SECURE: Stable Early Collision Understanding via Robust Embeddings in Autonomous Driving

Wenjing Wang[1](✉), Wenxuan Wang[2], and Songning Lai[3]

[1] Xiamen University Malaysia
[2] Xinjiang University
[3] HKUST(GZ)
DMT2309240@xmu.edu.my

Abstract. Deep learning has substantially advanced accident anticipation, but the robustness of these safety-critical systems under real-world perturbations remains a major challenge. We show that state-of-the-art models such as CRASH, despite their strong benchmark performance, can exhibit unstable predictions and latent representations under minor input perturbations, creating reliability risks in autonomous-driving scenarios. To address this problem, we introduce SECURE, Stable Early Collision Understanding via Robust Embeddings, a framework that formally defines and enforces robustness for accident anticipation. SECURE is built around four attributes: consistency and stability in both the prediction space and the latent feature space. We further propose a principled training strategy that fine-tunes a baseline model with a multi-objective loss, minimizing divergence from a reference model while penalizing sensitivity to adversarial perturbations. Experiments on the DAD and CCD datasets show that SECURE improves robustness under multiple perturbation settings while also improving clean-data performance.

Keywords: Accident Anticipation, Adversarial Robustness, Autonomous Driving, Robust Embeddings, Spatio-temporal Learning.

1 Introduction

The rapid development of deep learning has reshaped modern intelligent transportation systems (ITS) [1], [2], enabling vehicles and infrastructure to perceive complex environments, anticipate risks, and support timely decision-making. As ITS moves toward higher levels of automation, robustness and safety become foundational requirements rather than optional enhancements [3], [4]. Among the modules that support intelligent transportation, accident anticipation is especially safety-critical because predicting risk seconds before a hazardous event requires both accurate recognition and stable behavior under real-world disturbances [5]-[8].

However, reliable accident anticipation remains unresolved [9]-[11]. Beyond prediction accuracy, modern ITS must maintain stable outputs across a wide range of environmental changes. In high-stakes scenarios, minor perturbations, including illumination shifts, sensor noise, temporal jitter, and sampling irregularities, can cause substantial fluctuations in model predictions [4], [12]. Such instability threatens system reliability, complicates safety verification, and limits dependable operation in unconstrained environments [3].

Crash Recognition and Anticipation System Harnessing (CRASH), which integrates context-aware and temporal focus mechanisms, has recently become a strong representative model for ITS-related video risk forecasting [13]. CRASH achieves state-of-the-art performance on benchmark datasets and can identify hazardous patterns early in temporal sequences [13].

Although CRASH performs well under standard benchmark settings, our analysis reveals a limitation that has received insufficient attention: CRASH is unstable in both its predictive outputs and its latent representations. Minor input perturbations, such as sensor noise or small frame-level corruptions, can distort the representation space and shift the predicted probability trajectories (see Fig. 1). This instability can delay or destabilize time-to-accident estimates, raising safety concerns for intelligent transportation applications.

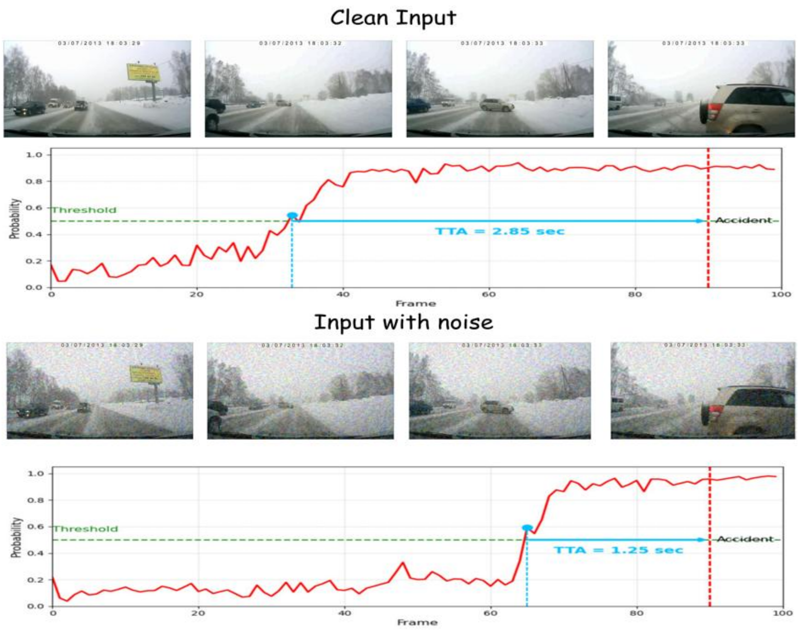

Fig. 1. CRASH produces different prediction trajectories for clean and perturbed inputs.

