IMPLEMENTATION OF CNN MODEL FOR GESTURE RECOGNITION BASED ON TENSORFLOW FRAMEWORK

Dr.T. Venkat Narayana Rao,
Professor, Department of C.S.E,
Sreenidhi Institute of Science and Technology,
Yamnampet, Hyderabad, T.S, India.

Kapa Vivek , Pabbaraju Surendra & Nagula Praveen
Student(s), Department of C.S.E,
Sreenidhi Institute of Science and Technology,
Yamnampet, Hyderabad, T.S, India.

Abstract

It application utilizes the open-source Tensorflow framework to construct a gesture identification model, the properties of the Tensorflow architecture, and presents a CNN model which is dependent on the Tensorflow platform. This study is designed to combine an established dataset with soul-collected data. The results of the experiment show that perhaps the prototype provides high precision recognition, good computational performance, high solidity, and thus can easily adjust the node framework, quickly determine the best configuration, and further perform the gesture recognition task. The system is categorized into 2 components: training and recognition. The structure of the node, as well as the operation of the two parts, are essentially the same. The main distinction is the original distribution of the variable weight. Within these, the preliminary values of the learning portion of the system and the values for each step of the convolution framework of the network are modified by the learning variables in order to minimize the specific output errors. The recognition component values directly use the system values acquired from the test, as well as the test collections, get projected during each stage of the node and the result is just the detection product. This paper focuses on how CNN model could be useful in building a system that could predict and classify the gestures.

Keywords: Convolution Neural Networks(CNN), tensorflow framework, deep learning and gesture.

I.INTRODUCTION

Data mining collects knowledge and observations from a vast volume of data. Data mining is a key step in the exploration of information through repositories. There are indeed a variety of repositories, info marts, and datacenters around the globe. Data Mining is more about collecting key information from a large number of databases. Data mining is often referred to as the Information Discovery Database. It has four main techniques, namely grouping, clustering,

regression, and collaboration. Machine learning strategies have the ability to effectively mine large quantities of data.

Data mining is primarily required in several fields to obtain valuable knowledge from a volume of data. Areas like the restorative sector, the commercial sector, and the insightful field have an endless sum of information, in this way these areas can be extracted through these techniques with valuable data. The classification strategy is used to categorize the complete set of data into two classifications expressly true and also no. The classify method is related to the data using the machine intelligence distinction calculation to be a basic decision tree and Naïve Bayes grouping models. All such models are used to improve the accuracy of the categorization. This demonstrates the success of both identification and expectation strategies.

1.1. ConvolutionNeural Networks(CNN)

A CNN / ConvNet is a Deep-Learning algorithm that can construct an input image, assign significance (learnable weights)to various aspects / items of the image and differentiate between one element and another.In contrast to other classification techniques, pre-processing in a ConvNet is now considerably easier. While normal-method filters are handmade , ConvNets can learn about these filters / features with adequate experience.The design of a ConvNet is similar to that of the communication system of Neurons in the Human Brain and was influenced by the structure of the Visual CortexIndividual neurons respond to a stimulus only in a small region of the visual field known as the receptive field. To complete the whole field of view, a selection of these fields overlaps. It is a modified type of neural network model optimized for dealing with 2-D image data, but it can also be used for 1-D and 3-D images. Convolution is a linear operation with many input weights on average, just like a traditional neural network. Since the process for 2-D input has been established, the dot product between the input data array and the 2-D Weight Set, known as a filter or kernel.The filter is smaller than the original data and also the kind of multiplication performed between both input filter-sized patch and a dot product is the filter[1][2].

1.2. Tensorflow Framework

The name of Tensorflow is taken directly from its central framework: the tensor. All of the simulations at Tensorflow include tensors. A tensor is an n-dimensional vector or matrix which represents all types of information. All attributes in a tensor retain a known (or partially known) shape of the same data type. The data form is a matrix or array dimensionality[3].tensor may derive from the input information or a measurement outcome. In TensorFlow, all operations within a graph are performed. The graph is a computational set that happens successively. Every operation is referred to as an op node and is connected. The graph shows the internode operations and relations. It does not show the values though. The node edge is the tensor, i.e., a way to inject data into the process.

