Face Emotion Recognition System Using Deep Learning

ASST. PROF M.KAVITHA[1], V. HARSHAN[2], M. PRAVEEN KUMAR[3], A. RAHUL[4]

[1234] *Dept. of Computer Science and Engineering, Government College of Engineering Srirangam, Tamilnadu, India*

---***---

Abstract - *This paper presents an enhanced system for real-time facial emotion detection, aiming to improve efficiency and accuracy through deep learning. The proposed approach utilizes VGG-19 transfer learning, a pre-trained convolutional neural network (CNN) architecture known for its depth and strong performance in image classification. VGG-19's pre-trained weights contribute to improved efficiency compared to simpler CNNs, allowing for effective feature extraction and classification of emotional expressions in real-time. This approach has the potential to benefit various applications in human-computer interaction and psychology by enabling accurate and timely emotion recognition*

Key Words: Facial Emotion Detection, Deep Learning, VGG-19 Transfer Learning, Real-Time Emotion Recognition, Efficiency, Human-Computer Interaction, Psychology

1.INTRODUCTION

Human communication encompasses speech, gestures, and emotions, vital for interpersonal interactions. AI systems capable of understanding human emotions are crucial, especially in healthcare, and e-learning where emotional understanding is paramount. Traditional emotion detection methods often fall short in real-time scenarios, necessitating models that can continuously interpret facial expressions for dynamic emotional assessment.

This paper proposes a real-time facial emotion recognition model leveraging AI and computer vision advancements. The model aims to enhance human-computer interactions across diverse applications by dynamically detecting and responding to emotions. Automatic Facial Expression Recognition (FER) has gained traction, driven by its potential in human-computer interaction and healthcare. While Ekman's discrete categorization model is widely used, its limitation in handling spontaneous expressions prompts the need for more comprehensive approaches.

Our focus is on categorical facial expression classification using the VGG-19 model, known for its depth and performance in image tasks. By employing pre-trained weights, our system achieves efficiency and accuracy for real-time emotion recognition. This work explores VGG-19 transfer learning's potential for facial emotion recognition while remaining adaptable to other models with suitable data.

1.1 RELATED WORK

Sharmeen M. Saleem Abdullah and Adnan Mohsin Abdulazeez [13] addressed the latest FER analysis. Numerous CNN architectures have been identified that have recently been proposed. They have provided various databases of random photographs obtained from the actual world and other laboratories to detect human emotions.

Hussein, E. S., Qidwai, U. and Al-Meer, M. [4] recommended a CNN model to understand face emotions with three continuum emotions. This model uses residual blocks and depth-separable convolutions inspired by Xception to minimize the sum of parameters to 33k. They use a convolutional neural FER network for emotional stability identification. CNN uses convolution operations to learn extract features from the input images, which reduces the need to extract features from images manually. The proposed model offers 81 percent total precision for invisible results. It senses negative and positive emotions, respectively, with a precision of 87% and 85%. However, the accuracy of neutral emotion detection is just 51%.

Jiang, P., Liu, G., Wang, Q., and Wu, J [5] introduced a new loss feature called the advanced softmax loss to eradicate imbalanced training expressions. The proposed losses guarantee that any class would have a level playing field and potential using fixed (unlearnable) weight parameters of the same size and equally allocated in angular space. The research shows that proposed (FER) methods are better than specific state-of-the-art FER methods. The proposed loss can be used as an isolated signal or used simultaneously with other loss functions. To sum up, detailed studies on FER2013 and the real-world practical face (RAF) databases have shown that ASL is considerably more precise and effective than many state-of-the-art approaches..

2 METHODOLOGIES

This approach utilizes the VGG19 architecture for Facial Emotion Recognition by preprocessing the dataset, training the model, and validating it for real-time deployment. It includes implementing a user interface for interaction and feedback loops for continual improvement.

CNNs, or Convolutional Neural Networks, are crucial in deep learning and particularly effective in computer vision tasks. They automatically learn relevant features from raw input data, making them ideal for image and video recognition. Structured to mimic human visual processing, CNNs consist of layers like input, convolutional, activation, pooling, fully connected, and output layers, each extracting increasingly abstract representations from input data..

2.2. Preprocessing Images:

The pixel values of grayscale images are converted from string format to numpy arrays and reshapes them to a consistent size of 48x48 pixels. These processed image arrays are then stacked together along the 0th axis to create a 4D numpy array representing the entire dataset, where each image is now in a standardized format suitable for further processing. Additionally, the grayscale images are converted to RGB format using OpenCV's `cv2.cvtColor` function to meet the input requirements of the VGG19 model. Following this conversion, the pixel values of the RGB images are normalized to a range of [0, 1] by dividing each pixel value by 255.0. This normalization step ensures uniformity and provides optimal training conditions for the deep learning model, enhancing its ability to learn meaningful patterns and features from the input images.

