

uTRAND: Unsupervised Anomaly Detection in Traffic Trajectories

Giacomo D'Amicantonio
Eindhoven University of Technology
Eindhoven, Netherlands
g.d.amicantonio@tue.nl

Egor Bondarau
Eindhoven University of Technology
Eindhoven, Netherlands
e.bondarev@tue.nl

Peter H.N de With
Eindhoven University of Technology
Eindhoven, Netherlands
p.h.n.de.with@tue.nl

Abstract

Deep learning-based approaches have achieved significant improvements on public video anomaly datasets, but often do not perform well in real-world applications. This paper addresses two issues: the lack of labeled data and the difficulty of explaining the predictions of a neural network. To this end, we present a framework called uTRAND, that shifts the problem of anomalous trajectory prediction from the pixel space to a semantic-topological domain. The framework detects and tracks all types of traffic agents in bird's-eye-view videos of traffic cameras mounted at an intersection. By conceptualizing the intersection as a patch-based graph, it is shown that the framework learns and models the normal behaviour of traffic agents without costly manual labeling. Furthermore, uTRAND allows to formulate simple rules to classify anomalous trajectories in a way suited for human interpretation. We show that uTRAND outperforms other state-of-the-art approaches on a dataset of anomalous trajectories collected in a real-world setting, while producing explainable detection results.

1. Introduction

In the domain of traffic surveillance, the task of anomaly detection has been subject to a progressive evolution. Initially, attempts have been made to devise expert systems characterized by complex sets of rules, aiming at emulating the intricacies of traffic dynamics. However, the implementation and scalability of such systems are impeded by the inherent complexity and variability of traffic behavior, often necessitating a disproportional number of exceptions to accommodate diverse scenarios. Recent advancements in computer vision are bolstered by the proliferation of sensor data from traffic monitoring infrastructure, which allow

for a paradigm shift in anomaly detection methodologies. These modern approaches leverage the computational capabilities of neural networks to analyze data streams obtained from traffic cameras. While exhibiting promising performance on standardized benchmarks, these methodologies have concurrently shown critical aspects such as privacy preservation, algorithmic biases, and false-positive mitigation.

This paper presents a methodological framework designed to address the aforementioned challenges in anomaly detection within traffic intersections called uTRAND (Unsupervised TRajjectory ANomaly Detector). Leveraging neural network architectures, the proposed approach detects and tracks moving traffic agents, by translating raw sensor data into a structured semantic-topological space. This framework allows to formulate simple anomaly detection rules. To validate the efficacy of uTRAND in practical scenarios, we experiment on a real-world intersection environment, leveraging a three-camera setup to capture anomalous traffic behavior. Comparative evaluations against existing approaches demonstrate the superior performance of the proposed framework in terms of anomaly detection accuracy and false-positive mitigation, while adhering to stringent deployment criteria delineated in this work.

2. Related Work

The field of traffic anomaly detection has seen significant contributions, categorized in three main groups: unsupervised (Section 2.1), weakly supervised (Section 2.2), and supervised (Section 2.3) approaches.

2.1. Unsupervised Anomaly Detection

Unsupervised methods rely on the assumption that any trajectory diverging from regular patterns indicates an

