IJARSCT ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53 Volume 4, Issue 2, April 2024

Research On Real Time Emotion Recognition Using Digital Image Processing Using ML

Prof. Bina Rewatkar, Ashwini Mandhare, Komal Chanekar, Sonal Barapatre, Harshali Kadu, Karan Nimje, Akash Tembhurne

UG Student, Department of Computer Science and Engineering

Professor, Department of Computer Science and Engineering

Nagarjuna Institute of Engineering Technology and Management, Nagpur, Maharashtra, India

Abstract*: Face recognition technology has garnered significant attention in recent years due to its wide range of applications in various fields such as security, surveillance, biometrics, and human-computer interaction. This abstract provides a comprehensive overview of the advancements in face recognition image processing techniques, methodologies, and applications. The abstract begins by elucidating the fundamental concepts underlying face recognition, including feature extraction, dimensionality reduction, and classification algorithms. Various approaches such as Eigenfaces, Fisher faces, and Local Binary Patterns (LBP) are discussed, highlighting their strengths and limitations.*

Moreover, recent developments in deep learning techniques, particularly convolutional neural networks (CNNs) and Siamese networks, have revolutionized face recognition by achieving remarkable accuracy and robustness. The abstract delves into the architecture and training procedures of these deep learning models, emphasizing their ability to learn discriminative features directly from raw pixel data. Furthermore, the abstract explores the challenges faced by face recognition systems, such as variations in pose, illumination, expression, and occlusion. Techniques for addressing these challenges, including data augmentation, normalization, and adversarial training, are examined.

Keywords: Machine learning, Emotion detection, Python-based framework, CNN (Convolutional Neural Network).

I. INTRODUCTION

This introduction provides an overview of the principles, techniques, and applications of face recognition in image processing. At its core, face recognition involves three main stages: face detection, feature extraction, and classification. Face detection algorithms locate and localize faces within an image, while feature extraction techniques extract discriminative facial features, such as edges, textures, or key landmarks. Finally, classification algorithms, ranging from traditional methods like Eigenfaces to state-of-the-art deep learning approaches like convolutional neural networks (CNNs), match the extracted features against a database of known faces for identification or verification. Over the years, significant advancements have been made in face recognition technology, propelled by the advent of deep learning. Deep learning-based approaches have demonstrated superior performance in handling variations in pose, illumination, facial expressions, and occlusions, thereby significantly enhancing the accuracy and robustness of face recognition systems. The applications of face recognition in image processing are diverse and far-reaching. In security and surveillance, it is employed for access control, monitoring, and forensic analysis.

II. LITERATURE SURVEY

1] Zixing Zhang, Fabien Ringeval, Eduardo Coutinho, Erik Marchi and Björn Schüller proposed some improvement in SSL technique to improve the low performance of a classifier that can deliver on challenging recognition tasks reduces the trust ability of the automatically labelled data and gave solutions regarding the noise accumulation problem - instances that are misclassified by the system are still used to train it in future iterations. They exploited the complementarity between audio-visual features to improve the performance of the classifier during the supervised phase. Then, they iteratively re-evaluated the automatically labelled instances to correct possibly mislabeled data and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

this enhances the overall confidence of the system's predictions. This technique gives the best possible performance using SSL technique where labelled data is scarce and/or expensive to obtain but still, there are various inherent limitations that limit its performance in practical applications.

2] Wei-Long Zheng and Bao-Liang Lu proposed EEG-based effective models without labelled target data using transfer learning techniques (TCA-based Subject Transfer) which is more accurate in terms of positive emotion recognition than other techniques used before. Their method achieved 85.01% accuracy. They used to transfer learning and their method includes three pillars, TCA-based Subject Transfer, KPCA-based Subject Transfer and Transudative Parameter Transfer. For data preprocessing they used raw EEG signals processed with a bandpass filter between 1 Hz and 75 Hz and for feature extraction, they employed differential entropy (DE) features. For evaluation, they adopted a leave-one subject-out cross-validation method. This method is limited in terms of negative and neutral emotion recognition. Yet a lot of improvement needed to recognize negative and neutral emotion more accurately.

