Machine Learning Engineer Nanodegree

Capstone Proposal

Rishab Kumar
June 3, 2017

Proposal

Domain Background

Saving T.V. shows and games for later view is an everyday thing most people are familiar with. Since people save the games to see the "games" alone and not the advertisements, advertisements waste both time and memory when saved. My life would be a thousand times better if T.V. had no advertisements. But because they exist, a evil way to lay them off should be developed because here is a little secret, "No one likes advertisements". For me, on a personal level, it's quite frustrating how in an 30 minute Indian TV show features up to 14 minutes of advertisement. It directly means that more than 45% of the storage used while storing a program is dedicated to thing I don't want on my system. The saving system has to be a little intelligent for sure. The system should know what to capture and what not to. I have tried my hands on implementing one such feature for the saving system. When a computer sits free, there are a thousand processes still going on in background. I wish to add one more. I aim to build a model using computer vision and deep learning that could distinguish between a cricket game(in fact any game) and advertisements so that I can see the game while not being anymore frustrated.

Problem Statement

Deep learning has been used to classify pre recorded video clips and even caption each clip, with each comprising a single action or subject. In this project I aim to

continuously classify video as it's captured. Continuous classification allows us to solve all sorts of interesting problems in real-time, like understanding what's in front of a car for autonomous driving applications to understanding what's streaming on a TV.

Now in this project I will be classifying what we see on our TV as either a cricket game or an advertisement using the power of Convolutional Neural Network(CNN). Using computer vision and deep learning , I propose to classify whether the program running on TV is cricket match or advertisement in real time. At each second image frame will be captured from TV and will be given as input to CNN for its appropriate classification. The so captured image frame will be resized as 100 x 100 RGB image(3 color channels) and an image shaped 100 x 100 x 3 will be the input.

Datasets and Inputs

For this project , I will be taking a clip of 20 minutes length of a recorded cricket match. It's a cricket match between India and Pakistan. In this, first an over is bowled and then an advertisement of approximately 30 seconds come in between. The match is recorded with 10 frames per second. 16 minutes of that recorded match is taken as training data and rest 4 minutes of the match is taken as testing data. So we have 9600 images for training data and 2400 images for testing data.The captured frames are labeled as 0 if its cricket match and 1 if it's an advertisement. (Image will be resized to 100 x 100 pixel and will be RGB colored image with 3 color channels)

Solution Statement

As we know , video is comprised of series of image frames and at each frame within a video, the frame itself holds important information (spatial). For this particular problem using only spatial features is sufficient for achieving high accuracy. I will be using 3-D CNN to solve this problem. Each image will be fed to our 3-D CNN model for image

classification and it will be well trained over 9600 images comprising of cricket match as well as advertisement. With the power and ability of 3-D CNN to capture important spatial features of an image, it will be easily able to classify input image between 2 categories.

Benchmark Model

Simple supervised learners and classifier ***Decision Trees*** will not be able to give good accuracy. As input feature vector will contain thousands of pixel values so it will be nearly impossible for traditional classifiers to classify the frames. So, using 3-D CNN will be easily able to capture the spatial information contained within the image and will be able to able to give high accuracy and good prediction. So, here I will be using Decision Tree as my benchmark model. I will first rain and test my data with ***Decision Tree*** which will fail to give good prediction and will be using it as benchmark.

Evaluation Metrics

The model shall be tested for accuracy, precision, recall and the time taken for prediction as well. Also it will be tested for its quick prediction in real time.

Project Design

This shall be the model that I propose. A colored frame with 3 channels (RGB) extracted from the video is input to the model. The image will be then resized to be a 100 x 100 pixel image with 3 color channels respectively. Then this array of shape 100 x 100 x 3 will be fed to a 3-D CNN with 2 hidden layers and one fully connected layer with the softmax function in the end for finding final probability of each class and giving the

prediction.

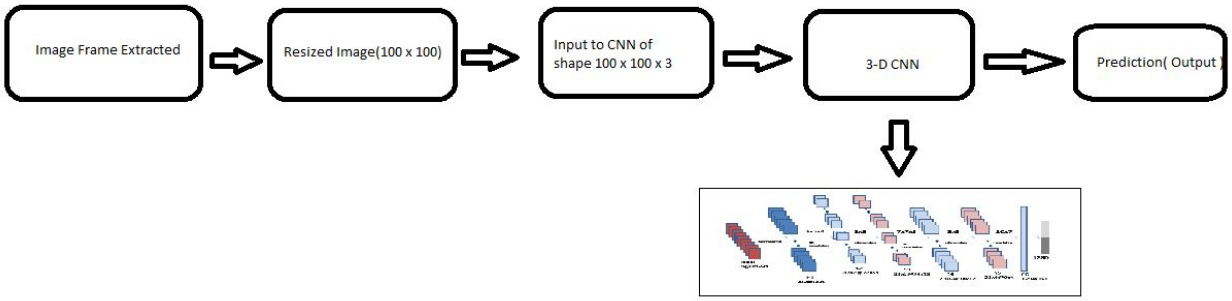
Image Frame Extracted
Resized Image(100 x 100)
Input to CNN of shape 100 x 100 x 3
3-D CNN
Prediction(Output)