

A Robust Face Expression Recognition system with Music Recommendations

Name:Vadlamudi Sai Lokesh

Dept:CSE

College:SRM

Email:sailokeshvadlamudi@gmail.com

Name:Kotha Sai Narasimha Rao

Dept:CSE

College:SRM

Email:kothasainarasimharao@gmail.com

Abstract— Human face recognition is one of the important tasks in social communications. face expressions directly reflect human emotion. Face expressions are helpful to identify the real feeling of the person during a certain situation. Face expressions are key for indirect communication. The proposed system is focused on creating a robust Architecture using VGG Net. The system considers a face expression dataset (FER 2013) for analysing the present model. The proposed architecture extracts and learns various unique expressions from the facial images available in FER 2013. The feature mapped values are further used to train and model the VGG net enabled deep convolution networks. The proposed model also considers a music recommendation framework. appropriate music recommendations are done to the various expressions made by the person. The system achieved a higher accuracy of 92% and compared with existing state of art approaches on performance metrics.

Keywords— Face expressions, Machine learning, Face recognition, Voila Jones algorithm, Music recommendation.

I. INTRODUCTION

Face recognition models are developed in recent years rapidly, since most of the deep learning applications engaged with today's world are formulated with face recognition frameworks. Despite many face recognition datasets available publicly, Live face recognition also created more interest in research. In the similar context, obtaining facial landmark detection is one of the methods to achieve face recognition system[1]. Face reconstruction is an art of image processing technology and deep learning technology. Creating accurate geometry is mandatory in face reconstruction. To formulate the face expressions accurately, face landmarks mapping is an impacted task. Identification of face geometry is important to conclude the facial expressions[2].

For synthesis of exact emotions of the face, significant models are developed in the existing systems. Self-organized automated mapping model is used in facial expressions identification. For realistic emotion extraction such as happy, sad, anger, disgust, confused, hurtled etc, the combination of face expressions are needed. Some researchers focused on creating basic emotions to be sequentially considered to form the real expression depicted in the face. Expressions reflect unique landmark changes in the face, such as Eye brow changes, smile, closed eyes etc. [3]. In today's technological world, interaction between Human-computers has increased. facial expressions act as the means of intelligent

human-computer interface. A face expression recognition system is the additional tool for making unique and accurate recommendations to the users based on demand. In certain cases, facial expressions are also utilized in the medical domain for acquiring the result of antidepressants given for the patients[4].

Face expression can be detected from images captured as well as video streams. attributes are important to model face expressions. pipelined facial expressions are collectively used to find the emotions of the person. Augmented reality systems are developed using human computer interfacing systems[5]. Face expressions show unique signs like, head poses, lighting poses etc[6].

The proposed model is created using VGG 16 architecture. the detection of face expression, with appropriate music recommendations are made based on the detected expressions. The (FER2013) Facial expressions recognition dataset is used here for analysis. The dataset contains publicly collected 50000 face expression images of various classes.

- The proposed system preprocesses the image dataset, and normalizes the data before making the analysis.
- The face expression recognition system is developed using convolution neural network architecture. Many existing frameworks have been developed in recent years. In that deep learning convolutional neural network system acts as the foremost method that gives better results.
- Here VGG-16 architecture is developed. The VGG-16 with modified layers, suitable only for the connected dataset is focused here. The Model is customized to form a clear attribute extraction from various face feature learnings.
- A customized music recommendation system also developed here based on the detected expressions such as happy, sad, disgust, excited etc.
- Further contribution of the paper is described below with the below organized method.

The presented paper is formulated as a detailed Literature study described in Section II. followed by system tool selection, problem statement in section III. Further the proposed System architecture, related modules were described in Section IV. The proposed results obtained using the VGG

network architecture are explained in Section V. The paper is concluded further with future enhancement.

II. BACKGROUND STUDY

H. Li, et al., (2021) The author presented an additive feature extraction based facial expression detection system. Adaptive weighted estimation based Cognitive model of feature in balance is discussed here. Percentage system utilised FCR + data set and achieved the accuracy of 88% using transfer learning approaches and compared with existing state of art approaches.[7].

H. Zhang et al., (2019) The percentage system utilized the FER2013 dataset of facial expression detection and implemented a convolutional neural network architecture. The normalized face feature fusions are extracted. The present system achieved the accuracy of 88 % and utilised for image based computing applications. [8].

