IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Identifying Human Emotions From Facial Expressions With Deep Learning

Mr B. Harsha Vardhan [1], Smt D. Madhuri [2], Smt M. Prashanthi [3]

[1]PG Scholar, Department of CSE, JNTUACEA Anantapuramu, India.

[2] Assistant Professor (Contract), Department of CSE, JNTUACEA Ananthapuramu, India.

[3] Assistant Professor (Contract), Department of CSE, JNTUACEA Ananthapuramu, India.

Abstract: This paper presents a Facial Emotion Detection System with Expression Recognition, utilizing Convolutional Neural Networks (CNNs) implemented with TensorFlow and Keras. The system operates in two phases: detecting faces in images and then recognizing their expressions. The combination of deep learning and emotion recognition has significant potential across various fields. This work focuses on "Image-based emotion detection using deep learning," leveraging CNNs to identify and classify facial expressions, which are key to understanding human emotions. CNNs are used for their ability to extract detailed features from images. Additionally, the wavelet transform is integrated to enhance the accuracy of recognizing subtle emotional nuances. The study utilizes a dataset from Kaggle, containing seven distinct facial expressions, for experimental analysis. Results indicate that the CNN method, improved with wavelet transform, significantly boosts accuracy in facial expression recognition. This innovative approach promises applications in human-computer interaction and mental health assessment, providing a pathway for precise and nuanced emotional interpretation of facial images.

Keywords: Emotions, Deep Learning, CNN**.**

Introduction:

The recognition of human facial expressions and emotions is highly significant today as it can interpret human behavior, feelings, and intentions. Traditional methods are limited by low speed and accuracy, whereas systems using deep learning have demonstrated superior performance[1].

As of today, there are approximately 6,909 spoken languages in the world. To convey 7% of communication, it is estimated that over 90,000 languages are needed to understand the remaining 93% of non-verbal communication. The Facial Action Coding System (FACS), introduced by Ekman and Friesen in 1978, defines seven major facial expressions—fear, anger, surprise, disgust, happiness, sadness, and neutral—that humans convey without words. This system is considered the benchmark for Facial Expression Recognition [2]. Traditional face recognition methods, such as template matching and feature extraction, face limitations in accuracy and efficiency. Consequently, researchers have explored new approaches, and the advent of Convolutional Neural Networks (CNNs) has revolutionized face recognition.

Using Convolutional Neural Networks (CNNs) for facial emotion detection has several disadvantages, despite their strengths. CNNs require large, high-quality annotated datasets, which can be costly and time-consuming to collect, and their performance heavily depends on data quality. They are computationally intensive, requiring powerful GPUs[2] and extensive memory, leading to high costs and long training times. CNNs are prone to overfitting, especially with small or non-diverse datasets, necessitating complex regularization techniques. They lack transparency, making their decision-making process difficult to interpret. CNNs are sensitive to input variations such as lighting and occlusions, requiring extensive preprocessing. Additionally, there are ethical and privacy concerns, as facial emotion detection involves capturing and analyzing images of individuals, raising issues of consent and potential misuse.

Face detection and facial expression recognition fall within the fields of biometrics and biomimetics. Predictions suggest a significant surge in these areas in the coming century, as computers and machines will be able to communicate more effectively with humans. Numerous researchers have approached face detection and expression recognition from various perspectives. In this technique, the Rectified Linear Unit [4] 1(ReLU) will be used in the second stage. While it's not necessary to fully understand CNNs to benefit from this lecture, reviewing the operation of ReLU layers and the concept of linearity in CNNs can be helpful. Understanding the Pooling Layer and the overall structure of CNNs will also enhance your comprehension.

Most available face detection and face alignment methods tend to overlook the inherent correlation[6] between these two tasks. Although several works attempt to address them jointly, They still have limitations. For instance, Chen et al. use a random forest approach to jointly perform alignment and detection based on pixel value difference features, but the performance is limited by the handcrafted features employed. Similarly, Zhang et al. leverage a multi-task CNN to enhance the accuracy of multi-view face detection, but their detection accuracy is constrained by the initial detection windows generated by a weak face detector.

