IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

AN OVERVIEW OF DEEP LERNING-BASED AUTOMATED FACIAL EXPRESSION RECOGNITION

[1]Priya Patel, [2]Twisha Patel

[1]Assistant Professor , [2]Assistant Professor,
[1]School of Engineering
[1]P P Savani University, Dhamdod, Kosamba, Gujarat, India.

Abstract

One of the areas of symmetry and a significant and promising area of computer vision and artificial intelligence is facial expression recognition (FER), which serves as the principal processing mechanism for non-verbal intentions. A thorough and organized summary of recent developments in FER is provided in this survey. The existing FER methods are first divided into two major categories, namely conventional approaches and deep learning-based approaches. We offer a broad structure of a traditional FER approach from a methodological perspective and examine the potential technologies that could be used in each component to emphasize the differences and similarities. Regarding deep learning-based approaches, four different types of cutting-edge FER approaches based on neural networks are described and examined. In addition, we summarize four FER-related dataset characteristics that may affect the selection and processing of FER techniques, as well as seventeen regularly used FER datasets. Following performance comparisons of several FER approaches on the benchmark datasets, evaluation methods and metrics are provided in the later portion to demonstrate how to evaluate FER algorithms. At the conclusion of the survey, we list certain issues and openings that require attention in the future.

Keywords

Facial expression recognition; Feature extraction; Classification; Deep learning

1. Introduction

One of the most significant components of biometry is facial expression, which has emerged over the past ten years as a new and active area of study because of how well it translates people's emotional states. Face expression continues to be the most expressive way for people to express their emotions because of its high level of directness, friendliness, convenience, and robustness. This latter analysis can be done through other features such as: voice [11], body gestures, social and contextual parameters of the situation [25], among others.
Due to the enormous amount of attention, it received, facial expression recognition (FER) has a wide range of applications nowadays. It is mostly utilized in human-machine interface (HCI) applications, including robotics, virtual reality, interactive gaming, and digital entertainment. Additionally, it is employed in applications for surveillance and law enforcement as well as mood and behavioural analysis in the medical field (Autism [17], mental illness [48], and pain evaluation [36]).

In his book The Expression of the Emotions in Man and Animals [7], Darwin established facial expression as a topic of study that has since been explored by a wide range of experts. Paul Ekman introduced the six fundamental emotions—happiness, sadness, anger, disgust, fear, and surprise—as well as the neutral emotion, which is considered in most works. These emotions became universal among people in the last few decades and are now regarded as the foundation of almost all research in this area [10].

The extremely complex Automated FER System (AFERS or FERS) allows robots to discern emotions automatically and without human assistance. This system takes in photos with faces as input, processes those images in the ways that will be discussed in the next section, and outputs the identified emotion. Even though it appears to be a pretty straightforward process for people, machine learning experts find it to be highly difficult. In this research, we first attempt to provide a succinct description of the FER principles enhanced by what scientists have learned over time, and then we present a comparison of the most recent deep learning-based efforts in the field.

The remainder of the essay is structured as follows: The background review of FER principles is presented in the following section. A thorough investigation of FER difficulties utilizing deep learning is presented in Section III. Section IV presents the observations and debate, followed by Section V's conclusion and suggested next steps.

2. Background Review

Figure 1 depicts the workflow of the FER system, in which we can see that the analysis is conducted in three distinct stages, depending on whether the inputs to the system are photos or dynamic sequences.

