Real-Time Emotion Detection Using Residual Masking Network

Ms.V.Bhavya[1], A.Hansika Reddy[2], B. Abhilash Reddy[3], D. Pranav Aditya[4], M. Sai Sri Harsha[5]

[1,2,3,4,5] *Department of Artificial Intelligence & Data Science, Vidya Jyothi Institute of Technology, Hyderabad, India*

Abstract- Human emotions are important for how we communicate and make decisions. The ability of machines to understand emotions in real time is becoming very important in areas like healthcare, education, and how people interact with computers. This paper introduces a real- time emotion detection system that uses a Residual Masking Network (RMN). It combines deep learning with attention mechanisms to make the system more accurate and efficient. The system takes live video input, detects facial features, and categorizes emotions into types such as happiness, sadness, anger, fear, surprise, and neutrality. Unlike older systems, this model uses residual learning to overcome some training issues and masking techniques to focus on important parts of the face. The system is designed for real-time performance with low delay and ensures privacy through secure data handling.

The experimental results show that this approach has higher accuracy, better performance under different conditions, and efficient real-time processing. This system offers a scalable and reliable solution for emotion recognition tasks.

I. INTRODUCTION

With the growth of artificial intelligence, machines are becoming better at understanding human behavior.

Emotion detection is a key part of this ability, as emotions affect decisions, communication, and interactions. Facial expressions are one of the most reliable ways to tell how someone is feeling. Research suggests that more than half of human communication is non-verbal, making facial expression recognition (FER) important in AI systems. Earlier methods relied on manual observation or simple algorithms, which were not very effective or accurate. Modern deep learning methods, especially convolutional neural networks (CNNs), have greatly improved emotion recognition. However, there are still challenges: the need for real-time processing, high computational costs, sensitivity to lighting and angles, and privacy issues. To address these, this paper proposes a system based on a Residual Masking Network, which improves feature extraction and ensures real-time performance.

I.1 SIGNIFICANCE OF REAL-TIME EMOTION DETECTION

Real-time emotion detection allows systems to react immediately to user emotions.

This is important in applications such as: Healthcare: monitoring patient emotions Education: understanding student engagement Customer Service: improving user experience Security: detecting suspicious behavior By providing instant feedback, the system improves interaction quality and decision-making.

I.2 THEORETICAL BACKGROUND

I.2.1 Facial Expression Recognition (FER)

FER involves identifying facial features such as eyes, mouth, and eyebrows to classify emotions.

I.2.2 Deep Learning Models

CNNs automatically extract features from images, so there's no need for manual feature engineering.

I.2.3 Residual Learning

Residual connections help deep networks train efficiently by preventing loss of gradients.

I.2.4 Attention Mechanism

Attention mechanisms help the model focus on important facial areas.

II. LITERATURE SURVEY

Emotion detection research has evolved from traditional machine learning methods to advanced deep learning techniques.

Earlier methods like SVM and k-NN required manual feature extraction and had limited accuracy. With the introduction of CNNs, feature extraction became automated, greatly improving performance.

Residual Networks (ResNet) introduced skip connections, enabling deeper models. Further developments led to Residual Masking Networks (RMN), which combine residual learning with attention mechanisms.

Despite improvements, challenges remain:

- Computational complexity
- Bias in datasets
- Environmental variations
- Privacy concerns

These issues highlight the need for efficient and scalable systems.

III. METHODOLOGY

The proposed system for real-time emotion detection is designed as a structured and modular pipeline that processes live video input, extracts meaningful facial features, and classifies human emotions using a deep learning model.

The methodology focuses on improving accuracy, reducing latency, and ensuring scalability while maintaining data privacy. Unlike traditional approaches that rely on manual intervention or simple models, this system integrates computer vision techniques with a Residual Masking Network (RMN) to achieve efficient and reliable performance.

III.1 EXISTING SYSTEM

In traditional emotion detection systems, identifying human emotions is usually done using old machine learning methods or manual observation. These systems often rely on manually created features, where specific facial traits like the distance between eyes or the curve of the lips are calculated by hand. While these methods worked well in controlled settings, they are not flexible and do not perform well in real-world situations.

Another big problem with these systems is that they can't work in real time. Many models take a lot of computing power and time to process, which leads to delays. This makes them unsuitable for applications like live monitoring, interactive systems, or security systems, where quick responses are needed.