H. Mo et al., (2019) The author presented a system for face alignment based expression detection framework. Proposed model attitude multiclass face feature extraction technique. the compromised method achieved a reduced error rate compared with other existing systems. If the number of images used for the training process increases then the performance of the prediction system increases. The presented system act as a basic model for deriving a deep learning based face expression detection system with top[9].

S. L. Happy et al., (2017) Spontaneous expression detection Framework is modelled here to stop the data set consists of various emotions of individuals comprised together is useful to detect the real emotions of the individuals pitstopDatabase consists of 428 Facial expressions collectively gathered from spontaneous detection of camera. Through the process of video clipping based facial expression detection and self assessment method is used here to make the validation process.[10].

J. Zhang et al., (2022) The author presented a system where the correspondence between two different scenarios of face expressions is detected. The presented approach tests the non professional users by making different facial expressions and the validation is processed with the existing frameworks. Using geometric and conceptual frameworks the presented model is validated.[11].

M. Garcia et al., (2020) The author presented a neural network architecture enabled facial expression detection system using a massive data set collection of facial images. Impact of environment age, physical factors, the images of various individuals are collected for the training purpose. The test data are collected in real time camera for the process to with the deep neural network architecture to detect the exact emotion present in the face to stop[12].

H. Yang, et al., (2021) Using the approach of generalized adversal network(GAN), Facial changes Obtained in a progressive way or detected by the proposed algorithm. The proposed system focused mainly on specific face expression changes framework, using MORPH and CACD databases, the static work achieved the accuracy of 99% accuracy. The

quantitative measures are compared with the existing state-of-the-art approaches[13].

III. SYSTEM DESIGN

Many existing frameworks are implemented for facial expressions detections. Accurate localization of facial expressions are mandatory to make the recommendations related to the expressions. Existing works with the issue of false positive rate, that depicts the expression classification that provides false results. Considering the Problem identified, it is highly important to make a robust framework for facial expression detection. The proposed model is developed using Python IDE, with VGG net architecture. The presented system considers the FER2013 dataset for training the model. Python is the High level computing tool used for Machine learning, Deep learning and Image Processing frameworks.

IV. METHODOLOGY

A. System architecture.

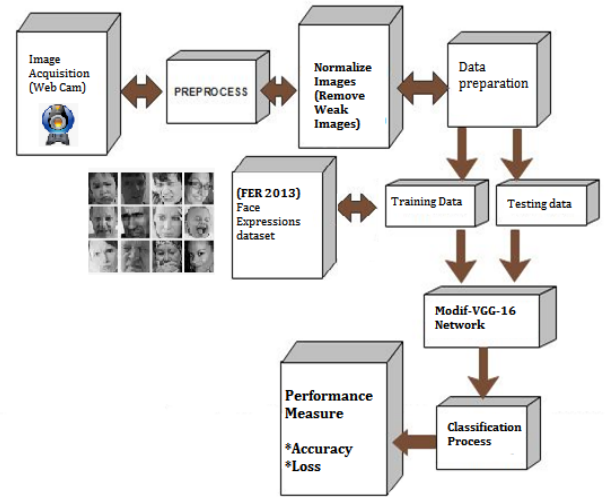


Fig. 1. System architecture of Proposed facial expression detector
Fig. 1. Shows the architecture of the proposed facial expression detector using VGG-16 Network architecture.

B. Image acquisition

Image capturing is done with the help of a webcam. The Live acquisition of video streams is given as input. Further, the face is detected from the streamed video. Once a face pattern is recognized by the system, image capturing is performed and saved into the backend.

C. Haar cascade technique

The Haar cascade algorithm is considered as one of the fastest feature extraction techniques in the image processing field. The edge detection, line detection and shape detection is the basic model extraction technique. The region of interest is first selected. based on the adjacent pixels present with the selected region, The features of the pixels associated are detected. The

differences between the pixel boxes are determined to make the training decision.

Cascade Filter

The Haar Cascade technique or the Viola Jones algorithm is based on Cascaded values of unique features. The features of relativity are determined by the weights obtained. These weights of the pixels selected in the region of interest are updated at every iterative analysis. The general form of face features cascading is depicted in the Fig. 2.