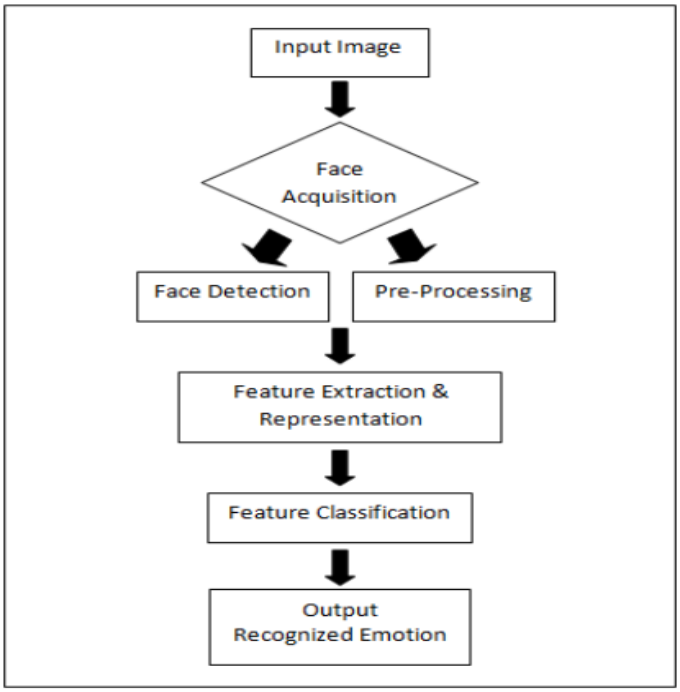

Fig. 1. Facial Expression Recognition System Workflow

2.1 Face Acquisition

The algorithm of Viola and Jones [55] is the most widely used method for face detection. Over the years, numerous techniques have been developed to detect faces in arbitrary scenes [19, 31, 42, 43, 46, 49]. Some of these methods can only detect faces in the frontal view, while others can detect faces in multi-views, such as side views.

Over time, it was believed that this stage was a pre-processing stage that included the adjustments we make to the various input images before feeding them to the FES. Pre-processing is based on suppressing undesired distortions and boosting image attributes to improve the quality of the input data the system will operate with. It may also involve image resizing, denoising, rotation correction, etc.

2.2 Feature Extraction and Representation

The process of extracting features from the input data follows the discovery of the face. We may classify the used methods for feature extraction based on the diversity of features that can be recovered during this procedure. Two major categories can be found in the literature for facial features.

Geometric features that describe the position and shape of the lips, eyes, brows, and nose when making a facial expression.

The Facial Action Coding System (FACS), developed by Ekman [13], contains 44 Action Units (AU), and another 20 were added in 2002 [10]. Each AU describes a group of facial muscles that cooperate to carry out movements associated with a particular facial expression. Geometric methods use a feature vector made up of facial feature points that have been extracted; some examples of this method's use may be found in [14, 44, 60, 16, 27, 51].

In contrast, Appearance Features depict the variations in the skin's texture or color of the face without taking into account how the muscles move. Although methods based on appearance elements are thought to take longer to analyze, studies have proven that they yield superior results because they examine pixel intensity, texture edges, color patterns, and wrinkles and furrows of the entire face or just a small portion of it. The local binary pattern (LBP) and its extensions are the appearance characteristics techniques that have achieved the most success [1, 26]. For other examples, see [51, 57, 16, 24, 20].

2.3 Feature Classification and Emotion Recognition

The goal of this phase is to select just the most discriminative features from the enormous set of features produced by the previous phase in order to speed up processing and increase the reliability of the results. Many new and innovative techniques have been created over the years by researchers; some of the most well-known include Support Vector Machines (SVM) [62], Bayes classifier [6], Fuzzy Techniques [3], Feature Selection [8], Artificial Neural Networks (ANN) [23], and others [64]. A FER system's whole process flow is displayed in

Even though recognition performances are getting better all the time, there are still some limitations in terms of efficiency (time, computational, and space complexity), accuracy (subjectivity, occlusion, pose, low resolution, scale, illumination variation, spontaneous vs. posed expressions), and efficiency (time). Researchers began looking for new approaches or concentrated on improving existing ones to get over these constraints, which led to the development of Deep Neural Networks (DNN), which are based on conventional ANN.

2.4 Deep Neural Networks

Deep Neural Networks (DNN) have the exact same structure as classical ANN; the only difference is that they use multiple and deeper hidden layers instead of just one, which increases the chance of learning bigger, more complex, and abstract features. Artificial Neural Networks are based on the idea of mathematically representing information processing in human brains; the majority of them rely on features selected or created by humans. The key feature of DNN is that they build a hierarchy of representations, each one becoming increasingly more abstract than the one before it, to help complete the assigned task [45]. Consider visual identification as an example. Instead of using features extracted using handcrafted off-the-shelf techniques like appearance and geometric based algorithms, one can use the entire image as an input.

The most popular and rapidly expanding machine learning research area over the past ten years has been deep neural networks (DNN).

Deep belief networks (DBN) and convolutional neural networks (CNN) are two of the deep learning architectures that are frequently used in the field of FER. In this paper, we are interested in CNN, which was first introduced in 1997 [28]. Its structure primarily relies on the alternate use of two types of basic layers, referred to as convolutional layers (C layers) and sub-sampling layer (S layers), commonly referred to as the Pooling layer, and at the end a fully [22] contains more thorough information about CNN.

