Progressive Risk Estimation for Accident Anticipation

Samet Hicsonmez[1,*] **Eray Çakar**[2,*] **Nermin Samet**[3,†] **Fatma Güney**[2,†]

[1]Independent Researcher, Paris, France
[2]Department of Computer Engineering, KUIS AI, Koç University, Istanbul, Turkey
[3]Valeo.ai, Paris, France

Abstract

Accident anticipation aims to recognize anomalous driving cues before a crash while avoiding false alarms during normal driving. Existing approaches typically formulate this task as binary classification, focusing on whether an accident will occur rather than *when* it will occur. We propose `PRE-ACT`, a framework that models accident risk as a continuously evolving signal that increases as the crash approaches. By explicitly enforcing temporal ordering and distance-to-accident awareness, our method progressively raises risk while suppressing premature alarms, leading to significant improvements on MM-AU subsets and Nexar. We further introduce a Separation Score to evaluate the global behavior of predicted risk curves beyond local temporal windows. Code and pretrained models are available at https://github.com/giddyyupp/PRE-ACT.

1 Introduction

Despite recent advances in autonomous driving [1–3], deploying these systems reliably in the real world remains challenging due to the lack of preventive mechanisms in safety-critical scenarios [4–8]. In this paper, we take a safety-first approach and propose a method to anticipate accidents early enough to prevent them. Given videos capturing the moments leading up to an accident, our goal is to develop an online monitoring system that recognizes early anomalous cues and issues warnings at each time step. Beyond timely and accurate detection of anomalous cues, a key challenge is minimizing false alarms to ensure the system remains reliable and practical for real-world deployment.

Accidents are inherently dynamic events that unfold over time, requiring models to capture spatio-temporal dependencies and extract meaningful representations from video. Prior work has addressed this by incorporating temporal modeling through recurrent architectures [9], graph-based representations [10], or visual attention cues [11]. However, most existing approaches ultimately formulate accident anticipation as a binary classification problem, typically at the frame level or over a short window [11], focusing on whether an accident will occur rather than *when* it will occur (Fig. 1).

In contrast, temporal proximity to an accident is a critical signal in anticipation. As an event approaches, subtle cues intensify and evolve, and correctly interpreting this progression requires understanding the temporal ordering and relative distance of frames to the approaching event. Modeling this temporal structure explicitly is therefore essential for timely and reliable accident anticipation. Motivated by this, we propose to model the increasing risk as an accident approaches and introduce a preference over video segments based on their temporal proximity to the event (Fig. 2).

Our method, `PRE-ACT`, advances the state of the art on commonly used accident anticipation datasets, including the CAP and DADA subsets of MM-AU [12]. It improves the mean AUC by over 10% on

*Equal contribution.
†Equal contribution for senior authorship.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

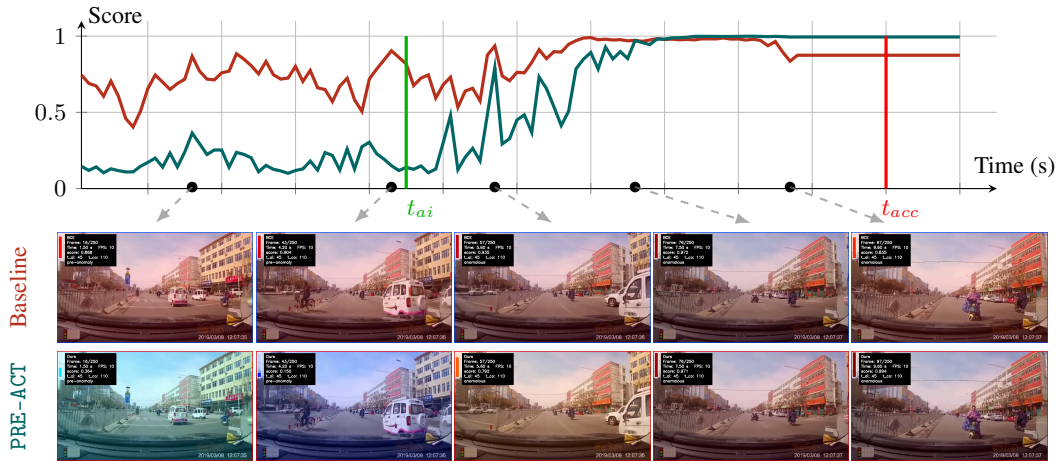

Figure 1: PRE-ACT**: From binary accident prediction to progressive risk anticipation.** Unlike the dominant paradigm of formulating accident anticipation as a binary classification of whether an accident will occur (baseline), we model risk as a temporally evolving signal that reflects *when* an accident is likely to occur. By explicitly enforcing temporal order and distance-to-crash awareness, our model reduces false alarms while progressively increasing risk as abnormal cues emerge. Scores are overlaid on frames as an alpha channel, with blue indicating safe and red indicating risky predictions.

CAP and nearly 30% on DADA, anticipates crashes 0.25 and 0.18 seconds earlier than prior methods, respectively, and surpasses the winner of the Nexar challenge [13] by +1 mAP.

Since standard metrics such as AUC, mean time-to-accident (mTTA), and mAP focus on local evaluation windows and may overlook the global behavior of the predicted risk curve, e.g., false alarms during normal driving, we introduce a novel separation score that evaluates global risk behavior by penalizing early false positives and encouraging higher risk after the anomaly onset.

In summary, our contributions are two-fold: **(i)** We reformulate accident anticipation as explicit continuous risk estimation, leading to significant improvements across all benchmarks and metrics. **(ii)** We identify a key limitation of existing evaluation metrics and introduce a novel Separation Score to capture the global risk behavior of anticipation models.

2 Related work

Accident anticipation. Accident anticipation from videos has been framed as an online video classification or anomaly detection problem. Early approaches relied on RNNs and temporal attention mechanisms to process sequential frames, predicting the likelihood of an accident [9, 14–19].

To capture the dynamic interactions that cause accidents, subsequent research [20–23] has focused heavily on spatiotemporal representation learning for accident anticipation. Graph-based methods, such as GSC [10] and Graph [24], explicitly model the dynamic interactions between the ego-vehicle and surrounding agents using dynamically updated adjacency matrices. Other works have sought to enrich the input space; for example, by adding 3D information from monocular depth [25]. Additionally, recent methods CAP [26], CRASH [27], and Mind-the-Gap [28] integrate human-inspired cognitive text descriptions or visual attention to boost performance. The underlying learning paradigm in these works generally remains frame-level binary risk scoring.

Recognizing the limitations of static, single-step anomaly scores, recent work has moved toward explicit temporal and future modeling. A notable example is TOP [11], which predicts accident probabilities at multiple future timestamps by outputting discrete classification logits for a fixed set of future frames. While this multi-horizon formulation improves anticipation performance and reduces false alarms, it still treats future accident occurrence as a binary target and does not explicitly model the progressive nature of risk. In contrast, we share TOP's motivation of anticipating future risk, but adopt a more structured formulation by changing the supervisory signal to a monotonically increasing risk function that captures how accident risk evolves as the crash approaches.

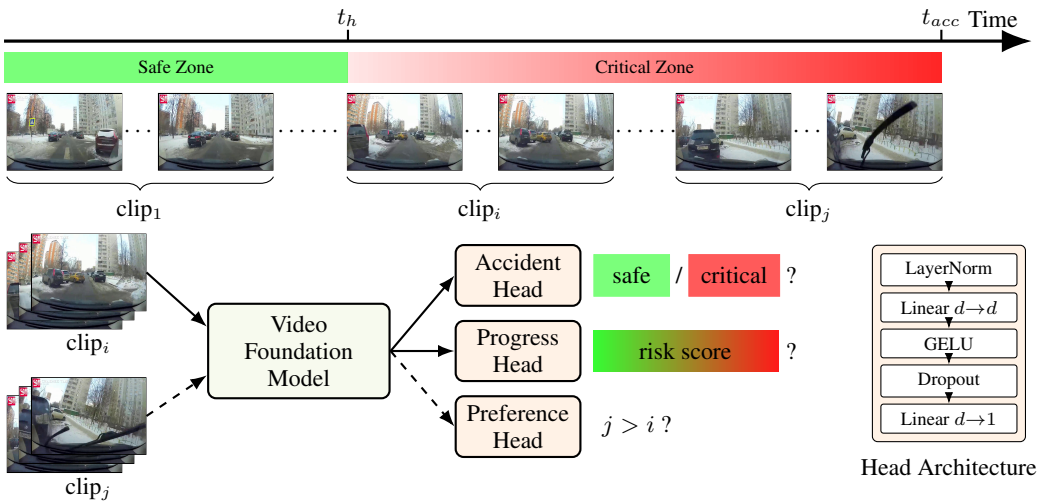

Figure 2: **Overview**. We define the period before the fixed anticipation horizon t_h as the safe zone, and the interval from t_h to the accident time t_{acc} as the critical zone with progressively increasing risk. During training, we extract spatio-temporal features using a video foundation model. Based on these features, the model classifies clips as safe or critical, assigns a continuous risk score, and ranks them using preference head outputs, encouraging higher scores for critical clips closer to the accident. Dashed arrows indicate components used only during training.