To address these challenges, we introduce SECURE, Stable Early Collision Understanding via Robust Embeddings, and propose a robustness analysis framework for accident anticipation. The framework systematically evaluates and improves the reliability of CRASH under noise perturbations by defining four core robustness properties.

These properties characterize the consistency and stability of predictions and latent representations. On this basis, we conduct a robustness analysis of CRASH and

SECURE, showing that the proposed framework improves stability under perturbations and provides a basis for further model optimization. The main contributions are as follows:

(1) Comprehensive Robustness Assessment: We conduct an in-depth robustness analysis of CRASH and identify reliability weaknesses that make its predictions sensitive to input and parameter perturbations.

(2) Formal Definition of SECURE: We define SECURE as a reliability-oriented accident anticipation framework and specify the criteria used to evaluate and improve robustness.

(3) Robust SECURE Construction Framework: We propose a training framework that improves robustness while preserving predictive performance by fine-tuning the parameters of the original model. Experiments validate the effectiveness of this framework.

2 Related Work

Traffic accident anticipation depends on whether a model can extract spatio-temporal features from complex and dynamic traffic scenes [3]. Early methods primarily used convolutional neural networks (CNNs) to extract visual features from dashcam footage [5], [14], [15], often combining them with sequential models such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks to preserve temporal coherence [16]-[18]. More recent studies introduce graph neural networks (GNNs) [8], [11], [19] and Transformer-based architectures [1], [20], [21] to model interactions among road users and capture spatial dynamics more effectively.

Because not all visual regions contribute equally to accident risk, attention mechanisms have become a standard component of accident anticipation models. Chan et al. [5] introduced dynamic soft attention to fuse object-level and frame-level features, while Karim et al. [7], [11] developed spatio-temporal attention networks to prioritize hazardous regions. Zeng et al. [22] used attention RNNs to localize potential danger zones. Interpretability has also been explored: Monjuru et al. [23] embedded Grad-CAM into a GRU to generate semantic feature maps, and Bao et al. [10], [24] studied generative models and visual explanations to improve transparency.

Despite strong benchmark accuracy, robustness in open-world ITS remains a critical but understudied challenge. Traffic environments require models to handle routine dynamic variation as well as rare events, sensor noise, and adversarial interference. Ma et al. [15] highlight the difficulty of autonomous-driving perception under unknown categories and rare events. The wide variability of traffic scenes [3] also creates long-tail distribution issues, where models trained on common data can fail to generalize to unusual anomalies [25]. Although some studies model uncertainty, such as the perceptual uncertainty graph of Bao et al. [9], most methods still assume relatively clean data distributions and do not rigorously stress-test environmental perturbations.

The safety-critical nature of accident prediction further requires resistance to interference. Han et al. [4] demonstrated physical backdoor attacks on lane detection systems, showing that subtle physical perturbations can mislead deep perception models.

However, most accident prediction frameworks focus on dynamic-object interactions [2], [7] or loss functions for early detection [6], while paying limited attention to visual noise, adversarial attacks, and domain shifts. As a result, the field still lacks a unified framework for benchmarking robustness and guiding mitigation under noise, attacks, and domain shift. Our work addresses this gap by evaluating and improving prediction stability and fidelity under perturbations.

3 Methods

3.1 Preliminaries: CRASH

CRASH is an accident anticipation model that predicts accident probabilities from dashboard videos. Given an input video $V = \{V_1, \ldots, V_T\}$, it outputs a frame-wise probability sequence $P = \{p_1, \ldots, p_T\}$, where $p_t \in [0,1]$ denotes the accident probability at frame t.

