A Survey Paper on Issues and Challenges in Facial Expression Recognition with solution.

[1]Nandani Sharma, [2]Peeyush Kumar Pathak, [3]Deepali Verma
[1]Research Scholar, [2]Assistant Professor, [3]Research Scholar
[1&2]Department Of Computer Science and Engineering, Goel Institute of Technology and Management,
Dr. A. P. J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh, India.
[3] Department Of Computer Science and Engineering, IITBHU, Varanasi, India.

Abstract: With the progress of Facial Expression Recognition (FER) from research laboratory- controlled to testing in-the-wild conditions and the new accomplishment of Deep learning strategies in different fields, deep neural Networks have progressively been utilized to learn discriminative portrayals for programmed FER. Ongoing Deep FER frameworks for the most part center on different significant issues and difficulties: Imbalance, intra-class varieties, Illumination and differentiation over-fitting, and so forth brought about by an absence of adequate preparing datasets like FRE2013, CK+, AffectNet, EmotioNet, MMI, and so on.

In this overview paper, Firstly present the accessible datasets that are generally utilized in the writing and give acknowledged information determination and assessment standards for these datasets and then focused on the issues and challenges with FER and provide the solution with respect to issues and challenges and summarized with competitive performances based on the static and dynamic images.

***Index Terms* - Facial Expression Recognition, Deep Neural Networks, FER2013, CK+, AffectNet, EmotioNet, MMI .**

JETIR

I. INTRODUCTION

Human visage plays a key role in personal identification, emotional expression and interaction. In the last decades, a number of popular research subjects related to face have grown up in the community of computer vision, such as emotion classification, facial landmark detection, face recognition, face verification, face-alignment, etc. [1] For the state-of-the-art in deep FER, we introduce existing novel deep neural networks and related training strategies that are designed for FER based on both static images and dynamic image sequences. Five types of data input (raw data, histogram equalization, isotropic smoothing, diffusion-based normalization, the Difference of Gaussian) [8], [7].

According to Shin, Minchul, Munsang Kim, and Dong-Soo Kwon et al. [7], learning methods are distinguishable from the traditional machine learning algorithms in that they perform the feature extraction and classification process simultaneously. Another advantage of using deep learning methods is that, since they extract features through an iterative weight update by back propagation and error optimization, the classifier could include critical and unforeseen features that humans hardly come up with. This process is called feature learning, and CNN is especially suitable for processing 2D image-based training datasets. CNN can be seen as a special type of multilayer perceptron (MLP), but CNN rather focuses on the local relationships between pixels by using receptive fields.

To train a deep convolutional network and verify its performance, we used multiple kinds of datasets. For the training dataset, the Facial Expression Recognition 2013 (FER-2013 [2]) dataset released for the ICMLW sub-challenge and the Static Facial Expression in the Wild (SFEW2.0 [18]) dataset released for the EmotiW2015 competition were used. FER-2013 contains 28709 training faces, 3589 private test faces, and 3589 public test faces. We used training faces and public test faces for training and left private test faces for the accuracy test. FER-2013 facial images are not exactly frontal, including a variation of the rotation and transition with a 48x48 grayscale format. Since FER-2013 [2] images were collected using the Google Image Search API, they contain a variety of facial expressions existing in real-world conditions. The SFEW2.0 [18] dataset consists of 944 training faces, 422 validation faces, and 372 test faces. The SFEW [31] dataset is a static subset of AFEW [30], which contains video clips extracted from movies. Although the emotions in movies are not very spontaneous, they provide facial expressions in a much more natural and versatile way than those found in laboratory controlled datasets. [12]

The majority of traditional methods have used handcrafted features or shallow learning (e.g., local binary patterns (LBP) LBP on three orthogonal planes (LBP-TOP), non-negative matrix factorization (NMF) and sparse learning) for FER. However, since 2013, emotion recognition competitions such as FER2013 [2] and Emotion Recognition in the Wild (EmotiW) [17], [18] have collected relatively sufficient training data from challenging real-world scenarios, which implicitly promote the transition of FER from lab-controlled to in-the-wild settings. Additionally, due to the dramatically increased chip processing abilities (e.g., GPU units) and well-designed network architecture, studies in various fields have begun to transfer to deep learning methods, which have achieved state-of-the-art recognition accuracy and exceeded previous results by a large margin [8].

