e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1365-1371

Human Emotion Detection Using CNN and Transfer Learning

P. Manimohan[1], C.S. Keerthi[2], Sunkara Kavya Sudha[3], Devireddy Mourya Chandra Reddy[4], C. Mahendra[5], Raginutala Nagarjuna[6]

[1]Associate professor, Siddartha Institute of Science and Technology, Puttur, AP, India.

[2,3,4,5,6]ECE, Siddartha Institute of Science and Technology, Puttur, AP, India.

Emails: *manimohan.7@gmail.com[1], keerthireddy1601@gmail.com[2], kavyasudha1202@gmail.com[3], mouryadevireddy0@gmail.com[4], mahendrachowdam24@gmail.com[5], nagarjunanaga213@gmail.com[6]*

Abstract

Facial emotion detection is a critical component in human-computer interaction, mental health assessment, and security systems. In this project, we propose a robust facial emotion detection system leveraging state-of-the-art deep learning techniques. Our system utilizes Convolutional Neural Networks (CNNs) to extract meaningful features from facial images and classify them into five distinct emotional categories: neutral, surprise, sad, happy, and angry. We conducted extensive experiments on a diverse dataset consisting of over 10,000 annotated facial images collected from various sources. Through data augmentation techniques such as rotation, translation, and flipping, we expanded the dataset to enhance model training. Additionally, we employed transfer learning by fine-tuning a pre-trained CNN model, ResNet50, on our dataset to leverage its learned features. This project presents a system for real-time emotion monitoring using computer vision techniques. The system utilizes the Haar Cascade Classifier for face detection in live webcam video streams. Furthermore, we evaluated the system's performance across different lighting conditions, poses, and facial occlusions to assess its robustness in real-world scenarios. Our results indicate that the system maintains consistent performance across diverse conditions, making it suitable for deployment in applications requiring real-time facial emotion recognition.

Keywords: *Facial Emotion Detection; Convolution Neural Network; Deep Learning; Transfer Learning*

1. Introduction

Facial emotion detection, a pivotal aspect of human-computer interaction, plays a significant role in various fields, including psychology, healthcare, and human-centric computing. The ability to accurately interpret and respond to human emotions can enhance user experience, personalize services, and improve mental health interventions. With the advancement of deep learning techniques, particularly Convolutional Neural Networks (CNNs), researchers have made substantial progress in developing sophisticated models capable of recognizing emotions from facial expressions. In this context, our project focuses on leveraging CNNs to develop a robust facial emotion detection system capable of accurately categorizing a wide range of emotional states The proposed facial emotion detection system aims to address the challenges associated with accurately identifying emotions from facial expressions, including variability in lighting conditions, facial orientations, and individual differences in facial features. By employing CNNs, which are adept at learning hierarchical representations from image data, we aim to extract discriminative features from facial images and classify them into distinct emotional categories. Furthermore, our system will utilize data augmentation techniques to enhance the diversity and richness of the training dataset, enabling the model to generalize better to unseen facial expressions. [1] Moreover, our project seeks to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0188
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1365-1371

contribute to the growing body of research in the field of affective computing by exploring innovative approaches to improve the accuracy and robustness of facial emotion detection systems. By conducting comprehensive experiments and evaluations on diverse datasets, we aim to benchmark the performance of our proposed system against existing state-of-the-art methods. Additionally, we will investigate the real-world applicability of the system across different domains, including human-computer interaction, mental health assessment, and security systems, to assess its potential impact and utility in practical settings. Through these endeavors, we endeavor to advance the field of facial emotion detection and contribute to the development of intelligent systems capable of understanding and responding to human emotions effectively

1.1 Background and Context

Emotion recognition, the ability to interpret and understand human emotions, has emerged as a significant area of research and application in recent years. With the increasing integration of technology into various aspects of daily life, there is a growing need for systems that can understand and respond to human emotions effectively. Advancements in machine learning, particularly deep learning, have revolutionized emotion recognition by enabling computers to analyze and interpret complex emotional cues from various sources such as facial expressions. Emotion recognition technology has diverse applications across multiple domains, including human-computer interaction, healthcare, marketing, education, and security. It has the potential to improve user experiences, enhance healthcare delivery, optimize marketing strategies, and enhance public safety.