TensorFlow uses a matrix structure. The graph collects and describes all computations of the series made during the training. The graph has many benefits to it:

1. Multiple CPUs or GPUs and on even mobile operating systems can be used.

2. The graph's portability allows the computations to be kept for immediate or later use. The graph can be stored for future implementation. All the simulations in the graph are connected together by tensors. A tensor has an edge and a node on it. The node holds the math operation and generates an array of endpoints. The edges of the edges explain the relationships between input / output nodes. TensorFlow is the strongest library of all, as it is structured to be open to all[3].

3. Tensorflow library integrates different APIs to be designed to deep learning architecture such as CNN or RNN on a scale.

II. ARCHITECTURE AND FLOW OF PROCESS

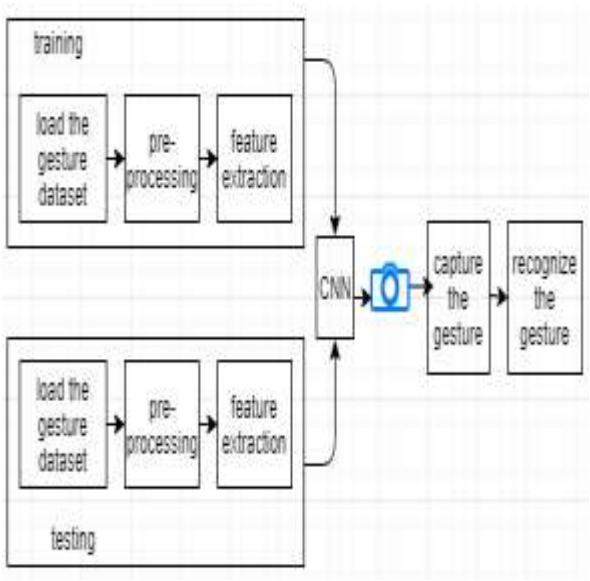

Fig.1. Flow Diagram

The above Fig.1. shows the flow of the system. The data collection is immediately loaded onto the training network. Preprocessing is performed before removing the item. Training takes place in the Convolutionary Neural Network. After training a webcam capture input image is provided. The picture given is checked to identify the gesture. A confusion matrix is generated with its mean precision as per the produced output. CNN is a common algorithm in machine learning. This is one of the Deep learning strategies and a learning algorithm used to conduct classification tasks through pictures, video, text, and sound. CNN's actually provide better results to recognize patterns in an image, leading to the identification of hand movements, faces, and any entity.

CNN 's benefit is that it doesn't need any selection of features to train the algorithm. The scaling and rotation are invariant of CNN.

III . Network Architecture

1. Image input layer: An imageInputLayer is the position where you set up the image dimension, where it uses 128-by-128-by-1. These numbers reflect height, distance, and channel count.

2. Convolution Layer: In this layer input arguments are filter height, no. of filters, and padding. If the size 10 filter is used here which specifies 10 x 10 filter. For instance if the no. of channels used is 10, it indicates that 10 neurons are connected. The padding of 1 states that its output image size is equivalent to that of an input data.

3. ReLu Layer: ReLU is a batch standardization layer that is positioned after a non lineary activation feature has been initialized. ReLU is an activation layer. Significance of this layer is to reduce sensitivity and increase the training speed.

4. Max Pooling Layer: One of the down-sampling techniques that is used for convolutionary layers is Max pooling layer. PoolSize is set to 2 in this architecture and the step size of the training function is 2.

5. Fully Connected Layer: Completely linked layers adopt layer of max pooling. The neurons on both layers are linked to the preceding cell in this system. For this layer for instance if the input argument is 10 which indicates 10 classes.

6. Softmax layer: Fully connected layers are accompanied by softmax layer which is the technique of normalization. This layer produces positive numbers as output, like the sum of those would be one. Such numbers are used for classification by classification layer.

7. Classification layer: The final layer of the architecture is classification layer. This layer classifies the classes based on softmax layer probabilities and also calculates cost function[3][2].

IV. PROPOSED SYTEM

Two elements of the model are preparation and recognition. The network architecture and the function of both parts is virtually identical. Initial allocation of the variable weight is the only difference. The preliminary weights of the learning component of the node are altered by the samples to eliminate real output errors, along with weights for each stage of the CNN structure.

