Vehicle overtaking hazard detection over onboard cameras using deep convolutional networks

Jorge García-González, Iván García-Aguilar, Daniel Medina, Rafael Marcos Luque-Baena, Ezequiel López-Rubio and Enrique Domínguez

Abstract The development of artificial vision systems to support driving has been of great interest in recent years, especially after new learning models based on deep learning. In this work, a framework is proposed for detecting road speed anomalies, taking as reference the driving vehicle. The objective is to warn the driver in real-time that a vehicle is overtaking dangerously to prevent a possible accident. Thus, taking the information captured by the rear camera integrated into the vehicle, the system will automatically determine if the overtaking that other vehicles make is considered abnormal or dangerous or is considered normal. Deep learning-based object detection techniques will be used to detect the vehicles in the road image. Each detected vehicle will be tracked over time, and its trajectory will be analyzed to determine the approach speed. Finally, statistical regression techniques will estimate the degree of anomaly or hazard of said overtaking as a preventive measure. This proposal has been tested with a significant set of actual road sequences in different lighting conditions with very satisfactory results.

Jorge García-González · Iván García-Aguilar · Rafael Marcos Luque-Baena · Ezequiel López-Rubio · Enrique Domínguez-Merino
Department of Computer Languages and Computer Science. University of Málaga, Bulevar Louis Pasteur, 35, 29071, Spain
Biomedical Research Institute of Málaga (IBIMA). C/ Doctor Miguel Díaz Recio, 28, 29010, Spain
e-mail: jorgegarcia@lcc.uma.es, ivangarcia@lcc.uma.es, rmluque@lcc.uma.es, ezeqlr@lcc.uma.es, enriqued@lcc.uma.es

Daniel Medina
Institute of Communications and Navigation, DLR. Kalkhorstweg 53, 17235, Neustrelitz, Germany, e-mail: daniel.ariasmedina@dlr.de

1 Introduction

Nowadays, data plays a critical role in just about every aspect of our lives. Therefore, it can be helpful to analyze and make use of it to improve each discipline. One of the most common causes of injury and death worldwide is traffic accidents because they are uncertain and can occur in any place and at any time.

Numerous systems have been developed that are either part of the road or installed in the vehicle itself to reduce the number of accidents. Advances included in the last group are, for instance, collision avoidance, lane detection, lane departure systems, parking assistants, etc. These systems are mainly composed of sensors and cameras located in strategic areas of the vehicle. Visual data contains rich information compared to other information sources. Thus, it can play a vital role in detecting accidents, traffic jams, and other anomalies. The data collected by these systems is processed using computer vision algorithms and machine learning. When a potential driving risk is detected, the time available to react to it is usually very short. Therefore, the effectiveness of collision avoidance systems depends directly on the time required to identify the anomaly. If a short amount of time is taken for possible crash detection, then a longer amount of time is available for driver warning or evasive actions to be performed. Applied to the field of object detection with convolutional neural networks, there is a multitude of pre-trained models available, which can be classified according to the time required. The first group follows the classical flow, based on setting up a series of candidate regions and then inferring each of them. Models such as EfficientNet [11] or CenterNet [14] stand out among others. However, this type of model is not feasible for real-time detection due to the high computational time required. The second group proposes detection algorithms focused on minimizing computing time and thus speeding up detection. Models such as Single Shot MultiBox Detector (SSD) [7], Faster R-CNN [10], or Yolo [1] can be highlighted. In the field of vehicle detection and autonomous driving, Convolutional Neural Networks (CNNs) have accomplished huge outcomes. Traffic analysis has been a recurring topic of interest for researchers in recent years. Numerous studies focused on detecting vehicles [3, 8, 9] or tracking and counting [4, 13] promoting the development of new proposals in this field.

