FACIAL EXPRESSION RECOGNITION USING DEEP LEARNING: A REVIEW

Neha A. Chinchanikar

Student, Dept. of Electronics & Communication, KLS Gogte Institute of Technology, Karnataka, India

---***---

Abstract - *Extensive attention facial expression recognition (FER) has received recently as facial expressions are considered as the fastest communication medium of any type of information. Facial expression recognition gives a better understanding towards a person's thoughts or views and analyzing them with the currently trending deep learning methods boosts the accuracy rate drastically compared to the traditional state-of-the-art systems. This article gives a brief about various application fields of FER and publicly available datasets used in FER and reviews the latest research in the field of FER using deep learning. Lastly, it concludes the efficient method among them.*

Key Words: Facial expression recognition, Feature extraction, Deep learning, Convolutional neural network

1. INTRODUCTION

Facial expressions or emotions are means of conveying non-verbal feelings or sentiments of a person. Facial expression recognition (FER) is a method to recognize expressions on one's face. We see numerous techniques available today to detect various human face expressions like angry, happy, sad, neutral, disgust, surprise, fear and few more which are difficult to be implemented. Many applications like human-computer interaction (computer responding/interacting with humans after analysing what human feels), computer forensics (in the case of lie detection), pain detection, the field of education (i.e. distance learning where teachers determine whether the student understood the course), games and entertainment (for asserting user experience) find its base in facial expression recognition systems. There are several traditional FER methods that require manual feature selection (like Histogram of Oriented Gradients (HoG), Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT), etc.) and then feed to a custom designed classifier to classify expressions. However, such methods fail to produce accurate results since the features are manually extracted and also it becomes cumbersome when dataset is huge. This is the place where deep learning prevails. Deep learning avoids the complexity of manually extracting features. In deep learning we try to replicate brain system with layered model structure to extract features from input data step by step thereby resulting in more abstract high-level feature representation[2].

Many researchers and developers have opted for deep networks such as Convolutional Neural Network(CNN) [1][2][3][4][6][7][9][11][13][14][16][17][21][22][25][26], Deep Belief Network (DBN)[12][23] and Recurrent Neural Network (RNN)[15] in the past and continue to research in the field of image processing and video processing.

There are two major approaches to go about facial emotion recognition 1) geometric based and 2) appearance based approach. In geometric based approach, different geometric parameters like position, angle, landmark points etc. are considered while in appearance based approach the entire input image is considered and features are extracted from the image that best represent the input image.

The daunting task in any deep learning model is fetching the data that suits your requirement. Majority of the standard datasets are composed of near-frontal face poses of good image quality and resolution with partial occlusions like sunglasses, hand covering the face, hat, beard, etc. So the major challenge in data acquisition is to gather the dataset in an unconstrained environment by taking into account different occlusions, age, head tilt, varying illumination, noisy data, changing background so that the system gives good performance under diverse conditions. When we gather such diverse data, preprocessing is a must to denoise the images and align/resize them. After preprocessing sometimes we happen to face dearth in the data, this is where data augmentation [4][13][15][25][26] comes to our rescue. It is one way by which we can overcome the problem of limited dataset availability and expand the dataset quantitatively.

FER can generally refer to a system that collectively does four major tasks 1) face detection, 2) preprocessing, 3)feature extraction, 4) expression classification. The objective of this paper is to get acquainted with the databases used in FER systems and give research fellows a consolidated paper that describes the work carried out in this field in recent years.

This paper is organized into four sections. The first section gives a brief on FER applications and the second section elucidates publicly available datasets used in FER systems, the third section includes a review of twenty six previous researches in the expression recognition using deep learning approaches. Finally, the last section concludes the effective approach among the different approaches discussed in section three.