Progress estimation. Prior works initially captured task progress through embedding-based visual representations [29–31] or via VQA-style binary success classifiers [32]. More recently, the focus has shifted toward foundation models to build generalist progress estimators. Recent work has explored fine-tuning VLMs on large-scale datasets of successful and failed trajectories to predict discretized progress scores [33], as well as modeling distance-to-goal values for precise robotic manipulation [34]. To improve generalization without relying heavily on human preference annotations, other approaches introduce auxiliary preference prediction objectives that compare heterogeneous trajectories alongside direct progress estimation [35]. Concurrently, a training-free line of research investigates zero-shot VLMs as temporal value estimators using in-context trajectory shuffling [36, 37], while alternative methods extract continuous progress signals directly from internal token probabilities to avoid the instability of numeric VLM outputs [38].

Inspired by these dense temporal formulations, our work adapts progress estimation to the domain of accident anticipation by inverting the concept: rather than tracking the closeness to task success, we track the temporal proximity to catastrophic failure. By modeling accident anticipation as a monotonically increasing "risk progression", we provide our system with the dense temporal supervisory signal necessary to learn the escalating danger in driving videos.

3 Method

In this paper, we develop a video monitoring system to anticipate accidents before they happen, i.e., in an online manner. Given the driving situation observed from the perspective of the ego-vehicle, our goal is to continuously increase the warning signal if the vehicle is on a trajectory likely to result in an accident. We hypothesize that there are observable cues in driving videos that intensify as an accident approaches; for example, a vehicle appearing increasingly larger before impact. To capture such cues, we first divide videos into temporal zones based on their proximity to an accident (Section 3.1). Then, we approach this problem as a video processing task and process each video relying on progress in video foundation models (Section 3.2). Finally, we propose a novel way of estimating the continuously increasing risk of an accident (Section 3.3). See Fig. 2 for an overview.

3.1 Problem formulation

We process driving videos online without assuming access to future information. At any time $t \in \{1, \ldots, T\}$ in a video of length T, we maintain a running state $\mathbf{s}_t$ that is evaluated in terms of its likelihood of leading to an accident $a_t \in [0, 1]$, with 1 representing moments that lead to an accident.

Accident anticipation datasets provide the time of the accident t_{acc}, but only a subset of them includes an anomaly onset label t_{ai}, denoting the first time step at which accident-related cues become visible. While this label is commonly used for evaluation, recent methods [11] avoid using it during training to reduce supervision. Moreover, such annotations are often subjective, dataset-dependent, and not consistently available across benchmarks; e.g., they are not provided on Nexar [13]. We, therefore, use a fixed anticipation horizon t_h to partition each sequence into two temporal zones:

- **Safe-Driving Zone:** The period of safe, nominal driving. This zone spans from the beginning of the sequence to a predefined critical horizon frame, denoted as $t \in [0, t_h)$.
- **Critical Zone:** The period immediately preceding the collision, during which the system is expected to anticipate the event. This zone is defined as $t \in [t_h, t_{\text{acc}})$.

For a safe video containing no accidents, the horizon and accident frames are equivalent to the total video length, $t_h = t_{\text{acc}} = T$, i.e., the critical zone is empty. Ideally, an accident anticipation system must trigger warnings within the critical zone ($t \geq t_h$) to enable timely reactions:

$$a_t = \begin{cases} 1 & t \geq t_h \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Equally, the system should strictly suppress warnings during the safe-driving zone ($t < t_h$) to prevent excessive false alarms. We propose a novel evaluation metric in Section 4 that considers both anticipation capacity in the critical zone and a low false alarm rate in the safe-driving zone.

3.2 Accident anticipation

To capture the dynamics leading to an accident, we construct the state $\mathbf{s}_t \in \mathbb{R}^{N \times H \times W \times 3}$ as a window of N recent frames. We then encode $\mathbf{s}_t$ into a set of spatio-temporal tokens $\mathcal{Z}_t$ using a video foundation model (FoMo):

$$\mathcal{Z}_t = \Phi_{\text{FoMo}}(\mathbf{s}_t), \quad \mathcal{Z}_t \in \mathbb{R}^{K \times D}. \tag{2}$$

where K denotes the number of tokens and D is the hidden dimension. To leverage the pre-training of the FoMo on large-scale datasets, we keep most layers frozen and finetune only the final layers.

To obtain a global representation of the segment, we apply average pooling over the K tokens, resulting in $\mathbf{z}_t$. We then process $\mathbf{z}_t$ with a lightweight MLP head to predict an accident score $\hat{a}_t$:

$$\mathbf{z}_t = \Phi_{\text{pool}}(\mathcal{Z}_t), \quad \mathbf{z}_t \in \mathbb{R}^D \tag{3}$$

$$\hat{a}_t = \Phi_{\text{MLP}}(\mathbf{z}_t). \tag{4}$$

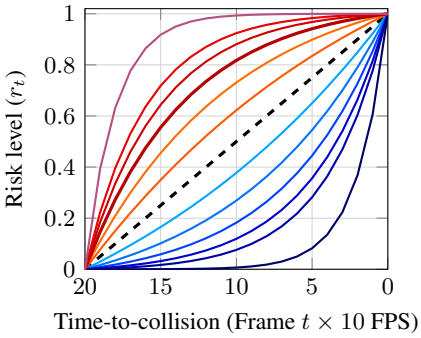

Figure 3: **Varying α.** Continuous risk curves with fast-increasing (red side) and slow-increasing (blue side) behavior obtained by varying α. The linear case is shown with a dashed line.

Segments in the critical zone are treated as positive examples, while negative samples are drawn from the safe-driving zone. See the model and the head architecture in Fig. 2.

3.3 Continuous risk estimation

Similar to prior work in accident anticipation [11], the model described in the previous section can be trained using a BCE loss. However, this formulation does not provide an explicit ordering between segments in terms of their proximity to an accident. To address this, we reformulate accident anticipation as a monotonically increasing function over time.

We define a temporal risk level r_t, which serves as our progress estimation target:

$$\tau = \frac{t - t_h}{t_{\text{acc}} - t_h}, \qquad r_t = \begin{cases} 0, & \text{if } t < t_h, \\ \dfrac{1 - \exp(-\alpha\tau)}{1 - \exp(-\alpha)}, & \text{if } t_h \leq t < t_{\text{acc}}. \end{cases} \tag{5}$$

The risk r_t is strictly 0 in the safe-driving zone. After the critical horizon t_h, it increases within the critical zone as a function of the normalized time τ, approaching 1 at t_{acc}. We explore both fast-

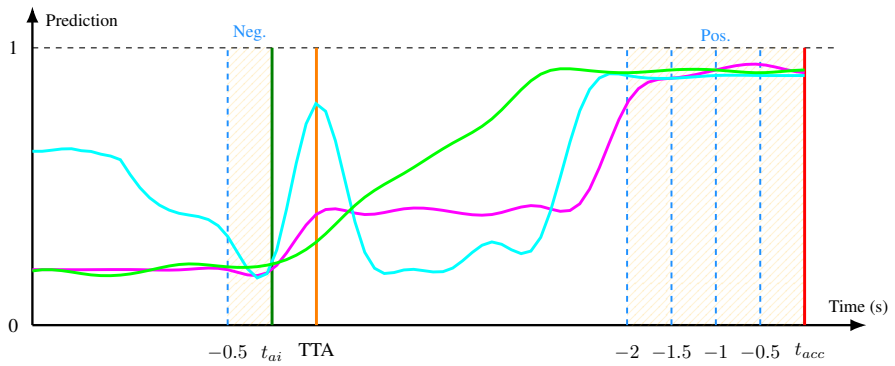

Figure 4: **Problem with current evaluation metrics.** Standard AUC evaluates predictions only within local temporal windows: negatives are sampled around anomaly onset ($t_{ai} - 0.5$) and positives close to the accident ($t_{acc} - 2$), illustrated by the yellow hatched regions. As a result, the cyan, magenta, and green curves achieve similarly high AUC (near 1) and identical TTA, since they rise relative to the pre-anomaly region at comparable times (orange line). However, their global behavior differs significantly. The cyan curve produces early false alarms before t_{ai}, while the magenta curve shows weak progression toward the accident. The green curve exhibits the desired behavior: low risk before t_{ai}, a clear increase after anomaly onset, and steadily rising risk as t_{acc} (red line) approaches.

and slow-increasing variants with different growth behavior by adjusting the non-zero curvature parameter, α, as shown in Fig. 3.

