Hand Gesture classification using Convolutional Neural Networks

RAJAVELU E
Signal Processing and Machine Learning
Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
Guwahati, India
e.rajavelu@iitg.ac.in

***Abstract*—Hand gestures can be a useful medium of human-computer interaction and can make the interaction easier. The goal of this project is to train a Machine Learning algorithm capable of classifying images of different hand gestures, such as a fist, palm, showing the thumb, and others. Since, CNN can learn complex and non-linear relationships among images, hand gesture classification using CNN was used in this project.**

I. Introduction

Hand gesture classification is the task of identifying and classifying different hand gestures. It is a challenging task due to the large variability in hand shapes, poses, and movements. However, recent advances in deep learning, particularly CNNs, have led to significant improvements in hand gesture classification accuracy. CNNs are a type of deep learning model that are well-suited for image classification tasks. CNNs learn to extract features from images that are relevant for classification. The features are then used to train a classifier to identify the different classes of images.

II. Methodology

A. Data Collection

This project uses the Hand Gesture Recognition Database available on Kaggle. It contains 20000 images with different hands and hand gestures. There is a total of 10 hand gestures of 10 different people presented in the data set. There are 5 female subjects and 5 male subjects.

B. Data preparation

The dataset is preprocessed using the following steps:

1) Resize all images to 50 x 50. This is done to ensure that all images have the same size, which makes it easier for the CNN model to learn the features from the images.
2) Shuffle the whole dataset to prevent bias of any sort. This is done to ensure that the training set and test set are representative of the entire dataset.
3) Reshape the images to a shape of 50x50x1. The CNN model expects the input images to have three dimensions: height, width, and channels.

Once the dataset is preprocessed, it is split into a training set and a test set using the train-test split technique. The training set is used to train the CNN model, and the test set is used to evaluate the performance of the model on unseen data.

C. Creating model

The CNN model used in this project consists of the following layers:

1) Input layer: The input layer takes the input image as input.
2) Convolutional layers: The convolutional layers extract features from the input image.
3) Pooling layers: The pooling layers reduce the dimensionality of the feature maps.
4) Fully connected layers: The fully connected layers classify the feature maps into different classes.

D. Training the model

Training of the Dense Neural Network section of the CNN based model uses the principle of Backpropagation which updates the weights in each iteration reducing the error using the stochastic gradient descent algorithm internally. Training internally includes finding all the first order derivatives of the cost function (error) w.r.t weights and then finding appropriate weights minimizing error ultimately.

E. Evaluate the model

Using some combinations of metrics, find the objective performance of the model. Observe and ensure the error and accuracy percentage improving iteration by iteration while training the data.

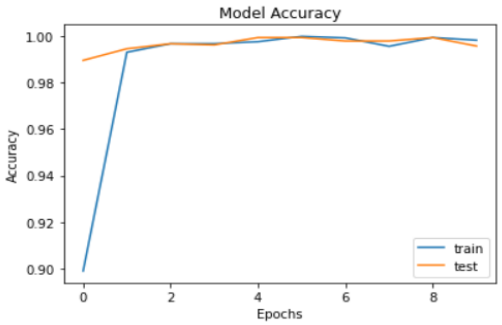

Fig. 1. Plot of Accuracy v No. of Epochs

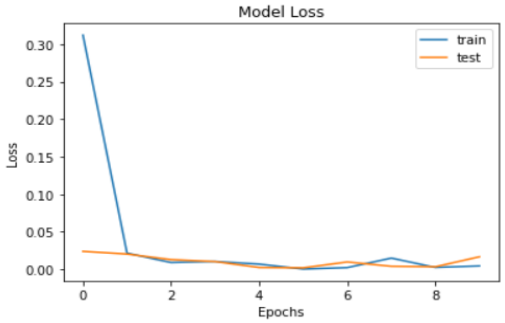

Fig. 2. Plot of Loss v No. of Epochs

F. Testing of model

Test the model by sending the testdata to the model and check the predicted output for sample testcase in testdata and print the sample output. And the accuracy of model is found by comparing each predicted output with corresponding actual output.

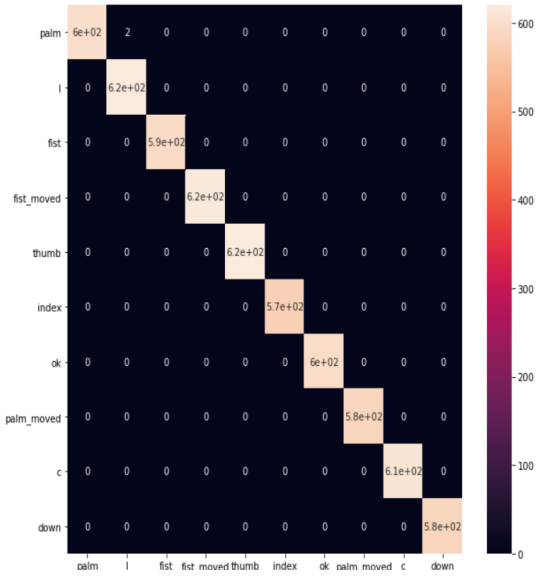

Fig. 3. Confusion Matrix

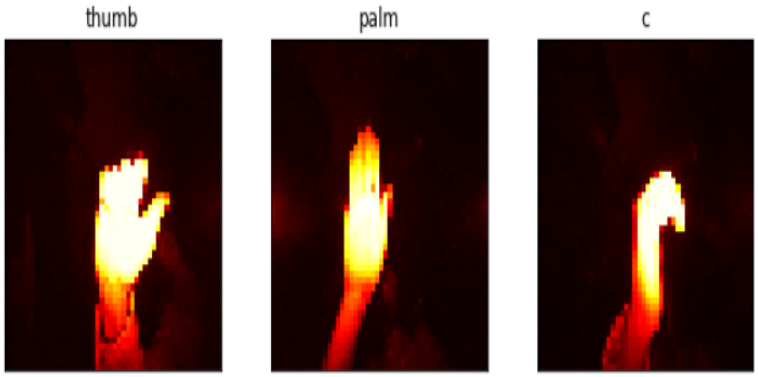

Fig. 4. Prediction done by the model

III. Conclusion

This project presented a hand gesture classification system using a CNN. The system was trained on the Leap Motion sensor hand gestures dataset and achieved an accuracy of 99.58% on the test set. This shows that the system is able to effectively classify hand gestures even in the presence of variations in hand shapes, poses, and movements.