ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

Face Emotion Detection System Using Deep Learning

M. Anuradha[1], K. Mokaksh[2],T. Meghana[3], V. Pujitha[4], N. Harsha[5]
Computer Science and Engineering Amrita Sai Institute of Science & Technology Paritala, Andhra Pradesh , India

ABSTRACT

Now-a-days with the continued development of artificial intelligence facial emotion recognition has become more popular. The emotion recognition plays a major role in interaction technology. In interaction technology the verbal components only play a one third of communication and the non-verbal components plays a two third of communication. A facial emotion recognition (FER) method is used for detecting facial expressions. Facial expression plays a major role in expressing what a person feels and it expresses inner feeling and his or her mental situation or human perspective. This paper aims to identify basic human emotions with the combination of gender classification and age estimation. The facial emotions such as happy, sad, angry, fear, surprised, neutral emotions are considered as basic emotions. Here proposes a real time facial emotion recognition system based on You Look Only Once (YOLO) version 2 architecture and a squeeze net architecture. The yolo architecture is a real time object detection system. Here it used for identify and detect faces in real time. These images are captured by using anchor boxes for accuracy. The second architecture is squeezing net and is used for gender classification and age estimation. It provides significant, accurate object detection and extracts high-level features that help to achieve tremendous performance to classify the image and detecting objects. Both the architectures provide accurate result than other methods with the large no of hidden layers and cross validation in the neural network.

Keywords: Facial Emotion Recognition, Deep Learning, Convolutional Neural Network (CNN), YOLOv2, Squeeze Net, Real-Time Face Detection, Facial Expression Classification, Age Estimation, Gender Classification.

I. INTRODUCTION

A face detection includes classifying image into two classes: one with faces (targets), and other containing the background (clutter) which needs to be removed. Commonalities exist between faces, they vary differently in terms of age, skin color and facial expression, this becomes difficult due to this commonality. The further problem is complicated by differing lighting conditions, image qualities and geometries, partial occlusion and disguise is also a possibility. A face detector should be able to detect the presence of any face under different set of lighting conditions in any background condition. The face detection analysis can be broken into two tasks. The first task is a classification task that takes some random regions of image as input and outputs a binary value of yes or no, indicating if there are any faces present in the image. The other task is the face localization task which is to take an image as input and output the location of any face or faces within that image as some bounding box/boxes with (x, y, width, height). Smart robots can be built by automatic facial expression application. These bots can be used in various applications like interactive games and service center. There are six universal expressions according to Ekman they are fear, disgust surprise, anger, sadness and happiness. Face variances can be observed to recognize these expressions. For example,

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

we can say a person is happy which can be identified as a gesture of smile by tightened eyelids and raised lips corners. A person's internal states, social communication and intentions are indicated by change in facial expressions. Many applications in many areas like human emotions analysis, natural Human computer interaction, image retrieval and talking bots have a large effect on them by automatic facial expression detection. Face Recognition with Histogram of Oriented Gradients using CNN detection has been an impacting issue in the technological community as human beings fined facial expressions one of the most natural and powerful means to express their intentions and emotions. Last stage of the system is facial expression detection.

There are basically three steps in training procedure in expression recognition systems named as feature learning, classifier construction and feature selection. Feature learning stage is first; feature selection is second and the last one is classifier construction. Only learned facial expressions variations among all features are extracted after feature learning stage. Facial expression is then represented by the best features which are chosen by feature selection. Not only maximizing inter class variation but they also should try to minimize the intra class variations of expressions not only maximizing inters class variation but they also should minimize the intra class variations of expressions. Because same expressions of different individuals in image are far from each other in pixel's space so minimizing the intra class variation of expressions is a problem. Techniques that can be used for facial detection are YOLO, SDD, RCNN, Faster RCNN.

II. RELATED WORK

1 EXISTING SYSTEM:

The emotion recognition plays a major role in interaction technology. In interaction technology the verbal components only play a one third of communication and the non-verbal components plays a two third of communication. A facial emotion recognition (FER) method is used for detecting facial expressions. Facial expression plays a major role in expressing what a person feels and it expresses inner feeling and his or her mental situation or human perspective

2 PROPOSED SYSTEM:

This paper aims to identify basic human emotions with the combination of gender classification and age estimation. The facial emotions such as happy, sad, angry, fear, surprised, neutral emotions are considered as basic emotions. Here proposes a real time facial emotion recognition system based on You Look Only Once (YOLO) version 2(FRAME PER SECOND) architecture and a squeezenet architecture. The yolo architecture is a real time object detection system. Here it used for identify and detect faces in real time. These images are captured by using anchor boxes for accuracy. The second architecture is squeezenet and is used for gender classification and age estimation. It provides significant, accurate object detection and extracts high-level features that help to achieve tremendous performance to classify the image and detecting objects. Both the architectures provide accurate result than other methods with the large no of hidden layers and cross validation in the neural network.