Finding traffic anomalies is a challenge due to the subjectivity of their definition. Based on the detection of anomalies in static highway cameras, *Earnest et al.* propose a framework for achieving a high Detection rate on general road-traffic surveillance footage [5]. The framework is based on local features such as trajectory intersection, velocity calculation, and their anomalies. It can detect accidents correctly with 71% detection rate on the accident videos obtained under various ambient conditions. According to the environmental conditions of the surroundings, *Wang et al.* sets up a framework for detecting different types of accidents in a mixed traffic environment with low-visibility conditions [12]. First, an image enhancement algorithm known as Retinex is applied. Next, a previously trained Yolov3 is applied to detect different situations including fallen pedestrians or vehicle rollovers. Finally, a set of features was developed from the Yolo results, based on which a decision model for crash detection was trained. In the field of anomaly detection in cameras installed in the car, we find works such as the one proposed by *Trung-Nghia et al.* who design an accident detection network called Attention R-CNN [6]. This network consists of two streams. The object characteristic stream employs the attention mechanism that exploits local and global contextual levels to recognize the object characteristic property. *Choi et al.* propose a car crash detection system, based on ensemble deep learning and multimodal data from dashboard cameras [2]. Deep learning techniques, gated recurrent unit (GRU), and convolutional neural network (CNN) are used to develop a car crash detection system. In addition, a weighted average ensemble set is used as an ensemble technique.

Our proposal is focused on detecting road speed anomalies, taking as reference the vehicle that is driving. Firstly, the information captured by the cameras installed in the car is processed to detect vehicles on the road. Each detected vehicle will be tracked over time, and its trajectory will be analyzed to determine the approach speed. Subsequently, statistical regression techniques are applied to determine whether or not an anomaly exists.

The following sections are structured as follows: section 2 shows the proposed methodology, section 3 on page 4 explains the experiments supporting our proposal and section 4 on page 9 explains our conclusions.

2 Methodology

In this section the proposed methodology for the detection of dangerously approaching vehicles is detailed. First of all, an object detection deep neural network must be applied to each incoming video frame, in order to generate a list of vehicle detections given by their bounding boxes. After that, an object tracking algorithm must be employed to obtain vehicle trajectories. Each trajectory is a list of bounding boxes associated to the same vehicle in successive video frames, allowing for intermediate frames where the vehicle has not been detected.

Let t be the time index within the video sequence. The angular diameter δ (in radians) of an object, also called apparent diameter, is given by:

$$\delta = 2 \arctan \frac{d}{2D} \tag{1}$$

where d is the actual diameter of the object (in meters) and D is the distance from the camera to the object (also in meters). If $D \gg d$, then we can apply the small angle approximation $\alpha \approx \arctan \alpha$ to obtain:

$$\delta = \frac{d}{D} \tag{2}$$

We have experimentally found that the small angle approximation yields good results for approaching vehicles. If we further assume that the object moves at a constant speed v relative to the camera, then (1) can be rewritten as:

$$\delta = \frac{d}{e_0 + vt} \tag{3}$$

where e_0 is the distance of the object at time $t = 0$. Then we can invert (3) to obtain:

$$\frac{1}{\delta} = \frac{e_0 + vt}{d} \tag{4}$$

Since d and e_0 are constants for each vehicle, (4) means that the inverse of the apparent diameter $\frac{1}{\delta}$ has a linear dependence with respect to time t. For each time instant that a vehicle is detected, its apparent diameter δ can be approximated as the square root of the area (in pixels) of the bounding box associated to the vehicle at that time instant. This allows computing the slope of the linear function of (4), which is the relative velocity v of the vehicle with respect to the camera, by linear regression. Please note that a sample for the linear regression is obtained for each video frame where the vehicle is visible. In order to enhance the estimation of v, the RANSAC (RANdom Sampling And Consensus) algorithm is employed to filter out the outlying measurements of $\frac{1}{\delta}$ within the linear regression. Finally, a threshold on v is defined so that approaching vehicles that have a excessively large velocity are detected as dangerous.

3 Experiments

3.1 Implementation Details

Figure 1 shows a schematic of our proposal. All the implementation of the method has been done using python[1] for general-purpose programming and pytorch[2] for the use of artificial neural networks. To ignore spurious detections, all trace detections with less than 30 elements are ignored. Likewise, to detect the relative velocity of a vehicle at time t, the detections of that vehicle from time $t - 15$ are used. The RANSAC application has been performed using the Scikit-Learn[3] library with 1 as the residual threshold value. 7 is used as a speed threshold to mark an overtaking as dangerous.

