Unsupervised detection of incoming and outgoing traffic flows in video sequences

Jose D. Fernández-Rodríguez[1,2,3 [0000−0003−3702−2230]], Pablo Carmona-Martínez[1], Rafaela Benítez-Rochel[1,3 [0000−0001−8500−0488]], Miguel A. Molina-Cabello[1,2,3 [0000−0002−8929−6017]], and Ezequiel López-Rubio[1,2,3 [0000−0001−8231−5687]]

[1] Department of Computer Languages and Computer Science
University of Málaga, Málaga, Spain
[2] ITIS Software, University of Málaga, Málaga, Spain
[3] Instituto de Investigación Biomédica de Málaga – IBIMA, Málaga, Spain
{josedavid, pcm6tq1}@uma.es ; {benitez,miguelangel,ezeqlr}@lcc.uma.es

Abstract. As traffic cameras become prevalent, and a considerable amount of traffic videos are stored for various purposes, new possibilities and challenges open in the automatic analysis of traffic scenes. Advances in deep learning also enable new ways to characterize traffic in such videos automatically. This work is motivated by the need to understand traffic flow without human supervision, especially the localization of road intersections in scenes from traffic cameras. For this purpose, a method is proposed that uses a deep learning neural network for vehicle detection, an object tracker to recover vehicle trajectories from the detections, and unsupervised machine learning techniques to detect potential incoming and outgoing traffic flows from the vehicle trajectories in the video sequences. A wide range of real and synthetic videos have been used to test the goodness of the proposal with satisfactory results, from traffic cameras at different heights and angles, different traffic patterns, and various weather conditions.

Keywords: Unsupervised learning · Object tracking · Object detection · Video surveillance · Deep learning

1 Introduction

In recent years, video surveillance has been widely used for traffic data collection, monitoring, or surveillance for all types of traffic scenarios, such as highways, intersections, and roundabouts ([6],[11],[9]). The main objective of the video surveillance system in traffic is to monitor driving behaviors. It can play a vital role in detecting/predicting congestion, accidents, and other anomalies apart from collecting statistical information about the status of road traffic. In this field, many studies have been conducted focusing on different aspects such as scene analysis ([5],[2]), vehicle detection and tracking ([15],[1],[10]), anomalous trajectory detection ([18],[25]), traffic monitoring ([3],[12]), emergency management ([13],[23]), event detection ([17],[14]), etc.

With the applications of deep convolutional neural networks and the vast amount of traffic camera video available to anyone, it is only a matter of time before intelligent systems that monitor traffic through video-based traffic surveillance become increasingly common.

Most of the works published so far can be considered a first step in using deep learning in traffic applications because most of them find several difficulties in urban traffic scenarios such as intersections. The recognition and tracking process complexity is higher in these environments because vehicles typically involve more acceleration/deceleration, waiting, and turning from different entry points than typical highway traffic. Also, vehicles tend to occlude each other or be occluded by roadside infrastructures [7].

Recently, some researchers have applied deep learning to overcome those specific traffic analysis difficulties at intersections. For example, in [16], a CNN-based tool was developed for the automatic extraction of vehicle trajectories at crossroads, but the applicability of the proposed method can be improved, and [20] uses deep-learning (Yolov4) for vehicle detection with a camera installed at an intersection but it presents severe limitations.

This paper is aimed at facilitating comprehension of traffic flow in intersections watched by static cameras without human supervision using deep neural networks. It proposes a method that detects the points along the border of the scene. In this way, points where vehicles enter and exit from the scene are automatically detected. This enables the automation of the analysis of road intersection videos without the need to determine the position of the roads manually. Autonomous recovery of this information allows further analysis of the videos, such as detecting anomalous trajectories relative to others. Still, this work is not concerned with this additional analysis.

The remainder of this paper is organized as follows. Section 2 provides a detailed description of the methodology used for detecting potential incoming and outgoing traffic flows in traffic videos at intersections. Section 3 describes the settings items and the data set used for the experiments and the assessment of the performance of the proposed method. Finally, conclusions are drawn in Section 4.

