International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Ensemble of Multi Feature Layers in CNN for Facial Expression Recognition using Deep Learning

Chintan B. Thacker, Ramji M. Makwana

Abstract: ***Facial Expression Recognition is an important undertaking for the machinery to recognize different expressive alterations in individual. Emotions have a strong relationship with our behavior. Human emotions are discrete reactions to inside or outside occasions which have some importance meaning. Involuntary sentiment detection is a process to understand the individual's expressive state to identify his intensions from facial expression which is also a noteworthy piece of non-verbal correspondence. In this paper we propose a Framework that combines discriminative features discovered using Convolutional Neural Networks (CNN) to enhance the performance and accuracy of Facial Expression Recognition. For this we have implemented Inception V3 pre-trained architecture of CNN and then applying concatenation of intermediate layer with final layer which is further passing through fully connected layer to perform classification. We have used JAFFE (Japanese Female Facial Expression) Dataset for this purpose and Experimental results show that our proposed method shows better performance and improve the recognition accuracy.***

Index Terms: ***Convolutional Neural Networks, Facial Expression Recognition, Feature Extraction, Feature Concatenation***

I. INTRODUCTION

Individual action examination and detection is immense and tricky research subject in the meadow of computer visualization and prototype identification. Behavioral examination is currently an important spot of explore in computer visualization which helps in solving many problems in indoor as well as outdoor surveillance system. The numbers of video surveillance systems have been increasing every day in order to monitor, track and analyze the behaviors in different areas. In day to day scenario using in different applications like visual traffic monitoring, home securities, luggage thief detection, people counting, in exam hall or in college/organization campus are applications where behavioral analysis is useful. The increasing need of behavioral analysis from surveillance in our basic life also increase research efforts behind improvement in this behavioral analysis field [1].

Manuscript published on November 30, 2019.
* Correspondence Author
Chintan B.. Thacker*, Computer Engineering Department, Gujarat Technological University, Ahmedabad, India
Dr. Ramji M. Makwana, M.D. AIIVINE PXL Pvt. Ltd, Rajkot, Gujarat, India

In our daily life communication plays a major role. Inside this period of innovation, individual Computer Interaction (HCI) and robotization, Emotion detection has turned into an essential pasture of learn. Non-verbal communication by human carried out through facial expressions and body gestures. 55% role is comprised of human body actions [2]. In setting up inter personal relations; role of facial expression is important.

Seven basic types of emotions are there which includes fear, neutral, anger, surprise, disgust, happy and sad. Every other emotions are after effect and diverseness of these basic emotions [3].

Some huge commitments are made in the research of Facial phrase detection with Local Dual Prototype. [4]. Researchers have studied various process for facial phrase detection using binary decision tree [5], with Convolutional Neural Networks [6], Sentiment categorization using NN and HMM [7], Sentiment study in illustration and auditory prompt [8], joining several core methods [9]. However these computational strategies contain a long ways following than individual precision like their establishment did not rely on the working concept of individual deep learning as well as training. Goal of our study is to observe the sentiment recognition in still imagery through the help of Convolutional Neural Network (CNN). CNN be a popular neural network for deep learning area that gives solutions of problems regarding large training is required in image recognition. Large amount of training is required to correctly identify or recognize emotion in a face. For example, it is difficult to determine whether a person's emotion is sad or angry without proper training. [10]

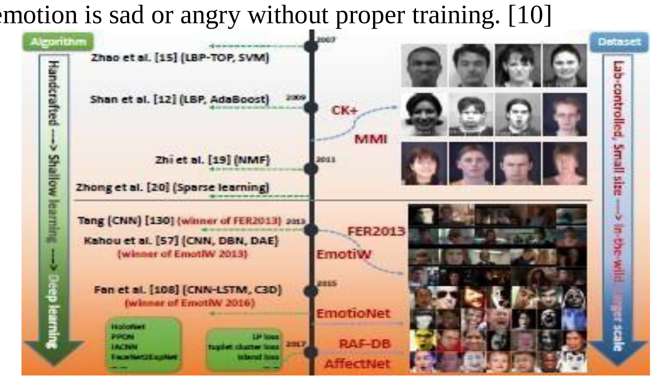

Fig 1. Facial Expression recognition in terms of database with methods [43]

Many reviews have been carried out by researchers on automatic emotion recognition recent years [11], [12], [13], [14]. These reviews have set up a lot of standard algorithmic pipelines for FER which will be helpful to carry out proper classification result analysis.

