Video Anomaly Detection

Priyesh Kaushik
Neelam Babel
Syed Haroon Perveez

Overview:

Video surveillance has been an area of significant interest in both academia and industry. Our focus is on problems, where we are given a set of nominal training video samples. Based on these samples we need to determine whether or not a test video contains an anomaly.

Anomalies in videos are broadly defined as events that are unusual and signify irregular behavior. Consequently, anomaly detection has broad applications in many different areas, including surveillance, intrusion detection, health monitoring, and event detection. Unusual events of interest in long video sequences, e.g. surveillance footage, often have an extremely low probability of occurring. As such, manually detecting these rare events, or anomalies, is a very meticulous task that often requires more manpower than is generally available.

This has prompted the need for automated detection and segmentation of sequences of interest. In contrast to the related field of action recognition where events of interest that are clearly defined, anomalies in videos are vaguely defined and may cover a wide range of activities. Our model, which is, trained using little to no supervision, including spatio-temporal features, and autoencoders, have been applied to evaluating anomalies.

Goal:

As part of this project, we propose a spatiotemporal architecture for anomaly detection in videos including crowded scenes.

Data:

http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html

Use Cases:

Anomaly detection has myriad applications such as –

- Intrusion detection
- Event detection
- Surveillance

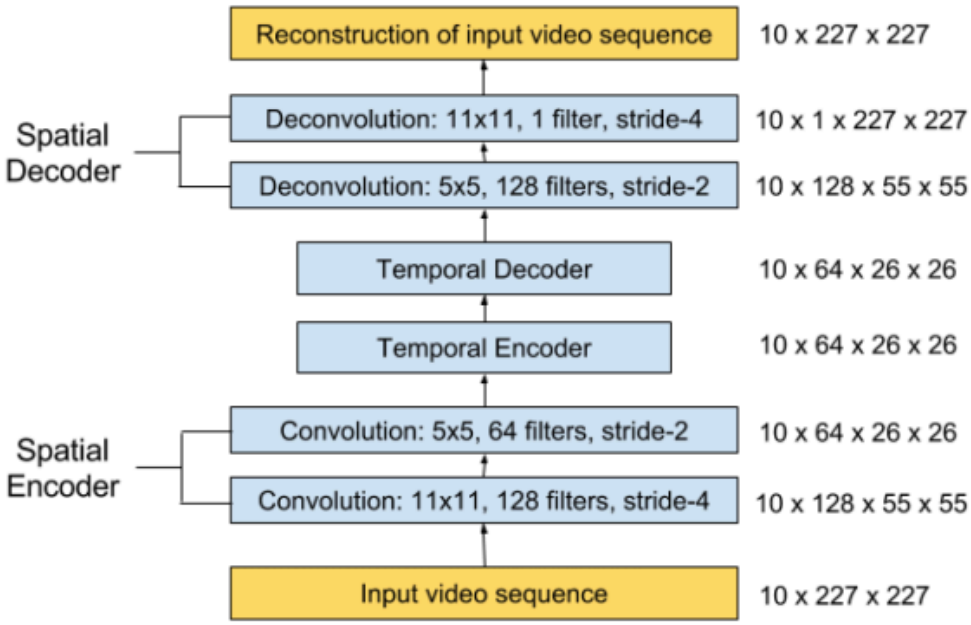
Process Outline:

Our approach consists of 3 main stages –

1. Preprocessing the data
2. Feature Learning using Convolution LSTM's -
   Our proposed architecture will have many stacked Conv-LSTM layers in an end-to-end trainable network. The 2 main elements of the design will be **Encoder** and **Decoder**.

   Encoder - The encoder accepts as input a sequence of reshaped frames in chronological order. By reshaping the input into a stack of non-overlapping patches, the model loses some detail but learns the more significant characteristics of the data.

   Decoder - We will have two decoders, one reconstructing the past input video sequence and the other predicting the future. The outputs of each layer are concatenated and summed through a convolutional filter to obtain the reconstruction of the input.



3. Evaluation -
   Once the model is trained, we can evaluate the model performance by feeding in test data to check if it's able to detect the abnormal events or not.

References:

https://arxiv.org/ftp/arxiv/papers/1612/1612.00390.pdf