Object and Context Features. A Cascade R-CNN is used to detect the top-n dynamic traffic agents (e.g., vehicles, pedestrians, cyclists). Their cropped regions are embedded by a VGG-16 backbone into object features $F_o \in \mathbb{R}^{n \times d}$. In parallel, a feature extractor encodes the full frame into global context features $F_c \in \mathbb{R}^d$, capturing lane markings, road layout, traffic signs, and other scene-level cues outside bounding boxes.

Object-aware and Context-aware Modules. The object-aware module enhances the detected object features F_o through an Object Focus Attention (OFA) mechanism, producing refined object-aware features $\tilde{F}_o$. In parallel, the context-aware module transforms the frame-level context features F_c into enhanced context-aware features $\tilde{F}_c$ through a lightweight feature refinement network.

Temporal Fusion and Prediction. Temporal features are obtained by concatenating the outputs of the previous modules and encoding them with a dual-layer GRU, producing the temporal hidden state h_t. An MLP then maps h_t to the accident probability p_t.

Training. The loss function combines an anticipation loss $\mathcal{L}_a$ and an enhancement loss $\mathcal{L}_e$. The anticipation loss encourages early prediction by applying a time-dependent penalty $e^{-\frac{1}{2}\max\left(\frac{\tau - t}{f}, 0\right)}$ to positive frames. It is defined as:

$$\mathcal{L}_a = \frac{1}{B} \sum_{v=1}^{B} \left[-l_v \sum_{t=1}^{T} e^{-\frac{1}{2}\max\left(\frac{\tau - t}{f}, 0\right)} \log(p_t) - (1 - l_v) \sum_{t=1}^{T} \log(1 - p_t) \right] \quad (1)$$

where B is the batch size, $l_v \in \{0,1\}$ represents the binary label of accident occurrence within each video (1 for an accident, 0 for none), T is the total number of frames per video, τ is the ground-truth accident time, f is the frames per second (fps) of the video, and p_t is the predicted accident probability at frame t.

To stabilize temporal representations, the enhancement loss $\mathcal{L}_e$ supervises an auxiliary prediction p_e.

$$p_e = \phi_{\mathrm{MLP}} \left(\phi_{\mathrm{MHA}} \left(\phi_{\mathrm{PE}} (h_1^2, h_2^2, \ldots, h_T^2) \right) \right) \quad (2)$$

where ϕ_{PE} denotes the position encoding mechanism, ϕ_{MHA} denotes the multi-head self-attention mechanism, and h_t^2 is the hidden state of the second-layer GRU at frame t.

This auxiliary prediction is derived from position-encoded second-layer GRU hidden states through a multi-head self-attention block and an MLP. The enhancement loss is defined as:

$$\mathcal{L}_e = \frac{1}{B}\sum_{v=1}^{B}[-l_v \log(p_e) - (1 - l_v)\log(1 - p_e)] \tag{3}$$

The overall loss combines $\mathcal{L}_a$ and $\mathcal{L}_e$ with homoscedastic uncertainty weighting:

$$\mathcal{L}_{Task} = \frac{\mu_1}{2\rho_1^2}\mathcal{L}_a + \frac{\mu_2}{2\rho_2^2}\mathcal{L}_e + \log(\rho_1\rho_2) \tag{4}$$

where μ_1, μ_2 are hyperparameters and ρ_1, ρ_2 are learnable uncertainty coefficients initialized to 1. This formulation jointly optimizes early accident anticipation and auxiliary temporal representation quality while adaptively balancing the two loss terms.

3.2 Robustness Issues

CRASH demonstrates state-of-the-art performance on video-based accident anticipation tasks. However, its sensitivity to input perturbations and parameter variations exposes reliability weaknesses. These weaknesses appear as unstable probability outputs, inconsistent attention to key agents, and fluctuating accuracy and time-to-accident estimates even when the input sequence changes only slightly.

This sensitivity undermines the reliability of CRASH in real-world autonomous-driving scenarios, where sensor noise, camera jitter, and environmental changes are unavoidable. The model's decisions may therefore deviate from the underlying visual evidence, reducing interpretability and credibility in safety-critical applications. By analyzing CRASH, we develop a general approach for improving reliability and robustness in video-based accident prediction.

