Prediction of Emotional Score of the multiple faces of a Photo Frame through Facial Emotion Recognition using the Deep Convolutional Neural Network

P V V S Srinivas[a] and Pragnyaban Mishra[b]

[a] PhD Research Scholar, Department of CSE, KoneruLakshmaiah Education Foundation (KLEF)

[b]Associate Professor Department of CSE, KoneruLakshmaiah Education Foundation (KLEF)

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: The facial movement of his or her is an important mark of understanding the emotion. The emotions are of different categories like angry, sad, neutral, happy, disgust, fear, surprise, etc. To classify the image into an appropriate class of emotion using the deep Convolutional neural network is a more scientific approach of classification. The classification is not only from the current *observations but* also from the past *evidence, i.e.* a training model is used to do this job. In this research article, we developed a model that derives the perception of a frame that consists of multiple faces by measuring each of its *facial* emotions. The developed model, therefore, claimed to be more efficient and robust against the variety of inputs.

Keywords: FER: Facial Emotion Recognition, CNN: Convolutional Neural Network, ANN: Artificial Neural Network, FER 2013, blob frame image

1. Introduction

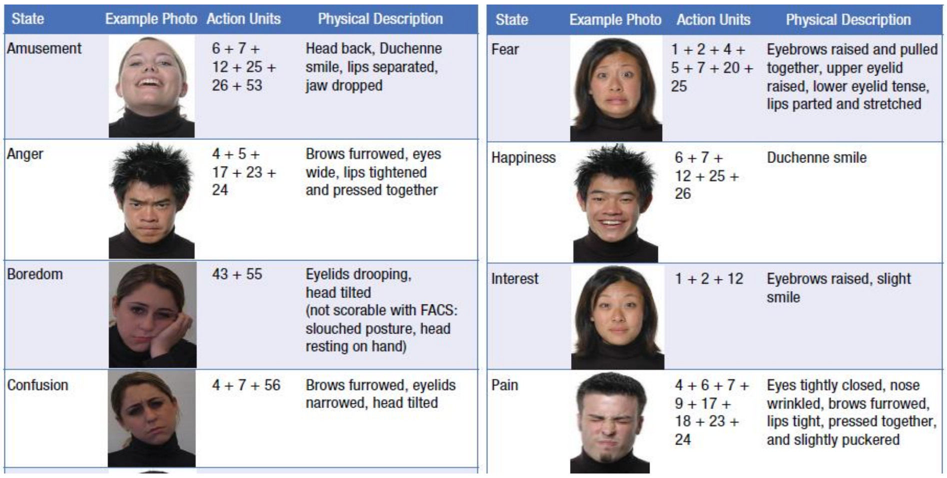
Facial emotion is a reflection of human thought. There are different emotional states like angry, sad, neutral, happy, disgust, fear, surprise, etc., which change from time to time when a human perceives inputs from the surroundings as a result there are changes observed in the facial muscles in a form of different curves.People very close to one another can recognize the feeling of others seeing the face. The expression of a face is a cognitive representation and contains the processed information about the inputs taken from the sense organs.For the last few years, automatic facial expression is used rapidly with an application of AI that is applied in the major sectors listed a few of them are, virtual reality, augmented reality, entertainment, human-computer interaction, and advanced driver assistant systems. However, the Face recognition in images is one of the most challenging research issues in tracking systems[1].

This paper explains the method of calculating the facial expression of multiple faces present in an image that is either a group photo, family photo, or frame and predicts the emotion score of that frame. For example, let us say in an image there are five faces of different expressions, so using our methodology the score of the image is calculated depending on the maximum quantum of the expression.

The remaining of the paper is ordered as follows: 2- Related research, 3- Model Development, 4- Experiment and Result Discussion and 5-Conclusion and Future Scope.

2. Related Research:

Darwin laid the foundation of emotion recognition by stating *"Emotional expressions are multimodal behavioral patterns of an individual"* [2] and claimed there are 40 emotional states a human express. Since many of the researchers have been contributedsignificantly to find out the true information and cognitive representation by measuring facial curves and expansion, contraction of facial muscles.The action units of the face emotion is an important observation to measure the kind of emotion. The authors [3, 4] explored 12 and 27 action units that appear in the faces. The mixing of any two of the facial expression [5] represents a different emotion is shown below.



