IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Literature Survey On Live Emotion Recognition System Using CNN

Suyog Udayshankar Gadage
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, India

Vishal Dilip Parandwal
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, India

Suyash Bajirao Gorde
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, India

Rohit Vasant Surve
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, India

Prof. A. S. Kamble
Department of Computer Engineering
Modern Education Society's College of Engineering
Pune, India

***Abstract*—Recently, the Facial Expression Recognition (FER) may be a vital side like in Security, Human-Computer Interaction, Lie Detection, etc. It's a system that facilitates to seek out human emotions (such as- Happiness, Anger, Sadness, Fear, Disgust, Surprise and Neutral). Moreover, Facial Expression Recognition (FER) is employed to 'Monitoring the Pilot Alertness' or 'Monitoring ATM user through camera fitted on ATM Machines'. Live Facial Expression Recognition may be a difficult drawback in Image Classification. Now a days the use of Deep Learning is achieving importance in Image Classification. This has led to increased efforts in solving the problem of Facial Expression Recognition using Convolutional Neural Networks (CNNs). A big challenge in Deep Learning is to style a specification that's straightforward and effective. An easy design is expeditious to train and very much simple to implement. A good effective architecture achieves good accuracy on the test data. In this paper, various datasets and techniques are used in recent papers are discussed for face expression Recognition.**

Keywords— Facial Expression Recognition, Deep Learning, Convolutional Neural Networks (CNNs), Image Classification, Human-Computer Interaction.

I. Introduction

The identification system has undergone great development since its invention. Related techniques are widely utilized and have many applications, such as Humanoid Robot [5], Emotion Detection on Physiological Signals [10]. Face plays an important role in social communication, this is a 'window' to human personality, emotions, and thoughts. According to the psychological research nonverbal part is the most informative channel in social communication. The verbal part contributes about 7% of the message, vocal – 34% and facial expression about 55% to the effect of the speaker's message. The variety of information in these face images makes face detection tough. For example, some of the conditions that should be accounted for, when detecting faces is:

Occlusion: faces may be partially occluded by other objects [9].

Recently, the convolutional neural network (CNN) has become the most popular deep learning architecture used to solve the FER problem. A Convolutional Neural Network (CNN) is a Deep Learning algorithm that 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 needed in an exceedingly CNN is far lower as compared to other classification algorithms. The convolutional neural network (CNN) architecture, the addition of a recursive method can enhance the features of an image, and it permits the CNN to automatically extract the image features to make the features more obvious.

II. Recognition System

A. Face Detection

Face Detection is the first and essential step for face recognition, and it is used to detect faces in the images. It is a part of object detection and can use in many areas such as security, biometrics, law enforcement, entertainment, personal safety, etc. It is used to detect faces in real-time for surveillance and trailing of persons or objects. It's wide utilized in cameras to spot multiple appearances with the frame Mobile cameras and DSLR's.

B. The CNN Architecture

In deep_learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analyzing visual imaging. 2012 was the first year that neural network grew to prominence as Alex Krizhevsky used them to win that year's ImageNet competition, dropping the classification error record from 26% to 15%, an outstanding improvement at that time.

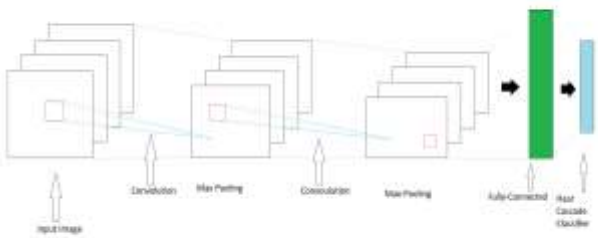
A convolutional neural network (CNN) contains several layers:

1. Convolutional: This layer accommodates a rectangular grid of neurons. It requires that the previous layer also be a

rectangular grid of neurons. Each neuron takes inputs from a rectangular section of the previous layer; the weights for this rectangular section are similar for each neuron in the convolutional layer. Thus, the convolutional layer is just an image convolution of the previous layer, wherever the weights specify the convolution filter.

2. Max-Pooling: After each convolutional layer, there may be a pooling layer. The pooling layer takes small rectangular blocks from the convolutional layer and sub-samples it to produce a single output from that block.

3. Fully-Connected: Finally, after several convolutional and max-pooling layers, the high-level reasoning in the neural network is done via fully connected layers. A fully connected layer takes all neurons within the previous layer.