3.3 The Concept of SECURE

In high-stakes accident anticipation, a SECURE framework denotes a model's ability to remain close to an optimal decision manifold while preserving invariance under perturbations. This notion emphasizes not only risk-assessment accuracy but also resilience to sensory noise and semantic alignment in latent representations. Enforcing these properties stabilizes the anticipation process and makes risk evolution less sensitive to environmental fluctuations.

To address robustness in accident anticipation, we propose that a SECURE framework should include four fundamental attributes organized across the output space and the latent space:

Consistency in Prediction Space (Cps). This attribute measures the alignment between the model's risk estimate and a theoretical optimal reference. Let $f(x;\theta)$ denote the trainable model's prediction for input x, and $f^*(x)$ denote the frozen model prediction. To satisfy this attribute, the divergence $D_{out}\big(f(x;\theta), f^*(x)\big)$ must be bounded by a constant γ_1. This constraint helps SECURE preserve the high-fidelity decision capability of CRASH.

Stability in Prediction Dynamics (Spd). This attribute measures prediction robustness under input perturbations without relying on external anchors. It requires SECURE to maintain intrinsic self-consistency. For an input x and its perturbed counterpart $x+\delta$, the forecasts must converge. The difference, measured by a metric D_{out}, must be bounded by γ_2.

Consistency in Latent Manifold (Clm). Robustness should also hold at the representation level. This attribute requires the high-dimensional features extracted by SECURE, $h(x;\theta)$, to align semantically with the CRASH feature manifold $h^*(x)$. The divergence $D_{feat}\big(h(x;\theta), h^*(x)\big)$ must not exceed a threshold β_1. This constraint encourages the model to capture semantic cues, such as motion patterns, rather than fitting noise.

Stability in Latent Dynamics (Sld). This attribute requires internal feature extraction to remain invariant to noise. The latent representation $h(x;\theta)$ and the perturbed representation $h(x+\delta;\theta)$ must stay within a compact neighborhood in the latent space. This constraint prevents minor input noise from amplifying into divergent semantic features.

3.4 Formal Definition of SECURE

Based on these attributes, we provide a formal definition. A CRASH model is considered $(\gamma_1, \gamma_2, \beta_1, \beta_2, \epsilon)$-SECURE if it satisfies the following four criteria for any input feature sequence $x = \{x_1, \ldots, x_T\} \in \mathcal{X}$ with target SECURE weights θ:

Consistency in Prediction Space (C_{ps}):

$$D_{out}\big(f(x;\theta), f^*(x)\big) \le \gamma_1 \tag{5}$$

Stability in Prediction Dynamics (S_{pd}):

$$D_{out}\big(f(x;\theta), f(x+\delta;\theta)\big) \le \gamma_2, \forall \| \delta \| \le \epsilon \tag{6}$$

Consistency in Latent Manifold (C_{lm}):

$$D_{feat}\big(h(x;\theta), h^*(x)\big) \le \beta_1 \tag{7}$$

Stability in Latent Dynamics (S_{ld}):

$$D_{feat}\big(h(x;\theta), h(x+\delta;\theta)\big) \le \beta_2, \forall \| \delta \| \le \epsilon \tag{8}$$

where D_{out} and D_{feat} represent distance metrics, namely Mean Squared Error, $\|\cdot\|$ denotes a norm constraint on the perturbation δ, and $\epsilon \ge 0$ defines the robustness radius.

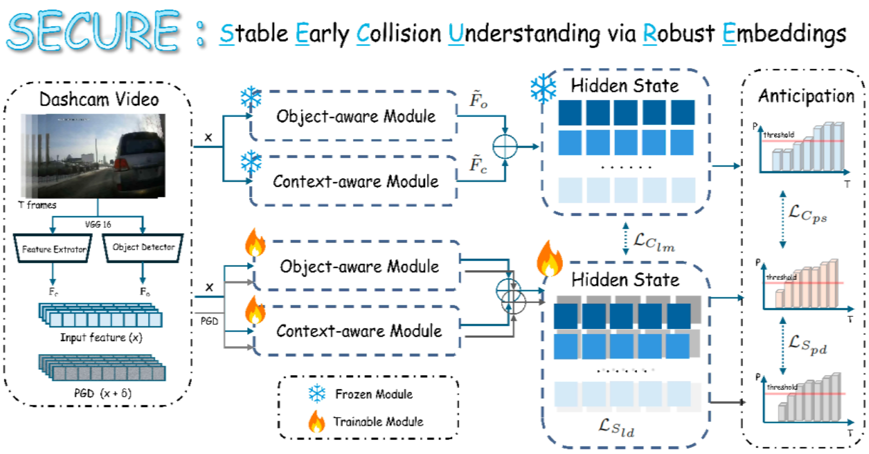

Fig. 2. Overview of the SECURE framework.

