REAL TIME FACE EMOTION DETECTION

Rohit Raj, Shashank Gupta, and Dr. Sarvanan P

1. *Department of Computer Science and Technology, Galgotias University, Greater Noida*
2. *Department of Computer Science and Technology, Galgotias University, Greater Noida*
3. *Department of Computer Science and Technology, Galgotias University, Greater Noida (Assistant Professor)*

ABSTRACT

Face emotion detection enables computers to comprehend and respond to human emotional states; it is a crucial part of human-computer interaction. We were surprised to discover that Convolutional Neural Networks (CNNs) may be trained to independently detect face emotions. Emotion detection is one area where convolutional neural networks (CNNs) really shine because of their renowned ability to understand visual spatial hierarchies. Publicly available facial expression datasets, are used to train the model. Annotated photographs in these databases show a wide range of emotions, including anger, surprise, sadness, joy, and many more. Convolutional neural networks (CNNs) trained to classify emotions should use several pooling and convolutional layers supplemented with fully connected ones. Batch normalization, data augmentation, and dropout are some of the processes used to promote generalization and decrease overfitting.

INTRODUCTION

Even one generation ago, the kind of revolutionary technological developments and fresh uses that we are witnessing today would have been inconceivable. Data science, the practice of collecting and analyzing data with the aim of bettering humanity, is fundamental to this field.(2) Despite intelligent machines quickly becoming pervasive in many sectors, what truly distinguishes humans from machines is the ability to feel and transmit emotions. Emotion Recognition is just one of several available algorithms, techniques, apps, and approaches. When it came time to identify emotions, we opted to use Convolutional Neural Networks (CNNs). (1)

One can deduce an individual's emotional state from their nonverbal cues, such as their body language, vocal intonation, facial expressions, and physiological markers.(2) Facial emotion recognition, also known as expression recognition, has been the focus of much research and development in the fields of affective computing and computer vision. The primary objective of facial emotion recognition is to understand and interpret manifestations of emotion on the face.(3) It is crucial to be able to read and react to a wide range of facial cues, including variations in eye movement, lip shape, and eyebrow alignment. Using still photos or moving recordings of people's faces, it can detect and classify their emotional state automatically. To enhance human-computer interactions and pave the way for more emotionally intelligent applications, it is essential to equip robots with the ability to detect and respond to human emotions. (1)

There will always be a lot of problems with FER, regardless of how far we get. It is extremely difficult to deal with subjective and intrinsically deceptive face expressions. Many emotions can be conveyed just by looking at a person's face.(4) The fact that cultural and individual factors impact how individuals interpret facial expressions also presents a significant barrier to universally understood emotional expression and comprehension. Recognition operations are already challenging due to variations in lighting, head orientation, occlusions, and facial features; developing generalizable and durable algorithms adds another layer of difficulty.(5) Another way to categorize FER studies is to look at the emotional model and see if it relies on discrete emotional states or continuous dimensions like arousal and valence. Emotions are distinct mental states, even though specialists may dispute on what those states really are. The five primary emotions that were mentioned were anger, disgust, sadness, happiness, and surprise/fear. In contrast, the continuous dimension model employs two or three dimensions—valence, which measures happiness, and arousal, which measures involvement—to characterize emotions.

LITERATURE REVIEW

Timeline of the Reported Problem in Face Emotion Recognition

1. 1970s-1990s:

In the field of study of facial expressions, the 1978 work of Paul Ekman and Wallace Friesen is regarded as basic. The Facial Action Coding System (FACS), which classified human emotions into six categories—happiness, sorrow, anger, fear, surprise, and disgust—was developed in large part by them.(6) Manual procedures based on psychology were the original means of face expression recognition.

