

Feature extraction and classification methods of facial expression: a surey

Moe Moe Htay

Faculty of Computer Science, University of Computer Studies, Mandalay (UCSM), Patheingyi, Myanmar

Article Info

Article history:

Received May 10, 2020

Revised Jun 1, 2020

Accepted Jul 24, 2020

Keywords:

Expression classification

Facial datasets

Facial features

Feature extraction

ABSTRACT

Facial expression is a significant role in affective computing and one of the non-verbal communication for human computer interaction. Automatic recognition of human affects has become more challenging and interesting problem in recent years. Facial Expression is the significant features to recognize the human emotion in human daily life. Facial expression recognition system (FERS) can be developed for the application of human affect analysis, health care assessment, distance learning, driver fatigue detection and human computer interaction. Basically, there are three main components to recognize the human facial expression. They are face or face's components detection, feature extraction of face image, classification of expression. The study proposed the methods of feature extraction and classification for FER.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Moe Moe Htay

Faculty of Computer Science

University of Computer Studies, Mandalay (UCSM)

Patheingyi, Myanmar

Email: moemoehtay@ucsm.edu.mm

1. INTRODUCTION

In artificial intelligent era, facial expression recognition (FER) is interesting and challenging task with the problems of limited dataset, different environments, pose, occlusion, person variation etc. FER systems have been applied many systems such as human-computer-interaction (HCI), games, animation of data-driven, surveillance, clinical monitoring etc., [1]. Ekman and Friesen, psychologists from America defined six universal facial expressions: fear, happiness, anger, disgust, surprise, and sadness and also explored action units based facial action coding system (FACS) to describe facial features of expressions [2]. Facial expressions convey nonverbal communication cues that play a significant role in interpersonal relations. Some literatures work adding on other emotions neutral, contempt, and many compound facial emotions. Some researchers employed on handcrafted features extracted using algorithms and others employed on complicated features extracted using deep learning methods. In this paper, we explored the feature extraction methods, feature descriptors, classification methods, methods of feature dimension reduction, frameworks of the facial expression recognition system and the comparison of the results. The remainder of the paper is organized as follows. In section 2, Literature of current FER system. Typical FER system is shown in section 3. After that, two types feature of facial images is discussed in section 4, and section 5 described facial databases for FER system. Section 6 describes the problem statement of FER system. In the last section, conclusion and future work is presented.

2. LITERATURE OF CURRENT FER SYSTEM

Used geometric feature extraction, regional local binary pattern (LBP) features extraction, fusion of both the features using autoencoders and self-organizing map (SOM)-based classifier. The average accuracy 97.55% of MMI and 98.95% of CK+ database. The accuracy of SOM-based classifier is significant improvement over SVM with 3.94% increase for CK+ and 4.36% for MMI dataset respectively [3]. Explored multiple feature fusion applying Histogram of oriented gradients from three orthogonal planes (HOG-TOP) with experimentation of three datasets CK+, GEMEP-FERA 2011, and acted facial expression in the wild (AFEW) 4.0 [4]. Presented a FER model using Haar cascades face components detection and neural network (NN) to train the eye and adding mouth features on JAFFE Japanese database. Comparison of the result of proposed method with Sobel edge detection methods is that the system has achieved more good accuracy. The problem of illumination and pose of the image and to make fully meet theory and practical requirements by integrating other biometric authentication methods and HCI perception methods is still existed [5]. Examined emotion recognition system using hybrid feature descriptors combining spatial Bag of features and spatial scale-invariant feature transform (SBoF-SSIFT) and classifiers of K-nearest neighbor. Codebook construction is applied after features extraction to represent large feature sets by grouping similar features into a specified cluster number. The experimentation accuracy has showed 98.33 and 98.5% on JAFFE and extended cohn-canade (CK+) dataset respectively. However, the recognition performance depends on the number of clusters for codebook generation, number of detected features, levels for image segmentation, and size of training dataset [1]. Implemented cognition and mapped binary pattern-based FER using basic emotion model and circumplex model on CK+ with 100 images for training and 50 images for testing. In the preprocessing step, unwanted information such as hair, ear, and background are removed from the facial image. LBP and pseudo 3D model are used to extract the facial contours and to segment face area into sub-regions. To reduce the dimension of the features mapped local binary pattern is employed and then used two classifiers of SVM and softmax. The result found that local features and expressions are correlated. Moreover, the two classifiers have a little difference in performance. The existence of occlusion, complex conditions, and micro-expression recognition will be conducted in future FER system [6]. Proposed a method Angled Local Directional Pattern (ALDP) for texture analysis of facial expression with six classifiers k-NN, SVM, DT, RF, Gaussian NB and Perceptron on CK+ dataset. Firstly, facial image was detected using Haar-like as [5] and then cropped and normalized the detected image. The accuracy improved 99% with ALDP method with no preprocessing [7]. Also proposed Grey Wolf optimization for feature selection and GWO-neural network (GWO-NN) for feature classification. The parts of face eyes, nose, mouth and ears are detected using Viola-John algorithm and then SIFT feature extraction is used feature points. The accuracy 89.79% on CK+ is less than [7] and achieved 91.22% [8]. Proposed a framework with high-dimensional features combination of appearance and geometric features. The system used deep sparse autoencoders (DSAE) to learn robust discriminative feature and active appearance model (AAM) to locate the facial landmarks 51 points. Three feature descriptors HoG, gray value and LBP are utilized to describe the local features. Linear dimension reduction method of PCA is used to compress the features and then give the map as the input of DASE. The accuracy of the proposed framework achieved 95.79% of CK+ dataset by using leave on subject out cross-validation method [9].

