

Peer Review

Review of: "VADSK: Video Anomaly Detection with Structured Keywords"

Khalid Housni¹

1. Faculty of Science, Department of Computer Science, Université Ibn-Tofail, Kenitra, Morocco

The article is well-written and uses clear and concise language that effectively communicates the author's ideas and methodology. It presents a novel approach to addressing the problem of video anomaly detection using structured keywords. The author proposes using a large language model (LLM), specifically the pre-trained Llama model developed by Meta, to detect video anomalies by leveraging the power of foundational models and natural language processing. The idea is to extract an image from a video sequence and then describe its content using the LLM in order to identify the activities and objects in the scene. Based on the generated image description, the author attempts to analyze the generated keywords and compute a score for each keyword to identify terms indicating anomalies and to determine whether the event displays normal or abnormal behavior. However, there are a few remarks:

- The results show that this approach does not outperform most advanced methods proposed in this field.
- Due to the temporal inference required for each image in a video, this method is not viable for real-time applications.
- This method does not take into account the temporal information of a video, which is essential in this type of problem. Indeed, an object may appear anomalous in one situation but not in another, depending on the spatio-temporal context.
- This approach seems unsuitable for real-world application in video anomaly detection systems.

Declarations

Potential competing interests: No potential competing interests to declare.