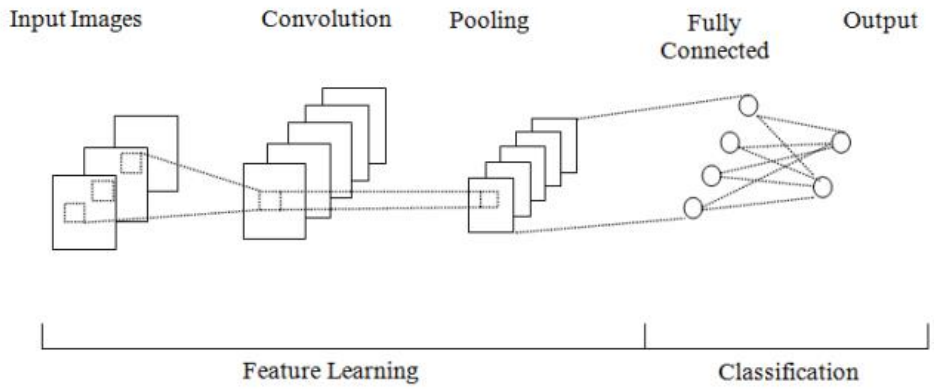

Fig. 2. Convolutional Neural Network Architecture

3. Deep Learning for Facial Expression Recognition

A detailed survey on the traditional procedures can be found in [11, 44, 27, 26, 20]. Over the past 20 years, a number of approaches have been created to address the issues of FER in computer vision, and every single one of them has improved the recognition performance. This section examines the most current and frequently mentioned deep learning based algorithms that have achieved high accuracy in the last five years. CNN in particular has played a significant role in the recent advancements in the area.

Using the real-image dataset of FER2013, which contains real facial photographs for the seven facial emotions, a recent study [5] examined how well CNN could improve FER accuracy. Their solution was created using the machine learning platform TensorFlow. Three convolutional layers, two max pooling layers, two fully connected layers, and dropouts at various levels make up the design. The authors also provided information on the architecture's various parameters. The technique achieved 91,12%, which is better than CNN augmentation setting.

Another model was proposed in[52], in which faces are identified from images in datasets using a Cascade Classifier. The obtained image is then converted to grayscale level, after which it is normalized, and finally enhanced using the Image Data Generator offered by the Keras API. The enhancement techniques used are horizontal flip, rotation, rescale, shear, and zoom. The CNN is then fed the updated dataset in order to forecast the emotion. Three convolutional layers with 32, 64, and 128 filters each, a kernal size of 3*3, and four fully connected layers made up their architecture. Multiple databases, including CK+, FER2013, MUG, KDEF & AKDEF, and Kinface W I & II, were used to collect the data. It needed only 120 epochs to reach an accuracy of 96,24% whereas CNN without data augmentation needed 260 epochs to reach only 92,95%.

A single deep CNN with convolutional layers and residual blocks was suggested by the authors in [21]. The system's accuracy was 93,24% and 95,23%, respectively, after training on the CK+ and JAFFE datasets. Combining CNN with residual blocks appears to improve overall outcomes and address the issue of facial expression recognition and categorization.

The model suggested in [37] has a three-layer deep learning architecture, with the first two layers consisting of extracting two different types of features—geometric and appearance-based (LBP)—and combining them. The authors claim their approach to be the first of its kind in this sector. The third layer classifies combined features using a self-organizing map (SOM) based classifier that combines the advantages of both supervised and unsupervised training algorithms. The system achieved respective scores of 98,95% and 97,55% on the two databases CK+ and MMI during validation.

In [34], authors suggested a technique that coupled certain picture preparation processes with CNN. The various steps are: (i) rotation correction; (ii) cropping; (iii) down-sampling; (iii) ensuring that the location of the facial component is the same in all images; and (iiii) intensity normalization (makes the brightness and contrast of all images the same in order to reduce the complexity of the network). A set of photos featuring faces, the locations of their eyes, and their expressions are used as the input for the pre-processing phase. The output is then sent to a CNN that has two convolutional layers, two sub-sampling layers, and a fully connected layer. The system tested

on three commonly used databases (CK+, JAFFE, and BV-3DFE) and evaluated the effects of each pre-processing step on the accuracy rate. It achieved the best accuracy on CK+ with 96,76%; the accuracy increased with the addition of each pre-processing phase.

