Computer Vision Project Documentation

Action Recognition

Team ID: 13

Team Members:

* Ahmed Araby kamel (section 1)
* Antonious Isaac
* Kirolos Albear
* Abd El-Eahman Rawhy
* Mario Naser Ebrahim

Model Architecture

C3D it’s an end to end model that take as input sequence of frames (raw images RGB) that represent consecutive time portion of the video to capture spatiotemporal features using 3D CNN layers, so in some sense the network is able to learn the required features and estimate the optical flow between frame.

Being end to end model and working on raw RGB data without any need for hand crafted features (e.g. optical flow) made it a good candidate for real time processing it also have high accuracy among it’s rivals.

The detailed architecture is:

* 5 groups each is formed of 3D convolutions and 3D max pooling layers followed by
* 4 fully connected layers including the output layer (classification layer), also there is a dropout in between the Fully connected layers.
* 3D kernels are 3X3X3.
* Number of kernels start with 64 to 512
* The model classify 5 different classes (in our case)

Data Preparation Process

To feed the model with the data either during training or testing, we have to convert the videos into collections of consecutive frames (possibly with gaps in between), the frames are RGB 112 pixel in width and 112 pixel in height also we have

* Preprocessing :
  + Normalized [0-1]
  + Centered image with Mean subtraction
* Augmentation:
  + Flipping the frames horizontally
  + Adding some noise

Some specific details about the followed techniques during training and testing

* During training:

We feed the model with sampled groups of consecutive frames randomly but within each group the frames are consecutive in term of time (16 frames per group in our case).

* During testing:

Same as training but instead of randomly selecting the frames we go throw all of them with gab = 8 frames in between the different groups of consecutive frames this decision was made to increase the accuracy of our model during prediction, also we have tried different size of gaps and 8 frames as a gap works best for us.

Results

Training accuracy: 99%

Testing accuracy in ladder board: 96%