2 Methodology

The proposed method to detect vehicle entry and exit locations in traffic videos depicting an intersection is given next. First of all, an object detection deep neural network $\mathcal{F}$ is employed to extract a set of object detections S_t in the current video frame $\mathbf{X}_t$ for each time instant t:

$$S_t = \mathcal{F}(\mathbf{X}_t) = \{(a_{t,i}, b_{t,i}, c_{t,i}, d_{t,i}) \mid i \in \{1, ..., N_t\}\} \tag{1}$$

where N_t is the number of object detections at time instant t. The upper left corner of the bounding box for the i-th object detection at time t is noted $(a_{t,i}, b_{t,i})$ while the lower right corner is noted $(c_{t,i}, d_{t,i})$.

Secondly, the output of the object detection network is supplied to a tracking method $\mathcal{G}$ that receives the current set of object detections S_t and the previous set of tracked objects Q_{t-1} and outputs the current set of tracked objects Q_t:

$$Q_t = \mathcal{G}\left(S_t, Q_{t-1}\right) = \{(\alpha_{t,j}, \beta_{t,j}, \gamma_{t,j}) \mid j \in \{1, ..., M_t\}\} \tag{2}$$

where M_t is the number of object tracks at time instant t. The centroid of the j-th tracked object at time t is noted $(\alpha_{t,j}, \beta_{t,j})$, and $\gamma_{t,j}$ is an integer that uniquely identifies the tracked object.

At the beginning of the video, the set of tracked objects is empty:

$$M_0 = 0,\ Q_0 = \emptyset \tag{3}$$

The tracked objects corresponding to all time instants up to the current time t are comprised in the set R_t:

$$R_t = \bigcup_{\tau=1}^{t} Q_\tau \tag{4}$$

Next, the set V_t of the first and last occurrences of all tracked objects up to the current time t is computed:

$$\begin{array}{l} V_t = \{(\alpha_{\tau,j}, \beta_{\tau,j}, \gamma_{\tau,j}) \in R_t \mid \\ (\forall\, (\alpha_{\tau',j}, \beta_{\tau',j}, \gamma_{\tau,j}) \in R_t \,:\, \tau' \leq \tau) \vee (\forall\, (\alpha_{\tau',j}, \beta_{\tau',j}, \gamma_{\tau,j}) \in R_t \,:\, \tau' \geq \tau)\} \end{array} \tag{5}$$

Then the elbow method for the k-means unsupervised clustering algorithm is run in order to obtain an optimal number of clusters k to cluster the set V_t. The elbow method selects the value of k that is associated with the maximum curvature of the curve of the Mean Quantization Error (MQE) as a function of k:

$$MQE_{k,t} = \frac{1}{|V_t|} \sum_{(\alpha_j, \beta_j, \gamma_j) \in V_t} \min_{h \in \{1,...,k\}} \left\| (\alpha_j, \beta_j) - \left(\bar{\alpha}_{h,k}, \bar{\beta}_{h,k}\right) \right\|^2 \tag{6}$$

where $|\cdot|$ stands for the cardinal of a set, $\|\cdot\|$ stands for the Euclidean norm of a vector, and $\left(\bar{\alpha}_{h,k}, \bar{\beta}_{h,k}\right)$ is the h-th cluster center obtained by the k-means algorithm for k clusters.

Let us note $\hat{k}$ the value selected by the elbow method. Then the associated set of cluster centers is:

$$C_t = \left\{ \left(\hat{\alpha}_h, \hat{\beta}_h\right) \mid h \in \left\{1, ..., \hat{k}\right\} \right\} \tag{7}$$

The set C_t is regarded as a concise representation of the set of observed centroids R_t. If vehicle detection and tracking have reasonably low error rates, the cluster centroids in C_t mark the places where vehicles enter and exit the scene in the camera viewport.