We introduce an additional MLP head on top of the FoMo features, as in Eq. (4), to predict the risk progress $\hat{r}_t$. This head is trained to approximate r_t using a smooth $\mathcal{L}_1$ loss.

Preference estimation. To further encourage temporal ordering between segments, we introduce an auxiliary preference ranking objective. Given a sampled segment, we randomly select another segment from the critical zone of the same video with a higher risk level, i.e., closer to the accident. Using another MLP head, we predict risk scores for both segments and enforce the higher-risk segment to receive a higher score via a margin ranking loss.

We jointly train the model to anticipate accidents, estimate the continuous risk progression toward them, and rank the temporal closeness of states to the accident. The overall objective $\mathcal{L}$ is defined as a weighted combination of binary classification ($\mathcal{L}_{\text{BCE}}$), progress estimation ($\mathcal{L}_1$), and preference ranking ($\mathcal{L}_{\text{rank}}$) losses:

$$\mathcal{L} = w_{\text{acc}}\,\mathcal{L}_{\text{BCE}} + w_{\text{prog}}\,\mathcal{L}_1 + w_{\text{pref}}\,\mathcal{L}_{\text{rank}} \tag{6}$$

4 Beyond AUC and TTA: Separation score for global risk evaluation

4.1 Current state of evaluation

AUC is a standard metric for accident anticipation; however, it requires defining positive and negative samples. Positives are defined at 1.5, 1.0, 0.5, and 0.0 seconds before the accident, where the last corresponds to the moment of impact. Negatives are sampled at a certain temporal distance from the accident, but not too far, i.e., from frames up to 0.5 s before the anomaly onset.

(mean) Time-To-Accident (mTTA) evaluates how early a model anticipates an accident. It is computed based on the first significant prediction peak after the ground-truth anomaly onset (t_{ai}), indicating the earliest time at which the model detects the impending crash.

What is missing in current metrics? Although these metrics provide useful measures of anticipation performance, they evaluate predictions only at selected temporal intervals. As shown in Fig. 4, this fixed-interval evaluation focuses on a narrow temporal window when comparing different methods. Moreover, these metrics often rely on dataset-specific thresholds that must be tuned for each dataset.

This local nature of evaluation can obscure important differences in model behavior over time. In particular, these metrics do not capture the overall shape of the predicted risk curve. As illustrated

Table 1: **Results on MM-AU.** We report results on the CAP [26] (top) and DADA [39] (bottom) splits of MM-AU [12]. Our method consistently outperforms prior work across almost all metrics by substantial margins, with improvements becoming more pronounced closer to the accident.

	Method	$AUC^{0.1}_{0.0s}$	$AUC^{0.1}_{0.5s}$	$AUC^{0.1}_{1.0s}$	$AUC^{0.1}_{1.5s}$	$mAUC^{0.1}$	$mTTA^{0.1}$
CAP	CAP [26]	0.042	0.040	0.030	0.037	0.036	0.637
	DRIVE [40]	0.129	0.117	0.108	0.123	0.116	0.395
	DSTA [9]	0.559	0.386	0.282	0.191	0.286	0.804
	GSC [10]	0.609	0.418	0.297	0.199	0.305	0.817
	TOP [11]	0.838	0.675	0.398	**0.214**	0.429	0.864
	PRE-ACT	**0.887**	**0.775**	**0.465**	0.202	**0.481**	**1.125**
DADA	CAP [26]	0.032	0.037	0.067	0.064	0.056	0.496
	DRIVE [40]	0.101	0.063	0.077	0.088	0.076	0.226
	DSTA [9]	0.473	0.328	0.221	0.135	0.228	0.695
	GSC [10]	0.514	0.350	0.238	0.139	0.242	0.703
	TOP [11]	0.790	0.567	0.288	0.140	0.332	0.885
	PRE-ACT	**0.861**	**0.751**	**0.398**	**0.170**	**0.440**	**1.069**

in Fig. 4, three different prediction curves can yield very similar AUC and TTA values because evaluation is restricted to local temporal windows. However, in accident anticipation, evaluating only a few isolated time snippets is insufficient. From a safety perspective, a reliable model should avoid producing high-risk predictions before the anomaly occurs, thereby minimizing false alarms, while increasing its predictions only after visual cues of an upcoming accident become observable.

4.2 A new separation metric for accident anticipation

An ideal model should assign low scores to pre-anomaly regions, i.e., up to the anomaly onset t_{ai}, and high scores afterward, i.e., from t_{ai} to t_{acc}, creating a clear separation between normal and anomalous temporal segments. Motivated by this observation, we propose a new metric, the *Separation Score*, for the global evaluation of risk prediction curves. The proposed metric compares the area under the predicted risk curve before and after the anomaly onset and combines them into a single score:

$$s_{\text{pre}} = \frac{1}{N_{\text{pre}}} \sum_{i=1}^{t_{\text{ai}}-1} \hat{a}_i, \qquad s_{\text{post}} = \frac{1}{N_{\text{post}}} \sum_{j=t_{\text{ai}}}^{t_{\text{acc}}} \hat{a}_j, \qquad s = \frac{1 - s_{\text{pre}} + s_{\text{post}}}{2}. \tag{7}$$

where $\hat{a}_i$ denotes the predicted risk score at time step i, and N_{pre} and N_{post} denote the number of pre- and post-anomaly predictions, respectively. We compute the average risk before and after the anomaly onset as s_{pre} and s_{post}, and combine them into a normalized separation score $s \in [0, 1]$, where higher values indicate better performance. The ideal case $s = 1$ corresponds to zero risk before the anomaly and unit risk after its onset, indicating perfect separation.

5 Experiments

5.1 Experimental setup

Datasets. We conduct experiments on three established crash anticipation benchmarks: MM-AU [12], DAD [14], and Nexar [13]. MM-AU contains two splits: DADA [39], a driver-attention-oriented dashcam accident dataset with 1,771 training and 198 test videos recorded at 30 FPS, and CAP [26], a larger-scale accident understanding dataset with richer scenario diversity, containing 8,959 training and 800 test videos with varying FPS values. We train and evaluate our models separately on the DADA and CAP splits. The DAD dataset consists of 1,284 training and 466 test videos with fixed five-second clips at 20 FPS. Although it does not provide exact accident-time annotations and only contains accidents between non-ego vehicles, it remains challenging. The Nexar Dashcam Collision Prediction Dataset contains 3,000 training videos and 1,354 test videos, with balanced accident and non-accident samples. Since the challenge permits the use of public datasets, we first train our model on CAP and then fine-tune it on Nexar.

Implementation details. We preprocess input videos by resizing frames to 224×224 and subsampling them at 10 FPS. Based on training data statistics, we set the horizon parameter t_h to 2 seconds. During training, we sample 5-frame clips from segments preceding the accident frame t_{acc}. To avoid an overwhelming number of negative samples, we limit the number of negative clips per video to 50. We set $\alpha = 3$ in the risk level function in Eq. (5). For the preference loss, we sample more risky windows from a temporal range of 0.5 to 1.5 seconds after the current segment, moving toward the accident. During inference, we apply the model in a sliding-window fashion over the video. Our default video encoder is the large version of VideoMAE [41], pre-trained on Kinetics-400 [42]. The task-specific heads consists of two linear layers with GELU activations [43]. We train the task heads together with the last four layers of the video encoder for one epoch using a batch size of 32 on a single NVIDIA A100 GPU. The learning rates are set to $1\text{e}{-5}$ for the video encoder and $1\text{e}{-4}$ for the task heads. We set loss weights in Eq. (6) as follows: $w_{\text{acc}} = 1.0$, $w_{\text{prog}} = 10.0$, and $w_{\text{pref}} = 0.1$.

