iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.51265

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Facial Emotion Recognition

Vikas Maurya[1], Ms. Anjali Awasthi [2], Rajat Singh[3], Deepak Maurya[4], Manorma Dwivedi[5]
[1]*UG Student of Department of Computer Science and Engineering, Babu Banarasi Das Institute ofTechnology and Management Lucknow, Uttar Pradesh, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, Babu Banarasi Das Institute ofTechnology and Management Lucknow, Uttar Pradesh, India*

Abstract: ***Facial emotion recognition (FER) has become an important topic in the fields of computer vision and artificial intelligence due to its great academic and commercial potential. Although FER can be performed using multiple sensors, this review focuses on studies using facial images exclusively, since visual expressions areone of the main channels of information in human communication. Automatic emotion recognition based onfacial expressions is an interesting research area that has been applied and applied in various fields such as safety, health and human-computer interface. Researchers in this field are interested in developing techniquesto interpret, encode facial expressions and extract these features for better prediction by computer. With the remarkable success of deep learning, different types of architectures of this technique are exploited to achieve better performance. The purpose of this paper is to conduct a study of recent work on automatic facial emotionrecognition (FER) via deep learning. We highlight these contributions, the architectures and the databases used, and we show the progress achieved by comparing the proposed methods and the obtained results. The purpose of this paper is to serve and guide researchers by reviewing recent work and providing insights to improve the field.***
Keywords: facial emotion recognition, deep neural networks, automatic recognition, database

I. INTRODUCTION

Automatic emotion recognition is a large and important research area that addresses two different subjects, which are psychological human emotion recognition and artificial intelligence (AI). Facial emotions are important factors in human communication that help us understand the intentions of others.

In general, people infer the emotional states of other people, such as joy, sadness, and anger, using facial expressions and vocal tone. The emotional state of humans can obtain from verbal and non-verbal information captured by the various sensors, for example from facial changes, tone of voice and physiological signals.

In 1967, Mehrabian 4 showed that 55% of emotional information were visual, 38% vocal and 7% verbal. Face changes during a communication are the first signs that transmit the emotional state, which is why most researchers are very interested by this modality.

In this paper, we provide a review of recent advances in sensing emotions by recognizing facial expressions using different deep learning architectures. We present recent results with an interpretation of the problems and contributions. It is organized as follows: in section two, we introduce some available public databases, section three; we present a recent state of the art on theFER using deep learning and we end in section four and five with a discussion and comparisons then a general conclusion with the future works.

II. LITERATURE REVIEW

For automatic FER systems, various types of conventional approaches have been studied. The commonalityof these approaches is detecting the face region and extracting geometric features, appearance features, or ahybrid of geometric and appearance features on the target face. For the geometric features, the relationship between facial components is used to construct a feature vector for training

Ghimire and Lee used two types of geometric features based on the position and angle of 52 facial landmarkpoints. First, the angle and Euclidean distance between each pair of landmarks within a frame are calculated,and second, the distance and angles are subtracted from the corresponding distance and angles in the first frame of the video sequence. For the classifier, two methods are presented, either using multi-class AdaBoostwith dynamic time warping, or using a SVM on the boosted feature vectors.

The appearance features are usually extracted from the global face region or different face regions containing different types of information.

©IJRASET: All Rights are Reserved

Happy et al. utilized a local binary pattern (LBP) histogram of different block sizes from a global face regionas the feature vectors, and classified various facial expressions using a principal component analysis (PCA).Although this method is implemented in real time, the recognition accuracy tends to be degraded because itcannot reflect local variations of the facial components to the feature vector. Unlike a global-feature-based approach, different face regions have different levels of importance. For example, the eyes and mouth containmore information than the forehead and cheek.