2. FACIAL EXPRESSION RECOGNITION APPLICATIONS

Facial expressions are the result of movement of muscles beneath the skin. They are predominant channels of conveying social information between individuals. Facial expression analysis provides objective and real-time information about how our faces intimate emotional content. FER has wide range of applications spread across fields like medicine, e-learning, monitoring, entertainment, law etc. The use of FER in each of the mentioned fields is as discussed below:

2.1 E-learning

In e-learning, lecturers will analyze student's ability to understand the themes by observing the emotions and change the teaching methodology and presentation as per the style of the learner. This helps in building a stronger education system through which students will profit a lot whether or not its distance learning.

2.2 Monitoring

Psychological studies have shown that driver's emotions play vital role in safe driving. The emotional condition of the driver helps in deciding the comfort and safety while driving a vehicle. According to the psychological study, expressions like anger, sadness, fear contribute to rash and speedy driving. Anger, aggression, fatigue, stress may increase the accident risks. Conjointly nervousness and sadness could also affect driving. Therefore if we have a FER system that constantly detects driver's expressions and if found to be one among the categories mentioned former can warn the driver and in turn prevent accidents.

FER will play most vital role in police work by analysing the expressions of an individual to detect the likeliness whether or not an individual is afraid while retreating cash from ATM, it will then conceive to stall dispensing cash.

Knowledge about the customer preferences and satisfaction can be monitored and analyzed by having a FER tool installed at the retail stores which provides data which when analysed may maximise user shopping experience [27].

2.3 Medicine

Many a times it happens that distance becomes a hurdle for patient check-up sessions when the patient resides at remote location or the patient is very ill to travel or the patient is old, to not let distance cause a break in the medicine therapy, FER systems can be the solution to such a scenario. Counselling and predicting mental state of the old age patient can be done remotely to see how the patient is feeling and how comfortable the patient is to the line of treatment.

Due to the impaired ability to process faces as seen amongst autistic children in turn accounts for problems that they showcase during social interaction. By building a FER application on mobile phone and handing it over to the autistic children will help them to detect the expressions and label it using an emoticon will help such children when they struggle to interpret feelings of an individual [21].

2.4 Entertainment

In video games to assert player experience [2] in real-time helps the developers to make the players emotionally attached to the game. To gain an understanding whether a game is successful in making the player experience enjoyable, we need to monitor and analyse the face expressions in real-time. This helps the developer to come up with a better solution.

2.5 Law

We know that lie detection is evaluation of verbal statement to reveal whether an individual is deceiving. If we combine lie detection techniques with FER it may yield much better results. It could be used for lie detection purposes in courts and criminal interrogation.

3. DATASETS FOR FER

Table -1: Brief on Publicly available FER databases

Databases	Description	Used in
	Image-based	
JAFFE	JAFFE is an acronym for Japanese Female Expression Database. A total of 213 images by 10 Japanese female models each with 7 facial expressions comprises the database.	[1], [3], [5], [6], [10], [13], [14], [18], [20]
FACES	It is composed of six facial expressions: neutrality, sadness, disgust, fear, anger and happiness images by 171 young (n=58), middle aged (n=56), and old (n=57) people resulting in a total of 2052 images.	[6]
FER 2013	Facial Expression Recognition (FER) 2013 dataset is available on Kaggle. It includes a train set of 28709 examples, public and private test set each of 3589 examples. Each image is of the size 48*48 pixels annotated with one among the seven basic expressions (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral).	[1], [2], [13], [17], [21]
VIVA-FACE	This dataset contains images of drivers with varying illumination clicked in natural driving environments, each image of resolution 984*544. This database was not explicitely designed for driver expression recognition so [5] took a survey of many people to categorise the dataset into custom emotions such as 'Happy', 'Talking', 'Makeup', 'Surprise', 'Neutral' and 'Distracted'. They then selected those images for which most people had consensus.	[5]
DriveFace	DriveFace is a database made up of images of different drivers. It contains 606 samples with resolution 640*480 of two male and two female drivers. Here too [5] took a survey of people to categorise the data into six expression labels as in VIVA-FACE since the data was unlabelled.	[5]
AffectNet	It is composed of nearly 1M colour face images collected from the Internet. Half the images collected are manually labelled for one of the seven basic facial expressions. Remaining images are automatically labelled by making use of ResNet model.	[18]
	Video-based	
CK+	Extended Cohn Kanade dataset is called CK+. It is composed of 327 image sequences of 118 people elucidated with seven facial expression classes (Anger, Contempt, Disgust, Fear, Happiness, Sadness and Surprise). 10 to 60 frames make up each image sequence beginning with a neutral expression and ending with a peak expression. Many traditional state-of-the-art methodologies have been assessed on this database.	[3], [4], [5],[6], [9], [10], [13], [14], [15], [18], [20], [22], [25], [26]
Oulu-CASIA	Oulu-CASIA database is composed of 480 image sequences, each starting off at a neutral face and ending with the apex expression. It is made up of 80 subjects with each subject's 6 emotion labels (Anger, Disgust, Fear, Happiness, Sadness and Surprise).	[4], [10], [20], [22], [25]
BU-3DFE	BU-3DFE, an acronym for Binghamton University 3D Facial Expressions has been used as benchmark dataset when dealt with static 3D FER. 100 subjects of which 56 are females and 44 males in the age ranging between 18 to 70 years old make	[7], [8], [11]