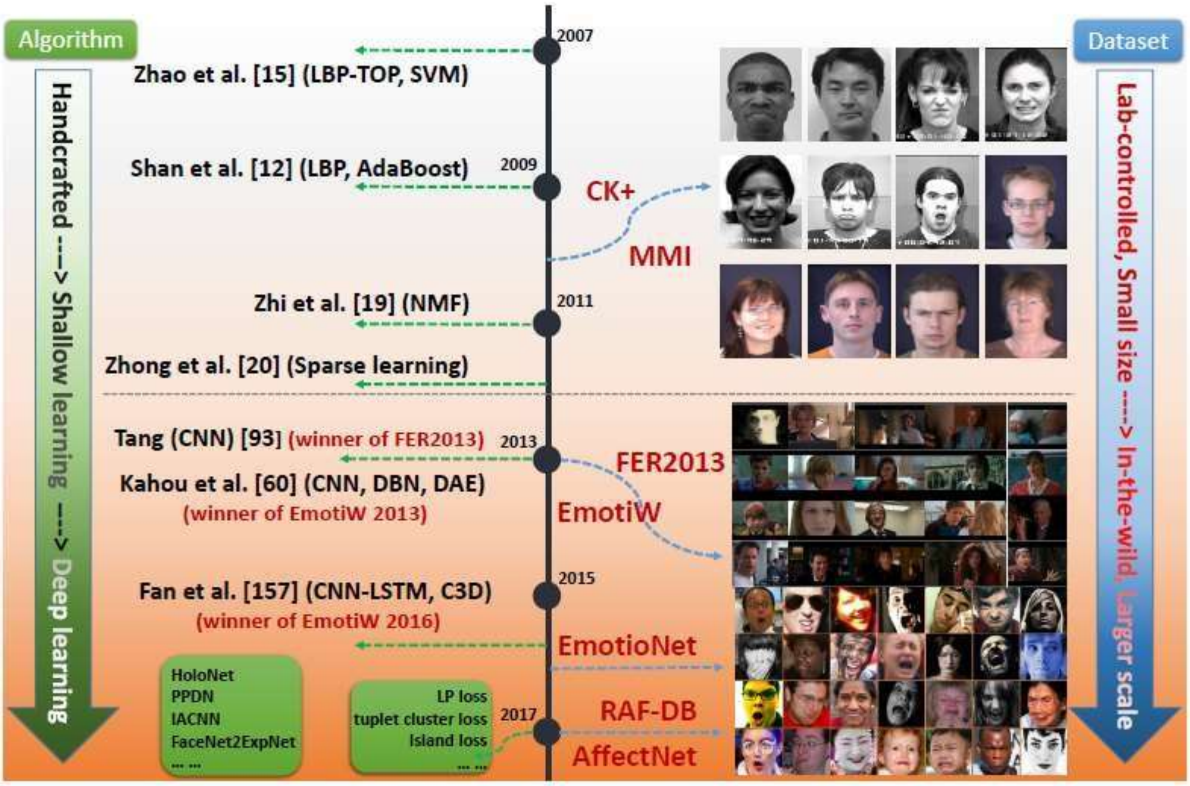

Algorithm
Dataset
2007
Zhao et al. [15] (LBP-TOP, SVM)
Shan et al. [12] (LBP, AdaBoost)
2009
CK+
MMI
Zhi et al. [19] (NMF)
2011
Zhong et al. [20] (Sparse learning)
Tang (CNN) [93] (winner of FER2013)
2013
FER2013
Kahou et al. [60] (CNN, DBN, DAE)
(winner of EmotiW 2013)
EmotiW
Fan et al. [157] (CNN-LSTM, C3D)
(winner of EmotiW 2016)
2015
EmotioNet
HoloNet
PPDN
IACNN
FaceNet2ExpNet
... ...
LP loss
tuplet cluster loss
Island loss
... ...
2017
RAF-DB
AffectNet
Handcrafted ---> Shallow learning ---> Deep learning
Lab-controlled, Small size ---> In-the-wild, Larger scale

Fig.1: methods and algorithm evolution for Facial Expression Detection

There are lots of research are done and continue yet. For facial expression detection or emotion detection various Researchers are using FER datasets for the detection process, compression of various approaches and other various purposes. In this survey paper, I include the mostly all expected types of issues of the datasets with respective their solutions using deep learning and CNN (Convolutional Neural Networks), image processing methodology and other mechanisms

II. A BRIEF DISCUSSION ABOUT FACIAL EXPRESSION RECOGNITION (FER) DATASETS

In this part, we examine freely accessible data sets that contain essential and basic expressions and that are broadly utilized in our surveyed papers for deep learning algorithm evolution. We likewise present recently delivered datasets that contain an enormous number of emotional images gathered from this present reality to profit the training of deep neural networks.

The details of the datasets are as following as:

1. **CK+ [13]:** In the Extended Cohn-Kanade (CK+) dataset contains 593 video sequences from a total of 123 different subjects, ranging from 18 to 50 years of age with a variety of genders and heritage. Each video shows a facial shift from the neutral expression to a targeted peak expression, recorded at 30 frames per second (FPS) with a resolution of either 640x490 or 640x480 pixels. Out of these videos, 327 are labeled with one of seven expression classes: anger, contempt, disgust, fear, happiness, sadness, and surprise. The CK+ database is widely regarded as the most extensively used the laboratory-controlled facial expression classification database available, and is used in the majority of facial expression classification methods.

2. **MMI [14]:** In the MMI dataset is also laboratory-controlled. In dissimilarity to CK+, sequences in MMI are onset-apex-offset labeled, i.e., the sequence begins with a neutral expression and reaches a peak near the middle before returning to the neutral expression. For experiments, the most common method is to choose the first frame (neutral face) and three peak frames in each frontal sequence to conduct person-independent 10-fold cross-validation.

3. **Oulu-CASIA [22]:** In the Oulu-CASIA dataset includes 2,880 image sequences together from 80 subjects. Each of the videos is captured with one of two imaging systems, i.e., near-infrared (NIR) or visible light (VIS), under three different illumination conditions. Similar to CK+, the first frame is neutral, and the last frame has the peak expression. Typically, only the last three peak frames and the first frame (neutral face) from the 480 videos collected by the VIS system under normal indoor illumination are employed for 10-fold cross-validation experiments.

4. **JAFFE [15]:** In the Japanese Female Facial Expression (JAFFE) database contains 213 samples of posed expressions from 10 Japanese females. Each person has 3˜4 images with each of six Essential facial expressions and one image with a neutral expression. Typically, all the images are used for the leave-one-subject-out experiment.

5. **FER2013 [2]:** In FER2013 is an unconstrained and large-scale dataset collected automatically by the Google image search API. All images were registered and resized to 48*48 pixels after rejecting incorrectly labeled frames and adjusting the cropped region. FER2013 contains 28709 training images, 3589 validation images and 3589 test images with seven expression labels.

6. **SFEW [31] and AFEW [30]:** In the Acted Facial Expressions in the Wild (AFEW) database contains video clips collected from different movies with spontaneous expressions, various head poses, occlusions and illuminations. AFEW is a temporal and multimodal database that provides vastly different environmental conditions in both audio and video. The AFEW is independently divided into three data partitions in terms of subject and movie/TV source, which ensures data in three sets, belong to mutually exclusive movies and actors. The Static visage Expressions with inside the Wild (SFEW) become created through deciding on static frames from the AFEW database. The most commonly used version, SFEW 2.0, has been divided into three sets: Train, Val and Test the expression labels of the training and validation units are publicly available, while the ones of the checking out set are held returned through the project organizer.

7. **Multi-PIE [19]:** In the CMU Multi-PIE dataset contains 755370 images from 337 subjects underneath 15 viewpoints and 19 illumination conditions in up to four recording sessions. Each facial image is labeled with one of six expressions. This dataset is typically used for multi-view facial expression analysis.

8. **BU-4DFE [21] and BU-3DFE [20]:** The Binghamton University 3D Facial Expression (BU-3DFE) dataset contains 606 facial expression sequences captured from 100 people. For each subject, six facial expressions are elicited in various manners with multiple intensities. Similar to Multi-PIE, this dataset is typically used for multi-view 3D facial expression analysis. To analyze the facial conduct from the static 3D space to a dynamic 3D space, BU- 4DFE was created that contains 606 3D facial expression sequences with a complete of approx. 60600 frame models.

9. **EmotioNet [25]:** In the EmotioNet, is a large-scale dataset with one million facial expression images collected from the Internet. A total of 950000 pictures were annotated by the automated action unit (AU) detection model in [43], and also the remaining 25000 pictures were manually annotated with eleven AUs. The second track of the EmotioNet Challenge provides six basic expressions and ten compound expressions [51], and 2478 images with expression labels are available.

10. **RAF-DB [28], [26]:** In the Real-world Affective Face Database (RAF-DB), is a real-world database that contains 29672 highly diverse facial images downloaded from the Internet. With manually Crowd-sourced annotation and reliable estimation, seven basic and eleven compound emotion labels are provided for samples. Specifically, 15,339 images from the basic emotion set are divided into two groups (12271 training samples and 3068 testing samples) for evaluation.