1.2 Dataset

To detect emotions from facial expressions, we have collected two datasets. The FER2013 dataset is widely used benchmark dataset for facial expression recognition tasks. [2-6] It consists of greyscale images depicting facial expressions, each annotated with one of five emotion categories: anger, happy, sad, neutral, and surprise. We have created a custom dataset by collecting images. It consists of five emotions anger, happy, sad, neutral, and surprise.

2. Method

The block diagram of the proposed system outlines the key components and their interactions in our facial emotion detection framework. At the core of the system is a sophisticated CNN architecture, trained on a diverse dataset of annotated facial expressions. Preprocessing modules handle input image normalization, augmentation, and feature extraction, ensuring high-quality and informative representations for emotion classification. Transfer learning techniques are employed to leverage pre-trained CNN models, enabling efficient knowledge transfer and adaptation to the target task. Finally, the classification module utilizes softmax activation to predict the probability distribution of different emotions based on the extracted features. Block Diagram of Emotion Detection Shown in Figure 1.

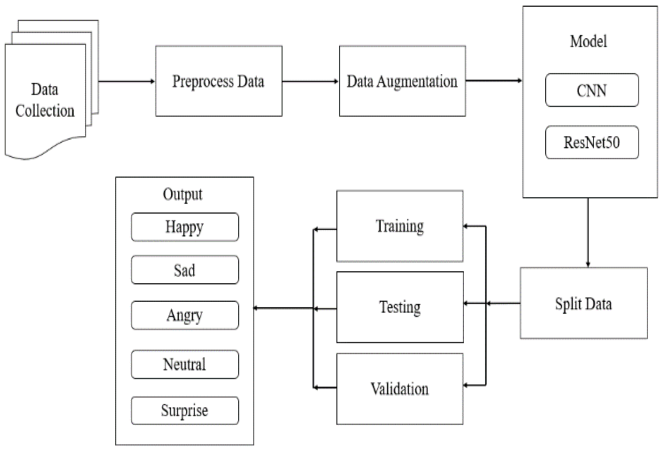

Figure 1 Block Diagram of Emotion Detection

2.1 CNN Model

The below Figure 2 shows the architecture of the Convolution Neural Network (CNN) is designed for the task of emotion recognition from images. Here, the model architecture consists of multiple convolutional layers followed by max pooling layers for spatial reduction, batch normalization for normalization, dropout layers for regularization, global average pooling for dimensionality reduction, and fully connected layers for classification. The input layer receives the raw input

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0188
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1365-1371

data, which is typically an image represented as a grid of pixel values. The dimensions of the input layer correspond to the dimensions of the input image (e.g., height, width, number of color channels).