Fig. 1. The initial image frame for the first video synthesized with CARLA (left) and for the video *Highway* from the 2014 CDNET dataset (right). Initial/final points from each detected vehicle trajectory are drawn over the frame. Points that are grouped within the same cluster (after applying K-means) are drawn in the same color. Cluster centroids are in black. Each centroid represents an area near the border where either a whole road or an individual lane gets out of frame.

3 Experimental Results

This section presents the results obtained by applying the previously described methodology in a series of experiments with various videos of road intersections from the perspective of a traffic camera.

3.1 Methods

The system is implemented using OpenCV to read image frames from recorded videos or retrieve a live stream from a traffic camera, if available. For vehicle detection, the deep learning network Yolov5 is used. Specifically, of the range of publicly available sub-models, yolov5x6 is used. This sub-model is trained on the COCO dataset, and has a mAP@0.5 score of 72.7 [22]. Objects of COCO classes *car*, *motorcycle*, and *truck* are considered to be vehicle detections. Any other detections are discarded. Vehicle detections are fed to Norfair, a publicly available state-of-the-art object tracker with standard assignment heuristics and Kalman filters [21]. Default parameters are used for these components.

For each vehicle trajectory, starting and ending points are retrieved, and clustered using the K-means algorithm. The number of clusters is estimated with the elbow method. This method consists of repeatedly computing clustering with different cluster numbers, measuring the clustering performance (see Section 2 for details), and taking the point of maximum curvature for the clustering metric. The objective is to get cluster centroids to mark the positions of the roads and/or road lanes in a road intersection.

3.2 Datasets

To test the proposed method, a wide range of videos selected from several well-known datasets has been used. Three videos have been synthesized using CARLA

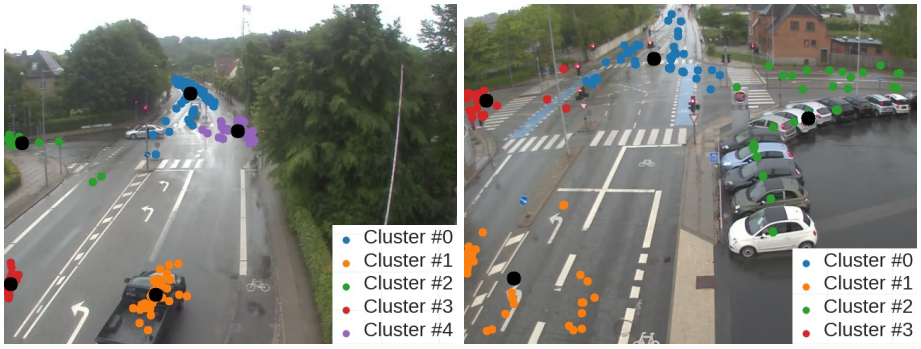

Fig. 2. The initial image frame for videos Hasserisvej-1 (left) and Hadsundvej-1 (right) from the AAU RainSnow Dataset. Left: the road end to the right of the image is obscured by a large tree, resulting in the detections corresponding to that road end being inside the intersection. Right: detected vehicles in a parking lot lead to an incorrectly placed cluster centroid (over the parking lot instead of over the road to its side). See the caption of the Figure 1 for details.

[8], and depict three different views of the same three-way intersection. These videos were scripted to include cars speeding, tailgating other vehicles, and unnecessarily switching lanes among dense traffic flow, to test the robustness of the proposed method when applied to videos with cars engaged in dangerous, nonconventional behaviors.

Apart from these synthetic videos, a total of 14 videos have been chosen from three state-of-the-art datasets. Seven videos are from the AAU RainSnow Traffic Surveillance Dataset[4] [4]. From this dataset, we take the following videos: Hadsundvej-1, Hadsundvej-2, Hasserisvej-1, Hasserisvej-2, Hasserisvej-3, Hjorringvej-2, and Ostre-3. These are videos recorded while raining and snowing, to validate the proposed method in bad weather conditions. For each intersection, all videos are from the same camera, but from different times and weather conditions.

