A Driver Warning System Based on Anticipation of Ego-Involved Accident Objects

Daehyeon Jeong
Gwangju Institute and Science and Technology
Gwangju, Republic of Korea
astra0314@gm.gist.ac.kr

Hoosang Lee
Gwangju Institute and Science and Technology
Gwangju, Republic of Korea
hoosang223@gm.gist.ac.kr

Juseong Park
Gwangju Institute and Science and Technology
Gwangju, Republic of Korea
jusungpark13@gist.ac.kr

Jeha Ryu*
Gwangju Institute and Science and Technology
Gwangju, Republic of Korea
ryu@gist.ac.kr

Abstract

For safe driving, potential accidents must be warned before an actual accidents happen. We present an on-going work for a driver warning system using an ego-centric camera such as dashcam. We propose a novel deep learning model for anticipating ego vehicle involved accident object that may cause an accident in the future as early and accurately as possible. The proposed method predicts the accident probability of each object surrounding an ego-vehicle using hidden states from a future object localization model; this model uses only motion features such as a bounding box and its differential motion. The proposed method can thus reduce effects of visual biases in datasets, and thereby generalize well to unseen data. The effectiveness of the proposed method was demonstrated on a real vehicle.

1. Introduction

For safe driving, advanced driver assistance systems (ADASs) based on the anticipation of traffic accidents is required. However, learning to anticipate accidents is extremely challenging, because accidents are diverse and typically occur suddenly. Recent deep- and machine-learning-based studies have shown that anticipating [1, 2, 3, 4, 5, 6, 7] accidents with high probability is possible using only a first-person camera (e.g., dashcam) in a vehicle.

Ego-involved accidents usually occur abruptly and quickly in close proximity; hence, a quick and urgent response is required to avoid them. On the contrary, non-ego-involved accidents usually occur over larger distances; therefore, quick attention is required to avoid them. Preventing ego-involved accidents should be a priority for drivers and agents because they require quicker actions.

Notably, most existing studies [1, 2, 3, 4, 5, 6, 7] concentrate only on accident anticipation tasks, which can determine whether traffic accidents will occur in each frame of the incoming video stream. However, they cannot localize the object that poses a threat to the ego-vehicle. The accident risk factor in [3] anticipated not for a specific object but for object categories (vehicles, pedestrians, and bicycles), which cannot localize specific risk factor that threatens the ego-vehicle. To avoid near-future accidents of the ego-vehicle, accident- involved object anticipation task must localize the object because avoidance actions are primarily decided based on the location of the collision-related object.

Another important issue is the generalization capability. Farhan et al. [7] discussed dataset biases [8] and then showed that the high AP and TTC in [4, 6] resulted mainly from visual bias. Therefore, existing methods may not be generalizable to unseen data. In addition, considering immediate warning or automatic avoidance of impending accidents, the inference of the anticipation model must be processed online for the vehicle. Therefore, fast inference and reduced memory are the main requirements.

To solve the aforementioned limitations (only accident anticipation task) and disadvantages (lack of generalization and slow inference) of the existing methods, this paper proposes a novel deep-learning-based method with a dashcam to anticipate a traffic accident-involved object. The main contributions of this study are summarized as follows: (1) We propose a novel deep-learning-based traffic-accident-involved object anticipation model that predicts ego-involved accidents. (2) Unlike previous methods, the proposed model does uses only the concatenation of the

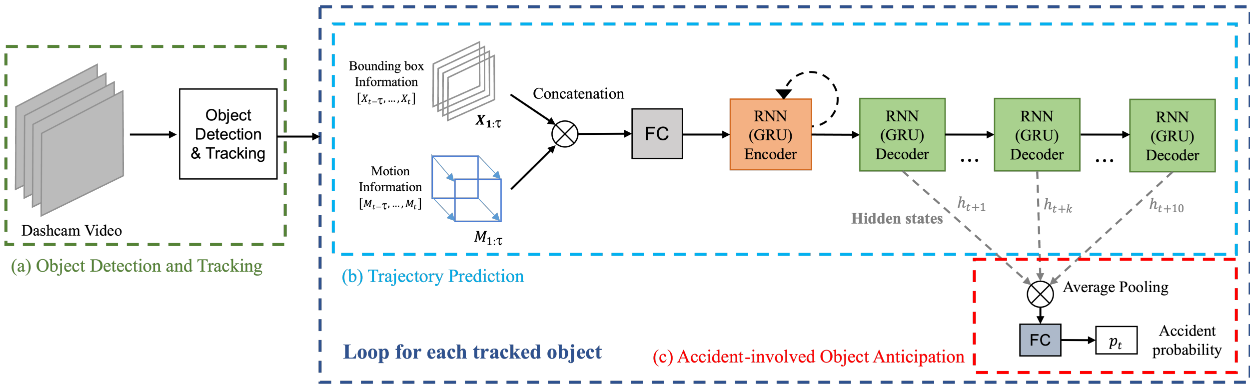

Figure 1. Overview of the proposed traffic accident-involved object anticipation framework.