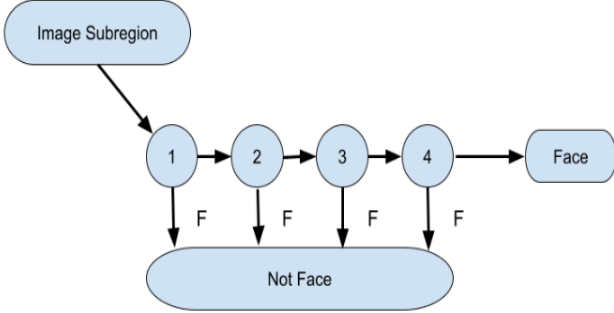


Fig. 2. Face feature Cascading process

Fig. 2. Shows the Face feature cascade process. Through Viola Jones algorithm, unique features of the face are extracted one by one and formulated into common robust features. The Cascade prediction is implemented using the OpenCV library.

D. VGG-16 Architecture

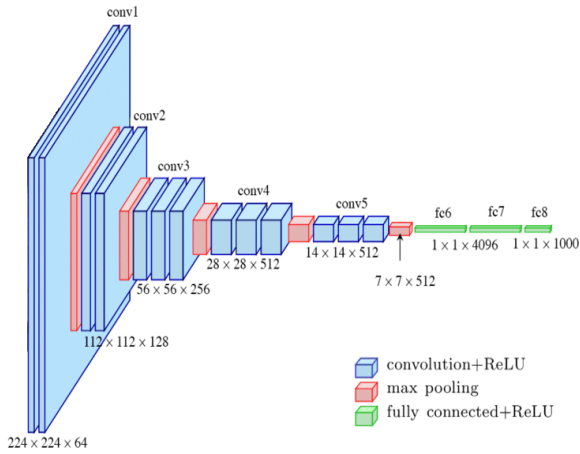


Fig. 3. VGG-16 Architecture

Fig. 3. shows the VGG-16 architecture with deep extraction of feature maps. The presented model utilized an input layer of 224x224x64, followed by the Convolution layer of 112x112x128 unique features mapped. a further number of convolution layers are mapped together in order to extract deep features. The extension of the features depends on the complexity of the input images. Convolution deep layer of 28x28x512 is formulated, finally with the very unique convolution filter of size 14x14x512 is modelled. The output layer holds the very robust feature of the input test image,

with the size of 1x1x1000 at the output layer. the Max pooling layer is connected between the output layer and ReLu layer. \

E. Implementation Summary

The proposed real time face expression detection framework is developed here. FER2013 face expression dataset is considered for analysis. The present approach is divided into two phases of operations. Such as Training process and Testing process.

- In the training process, the FER2013 dataset images are independently fetched into the VGG-16 Deep convolutional neural network architecture.
- The feature extraction process through Viola-Jones algorithm with Haar Cascade model is implemented for all the database images.
- Further, Learned features are saved for making the validation process with Test image collected from live webcam,
- The robust model is created using 80% of training images randomly selected from the FER2013 database, with 20% of testing image from the same folder.
- Based on the accuracy obtained from the training data and testing data, the Model performance is evaluated.
- The Live test image captured is fetched into the created model and further the interpretation is expanded.
- If the face is recognized with Haar Cascade algorithm, expressions are detected using VGG-16 network.
- A small Validating routine with saddle point estimation framework is modelled here. it considers the accuracy of prediction obtained and further based on the classified expressions in the face image, it recommends a Music to be played in the backend Speaker device.
- The Music list is a predefined pattern of songs with mixed emotions. Based on the expression, for example if a Happy expression is detected then, a song for excitement and happiness is played.

F. Evaluation metrics

The performance evaluations of the proposed system are measured using the Confusion matrix. The accuracy is calculated with respect to the obtained True positive Value, True negative value, False positive value and false negative value. Accuracy is calculated by the below formula. The loss function is calculated with the difference in predicted result with respect to the actual expected result for the given set of training and testing data.

$$Accuracy = \frac{TPV+TNV}{(TPV+TNV+FPV+FNV)}$$

V. RESULTS AND DISCUSSIONS

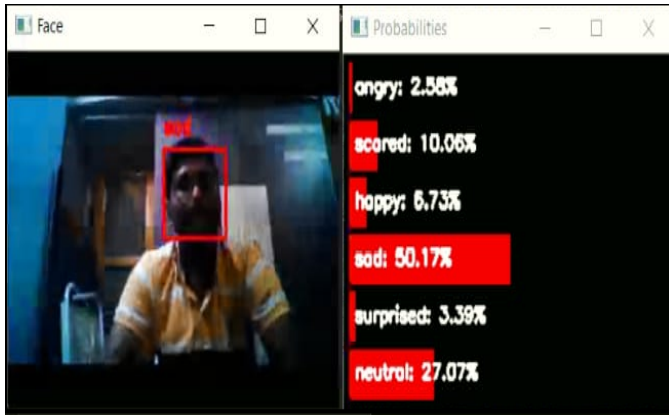


Fig. 4. Simulation result of expression detection -Sad

Fig. 4. Shows the Simulation result of Live video streaming, where face is detected and shown with a bounding box. Further the face expression with obtained accuracy value is simultaneously shown. At an accuracy of 50.17% as Sad is detected and 27.07% neutral is detected.