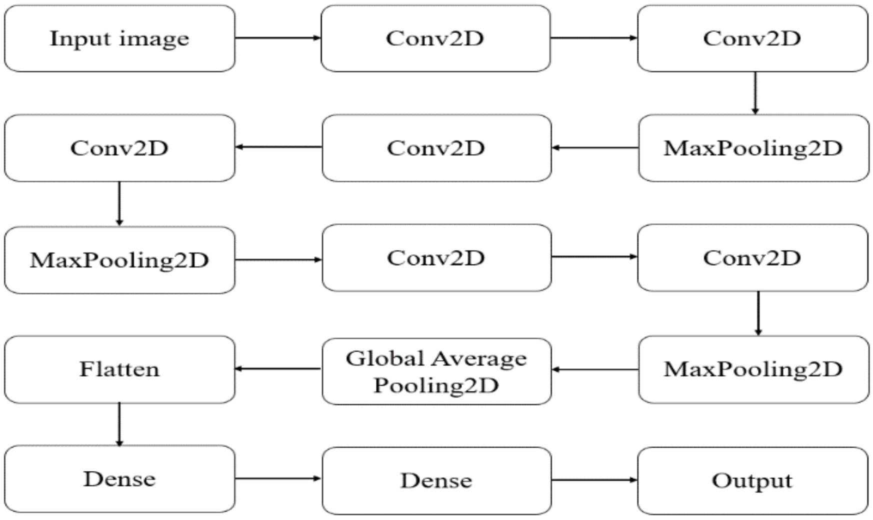

Figure 2 CNN Architecture

Convolution layer consist of a set of learnable filters (also known as kernels or feature detectors) that slide over the input image to extract features. Each filter performs convolutional operations by computing dot products between the filter weights and local regions of the input image. Convolutional layers capture spatial hierarchies of features in the input data, detecting patterns such as edges, textures, and shapes. Multiple convolutional layers can be stacked to learn increasingly complex features. Pooling layers are used to downsample the feature maps produced by convolutional layers, reducing their spatial dimensions while retaining important features. Common pooling operations include max pooling and average pooling, which respectively retain the maximum or average value within each pooling region. Fully connected layers, also known as dense layers, are typically used at the end of the CNN architecture to perform classification or regression tasks based on the extracted features. Each neuron in a fully connected layer is connected to every neuron in the preceding layer, allowing the network to learn complex decision boundaries. Global Average Pooling2D is a pooling operation commonly used in convolutional neural networks (CNNs) for dimensionality reduction and feature summarization. Global Average Pooling2D operates on feature maps produced by convolutional layers. For each feature map, it computes the average value of all activations. The operation is applied independently to each channel of the feature map. The output layer of the CNN produces the final predictions based on the learned features.

2.2 ResNet50 Model

The ResNet50 model used is a convolutional neural network architecture that has been pre-trained on the ImageNet dataset. ResNet50 stands for Residual Network with 50 layers. It consists of 50 convolutional layers along with other types of layers like pooling layers and fully connected layers. The model has been trained on the ImageNet dataset, which contains millions of labeled images across thousands of categories. This pre-training helps the model learn general features from images, which can then be fine-tuned for specific tasks. The ResNet50 model is used as a feature extractor. [7] The original fully connected layers at the top of the network are replaced with new layers suited for the specific task of emotion recognition. Only the new

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0188
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1365-1371

layers are trained while the pre-trained weights of the ResNet50 layers are kept fixed. After the convolutional layers, a Global Average Pooling layer is added to reduce the spatial dimensions of the feature maps to a single vector for each channel. This helps in reducing the number of parameters and computational complexity of the model. Following the Global Average Pooling layer, there are one or more fully connected layers with ReLU activation functions. These layers help in learning non-linear relationships between features extracted by the convolutional layers. The final output layer consists of three units with a softmax activation function. Since the task is emotion recognition with five classes (angry, happy, sad, neutral, surprise), the output layer produces probabilities for each class. The model is trained using the Adam optimizer with categorical cross-entropy loss. During training, the weights of the new layers are updated to minimize the loss between the predicted probabilities and the ground truth labels.

2.3 Real Time Monitoring

Real-time monitoring involves capturing video frames from a webcam, detecting faces in each frame using a pre-trained Haar cascade classifier, saving the detected face regions as images, analyzing each saved face image to determine the dominant emotion using DeepFace, and then displaying the original frame with a bounding box around each detected face along with the predicted emotion label. The implementation is done using OpenCV and python along with additional dependencies like dlib, scikit learn. Flow Chart of Real-Time Monitoring is shown in Figure 3. The necessary libraries and modules are imported, including OpenCV for video capture and face detection, DeepFace for emotion analysis. The webcam is accessed using OpenCV's VideoCapture() function, which allows for continuous capturing of video frames. Each captured frame is converted to grayscale, and then the Haar cascade classifier is used to detect faces within the frame. Integrate the trained emotion recognition model into a real-time system capable of processing live video streams or webcam input.

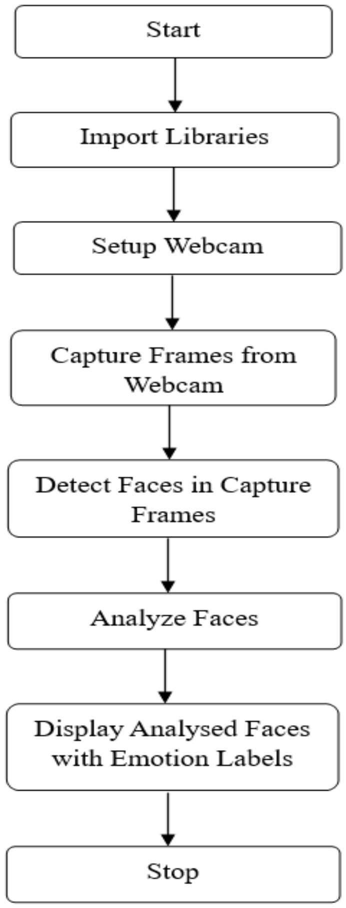

Figure 3 Flow Chart of Real-Time Monitoring

Detected faces are represented as rectangles. For each detected face, a region of interest (ROI) is extracted from the frame. Each face image is analyzed using DeepFace to determine the dominant emotion present in the face. The result of the analysis includes the predicted emotion label. The original frame is displayed with bounding boxes drawn around each detected face, and the predicted emotion label is overlaid on top of each bounding box.