3.5 Constructing a SECURE Framework

To construct a $(\gamma_1, \gamma_2, \beta_1, \beta_2, \epsilon)$-robust accident anticipation model from CRASH, we formulate a multi-objective optimization problem aligned with the attributes defined above and illustrated in Fig. 2. The resulting objective balances four aspects: Consistency in Prediction Space (C_{ps}), Stability in Prediction Dynamics (S_{pd}), Consistency in Latent Manifold (C_{lm}), and Stability in Latent Dynamics (S_{ld}).

Given the definitions of C_{ps}, S_{pd}, C_{lm}, and S_{ld}, we aim to find the optimal model parameters θ that satisfy the following optimization problem. Formally,

$$\min_{\theta} \left[\mathcal{L}_{C_{ps}}(f(\cdot)) + \mathcal{L}_{S_{pd}}(f(\cdot)) + \mathcal{L}_{C_{lm}}(h(\cdot)) + \mathcal{L}_{S_{ld}}(h(\cdot)) \right] \tag{9}$$

Each loss term corresponds to one attribute of the SECURE framework:

$$\mathcal{L}_{C_{ps}}(f(\cdot)) = \mathbb{E}_x\left[D_{\text{out}}\left(f(x;\theta), f^*(x)\right)\right], \tag{10}$$

$$\mathcal{L}_{S_{pd}}(f(\cdot)) = \mathbb{E}_{x,\delta}\left[D_{\text{out}}\left(f(x;\theta),\ f(x+\delta;\theta)\right)\right], \tag{11}$$

$$\mathcal{L}_{C_{lm}}(h(\cdot)) = \mathbb{E}_x\left[D_{\text{feat}}\left(h(x;\theta), h^*(x)\right)\right], \tag{12}$$

$$\mathcal{L}_{S_{ld}}(h(\cdot)) = \mathbb{E}_{x,\delta}\left[D_{\text{feat}}\left(h(x;\theta),\ h(x+\delta;\theta)\right)\right] \tag{13}$$

where D_{out} and D_{feat} are measures of divergence in the output and feature spaces, respectively. The perturbation δ is constrained by $\| \delta \| \leq \epsilon$, and $\mathbb{E}_{x,\delta}$ denotes the expectation over both the input space and the perturbation space. $f^*(x)$ and $h^*(x)$ represent the mappings of the frozen CRASH model, while $f(x;\theta)$ and $h(x;\theta)$ denote the predicted output and hidden state representations of SECURE. To achieve this, we incorporate the following loss terms into the optimization process:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Task}} + \lambda_c^{out}\mathcal{L}_{C_{ps}} + \lambda_s^{out}\mathcal{L}_{S_{pd}} + \lambda_c^{feat}\mathcal{L}_{C_{lm}} + \lambda_s^{feat}\mathcal{L}_{S_{ld}} \tag{14}$$

where $\mathcal{L}_{\text{Task}}$ is the base loss function of CRASH, and the λ terms are regularization coefficients that control the influence of each auxiliary robustness term.

Inspired by the Projected Gradient Descent (PGD) method described by Madry et al. [12], we iteratively update the perturbation δ and the model parameters θ so that the model remains robust to adversarial perturbations while maintaining consistency in both prediction and latent spaces. The iterative process for the p-th update is defined as follows.

$$\delta_p = \delta_{p-1}^* + \frac{\alpha_p}{|\mathcal{B}_{p-1}|} \sum_{x \in \mathcal{B}_{p-1}} \nabla_{\delta_{p-1}^*} \left(\mathcal{L}_{C_{ps}} + \mathcal{L}_{S_{pd}} + \mathcal{L}_{C_{lm}} + \mathcal{L}_{S_{ld}} \right),$$

where

$$\delta_p^* = \arg\min_{\|\delta\| \le \epsilon} \| \delta - \delta_p \|$$

projects the perturbation back into the ϵ-ball to satisfy the norm constraint. Here, $\mathcal{B}_{p-1}$ denotes the batch of samples at iteration $p-1$, and α_p is the PGD step size.