Presented three models of differential geometric fusion network (DGFN) with extraction of handcrafted features, deep facial sequential network (DFSN) based on CNN with auto-extracted features, and DFSN-1 combination of the advantages of DGFN and DFSN by mapping and concatenation of handcrafted and auto-extracted features. DFSN-1 achieved the best performance among the three models on all of CK+, Oulu-CASIA and MMI dataset [10]. Used deep convolutional neural network (DCNN) using caffe framework and Telsa K20Xm GPU. The frontal face is detected and cropped applied by openCV in facial images preprocessing from CK+ and JAFFE. The accuracy of experiment achieved 97% with leave-one-subject-out cross validation on CK+ and 98.12% with 10-folds cross validation on JAFFE [11]. Presented three models of differential geometric fusion network (DGFN) with extraction of handcrafted features, deep facial sequential network (DFSN) based on CNN with auto-extracted features, and DFSN-1 combination of the advantages of DGFN and DFSN by mapping and concatenation of handcrafted and auto-extracted features. DFSN-1 achieved the best performance among the three models on all of CK+, Oulu-CASIA and MMI dataset [10]. Used deep convolutional neural network (DCNN) using caffe framework and Telsa K20Xm GPU. The frontal face is detected and cropped applied by openCV in facial images preprocessing from CK+ and JAFFE. The accuracy of experiment achieved 97% with leave-one-subject-out cross validation on CK+ and 98.12% with 10-folds cross validation on JAFFE [11]. Reviewed analysis of 22 Local Binary Pattern variances on JAFFE and CK databases using the simple parameter-free nearest neighbor classifier (1-NN). For JAFFE database, the highest recognition accuracy achieved 97.14% by using dLBP α , ELGS and LTP, while CK database, the highest recognition rate of 100% by using AELTP, BGC3, CSALTP, dLBP α , nLBPd, STS, and WLD descriptors. The basic LBP descriptor achieved the acceptable performance of 95.71% on JAFFE and 99.28% of CK database. The study can be extended including other problems and other datasets.

Used DCNN adding data augmentation, cross entropy and L2 multi-class SVM [12]. In [13], weighted center regression adaptive feature mapping (W-CR-AFM) for feature distribution and CNN for feature training on CK+, Radboud Faces database (RaFD), Amsterdam dynamic facial expression set (ADFES) and proprietary database. Different of other papers, spatial normalization and feature enhancement preprocessing methods are used. The recognition obtained 89.84%, 96.27%, 92.70% for CK+, RaFD and ADFES respectively. Address illumination problem of real-world facial images using fast fourier transform and contrast limited adaptive histogram equalization (FFT+CLAHE) for poor illumination and then applied merged binary pattern code (MBPC). PCA is used as a method of feature dimension reduction and k-NN as a classifier on SFEW dataset [14]. Released a new database iCV-MEFED at FG work-shop. Multi-modality CNN is compared with CNN for micro emotion recognition in the paper. The proposed network extracted firstly visual and geometrical information of features then concatenated these into a long vector. The feature vector is fed to the hinge loss layer. The framework is better performance than CNN with the misclassification of 80.212137 using caffe [15]. Also proposed another three works of the work-shop. The first winner method using CNN with geometric representation of landmark displacement leading better results compared with texture-only information. The recognition accuracy achieves 51.84% for seven expressions and 13.7% for compound emotion with the performance of average time 1.57 ms using GPU or 30ms using CPU [16].