Another six videos are taken from the Ko-PER Intersection dataset [19]. These are named Seq1_SK_1, Seq1_SK_4, Seq2_SK_1, Seq2_SK_4, Seq3_SK_1, and Seq3_SK_4, respectively. Of these, videos whose name is the same except for the ending number were recorded at the same time in the same intersection, but from cameras in different places and orientations. This is useful to test the proposed method with different viewing angles of the same vehicles. All videos depict four-way intersections, but some recordings do not include incoming/outgoing vehicles in one of the four ways.

While the proposed method has been conceived to be applied to videos from road intersections, we also test it in one video showing a non-branching road segment: the video titled *Highway* from the 2014 CDNET datasetDataset [24] (`http://changedetection.net/`).

[4] `https://www.kaggle.com/datasets/aalborguniversity/aau-rainsnow`

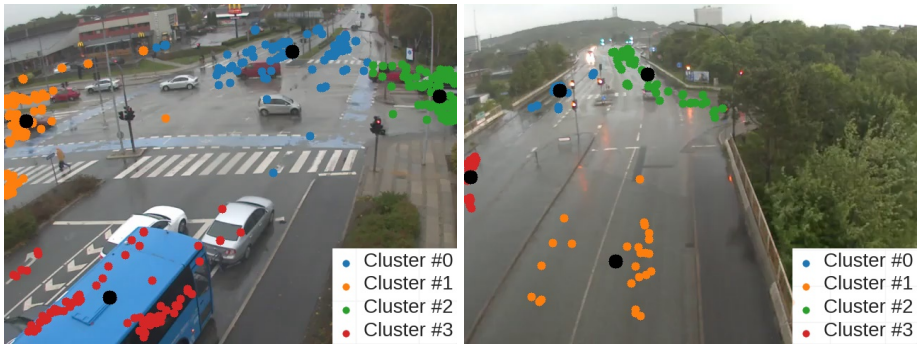

Fig. 3. The initial image frames for videos Hjorringvej-2 and Ostre-3, from the AAU RainSnow Dataset, with results obtained using yolov5x6 for detections. See the caption of Figure 1 for details.

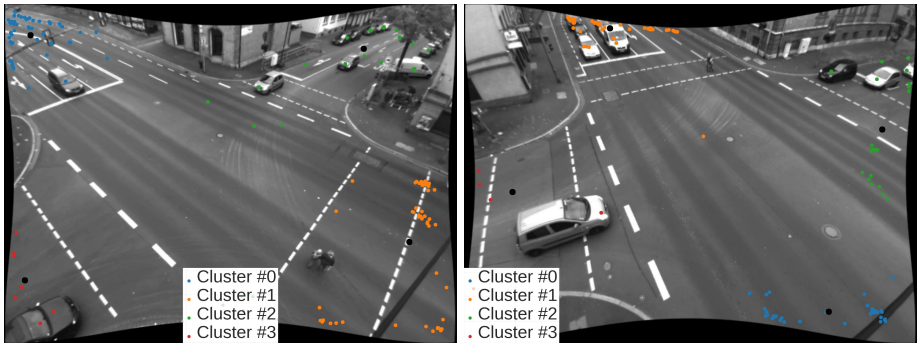

Fig. 4. The initial image frames for videos Seq1_SK_1 and Seq1_SK_4. See the caption of Figure 1 for details.

3.3 Results

The same three-way intersection is used with several camera views in the case of the three videos synthesized using CARLA. Only the first video is shown in the left side of Figure 1 (for a summary of results in the other two videos, see Table 1). For the first video (Figure 1) there are several false starting/ending points of trajectories in the middle of the image. This happens because at that point cars pass below a semaphore pole, and sometimes the method loses track of vehicle trajectories in these circumstances. The right side of Figure 1 shows the result of applying the proposed method to a traffic camera video of a straight road segment. One cluster marks the far end of the road, while two clusters are placed at the nearest road ending, corresponding to the two lanes of the road.