Still many researchers have focused on traditional methods compared to deep learning methods and trying to improve the result based on current challenges in FER. Recently review on FER with the help of deep learning carried out which contains brief review on FER datasets and technical details of different methodologies. Also systematic review represented on different datasets with static images and real time videos covered with its methods as shown in figure 1. [15] It will be very useful for any newcomer to analyze important information in this area.

Also it will give modern analysis of deep learning methods inside FER. Below figure 2 shows the working mechanism of CNN for facial expression recognition which is made of Convolutional Layers followed by Fully Connected Layers and Classification Layer to classify the emotions.

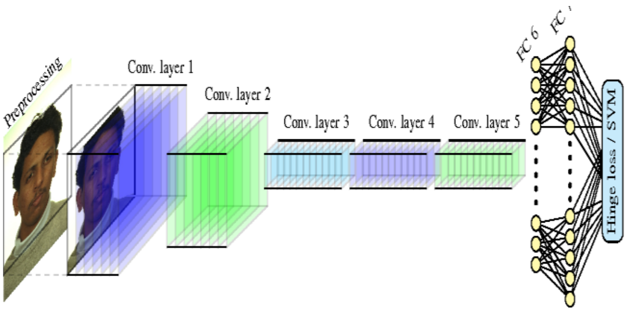

Fig 2. Working mechanism of CNN architecture with its layer details for Facial Expression Recognition [44]

Some of the problems identified in FER using deep learning though it has powerful feature of learning ability. A huge quantity of training data require in deep neural network to avoid over fitting problem. To prepare renowned neural network through deep structural design to facilitate attained a large amount significant results in object recognition, existing facial expressions databases are not adequate to train this networks. Also subject varieties exist because of various characteristics like age, masculinity, cultural surroundings and intensity of articulacy [16]. Also pose variations, enlightenment in addition to occlusions are universal factors in unimpeded emotion recognition circumstances. These aspects be involved with facial appearances scenario and consequently to perform better classification, necessity of deep systems is there with huge intra-class changeability in addition to study efficient classification phrase scenario.

Rest of the article is structured like follows. Section II presents Related Work, Section III presents Proposed work for FER by ensemble the features with its layers, Section IV represents Experimental Results compared with pre-trained architecture and Section V represents Conclusion along with Future work and References

II. RELATED WORK

Different kinds of conservative advances contain carried out for Automatic FER systems. To generate a feature vector for training, association among facial apparatus is employed for geometric characteristics found lying on place and viewpoint of 52 degree of facial marker spots. Here primary viewpoint and Euclidean distance is calculated involving every duo of landmarks inside a framework and then distance along with angle values be deducted as of the matching space plus angle values of primary frame in record string. Two classifiers techniques are used here: multi class AdaBoost in the company of dynamic time warping and SVM on the boosted feature vectors. [38]

Diverse face expanses contain diverse styles of detail so look features are habitually mined on or after universal face area. Happy et al.[39] used an approach of Local Binary Pattern (LBP) histogram with dissimilar chunk ranges as of a universal facade region as a characteristic vector plus after that categorized diverse facial expression via Principal Component Analysis (PCA). Though this technique is applied in instantaneous environment, its precision is corrupted as of not able to mirror local differences of facial sections to characteristic vector. Diverse face regions contain poles apart intensities of significance. For instance, compare to forehead plus cheek, eyes in addition to mouth contains additional information. Ghimire et al.[40] divided whole face region hooked on domain precise local expanses to extract appearance features and using an incremental search method, important local regions were identified which provides improvement in recognition accuracy and reduction in feature dimensions.