	up this dataset. Each subject has 25 samples of 7 expressions i.e. one neutral sample and 24 samples for six prominent expressions namely anger, disgust, fear, sadness, happiness and surprise. It consists of 2500 geometric shape images and 2500 2D texture images.	
Bosphorus 3D	Widely used for 3D face recognition under adverse conditions, 3D facial action unit detection. It contains 105 subjects and 4666 pairs of 3D face expressions, poses and occlusions.	[7]
MUGFE	This database contains 1032 video snaps of a total of 86 subjects (35 women and 51 men) of Caucasian origin aged between 20 to 35 years for six basic expressions. Each video clip has 50-160 frames starting with neutral to peak expression.	[10]
Multi-PIE	Like BU-3DFE, this is a non-frontal database containing six facial expressions (disgust, neutral, scream, smile, squint & surprise). The expressions were captured under 15 view points and 19 illumination variations by 337 subjects (235 males & 102 females).	[11]
BAUM-1S	It contains six basic expressions (joy, anger, sadness, surprise, disgust, fear) with boredom and contempt, also it contains four mental states i.e. thinking, unsure, concentrating, bothered. This database is collected from 31 turkish individuals resulting in a total of 1222 video clips.	[12]
RML	It contains 720 video clips from 8 persons and has 6 basic expressions (angry, sadness, disgust, joy, fear and surprise).	[12]
MMI	Comprises of 2894 video samples captured from 30 subjects aged between 19 to 62. Of this 2894, 213 video sequences are annotated with six basic expressions.	[5], [12], [25]

4. LITERATURE REVIEW

This literature reviews the latest research in the field of facial expression recognition. It provides an insight on the face detection methods, architecture of the model used for feature extraction and classification along with the accuracies obtained by the following researchers in the area of FER.

4.1 Analysis based on 2D images

Subarna B. and Daleesha M Viswanathan [1] proposed a deep convolutional spatial neural network (DCSNN) which is trained on FER-2013 and JAFFE and tested it in live webcam for real-time face expression recognition. The primary step of face detection is done with Viola Jones algorithm using haar features. This DCSNN is made up of three convolution, two pooling, one fully connected and a softmax layer with Rectified Linear unit (ReLu) activation function to classify the expressions using the probability function. In this paper they have shown DCSNN as an alternative solution to the traditional FER methods.

Xiao Lin and Kiju Lee [2] presented a novel optimized solution in image processing with two preprocessing filters i.e. brightness and contrast filter and edge extraction filter combined with CNN and Support Vector Machine (SVM) for emotion classification. The preprocessing filter parameters are automatically tuned after analyzing the outcomes from CNN. They have claimed an accuracy of 98.19% using CNN. With such high accuracy and efficiency, they have demonstrated that their system has great potential in embedded applications like asserting user experience in games etc.