Regarding the videos from the AUU RainSnow Dataset, Hadsundvej-1 and Hadsundvej-2 (Figure 2 shows the first one) show a four-way intersection with a parking lot in one of the intersection's corners, full of cars. These cars are

Table 1. Summary of results. Each row corresponds to the performance of the proposed method for a video. Performance is measured by counting the number of cluster centroids placed over a road end. A false negative means that a road end in the video had no cluster centroid over it. Percentages are expressed over the ground truth (i.e., the number of road ends with vehicle traffic in the scene). A false positive means that a cluster centroid is placed inside or around the road intersection instead of over a road end. The last row shows aggregate figures for all videos.

video	number of road ends with vehicles entering/exiting the scene				
	ground truth	detected with 1 cluster	detected with > 1 cluster	false negatives	false positives
CARLA #1	3	3 (100%)	0 (0%)	0 (0%)	0
CARLA #2	3	2 (67%)	1 (33%)	0 (0%)	0
CARLA #3	3	3 (100%)	0 (0%)	0 (0%)	0
Hadsundvej-1	4	3 (75%)	0 (0%)	1 (25%)	1
Hadsundvej-2	4	3 (75%)	0 (0%)	1 (25%)	1
Hasserisvej-1	4	3 (75%)	1 (25%)	0 (0%)	0
Hasserisvej-2	4	3 (75%)	1 (25%)	0 (0%)	0
Hasserisvej-3	4	3 (75%)	1 (25%)	0 (0%)	0
Hjorringvej-2	4	4 (100%)	0 (0%)	0 (0%)	0
Ostre-3	3	0 (0%)	2 (67%)	1 (33%)	0
Seq1_SK_1	4	4 (100%)	0 (0%)	0 (0%)	0
Seq1_SK_4	4	4 (100%)	0 (0%)	0 (0%)	0
Seq2_SK_1	3	2 (67%)	1 (33%)	0 (0%)	0
Seq2_SK_4	3	2 (67%)	1 (33%)	0 (0%)	0
Seq3_SK_1	4	4 (100%)	0 (0%)	0 (0%)	0
Seq3_SK_4	4	3 (75%)	0 (0%)	1 (25%)	0
Highway	2	1 (50%)	1 (50%)	0 (0%)	0
TOTAL	60	47 (78%)	9 (15%)	4 (7%)	2

also detected by the proposed method, and their presence distorts the clustering enough to displace one of the cluster centroids to the point that it is over the parking lot rather than over one of the roads, so it does not detect the position of the road in any of these two videos. Furthermore, the off-road cluster centroid must be considered a false positive. It should also be noted that both videos are shot from the same camera but at different times. Relatively worse weather in Hadsundvej-2 leads to vehicles in the farthest road ending remaining undetected, shifting the cluster centroid corresponding to that road ending toward the center of the intersection.

Videos Hasserisvej-1 to Hasserisvej-3 (Figure 2 shows the first one) show another four-way intersection, this one with one of the roads obstructed by a big tree. Even in this case, the proposed method correctly places a cluster centroid in the position where the vehicles enter/exit the intersection, next to the tree blocking the view of the corresponding road ending. In all three videos, two cluster centroids (one per lane) are detected at the road ending closest to the camera. Finally, Hjorringvej-2 shows a four-way intersection and Ostre-3 a three-way one (Figure 3). In the first case, one cluster centroid is placed over each

road ending, while in the second case, the main road has two cluster centroids at each road ending (marking the lanes), and the detections for the other road ending are erroneously merged into one of these clusters, leaving that road ending unmarked.

With respect to the videos from the Ko-PER dataset, for Seq1_SK_1 and Seq1_SK_4 (Figure 4), the 4 clusters roughly correspond to the road endings. As can be seen, there are some initial/final trajectory points inside the road intersection, because of tracking errors, but these do not significantly influence the placement of the cluster centroids in most cases. On the other hand, for Seq1_SK_4 (right image), the orange cluster contains both points from two road endings, but the resulting cluster centroid is still placed over one of them (the one to the upper right). From the same dataset, clusters were also detected for videos Seq2_SK_1, Seq2_SK_4, Seq3_SK_1 and Seq3_SK_4 (not shown in any Figure, but results summarized in Table 1).