III.METHODOLOGY

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

Here's the corrected and professionally formatted version of your content, suitable for inclusion in a project report or research paper:

1. Problem Definition

The objective is to recognize and classify human facial emotions from static images or video frames using a deep learning-based approach. The system aims to identify emotions such as happiness, sadness, anger, surprise, disgust, fear, and neutrality.

2. Data Collection

Utilize publicly available facial expression datasets, including:

- **FER-2013** (Facial Expression Recognition 2013)
- **CK+** (Extended Cohn-Kanade Dataset)
- **AffectNet**

These datasets provide thousands of labeled facial expression images captured under various lighting conditions and facial orientations.

3. Data Preprocessing

- Convert images to grayscale *(optional, depending on model architecture)*
- Resize all images to a uniform resolution (e.g., 48x48 or 64x64 pixels)
- Normalize pixel values to the **[0, 1]** range
- Apply data augmentation (e.g., rotation, zoom, flipping) to enhance dataset diversity and reduce overfitting
-

4. Model Design and Architecture

Adopt a Convolutional Neural Network (CNN) for emotion recognition due to its proven efficacy in image classification. The model may include:

- **Convolutional layers** for feature extraction
- **Pooling layers** for dimensionality reduction
- **Dropout layers** to prevent overfitting
- **Fully connected (dense) layers** for high-level feature interpretation
- A **Softmax output layer** for multi-class emotion classification

Optionally, fine-tune pre-trained models such as VGG16, ResNet, or MobileNet to leverage transfer learning for improved accuracy.

5. Model Training

- Split the dataset into training, validation, and **testing** subsets (e.g., 70% / 15% / 15%)
- Use appropriate loss functions such as categorical cross-entropy Apply optimizers like Adam or **SGD**
- Integrate callbacks including:
  - **Early stopping** to prevent overfitting
  - **Reduce LR On Plateau** to dynamically adjust learning rates during training

6. Model Evaluation

Evaluate the model's performance using:

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

- **Accuracy**
- **Precision**, **Recall**, and **F1-Score**
- **Confusion matrix** to analyze per-class classification performance
- Apply **cross-validation** or **k-fold validation** for increased robustness and generalizability

7. Deployment (Optional)

Deploy the trained model into a real-time system using:

- **Input methods**: Webcam or image upload interface
- **Backend frameworks**: Flask, FastAPI, or Django
- **Frontend interface**: To display predicted emotions in real-time or upon image submission

8. Conclusion and Future Work

- Summarize the model's performance, strengths, and limitations
  Suggest areas for future improvement, such as:
  - Incorporating **3D facial data**
  - Performing **temporal emotion analysis** in videos
  - Exploring **multimodal emotion detection** (e.g., combining facial expressions with voice or physiological signals)

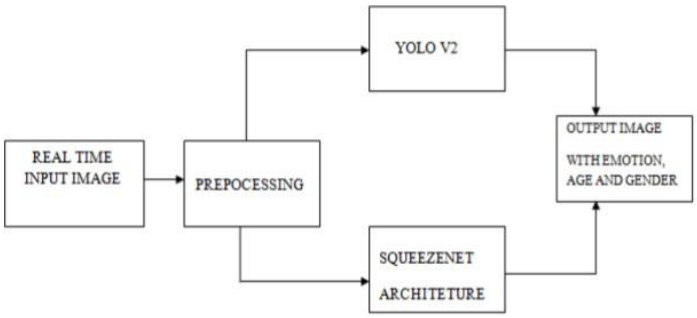

Figure 1 : Shows the system Architecture

DATA FLOW DIAGRAM:

1.The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

2.The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

3.DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 2
www.ijisea.org

4.DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

III. CONCLUSION

The FlixLens system plays a crucial role in modern content exploration by leveraging datadriven technologies to enhance content discovery, personalized recommendations, and user engagement. By analyzing viewing patterns and genre preferences, this system assists users in identifying relevant entertainment options, enabling a more tailored streaming experience. This approach not only improves content accessibility but also helps in reducing the time spent searching for suitable shows or movies. One of the most significant advantages of this system is its potential to improve content recommendation accuracy. Many users, particularly those overwhelmed by vast content libraries, struggle to find engaging content due to the sheer volume of options available. By providing an interactive and data-driven recommendation tool, Flix Lens ensures that users receive curated suggestions without the need for extensive manual browsing. This enhances the streaming experience while reducing frustration associated with content selection. Additionally, predictive content models contribute to better audience engagement by optimizing the discovery process. Personalized recommendations based on viewing habits allow users to explore new genres and trending titles, thereby broadening their entertainment choices. By encouraging intelligent content curation, this system plays a crucial role in increasing platform retention and enhancing user satisfaction. Despite its benefits, the effectiveness of predictive content systems depends on multiple factors. The accuracy of recommendations is directly linked to the quality and diversity of the viewing data used. Ensuring proper data collection, storage, and validation is essential for minimizing irrelevant suggestions and improving system reliability. Furthermore, integrating such systems with streaming analytics and user feedback mechanisms is necessary to refine recommendations and enhance personalization. In conclusion, the Flix Lens system has the potential to significantly transform content exploration by enabling precise recommendations, improved engagement, and an enhanced streaming experience. By relying on data-driven insights and ensuring seamless integration with streaming platforms, this technology can pave the way for a more efficient, engaging, and personalized entertainment ecosystem. Addressing the associated challenges and fostering collaboration among content providers will be crucial in maximizing its impact, ultimately contributing to a smarter and more enjoyable streaming experience.

