DOI: 10.24996/ijs.2023.64.3.32

ISSN: 0067-2904

Facial Expression Recognition Based on Deep Learning: An Overview

Salar Jamal Abdulhameed Al-Atroshi[*], Abbas M. Ali

Department of Software and Informatics Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Iraq

Received: 12/1/2022 Accepted: 25/7/2022 Published: 30/3/2023

Abstract:

Recognizing facial expressions and emotions is a basic skill that is learned at an early age and it is important for human social interaction. Facial expressions are one of the most powerful natural and immediate means that humans use to express their feelings and intentions. Therefore, automatic emotion recognition based on facial expressions become an interesting area in research, which had been introduced and applied in many areas such as security, safety health, and human machine interface (HMI). Facial expression recognition transition from controlled environmental conditions and their improvement and succession of recent deep learning approaches from different areas made facial expression representation mostly based on using a deep neural network that is generally divided into two critical issues. These are a variation of expression and overfitting. Expression variations such as identity bias, head pose, illumination, and overfitting formed as a result of a lack of training data. This paper firstly discussed the general background and terminology utilized in facial expression recognition in field of computer vision and image processing. Secondly, we discussed general pipeline of deep learning. After that, for facial expression recognition to classify emotion there should be datasets in order to compare the image with the datasets for classifying the emotion. Besides that we summarized, discussed, and compared illustrated various recent approaches of researchers that have used deep techniques as a base for facial expression recognition, then we briefly presented and highlighted the classification of the deep feature. Finally, we summarized the most critical challenges and issues that are widely present for overcoming, improving, and designing an efficient deep facial expression recognition system.

Keywords: Facial expression datasets, Facial expression recognition, Convolutional neural network, Deep belief network, Deep learning.

التعرف على تعبيرات الوجه المعتمدعلى التعلم العميق: نظرة عامة

سالار الاتروشي ، عباس علي

قسم هندسة البرمجيات والمعلوماتية ، كلية الهندسة ، جامعة صلاح الدين ، أربيل ، العراق

الخلاصة:

التعرف على تعابير الوجه والعواطف هي مهارة أساسية يتم تعلمها في سن مبكرة ومهمة للتفاعل الاجتماعي البشري. تعابير الوجه هي واحدة من أقوى الوسائل الطبيعية والفورية التي يستعملها البشر للتعبير عن مشاعرهم ونواياهم. لذلك ، أصبح التعرف التلقائي على المشاعر المستند إلى تعابير الوجه مجالًا مثيرًا

[*]Email: salar.atroshi@su.edu.krd

للاهتمام في البحث ، والذي تم تقديمه وتطبيقه في العديد من المجالات مثل الامنية، السلامة الصحية وواجهة الإنسان للآلة (HMI). انتقال التعرف على تعبيرات الوجه من الحالة البيئية الخاضعة للرقابة وتحسينها وتعاقب طرق التعلم العميق من مناطق مختلفة جعل تمثيل تعبيرات الوجه يعتمد في الغالب على استعمال شبكة عصبية عميقة والتي تنقسم عمومًا إلى أمرين حاسمين هما تباين التعبير وعدم قدرة الشبكة على التعلم. تشكل أختلاف التعابير مثل تحيز الهوية ووضعية الرأس والإضاءة وعدم قدرة الشبكة على التعلم نتيجة لنقص بيانات التدريب. تطرق هذا البحث أولاً الى الخلفية العامة والمصطلحات المستعملة في التعرف على تعابير الوجه في مجال الرؤية باستعمال الحاسوب ومعالجة الصور وثانيًا ، ناقشنا المراحل الرئيسية للتعلم العميق. بعد ذلك من أجل التعرف على تعبيرات الوجه لتصنيف المشاعر ، يجب أن تكون هناك مجموعة بيانات لمقارنة الصورة بمجاميع البيانات لتصنيف المشاعر. إلى جانب ذلك ، قمنا بتلخيص ومناقشة ومقارنة أساليب مختلفة للباحثين التي استعملت تقنيات التعلم العميق كأساس للتعرف على تعابير الوجه ، ثم قدمنا بإيجاز وأبرزنا تصنيف السمات العميقة ، وأخيراً قمنا بتلخيص أهم التحديات والقضايا التي تطرح على نطاق واسع للتغلب على وتحسين وتصميم نظام فعال للتعرف على تعابير الوجه العميقة.

1. Introduction

Facial expression is one of the most important powerful and natural expressions for human being in order to transmit their emotional and universal signal. There are a lot of research conducted on analyzing facial expression due to it is critically significant in the broad field such as medical treatment, robotics, and driver fatigue surveillance also with most areas of computer vision and machine learning too. Researchers have explored numerous researches on facial expression for encoding information about face from facial representation as shown in Figure 1. Ekman and Friesen [1] at the beginning of the twenty century identified six essential emotions based on the study of cross culture [2] that represented essential perceiving of human emotion in the same way regardless of culture. The prototypical facial expressions are classified into six classes that are disgust, anger, fear, happiness, surprise, and sadness [3]. Today neuroscience and psychology researchers are disagreed on considering six essential emotions to be culture specific rather than not be universal [4].

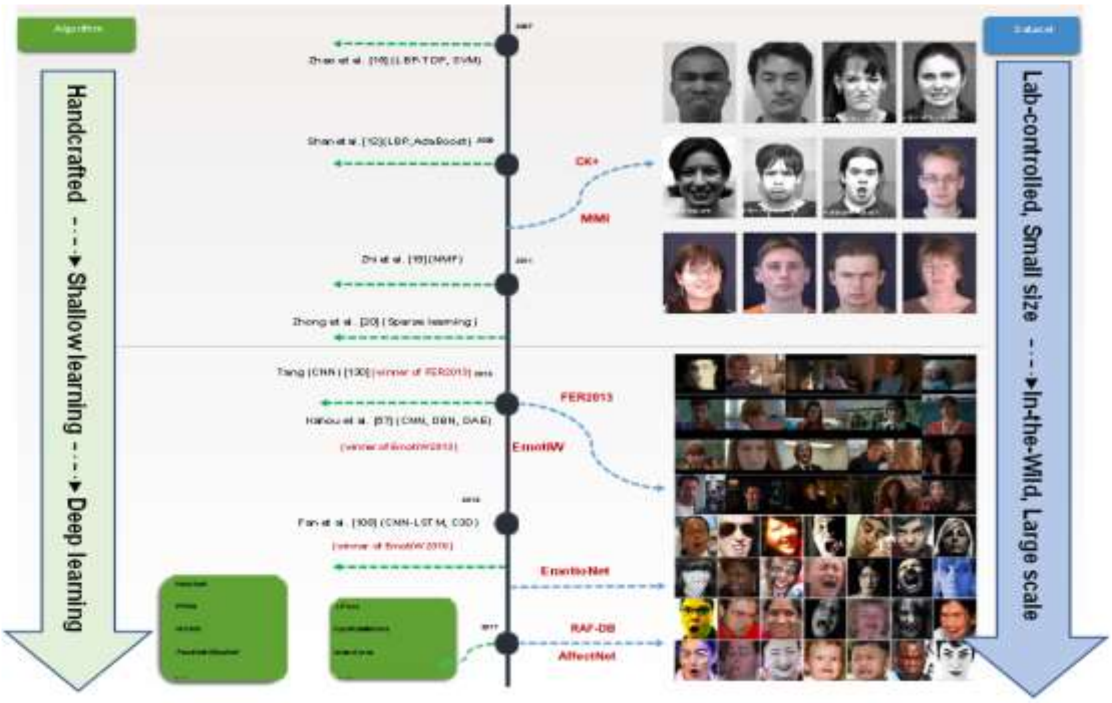
Algorithm
Dataset
Handcrafted ---▸Shallow learning ---▸Deep learning
Lab-controlled, Small size ---▸In-the-Wild, Large scale
Zhao et al. [18] (LBP-TOP, SVM)
Shan et al. [12](LBP, AdaBoost)
Zhi et al. [19] (NMF)
Zhong et al. [20] (Sparse learning)
Tang (CNN) [130] (winner of FER2013)
Kahou et al. [57] (CNN, DBN, DAE)
(winner of EmotiW2013)
Fan et al. [100] (CNN-LSTM, C3D)
(winner of EmotiW2016)
CK+
MMI
FER2013
EmotiW
EmotioNet
RAF-DB
AffectNet

Figure 1: Evolution of facial expression based on deep learning [5].

The emotion system of recognition takes data from different types. There are two types of emotion recognition; generally, emotion recognition and expression recognition which differs from each other. Human stream facial image provides extreme information for recognizing expressions aside from taking photos by the camera, electromyography, electrocardiogram, and physiological signal, as well as they are used as an extra data source from real world FER scenarios [6]. In 2013, the emotion recognizer in wild (EmotiW) [7-9] with expression recognition like FER2013 [10] were in competition by gathering plenty of training data from real world which directly impacts on transferring lab controlled for wild setting. Furthermore, by increasing chip processing power such as GPU, improved network architecture design, and considering various case studies represented by use of deep learning for face recognition will greatly effect on recognition state of art and increases results by a big margin. The deep learning approach is widely used for solving challenges of facial expressions [11-14].

In this paper, we summarized various studies made by researchers on facial expression recognition based on deep learning and review it. The review covered classifying six basic emotions (disgust, anger, fear, happiness, surprise, and sadness). In addition, we mention the common datasets used for FER and also include a section on challenges that are faced up by facial expressions based on deep learning. Our work is to extend and assist researchers, scientists, and other fields for having clear information, idea, and advise them for future co-related work.

2. Facial Expression Recognition Background

In the decades of studying facial expression recognition, two modules (Valence arousal space (V-A space) [15], and action units (AUs) [16] are the popular model. The first model V-A space model is a general approach broadly utilized in the field of recognizing task emotions such as visual, audio, and psychological signals as shown in Figure 2. This module determines emotion classification depending on the dimension of emotion value (Valence and Arousal….). AUs encode basic movement of facial muscles and AUs combination has to be used for facial expression recognition. In [17] proposed a framework to use AUs in order to estimate the intensity of V-A. Facial expression recognition is categorized into two different groups based on features extracted manually or produced through neural network output, like, methods of classical facial expression recognition and deep learning based on facial expression recognition methods [18].