3.2 Data

Due to the lack of a specific dataset for this problem, a set of videos have been obtained using a camera installed on the rear of a car. The dataset contains 4 150-

[1] https://www.python.org/

[2] https://pytorch.org/

[3] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RANSACRegressor.html

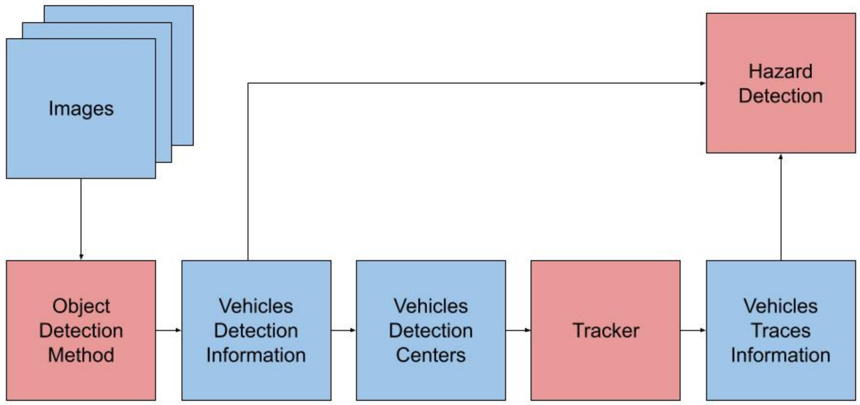

Fig. 1: Proposal scheme. Blue squares represent information while red squares represent sub-methods. Images from the sequence are provided to object detection method to get classes, bounding boxes, and confidences. Vehicles centers are given to the tracker to obtain a relation between vehicles appearing in different images. Both trace information and bounding boxes are then given to the proposed hazard detection method.

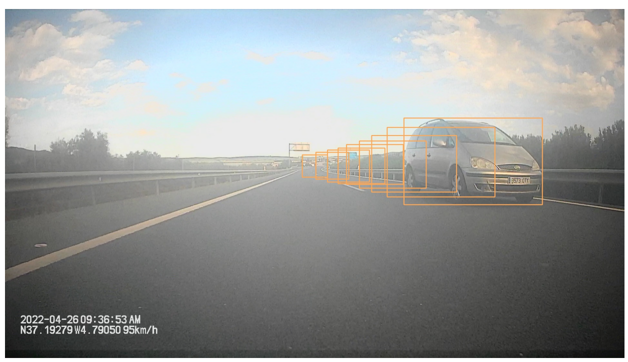

Fig. 2: Object detection bounding box evolution for a tracked overtaking car. Only one in ten bounding box is printed to allow a better readability.

second videos with shape 1920x1090, 30 frames per second, and 23 overtaking vehicles (both car and trucks). Overtakings have been manually annotated with their time window and labeled as dangerous or safe based on the relative speed of the recording car.

