial features detection, face alignment, embedding, and classification. We are using deep learning methods for the fiducial points extraction and embedding. Support Vector Machine (SVM) is used for classification. This system is capable to run in real-time therefore it is easy to send an email alert with the students name to the HOD and the student so that the student will get to know he/she is bunking the lectures and the camera captured him/her The warning email will help students to attend lectures and hence it will decrease the most concerned problem i.e. Bunking Lectures. This System is based on Face Detection Algorithm and the details of the system will be shared totally throughout this paper with the comparative study of other similar algorithms.

1. Introduction

THIS work is moreover the covering lines on the field of managing the location of students in a college campus and making them attend lectures by using the algorithm, where creating a Convolutional Neural-network for facial feature extraction and identification, creating a database for training the neural-network and testing it on different test cases is done. Creating a database for an IT student is normally easy but the new thing which is added in this work is the Multiview Face Detection Algorithm which can detect faces using a surveillance camera frame by frame.[11] Use of Deep Convolution Network and Multiple-Layer Perceptron[1] is highly used in recent years of reaches related to computer vision which can achieve human-level efficiency and with the processing speed and vast memory of an artificially intelligent system it has become easy and feasible to use.

A. Research in Past

By comparing different famous and recently used algorithms that support the use of DNN we can identify a simple and efficient system that helps in deploying an easy to manage bunk alert system with the department student's dataset which can be integrated with a lightweight alert system.

a) OpenFace

OpenFace is a *facial behavior analysis* opensource system that uses state-of-the-art algorithms in real-time to 1) detect facial landmarks, 2) estimate head pose 3) eye-gaze estimation and provide us with 4) facial action unit recognition. The OpenFace facial behavior analysis pipeline includes supplying an input image in which face is detected and from it facial landmarks are extracted from the help of which eye gaze estimation, head pose estimation, face alignment, and appearance extraction, feature fusion, and person normalization is done then dimension of face alignment and appearance is reduced and added with feature fusion and person normalization to provide us with the facial action unit which shows the highest possibility of a certain face expression.[4]

b) FaceNet

FaceNet solves the problem of *pose and illumination variance*, with a scale of 0-4 where 4 represents the image of a different person and below 1.1 represents a match with the correct identity of a person. It uses the Euclidean space formula to store images in a database where distance is directly proportional to the measure of face similarity. It uses a triplet loss function where 3 images are compared from which 2 images contain the same person and the 3rd person image is used to create a wide gap between the measurements of datasets images. It helps with the clustering of the same person image dataset in a folder. In this system, inputs are provided in batches to a Deep ConvNet which is further normalized using L2 normalization then the embedding process comes in the process where triplet loss is checked and then images of the same person are clustered. It achieves 99.63% accuracy and 95.12% accuracy in the YouTubes dataset and provides greater representation efficiency.[3]

c) DeepFace

DeepFace uses 3D face modeling to train the Deep ConvNet which is 9 layers deep and uses piecewise affine in alignment and representation process and then the image can be classified. There are roughly seven steps in the alignment of a face in DeepFace those are as follows: 1) detecting a face, 2) cropping the face for 2D alignment, 3) extracting 67 fiducial points for delaunay triangulation and addition of triangles to the contour of the face to avoid discontinuity, 4) reference 3D shape is transformed into 2D aligned crop image, 5) triangle visibility concerning 2D-3D camera (darker triangle used to show less visibility), 6) 67 fiducial points induced by 3D model for directing piece-wise affine wrapping, which provides us with 7) final frontalization crop.

It achieves an accuracy of 97.35% for Labeled face in the Wild and for the YouTube dataset, it reduces 50% error compared to the current art-of-state algorithm.[6]

2. Proposed System

The system is composed of five main steps: face segmentation, facial features detection, face alignment, embedding, and classification. We are using deep learning methods for the fiducial points extraction and embedding.[1] Support Vector Machine (SVM) is used for classification.

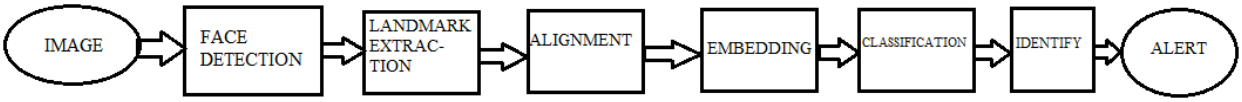

Fig2.1: Simple block diagram of the system's workflow.[1]

Face Detection

Before recognizing a person in an image, it is necessary to locate its face. Although a classical technique from 2004 known as Viola-Jonas [2] has been broadly used, better techniques using deep learning are rising. This work uses a Histogram of Oriented Gradient (HOG) method the Max-Margin Object Detection (MMOD), implemented by using the dlib library. The HOG [11] is a scale, rotation, and illumination invariant descriptor and has been mostly applied in image processing and computer vision applications. MMOD, which is similar to an SVM, is trained in HOG feature space to detect objects with high accuracy even when the image is in a good quality [11]. The segmented face is delivered to the facial feature extraction step.