Metrics. We evaluate our models using the metrics described in Section 4.1. We report the *Separation Score* only for our method, as pretrained models from prior work are not publicly available. Additionally, we report results under an FPR ≤ 0.1 constraint, following [11], since excessive false alarms can reduce trust in the system and limit its practical usefulness.

5.2 Quantitative results

MM-AU [12]. We compare our method with prior work in Table 1 on the CAP [26] and DADA [39] splits of MM-AU, using the baseline results reported in [11]. Our method surpasses the current state-of-the-art TOP [11] by over $\mathbf{10}\%$ on CAP and nearly $\mathbf{30}\%$ on DADA in terms of mAUC. The improvement on CAP further increases to almost $\mathbf{15}\%$ at shorter horizons ($\text{AUC}_{0.5s}$), where positive frames are closer to the collision, highlighting the effectiveness of our risk formulation in emphasizing near-accident scenarios. Our method also anticipates accidents earlier than prior approaches, as reflected by consistent gains in mTTA, improving by **0.25 s** on CAP and **0.18 s** on DADA.

Nexar [13]. We report our results on Nexar in Table 2, comparing against the winner and runner-up as listed on the challenge leaderboard. Our method demonstrates strong performance on Nexar, outperforming the previous winner by $+\mathbf{1}$ mAP. While Nexar officially uses mAP as the primary evaluation metric, we additionally report results with our standard metrics in App. A.8.

Table 2: **Results on Nexar.**

Method	mAP
Runner-up	0.875
Winner	0.885
PRE-ACT	**0.895**

Table 3: **Results on DAD.**

Method	mAP	mTTA
AdaLEA [16]	0.532	3.43
CCAF-Net [44]	0.718	**4.15**
MASTTA [45]	0.702	3.96
TAP [46]	0.727	3.98
PRE-ACT	**0.741**	4.12

DAD [14]. We present results on DAD in Table 3 using the common metrics; mAP and mTTA. PRE-ACT improves mAP by **1.4** points and achieves competitive performance on mTTA.

Cross-dataset generalization. The CAP [26] subset contains substantially more training samples than DADA [39] and Nexar [13], providing a richer source of supervision. To evaluate cross-dataset generalization, we train our model on CAP and directly test it on DADA and Nexar without any further fine-tuning. Results are reported in Table 4, together with performance differences relative to models trained and evaluated on each target dataset.

Table 4: **Cross-dataset generalization.** We train our model on the largest subset, CAP, and evaluate it on DADA and Nexar. As indicated by the performance differences (diff.; increase and decrease) relative to models trained on each dataset's own training set, our model generalizes well to DADA, with a slight performance drop on Nexar due to a larger domain gap.

Train	Test	$\text{AUC}^{0.1}_{0.0s}$	$\text{AUC}^{0.1}_{0.5s}$	$\text{AUC}^{0.1}_{1.0s}$	$\text{AUC}^{0.1}_{1.5s}$	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	s
CAP	DADA	0.897	0.799	0.413	0.178	0.464	1.155	0.760
	diff.	0.036	0.048	0.015	0.008	0.024	0.086	0.020
	Nexar	-	0.529	0.336	0.218	0.361	1.020	0.650
	diff.	-	0.082	0.236	0.151	0.149	0.265	0.049

Table 5: **Ablation on progress and preference estimation.** Compared to the BCE baseline, progress estimation (Prog.) significantly improves performance across all metrics. Adding preference estimation (Pref.) provides slight additional gains, achieving the best overall results.

BCE	Prog.	Pref.	$AUC^{0.1}_{0.0s}$	$AUC^{0.1}_{0.5s}$	$AUC^{0.1}_{1.0s}$	$AUC^{0.1}_{1.5s}$	$mAUC^{0.1}$	$mTTA^{0.1}$	s
✓	✗	✗	0.807	0.690	0.392	0.176	0.419	1.105	0.694
✓	✓	✗	0.884	0.773	0.463	0.197	0.477	1.128	0.723
✓	✓	✓	0.887	0.775	0.465	0.202	0.481	1.125	0.724

Table 6: **Comparison to BCE-only baselines.** `PRE-ACT` consistently outperforms binary supervision across all datasets.

Dataset	Loss	$AUC^{0.1}_{0.0s}$	$AUC^{0.1}_{0.5s}$	$AUC^{0.1}_{1.0s}$	$AUC^{0.1}_{1.5s}$	$mAUC^{0.1}$	$mTTA^{0.1}$
DADA	BCE	0.739	0.534	0.246	0.136	0.305	0.801
	`PRE-ACT`	**0.861**	**0.751**	**0.398**	**0.170**	**0.440**	**1.069**
Nexar	BCE	N/A	0.573	0.552	**0.389**	0.504	1.248
	`PRE-ACT`	N/A	**0.613**	**0.572**	0.369	**0.520**	**1.285**
DAD	BCE	0.577	0.575	0.386	0.251	0.404	1.840
	`PRE-ACT`	**0.636**	**0.610**	**0.394**	**0.264**	**0.423**	**1.871**

Training on CAP transfers well to DADA, yielding improvements across most metrics. This indicates that the scale and diversity of CAP enable the model to learn transferable accident-related cues. In contrast, performance degrades slightly on Nexar, likely due to a larger domain gap in visual appearance and driving conditions, e.g., differences in geography, camera setup, and traffic patterns.

5.3 Ablation study

We perform ablations on the CAP subset using a vanilla model trained with BCE loss as the baseline. Our analysis studies the contributions of progress and preference estimation, the effect of different risk functions, and the ability of the proposed Separation Score to capture differences in the global behavior of models. We present further experiments and analysis in the Appendix.

Contribution of progress and preference estimation. In Table 5, we start with the BCE-only baseline in the first row. Our BCE-only baseline already achieves performance comparable to the SOTA model TOP [11] (Table 1). In particular, it obtains a similar mAUC (0.419 vs. 0.429) while achieving substantially earlier anticipation in terms of mTTA (1.105 vs. 0.864). Adding progress estimation (second row) yields substantial improvements across all metrics, highlighting the benefit of modeling a continuously increasing risk signal. Incorporating preference estimation provides slight additional gains, particularly at the earlier horizon of 1.5 seconds before the accident. The best overall performance is achieved when both are used together, as shown in the last row.

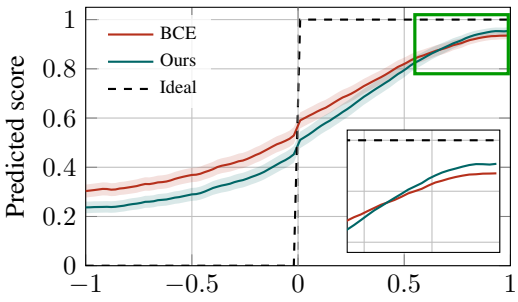

Figure 5: **Separation profile on CAP.** BCE inflates pre-onset scores, whereas our model provides a smoother rise and improved accuracy near the accident, as shown zoomed-in.

We further extend the comparison of our proposed loss formulation vs. BCE-only baselines on other datasets in Table 6.

Risk functions. We experiment with different risk functions by varying the α parameter from Eq. (5), as illustrated in Fig. 3, including a linear baseline. The linear risk function provides a strong

and competitive baseline across metrics. We observe a clear trade-off induced by the shape of the risk function: exponential functions with positive α improve performance at longer anticipation horizons, reflected by higher $\text{AUC}_{1.0s}$ and $\text{AUC}_{1.5s}$, whereas negative α values yield stronger performance at shorter horizons, particularly on $\text{AUC}_{0.0s}$. This provides control over the anticipation behavior of the model. We choose $\alpha = 3$, which achieves the best overall performance in terms of both mAUC and mTTA. Detailed results are presented in Section A.4.

Separation score. In Fig. 5, we plot the normalized mean risk scores of the BCE-only baseline and our model on the CAP subset. The curves are centered at 0, corresponding to anomaly onset, and normalized to $[-1, 1]$, representing the start of the video and the moment of the accident, respectively. Ideally, an accident anticipation model should assign low scores before anomaly onset and then gradually increase them toward the accident. The BCE-only model predicts higher scores before anomaly onset, indicating increased false alarm spikes during safe-driving segments. In contrast, our model exhibits a more controlled increase after onset and achieves higher scores closer to the accident, as highlighted in the zoomed-in region.

