Computer Vision 2020 Project

Action Recognition Using Deep Learning

Action recognition task involves the identification of different actions from video clips (a sequence of 2D frames) where the action may or may not be performed throughout the entire duration of the video. This seems like a natural extension of image classification tasks to multiple frames and then aggregating the predictions from each frame.

However, despite the huge success of deep learning architectures in image classification, progress in architectures for video classification and representation learning has been slower.

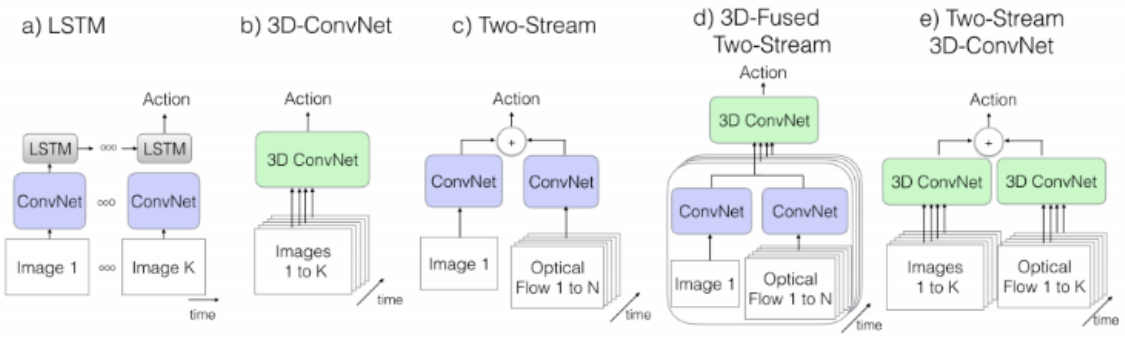
There are two main approaches to neural networks that are appropriate for time series classification and that have been demonstrated to perform well on activity recognition.

They are Convolutional Neural Network Models and Recurrent Neural Network Models.

Some Action Recognition Approaches that you can use:

- Single Stream Network: One Network for Spatial information.
- Two Stream Network: Involves the use of two separate networks and fusing their results for predicting classification score.
  →Spatial Network
  →Temporal Network

There are several techniques that build on these ideas as seen in the following figure:



Both types of neural networks suitable for time series classification **require that data be prepared in a specific manner in order to fit a model.** That is, in a ‘supervised learning‘ way that allows the model to associate signal data with an activity class.

In this project, you are given a dataset consisting of video sequences for some realistic sports videos such as: Basketball, Diving, Jumping. Your task is to perform action recognition and classification on this dataset using deep learning techniques.

The input: a video clip of a human performing a sport activity

The output: is to print the activity name.

Implementation Hint: NNs require the input to be of a fixed size. So, variable length videos must be handled accordingly before sending them to the networks.

References: Deep Learning for Videos: A 2018 Guide to Action Recognition

Project Notes:

- The Project results submission will be through a competition on Kaggle. You can find the competition link [here].
- The competition will be open until **14/5/2020**.
- You can register Project teams through [this form].
- Team Registration ends: **27/4/2020**.
- Minimum number of members is 3 and the maximum is 5

Final Project Deliverables:

1 - Complete action recognition model on the given dataset using deep learning techniques. (You can use the above techniques for guidance or any other deep learning technique).

2 – A Report that includes description of:

- Your data preparation process.
- Your neural network model architecture.
- Activity classification accuracy for this model.

Action Recognition Dataset

Description:

UCF is an action recognition data set of realistic action videos, collected from YouTube.

The dataset provided to you contains 5 categories **(Basketball, Diving, Jumping, Tennis, and Walking)**.

The goal is to build a model than can recognize the category of each video in the testing set.

- **You can download the Training set from** here.
- **You can download the Testing set from** here.
- The training set contains 5 folders, around 100 video for each category. The testing set contains 126 videos.
- On Kaggle you will find two files: **names.csv** and **sample.csv**
- **"submit.csv"** is a sample submission file for Kaggle in the correct format.
- **"names.csv"** contains the IDs of the five categories.