The component recognition units use the device weights directly from the testing, the test sets are directed through every surface of the frame and the result is the detection product.Due to the extensive availability of digital camera systems, many researchers are researching gesture detection applications. Still due to the complexity of recognition of gestures, there are many challenges. This issue is then solved by deep learning using the Convolutionary Neural Network. Deep learning is more effective when it comes to recognizing pictures.The ASL dataset that contains the hand gestures (0-9) is used here. In general, an image is pre-processed which itself plays a crucial role in retrieving the gesture from a static image (i.e. background subtraction, binarization of images). After binarization, the function is derived from all images. Neurons with learning weights and biases are the convergent neural network. Each neuron receives multiple inputs and takes over a weighted amount. The activation function is then passed and the output responds to it.

4.1. **User**

The user need to download or make the dataset containing the images of different gestures each of them in separate folders with large number of samples.

Step 1: Firstly , the dataset is imported in to the particular folder

Step 2: The path of that folder is provided as the input by the user in order to train the model using the training data

Step 3: Later on to test the model's exactness the user provides the path of an image which is considered as test data

Step 4: Then, trained model would predict it and gives the output of the concerned action.

Step 5: If not it would change the weights and repeat step 4.

4.2. **CNN METHOD**

Step 1: First ,the dataset is imported to a specific location and stored in x_data ,y_data variables,location is stored in path variable.
Step 2: Once it is trained we will give our own inputs to test against training data to get desired outputs .x_train ,y_train are used to storetrained data.
Step 3: Weights are randomly initialized in the network.
Step 4: Do it for each value in training samples.Do Neural–net-output(model,e)forward pass
Step 5: Calculate the error at each outputs
Step 6: Calculate the validation_data for each.Hidden Layer to output layer weights.This is backward pass.
Step 7: Calculate the validation_data for each Hidden Layer to output layer weights(backward pass continues)
Step 8: Update all the weights in the end of the network.

Step 9: Until all the inputs are correctly classified or exit condition satisfied.
Step 10: return model

4.3. **Sequential model**

The stpes involved in this model are:
Step 1:Define the model
Step2: Compile the model
Step3: Fit the model
Step4: Assess the pattern
Step5: Draw Conclusions

V. Working of Models

In order to built the model, Firstly we need the large dataset consisting of images of each **g**esture and stored in separate directories. These directories must be saved using the numbers 00 to 09(ASL dataset) depending up on the number of distinct gestures There by in each directory large number of images of that respective gesture must be stored using the serial numbers starting from 00 in png format. We need to decide the ratio of training data and test data where training data is used for training ,building the model and test data is used is used to determine the accuracy of the model. Then we will initialize the image size as 150 in the image input layer[4].The input arguments for the layer is 64 filters ,kernel size=(5,5) and the activation is 'relu' which helps in lowering down the sensitivity and pace of training is increased. The max pool size is set to 2.In order to test we need to provide the path of the image which has gesture as an input to the model. The normalization is done with the help of softmax layer that generates a set of positive numbers as output whose sum is one and this is used by the classification layer. This classify the classes based on the probabilities obtained from softmax layer. Based on this which of the position in the matrix had got the result as one that corresponding action is the result of the given input.

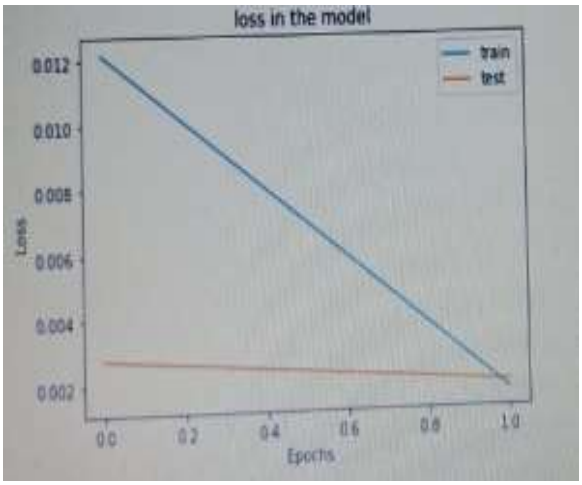

Fig.2. Loss in the model

The above Fig.2. shows the loss in the model for both the train and test data.Intilally at the beginning of the training the loss is quite high but as the epoch are being increased the loss in the model is gradually decreased.On the other hand the loss in the training data is significantly low .