State	Example Photo	Action Units	Physical Description	State	Example Photo	Action Units	Physical Description
Amusement		6 + 7 + 12 + 25 + 26 + 53	Head back, Duchenne smile, lips separated, jaw dropped	Fear		1 + 2 + 4 + 5 + 7 + 20 + 25	Eyebrows raised and pulled together, upper eyelid raised, lower eyelid tense, lips parted and stretched
Anger		4 + 5 + 17 + 23 + 24	Brows furrowed, eyes wide, lips tightened and pressed together	Happiness		6 + 7 + 12 + 25 + 26	Duchenne smile
Boredom		43 + 55	Eyelids drooping, head tilted (not scorable with FACS: slouched posture, head resting on hand)	Interest		1 + 2 + 12	Eyebrows raised, slight smile
Confusion		4 + 7 + 56	Brows furrowed, eyelids narrowed, head tilted	Pain		4 + 6 + 7 + 9 + 17 + 18 + 23 + 24	Eyes tightly closed, nose wrinkled, brows furrowed, lips tight, pressed together, and slightly puckered

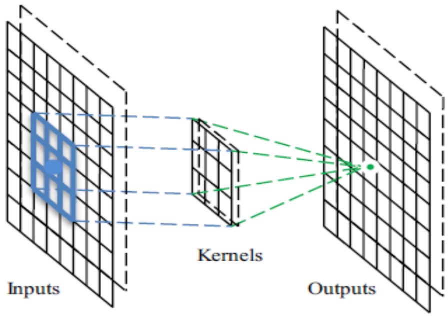
Figure 2.1: Mixture of multiple AU represents different emotion

In the literatureFER is obtained inthree fundamental phases viz.i) Face Detection, ii) Feature separation, and iii) Classification of the emotions.The development of the new and innovative models developed by the researchers from time to time paid more attention on accuracy. Predict the expression more accurately that improves the performance are the key contribution of authors available in the literature. The Hidden Markov Model (HMM) is used to find out the expression [6] used a classification algorithm to classify the emotions by measuring the distance between eyebrows and iris into several categories like anger, fear, happiness and disgust. The ANN classified facial images *into* different classes by training the ANN [7]. [8] used two different methods FLF and DLF of FER that categorizes into classes of positive, neutral, and negative by tracking eye movement and claimed the achievement of accuracy to 88.64% and 88.35%.

From the recent past, the use of deep learning [9] is the most interesting and widely used in designing modelsthat dominated all the mathematical models due to its robustness and ease of use [10]. Prior to the use of deep learning the researchers used different approaches to classify the images into appropriate classes like brain tumor identification [11, 12], Plant disease identification [13, 14] and other applications [15, 16,28] etc.Authors of [17] [18] [19] developed their own model and applied convolutional neural network architecture for FER and claimed suitable performance with references to different data set applied. The FER2013 dataset is [20] have been recognized as a best input image dataset to define and validate the model in most of the researchesFER. The description of the FER2013 dataset is explained in the experiment and result section.Few of the researchers associated this mechanism to any of the application like [21] the FER of patients of medical, hotels, customer service[22],tourist satisfaction[23], robustsurveillance systems [24], brain tumor image classification[25], dance action identification [26], etc.

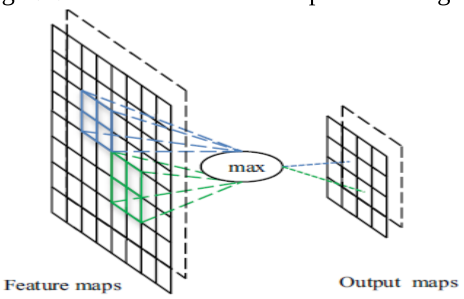
2.1 Deep CNN:

Deep CNN is the most widely used ANN for deep classification [10, 27]. The DCNN consists of layers namely convolutional, pooling, and fully connected.



