Table of Contents

Chapter 1

Introduction : Behaviour recognition is the identification and classification of activities on the basis of observation sent through facial expression. In facial expression one can find out or identify the behaviour of a human being .For example if a person is smiling then we can say that either he is happy or if he is unnecessarily smiling indicates he may be mentally disordered. Hence by identifying various expressions of face of a person we may find out his state of mind .A person expresses his emotions like happiness, confusion, sadness, surprise ,anger, excitement, disgust, fear, Desire, contempt, or neutral.

Fig.1 Eight classes of basic emotion [1]

Machine learning systems are infiltrating our lives and are beginning to become important in our education systems [2].With the help of machine learning after putting the images captured one can compare them with the existing expression and may find out approximate state of that person. A human facial expressions are generated through the brain which indicate the interest of him in the running activity and they may be arise due to his past experiences, in a positive or negative way. Due to facial expression our cortex is caused to pay attention since the brain is only responsible to keep us alive and well.

Facial expression data provides crucial insights that will help us to gain insight in complex human behaviour in greater depth. This can play a vital role in various kind of matters, For example the interest of students in online classes,interest of employees in an organization ,interest of a student in chosen field and also interest of a family member in the family etc.

Deep learning is a branch of machine learning which is totally based on artificial neural networks, as neural network is going to imitate the human brain therefore deep learning is also a kind of imitate of human brain. Deep learning methods are able to support very large datasets of faces and learn rich and compressed representations of faces, allowing modern models to first perform as-well and later to outperform the face recognition capabilities of humans. There are several reasons for which we must have to perform behaviour recognition through facial expression as follows:

- For restricting access to a resource to one person, which is known as face authentication.
- For certifying that the person matches their ID, which is known as face verification.
- For assigning a name to a face, which is known as face identification.

Face recognition is often described as a process that first involves four steps; they are: face detection, face alignment, feature extraction, and finally face recognition.

Face Detection. Locate one or more faces in the image and mark with a bounding box. It is a problem of object recognition that requires that both the location of each face in a photograph is identified (e.g. the position) and the extent of the face is localized (e.g. with a bounding box). Object recognition itself is a challenging problem, although in this case, it is similar as there is only one type of object, e.g. faces, to be localized, although faces can vary wildly. Further, because it is the first step in a broader face recognition system, face detection must be robust. Face detection methods that can be broadly divided into two main groups:

- Feature-Based.
- Image-Based.

The feature-based face detection uses hand-crafted filters that search for and locate faces in photographs based on a deep knowledge of the domain. They can be very fast and very effective when the filters match, although they can fail dramatically when they don't, e.g. making them somewhat fragile. Image-based face detection is holistic and learns how to automatically locate and extract faces from the entire image. Neural networks fit into this class of methods.

Face Alignment. Normalize the face to be consistent with the database, such as geometry and photometric.

Feature Extraction. Extract features from the face that can be used for the recognition task.

Face Recognition. Perform matching of the face against one or more known faces in a prepared database.

Matching requires that the candidate matching face image be in some set of face images selected by the system. Similarity detection requires in addition to matching that images of faces be found which are similar to a recalled face this requires that the similarity measure used by the recognition system closely match the similarity measures used by humans Transformation applications require that new images created by the system be similar to human recollections of a face.

The 2011 book on face recognition titled "Handbook of Face Recognition" describes two main modes for face recognition, as:

- **Face Verification**. A one-to-one mapping of a given face against a known identity (e.g. *is this the person?*).
- **Face Identification**. A one-to-many mapping for a given face against a database of known faces (e.g. *who is this person?*).

A face recognition system is expected to identify faces present in images and videos automatically. It can operate in either or both of two modes: (1) face verification (or authentication), and (2) face identification (or recognition) [3].

In all tasks, the input is a photo that contains at least one face, most likely a detected face that may also have been aligned. The output varies based on the type of prediction required for the task; for example:

- It may then be a binary class label or binary class probability in the case of a face verification task.
- It may be a categorical class label or set of probabilities for a face identification task.
- It may be a similarity metric in the case of a similarity type task.

Perhaps one of the more widely known and adopted "machine learning" methods for face recognition was described in the 1991 paper titled "Face Recognition Using Eigenfaces." Their method, called simply "*Eigenfaces*," was a milestone as it achieved impressive results and demonstrated the capability of simple holistic approaches.

Face images are projected onto a feature space ("face space") that best encodes the variation among known face images. The face space is defined by the "eigenfaces", which are the eigenvectors of the set of faces; they do not necessarily correspond to isolated features such as eyes, ears, and noses [4].

Artificial Intelligence has become a part of each and every human life. Hence it also provides various research fields to perform the research in several aspects using deep learning.