2.3. Cross-Validation:

To assess the performance and generalization capabilities of the CNN models, cross-validation is employed. The dataset is divided into multiple folds or subsets. The training and evaluation procedure is reiterated numerous times, with every fold being utilized as the validation set once. This approach ensures a comprehensive evaluation of the CNN models' performance across different subsets of the data. Cross-validation helps in validating the model's performance robustness and ensures that it can generalize well to unseen data, reducing the risk of overfitting and providing more reliable performance metrics.

2.4. Load VGG-19 Model:

The pre-trained VGG-19 model is loaded using TensorFlow's Keras API, initialized with weights pre-trained on the ImageNet dataset. Figure 1 illustrates the system architecture, showcasing the loading process without the fully connected layers typically used for ImageNet classification. These layers will be replaced with custom ones for the emotion recognition task. After loading, you can use `model.summary()` to examine the model's architecture, layer types, output shapes, and trainable parameters. Understanding these details is crucial for customizing the model and preparing it for training on your specific dataset.

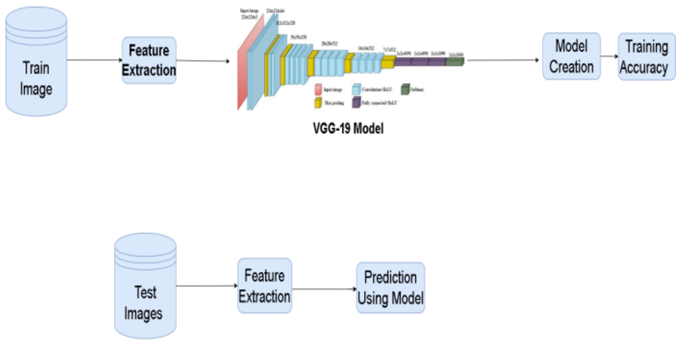

Fig. 1. System Architecture

2.5. Feature Extraction:

The module accesses the intermediate features learned by the VGG19 model, specifically targeting the second-to-last layer before the classification layers. These intermediate features capture rich representations of facial expressions and features crucial for emotion recognition. After obtaining the intermediate features, global average pooling is applied to summarize and condense the feature maps into a fixed-length feature vector for each image. A custom output layer, typically a Dense layer with softmax activation, is added to map the extracted features to different emotion classes, facilitating emotion recognition based on the VGG19-derived features. This approach leverages the power of transfer learning, using pre-trained deep learning models to extract meaningful features and then customizing the output layers for specific tasks like emotion recognition.

2.6. Emotion Recognition:

The process of emotion prediction using the trained model involves several steps. Initially, the preprocessed sample image data is fed into the trained model using `model.predict(input_data)`, where the model calculates the probabilities associated with each emotion class based on the image features learned during training. Subsequently, the predicted probabilities array is examined to determine the emotion with the highest probability, which serves as the predicted emotion label. This label is then mapped to a human-readable emotion category using a dictionary, providing the final prediction of the emotional state depicted in the sample image. Finally, the predicted emotion label is extracted and displayed alongside the original image, offering a clear representation of the model's prediction regarding the emotional state conveyed in the image. This comprehensive process allows for accurate and interpretable emotion recognition outcomes based on deep learning techniques and model predictions.

3. EXPERIMENTAL EVALUATION

3.1 Environment Specifications:

Table 1 outlines the experimental configuration employed in the study. The research work takes place on an AMD Ryzen 5 3500U CPU with 12GB of RAM and an integrated graphics card. The models are constructed using Python and executed using deep learning frameworks such as Keras and TensorFlow.

Process name	S.N	Action
Input	1.	Collected images of 10 classes of facial expressions, including 7 classes representing different emotions
Environment Configuration	2.	Anaconda, Jupyter Notebook
	3.	Import all necessary libraries and packages
Directories Configuration	4.	Load the images
	5.	Load the directories for training, testing and create validation on 20% training data
Training and Testing	6	Developed transfer learning models trained on the ImageNet dataset and CNN models
	7	Fine-tuned by adding the fully connected (FC) layer with SoftMax activation for face emotion detection using VGG19.
Model Compilation	8	The model complies with Adam optimizer and a learning rate 0.001
	9	Set 70 epochs for model training
	10	As model checkpoint, use the validation loss to monitor
	11	Save model
Performance Report	12	Generate classification report
	13	Generate model accuracy and loss reports
Prediction	14	Load the best model
	15	Predict the type of Emotions and generate solution class.

TABLE 1 Experimental Setup

3.2 Dataset:

A carefully curated dataset of human facial expression images is collected. The dataset comprises labeled images depicting a range of facial expressions, including happiness, sadness, anger, and others as in fig.2. These images are utilized as input data for training and evaluating convolutional neural network (CNN) models. The Fer2013 dataset stands as a widely employed resource for facial expression recognition tasks.. The dataset contains 35,887 grayscale facial images containing 7 different emotions (anger, disgust, fear, happiness, neutral, sad, and surprise).

Fig. 2. FER-2013 Samples

3.3 Result Analysis:

Upon thorough analysis of the results, it is apparent that the VGG19 model achieved the highest accuracy of 88%, surpassing both the custom CNN model with 76% accuracy as shown in Fig6. All models were trained using data augmentation techniques to enhance their performance. Despite the computational resources required by VGG19 compared to CNN, its deeper architecture and feature extraction capabilities contribute significantly to its performance. Therefore, based on these findings, the proposed system opts for VGG19 as the preferred model, prioritizing accuracy over computational efficiency in this scenario.