Ghimire et al. extracted region-specific appearance features by dividing the entire face region into domain-specific local regions. Important local regions are determined using an incremental search approach, which results in a reduction of the feature dimensions and an improvement in the recognition accuracy. for hybrid features, some approaches have combined geometric and appearance features to complement the weaknessesof the two approaches and provide even better results in certain cases. In video sequences, many systems are used to measure the geometrical displacement of facial landmarks between the current frame and previous frame as temporal features, and extracts appearance features for thespatial features. The main difference between FER for still images and for video sequences is that the landmarks in the latter are tracked frame-by-frame and the system generates new dynamic features through displacement between the previous and current frames. Similar classification algorithms are then used in the video sequences, as described in Figure 1. To recognize micro-expression, high speed camera is used to capture video sequences of the face.

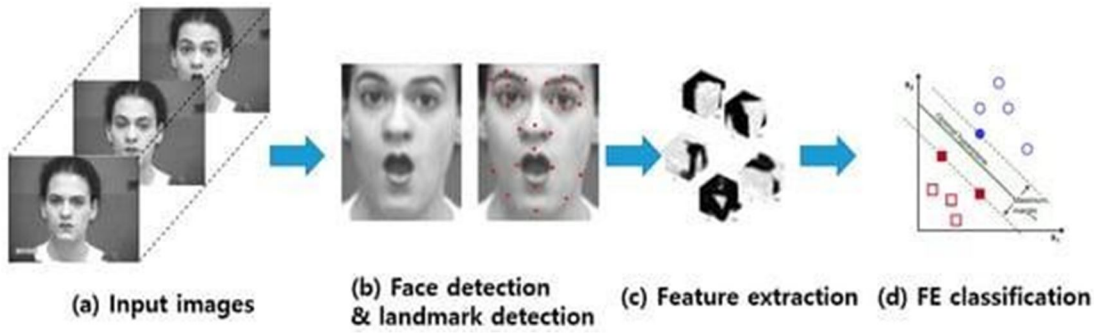


Figure 1. Procedure used in conventional FER approaches: From input images (a), face region and facial landmarks are detected (b), spatial and temporal features are extracted from the face components and landmarks (c), and the facial expression is determined based on one of facial categories using pre-trained pattern classifiers (face images are taken from CK+ dataset (d).

Polikovsky et al. presented facial micro-expressions recognition in video sequences captured from 200 framesper second (fps) high speed camera. This study divides face regions into specific regions, and then 3D- Gradients orientation histogram is generated from the motion in each region for FER.

Apart from FER of 2D images, 3D and 4D (dynamic 3D) recordings are increasingly used in expression analysis research because of the problems presented in 2D images caused by inherent variations in pose and illumination. 3D facial expression recognition generally consists of feature extraction and classification. One thing to note in 3D is that dynamic and static system are very different because of the nature of data. Static systems extract feature from statistical models such as deformable model, active shape model, analysis of 2Drepresentations, and distance-based features. In contrast, dynamic systems utilize 3D image sequences for analysis of facial expressions such as 3D motion-based features. For FER, 3D images also use the similar conventional classification algorithms. Although 3D-based FER showed higher performance than 2D-based FER, 3D and 4D-based FER also has certain problems such as a high computational cost owing to a high resolution and frame rate, as well as the amount of 3D information involved. Some researchers have tried to recognize facial emotions using infrared images instead of visible light spectrum (VIS) image because visiblelight (VIS) image is variable according to the status of illumination.

Zhao et al. used near-infrared (NIR) video sequences and LBP-TOP (Local binary patterns from three orthogonal planes) feature descriptors. This study uses component-based facial features to combine geometricand appearance information of face. For FER, a SVM and sparse representation classifiers are used.

Shen et al. used infrared thermal videos by extracting horizontal and vertical temperature difference from different facial sub-regions. For FER, the Adaboost algorithm with the weak classifiers of k-Nearest Neighboris used.

Szwoch and Pieniążek recognized facial expression and emotion based only on depth channel from Microsoft Kinect sensor without using camera. This study uses local movements within the face area as the feature and recognized facial expressions using relations between particular emotions. Sujono and Gunawanused Kinect motion sensor to detect face region based on depth information and active appearance model (AAM) to track the detected face. To role of AAM is to adjust shape and texture model in a new face, whenthere is variation of shape and texture comparing to the training result. To recognize facial emotion, the change of key features in AAM and fuzzy logic based on prior knowledge derived from FACS are used. Weiet al. proposed FER using color and depth information by Kinect sensor together. This study extracts facial feature points vector by face tracking algorithm using captured sensor data and recognize six facial emotionsby random forest algorithm.