3] Pal, Shantanu, et al. In this paper, the author presents the development and progress in sensors and technologies to detect human emotions. They review the state-of-the-art sensors used for human emotion recognition and different types of activity monitoring. They present the design challenges and provide practical references of such human emotion recognition systems in the real world. Finally, they discuss the current trends in applications and explore the future research directions to address issues, e.g., scalability, security, trust, privacy, transparency, and decentralisation.

4] A. Stuhlsatz, et al. The author, in this paper, compares the performance of GerDA features and their subsequent linear classification with previously reported benchmarks obtained using the same set of acoustic features classified by Support Vector Machines (SVMs). Our results impressively show that low-dimensional GerDA features capture hidden information from the acoustic features leading to a significantly raised unweighted average recall and considerably raised weighted average recall.

5] Y. Fan, X. Lu, D. Li, and Y. Liu. proposed a method for video-based emotion recognition in the wild. They used CNN-LSTM and C3D networks to simultaneously model video appearances and motions [16]. They found that the combination of the two kinds of networks can give impressive results, which demonstrated the effectiveness of the method. In their proposed method they used LSTM (Long Short-Term Memory) - a special kind of RNN, C3D – A Direct Spatio-Temporal Model and Hybrid CNN-RNN and C3D Networks.

III. METHODOLOGY

The proposed system aims to develop a Python-based real-time emotion recognition framework utilizing digital image processing and machine learning techniques.

This system will enable real-time video streaming through an inbuilt camera, allowing it to operate seamlessly on live streaming data.

The core of the system will be powered by CNN algorithms, which will be employed to accurately detect and categories facial expressions, thereby providing a comprehensive solution for real-time emotion detection in various interactive applications.

A CNN is a DL algorithm which takes an input image, assigns importance (learnable weights and biases) to various aspects/objects in the image and is able to differentiate between images. The preprocessing required in a CNN is much lower than other classification algorithms. Figure shows the CNN operations.

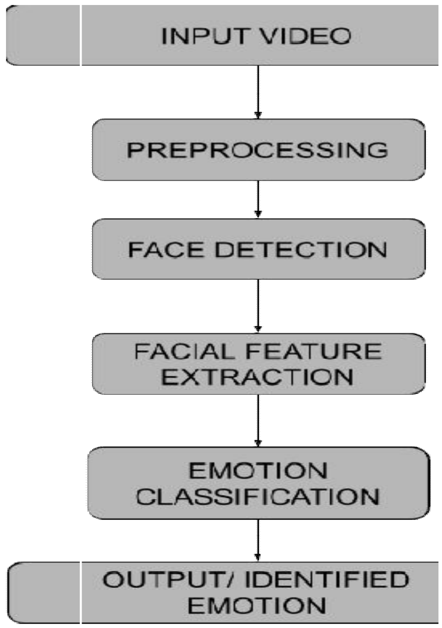
The architecture of a CNN is analogous to that of the connectivity pattern of neurons in the human brain and was inspired by the organization of the visual cortex.

One role of a CNN is to reduce images into a form which is easier to process without losing features that are critical for good prediction. This is important when designing an architecture which is not only good at learning features but also is scalable.

to massive datasets. Tialhe main CNN operations are convolution, pooling, batch normalization and dropout which are described below.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

Flowchart



IV. PROPOSED SYSTEM

The proposed system aims to develop a Python-based real-time emotion recognition framework utilizing digital image processing and machine learning techniques. This system will enable real-time video streaming through an inbuilt camera, allowing it to operate seamlessly on live streaming data. The core of the system will be powered by CNN algorithms, which will be employed to accurately detect and categories facial expressions, thereby providing a comprehensive solution for real-time emotion detection in various interactive applications.