ACKNOWLEDGEMENT(S):

It gives me immense pleasure to thank my guide Dr. M. SREEDEVI Ph.D Professor , CSE for having been a source of constant inspiration, precious guidance and generous assistance during the project work. I deem it as a privilege to have worked under her able guidance. Without her close monitoring and valuable suggestions, this work would not have taken this shape. I feel that his help is substitutable and unforgettable.

I take the privilege to record my thanks to H.O.D of CSE Dr. P CHIRANJEEVI Ph.D. for his encouragement, cooperation and valuable suggestions to complete project in time.

REFERENCES:

[1] Jumani, S.Z., Ali, F., Guriro, S., Kandhro, I.A., Khan, A. and Zaidi, A., 2019. Facial Expression Recognition with Histogram of Oriented Gradients using CNN. Indian Journal of Science and Technology, 12, p.24.

[2] J. R. Barr, L. A. Cament, K. W. Bowyer, and P. J. Flynn. Active clustering with ensembles for social structure extraction. In Winter Conference on Applications of Computer Vision, pages 969–976, 2014

[3] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[4] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).R. Goh, L. Liu, X. Liu, and T. Chen. The CMU face in action (FIA) database. In International Conference on Analysis and Modelling of Faces and Gestures, pages 255–263. 2005.

[5] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In Computer Vision and Pattern Recognition, pages 529–534, 2011

[6] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patchbased probabilistic image quality assessment for face selection and improved video-based face recognition. In Computer Vision and Pattern Recognition Workshops, pages 74–81, 2011.

[7] B. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, P. J. Flynn, and S. Cheng. The challenge of face recognition from digital point-and-shoot cameras. In Biometrics: Theory Applications and Systems, pages 1–8, 2013

[8] N. D. Kalka, B. Maze, J. A. Duncan, K. A. O Connor, S. Elliott, K. Hebert, J. Bryan, and A. K. Jain. IJB– S: IARPA Janus Surveillance Video Benchmark. In IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2018.

[9] M. Singh, S. Nagpal, N. Gupta, S. Gupta, S. Ghosh, R. Singh, and M. Vatsa. Cross-spectral crossresolution video database for face recognition. In IEEE International Conference on Biometrics Theory, Applications and Systems, 2016 [10] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2879–2886.

[10] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun, “Joint cascade face detection and alignment,” in Proc. Eur. Conf. Comput. Vis., 2014, vol. 8694, pp. 109–122. [12] D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 2650– 2658

[11] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.

[13] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE

[14] Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1915–1929. [16] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.

[15] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 580–587.

[16] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, International Conference on Learning Representations.

[17] DESivaram, M., Porkodi, V., Mohammed, A.S. and Manikandan, V., 2019. Detection Of Accurate Facial Detection Using Hybrid Deep Convolutional Recurrent Neural Network. ICTACT Journal on Soft Computing, 9(2).

[18] Sivaram, M., Porkodi, V., Mohammed, A.S. and Manikandan, V., 2019. Detection of Accurate Facial Detection Using Hybrid Deep Convolutional Recurrent Neural Network. ICTACT Journal on Soft Computing, 9(2).

[19] Shafiee, M.J., Chywl, B., Li, F. and Wong, A., 2017. Fast YOLO: a fast you only look once system for real-time embedded object detection in video. arXiv preprint arXiv:1709.05943.

[20] Ravidas, S. and Ansari, M.A., 2019. An Efficient Scheme of Deep Convolution Neural Network for Multi View Face Detection. International Journal of Intelligent Systems and Applications, 11(3), p.53.

[21] Gupta, S., Gupta, N., Ghosh, S., Singh, M., Nagpal, S., Vatsa, M. and Singh, R., 2019. FaceSurv: A benchmark video dataset for face detection and recognition across spectra and resolutions. challenge, 14, p.18.

[22] Guo, G., Wang, H., Yan, Y., Zheng, J. and Li, B., 2019. A Fast Face Detection Method via Convolutional Neural Network. Neurocomputing.