bounding box information (location and size) and its differential motion information (velocity) of the detected objects. In addition, the ego-vehicle motion is not used. (3) We collect new traffic accident data from South Korea (KAD) to demonstrate the generalization capability of the proposed model to a dataset that is not introduced in the training process.

2. Method

Fig. 1 shows the deep learning framework for the proposed traffic accident-involved object anticipation model. It comprises three stages: (a) In the first stage, given dashcam videos, objects are detected as bounding boxes and then tracked. (b) In the second stage, the future locations of the detected and tracked objects are predicted individually using both the bounding box features and the differential motion of the bounding boxes. (c) In the third stage, the agent computes the accident probability of each object using the averaged hidden states in the trajectory-prediction stages. An ego-involved accident can thereby be predicted from the future tracking-related hidden states to the fully connected layer, outputting the accident probability.

Notably, the proposed architecture is similar to that in [9], where the object detector, tracker, and predictor for Future Object Localization (FOL) were used to detect accidents using performance metrics such as the frame-level Area Under the Curve (AUC). However, this method only detects accidents after the accident has occurred and does not predict accidents in advance because it detects the anomalies at the current time using several previous data, even though their algorithms may also find the most anomalous object in the frame.

There are three main differences between the proposed architecture and the method in [9]. (i) We have added a new anticipator model after FOL. (ii) We use differential motion from the detected objects instead of expensive and high-memory optical flows for the FOL to reduce memory and computation costs. (iii) We do not use ego-vehicle motion prediction, because previous numerical studies had shown minor differences [10], [11].

2.1. Trajectory Prediction

An object must be detected and tracked using appropriate methods to extract its future trajectory-related hidden states. Any object detectors such Mask R-CNN [8] and any trackers (e.g. deep SORT [12]) can be used to detect and track objects in a particular frame of a dashcam video. The bounding box X_t is composed of four parameters (location and size of the bounding box). In addition, two bounding boxes, X_{t-1} and X_t, of the tracked object are used to calculate the differential motion M_t at time t. Then, both are provided to the trajectory prediction model to predict the future trajectory of the tracked bounding box information $\hat{X}_{t,t+1}, ..., \hat{X}_{t,t+\delta}$, where δ is the number of predicted future frames and $\hat{X}_{t,t+i}$ is the predicted bounding box of the tracked object at time, (i = 1, 2, ..., δ) at current time.

2.2. accident-involved Object Anticipation

The proposed neural network model has been trained through supervised learning using both accident and non-accident videos. Because the model predicts the accident probability for each object, each object has either a positive label for an ego-involved accident object or a negative label for the others.

3. Experiments

3.1. Dataset for Accident-involved Object Anticipation

To train and evaluate the proposed model, three datasets were used: (1) The Detection of Traffic Anomaly dataset (DoTA) [13] from all over the world. This dataset comprises 4,000+ videos in various weather and lighting conditions and various types of traffic accidents involving both

Dataset	# of training video (accident)	# of test video (accident)
DoTA [13]	1,702(851)	738(369)
KAD*	-	898 (449)
CCD* [4]	-	112 (56)

Table 1. Datasets for training and testing

	Accident object-involved anticipation		Accident anticipation	
Architecture	mAP (%)	ATTC (sec)	mAP (%)	ATTC (sec)
Ours	82.18	0.969	90.57	1.114
DSA-RNN	71.62	1.232	78.03	2.535

Table 2. Comparison between architectures

ego-vehicle accidents and other accidents. (2) Similar to the DoTA, the CCD [4] comprises 1,500 accident videos and 3,000 non-accident videos from BDD100K [14]. (3) The new Korean Accident Dataset (KAD) is recently collected for 600+ videos of South Korean traffic environments from YouTube. This dataset is intended to test the generalization capability of the proposed model to unseen data in a specific country, such as South Korea. Testing the model that had been trained only on the DoTA without additional training shows a good generalization capability to other unseen datasets.