Facial Feature Extraction

In this step, a regressor is employed to extract critical facial key points on eyes, eyebrow, nose, lips, and others. There are many ways to perform this task, but computer vision-based methods are less expensive and intrusive. Nowadays, the algorithms using deep learning have better results. This work tests the use of Multi-Layer Perceptron (MLP) and Convolutional (CNN) neural networks, described as follows.

1) Multi-Layer Perceptron Neural Network: In an MLP network, with two layers, with N1 neurons in the hidden layer and N neurons at the output, for an input $\boldsymbol{x} \in \mathbb{R}^{(C*H*W)}$ representing an image with C channels, height H and width W, the output is:

$$\boldsymbol{h}=\boldsymbol{W}_{xh}\,\boldsymbol{x}+\boldsymbol{w}_{h}, \quad z=\sigma(h), \quad \boldsymbol{y}=\boldsymbol{W}_{zl}\,\boldsymbol{z}+\boldsymbol{w}_{l}, \qquad (1)$$

$\boldsymbol{W}_{xh} \in \mathbb{R}^{(C*H*W)\ x\ N_1}$, $\boldsymbol{w}_{h} \in \mathbb{R}^{N_1}$ are the weights at the input layer, $\sigma(\cdot)$ is differentiable nonlinear function, $\boldsymbol{W}_{zl} \in \mathbb{R}^{N\ x\ N_1}$ and $\boldsymbol{w}_{l} \in \mathbb{R}^{N}$ is the weight at the output layer. The training consists of minimizing a cost function concerning the network parameters. Since it is a nonlinear problem, it is highly recommended to optimize the function by using small batches from the training set [3].

2) Convolutional Neural Network: MLP networks do not take spatial information from images into account to reduce the number of parameters. Since CNN's use filters to find local structures (e.g., edges in images), they have been applied in many image processing applications and other areas in which data has high dimension and many spatial structures. CNN's perform this task with less computational cost and parameters than MLPs.[3] CNN's usually have two main components: convolutional layers and pooling layers.

CNN's operate over tensors rather than vectors and do so through convolution, that is the reason why they are called convolutional networks. A convolutional layer is composed of F filters f_f of the same size, with width W_f, height H_f and the same number of input channels. A tensor with all parameters has a model $\boldsymbol{F} \in \mathbb{R}^{F\ x\ C\ x\ H_f\ x\ W_f}$. It is common to use $W_f = H_F = K$ (kernel size) with values K = 3, 5, or 7. The inner product between each filter f_f, and each position of the image generates a point in the features map $\boldsymbol{a}_f$. These maps are stacked to compose the output $\boldsymbol{a} \in \mathbb{R}^{F\ x\ H_o\ x\ W_o}$ whose width and height depend on the size of the input (image), the padding P and the stride S. [1]

The pooling layer has an important role in CNN since it aggregates information. Similarly, to the convolutional layers, it has a kernel size, padding, and stride. However, it does not have weights as it only applies the same function and operates on each channel independently. The most common types of pooling are max-pooling (maximum value inside a window) and average pooling (average value inside a window). Usually, the CNN's architectures are composed of blocks with convolutional layers, a nonlinear activation function (e.g., ReLU [1]), regularization layer (e.g., dropout [1] or batch-norm [1]) and periodic pooling layer between the blocks. In the end, an MLP network, also known as a Fully Connected (FC) layer, or a global average pooling layer [20] is added to generate the output. Many hyperparameters must be chosen, such as the number of layers, kernel size, stride, etc. As for many machine learning algorithms, there is no recipe to build a ConvNet. However, some architectures became popular: LeNet, AlexNet, VGG, ResNets, and the DenseNets. It is worthy to mention that ResNets is employed in many state-of-the-art computer vision algorithms due to its simplicity and high generalization capability.[1] In this work, we use this CNN architecture due to the properties mentioned.

Face Alignment

The objective of this step is to standardize the input of the classifier to generate a simpler classification model. Since the facial feature's location is known, it consists of aligning the faces from the image in such a way that the nose, eyes, and mouth are aligned with the centre of the image as much as possible. To do so, an affine transformation is employed. The affine transform does not distort the relative positions of the facial landmarks, i.e., parallel straight lines remain parallel after transformation. The *getAffineTransform* function from OpenCV library returns the rotation and translation necessary to take the original points to the desired ones (an average mask calculated from the points in the training set).[1] The *warpAffine* function,

also from the OpenCV library, applies the transformation, which also scales the resulting image.