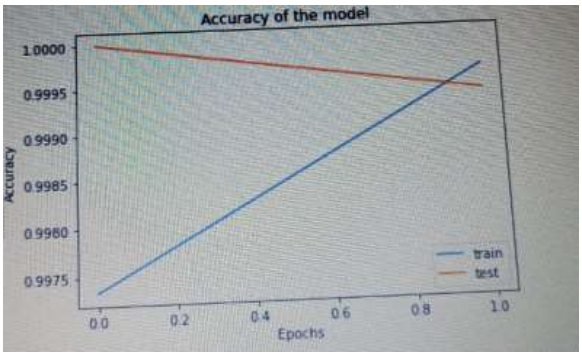

Fig.3. Accuracy of the Model

The above Fig.3. shows the accuracy of the model for both the test and train data.During testing the accuracy increases gradually with respect to the epoch size.While for the test data the accuracy is significantly high.

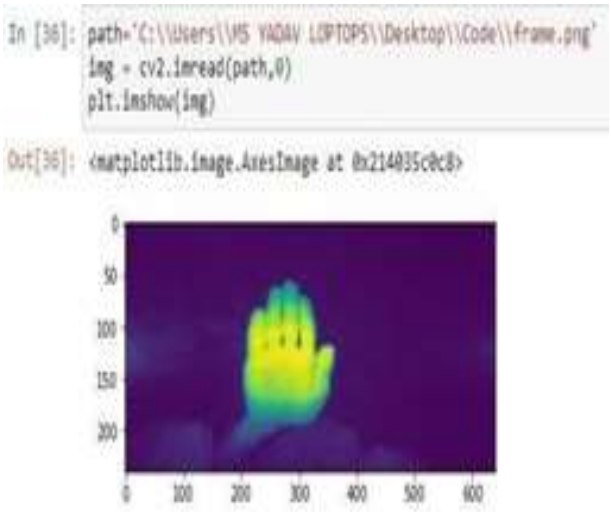

Fig.4. Input to the model

Fig.5. Output of the model

The image as shown in the above Fig.4. is provided as an input to the model. Then it undergoes different series of steps as mentioned in the working. Thereby, finally we get the resultant value as one for one respective position in the matrix that will be mapped with the corresponding action .In this case the model generated the numerical value one at result[0][0] position. As the result[0][0] position is mapped with the action "stop" as shown in the Fig.5. .Hence, the model had returned the output as "Stop"[5].

VI.APPLICATIONS

Numerous ideas for applications of these have come up in the past few years Although many of them are still at developing stage, they prove to be a promising technology in future that will make many of things much simpler to use. In future gesture recognition can eliminate the necessity to have a physical contact with input devices. Few of them are:

1. New intervention in Gaming, It can be used in gaming consoles like Xbox, etc.
2. Automated homes where all the gadgets could be controlled by the gestures.
3. It also helps the doctors to manipulate the images without having any physical contact with the keyboards or the monitors.
4. Provides aid to the physically disabled people to perform their tasks with the help of gestures.

VII. CONCLUSION

This system implements convolution neural networks that use hidden layers and assign initial weights and change them accordingly until the classification is achieved . The image that is given as input goes through different layers and then the classification layer confines the image to a particular category. The accuracy obtained using CNN is very much high as compared to other algorithms that are used for image detection or classification. Moreover, the accuracy of the model increases as the number of epochs are increased. For instance, in our case, the accuracy of epoch1 was 0.8871 and for epoch2 it was 0.9991. In future gesture recognition can eliminate the necessity to have physical contact with input devices like mouse and keyboard.

REFERENCES

[1]. Understanding of a convolutional neural network by Saad Albawi ; Tareq Abed Mohammed ; Saad Al-Zawi IEEE,2017.

[2]. Convolutional neural networks for image classification by Nadia Jmour ; Sehla Zayen ; Afef Abdelkrim IEEE ,2018.

[3] Data classification with deep learning using Tensorflow by Fatih Ertam ; Galip Aydın IEEE,2017.

[4] Hand gesture recognition based on shape parameters by Meenakshi Panwar IEEE,2012.

[5] Hand gesture recognition using deep learning by Soeb Hussain ; Rupal Saxena ; Xie Han ; Jameel Ahmed Khan ; Hyunchul Shin IEEE,2017.