Convolutional Layer: In this layer, a kernel sometimes called filter slides over the image and useful for the following advantages like reducesthe number of parameters, representing the correlation between neighbored pixels, Invariance to the location of the object.

Fig 2.2: Kernel slides on the top of the image



Pooling layer: This layer is used after the convolutional layer where the dimension of the image is reduced. Average pooling and max pooling a widely used mechanism in this respect.

Fig 2.3: Applying max pooling layer for Dimensional Reduction

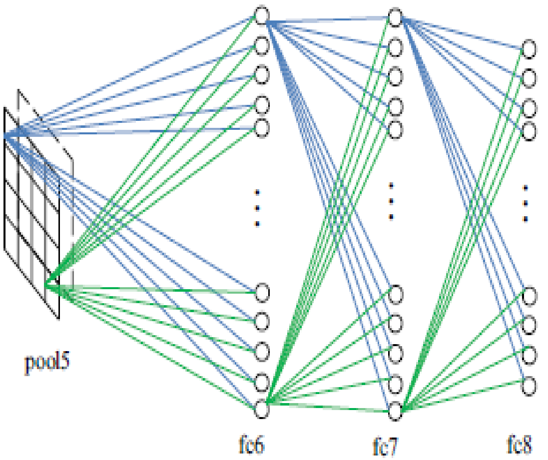


Fig 2.4: Fully Connected NN used for classification

Fully Connected Network: This is a network coversthe major portion of DCNN it is a conventional feed forward NN. After the image passes through multiple convolution layers, pooling layers its dimension shirked finally the left out is transformed into the vector and entered the fully connected NN as an input. The ANN is trained to recognize the image into its appropriate class. The training of the model shall be performed untilthe test accuracy is high. The training of the model should be stopped the moment the test accuracy starts declining.

In the given below figure the input is an image to DCNN which passes through convolution layer and pooling layers interchangeable and assign to ANN after converting it into vector. The ANN is a modeled to recognize the object.

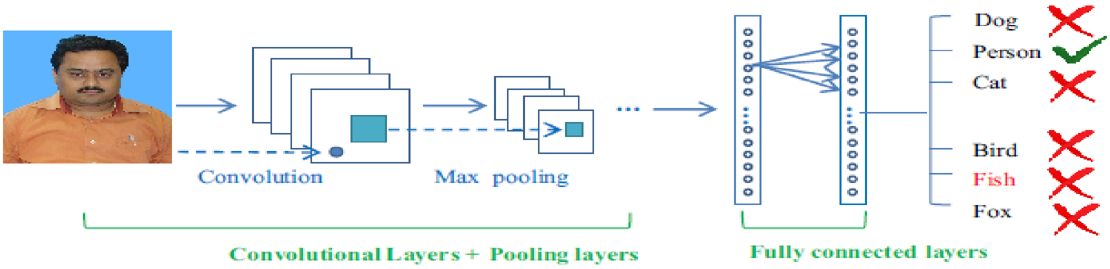


Fig 2.5: Different Layers of Convolutional Neural Network for classification.

Some of the predefined CNNarchitectures are as described below:

Sl No	Network Name	Network Description	Inception Yr
1	AlexNet	5 Convolution Layer and 3 Fully Connected Layer	2012
2	Clarifai	5 Convolution Layer and 3 Fully Connected Layer	2013
3	SPP	5 Convolution Layer and 3 Fully Connected Layer	2014
4	VGG	15 Convolution Layer and 3 Fully Connected Layer	2014
5	GoogleNet	21 Convolution Layer and 1 Fully Connected Layer	2014

Table 2.1: Different Convolutional Neural Network and its Architecture.

The FER2013 dataset is made available by Kaggle is the biggest data source to train the model as well used to validate and test the model. Given below [19] is a representation of facial image along with labels explained in the FER 2013 dataset.

Fig 2.6: Sample Representation of Emotion Facial Data Available in FER2013 Dataset

3. Methodology:

In this research work the frame consists of multiple faces are processed and the knowledge of the frame is assessed by processing individually extracted faces. The entire research is divided into three stages.