For this a computer system must have to understand the human emotions verbal and nonverbal using various sensors.These sensors may use facial expression changes, tone of voice and physiological signals. On the basis of these emotions one can recognize the behaviour of a human being. Hence it is needed to design and develop a system that can measure the emotional state of a person based on any of these emotions like gestures (body movements and postures), facial expression, auditory characteristics and emotions expressed in the text. If the sensors are fed with predetermined emotion then the body signals and facial expressions may be recorded pragmatically in real-time.The ultimate goal may be to improve the system's decisions so that they can react accordingly to recognized emotions, which will allow better human-machine interaction. For several applications the facial animation systems has become more stunning in the information technology era because face-to-face communication is the most natural way of human interaction.

Basically two methods are used in different way for recognizing face expression,**PCA and CNN Architecture.**

In 1991, Turk and Pentland [5] suggested an approach to face recognition that uses dimensionality reduction and linear algebra concepts to recognize faces which is PCA (Principal Component Analysis) , a dimensionality reduction technique that was proposed by Pearson in 1901. It uses Eigenvalues and EigenVectors to reduce dimensionality and project a training sample/data on small feature space.

CNN Architecture: A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field(see fig.2). A collection of such fields overlap to cover the entire visual area [6].

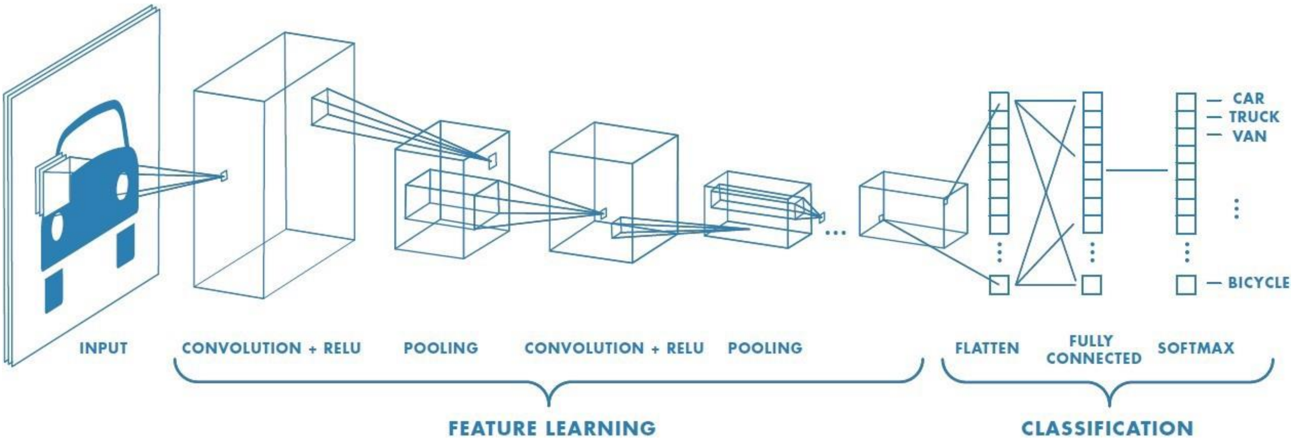

Fig.2 CNN Architecture

Hence our main objective is to identify the behaviour of a human using already existing methods in deep learning whether he is in a classroom, organization or market place etc.So that we may come to know about the attention of a human to talking ,learning, working among many humans which may identify his/her behaviour.

Chapter 2

Literature Review

Several traditional methods exist are used for the extraction facial features such as geometric and texture features for example local binary patterns LBP, facial action units FAC,local directional patterns LDA, Gabor wavelet . In recent years, deep learning has been very successful and efficient approach thanks to the result obtained with its architectures which allow the automatic extraction of features and classification such as the convolutional neural network CNN and the recurrent neural network RNN; here what prompted researchers to start using this technique to recognize human emotions. Several efforts are made by researchers on the development of deep neural network architectures, which produce very satisfactory results in this area.

One of the success factors of deep learning is the training the neuron network with examples, several FER databases now available to researchers to accomplish this task, each one different from the others in term of the number and size of images and videos, variations of the illumination, population and face pose.

Albert Mehrabian in 1967 [7] conducted a study by initiating a “3V rule” which is known as 7%-38%-55% rule.According to the rule 7% of the communication is verbal,38 % is vocal and 55% is visual.Through this study it is justified that nonverbal communication has very interesting and important role.

Taking out the features from one face to another is a difficult and sensitive task in order to have a better classification. In 1978 Ekman and Freisen [8] had developed FACS (Facial Action Coding System) in which facial movements are described by Action Units AUs.They broke down the human face into 46 AUs action units each AU is coded with one or more facial muscles.

