Rethinking Video Anomaly Detection - A Continual Learning Approach - Supplementary Material

1. Dataset Description

In existing datasets (UCSD, Avenue, ShanghaiTech), the definition of a nominal event is restricted to a person walking, and the rest is considered anomalous. This is an unrealistic and trivial problem formulation since in a practical setup there are several non-walking activities that can be considered nominal. Hence, in the proposed NOLA dataset, the nominal and anomalous events are defined based on what a human surveillance operator might consider in a real-life scenario. For example, we consider events such as person loitering at an odd hour, a person carrying a snake, a vehicle moving in the wrong direction, etc. as anomalous. The new proposed dataset was collected from just one scene over a week, where a single camera moves after a fixed time duration. In the existing work, we only use the temporal annotations, but spatial annotation for a section of the dataset will also be made available. Due to the size of the dataset, it is not feasible to spatially annotate each and every frame. However, the entire stream of the CCTV feed will be publicly available.

2. Results

The reason for considering [1, 2] for comparison purposes is that only a few recently proposed state-of-the-art approaches have their codes publicly available. To make each model capable for continual learning, we followed the commonly adapted incremental learning approach, in which each model is trained on a batch of new data as it arrives, and then evaluated on the testing data. We believe it makes it a fair comparison because the same incremental training and evaluation is used for our approach and the existing ones. While we do use the day and time as input features, we were unable to find a straightforward way of including them in either of the existing approaches.

We are able to significantly outperform the existing approaches on the UCSD dataset because of class imbalance. In 9 out of 12 videos in the UCSD testing dataset, the primary anomalous activity is a person riding a bike, which the other state-of-the-art algorithms are unable to continuously learn and classify as a nominal activity. On the other hand, ShanghaiTech consists of several other anomalous activity types, due to which the performance gain is not so significant.

In Fig. 2, the optical flow approach used simply computes the mean and standard deviation for each frame in a video and compares it to a fixed threshold. The results illustrate the shortcomings of existing datasets since the anomalies present in them are primarily optical flow based.

References

[1] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6536–6545, 2018. 1

[2] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14372–14381, 2020. 1