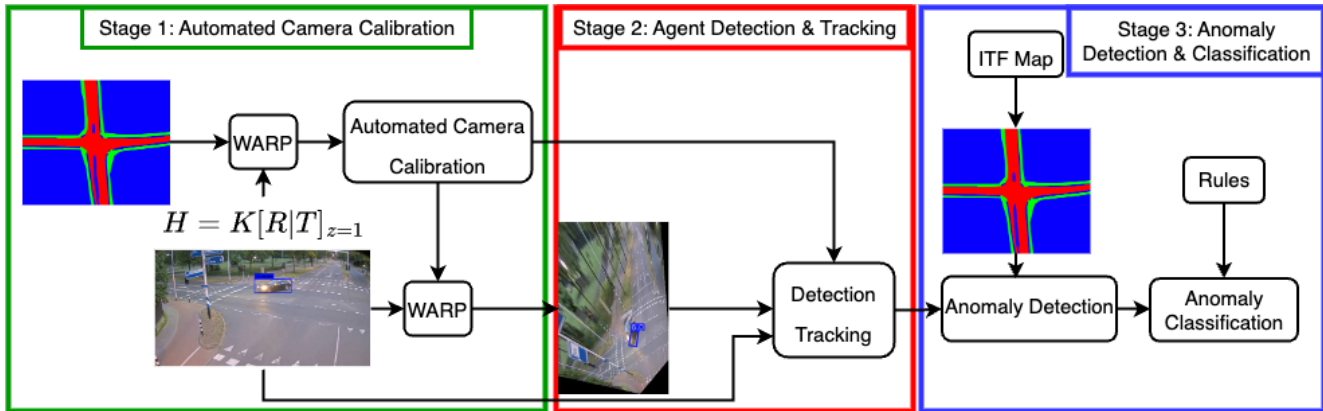


Figure 1. Overview of uTRAND. At the first stage, the cameras are calibrated and the videos are warped, obtaining their BEV videos. At the second stage, the traffic agents are detected and tracked in the camera video and translated to the BEV video. In the last stage, the intersection is split in semantic patches (nodes). The previously detected agents are associated with the patch they occupy at each frame. The links between the nodes and normal behavior of an agent in that node are learned by analyzing a few hours of camera videos through the framework, modeling the dynamics of the intersection. An agent that deviates from the modeled behaviors is detected as anomalous. The anomaly can further be classified by using simple rules that do not require domain knowledge.

anomaly. Common approaches to the anomaly detection task [3, 21, 34] define the problem as a clustering task, employing high-dimensional representations to distinguish regular and anomalous trajectories. Despite varied implementations and conceptual approaches, these methods collectively encounter challenges related to computational complexity. Notably, D’Acerno et al. [8] have addressed these computational issues. Such methodologies find their applications in both anomaly detection and classification tasks.

Concerning unsupervised deep learning methodologies, the Variational Auto-Encoder (VAE) proposed by Kingma et al. [16, 17] is a prominent model for anomaly detection. Approaches based on VAEs leverage unsupervised training to acquire latent representations of dataset distributions, thereby enabling the identification of anomalies by detecting deviations from learned representations [4, 20, 29, 44]. While some studies have applied these techniques to traffic video analysis [18, 33, 35], privacy-related concerns arise due to the utilization of visual features. More recently, this has resulted in efforts to address such issues [12, 24, 25, 42]. Another avenue of research focuses on frame reconstruction within videos, where anomalies are detected based on discrepancies between generated and original frames [4, 12, 15]. However, these methods may fail when considering subtle differences in trajectories.

2.2. Weakly-Supervised Anomaly Detection

In weakly-supervised anomaly detection methodologies, the task of identifying anomalies within video data often adopts a Multiple Instance Learning (MIL) framework.

Each video instance is annotated with a label indicating the presence of anomalies, while models are trained to predict the onset and duration of these anomalies, which ensures accurate classification based on both aspects. This approach has been extensively explored in the literature, as evidenced by studies [13, 38, 40, 43]. Despite leveraging the temporal dimension of videos in various ways, these methods are frequently constrained by limitations in visual representations. A common challenge encountered in such approaches is the discrepancy between training and deployment scenes, wherein background pixels may hamper model performance and generalization, even when not directly implied from the detected anomalous behavior.

Recent advancements have witnessed the application of transformer architectures to this domain. For instance, in [19], attention mechanisms are employed to capture correlations between video-level and snippet-level anomalies. Similarly, [6] proposes to utilize transformer networks to mitigate scene-inconsistencies, which demonstrates higher efficacy on prevalent public datasets.

2.3. Supervised Anomaly Detection

Supervised anomaly detection frameworks typically involve training classifier models with anomaly-level labels. A common class of methods in this field are algorithms such as Support Vector Machines (SVM) [2, 30], operating on the premise of discerning anomalous trajectories from regular ones. However, these methods often lack the required flexibility to effectively handle diverse anomaly types, such as discrepancies in trajectory length. While some approaches adopt direct supervision [36, 37], others



Figure 2. The agents are initially detected in the camera view. YoloV8 estimates a bounding box for each agent, assigns to it an ID, and tracks it across frames. Subsequently, the detected bounding boxes are projected in the BEV, where the three dimensional bounding boxes are detected, carrying over the id assigned to the agent in the camera view.

explore semi-supervised or self-supervised learning strategies [11, 14, 41]. However, the acquisition and labeling of a sufficiently large dataset for effective model training can be prohibitively expensive. Furthermore, the labeled classes in the data define a specific set of classes which are detected by the model, while the others require a general out-of-dataset class or a dedicated fine-tuning to be detected. This aspect is relevant to control the occurrence of false positives.