The automatic FER is the most studied by researchers compared to other technologies to statistics which made by Philipp et al. **[9]**, but this task is not very easy because each person presents his emotion by his way. For finding the exact behave on the basis of facial expression we a system may face several obstacles and challenges which should not be neglected easily like the variation of head pose, luminosity, age, gender and the background, as well as the problem of occlusion caused by Sunglasses, scarf, skin illness…etc.

C.Shan et al. **[10]**.in 2009 had used effective and efficient method Local Binary Pattern(LBP), statistical feature for recognizing facial expression which was person independent.They formulated Boosted-LBP to draw out the most distinguish LBP features and obtained the best performance through Support Vector Machine classifiers with Boosted-LBP.They also investigated LBP features for low-resolution facial expression recognition, which is a critical problem but not seen often in the existing work.

Deepak Jain et al. **[11]** proposed a novel deep CNN model containing two residual blocks, each having four- convolution layer.The model was trained on JAFFE and CK+ databases after a pre-processing step, which allowed cropping and normalizing the intensity of the images.

Mollahosseini et al. **[12]** proposed a deep CNN for FER across several available databases. After extracting the facial landmarks from the data, the images reduced to 48x 48 pixels. Then, they applied the augmentation data technique. The architecture used consist of two convolution-pooling layers, then add two inception styles modules, which contains convolutional layers size 1x1, 3x3 and 5x5. They showed the ability to use technique the network-in-network, which allow increasing local performance due to the convolution layers applied locally, and this technique also made it possible to reduce the over-fitting problem.

Lopes et al. **[7]** studied the impact of each image pre-processing operation in order to have a better emotion classification. They followed the steps as Data augmentation, rotation correction, cropping, down sampling with 32x32 pixels and intensity normalization before CNN, which consist of two convolution-pooling layers ending with two fully connected with 256 and 7 neurons. The best weight obtained at the training stage used at the test stage. This experience was evaluated in three accessible databases: CK+, JAFFE, BU-3DFE. Researchers shows that combining all of these pre-processing steps is more effective than applying them separately.

.

Mohammadpour et al. **[13]** implemented pre-processing techniques. They proposed a novel CNN for recognizing AUs of the face. For the network, they used two convolution layers, each followed by a max pooling and ending with two fully connected layers that indicate the numbers of AUs activated.

In 2018, for the disappearance or explosion gradient problem Cai et al. **[14]**proposed a novel architecture CNN with Sparse Batch normalization SBP. The property of this network is to use two convolution layers successive at the beginning, followed by max pooling then SBP, and to reduce the over-fitting problem, the dropout applied in the middle of three fully connected layers. For the facial occlusion problem Li et al. **[15]** present a new method of CNN, firstly the data introduced into VGGNet network, then they apply the technique of CNN with attention mechanism ACNN. This architecture trained and tested in three large databases FED-RO, RAF-DB and AffectNet.

Detection of the essential parts of the face was proposed by Yolcu et al. **[16]**.They used three CNN with same architecture each one detect a part of the face such as eyebrow, eye and mouth. Before introducing the images into CNN, they go through the crop stage and the detection of key-point facial. The iconic face obtained combined with the raw image was introduced into second type of CNN to detect facial expression. Researchers show that this method offers better accuracy than the use of raw images or iconize face alone (See Fig.3).

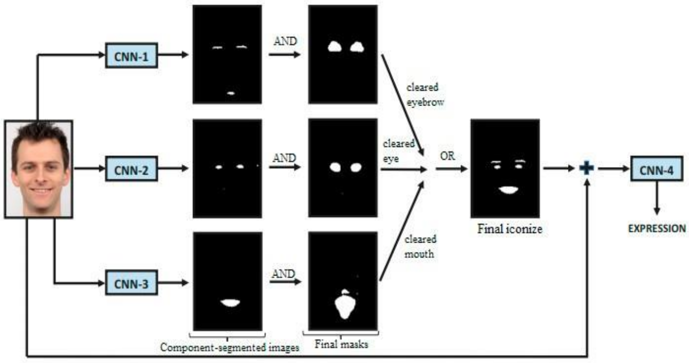

Fig. 3. Different deep learning methods proposed by Yoclu et al. [16]

Agrawal et al. **[18]** in 2019 studied the FER2013 database for recognizing the rate of influence variation of the CNN parameters.Firstly they defined at 64x64 pixels,then changed the size and number with their type of optimizer chosen(adam,SGD,adadelta) on a simple CNN containing two convolution layers.The first layer perform the max pooling and second layer performed classification using softmax function.Through their study,the researchers established two models of CNN having 65.23% and 65.77% of accuracy.The main characteristics of these model was that they do not contain fully connected layers dropout and the similar filter size is used.

