Theory and Practice of Science and Technology
2025, VOL. 6, NO. 2, 51-54
DOI: 10.47297/taposatWSP2633-456911.20250602

# Artificial Intelligence for Facial Emotion Recognition: A Review

Waikan Chong

Intelligent Science and Engineering, Jinan University, ShanDong, 519000, China

**ABSTRACT**
With the proliferation of online platforms, an increasing number of people are immersing themselves in the virtual world, leading to reduced real-world social interactions, more introverted personalities, and even psychological conditions such as depression. Therefore, the implementation of facial emotion recognition technology is significant for early detection of negative emotions and providing timely humanistic care. Emotion recognition not only helps individuals better understand their own emotions but also promotes positive self-perception and mental health. By analyzing facial expressions, one can effectively determine a person's emotional state. Research on facial expression recognition began in the 1970s, and with the development of image processing and pattern recognition technologies, it has gradually transitioned from traditional methods to automated recognition based on deep learning. This paper reviews recent research achievements in the field of emotion recognition, focusing on how to improve the accuracy and efficiency of emotion recognition to promote its widespread application in real life. By analyzing three representative studies, this paper summarizes the current shortcomings of emotion recognition technology, including insufficient robustness, real-time issues, limitations of datasets, inadequate model complexity and optimization, and insufficient multi-modal fusion. In response to these shortcomings, this paper proposes corresponding directions for improvement, such as enhancing environmental adaptability, optimizing model structure, expanding dataset scale, improving loss functions, and adopting multi-modal fusion. These directions aim to improve the adaptability, real-time performance, and accuracy of emotion recognition technology in complex environments, while enhancing its generalization ability and practicality. These research progress and suggestions for improvement provide valuable references for the future development of emotion recognition technology.

**KEYWORDS**
Artificial Intelligence; Deep Learning; Convolutional Neural Networks; Emotion Recognition

## 1 Introduction

With the prevalence of online platforms, more and more people are reluctant to engage in real-life interactions, preferring to immerse themselves in the virtual world. This trend can lead to increasingly introverted personalities and a reluctance to express emotions, which may in turn trigger psychological issues such as depression. Therefore, the implementation of facial emotion recognition technology is crucial for the early detection of negative emotions and the provision of timely humanistic care.

Emotion recognition refers to the ability to accurately perceive and understand one's own and others' emotions. When individuals can identify their own emotions, they can better understand their inner needs. It also helps individuals build a more accurate self-image. If a person can consistently recognize positive emotions such as confidence and optimism in different situations, they are likely to form a positive self-perception, which in turn motivates them to continuously strive for better development in life.

Emotion recognition primarily relies on the analysis of facial expressions. Emotional fluctuations can alter the state of facial muscles and features, leading to changes in facial expressions. Therefore, by analyzing facial expressions, one can determine a person's emotional state. Research on facial expression recognition began in the 1970s, when Ekman and other professors concluded that six types of facial expressions correspond to six types of emotions across different countries and cultures, laying the foundation for the feasibility of emotion recognition through facial expressions. However, for more than a decade afterward, progress in expression recognition was slow due to limitations in face detection and alignment algorithms, as well as computational power. It was not until the 1990s that advancements in image processing and pattern recognition technologies made high-automation expression recognition possible, leading to the development of both traditional and deep learning-based expression recognition methods.

This paper reviews recent research achievements in the field of emotion recognition, with a focus on how to better improve the accuracy and efficiency of emotion recognition to facilitate its widespread application in real life.

## 2 Literature Review

In the study "Design and Implementation of a Real-Time Monitoring System for Adolescent Mood Assessment Based on Facial Recognition," the researchers employed a deep learning-based Convolutional Neural Network (CNN) as the core algorithm. The system was developed using Python, the OpenCV computer vision library, and the OpenVino deep learning engine. The front end was implemented using a We-Chat mini-program, while the back end established a student facial expression database for storing and analyzing emotional data. Embedded facial recognition programs were installed in key locations, such as the entrance of the psychological counseling room, to capture video frames and perform frame-by-frame analysis. The system was tested with a dataset of 40 facial images over 80 trials. Using the OpenCV computer vision library to detect facial positions and pass facial information to the expression recognition module, the system identified five basic emotions (neutral, angry, sad, happy, and surprised). By calculating the probabilities of the most and second-most likely emotions, the system achieved an accuracy rate of 96.25%. Issues encountered during the experiment included errors or ambiguities caused by factors such as hair, glasses, and caps obstructing facial features, as well as excessive background exposure. The study introduced several innovative aspects in the design of emotion recognition: the use of the OpenVino acceleration engine to enhance deep learning performance and optimize the performance bottleneck of the CNN network on edge devices; the system also utilized facial feature codes for individual recognition and storage, enhancing human-computer interaction.