[47] presents another system made up of four modules: input, pre-processing, recognition, and output. It was evaluated on the JAFFE and CK+ databases, and when compared to the KNN method, it earned scores of 76,7442% and 80,303%, respectively. Face detection was integrated into the system utilizing OpenCV's Haar-like features and histogram equalization. The four layers of the system architecture—two convolutional and two sub-sampling layers—are followed by a SoftMax classifier for multiclassification. Following the application of Histogram Equalization to make gray scale values and contrast more uniform across all images, the authors concluded that CNN gives better results in solving the problem of facial expression recognition when compared to methods that are based on KNN. The authors preferred Haar-like features to capture the useful portion of facial expression and remove most of the background information.

The authors of [12] created a CNN with variable depth, allowing the user to select the quantity of convolutional and fully connected layers as well as batch normalization dropout and maximum pooling layers. Users can also specify the quantity of filters, strides, and zero padding, with default values available in the event that they do not. The test and training were conducted using a database made available by the Kaggle Website and comprised of grayscale photos of faces. The authors constructed the system in Torche with high GPU capacity integration. They use a method that combines features obtained from the convolutional layer with those from the Histogram of Oriented Gradient (HOG) and feeds these characteristics into fully connected layers. In order to determine which architecture is optimal for recognition, the system was trained using each of them for 30 epochs with a batch size of 128. The first shallow architecture, which had a validation score of 55% and a test score of 54%, consisted of two convolutional layers, one fully connected layer, and a hidden layer made up of 512 neurons. The second CNN is more complex, featuring four convolutional layers, two fully connected layers, and a first hidden layer made up of 256 neurons and a second layer of 512 neurons. It achieved 65% on the validation set and 64% on the test set.

Although the authors trained more complex designs with five and six convolutional layers, the outcomes were unsatisfactory. They also tried combining HOG features with features from the convolutional layers of both architectures, but it appears that the hybrid features had no effect on the model's accuracy. As a result, they came to the conclusion that a not-very-deep architecture is adequate to achieve good accuracy and address the issue.

A real-time model was published in [2], where the authors provided two distinct architectures for a real-time vision system that recognizes faces, classifies them based on CNN, first by gender, then by emotion.

The first one depends on completely removing the layers that are fully connected. The second also gets rid of completely connected layers and combines residual models and depth-wise separable convolutions. It takes 0.22 milliseconds to classify the gender and emotions. On the IMDB gender dataset, the first design classified gender with an accuracy of 96%, while the second one classified gender with an accuracy of 95%. Both architectures classified emotions with an accuracy of 66% on the FER2013 dataset.

The work presented [38] uses a deep multi-layer network for saliency prediction to obtain intensity maps, which are then fed into a CNN-based AlexNet model. The model was trained on the ILSVRC2012 dataset, which includes the posed databases CFEE and RaFD. Faces were cropped using the Viola and Jones algorithm.

a 3D creation When extracting spatial and temporal data from video pictures in video sequences, ResNet architecture, which was first introduced in [18], is the main goal.

Facial Landmarks were added as a second input to help extract the main facial features (eyebrows, lip corners, and eyes), and they improved their method by using it. The model was tested using subject-independent tasks in which each database was divided into training and validation sets and cross validation tasks in which the system was trained on one database and tested on another. The databases used were CK+, MMI, FERA, and DISFA. In the subject independent task, they attained the following accuracy levels: 93,21%, 77,50%, 77,42% and 58,00%. They also attained the following accuracy levels on the cross validation task: 67,52%, 54,76%, 41,93%, and 40,51%. Although it appears that subject-independent evaluation produces better results, the authors claim that their method beats several cutting-edge techniques in both tasks.

Each of the two convolutional layers in the DNN architecture described in [40] was followed by a maximum pooling layer, and then inception layers. The system classifies input photos into the six main emotions using data

from seven distinct databases; the databases' subject-independent accuracy ratings are: CMU MultiPIE 94,7%, MMI 77,9%, CK+ 93,2%, DISFA 55,0%, FERA76,7%, SFEW 47,7%, and FER2013 66,4%. The authors noted that using inception layers with CNN rather than traditional CNN increases classification accuracy on both subject independent and cross database evaluation tasks, and they verified that the results in the subject independent tests were either comparable to or better than the state of the art at the time.