Commonly, conventional approaches determine features and classifiers by experts. For feature extraction, many well-known handcrafted feature, such as HoG, LBP, distance and angle relation between landmarks are used and the pre-trained classifiers, such as SVM, AdaBoost, and random forest, are also used for FE recognition based on the extracted features. Conventional approaches require relatively lower computing power and memory than deep learning-based approaches. Therefore, these approaches are still being studiedfor use in real-time embedded systems because of their low computational complexity and high degree of accuracy However, feature extraction and the classifiers should be designed by the programmer and they cannot be jointly optimized to improve performance.

One of the success factors of deep learning is the training the neuron network with examples, several FERdatabases now available to researchers to accomplish this task, each one different from the others in term ofthe number and sizeof images and videos, variations of the illumination, population and face pose. Some presented in the Table.1 in whichwe will note its presence in the works cited in the following section.

III. FACIAL EMOTION RECOGNITION USING DEEP LEARNING

Despite the notable success of traditional facial recognition methods through the extracted of handcrafted features, over the past decade researchers have directed to the deep learning approach due to its high automatic recognition capacity. In this context, we will present some recent studies in FER, which show proposed methods of deep learningin order to obtain better detection. Train and test on several static or sequential databases.

Mollahosseini et al. propose deep CNN for FER across several available databases. After extracting the facial landmark from the data, the images reduced to 48x 48 pixels. Then, they applied the augmentation data technique. The architecture used consist of two convolution-pooling layers, then add two inception styles modules, which contains convolutional layers size 1x1, 3x3 and 5x5. They present the ability to usetechnique the network-in-network,which allow increasing local performance due to the convolution layers applied locally, and this technique also makeit possible to reduce the over-fitting problem.

Lopes et al. 24 Studied the impact of data pre-processing before the training the network in order to have a betteremotion classification. Data augmentation, rotation correction, cropping, down sampling with 32x32 pixels and intensity normalisation are the steps that were applied before CNN, which consist of two convolution-pooling layers ending with two fully connected with 256 and 7 neurons. The best weight gainedat the training stage are used at the test stage. This experience was evaluated in three accessible databases: CK+, JAFFE, BU-3DFE. Researchers showsthat combining all of these pre-processing steps is more effective than applying them separately.

These pre-processing techniques also implemented by Mohammad pour et al. They propose a novel CNN for detecting AUs of the face. For the network, they use two convolution layers, each followed by a max pooling and ending with two fully connected layers that indicate the numbers of AUs activated.

In 2018, for the disappearance or explosion gradient problem Cai et al. propose a novel architecture CNN with Sparse Batch normalization SBP. The property of this network is to use two convolution layers successive at thebeginning, followed by max pooling then SBP, and to reduce the over-fitting problem, thedropout applied in the middle of three fully connected. For the facial occlusion problem Li et al. present a new method of CNN, firstlythe data introduced into VGG Net network, then they apply the technique of CNN with attention mechanism ACNN. This architecture trained and tested in three large databases FED-RO, RAF-DB and Affect Net.

Detection of the essential parts of the face was proposed by Yolcu et al. They used three CNN with same architecture each one detect a part of the face such as eyebrow, eye and mouth. Before introducing the images into CNN, they go through the crop stage and the detection of key-point facial. The iconic face obtained combined with the raw image was introduced into second type of CNN to detect facial expression.Researchers show that this methodoffers better accuracy than the use raw images or iconize face alone (See Fig.1.a).

In 2019, Agrawal et Mittal make a study of the influence variation of the CNN parameters on the recognition rate using FER2013 database. First, all the images are all defined at 64x64 pixels, and they make a variationin size and number of filters also the type of optimizer chosen (adam, SGD, adadelta) on a simple CNN, which contain two successive convolution layers, the second layer play the role the max pooling, then a soft max function for classification. According to these studies, researchers create two novel models of CNN achieve average 65.23% and 65.77% of accuracy, the particularity of these models is that they do notcontain fully connected layers dropout, and the same filter size remains in the network.