After obtaining the worst-case perturbation δ^* after P iterations, the model parameters are updated from θ^{t-1} to θ^t using batched gradients derived from the total objective $\mathcal{L}_{\text{Total}}$. This update encourages consistency in prediction space, stability in prediction dynamics, consistency in the latent manifold, and stability in latent dynamics, yielding more dependable accident anticipation across varying conditions.

4 Experiments

4.1 Experimental Setup

Datasets and Baselines. We evaluate the framework on two benchmark datasets, the Dashcam Accident Dataset (DAD) [5] and the Car Crash Dataset (CCD) [9], following their standard training and testing splits. We compare SECURE with the state-of-the-art accident anticipation baseline CRASH to evaluate the effectiveness of the proposed consistency and stability constraints.

Implementation Details. We use Projected Gradient Descent (PGD) attacks. The attack hyperparameters are configured as follows: perturbation budget $\epsilon = 0.01$, step size $\alpha = 0.002$, and $P = 20$ attack iterations. All experiments are conducted on an NVIDIA GeForce RTX 4090 GPU. We train the model using the Adam optimizer with an initial learning rate of 1×10^{-4} and a batch size of 10. The hyperparameters for the robustness loss terms are set as follows: output stability $\lambda_s^{\text{out}} = 50.0$, output consistency $\lambda_c^{\text{out}} = 50.0$, and feature-level $\lambda_s^{feat} = \lambda_c^{feat} = 0.01$.

Table 1. Performance comparison on the complete datasets. Bold and underlined values indicate the best and second-best results in each category, respectively.

Model	DAD [5]		CCD [9]	
	AP(%)↑	mTTA(s)↑	AP(%)↑	mTTA(s)↑
DSA [5]	48.1	1.34	99.6	4.53
L-RAI [22]	51.4	3.01	98.9	3.32
AdaLEA [6]	62.3	2.94	99.2	3.45
DSTA [7]	59.2	2.60	98.8	4.53
UString [9]	58.0	2.85	99.5	4.52
GSC [19]	60.4	2.55	99.3	3.58
CRASH [13]	67.2	3.02	99.5	4.55
SECURE	**68.7**	**3.13**	**99.7**	**4.58**

4.2 Performance Results

Table 1 reports the test results of SECURE and the baseline CRASH without added noise perturbations. Following the CRASH evaluation protocol, we use Average Precision (AP) and Mean Time-to-Accident (mTTA) as the performance metrics. AP measures prediction quality across decision thresholds by computing the area under the precision-recall curve. mTTA measures timeliness by calculating the average interval between the first risk prediction above a predefined threshold and the actual accident time. A higher mTTA indicates earlier accident anticipation and more available reaction time.

As shown in Table 1, SECURE outperforms CRASH on clean data without added perturbations. This improvement suggests that real-world driving data already contain inherent noise, and that reducing the effect of such noise during training can improve prediction quality. Importantly, SECURE does not trade clean-data performance for robustness; instead, it improves performance while maintaining stronger stability.

4.3 Robustness Results

To evaluate robustness, we introduce two perturbation types during testing:

(1) Input Perturbation. Gaussian noise is added to each input frame-level feature and expressed as $x_t{}' = x_t + \mathcal{N}(0, \sigma)$.

(2) Latent Perturbation. Because the second-layer GRU produces the model's key temporal latent representation, Gaussian noise is injected into its parameters $\theta_{\mathrm{GRU2}}{}' = \theta_{\mathrm{GRU2}} + \mathcal{N}(0, \sigma)$.

These perturbations are applied during evaluation to assess model stability under external input noise and internal temporal-module disturbance. Table 2 reports the performance of the baseline model and SECURE under both input perturbation and intermediate-layer perturbation settings.

Table 2. Comparison of CRASH and SECURE on the DAD and CCD datasets under different perturbations. Values in parentheses denote the standard deviation of Gaussian noise.