Stage1: Extracting the faces from the frame:

A box of a fixed dimension detects faces one after the another from the frame by using blobfromimage library of the cv2 module

Stage2: Determine the correct facial expression of all the extracted faces.

A CNN that consists of fully convolutional layers, max polling, batch Normalization, Dropout, and Dense layers was developed and trained on Fer2013 dataset and a model was generated to classify the emotion of the facial images

Stage3: Findingthe meaning of the group photo.

After all the faces emotions of the frame are recognized then, insight behind the frame is visualized using visualization tool and techniques.

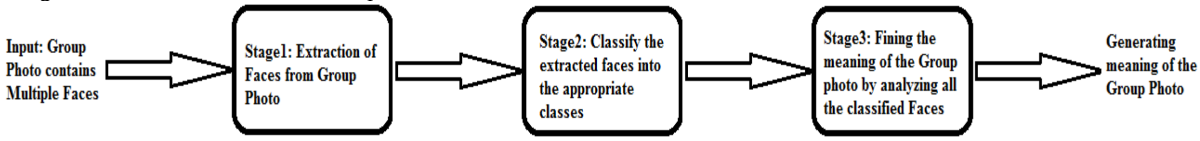


Fig 3.1: Model Description for finding Group photo facial expression

3.1 Algorithm for the Proposed Model

The research is carried out in the following steps:

Step1:Developed a model called *CNN_MODEL* which has been trained with the available dataset like FER2013 etc.

Step2: Identify the faces from the group photo images let's say these are $f_1, f_2, f_3, ..., f_n$.

Step3: Apply all the extracted faces $f_1, f_2, f_3 ..., f_n$to the *CNN_MODEL* to get the class label information $C_1, C_2, C_3, ...$

C_n.of distinct k classes.

Step4:Find out the max class level information C_{res} from the $C_1, C_2, C_3 ... C_n$

Step5:C_{res} is the Class level result of the group photo image.

Above steps of the algorithm is explained with given below example:

Step-1: *CNN_MODEL is developed as follows:*

Layer (type)	Output Shape	Param #
input (InputLayer)	(None, 48, 48, 1)	0
conv1_1 (Conv2D)	(None, 48, 48, 64)	640
.		
.		
conv5_3 (Conv2D)	(None, 3, 3, 512)	2359808
batch_normalization_16 (Batc	(None, 3, 3, 512)	2048
conv5_4 (Conv2D)	(None, 3, 3, 512)	2359808
pool5_1 (MaxPooling2D)	(None, 1, 1, 512)	0
drop5_1 (Dropout)	(None, 1, 1, 512)	0
flatten (Flatten)	(None, 512)	0
output (Dense)	(None, 7)	3591

Total params: 13,111,367
Trainable params: 13,103,431
Non-trainable params: 7,936

Step 2:*Extracting faces from the group photo image*

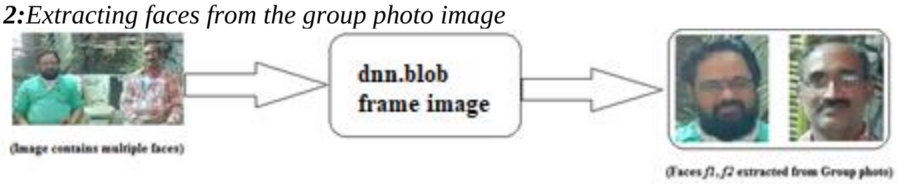


Fig 3.2 Example of face extraction from the frame

Step 3: *Categorizing the extracted faces into appropriate class levels.*

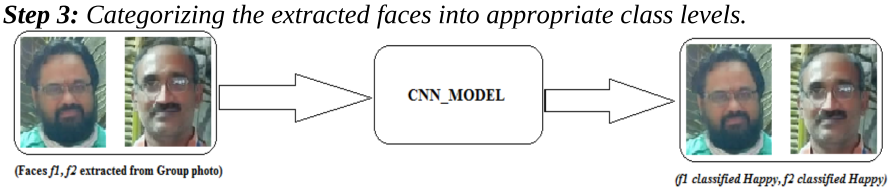


Fig 3.3 Example of FER from the extracted faces

Step4: *Set the C_{res} with the highest frequency.*

Happy → C_1, *Happy* → C_2 So C_{res} is set to 2 that is the highest frequency.