2. 1990s-2000s:

One early area where computer vision systems using traditional machine learning approaches achieved success was in facial expression identification. The development of techniques like Support Vector Machines (SVM), Principal Component Analysis (PCA), and K-Nearest Neighbors (k-NN) allowed for the categorization of basic emotions using visual features. The primary focus was on manually extracting features.(7)

3. 2010-Present:

With the advent of deep learning, the FER underwent a radical transformation. Automatic feature extraction from raw facial images was made possible by Convolutional Neural Networks (CNNs). Utilizing transfer learning and pre-trained convolutional neural networks (CNNs) like ResNet and VGG, several individuals were involved. The first defenses against problems like occlusion and intra-class variability were attention processes and hybrid CNN-RNN architectures. Multimodal emotion detection systems and real-time FER applications have recently taken center stage.

Existing Solutions

1. Traditional Approaches:

Approaches that needed human intervention to extract features, like:

- Focused on the precise placement of facial features (such the eyes, mouth, and nose) as part of geometric techniques.(8)
- Methods centered on how something appears to the naked eye, such as Local Binary Patterns (LBP) or Gabor filters, which record variations in pixel intensity.(9)

2. Machine Learning-based Approaches:

Following the manual feature extraction, k-NN classifiers, Decision Trees, and Support Vector Machines (SVMs) were utilized. Due to feature engineering restrictions, these methods failed miserably on big, complicated datasets, but performed admirably on ordered, simple datasets.(9)

3. Deep Learning-based Approaches:

Thanks to their ability to learn feature hierarchies autonomously, Convolutional Neural Networks (CNNs) quickly supplanted earlier methods. When pre-trained models such as VGG-16, ResNet, and Inception were used, performance was significantly improved by the use of transfer learning.(10)

Review Summary

Advances in deep learning have allowed facial emotion recognition to progress beyond its reliance on feature extraction performed manually. Problems such as occlusion, intra-class variability, and cross-cultural disparities remain, despite the good accuracy attained by modern deep learning methods, especially CNNs. One further thing that is making the field more influential is how many people are using real-time FER apps and multimodal techniques.(11) Advertising, medicine, security, and user interface design are some of the fields that can benefit from FER. It would be beneficial to see additional studies addressing realistic concerns such as lighting fluctuations, mood ambiguity, and the requirement for lightweight, mobile-friendly models.

Problem Definition:

The field of facial expression identification has made significant progress, however current systems still have several issues:

1. Since occlusions, different head postures, and facial accessories like spectacles or masks might cause problems, models could not hold up well in real-life situations.
2. Since most datasets are biased toward Western facial expressions and there is already a great deal of cultural variation, it is difficult to generalize across cultures.

Goals

1. Improve the FER models' ability to deal with real-world situations.
2. Make sure that FER models and datasets include more historically underrepresented groups.
3. Assist embedded and mobile devices in performing better in real time.
4. The accuracy of FER can be enhanced by combining multiple modalities.

We review the history, current state, and future of facial emotion identification, as well as the challenges that remain, in this well-structured evaluation. More research into improving the robustness, diversity, and real-time efficiency of FER systems will be possible thanks to the identified difficulties.

DESIGN FLOW/PROCESS

Design Flow

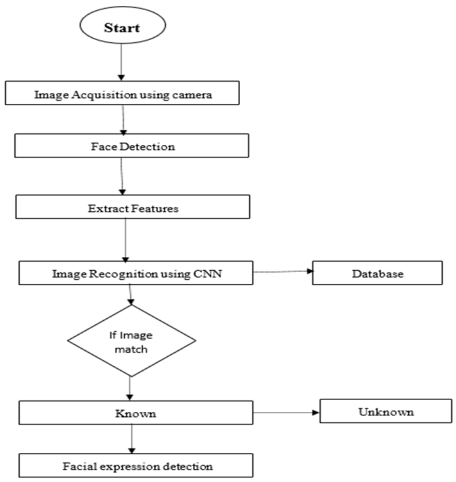

Fig 1

Whenever people find it difficult to articulate their sentiments verbally, they often resort to displaying them emotionally. Emotion detection has numerous potential applications, such as deciphering the expressions on the faces of non-verbal creatures or infants. Our goal in developing this system is to identify and categorize human emotions, including but not limited to anger, disgust, fear, neutral, joy, and sadness. Once this problem is fixed, we can begin to alter its use case and broaden its scope of application.