Kim et al. **[19]** proposed a spatio-temporal architect with a combination of CNN and LSTM after studying variation in facial expression during emotional state. At first time, CNN learn the spatial features of the facial expression in all the frames of the emotional state followed by an LSTM applied to preserve the whole sequence of these spatial features.

Yu et al. **[20]** introduced a novel architecture called Spatio-Temporal Convolutional with Nested LSTM (STC-NLSTM), this architecture based on three deep learning sub network such as: 3DCNN for extraction spatio- temporal features followed by temporal T-LSTM to preserve the temporal dynamic, then the convolutional C-LSTM for modelled the multi-level features.

Liang et al. **[21]**, proposed a Deep convolutional BiLSTM architecture.They established two DCNN, in which the first one was designated for spatial features and the other for extracting temporal features in facial expression sequences.These features were put together at level on a vector with 256 dimensions, and for the classification into one of the six basic emotions, researchers used BiLTSM network.They used the Multitask cascade convolutional network for detecting the face at the pre-processing stage, then they used the technique of data augmentation to broaden database (See Fig.4).

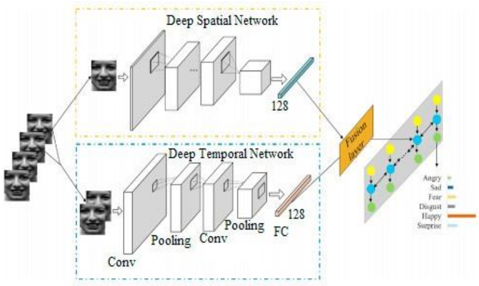

Fig.4 Different deep learning methods proposed by Liang et al. [17]

Lang He et al.,2020 [22] have proposed a deep local global attention convolutional neural network (DLGA CNN)model for recognising depression with attention mechanism from videos.They collected the depression databases AVEC2013 and AVEC2014. This proposed model adopts CNN with attention mechanism as well as weighted spatial pyramid pooling (WSPP) to learn deep and global re presentation .They introduced local attention based (LA-CNN) and global attention based (GA-CNN) in combination for capturing the complementary information .With the help of a novel framework named DLGA CNN many comprehensive experiments were performed on depression databases.

Shiquin et al.,2021 [23] have studied the tourist emotion relying on self-reports instead of explicit expressed emotion.Their main aim was to compare residents'emotional responses towards tourist facial expression and self reports .Then they interpreted identified discrepancies by investigating the psychological mechanism behind the two expression modes. They found that facial expression conveyed more desires derived emotions like happiness sadness and anger while self reports stressed on stereotype elicited emotions partially disgust.For this they proposed dual process model to interpret the emotional expressed discrepances.Hence they enhanced the theorization of tourism studies on motion. Though the identification of the discrepancies they demonstrated that facial expressions and words can tell different stories.

Rebecca Shankland et al.,2020 [24],analysed how mindfulness practice may reduce the use of prior knowledge during the recognition of emotional facial expressions .On the basis of predictive brain model they hypothesized that mindfulness practice would shorten the top down processing of low spatial frequency information.They performed the experiment for comparing the performance of a mindfulness group (n=32)and a waitlist control group(n=30) in an emotional stroop task before and after an 8 week training course.The result of this experiment helped in learning the effect of mindfulness based interventions on global attentional control.

A modern critical survey was provided by W.Zhao et al. [25] on still and video-based face recognition research.They categorized already existing recognition techniques with detailed descriptions of each method. They also presented some psychophysical studies, system evaluation, and issues of illumination and pose variation.

Since the face detection is the first step in face recognition technique, hence ErikHjelmås et al. [26] had presented an exhaustive and condemnatory survey of face detection algorithms.They classified the algorithms on either feature-based or image-based and are discussed in terms of their technical approach and performance. They also presented some proposed applications and possible application areas.

R.Chellappa et al. [27] explained the various applications of face recognition technique in commercial and law enforcement sectors through psychophysics community. They gave a review on the techniques for segmentation/location of the face, feature extraction and recognition and also summarized Global transform and feature based methods using statistical, structural and neural classifiers.

For the detection and identification of a persons's face ,M.A.Turk et al. [3] used an approach by treating a 2D recognition problem and then described it by a small set of 2-D charcteristics views.They projected the face images onto a face space that best encodes the variation among known face images. The face space is defined by the 'eigenfaces', which are the eigenvectors of the set of faces. Through this framework it became possible to learn to recognize new faces in a separate manner.