Deepak Jain et al. propose a novel deep CNN which contain two residual blocks, each one contain four-convolution layer. These model trains on JAFFE and CK+ databases after a pre-processing step, which allows cropping and normalizing the intensity of the images.

Kim et al. 31 studies variation facial expression during emotional state, they propose a spatio-temporal architectwith a combination between CNN and LSTM. At first time, CNN learn the spatial features of the facial expression inall the frames of the emotional state followed by an LSTM applied to preserve the wholesequence of these spatial features. Also Yu et al. Present a novel architecture called Spatio-Temporal Convolutional with Nested LSTM (STC-NLSTM), this architecture based on three deep learning sub network such as: 3DCNN for extraction spatio-temporal features followed by temporal T-LSTM to preserve the temporal dynamic, then the convolutional C-LSTMfor modelled the multi-level features.

Deep convolutional Bi LSTM architecture was proposed by Liang et al. they create two DCNN, one of whichis designated for spatial features and the other for extracting temporal features in facial expression sequences, these features fused at level on a vector with 256 dimensions, and for the classification into one of the six basic emotions, researchers used Bi LTSM network. For the pre-processing stage, they used the Multitask cascade convolutional network for detecting the face, then applied the technique of data augmentation to broaden database (See Fig.1.b).

All of the researchers cited previously classifying the basic emotions: happiness, disgust, surprise, anger,fear, sadness. and neutral, Fig 3. Present some different architecture proposed by the researchers who mentioned above

(a) (b)

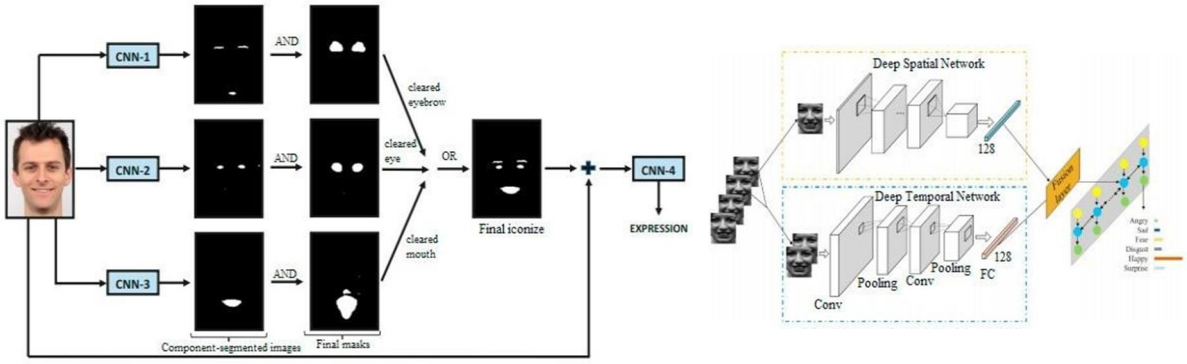


Fig. 1. Different deep learning methods proposed by Yoclu et al. and (b). Liang et al.

IV. DISCUSSION AND COMPARISON

In this paper, we clearly noted the significant interest of researchers in FER via deep learning over recent years. The automatic FER task goes through different steps like: data processing, proposed model architecture and finally emotion recognition.

The preprocessing is an important step, which was present in all the papers cited in this review, that consist several techniques such as resized and cropped images to reduce the time of training, normalization spatialand intensity pixelsand the data augmentation to increase the diversity of the images and eliminate the over-fitting problem. All these techniques are well presented by lopes et al.