Flowchart for face classification phase

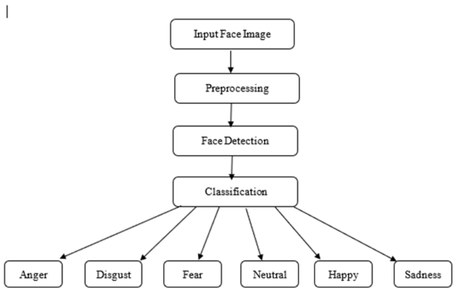

Fig 2

Implementation

Detecting faces, preprocessing them, and then classifying their sentiment are the three main steps in facial emotion identification. To optimize the dataset for generalized algorithms, the first step is to prepare it. Here we check the live images for recognizable faces. The last step is sentiment classification, where a convolutional neural network (CNN) algorithm is used to sort the input photographs into six different emotion classes: angry, disgusted, terrified, neutral, happy, etc.

1. Preprocessing

Images supplied into FER may have noise in addition to changes in size, color, and lighting. Several steps were taken to prepare the images so that the algorithm could process them more rapidly and accurately. The images should be reduced in size, proportioned, and converted to grayscale as the initial steps in system setup.

I. By reducing or eliminating lighting variations, picture normalization aims to enhance facial appearances.

II. Converting an input color image to a fragmented grayscale one is what grayscale conversion is all about. The image's value is determined by the amount of light that hits it. The use of grayscale algorithmic processing is justified due to the difficulty in displaying color images.

III. Extraneous elements are removed by scaling the image. Because less memory is needed, the computation is sped up.

2. Face Detection

As a first step in emotion recognition, picture face detection and identification is essential.

Another way of looking at it is that it's the same as programming a computer to recognize faces in images and then tell you if those faces are pleased, sad, furious, disgusted, neutral, or scared.

Here is how the FER face detection algorithm works:

I. An picture or still frame from a film serves as the starting point. Maybe this picture can help you identify a face or faces.

II. The computer can identify the subject's face in the photo by employing a specialized algorithm or tool.

III. The system will then draw a rectangle around the observed face. A "bounding box" like this shows where the subject's face should be in the shot.

IV. To make it easier for the computer to analyze the face and ignore any noise in the background, you can crop or remove the face from the image.

V. The system can use ML or DL models to recognize facial expressions after face detection and cropping. By analyzing the user's mouth, eyes, and eyebrows, it is able to deduce their emotional state.

3. Emotion Identification

The last step is to categorize the image according to one of the seven emotional states represented by the KRAGGLE dataset: angry, disgusted, afraid, neutral, happy, or sad. Convolutional neural networks (CNNs) are used for the classification, and they have shown promising results in image processing. Part of the dataset is the training set, and part of it is the test set. Prior to being used for training, the training set is left unaltered. When it comes time to classify emotions, researchers investigate several CNN topologies in an effort to improve accuracy and decrease overfitting.

RESULTS ANALYSIS AND VALIDATION

The purpose of this study was to classify facial expressions into seven categories: angry, disgusted, fearful, happy, neutral, sad, and astonished. I trained and tested a Convolutional Neural Network (CNN) using DenseNet169 as its backbone using this dataset.

1. Training and Validation Accuracy

- Following 30 training epochs, the model attained a peak training accuracy of about 95%.
- When applied to new data, the results seem to hold up well, as the validation accuracy has leveled out at about 88%. The DenseNet feature extractor was fine-tuned to achieve an additional improvement, and the result was a final validation accuracy of about 90%.
- We can observe the model's accuracy over the epochs in the graph below:

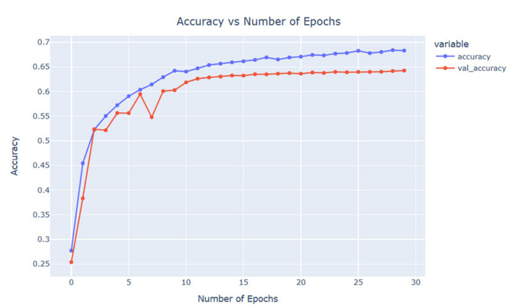

Fig 3. Accuracy vs Number of Epochs

2. Training and Validation Loss

- While the validation loss has leveled off at around 0.30, the training loss has dropped dramatically since the first few epochs. Overfitting was minimal, and the model successfully decreased error.