Feng-Ju Chang et al. [28] had described a deep learning based method for estimating 3D facial expression coefficients(see fig.5). The modern methods of CNN were able to estimate shapes for occluded faces appearing in extraordinary in-the-wild viewing conditions. They retrained those methods by showing that facial expressions can also be estimated by a robust, deep, landmark-free approach. They applied ExpNet CNN to the intensities of a face image and regresses a 29D vector of 3D expression coefficients.They showed that the ExpNet produces expression coefficients which better differentiate between facial emotions than those obtained using state of the art, facial landmark detection techniques.

Fig. 5: Deep 3D face modeling with expressions. [28]

Annemarie J. Nanne et al. [29] followed a two-step approach to examine sender presence effects in various levels of tie strength. They found that a happy facial expression significantly increases like intention and brand attitude.

D Y Liliana [30] performed a task of detecting the occurrence of facial Action Units (AUs) as a subpart of Facial Action Coding System (FACS) which represents human emotion. He employed a regularization method called “dropout” in the CNN fully-connected layers to reduce overfitting using the extended Cohn Kanade (CK+) dataset .The system performance gained average accuracy rate of 92.81%. The system has been successfully classified eight basic emotion classes. The result was the mean square error declining as the number of training data increasing. From the experiment it was concluded that the mean square error declines as the training data grows.

Facial available databases

One of the success factors of deep learning is the training the neuron network with examples, several FER databases now available to researchers to accomplish this task, each one different from the others in term of the number and size of images and videos, variations of the illumination, population and face pose. Some presented in the Table.1 in which we will note its presence in the works cited in the following section. **[30]**

Table 1. A summary of some FER databases

Databases	Description
The Color FERET Database, USA	The FERET database was collected in 15 sessions between August 1993 and July 1996. The database contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images.
SCface - Surveillance Cameras Face Database	Database contains 4160 static images (in visible and infrared spectrum) of 130 subjects
SCfaceDB Landmarks	The database is comprised of 21 facial landmarks (from 4160 face images) from 130 users annotated manually by a human operator.
Multi-PIE	It contains 337 subjects, captured under 15 view points and 19 illumination conditions in four recording sessions for a total of more than 750,000 images.
The Yale Face Database	Contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy,

	left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.
The Yale Face Database B	Contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured.
PIE Database, CMU	A database of 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions.
Project - Face In Action (FIA) Face Video Database, AMP, CMU	Three out of the six cameras have smaller focus length, and the other three have larger focus length. Plan to capture 200 subjects in 3 sessions in different time period. For one session, both in-door and out-door scenario will be captured. User-dependent pose and expression variation are expected from the video sequences.
AT&T "The Database of Faces" (formerly "The ORL Database of Faces")	Ten different images of each of 40 distinct subjects.
Cohn-Kanade AU Coded Facial Expression Database	Subjects in the released portion of the Cohn-Kanade AU-Coded Facial Expression Database are 100 university students.
Image Database of Facial Actions and Expressions - Expression Image Database	24 subjects are represented in this database, yielding between about 6 to 18 examples of the 150 different requested actions.
NIST Mugshot Identification Database	There are images of 1573 individuals (cases) 1495 male and 78 female. The database contains both front and side (profile) views when available.
NLPR Face Database	450 face images. 896 x 592 pixels. JPEG format. 27 or so unique people under with different lighting/expressions/backgrounds.
M2VTS Multimodal Face Database (Release 1.00)	Database is made up from 37 different faces and provides 5 shots for each person.

The Extended M2VTS Database, University of Surrey, UK	Contains four recordings of 295 subjects taken over a period of four months.
The AR Face Database, The Ohio State University, USA	4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf).
The University of Oulu Physics-Based Face Database	Contains 125 different faces each in 16 different camera calibration and illumination condition, an additional 16 if the person has glasses. Faces in frontal position captured under Horizon, Incandescent, Fluorescent and Daylight illuminant. Includes 3 spectral reflectance of skin per person measured from both cheeks and forehead. Contains RGB spectral response of camera used and spectral power distribution of illuminants.
Japanese Female Facial Expression (JAFFE) Database	The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.
BioID Face DB - HumanScan AG, Switzerland	The dataset consists of 1521 gray level images with a resolution of 384x286 pixel. Each one shows the frontal view of a face of one out of 23 different test persons.
Psychological Image Collection at Stirling (PICS)	This is a collection of images useful for research in Psychology, such as sets of faces and objects. The images in the database are organised into SETS, with each set often representing a separate experimental study.
The Sheffield Face Database (previously: The UMIST Face Database)	Consists of 564 images of 20 people. Each covering a range of poses from profile to frontal views. Subjects cover a range of race/sex/appearance. Each subject exists in their own directory labelled 1a, 1b, ...
Face Video Database of the Max Planck Institute for Biological Cybernetics	This database contains short video sequences of facial Action Units recorded simultaneously from six different viewpoints, recorded in 2003 at the Max Planck Institute for Biological Cybernetics.
Caltech Faces	450 face images. 896 x 592 pixels. JPEG format. 27 or so unique people under with different lighting/expressions/backgrounds.