Many researchers have identified different feature extraction methods and classifiers for conventional approaches. For facial expression recognition well known methods for characteristic mining like Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), distance along with angle relation flanked by facial landmarks plus classifiers for instance Support Vector Machine (SVM), AdaBoost, Random Forest are employed founded on mined characteristics. Benefits of conservative approaches are that they oblige inferior computing control with remembrance compared to Deep learning based procedures. Thus these procedures are tranquil mortal employed in real time organizations since of their lower computational difficulty along with higher accuracy [60].

Now days, deep learning area has revealed tremendous presentation in the computer vision meadow other than very less amount of work has been carried out for Facial Expression Recognition [17]. Deep Belief Network (DBN) has been utilized widely to achieve facial expression detection chores. No image pre- processing is required while analyzing the facial picture through DBN, directly face image be able to be concerned [18].

For instance, Ranzato et al. used gated Markov Random field approach to study discrimination facial expression [19]. In addition to this, DBN identifies facial expression on huge set of facial imageries. In addition it is used on pre-trained guiding data examples wherever a Gabor Filter was applied with deep structure framework to carry out feature extraction. An additional technique for facial expression detection is AU-aware deep network where appearance differences can be described by several limited facial Action Units (AU) with followed by it will be used by DBN to discover several characteristics for executing ultimate facial expression recognition [20]. Recently many researchers have been found out that facial expression recognition can perform well with Convolutional Neural Network (CNN) for candid and non- posed pictures through multi-scale CNN approach and also via uniting a Facial Coding System (FACS) through CNN [22]. Another approach carried out is to combine CNN and SVM classifier with deep network assembled through a heap of SVMs [23].

Retrieval Number: D8940118419/2019©BEIESP
DOI:10.35940/ijrte.D8940.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

CNN has additionally demonstrated its superior performance within challenges. It was used with the champions of a 2014 challenge demonstrating extensive enhancement in their outcomes as of the preceding challenge [24]. A different viewpoint employed in CNN in the company of Softmax classifier instead of SVM classifier for succeeding the FER confront in 2013 [25]. Ensemble techniques have been designed for FER except not into deep learning framework. However they have been consolidated with DBN for time series forecasting and object tracking purpose [26]. Additionally Restricted Boltzmann Machines (RBM) is coordinated to make a lot increasingly precise individual classifier.

Though, these group training strategies aren't suitable meant for FER [27].

For enhanced performance of FER, different Deep Belief Network (DBN) models have been combined with audio-video signals to extract important features for Emotion recognition. Also boosted DBN approach suggested discovering and picking important characteristics as well as designing the required categorizer [28]. With the help of different deep learning techniques, complex approach is used to recognize emotion from videos on different modes. For this approach, DBN and CNN are used to implement the representation of audio stream and to capture visual information from detected faces respectively [29]. In other approach different CNNs were combined where individual CNN was implemented for varying input features with varying initializing weights of Neural Network with the use of Fusion function across all classifiers to calculate maximum value of the outputs. Another study of CNN by applying pre processing of inputs with different methods then by averaging results of all CNNs final recognition results were generated [30]. Further improvement is by Fusion and adjusting weighting method where based on accuracy of validation set weight of each network defined [31].

Some ensemble methods of CNNs were proposed for focused appliances. For instance, during a learning by Lyksborg et al., to segment tumor tissue 3 CNNs were guided on 46*46 images [32], while Haibo et al. to perform biological cell detection a cascaded procedure united by a CNN model along with handcrafted characteristics [33]. Another method proposed for ensembles of CNNs to detect polyp with specialization of individual CNN as a feature. Though the entire assembly techniques of DBNs plus CNNs were not found efficient in support of Facial Expression Recognition [34].