Results are summarized in Table 1. Each row show results for a video:

- The number of road endings in the video with vehicles entering or exiting through that end. This can be considered the ground truth.
- The number of road endings with exactly one cluster centroid placed over it.
- The number of road endings with more than one cluster centroid placed over it.
- The number of road endings that remain undetected (i.e. with no cluster centroid placed over it). This can be considered the number of false negatives.
- The number of cluster centroids that are placed inside or around the road intersection and cannot be considered to correspond to any of the road endings. This can be considered the number of false positives.

While the first of these items is the ground truth, the rest are results from the proposed method. Globally, 60 road endings with vehicle traffic are present in the 17 videos. Of these, the method places a single cluster centroid in around three-quarters of them, two cluster centroids in around one in ten, and no cluster centroids (false negatives) in around one in twenty.

4 Conclusions

Traffic surveillance videos from road intersections show traffic patterns that can be used to extract information about the scene. In particular, in this work, a method has been proposed that can automatically recover those potential incoming and outgoing traffic flows. This method works by detecting the position of the vehicles in each image frame. Then, from these frame-by-frame detections, vehicle trajectories are reconstructed by using a tracker method. By applying unsupervised clustering to these trajectories, the resulting cluster centroids enable an automatic understanding of the incoming and outgoing flows.

The proposed method has been tested with a set of videos, both from publicly available datasets and synthetically generated. Experimental results demonstrate that using better object detection models reduces the number of false negatives

and false positives in the placement of cluster centroids in the road ends. Since the method is robust to false starting/ending points of the vehicle trajectories, it also works when vehicles follow nonconventional trajectories (speeding, tailgating, repeatedly switching lanes) and in bad weather conditions (raining and snowing).

Acknowledgements

This work was partially supported by the Ministry of Science and Innovation of Spain, grant number PID2022-136764OA-I00, the Autonomous Government of Andalusia (Spain) under project UMA20-FEDERJA-108, and the University of Malaga (Spain) under grants B1-2021_20 and B1-2022_14. The authors acknowledge the grant of the Universidad de Málaga (UMA) and the Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND.

References

1. Abbas, A., Sheikh, U., Al-Dhief, F., Haji Mohd, M.N.: A comprehensive review of vehicle detection using computer vision. TELKOMNIKA (Telecommunication Computing Electronics and Control) **19**, 838–850 (06 2021). https://doi.org/10.12928/TELKOMNIKA.v19i3.12880
2. Abbas, Q., Ibrahim, M., M., J.: Video scene analysis: an overview and challenges on deep learning algorithms. Multimedia Tools and Applications **77**(16), 20415 –20453 (2018). https://doi.org/- 10.1007/s11042-017-5438-7
3. Abbasi, M., Shahraki, A., Taherkordi, A.: Deep learning for network traffic monitoring and analysis (ntma): A survey. Computer Communications **170**, 19–41 (2021)
4. Bahnsen, C.H., Moeslund, T.B.: Rain removal in traffic surveillance: Does it matter? IEEE Transactions on Intelligent Transportation Systems pp. 1–18 (2018). https://doi.org/10.1109/TITS.2018.2872502
5. Chong, Y.S., Tay, Y.H.: Modeling representation of videos for anomaly detection using deep learning: A review. ArXiv **abs/1505.00523** (2015)
6. Datondji, S.R.E., Dupuis, Y., Subirats, P., Vasseur, P.: A survey of vision-based traffic monitoring of road intersections. IEEE Transactions on Intelligent Transportation Systems **17**, 2681–2698 (2016)
7. Datondji, S.R.E., Dupuis, Y., Subirats, P., Vasseur, P.: A survey of vision-based traffic monitoring of road intersections. IEEE Transactions on Intelligent Transportation Systems **17**, 2681–2698 (2016)
8. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning. pp. 1–16 (2017)
9. Ercan Avsar, Y.O.: Moving vehicle detection and tracking at roundabouts using deep learning with trajectory union. Multimedia Tools and Applications **17**, 6653–6680 (2021)

