

# Synopsis

Sriram Anupindi,Sharath R,Krishna Chaitanya S,Nirupam K N

February 4, 2016

Data is a set of values of qualitative or quantitative variables which is measured,collected ,reported and analyzed. Data can appear in many forms ranging from the traditional row-columns to images and videos. Due to the social media boom there has been a large increase in the number of images and videos generated and shared. This rapid increase in multimedia has peaked the interest of data scientist world over introducing the concept of "Semantic Analysis of Multimedia".

Multimedia consists of audio,video and images our project will focus on the automated analysis of video which is termed as Video Content Analysis(VCA).VCA deals with the analysis of video streams in order to perform a variety of tasks ranging from real-time analysis of video for immediate detection of interested events to analysis of pre-recorded video for the purpose of extracting events and data.

The task of Video analysis can be further classified as Activity recognition, Image description and Video description. Activity Recognition refers to detection of an activity. Image description is concerned with identifying the objects present in the image and presenting them in the form of text. Video Description is a more complex task it provides a textual description of a clipped video. Analysis of video can be carried out in different ways; one such way is Deep Learning.

Deep Learning is a branch of machine learning based on a set of algorithms which attempt to model high-level abstractions in data by utilizing the concept of multiple processing layers which have complex structures. In a traditional Neural Network approach a significant amount of redundant storage is generated for images and videos, this is overcome by introducing the concept of Convolutional Neural Networks(CNN).CNN's are composed of several layers of processing, each involving linear as well as non-linear operators, which are learnt jointly in an end-to-end manner to solve particular tasks.The CNN approach is the dominant approach for feature extraction from audio-visual and textual data.

However our challenge in extracting features doesnt get easily solved with CNN. The problem with CNN is, it treats each frame of a video as an independent entity. So RNN is used which considers the previous classes along with the inputs. The tradeoff between the conceptual and implantation complexity arises between the CNN and RNN approaches. The main motive of the project is to come up with an effective implementation using the available constructs. Many frameworks exist in implementing neural networks like Caffe, Keeras, etc. Caffe is a deep learning framework made with expression, speed and modularity in mind. It was developed by the Berkeley Vision and Learning Center (BVLC). Keeras is another framework which is quicker than Caffe. Written in Python, it runs on top of Theano and TensorFlow. There a variety of data sets available such as UCF-101,CCV,Hollywood 2 etc these data sets form the basis for video analysis.

Video Content Analysis covers a large domain of applications ranging from traffic to sports analysis.The Caltech Pedestrian dataset services as a reference for the pedestrian application.Abnormality and intrusion detection are popular applications which can be used for security purposes and theft prevention. Behavioural analytics on public video footage is a popular domain of VCA. The above applications provide an insight into the goals and aspirations of our project.