Table 7: **Separation scores.**

	Model	$s_{\text{pre}} \downarrow$	$s_{\text{post}} \uparrow$	$s \uparrow$
CAP	BCE	0.420	**0.808**	0.694
	PRE-ACT	**0.344**	0.792	**0.724**
DADA	BCE	0.372	0.763	0.696
	PRE-ACT	**0.319**	**0.798**	**0.740**

In Table 7, we report separation scores on CAP and DADA. Our model consistently assigns lower s_{pre} scores on both datasets, indicating fewer false positives in the safe-driving zone. Although the BCE-only obtains a slightly higher s_{post} on CAP, its substantially higher s_{pre} leads to a lower final score. In contrast, our model achieves cleaner separation between s_{pre} and s_{post}, resulting in consistently higher Separation Score s on both datasets.

Varying backbones. In Table 8, we evaluate PRE-ACT with backbones of varying capacities. Performance consistently improves with stronger video representations, with VideoMAE-L achieving the best overall results. Nevertheless, PRE-ACT remains effective with substantially smaller models, while providing a favorable efficiency–accuracy trade-off: VideoMAE-B reduces the parameter count by more than $3\times$ and increases inference speed from 14.4 to 47.5 FPS, whereas the lightweight X3D-M [47] variant reaches 108.8 FPS with only 6.1M parameters. Moreover, PRE-ACT with VideoMAE-B still achieves better performance than the previous SoTA model, TOP [11], with fewer parameters. These results show that our formulation is not tied to a specific backbone and can be adapted to different computational budgets.

Table 8: **Performance across varying backbone size.** FPS is measured using the standard sliding-window inference pipeline for each method.

Backbone	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	mAUC	mAP	# Params (M)	FPS
TOP [11]	0.429	0.864	0.831	N/A	115	N/A
VideoMAE-L	0.481	1.125	0.833	0.837	306	14.44
VideoMAE-B	0.436	1.101	0.821	0.824	90	47.5
VideoMAE-S	0.358	1.000	0.803	0.797	22	94.8
X3D-M	0.348	0.928	0.780	0.781	6.1	108.8

5.4 Qualitative Results

Figure 6 compares PRE-ACT with the BCE-only baseline on two CAP videos. In the first example, the baseline initially increases the risk estimate after anomaly onset t_{ai}, but subsequently decreases it over part of the critical horizon window despite the approaching collision. PRE-ACT maintains lower risk before anomaly onset and exhibits a more consistent upward trend toward the collision.

In the second example, the BCE-only baseline produces pronounced risk peaks in the safe-driving zone. PRE-ACT substantially suppresses these premature alarms while retaining a sharp increase shortly after t_{ai}. Although both methods ultimately assign high risk near the collision, their predictions differ substantially during the entire driving sequence.

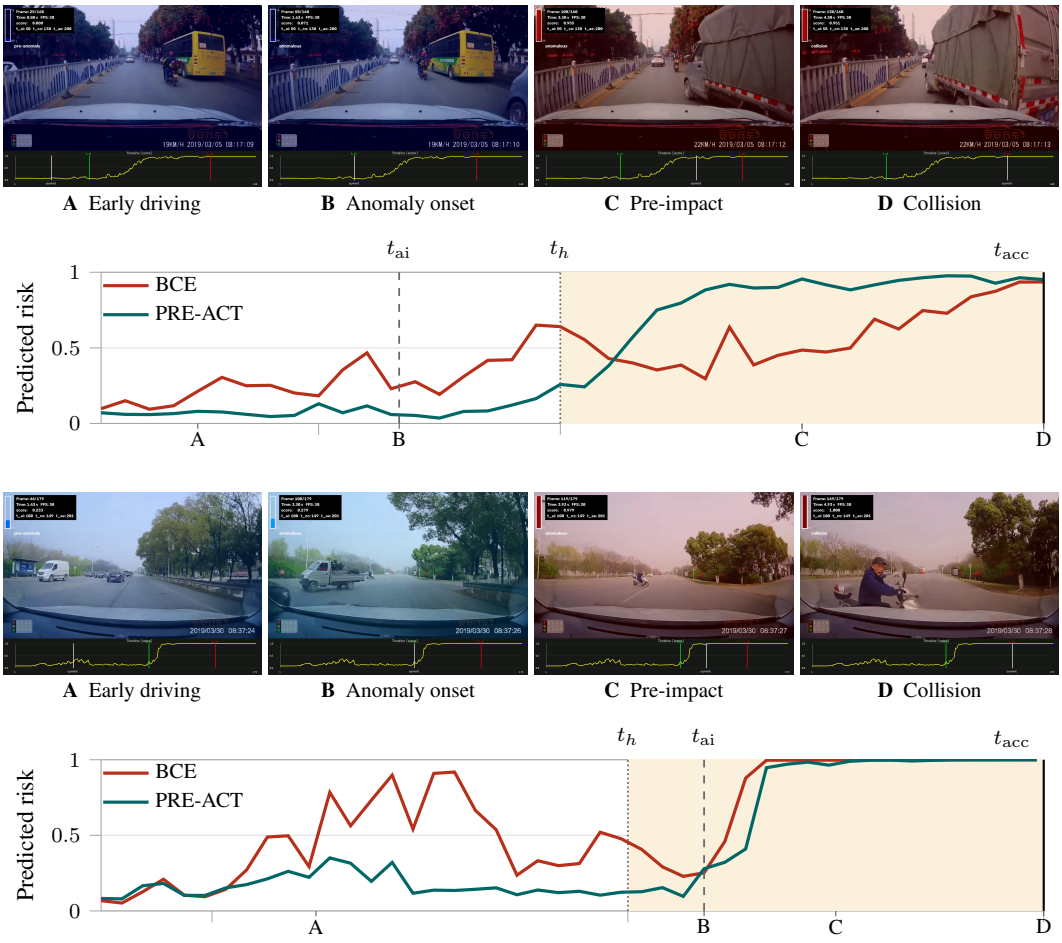

Figure 6: **Qualitative results of PRE-ACT and baseline BCE-only model on two CAP videos.** Shaded regions mark the *critical zone*.

Additional *qualitative results* and *failure cases* are available in the repository linked in the abstract. The observed failures primarily involve poor visibility, including snowy conditions and strong frontal sunlight, which can obscure the visual cues needed for reliable risk estimation.

6 Conclusion

We reformulated accident anticipation from binary classification to continuous risk estimation by modeling risk as a temporally evolving signal that increases as the crash approaches. To better capture global model behavior, we additionally introduced the Separation Score, which evaluates predicted risk curves beyond local temporal windows. Our method significantly advances SOTA on the CAP and DADA subsets of MM-AU, and outperforms the Nexar challenge winner. We identified that most gains come from explicit continuous risk progression estimation, while preference-based ranking provides additional benefits for longer-horizon anticipation. Our ablations on risk functions further reveal a trade-off between short- and long-horizon anticipation behavior. Finally, our Separation Score reveals important differences in false alarm behavior and pre-accident risk evolution that are not captured by existing metrics.

Limitations. A key limitation of our work is that we study accident anticipation as a video monitoring problem, whereas real-world driving safety also depends on factors such as driver state, vehicle dynamics, and sensor reliability. In addition, our CAP-to-Nexar transfer experiments highlight the challenges of cross-domain generalization, which should be studied more systematically given the difficulty of collecting large-scale real crash data.

Acknowledgments and Disclosure of Funding

Eray Çakar and Fatma Güney are funded by the European Union (ERC, ENSURE, 101116486). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] Xiaosong Jia, Junqi You, Zhiyuan Zhang, and Junchi Yan. Drivetransformer: Unified transformer for scalable end-to-end autonomous driving. *arXiv preprint arXiv:2503.07656*, 2025. 1

[2] Ke Guo, Haochen Liu, Xiaojun Wu, Jia Pan, and Chen Lv. ipad: Iterative proposal-centric end-to-end autonomous driving. *arXiv preprint arXiv:2505.15111*, 2025.

[3] Ellington Kirby, Alexandre Boulch, Yihong Xu, Yuan Yin, Gilles Puy, Éloi Zablocki, Andrei Bursuc, Spyros Gidaris, Renaud Marlet, Florent Bartoccioni, et al. Driving on registers. *CVPR*, 2026. 1

[4] Nazir Nayal, Misra Yavuz, Joao F Henriques, and Fatma Güney. Rba: Segmenting unknown regions rejected by all. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 711–722, 2023. 1

[5] Tomáš Vojíř, Jan Šochman, and Jiří Matas. Pixood: Pixel-level out-of-distribution detection. In *European Conference on Computer Vision*, pages 93–109, 2024.