EQUINOX HID Face Database	Equinox is collecting an extensive database of face imagery in the following modalities: coregistered broadband-visible/LWIR (8-12 microns), MWIR (3-5 microns), SWIR (0.9-1.7 microns).
VALID Database	The database consists of five recording sessions of 106 subjects over a period of one month. One session is recorded in a studio with controlled lighting and no background noise, the other 4 sessions are recorded in office type scenarios
Georgia Tech Face Database	The database contains images of 50 people and is stored in JPEG format.
Indian Face Database	The database contains a set of face images taken in February, 2002 in the IIT Kanpur campus. There are eleven different images of each of 40 distinct subjects. The following orientations of the face are included: looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, looking down. Available emotions are: neutral, smile, laughter, sad/disgust.
YMU (YouTube Makeup) Dataset	The dataset consists of 151 subjects, specifically Caucasian females, from YouTube makeup tutorials. Images of the subjects before and after the application of makeup were captured.
VMU (Virtual Makeup) Dataset	The VMU dataset was assembled by synthetically adding makeup to 51 female Caucasian subjects in the FRGC dataset.
MIW (Makeup in the "wild") Dataset	The MIW dataset contains 125 subjects with 1-2 images per subject. Total number of images is 154 (77 with makeup and 77 without makeup). The images are obtained from the internet and the faces are unconstrained.
3D Mask Attack Database (3DMAD)	The data is collected in 3 different sessions for all subjects and for each session 5 videos of 300 frames are captured.
MIT-CBCL Face Recognition Database	The MIT-CBCL face recognition database contains face images of 10 subjects.
McGill Real-world Face Video Database	This database contains 18000 video frames of 640x480 resolution from 60 video sequences, each of which recorded from a different subject (31 female and 29 male).
SiblingsDB Database	The SiblingsDB contains two different datasets depicting images of individuals related by sibling relationships. The first, called HQfaces,

	contains a set of high quality images depicting 184 individuals (92 pairs of siblings).
FaceScrub - A Dataset With Over 100,000 Face Images of 530 People	It comprises a total of 107,818 face images of 530 celebrities, with about 200 images per person. As such, it is one of the largest public face databases.
Indian Movie Face database (IMFDB)	Indian Movie Face database (IMFDB) is a large unconstrained face database consisting of 34512 images of 100 Indian actors collected from more than 100 videos.
10k US Adult Faces Database	It is a database of 10,168 natural face photographs of all different individuals, and major celebrities removed.
Denver Intensity of Spontaneous Facial Action (DISFA) Database	This database contains stereo videos of 27 adult subjects (12 females and 15 males) with different ethnicities. The database also includes 66 facial landmark points of each image in the database.
BU-3DFE Database (Static Data)	BU-3DFE (Binghamton University 3D Facial Expression) includes 100 subjects with 2,500 facial expression models.
BP4D-Spontanous Database	The database includes 41 participants (23 women, 18 men). Each participant is associated with 8 tasks. For each task, there are both 3D and 2D videos. As well, the metadata include manually annotated action units (FACS AU), automatically tracked head pose and 2D/3D facial landmarks. The database is in the size of about 2.6 TB (without compression).
DMCSv1 Database - Multimodal Biometric Database of 3D Face and Hand Scans	The database contains 3D face and hand scans. It was acquired using the structured light technology.
CAFE - The Child Affective Face Set	The set is made up of 1200 photographs of over 100 child models (ages 2-8) making 7 different facial expressions - happy, angry, sad, fearful, surprise, neutral, and disgust.
UFI - Unconstrained Facial Images	Two different partitions of the database are available. The first one contains the cropped faces that were automatically extracted from the photographs using the Viola-Jones algorithm. The images in the second partition have more background, the face size also significantly differs and the faces are not localized.