Employed deep emotional attention model using cross channel CNN by adding attention modulator on the bimodal face and body (FABO) benchmark database. The system applied CNN to learn the location of face expressions in a cluttered scene. The study has shown that the experimentation of one expression attention mechanism and two expression attention mechanism. The accuracy of the framework with attention is better than that of without attention [17]. Proposed a robust facial landmark extraction method by combining data-driven of fully convolution network (FCN) and model-driven of pre-trained point distribution model (PDM) with three steps estimation-correction-tuning (ECT). The computation of response maps of global landmark estimation is trained by FCN and then the maximum points of the maps are fitted with PDM to generate initial facial shape. In the final, a weighted version of regularized landmark mean-shift (RLMS) is applied to fine-tune the facial shape iteratively [18].

Designed to learn NN architecture with three loss functions fully supervised, weakly supervised and hybrid regularization. The experimentation of the proposed model has achieved promising results on CK+, JAFFE under lab-environment and SFEW in the wild [19]. Proposed transductive deep transfer learning (TDTL) architecture to address the problem of cross-database non-frontal facial expression recognition applying VGGface 16-Net on BU-3DEF and Multi-PIE datasets. The study found that feature representation with VGG network is better than traditional handcrafted features such like SIFT and LBP to represent complicated features [20]. [21] Also used the two datasets for the experimentation to address the problem of cross-domain and cross-view of facial expressions using transductive transfer regularized least-square regression (TTRLR) model, color SIFT (CSIFT) features with 49 landmarks and SVM classifiers. The two databases have only four identical categories neutral, surprise, happy and disgust. The experimentation of the study conducted two kinds cross-domain and same view and cross-view and same domain. PCA algorithm also applied to reduce the features dimension.

The studies in references [3], [5], [6] classified six universal emotions as happiness, angry, sadness, surprise, fear, and disgust. In [8], [12], [14], [22] have classified one more class as neutral and [7], [16], [22] have done contempt class. All of eight classes have been classified by the studies in [10], [9], [15]. However, [20] and [21] have worked on neutral, happiness, surprise and disgust expressions. Chen *et al.* [4] employed with 5 classes of GEMEP-FERA 2011 database and 7 classes of CK+ and AFEW. Li *et al.* [23] explained seven basic emotions and 11 compound emotions sadly angry, sadly surprised, sadly fearful, happily surprised, happily disgusted, sadly disgusted, fearfully surprised, fearfully angry, angrily surprised, angrily disgusted and disgustedly surprised. Ferreira *et al.* [19] has worked classification 6 universal classes of JAFFE, SFEW with classes of 6 basic and neutral, and CK+ with 8 classes including contempt.

3. TYPICAL FER SYSTEM

Typical FER system is showed in the following system flow Figure 1. In the detection of face consists of three works: locate the face, crop the face, and scale the face. Features extraction methods, dimension reduction method and classification methods could be selected.

4. FEATURES OF FACIAL IMAGES

Most of the FER system used geometrical features or visual features or both of these features to extract the features from the images of faces.