Step5: *The category of this photo is categorized into the class of Happy because the highest frequency belongs to Happy category.*

3.2 Model Description:

- Let the FER image dimension is defined as: $Dim(I) = (I_h, I_w, n_c)$
  WhereI_h: Image Height, I_w: Image width, n_c: Number of Channels.
- There is a filter *K* for each of the channel of the FER image and is defined as: $Dim(K) = (f_i, f_j, n_c)$
  Where f_i, f_j, n_c are the filter lenght, width and no of channels.The length and width of kernel *K* is uniform hence can be represented as *f*
- The filter *K*slided over the image *I* as a result the feature map *con(I, K)* is as follows:
$$con(I, K) = \sum_{i=1}^{I_h} \sum_{j=1}^{I_w} \sum_{k=1}^{n_c} K_{i,j,k} I_{x+i-1,y+j-1,k}$$
- The dimension of the convoluted feature map is as follows:
$$Dim\big(con(I, K)\big) = \left(lowInt\left(\frac{I_h + 2p - f}{s} + 1\right), lowInt\left(\frac{I_w + 2p - f}{s} + 1\right)\right)$$
  Where*s=1, p=0* is the stride and padding used in the convolution layer for this experiments.
- The feature map extracted from the convolutional layer undergone through pooling layer as a result the feature map is down sampled through summing up the information.

  Let φ be the pooling function of *Average pooling* or *Max pooling*used in pooling layer, for each of the channel there is a pooling filter applied to it as a result the feature map is dimensionally reduced with new dimension (I_h^l, I_w^l, n_c^l)
- The convolution layer, pooling layer, batch normalization dropout is used interchangeablyIn this CNN model there are 17 number of convolutional layers, 10 number of pooling layers, 10 numbers of dropping layers, and 15 numbers of batch normalization layers available.
- The feature map is converted to vector map and assign to fully connected Neural network which in turn classify the faces into appropriate classes.
- The output of the classification will be measured and express the meaning of the photo.

3.3 Face Extraction from Frame

```
for file in os.listdir(base_dir + 'images'):
    file_name, file_extension = os.path.splitext(file)
    if (file_extension in ['.png','.jpg']):
        image = cv2.imread(base_dir + 'images/' + file)

        (h, w) = image.shape[:2]
        blob = cv2.dnn.blobFromImage(cv2.resize(image, (300, 300)), 1.0, (300, 300), (104.0, 177.0, 123.0))

        model.setInput(blob)
        detections = model.forward()

        # Identify each face
        for i in range(0, detections.shape[2]):
            box = detections[0, 0, i, 3:7] * np.array([w, h, w, h])
            (startX, startY, endX, endY) = box.astype("int")

            confidence = detections[0, 0, i, 2]

            # If confidence > 0.5, save it as a separate file
            if (confidence > 0.5):
                count += 1
                frame = image[startY:endY, startX:endX]
                cv2.imwrite(base_dir + 'faces/' + str(i) + '_' + file, frame)
```

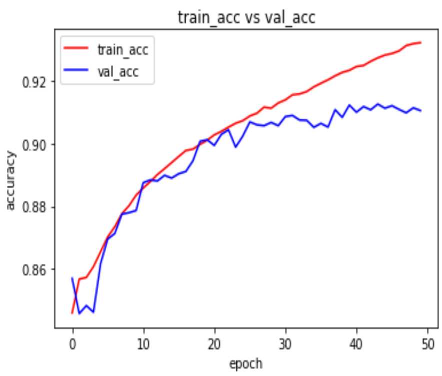
Fig 3.4 Code Snippet for extracting faces from frame

Fig 3.5 Sample output after applying face extraction Algorithm, Image source Google

3.4 CNN Model for FER

The CNN model is developed after training the model by theFER-2013 dataset and tested with the test data set. This module is with reference to the step-01 of section 3.1 which follows the measures given below.