A significant limitation of neural networks is their inherent lack of explainability [28, 41]. Recently, hybrid frameworks have emerged to address this issue. For instance, in [1, 27], neural networks are utilized for pattern and feature extraction from video data, while anomaly detection is performed using more insightful methods like decision trees. The framework introduced in this paper aligns with this hybrid paradigm, offering both feature extraction capabilities and enhanced explainability, while preserving the privacy of citizens when detecting an anomaly.

3. Methodology

The proposed trajectory-based anomaly detection framework uTRAND is designed to meet the following criteria.

- *Human-interpretable predictions*: the system is designed to produce predictions that are comprehensible to human users.
- *Reduced dependency on labeled data*: the framework operates autonomously without necessitating extensive collections of labeled video data, facilitating its deployment in real-world conditions.

To address the above criteria, we shift the detection of anomalies in trajectories from the camera video domain to a

semantic graph representation derived from the bird’s-eye-view (BEV) of the intersection. The framework learns regular behavioral patterns of agents traversing intersections through unsupervised analysis of unlabeled data, thereby enabling the detection of anomalous trajectories. Moreover, this approach facilitates the formulation of straightforward rules for classifying different anomaly types. The architecture of uTRAND is depicted in Figure 1 and comprises of three primary components: automated camera calibration, agent detection and tracking, and anomaly detection.

3.1. Automated Camera Calibration

The initial step in the framework involves the automated calibration of cameras located within the intersection. The calibration is performed by a Graph Neural Network (GNN)-based technique, which is described in [9]. Subsequently, the BEV representation of the intersection is segmented in three semantic regions: road, bicycle path, and walkable terrain, as depicted in Figure 3. For the sequel of this paper, we will refer to the semantically segmented BEV of the intersection as the BEV intersection. To facilitate this segmentation, intrinsic and extrinsic parameters of a Pan-Tilt-Zoom (PTZ) camera model are sampled. The BEV is then warped using the homography transformations obtained from the sampled parameters, to generate synthetic images of different viewpoints of the same intersection.

The GNN is trained on these semantically-segmented synthetic images to predict the homography matrices of each camera, efficiently mapping pixels from the camera views to the BEV. The homography projection operates under the assumption that every pixel lies on a planar surface, which is valid within the semantic domain and does not

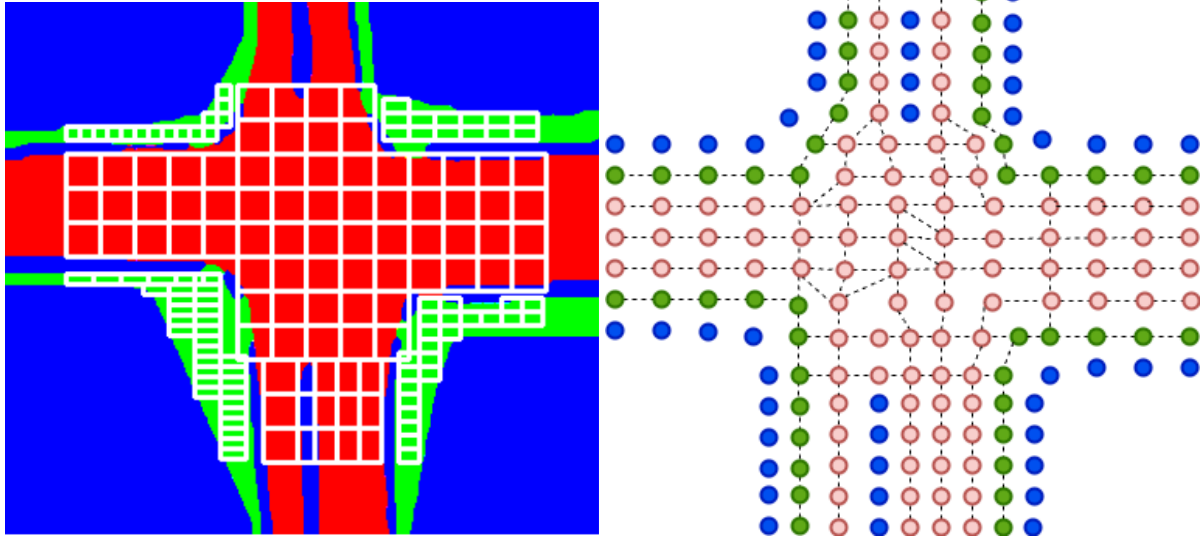


Figure 3. The intersection is split into patches according to the ITF map. The nodes can be of four different classes: road, bicycle lane, curb and crosswalk. The framework establishes connections between the nodes in which traffic agents move more often.

compromise the method’s generality. Leveraging the estimated homographies, the camera videos are warped to generate the BEV representations of each camera field-of-view, thereby enabling comprehensive analysis and anomaly detection within the intersection environment.