III. LITERATURE SURVEY

The main aim of this research work is to classify the various facial expressions of the human face. A survey on various existing research works of facial expression recognition is reviewed and discussed.

A. Real-Time Facial Expression Recognition Based on CNN

This technique is purposed by Keng-Cheng Liu, Chen-Chien Hsu, Wei-Yen Wang, and Hsin-Han Chiang [1]

In this paper, the proposed method is for improving the robustness of real-time facial expression recognition. One major problem is that when the camera is capturing the real-time images (these images are in high speed), changes in image characteristics may occur at certain moments due to light and other factors. This can lead to the incorrect recognition of human facial expressions. To avoid errors in real-time facial expression recognition, an 'Average Weighting Method' is proposed in this paper, with traditional convolutional neural networks (CNN). Moreover, because of the average weighting method results become more robust. By using traditional convolutional neural network (CNN) the current recognition result and the previous recognition result are considered both by the average weighting method. Each image has 7 weights indicating different expressions, which are stored and averaged with the previous image (from the dataset). The largest output is taken as the predicted expression. As a result, the accuracy of facial expression recognition is improved. Experimental results have shown that the proposed facial expression recognition system is more reliable than the traditional CNN approach.

The average weighting method, avoid potential errors and reduced influence of noise from the environment. But the computational cost is high.

B. Facial expression recognition with FRR-CNN

This technique is purposed by Siyue Xie and Haifeng Hu [2]

The method proposed in this paper is used to reduce the redundancy of the convolutional neural network (CNN). Different from traditional CNN, convolutional kernels of Feature redundancy-reduced CNN (FRR-CNN) is induced to be divergent by presenting a more discriminative mutual difference among feature maps of the same layer, which results in generating less redundant features and yields a more compact representation of an image. By using TI-pooling, identity can be seen as some transformation. A transformation set can thus be constructed by images from the same class of expression. To obtain a new feature vector independent from the known variations and match the strategy of TI-pooling. The siamese network is introduced with multiple channels to enhance the power of generating discriminative features. In our model, images from the same transformation set are set as the inputs of identical parallel channels. Then, each image passes through the convolutional module and yields a concatenate feature vector. Weight sharing is implemented among all the channels. TI-pooling is implemented across all these channels. Which can be formulated as the following expression:

$$g_{TI}^{n}(x) = max\{f_{1}^{n}(x), f_{2}^{n}(x), \dots, f_{k}^{n}(x), \dots\} \quad (1)$$

The databases are used: CK+ (Cohn-Kanade) and JAFFE (Japanese Female Facial Expression). CK+ database is a public expression database, which consists of 593 sequences. As for JAFFE, it consists of 213 expressional images of ten Japanese female subjects. As implementing FRR-CNN the accuracy becomes 92.06% for CK+ database.

The main advantage is that the siamese architecture is combined with CNN to handle Ti-pooling for obtaining a more descriptive representation of the image. But the framework of FRR-CNN is more complex.

C. Recognizing Facial Expressions Using a Shallow Convolutional Neural Network

This technique is purposed by Si Miao, Haoyu Xu, Zhenqi Han, and Yongxin Zhu [3]

Generally, facial expressions could be classified into two categories: static facial expressions and micro-expressions. Unlike other techniques, this paper proposed to classify facial expressions having both categories static facial expressions and micro-expressions. The proposed method is tested on five public datasets: FER2013, FERPlus, CASME, CASME II, and SAMM. The details of datasets are listed below:

FER2013: The Fer2013 dataset consists of 35887 grayscale face images of size 48×48 used for Kaggle challenge. Each image is labelled with one of the seven kinds of expressions (angry, disgust, fear, happy, sad, surprise and neutral).

FER-Plus: To solve the noisy label problem in Fer2013, tagged the images again and used probability distribution instead of a unique tag to determine the category of each image.

CASME: There are eight kinds of expressions in Casme: tense, happiness, repression, surprise, disgust, fear, contempt and sadness. There are 19 subjects, 189 videos in the dataset. The distribution is: tense (69 videos), happiness (9 videos), repression (38 videos), surprise (20 videos), disgust (44 videos), fear (2 videos), contempt (1 video), sadness (6 videos).

CASME II: Seven kinds of expressions are included in CASME II, they are happiness, others, disgust, repression, surprise, fear, and sadness. There are 26 subjects, 255 videos and 16781 frames in the dataset. The distribution is happiness (32 videos, 2319 frames), others (99 videos, 6336 frames), disgust (63 videos, 4153 frames), repression (27 videos, 2150 frames), surprise (25 videos, 1514 frames), fear (2 videos, 66 frames), sadness (7 videos, 243 frames).