The final network was created by connecting all the subnets together. In [32], authors suggested a CNN model based on structured subnets. Each subnet is a compact CNN model trained individually. The three planned subnets each have three, four, or five convolutional layers, with the other parameters being the same. The system's overall process consists of two stages: i) feeding the input data to the subsets. ii) Determine the emotion based on step i's final outcome. A fully connected layer is added at the end to combine the information gathered by the three subsets, and a Softmax layer is then utilized as the output layer for the entire network. The accuracy of the entire model was 65,03%, and the best individual subnet got 62,44%.

The system suggested in the work published in [29] employs a webcam and can distinguish persons and faces at a distance of 2-3 m in a TV setting, as well as human emotions based on the six main emotions. Their model was created using CAFFE and is based on CUDA. It has two fully connected layers, two convolutional layers, and a rectified liner unit (ReLu) activation function. The system was tested using real-time images after being trained using FER2013. It demonstrated good accuracy and was able to identify the six primary emotions as well as three secondary ones (excited, bored, and concentration). As a result, it appears to be successful and can be applied to a variety of fields, including interactive TV, intelligent vehicles, and others.

In order to capture the local appearance variations brought on by facial expressions, authors of [33] proposed an action unit-inspired deep network. They built a convolutional layer, a max-pooling layer, and a micro action pattern representation, combined the various maps, and at the end used a multi-layer learning process to produce high level features that are used for the expression recognition. The method appears to have produced good accuracy on three databases—CK+, MMI, and SFEW—when authors did cross-database validation, compared their results with those of state-of-the-art techniques, and also with those from handcrafted techniques.

The algorithm used in [30] is a CNN-based one that can identify emotions from static facial photographs. In order to reduce network effort and complexity, their solution involves sending pre-processed images to CNN as input rather than RGB images. The images are then converted to LBP codes to account for illumination variations. The Emotion Recognition in the Wild Challenge (EmotiW 2015) and Static Facial Expression Recognition sub-Challenge (SFEW) are used to assess the model, which was trained on CASIA webface pictures.This approach at the time produced results that were 15,36% better than the baseline findings and increased the system's performance by 40% by using inputs other than RGB images to feed the CNN network.

The reader may refer to [56, 59, 58, 61, 63, 61, 50, 4] for further works and approaches in the field of image processing utilizing convolutional neural networks; these publications demonstrate the effectiveness of CNN in a variety of fields, including: face recognition and gender recognition. Since the results of certain older investigations have been superseded by more contemporary ones, it is difficult to evaluate them now.

4. Comparative Research and Analysis

We discovered the following things after reading a small number of publications, primarily the most recent research in the field of facial expression identification, specifically facial expression recognition based on deep learning and particularly on convolutional neural networks:

The majority of the works that are being presented are based on the CNN architecture with various depth sizes. Because of its classification ability, CNN consistently demonstrates its effectiveness in tackling image recognition challenges, even if it necessitates a sizable training dataset and a powerful GPU.

Pre-processing operations like face detection, cropping, illumination normalization, resizing, and flipping, to name a few, are carried out in almost every work; they can help increase accuracy, reduce architecture complexity, and shorten training time.

Data augmentation is a crucial step to enlarge and expand the dataset by exercising various preset processes such as horizontal and vertical flipping, rescaling, zooming, and rotating. Convolotional neural networks require a significant quantity of data, but if the dataset provided is not large enough, this is a problem.

Although deep learning is particularly popular for tackling facial expression recognition problems, this does not necessarily imply that deepening the architecture would boost recognition rates [12], occasionally a shallow design may suffice.