Facial Emotion Recognition (FER) [7] combines two significant fields: psychology and technology. In psychology, extensive research has been conducted on facial responses to emotional changes. In parallel, the technology leverages image processing concepts (computer vision) and machine learning techniques to automate the recognition process. FER's general architecture consists of three main phases: pre-processing, feature extraction, and classification or recognition. Each phase sequentially performs its specific tasks on a given FER database, establishing the ground truth needed for the system to achieve its goal.

Pooling layers, typically situated after convolutional layers, compress the output feature data from these layers. This compression enhances the output results and reduces the likelihood of overfitting in the neural network. Additionally, pooling operations enable further feature extraction from the image without compromising essential information acquisition. CNNs, widely used in applications such as image classification, object detection, segmentation, facial recognition, and medical image analysis, offer significant advantages. These include parameter sharing, where the same filter is applied across different image parts, reducing parameters and computational complexity, and local connectivity, where filters are applied to local input regions, allowing the network to learn spatial hierarchies. CNNs have revolutionized computer vision and are fundamental to many modern AI applications.

Detecting human emotions using facial expressions with Convolutional Neural Networks (CNNs) involves several steps, leveraging the powerful feature extraction and pattern recognition capabilities of CNNs. The key steps include data collection and preprocessing, model architecture design, training, evaluation, and deployment. Initially, a large dataset of facial images annotated with corresponding emotions (e.g., happiness, sadness, anger, surprise, fear, disgust, and neutral) is collected. The images are then normalized, resized, and augmented to increase the diversity of training samples and reduce overfitting.

After training, the model is evaluated on a test dataset to assess its accuracy, precision, recall, and F1 score. A confusion matrix is used to understand the model's performance across different emotion classes. Fine-tuning involves adjusting the model's hyperparameters, architecture, and training process based on evaluation results to improve performance. Once the model is fine-tuned, it can be deployed in real-world applications such as emotion recognition systems in video conferencing, security systems, or interactive user interfaces. Applications of emotion detection using CNNs span customer service, healthcare, marketing, and human-computer interaction, where understanding human emotions is crucial. By leveraging CNNs, emotion detection systems can achieve high accuracy and robustness, making them valuable tools in various fields.

Using Convolutional Neural Networks (CNNs) for facial emotion detection offers significant advantages, including automatic feature extraction where CNNs learn hierarchical features from raw images, capturing complex patterns without manual engineering. They provide spatial invariance, recognizing facial features regardless of position, and robustness to variations in scale, rotation, and distortions. Parameter sharing in CNNs reduces the number of parameters, enhancing efficiency and generalization. CNNs build hierarchical representations, recognizing complex patterns critical for emotion detection. They are scalable, adaptable, and integrate well with modern deep learning frameworks like TensorFlow, PyTorch, and Keras, facilitating ease of use and leveraging pre-trained models for transfer learning. CNNs achieve high accuracy and real-time processing capabilities, essential for interactive applications, and incorporate regularization techniques to prevent overfitting, ensuring robust performance on unseen data. These advantages make CNNs the preferred choice for highly accurate, efficient, and scalable facial emotion detection solutions.

Improving CNNs in facial emotion detection involves strategies such as enhancing data quality through augmentation and diversified datasets, optimizing model architecture with deeper networks and transfer learning, and improving training techniques with regularization and early stopping. Data preprocessing, including normalization and face alignment, helps ensure consistent input and accuracy. Advanced techniques like attention mechanisms and multi-task learning enhance feature extraction and understanding. Addressing ethical concerns involves obtaining informed consent and mitigating biases. Post-processing methods such as smoothing and ensemble techniques further refine predictions for improved robustness and accuracy.

