International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 587-589 www.ijaem.net ISSN: 2395-5252

Hand gesture recognition using TensorFlow and Machine Learning

Dr. S. P. Maniraj, Ayush Gupta, Mullamuri Venkata Yaswanth, Prateek Soumya

Department of Computer Science and Engineering
SRM University, Ramapuram
Tamil Nadu, India

Submitted: 15-11-2022 Accepted: 25-11-2022

ABSTRACT - This report is an insinuation of application of hand gesture recognition with words to sentence conversion for the deaf or mute individuals. For example, a certain individual will look into the camera and make the signs that are to be recognised and the model will recognise them and correctly identify the words that are meant to be inferred. From those words that are formed by the individual's actions in the computer visions are then transformed into sentence that most likely were inferred by the individual. One of the primary objectives of gesture recognition is to create systems, which can identify specific gestures and use them to convey information or to control a device. Though, gestures need to be modelled in the spatial and temporal domains, where a hand postures the static structure of the hand and a gesture is the dynamic movement of the hand.

Index Terms – LSTM; RNN; NLP; TENSORFLOW

I. INTRODUCTION

Image detection or object detection has found its use sprawling use ranging from face-mask in the covid-19 pandemic to anomaly detection in the medical field. A hand gesture recognition may provide an entirely new way of communication through the eye of the computer. In recent years, hand gesture recognition or abbreviated as HGR has found in a number of applications as discussed before and they have been exclusively developed with the technologies of human-computer interaction or abbreviated as HCI.

Recurrent Neural Networks (RNNs) have achieved excellent results in speech recognition and have promoted the application in the field of gesture recognition. Recently, RNNs can model the long terms contextual information of temporal sequences, and have been successfully applied to processing sequential data with variable length such as language modelling and video analysis. At the same time, the method based on 3D skeleton data has the advantages of simplicity, view-independent and good robustness. Several methods based on RNNs have been proposed for 3D skeleton data.

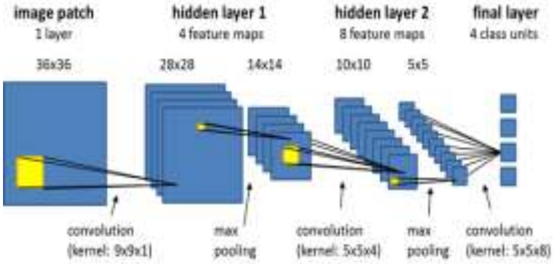
American Sign Language (ASL) is one of the primary utility for people with deaf and mute disability. Communication is one of the most basic needs for an individual. It is indeed a process of expressing an individual's thoughts and expressions.

II. DEEP LEARNING

Different levels of abstraction can be learned by adding a different layer or hidden layer which can be facilitated by Deep Learning methods. Deep Learning allows computers to learn complicated concepts by breaking them down into simpler concepts and standardizing the information obtained in each process in order to obtain a much broader comprehension. In the case of artificial neural networks, deep learning or hierarchical learning is about assigning credits in many computational steps in a precise way, to transform the aggregate activation of the network. That is, non-linear operations are used under different architectures in order to allow neural network to learn very complex functions that better represent the desired characteristics. Various kinds of architecture for artificial neural networks(ANN) such as Convolutional Neural Network (CNN), Feedforward Neural Network (FNN), Recurrent Neural Network (RNN) and many such others were created which gave scope to wide range of applications.

A. Recurrent Neural Network(RNN)

We developed a special type of Recurrent Neural Network (RNN) based deep learning approach called Long Short Term Memory (LSTM) to detect gestures made through hand. Sequential data is processed using RNN which is one of the part of neural networks family. Similar to Convolutional Network that is a neural network which is specialized in processing a rack of X values such as an image, an RNN is a neural network that is specialized to process a succession of values x1, x2, ...,xn. Similar to Convolution networks which are capable to scale and process the images with a very large height and width, RNN can also be scaled to successions much longer than that would be practical for unsecured networks. Even RNN can also process variable length sequences. However, there is a mathematical challenge to establish the long-term dependencies in recurrent networks. It is a problem that the gradients in RNN tend to discharge and it is a problem of vanishing gradient. The prime phase is the one that usually offers more even if it is assumed that the parameters included in this layer of RNN(Recurrent Neural Network) are stable (can store memories, with gradients that do not exploit), the difficulties with similar long-term dependencies persist because of petite or minor weights given by the long term interface, which involve the reduplication of various Jacobs compared with the Short-term. To deal with this issue, we have created new design focusing on the primary idea of the recurrent network. Unlike regular Neural Networks, in the layers of RNN, the neurons are arranged in 3 dimensions: width, height, depth. The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner.



1. Recurrent Layer

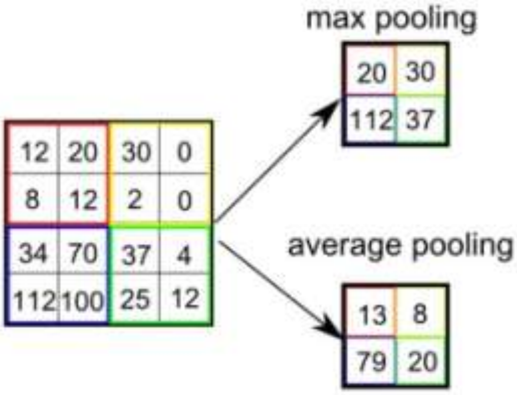
In convolution layer I have taken a small window size [typically of length 5*5] that expands to the depth of the input matrix A tiny window dimension has been taken in complexity layer 1. The layer consists of attainable filters of window size. During every repetition I slid the window by stride size [typically 1], and enumerated the dot product of filter entries and input values at a given location. As I resume this process we will create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position.