Despite the high accuracy rate of 96.25% demonstrated in the tests, the system's robustness in complex environments remains to be improved, as low lighting, low resolution, and obstructions such as hair, glasses, and caps can affect the accuracy of recognition. Additionally, although the system is designed for real-time monitoring, delays in data processing and analysis may impact its real-time performance, especially when handling large-scale data. The study only mentions that the system is currently implemented using Python, OpenCV, and OpenVino, without addressing whether it supports other technical frameworks or platforms for expansion, which may limit further development and optimization of the system.

In the study "Research on Human Abnormal Emotion Recognition Algorithm Based on Video Images," the authors utilized not only the Convolutional Neural Network (CNN) algorithm but also the MLP hierarchical architecture, combining CNN with MLP to extract facial expression and body movement features. Through convolutional layers, activation functions (RELU), and pooling layers, global features were extracted. The DNET was employed to connect feature maps from different layers, enhancing feature recognition effects and reducing the vanishing gradient phenomenon. The experiment established a training expression sample library and a test face library containing 100 samples. After a series of image processing, emotions were classified and recognized. The purity of the samples was measured using information entropy, with abnormal emotions set as the target emotion and others as interfering emotions. The classification performance was assessed using the False Recognition Probability (FRP).

During the experiment, data for facial recognition tests were collected using a single camera with a fixed lens, ensuring differences in lighting and lens positions across various scenes. A pre-trained DNET network was used to extract features, with 100 iterations. As the number of iterations increased, the error rate remained below 5%. The error rate was relatively higher in the control groups. The study also attempted to add Gaussian and Poisson noise to the training set, and the experimental group demonstrated the lowest error rate, showing good noise resistance. After adding noise, the image recognition results of the experimental group were clearer and more detailed.

The innovation of this experiment lies in the combination of deep learning algorithms and feature extraction, which improves the accuracy of feature recognition. The experiment also verified the recognition effect of the algorithm after adding noise, demonstrating good noise resistance.

However, the experiment had the following shortcomings: 1. Pixel-level recognition accuracy needs improvement. 2. The validation and optimization of the loss function are insufficient, which may affect the model's convergence speed and recognition effect. 3. Despite the low error rate, there is still room for further optimization, especially in terms of recognition performance in complex environments.

In the study "Research on Facial Emotion Recognition Based on Features," the researchers introduced the classic LeNet-5 network to extract deep digital features of facial expressions, implementing emotion recognition through convolutional, pooling, and fully connected layers. The Viola-Jones algorithm was used to predict facial regions, accurately locating facial areas. This algorithm, based on Haar-like features and the Adaboost algorithm, quickly calculates feature values through integral images for efficient detection. By using Haar-like features to locate the centers of the eyes and combining anthropometric theory to position and segment emotional organs such as eyebrows, eyes, and mouth, a feature face was constructed to reduce information redundancy in non-emotional facial areas. Additionally, Gaussian filtering was applied to denoise images, and Gamma correction was used to adjust the uniformity of lighting, enhancing image quality. The authors collected a data-set of 212 images featuring seven types of facial expressions (anger, disgust, fear, happiness, sadness, surprise, and neutrality) from ten Japanese females as the JAFFE study dataset, with each image labeled with an emotion tag.

The experiment utilized Matlab 2018b and an Intel i7-4702MQ CPU as the development environment, randomly selecting 80% of the images for the training set and 20% for the test set. The initial learning rate was set at 0.001, with a maximum of 250 iterations. The results showed that the feature face model achieved an average recognition accuracy of 90.12%, outperforming the geometric feature model (53.75%) and the full-face feature model (87.46%). The feature face model excelled in recognizing multiple emotions, with only disgust being misidentified as sadness, while the recognition accuracy for other emotions was 100%.

The innovation of this experiment lies in the introduction of the LeNet-5 Convolutional Neural Network, which automatically extracts deep digital features, thereby enhancing recognition performance. The study also conducted a detailed analysis of facial expression features, combining anthropometric theory to accurately locate emotional organs and construct a feature face, reducing interference from non-emotional information.

However, the experiment had the following shortcomings: Although the feature face model improved recognition accuracy, its high model complexity may lead to low computational efficiency, making it difficult to run efficiently on edge devices. The experiment was only conducted on the JAFFE dataset, which has a limited sample size and may affect the model's generalization ability.