Minjun Wang et al. [3] proposed a CNN for the task of FER which is based on depth volume of the CNN layers. The experiments were carried out on CK+ and JAFFE dataset using a CNN with five continuous convolution, three max pooling, one fully connected and one output layer to recognise the expressions using softmax classifier. Based on the accuracy table provided, the algorithm used in this paper gives overall accuracy of 87.2% for both the datasets which is high compared to methods like HOG, LBP and traditional CNN.

Jiayu Dong et al. [4] presented a compact CNN with dense connections to mitigate the overfitting problem that occurs when working with finite sized training set. The architecture consists of only four convolutional layers with dense connections, one softmax layer, no connected layers for efficient feature sharing with ReLu activation function. They proposed a dynamic FER system trained on the CK+ and Oulu-CASIA database with an accuracy of 97.25% and 83.33% respectively.

Bindu Verma and Ayesha Choudhary [5] have come up with a real-time framework for recognising the driver emotion using deep learning and Grassmann manifolds. The Region of interest (RoI) was found using haar-cascades and the Deep Neural Networks (DNNs). AlexNet and VGG16 were fine tuned to extract the features which were later fed to Grassmann Graph embedding Discriminant Analysis (GGDA) to produce an accuracy of 98.8% on JAFFE, 98.8% on MMI and 97.4% on CK+ databases which seem to outperform than the state-of-the-art systems.

Atul Sajjanhar et al. [6] have presented a comparative study between the accuracies obtained on the CNN trained from scratch and the pretrained models like inception v3, VGG 19, VGGFace and VGG 16 on publicly available databases like CK+, JAFFE and FACES. The CNN built from scratch has two convolution, two max pooling and two dense layers. Training and testing is carried out on different types of appearance based face images like raw images, difference images and LBP. The outcomes presented show that accuracy obtained on VGGFace is higher compared to the other mentioned pretrained models.

Heechul Jung et al. [9] have developed a real-time system for FER using two deep networks i.e. DNN and CNN on the CK+ database. The DNN consisted of three fully connected layers and a softmax output layer to categorise expressions while there were three convolution, three max pooling, two fully connected and softmax layer with ReLu activation in CNN. They have obtained recognition rate of 72.78% and 86.54% for DNN and CNN respectively and hence concluded that CNN performs better compared to DNN.

Min Peng et al. [10] have come up with DNN to identify the micro expressions using transfer learning approach. Face RoI was segmented using AAM [11] & normalized to 224 *224 pixels. Since they had a small database they fine-tuned ResNet10 on four macro expression datasets and later on micro expression dataset eventually concluded that their approach gives high weighted Average Recall (WAR) and Unweighted Average Recall(UAR) than the traditional (LBP-TOP, HOG3D) ones. The average accuracy of the model was 74.70% with F1 score of 0.64.

Shiqing Zhang et al. [12] proposed a hybrid deep learning model to affectively learn the features in a video sequence for FER. This hybrid model consists of two CNNs, spatial CNN to process static face images and another one called temporal CNN to process optical flow images. These CNNs are fine-tuned using pretrained model and the joint features obtained are fed to the Deep Belief Network (DBN) for the classification task. DBN consists of one visible layer and two hidden layers fed together to Restricted Boltzmann Machines (RBFs) with one softmax classification layer. The accuracies obtained for BAUM-1S, RML and MMI were 55.85%, 73.73% and 71.43% respectively.

Soodamani Ramalingam and Fabio Garzia [13] have proposed transfer learning method using deep neural networks VGG16 and VGG19 for the task of expression prediction when small dataset issue emerges. They have carried out data augmentation- rotation, shifting and scaling on FER 2013 dataset. Pretrained model VGG16 is retrained with augmented FER 2013 and transfer learning is done using VGG19 on Jaffe and CK+. This paper describes the step-by-step procedure on how to use pretrained models for FER. Overall accuracy mentioned was 93%.