A novel proposed method with CNN which has 2 Convolutional layers in which each layer pursued by Max Pooling along with 4 Inception levels are applied in favor of Facial Expression Recognition which is not related to Ensemble Learning method [35]. Based on hierarchical committee of CNN is newly proposed method of FER which has improved 2.14% accuracy with its classifier whereas ensemble methods improved accuracy by 2.81% which proves that ensemble method is more effective. But based on its hierarchical architecture which is composed of CNNs making it more complex [36].

Precision, Recall, Accuracy and F1-Score are four major parameters used to evaluate metrics of FER. Precision (P) is described the same as (True Positive) / (True Positive + False Positive) and Recall (R) is described the same as (True Positive) / (True Positive + False Negative) wherever True Positive is the numeral of true positives and False positive is the numeral of false positives in given dataset. Divisions of automatic footnotes of emotion to facilitate are appropriately identified is known as Precision. Accurate detections of emotion in excess of actual numeral of imagery through emotion are known as Recall. Accuracy is described by the proportion of true outcomes via entire numeral of images mentioned below in Eq. (1) to find the same [61].

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Images}} \quad (1)$$

III. PROPOSED METHOD

Here in this paper we have proposed an architecture in which we are going to combine features of intermediate layer of Inception V3 architecture with features of final layer before providing to fully connected layers. Here from pre-trained architecture we are going to extract High level features from inter mediate layers which contains trained and important features which will be further helpful to classify our images more accurately. Initially we have to provide images as an input to pre-trained CNN architecture. Here we have to divide input database images into training and testing purpose. During training process architecture will carried out Pre-processing on images and then it will be provided for Convolutional operations as per the basic architecture of Inception V3 for the further process as shown in figure 3.

During validation we have to calculate results of pre-trained architecture for its different network parameters like Batch Size and Learning Rates and we have to measure its classification accuracy. Further as per our proposed architecture shown in figure 4, at initial stage we have to do the same thing as we have carried out in pre-trained architecture but we have to take intermediate feature for concatenation with final layer which contains high level features for input facial images to carried out efficient recognition process with good classification accuracy. Both the said processes are explained below in detail:

A) Pre-Trained InceptionV3 CNN Architecture:

InceptionV3 is a Convolutional Neural Network which is trained on millions of images from the ImageNet database. Also it is an extended version of GoogleNet architecture. It contains concept of Inception Modules to reduce the number of connections/parameters. It has 11 Inception modules and an image input size is 299*299.

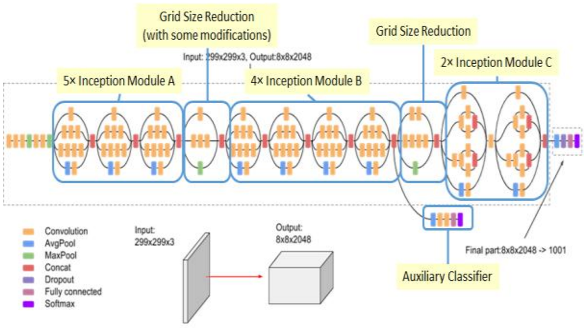

Fig 3. Basic Architecture of Inception V3 [45]

B) Proposed Architecture:
As discussed earlier here in proposed architecture we are going to concatenate intermediate layer which contains high level features with final layer which will further provided to Fully Connected Layers (FC) and Classification Layer (Softmax we have used here).

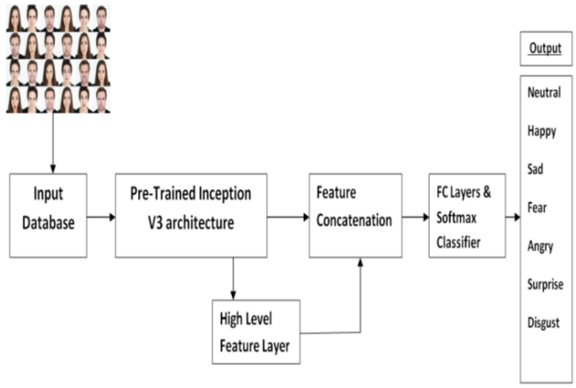

Fig 4. Proposed Architecture of Ensemble of Features by Concatenating layers in Inception V3

IV. EXPERIMENTS

To implement existing and proposed architecture, all Training and Testing processes were performed on NVIDIA GeForce GTX 1050 Ti 4GB GPU with i7 8th generation Windows system. Also we have used Keras-Tensorflow and Anaconda Environment with Python programming language by using PyCharm Community Edition as an IDE tool.

