Abstract

Abnormality detection augments a visual inspector's performance by drawing human attention towards the object of interest. This work focuses on devising such pipelines with the capability to detect and localize abnormality in the presence of diverse objects through automated machine learning.

Image and video signals are high dimensional data and are very structured in nature. These structural associations are utilized for detecting abnormality. We developed a framework that jointly learns both appearance and motion representations directly from the video. Further, analysis and visualization of spatiotemporal features learned at various hidden layer by deep architecture are also presented. Recently, deep architecture is also found to be vulnerable to adversarial attacks; most of them are focusing on the classification task. We extended such concept for abnormality detection, followed by providing a feasible solution as a defense mechanism.

Abnormality detection is also context-dependent. The same object can be considered as normal in one scenario while abnormal in other. In this regard, we developed an efficient image embedding technique that combines deep architecture with shallow network. The proposed approach is motivated by the word2vec model and extended for localizing contextual abnormality in the images.

The context captured in a frame depends on the size of the context window, patch-based processing of large input frame is time-consuming and unable to capture global context. For this, we developed frame-level global context modeling for abnormality localization. It takes the benefit of both convolutional neural networks and the adversarial autoencoder. The proposed model operates on the entire frame at a time and demands less number of trainable parameters.

Moreover, a similar region can occur between different objects and leads to ambiguity. Patch-based processing encodes limited information compared to object-level information we can get. From this perspective, we developed an efficient local-instance and context dictionary-based technique; it incorporates high-level semantic knowledge of objects and their association with different objects available across the complete frame in an unconstrained environment.

We further observed that the existing performance evaluation technique suffers from a few limitations; this has been thoroughly investigated. Finally, a solution to overcome these challenges is presented.