Ji-Hae Kim et al. [14] have designed a robust and efficient FER system using hierarchical deep neural network, including an appearance based network (first CNN) which extracts the LBP features and a geometric based network (second CNN) that extracts the difference between the apex and the neutral expression image thereby constructing a robust feature which is given for classification task. Their algorithm seems to have higher accuracy compared to the other state-of-the-art architectures on CK+ and Jaffe datasets. For CK+ the accuracy recorded was 96.46% and for Jaffe 91.27%.

Tong Zhang et al. [15] unlike [12] here the spatial and temporal features are exploited in a RNN called spatial-temporal RNN (STRNN) for EEG signal-based and facial image-based emotion recognition. SJTU Emotion EEG Dataset (SEED) and CK+ were used with 7º clockwise and 12º anti-clockwise rotation. To have a good discriminant ability of emotions, sparse projection was introduced. Model accuracy was demonstrated as 95.4%.

Kejun Wang et al. [16] have fine-tuned the CASME and Large MPI databases to obtain 12 emotion dataset which they fed to the convolutional block CNN (CBCNN) in which small convolution kernels are used in multiple layers rather than having large kernels so as to have a system highly sensitive towards image details and improve accuracy rate. Seven convolution, four max pooling and a fully connected layer composed the CBCNN giving an average error rate 0.132 and average training time 48.69s.

Burhanudin Ramdhani et al. [17] presented a CNN to carry out FER on two kinds of databases, FER2013 and self-created database for four expression recognition to analyse consumer satisfaction. Haar cascade classifier was used for frontal face detection with the CNN having three convolution, three max pooling, two fully connected and one output layer. The accuracy obtained on self-created dataset is 63.41% which is not on par with the traditional architectures for FER.

Nehemia Sugianto et al. [18] had carried out research on measuring customer satisfaction in video surveillance at the airport using deep Residual Network (ResNet50). Viola Jones algorithm was used for detecting faces from surveillance videos. Since most of the passengers carry only neutral and happy faces, to obtain a balanced dataset was a challenge and hence opted for transfer learning to categorise the emotions in three categories (positive, negative and neutral). Their proposed technique demonstrated an accuracy of 91.41%, 85.34% and 83.10% on CK+, Jaffe and AffectNet dataset.

Hermawan Nugroho et al. [19] presented a smart home system for detecting pain especially in elderly people using FER deep learning method. They tuned the available models like OpenFace and FaceNet, which were originally designed for face recognition to detect normal and in pain expressions and high accuracy rate of 93% by this model makes it a preferred solution to be embedded in smart home system.

Peng Wang et al. [20] to counteract the low accuracy rate issue by most traditional FER models, they have come up with a shallow residual network to analyse the mental state of the soldiers using facial expression recognition. Res-Hyper Net structure composed of three convolution layers, average pool before softmax prediction. The experiment was conducted on three standard databases and a self-created dataset and 95.52% accuracy mentioned in the paper shows a great scope for residual network in the FER task.

Md Inzamam Ul Haque and Damian Valles [21] developed a FER system for the children affected with autism using deep convolutional neural network (DCNN) trained on FER2013 dataset. They have later modified the dataset based on variations in illumination and compared the accuracies given by these datasets. According to the results presented in this paper, brighter image dataset gives satisfactory results since it shows 63.11% compared to 53.10% accuracy in the darkest dataset.

Chieh-Ming Kuo et al. [22] demonstrated a frame to sequence methodology ideally designed for portable devices wherein temporal data from the face images was taken into consideration. A CNN with four convolution, two pooling, one fully connected and a softmax layer was deployed with ReLu activation and results demonstrated that their approach gave high accuracy on standard databases. Results provided show a demarcable accuracy of 95.33%.