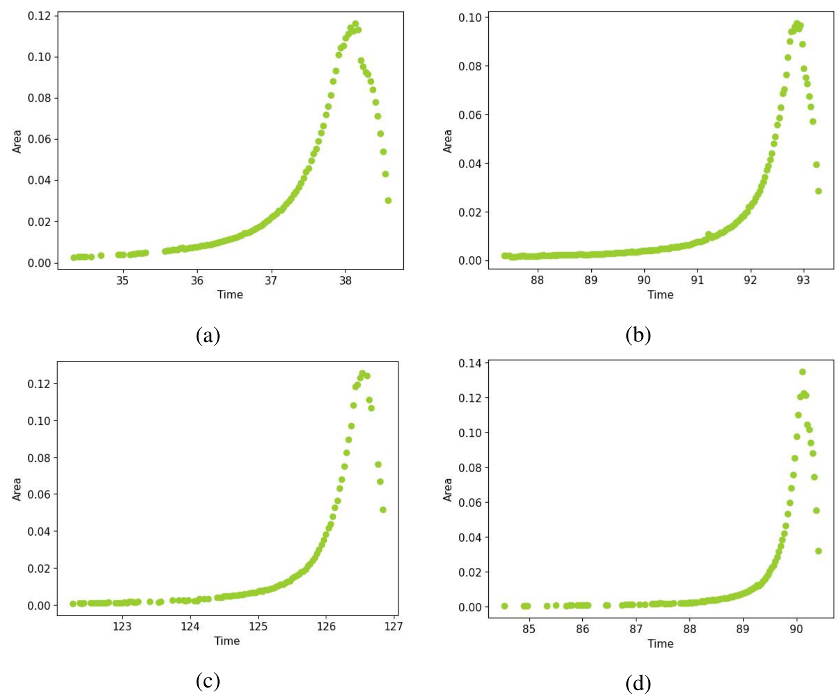

Fig. 3: Ratio of vehicle bounding box to the total area of the image for four different overtaking vehicles.

3.2.1 Tracker and Object detection

As a tracker algorithm, incoming frames bounding box centers are related to the last tracked objects using a linear sum assignment[4] to get the minimum weight matching in bipartite graphs. Our tracker also contemplates occasional disappearances of objects due to object detection failures.

In order to perform the object detection task, YOLO v5 [1] model has been applied. YOLO is a convolutional layer-based method using a single detection pass to split the image into N regions with size $S \times S$ to contain objects and classes proposals later unified using Non Maximum Suppression. The well known *ultralytics*[5] implementation of this model has been used with its pre-trained weights.

---

[4] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.linear_sum_assignment.html`

[5] `https://github.com/ultralytics/yolov5`

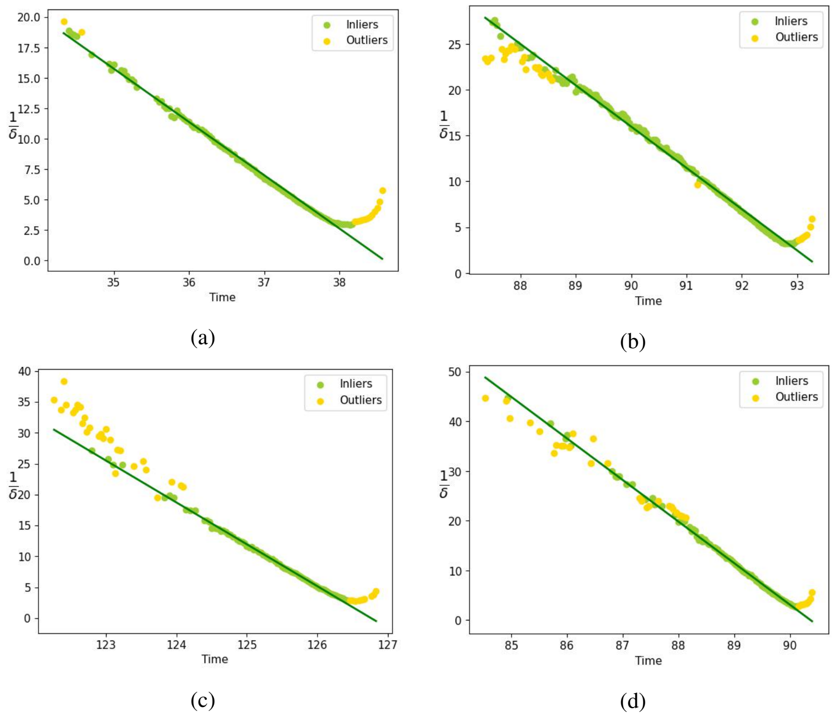

Fig. 4: Figure 3 data converted to lines and RANSAC effect. Green dots are points taken into account by RANSAC, yellow ones are points taken as outliers by RANSAC algorithm and not used to get the linear regression. Green line is the predicted line by the RANSAC regressor for the same time.

3.3 Evaluation

In order to study our method performance, two well known metrics as accuracy and precision have been selected.