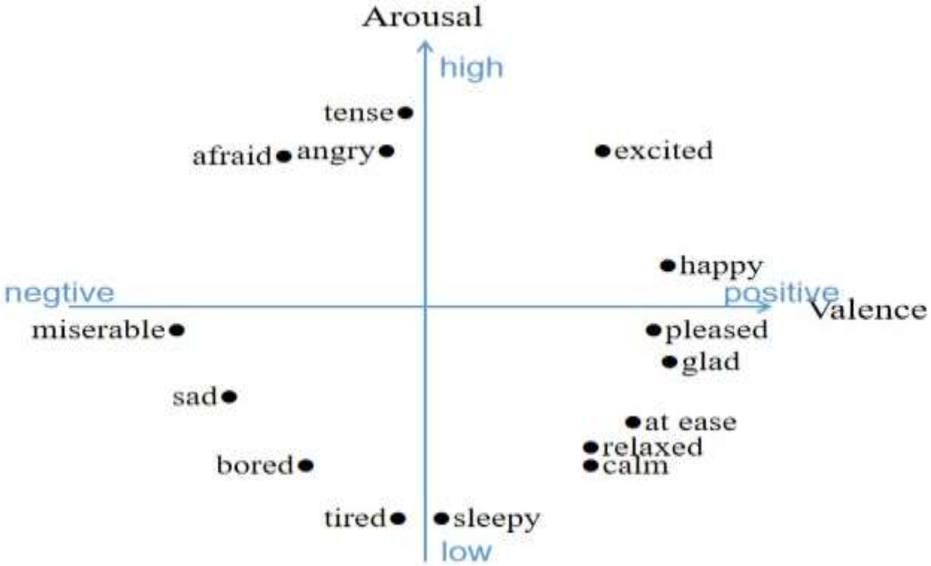

Figure 2: Various emotions and Valence-Arousal Space [18].

2.1. Terminology

In this section we discuss terminology that supports facial expressions that are related to approaches which includes facial land action units, facial landmarks, facial action coding systems deal with how to change facial action to emotion, components, and basic and micro expressions. These are various important definitions in emotion classification. Current studies in facial expression are depending on adjusting these terms and concepts [18].

2.1.1. Facial Landmarks

Facial landmarks are highlighted in facial areas like nose, alae, ending of eyebrows, and corner of the mouth. Positions of facial landmarks surrounding components of face determine capturing as a result of facial expression and head movements. The human faces feature vector is established from point to point that corresponded to facial landmarks as shown in Figure 3 [19, 20].

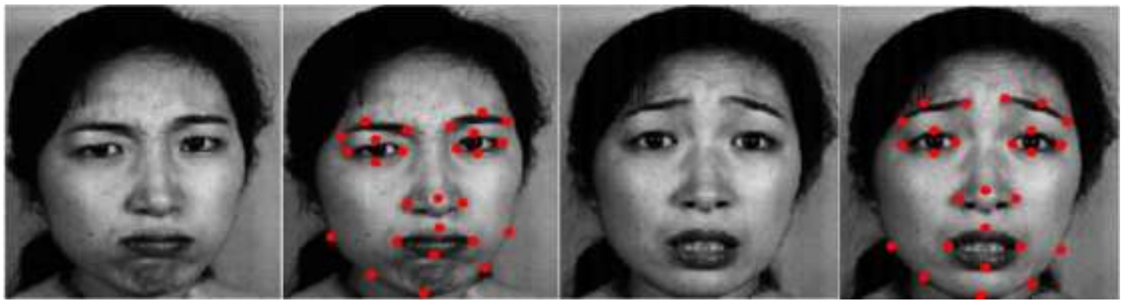

Figure 3: Facial landmarks example [21].

2.1.2. Facial Action Units

Forty-six action units encoded essential movements of groups or individual muscles that were observed during facial expressions and created specific emotions [16]. Figure 4 demonstrates some examples. The facial expression recognition system has classified those expressions into various categorizations by investigating combinations of faces that were detected face AUs. Such example is shown in Figure 4 below. The image is notated with 1,2,5,25 AUs then it is categorized as the emotion in Awed categorization [18].

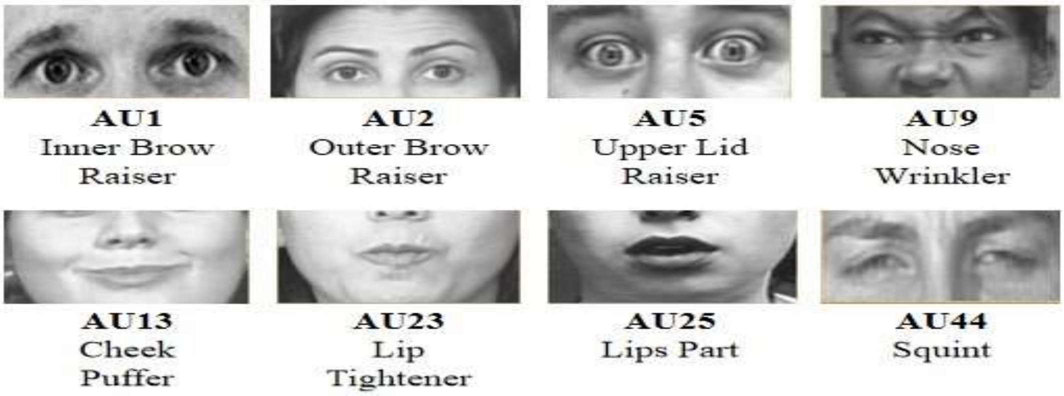

Figure 4: Some examples of Facial Action Units [22].

2.1.3. Facial Action Coding System

Friesen and Ekman psychologists described the critical relationships among expressions and facial muscle movements through biofeedback and observations [23]. Based on anatomical features, initially separating the face into various independent and interrelation AUs, then analyzing the behavior of AUs, their FACS categorizing most human expressions in real world and became a basic reference for muscle movements in facial expressions [18]. The basic and compound of typical AUs are represented in table 1 below.

Table 1: Basic and compound emotions category in prototypical [24].

Category	AUs	Category	AUs
Sad	25,12	Fearfully angry	4, 25, 20
Happy	15, 4	Sadly disgusted	10, 4
Fearful	1,4,25,20	Fearfully surprised	2,5,1,25,20
Angry	7,4,24	fearfully disgusted	4,1,10,25,20
Disgust	10,9,17	Disgusted surprised	2,5,1,10
Surprised	2,1,26,25	Angrily disgusted	25,26,4
Happily sad	6,4,25,12	Happily fearfully	2,1,12,26,25
Happily surprised	2,1,12,25	Angrily disgusted	10,4,17
Happily disgusted	12,10,25	Awed	2,5,1,25
Sadly fearful	4,1,15,25	Appalled	9,4,10
Sadly angry	7,4,15	Hatred	7,4,10
Sadly surprised	4,1,26,25		

2.1.4. Basic Emotions

Human emotions are categorized into six basic emotions that are including surprise, happiness, sadness, anger, disgust, and fear which proposed in [25]. In general datasets of facial expression recognition are labeled with these six emotions [18].

2.1.5. Compound Emotions

This type combing two essential emotions and produce twenty-two emotions [26] that included seven essential emotions (one neutral and six main emotions), twelve compound emotions represented typically via humans, and three extra emotions (Hatred, Awed, and Appalled) [18].

2.1.6. Micro Expressions

This type [27] demonstrates more subtle and spontaneous facial movements that continuously appear. Their aim to explain the potential and true expressions of an individual in a limited time. Micro expression period is so short and remains for 1/25 to 1/3 sec. Micro expression is typically used in police investigation and psychology [18].

3. Facial Expression Recognition Using Deep Learning

We explained three essential steps that are required for deep facial expression recognition including pre-processing, deep learning feature, and feature classification deep learning. We briefly summarized broadly utilized approaches of each above steps as shown in Figure 5.

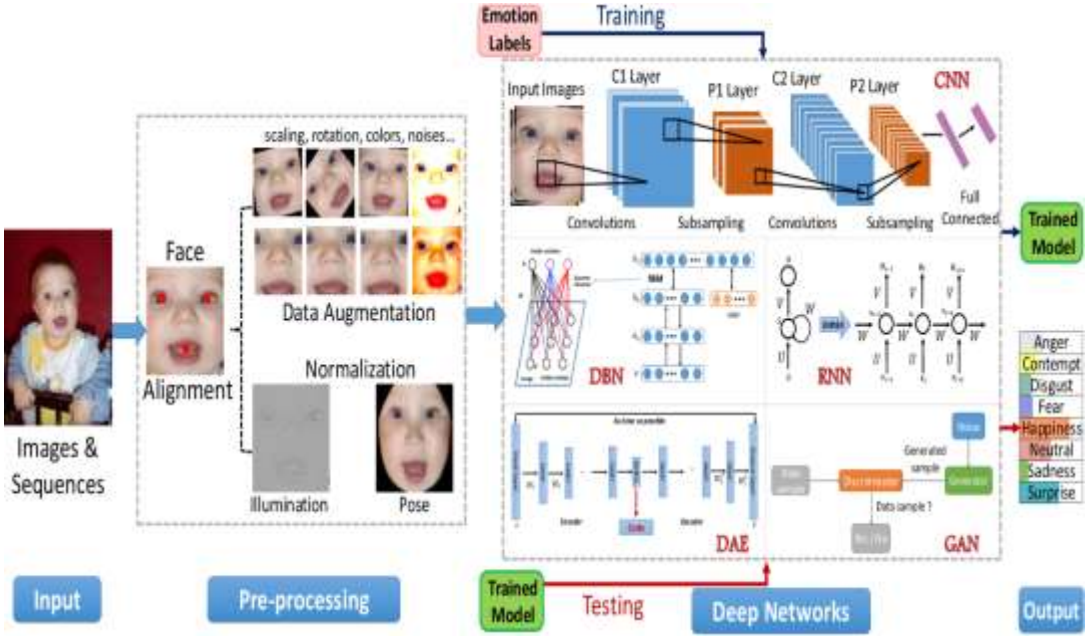

Figure 5: Deep Facial Expression System Pipeline [5].

3.1. Preprocessing

Preprocessing is the main step in deep learning before training the deep neural network for aligning and normalizing visual semantic information from the face that is formed as a result of different variations that are: various backgrounds, head poses, and illuminations [5].

3.1.1. Face Alignment

The initial step before training data is face detection, then removing background and non-face regions. A lot of methods are proposed for face detection as in [28] utilized localization landmarks for face alignment that improved the performance of FER.

3.1.2. Data Augmentation

Data training of deep neural networks require a huge data for ensuring generalizability of recognition tasks. Nevertheless, public databases that are available for facial expression recognition do not have an extremely large number of images in them. Hence, data augmentation is a critical way for deep learning facial expression recognition and is categorized into two types that are offline data augmentation and on the fly data augmentation [5].