3.2. Agent Detection and Tracking

The second component detects and tracks traffic agents in the camera videos and projects their bounding boxes in the BEV videos obtained by warping the camera video, as shown in Figure 2. The agents detected and tracked are vehicles, pedestrians and bicyclists. To this end, the approach proposed by [5] is updated to use the more efficient YoloV8 model [31]. Each agent detected in the camera videos is assigned a unique ID and is tracked across frames. The model leverages the estimated homography to warp the feature maps of the camera video produced by different layers of the model. These feature maps are concatenated with the feature maps obtained from the BEV video. As a result, the model produces a bounding box for each agent in the BEV video, representing the base of the 3D bounding box, and an orientation angle indicating the direction of movement. At the end of this stage, the agents are detected, tracked and localized in the BEV of the intersection at each frame.

3.3. Anomaly Detection

In the final stage of the framework, the BEV intersection undergoes an automated partitioning into patches utilizing the Intersection Topology Format (ITF) maps [7], as shown in Figure 3. The ITF map models the topological layout of

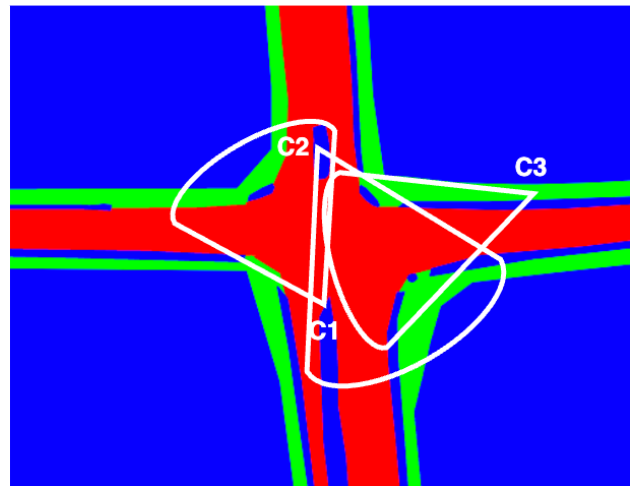


Figure 4. Semantically segmented bird’s-eye-view of the intersection (called BEV intersection). The fields-of-view of the three cameras overlap at the center of the intersection.

the intersection, delineating features such as ingress and egress points of road lanes and bicycle paths, expressed in world coordinates. It is also possible to perform an intersection partitioning into patches by utilizing the segmentation masks of the road markings [39]. Leveraging the camera calibration parameters acquired in the initial stage, these world coordinates are accurately projected onto the BEV videos, thereby facilitating the mapping of world coordinates to pixels. Subsequently, these pixels