Table 1. List of Presented Papers

Paper	Year	Dataset	Accuracy	Architecture
[5]	2019	FER2013	91,12%	3 C layers, 2 max P layers, 2 FC layers, dropouts at different levels
[52]	2019	Multiple Databases	96,24%	3 C layers, 4 FC layers, data augmentation
[21]	2019	CK+ JAFFE	93,24% 95,24%	C layers combined with residual blocks
[37]	2018	CK+ MMI	98,95% 97,55%	3 C layers, geometric and appearance features, classification is done using self-organizing map (SOM)
[34]	2017	CK+, JAFFE , BV3DFE	96,76%	2 C layers, 2 sub-sampling layers, FC layer, pre-processing steps
[47]	2017	JAFFE CK+	76,7442% 80,303%	2 C layers, 2 sub sampling layers, Softmax classifier, Haar like features
[12]	2017	Kaggle Website	64%	CNN combined with HOG, Architecture 1 (2 C layers, 1 FC layer), Architecture 2 (4 C layers, 2 FC layers), parameters defined by the user
[2]	2017	FER2013	66%	Gender and emotion classification, Architeture 1 (eliminating FC layers), Architecture 2 (elimination combined with residual models)
[38]	2017	CFEE RaFD	95,71% 74,79%	CNN architecture based on AlexNet
[18]	2017	CK+ , MMI FERA , DISFA	93,21% ; 77,50% 77,42% ; 58,00%	CNN combined with facial landmarks
[40]	2016	CMU MultiPIE, MMI CK+, DISFA FERA, SFEW, FER2013	94,7% ; 77,9% 93,2% ; 55,0% 76,7% ; 47,7%; 66,4%	2 C layers, max pooling layer, inception layer
[29]	2016	FER2013	-	3 C layers, 2 FC layers, rectified liner unit (ReLu) activation function
[33]	2015	CK+, MMI SFEW	93,70% ; 75,85% 30,14%	1 C layer, 1 max pooling layer, micro action pattern representation
[30]	**2015**	CASIA webface images EmotiW	15,36% improvement	CNN with pre-processed images

		2015, SFEW		

Cross validation is intended for simple models with few parameters, but for CNN known to have huge number of parameters performing a cross validation will be extremely time consuming and therefore decrease the p-value. All the performed validation tasks use subject independent and cross database validation, they both give good results, but for most of the presented works it seems that the highest accuracy is obtained when performing a subject independent validation.

In order to demonstrate the effectiveness of their system in real-time applications like Human Computer Interaction (HCI) evaluation and advertisement services, some studies employed real-time photos captured during the test phase of real-time face acquisition systems.

The majority of the datasets gathered by the authors are either one or a combination of the extended Cohn Kanade + (CK+) [35] databases, which contain posed, spontaneous, and smiling photos. The Japanese Female Facial Expression (JAFFE) [39] exclusively includes photos that have been staged. The Facial Expression Recognition 2013 dataset (FER2013) [15] was developed by searching for pictures of faces that represented various emotions using the Google image search API. Check [41] for further information on the available databases in the area.

We concentrated on looking into deep learning-based research from the previous five years. Based on the results we found in various articles, it appears that deep learning is the new approach to the problem of recognizing facial and emotional expressions because various authors contrasted their findings with those of cutting-edge techniques.

The different datasets used in each research, as well as the degrees of accuracy attained by each method, are listed in Table 1 along with the several presented studies, arranged by year of publication. It appears that CNN architectures and more generally deep learning satisfy the objective of recognizing facial expressions quite well given the increase in accuracy gained using CNN models over time.

5. Conclusion and Future Work

In this study, we looked at the most current and frequently referenced papers in the field of FER as identified by Google Scholer; the majority of these works are based on convolutional architectures and deep learning architectures. We noticed that deep learning methods are gaining popularity among researchers as a result of the latter's success in achieving high accuracy over the previous five years. As a result, deep learning is regarded as the new generation for solving FER problems due to its effectiveness in feature extraction and classification tasks. We are currently working on creating a deep architecture based on CNN that will surpass the accuracy of the present day.

References

[1] Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence (12), 2037–2041 (2006)

[2] Arriaga, O., Valdenegro-Toro, M., Pl¨oger, P.: Real-time convolutional neural networks for emotion and gender classification. arXiv preprint arXiv:1710.07557 (2017)

[3] Chakraborty, A., Konar, A., Chakraborty, U.K., Chatterjee, A.: Emotion recognition from facial expressions and its control using fuzzy logic. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 39(4), 726–743 (2009)

[4] Chang, W.J., Schmelzer, M., Kopp, F., Hsu, C.H., Su, J.P., Chen, L.B., Chen, M.C.: A deep learning facial expression recognition based scoring system for restaurants. In: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). pp. 251–254. IEEE (2019)

[5] Christou, N., Kanojiya, N.: Human facial expression recognition with convolution neural networks. In: Third International Congress on Information and Communication Technology. pp. 539–545. Springer (2019)

[6] Cohen, I., Sebe, N., Gozman, F., Cirelo, M.C., Huang, T.S.: Learning bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. vol. 1, pp. I–I. IEEE (2003)

[7] Darwin, C., Prodger, P.: The expression of the emotions in man and animals. Oxford University Press, USA (1998)