To avoid potential capture bias in the datasets, unlike the training and testing datasets in existing approaches [3, 4], we created non-accident datasets by extracting non-accident driving frames from only the non-ego-involved accident videos; the frames were extracted from the first frame to the frame just before the anomaly started. For data balancing, we used the same number of non-accident and ego- involved accident videos for training (1,702 (851 accident + 851 non-accident)) and testing (738 (369 accident + 369 non-accident)). Because the KAD consists of only ego-involved accident videos, during testing, we used 449 non-accident videos from BDD100K [15] and 449 accident videos from the KAD. In the test with CCD, we used a combination of 56 ego-involved accident videos and 56 non-accident videos that were randomly selected from the 486 non-accident videos in CCD [4]. Table I summarizes the datasets. All the video frames were extracted at 10 frames per second.

3.2. Implementation

The proposed accident-involved object anticipation model was trained using the Adam optimizer and AdaLEA [3] as a loss function for 100 training epochs, with a learning rate of 0.0001. A batch size of one was used because each tracked object has a different number of tracked frames. To select the best accident anticipation model during training, the harmonic mean between the AP and ATTC for the test data (DoTA) was used, because the two metrics are equally important performance measures. An NVIDIA RTX 3090 GPU and PyTorch were used for computation on Ubuntu 18.04. The inference time from object detection to the final accident anticipation was 84 ms (mostly consisting of the object detection time: 63.9 ms) for ten tracked objects. Ten objects surrounding an ego-vehicle is sufficiently representative of urban driving scenarios; thus, it can run in real time because the provided video stream has 10 frames per second.

3.3. Evaluation Metric

The proposed accident-involved object anticipation model predicts the probability (p_t in the range of 0–1) of a future accident in each frame for each object. The time-to-collision (TTC) is defined by the difference between the time of the accident and the time judged when the accident will occur. The TTC is obtained only for true positive (TP) data. If p_t is greater than a certain threshold, then an accident may occur after the predicted TTC; otherwise, an accident may not occur. To evaluate the performance of the binary classification problem, we used the average precision (AP). In addition, mAP is the mean of the AP for multiple object categories (vehicles, pedestrians, bicycles, etc.).

4. Results

The proposed approach for accident-involved object anticipation can also be applied to accident anticipation tasks in which the accident of an entire scene (a video clip) is anticipated, as in [1]. In the accident anticipation task, if the probability of any one object is larger than a preset threshold value, the video clip is judged as an accident.

Table II compares the two types of anticipation, in which the accident anticipation task showed both higher mAP and ATTC than the other. This is because accident anticipation is judged when there is any one object whose accident probability exceeds the preset threshold among the total detected objects, even though this object is not a true colliding object, whereas accident-involved object anticipation is judged when only a true colliding object whose accident probability exceeds the preset threshold. In other words, accident anticipation is determined by a greater number of true-positive counts. Table II also shows a comparison between the existing approach DSA-RNN and the proposed method, which shows significantly improved mAP but degraded ATTC because mAP and ATTC have a trad-off relationship.

Table III lists the generalization capabilities of the pro- posed method that was trained and tested on the DoTA dataset. The test results using KAD with the proposed method trained on DoTA showed slightly lower AP (73.56%) than the DoTA test, but the ATTC increased to

Dataset	AP (%)	ATTC (s)
DoTA [10]	82.18	0.969
KAD*	73.56	1.229
CCD* [4]	64.41	1.068

Table 3. Generalization capability

1.229 s. In the case of CCD, the AP was 64.41% with an ATTC of 1.068 s. In general, object detection and tracking are affected by the video (image) quality. The number of detected and tracked objects in low-quality images was smaller than that in relatively higher-quality images [15]. The performance difference among the datasets in Table III enforces this fact because the video quality (in terms of bitrate) is different: KAD (3989 kbps), CCD (916 kbps) and DoTA (3903 kbps). Therefore, a lower bit rate in the CCD may result in a lower performance.

5. Discussion

There are two typical failure cases: (1) low accident probability for a positive (accident-involved) object (worst case) and (2) high accident probability for a negative (non-accident-involved) object (false alarm). The reason for the worst-case (1) is that the positive object (with a high accident probability) was not detected instantaneously close to the time of the accident. Therefore, the probability of an accident decreases to zero. In contrast, the reason for false alarm case (2) is that, when the negative object closely crosses the front of the ego-vehicle or passes the side of the ego-vehicle at a close range, the training dataset states that these should be treated as accidents. However, in the real test data, there were no accidents (collisions) because the drivers may have performed collision avoidance actions (near-miss case). However, this false alarm may be helpful in careful driving. Because the proposed accident-involved object anticipation model does not consider the driver's action as an input and the actual collision cases are labeled as positive, these types of false alarms may not be predicted, which is a subject for future research.

