Interpretable Video Anomaly Detection for Enhancing Cyclist Safety

Çağkan Gürsoy, Maya Aghaei

NHL Stenden University of Applied Sciences

Cyclist safety is a critical concern in urban environments, involving interactions with other road users such as cyclists, vehicles, pedestrians, and objects on the road. Monitoring cyclists' behavior in such a setting is essential to identify risks and requires innovative approaches to mitigate them in real-time. This research investigates the application of video anomaly detection (VAD) algorithms to detect unsafe cyclist behaviors and interactions in urban settings. Current manual methods for analyzing cyclist behavior in video footage require significant expertise and are not scalable for large datasets. To address these challenges, we propose an automated system that leverages VAD to efficiently extract the most relevant and anomalous segments or objects from long video recordings, enabling researchers and traffic safety professionals to focus on detailed analysis of shorter, high-priority clips.

This study builds on the methodology of Reiss and Hoshen's[1] work on attribute-based representations, which demonstrated the effectiveness of combining interpretable features with deep embeddings to enhance anomaly detection in video data. Our approach integrates velocity as an interpretable feature alongside deep representations to achieve robust and interpretable anomaly detection. Object detection using an off-the-shelf detector and optical flow estimation were employed to extract these features, which were then analyzed using density-based methods. Unique models were trained for each feature type: a Gaussian Mixture Model (GMM) for velocity features and k-Nearest Neighbors (kNN) for deep representations. The anomaly scores from these models were aggregated to produce a final anomaly score, enabling detection at both frame and object levels.

We have retrained and evaluated this method on a real-world dataset containing footage of the cyclist path and the road, annotated with cyclist safety anomalous labels. Notably, the proposed method achieved an AUROC of 94.6% in frame-level evaluations. However, object-level evaluations yielded an AUROC of 55.5%, reflecting a significant drop in performance. This drop can be primarily attributed to the use of an off-the-shelf object detector which introduces excessive noise and inaccuracies in feature extraction. Addressing these limitations remains a priority for future research.

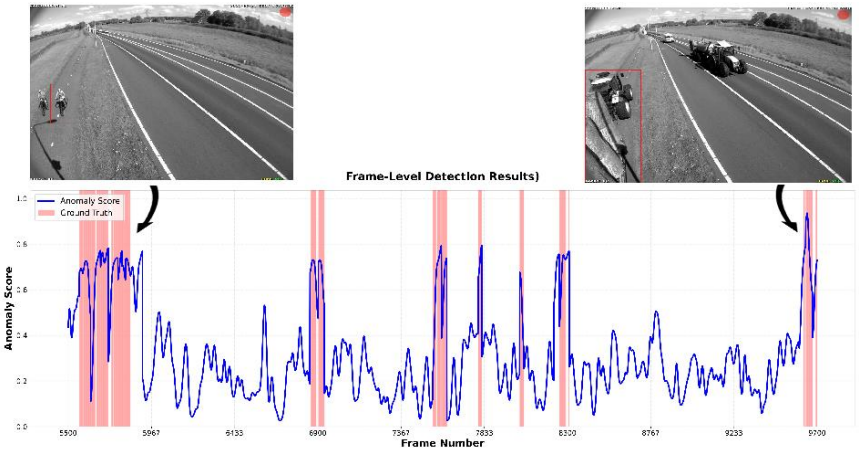

Figure 1. *Frame-Level Anomaly Detection Results: Comparison of Anomaly Scores and Ground Truth. Left shows the example of a cyclist riding in the wrong side of the bike lane. Right shows the example of a vehicle riding in the bike lane.*

[1] Tal Reiss and Yedid Hoshen. Attribute-based representations for accurate and interpretable video anomaly detection. arXiv preprint arXiv:2212.00789, 2022