Several methods and contributions presented in this review was achieved high accuracy. Mollahosseini et al. showed the important performance by adding inception layers in the networks. Mohammad pour et al. 25 prefer toextract AU from the face than the classification directly the emotions, Li et al. is interested in the study the problem of occlusion images, also for to get network deeper, Deepak et al. propose adding theresidual blocks. Yolcu et al. 28 shows the advantage of adding the iconized face in the input of the network, enhance compared withthe training just with the raw images. For Agrawal et Mittal. offers two new CNN architecture after an in-depth study the impact of CNN parameters on the recognition rate. Most of these methods presented competitive results over than 90%.

For extraction the spatio-temporal features researchers proposed different structures of deep learning such as a combination of CNN-LSTM, 3DCNN, and a Deep CNN. According to the results obtained, the methods proposed byYu et al. and Liang et al. 33 achieve better precision compared to the method used byKim et al. 31. With a rate higher than 99%.

Researchers achieve high precision in FER by applying CNN networks with spatial data and for sequential data, researchers used the combination between CNN-RNN especially LSTM network, this indicate that CNN is the network basic of deep learning for FER. For the CNN parameters, the Soft max function and Adam optimization algorithm are the most used by researchers. We also note in order to test the effectiveness of the proposed neural network architecture, researchers trained and tested their model in several databases, and we clearly see that the recognition rate varies from one database to another with thesame DL model (See Table.1). Table1 summarize all the articles cited above, lists the architecture, the database and the recognition rate.

Table 1 Comparison between presented works.

Authors	databases	The architecture used	Recognition rate
Mollahosseini et al.	MultiPie10, MMI11, DISFA34, FERA12, SFEW13, CK+14, FER201315	CNN	94.7%, 77.9%, 55%, 76.7%, 47.7%, 93,2%, 61.1%
Lopes et al.	CK+14, JAFFE16, BU-3DFE17	CNN	96.76% for CK+
Mohammad pour et al.	CK+14	CNN	97.01%
Cai et al.	JAFFE16, CK+14	SBN-CNN	95.24%, 96.87%
Li et al.	RAF-DB21, AffectNet20	ACNN	80.54%, 54.84%
Yolcu et al.	RafD22	CNN	94.44%
Agrawal et Mittal.	FER201315	CNN	65%
Deepak jain et al.	JAFFE16, CK+14	CNN	95.23%, 93.24%
Kim et al.	MMI11, CASME II 18	CNN-LSTM	78.61%, 60.98%
Yu et al	CK+14, Oulu-CASIA19, MMI11, BP4D35	STC-NLSTM	99.8%, 93.45%, 84.53%
Liang et al.	CK+14, Oulu-CASIA19, MMI11	DCBiLSTM	99.6%, 91.07%, 80.71%

V. CONCLUSION AND FUTURE WORK

This paper briefly reviews the FER method, allowing us to understand the recent developments in the field. As we describe, these methods can be divided into two main streams: Traditional FER methods consist of three steps, namely face and facial component detection, feature extraction, and expression classification. We describe different architectures of CNNs and CNN-LSTMs recently proposed by different researchers, and propose several different databases containing spontaneous images collected from the real world and others formed in the laboratory in order to accurately detect human emotions. We also present a discussionshowing the high ratios obtained by the researchers, which highlights that today's machines will be more capable of interpreting emotions, implying that human-computer interactions are becoming more and more natural. FERs are one of the most important ways to provide information about emotional states, but they have always been limited by learning only the six basic and neutral emotions. It clashes with what exists in everyday life, which has more complex emotions. This will push researchers to build larger databases and create powerful deep learning architectures to identify all primary and secondary emotions in future work.Furthermore, emotion recognition today has evolved from single-modal analysis to multi-modality in complex systems.

REFERENCES

[1] C. Marechal et al., « Survey on AI-Based Multimodal Methods for Emotion Detection », in High- Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST ActionIC1406 cHiPSet, J. Kołodziej et H. González-Vélez, Éd. Cham: Springer International Publishing, 2019, p. 307-324.