Fuhua Lin et al. [23] came up with a different deep learning model to detect the engagement of online learners using FER for two outputs i.e. not-engaged and engaged and for three outputs i.e. moderately engaged, highly engaged and not-engaged. Two DBN models with three hidden layers in each were designed. The different approach taken here was to use Kernel Principal Component Analysis (KPCA) and Local Directional Pattern (LDP) to extract the features. Two level engagement showed 90.89% and three level as 87.25% accuracy.

Dung Nguyen et al. [24] used transfer learning approach to avoid the overfitting problem posed in many real time applications. They have used PathNet model to transfer knowledge by retraining the last layer of the model on SAVEE and eNTERFACE resulting in an accuracy of 93.75% and 87.5% on SAVEE and eNTERFACE respectively.

Wenqi Wu et al. [25] proposed a novel method on FER by having special landmark detection networks for various pose faces. Also, they have introduced improved center loss to maximise the distance between non-correlating expressions thereby resulting in superior performing network as compared to standard architectures. Deep CNN with five convolution and two pool layers was used for landmark detection and FER. Reported accuracy rates were 98.22%, 83.10%, 87.39% and 95.97% for CK+, MMI, Oulu-CASIA and CASIA-MFE respectively.

Ming Li et al. [26] came up with an interesting approach wherein they developed a joint feature from identity and emotion to exploit the fact that different people express differently and capturing the correlation between the two could prove to be an important factor when dealing with FER. In this model, one CNN was developed for identity prediction which had four convolutional layers and the other for emotion with deep ResNet architecture designed for FER. Features from both the networks are concatenated and fed to FC layers. Their approach showed an appreciable accuracy of 99.31% and 84.29% on CK+ and FER+ datasets.

4.2 Analysis based on 3D images

Huibin Li et al. [7] proposed a deep fusion convolutional neural network (DF-CNN) for 2D+3D facial expression recognition. The network consists of three subnets namely a feature extraction layer, a feature fusion layer and a softmax layer. Each 3D image is transformed into six types of 2D face attribute maps and then fed to DF-CNN. After this the expression recognition is carried out in two ways by using linear SVM and a softmax prediction. As this network amalgamates feature and fusion learning it gives the best results on expression prediction on 3D images.

Qian Li et al. [8] have used Orthogonal Tensor Marginal Fisher Analysis (OTMFA) on geometric maps for 3D face expression prediction. In this system they have extracted five geometric 2D maps to derive low dimensional tensor subspaces using OTMFA. Each map is trained and tested on multiclass SVM classifiers and the output obtained by each of these classifiers is aggregated for final recognition of expression. Experiment was conducted on BU-3DFE database to achieve an average accuracy of 88.32% and 84.27% which is comparable with the state-of-the art networks.

Tong Zhang et al. [11] have designed a DNN to deal with the multi-view FER issue. The Landmark points were extracted using SIFT such that the model derived 83 and 68 key points from each face in BU3D-FE and Multi-PIE respectively. The DNN model was based on two projection, one convolution, two fully connected & a softmax layer. The proposed method was evaluated on BU-3DFE and Multi-PIE databases to get an average recognition accuracy of 80.1% and 85.2% respectively.

5. CONCLUSION

Facial expression recognition using deep learning has boosted the performance of the system compared to the conventional methods like HoG, LBP etc. When we deal with 2D face images be it static image-based or dynamic video-based, it is seen that CNN with dense connections gives good results and if combined with facial landmarks, or other features gives much better understanding of human expression prediction and increases the accuracy of the system. When it comes to 3D images using a combination of feature learning and fusion learning yields excellent performance. Also having a hybrid model which takes into account the identity and emotion features could lead to better results.

