International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1971-1974 www.ijaem.net ISSN: 2395-5252

Real Time Hand Gesture Detection Using Neural Networks

Ashu Berwal, Rahul
Student ,M.Tech , Dept. of Computer Science , Shri Baba Mastnath College of Engineering & Technology Rohtak , Haryana , India
Asst. Prof. , Dept. of Computer Science , Shri Baba Mastnath College of Engineering & Technology Rohtak , Haryana , India

Submitted: 01-07-2021 Revised: 13-07-2021 Accepted: 16-07-2021

ABSTRACT—Real-time recognition of dynamic hand gestures from video streams may be a challenging task since (i) there's no indication when a gesture starts and ends within the video, (ii) performed gestures should only be recognized once, and (iii) the whole architecture should be designed considering the memory and power budget. In this work, we address these challenges by proposing a hierarchical data structure enabling offline-working convolutional neural network (CNN) architectures to work online efficiently by using sliding windowapproach. The proposed architecture consists of two models: (1) A detector which may be a lightweight CNN architecture to detect gestures and (2) a classifier which may be a deep CNN to classify the detected gestures. In order to guage the single-time activations of the detected gestures, we propose to use Levenshtein distance as an evaluation metric since it can measure misclassifications, multiple detections, and missing detections at the same time.

I. INTRODUCTION

A Neural Network refers to a computing system made up of simpleand highly interconnected processing elements operating side by side. The function of neural network is determined by the connections between these elements. We can train our neural network to perform a particular function by adjusting the values of the connections between elements.

Neural networks are adjusted, or trained majority of times, so that a particular input to a dynamic system leads to a particular target output. In today's world of Artificial intelligence Neural networks have a great application in different fields including pattern recognition, identification, classification, speech, vision and control systems. Neural networks can be trained to solve problems that are difficult for traditional computers and human beings.

The test input is given to the input for testing, the input is compared with the trained set of inputs for comparison and the most matching set is returned as output. The field of Neural Networks came into existence some fifty years ago but found solid application only in the past fifteen years and is still developing as the potential of Artificial intelligence is very huge, it is very different from the field of simple programming where we apply simple logic including basic mathematics and design procedures to execute a bunch of code to get desired result.

Neural Network is a sub-part of Artificial intelligence, there are two modes of learning: Supervised and Unsupervised

Supervised Learning

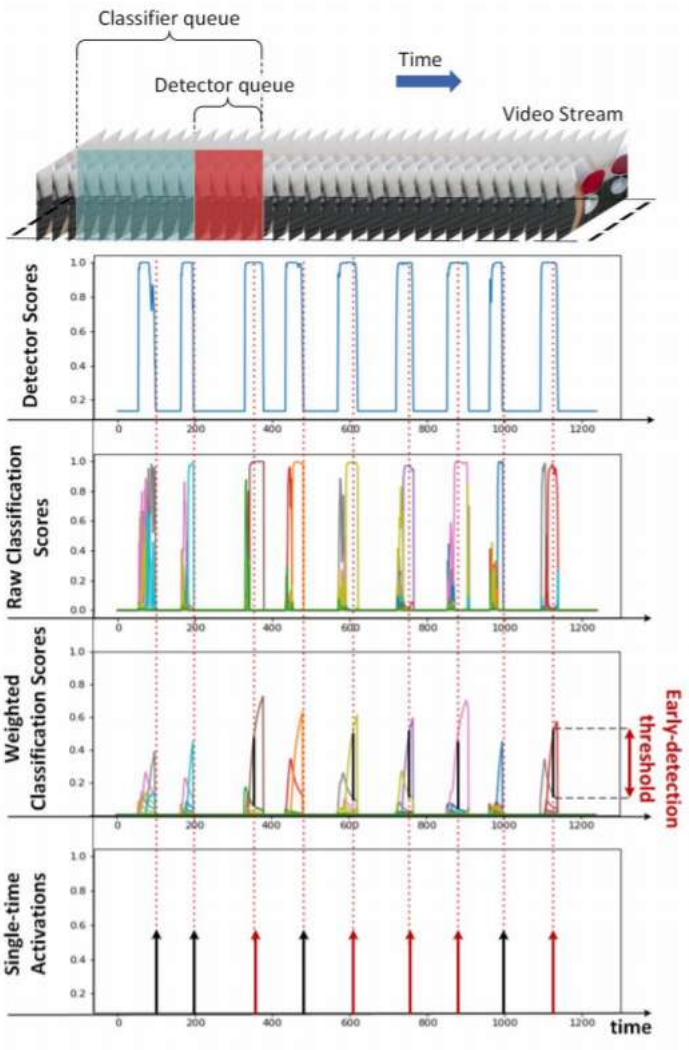
When we provide solutions for each input given to the system, such solutions are called target or output. So, if we are training our system for each input with its matching target it is called Supervised Learning. After sufficient training our system will point to the specific target for each input provided. Due to the supervised learning accuracy level will be high compared to the unsupervised learning. For example we get a a number of outputs for a single input.So we prefer supervised learning over unsupervised learning.

Unsupervised Learning

Contrary to the supervised Learning the system is trained for a limited set of inputs, as soon as any matching input is given to the system the output is displayed accordingly.Due to the limited set of inputs accuracy level will be not high compared to the supervised learning.It is a type of learning in which models are trained using unlabeled datasheet and are allowed to act on that data without any supervision. Unsupervised Learning algorithm will perform this task by

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1971-1974 www.ijaem.net ISSN: 2395-5252

clustering the image dataset into the groups according to similarities between the images. Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs .



Advantages of Neural Computing

There are many advantages of neural networking which a programmer or user realizes in their work.