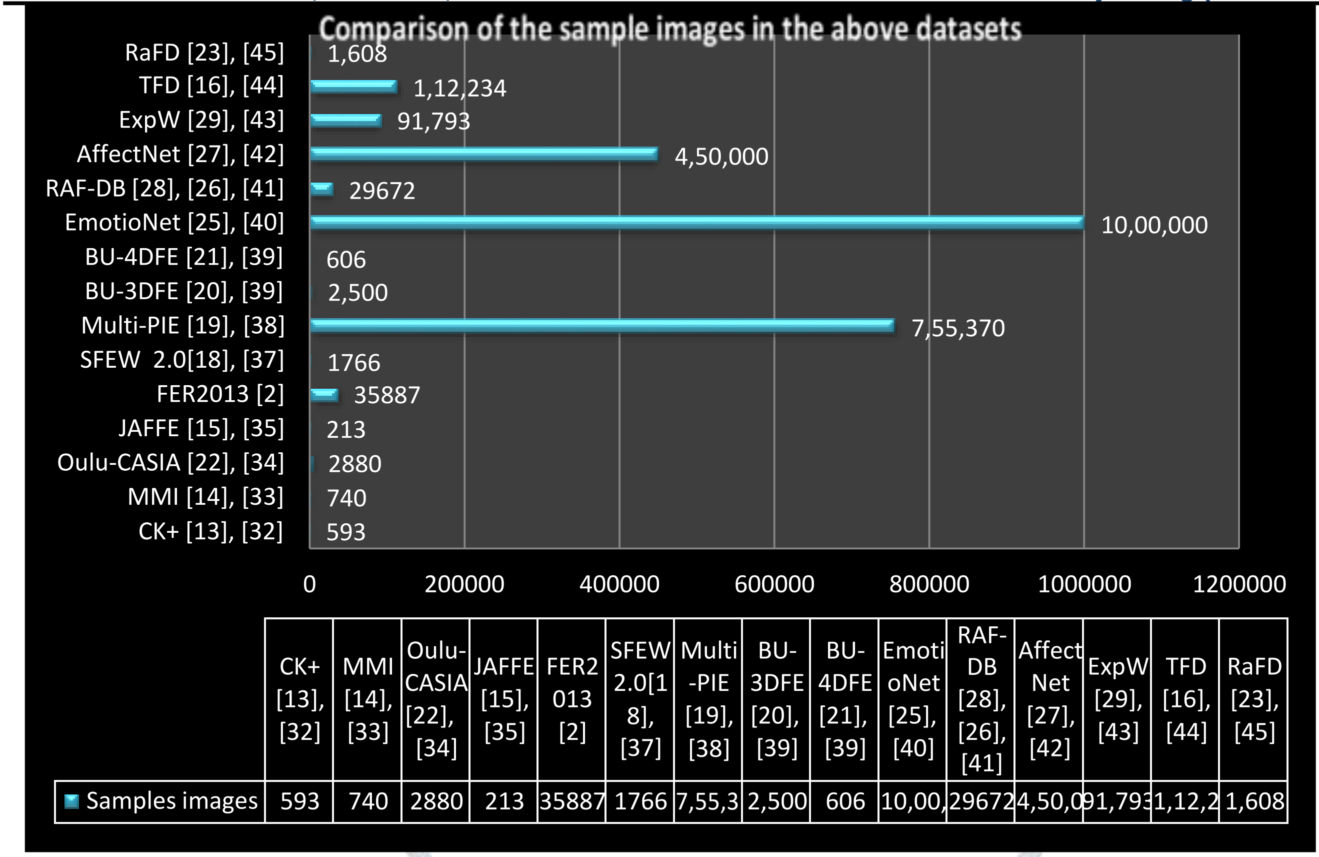
11. **AffectNet [27]:** In the AffectNet , contains more than one million images from the Internet that were obtained by querying various search engines using emotion-related tags. It is by far the largest database that provides facial expressions in two different emotion models (categorical model and dimensional model), of which 450,000 images have manually annotated labels for eight basic expressions.

12. **ExpW [29]:** In the Expression in-the-Wild Database (ExpW), contains 91793 faces downloaded with help of Google image search. Every face images was manually annotated as one of the
13. 7 essential expression categories. Non-face images or pictures were removed in the annotation procedure.

S. No.	Database	Samples images	Samples videos	Subj.	CC	EC	BED	Access Link
1	**CK+ [13] [32]**	593 seq.	_	123	L	P & S both	7 added contempt	https://www.kaggle.com/shawon10/ck-facial-expression-detection/data
2	**MMI [14] [33]**	740	2900	25	L	P only	7	https://mmifacedb.eu/
3	**Oulu-CASIA [22], [34]**	2880 seq.	_	80	L	P only	6 without neutral	https://www.oulu.fi/cmvs/node/41316
4	**JAFFE [15], [35]**	213	_	10	L	P only	7	http://www.kasrl.org/jaffe.html
5	**FER2013 [2]**	35887	_	_	W	P & S both	7	https://www.kaggle.com/msambare/fer2013
6	**AFEW 7.0[17], [36]**	_	1809	_	M	P & S both	7	https://sites.google.com/site/emotiwchallenge/
7	**SFEW 2.0[18], [37]**	1766	_	_	M	P & S both	7	https://cs.anu.edu.au/few/emotiw2015.html
8	**Multi-PIE [19], [38]**	755370	_	337	L	P only	grin, astonished, peer, hatred, shout and disinterested	http://www.flintbox.com/public/project/4742/
9	**BU-3DFE [20], [39]**	2500 3D	_	100	L	P only	7	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
	BU-4DFE [21], [39]	606 3D seq.	_	101	L	P only	7	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
10	**EmotioNet [25], [40]**	1,000,000	_	_	I	P & S both	23 compound expressions	http://mohammadmahoor.com/databases-codes/
11	**RAF-DB [28], [26], [41]**	29672	_	_	I	P & S both	7 and 12 compound expressions	http://www.whdeng.cn/RAF/model1.html
12	**AffectNet [27], [42]**	450,000 (labeled)	_	_	I	P & S both	8	http://mohammadmahoor.com/databases-codes/
13	**ExpW [29], [43]**	91,793	_	_	I	P & S both	7	http://mmlab.ie.cuhk.edu.hk/projects/socialrelation/index.html
14	**TFD [16], [44]**	112,234	_	_	L	P only	7	josh@mplab.ucsd.edu
15	**RaFD [23], [45]**	1,608	_	67	L	P only	7 added contempt	http://www.socsci.ru.nl:8180/RaFD2/RaFD
16	**KDEF [24], [46]**	4,900	_	70	L	P only	7	http://www.emotionlab.se/kdef/