Dataset	Perturba-tion	Model	AP (%) ↑	mTTA (s) ↑
DAD	**Clean**	CRASH	67.21	3.020
		SECURE	**68.67**	**3.127**
	IP (0.1)	CRASH	65.83	3.053
		SECURE	**68.42**	**3.120**
	IP (0.2)	CRASH	63.37	2.964
		SECURE	**68.32**	**3.062**
	LP (0.1)	CRASH	64.41	2.984
		SECURE	**68.19**	**3.083**
	LP (0.2)	CRASH	44.73	3.119
		SECURE	**57.61**	**3.207**
CCD	**Clean**	CRASH	99.53	4.554
		SECURE	**99.73**	**4.583**
	IP (0.1)	CRASH	99.29	4.540
		SECURE	**99.67**	**4.483**
	IP (0.2)	CRASH	99.17	3.965
		SECURE	**99.48**	**4.481**
	LP (0.1)	CRASH	98.88	3.074
		SECURE	**99.29**	**3.191**
	LP (0.2)	CRASH	69.26	3.413
		SECURE	**81.45**	**3.430**

Under input perturbation (IP), the performance of CRASH decreases substantially, indicating that input noise affects its predictive precision. On DAD, AP decreases from 67.21 to 63.37. On CCD, mTTA decreases from 4.554 to 3.965. Under latent perturbation (LP), the degradation is larger, especially for AP: CRASH drops from 67.21 to 44.73 on DAD and from 99.53 to 69.26 on CCD.

Compared with the larger performance fluctuations of CRASH under perturbations, SECURE behaves more stably. In AP, SECURE shows smaller degradation across both input and latent perturbation settings, while generally achieving higher mTTA than CRASH. These results indicate that the PGD-based adversarial optimization in SECURE improves robustness to both external feature noise and internal parameter disturbances without sacrificing accuracy.

4.4 Ablation Study

Table 3. Ablation results of SECURE on the DAD dataset under IP (0.2).

$\mathcal{L}_{Task}$	$\mathcal{L}_{C_{ps}}$	$\mathcal{L}_{S_{pd}}$	$\mathcal{L}_{C_{lm}}$	$\mathcal{L}_{S_{ld}}$	**AP(%) ↑**	**mTTA(s)↑**
✓	×	×	×	×	63.37	2.964
✓	✓	×	×	×	63.47	3.036
✓	✓	✓	×	×	64.80	3.015
✓	✓	✓	✓	×	65.29	3.001
✓	✓	✓	✓	✓	**68.32**	**3.062**

In the ablation study, we evaluate each component in Equation (14) to assess its contribution to model performance. We designate the first term, L_{Task}, as the principal loss and then evaluate different configurations of L_{Cps}, L_{Spd}, L_{Clm}, and L_{Sld} by selectively omitting these regularization terms.

The results in Table 3 show that each regularization term in the objective function contributes to model performance. Each component provides a distinct robustness benefit, and removing any term weakens the overall framework.

5 Conclusion

In this paper, we address the under-explored problem of robustness in accident anticipation models. We introduce SECURE, a principled framework that formalizes model reliability through four attributes: consistency and stability in both prediction and latent spaces. To construct a SECURE model, we propose a multi-objective adversarial optimization scheme that reduces sensitivity to perturbations while preserving prediction fidelity. Experiments on the DAD and CCD datasets show that SECURE improves stability under perturbations and achieves strong performance on clean data. This work provides a foundation for more trustworthy accident prediction systems and a practical framework for evaluating robustness in safety-critical autonomous-driving modules.

References

1. Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., Feichtenhofer, C.: Multiscale vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6824-6835 (2021)
2. Liao, H., Shen, H., Li, Z., Wang, C., Li, G., Bie, Y., Xu, C.: GPT-4 enhanced multimodal grounding for autonomous driving: Leveraging cross-modal attention with large language models. Communications in Transportation Research 4, 100116 (2024)
3. Li, Z., Liao, H., Tang, R., Li, G., Li, Y., Xu, C.: Mitigating the impact of outliers in traffic crash analysis: A robust bayesian regression approach with application to tunnel crash data. Accident Analysis & Prevention 185, 107019 (2023)
4. Han, X., Xu, G., Zhou, Y., Yang, X., Li, J., Zhang, T.: Physical backdoor attacks to lane detection systems in autonomous driving. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 2957-2968 (2022)