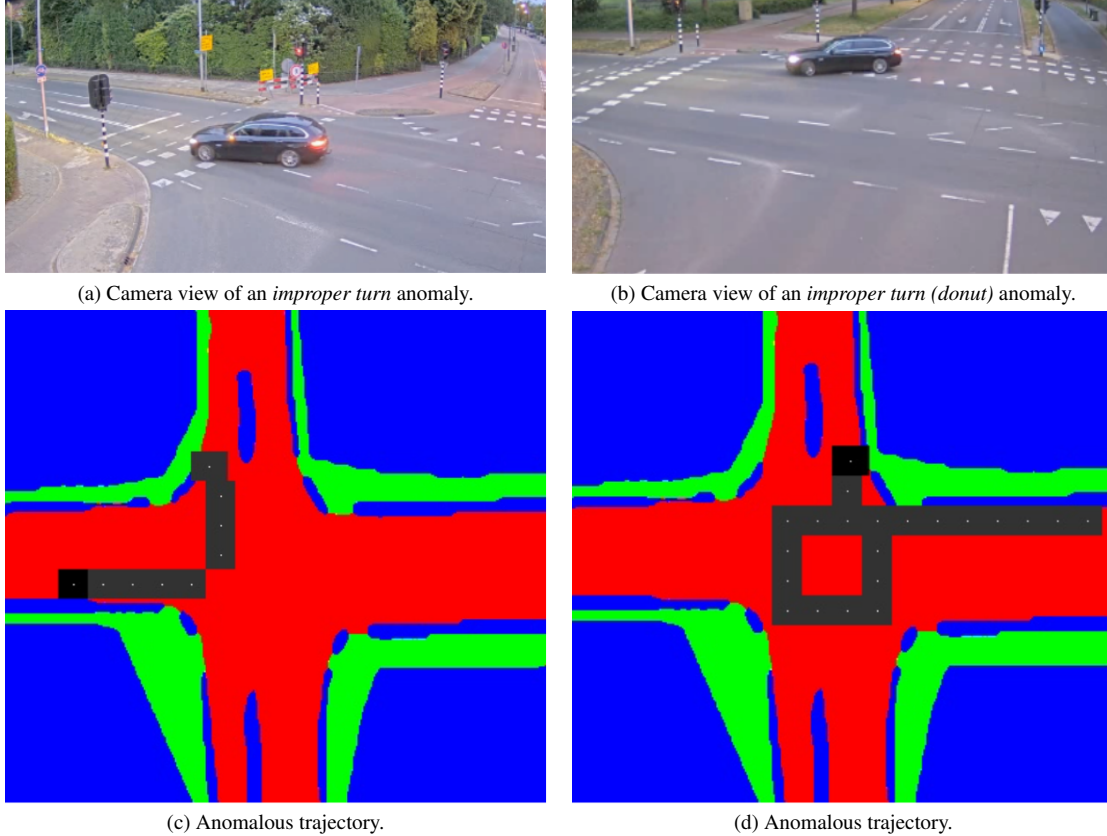


Figure 5. Camera views on the same intersection and the corresponding trajectory visualization. Top row: two frames of a vehicle performing improper turns. In subfigure 5a, the car is turning right and then drives in the opposite lane, while in subfigure 5b the same vehicle is driving in a circle in the intersection. Bottom row: the uTRAND’s visualization of the trajectories of the two actions. Both trajectories are correctly detected and classified as anomalies.

are utilized to automatically partition the semantic regions of the intersection into discrete patches. Each patch of the BEV intersection corresponds to a segment of a lane, bicycle path, or curb. Patches associated with crosswalks constitute an additional type of a patch. Agent association with patches is achieved by calculating the Intersection over Union (IoU) between the BEV bounding box of an agent and the patch itself. These patches are conceptualized as nodes within a graph representing the entire intersection. The relationships between nodes are learned through observation of agent movements within the intersection, establishing links between nodes, as an agent moves from one to another. This enables the framework to learn and model the agents behaviors within the intersection, characterized by three attributes for each node:

- $S_{node_i} = \{node_k^{t+1}, \dots, node_n^{t+1}\}$ the set of nodes to which an agent normally proceeds at (video) frame time $t + 1$ given $node_i$ at time t ,
- T_{avg} is average time spent in a node,

- A is a set that contains the types of agents allowed in the node.

These attributes are learned by processing several hours of camera videos, using the described pipeline. As illustrated in Figure 4, there is an overlap in the field of view (FOV) of the cameras. Consequently, if the cameras are synchronized, it becomes feasible to track an agent’s trajectory, as it transitions between the FOVs of different cameras. This capability enables the comprehensive analysis of an agent’s complete trajectory within the intersection. In this context, if the trajectory of an agent deviates from the learned regular behavioral model, the trajectory is automatically identified as anomalous in real time. In conclusion, the framework facilitates the formulation of rules that categorize anomalous behavior in simple terms using the previously defined attributes. For example, an unlawful turn by an agent from $node_1$ to $node_2$ at time t can be classified by the rule:

$$node_2 \notin S_{node_1}. \quad (1)$$

Furthermore, an agent in an unlawful position, e.g. a bicycle

on the road at $node_k$, can be classified in the following way:

$$bicycle \notin A_{node_k}, \quad (2)$$

where A_{node_k} is the set of agent types allowed in $node_k$. Finally, an agent that stays in $node_l$ from t_i to t_n can be classified as unlawfully stopping via the rule:

$$\text{card}\{t_i, \dots, t_n\} > T_{avg}^{node_l}. \quad (3)$$

The previous examples provide a set of rules to classify three types of anomalous trajectories. The formulation of such rules defines the framework to detect and classify anomalous trajectories of agents without explicitly labeling videos of anomalous and normal trajectories. Consequently, uTRAND is part of the unsupervised learning paradigm.

