Towards Interpretable Video Anomaly Detection

Anonymous WACV 2023 Algorithms Track submission

Paper ID ****

1. Proof of Theorem 1

Proof. Consider a circle $\mathcal{S}_t \in \mathbb{R}^2$ centered at $\boldsymbol{z}_t$ with radius δ_t, the kNN distance of $\boldsymbol{z}_t$ with respect to the training set $\mathcal{Z}$. The maximum likelihood estimate for the probability of a point being inside $\mathcal{S}_t$ under f_0 is given by $k/|\mathcal{Z}|$. It is known that, as the total number of points grow, this binomial probability estimate converges to the true probability mass in $\mathcal{S}_t$ in the mean square sense [1], i.e., $k/|\mathcal{Z}| \xrightarrow{L^2} \int_{\mathcal{S}_t} f_0(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}$ as $|\mathcal{Z}| \to \infty$. Hence, the probability density estimate $\hat{f}_0(\boldsymbol{z}_t) = \frac{k/|\mathcal{Z}|}{2\pi\delta_t^2}$ converges to the actual probability density function, $\hat{f}_0(\boldsymbol{z}_t) \xrightarrow{p} f_0(\boldsymbol{z}_t)$ as $|\mathcal{Z}| \to \infty$, since $\mathcal{S}_t$ shrinks and $\delta_t \to 0$. Similarly, considering a circle $\mathcal{S}_\alpha \in \mathbb{R}^2$ around $\boldsymbol{z}_\alpha$ which includes k points with its radius d_α, we see that as $|\mathcal{Z}| \to \infty$, $d_\alpha \to 0$ and $\hat{f}_0(\boldsymbol{z}_\alpha) = \frac{k/|\mathcal{Z}|}{2\pi d_\alpha^2} \xrightarrow{p} f_0(\boldsymbol{z}_\alpha)$. Assuming a uniform distribution $f_1(\boldsymbol{z}) = f_0(\boldsymbol{z}_\alpha)$, $\forall \boldsymbol{z}$, we conclude with $\log \frac{\frac{k/|\mathcal{Z}|}{2\pi d_\alpha^2}}{\frac{k/|\mathcal{Z}|}{2\pi\delta_t^2}} = \log \delta_t^2 - \log d_\alpha^2 \xrightarrow{p} \log \frac{f_1(\boldsymbol{z}_t)}{f_0(\boldsymbol{z}_t)}$ as $|\mathcal{Z}| \to \infty$. □

2. Online Anomaly Detection

The classical frame-based formulation for video anomaly detection does not evaluate the quick detection performance, which is critical in general in many anomaly detection applications, including video surveillance. To this end, we here consider a recently proposed online event-based formulation [2], which defines a sequence of successive anomalous frames as an anomalous event. In [2], a new performance metric, Average Precision Delay (APD), was proposed to evaluate the quick detection performance, which covers the classical AUC metric as a special case. While the cost is 1 for false negative (misdetection) and 0 for true positive for computing AUC, the APD metric uses a more detailed cost function, which penalizes each true positive with its detection delay (number of frames detection lags the anomaly onset) and each false negative with the maximum tolerable delay. APD measures the area under the alarm precision (number of true alarms/number of all alarms) vs. normalized average detection delay curve.

In this setup, we present our results only on the ShanghaiTech dataset as the CUHK Avenue dataset has fewer than 50 anomalous events, which is not enough for a reliable average performance comparison. A common technique used by several recent works [4, 3, 7, 9] is to normalize the computed statistic for each test video independently, including the ShanghaiTech dataset. However, this methodology cannot be implemented in an online (real-time) system as it requires the prior knowledge of the minimum and maximum values the statistic might take. Moreover, many recent methods [3, 5, 8] do not have their implementation details/code publicly available, while others are end-to-end [8, 10, 11] and cannot be implemented to work in an online fashion. Hence, we compare our method with the online versions of [4, 7, 6].