C) Database Details:
In many research JAFFE database is widely used which contains 10 Japanese feminine's expressions through 7 different Facial Expressions in addition to contain Total 213 Gray Imagery with resolution of 256*256 for each image. It contains 7 basic facial expressions which include Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise. Sample of JAFFE database is revealed inside figure 5.

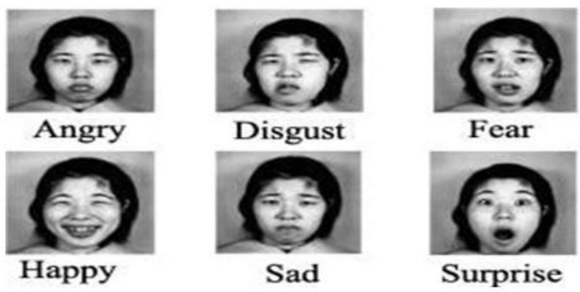

Fig 5. Example imagery from JAFFE Database [37]

Many researchers are dividing the database in 70-30% ration or 80-20% ratio in their evaluation process. JAFFE database contains total 213 images so here we have divided JAFFE dataset into training and testing purpose as mentioned in below Table 1.

Table 1. JAFFE Databases Description

Name of Database	Total No. of Images	No. of Training Images	No. of Testing Images
Japanese Female Facial Expressions (JAFFE) [37]	213	153	60

D) Network Parameters:
To implement these pre-trained and proposed architectures in Keras-Tensorflow Anaconda environment, we have used some network parameters which are necessary for the process. Parameters values and details are mentioned below:
Epochs: Entire dataset is passed (forward & backward) through Neural Network
Batch Size: Total number of training samples present in a single batch
Learning Rate: Parameters that controls adjusting the weights of our network with respect to loss gradient
Optimizer: Used to find the parameters to minimize loss function
Classifier: To classify your image into resultant output
In our implementation we have used Epochs 100, Batch size 8,16,32 and Learning Rate varies from 10^-1 to 10^-5. We have used Stochastic Gradient Descent (SGD) as an optimizer and Softmax as a Classifier.

E) Evaluation & Results:
We have generated results of existing pre-trained architecture and on proposed work in terms of accuracy mentioned below in Table 2 and Table 3. We found that in proposed work we got some better performance.

Table 2. Result in terms of Accuracy carried out on existing pre-trained Inception V3 CNN architecture for different Batch Size & Learning Rate with JAFFE dataset

Implementation Result of InceptionV3 CNN Model Accuracy (%) on JAFFE Dataset (Without Feature Selection)					
Max. Epochs = 100	Learning rate				
Batch Size	0.1	0.01	0.001	0.0001	0.00001
8	73.33	63.33	70	60	73.33
16	60	61.66	63.33	63.33	65
32	50	50	51.66	55	60

Table 3. Result in terms of Accuracy carried out on proposed Inception V3 CNN architecture for different Batch Size & Learning Rate with JAFFE dataset

Implementation Result of InceptionV3 CNN Model Accuracy (%) on JAFFE Dataset (With Feature Selection)					
Max. Epochs = 100	Learning rate				
Batch Size	0.1	0.01	0.001	0.0001	0.00001
8	80	66.66	76.66	73.33	75
16	66.66	63.33	63.33	66.66	68.33
32	63.33	55	51.66	58.33	63.33

Above experimental results for pre-trained Inception V3 CNN architecture is shown in below figure 6 and results for proposed architecture with feature concatenation is shown in figure 7. Both the figures contains accuracy results in (%) shown with the help of bar chart.

