Video Classification with a CNN-RNN Architecture

Abstract:

Video classification task has gained a significant success in the recent years. Specifically, the topic has gained more attention after the emergence of deep learning models as a successful tool for automatically classifying videos. In recognition of the importance of the video classification task and to summarize the success of deep learning models for this task, this paper presents a very comprehensive and concise review on the topic.

There are a number of existing reviews and survey papers related to video classification in the scientific literature. However, the existing review papers are either outdated, and therefore, do not include the recent state-of-art works or they have some limitations. In order to provide an updated and concise review, this paper highlights the key findings based on the existing deep learning models. The key findings are also discussed in a way to provide future research directions. This review mainly focuses on the type of network architecture used, the evaluation criteria to measure the success, and the data sets used.

To make the review self-contained, the emergence of deep learning methods towards automatic video classification and the state-of-art deep learning methods are well explained and summarized. Moreover, a clear insight of the newly developed deep learning architectures and the traditional approaches is provided, and the critical challenges based on the benchmarks are highlighted for evaluating the technical progress of these methods.

Existing System:

The importance of accurate video classification tasks can be realized by the large amount of video data available online. People around the world generate and consume a huge amount of video content. Currently, on YouTube only, over 1 billion hours of video is being watched by different people on every single day. In recognition of the importance of video classification tasks, a combined effort is being made by the researchers to propose an accurate video classification framework. Companies like Google AI are investing in different competitions to solve the challenging problem under constrained conditions.

To further advance the progress of automatic video classification tasks, Google AI has released a public dataset called YouTube-8M with millions of video features and more than 3700 labels. All these efforts being made demonstrate the need for a powerful video classification model. An Artificial Neural Network (ANN) is an algorithm based on the interconnected nodes to recognize the relationships in a set of data. Algorithms based on ANNs have shown great success in modelling both the linear and the non-linear relationships in the underlying data.

Proposed System:

Basic Deep Learning Architectures for Video Classification The two most widely used deep learning architectures for video classification are Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). CNNs are mostly used to learn the spatial information from videos, whereas RNNs are used to learn the temporal information from videos. As, the main difference between these two architectures is the ability to process temporal information or data that comes in sequences. Therefore, both these network architectures are used for completely different purposes in general.

However, the nature of video data with the presence of both the spatial and the temporal information demands the use of both these network architectures to accurately process the two-stream information. Video Classification based on different Modalities Uni-Modal Multi-Modal Text Audio Visual Combination of Text, Audio, and Visual information. The architecture of a CNN applies different filters in the convolutional layers to transform the data. RNNs on the other hand reuse the activation functions to generate the next output in a series from the other data points in the sequence. However, the use of only 2D CNNs alone limits the understanding of video to only the spatial domain. RNNs on the other hand have the ability to understand the temporal content of a sequence. Both these basic architectures, and their enhanced versions, are applied in several studies for the task of video classification.

Software Tools:

1. VS Code
2. Jupyter Notebook
3. Colab
4. Python3
5. TensorFlow
6. Anaconda
7. Keras

Hardware Tools:

1. Laptop
2. Operating System: Windows 11
3. ROM: 16GB
4. RAM: 5GB
5. Fast Internet Connectivity

Applications:

1. Video Tagging to the Tag Cloud in YouTube
2. Categorising Videos for google ads