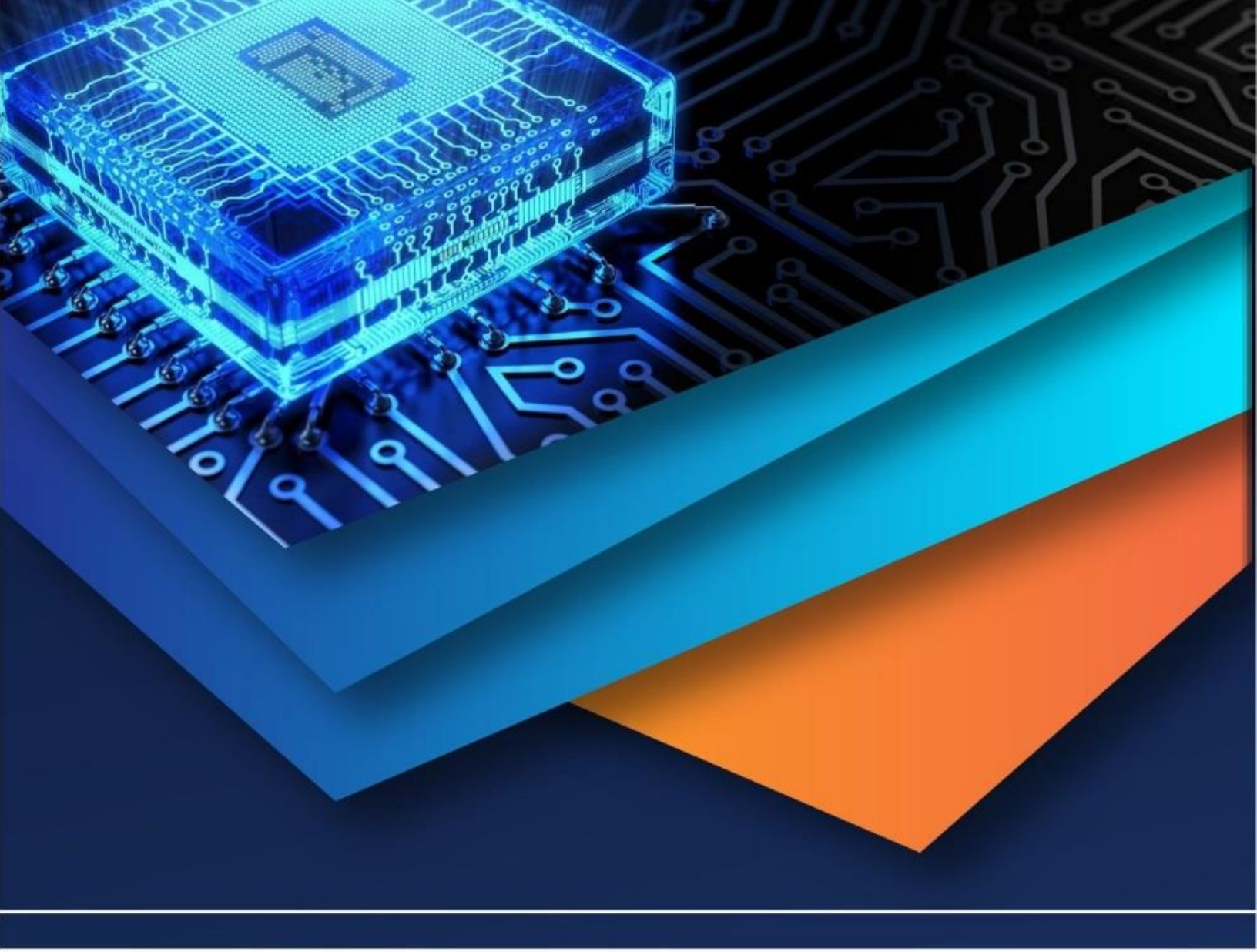
[2] M. H. Alkawaz, D. Mohamad, A. H. Basori, et T. Saba, « Blend Shape Interpolation and FACS for Realistic Avatar », 3D Res., vol. 6, n° 1, p. 6, janv. 2015, doi: 10.1007/s13319-015-0038-7.