3.1.3. Face Normalization

The variance of illumination and head poses have great effects on images and on the performance of facial expression recognition. Wherefore, illumination and pose have to be normalized [5].

3.2. Feature Learning Based on Deep Learning

Deep learning is represented as a hierarchical architecture of multiple non-linear transformations. Briefly, we described and discussed some of the deep learning approaches that have been utilized for facial expression recognition which are: convolutional neural network, generative adversarial network, deep belief network, and self-organizing map.

3.2.1. Convolutional Neural Network

A convolutional neural network (CNN) is the end to end approach and an enhancement of an artificial neural network. Characteristics of CNN include weight sharing, local connectivity that resulted in a fewer number of parameters, increasing training speed, and the effect of regularization, below is an example of FER based CNN approach as shown in Figure 6.

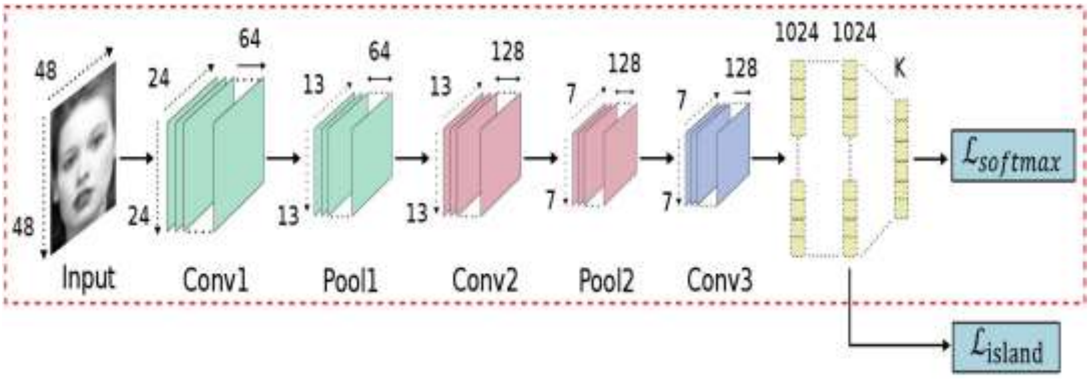

Figure 6: Convolutional Neural Network [29].

Fathallah *et al.* [30] used deep learning techniques for recognizing facial expression and classifying it into several emotions. They used a convolutional neural network (CNN) which is made of three essential structure layers including the convolutional layer, pooling layer, and fully connected layer. Their proposed scheme composed of four convolutional layers for extracting features preceded by three max pooling layers with SOFTMAX output that demonstrated six emotional classes. They trained their method by using a visual geometry group for creating the initial model and in the next step, they enhanced training using the previous first model as shown in Figure 7. Their experiment showed that the recognition rate of five classes which are (disgust, happy, neutral, sad, and surprise) is much higher than compering to angry emotions recognizing that is much more difficult because of database characteristics as shown below in Figure 7. Also, they noticed that uncorrected classification of emotions such as neutral and sad, especially in sad emotion as a result of unclear features in this class.

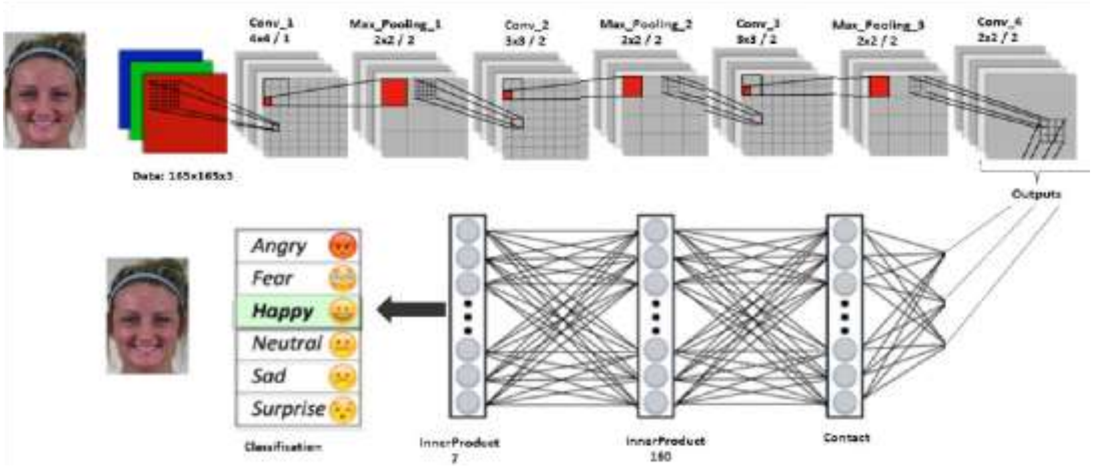

Figure 7: Facial expression based convolutional neural network [30].

Verma and Verma [31] used a convolutional neural network (CNN) for predicting emotions in images by analyzing facial expressions. Their proposed approach is composed of two CNN which are primary CNN (P-CNN) and secondary CNN (S-CNN). P-CNN is used for analyzing primary emotions in images such as sad or happy. Their architecture consists of three FCL (fully connected layer) and three convolutional layers with 1024 neurons for each layer. The final layer consists of two neuron layers with SOFTMAX are applied for classifying images. While second CNN is used for predicting secondary emotions in the image based on the result of P-CNN. Their architecture made of five convolutional layers that are linked with max pooling layer also with three FCL. After that, two dropout layers with the rate of 0.2 are inserted into the network after the first and second intensive layers to decrease the training time and avert overfitting. The last intensive layer includes four neurons for the classification of the secondary emotions of neutral, surprise, fear, or anger. They applied 0.2 which stands for dropout rate for getting higher accuracy. They implemented their experiment only for real time images that are applied by using a camera, not real time video. For their evaluation, they used two different datasets such as JAFFE and FER2013 that got an accuracy of 94.12% with 97.07% respectively.

Talegaonkar *et al.* [32] used CNN for detecting facial expressions in real time and analyzing user emotions during watching video lectures or movie trailers and classifying expressions into seven essential emotions. Their proposed scheme consists of three phases that are preprocessing, face detection, and emotion detection. The first phase is preprocessing of images used for reducing noise and invariance by applying three sequential steps which are normalization, grayscale and resizing. In the second phase, they used Viola Jones detectors for image detection. In the last phase, they utilized various CNN for classifying images into seven essential emotions for obtaining high accuracy and decreasing overfitting. Their CNN architecture used 6 convolutional Layers with max pooling layer for reducing image size and computational also using FCL with activation function (RELU) used for reducing overfitting. They did three experiments and noticed that by increasing the number of Epoch accuracy increased also overfitting will be increased too. Their system did not have abilities to correctly classify disgust and fear as clearly seen in their confusion matrix.

Lee *et al.* [33] proposed a photoplethysmogram approach that combined strategies of statistical features with deep learning. Statistical features were chosen by the person correlation method with deep features extracted by two convolutional neural networks as shown in Figure 8. A photoplethysmogram (PPG) signal is obviously achieved through most devices and procedures for reducing this signal are much easier than for other signals of physiological. Normal to normal (NN) heart rate is used for extracting time domain features. After that they selected features that are highly correlated with emotion through the correlation of persons. Statistical features are combined with deep learning features extracted from the convolutional neural network (CNN). The NN interval and PGG signal are utilized as input to CNN for extracting features and overall merged features are used for classifying the arousal and valence basic parameters of emotion. In their experiment observed that CNN architecture improved due to a fully connected would able to recognize emotions well. In addition, introducing statistical features increased their performance, thus increasing a rate by 3% and improving in terms of performance and short duration time interval.

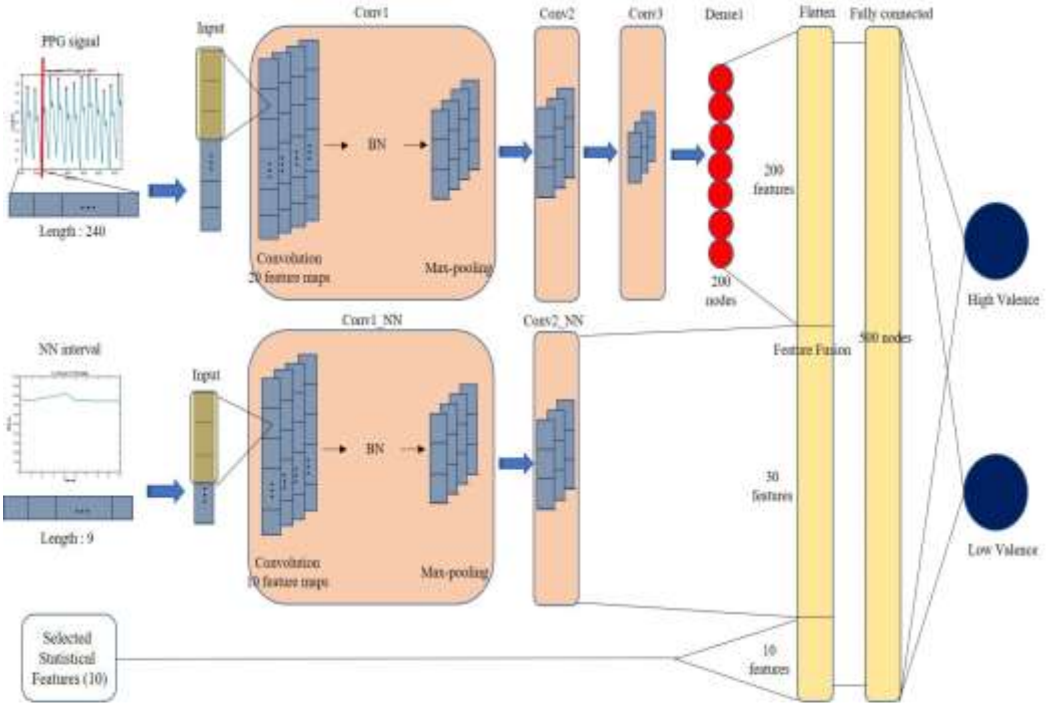

Figure 8: Facial expression based on statistical features and deep learning [33].

Mayya *et al.* [34] proposed an approach for recognizing facial expression based on a deep convolutional neural network (DCNN). They utilized combination layers (convolutional, rectified linear, local response normalization, and pooling operation). Their proposed scheme architecture initially consists of a convolution layer for extracting features of lower level edge, then followed by a RELU layer for increasing nonlinearity and a pooling layer for reducing the dimension of the image. After pooling facial edge was obtained in the fourth layer, local response normalization was used for normalizing brightness in order to increase visibility of features and decreasing irrelevant features. Subsequent to feature extraction, SVM has been used for classification and cross validation methods for estimating the performance. Their methods do not require any preprocessing for feature extraction and their system had capabilities for decreasing the time of feature extraction since they utilized general purpose of graphic purpose unit, they improved state of art by utilizing DCNN.