10. Fernández, J.D., García-González, J., Benítez-Rochel, R., Molina-Cabello, M.A., López-Rubio, E.: Anomalous trajectory detection for automated traffic video surveillance. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Adeli, H. (eds.) Bio-inspired Systems and Applications: from Robotics to Ambient Intelligence. pp. 173–182. Springer International Publishing, Cham (2022)
11. Jahongir Azimjonov, A.O.: A real-time vehicle detection and a novel vehicle tracking systems for estimating and monitoring traffic flow on highways. Advanced Engineering Informatics **50** (2021)
12. Jain, N.K., Saini, R., Mittal, P.: A review on traffic monitoring system techniques. Soft Computing: Theories and Applications pp. 569–577 (2019)
13. Lopez-Fuentes, L., van de Weijer, J., González-Hidalgo, M., Skinnemoen, H., Bagdanov, A.: Review on computer vision techniques in emergency situations. Multimedia Tools and Applications **77**(13), 17069–17107 (Jul 2018). https://doi.org/10.1007/s11042-017-5276-7
14. Mo, X., Sun, C., Zhang, C., Tian, J., Shao, Z.: Research on expressway traffic event detection at night based on mask-spynet. IEEE Access **10**, 69053–69062 (2022)
15. Molina-Cabello, M.A., Luque-Baena, R.M., Lopez-Rubio, E., Thurnhofer-Hemsi, K.: Vehicle type detection by ensembles of convolutional neural networks operating on super resolved images. Integrated Computer-Aided Engineering **25**(4), 321–333 (2018)
16. Osama Abdeljaber, Adel Younis, W.A.: Extraction of vehicle turning trajectories at signalized intersections using convolutional neural networks. Arabian Journal for Science and Engineering pp. 8011–8025 (2020)
17. Pramanik, A., Sarkar, S., Maiti, J.: A real-time video surveillance system for traffic pre-events detection. Accident Analysis & Prevention **154**, 106019 (2021)
18. Raja, R., Sharma, P.C., Mahmood, M.R., Saini, D.K.: Analysis of anomaly detection in surveillance video: recent trends and future vision. Multimedia Tools and Applications pp. 1–17 (2022)
19. Strigel, E., Meissner, D., Seeliger, F., Wilking, B., Dietmayer, K.: The ko-per intersection laserscanner and video dataset. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 1900–1901. IEEE (2014)
20. Tak, S., Lee, J.D., Song, J., Kim, S.: Development of ai-based vehicle detection and tracking system for c-its application. Journal of Advanced Transportation **2021**, 1–15 (08 2021). https://doi.org/10.1155/2021/4438861
21. Tryolabs: Reference repository for Norfair. `https://github.com/tryolabs/norfair/`
22. Ultralytics: Reference repository for YoloV5. `https://github.com/ultralytics/yolov5`
23. Vivacqua, A.S.: Preparing a smart environment to decision-making in emergency traffic control management. In: Information Technology in Disaster Risk Reduction: Third IFIP TC 5 DCITDRR International Conference, ITDRR 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 20–21, 2018, Revised Selected Papers. vol. 550, p. 12. Springer Nature (2019)
24. Wang, Y., Jodoin, P.M., Porikli, F., Konrad, J., Benezeth, Y., Ishwar, P.: Cdnet 2014: An expanded change detection benchmark dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 387–394 (2014)
25. jun Zhao, X., Su, J.R., hui Cai, J., Yang, H., Xi, T.: Vehicle anomalous trajectory detection algorithm based on road network partition. Appl. Intell. **52**, 8820–8838 (2022)