## 3 Conclusion

### 3.1 Shortcoming

After reviewing the above research achievements in emotion recognition, the following overall shortcomings have been identified:

#### 3.1.1 Insufficient Robustness

Environmental Factors: Under conditions of low lighting, low resolution, and complex backgrounds, the accuracy of emotion recognition significantly decreases. For example, obstructions such as hair, glasses, and caps can affect recognition results.

Noise Interference: Although some studies have made progress in noise handling, the overall noise resistance still needs improvement, especially in high-noise environments.

Adaptability to Complex Scenes: Existing methods perform well in single scenes (such as laboratory environments) but lack adaptability in real-world complex scenarios, struggling to cope with varying lighting, poses, and expressions.

#### 3.1.2 Real-Time Issues

Data Processing Delays: When handling large-scale data, delays in data processing and analysis can impact the system's real-time performance, especially in applications requiring rapid responses.

Computational Efficiency: Some models, with their high complexity, consume significant computational resources, making it difficult to meet real-time requirements.

#### 3.1.3 Dataset LimitationsLimited Sample Size

Most studies validate their methods on specific datasets (such as JAFFE), which have limited sample sizes and may affect the model's generalization ability.

Lack of Diversity: Existing datasets lack diversity in terms of race, age, and gender, failing to cover a broader range of populations and emotion types.

#### 3.1.4 Model Complexity and Optimization

High Model Complexity: Some models, such as the feature face model, achieve high recognition accuracy but have high complexity, leading to low computational efficiency and difficulty in running efficiently on edge devices.

Insufficient Loss Function Validation: The validation and optimization of loss functions are inadequate, which may affect the model's convergence speed and recognition effect.

#### 3.1.5 Insufficient Multi-modal Fusion

Single Modality: Current research mainly focuses on facial expression recognition, lacking the fusion of multi-modal information such as voice and posture, which limits the comprehensiveness and accuracy of emotion recognition.

### 3.2 Suggestions for Improvement

Based on the above shortcomings, the following suggestions are proposed to guide future research in emotion recognition:

### 3.2.1 Enhancing Robustness

Environmental Adaptation Optimization: Develop more advanced preprocessing techniques, such as adaptive lighting correction and background segmentation, to improve the model's adaptability in complex environments.

Obstruction Handling: Introduce attention mechanisms or Generative Adversarial Networks (GANs) to automatically repair obstructed facial areas, enhancing recognition effects.

### 3.2.2 Improving Real-Time Performance

Model Optimization: Adopt lightweight neural network architectures (such as MobileNet and ShuffleNet) to reduce model complexity and improve computational efficiency.

Hardware Acceleration: Utilize dedicated hardware (such as GPUs and FPGAs) to accelerate model inference and enhance real-time performance.

Edge Computing: Migrate some computational tasks to edge devices to reduce data transmission delays and improve system response speed.

### 3.2.3 Expanding Datasets

Data Augmentation: Increase the diversity and scale of datasets through techniques such as rotation, scaling, and color adjustment.

Multi-Source Datasets: Integrate multiple datasets covering different races, ages, genders, and emotion types to enhance model generalization ability.

Dynamic Datasets: Develop dynamic datasets for real-time data collection and updating to adapt to different scenarios and populations.

### 3.2.4 Model Optimization and Validation

Loss Function Improvement: Design more rational loss functions, such as those combining class imbalance, to improve model convergence speed and recognition effect.

Model Evaluation: Use more comprehensive evaluation indicators (such as precision, recall, and F1 score) to assess model performance from multiple angles.

Model Compression: Compress model size through techniques such as pruning and quantization to improve model efficiency on edge devices.

### 3.2.5 Multi-modal Fusion

Multi-modal Feature Extraction: Develop multi-modal feature extraction methods, such as combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to extract deep features from facial expressions, voice, and posture.

Multi-modal Fusion Models: Construct multi-modal fusion models, such as multi-modal attention networks, to dynamically adjust the weights of different modalities through attention mechanisms, enhancing the accuracy of emotion recognition.

Multi-modal Data Synchronization: Address the synchronization issue of multi-modal data to ensure temporal consistency among different modalities, improving fusion effects.

## References

[1] Wang Zijian, Meng Xiaoyan. Design of teenagers' mood assessment system based on face recognition to realize real-time monitoring. 1009-3044(2023)27-0026-03.

[2] Research on Human Abnormal Emotion Recognition Algorithm Based on Video image.

[3] Lu Jinye, Zheng Fangyuan, Wang Juntao, Ma Yuhong. Research on facial emotion recognition based on feature faces.