4. Experiments

The proposed framework is evaluated on video data collected at a traffic intersection, of which the BEV intersection is shown in Figure 4. The BEV intersection presents bicycle paths and curbs on the side of every road. Three cameras were mounted on light poles and cover the majority of the intersection. The model is set to learn the attributes of each node within the intersection, by observing the trajectories of traffic agents during a 12 hour temporal interval, covering the afternoon and evening from 12 pm to 12 am. This interval encompasses a diverse range of typical activities, including vehicular, cyclist and pedestrian moving in the intersection. Each node is connected to

Table 1. Amount of anomalous actions recorded. The anomalies are split in three classes: *agent in improper zone* (i.e. pedestrians and bicyclists on the road, vehicle parked on the curb), *unlawful stop* (i.e. car stopping at the center of the intersection) and *improper turn* (vehicles turning in the wrong direction, changing lanes, driving in a circle or doing U-turns).

| Anomaly | Amount |
|------------------------|-----------|
| Agent in improper zone | 24 |
| Unlawful stop | 3 |
| Improper turn | 14 |
| Total | 41 |

all neighboring nodes, meaning that each node is connected to all adjacent patches in the intersection. The inter-node links are uniformly initialized with unity weights. After this initialization, each traversal of an agent between two nodes contributes to a unity increment to the corresponding link’s weight. Upon observation and learning of trajectories within the 12-hour dataset, links characterized by a weight falling below an empirically determined threshold are pruned. This empirically established threshold is essential for mitigating localization inaccuracies, encountered

Table 2. Comparison between uTRAND and previous works. The baseline methods have been chosen based on their performances on public datasets. The F_1 score is the chosen comparison metric.

| Method | Pre-training | F_1 (%) |
|---------------|------------------|--------------|
| MONAD[10] | UCSD Ped2[26] | 62.35 |
| HF2VAD[23] | ShanghaiTech[22] | 69.62 |
| AI-VAD[32] | ShanghaiTech[22] | 72.49 |
| uTRAND | - | 82.89 |

Table 3. Performance of uTRAND on the classification task. The accuracy of the framework is reported along with the False Positive rates (FP) for each class.

| Anomaly | Accuracy (%) | FP |
|------------------------|--------------|------|
| Improper turn | 71.43 | 0.46 |
| Unlawful stop | 66.67 | 0.28 |
| Agent in improper zone | 91.16 | 0.54 |

during the second stage of the framework. Specifically, in crowded scenes the orientation of the estimated bounding boxes can change across consecutive frames while the direction of movement of an agent remains the same, resulting in an erroneous assignment during the third stage of the framework.

To evaluate the efficacy of the framework, various anomalous behaviors have been authorized, played and recorded. The dataset comprising anomalous trajectories contains 41 instances, as detailed in Table 1. The performance of uTRAND is compared with three anomaly detection methods in Table 2. The selected benchmark methods are also based on the unsupervised learning paradigm, as mentioned in Section 2.

The uTRAND framework outperforms other approaches by a significant F_1 score margin, showing a notable performance improvement. Additionally, anomalous trajectories are visually depicted, as illustrated in Figure 5. The visualization produced by the framework portrays the logic of the anomaly detection, therefore it is intuitive for a human to understand why a trajectory is detected as anomalous. This is an important advantage over other methods that are based on estimation of an anomalous score computed by a neural network. In fact, such scores can be influenced by a significant amount of spurious correlations that the model may infer from a training dataset, thereby reducing the interpretation capability of the results.

Moreover, with the rules expressed in Section 3.3, the uTRAND framework exhibits the capability to classify anomalous actions. The performance of the proposed method on the classification task is shown in Table 3. It is important to note that the instances of false positive iden-

tifications are heavily dependent upon the precision of the camera calibration and agent detection stages. The straightforward formulation of the classification rules permits the extension of the framework’s final stage to classify additional types of anomalous trajectories neither by re-training the model from scratch nor employing any transfer learning and fine-tuning techniques. This capability of the framework represents a significant advantage over methods relying on neural networks to classify actions.

5. Conclusion

This study proposes a three-stage framework called uTRAND, for detecting and classifying abnormal traffic trajectories within an intersection. The framework leverages on the intersection’s topological layout by transposing the problem domain onto a semantic graph structure. We demonstrate that within this domain, accurate detection of abnormal trajectories is achievable, yielding interpretable predictions. The classification of anomalies relies on straightforward principles, rendering the framework adaptive to real-world scenarios and being interpretable for end-users. Further exploration in this field could benefit from more accurate object detection and tracking algorithms. While uTRAND demonstrates efficacy in detecting and classifying single-agent anomalous trajectories, additional efforts are necessary to tailor this approach towards multi-agent trajectory analysis. One possible research direction entails the application of neuro-symbolic or causality-based methodologies to address this challenge.