2. Pooling Layer

We use a pooling layer to reduce the size of the activation matrix and ultimately reduce the learnable variables. There are two types of pooling:

a. Max Pooling: In max pooling we take a window size [for example window of size 2*2], and only taken the maximum of 4 values.

b. Average Pooling: In average pooling we take average of all Values in a window.



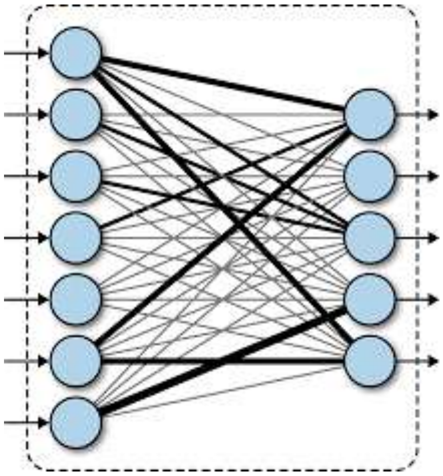
12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

20	30
112	37

13	8
79	20

c. Fully Connected Layer

In convolution layer neurons are connected only to a local region, while in a fully connected region, well connect the all the inputs to neurons.



4. Final Output Layer

After extracting values form the previous fully connected layer, we will converge them with the last level of neurons which will play the prescient in predicting each image of various class.

B. Algorithm

Full-connected neural networks (Full-connected NN) is the simplest structure of neural networks of deep learning method. There are connections between different layers but no correlation of nodes in the same layer in Full connected NN. The timing information of time series signals cannot be accurately expressed. Recurrent neural networks (RNN) with cycle of iterative function stores timing related information effectively. The connection of hidden layers in RNN is iteration cycle, which has weight connection between the current state and last moment state. According to the unique design of RNN, information of recent input events can be stored by the feedback connection, which is called short term memory. Long-short term memory (LSTM) is used widely to deal with the loss of long term memory in the process of training sequence signal. Based on the above analysis, LSTM is proposed to optimize the networks structure.

Training:

We have converted the input image which is obtained from the computer vision module called "OpenCV" into Gaussian blur which removes the unnecessary disturbance. We have also applied adaptive threshold to extract hand from the image imported by the computer vision. The image is then after fed to model for training and testing. The above mentioned prediction layer estimates the category of classes. Hence, the output will be normalized into 0 and 1. SoftMax function was used to achieve this feature.

III. FEASIBILITY STUDY

To analyse the strengths and weaknesses of the proposed system, a feasibility check is carried out.. The application of usage of a live translator. The feasibility study is carried out in three forms.

A. Economic Feasibility

The proposed system does not require any high cost equipment. This project can be developed within the available software.

B. Technical Feasibility

The proposed system is completely a Machine learning model. The main tools used in this project are Anaconda prompt, Visual studio, Kaggledata sets, Jupyter Notebook And the language used to execute the process in Python. The tools mentioned in this paper are open source and competitive technical proficiency is required to use them. From this we can conclude that the project is technically feasible.

C. Social Feasibility

Social feasibility is a determination of whether project will be acceptable or not. Our project is Eco-friendly for society and there is no social issues. Our project must not threatened by the system instead must accept it as a necessity. Since our project is applicable for every individuals in the society to take care about the society and environment. The level of the acceptance of System is very high and it depends on the methods deployed in the system. Our system is highly familiar with the society.

REFERENCES

[1] George Sung, Kanstantsin Sokal, Esha Uboweja, "On-device Real-time Hand Gesture Recognition", 2021.

[2] A.Mohanarathinam, K.G Dharani, R. Sangeetha, "Study on Hand Gesture Recognition by using Machine Learning", 2020.

[3] Naoto Ageishi, Fukuchi Tomohide, Abderazek, Ben Abdallah, "Real-time Hand-Gesture Recognition based on Deep Neural Network", 2021.

[4] Enea Ceolini, Sumit Bam Shrestha, Elisa Donati, "Hand-Gesture Recognition Based on EMG and Event-Based Camera Sensor Fusion: A Benchmark in Neuromorphic Computing", 2020.

[5] Zhang Chen, Kim, Johee, "Video Object Detection With Two-Path Convolutional LSTM Pyramid", 2020.

[6] K. Cheng, Z. Yang, Q. Chen and Y.-W. Tai, "Fully Convolutional Networks for Continuous Sign Language Recognition", pp. 697-714, 2020.

[7] A. Krizhevsky, I. Sutskever and G. Hinton, "Imagenet classification with deep convolutional neural networks", Neural Information Processing Systems, vol. 25, 2012.

[8] S.-J. Ryu, J.-S. Suh, S.-H. Baek, S. Hong and J.-H. Kim, "Feature-based hand gesture recognition using an FMCW radar and its temporal feature analysis", IEEE Sensors J., vol. 18, no. 18, pp. 7593-7602, Sep. 2018.

[9] W. J. Zhang, J. C. Wang and F. P. Lan, "Dynamic hand gesture recognition based on short-term sampling neural networks", IEEE/CAA J. Autom. Sinica, vol. 8, no. 1, pp. 110-120, Jan. 2021.

[10]S. Tripathi, K. Nguyen, T. Guha, B. Du and T. Q. Nguyen, "Sg2caps: Revisiting scene graphs for image captioning", arXiv preprint, 2021.