[8] Dash, M., Liu, H.: Feature selection for classification. Intelligent data analysis 1(1- 4), 131–156 (1997)

[9] Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. Journal of personality and social psychology 17(2), 124 (1971)

[10] Ekman, R.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)

[11] El Ayadi, M., Kamel, M.S., Karray, F.: Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition 44(3), 572– 587 (2011)

[12] Eusebio, J.M.A.: Convolutional neural networks for facial expression recognition (2016)

[13] Friesen, E., Ekman, P.: Facial action coding system: a technique for the measurement of facial movement. Palo Alto 3 (1978)

[14] Ghimire, D., Lee, J.: Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors 13(6), 7714–7734 (2013)

[15] Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., et al.: Challenges in representation learning: A report on three machine learning contests. In: International Conference on Neural Information Processing. pp. 117–124. Springer (2013)

[16] Happy, S., Routray, A.: Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing 6(1), 1–12 (2014)

[17] Harms, M.B., Martin, A., Wallace, G.L.: Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychology review 20(3), 290–322 (2010)

[18] Hasani, B., Mahoor, M.H.: Facial expression recognition using enhanced deep 3d convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 30–40 (2017)

[19] Heiselet, B., Serre, T., Pontil, M., Poggio, T.: Component-based face detection. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)

[20] Jafri, R., Arabnia, H.R.: A survey of face recognition techniques. Jips 5(2), 41–68 (2009).

[21] Jain, D.K., Shamsolmoali, P., Sehdev, P.: Extended deep neural network for facial emotion recognition. Pattern Recognition Letters 120, 69–74 (2019)

[22] Karpathy, A., et al.: Cs231n convolutional neural networks for visual recognition. Neural networks 1 (2016)

[23] Kobayashi, H., Hara, F.: Recognition of six basic facial expression and their strength by neural network. In: [1992] Proceedings IEEE International Workshop on Robot and Human Communication. pp. 381–386. IEEE (1992)

[24] Koelstra, S., Pantic, M., Patras, I.: A dynamic texture-based approach to recognition of facial actions and their temporal models. IEEE transactions on pattern analysis and machine intelligence 32(11), 1940–1954 (2010)

[25] Kosti, R., Alvarez, J.M., Recasens, A., Lapedriza, A.: Emotion recognition in context. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1667–1675 (2017)

[26] Kristensen, R.L., Tan, Z.H., Ma, Z., Guo, J.: Binary pattern flavored feature ex-tractors for facial expression recognition: An overview. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 1131–1137. IEEE (2015)

[27] Kumari, J., Rajesh, R., Pooja, K.: Facial expression recognition: A survey. Procedia Computer Science 58, 486–491 (2015)

[28] Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: A convolutional neural-network approach. IEEE transactions on neural networks 8(1), 98– 113 (1997)

[29] Lee, I., Jung, H., Ahn, C.H., Seo, J., Kim, J., Kwon, O.: Real-time personalized facial expression recognition system based on deep learning. In: 2016 IEEE Inter- national Conference on Consumer Electronics (ICCE). pp. 267–268. IEEE (2016)

[30] Levi, G., Hassner, T.: Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of the 2015 ACM on inter- national conference on multimodal interaction. pp. 503–510. ACM (2015)

[31] Li, S.Z., Zou, X., Hu, Y., Zhang, Z., Yan, S., Peng, X., Huang, L., Zhang, H.: Real-time multi-view face detection, tracking, pose estimation, alignment, and recognition. IEEE CVPR Demo Summary (2001)

[32] Liu, K., Zhang, M., Pan, Z.: Facial expression recognition with cnn ensemble. In:2016 international conference on cyberworlds (CW). pp. 163–166. IEEE (2016)

[33] Liu, M., Li, S., Shan, S., Chen, X.: Au-inspired deep networks for facial expression feature learning. Neurocomputing 159, 126–136 (2015)

[34] Lopes, A.T., de Aguiar, E., De Souza, A.F., Oliveira-Santos, T.: Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recognition 61, 610–628 (2017)

[35] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The ex-tended cohnkanade 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. pp. 94–101. IEEE (2010)

[36] Lucey, P., Cohn, J.F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., Prkachin, K.M.: Automatically detecting pain in video through facial action units. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 41(3), 664– 674 (2010)

[37] Majumder, A., Behera, L., Subramanian, V.K.: Automatic facial expression recognition system using deep network-based data fusion. IEEE transactions on cybernetics 48(1), 103–114 (2016)