References

- [1] Armstrong Aboah. A vision-based system for traffic anomaly detection using deep learning and decision trees. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4207–4212, 2021. 3
- [2] Praveen Batapati, Duy Tran, Weihua Sheng, Meiqin Liu, and Ruili Zeng. Video analysis for traffic anomaly detection using support vector machines. *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, 2015:5500–5505, 2015. 2
- [3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), 2009. 2
- [4] Wenxiao Chen, Haowen Xu, Zeyan Li, Dan Pei, Jie Chen, Honglin Qiao, Yang Feng, and Zhaogang Wang. Unsupervised anomaly detection for intricate kpis via adversarial training of vae. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 1891–1899. IEEE, 2019. 2
- [5] Yiqiang Chen, Feng Liu, and Ke Pei. Monocular vehicle 3d bounding box estimation using homography and geometry in traffic scene. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1995–1999, 2022. 4
- [6] Yingxian Chen, Zhengzhe Liu, Baoheng Zhang, Wilton Fok, Xiaojuan Qi, and Yik-Chung Wu. Mgnfn: Magnitude-contrastive glance-and-focus network for weakly-supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 387–395, 2023. 2
- [7] CROW. Intersection topology format (itf). topology guidelines version 2.1. Technical report, Crow, 2018. 4
- [8] Antonio d’Acierno, Alessia Saggese, and Mario Vento. Designing huge repositories of moving vehicles trajectories for efficient extraction of semantic data. *IEEE Trans. on Intelligent Transportation Systems*, 16(4):2038–2049, 2015. 2
- [9] Giacomo D’Amicantonio, Egor Bondarev, et al. Automated camera calibration via homography estimation with gnns. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5876–5883, 2024. 3
- [10] Keval Doshi and Yasin Yilmaz. Online anomaly detection in surveillance videos with asymptotic bounds on false alarm rate. *CoRR*, abs/2010.07110, 2020. 6
- [11] Mariana-Iuliana Georgescu, Antonio Barbalau, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, and Mubarak Shah. Anomaly detection in video via self-supervised and multi-task learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12742–12752, 2021. 3
- [12] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. *CoRR*, abs/1904.02639, 2019. 2
- [13] Xing Hu, Jian Dai, Yingping Huang, Haima Yang, Liang Zhang, Wenming Chen, Genke Yang, and Dawei Zhang. A weakly supervised framework for abnormal behavior detection and localization in crowded scenes. *Neurocomputing*, 383:270–281, 2020. 2
- [14] Chao Huang, Jie Wen, Yong Xu, Qiuping Jiang, Jian Yang, Yaowei Wang, and David Zhang. Self-supervised attentive generative adversarial networks for video anomaly detection. *IEEE transactions on neural networks and learning systems*, 2022. 3
- [15] 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. *CoRR*, abs/1812.04960, 2018. 2
- [16] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *CoRR*, abs/1906.02691, 2019. 2
- [17] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022. 2
- [18] Santhosh Kelathodi Kumaran, Debi Prosad Dogra, Partha Pratim Roy, and Adway Mitra. Video trajectory classification and anomaly detection using hybrid cnn-vae. *arXiv preprint arXiv:1812.07203*, 2018. 2
- [19] Shuo Li, Fang Liu, and Licheng Jiao. Self-training multi-sequence learning with transformer for weakly supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1395–1403, 2022. 2
- [20] Shuyu Lin, Ronald Clark, Robert Birke, Sandro Schönborn, Niki Trigoni, and Stephen Roberts. Anomaly detection for