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{5}$$

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

with TP (True Positive) as correct positive identification, TN (True Negative) as correct negative identification, FP (False Positive) as type I error, and FN (False Negative) as type II error. Accuracy provides an estimate of how close our predictions are to previously observed data while precision indicates the reliability of the system in identifying dangerous situations. Given a vehicle trace, if its estimated speed has

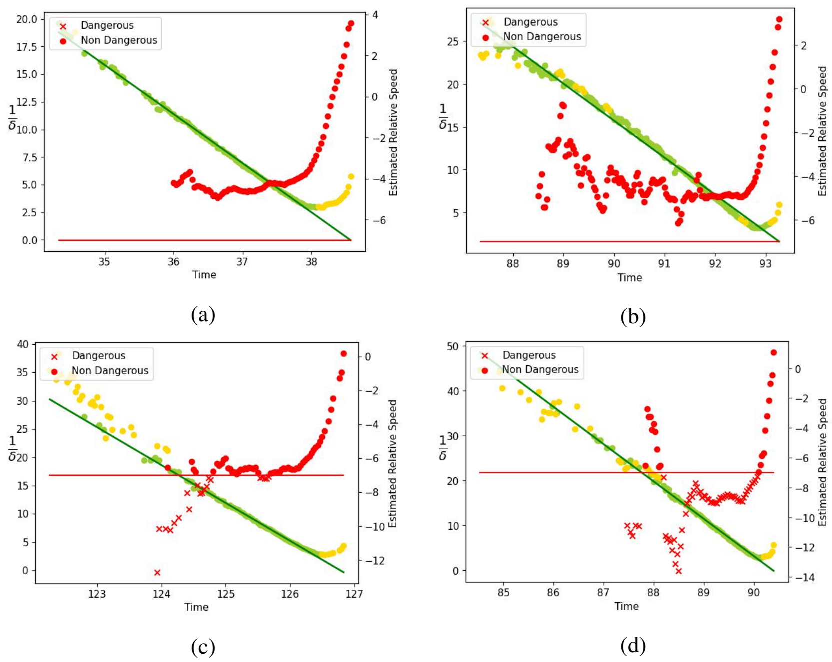

Fig. 5: Figure 4 with the danger labels assigned to each time t by our method. Red lines are thresholds to consider a relative speed as dangerous. Red dots are non dangerous speeds while red X's are dangerous detected speeds. Two are approximations considered hazardous, and two are approximations considered non-hazardous. Red lines, red dots and red X's are to be measured with the right y-axis.

been designated as hazardous at some point in the trace, the approach of that vehicle is considered hazardous.

3.4 Results

Figure 3 shows the ratio of the vehicle bounding box to the total area of the image for four different overtaking vehicles. The bounding box area increases while the overtaking is starting and then drops as the vehicle is coming out of the camera angle. Figure 4 shows the data for same four vehicles converted to a line using equations described in section 2. As can be observed, as the vehicle approaches the data is more stable and continuous until the vehicle begins to leave the visible region. The same figure also shows the effect of RANSAC on the data and its usefulness in preventing the slope from being affected by outliers.

Figure 5 shows the same data as 4 with the hazard output. As can be seen, even with RANSAC, older data tend to be less stable.

TP	TN	FP	FN	ACC	PPV
9	11	3	0	0.8695	0.75

Table 1: Quantitative results. The data shown in the first four columns are the sum of the results for the four videos studied because the accuracy and precision are calculated on the complete dataset.

Table 1 shows a summary of our results. The hyperparameter adjustment has taken into account the preference of type I errors over type II errors given the application of the system. This system is initially conceived as a possible aid to the driver, not as a substitute for him, so it would always be supervised and in the event of a hypothetical overtaking indicated as dangerous, the driver would be in charge of verifying and acting accordingly. This approach aligns with the absence of False Negatives, but the presence of false positives, decreasing precision but maintaining a better accuracy as shown in Figure 1.