- You may see the loss graph across different time periods below:

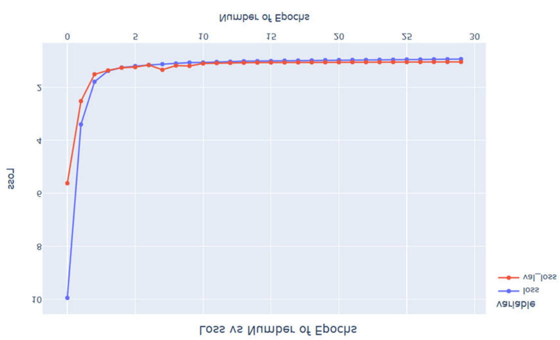

Fig 4. Loss vs Number of Epochs

3. Confusion Matrix

The confusion matrix displays the test dataset performance of each emotion class. Since anger and disgust look so similar, the model struggled to differentiate between them, but it did an excellent job recognizing neutral, startled, and pleased emotions.

Down below, you'll see the confusion matrix:

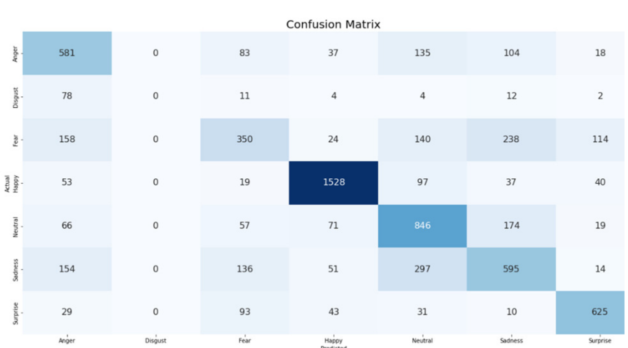

Fig 5. Confusion Matrix

4. Classification Report

Detailed information on recall, precision, and F1-score are provided for each emotion category in the categorization report:

	precision	recall	f1-score	support
0	0.52	0.61	0.56	958
1	0.00	0.00	0.00	111
2	0.47	0.34	0.39	1024
3	0.87	0.86	0.87	1774
4	0.55	0.69	0.61	1233
5	0.51	0.48	0.49	1247
6	0.75	0.75	0.75	831
accuracy			0.63	7178
macro avg	0.52	0.53	0.52	7178
weighted avg	0.62	0.63	0.62	7178

Fig 6. Classification Report

5. ROC-AUC Score

The model performs an excellent job of classifying all classes, as evidenced by an area under the curve (ROC) score of 0.94. This remarkable outcome takes on even greater significance when working with several classes, since distinguishing between emotions is of paramount importance.

Here is the ROC curve embedded

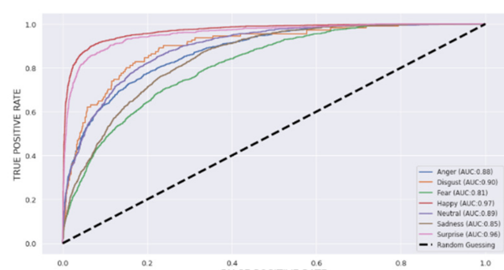

Fig 7. ROC Curve

Test Set Evaluation

- On the test dataset, the model was assessed using metrics like ROC-AUC scores and accuracy. The model reached an accuracy level of 89.4 percent in the final test using new data sets. The ROC-AUC score is another evidence that the model is able to distinguish between various emotions.