Figure 2 shows a visual warning display when a warning threshold is over a prescribed value. For quicker warning, a sound warning (e.g. left, center, right) may also be issued independently without visual warning. We are now developing on-vehicle implementation by using NVIDIA NX module and are demonstrating on any real roads.

6. Conclusion and Future Work

This study proposes an ego-involved accident object anticipation method. Experimental results with actual datasets showed good performance, even without additional training for the unseen data, demonstrating the generalizability of the proposed approach. Future works will investigate anticipation of non-ego-vehicle-involved accidents as the proposed trajectory prediction and accident anticipation model processes the tracked objects one by one rather than all together, which may require more memory and processing time as well as a more complicated anticipation algorithm.

Figure 2. Visual warning in display.

References

[1] Fu-Hsiang Chan, Yu-Ting Chen, Yu Xiang, and Min Sun. Anticipating accidents in dashcam videos. In *Asian Conference on Computer Vision*, pages 136–153. Springer, 2016.

[2] K. H. Zeong, S. H. Chou, F. H. Chan, J. C. Niebles, and M. Sun. Agent-centric risk assessment: Accident anticipation and risky region localization. In *in Proc. IEEE CVPR, Honolulu, HI, USA*, pages 2222–2230, 2017.

[3] T. Suzuki, H. Kataoka, Y. Aoki, and Y. Satoh. Anticipating traffic accidents with adaptive loss and large-scale incident db. In *in Proc. IEEE CVPR, Salt Lake City, UT, USA*, page 3521–3529, 2018.

[4] W. Bao, Q. Yu, and Y. Kong. Uncertainty-based traffic accident anticipation with spatio-temporal relational learning. In *in Proc. 28th ACM Int. Conf. Multimedia, New York, NY, USA*, page 2682–2690, 2020.

[5] W. Bao, Q. Yu, and Y. Kong. Drive: Deep reinforced accident anticipation with visual explanation. In *in Proc. IEEE CVPR, Virtual*, page 7619–7628, 2021.

[6] M. M. Karim, R. Qin Y. Li, and Z. Yin. A dynamic spatial-temporal attention network for early anticipation of traffic accidents. 23(7), 2021.

[7] F. Mahmood, D. Jeong, and J. Ryu. A new approach to traffic accident anticipation with geometric features for better generalizability. *IEEE Access*, 11, 2023.

[8] K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask r-cnn. In *in Proc. IEEE ICCV, Venice, Italy*, page 2961–2969, 2017.

[9] T. Tommasi, B. Caputo N. Patricia, and T. Tuytelaars. A deeper look at dataset bias. In *in Proc. Pattern Recognit.: 37th German Conf., Aachen, Germany*, page 504–516, 2015.

[10] Y. Yao, M. Xu, Y. Wang, D. J. Crandall, and E. M. Atkins. Unsupervised traffic accident detection in first-person videos. *IEEE/RSJ IROS, Macau, China*, page 273–280, 2019.

[11] Y. Yao, M. Xu, C. Choi, D. J. Crandall, E. M. Atkins, and B. Dariush. Egocentric vision-based future vehicle localization for intelligent driving assistance systems. *IEEE/ICRA, Montreal, Canada*, page 9711–9717, 2019.

[12] N. Wojke, A. Bewley, and D. Paulus. Simple online and realtime tracking with a deep association metric. In *in Proc. IEEE ICIP, Beijing, China*, page 3645–3649, 2017.

[13] Y. Yao, X. Wang, M. Xu, Z. Pu, E. Atkins, and D. J. Crandall. When, where, and what? a new dataset for anomaly detection in driving videos. *arXiv:2004.03044*, 2020.

[14] F. Yu et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *in Proc. IEEE CVPR, Virtual*, page 2636–2645, 2020.

[15] H. Hsiang, K. C. Chen, P.Y. Li, and Y. Y. Chen. Analysis of the effect of automotive ethernet camera image quality on object detection models. In *in Proc. 2nd Int. Conf. Artif. Int. Inf. Commun,Fukuoka, Japan*, page 21–26, 2020.