Hand Gesture Recognition for Emoji Prediction

Jatin Gupta1 , Shreya Goel2 , Sanyam Varshney3 , Ms. Shafali Dhall4 , Ms. Neha Gupta5 1, 2, 3, 4Student, 5Assistant Professor, Department of Information Technology, Bharati Vidyapeeth's College of Engineering, New Delhi

Abstract:
Emojis are ideograms and smileys visual symbols that are used widely in wireless communication. They present rich novel possibilities of representation and interaction as a new modality. They exist in various genres, including hand gestures, human faces, figures, and signs. Hand gestures, which are the most common and intuitive non-verbal means of communication when we are using a computer, and related work, has recently sparked an interest. Hands often appear in images, videos, and their appearances and pose can give important clues about what people are doing. A combination of hand gestures and emojis can communicate and express the message very conveniently. Considering the positive outcomes of image recognition from specific deep learning methods, we suggest an emoji predictor in real-time. This project consists of a hand gesture recognition method and emoji generator using filters to detect hands and Convolutional Neural Network (CNN) for training the model. Here, a database is being created of hand gestures to train the system. The prediction will be focused on the hand movements and capture the changes by preparing for different hand movement positions to the highest degree of precision.
Keywords: Deep Learning, Convolutional Neural Network (CNN), Hand Detection, Gesture Recognition, Morphological Operation, Contours, Gaussian Blur, Emoji Prediction.

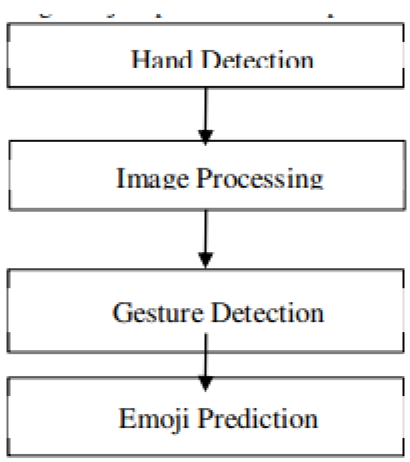
Introduction:
Deep learning is a prominent field within machine learning, excelling in tasks such as image recognition, object detection, language translation, and trend prediction by utilizing multiple layers to extract higher-level features from raw data. Hand and gesture recognition, a subfield, focuses on identifying gestures that encompass various body language and sign language cues. This research leverages image-based data for 11 different hand gestures, utilizing convolutional neural networks (CNNs) for classification. The study demonstrates the effectiveness of deep CNN architectures for static gesture recognition. As emojis have become integral to social communication, there's a growing need for instant emoji generation based on gestures. The paper discusses the methodology, approach, and experiments conducted, ultimately presenting the outcomes and offering insights for future research.

Concept used:
i) Morphological operations: Morphological operations in image processing involve the application of structural elements to process image shapes. They encompass operations like erosion and dilation, which respectively reduce object attributes and expand object areas. Additionally, opening combines dilation followed by erosion, while closing involves the reverse sequence, aimed at noise reduction and segmentation in image processing.
ii) Contour extraction:
Contours are curves representing areas of consistent color or intensity in an image, crucial for object detection and recognition. To improve contour accuracy and reduce noise in low-quality images, a polygon approximation technique can be applied, which simplifies the contour shape

by eliminating points farther from the curve than a set epsilon value. Identifying contours involves locating white objects on a black background in the image.
iii) Convolutional Neural networks:
Convolutional neural networks (CNNs) are instrumental for image classification, object recognition, and detection, comprising convolution, grouping, and pooling layers. Their architecture is customized based on the number of layers, neurons per layer, and activation methods. In the convolution layer, kernels process input images, generating feature maps. Following this, the pooling layer enhances training efficiency by reducing feature map size. Multi-layer neural perceptrons categorize objects based on extracted features. CNNs excel in automatic feature extraction and optimization during training through backpropagation, making them a pivotal tool for image analysis and machine learning tasks.

Methodology:
The techniques for segmentation, filters, and morphological manipulations to enhance the necessary design details

A) Dataset: In this project, train the models eleven emojis, and a comparison of the target (output) and training images can be seen. There are 1,200 corresponding image specific training images to every eleven emoji that maintain the stability of our training setup
B) Overview: The hand is detected by using the subtraction process of background, and the effect is converted into an image of contour. The gestures are then identified to make the Emoji prediction easier. The corresponding Emoji is predicted and represented in the frame.



C) Hand detection: A camera continuously monitors user-generated real-time hand gestures, capturing and forwarding images for processing. These images are compared with the dataset to predict the corresponding emoji for each gesture. All hand images are captured under consistent environmental conditions with identical backgrounds, and they are resized to a standardized 350x350 format for accurate analysis.
D) Image processing: The dataset augmentation involved ImageDataGenerator with horizontal flips. Input images from the webcam were resized to 50x50 to match the dataset's scale. They were then processed by converting to HSV and applying a black-white mask to align with dataset input requirements. Gaussian Blur improved

precision, and morphological operations, like dilation and closing, removed background noise and enhanced hand gesture shapes for analysis.

E) Gesture detection: The model utilizes a CNN architecture with convolutional, ReLu, and max-pooling layers in two sets, followed by a fully connected layer, SoftMax, and another ReLu layer for label prediction. CNNs automatically learn and extract features, making them highly effective for image classification. Preprocessing involves resizing, thresholding, and centering hand images, ensuring a level of invariance in scale and translation during CNN learning, enabling real-time gesture recognition.

F) Emoji Prediction: value is assigned while predicting a gesture used to blend the image to predict the corresponding Emoji. Images are given different weights, so it provides a feeling of transparency or blending.

Conclusion:

The project has achieved impressive test results with accuracy close to 1.0, highlighting the strength of deep learning in image recognition. To further enhance the model, considerations include implementing pre-trained image segmentation methods with a lighting threshold for noise reduction, exploring a wider range of lighting conditions and image quality, using multiple successive frames as inputs, and expanding the database to accommodate various hand configurations for a more tailored approach.

Issues and future scopes:

Real-time gesture recognition has a wide range of applications, including medical use for individuals with physical disabilities and commercial applications in consumer shops and homes. It enables interactions through hand gestures and emoji prediction for tasks like robot control and office or household applications. However, there are several potential issues to address. These include the need for robust classifiers to handle dynamic gestures and the development of complex motion recognition for advanced human-computer interaction systems. Additionally, ensuring privacy and data security, addressing real-time processing demands, and improving accuracy in recognizing intricate gestures are important considerations.