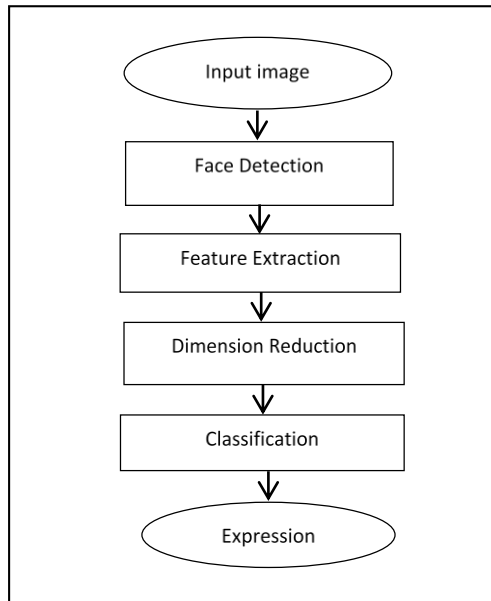


Figure1. Typical FER System

4.1. Geometrical features

Geometrical methods can estimate facial landmarks location or some components of facial images such as the eyebrows, the mouth, and the nose and these features can be measured by distances, curvatures, deformations, and other geometric properties to represent the geometric facial features as they are sensitive to noise [3], [4], [8], [15], [16]. The paper [8] described facial point extraction method to extract the points of eye, nose, mouth, and ears based on Viola-Jones object detection algorithm. Four key regions of face are used to extract geometric features with four steps: detect face, detect eyes, locate eye center then get eye region height, and estimate nose and lips regions. In the paper [16], facial landmark displacement method is applied to extract geometrical information. Affective geometric features are extracted using the warp transformation of facial landmarks to capture the configuration of facial landmark in [4]. Facial landmark with 68 points is described as geometrical representation of face [15].

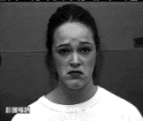
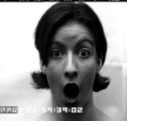
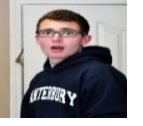
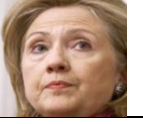
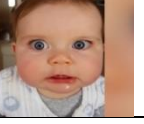
4.2. Apperance features

Appearance methods such as scale invariant feature transform (SIFT), Gabor appearance, local phase quantization can detect the multi-scale, multi-direction of the local texture changes on either specific regions or the whole face to encode the texture [3], [4], [7], [8], [15]. In [6], mapped local binary pattern with four neighborhoods is used to describe the change of local texture features and then face is divided six regions such as forehead, eyes, nose, mouth, left cheek and right cheek using pseudo 3D model. The paper [7] described the texture feature using angled local directional pattern considering the center pixel. In reference [8], SIFT method is applied to extract the unique and precise informative face features. The paper [3] used local binary pattern to extract local texture feature of four basic regions of face: two eyes, nose and mouth. To extract the dynamic texture features from the video, [4] used histogram of oriented gradients from three orthogonal planes (HOG-TOP). The visual features are extracted from the color image using convolutional neural network (CNN) as a feature descriptor in [15]. The effects of the approaches are time-consuming, and the characteristic dimension is huge, so the dimensionality reduction methods are used to affect the accuracy of facial expression recognition.

5. FACIAL DATASETS

Facial expression datasets have two types of creation of images: posed expressions images and spontaneous expressions images datasets. Researchers acquired facial images in three ways such as peak expression images only, image sequences portraying an emotion from neutral to its peak, and video clips with emotional annotations. The two widely used datasets are CK+ and JAFFE [24]-[27]. The real-world facial databases are FER-2013, FERG-DB, SFEW2.0 (static facial expression in the wild), RAF-DB (real world affective face database) and AffectNet database. Sample images of basic facial expression are described in Table 1 for each dataset.

Table 1. Sample image of facial image datasets

Dataset	Sample Images					
	Happy	Sad	Surprise	Fear	Anger	Disgust
CK+						
JAFPE						
FER-2013						
FERG-DB						
SFEW						
RAF-DB						
AffectNet						

5.1. Extended cohn-canade dataset (CK+)

CK+ data set have been widely used in many years in facial expression system. This data set comprises of 593 sequences of image vary in duration from 10 to 60 frames collected from 123 subjects. The age range of subjects is 18-50 years, where 31% are men and 69% are women. The images express seven categories of expressions: happy, sad, surprise, anger, fear, disgust, and neutral that cover the basic emotions. Each image has 640*640- or 490-pixels resolution [25].

5.2. Japanese female facial expression dataset (JAFPE)

JAFPE data set is also widely used in expression recognition of human emotion. This dataset consists of 213 images of 10 Japanese females including seven expressions: six basic (happy, surprise, sad, anger, fear and disgust) and neutral. Each image has the resolution of 256*256 pixels [26].

5.3. FER 2013 dataset

FER-2013 data set contains 28,000 images that are labeled. The dataset is created in 2013 for learning focused on three challenges: the black box learning, the facial expression recognition challenges and the multimodal learning challenges. The images are 48*48 pixels grayscale of faces in seven expressions: six basic expression and neutral [28].