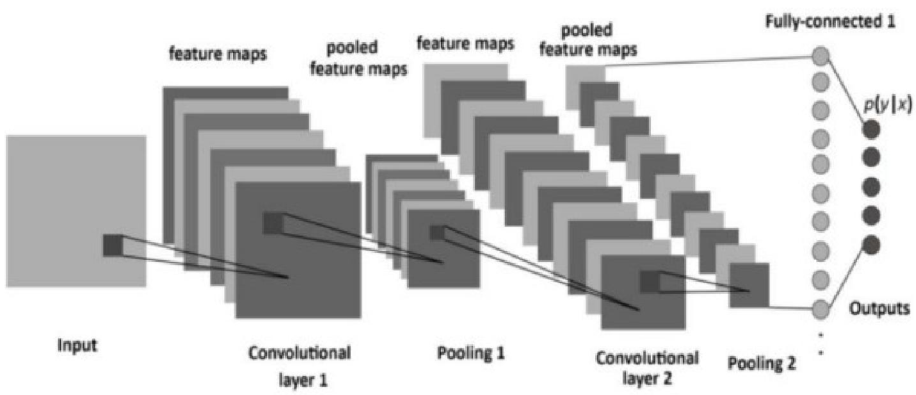


Fig. a. Convolutional Neural Network

Convolution operation

The objective of the convolution operation is to extract high level features such as edges from an input image. The convolution layer functions are as follows. The first convolutional layer(s) learns features such as edges, color, gradient orientation and simple textures. The next convolutional layer(s) learns features that are more complex textures and patterns. The last convolutional layer(s) learns features such as objects or parts of objects. The element involved in carrying out the convolution operation is called the kernel. A kernel filters everything that is not important for the feature map, only focusing on specific information. The filter moves to the right with a certain stride length till it parses the complete width. Then, it goes back to the left of the image with the same stride length and repeats the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

process until the entire image is traversed. Figure 3.5 presents an image with dimensions 5 × 5 (shown in green) and the following 3 × 3 kernel filter.

$$\begin{matrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{matrix}$$

The stride length is chosen as one so the kernel shifts nine times, each time performing a matrix multiplication of the kernel and the portion of the image under it.

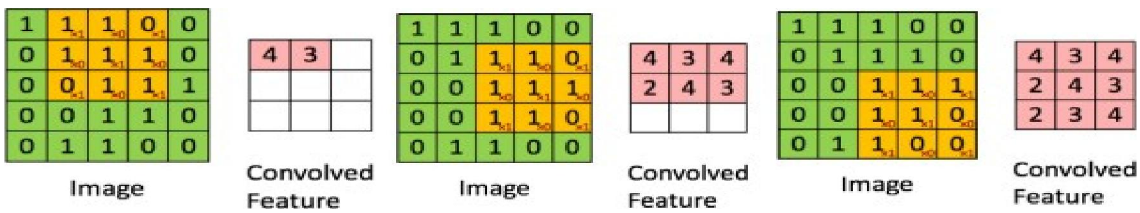


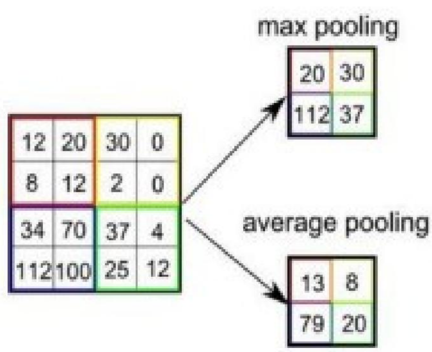
Figure Convolving a 5 × 5 image with a 3 × 3 kernel to get a 3 × 3 convolved.

The convolved feature can have the same dimensions as the input or the kernel. This is done by the same or valid padding. Same padding is when the convolvedfeature has the dimensions of the input image and valid padding is when this feature has the dimensions of the kernel.