[3] P. V. Rouast, M. Adam, et R. Chiong, « Deep Learning for Human Affect Recognition: Insights and New Developments », IEEE Trans. Affect. Comput., p. 1-1, 2018, doi: 10.1109/TAFFC.2018.2890471.

[4] C. Shan, S. Gong, et P. W. McOwan, « Facial expression recognition based on Local Binary Patterns:A comprehensive study », Image Vis. Comput., vol. 27, n° 6, p. 803-816, mai 2009, doi: 10.1016/j.imavis.2008.08.005.

[5] T. Jabid, M. H. Kabir, et O. Chae, « Robust Facial Expression Recognition Based on Local Directional Pattern », ETRI J., vol. 32, n° 5, p. 784-794, 2010, doi: 10.4218/etrij.10.1510.0132.
[6] S. Zhang, L. Li, et Z. Zhao, « Facial expression recognition based on Gabor wavelets and sparse representation », in 2012 IEEE 11th International Conference on Signal Processing, oct. 2012, vol. 2, p. 816-819, doi: 10.1109/ICoSP.2012.6491706.
[7] R. Gross, I. Matthews, J. Cohn, T. Kanade, et S. Baker, « Multi-PIE », Proc. Int. Conf. Autom. Face Gesture Recognit. Int. Conf. Autom. Face Gesture Recognit., vol. 28, n° 5, p. 807-813, mai 2010, doi:10.1016/j.imavis.2009.08.002.
[8] M. Pantic, M. Valstar, R. Rademaker, et L. Maat, « Web-based database for facial expression analysis », in 2005 IEEE International Conference on Multimedia and Expo, juill. 2005, p. 5 pp.-, doi:10.1109/ICME.2005.1521424.
[9] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, et K. Scherer, « The first facial expression recognition and analysis challenge », in Face and Gesture 2011, mars 2011, p. 921-926, doi: 10.1109/FG.2011.5771374.
[10] A. Dhall, R. Goecke, S. Lucey, et T. Gedeon, « Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark », in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), nov. 2011, p. 2106-2112, doi: 10.1109/ICCVW.2011.6130508.
[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, et I. Matthews, « The Extended Cohn-KanadeDataset (CK+): A complete dataset for action unit and emotion-specified expression », in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition- Workshops, juin 2010, p. 94-101, doi: 10.1109/CVPRW.2010.5543262.
[12] I. J. Goodfellow et al., « Challenges in Representation Learning: A Report on Three Machine LearningContests », in Neural Information Processing, Berlin, Heidelberg, 2013, p. 117-124, doi: 10.1007/978-3-642-42051-1_16.
[13] M. Lyons, M. Kamachi, et J. Gyoba, « The Japanese Female Facial Expression (JAFFE) Database ». Zenodo, avr. 14, 1998, doi:10.5281/zenodo.3451524.
[14] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, et M. J. Rosato, « A 3D facial expression database for facial behavior research », in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), avr. 2006, p. 211-216, doi: 10.1109/FGR.2006.6.
[15] W.-J. Yan et al., « CASME II: An Improved Spontaneous Micro-Expression Database and the BaselineEvaluation », PLoS ONE, vol. 9, n° 1, janv. 2014, doi: 10.1371/journal.pone.0086041.
[16] G. Zhao, X. Huang, M. Taini, S. Z. Li, et M. Pietikäinen, « Facial expression recognition from near- infrared videos », Image Vis. Comput., vol. 29, n° 9, p. 607-619, août 2011, doi: 10.1016/j.imavis.2011.07.002.
[17] A. Mollahosseini, B. Hasani, et M. H. Mahoor, « AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild », IEEE Trans. Affect. Comput., vol. 10, n° 1, p. 18-31, janv. 2019, doi: 10.1109/TAFFC.2017.2740923.
[18] S. Li, W. Deng, et J. Du, « Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild », 2017, p. 2852-2861.
[19] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, et A. van Knippenberg, « Presentation and validation of the Radboud Faces Database », Cogn. Emot., vol. 24, n° 8, p. 1377-1388,déc. 2010, doi: 10.1080/02699930903485076.