SAMM: Compared with CASME and CASME II which only consist of Chinese subjects, SAMM is a novel dataset whose subjects are selected from a diverse range of age and ethnicity. We classify the SAMM into three categories: Positive (Happiness), Negative (Anger, Fear, Disgust, Contempt) and Surprise. These three categories contain 26, 92, 15 samples respectively, hence SAMM is an unbalanced database.

Accuracy for five public datasets, FER2013- 69.10%, FERPlus- 86.54%, CASME- 63.33%, CASME II- 71.14%, SAMM- 86.47%.

This technique recognizes static expressions as well as micro-expressions. But follows the traditional way of CNN.

D. Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network

This technique is purposed by Huibin Li, Jian Sun, Zongben Xu and Liming Chen [4]

This proposed method is especially for 2D+3D facial expression recognition and using Deep Fusion convolutional neural network (DF-CNN). DF-CNN contains a feature extraction subnet, a feature fusion subnet, and a softmax layer. Especially, each textured 3D face scan is represented as six types of 2D facial attribute maps (i.e., geometry map, three normal maps, curvature map, and texture map), all of which are jointly fed into DF-CNN for feature learning and fusion learning, resulting in a highly concentrated facial representation (32-dimensional). The architecture of DF-CNN is formed by a feature extraction subnet, a feature fusion subnet, and a softmax layer (softmax layer is classifier). The main building blocks of feature extraction subnet include the convolutional layers and ReLU nonlinearity, while the reshape layer and fusion layers are the main components of feature fusion subnet. Databases used for this technique:

1. BU-3DFE: The BU-3DFE (Binghamton University 3D Facial Expression) Database has been the benchmarking for static 3D FER. It includes 100 subjects (56 females and 44 males), with age ranging from 18 to 70 years old, and with a variety of racial ancestries (e.g., White, Black, East-Asian). Each subject has 25 samples of seven expressions: one sample for neutral, and other 24 samples for six prototypical expressions (anger, disgust, fear, happiness, sadness, and surprise), each includes four levels of intensity. As a result, this database consists of 2,500 2D texture images and 2,500 geometric shape models. There are two subsets:
   i. BU-3DFE subset I: This subset is the standard dataset used for 3D FER. It contains 1,200 2D and 3D face pairs (i.e., 7,200 2D facial attribute maps) of 100 subjects with 6 prototypical expressions and two higher levels of expression intensity.

   ii. BU-3DFE subset II: This subset includes all samples of BU-3DFE except the 100 neutral samples. It contains 2,400 2D and 3D face pairs (i.e., 14,400 2D facial attribute maps) of 100 subjects with 6 prototypical expressions of four levels of intensity.

2. Bosphorus 3D Face Database: The Bosphorus 3D Face Database has been widely used for 3D face recognition under adverse conditions, 3D facial action unit detection, 3D facial landmarking, etc. It contains 105 subjects and 4,666 pairs of 3D face models and 2D face images with different action units, facial expressions, poses and occlusions. In this dataset, there are a total of 65 subjects performing the six prototypical expressions with a near frontal view. Each person has only one 2D (or 3D) sample for each expression, resulting in 390 2D and 3D face pairs.

Accuracy for BU-3DFE database: BU-3DFE subset I- 86.86%, BU-3DFE subset II- 81.33%.

The main advantage is that this technique can work on 2D and 3D faces. But due to Deep fusion speed is comparatively slow.

E. CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot

This technique is purposed by Tzuu-Hseng S. Li, Ping-Huan Kuo, Ting-Nan Tsai, and Po-Chien Luan [5]

For the human-robot interaction robots must be able to recognize human emotions, this method is proposed especially for a humanoid robot. The robot is equipped with a camera to capture user facial images, and it uses this system to recognize users emotions and responds appropriately. First, a convolutional neural network (CNN) is used to extract visual features by learning on a large number of static images. Second, a long short-term memory (LSTM) recurrent neural network is used to determine the relationship between the transformation of facial expressions in image sequences and the seven basic emotions. Third CNN and LSTM are combined to use their advantages on the proposed model.

CNN is used to capture appearance feature because the inputs are cropped region of interest of the image, which is also in the region of the detected face. It is difficult for LSTMs to learn such high-dimensional data. Therefore, the solution to this problem is to take advantages of CNN because it can sub-instance a high-dimensional image without losing important information. We use a few previous layers of the CNN as the feature extractor.