Embedding

Given the aligned images, the next step consists of recognizing the person that is in the current image. However, feeding the classifier with the raw pixels of the image is not efficient. Therefore, an embedding process is employed. It consists of extracting information from each image and creates an array, in a lower-dimensional space, which better describes this image. At this point, another deep learning method is employed to extract such an array. Differently from the previous networks, the training objective is to minimize the so-called triplet loss – at each iteration, the network is fed with three images: two distinct images of the same person and an image of a different person. The triplet loss is defined as

$$Loss = \sum_{i=1}^{N} \left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha \right]_+ \quad (2)$$

Fig: The Triplet Loss equation.[3]

Where $f(x) \in \mathbf{R}^d$ is the function that transforms the image into an array of representations, $x^a{}_i$ is an image of a person, $x^p{}_i$ is another image from the same person, $x^n{}_i$ is an image from a different person, α is a margin to be forced between the positive and negative pairs[3]. The network is then trained to create representation arrays that are close to images of the same people and as far as possible for different people. By repeating this process thousands of times for hundreds of thousands of images containing tens of thousands of people, the network is capable of learning how to generate a good embedded representation for each person.

Classification

Due to the alignment and coding processes, the last step was greatly simplified. Ideally, images of the same person have similar feature vectors while images from different people have different ones. Therefore, in this last step, a simple machine learning algorithm is applied to classify each vector as belonging to a person or not. The SVM algorithm was chosen because it is fast for both training and inference [11]. Also, the Database that is big data we will see what other face recognition methods have been discovered lately. Early face recognition algorithms used simple geometric models, but the recognition method has now developed into a science of sophisticated mathematical illustrations and matching processes. Since the initial 1950s meanwhile digital computers were born and the world earned significant processing capability, computer experts have attempted to inducing knowledge and the senses to the computer. During the 1980s, work on face recognition endured mainly dormant. Troubled with the fears expressed in George Orwell's 1984, maximum constituents of the community were very concerned about the usage of a computer system that is competent in recognizing them wherever they go. Since the 1990s, the research interest in face recognition has grown significantly as a result of the following facts:

1. The increase in emphasis on civilian/commercial research work's, the re-emergence of neural network classifiers with emphasis on real-time computation and adaptation.
2. The surveillance-related availability of real-time hardware.
3. The increasing need for surveillance-related applications due to terrorist and drug trafficking activities, etc.

In this work all code is developed in Python, using Pytorch, sklearn, dlib and opencv2 libraries.

Alert System

After the identity of the student classified and if the student is found to be roaming while lectures are conducted then a bunk alert will be send to the head of department and the student itself and then the HOD will decide if the student has a valid reason to be outside of the lecture else delete the alert from the system, if not further actions will be taken by the HOD.

Table 1

SHOWING THE COMPARATIVE STUDY OF RECENT RESEARCHES IN THE FIELD OF COMPUTER VISION AND CLASSICAL ALGORITHM WHICH ARE USED FOR IMAGE DETECTION WITH LOW PICTURE CLARITY.

Author Name.	Published in Year.	Technique.	Pros.	Cons.
Steve Lawrence C.Lee Giles Andrew D Back	1997	Convolutional neural network	It provides partial invariance to translation, rotation, scale, and deformation.96.2% error.	Computation complexity. An error of 3.8% .
Paul viola, Michael Jones	2001	The boosted cascade of simple features.	It detects things which are Objects from region, not background region. 99% is accurate.	It detects without restore to image difference and skin color detection.1 %error is detected.
Florian Schroff, Dmitry Kalenichenko, James Philbin	2015	A unified embedding for face recognition and clustering	It gives greater representation efficiency. It can be used only for 128 bytes per face.	There is an error that occurred regarding datasets.

Tabas Baltrusaitis, Peter Robinson, Louis Philippe Morency	2016	Open source facial. Behavior analysis toolkit	It is capable of facial landmark detection, head pose estimation, facial action unit recognition, and eye gaze estimation.	It cannot detect the face if there are not specific features like a facial landmark.
Wesley L. Passos Igor M. Quintanilha Gabriel M. Araujo	2018	Real-time DL based	It is useful in fiducial points extraction and embedding. It is also useful for the classification task.99.9% accuracy.	Showing error in facial features detection and face recognition. 0.123% error is detected.
Bienvenido, bartido, abad jr	2018	Image processing technique	Identification rate is high with threshold measurement	Normal and illustration normalization should be observed when setting a threshold to avoid false identification.
Chung Hua Chu, Yu Kai Feng	2018	Polynomial neural network classifier on data preprocessing technique	Authentication by eye blink (nonbiometric detection), 99% is accurate.	Computational complexity.
I. Gallo, S. Nawaz, A. Calefati	2018	Convolutional neural network using pipeline	A generic pipeline is capable of creating, cleaning and recognition from videos and images.99.33% accuracy obtained.	Video with low pose variability not detect well. 0.66% of error is detected.
Dr. Priya Gupta, Nidhi Saxena, Meetika Sharma, Jagriti Tripathi	2018	Deep neural network and CovNet technique	97.05 % accuracy is obtained.	2.95% error is detected.
Muhammad Imran Razzak, Saeeda Naz ,and Ahmad Zaib	2019	Deep learning for image processing	It is used for medical image segmentation and classification basically in the health sector.	It is limited in subjectivity, the complexity of the image, extensive variations.