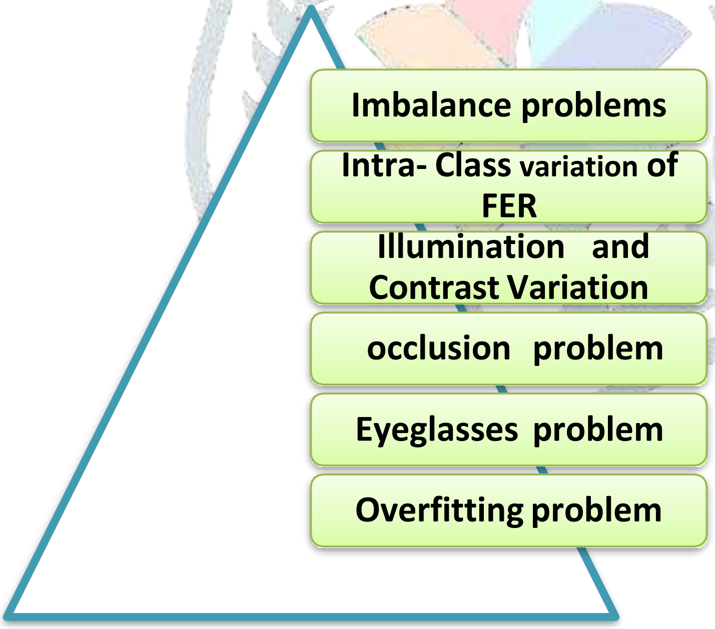
P = posed; S = spontaneous; CC = Collection condition; EM = Elicitation method, Subj. = Subject, BED = Basic Expression Distribution, seq. = sequences, L = Lab, W = Web, M = Movies, I = Internet

Table 1: Information about Facial Expression Recognition Datasets [8]



III. ISSUES AND CHALLENGES OF FACIAL EXPRESSION DETECTION

FER issues and Challenges as following as:



1. **IMBALANCE PROBLEM:**

Data for face analysis usually exhibit highly-skewed category distribution, i.e., most data and information belong to a couple of majority categories, whereas the minority categories solely contain a scarce quantity of instances. For solution, class imbalance problem is divided into two parts. First is Data Re-Sampling and second is Cost- Sensitive Learning [3]. According to Huang, Chen, Yining Li, subgenus Chen amendment Loy, and Xiaoou Tang, et al. [3], planned to mitigate this issue, up to date deep learning ways usually follow classical methods like class re-sampling or cost-sensitive Learning. While not handling the imbalance issue standard ways tend to be biased toward the bulk category with poor accuracy for the minority category. They are simple to organize angular margins between the cluster distributions on a hyper sphere diverse. Such learned Cluster-based massive Margin native Embedding (CLMLE), once combined with an easy k-nearest cluster algorithmic program, shows important enhancements in accuracy over existing ways on each face recognition and face attribute prediction tasks that exhibit unbalanced category distribution. [3]

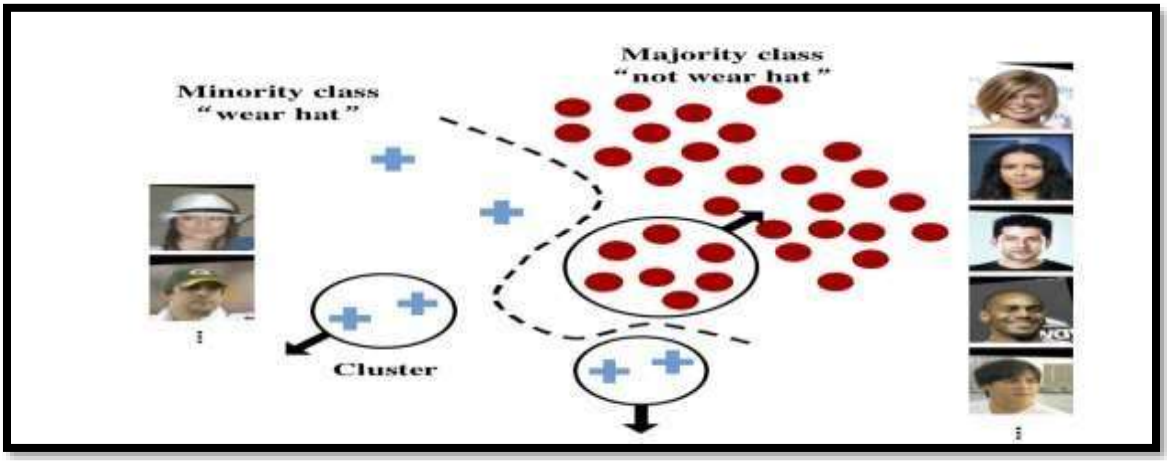


Fig. 2: Example of sophistication imbalance for the binary face attributes “wear hat” method aims to separate the cluster distributions each inside and between categories. This effectively reduces the class imbalance in native neighborhoods and forms balanced native class boundaries that area unit insensitive to the unbalanced size of remaining class samples. [3]