[6] Silvio Galesso, Philipp Schröppel, Hssan Driss, and Thomas Brox. Diffusion for out-of-distribution detection on road scenes and beyond. In *European Conference on Computer Vision*, pages 110–126, 2024.

[7] Youssef Shoeb, Azarm Nowzad, and Hanno Gottschalk. Out-of-distribution segmentation in autonomous driving: Problems and state of the art. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4310–4320, 2025.

[8] Simon Gerstenecker, Andreas Geiger, and Katrin Renz. Fail2drive: Benchmarking closed-loop driving generalization. *arXiv preprint arXiv:2604.08535*, 2026. 1

[9] Muhammad Monjurul Karim, Yu Li, Ruwen Qin, and Zhaozheng Yin. A dynamic spatial-temporal attention network for early anticipation of traffic accidents. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):9590–9600, 2022. 1, 2, 6

[10] Tianhang Wang, Kai Chen, Guang Chen, Bin Li, Zhijun Li, Zhengfa Liu, and Changjun Jiang. Gsc: A graph and spatio-temporal continuity based framework for accident anticipation. *IEEE Transactions on Intelligent Vehicles*, 9(1):2249–2261, 2023. 1, 2, 6

[11] Tianhao Zhao, Yiyang Zou, Zihao Mao, Peilun Xiao, Yulin Huang, Hongda Yang, Yuxuan Li, Qun Li, Guobin Wu, and Yutian Lin. Accident anticipation via temporal occurrence prediction. In *Advances in neural information processing systems*, 2025. 1, 2, 4, 6, 7, 8, 9

[12] Jianwu Fang, Lei-lei Li, Junfei Zhou, Junbin Xiao, Hongkai Yu, Chen Lv, Jianru Xue, and Tat-Seng Chua. Abductive ego-view accident video understanding for safe driving perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22030–22040, 2024. 1, 6, 7, 17

[13] Daniel Moura, Shizhan Zhu, and Orly Zvitia. Nexar dashcam collision prediction dataset and challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2583–2591, 2025. 2, 4, 6, 7, 16, 17

[14] 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, 6, 7

[15] Kuo-Hao Zeng, Shih-Han Chou, Fu-Hsiang Chan, Juan Carlos Niebles, and Min Sun. Agent-centric risk assessment: Accident anticipation and risky region localization, 2017. URL https://arxiv.org/abs/1705.06560.

[16] Tomoyuki Suzuki, Hirokatsu Kataoka, Yoshimitsu Aoki, and Yutaka Satoh. Anticipating traffic accidents with adaptive loss and large-scale incident db. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 7

[17] Wentao Bao, Qi Yu, and Yu Kong. Uncertainty-based traffic accident anticipation with spatio-temporal relational learning. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2682–2690, 2020.

[18] Mishal Fatima, Muhammad Umar Karim Khan, and Chong-Min Kyung. Global feature aggregation for accident anticipation. In *2020 25th International conference on pattern recognition (ICPR)*, pages 2809–2816, 2021.

[19] Wenfeng Song, Shuai Li, Tao Chang, Ke Xie, Aimin Hao, and Hong Qin. Dynamic attention augmented graph network for video accident anticipation. *Pattern Recognition*, 147:110071, 2024. 2

[20] Farhan Mahmood, Daehyeon Jeong, and Jeha Ryu. A new approach to traffic accident anticipation with geometric features for better generalizability. *IEEE Access*, 11:29263–29274, 2023. 2

[21] Inpyo Song and Jangwon Lee. Real-time traffic accident anticipation with feature reuse. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 2312–2317, 2025.

[22] Patrik Patera, Yie-Tarng Chen, and Wen-Hsien Fang. Spatio-temporal adaptation with dilated neighbourhood attention for accident anticipation. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 2452–2458, 2024.

[23] Zongyao Li, Satoshi Yamazaki, and Jianquan Liu. Dual-stream spatio-temporal accident anticipation and detection. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 2630–2635, 2025. 2

[24] Nupur Thakur, PrasanthSai Gouripeddi, and Baoxin Li. 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)*, pages 7533–7541, 2024. 2

[25] Haicheng Liao, Yongkang Li, Zhenning Li, Zilin Bian, Jaeyoung Lee, Zhiyong Cui, Guohui Zhang, and Chengzhong Xu. Real-time accident anticipation for autonomous driving through monocular depth-enhanced 3d modeling. *Accident Analysis & Prevention*, 207:107760, 2024. 2

[26] Jianwu Fang, Lei-Lei Li, Kuan Yang, Zhedong Zheng, Jianru Xue, and Tat-Seng Chua. Cognitive accident prediction in driving scenes: A multimodality benchmark. *arXiv preprint arXiv:2212.09381*, 2022. 2, 6, 7, 16

[27] Haicheng Liao, Haoyu Sun, Huanming Shen, Chengyue Wang, Chunlin Tian, KaHou Tam, Li Li, Chengzhong Xu, and Zhenning Li. Crash: Crash recognition and anticipation system harnessing with context-aware and temporal focus attentions. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 11041–11050, 2024. 2

[28] Hoe Sung Ryu and Christian Wallraven. Mind the gap: Quantifying and aligning human-ai visual attention for accident anticipation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 37887–37895, 2026. 2

[29] Yecheng Jason Ma, Shagun Sodhani, Dinesh Jayaraman, Osbert Bastani, Vikash Kumar, and Amy Zhang. Vip: Towards universal visual reward and representation via value-implicit pre-training, 2023. URL https://arxiv.org/abs/2210.00030. 3

[30] Yecheng Jason Ma, William Liang, Vaidehi Som, Vikash Kumar, Amy Zhang, Osbert Bastani, and Dinesh Jayaraman. Liv: Language-image representations and rewards for robotic control, 2023. URL https://arxiv.org/abs/2306.00958.

[31] Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. R3m: A universal visual representation for robot manipulation, 2022. URL https://arxiv.org/abs/2203.12601. 3

[32] Yuqing Du, Ksenia Konyushkova, Misha Denil, Akhil Raju, Jessica Landon, Felix Hill, Nando de Freitas, and Serkan Cabi. Vision-language models as success detectors, 2023. URL https://arxiv.org/abs/2303.07280. 3

[33] Tony Lee, Andrew Wagenmaker, Karl Pertsch, Percy Liang, Sergey Levine, and Chelsea Finn. Roboreward: General-purpose vision-language reward models for robotics, 2026. URL https://arxiv.org/abs/2601.00675. 3

[34] Huajie Tan, Sixiang Chen, Yijie Xu, Zixiao Wang, Yuheng Ji, Cheng Chi, Yaoxu Lyu, Zhongxia Zhao, Xiansheng Chen, Peterson Co, Shaoxuan Xie, Guocai Yao, Pengwei Wang, Zhongyuan

Wang, and Shanghang Zhang. Robo-dopamine: General process reward modeling for high-precision robotic manipulation, 2025. URL https://arxiv.org/abs/2512.23703. 3

[35] Anthony Liang, Yigit Korkmaz, Jiahui Zhang, Minyoung Hwang, Abrar Anwar, Sidhant Kaushik, Aditya Shah, Alex S. Huang, Luke Zettlemoyer, Dieter Fox, Yu Xiang, Anqi Li, Andreea Bobu, Abhishek Gupta, Stephen Tu, Erdem Biyik, and Jesse Zhang. Robometer: Scaling general-purpose robotic reward models via trajectory comparisons, 2026. URL https://arxiv.org/abs/2603.02115. 3

[36] Paweł Budzianowski, Emilia Wiśnios, Michał Tyrolski, Gracjan Góral, Igor Kulakov, Viktor Petrenko, and Krzysztof Walas. Opengvl – benchmarking visual temporal progress for data curation, 2026. URL https://arxiv.org/abs/2509.17321. 3

[37] Yecheng Jason Ma, Joey Hejna, Ayzaan Wahid, Chuyuan Fu, Dhruv Shah, Jacky Liang, Zhuo Xu, Sean Kirmani, Peng Xu, Danny Driess, Ted Xiao, Jonathan Tompson, Osbert Bastani, Dinesh Jayaraman, Wenhao Yu, Tingnan Zhang, Dorsa Sadigh, and Fei Xia. Vision language models are in-context value learners, 2024. URL https://arxiv.org/abs/2411.04549. 3