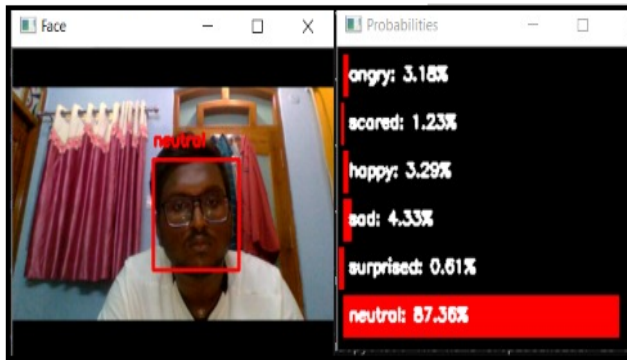


Fig. 5. Simulation result of expression detection- Neutral

Fig. 5. Shows the Simulation result face expression with obtained accuracy value is simultaneously shown. At an accuracy of 87.36% as expression Neutral is detected.

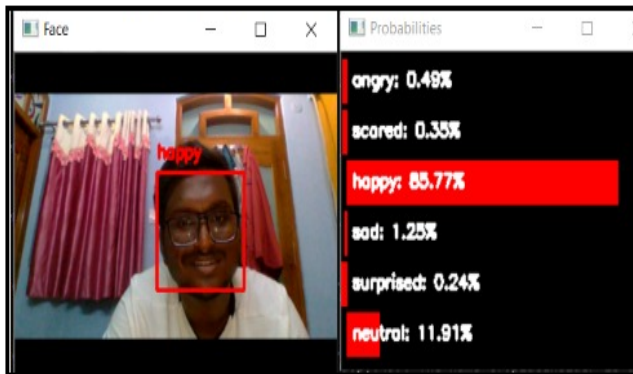


Fig. 6. Simulation result of expression detection -Happy

Fig. 6. Shows the Simulation result face expression with obtained accuracy value is simultaneously shown. At an accuracy of 85.77% as expression Happy is detected.

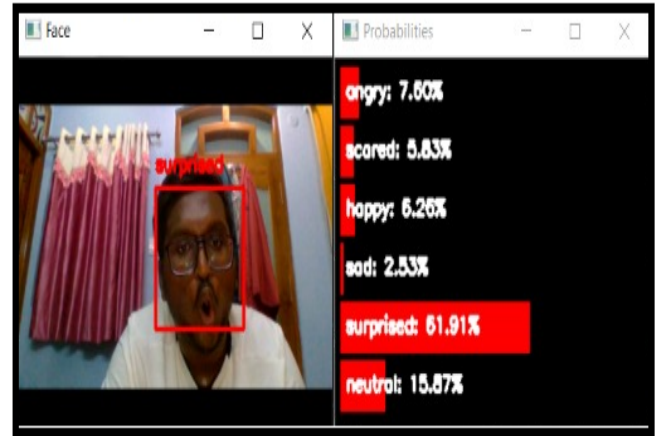


Fig. 7. Simulation result of expression detection - Surprised

Fig. 7. Shows the Simulation result face expression with obtained accuracy value is simultaneously shown. At an accuracy of 51.91% as expression Surprised is detected.

VI. CONCLUSION

Emerging technologies utilized artificial intelligence enabled systems for many products. Various core products depend on face recognition for security and sourcing. Face recognition systems act as one of the foremost needs in many real time emerging products. Facial expressions are helpful for making automated reviews of online courses, products, also to detect the emotional effect of the individuals. expressions are a direct impact of emotions. The proposed framework is modelled using Modif-VGG16 network for facial expression detection as well as Saddle point recommendation framework for music suggestions. The proposed structure is validated with the FER-2013 dataset. The System achieved the accuracy of 92% on an average and further the system can be improved with more real time dataset with the help of transfer learning approaches.