The number of epochs used:50, Test size:7178, Train size: 28709



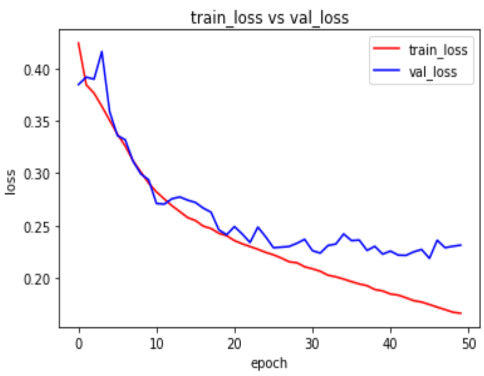
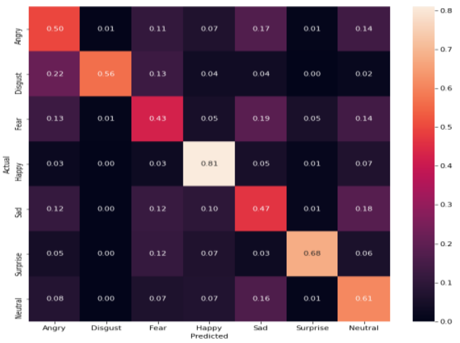


Fig 3.6 Train vs Test Accuracy Fig 3.7 Train vs Test Loss



The output of the classification will be measured based on the performance metrics as follows:

Accuracy = $\frac{T_{Pr}+T_{Nr}}{T_{Pr}+T_{Nr}+F_{Pr}+F_{Nr}}$

Whereas T_{Pr} called as True positive ratio, T_{Nr} is called as True negative ratio, F_{Pr} called as False positive ratio, F_{Nr} called as False negative ratios Respectively

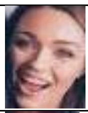
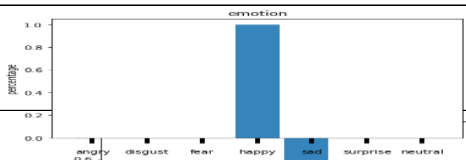
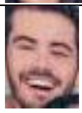
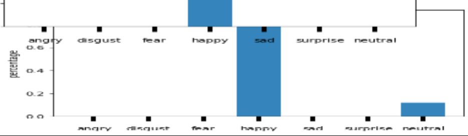
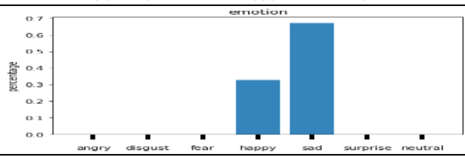
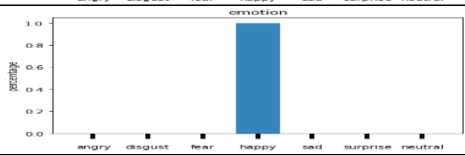
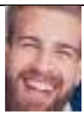
Fig 3.8 Confusion matrix of the model

4 Experiment and Result:

Following is the demonstration of frame emotion analysis:

Original Frame with Multiple Faces	Faces Detected By Applying 3.3
	

Fig 4.1 Frame Before and After Applying to Snippet Mentioned 3.3

S.No	Extracted Face	Amount of FER In The Extracted Face
1		
2		
3		
4		
5		

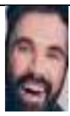
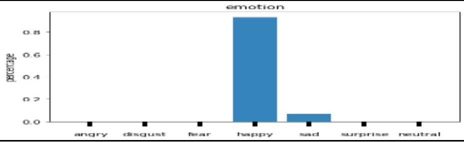
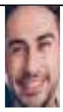
6		
7		
8		

Fig 4.2 Faces Recognized and the Amount Obtained After Applying to The Model

In the above development the count of happy faces is the highest that is 6 out of 8 which set off to the frame is a happy frame.