The databases are used: AFFECTNET and CK+ (Cohn-Kanade). The AffectNet Database contains approximately 450,000 manually annotated and approximately 500,000 automatically annotated colour images of various sizes. They are labelled as neutral, happy, sad, surprise, fear, disgust, anger, and contempt. However, the dataset is highly imbalanced; there are more than 100,000 images with happy expressions but less than 5,000 images with disgust expressions.

The Cohn-Kanade (CK+) dataset includes 123 subjects and 593 image sequences. Image sequences are captured using a camera located in front of the subject. All images are 640 × 480 grayscale images. The subjects are mostly Euro-American females. Among the 593 image sequences, 327 are labelled with seven expressions: happy, sad, surprise, fear, disgust, anger, and contempt. The sequences have different lengths and must contain at least 10 images.

Accuracy for databases: CK+- 90.51%.

As accuracy specifies that this method has an advantage over different techniques because of the combination of CNN and LSTM. Fusion makes this method complex and troublesome to implement.

F. Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images

This technique is purposed by Biao Yang, Jinmeng Cao, Rongrong Ni, and Yuyu Zhang [6]

A weighted mixture deep neural network (WMDNN) is proposed in this paper to automatically extract the features that are essential for facial expression recognition (FER) tasks. Pre-processing approaches that are used to restrict the region for FER are face detection, rotation rectification, and data augmentation. There are mainly two facial channels: facial grayscale images and local binary pattern (LBP) facial images are proposed by WMDNN. Facial grayscale images are extracted by partial VGG16 network while local binary pattern (LBP) are extracted by a shallow neural network [3]. Then at the end, the facial images are classified into one of the seven emotions (happy, sad, anger, disgust, scared, surprise, and neutral). So, this is the concept of WMDNN double-channel facial images.

The databases are used: CK+ (Cohn-Kanade), JAFFE (Japanese Female Facial Expression), and Oulu-CASIA.

1. CK+: This fully annotated dataset includes 593 sequences that represent seven expressions (happiness, sadness, surprise, disgust, fear, anger, and neutral) of 123 subjects (males and females). All images are 640 × 480 grayscale images.

2. JAFFE: This fully annotated dataset includes 213 samples of 10 Japanese females. The dataset also contains the six basic expressions and a neutral expression.

3. Oulu-CASIA: A total of 10,800 labelled samples are captured from 80 subjects (a mix of male/female and glasses/without glasses).

Accuracy for databases: CK+- 97.02%, JAFFE: 92.21%, Oulu-CASIA: 92.89%.

The main advantage of this method is unlike other FER systems this uses both facial images: Greyscale and LBP.
But due to the double-channel concept computational cost is high.

G. A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing

This technique is purposed by Hongli Zhang, Alireza Jolfaei, and Mamoun Alazab [7].

In this paper, is especially avoid the complex process of explicit facial expression recognition. A facial expression recognition technique is predicated on a convolutional neural network (CNN) and an image edge detection. The extracted edge information is to preserve the edge structure information of the texture image. Then, the dimensionality reduction of the extracted implicit features is processed by the maximum pooling method. Human-computer interaction technology [5] refers to a kind of technology which takes computer equipment as the medium, to realize the interaction between human and computer. The facial expression recognition, as an important means of intelligent human-computer interaction, has a broad application background. It has been applied in the fields of assistant medicine, distance education, interactive games and public security. The facial expression recognition extracts the information representing the facial expression features from the original input facial expression images through computer image.

Image edge detection: The edge information of an image is often reflected in the area where the gradient information of the image changes dramatically. The edge of the image cannot be ignored in the process of texture synthesis. Some edge information is lost, that's why it affects the accuracy of the model. The extraction of the edge of the image is proposed in this technique.

The dataset is used by scientifically mixing the Fer-2013 facial expression database with the LFW data set. Accuracy for the dataset is 88.56%.

The advantage of this proposed method is that extracts facial features effectively and also extract the edge of the image. But only high accuracy for FER-2013 and LFW datasets.

H. Facial expression recognition using feature additive pooling and progressive fine-tuning of CNN

This technique is purposed by T.H. Kim, C. Yu, and S.W. Lee [8].