Our proposed algorithm achieves a better performance than the other algorithms in terms of quick detection and achieving high precision in alarms, as indicated by Table 2 in terms of the APD value.

Online Detection	
Method	**APD**
Liu et al. [4]	0.504
Morais et al. [7]	0.324
Luo et al. [6]	0.447
Ours	**0.675**

Table 1. Online detection comparison in terms of the APD metric [2] on the ShanghaiTech dataset. Higher APD value represents a better online anomaly detection performance.

3. Computational Complexity

For global monitoring, we consider the scenario in which the object is variable and the subject is fixed for positive/negative pairs. To compute the loss, we sample N subjects and K objects, which makes the computational complexity as $\mathcal{O}(NK)$. In practice, we are able to run the global monitoring branch at 10 fps. This can be further improved by reducing the number of subject/object pairs to be considered. The local monitoring branch can be run in par-

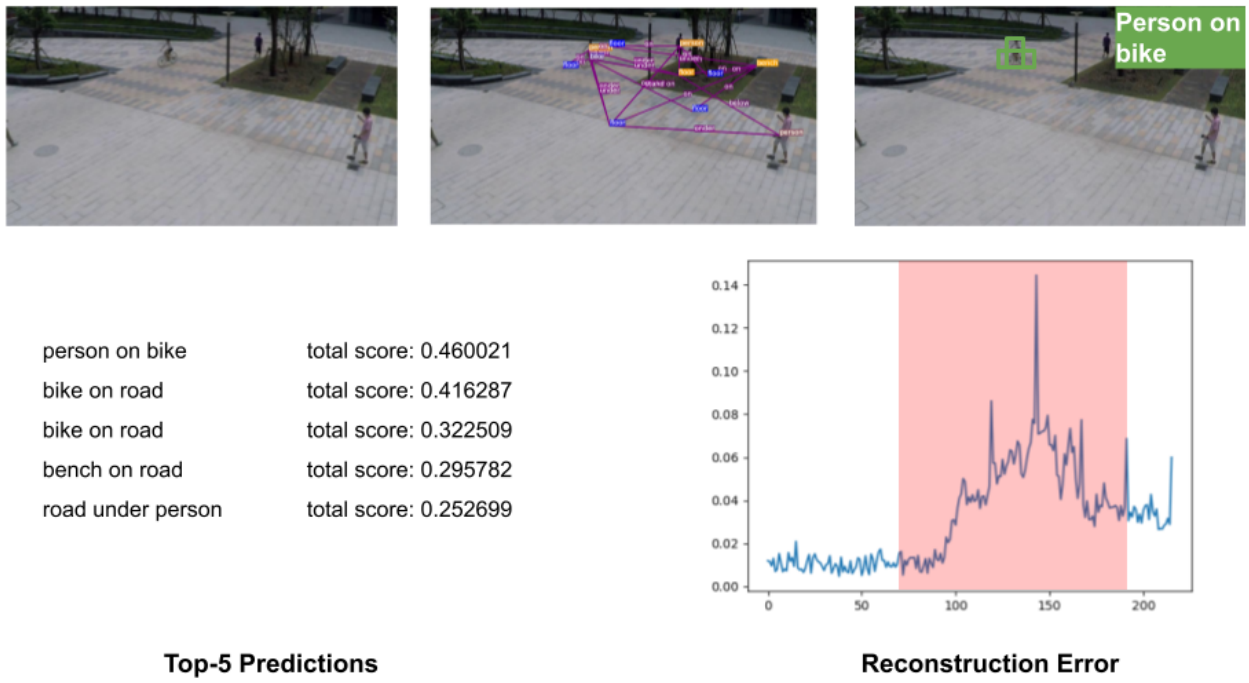

Figure 1. Pipeline level output from the local and global branches.

allel at 12 fps, which can be further improved by using a lightweight pose tracking model. The sequential anomaly detection algorithm has negligible overhead.