[38] Shirui Chen, Cole Harrison, Ying-Chun Lee, Angela Jin Yang, Zhongzheng Ren, Lillian J. Ratliff, Jiafei Duan, Dieter Fox, and Ranjay Krishna. Topreward: Token probabilities as hidden zero-shot rewards for robotics, 2026. URL https://arxiv.org/abs/2602.19313. 3

[39] Jianwu Fang, Dingxin Yan, Jiahuan Qiao, Jianru Xue, and Hongkai Yu. Dada: Driver attention prediction in driving accident scenarios. *IEEE transactions on intelligent transportation systems*, 23(6):4959–4971, 2022. 6, 7, 16

[40] Wentao Bao, Qi Yu, and Yu Kong. Drive: Deep reinforced accident anticipation with visual explanation. In *ICCV*, pages 7619–7628, 2021. 6

[41] Zhan Tong, Yibing Song, Jue Wang, and Limin Wang. Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training. *Advances in neural information processing systems*, 35:10078–10093, 2022. 7, 14, 16

[42] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017. 7

[43] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. 7

[44] Wei Liu, Yafei Li, Tao Zhang, Yixiang Gao, Longsheng Wei, and Jun Chen. Ccaf-net: Cascade complementarity-aware fusion network for traffic accident prediction in dashcam videos. *Neurocomputing*, 624:129285, 2025. 7

[45] Patrik Patera, Yie-Tarng Chen, and Wen-Hsien Fang. A multi-modal architecture with spatio-temporal-text adaptation for video-based traffic accident anticipation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 7

[46] Yuanhong Zhong, Ge Yan, Ruyue Zhu, Ping Gan, and Xuerui Shen. Early traffic accident anticipation via feature consistency representation and soft label regression. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(11):1–21, 2025. 7

[47] Christoph Feichtenhofer. X3d: Expanding architectures for efficient video recognition. In *CVPR*, 2020. 9

[48] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 16

[49] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014. 16

[50] Limin Wang, Bingkun Huang, Zhiyu Zhao, Zhan Tong, Yinan He, Yi Wang, Yali Wang, and Yu Qiao. Videomae v2: Scaling video masked autoencoders with dual masking. In *CVPR*, pages 14549–14560, 2023. 16

[51] Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025. 16

A Appendix

In this appendix, we provide:

- additional details on our model size and inference speed (App. A.1),
- sensitivity to loss weights (App. A.2),
- effect of using different anticipation horizon values (App. A.3),
- ablations on varying risk function (App. A.4),
- performance under different seeds (App. A.5),
- ablation experiments with varying FPR (App. A.6),
- ablation experiments with varying backbones (App. A.7),
- Nexar results using standard anticipation metrics (App. A.8).

A.1 More details on model

Our model, `PRE-ACT`, has 307M total parameters, of which 53M are trainable. We train our models on a single NVIDIA A100 GPU for 2000 iterations with a batch size of 32. During inference, our model achieves an FPS of 14.44 on the same GPU.

Our backbone video foundation model (VideoMAE [41]) takes, by default, 16 frames as input. In order to feed 5 frames (i.e. 0.5 second clips), we devised a padding scheme. We repeat each frame 3 times, except for the last one, which we repeat 4 times.

A.2 Sensitivity to loss weights

We select the weights of progress and preference losses to balance the magnitudes of the three loss terms. We additionally conduct a sensitivity analysis by fixing the weight of the BCE loss (as 1.0) and varying the other two weights. As summarized in Table 9, the mean and standard deviation across five weight configurations show consistently low variation across all metrics, demonstrating that PRE-ACT is not sensitive to the precise choice of loss weights.

Table 9: **Varying weights of progress (w_{prog}) and preference (w_{pref}) losses.** In all experiments, w_{acc} is kept as 1.0.

w_{prog}	w_{pref}	$\text{AUC}^{0.1}_{0.0s}$	$\text{AUC}^{0.1}_{0.5s}$	$\text{AUC}^{0.1}_{1.0s}$	$\text{AUC}^{0.1}_{1.5s}$	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	mAUC	mAP
10	0.1	0.887	0.775	0.465	**0.202**	0.481	1.125	0.833	0.837
5	0.1	0.880	0.765	0.450	0.192	0.469	1.117	0.829	0.835
20	0.1	**0.890**	**0.780**	**0.470**	**0.202**	**0.484**	**1.133**	**0.836**	**0.839**
10	0.05	0.887	0.775	0.465	**0.202**	0.481	1.125	0.833	0.837
10	0.2	0.887	0.774	0.464	0.201	0.480	1.125	0.833	0.837
Mean (std)		0.886 (0.004)	0.774 (0.005)	0.463 (0.008)	0.200 (0.004)	0.479 (0.006)	1.125 (0.006)	0.833 (0.002)	0.837 (0.001)

A.3 Varying anticipation horizon

As explained in the main text, we select the anticipation horizon length based on the training-set statistics of CAP, the largest benchmark with approximately 9,000 training videos and human-annotated anomaly onset. The mean/median intervals between anomaly onset and collision are 2.10/1.50 s in training and 1.88/1.50 s in testing, motivating our default t_h=2 seconds.

Table 10: **Performance across varying anticipation-horizon (t_h).**

H	$\text{AUC}^{0.1}_{0.0s}$	$\text{AUC}^{0.1}_{0.5s}$	$\text{AUC}^{0.1}_{1.0s}$	$\text{AUC}^{0.1}_{1.5s}$	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	mAUC
1.0	**0.916**	**0.780**	0.362	0.142	0.428	1.078	0.815
2.0	0.887	0.775	**0.465**	**0.202**	**0.481**	1.125	**0.833**
3.0	0.865	0.751	0.437	0.179	0.456	**1.128**	0.818
4.0	0.859	0.751	0.457	0.189	0.466	1.112	0.815
5.0	0.867	0.754	0.437	0.169	0.453	1.119	0.819

Table 11: **Results with varying risk functions.** Fast-increasing risk functions with positive α improve longer-horizon anticipation, i.e., ≥ 1 sec, while slow-increasing functions with negative α perform better at shorter horizons, i.e., < 1 sec. $\alpha = 3$ achieves the best overall performance.

Increase	α	$\text{AUC}^{0.1}_{0.0s}$	$\text{AUC}^{0.1}_{0.5s}$	$\text{AUC}^{0.1}_{1.0s}$	$\text{AUC}^{0.1}_{1.5s}$	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	s
Fast	10	0.858	0.750	0.459	0.203	0.471	1.120	0.722
	5	0.877	0.768	**0.467**	**0.204**	0.480	**1.125**	0.724
	4	0.882	0.772	0.466	**0.204**	**0.481**	**1.125**	0.724
	3	0.887	0.775	0.465	0.202	**0.481**	**1.125**	0.724
	2	0.893	**0.778**	0.461	0.198	0.479	1.124	0.725
	1	0.898	**0.778**	0.454	0.193	0.475	1.122	0.726
Linear	N/A	0.901	**0.778**	0.442	0.190	0.470	1.118	0.726
Slow	-1	**0.902**	0.776	0.435	0.187	0.466	1.115	**0.727**
	-2	0.901	0.774	0.430	0.185	0.463	1.114	**0.727**
	-3	0.899	0.769	0.423	0.182	0.458	1.112	**0.727**
	-4	0.899	0.766	0.422	0.184	0.458	1.113	**0.727**
	-5	0.897	0.763	0.420	0.185	0.456	1.111	0.726
	-10	0.885	0.747	0.422	0.188	0.452	1.107	0.724

We additionally evaluate horizons from 1 to 5 seconds, and present the results in Table 10. The 2-second horizon achieves the best overall mAUC and mTTA. A shorter horizon improves detection immediately before impact but substantially weakens early anticipation, whereas longer horizons provide no consistent benefit because they extend supervision to periods before reliable visual cues emerge. Notably, mTTA varies by only 0.05 seconds across all settings, showing that PRE-ACT is robust to the fixed choice of horizon.

A.4 Risk Functions

In the main text, we analyze how different risk functions affect the prediction performance. Here, we present quantitative results on the CAP dataset in Table 11. The linear risk function provides a strong and competitive baseline across metrics. We observe a clear trade-off induced by the shape of the risk function: exponential functions with positive α improve performance at longer anticipation horizons, reflected by higher $\text{AUC}_{1.0s}$ and $\text{AUC}_{1.5s}$, whereas negative α values yield stronger performance at shorter horizons, particularly on $\text{AUC}_{0.0s}$. This provides control over the anticipation behavior of the model. We choose $\alpha = 3$, which achieves the best overall performance in terms of both mAUC and mTTA.