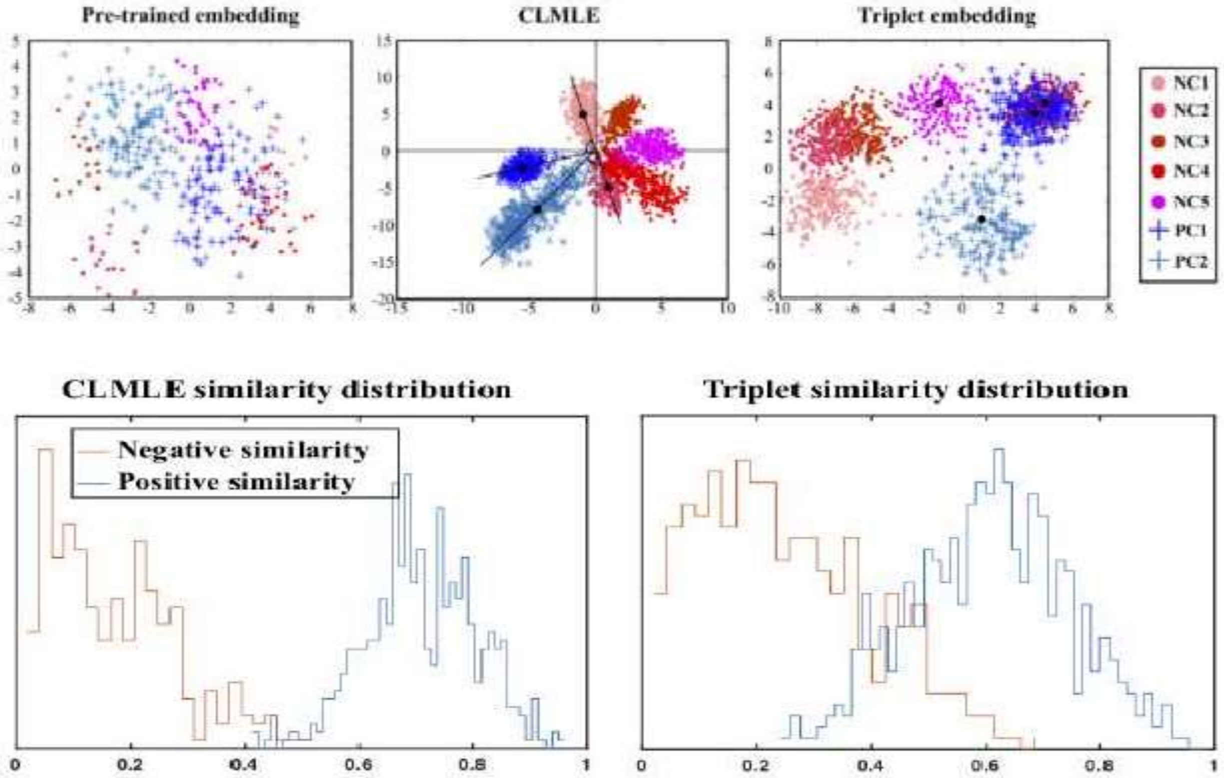


Fig. 3: The 2-D feature space having t-SNE and pair-wise feature similarity for one binary face attribute from the CelebA dataset. They show two Positive Clusters (PC) and five Negative Clusters (NC) to represent the class imbalance. The embedding of a pre-trained model, our CLMLE, and triplet embedding area unit compared. We able to see that between-class clusters (with totally different colors) are well separated in CLMLE, however they're overlapped in triplet embedding, resulting in overlapping binary score distributions. [3]

According to Ngo, Quan T., and Seokhoon Yoon et al. [4] proposed to a new loss function called the weighted-cluster loss, which integrates the advantages of the weighted-softmax approach and the auxiliary loss approach to mitigate the imbalance issue in FER.

2. INTRA- CLASS VARIATION OF FER:

FER is still a challenging problem due to the intra-class variation caused by subject identities [5]. Learning discriminative features for Facial Expression Recognition (FER) in the wild using Convolutional Neural Networks (CNNs) is a non-trivial task due to the significant intra-class variations and inter-class similarities. Deep Metric Learning (DML) approaches such as center loss and its variants jointly optimized with softmax loss have been adopted in many FER methods to enhance the discriminative power of learned features in the embedding space. [6]

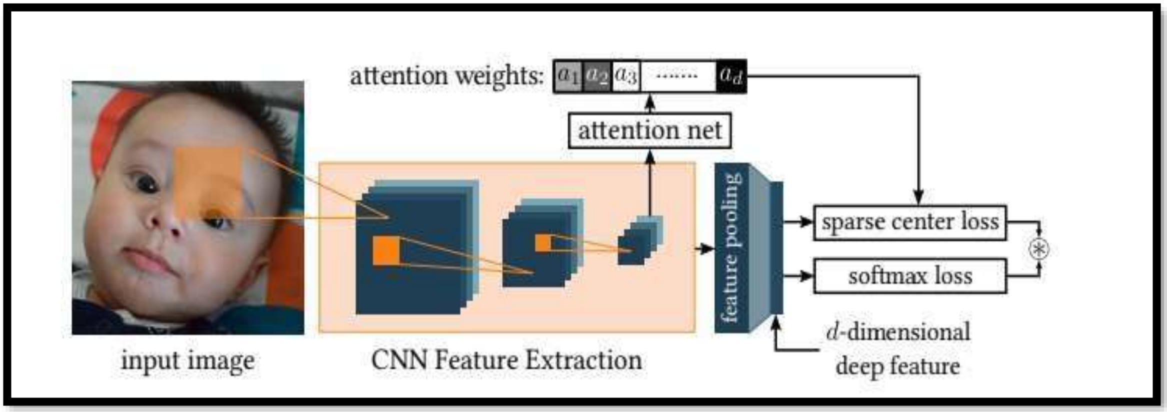


Fig. 4: The high-level overview of our proposed Deep Attentive Center Loss (DACL) method: A Convolutional Neural Network (CNN) yields spatial convolutional features and a feature pooling layer extracts the final d-dimensional deep feature vector for softmax loss and sparse center loss [6]. According to Ngo, Quan T., and Seokhoon Yoon et al. [4] proposed to weighted-cluster loss function simultaneously improves the intra-class compactness and the inter-class separability by learning a class center for each emotion class. Usually, the weighted-softmax loss approach is working to handle this drawback by weight the loss term for every feeling category supported its relative proportion within the training set [4]. A self-difference convolutional network (SD-CNN) is proposed to address the intra-class variation issue in FER. [5]

3. ILLUMINATION AND CONTRAST VARIATION:

Illumination or contrast influences the result accuracy greatly, depending on how it is dealt with Histogram equalization (Hist-eq) is a contrast enhancement technique that usually increases the global contrast of images. This method is effective when the background and foreground's brightness are almost the same. [7]

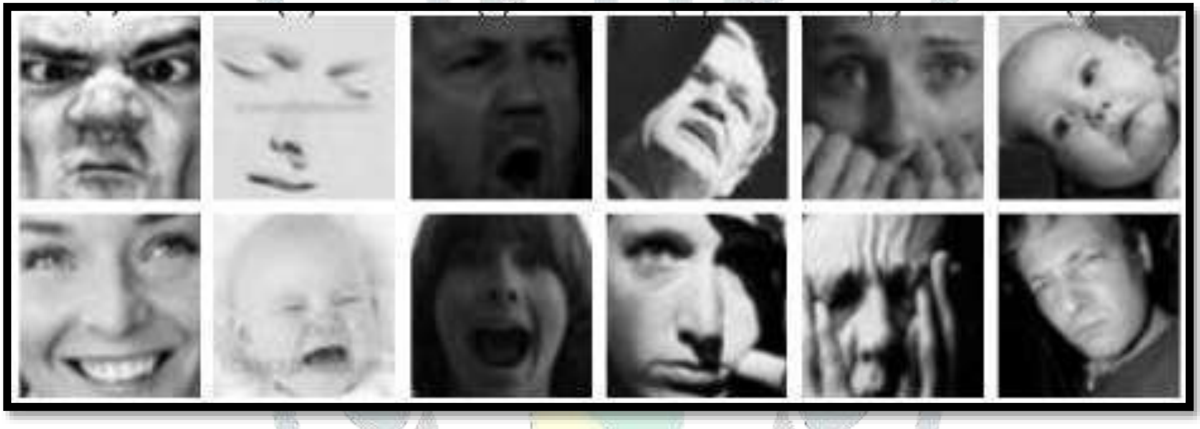


Fig. 5: Illumination and contrast variation [7]

Illumination and contrast can vary in different images even from the same person with the same expression, especially in unconstrained environments, which can result in large intra-class variances. Various algorithms, such as isotropic diffusion (IS)-based normalization, discrete cosine transform (DCT)-based normalization, Difference of Gaussian (DoG) and homomorphic filtering-based normalization, can be used for illumination normalization. [8]

4. OCCLUSION PROBLEM:

Facial expression recognition remains difficult because of the existence of partly occluded faces. It's non-trivial to deal with the occlusion issue as a result of occlusions varies within occluders and their positions. The occlusions are caused by hair, glasses, scarf, respiration mask, hands, arms, food, and different objects that might be placed before of the faces in way of life. According to Li, Yong, Jiabei Zeng, Shiguang Tai Long, and Xilin Chen, et al. [9] proposed to a convolution neutral network (CNN) paying attention mechanism (ACNN) that may understand the occlusion regions of the face and specialize in the foremost discriminative un-occluded regions. ACNN is AN end-to-end learning framework. It combines the multiple representations from facial regions of interest (ROIs).