4. Effect of Individual Branches

In Table 2, we compare the branch-wise performance for Avenue dataset. We notice that the local monitoring branch significantly outperforms the global monitoring branch. However, this can be attributed to a small percentage of anomalies occurring in the Avenue dataset which involve human-object interaction. For example, out of 21 test videos, 2 videos consist of a person dancing, which cannot be detected using the global monitoring branch but can be detected through the pose-based local monitoring branch. On the other hand, the global branch is able to successfully detect anomalous cases such as a "person on grass", which does not involve an anomalous pose, and thus is ignored by the local branch.

We also show the pipeline level output from the global and local branches in Fig. 1. The output of the global branch is in terms of the predicted triplet and their scores whereas the output from the local branch is the reconstruction error. We see that the global branch outputs "bike on road" multiple times due to associating with two different sections of the crossroads. Here, the anomaly begins at $t = 57$, but the local branch detects it much later since the person is very small and pose estimation is not accurate. On the other hand, the global branch is able to detect it much earlier and at the same time gives an interpretable output. This shows the efficacy of a dual-branch approach.

Effect of each branch	
Branch	**AUC**
Global Monitoring Branch	0.46
Local Monitoring Branch	0.71

Table 2. Impact of each branch on the overall performance.

5. Additional Qualitative Interpretability Results

We provide more interpretability results in Fig. 1 for the ShanghaiTech dataset. Each row shows interpretation provided for a successfully detected anomaly. The ground truth in the dataset only provides labels (nominal or anomalous), but not the root cause. It is seen in Fig. 2 that the interpretations provided by the algorithm (in the last column) explain the root cause. In the last example (third row), the algorithm raises alarm due to two different object-object interactions

Figure 2. Anomaly interpretation examples for the proposed approach on the ShanghaiTech dataset. Each row corresponds to a correct detection case. In the last row, the algorithm provides two interpretations for the detected anomaly.

over the course of anomalous frame sequence. In addition to the obvious "person on bike" interpretation, the "bike on road" predicate also triggers alarm in several frames. Interestingly, both the bike itself and its reflection on the glass window contribute to the bike-on-road alarm.

References

[1] Alan Agresti. *An introduction to categorical data analysis*. Wiley, 2018. 1

[2] Keval Doshi and Yasin Yilmaz. A modular and unified framework for detecting and localizing video anomalies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3982–3991, 2022. 1

[3] Radu Tudor Ionescu, Fahad Shahbaz Khan, Mariana-Iuliana Georgescu, and Ling Shao. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7842–7851, 2019. 1

[4] 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

[5] Yiwei Lu, Frank Yu, Mahesh Kumar Krishna Reddy, and Yang Wang. Few-shot scene-adaptive anomaly detection. *arXiv preprint arXiv:2007.07843*, 2020. 1

[6] Weixin Luo, Wen Liu, Dongze Lian, Jinhui Tang, Lixin Duan, Xi Peng, and Shenghua Gao. Video anomaly detection with sparse coding inspired deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1

[7] Romero Morais, Vuong Le, Truyen Tran, Budhaditya Saha, Moussa Mansour, and Svetha Venkatesh. Learning regularity in skeleton trajectories for anomaly detection in videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11996–12004, 2019. 1

[8] Guansong Pang, Cheng Yan, Chunhua Shen, Anton van den Hengel, and Xiao Bai. Self-trained deep ordinal regression for end-to-end video anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12173–12182, 2020. 1

[9] 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

[10] Bharathkumar Ramachandra and Michael Jones. Street scene: A new dataset and evaluation protocol for video anomaly detection. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2569–2578, 2020. 1

[11] Royston Rodrigues, Neha Bhargava, Rajbabu Velmurugan, and Subhasis Chaudhuri. Multi-timescale trajectory prediction for abnormal human activity detection. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2626–2634, 2020. 1