Senthil IRTT Face Database Version 1.1	This database contains IRTT (Institute of Road and Transport Technology) students of both colour and gray scale facial images. There are 317 facial images for 13 IRTT students. Each subject have variety of face expressions, little makeup, scarf, poses and hat also.
Senthil IRTT Face Database Version1.2	There are 100 facial images for 10 IRTT girl students (all are female) with 10 faces per subject with age factor around 23 to 24 years.
Senthil IRTT Video Face Database 1.0	This IRTT student video database contains one video in .mp4 format. The video duration is 55.938 seconds and contains 30 frames with resolution of 720x1280.
VT-AAST Bench-marking Dataset	Part one is a set of 286 color photographs that include a total of 1027 faces
SEAS-FR-DB (School of Engineering & Applied Science - Face Video Database)	The database is a useful input for offline as well as online (Real-Time) Video scenarios. he database has primarily three classified subjects (3) and two not-classified (unknown/un-labeled) subjects available in 30 fps - High Definition Video (Full HD - 1080p) video developed at School of Engineering & Applied Science, Ahmedabad University in 2016.
Facial Expression Research Group Database (FERG-DB)	The database contains facial expression images of six stylized characters. The images for each character is grouped into seven types of expressions - anger, disgust, fear, joy, neutral, sadness and surprise.
Specs on Faces (SoF) Dataset	The SoF dataset is a collection of 42,592 (2,662×16) images for 112 persons (66 males and 46 females) who wear glasses under different illumination conditions.
Large Age-Gap Database (LAG)	Large Age-Gap (LAG) dataset is a dataset containing variations of age in the wild, with images ranging from child/young to adult/old. The dataset contains 3,828 images of 1,010 celebrities. For each identity at least one child/young image and one adult/old image are present.
Sejong Face Database: A Multi-modal disguised face database (SFD)	The database contains a total of 24,500 facial images.
CyberExtruder Ultimate Face Matching Data Set	The CyberExtruder Ultimate Face Matching Data Set contains 10,205 images of 1000 people scraped from the internet.

The Makeup Induced Face Spoofing (MIFS) dataset	The Makeup Induced Face Spoofing (MIFS) dataset consists of 107 makeup-transformations taken from random YouTube makeup video tutorials.
RAVDESS: Ryerson Audio-Visual Database of Emotional Speech and Song	The database is gender balanced consisting of 24 professional actors, vocalizing lexically-matched statements in a neutral North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and song contains calm, happy, sad, angry, and fearful emotions. Each expression is produced at two levels of emotional intensity, with an additional neutral expression.
Disguised Faces in the Wild	The Disguised Faces in the Wild (DFW) dataset has been prepared in order to address these limitations. The proposed DFW dataset consists of 11,157 images of 1,000 subjects. The dataset contains a broad set of unconstrained disguised faces, taken from the Internet.
BAUM-1: Bahcesehir University Multimodal Face Database of Spontaneous Affective and Mental States	BAUM-1 is a spontaneous audio-visual affective face database of affective and mental states.
NMAPS - NMAmit Photo Sketch database	Database contains 208 South Indian images gathered at the computer science research lab of NMAM Institute of Technology.
Grammatical Facial Expressions Data Set	The dataset is organized in 36 files: 18 datapoint files and 18 target files, one pair for each video which compose the dataset.The name of the file refers to each video: the letter corresponding to the user (A and B), name of grammatical facial expression and a specification (target or datapoints).
Indian Semi Acted Facial Expression Database (iSAFE)	The dataset contains 395 clips of 44 volunteers between 17 to 22 year of age. All the clips are manually splitted from the video recorded during stimulent clips are watched by volunteers.
VIP_attribute Dataset	Images in the VIP_attribute dataset are obtained in 2017 from the WWW corresponding to 513 female and 513 male subjects (mainly actors, singers and athletes). The images include the frontal pose of the subjects.

EURECOM Visible and Thermal paired Face database	The database was collected from 50 subjects of different age, sex and ethnicity, resulting a total of 2100 images.

Chapter 3

Conclusion

The improvements in Artificial Intelligence have been helping the visually challenged in numerous ways by creating bit by bit improvement in various pieces of life of the people who are studying,teaching or working in an organization etc. To help such communities and bring the world closer together, we must eliminate the obstacles for preventing these communities from participating in day-to-day social interactions. Hence Behaviour Recognition need to be performed. Several techniques like FER, CNN, PCA etc. are analyzed rigorously. For this task we use primary as well as secondary database to perform the task of facial recognition which will help in identifying the behaviour of a human being.If we come to know the behave of a human being then we may alert at early age what would be the behave of a human and then how we tackle that human.

References

[1] D. Y. Liliana, " Emotion recognition from facial expression using deep convolutional neural network," in *International Conference of Computer and Informatics Engineering (IC2IE)*.

[2] A. F. ,. M. ,. M. W. D. Z. Mary E. Webb, "Machine learning for human learners: opportunities, issues,tensions and threats," *Education Tech Research Dev,* 2020.

[3] A. K. J. Stan Z. Li, Handbook of Face Recognition, 2nd ed., A. K. Jain, Ed., 2011.