The accuracy and loss graphs for VGG19 model along with confusion matrices aiding in identifying correct predictions for each emotion class, are provided herein. Refer to Fig.3 for the accuracy and loss graphs of VGG19. Refer to Fig.4 for the confusion matrix of VGG19 Additionally, the classification report corresponding to each model is included within the Fig.5, facilitating a detailed assessment of the model's performance across different emotion classes. These visual representations serve as valuable tools for comprehensively evaluating and comparing the efficacy of each model in the context of facial emotion recognition.

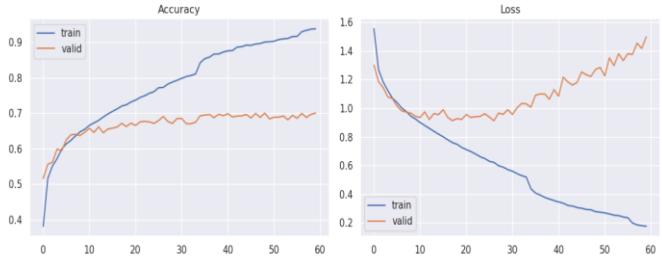

Fig.3. Accuracy and Loss Graph.

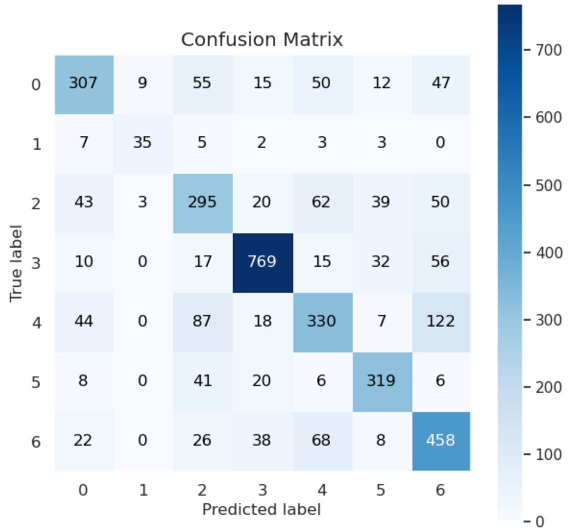

Fig.4. Confusion matrix

```
113/113 [==============================] - 1s 10ms/step
              precision    recall  f1-score   support

           0       0.56      0.70      0.62       495
           1       0.79      0.67      0.73        55
           2       0.62      0.53      0.57       512
           3       0.89      0.85      0.87       899
           4       0.59      0.56      0.57       608
           5       0.80      0.80      0.80       400
           6       0.64      0.66      0.65       620

    accuracy                           0.69      3589
   macro avg       0.70      0.68      0.69      3589
weighted avg       0.70      0.69      0.69      3589
```

Fig.5. Classification Report

3.4 Deployment:

The facial emotion recognition project is deployed via a web-based platform built on the Flask framework and integrated with React and OpenCV. Users access the application through a web browser, where they can upload images for emotion analysis. The Flask backend handles image processing tasks using the trained VGG19 model, while the frontend displays results such as predicted emotions and confidence scores. Utilizing Flask, React, and OpenCV enables seamless interaction between users and the application, providing a user-friendly interface for emotion recognition tasks

4. CONCLUSIONS

In this system implementing face emotion recognition using deep learning with the VGG19 architecture presents a promising approach for accurately detecting and classifying emotions from facial images. By following a systematic approach that involves data collection, preprocessing, model architecture selection, transfer learning, training, evaluation, and deployment, it's possible to develop a robust and effective emotion recognition system. Transfer learning with pre-trained VGG19 models enables leveraging knowledge learned from large-scale image classification tasks, which can significantly enhance the performance of the emotion recognition model, especially when training data is limited. Throughout the development process, careful attention should be paid to data preprocessing, augmentation, hyperparameter tuning, and model evaluation to ensure the model generalizes well to unseen data and accurately predicts emotions across various facial expressions and environmental conditions. Ultimately, the successful deployment of a face emotion recognition system can open up possibilities for applications in diverse fields, including human-computer interaction, healthcare, entertainment, and security, contributing to advancements in technology and enhancing user experiences.

REFERENCES

[1] Sharmeen M. Saleem Abdullah, Adnan Mohsin Abdulazeez. Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review in JOURNAL OF SOFT COMPUTING AND DATA MINING VOL. 2 NO. 1 (2021) 53-65.

[2] Hussein, E. S., Qidwai, U. and Al-Meer, M. Emotional Stability Detection Using Convolutional Neural Networks, 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 136-140.

[3] Jiang, P., Liu, G.,Wang, Q. and Wu, J. (2020). Accurate and Reliable Facial Expression Recognition Using Advanced Softmax Loss with Fixed Weights. IEEE Signal Processing Letters, 27, 725-729.