3. Discussion

A. Reasons why student bunk lectures

- Hanging around with friends: Most of the students would hang out with their friends in the college campus or anywhere near the college campus rather than attending their lectures because student's can easily absorb in peer pressure of not attend the lecture.
- Explore college campus: Colleges have big campus where students can hang out and bunk lectures, they can go and sit in the canteen, library, college ground, or college gym.
- The subject is not interesting: If a student doesn't find a subject interesting, they tend to bunk those subjects' lectures.
- Play Sports: Students who like to spend playing rather than attending lectures would spend their time in college grounds and gym and miss their lectures.

B. Purpose of this System

There may be many reasons to give but not attending lectures and staying back in studies can lead to major damaging to a student's life. This system is designed to maintain discipline concerning attending lectures while a student is still in college campus when lectures are being conducted it is transparent because it gives an alert message to the student who is bunking the lecture. And a log is maintained to determine the bunk rate which can be integrated with an analysis system to determine in which teachers lecture students bunk the lectures most.

4. Conclusion

Face recognition with a Bunk alert system is built using python, the HOG algorithm, Max-Margin Object Detection. ResNet to build ConvNet and is employed to locate facial fiducial points (or facial features) and embedding.

1) It finally concluded that many face recognition systems don't work properly but using a simplified technique of learning like triplet loss will help achieve greater efficiency.
2) Face recognition is an easy integration process as it gives high accuracy rates. Also, it requires memory storage for storing the data that is the images that can be stored in a folder like structure for embedding.
3) The best results can be obtained using a customized architecture of the residual network with 18 layers, drastically reducing the number of parameters from 19M to 700k. Due to the employed preprocessing (face detection, fiducial points detection, face alignment, and embedding), the classification model is simple but efficient.

5. Future Scope

Summing it up in every new face recognition technology renders immense aspects and hopes for future growth. It's extremely possible that in a couple of years such practices would be able to prepare signs, expressions, motion patterns, palm & ear prints, voice, and scent signatures. And most of the cons can be beaten by simple actions from the developer's side.

References

1. Wesley L. Passos, Igor M. Quintanilha, Gabriel M. Araujo"Real-Time Deep-Learning-Based System for Facial Recognition" under XXXVI SIMPOSIO BRASILEIRO DE TELECOMUNICAC¸OES E PROCESSAMENTO DE SINAIS - SBrT2018, 16-19 DE SETEMBRO DE 2018, CAMPINA GRANDE, PB
2. Paul Viola, Michael Jones "Rapid Object Detection using a Boosted Cascade of Simple Features" under ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001.
3. Florian Schroff, Dmitry Kalenichenko, James Philbin "FaceNet: A Unified Embedding for Face Recognition and Clustering"
4. Tadas Baltrusaitis ˇ, Peter Robinson, Louis-Philippe Morency "OpenFace: an open source facial behavior analysis toolkit"
5. Muhammad Imran Razzak, Saeeda Naz and Ahmad Zaib "Deep Learning for Medical Image Processing: Overview, Challenges and Future"
6. Yaniv Taigman. Ming Yang, Marc'Aurelio Ranzato, Lior Wolf "DeepFace: Closing the Gap to Human-Level Performance in Face Verification"
7. Chung-Hua Chu and Yu-Kai Feng "Study of Eye Blinking to Improve Face Recognition for Screen Unlock on Mobile Devices" under J Electr Eng Technol.2017; 13(?): 1921-718
8. Bienvenido B. Abad, Jr. "PROPOSED IMAGE PRE-PROCESSING TECHNIQUES FOR FACE RECOGNITION USING OPENCV"
9. Dr. Priya Gupta, Nidhi Saxen, Meetika Sharma, Jagriti Tripathia "Deep Neural Network for Human Face Recognition" under I.J. Engineering and Manufacturing, 2018, 1, 63-71
10. Dr. Priya Gupta, Nidhi Saxen, Meetika Sharma, Jagriti Tripathia "Deep Neural Network for Human Face Recognition" under I.J. Engineering and Manufacturing, 2018, 1, 63-71
11. Navneet Dalal and Bill Triggs "Histograms of Oriented Gradients for Human Detection"