Literature Survey:

In [1], the authors proposed a A CNN (Convolutional Neural Network) model is employed to detect human expressions and emotions from the Electrocardiogram (ECG) dataset, which can also be applied to categorize brain signals. The system achieves approximately 83% accuracy on testing. It utilizes a CNN-based emotion recognition system with feature map and max pooling layers, along with a fully connected output layer, to identify emotions in individuals.

In [2], the authors initial stage of proposed system involves extracting pertinent facial regions to facilitate both feature extraction and classification processes. This pre-processing step employs a dual approach using the Viola-Jones algorithm. Firstly, the photograph undergoes cropping using Haar Feature Selection to isolate the facial area and discard irrelevant surrounding data. Subsequently, the cropped image is converted into grayscale.

The authors in [3] explained by Utilizing the Viola-Jones algorithm [3], a straightforward MATLAB script was developed to crop the face image by detecting the coordinates of the top-left corner, as well as the height and width of the bounding rectangle around the face. The resulting cropped image is then saved in a designated folder for subsequent use in the facial expression recognition stage.

In [4], the author has explained Neural networks are organized in layers made up of interconnected nodes containing activation functions. Input layers present patterns to hidden layers, where processing occurs via weighted connections. Facial expression recognition involves three stages: image preprocessing (face and parts detection using the Viola-Jones algorithm), feature extraction, and classification using CNN. Keras, an open-source neural network library in Python, aids in preprocessing, modeling, evaluating, and optimizing. The system uses a sequential model with layers like convolution, pooling, flattening, and dense layers. Image preprocessing detects and normalizes face parts, reducing image scale for faster training. Convolution and pooling layers enhance accuracy, with emotions classified from a large dataset of images.

Results:

As the number of epochs increases, the accuracy of the data improves while the data loss decreases, as illustrated in Figure 10.1. In this study, I ran 250 epochs, resulting in the system accurately and efficiently recognizing emotions.

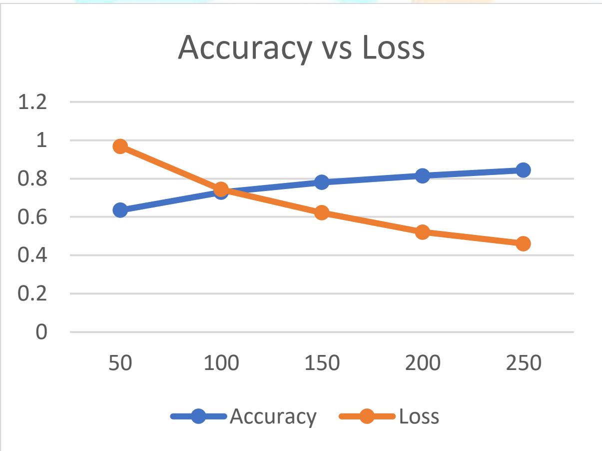

Fig.1: Accuracy Vs Loss

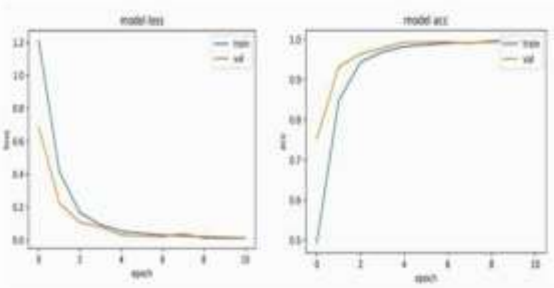

Fig.2: Model Loss and Model Accuracy.

The dataset used for our model, available in the Kaggle repository, classifies emotions such as joyful, disgusted, depressed, scared, angry, surprised, and neutral. Each image in the dataset varies in pixel size. As shown in Table 10.1, each class has an equal number of training samples, ensuring the dataset remains balanced and unbiased. This balance is consistently maintained across the training, testing, and validation sets.

Table.1: Details of training samples.