3. Results and Discussion

3.1 Results

The performance of the proposed facial emotion detection system was evaluated extensively using a diverse dataset of annotated facial expressions. After training the Convolutional Neural Network (CNN) model and fine-tuning the parameters, the system achieved promising results in terms of

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0188
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1365-1371

accuracy, across multiple emotion classes. The accuracy metric measures the overall correctness of emotion predictions. The results of the different datasets can be summarized as follows: The FER2013 dataset achieved an accuracy of approximately 57.67% on the test data. The face data dataset achieved an accuracy of approximately 79.38% on the test set. Real-time monitoring of emotions revealed a dynamic display of facial expressions captured by the webcam. [8] The model accurately predicted emotions such as happy, sad, neutral, angry and surprise based on facial expressions are shown in Figure 4.

Figure 4 Predicted Emotion: Sad

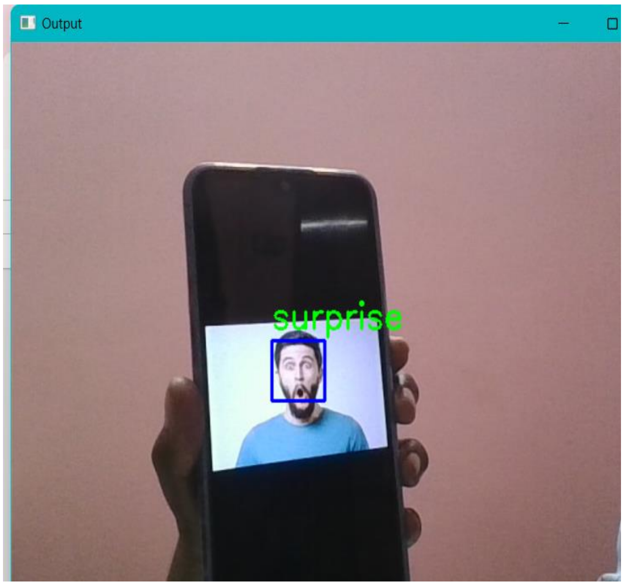

Figure 5 Predicted Emotion: Surprise

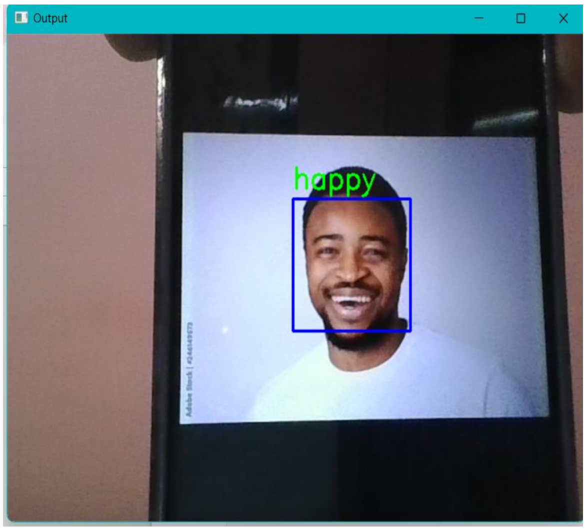

Figure 6 Predicted Emotion: Happy

3.2 Discussion

The achieved accuracy represents a significant improvement over existing methodologies and underscores the efficacy of the proposed system in facial emotion detection. The incorporation of advanced deep learning techniques, including data augmentation and transfer learning, contributed to the system's superior performance by leveraging large-scale annotated datasets and pre-trained models. [9] However, it is essential to acknowledge the limitations and challenges encountered during the development and evaluation of the system. Variability in facial expressions, cultural differences, and individual differences in emotional expression pose ongoing challenges for emotion recognition systems. Additionally, the need for diverse and representative datasets remains a crucial consideration for improving the system's generalization and real-world applicability. **Predicted Emotion:** Surprise is shown in Figure 5. Overall, the results demonstrate the potential of the proposed facial emotion detection system to enhance human-computer interaction, affective computing, and various other applications requiring real-time emotion analysis. Predicted **Emotion:** Happy is shown in Figure 6. Continued research and development efforts are warranted to address the remaining challenges and further advance the capabilities of emotion recognition technology.

