Human Action recognition using Neural Network

ABSTRACT:

One of the most popular research areas in the field of computer vision,machine learning and pattern recognition is human action recognition and over the years a number of research have been conducted on this area and researchers are still pursuing knowledge of this area. Human action recognition captured significant attention among researchers due to its vast applicability such as surveillance, robotics, health care, video searching and human-computer interaction. In recent times, deep learning based methods have taken a significant leap, especially in image analysis, object recognition, and speech recognition. As a result, researchers all over the world are shifting towards deep learning from the old traditional feature-based approaches. Human action recognition is somewhat complicated and complex work due to different kinds of challenges like occlusions, viewpoint variation, execution rate, and camera motion etc. Considering all of these, in this paper, we proposed a method for human action recognition using deep neural networks. Deep neural networks such as CNN, RNN can have a significant impact on human action recognition and this is our aim to achieve.

INTRO:

In human-computer interaction, human action recognition has many applications such as surveillance, video indexing and retrieval. Recognizing human action from a video stream is quite a challenging task which has been the focal point of research for a long time. Whenever a human being does something, from those actions which are a combination of some body movement, varying patterns can be generated as spatio-temporal appearances that can be used as feature descriptors. But it’s not just generating patterns of movement of different body parts when it comes to human action recognition, rather it involves a description of a person’s intention, emotion and thoughts. So, it is also very important to analyze and understand human behavior. Still image classification, video data containing temporal information are important for action recognition. Moreover, video data includes natural data augmentation and video frame classification are also important for the task. There are several parts of action recognition such as video classification, resolution, segmentation etc and much work has been done on these topics

which will be discussed in the next section of this paper. In this paper, we propose a human action recognition system using neural networks.

Literature review:

3d Convolution architecture is used as it needs to handle video input. In this paper [1] they used The architecture has 10 layers which contain the input too. They also mentioned a parameter is 17,169. Use of online backpropagation with momentum algorithms adapted to weigh sharing is used. Furthermore, it trained on the KTH dataset. According to [1] they propose to learn to label to learn the entire sequence, this helps as an advantage for the temporal evolution of the features. This paper [1] also mentions Recurrent Neural Network (RNN) is on of the highly used machine learning techniques for temporal analysis of data. They used mirror and flipped version of training sample to increase the training parameters. LSTM is important as it boosts performance by 3%. The paper [1] further mentions that KTH! Is better than KTH2 with KTH1 (94.4%) and KTH2(92.2). Lastly, this paper [1] mention 3D ConvNet is directly connected with LSTM sequence classifier. It helps reduce the computation time. The flaw was when it comes to calculating retro propagated error. Using challenging dataset is still a work is still a work in progress.

In this paper [2] the proposed using pre trained Convolutional Neural Network (CNN) model as source architecture also using hybrid Support Vector Machine and K Nearest Neighbor (KNN) for extracting features. According to [2] KTH and UCF sport dataset were used for training the model. KTH achieves the highest accuracy of 98.2% and UCF sports 91.5%. The use if transfer Learning was used as it is very powerful and useful when the training data is not sufficient. Deep learning model needs lots of training time and computational resources which is saved thanks to transfer learning. The paper [2] mentions the problems arises when it comes IXMAS, UCF -50, UCF-101 dataset. Hybrid classifier has advantage over single classifier in boosting the accuracy of the classification system.

Human action recognition is a vast and very popular research topic in this modern era. In this paper [3] they review many types of state of the art deep learning model based techniques which are being implemented for human action recognition. Moreover, they used three different types of dataset which are Single viewpoint, Multiple viewpoint and RGB depth videos. For each dataset they implemented different techniques. For instance, single videpoint according to [3] CNN with RNN were implemented with a high performance score of 94/4%. On the other hand,

CNN is best known for learning spatiotemporal features. For the Multiple viewpoint dataset MOCAP was implemented it is basically CNN but it deals with motion capture information. Advantages of MOCAP is that it can acheive high classification accuracy. The primary descriptors for hierarchical statistical methods to detect or predict human activities include RGB datasets, skeletal structure, gestures, and body motion. According to the study, weighted hierarchical depth motion maps are used to detect human activities using three channel deep CNNs. Challenge arise of recognizing or tracking several people in a video has been addressed for years. Mainly other issues as mentioned in this paper [3] the dataset is very simple and performs simple tasks. For example, punch can be defined as a fight or greeting between friends, depending on strength, blow speed, and other atomic-level action, perhaps a smile. In order to recognize such actions, we need to analyze several networks from
different streams. For each stream, different networks can track
different objects such as face or short movements etc.

Challenges:

- Unlike the conventional artificial feature extraction methods, the deep learning strategies require a large quantity of data to train the network.
- Otherwise, it is simple to reason the network to overfit, and the learned features aren't representative, resulting in unsatisfactory action recognition results.
- Many datasets cannot offer sufficient data, resulting in inadequate training data.
- For the problem of insufficient training data, one viable approach is to enlarge the datasets and increase the quantity of data.
- Another viable approach is to apply a pre-training network through transfer learning to prevent overfitting.
- Data growth can increase the statistics successfully.
- Common data growth strategies have flip, rotation, scale transformation, random picking, color dithering, and Gaussian noise.

- The transfer learning normally makes use of a predecessors trained action recognition network as a feature extractor to extract features or immediately train new data with a predecessors trained action recognition method, and fine-tuning the network.

- Scalability of calculations, non-convex optimization, disentangling underlying factors of variation, and lack of appropriate representation learning measure.

- Moreover, there are a few different challenges inclusive of the need for a professional to design the proper network due to a couple of hyper parameters, modelling numerous elements of versions and the interactions among them, big quantity of local minima, overall performance of Stochastic Gradient Descent (SGD) in lots of layers, overfitting problem, and the want for a terrific deal of data.

Conclusion:

Deep learning techniques have these days been introduced withinside the video-based human action recognition research area. They have been extensively utilized in different areas, which includes speech recognition, language processing and recommendation systems etc. There are many benefits to hierarchical statistical approaches, which includes raw data input, self-learnt functions and a high-level or complicated action recognition, hence, deep learning techniques have acquired a great deal of interest. Based on those benefits, researchers should lay out a real-time, adaptive and high performing recognition system. However, those approaches additionally have numerous drawbacks, which includes the want to generate large datasets, the overall performance relies upon the dimensions of the network weights and hyper-parameter tuning is non-trivial etc

Reference

1.
https://liris.cnrs.fr/Documents/Liris-5228.pdf?fbclid=IwAR3tNeAYLPnZdzLBB4POsfSQJOPz016MR4xILa7wCY9-7VDj7uyYpdojsSA

2.
https://eprints.lancs.ac.uk/id/eprint/87182/1/Human_Action_Recognition_using_Transfer_Learning_with_Deep_Representations_Revised.pdf?fbclid=IwAR2fvylQu8CdnWxRMBejYcAso-uBrW-yAxuf7flU5H5mXIccj1OXDGiO_DA

3.
https://opus.lib.uts.edu.au/bitstream/10453/106067/4/bare_conf_NS%20(4).pdf?fbclid=IwAR2kqJpcnBwQFVpoEg2pXxIAvsmxZHYI3eMouP1F5QfjgmMAHB4ljpgyMoo

.