Gesture recognition is a powerful tool for harnessing the information in the data. Neural net learns from the existing dataset.

Neural Networking is quite different from simple programming as the system is developed from learning rather than simple codes of a program.

Neural Networks are very brilliant at adapting to constantly changing environment, the conventional programming system is designed for a particular context and rules and as soon as the environment changes they are not valid. So, neural networks are very adaptive to the changing environment.

Performance of neural networks is quite good as that of traditional modelling. Neural Networks build data which is more attractive and understandable in less time.

Neural networks now operate well on modest computers and hardware requirements are quite acceptable. Neural networks require intensive computation, but now the routines are optimized which could be run on personal computers.

Neural networks can be used to build computing models which are not possible through conventional approaches. As neural networks can model data which is very complex, it is easier to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1971-1974 www.ijaem.net ISSN: 2395-5252

use neural nets in place of other conventional approaches.

Neural network can hold large amount of data which is not possible for a simple network.

Limitations of Neural Computing

There are some limitations to use of neural networks which are necessary to completely understand neural networks.

It is difficult to explain the model it has built in a useful way. This is sometimes important at some instances when we need to explain the processes to some analyst or people involved in Artificial Intelligence. To explain all things to the analyst and the people involved it require large amount of time and its wastage which is not beneficial for the system.

As with most of the conventional systems, we cannot just throw data into the neural network, it requires proper input which would be used to compare with trained data and will throw appropriate result. So, improper data input affects the output of the system. Good results are produced for proper data input.

II. RELATED WORK

We have plethora approaches by the help of CNNs and video data can extract spatio temporal information. Static images , video analysis these applied 2D CNNs because of success of 2D CNNs . In video frames are treated as multi-channel inputs to 2D CNNs. Temporal Segment Network (TSN) divides video into several segments, extracts information from color and optical flow modalities for every segment using 2D CNNs, then applies spatio-temporal modeling for action recognition. A convolutional long STM (LSTM) architecture is proposed in , where the authors extract first the features from video frames by a 2D CNN then apply LSTM for global temporal modeling. The strength of of these approaches comes from the very fact that there are many very successful 2D CNN architectures, and these architectures are often pretrained using the very large-scale ImageNet dataset. Although 2D CNNs perform pretty much on video analysis tasks, they're limited to model temporal information and motion patterns

The real-time systems for hand gesture recognition requires to use detection and classification simultaneously on continuous stream of video. Plenty of works addressing classification and detection individually. In authors apply histogram of oriented gradient (HOG) algorithm along side an SVM classifier. Author now having command on unique radar system to segment gestures and detect. In our work, we've trained a light-weight weight 3D CNN for gesture detection. Moreover, in human computer interfaces, performed gestures must be recognized just one occasion (i.e.singletimeactivations) by the computers. That is an extremely evaluative and disparaging issue haven't been addressed good yet. Connectionless Temporal Classification(CFC) single time activations doesn't work but authors apply here to detect consecutive similar gestures. To the best of our knowledge, in this study, it is the first time singletime activations are performed for deep learning based hand gesture recognition.

III. METHODOLOGY

In this section, we elaborate on our two-model hierarchical architecture that permits the-state-of-the-art CNN models to be utilized in real-time gesture recognition applications as efficiently as possible. Before training details are elaborated , narrate the architecture. Finally, we provides a detailed explanation for the used post processing strategies that allow us to possess single-time activation per gesture in real-time.

Detector: the aim of the detector is to differentiate between gesture and no gesture classes by running on a sequence of images, which detector queue masks. Its top and unique role is to act as a switch for the classifier model, meaning that if it identifies a gesture, then the classifier is activated and fed by the frames in the classifier queue.

Classifier: Since we don't have any limitation regarding the dimensions or complexity of the model, any architecture providing an honest classification performance are often selected as classifier. This leads us to use two recent 3D CNN architectures (C3D and ResNext-101) as our classifier model. However, it's important to notice that our architecture is independent of the model type.

Post-processing: In dynamic hand gestures, it's possible that the hand gets out of the camera view while performing gestures. Any error of the proposed architecture lessen the general performance even so the foregoing predictions of the detector are accurate. In order to form use of previous predictions, we add the raw softmax probabilities of the previous detector predictions into a queue (qk) with size k, and apply filtering on these raw values and acquire final detector decisions. With this approach, detector increases its confidence in deciding , and clears out most of the misclassifications in consecutive predictions. The size of the queue (k) is chosen as 4, which achieved the simplest results for stride s of 1 in our experiments.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 1971-1974 www.ijaem.net ISSN: 2395-5252

Single-time Activation: In real-time gesture recognition systems, it's extremely important to possess smaller response time and single-time activation for every gesture. Pavlovic et al. states that dynamic gestures have preparation, nucleus (peak or stroke) and retraction parts . Out of all parts, nucleus is that the most discriminative one, since we will decide which gesture is performed in nucleus part even before it ends

IV. CONCLUSION

This paper presents a completely unique two-model hierarchical architecture for real-time hand gesture recognition systems. The proposed architecture provides resource efficiency, early detections and single time activations, which are critical for real-time gesture recognition applications. We are addressing on two dynamic datasets of hand gesture, and fulfil exact outcome for twain of them. For real-time evaluation, we've proposed to use a replacement metric, Levenshtein accuracy, which we believe may be a suitable evaluation metric since it can measure misclassifications, multiple detections and missing detections at the same time. Moreover, we've applied weighted-averaging on the category probabilites over time, which improves the general performance and allows early detection of the gestures at an equivalent time.

By the help of highest two average class probabilities as a confidence measure difference we have obtained single time activation on every single gesture. However, we might wish to investigate more on the statistical hypothesis testing for the arrogance measure of the single-time activations as a future work.