REFERENCES

[1] Subarna B and Daleesha M Viswanathan, "Real Time Facial Expression Recognition Based On Deep Convolutional Spatial Neural Networks" International Conference on Emerging Trends and Innovations in Engineering and Technological Research (ICETIETR) 2018
[2] Xiao Lin and Kiju Lee, "Optimized Facial Emotion Recognition Technique for Asserting User Experience" IEEE Games, Entertainment, Media Conference (GEM) 2018
[3] Minjun Wang et al., "Face Expression Recognition Based on Deep Convolution Network" 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) 2018
[4] Jiayu Dong et al., "Dynamic Facial Expression Recognition Based on Convolutional Neural Networks with Dense Connections" 24th International Conference on Pattern Recognition (ICPR) 2018
[5] Bindu Verma and Ayesha Choudhary, "A Framework for Driver Emotion Recognition using Deep Learning and Grassmann Manifolds" 21st International Conference on Intelligent Transportation Systems (ITSC) 2018
[6] Atul Sajjanhar et al., "Deep Learning Models for Facial Expression Recognition" Digital Image Computing: Techniques and Applications (DICTA) 2018
[7] Huibin Li et al., "Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network" IEEE Transactions on Multimedia, Vol. 19. No. 12, December 2017
[8] Qian Li et al., "3D Facial Expression Recognition using Orthogonal Tensor Marginal Fisher Analysis on Geometric Maps" International Conference on Wavelet Analysis and Pattern Recognition 2017
[9] Heechul Jung et al., "Development of Deep Leaning-based Facial Expression Recognition System" 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) 2015
[10] Min Peng et al., "From Macro to Micro Expression Recognition: Deep Learning on Small Datasets Using Transfer Learning" 13th IEEE International Conference on Automatic Face & Gesture Recognition 2018
[11] Tong Zhang et al., "A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition" IEEE Transactions on Multimedia, Vol. 18, Issue. 12, December 2016

[12] Shiqing Zhang et al., "Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning" IEEE Access, Volume: 7, Pages: 32297-32304
[13] Soodamani Ramalingam and Fabio Garzia, "Facial Expression Recognition using Transfer Learning" International Carnahan Conference on Security Technology (ICCST) 2018
[14] Ji-Hae Kim et al., "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure" IEEE Access, Volume: 7, Pages: 41273-41285
[15] Tong Zhang et al., "Spatial-Temporal Recurrent Neural Network for Emotion Recognition" IEEE Transactions on Cybernetics, Vol. 49, No. 3, March 2019
[16] Keiun Wang et al., "Facial expression recognition based on deep convolutional neural network" IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems 2018
[17] Burhanudin Ramdhani et al., "Convolutional Neural Networks Models for Facial Expression Recognition" International Symposium on Advanced Intelligent Informatics (SAIN) 2018
[18] Nehemia Sugianto et al., "Deep Residual Learning for Analyzing Customer Satisfaction using Video Surveillance" 15th IEEE International Conference on Advanced Video and Signal Bases Surveillance (AVSS) 2018
[19] Hermawan Nugroho et al., "On the Development of Smart Home Care: Application of Deep Learning for Pain Detection" IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES) 2018
[20] Peng Wang et al., "Facial Expression Recognition Algorithm Using Shallow Residual Network for Individual Soldier System" Chinese Automation Congress (CAC) 2018
[21] Md Inzamam Ul Haque and Damian Valles, "A Facial Expression Recognition Approach Using DCNN for Autistic Children to Identify Emotions" IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) 2018
[22] Chieh-Ming Kuo et al., "A Compact Deep Learning Model for Robust Facial Expression Recognition" IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2018
[23] Fuhua Lin et al., "A Deep Learning Approach to Detecting Engagement of Online Learners" IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovations 2018
[24] Dung Nguyen et al., "Meta Transfer Learning for Facial Emotion Recognition" 24th International Conference on Pattern Recognition (ICPR) 2018
[25] Wenqi Wu et al., "Facial Expression Recognition for Different Pose Faces Based on Special Landmark Detection" 24th International Conference on Pattern Recognition (ICPR) 2018
[26] Ming Li et al., "Facial Expression Recognition with Identity and Emotion Joint Learning" IEEE Transaction on Affective Computing, Vol. 14, No. 8, August 2015
[27] Andrea Generosi et al., "A deep learning-based system to track and analyze customer behaviour in retail store" IEEE 8th International Conference on Consumer Electronics 2018