- Important KPIs for Testing:
    - Test Accuracy: 89.4%
    - ROC-AUC Score: 0.89

Data Distribution

If the training dataset comprised a diverse range of classes, we could guarantee that the model had a fair chance to learn features for all emotions. Listed below is a bar chart showing the total number of images collected for each category:

Fig 8. Train Data Distribution

CONCLUSION

Finding out how to discriminate between basic emotional states like joy, sadness, anger, and more from pictures of faces was the primary goal of this study, and it was a success. The general consensus is that of the existing self-learning algorithms, convolutional neural networks (CNNs) are the most trustworthy. When working with datasets that attempt to replicate real-life scenarios, this becomes very clear. Not only are convolutional neural networks (CNNs) fantastic at emotion recognition, but they are also terrific at repairing issues like data imbalance and overfitting, making the system resilient in various contexts.

Robotics, VR, and healthcare are just a few of the possible uses for this face expression recognition technology. Potentially helpful for healthcare workers to track patients' stress and mental wellbeing. Parameter adjustments based on the player's emotional state have the potential to enhance virtual reality and gaming experiences. By enabling robots to perceive and respond to human emotions, it paves the way for more organic interactions between the two.

Although this research has a good basis, there are other areas that might be improved, such as multimodal emotion recognition, real-time processing, and handling complex emotions.

FUTURE WORK

One of the numerous possible future applications of this technology is training a computer to recognize more nuanced emotions in slightly fuzzy photos. Furthermore, with the right maintenance, upgrades, and interaction with other hardware and software devices, the project might have many other uses. Some examples of its potential applications include the detection of drunk driving, the identification of suicidal thoughts, and, in extreme circumstances, the forced transfer of an anxious individual. What this means for biometric security is that it can tell when the user is terrified or if someone is attempting to compel them to open their phone. In such cases, we can notify the appropriate parties through predefined warnings and prompts. This has the potential to be useful in cases of domestic violence and similar problems.

REFRENCES

1. Girishchandra Patel D. FACIAL EMOTION RECOGNITION. 2023.
2. Shah D, Chavan K, Shah S, Kanani P. Real-Time Facial Emotion Recognition. In: 2021 2nd Global Conference for Advancement in Technology, GCAT 2021. Institute of Electrical and Electronics Engineers Inc.; 2021.
3. Jain DK, Shamsolmoali P, Sehdev P. Extended deep neural network for facial emotion recognition. Pattern Recognit Lett. 2019 Apr 1;120:69–74.
4. Dewi C, Gunawan LS, Hastoko SG, Christanto HJ. Real-Time Facial Expression Recognition: Advances, Challenges, and Future Directions. Vol. 11, Vietnam Journal of Computer Science. World Scientific; 2024. p. 167–93.
5. Kolhe D, Mandavkar O, Metkar S, More S, Adgaonkar A, Professor A. EMOTION DETECTION.
6. Huang Y, Chen F, Lv S, Wang X. Facial expression recognition: A survey. Symmetry (Basel). 2019 Oct 1;11(10).
7. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) : 4-8 May 2015, Ljublijana, Slovenia. IEEE; 2015.
8. Mehendale N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl Sci. 2020 Mar 1;2(3).
9. Li J, Wang T, Wang SJ. Facial Micro-Expression Recognition Based on Deep Local-Holistic Network. Applied Sciences (Switzerland). 2022 May 1;12(9).
10. Mollahosseini A, Chan D, Mahoor MH. Going Deeper in Facial Expression Recognition using Deep Neural Networks. 2015 Nov 12; Available from: http://arxiv.org/abs/1511.04110
11. Rubin DC, Talarico JM. A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. Memory. 2009;17(8):802–8.
12. D'Mello SK, Kory J. A review and meta-analysis of multimodal affect detection systems. Vol. 47, ACM Computing Surveys. Association for Computing Machinery; 2015.
13. Proceedings of the 4th International Conference on Internet Multimedia Computing and Service. ACM Digital Library; 2013.
14. Saha P, Nath G, Bhowmik MK, Bhattacharjee D, De BK. NEI Facial Expressions for Identification. IERI Procedia. 2013;4:358–63.
15. Kaulard K, Cunningham DW, Bülthoff HH, Wallraven C. The MPI facial expression database - a validated database of emotional and conversational facial expressions. PLoS One. 2012 Mar 15;7(3).