Moreover, traditional systems struggle with changes in lighting, the angle of the face, and background noise. These factors greatly affect how accurate and reliable the system is. Also, most early systems were not built to handle large amounts of data or multiple users at once.

III.2 PROPOSED SYSTEM

To solve the problems with existing methods, the proposed system introduces a real-time emotion detection framework based on the Residual Masking Network. The system is designed to automate the whole process—from capturing facial data to predicting emotions—so it is both fast and accurate.

The system uses a centralized processing approach, where all input data is handled through a single flow. This allows different parts of the system, like face detection, feature extraction, and emotion classification, to work together smoothly. The system can process live video and provide instant results without any noticeable delay.

A major advantage of the proposed system is its ability to focus on important parts of the face while ignoring extra details. This is done through the masking mechanism in the RMN, which improves feature extraction and classification accuracy. The system also uses optimized processing techniques and GPU acceleration to ensure real-time performance.

III.2.1 SYSTEM WORKFLOW

The system starts by capturing real-time video input using a webcam or other camera. Each frame from the video is processed separately to find human faces. Once a face is found, it is separated from the background and moved to the preprocessing stage.

During preprocessing, the face is resized, normalized, and prepared for feature extraction. This step ensures that the input data is consistent, which is important for accurate model performance. The system then identifies facial landmarks like the eyes, nose, and mouth using advanced computer vision techniques.

After extracting features, the processed patterns.

In emotion detection, residual learning helps the model catch small facial changes, like slight eye movements or lip curves, which are important for telling similar emotions apart.

The masking mechanism acts as an attention module that highlights the most important areas of the face. Emotions are usually shown through particular areas like the eyes, mouth, and eyebrows. The masking module creates attention maps that focus on these regions while ignoring irrelevant details like background noise or hair.

By combining residual learning with masking, the RMN allows the model to extract meaningful features efficiently. This results in higher accuracy and faster processing compared to traditional CNN models data is sent to the Residual Masking Network. The RMN analyzes the facial features and finds patterns linked to different emotions. Based on these features, the model predicts the most likely emotion.

Finally, the detected emotion is shown on the screen in real time, along with visual cues like bounding boxes and labels. The system keeps processing new frames continuously, ensuring a smooth and uninterrupted detection process.

III.2.2 RESIDUAL MASKING NETWORK (RMN)

The Residual Masking Network is the key part of the proposed system and plays a major role in improving emotion detection. It combines two advanced deep learning ideas: residual learning and attention-based masking.

Residual learning helps overcome the issue of vanishing gradients in deep neural networks. As networks grow deeper, it becomes hard to pass gradients back through each layer, which can slow down training. Residual connections fix this by allowing the input of a layer to skip over some layers and be directly added to the output. This makes training more stable and helps the network learn complex

III.2.3 DATA PROCESSING AND MANAGEMENT

Data processing is a vital part of the system, as the quality of the input data greatly affects the model's performance. The system begins by capturing raw video frames, which are then converted into a format suitable for analysis. Each frame goes through preprocessing steps such as resizing, normalization, and noise reduction to ensure consistency.

Facial landmark detection is done using advanced frameworks that find key points on the face. These landmarks provide important details about facial structure and movement, which are essential for emotion recognition.

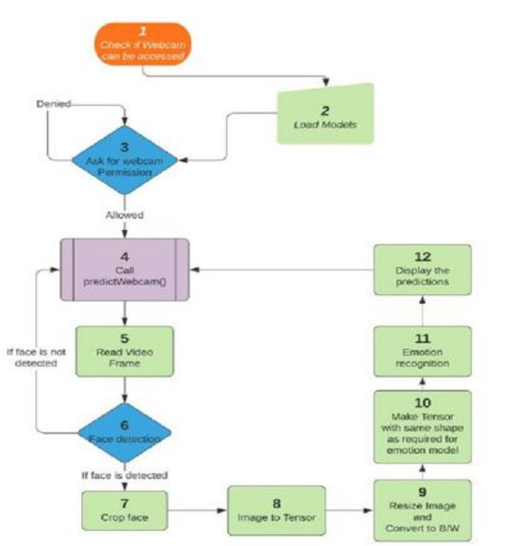
The processed data is stored temporarily in memory and sent through the model for analysis. Efficient data handling techniques are used to reduce processing time and ensure smooth operation. The system is also built to handle large volumes of data, making it scalable for real-world use.