5. Chan, F.-H., Chen, Y.-T., Xiang, Y., Sun, M.: Anticipating accidents in dashcam videos. In: Computer Vision - ACCV 2016, pp. 136-153. Springer (2017)
6. Suzuki, T., Kataoka, H., Aoki, Y., Satoh, Y.: Anticipating traffic accidents with adaptive loss and large-scale incident DB. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3521-3529 (2018)
7. Karim, M.M., Li, Y., Qin, R., Yin, Z.: A dynamic spatial-temporal attention network for early anticipation of traffic accidents. IEEE Transactions on Intelligent Transportation Systems 23(7), 9590-9600 (2022)
8. Thakur, N., Gouripeddi, P., Li, B.: Graph(graph): A nested graph-based framework for early accident anticipation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 7533-7541 (2024)
9. Bao, W., Yu, Q., Kong, Y.: Uncertainty-based traffic accident anticipation with spatio-temporal relational learning. In: Proceedings of the 28th ACM International Conference on Multimedia (MM '20) (2020)
10. Bao, W., Yu, Q., Kong, Y.: DRIVE: Deep reinforced accident anticipation with visual explanation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 7619-7628 (2021)
11. Karim, M.M., Yin, Z., Qin, R.: An attention-guided multistream feature fusion network for early localization of risky traffic agents in driving videos. IEEE Transactions on Intelligent Vehicles (2023)
12. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. In: International Conference on Learning Representations (2018)
13. Liao, H., Sun, H., Shen, H., Wang, C., Tian, C., Tam, K.L., Li, Z., et al.: CRASH: Crash recognition and anticipation system harnessing context-aware and temporal focus attentions. In: Proceedings of the 32nd ACM International Conference on Multimedia, pp. 11041-11050 (2024)
14. Fang, J., Qiao, J., Bai, J., Yu, H., Xue, J.: Traffic accident detection via self-supervised consistency learning in driving scenarios. IEEE Transactions on Intelligent Transportation Systems 23(7), 9601-9614 (2022)
15. Ma, Z., Yang, Y., Wang, G., Xu, X., Shen, H.T., Zhang, M.: Rethinking open-world object detection in autonomous driving scenarios. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 1279-1288 (2022)
16. Fatima, M., Khan, M.U.K., Kyung, C.-M.: Global feature aggregation for accident anticipation. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 2809-2816. IEEE (2021)
17. Takimoto, Y., Tanaka, Y., Kurashima, T., Yamamoto, S., Okawa, M., Toda, H.: Predicting traffic accidents with event recorder data. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility, pp. 11-14 (2019)
18. Ye, M., Peng, X., Gan, W., Wu, W., Qiao, Y.: AnoPCN: Video anomaly detection via deep predictive coding network. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 1805-1813 (2019)
19. Wang, T., Chen, K., Chen, G., Li, B., Li, Z., Liu, Z., Jiang, C.: GSC: A graph and spatio-temporal continuity based framework for accident anticipation. IEEE Transactions on Intelligent Vehicles (2023)
20. Feng, J.-C., Hong, F.-T., Zheng, W.-S.: MIST: Multiple instance self-training framework for video anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14009-14018 (2021)

21. Li, K., Wang, Y., He, Y., Li, Y., Wang, Y., Wang, L., Qiao, Y.: UniFormerV2: Spatiotemporal learning by arming image ViTs with video uniformer. arXiv preprint arXiv:2211.09552 (2022)
22. Zeng, K.-H., Chou, S.-H., Chan, F.-H., Niebles, J.C., Sun, M.: Agent-centric risk assessment: Accident anticipation and risky region localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
23. Karim, M.M., Li, Y., Qin, R.: Towards explainable artificial intelligence (XAI) for early anticipation of traffic accidents. arXiv e-prints, arXiv:2108 (2021)
24. Yao, Y., Wang, X., Xu, M., Pu, Z., Wang, Y., Atkins, E., Crandall, D.: DOTA: Unsupervised detection of traffic anomaly in driving videos. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
25. Thakare, K.V., Dogra, D.P., Choi, H., Kim, H., Kim, I.-J.: RareAnom: A benchmark video dataset for rare type anomalies. Pattern Recognition 140, 109567 (2023)