For different batch sizes 8,6,32 we have applied different learning rates with required network parameters as discussed in above section and achieved the result. Proposed work figure 7 shows that how recognition accuracy is increasing compare to figure 6 of pre-trained architecture model.

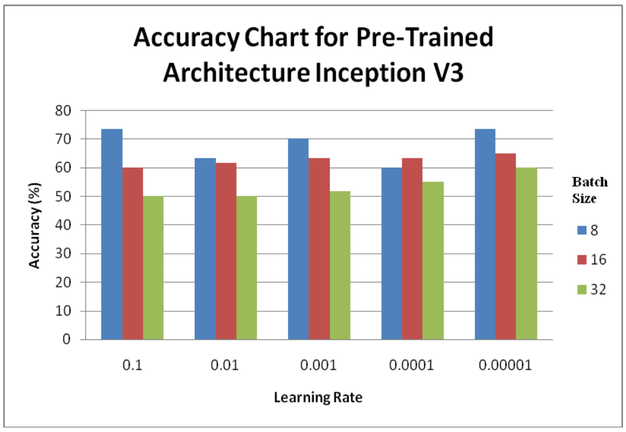

Fig 6. Evaluation results of Accuracy under various batch size and learning rate values for pre-trained Inception V3 architecture.

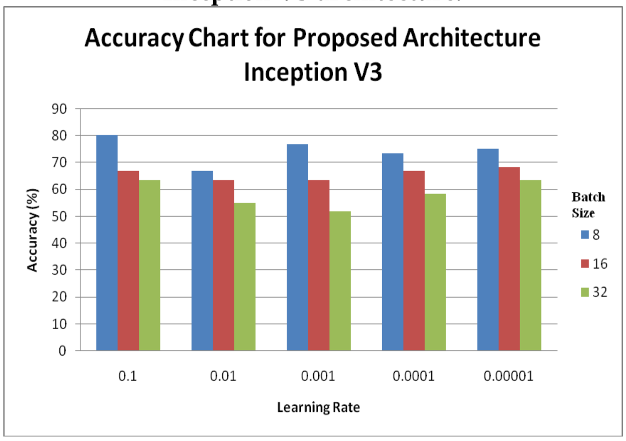

Fig 7. Evaluation results of Accuracy under various batch size and learning rate values for our proposed Inception V3 architecture.

As we are getting improved accuracy in our proposed architecture by concatenation features and for that comparison chart of Batch size 8 for both pre-trained and proposed architecture is shown in figure 8.

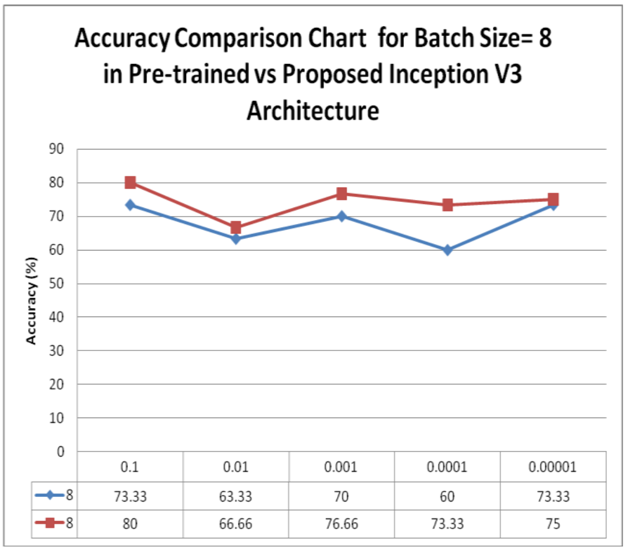

Fig 8. Comparison of Accuracy for Batch size 8 of pre-trained and proposed architecture

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed feature concatenation architecture with Deep Learning approach. In the experiment and results our proposed architecture shows improved performance compare to pre-trained CNN architecture which does not do any feature concatenation. For different batch size and learning rate criteria in both the architectures we are getting improved accuracy from 60-73.3% range to 66.6-80% and hence providing that this proposed method providing good classification accuracy also. In future work regarding this method is to extend this work by combining together with low and high level features by applying different real-time datasets and with different CNN architectures like ResNet50 or MobileNet. Further you can also combine features from different CNN architectures by creating Multi CNN ensemble for expression recognition.