Pooling operation

The pooling layer reduces the spatial size of a convolved feature. This is done todecrease the computations required to process the data and extract dominant features which are rotation and position invariant. There are two types of pooling, namely max pooling and average pooling. Max pooling returns the maximum value from the portion of the image covered by the kernel, while average pooling returns the average of the corresponding values. Figure 3.6 shows the outputs obtained by performing max and average pooling on an image.

Fully connected layer

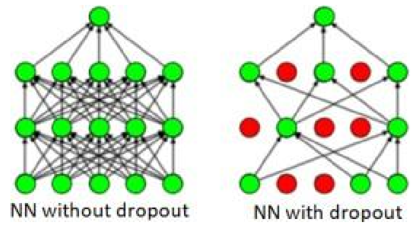


Neurons in a fully connected layer have connections to all neurons in the previous layer. This layer is found towards the end of a CNN. In this layer, the input from the previous layer is flattened into a one-dimensional vector and an activation function is applied to obtain the output.

Dropout

Dropout is used to avoid overfitting. Overfitting in an ML model happens when the training accuracy is much greater than the testing accuracy. Dropout refers to ignoring neurons during training so they are not considered during a particular forward or backward pass leaving a reduced network. These neurons are chosen randomly and an example is shown in Figure The dropout rate is the probability of training a given node in a layer, where 1.0 means no dropout and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024



V. IMPLEMEMTATION

Step 1: Install Required Libraries

In the initial step of implementing "Real-time Emotion Recognition using Digital Image Processing with ML," the installation of TensorFlow libraries is crucial. TensorFlow is a widely-used open-source machine learning framework that facilitates the development and deployment of machine learning models. By executing the command pip install TensorFlow, installing TensorFlow and its necessary dependencies, ensuring that the project can leverage the framework for the subsequent stages of image processing and machine learning tasks. This foundational step is essential for seamlessly integrating TensorFlow into this project environment, enabling the implementation of advanced emotion recognition models.

Step 2: Import Libraries

In the second step of implementing "Real-time Emotion Recognition using Digital Image Processing with ML," the focus is on importing the necessary libraries, including TensorFlow. Once TensorFlow is successfully installed in the project, the next step involves importing relevant libraries to facilitate image processing and machine learning operations. The import statements, such as import TensorFlow as if and from koras. preprocessing import image, enable access to the functionalities provided by TensorFlow and Kera's, an open-source deep learning library. These imported libraries play a pivotal role in handling image data, constructing and deploying machine learning models, thereby laying the foundation for subsequent stages of the real-time emotion recognition project.

Step 3: Load the Pre-trained Emotion Recognition Model

In the third step of the "Real-time Emotion Recognition using Digital Image Processing with ML" project, a pre-trained emotion recognition model is loaded into the system. This involves using the TensorFlow and Kera's libraries to access and load a previously trained neural network model, specifically designed for recognizing emotions in images. The pre-trained model has learned patterns and features from a relevant dataset, making it capable of inferring emotional states from facial expressions. By loading this model, the project establishes a foundation for real-time emotion analysis, allowing subsequent frames from live video feeds to be processed and classified swiftly, contributing to the overall efficiency and accuracy of the emotion recognition system.

Step 4: Start Live Testing

In the fourth step of "Real-time Emotion Recognition using Digital Image Processing with ML," we initiate live testing by implementing a graphical user interface (GUI) to interactively capture video frames from a camera source and display the real-time emotion prediction. This involves using a library such as Tintern for GUI development in Python.

VI. RESULT

In the Result the Python-based project employing the CNN algorithm for real-time emotion recognition, the results are evident in the accurate classification of emotions in two videos featuring a boy expressing sadness and happiness.