- time series using vae-lstm hybrid model. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4322–4326. Ieee, 2020. 2
- [21] Sophia W.T.T. Liu, Henry Y.T. Ngan, Michael K. Ng, and Steven J. Simske. Accumulated relative density outlier detection for large scale traffic data. *Electronic Imaging*, 30(9): 239–1–239–1, 2018. 2
- [22] W. Liu, D. Lian W. Luo, and S. Gao. Future frame prediction for anomaly detection – a new baseline. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 6
- [23] Zhian Liu, Yongwei Nie, Chengjiang Long, Qing Zhang, and Guiqing Li. A hybrid video anomaly detection framework via memory-augmented flow reconstruction and flow-guided frame prediction. *CoRR*, abs/2108.06852, 2021. 6
- [24] Zhian Liu, Yongwei Nie, Chengjiang Long, Qing Zhang, and Guiqing Li. A hybrid video anomaly detection framework via memory-augmented flow reconstruction and flow-guided frame prediction. *CoRR*, abs/2108.06852, 2021. 2
- [25] Weixin Luo, Wen Liu, and Shenghua Gao. Remembering history with convolutional lstm for anomaly detection. *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pages 439–444, 2017. 2
- [26] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. Anomaly detection in crowded scenes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1975–1981, 2010. 6
- [27] Rifkat Minnikhanov, Maria Dagaeva, Igor Anikin, Tikhon Bolshakov, Alisa Makhmutova, and Kamil Mingulov. Detection of traffic anomalies for a safety system of smart city. In *CEUR Workshop Proceedings*, pages 337–342, 2020. 3
- [28] Quoc Phong Nguyen, Kar Wai Lim, Dinil Mon Divakaran, Kian Hsiang Low, and Mun Choon Chan. Gee: A gradient-based explainable variational autoencoder for network anomaly detection. In *2019 IEEE Conference on Communications and Network Security (CNS)*, pages 91–99. IEEE, 2019. 3
- [29] Zijian Niu, Ke Yu, and Xiaofei Wu. Lstm-based vae-gan for time-series anomaly detection. *Sensors*, 20(13):3738, 2020. 2
- [30] Claudio Piciarelli, Christian Micheloni, and Gian Luca Foresti. Trajectory-based anomalous event detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1544–1554, 2008. 2
- [31] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. Real-time flying object detection with yolov8, 2023. 4
- [32] Tal Reiss and Yedid Hoshen. Attribute-based representations for accurate and interpretable video anomaly detection, 2022. 6
- [33] Pankaj Raj Roy and Guillaume-Alexandre Bilodeau. Road user abnormal trajectory detection using a deep autoencoder. In *Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13*, pages 748–757. Springer, 2018. 2
- [34] K. K. Santhosh, D. P. Dogra, and P. P. Roy. Anomaly detection in road traffic using visual surveillance: A survey. *ACM Comput. Surv.*, 53(6), 2020. 2
- [35] Kelathodi Kumaran Santhosh, Debi Prosad Dogra, Partha Pratim Roy, and Adway Mitra. Vehicular trajectory classification and traffic anomaly detection in videos using a hybrid cnn-vae architecture. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):11891–11902, 2021. 2
- [36] Mohammad Ibrahim Sarker, Cristina Losada-Gutiérrez, Marta Marron-Romera, David Fuentes-Jiménez, and Sara Luengo-Sánchez. Semi-supervised anomaly detection in video-surveillance scenes in the wild. *Sensors*, 21(12):3993, 2021. 2
- [37] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6479–6488, 2018. 2
- [38] Yu Tian, Guansong Pang, Yuanhong Chen, Rajvinder Singh, Johan W. Verjans, and Gustavo Carneiro. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4975–4986, 2021. 2
- [39] Julien A. Vijverberg, Bart Beers, and Peter H. N. De With. Fusing segmentation and domain-knowledge model to extract intersection topologies from aerial/terrestrial orthographic images. In *International Symposium on Electronic Imaging 2024*. SPIE, 2024. 4
- [40] Jie Wu, Wei Zhang, Guanbin Li, Wenhao Wu, Xiao Tan, Yingying Li, Errui Ding, and Liang Lin. Weakly-supervised spatio-temporal anomaly detection in surveillance video. *arXiv preprint arXiv:2108.03825*, 2021. 2
- [41] Jhih-Ciang Wu, He-Yen Hsieh, Ding-Jie Chen, Chiou-Shann Fuh, and Tyng-Luh Liu. Self-supervised sparse representation for video anomaly detection. In *Computer Vision – ECCV 2022*, pages 729–745, Cham, 2022. Springer Nature Switzerland. 3
- [42] Guang Yu, Siqi Wang, Zhiping Cai, En Zhu, Chuanfu Xu, Jianping Yin, and Marius Kloft. Cloze test helps: Effective video anomaly detection via learning to complete video events. *CoRR*, abs/2008.11988, 2020. 2
- [43] Chen Zhang, Guorong Li, Qianqian Xu, Xinfeng Zhang, Li Su, and Qingming Huang. Weakly supervised anomaly detection in videos considering the openness of events. *IEEE Transactions on Intelligent Transportation Systems*, 23(11): 21687–21699, 2022. 2
- [44] Yu Zhou, Xiaomin Liang, Wei Zhang, Linrang Zhang, and Xing Song. Vae-based deep svdd for anomaly detection. *Neurocomputing*, 453:131–140, 2021. 2