REFERENCES

1. Sreeja Sankaran and Anoop B.K.," Review on vision based Human Activity Analysis", International Journal of Computer Applications, Vol.99, No.2,August 2014
2. Carton, J.S., Kessler, E.A., Pape, C.L.: Nonverbal decoding skills and relationship well-being in adults. J. Nonverbal Behav. 23(1), 91–100 (1999)
3. Izard, C.E.: Human Emotions. Springer, New York(2013)
4. Happy, S.L., George, A., Routray, A.: A real time facial expression classification system using Local Binary Patterns. In: Intelligent Human Computer Interaction (IHCI), 4th International Conference, pp. 1–5. IEEE (2012)
5. Lee, C.C., Mower, E., Busso, C., Lee, S., Narayanan, S.: Emotion recognition using a hierarchical binary decision tree approach. Speech Commun. 53(9), 1162–1171 (2011)
6. 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 Recogn. 61, 610–628 (2017)
7. Hu, T., De Silva, L.C., Sengupta, K.: A hybrid approach of NN and HMM for facial emotion classification. Pattern Recogn. Lett. 23(11), 1303–1310 (2002)
8. Sebe, N., Cohen, I., Gevers, T., Huang, T.S.: Emotion recognition based on joint visual and audio cues. In: 18th International Conference on Pattern Recognition, ICPR, vol. 1, pp. 1136–1139. IEEE, August 2006
9. Liu, M., Wang, R., Li, S., Shan, S., Huang, Z., Chen, X.: Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. In: Proceedings of the 16th ACM International Conference on Multimodal Interaction, pp. 494–501 (2014)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep Convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
11. Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE transactions on pattern analysis and machine intelligence, vol. 31, no. 1, pp. 39–58, 2009.
12. E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 6, pp. 1113–1133, 2015.
13. M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 12, pp. 1424–1445,2000.
14. B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," Pattern recognition, vol. 36, no. 1, pp. 259–275, 2003.