5.4. FERG-DB dataset

FERG-DB stands for facial expression research group database that consists of face images of six stylized characters grouped into seven types of expressions: six basic expressions and neutral. The dataset includes 555767 images [29].

5.5. Static facial expression in the wild dataset (SFEW)

The images in the SFEW are extracted from a temporal facial expressions database Acted Facial Expressions in the Wild (AFEW) which has been extracted from movies. The database contains 700 images that have been labeled into six basic expressions [15].

5.6. Real-world affective face database (RAF-DB)

RAF-DB database is a large-scale facial expression database that includes facial images downloaded from internet. The dataset is annotated seven-dimensional expression distribution vectors for each image [9].

5.7. AffectNet dataset

AffeNet is a largest database of facial expression in the real-world and contains more than 1,000,000 facial images downloaded from the internet search by six different languages with 1250 emotion related keywords. The database defined eleven categories of expression: six basic expressions, neutral, contempt, none, uncertain, and non-face [15].

6. PROBLEM STATEMENT

FER system is need to develop under the problem of illumination, lighting, pose, aging, occlusion for the real-world expression classification system. The major challenges of the study include:

- Most of researches classify basic emotions but fine-grain emotion is relatively small.
- The reaearch works on mocro-expression and compound emotion recognition system are limited.
- Mathematical model is needed to be developed for extraction more discriminant features facial images in the wild.
- Real time facial expression recognition systems should be developed to meet practical application.
- Deep learning model also need to create for improving facial feature extraction and classification.

7. CONCLUSION AND FURTURE WORK

Facial expression recognition is an active research area and more interesting for researcher under the problem of occlusion, brightness, viewing angle, pose, and background in the real-life images, sequence of images and videos. This review paper has presented methods of preprocessing, feature extraction and classification scheme. The FER research goes on to meet real-life applications for driver drowsiness recognition, assistant of distance learning, clinical patient monitoring and teaching robot, health care system for autism children. In the future, FER system will be developed for fined grained facial expressions recognition and compound emotions recognition by using facial images.

REFERENCES

- [1] T. Kalsum, S. M. Anwar, M. Majid, B. Khan, and S. M. Ali, "Emotion Recognition from Facial Expressions using Hybrid Feature Descriptors," *IET Image Processing*, vol. 12, no. 6, pp. 1004-1012, Jan. 2018, doi: 10.1049/iet-ipr.2017.0499.
- [2] P. Ekman and W. V. Friesen, "Facial action coding system a technique for the measurement of facial movement," *Palo Alto: Consulting Psychologists Press*, 1978.
- [3] A. Majumder, L. Behera, and V. K. Subramanian, "Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 103-114, Jan. 2018, doi: 10.1109/TCYB.2016.2625419.
- [4] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial Expression Recognition in Video with Multiple Feature Fusion," *IEEE Transactions on Affective Computing*, vol. 9, pp. 38-50, 1 Jan.-March 2018, doi: 10.1109/TAFFC.2016.2593719.
- [5] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Sinnggh, and A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment," *Procedia Computer Science*, vol. 125, pp. 2-10, 2018, doi: 10.1016/j.procs.2017.12.003.
- [6] C. Qi *et al.*, "Facial Expressions Recognition Based on Cognition and Mapped Binary Patterns," *IEEE Access*, vol. 6, pp. 18795-18803, 2018, doi: 10.1109/ACCESS.2018.2816044.
- [7] A. M. M. Shabat and J. Tapamo, "Angled local directional pattern for texture analysis with an application to facial expression recognition," *IET Computer Vision*, vol. 12, no. 5, pp. 603-608, 2018, doi: 10.1049/iet-cvi.2017.0340.
- [8] N. P. N. Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis, and R. Boothalingam R. "Grey Wolf optimisation-based feature selection and classification for facial emotion recognition," *IET Biometrics*, vol. 7, no. 5, pp. 490-499, 2018, doi: 10.1049/iet-bmt.2017.0160.
- [9] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643-649, 2018, doi: 10.1016/j.neucom.2017.08.043.
- [10] Y. Tang, X. M. Zhang, and H. Wang, "Geometric-Convolutional Feature Fusion Based on Learning Propagation for Facial Expression Recognition," *IEEE Access*, vol. 6, pp. 42532-42540, 2018, doi: 10.1109/ACCESS.2018.2858278.