Oyedotun *et al.* [35] used a combination of deep map latent and RGB representations by using the deep learning technique for recognizing facial expressions. The deep pipeline used DCNN which consists of three max pool layers and five convolution layers for training deep map images, while the RGB pipeline was used for extracting features by utilizing VGG19 and RESNET50 as shown in Figure 9. Then, it is constructed to fully connected layer and SOFTMAX layer with six units for categorizing facial expressions. In their experiment, they used global average pooling for reducing the dimension of data and also, they used batch normalization for enhancing optimization and generalization. They noticed that dropout and batch normalization did not provide any enhancement and their created pipeline utilized hidden unit and activity function of rectified linear function for enhancing gradient to improve convergence speed. Their validation showed that their combined approaches are much improved over using separate modules and the effectiveness of the joint approach on facial expression is clearly seen.

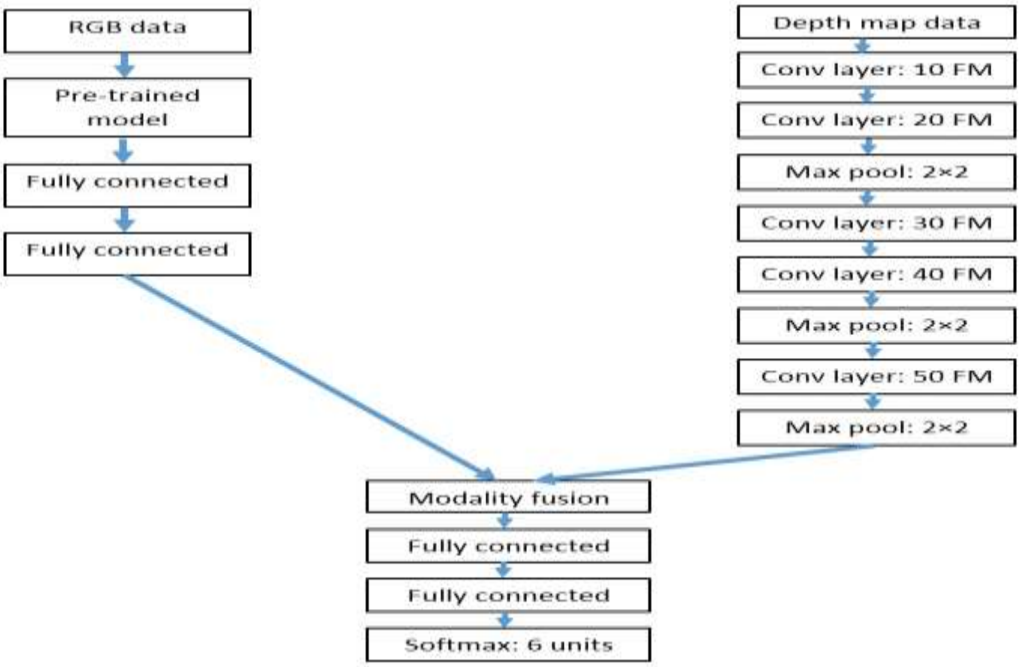

Figure 9: Representation of RGB and Depth map latent fusion pipeline: PL-fusion [35].

Kandeel *et al.* [36] suggested two models based on Convolutional Neural Network (CNN) for Facial Expression Recognition (FER). With less computing complexity, one of these models obtained 100% accuracy for the JAFFE and CK+ benchmark datasets. With the first model, they used image augmentation and image enhancement techniques as shown in Figure 10. The other CNN model is an enhanced version of the first model that has been verified for the more difficult FER2013 dataset, for which we received a score of 69.32 percent. We demonstrate the higher accuracy and efficiency of the suggested approaches by comparing them to recent state-of-the-art approaches to FER.

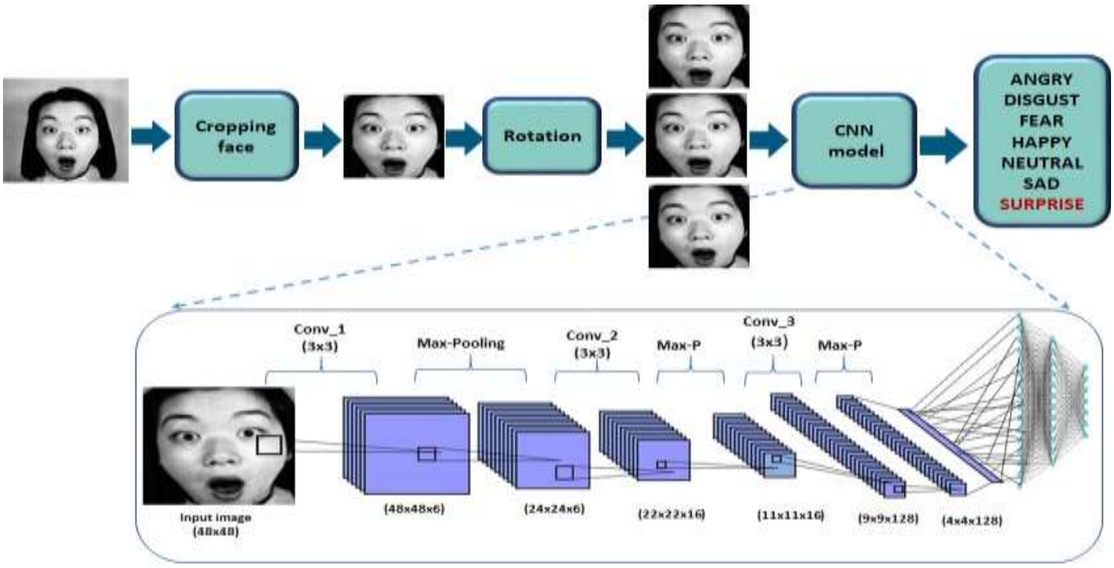

Figure 10: Major phases of the first proposed model architecture [36].

Liu *et al.* [37] proposed a new deep model to enhance facial expression classification accuracy. The following are the advantages of the suggested model: 1) A pose-guided face alignment method is suggested to decrease intra-class variation and solve the effects of environmental noise; 2) A hybrid feature extraction method is proposed to achieve high-level discriminative facial features that improve its performance in classification networks; 3) A lightweight backbone is structured that merges the ResNet and the VGG-16 to obtain low data and low computational during the training. Finally, they perform a chain of experiments on four benchmark datasets, including the CK+, the JAFFE, the Oulu-CASIA, and the AR, to evaluate the proposed model. Their results demonstrate that the suggested model achieves state-of-the-art recognition rates of 98.9%, 96.8%, 94.5%, and 98.7%, respectively. The suggested model has comparable performance in a variety of tasks when compared to standard approaches and other advanced deep learning approaches. Figure 11 illustrates the stages of the proposed model.

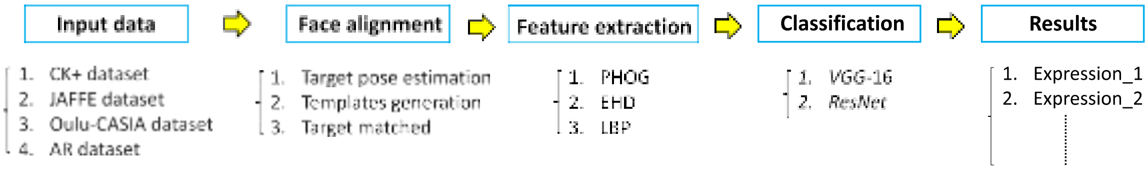

Figure 11: The proposed model structure [37].

3.2.2. Generative Adversarial Networks

It is a combination of two neural network deep learning techniques that are discriminator and generator. The generator produces artificial data, on the other hand discriminator aids in discerning false and real data.

Zhang *et al.* [38] proposed a scheme for generating a facial image with various expressions using a deep learning model that combines three related tasks face synthesis, face alignment, and facial expression recognition. The deep learning model utilized combined geometry code, expression code, and generated code for creating various poses of facial expression recognition. Their proposed scheme had the advantage of improving by complementing joint tasks and separating local and global identities from various geometry codes and expression codes. Therefore, the combined approach has the ability for creating facial images with various expressions under arbitrary geometry codes with increasing performance, however generating facial expressions of various emotions faces several difficulties including change in pose, unlimited facial expressions, incomplete training data, and variation in illustration of original images. Facial expression recognition tasks provide expression code for face synthesis and face alignment, while on the other hand face alignment provides geometry code for facial expression code, facial synthesis (face synthesis) provides adequate training data for facial expression, and facial alignment is used for designing the effective framework and solving challenges as shown in Figure 12. Their target for localizing landmarks, extracting features, classifier learning, and generating sharp images respectively. In their evaluation experiments, they noticed that inappropriate weight effects decrease the performance of face alignment and face expression recognition due to they have an impact on the quality of image generation. So, it directly affects facial expression recognition and facial alignment also observed that there is a strong correlation between facial alignment and facial expression recognition. Their results show that complex conflict is clearly seen between categorized emotions due to inappropriate weight. On other hand, scream and smile emotions were recognized obviously as compared to other expressions, furthermore, the most difficult recognizable is disgust because of their confliction matrix with a squint as they had equal muscle location around eyes also, they observe that expression code is essential for maintaining better expression for facial images.