15. T. Zhang,"Facial expression recognition based on deep learning: A Survey" in International Conference on Intelligent Systems and Applications, Springer 2017, pp. 345-352
16. M.F.Valstar, M.Mehu, B.Jiang, M.Pantic and K.Scherer,"Metaanalysis of the first facial expression recognition challenge," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol.42, no.4, pp. 966–979, 2012.
17. e S, Wang S, Wuwei, Fu L, Ji Q. Facial expression recognition using deep Boltzmann machine from thermal infrared images. In: Humaine association conference on affective computing and intelligent interaction; p. 239–244. 2013.
18. Susskind JM, Hinton GE, Movellan JR, Anderson A K. Generating facial expressions with deep belief nets. In: Kordic V, editor. Affective computing, emotion modelling, synthesis and recognition; 2008. p.421-440.
19. Ranzato M, Mnih V, Susskind J, Hinton G. Modeling natural images using gated mrfs. IEEE TPAMI. 2013;35(9):2206–16.
20. Liu M, Li S, Shann S, Chen X. AU-inspired deep networks for facial expression feature learning. Neurocomputing. 2015;159:126–136.
21. Rifai S, Bengio Y, Courville A, Vincent P, Mirza M. Disentangling factors of variation for facial expression recognition. In: ECCV; p. 808–822. 2012.
22. Liu M, Li S, Shann S, Chen X. AU-inspired deep networks for facial expression feature learning. Neurocomputing. 2015;159:126–136.
23. Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, Bartlett M. The computer expression recognition toolbox (CERT). In: IEEE Int'l Conf. on automatic face and gesture recognition; p. 1–2. 2011.
24. Liu M, Wang R, Huang Z, Shan S, Chen X. Partial least squares regression on grassmannian manifold for emotion recognition. In: 15th ACM on International conference on multimodal interaction; p. 525–530. 2013.
25. Tang Y. Deep learning using linear support vector machines. In: Workshop on challenges in representation learning in ICML. 2013
26. Qiu X, Zhang L, Ren Y, Suganthan PN. Ensemble deep learning for regression and time series forecasting. In: 2014 IEEE Symposium on computational intelligence in ensemble learning; p. 1–6.2014.
27. Zhang C-X, Zhang J-S, Ji N-N, Guo G. Learning ensemble classifiers via restricted Boltzmann machines. Pattern Recogn Lett.2014;36:161–170.
28. Kim Y, Lee H, Provost EM. Deep learning for robust feature generation in audio visual emotion recognition. In: ICASSP; p.3687–3691. 2013.
29. Kahou SE, Bouthillier X, Lamblin P, et al. EmoNets: multimodal deep learning approaches for emotion recognition in video. J Multimodal User Interf. 2016;10:99–111.
30. Ciresan DC, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In: IEEE Conference on computer vision and pattern recognition (CVPR); p. 3642–3649.2012.
31. Fraz¨ao X, Alexandre LA.Weighted convolutional neural network ensemble. In: CIARP; p. 674–681. 2014.
32. Lyksborg M, Puonti O, Agn M, Larsen R. An ensemble of 2D CNN for tumor segmentation. Lect Notes Comput Sci. 2015;9127(1):201–211.
33. Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, Tomaszewski J, Gonzalez F, Madabhushi A. Cascaded ensemble of CNN and handcrafted features for mitosis detection. Proc SPIE.2014;9041(2): 90410B-90410B-10.
34. Tajbakhsh N, Gurudu SR, Liang J. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In: 12th IEEE International symposium on biomedical imaging; p. 16–19. 2015.
35. Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter conference on applications of computer vision (WACV); p.1–10. 2016.
36. Kim B-K, Roh J, Dong S-Y, Lee S-Y. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. J Multimodal User Interf. 2016;1–17.
37. http://www.kasrl.org/jaffe.html
38. Ghimire, Deepak, and Joonwhoan Lee. "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines." Sensors13.6 (2013): 7714-7734.
39. Happy, S. L., Anjith George, and Aurobinda Routray. "A real time facial expression classification system using local binary patterns." 2012 4th International conference on intelligent human computer interaction (IHCI). IEEE, 2012.
40. Ghimire, Deepak, et al. "Facial expression recognition based on local region specific features and support vector machines." Multimedia Tools and Applications 76.6 (2017): 7803-7821
41. Suk, Myunghoon, and Balakrishnan Prabhakaran. "Real-time mobile facial expression recognition system-a case study." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014.
42. Fabian Benitez-Quiroz, C., Ramprakash Srinivasan, and Aleix M. Martinez. "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
43. Li, Shan, and Weihong Deng. "Deep facial expression recognition: A survey." arXiv preprint arXiv:1804.08348 (2018).
44. van de Wolfshaar, J., Karaaba, M.F. and Wiering, M.A., 2015, December. Deep convolutional neural networks and support vector machines for gender recognition. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 188-195). IEEE.
45. Review: Inception-V3 architecture https://www.medium.com

AUTHORS PROFILE

Chintan B. Thacker is a Research Scholar at Gujarat Technological University. He has received his Master Degree in 2012 from Gujarat Technological University. His research interest includes Computer Vision, Deep Learning, Machine Learning, and Expression Recognition.

Dr. Ramji M. Makwana is a Managing Director of AIIVINE PXL Pvt. Ltd. He received Ph.D. degree from S. P. University, Vallabh Vidyanagar in 2011. He has authored several papers in major computer vision and multimedia conferences and journals. His research interests include Data mining, Soft computing and deep learning with applications on computer vision tasks, like object recognition, action recognition and Object tracking