Category	**Number of Items**
Number of classes	7
Number of Trainings Images	28821
Number of Validations in Images	7066
Total number of Images	35887

Fig.3: Identified as Happy

Fig.4: Identified as Sad

Fig.5: Identified as Angry

Fig.6: Identified as Surprise

Conclusion:

The Real-Time Emotion Detection System is a comprehensive application designed to analyze and detect human emotions in real time using a webcam. It leverages computer vision techniques and a pre-trained deep learning model to identify and classify emotions from facial expressions. The system's testing strategies ensure robustness, accuracy, and user-friendliness. During the unit testing phase, individual components like frame capture, model loading, and feature extraction were validated. White box testing examined the internal code to uncover hidden issues, ensuring functions like data preprocessing, emotion prediction, and label mapping work accurately. Black box testing evaluated the system's behavior based on input and output interactions, testing its response to various user inputs without knowledge of its internal structure. Integration testing ensured seamless interaction between components, such as video capture, face detection, feature extraction, and emotion prediction. Functional testing confirmed the system meets its defined requirements and operates as anticipated from the user's standpoint. Acceptance testing evaluated its functionality, performance, and user satisfaction, confirming alignment with business needs and readiness for deployment. The system accurately detects emotions, is robust in various scenarios, and features a user-friendly interface. Future enhancements may focus on scalability, performance optimization, and feature expansion to improve the system's capabilities and user experience. Overall, the Real-Time Emotion Detection System is robust, accurate, and user-friendly, successfully meeting its objectives with a solid foundation for future enhancements.

Future Scope:

The Real-Time Emotion Detection System holds significant potential for further development and expansion. Advanced Emotion Recognition can be achieved by refining the deep learning model architecture, collecting diverse training data, and exploring state-of-the-art techniques. Real-time Feedback and Adaptation features could offer personalized recommendations based on detected emotions. Multi-modal Emotion Detection, incorporating facial expressions, voice tone analysis, and body language recognition, would enhance accuracy. Implementing Personalized User Profiles would tailor responses to individual characteristics. Enhancing Privacy-preserving Features is crucial for data protection. Extending Cross-platform Compatibility would broaden accessibility. Exploring Real-world Applications beyond emotion detection, such as mental health monitoring or market research, would leverage the system's capabilities. Continuous Improvement mechanisms and Collaborative Research can further advance the field. Exploring Commercialization and Deployment opportunities across sectors requires robust business models and partnerships. Overall, the future scope for the Real-Time Emotion Detection System is promising, contingent upon continued research, development, and collaboration efforts.

Reference:

[1] P.A.Ramesh, Sivakumar Depuru, et.al., “Human emotion recognition System Using Deep learning Technique”, Journal of Pharmaceutical Negative Results, Volume 13, Issue 4, 2022.

[2] Geerish Suddul, Phavish Babajee, et.al, “Identifying Human emotions from Facial Expressions with Deep learning”,Zooming Innovation in Consumer Technologies Conference , 2020.

[3] Roger Achkar, Michael Owayjan, “Face Detection with Expression Recognition using Artificial Neural Networks”, Middle East Conference on Biomedical Engineering, 2016.

[4] T. Chandra Sekhar Rao, D.Sri Hari, “Face Recognition Using Computer Vision and CNN Algorithm”, International Journal of Intelligent Systems and applications in Engineering, 2023.

[5] Feng Yanyan, Zhou Yue, “ Facial Expression Recognition Based on Convolutional Neural Network”, 2019

[6] Zhang Z, Zhang K, Li Z, et al. “Joint Face Detection and Alignment using Multi-task Cascaded

[7] Mahoor M H, Hasani B , “ Spatio-Temporal Facial Expression Recognition Using Convolutional Neural Networks and Conditional Random Fields” IEEE International Conference on Automatic Face & Gesture Recognition, IEEE, 2017, Pages: 790-795.

[8] Baoye Song, Pengu Lu, “Human face recognition based on Convolutional Neural Network and Augmented dataset”, System Science and Control Engineering, Volume:9, Pages:29-37