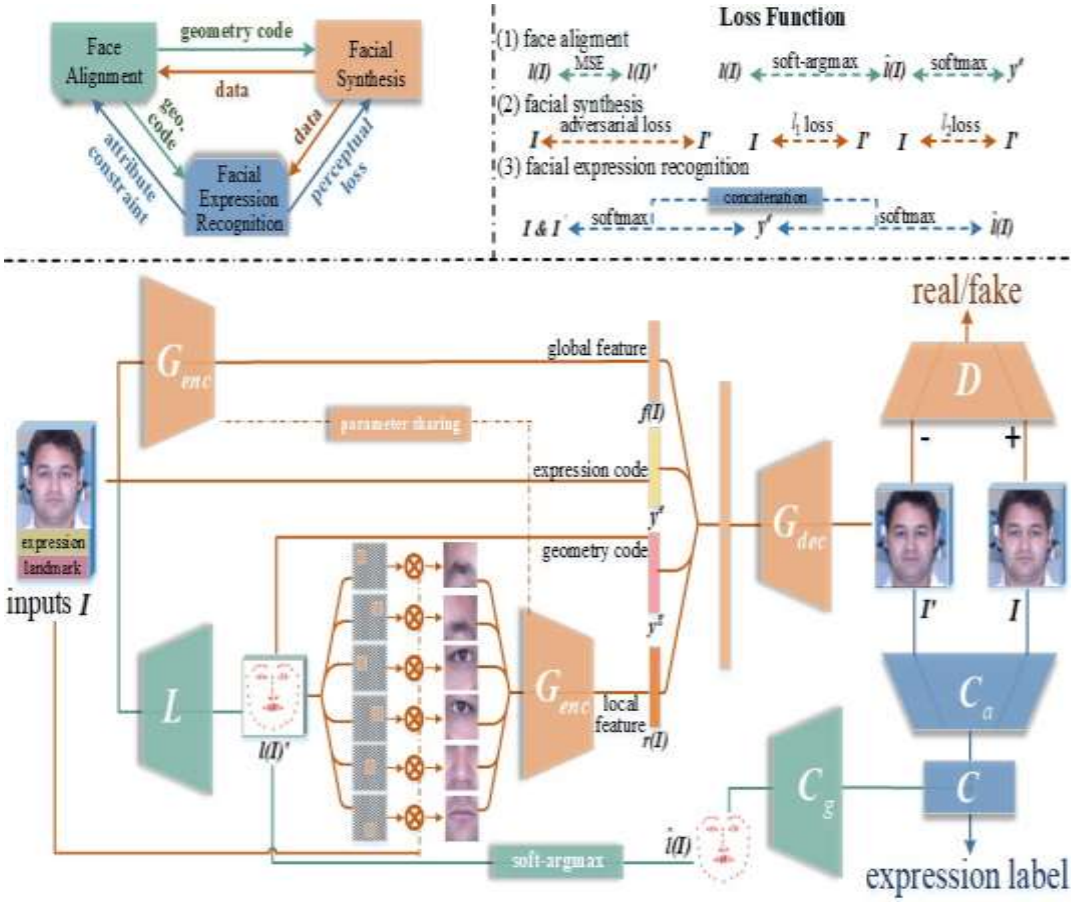

Figure 12: Facial expression recognition based on Generative Adversarial Networks [38].

3.2.3. Deep Belief Network (DBN)

It is based on Restricted Boltzmann Machine (RBM) and its extracting feature of the input signal is abstracted and unsupervised. The facial expression recognition method based on DBN has the ability to learn abstract information of facial images automatically and is considered an effective approach among other deep learning techniques, Figure 13 shows the architecture of a deep belief network.

Zhao *et al.* [39] used a face parse tree and deep learning technique for recognizing facial expressions. They utilized the idea of various components that were active in expression. They trained their model parse tree by using a deep belief network and using logistic regression adjusting. In their scheme detectors first detect the face, after that detect nose, eyes, and mouth in hierarchical form. They used a stack encoder for training deep architecture for recognizing facial expression with respect to the limited feature of detected components and the parsing component eliminate excessive information of expression recognition that location alignment and other technical artificial treatment were not required. Their experimentation results showed that computation complexity is much lesser than their compared methods, thus it directly impacts increasing performance and it easily notice that they used various weights for various portions of face image since each portion has contains different information about expression also noticed that eyes and mouth were easily recognizable as respecting to others scheme were tested and validated the recognizing shapes of different faces. Figure 14 shows the general structure of facial expression recognition by deep learning.

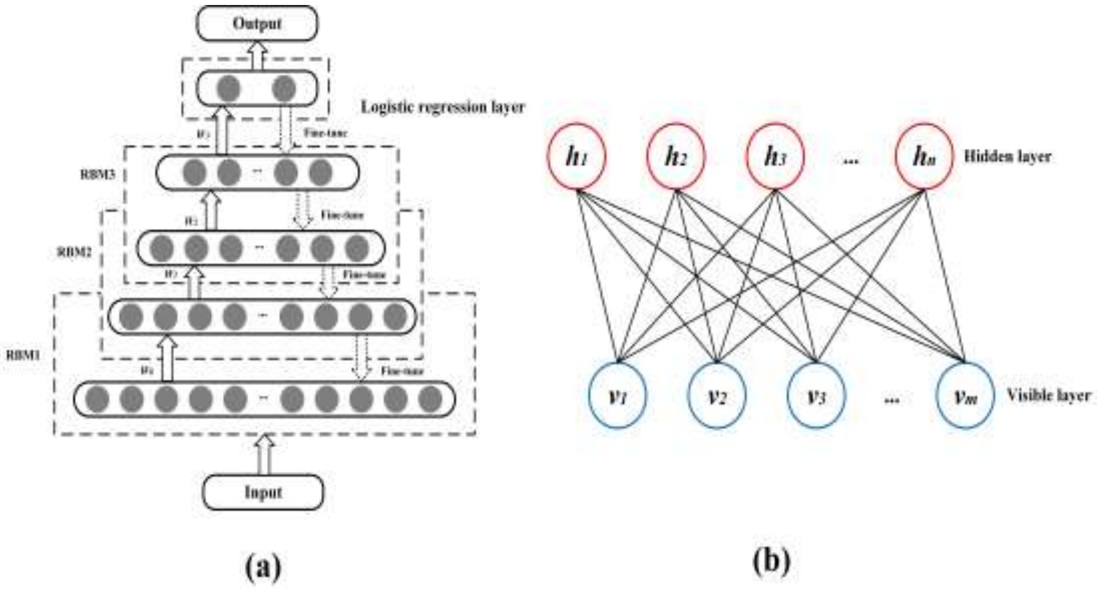

Figure 13: (a) Deep Belief Network and **(b)** Restricted Boltzmann Machine Architecture [40].

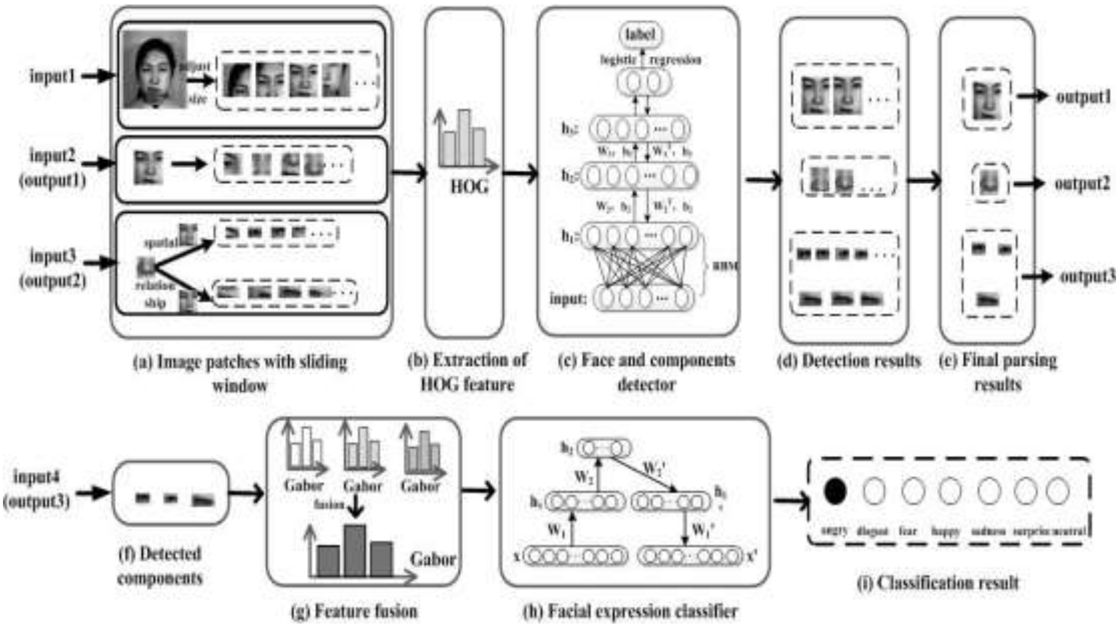

Figure 14: Facial expression recognition by deep learning [39].

Guo *et al.* [41] built a database for depression and asked participants for choosing five various mood-elicitation tasks using Kinect. Facial action units (AUs) and facial feature points (FFPs) were obtained from the database. They extracted facial features from facial expressions utilizing deep learning models depending on FFPs and AUs, named AU-5BDN, 5DBN, and 5DBN-AU. Deep learning consists of multiple layers, their main purpose for learning high level abstracted features, good representation, and using a facial feature as input for creating various hidden layers for three different emotions for males and females respectively. By increasing facial points data recognition rate becomes higher with respecting to other DBN layers. Their results evaluation shows that the performance of AU-5DBN is better than 5DBN-AU and single AU in recognizing emotional tasks. So, it means that AU coupled with facial feature points can improve robustness of extracting features, the performance of AU-5DBN is better than 5DBN-AU due to facial points number is far bigger than AU so, AU has to be linked with a lower level for better improvement. On other hand, the best semantic feature obtained by reducing facial feature part layer by layer then linked with AU-5DBN also presented that the recognition rate of stimulating reading text and question answered were lower due to recording facial expression during speaking in both masks, features were mixed with other factors and shown that female rate classification is much bigger than male rate classification.

3.2.4. Self-Organizing Map

The self-Organizing Map operates with the aid of unsupervised data that decreases the number of random variables in the approach. It has an output of fixed two dimensions each synapse linked with two output and input nodes as shown in Figure 15.

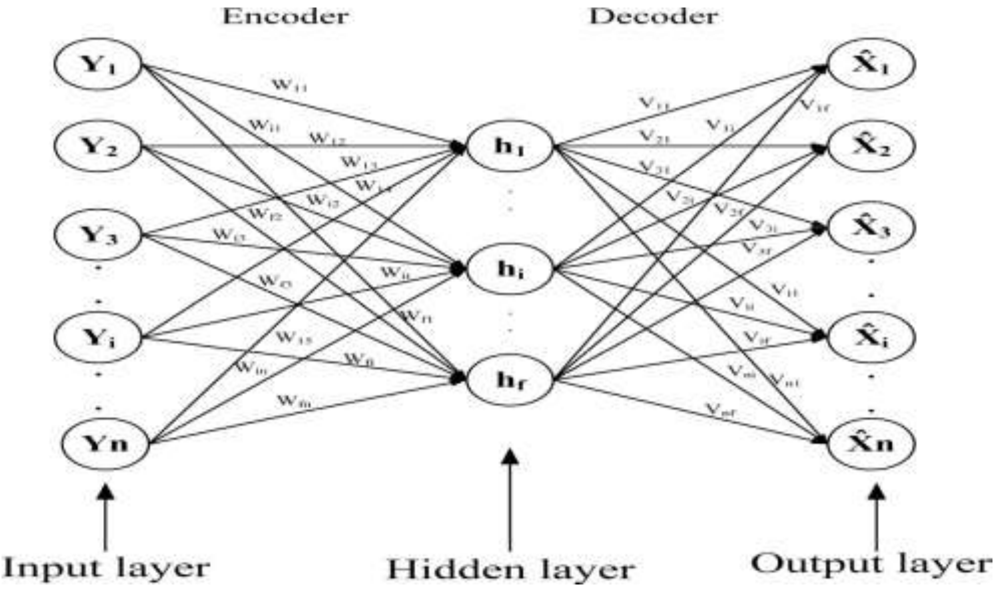

Figure 15: The architecture of an auto-encoder with a single hidden layer [42].