- [11] V. Mayya, R. M. Pai, and M. M. M. Pai, "Automatic facial expression recognition using DCNN," *Procedia Computer Science*, vol. 93, pp. 453-461, 2016, doi: 10.1016/j.procs.2016.07.233.
- [12] D. V. Sang, N. V. Dat, and D. P. Thuan, "Facial expression recognition using deep convolutional neural networks," *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, 2017, pp. 130-135, doi: 10.1109/KSE.2017.8119447.
- [13] B. Wu and C. Lin, "Adaptive Feature Mapping for Customizing Deep Learning Based Facial Expression Recognition Model," *IEEE Access*, vol. 6, pp. 12451-12461, 2018, doi: 10.1109/ACCESS.2018.2805861.
- [14] A. Munir, A. Hussain, S. A. Khan, M. Nadeem, and S. Arshid, "Illumination invariant facial expression recognition using selected merged binary patterns for real world images," *Optik*, vol. 158, pp. 1016-1025, 2018, doi: 10.1016/j.ijleo.2018.01.003.
- [15] J. Guo *et al.*, "Multi-modality network with visual and geometrical information for micro emotion recognition 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 814-819, doi: 10.1109/FG.2017.103.
- [16] J. Guo *et al.*, "Dominant and Complementary Emotion Recognition From Still Images of Faces," *IEEE Access*, vol. 6, pp. 26391-26403, 2018, doi: 10.1109/ACCESS.2018.2831927.
- [17] P. Barros, G. I. Parisi, C. Weber, and S. Wermter, "Emotion-modulated attention improves expression recognition: A deep learning model," *Neurocomputing*, vol. 253, pp. 104-114, 2017, doi: 10.1016/j.neucom.2017.01.096.
- [18] H. Zhang, Q. Li, Z. Sun, and Y. Liu, "Combining Data-Driven and Model-Driven Methods for Robust Facial Landmark Detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2409-2422, Oct. 2018, doi: 10.1109/TIFS.2018.2800901.
- [19] P. M. Ferreira, F. Marques, J. S. Cardoso and A. Rebelo, "Physiological Inspired Deep Neural Networks for Emotion Recognition," *IEEE Access*, vol. 6, pp. 53930-53943, 2018, doi: 10.1109/ACCESS.2018.2870063.
- [20] K. Yan, W. Zheng, T. Zhang, Y. Zong, and Z. Cui, "Cross-database non-frontal facial expression recognition based on transductive deep transfer learning," *Computer Vision and Pattern Recognition*, 2018.
- [21] W. Zheng, Y. Zong, X. Zhou, and M. Xin, "Cross-Domain Color Facial Expression Recognition Using Transductive Transfer Subspace Learning," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 21-37, 2018, doi: 10.1109/TAFFC.2016.2563432.
- [22] I. Tautkute, T. Trzcinski, and A. Bielski, "I Know How You Feel: Emotion with Facial Landmarks," *Computer Vision and Pattern Recognition*, 2018.
- [23] S. Li and W. Deng, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356-370, Jan. 2019, doi: 10.1109/TIP.2018.2868382.
- [24] C. Loob *et al.*, "Dominant and Complementary Multi-Emotional Facial Expression Recognition Using C-Support Vector Classification," *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 833-838, doi: 10.1109/FG.2017.106.
- [25] 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," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.
- [26] A. Dhall, R. Geoecke, S. Lucey, and T. Gedeon, "Static facial expressions in the wild: data and experiment protocol," *CVHCI Google Scholar*. [Online] <https://fipa.cs.kit.edu/download/SFEW.pdf>.
- [27] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The Japanese female facial expression (JAFFE) database," *Proceedings of third international conference on automatic face and gesture recognition*, pp. 14-16, 1998.
- [28] I. J. Goodfellow *et al.*, "Challenges in Representation Learning: A report on three machine learning contests," *International Conference on Neural Information Processing Springer Berlin Heidelberg*, pp. 117-124, 2013.
- [29] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Modeling stylized character expressions via deep learning," *Asian Conference on Computer Vision*, Springer, Cham, pp. 136-135, 2016.