A.5 Multirun Evaluation

In the main text, we present single run results using a fixed seed of 42 for each experiment. To assess variability, we train and evaluate PRE-ACT on CAP using four additional random seeds. As shown in Table 12, the standard deviations across the five runs are small for all metrics, and some runs outperform the default seed on multiple metrics. These results demonstrate that the reported gains are stable and are not attributable to a favorable seed.

Table 12: **Multirun training results of our method on CAP.**

Seed	$\text{AUC}^{0.1}_{0.0s}$	$\text{AUC}^{0.1}_{0.5s}$	$\text{AUC}^{0.1}_{1.0s}$	$\text{AUC}^{0.1}_{1.5s}$	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	mAUC	mAP
42 (default)	0.887	0.775	0.465	0.202	0.481	1.125	0.833	0.837
123	0.865	0.741	0.424	0.198	0.454	1.079	0.819	0.825
2024	0.874	0.760	0.459	0.188	0.469	1.150	0.835	0.838
3407	0.887	0.787	0.471	0.195	0.485	1.125	0.833	0.839
9876	0.871	0.742	0.419	0.191	0.451	1.110	0.827	0.829
Mean	0.877	0.761	0.448	0.195	0.468	1.118	0.829	0.834
(std)	(0.010)	(0.020)	(0.024)	(0.006)	(0.015)	(0.026)	(0.007)	(0.006)

A.6 Complementary results with varying FPR

So far, we have evaluated our method at FPR ≤ 0.1. In Table 13, we further report performance under both stricter (FPR ≤ 0.01) and more lenient (FPR ≤ 1) constraints. When the FPR is constrained to low values, performance drops substantially, particularly for long-horizon predictions. Nevertheless, the mTTA results indicate that our method can still issue warnings within a reasonable time window, achieving above 0.8 s on both datasets even under the strictest setting.

Table 13: **Performance across varying FPR.** Despite drops at low FPR, our method maintains timely warnings, achieving 0.8 s mTTA at FPR ≤ 0.01.

Dataset	λ	$\text{AUC}^{\lambda}_{0.0s}$	$\text{AUC}^{\lambda}_{0.5s}$	$\text{AUC}^{\lambda}_{1.0s}$	$\text{AUC}^{\lambda}_{1.5s}$	mAUC^{λ}	mTTA^{λ}
CAP [26]	0.01	0.721	0.500	0.106	0.025	0.210	0.817
	0.1	0.887	0.775	0.465	0.202	0.481	1.125
	1.0	0.976	0.945	0.853	0.701	0.833	1.820
DADA [39]	0.01	0.732	0.503	0.136	0.038	0.226	0.826
	0.1	0.861	0.751	0.398	0.170	0.440	1.244
	1.0	0.960	0.922	0.782	0.631	0.778	1.597

A.7 Experiments with varying backbones

In Table 14, we evaluate our method using an image foundation model and an alternative video foundation model. For the image model, we use DINOv2 [48] with GRU [49] layers to incorporate temporal information. For the video model, we replace the VideoMAE [50] backbone with V-JEPA [51]. All variants are trained with the same hyperparameters, as described in Section 5.1. As expected, the image backbone performs substantially worse than both video backbones, highlighting the importance of explicit spatio-temporal representations. V-JEPA achieves comparable performance to our model and even outperforms it in mTTA on DADA, indicating that our method is not tied to a specific video encoder.

Table 14: **Varying backbone.** Video models (V-JEPAv2 and VideoMAE) clearly outperform the image backbone (DINOv2), emphasizing spatio-temporal features; similar results across video encoders show backbone-agnostic behavior for our method.

Dataset	Backbone	$\text{AUC}^{0.1}_{0.0s}$	$\text{AUC}^{0.1}_{0.5s}$	$\text{AUC}^{0.1}_{1.0s}$	$\text{AUC}^{0.1}_{1.5s}$	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$
CAP [26]	DINOv2 [48]	0.837	0.573	0.272	0.136	0.327	1.069
	V-JEPAv2 [51]	0.838	0.733	0.418	0.188	0.446	1.288
	VideoMAE [41]	**0.887**	**0.775**	**0.465**	**0.202**	**0.481**	**1.125**
DADA [39]	DINOv2 [48]	0.674	0.358	0.179	0.117	0.218	0.850
	V-JEPAv2 [51]	0.596	0.516	0.314	**0.177**	0.336	**1.250**
	VideoMAE [41]	**0.861**	**0.751**	**0.398**	0.170	**0.440**	1.244

A.8 Complementary results on Nexar challenge

In the main text, we present the Nexar challenge results[3] using the only official metric, mAP. In Table 15, we share the results using standard metrics, AP for different horizons, mAUC, mTTA, and Separation Score.

Table 15: Results using standard metrics on the Nexar [13] dataset.

Method	$\text{AP}_{0.5s}$	$\text{AP}_{1.0s}$	$\text{AP}_{1.5s}$	mAP	mAUC	$\text{mAUC}^{0.1}$	$\text{mTTA}^{0.1}$	s
`PRE-ACT`	0.919	0.911	0.854	0.895	0.895	0.510	1.285	0.665

[3] https://www.kaggle.com/competitions/nexar-collision-prediction/leaderboard

Nexar [13] dataset differs from MM-AU [12] in two main ways. First, the positive videos (i.e. videos with accident) in the test set are trimmed either 0.5, 1.0 or 1.5 seconds before the accident happens. Hence, they do not include the accident frames. Second, the dataset misses the “anomaly onset" labels for the positive videos in the test set.

To calculate the mTTA metric (recall that it requires a t_{ai} annotation), we artificially set the “anomaly onset" label to the same point as t_h. Specifically, we set t_{ai} exactly 2.0 seconds before the accident happens.

A.9 Licenses

All datasets and code used in this work are utilized in accordance with their respective licenses and terms of use. No data is used in violation of applicable licensing restrictions.

VideoMAE license. The official VideoMAE implementation is released primarily under the Creative Commons Attribution–NonCommercial 4.0 International License (CC BY-NC 4.0). This license permits use, distribution, and modification of the material for non-commercial purposes, provided that appropriate credit is given. The repository also notes that some incorporated components are released under separate licenses, including Apache 2.0 and MIT, so those portions remain subject to their original terms.

DINOv2 license. The official DINOv2 code and model weights are released under the Apache License 2.0. This is a permissive license that allows use, modification, and redistribution, including for commercial purposes, provided that the required copyright notice and license terms are preserved. The repository also includes additional subcomponents with separate licensing terms, but the main DINOv2 code and weights are distributed under Apache 2.0.

V-JEPA2 license. The official V-JEPA2 repository states that the majority of the project is released under the MIT License. This license permits use, modification, distribution, and sublicensing with minimal restrictions, provided that the copyright notice and license text are retained. The repository further notes that several utility files are released separately under the Apache License 2.0, so those components remain governed by their original terms.

MM-AU dataset license. The MM-AU dataset is distributed under the Creative Commons Attribution–NonCommercial 4.0 International License (CC BY-NC 4.0). This license permits sharing and adaptation of the dataset for non-commercial purposes, provided that appropriate credit is given. Commercial use is not permitted unless additional permission is obtained from the dataset providers.

Nexar Crash Prediction dataset license. The official paper for the Nexar Dashcam Collision Prediction Dataset and Challenge states that the dataset is released under an open license with restrictions on unethical use. Based on that description, the dataset appears intended to support research and responsible innovation, but the exact license name and full legal terms are not specified in the paper excerpt alone.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction (Sec. 1) summarize the main contributions and scope of this paper reflecting the results in Sec. 5.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations of the work are discussed in the second paragraph of Sec. 6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

Justification: This paper does not present theoretical results.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All necessary implementation details are discussed in Sec. 5.1 and Appendix. The code is also publicly available.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: All data used in this work is publicly available. The code is be also made publicly available with clear instructions for reproducibility.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: All necessary training and test details are discussed in Sec. 5.1 and Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The results in Sec. 5 are accompanied by statistical significance metrics whenever applicable.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Compute resources are reported in Sec. 5.1 alongside implementation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The authors have reviewed and understood the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Potential impacts are discussed in Sec. 1.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The original papers are cited in the main paper and the licenses are reported in the Appendix.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The new assets are released with proper documentation.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve such studies.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve such studies.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: This research does not involve LLM usage in the core method development.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.