Gesture Recognition System using Computer Vision

Shreya Matta[1], Omkar Bare[2], Manish Jha[3], Bhavna Arora[4]

[1,2,3] *Student, Department of Computer Engineering, Atharva College of Engineering, Mumbai*
[4] *Professor, Department of Computer Engineering, Atharva College of Engineering, Mumbai*

---***---

Abstract - *By making computers more receptive to user needs, Human Computer Interaction improves the interaction between users and computers. Interaction with today's personal computers is not just confined to keyboard and mouse interaction. A host of sensory modes contribute to human interaction, such as gestures and facial expressions. As well as providing an overview of various methods that can be used to recognize hand and face gestures, the paper highlights the benefits and lists some of the limitations while implementing these techniques. We have further presented in this paper, a cost effective system for recognizing gestures and translating them into actions.*

Key Words: **Gesture Recognition, Neural Neural, Convolutional Neural Network, Transfer Learning, Pattern recognition, artificial neural networks, machine learning, image analysis**

1. INTRODUCTION

The purpose of building gesture recognition systems is to create a natural interface between humans and computers, where gestures recognized by the system can be used to perform certain tasks. Forming hand gestures that are easily understood and well interpreted by a computer is considered to be the key to gesture interaction. Human Computer Interaction systems should be designed based on two key characteristics: functionality and usability. The functionality of a system refers to the set of services a system provides its users, while the usability of a system refers to its ability to operate and accomplish specific tasks. When a system achieves a suitable balance between these concepts, it is considered powerful and influential. This paper discusses the advancement of gesture recognition systems and the stages necessary to build a complete system with a less error-prone algorithm. Using Convolutional Neural Networks (CNN) in real-time, we develop a gesture recognition system that converts video of a user's gesture into specific tasks to be performed.

There are three real time tasks we need to accomplish:

1. To obtain a video of the user
2. Analyzing each frame of the video to determine its gesture.
3. Completing a related task by reconstructing it.

1.1 Motivation

According to a report by the World Health Organization and the World Bank, 70 million people are disabled in some way. Individuals who are unable to perform tasks due to a physical ailment, injury, or other handicap reasons need a system which can do tasks using their hand and face gestures. We aim to develop a system which can solve this problem. [1]

1.2 Related work

ASL recognition isn't a new problem in computer vision. Researchers have utilized classifiers from a range of categories during the last two decades, which we can group together. linear classifiers, neural networks, and neural networks.

ASL translation has been tackled using neural networks. The ability of neural networks to learn the most relevant classification features is maybe their most significant benefit. They do, however, need a lot more time and data to train. The majority of them have been shallow to date. [4]

A major research initiative was conducted over two decades to develop human computer interaction. All of them were aimed to fetch the classified facial expressions. However, the accuracy achieved was not quite decent. [2]

2. PROPOSED SYSTEM

2.1 Construction of classifier for gestures

2.1.1 Neural Network

Neural networks performance on large-scale video classification is outstandingly well. We discovered that CNN architectures are capable of learning powerful features from poorly labeled data that outperform feature-based approaches, and that these advantages are remarkably resistant to changes in the architectures' connection over time. Interpretable errors are revealed through a qualitative assessment of network outputs and confusion matrices. [3]

For face emotion recognition we are using a novel technique called “facial emotion recognition using convolutional neural networks (FERC)”. With single-level

CNN, FERC varies from commonly used techniques, enhancing accuracy. Furthermore, prior to the formation of EV, a novel background removal approach is used to avoid dealing with many issues that may arise (for example distance from the camera). The extended Cohn–Kanade expression, Caltech faces, CMU, and NIST datasets were used to test FERC with over 750K photos.[3]

2.1.2 Transfer Learning

A convolutional neural network (CNN or ConvNet) is used to classify ASL letters. CNNs are machine learning algorithms that have had tremendous success in a range of tasks involving video and image processing. Since 2012, the discipline of picture categorization, object localization, and object detection has seen a surge in growth and applications.

The ability of CNN to learn features as well as the weights associated with each feature is a major benefit of using such algorithms. CNNs, like other machine learning algorithms, aim to improve a specific objective function, such as the loss function. We are used transfer learning

2.1.3 MobileNet

MobileNet is a CNN class that was open-sourced by Google, and it provides us with an ideal starting point for training our classifiers.

Depthwise separable convolutions are used by MobileNet. When compared to a network with regular convolutions of the same depth in the nets, it dramatically reduces the number of parameters. As a result, lightweight deep neural networks are created.

3. DATASET AND FEATURES:

The training data set contains 87,000 images which are 200x200 pixels. There are 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE and NOTHING. These 3 classes are very helpful in real-time applications, and classification.

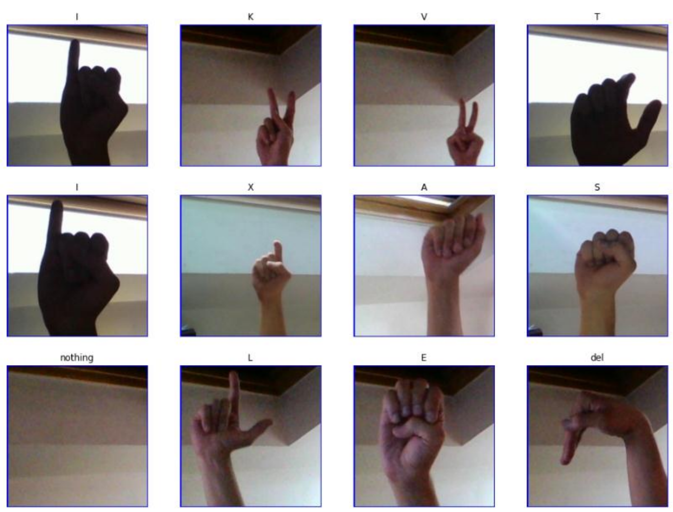

Fig -1: ASL dataset

3. RESULTS

Our training and validation accuracy for hand gesture classifier is as shown:

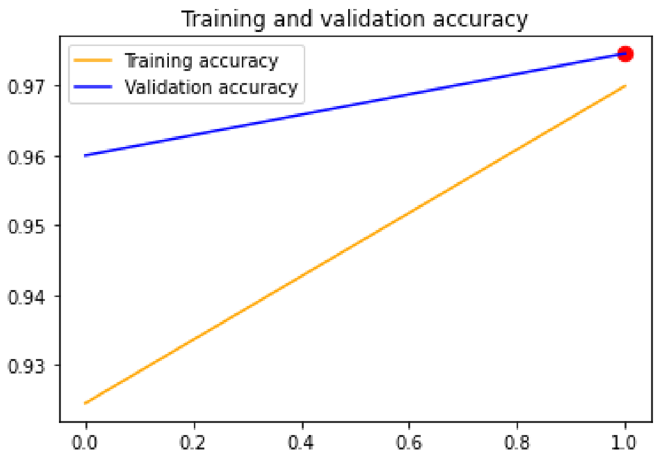

Fig -2: Accuracy graph for hand classifier

Similarly, training and validation loss is as depicted below:

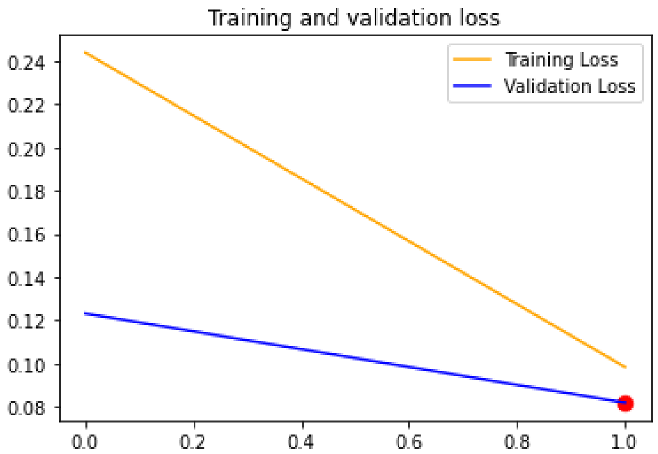

Fig -3: Loss graph for hand classifier

The system recognizes and classifies the hand gesture being made. The recognized sign is then shown at the top left of the application. The output of our implementation can be seen in the figure:

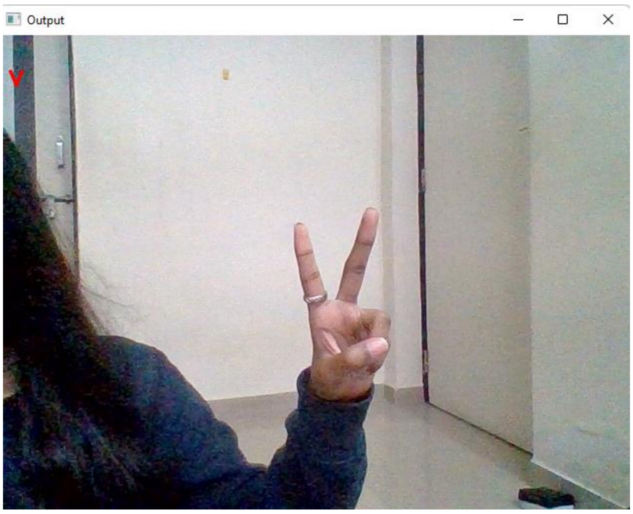

Fig -4: Implementation of hand recognition

For the face, the application recognizes the expression being made and displays it on the screen above the face. The implementation of our application is as shown below:

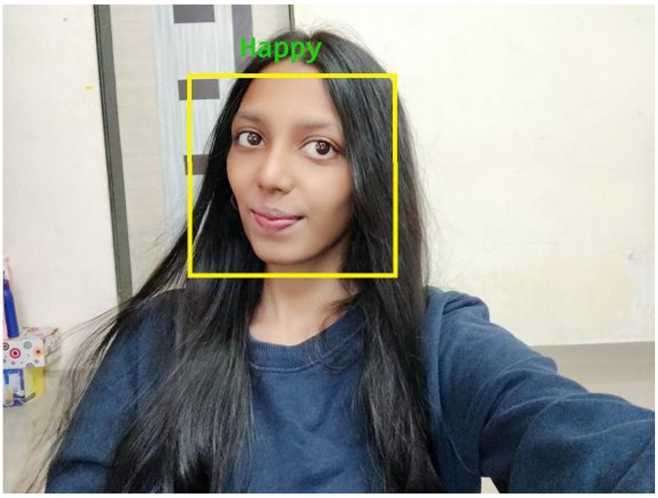

Fig -5: Implementation of face recognition

4. CONCLUSION

We implemented and trained an American Sign Language translator using transfer learning based on a Convolutional Neural Network (CNN). Besides that, we constructed a CNN-based facial expression recognition model. For both objectives, we were able to create a robust model.

REFERENCES

[1] Analysis of the impact of technological innovation on healthcare services- José Figueiredo, Vasco Eiriz

[2] Facial Expression Recognition - Chandra Bhushan Singh , Babu Sarkar , Pushpendra Yadav

[3] Facial emotion recognition using convolutional neural networks (FERC) - Ninad Mehendale.

[4] Large-scale Video Classification with Convolutional Neural Networks - Andrej Karpathy, George Toderici, Sanketh Shetty.