In this paper, the proposed method is trying to combine the advantages of the convolutional neural network (CNN) and visual geometry group (VGG). Besides, features (captured image) additive pooling is done by CNN, progressive fine-tuning of the convolutional neural network (CNN) for facial expression recognition in a static image is introduced. The network is proposed that partially employs the visual geometry group (VGG)-face model pre-trained on a VGG-face dataset. The proposed method is based on VGG-16 model (a variant of VGG-face model) using VGG-face dataset.

To avoid the issue of choosing the appropriate layers to be re-trained, we present a network architecture consisting of convolution-block branches from the five VGG-face models trained on facial expression datasets with different fine-tuning levels.

The datasets are used: CK+ (Cohn-Kanade), KDEF (Karolinska directed emotional face), and JAFFE (Japanese female facial expression)

1. CK+: The Cohn-Kanade (CK+) dataset includes 123 subjects and 593 image sequences. Among the 593 image sequences, 327 are labelled with seven expressions: happy, sad, surprise, fear, disgust, anger, and contempt.

2. KDEF: The Karolinska Directed Emotional Faces (KDEF) is a set of totally 4900 pictures of human facial expressions. The set of pictures contains 70 individuals displaying 7 different emotional expressions. Each expression is viewed from 5 different angles.

3. JAFFE: This fully annotated dataset includes 213 samples of 10 Japanese females. The dataset also contains the six basic expressions and a neutral expression.

Accuracy for datasets: CK+- 93.02%, KDEF- 89.52%, JAFFE- 50.70%.

The main advantage is to improve accuracy when large data distributions in training and testing datasets. But because of additive pooling, CNN structures get complex.

I. Occlusion aware facial expression recognition using CNN with attention mechanism

This technique is proposed by Yong Li, Jiabei Zeng, Shiguang Shan, and Xilin Chen[9].

In this paper, the convolution neural network is proposed with the Attension mechanism (ACCN) that can perceive the occlusion from the face and only focus on the most discriminative unoccluded regions. ACNN needs not to use the explicitly handle occlusions and unify representation learning which avoids propagating detects error afterwards and occlusion pattern encoding in an end to end CNN.

1. pACNN: Patch Attention technique is focused on discriminative and representative patches.
   - i. Region Decomposition: Face is divided into multiple local patches to find occlusions.
   - ii. Occlusion perception with Gate Unit: In Gate-Unit each small patch is weighed differently according to its occlusion condition.

2. gPCNN: Global-local based ACNN take global face the consideration.
   - i. Integration with full face region: It will infer the local details and global context cues from image concurrently.
   - ii. Global-Gated Unit: GG-Unit in gACNN to automatically weigh the global facial representation.

Accuracy for this technique is 85.07%.

The main advantage is this technique can work on both unoccluded and occluded faces. But this technique has a lower accuracy than other techniques.

J. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)

This technique is purposed by Luz Santamaria-Granados, Mario Munoz-Organero, Gustavo Ramirez-González, Enas Abdulhay, and N. Arunkumar [10]

In facial expression detection using deep convolution neural network on physiological signals dataset (AMIGOS).

In this paper, the focus is on models of deep Convolutional Neural Network (DCNN) than the tradition machine learning algorithms. By using the AMIGOS dataset the experimental tests for classification of the emotional dimensions of arousal: measures how calming and exciting is information and valence: positive and negative affectivity made. For the Features of human emotional behaviour, a deep learning approach applies non-linear transformations to physiological signals.

In this paper, the validation of the supervised learning algorithm and deep learning for facial expression recognition is proposed. To reduce the noise and discover particular morphological patterns CNN layers are used in the model. For detection the emotion Heart Rate Variability (HRV), Blood Volume Pulse (BVP), Skin Temperature (SKT), Electrocardiogram (ECG), and Electrodermal Activity (EDA) is used.

The Arousal and Valence of this proposed technique using DCNN with AMIGOS Dataset are: 76% and 75%.

The advantage is that uses physiological signals (AMIGOS Dataset) improves the accuracy of CNN. But this technique only uses AMIGOS Dataset.