These objects might block the attention, mouth, a part of the cheek, and the other a part of the face. Convolution Neural Network paying attention mechanism (ACNN), mimicking the approach that human acknowledge facial expressions [9].

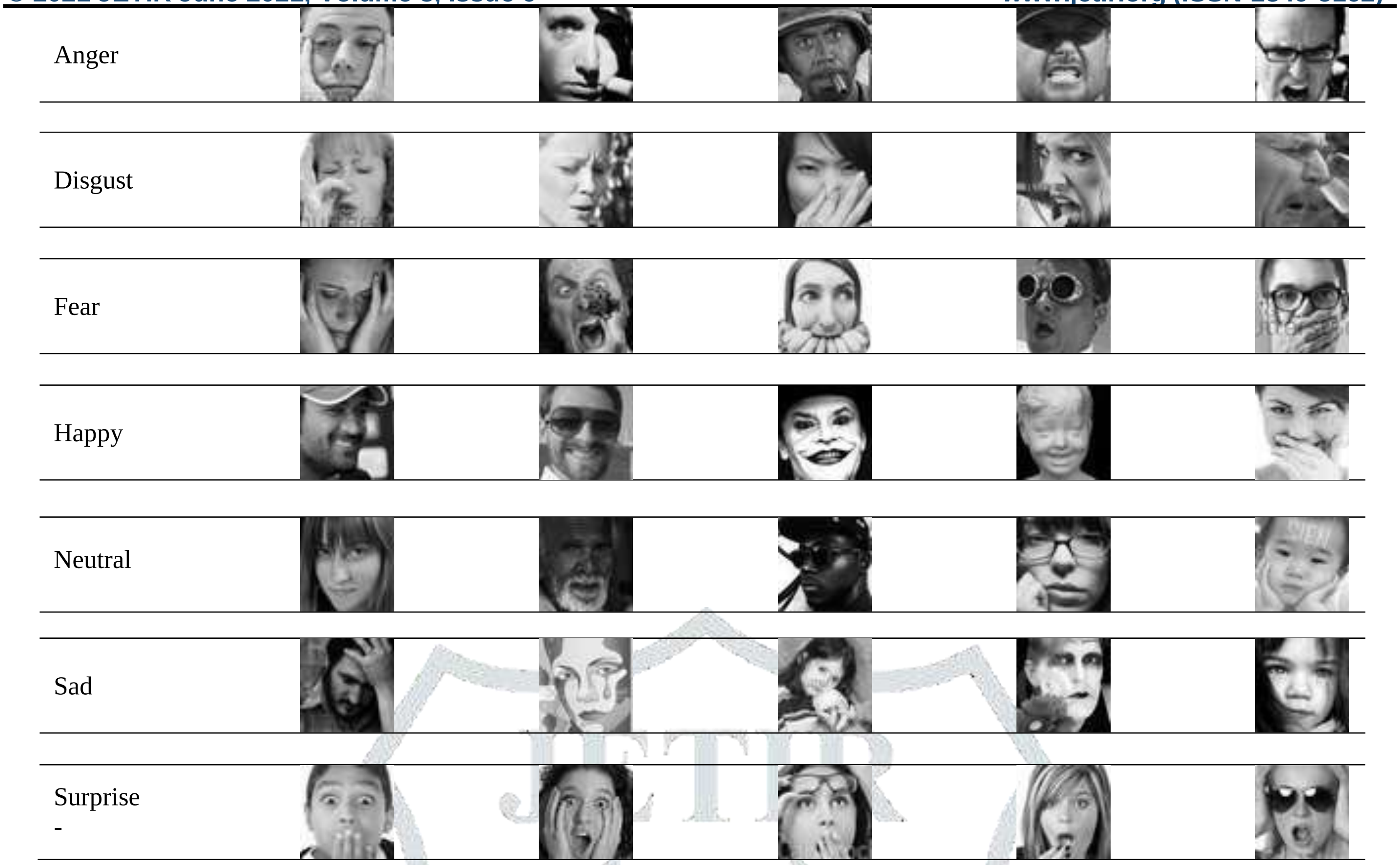
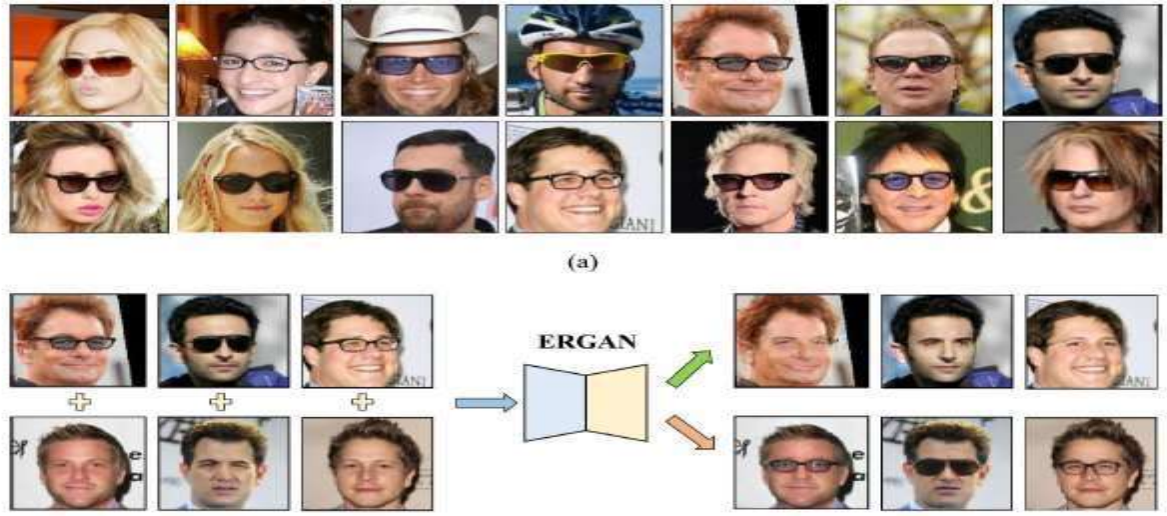


Fig. 6: some images taken from the FER2013 dataset of different occlusion class.