We have taken different frames from google source and applied to our model. The observation is as tabulated below

Sl No	Name of the Frame	Nos. of Faces Detected	Nos. of Happy Faces	Nos. of Angry Faces	Nos. of Disgust Faces	Nos. of Fear Faces	Nos. of Sad Faces	Nos. of Surprise Faces	Nos. of Neutral Faces	Max Face	Count	% Max Faces	Summary of the Frame
1	Frame1	8	6	0	0	0	1	0	1	Happy	6	75	Frame1 is categorized into a happy frame with percentage of Happiness-75%
2	Frame2	5	2	0	0	0	0	0	3	Neutral	3	60	Frame2 is categorized into a neutral frame with percentage of Neutralness-60%
3	Frame3	17	8	2	0	0	3	0	4	Happy	8	47.1	Frame3 is categorized into a partial happy frame with percentage of Happiness-47.1%
4	Frame4	5	0	1	0	0	1	0	3	Neutral	3	60	Frame4 is categorized into a neutral frame with percentage of Neutralness-60%
5	Frame5	8	3	0	0	0	4	0	1	Sad	4	50	Frame5 is categorized into a neutral frame with percentage of Sadness-50%

5. Conclusion and Future Scope

The described model is a well-trained model to extract multiple faces from a frame and get the internal sense by characterizing each of the faceusing Deep CNN. This research may be extended and applied to the motion scene to get its overall rating after analyzing the maximum faces belongs to a class.

References:

1. Zafar, U.; Ghafoor, M.; Zia, T.; Ahmed, G.; Latif, A.; Malik, K.R.; Sharif, A.M. “Face recognition with Bayesian convolutional networks for robust surveillance systems”. EURASIP J. Image Video Process. 2019, 10.
2. Keltner, D. Born to Be Good: The Science of a Meaningful Life; WW Norton & Company: New York, NY, USA, 2009.
3. AitorAzcarate, Felix Hageloh, Koen van de Sande, Roberto Valenti “Automatic facial emotion recognition” June 2005