TABLE I. LITERATURE SURVEY

SR. NO.	*PAPER NAME*	*ADVANTAGES*	*DISADVANTAGES*	*DESCRIPTION*
1.	Real-Time Facial Expression Recognition Based on CNN.	The average weighting method, avoid potential errors and reduced influence of noise from environment.	The computational cost is high.	The average method is used after the basic CNN architecture.
2.	Facial expression recognition with FRR-CNN.	The siamese architecture is combined with CNN to handle Ti-pooling for obtaining a more descriptive representation of image.	The framework of FRR-CNN is more complex.	Feature Redundancy Reduced CNN (FRR-CNN) used siamese architecture and Ti-pooling.
3.	Recognizing Facial Expressions Using a Shallow Convolutional Neural Network.	Recognizes Static expressions as well as micro-expressions.	Follows traditional way of CNN.	Use Basic CNN procedure for facial expression recognition.
4.	Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network.	Recognizes 2D and 3D facial expressions.	Due to deep fusion speed is comparatively slow.	3D faces are converted into various 2D maps and then CNN basic procedure.
5.	CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot.	Combination of CNN and LSTM improves accuracy.	Because of fusion difficult to implement.	Combine CNN and LSTM for HRI.
6.	Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images.	Using of Double-Channel Facial images improves accuracy.	High Computational Cost due to Double-Channel.	Two facials images: Greyscale and LBP.
7.	A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing.	Extracts facial features effectively.	Only high accuracy of FER-2013 and LFW Datasets.	Extracts Image Edge and Main image both.
8.	Facial expression recognition using feature additive pooling and progressive fine-tuning of CNN.	Improve accuracy when large data distributions in training and testing datasets.	Because of additive pooling CNN structures get complex.	This method of CNN for Facial expression recognition in a static image.
9.	Occlusion aware facial expression recognition using CNN with attention mechanism.	Improve accuracy on both unoccluded and occluded faces.	Only improves occluded faces.	Occlusion perception for each region of interest in CNN.

SR. NO.	PAPER NAME	ADVANTAGES	DISADVANTAGES	DESCRIPTION
1.	Real-Time Facial Expression Recognition Based on CNN.	The average weighting method, avoid potential errors and reduced influence of noise from environment.	The computational cost is high.	The average method is used after the basic CNN architecture.
10.	Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS).	Uses physiological signals (AMIGOS Dataset) improves accuracy of CNN.	Only uses AMIGOS Dataset	Using of Deep-CNN on AMIGOS Dataset.

IV. CONCLUSION

This paper shows a survey of the different facial recognition techniques currently in use. Some of these techniques are as follows: FRR-CNN [2], Double Channel Facial Images [6], Image Edge Computing [7], Feature Additive Pooling and Progressive Fine-Tuning [8], Occlusion aware [9]. The efficiencies of the survey techniques were found to be 75-95%. We will be using this literature survey as a basis for implementing [4] modal in the future.

V. ACKNOWLEDGMENT

This work was supported by Prof. A. S. Kamble (Department of Computer Engineering) of Modern Education Society's College of Engineering (MESCOE). Furthermore, we indebted to our Head of Department Dr.(Mrs) N. F. Shaikh, whose constant encouragement and motivation inspired us to do our best.

VI. REFERENCES

[1] Keng-Cheng Liu, Chen-Chien Hsu, Wei-Yen Wang, and Hsin-Han Chiang, "Real-Time Facial Expression Recognition Based on CNN", International Conference on System Science and Engineering (ICSSE), 5th September 2019.

[2] Siyue Xie and Haifeng Hu, "Facial expression recognition with FRR-CNN", IEEE, 16th February 2017.

[3] Si Miao, Haoyu Xu, Zhenqi Han, and Yongxin Zhu, "Recognizing Facial Expressions Using a Shallow Convolutional Neural Network", IEEE Access, 5th June 2019.

[4] Huibin Li, Jian Sun, Zongben Xu, and Liming Chen, "Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network", IEEE, 8th June 2017.

[5] Tzuu-Hseng S. Li, Ping-Huan Kuo, Ting-Nan Tsai, and Po-Chien Luan, "CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot", IEEE Access, 11th July 2019.

[6] Biao Yang, Jinmeng Cao, Rongrong Ni, and Yuyu Zhang, "Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images", IEEE Access, 15th December 2017.

[7] Hongli Zhang, Alireza Jolfaei and Mamoun Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing", IEEE Access, 28th October 2019.

[8] T.H. Kim, C. Yu, and S.W. Lee, "Facial expression recognition using feature additive pooling and progressive fine-tuning of CNN", IEEE, 11th May 2018.

[9] Yong Li, Jiabei Zeng, Shiguang Shan, and Xilin Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism", IEEE Transactions on Image Processing, 8th August 2018.

[10] Luz Santamaria-Granados, Mario Munoz-Organero, Gustavo Ramirez-González, Enas Abdulhay, and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)", IEEE Access, 23rd November 2018.