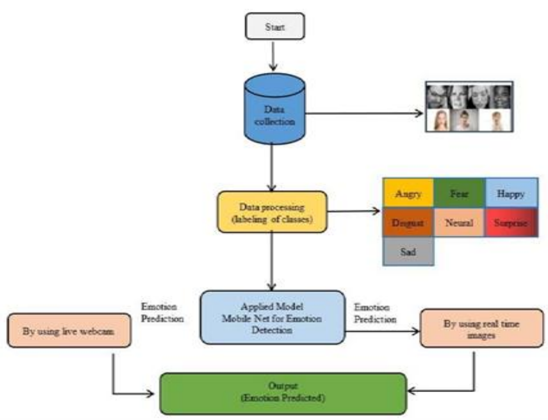
III.2.4 SECURITY AND ACCESS CONTROL

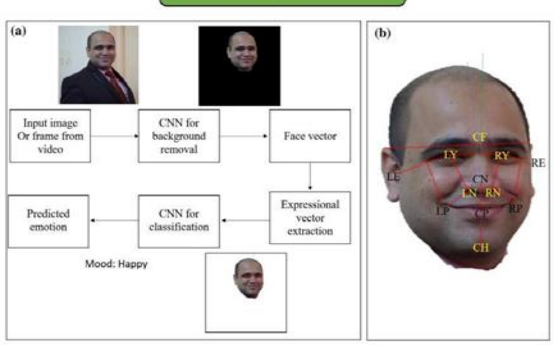
Since the system handles sensitive facial data, security and privacy are very important. The system is built to process only the necessary features for emotion detection, without keeping raw images or personal details.

To protect the system, strong authentication methods are used so only authorized users can access it. Role-based access control is also in place to limit who can perform administrative tasks. Data is sent securely using encryption, which helps prevent unauthorized access or leaks.

By using privacy-focused methods, the system follows ethical standards and data protection laws.







IV. SYSTEM EVALUATION

The evaluation of this system looks at how well it performs, how efficient it is, and how easy it is to use in real-life situations.

Unlike models that are just ideas, this system is made for actual use, so testing it in real time is important.

IV.1 FUNCTIONAL EVALUATION

This part checks if the system does what it's supposed to. The system was tested to see if it can find faces, get features, and correctly classify emotions.

The results show that the system can find faces in real time and assign the right emotion labels.

Each face has a unique prediction, ensuring accurate tracking. The system works consistently even when there are multiple faces.

IV.2 PERFORMANCE EVALUATION

This checks how fast and smooth the system runs. The system has low delay, with processing times usually under 50 milliseconds per frame. This makes real-time use possible without noticeable lags.

Using a GPU improves how quickly the system works, allowing it to handle many inputs at once. The central design helps data move smoothly, reducing delays and boosting performance.

IV.3 USABILITY AND TRANSPARENCY

Usability is measured by how easy it is to use and how good the user experience is. The system has a simple and easy-to-use interface, letting people see emotions in real time without needing technical skills.

Transparency is achieved through real-time updates and clear visual feedback. This makes it easy for users to understand the system's results, which builds trust and reliability

IV.4 COMPARATIVE ANALYSIS

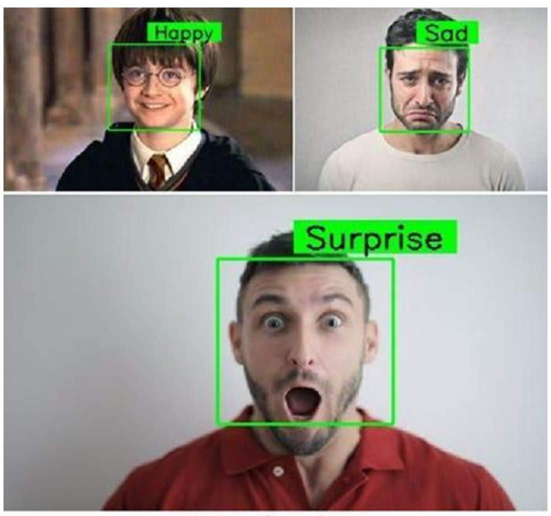
Compared to traditional methods, this system offers much better accuracy, speed, and ability to handle more users. Using RMN helps the system perform better than standard CNN models by focusing on key features and cutting out noise.