Hussien et al. [43] used a deep learning model for recognizing facial expressions and improving challenges such as accuracy faced up by three spontaneous databases that were utilized in their approach. Their enhancement approach aimed for solving issues including misalignment issues for training images, decreasing the complexity of feature extraction and using auto encoding for deep learning. Feature extraction for facial expression used two crucial ways of facial detection for detecting faces that were used for preprocessing faces from the reference image. While landmarking was used for detecting points in the image, then used to solve alignment based on landmarking for the reference image and express image. After solving misalignment issues, geometric features were extracted with using Discriminative Response Map Fitting (DRMF) that directly reduce complexity. Finally, they used auto encoding for deep learning, back propagation for closing target values in an input feature vector, and used a self-organizing map (SOM) based classifier that consists of two steps of mapping and training. In the training step, inputs used for creating maps however in the second step automatically vectors were classified in mapping, SOM, and encoder was complemented each other as well as they decided the final decision of classification. They showed that the accuracy of the database much improved by using a mixture of extracted features with approaches of normalization, utilization of translation, rotation, and scaling as compared to other art methods, and their results much improved in terms of recognition of facial emotion.

3.3. Classification of Facial Expression

The last stage of deep learning is classifying an image into one of the main categories of emotion. It does not like classical approaches of feature extraction and feature classification stages. They are not dependent on each other. Facial expression recognition end to end is performed by deep networks and at the final step, a loss layer is added for regulating the error of back propagation, after the network output the image that was given for prediction. In convolutional neural network loss is mostly utilized function that reduces cross entropy among ground truth and portability of class that was estimated by deep approach. [44] presented the advantage of utilizing a linear support vector machine for using end to end training that directly impacts reducing margin-based loss rather than cross-entropy. On another hand, [45, 46] deep neural forest adaptation (NFs) investigated replacing results obtained with SOFTMAX loss with NFs. Furthermore, the learning method of end to end, others have proposed [47, 48] for employing deep neural network, especially CNN as a tool extractor for a feature and after that performing independent extra classifiers, like random forest and support vector machine for extracting representations. [49, 50] revealed that computed descriptors on DCNN features and Gaussian kernel classification on positive symmetric definition manifold are extremely better than classical SOFTMAX classification as shown in Figure 16.

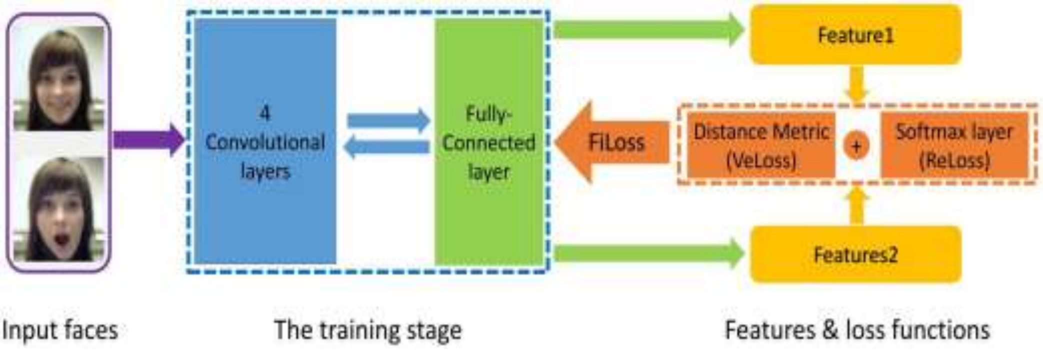

Figure 16: Loss layer for facial expression recognition [51].

Table 2 below illustrates the outline of the literature review on facial expressions based on deep learning.

Table 2: Summary of review literature on facial expression based on deep learning, P= posed; S= spontaneous; Elicit. = Elicitation method.

Author	Methods	Database	Samples	Subjects	Elicit.	Data Groupe	Extra	Resolution	Accuracy	Emotion Number
Zhao *et al.* [39]	DBN FP+ SAE	Japanese Female Facial Expression (JAFFE) database + (CK+) database	213 + 593 images	10 123	P&S	7 Folds	SVM	256×256 640 × 480, 640 × 490	90.47 % 91.11 %	6 Basic Emotions+ Neutral
Zhang *et al.* [38]	GAN	Multi-PIE+BU-3DFE+SFEW datasets	755,370 +2500 +1766 images	337 100 N/A	5P 35P P	5 Folds	SVM	N/A 1040 × 1329 N/A	92.91 % 82.6% 29.10 %	6 Basic Emotions+ Neutral

Guo *et al.* [41]	DBN, AU-5DBN,5DBN-AU	Kinect	1347 Facial points	3	P&S	10 Folds		200×200	80% 86% 90%	6 Basic Emotions
Oyedotun *et al.* [35]	Depth Map+ RGB: PL-Fusion-VGG19,ResNet50	BU-3DFE	2500 images	60	P&S	10 Folds		64×64	87.08 % 89.31 %	6 Basic Emotions
Fathallah *et al.* [30]	CNN	CK+, MUG, and RAFD	37000 images	86 67	P	5 Folds	Adaboost	224×242	99.33 % 87.65 % 93.33 %	6 Basic Emotions
Hussien *et al.* [43]	Self-Organizing Map	MMI+VD-MFP+BINED	500 images	50	P&S	10 Folds		240×240	97.0% 94.1% 93.7%	6 Basic Emotions
Verma and Verma [31]	CNN P-CNN S-CNN	JAFFE+ FER2013	213+ 28,709 images	50	P&S	10 Folds		48×48	94.12 % 97.07 %	6 Basic Emotions+ Neutral
Talegaonkar *et al.* [32]	CNN	FER2013	35,887 images	50	P	10 Folds		48×48	89.78 %	6 Basic Emotions+ Neutral
Lee *et al.* [33]	CNN PPG	DEAP	4800 images	32	P	6 Folds	Adam	20×40	82.1% 80.9%	6 Basic Emotions
Mayya *et al.* [34]	DCNN	CK+ +JAFFE	327+ 213 images	118	P&S	10 Folds	SVM	256×256	96.02 % 98.12 %	6 Basic Emotions + Contempt
Kandeel *et al.* [36]	CNN	JAFFE + CK+ +FER2013	213+ 593+ 35,887 images	10 123 N/A	P	5 Folds		48 × 48	100% 100% 69.32	6 Basic Emotions+ Neutral
Liu *et al.* [37]	CNN	CK+ + JAFFE +Oulu-CASIA + AR	593+ 213+ 2880 + 4000 + images	123 10 80 126	P	10 Folds		256×256	98.9% 96.8% 94.5% 98.7%	6 Basic Emotions + Contempt+ Neutral 6 Basic Emotions+ Neutral 6 Basic Emotions 13 Emotions

4. Datasets

This section presents Some popular datasets related to FER.

4.1. Karolinska Directed Emotional Face (KDEF)

The dataset of KDEF includes 4900 pictures of facial expressions of humans. The dataset comprises 70 individuals, each showing seven unique expressions taken from five different angles. The first size of each face image is 562 pixels × 762 pixels [52].

4.2. Compound Emotion Dataset (CE)

CE dataset involves 5060 images, containing 22 compound emotions (CEs) and basic emotions (BEs) of 230 subjects applied to an average age of 23, including many races. Occlusion of the Facial is reduced, without glasses. Male subjects are needed to be shaved, and all subjects are additionally asked to completely uncover their eyebrows. They are color images with 3000 pixels × 4000 pixels resolution [26].

4.3. NVIE Dataset

NVIE dataset is an infrared and characteristic apparent facial expression, which contains both posed and spontaneous expressions of in excess of 100 subjects, with illumination, given from three different directions. The dataset of pose involves the summit expressional images with and without glasses. It is labeled with six facial emotions, Arousal–Valence label, and expression intensity [53].

4.4. Japanese Female Facial Expressions (JAFFE)

JAFFE dataset includes 213 images of seven facial expressions (one neutral with six basic emotions) posed via ten Japanese females. Every one of the images is evaluated dependent on six emotional adjectives utilizing 60 Japanese subjects. The original images are 256 pixels × 256 pixels [21].

4.5. FER2013 Face Dataset

The FER2013 dataset is really supplied for a Kaggle competition. The dataset contains 35,887 face images, including 28,709 training sets, 589 test sets, and, 3589 verification sets, which are all grayscale images of 48 pixels × 48 pixels. These samples are partitioned into seven categories on an essentially average distribution, i.e., angry, disgusting, amazed, sad, neutral, happy, and fearful. Each sample in the dataset has a huge contrast in age, facial direction, or other aspects, which is near to the real-world status [10].

4.6. CMU Multi-PIE Database (Multi-PIE)

The CMU Multi-PIE dataset is utilized for research in face recognition across illumination and pose. It includes 337 subjects, captured under 19 illumination conditions and 15 viewpoints in four recording sessions for an aggregate of in excess of 750,000 images. The labels are AAM-style with somewhere in the range of 39 and 68 feature points [54].

4.7. Denver Intensity of Spontaneous Facial Action Dataset (DISFA)

DISFA dataset includes 130,000 videos of 27 subjects of various ethnicities and gender. The images are obtained at high resolution (1024 pixels × 768 pixels), and all video frames are scored in a manual way with the intensity of AU's (0–5 scale). A sum of 66 facial landmarks of each image in the dataset is labeled [55].

4.8. MMI Facial Expression Dataset

The MMI Facial Expression dataset incorporates more than 2900 videos and high-resolution still images of 75 subjects. Every one of the AUs in the videos is totally annotated. The size of the original face images is 720 pixels × 576 pixels [56].