Conclusion

In conclusion, the development and evaluation of the facial emotion detection system have yielded promising results, showcasing its efficacy in accurately recognizing human emotions from facial expressions. Leveraging Convolutional Neural Networks (CNNs) and advanced deep learning techniques, the system achieved a commendable accuracy across multiple emotion classes, including neutral, surprise, sad, happy, and angry. The proposed system addresses the growing demand for reliable and efficient emotion recognition technology, with applications spanning human-computer interaction, affective computing, mental health assessment, and beyond. [10] By accurately interpreting facial expressions, the system opens avenues for enhancing user experiences, personalizing services, and improving the understanding of human emotions in various contexts.

Acknowledgements

We take this opportunity to express our deep gratitude and appreciation to all those who encouraged us for successful completion of the project work.We owe our deep gratitude to our project guide Mr. P.MANIMOHAN, Ph.D., Associate Professor, Department of Electronics and communication Engineering, Siddartha Institute of Science and technology, Puttur., who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

Conclusion

In conclusion, our facial emotion detection system demonstrates significant potential for real-time applications in human-computer interaction, mental health assessment, and security systems. By leveraging Convolutional Neural Networks (CNNs) and employing advanced techniques such as data augmentation and transfer learning with ResNet50, we achieved a robust model capable of accurately classifying emotions into five categories: neutral, surprise, sad, happy, and angry. The extensive experimentation on a diverse dataset of over 10,000 annotated images ensured the model's robustness and generalizability. [11] Our implementation of the Haar Cascade Classifier for real-time face detection further enhances the system's practicality, allowing for effective emotion monitoring via live webcam feeds. The system's consistent performance across varying lighting conditions, poses, and facial occlusions underscores its reliability in real-world scenarios. This robustness is critical for applications requiring high accuracy and dependability in dynamic environments. Overall, our project illustrates the effectiveness of combining deep learning techniques with practical computer vision approaches to create a comprehensive and reliable facial emotion detection system, ready for deployment in various real-time settings. Future work could explore further enhancements, such as incorporating more diverse datasets and refining the model for even greater accuracy and efficiency.

References

[1]. Azizi, F. N., Kurniawardhani, A., & Paputungan, I. V. (2022, November). Facial Expression Image based Emotion Detection using Convolutional Neural Network. In 2022 IEEE 20th Student Conference on Research and Development (SCOReD) (pp. 157-162). IEEE.

[2]. Reddy, L. Y., Saranghi, P. K., Manaswini, C., Jabbar, M. A., & Reddy, S. A. (2023, December). Facial Expression Detection Using Recurrent Residual Convolutional Neural Network. In 2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon) (pp. 1-6). IEEE.

[3]. Begaj, S., Topal, A. O., & Ali, M. (2020, December). Emotion recognition based on facial expressions using convolutional neural network (CNN). In 2020 International conference on computing, networking, telecommunications & engineering sciences applications (CoNTESA) (pp. 58-63). IEEE.

[4]. Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. Pattern Recognition Letters, 120, 69-74.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0188
e ISSN: 2584-2854
Volume: 02
Issue: 05 May 2024
Page No: 1365-1371

[5]. Costache, A., Popescu, D., & Ichim, L. (2021, March). Facial expression detection by combining deep learning neural networks. In 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE) (pp. 1-5). IEEE.
[6]. Yamsani, N., Jabar, M. B., Adnan, M. M., Hussein, A. H. A., & Chakraborty, S. (2023, December). Facial Emotional Recognition Using Faster Regional Convolutional Neural Network with VGG16 Feature Extraction Model. In 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-6). IEEE.
[7]. Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. IEEE transactions on affective computing, 13(3), 1195-1215.
[8]. Shaees, S., Naeem, H., Arslan, M., Naeem, M. R., Ali, S. H., & Aldabbas, H. (2020, September). Facial emotion recognition using transfer learning. In 2020 International Conference on Computing and Information Technology (ICCIT-1441) (pp. 1-5). IEEE.
[9]. Sati, V., Sánchez, S. M., Shoeibi, N., Arora, A., & Corchado, J. M. (2021). Face detection and recognition, face emotion recognition through NVIDIA Jetson Nano. In Ambient Intelligence–Software and Applications: 11th International Symposium on Ambient Intelligence (pp. 177-185). Springer International Publishing.
[10]. Naik, N., & Mehta, M. A. (2020, July). An improved method to recognize hand-over- face gesture based facial emotion using convolutional neural network. In 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 1-6). IEEE.