[38] Mavani, V., Raman, S., Miyapuram, K.P.: Facial expression recognition using visual saliency and deep learning. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2783–2788 (2017)

[39] Michael, J., Lyons, M.K., Gyoba, J.: Japanese female facial expressions (jaffe). Database of digital images (1997)

[40] Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter conference on applications of computer vision (WACV). pp. 1–10. IEEE (2016)

[41] Mollahosseini, A., Hasani, B., Mahoor, M.H.: Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing 10(1), 18–31 (2017)

[42] Pentland, A., Moghaddam, B., Starner, T., et al.: View-based and modular eigenspaces for face recognition (1994)

[43] Rowley, H.A., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Transactions on pattern analysis and machine intelligence 20(1), 23–38 (1998)

[44] Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L.: Static and dynamic 3d facial expression recognition: A comprehensive survey. Image and Vision Computing 30(10), 683–697 (2012)

[45] Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks 61, 85–117 (2015)

[46] Schneiderman, H., Kanade, T.: A statistical approach to 3D object detection applied to faces and cars. Carnegie Mellon University, the Robotics Institute (2000)

[47] Shan, K., Guo, J., You, W., Lu, D., Bie, R.: Automatic facial expression recognition based on a deep convolutional-neural-network structure. In: 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA). pp. 123–128. IEEE (2017)

[48] Sprengelmeyer, R., Young, A., Mahn, K., Schroeder, U., Woitalla, D., Bu¨ttner, T., Kuhn, W., Przuntek, H.: Facial expression recognition in people with medicated and unmedicated parkinsons disease. Neuropsychologia 41(8), 1047–1057 (2003)

[49] Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Transactions on pattern analysis and machine intelligence 20(1), 39–51 (1998)

[50] Tang, J., Zhou, X., Zheng, J.: Design of intelligent classroom facial recognition based on deep learning. In: Journal of Physics: Conference Series. vol. 1168, p. 022043. IOP Publishing (2019)

[51] Tian, Y., Kanade, T., Cohn, J.F.: Facial expression recognition. In: Handbook of face recognition, pp. 487–519. Springer (2011)

[52] Uddin Ahmed, T., Hossain, S., Hossain, M.S., Ul Islam, R., Andersson, K.: Facial expression recognition using convolutional neural network with data augmentation. In: Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) (2019)

[53] Valstar, M.F., Almaev, T., Girard, J.M., McKeown, G., Mehu, M., Yin, L., Pantic, M., Cohn, J.F.: Fera 2015-second facial expression recognition and analysis challenge. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). vol. 6, pp. 1–8. IEEE (2015)

[54] Valstar, M.F., Jiang, B., Mehu, M., Pantic, M., Scherer, K.: The first facial expression recognition and analysis challenge. In: Face and Gesture 2011. pp. 921–926. IEEE (2011)

[55] Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision 57(2), 137–154 (2004)

[56] Wu, Y., Hassner, T., Kim, K., Medioni, G., Natarajan, P.: Facial landmark detection with tweaked convolutional neural networks. IEEE transactions on pattern analysis and machine intelligence 40(12), 3067–3074 (2017)

[57] Yang, J., Zhang, D., Frangi, A.F., Yang, J.y.: Two-dimensional pca: a new approach to appearance-based face representation and recognition. IEEE transactions on pattern analysis and machine intelligence 26(1), 131–137 (2004)

[58] Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. pp. 435–442. ACM (2015)

[59] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., Dobaie, A.M.: Facial expression recognition via learning deep sparse autoencoders. Neurocomputing 273, 643–649 (2018)

[60] Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. IEEE Transactions on Affective Computing 2(4), 219–229 (2011)

[61] Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., Yan, K.: A deep neural network- driven feature learning method for multi-view facial expression recognition. IEEE Transactions on Multimedia 18(12), 2528–2536 (2016)

[62] Zhang, Y.D., Yang, Z.J., Lu, H.M., Zhou, X.X., Phillips, P., Liu, Q.M., Wang, S.H.: Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access 4, 8375–8385 (2016)

[63] Zhao, X., Shi, X., Zhang, S.: Facial expression recognition via deep learning. IETE technical review 32(5), 347–355 (2015)

[64] Zhao, X., Zhang, S.: A review on facial expression recognition: Feature extraction and classification. IETE Technical Review 33(5), 505–517 (2016)