4. Yongmian Zhang, Qiang Ji "Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion" Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003)
5. Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements" Psychological Science in the Public Interest 2019, Vol. 20(1) 1–68
6. SANCHETI, G., NAGAR, R., & AGRAWAL, V. (2014). Prediction of Deflection in Post-Tensioned Slabs at Conceptual Stage of Design by Applying Resubstitution Validation Technique.
7. Rajakumari, B.; SenthamaraiSelvi, N. "HCI and eye tracking: Emotion recognition using hidden markov Model". Int. J. Comput. Sci. Eng. Technol. 2015, 6, 90–93.
8. Bahreini, K.; Nadolski, R.; Westera, W. T"owards multimodal emotion recognition in e-learning environments." Interact. Learn. Environ. 2016, 24, 590–605.
9. Wang, Y.; Lv, Z.; Zheng, Y. "Automatic emotion perception using eye movement" information for e-healthcare systems. Sensors (Basel) 2018, 18, 2826.
10. Moulana Mohammed, M. Venkata Sai Sowmya, Y. Akhila, B. Naga Megana, Visual Modeling of Data using Convolutional Neural Networks, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019, PP.No. 4938-4942.
11. YanmingGuo, YuLiu, ArdOerlemans, SongyangLao, SongWu, MichaelS.Lew "Deep learning for visual understanding: A review" Neurocomputing187 (2016) 27–48.
12. Kumar, B. H., & Chitra, P. SURVEY PAPER OF SCRIPT IDENTIFICATION OF TELUGU LANGUAGE USING OCR.
13. Praveena M., Rohini G., Reddy G.T., Nikhil K.H.S. (2019), 'Automatic brain tumor identification using clustering of k-means algorithm in image processing', Journal of Advanced Research in Dynamical and Control Systems, 11(7), PP.621-630.
14. Pradeepini G., Sekhar Babu B., Tejaswini T., Priyanka D., Harshitha M. (2018),'A comparative study on brain tumor diagnosis techniques using MRI image processing',International Journal of Engineering and Technology(UAE),7 (0),PP. 486-489
15. Vamsidhar E., Rani P.J., Babu K.R. (2019), 'Plant disease identification and classification using image processing', International Journal of Engineering and Advanced Technology, 8(0), PP.442-446.355)
16. Lakshmi Praneetha S.K., Anusha K., Geetha Viharika R., Divya Sree M., Vidyullatha P. (2019), 'Automated leaf disease detection in corn species through image analysis', International Journal of Advanced Trends in Computer Science and Engineering, 8(6), PP.2893-2899.
17. KEKAN, A. H., & KUMAR, B. R. (2019). Crack depth and crack location identification using artificial neural network. *Int. J. Mech. Product. Eng. Res. Develop*, *9*(2), 699-708.
18. Ramya Keerthi P., Niharika B., Dinesh Kumar G., Sai Venakat K., Sheela Rani C.M. (2019), 'Reorganization of license plate characteristics using image processing techniques', International Journal of Recent Technology and Engineering, 7(6), PP.1260-1264.
19. Inthiyaz, Syed, B. T. P. Madhav, and P. V. V. Kishore. "Flower image segmentation with PCA fused colored covariance and gabor texture features based level sets." *Ain Shams Engineering Journal* 9.4 (2018)
20. Kuo, C.-M.; Lai, S.-H.; Sarkis, M. "A compact deep learning model for robust facial expression recognition" . In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 2121–2129.
21. Wu, X.; Yuan, P.;Wang, T.; Gao, D.; Cai, Y. "Race Classification from Face using Deep Convolutional Neural Networks" . In Proceedings of the 2018 3rd International Conference on Advanced Robotics and Mechatronics (ICARM), Singapore, 18–20 July 2018; pp. 1–6.
22. Singh, S., Nasoz, F. (2020). Facial Expression Recognition with Convolutional Neural Networks. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)* 324-328. Las Vegas, NV: Institute of Electronics and Electrical Engineers.
23. YEDILKHAN, A., MURAT, K., ALIYA, K., Ainur, K., & Beibut, A. (2019). Predicting heating time, thermal pump efficiency and solar heat supply system operation unloading using artificial neural networks. *International Journal of Mechanical and Production Engineering Research and Development*, *9*(6), 221-232.
24. Goodfellow, Ian J., Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski et al. "Challenges in representation learning: A report on three machine learning contests." In International conference on neural information processing, pp. 117-124. Springer, Berlin, Heidelberg, 2013.
25. Hsu, G.-S.J.; Kang, J.-H.; Huang,W.-F. "Deep hierarchical network with line segment learning for quantitative analysis of facial palsy" . IEEE Access 2019, 7, 4833–4842.

26. Pantano, Eleonora. "Non-verbal evaluation of retail service encounters through consumers' facial expressions." *Computers in Human Behavior* (2020): 106448.
27. González-Rodríguez, M. Rosario, M. Carmen Díaz-Fernández, and Carmen Pacheco Gómez. "Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions." *Telematics and Informatics* (2020).
28. ABD EL NAEEM, A., GHAZALY, N. M., & ABD EL-JABER, G. T. IDENTIFICATION OF UNBALANCE SEVERITY THROUGH FREQUENCY RESPONSE FUNCTION AND ARTIFICIAL NEURAL NETWORKS.
29. Zafar, Umara, et al. "Face recognition with Bayesian convolutional networks for robust surveillance systems." *EURASIP Journal on Image and Video Processing* 2019.
30. MoulanaMohammed, Sai SreeNalluru, Sandhya Tadi, Rachana Samineni. Brain tumor image classification using convolutional neural networks. International Journal of Advanced Science and Technology, 29(05), 928 – 934, (2020).
31. Kishore, P. V. V., et al. "Indian classical dance action identification and classification with convolutional neural networks." *Advances in Multimedia* 2018
32. Kavitha, K., and B. Thirumala Rao. "Evaluation of distance measures for feature-based image registration using AlexNet." *arXiv preprint arXiv:1907.12921* (2019).
33. TALREJA, S. (2016). Stochastically optimized handwritten character recognition system using Hidden Markov Model.
34. Madhuri, N. Phani, A. Meghana, PVRD Prasada Rao, and P. Prem Kumar. "Ailment Prognosis and Propose Antidote for Skin using Deep Learning." International Journal of Innovative Technology and Exploring Engineering,2019.