[20] A. Mollahosseini, D. Chan, et M. H. Mahoor, « Going deeper in facial expression recognition using deep neural networks », in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV),mars 2016, p. 1-10, doi: 10.1109/WACV.2016.7477450.
[21] A. T. Lopes, E. de Aguiar, A. F. De Souza, et T. Oliveira-Santos, « Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order », Pattern Recognit., vol. 61, p. 610-628, janv. 2017, doi: 10.1016/j.patcog.2016.07.026.
[22] M. Mohammadpour, H. Khaliliardali, S. M. R. Hashemi, et M. M. AlyanNezhadi, « Facial emotion recognition using deep convolutional networks », in 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), déc. 2017, p. 0017-0021, doi: 10.1109/KBEI.2017.8324974.
[23] J. Cai, O. Chang, X. Tang, C. Xue, et C. Wei, « Facial Expression Recognition Method Based on SparseBatch Normalization CNN », in 2018 37th Chinese Control Conference (CCC), juill. 2018, p. 9608-9613, doi:10.23919/ChiCC.2018.8483567.
[24] Y. Li, J. Zeng, S. Shan, et X. Chen, « Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism », IEEE Trans. Image Process., vol. 28, n° 5, p. 2439-2450, mai 2019, doi:10.1109/TIP.2018.2886767.
[25] G. Yolcu et al., « Facial expression recognition for monitoring neurological disorders based on convolutional neural network », Multimed. Tools Appl., vol. 78, n° 22, p. 31581-31603, nov. 2019, doi:10.1007/s11042-019-07959-6.
[26] A. Agrawal et N. Mittal, « Using CNN for facial expression recognition: a study of the effects of kernelsize and number of filters onaccuracy », Vis. Comput., janv. 2019, doi: 10.1007/s00371-019-01630-9.
[27] D. K. Jain, P. Shamsolmoali, et P. Sehdev, « Extended deep neural network for facial emotion recognition », Pattern Recognit. Lett., vol. 120, p. 69-74, avr. 2019, doi: 10.1016/j.patrec.2019.01.008.
[28] D. H. Kim, W. J. Baddar, J. Jang, et Y. M. Ro, « Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition», IEEE Trans. Affect. Comput., vol. 10, n° 2, p. 223-236, avr. 2019, doi: 10.1109/TAFFC.2017.2695999.
[29] Z. Yu, G. Liu, Q. Liu, et J. Deng, « Spatio-temporal convolutional features with nested LSTM for facial expression recognition », Neurocomputing, vol. 317, p. 50-57, nov. 2018, doi: 10.1016/j.neucom.2018.07.028.
[30] D. Liang, H. Liang, Z. Yu, et Y. Zhang, « Deep convolutional BiLSTM fusion network for facial expression recognition », Vis. Comput., vol. 36, n° 3, p. 499-508, mars 2020, doi: 10.1007/s00371-019-01636-3.
[31] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, et J. F. Cohn, « DISFA: A Spontaneous Facial Action Intensity Database », IEEE Trans. Affect. Comput., vol. 4, n° 2, p. 151-160, avr. 2013, doi: 10.1109/T-AFFC.2013.4.
[32] M. F. Valstar et al., « FERA 2015 - second Facial Expression Recognition and Analysis challenge », in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), mai 2015, vol. 06, p. 1-8, doi: 10.1109/FG.2015.7284874.
[33] M. Pantic et L. J. M. Rothkrantz, « Toward an affect-sensitive multimodal human-computer interaction », Proc. IEEE, vol. 91, n° 9, p. 1370-1390, sept. 2003, doi: 10.1109/JPROC.2003.817122.
[34] S. Zhang, S. Zhang, T. Huang, et W. Gao, « Multimodal Deep Convolutional Neural Network for Audio-Visual Emotion Recognition », in Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, New York, NY, USA, 2016, p. 281–284, doi: 10.1145/2911996.2912051.
[35] F. Ringeval et al., « Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data », Pattern Recognit. Lett., vol. 66, p. 22-30, nov. 2015, doi: 10.1016/j.patrec.2014.11.007.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)