[4] J. Brownlee, "Gentle Introduction to Deep Learning for Face Recognition," *Deep Learning for Computer Vision,* 31 May 2019 .

[5] A. P. Matthew Turk, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.

[6] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," *towards Data Science,* 15 December 2018.

[7] E. d. A. A. F. D. S. e. T. O.-S. “. e. r. w. C. N. N. C. w. f. d. a. t. t. s. o. T. Lopes, *Pattern Recognition,* vol. 61, pp. 610-628, January 2017.

[8] D. M. A. H. B. e. T. S. M. H. Alkawaz, "Blend Shape Interpolation and FACS for Realistic Avatar," *3D Res.,,* vol. 6, no. 1, p. 6, Januaery 2015.

[9] M. A. e. R. C. P. V. Rouast, *Deep Learning for Human Affect Recognition: Insights and New Developments*, 2018, p. 1-1.

[10] S. G. e. P. W. M. C. Shan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image Vis. Comput.,* vol. 27, no. 6, pp. 803-816, May 2009.

[11] P. S. e. P. S. D. K. Jain, "Extended deep neural network for facial emotion recognition," *Pattern Recognit. Lett.,* vol. 120, pp. 69-74, April 2019.

[12] D. C. e. M. H. M. Mollahosseini, "Going deeper in facial expression recognition using deep neural networks," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[13] H. K. S. M. R. H. e. M. M. A. M. Mohammadpour, "Facial emotion recognition using deep convolutional networks," in *IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI),,* 2017.

[14] O. C. X. T. C. X. e. C. W. J. Cai, "Facial Expression Recognition Method Based on Sparse Batch Normalization CNN," in *37th Chinese Control Conference (CCC)*, 2018.

[15] J. Z. S. S. e. X. C. “. v. 2. n. 5. p. 2. m. 2. Y. Li, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," *IEEE Trans. Image Process,* vol. 28, no. 5, pp. 2439-2459, 2019.

[16] G. Yolcu et al., "Facial expression recognition for monitoring neurological disorders based on convolutional neural network," *Multimed. Tools Appl,* vol. 78, no. 22, pp. 31581-31603, 2019.

[17] H. L. Z. Y. e. Y. Z. “. c. B. f. n. f. f. e. r. V. C. ,. [12] D. Liang, vol. 36, no. 3, pp. 499-508, March 2020.

[18] A. e. N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *Vision Computing,* January 2019.

[19] W. J. B. J. J. e. Y. M. R. D. H. Kim, "Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition," *IEEE Trans. Affect. Comput.,* vol. 10, no. 2, pp. 223-23.

[20] G. L. Q. L. e. J. D. Z. Yu, "Spatio-temporal convolutional features with nested LSTM for facial expression recognition," *Neurocomputing,* vol. 317, pp. 50-57, 2018.

[21] H. L. Z. Y. e. Y. Z. D. Liang, "Deep convolutional BiLSTM fusion network for facial expression recognition," *Vision Computing,* vol. 36, pp. 499-508, March 2020.

[22] J. C.-W. C. ,. Z. W. Lang He, "Automatic depression recognition using CNN with attention mechanism from videos," *Neurocomputing* , Vols. 422,, pp. 165-175, January 2021.

[23] N. C. C. H. H. Shiqin Zhang, "Facial expressions versus words: Unlocking complex emotional responses of residents toward tourists". *“Tourism Management 83 (2021) 104226.*

[24] R. S. &. P. F. &. I. K. &. M. Mermillod, "Mindfulness and De-automatization: Effect of Mindfulness-Based Interventions on Emotional Facial Expressions Processing”,," September 2020.

[25] R. C. J. P. W. Zhao, "Face recognition: A literature survey," *ACM Computing Surveys,* vol. 35, no. 4, pp. 399-458, December 2003.

[26] B. K. ErikHjelmås, "Face Detection: A Survey," *Computer Vision and Image Understanding,* vol. 83, no. 3, pp. 236-274, September 2001.

[27] W. C. S. R.Chellappa, "Human and machine recognition of faces: a survey," vol. 83, no. 5, pp. 705-741, May 1995.

[28] A. T. T. H. M. ,. N. ,. M. Feng-Ju Chang, "ExpNet:Landmark-Free, Deep, 3D Facial Expressions," in *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018.

[29] M. L. A. ,. G. v. N. Annemarie J. Nanne, "The role of facial expression and tie strength in sender presence effects on consumers' brand responses towards visual brand-related user generated content," *Computers in Human Behaviour,* vol. 117, 2021.

[30] D. Y. Liliana, " " Emotion recognition from facial expression using deep convolutional neural network," in *International Conference of Computer and Informatics Engineering (IC2IE).*