REFERENCES

- [1] F. Liu, Q. Zhao, X. Liu and D. Zeng, "Joint Face Alignment and 3D Face Reconstruction with Application to Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 664-678, 1 March 2020, doi: 10.1109/TPAMI.2018.2885995.
- [2] M. Emambakhsh and A. Evans, "Nasal Patches and Curves for Expression- Robust 3D Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 5, pp. 995-1007, 1 May 2017, doi: 10.1109/TPAMI.2016.2565473.
- [3] F. Qu, S. Wang, W. Yan, H. Li, S. Wu and X. Fu, "CAS(ME)S²: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition," in IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 424-436, 1 Oct.-Dec. 2018, doi: 10.1109/TAFFC.2017.2654440.
- [4] M. Wairagkar et al., "Emotive Response to a Hybrid-Face Robot and Translation to Consumer Social Robots," in IEEE Internet of Things Journal, vol. 9, no. 5, pp. 3174-3188, 1 March 2022, doi: 10.1109/IIOT.2021.3097592.

- [5] S. Tulyakov, L. A. Jeni, J. F. Cohn and N. Sebe, "Viewpoint-Consistent 3D Face Alignment," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 9, pp. 2250-2264, 1 Sept. 2018, doi: 10.1109/TPAMI.2017.2750687.
- [6] R. Zatarain-Cabada, M. L. Barrón-Estrada, F. González-Hernández and H. Rodríguez-Rangel, "Building a Face Expression Recognizer and a Face Expression Database for an Intelligent Tutoring System," 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), 2017, pp. 391-393, doi: 10.1109/ICALT.2017.141.
- [7] C. Petpairote, S. Madarasmi and K. Chamnongthai, "A pose and expression face recognition method using transformation based on single face neutral reference," 2017 Global Wireless Summit (GWS), 2017, pp. 123-126, doi: 10.1109/GWS.2017.8300485.
- [8] H. Li, N. Wang, X. Ding, X. Yang and X. Gao, "Adaptively Learning Facial Expression Representation via C-F Labels and Distillation," in *IEEE Transactions on Image Processing*, vol. 30, pp. 2016-2028, 2021, doi: 10.1109/TIP.2021.3049955.
- [9] H. Zhang, A. Jolfaei and M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing," in *IEEE Access*, vol. 7, pp. 159081-159089, 2019, doi: 10.1109/ACCESS.2019.2949741.
- [10] H. Mo, L. Liu, W. Zhu, S. Yin and S. Wei, "Face Alignment With Expression- and Pose-Based Adaptive Initialization," in *IEEE Transactions on Multimedia*, vol. 21, no. 4, pp. 943-956, April 2019, doi: 10.1109/TMM.2018.2867262.
- [11] S. L. Happy, P. Patnaik, A. Routray and R. Guha, "The Indian Spontaneous Expression Database for Emotion Recognition," in *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 131-142, 1 Jan.-March 2017, doi: 10.1109/TAFFC.2015.2498174.
- [12] J. Zhang, K. Chen and J. Zheng, "Facial Expression Retargeting From Human to Avatar Made Easy," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1274-1287, 1 Feb. 2022, doi: 10.1109/TVCG.2020.3013876.
- [13] M. Garcia Villanueva and S. Ramirez Zavala, "Deep Neural Network Architecture: Application for Facial Expression Recognition," in *IEEE Latin America Transactions*, vol. 18, no. 07, pp. 1311-1319, July 2020, doi: 10.1109/TLA.2020.9099774.
- [14] H. Yang, D. Huang, Y. Wang and A. K. Jain, "Learning Continuous Face Age Progression: A Pyramid of GANs," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 499-515, 1 Feb. 2021, doi: 10.1109/TPAMI.2019.2930985.
- [15] V. Wati, K. Kusrini and H. A. Fatta, "Real Time Face Expression Classification Using Convolutional Neural Network Algorithm," 2019 International Conference on Information and Communications Technology (ICOIACT), 2019, pp. 497-501, doi: 10.1109/ICOIACT46704.2019.8938521.
- [16] M. Phankokkrud and P. Jaturawat, "Influence of facial expression and viewpoint variations on face recognition accuracy by different face recognition algorithms," 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017, pp. 231-237, doi: 10.1109/SNPD.2017.8022727.
- [17] X. Chai, J. Chen, C. Liang, D. Xu and C. -W. Lin, "Expression-Aware Face Reconstruction via a Dual-Stream Network," in *IEEE Transactions on Multimedia*, vol. 23, pp. 2998-3012, 2021, doi: 10.1109/TMM.2021.3068567.
- [18] B. Vishnudharan and K. Anusudha, "A discriminative model for facial expression recognition using local directional number pattern," 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2016, pp. 349-352, doi: 10.1109/ICCICCT.2016.7987972.