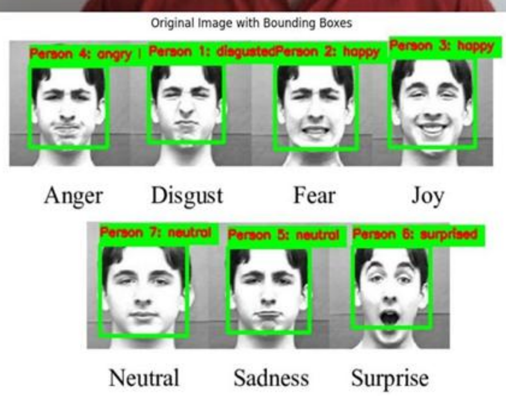
V. RESULTS

The results show the system is effective at detecting emotions in real time.

It performs well in different situations and lighting conditions, showing it's reliable and adaptable. The real-time feedback lets users see emotional changes instantly, making it good for interactive uses.

The clear workflow and optimized processing ensure the system runs well without any loss in performance.





VI. CONCLUSION AND FUTURE SCOPE

VI.1 CONCLUSION

This project created a real-time emotion detection system using facial expressions and a Residual Masking Network (RMN).

The system is accurate, fast, and inclusive, solving problems like delays, biased data, and privacy issues. It uses advanced deep learning and tools like OpenCV and MediaPipe for facial landmark detection, showing strong performance in various real-world uses like mental health, education, customer service, and security. The project shows how emotion detection can connect human feelings with AI, making a big impact in many industries.

VI.2 FUTURE SCOPE

The project can be improved in these ways:

1. Multimodal Emotion Recognition: Add data from other sources like voice and body language to make emotion detection more accurate. Build a system that can analyze emotions from multiple channels.

2. Edge Computing: Use devices like smartphones or IoT gadgets to run the system offline, making it less dependent on cloud services.

3. Improved Robustness: Improve performance in low-light and when faces are partially covered by using more varied training data and better preprocessing techniques.

4. Scalability and Optimization: Make the system work well on low-power devices while keeping real-time performance. Expand the model to manage large crowds for public safety and events.

5. Ethical and Regulatory Compliance: Improve privacy by using federated learning and making sure the system complies with global AI and data protection rules.

REFERENCES

[1] B. C. Ko, "A brief review of facial emotion recognition based on visual information" Sensors (Switzerland), vol. 18, no. 2, 2018.

[2] Y. Wu, H. Liu and H. Zha, "Modeling facial expression space for recognition" in IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2005), 2005, pp. 1968-1973.

[3] J. H. Yu, K. E. Ko and K. B. Sim, "Facial point classifier using convolution neural network and cascade facial point detector" Journal of Institute of Control, Robotics and Systems, vol. 22, no. 3, pp. 241-246, 2016.

[4] W. Zhang, Y. Zhang, L. Ma, J. Guan and S. Gong, "Multimodal learning for facial expression recognition" Pattern Recognition, vol. 48, no. 10, pp. 3191–3202, 2015.

[5] "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches" in 14th Symposium on Neural Networks and Applications (NEUREL), IEEE, 2018, by Aneta Kartali, Miloš Roglić, Marko Barjaktarović, Milica Đurić-Jovičić, Milica M. Janković.

[6] "A Deep Learning Approach for Real- Time Facial Emotion Recognition" Published in the 10th International Conference on System Modeling & Advancement in Research Trends (SMART) in 2022, Rupali Gill and Jaiteg Singh.

[7] "Comprehensive Review and Analysis on Facial Emotion Recognition: Performance Insights into Deep and Traditional Learning with Current Updates and Challenges" in Computers, Materials & Continua, vol. 82, no. 1, 2025, by A. Rehman, M. Mujahid, A. Elyassih, B. AlGhofaily, and S. A. O. Bahaj.

[8] "Real Time Facial Emotion Recognition using Deep Learning Models" in International Journal of Computer Applications, vol. 186, no. 29, 2023, by Priya Singh and Aman Gupta.

[9] "Dual Integrated Convolutional Neural Network for Real-Time Facial Expression Recognition in the Wild" in The Visual Computer, vol. 38, 2022, by S. Saurav, R. Saini, and S. Singh