In the first video portraying sadness, the CNN algorithm effectively analyses facial features in real-time and successfully categorizes the boy's expressions as "Sad." This demonstrates the robustness of the CNN model in capturing subtle nuances associated with sadness, such as changes in facial muscle patterns and overall expression.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

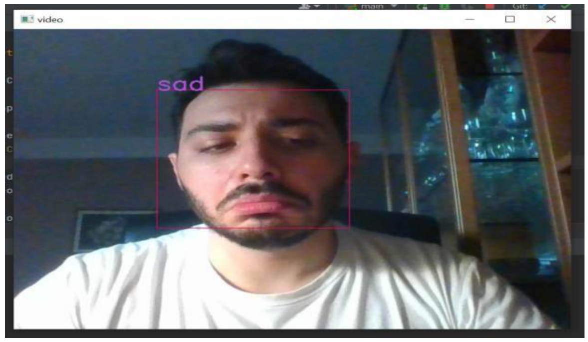


Fig. a. shows Sad emotions

Conversely, in the second video depicting happiness, the CNN algorithm appropriately recognizes and labels the boy's facial expressions as "Happy." The model demonstrates its ability to detect positive emotional cues, including smiling and uplifted features, showcasing its adaptability to varied emotional states.

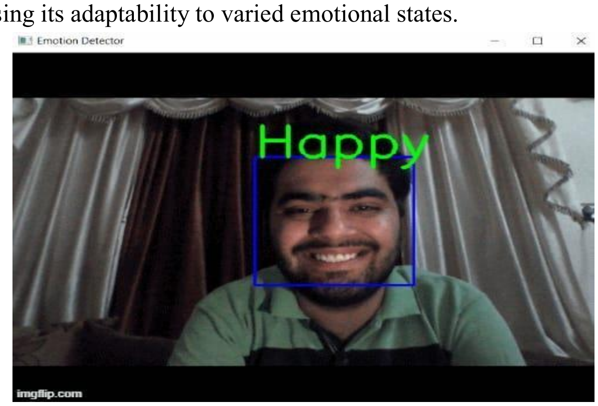


Fig. b. Shows happy emotions

VII. CONCLUSION

The development of a real-time emotion recognition system using digital image processing and machine learning holds significant promise for a wide array of applications. We explored a novel way of classifying human emotions from facial expressions. Thus, a neural network-based solution combined with image processing was proposed to classify the six universal emotions: (joy, anger, sadness, disgust, surprise, fear and neutral) in video streams. From enhancing human-computer interaction to revolutionizing the fields of mental health monitoring, personalized user experiences, and market research. While there are challenges to overcome, including privacy concerns and technical limitations, continuous research and advancement in this domain are likely to lead to even more accurate, efficient, and ethically responsible systems in the future, ultimately reshaping the way we interact with technology and understand human emotions.

REFERENCES

[1]. Zixing Zhang, Fabien Ringeval, Fabien Ringeval, Eduardo Coutinho, Erik Marchi and Björn Schüller, Semi-Supervised Learning (SSL) technique. Published in: 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016-3-20 - 2021-3-25.

[2]. Wei-Long Zheng1 and Bao-Liang Lu, Personalizing EEG- Based Affective Models with Transfer Learning, Published in: 2021. Proceedings of the -Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

[3]. Pal, Shantanu, Subhas Mukhopadhyay, and Nagender Suryadevara. "Development and progress in sensors and technologies for human emotion recognition." Sensors 21.16 (2021): 5554.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, April 2024

[4]. Stuhlsatz, C. Meyer, F. Eyben, T. ZieIke, G. Meier, and Schuller, — Deep neural networks for acoustic emotion recognition: raising the benchmarks, Published in: 2021.in Acoustics, Speech and Signal Processing (ICASSP), 2018 IEEE International Conference on. IEEE, pp. 5688–5691.

[5]. Y. Fan, X. Lu, D. Li, and Y. Liu. Video-based Emotion Recognition Using CNN-RNN and C3D Hybrid Networks. Proceedings of ICMI 2020 Proceedings of the 18th ACM International Conference Published in: 2020. on Multimodal Interaction, Pages 445-450, Tokyo, Japan — November 12 - 16.