4.9. Oulu-CASIA NIR-VIS Database (Oulu-CASIA)

The Oulu-CASIA NIR-VIS database comprises 2880 image successions with six fundamental expressions from 80 individuals somewhere in the range of 23 and 58 years of age. All expressions are captured in the frontal direction with three different conditions of illumination unique: normal, dark, and weak. Subjects were approached to make a facial expression as indicated by an expression example displayed in picture sequences. The image resolution is 320 × 240 pixels and imaging hardware works at the rate of 25 frames per second [57].

4.10. Radboud Faces dataset (RaFD)

RaFD dataset contains 1,608 images from 67 subjects with three different gaze directions, i.e., right, left, and front. Each sample is labeled with one of eight emotions: anger, contempt, fear, sadness, happiness, surprise, and disgust with neutral [58].

4.11. Expression in-the-Wild Database (ExpW)

ExpW database includes 91,793 faces downloaded utilizing Google image search. Every one of the face images was commented on in a manual way as one of the seven essential emotion categories. Non-face images were eliminated in the comment process [59].

4.12. AffectNet Dataset

AffectNet includes more than 1,000,000 images from the Internet that were gotten by querying various search engines utilizing expression related labels. It can be considered as one of the widest datasets that give facial expressions which are classified into two different emotion models (dimensional model and categorical model), of which 450,000 images have commented in manual way labels for eight fundamental emotions [60].

4.13. GEMEP-FERA

The GEneva Multimodal Emotion Portrayals (GEMEP) is an assortment of audio and video recordings featuring 10 entertainers depicting 18 emotional states, with various verbal substances and various methods of articulation. This corpus comprises in excess of 7000 audio and video emotion portrayals, clarifying 18 emotions (involving subtle emotions that are scarcely studied), depicted by 10 expert entertainers who are instructed by a professional director [61].

4.14. Binghamton-Pittsburgh 3D Dynamic Spontaneous (BP4D-Spontaneous)

BP4D-unconstrained is organized by 41 members (18 males and 23 females, their ages between 18 and 29 years old). The protocol of expression elicitation is intended to inspire the emotions of members in an effective way. Eight errands are covered with an interview process and a progression of exercises to evoke eight emotions. For each errand, there are both 2D and 3D videos. In the interim, the meta-data incorporates head pose tracked automatically, action units (AU) commented in the manual way and 2D/3D facial landmarks. The original size of each face image is 1040 pixels × 1329 pixels [62].

4.15. Broadened Cohn–Kanade Dataset (CK+)

CK+ dataset is an augmentation of the CK dataset, involving 593 video sequences and static images of seven facial expressions (one contempt and 6 essential expressions). The static images are presented in a lab circumstance, and the videos are shot in a similar circumstance. The age of 123 subjects is between 18 and 30. The resolution of the images is 640 pixels × 480 pixels and 640 pixels × 490 pixels, and the grey value is 8-bit precision [22]. Table 3 shows the important details for the most popular datasets used in the facial expression recognition field.

Table 3: Facial Expression Recognition Datasets Overview.

Dataset	Resolution	Samples	Subjects	Condition	Elicit.	Access Source	Emotions Number
KDEF [52]	562 × 762	4900 static images	70	Lab	P	http://www.emotionlab.se/kdef/	6 Basic Emotions+ Neutral
CE [26]	3000 × 4000	5060 static images	230	Lab	P	http://cbcsl.ece.ohio-state.edu/dbform_compound.html	22 Basic and Compound Emotions
JAFFE [21]	256 × 256	213 static images	10	Lab	P	http://www.kasrl.org/jaffe.html	6 Basic Emotions + Neutral
FER2013 [10]	48 × 48	35,887 static images	N/A	Web	P&S	https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data	6 Basic Emotions + Neutral
Multi-PIE [54]	N/A	755,370 static images	337	Lab	P	http://www.flintbox.com/public/project/4742/	5 Basic Emotions + Neutral and Facial landmarks
DISFA [55]	1024 × 768	130,000 stereo videos	27	Lab	P	http://www.engr.du.edu/mmahoor/DISFA.htm	Action units' intensity scale and 66 Facial landmarks
MMI [56]	720 × 576	740 static images and 2900 video sequence	75	Movies	P	http://mmifacedb.eu/	6 Basic Emotions+ Neutral and Action units.
Oulu-CASIA [57]	320 × 240	2880 image sequence	80	Movies	P	http://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA	6 Basic Emotions
RaFD [58]	N/A	1,608 static images	67	Lab	P	http://www.socsci.ru.nl:8180/RaFD2/RaFD	6 Basic Emotions + neutral and contempt
ExpW [59]	Original web images	91,793 static images	N/A	Web	P&S	http://mmlab.ie.cuhk.edu.hk/projects/socialrelation/ind	6 Basic Emotions + Neutral
AffectNet [60]	N/A	450,000 static images	N/A	Web	P & S	http://mohammadmahoor.com/databases-codes/	6 Basic Emotions + Neutral
GEMEP-FERA [61]	N/A	7000 audio-video	10	Web	S	https://gemep-db.sspnet.eu/	18 Emotions
BP4D-Spontaneous [62]	1040 × 1329	328 videos	41	Lab	S	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html	8 Emotions, Action units and Facial landmarks
CK+ [22]	640 × 480, 640 × 490	593 static images and video sequence	123	Lab	P	http://www.pitt.edu/~emotion/ck-spread.htm	6 Basic Emotions + Neutral and contempt

5. Research Challenges

In this section, we discuss and mention the main challenges that stand against the researchers during their research.

5.1. Datasets of Facial Expression

Most of the facial expression works of literature presented here focused on the main issues and challenges is in conditions of a wild environment. Most of them suggested employing deep learning approaches for handling these difficult problems such as the variance of illumination and non-frontal head pose, occultation, low intensity recognizing expression and identity bias. The deep network approach required huge data for training and the most critical challenge that deep facial expressions face is the lack of training data in terms of quantity and quality [5].

5.2. Imbalanced Distribution and Dataset Bias

Annotation of inconsistency and data bias is most common between facial expressions and datasets because of various collecting conditions and notation subjections. Recent studies evaluated their approaches using a special dataset and obtained satisfactory performance [63]. On the other hand, approaches that were evaluated using database protocol and without generalizing on unseen test data and cross-dataset performance settings are mostly deteriorated because of existing discrepancies. The main issue in facial expression is unbalancing among class distribution in facial expression because sample acquirement such as happy face can easily be annotated while it is vice versa for anger, fear, and disgust [64, 65].

5.3. Integrating Other Affective Models

Facial expression recognition and classification model had been broadly researched and acknowledged, prototypical expressions definition covers only a few parts of specific classifications and does not have abilities to capture that full repertoire of expressive attitudes for realistic interaction. Other two issues are visual attention based network and highlighting the main AUs related area to facial expression recognition task through the mechanism of attention and allowing model for learning representation of discriminative expression [66, 67].

5.4. The Effect of Multi-modal on Recognition

Human expression attitudes in real applications include encoding from various perspectives and one modality for facial expression during the progression of user generated content on social media. A huge amount of data is uploaded by users from different platforms and analyzing multimodal sentiment becomes crucial and popular in the processing of these various analyzing modalities' opinions (negative and positive) towards a certain entity. For joining effective information from various modalities, a new analysis of multimodal sentiment based on deep neural network Poria *et al.* proposed various multi sensors data fusions model [68]. Furthermore, fusion with other multimodal for example depth information from 3D models, infrared images, and physiological data becomes a crucial research direction because of the huge complementary for facial expressions and their critical advantage for applications such as human computer interaction [69-71].

5.5. Visual Privacy

Growing privacy concerns in camera-equipped systems, such as the real-time FER for smartphones, are a key roadblock. Even though a few attempts have been made in [72-74], many proposed FER methods rely on high-resolution photographs, with little or no consideration paid to safeguard their users' visual privacy. As a result, more reliable and precise privacy protection measures are required for FER systems to strike a balance between data utility and privacy.

6. Conclusion

In recent years, facial expression recognition becomes much more attractive and popular from the view of academic researchers, In the last decade, they have progressed to the point that the majority of current research has focused on facial expression recognition. This survey provided and presented an in-depth look at the latest developments and improvements in facial expression recognition techniques. We initially introduced and described background concepts of facial expression, after that, we explained the critical steps required for recognizing expression that include preprocessing, facial expression based on deep learning, and facial expression classification. In preprocessing we briefly described the essential concepts of facial expression with some critical terms that are required in the field of computer vision and image processing, then we discussed, illustrated, and summarized recent research on facial expression recognition based on deep learning with their comparison in tabular form next we illustrated and presented classification methods of facial expression, we noticed that 83% of researches used CNN and DBN techniques for recognizing emotions, whereas 17% of them used SOM and GAN with using different datasets from net and laboratory that involve various pose and spontaneous. In addition, fifteen FER datasets are represented and explained. Finally, we highlighted and described the critical issues and challenges that formed in facial expression based on deep learning with their solutions and future research direction. The intent of this survey is to give an orderly and extensive study of the work done in the field of FER and to motivate more researchers in this area.

References

[1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J Pers Soc Psychol*, vol. 17, no. 2, pp. 124-9, Feb 1971, doi: 10.1037/h0030377.

[2] P. Ekman, "Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique," *Psychol Bull*, vol. 115, no. 2, pp. 268-87, Mar 1994, doi: 10.1037/0033-2909.115.2.268.

[3] D. Matsumoto, "More evidence for the universality of a contempt expression," *Motivation and Emotion*, vol. 16, no. 4, pp. 363-368, 1992, doi: 10.1007/BF00992972.

[4] R. E. Jack, O. G. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7241-7244, 2012, doi: 10.1073/pnas.1200155109.

[5] S. Li and W. Deng, "Deep facial expression recognition: A survey," in *IEEE Transactions on Affective Computing*, 2020, doi: 10.1109/TAFFC.2020.2981446.

[6] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review," in *2011 IEEE 7th international colloquium on signal processing and its applications*, 2011, pp. 410-415, doi: 10.1109/CSPA.2011.5759912.

[7] 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*, 2017, pp. 524-528, doi: 10.1145/3136755.3143004.

[8] A. Dhall, R. Goecke, J. Joshi, J. Hoey, and T. Gedeon, "Emotiw 2016: Video and group-level emotion recognition challenges," in *Proceedings of the 18th ACM international conference on multimodal interaction*, 2016, pp. 427-432, doi: 10.1145/2993148.2997638.

[9] 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*, 2015, pp. 423-426, doi: 10.1145/2818346.2829994.