Consideration of local patches makes identification of relevant regions relatively easier since, for example, the object of occlusion could cover certain local patches, thereby making them less relevant eventually as the machine learns. Local attention is generally combined with global context obtained from the undivided attention weighted input for performance gain. [10]

5. EYEGLASS

Eyeglass removal is challenging in removing different kinds of eyeglasses, e.g., rimless glasses, full-rim glasses and sunglasses, and recovering appropriate eyes. Due to significant visual variants, conventional methods lack scalability. Most existing works focus on the frontal face images in the controlled environment, such as the laboratory, and need to design specific systems for different eyeglass types to shows the model for removal of eyeglasses proposed Eyeglasses Removal Generative Adversarial Network (ERGAN) [11].



(a)

(b)

Fig.7: Examples of dissimilar type of eyeglasses in the wild. Different types of eyeglasses have significant visual variants, such as color, style, and transparency. Besides, the faces in the wild are usually in the arbitrary poses with various lighting and backgrounds. (b) A brief pipeline of the ERGAN method. The proposed Eyeglasses Removal Generative Adversarial Network (ERGAN) at the same time takes the 2 relevant tasks, i.e., carrying and removing glasses, into thought [11].

6. OVER-FITTING PROBLEM:

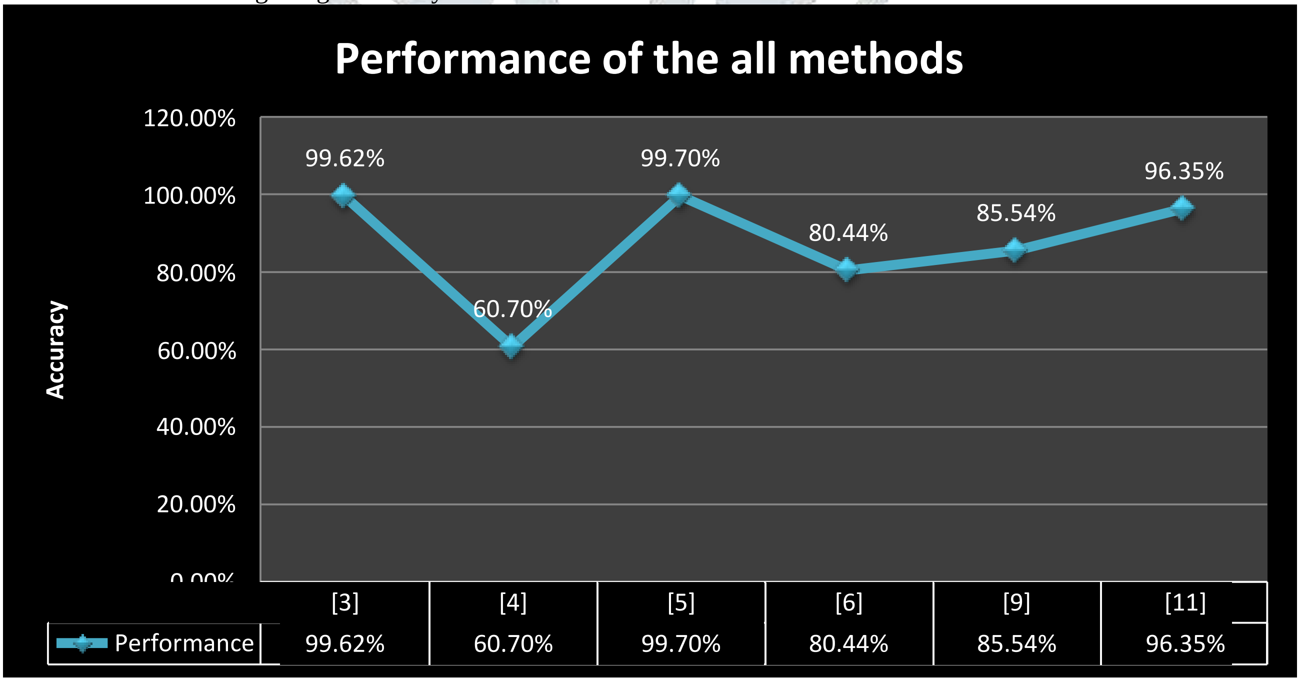
Over fitting caused by an absence of enough training data and information and expression-unrelated variations, like illumination, head create and identity bias [8]. To mitigate this drawback, several studies used further task-oriented information to pre-train their self-built networks from scratch or fine-tuned on well-known pre-trained models (e.g., AlexNet , VGG ,VGG-face and GoogleNet) Indicated that the utilization of further information will facilitate to get models with high capability while not over-fitting, thereby enhancing the FER performance [8], [4]. To avoid over fitting, we have a tendency to another a dropout layer to each single convolutional and totally connected layer and to the ReLU layer further [7].

Table-2: Performance outline of representative strategies for static and dynamic based mostly deep facial expression recognition on the foremost wide evaluated datasets.

Issues/ Problems	Methods	Networks Types	Datasets	Expressions /Classes	Classifier	Performance %
Imbalance and over-fitting problems	[3]	ACNN	LEW CelebA	6 classes	kNN	99.62% 88.78%
	[4]	DCNN	AffectNet	8 classes	SVM	60.70%
Intra- Class variation of FER and over-fitting problems	[4]	DCNN	AffectNet	8 classes	SVM	60.70%
	[5]	SD-CNN	CK+ Oulu-CASIA	8 classes 6 classes	6 DiffNets	99.7% 91.3%
	[6]	CNN	RAF-DB AffectNet	7 Classes 8 classes	NN	80.44% 65.20%
Illumination ,Contrast Variation and over-fitting problems	[8], [7]	CNN	FER2013, CK+,KDEF, JAFFE, SFEW2.0 **	7 & 8 classes	SVM	High and Good **
occlusion problem	[9]	ACNN	RAF-DB AffectNet	7 Classes 8 classes	NN	85.54% 54.84%
	[10]	ACNN	FERPlus, RAF-DB, AffectNet, SFEW	7 & 8 classes	NN	**
Eyeglasses	[11]	ERGAN	CelebA	6 classes	NN	96.35%

ACNN = Convolutional Neural Network with Attention Mechanism, DCNN= Deep Convolutional Neural Network, SD-CNN= Self-Difference Convolutional Neural Network, CNN= Convolutional Neural Network, ERGAN= Eyeglasses Removal Generative Adversarial Network, kNN= k-Nearest Neighbor, SVM= Support Vector Machine, NN= Neural Networks, **= Too much Data and information Available

The above table implies that there are various methods available but here using of the deep learning networks with high accuracy and performance. A Convolutional neural network is the best network using attention mechanism because using ACNN the performance of the Experiments are improving as compare to others. After using of ACNN model to many issues are resolved and got high accuracy.