4 Conclusions

In this paper, a method to support driving by detecting hazards during vehicle overtaking has been proposed. The proposal takes advantage of the evolution of the bounding boxes provided by an object detection method based on convolutional neural networks to estimate the relative speed of another vehicle overtaking the recording vehicle. To test it, a set of overtaking data has been collected and manually annotated. Based on the results obtained, the proposal is considered promising as a possible future driver assistance system that will allow the driver to act in the prevention of a possible accident. As future work, a larger data set of overtaking data should be used to study and correct the weaknesses of the method.

Acknowledgment

This work is partially supported by the Ministry of Economy and Competitiveness of Spain under grant TIN2016-75097-P and UMA grant B1-2019_01. It is also partially supported by the Ministry of Science, Innovation and Universities of Spain under grant RTI2018-094645-B-I00, project name Automated detection with low-cost hardware of unusual activities in video sequences and by the Autonomous Government of Andalusia (Spain) under project UMA18-FEDERJA-084, project name Detection of anomalous behavior agents by deep learning in low-cost video surveillance intelligent systems. All of them include funds from the European Regional Development Fund (ERDF). The work has been also supported by the University of Málaga through its Research Plan (Plan Propio de Investigación UMA). The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. They also gratefully acknowledge the support of NVIDIA Corporation with the donation of two Titan X GPUs used for this research. Iván García-Aguilar is funded by a scholarship from the Autonomous Government of Andalusia (Spain) under the Young Employment operative program [grant number SNGJ5Y6-15]. The authors acknowledge the funding from the Universidad de Málaga.

References

1. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection (2020). URL https://arxiv.org/abs/2004.10934
2. Choi, J.G., Kong, C.W., Kim, G., Lim, S.: Car crash detection using ensemble deep learning and multimodal data from dashboard cameras. Expert Syst. Appl. **183**(C) (2021). DOI 10.1016/j.eswa.2021.115400
3. García-Aguilar, I., Luque-Baena, R.M., López-Rubio, E.: Improved detection of small objects in road network sequences using CNN and super resolution. Expert Systems **39**(2) (2021). DOI 10.1111/exsy.12930
4. Gomaa, A., Minematsu, T., Abdelwahab, M.M., Abo-Zahhad, M., ichiro Taniguchi, R.: Faster CNN-based vehicle detection and counting strategy for fixed camera scenes. Multimedia Tools and Applications (2022). DOI 10.1007/s11042-022-12370-9
5. Ijjina, E.P., Chand, D., Gupta, S., Goutham, K.: Computer vision-based accident detection in traffic surveillance. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6 (2019). DOI 10.1109/ICCCNT45670.2019.8944469
6. Le, T.N., Ono, S., Sugimoto, A., Kawasaki, H.: Attention r-cnn for accident detection. In: 2020 IEEE Intelligent Vehicles Symposium (IV), pp. 313–320 (2020). DOI 10.1109/IV47402.2020.9304730
7. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single shot MultiBox detector. In: Computer Vision – ECCV 2016, pp. 21–37. Springer International Publishing (2016). DOI 10.1007/978-3-319-46448-0_2
8. Luque, R., Domínguez, E., Palomo, E., Muñoz, J.: A neural network approach for video object segmentation in traffic surveillance. Lecture Notes in Computer Science **5112 LNCS**, 151–158 (2008)
9. Molina-Cabello, M., Luque-Baena, R., López-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). DOI 10.3233/ICA-180577
10. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks (2015). DOI 10.48550/ARXIV.1506.01497
11. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning (2019). URL https://arxiv.org/abs/1905.11946
12. Wang, C., Dai, Y., Zhou, W., Geng, Y.: A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition. Journal of Advanced Transportation **2020**, 1–11 (2020). DOI 10.1155/2020/9194028
13. Youssef, Y., Elshenawy, M.: Automatic vehicle counting and tracking in aerial video feeds using cascade region-based convolutional neural networks and feature pyramid networks. Transportation Research Record: Journal of the Transportation Research Board **2675**(8), 304–317 (2021). DOI 10.1177/0361198121997833
14. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points (2019). URL https://arxiv.org/abs/1904.07850