[10] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International conference on neural information processing*, 2013, pp. 117-124: Springer, doi: 10.1007/978-3-642-42051-1_16.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[12] R. S. Muhammad and M. I. Younis, "The limitation of pre-processing techniques to enhance the face recognition system based on LBP," *Iraqi Journal of Science*, vol. 58, no. 1B, pp. 355-363, 2017.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014, doi: 10.48550/arXiv.1409.1556.

[14] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[15] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology," *Dev Psychopathol*, vol. 17, no. 3, pp. 715-34, Summer 2005, doi: 10.1017/S0954579405050340.

[16] Y.L. Tian, T. Kanade, and J. F. Colin, "Recognizing action units for facial expression analysis," in *Multimodal interface for human-machine communication*: World Scientific, 2002, pp. 32-66, doi: 10.1142/9789812778543_0002.

[17] W.Y. Chang, S.H. Hsu, and J.H. Chien, "FATAUVA-Net: An integrated deep learning framework for facial attribute recognition, action unit detection, and valence-arousal estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 17-25.

[18] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry*, vol. 11, no. 10, p. 1189, 2019, doi: 10.3390/sym11101189.

[19] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115-142, 2019, doi: 10.1007/s11263-018-1097-z.

[20] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European conference on computer vision*, 2014, pp. 94-108: Springer, doi: 10.1007/978-3-319-10599-4_7.

[21] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings Third IEEE international conference on automatic face and gesture recognition*, 1998, pp. 200-205, doi: 10.1109/AFGR.1998.670949.

[22] 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 *2010 IEEE computer society conference on computer vision and pattern recognition-workshops*, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[23] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978, doi: 10.1037/t27734-000.

[24] C. Fabian 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 the IEEE conference on computer vision and pattern recognition,* 2016, pp. 5562-5570.

[25] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169-200, 1992, doi: 10.1080/02699939208411068.

[26] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proc Natl Acad Sci U S A*, vol. 111, no. 15, pp. E1454-62, Apr 15 2014, doi: 10.1073/pnas.1322355111.

[27] P. Ekman, "Darwin, deception, and facial expression," *Ann N Y Acad Sci,* vol. 1000, no. 1, pp. 205-21, Dec 2003, doi: 10.1196/annals.1280.010.

[28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, 2001, vol. 1, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[29] J. Cai, Z. Meng, A. S. Khan, Z. Li, J. O'Reilly, and Y. Tong, "Island loss for learning discriminative features in facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 302-309, doi: 10.1109/FG.2018.00051.

[30] A. Fathallah, L. Abdi, and A. Douik, "Facial expression recognition via deep learning," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 2017, pp. 745-750, doi: 10.1109/AICCSA.2017.124.

[31] G. Verma and H. Verma, "Hybrid-deep learning model for emotion recognition using facial expressions," *The Review of Socionetwork Strategies*, vol. 14, no. 2, pp. 171-180, 2020, doi: 10.1007/s12626-020-00061-6.

[32] I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, and A. Kulkarni, "Real time facial expression recognition using deep learning," in *Proceedings of International Conference on Communication and Information Processing (ICCIP)*, 2019, https://dx.doi.org/10.2139/ssrn.3421486.

[33] M. Lee, Y. K. Lee, M. T. Lim, and T. K. Kang, "Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features," *Applied Sciences*, vol. 10, no. 10, p. 3501, 2020.

[34] V. Mayya, R. M. Pai, and 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.

[35] O. K. Oyedotun, G. Demisse, A. El Rahman Shabayek, D. Aouada, and B. Ottersten, "Facial expression recognition via joint deep learning of rgb-depth map latent representations," in *Proceedings of the IEEE international conference on computer vision workshops*, 2017, pp. 3161-3168.

[36] A. Kandeel, M. Rahmanian, F. Zulkernine, H. M. Abbas, and H. Hassanein, "Facial expression recognition using a simplified convolutional neural network model," in *2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2021, pp. 1-6, doi: 10.1109/ICCSPA49915.2021.9385739.

[37] J. Liu, Y. Feng, and H. Wang, "Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning," in *IEEE Access*, vol. 9, pp. 69267-69277, 2021, doi: 10.1109/ACCESS.2021.3078258.

[38] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment," *IEEE Trans Image Process*, vol. 29, pp. 6574-6589, May 8 2020, doi: 10.1109/TIP.2020.2991549.

[39] X. Zhao, X. Shi, and S. Zhang, "Facial expression recognition via deep learning," *IETE technical review*, vol. 32, no. 5, pp. 347-355, 2015, doi: 10.1080/02564602.2015.1017542.

[40] C. Ou, J. Yang, Z. Du, X. Zhang, and D. Zhu, "Integrating cellular automata with unsupervised deep-learning algorithms: A case study of urban-sprawl simulation in the Jingjintang urban agglomeration, China," *Sustainability*, vol. 11, no. 9, p. 2464, 2019, doi: 10.3390/su11092464.

[41] W. Guo, H. Yang, and Z. Liu, "Deep neural networks for depression recognition based on facial expressions caused by stimulus tasks," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2019, pp. 133-139, doi: 10.1109/ACIIW.2019.8925293.

[42] S. R. Chiluveru and M. Tripathy, "A real-world noise removal with wavelet speech feature," *International Journal of Speech Technology*, vol. 23, no. 3, pp. 683-693, 2020, doi: 10.1007/s10772-020-09748-1.

[43] H. Hussein, F. Angelini, M. Naqvi, and J. A. Chambers, "Deep-learning based facial expression recognition system evaluated on three spontaneous databases," in *2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC)*, 2018, pp. 270-275, doi: 10.1109/ISIVC.2018.8709224.

[44] Y. Tang, "Deep learning using linear support vector machines," *arXiv preprint arXiv:1306.0239*, 2013, doi: 10.48550/arXiv.1306.0239.

[45] A. Dapogny and K. Bailly, "Investigating deep neural forests for facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 629-633, doi: 10.1109/FG.2018.00099.

[46] P. Kontschieder, M. Fiterau, A. Criminisi, and S. R. Bulo, "Deep neural decision forests," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1467-1475.

[47] J. Donahue *et al.*, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, 2014, pp. 647-655: PMLR.

[48] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806-813.

[49] D. Acharya, Z. Huang, D. Pani Paudel, and L. Van Gool, "Covariance pooling for facial expression recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 367-374.

[50] N. Otberdout, A. Kacem, M. Daoudi, L. Ballihi, and S. Berretti, "Deep covariance descriptors for facial expression recognition," *arXiv preprint arXiv:1805.03869*, 2018, doi: 10.48550/arXiv.1805.03869.

[51] K. Zhang, Y. Huang, Y. Du, and L. Wang, "Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks," *IEEE Trans Image Process*, vol. 26, no. 9, pp. 4193-4203, Sep 2017, doi: 10.1109/TIP.2017.2689999.

[52] D. Lundqvist, A. Flykt, and A. F. D. O. C. Öhman, "Psychology section Karolinska Institutet “The Karolinska directed emotional faces (KDEF)," *Cognition and emotion*, vol. 91, p. 630, 1998.

[53] S. Wang *et al.*, "A natural visible and infrared facial expression database for expression recognition and emotion inference," in *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 682-691, 2010, doi: 10.1109/TMM.2010.2060716.

[54] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Proc Int Conf Autom Face Gesture Recognit*, vol. 28, no. 5, pp. 807-813, May 1 2010, doi: 10.1016/j.imavis.2009.08.002.

[55] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," in *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151-160, 2013, doi: 10.1109/T-AFFC.2013.4.

[56] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE international conference on multimedia and Expo*, 2005, p. 5 pp.-, doi: 10.1109/ICME.2005.1521424.

[57] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. PietikäInen, "Facial expression recognition from near-infrared videos," *Image and vision computing*, vol. 29, no. 9, pp. 607-619, 2011, doi: 10.1016/j.imavis.2011.07.002.

[58] O. Langner *et al.*, "Presentation and validation of the Radboud Faces Database," *Cognition and emotion*, vol. 24, no. 8, pp. 1377-1388, 2010, doi: 10.1080/02699930903485076.

[59] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "From facial expression recognition to interpersonal relation prediction," *International Journal of Computer Vision*, vol. 126, no. 5, pp. 550-569, 2018, doi: 10.1007/s11263-017-1055-1.

[60] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," in *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18-31, 2017, doi: 10.1109/TAFFC.2017.2740923.

[61] M. F. Valstar, M. Mehu, J. Bihan, M. Pantic, and K. Scherer, "Meta-Analysis of the First Facial Expression Recognition Challenge," *IEEE Trans Syst Man Cybern B Cybern*, vol. 42, no. 4, pp. 966-79, Aug 2012, doi: 10.1109/TSMCB.2012.2200675.

[62] X. Zhang *et al.*, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692-706, 2014, doi: 10.1016/j.imavis.2014.06.002.

[63] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and vision Computing*, vol. 27, no. 6, pp. 803-816, 2009, doi: 10.1016/j.imavis.2008.08.005.

[64] S. Li and W. Deng, "Deep emotion transfer network for cross-database facial expression recognition," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3092-3099, doi: 10.1109/ICPR.2018.8545284.

[65] X. Wei, H. Li, J. Sun, and L. Chen, "Unsupervised domain adaptation with regularized optimal transport for multimodal 2d+ 3d facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 31-37, doi: 10.1109/FG.2018.00015.

[66] P. Ekman, "Facial action coding system (FACS)," *A Human Face, Salt Lake City*, 2002.

[67] E. L. Rosenberg and P. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 2020.

[68] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 2539-2544.

[69] F. Ringeval *et al.*, "Avec 2017: Real-life depression, and affect recognition workshop and challenge," in *Proceedings of the 7th annual workshop on audio/visual emotion challenge*, 2017, pp. 3-9, doi: 10.1145/3133944.3133953.

[70] M. Soleymani *et al.*, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3-14, 2017, doi: 10.1016/j.imavis.2017.08.003.

[71] M. Valstar *et al.*, "Avec 2016: Depression, mood, and emotion recognition workshop and challenge," in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 3-10, doi: 10.1145/2988257.2988258.

[72] J. Chen, J. Konrad, and P. Ishwar, "Vgan-based image representation learning for privacy-preserving facial expression recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 1570-1579.

[73] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232-243, 2005, doi: 10.1109/TKDE.2005.32.

[74] Y. Rahulamathavan and M. Rajarajan, "Efficient privacy-preserving facial expression classification," in *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 3, pp. 326-338, 2015, doi: 10.1109/TDSC.2015.2453963.