IV. CONCLUSION:

In this paper, we have an inclination to investigate a additional economical network structure and information preprocessing methodology for establishing a baseline structure for countenance recognition. For the preprocessing methodology of the input image, the Hist-eq methodology showed the foremost reliable performance for all the network models.

Here, we provide the description about the most of used datasets by researchers and also focused on the issues and Challenges of FER datasets and also provide solution and methodology to resolve this issues and solutions in concise manner. In table 1, providing the very helpful information about the mostly preferable datasets which help to other's research work and in table 2, Performance outline of representative strategies for static and dynamic based mostly deep facial expression recognition on the foremost wide evaluated datasets for summarizing this survey paper. Finally, the DCNN and ACNN are most effective and mostly used techniques to get the better performance and accuracy approximately 99% or higher using various deep learning techniques by researchers.

V. REFERENCES

1. Wang, Xiang, Kai Wang, and Shiguo Lian. "A survey on face data augmentation for the training of deep neural networks." Neural computing and applications (2020): 1-29.
2. I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al.,"Challenges in representation learning: A report on three machine learning contests," in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124. https://www.kaggle.com/msambare/fer2013
3. Huang, Chen, Yining Li, Chen Change Loy, and Xiaoou Tang. "Deep imbalanced learning for face recognition and attribute prediction." IEEE transactions on pattern analysis and machine intelligence 42, no. 11 (2019): 2781-2794.
4. Ngo, Quan T., and Seokhoon Yoon. "Facial Expression Recognition Based on Weighted-Cluster Loss and Deep Transfer Learning Using a Highly Imbalanced Dataset." Sensors 20, no. 9 (2020): 2639
5. Liu, Leyuan, Rubin Jiang, Jiao Huo, and Jingying Chen. "Self-Difference Convolutional Neural Network for Facial Expression Recognition." Sensors 21, no. 6 (2021): 2250.
6. Farzaneh, Amir Hossein, and Xiaojun Qi. "Facial expression recognition in the wild via deep attentive center loss." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2402-2411. 2021. https://openaccess.thecvf.com/content/WACV2021/html/Farzaneh_Facial_Expression_Recognition_in_the_Wild_via_Deep_Attentive_Center_WACV_2021_paper.html
7. Shin, Minchul, Munsang Kim, and Dong-Soo Kwon. "Baseline CNN structure analysis for facial expression recognition." In 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 724-729. IEEE, 2016.
8. Li, Shan, and Weihong Deng. "Deep facial expression recognition: A survey." IEEE Transactions on Affective Computing (2020).
9. Li, Yong, Jiabei Zeng, Shiguang Shan, and Xilin Chen. "Occlusion aware facial expression recognition using CNN with attention mechanism." IEEE Transactions on Image Processing 28, no. 5 (2018): 2439-2450.
10. Gera, Darshan, and S. Balasubramanian. "Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition." Pattern Recognition Letters 145 (2021): 58-66.
11. Hu, Bingwen, Zhedong Zheng, Ping Liu, Wankou Yang, and Mingwu Ren. "Unsupervised eyeglasses removal in the wild." IEEE Transactions on Cybernetics (2020).
12. Yu, Z., & Zhang, C. (2015, November). Image based static facial expression recognition with multiple deep network learning. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 435-442). ACM.
13. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010, pp. 94–101.
14. M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect, 2010, p. 65.
15. M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, 1998, pp. 200–205.
16. J. M. Susskind, A. K. Anderson, and G. E. Hinton, "The toronto face database," Department of Computer Science, University of Toronto, Toronto, ON, Canada, Tech. Rep, vol. 3, 2010.
17. A. Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, "Video and image based emotion recognition challenges in the wild: Emotiw 2015," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015, pp. 423–426.

18. A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, “From individual to group-level emotion recognition: Emotiw 5.0,” in Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017, pp. 524–528.
19. R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-pie,” Image and Vision Computing, vol. 28, no. 5, pp. 807–813, 2010.
20. L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, “A 3d facial expression database for facial behavior research,” in Automatic Face and Gesture Recognition, 2006. 7th International Conference on. IEEE, 2006, pp. 211–216.
21. L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, “A high-resolution 3d dynamic facial expression database,” in The 8th International Conference on Automatic Face and Gesture Recognition. Amsterdam, The Netherlands. IEEE, 2008.
22. G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietik¨aInen, “Facial expression recognition from near-infrared videos,” Image and Vision Computing, vol. 29, no. 9, pp. 607–619, 2011.
23. O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg, “Presentation and validation of the radboud faces database,” Cognition and Emotion, vol. 24, no. 8, pp. 1377–1388, 2010.
24. D. Lundqvist, A. Flykt, and A. O¨ hman, “The karolinska directed emotional faces (kdef),” CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, no. 1998, 1998.
25. C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, “Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild,” in Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.
26. S. Li and W. Deng, “Reliable crowdsourcing and deep localitypreserving learning for unconstrained facial expression recognition,” IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 356–370,Jan 2019.
27. A. Mollahosseini, B. Hasani, and M. H. Mahoor, “Affectnet: A database for facial expression, valence, and arousal computing in the wild,” IEEE Transactions on Affective Computing, vol. PP, no. 99, pp. 1–1, 2017.
28. S. Li, W. Deng, and J. Du, “Reliable crowdsourcing and deep locality preserving learning for expression recognition in the wild,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017, pp. 2584–2593.
29. Z. Zhang, P. Luo, C. L. Chen, and X. Tang, “From facial expression recognition to interpersonal relation prediction,” International Journal of Computer Vision, vol. 126, no. 5, pp. 1–20, 2018.
30. A. Dhall, R. Goecke, S. Lucey, T. Gedeon et al., “Collecting large, richly annotated facial-expression databases from movies,” IEEE multimedia, vol. 19, no. 3, pp. 34–41, 2012.
31. A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, “Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark,” in Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011, pp. 2106–2112.
32. https://www.kaggle.com/shawon10/ck-facial-expression-detection/data
33. https://mmifacedb.eu/
34. https://www.oulu.fi/cmvs/node/41316
35. http://www.kasrl.org/jaffe.html
36. https://sites.google.com/site/emotiwchallenge/
37. https://cs.anu.edu.au/few/emotiw2015.html
38. http://www.flintbox.com/public/project/4742/
39. http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html
40. http://mohammadmahoor.com/databases-codes/
41. http://www.whdeng.cn/RAF/model1.html
42. http://mohammadmahoor.com/databases-codes/
43. http://mmlab.ie.cuhk.edu.hk/projects/socialrelation/index.html
44. josh@mplab.ucsd.edu
45. http://www.socsci.ru.nl:8180/RaFD2/RaFD
46. http://www.emotionlab.se/kdef/