BADAS: Context Aware Collision Prediction Using Real-World Dashcam Data

Roni Goldshmidt
roni.goldshmidt@getnexar.com

Hamish Scott
hamish.scott@getnexar.com

Lorenzo Niccolini
lorenzo.niccolini@getnexar.com

Shizhan Zhu
shizhan.zhu@getnexar.com

Daniel Moura
daniel.moura@getnexar.com

Orly Zvitia
orly.zvitia@getnexar.com

October 17, 2025

Abstract

Existing collision prediction methods often fail to distinguish between ego-vehicle threats and random accidents non-involving ego-vehicle, leading to excessive false alerts in real-world deployment. We present BADAS, a family of collision prediction models trained on Nexar's real-world dashcam collision dataset—the first benchmark designed explicitly for ego-centric evaluation. We re-annotate major benchmarks to identify ego involvement, add consensus alert-time labels, and synthesize negatives where needed, enabling fair AP/AUC and temporal evaluation. BADAS uses a V-JEPA2 backbone trained end-to-end and comes in two variants: BADAS-Open (trained on our 1.5k public videos) and BADAS1.0 (trained on 40k proprietary videos). Across DAD, DADA-2000, DoTA, and Nexar, BADAS achieves state-of-the-art AP/AUC and outperforms a forward-collision ADAS baseline while producing more realistic time-to-accident estimates. We release our BADAS-Open model weights and code, along with re-annotations of all evaluation datasets to promote ego-centric collision prediction research.

1. Introduction

Collision prediction is fundamental to Advanced Driver Assistance Systems (ADAS) and autonomous vehicles, yet current approaches fail to meet real-world deployment requirements. Despite decades of research, existing methods struggle with excessive false alarms and miss critical ego-vehicle threats. We present BADAS (V-JEPA2 [1] Based Advanced Driver Assistance System), a new approach that achieves state-of-the-art performance by combining modern video foundation models with high-quality, ego-centric real-world driving data. As shown in Figure 1, BADAS significantly outperforms both academic methods and commercial ADAS systems across major benchmarks, demonstrating the power of aligning training data with actual deployment scenarios.

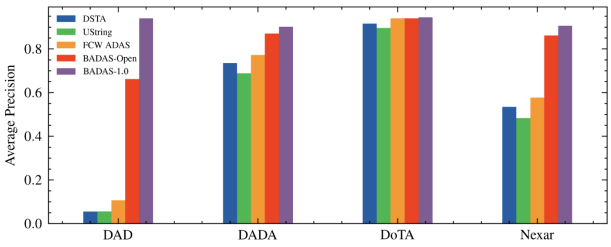

Figure 1. BADAS achieves state-of-the-art performance on collision prediction benchmarks. Our approach substantially outperforms existing academic methods and commercial vision-based forward collision warning systems by leveraging ego-centric real-world data and modern video foundation models when evaluated on ego-vehicle involved collisions.

The core challenge in collision prediction lies not in the algorithms themselves, but in the fundamental misalignment between training data and actual driving scenarios. More critically, existing real-world datasets like DAD [2], DoTA [3], and DADA-2000 [4] suffer from a conceptual flaw: they treat all visible accidents equally, training models to detect any collision within the camera's field of view. This approach generates excessive false alarms in deployment—a vehicle accident two lanes over triggers the same alert as an imminent frontal collision threatening the ego vehicle.

Figure 2 illustrates this critical distinction between ego-centric and general collision prediction. While visually salient accidents in adjacent lanes may capture attention, they are irrelevant to the ego vehicle's safety and should not trigger warnings. Our systematic re-annotation of existing benchmarks reveals the severity of this problem as shown in Table 1.

Figure 2. The critical distinction between ego-centric and general collision prediction. Top row: Ego-vehicle involved events requiring immediate driver intervention. Bottom row: Non-ego accidents that are visually prominent but irrelevant to ego-vehicle safety. BADAS focuses exclusively on ego-relevant scenarios to minimize false alarms while maintaining high sensitivity to actual threats. Examples from the DAD dataset [2].

Table 1. Re-annotation results of major collision-prediction datasets. The high percentage of non-ego accidents (40–92%) suggests that these datasets should be used with caution when developing or evaluating ADAS-related methods. *Pos-Ego*: ego-involved accidents. *Pos-Not Ego*: non-ego accidents. *Less-2s*: samples filtered out due to insufficient anticipation time—specifically, cases where the collision occurs within 2 seconds from the beginning of the video. Our model requires a full 2 seconds prediction horizon; thus, such cases are invalid for anticipation. *Negative*: normal driving videos. *%Not Ego*: percentage of non-ego accidents among retained positive samples.

Dataset	#Pos-Ego	#Pos-Not Ego	#Less-2s	#Negative	% Pos-Not Ego
DADA-2000 [4]	75	51	2	0	40.5
DAD [2]	13	150	2	301	92.0
DoTA [3]	327	255	16	0	43.8
Nexar [5]	672	0	0	672	0.0

Beyond the ego-centric issue, existing datasets exhibit additional limitations, posing challenges for effective ADAS development. DoTA and DADA-2000 contain only positive examples, making it impossible to evaluate false positive rates—a crucial metric for user acceptance. They also lack consistency in defining alert timing: DAD bases alerts on abnormal behavior onset, DoTA uses subjective "inevitability" judgments, and DADA-2000 triggers alerts upon 3rd party vehicle appearance regardless of actual risk.

The recently introduced Nexar Dashcam Collision Prediction Dataset [5, 6] addresses these fundamental problems through a paradigm shift in data collection. Unlike existing datasets, Nexar comprises 1.5k real-world dashcam videos from actual drivers experiencing genuine collisions and near-misses. This dataset provides three key advantages: (1) exclusive focus on ego-vehicle involved events, eliminating irrelevant accidents, (2) inclusion of near-collision events where accidents are avoided through emergency maneuvers—capturing the much more common successfully-resolved dangerous situations that provide rich training signals, and (3) standardized, consensus-based alert timing annotations from 10 human annotators that enable consistent temporal evaluation. The real-world nature of this data is crucial: it captures the true distribution of driving scenarios, including rare events that synthetic datasets cannot replicate and that drivers actually encounter on the road.

Our approach leverages recent advances in video foundation models, particularly V-JEPA2 [1], which has shown strong capabilities in understanding temporal dynamics and visual patterns. These architectures excel when trained on data that connects visual patterns to concrete outcomes—precisely what ego-centric collision and near-collision events provide. By fine-tuning V-JEPA2 on the Nexar high-quality ego-relevant dataset, we harness its temporal reasoning capabilities for collision prediction. The combination of modern architectures with appropriate real-world data yields the significant performance gains shown in Figure 1, substantially surpassing both task-specific architectures and commercial ADAS systems.

The contributions of this paper are as follows:

1. **Ego-centric problem reformulation**: We redefine collision prediction as an ego-centric task and systematically re-annotate major benchmarks (DAD, DoTA, DADA-2000) to identify ego-vehicle involvement, revealing that 40-92% of their accidents are irrelevant for ADAS applications. These annotations are publicly released to benefit the community.
2. **Standardized temporal evaluation**: We establish a coherent definition of alert timing through 10-annotator consensus and contribute precise temporal annotations for all test sets, addressing the inconsistent and subjective definitions across existing datasets. Our results in 8 emphasize that longer estimated time to accident may represent early anticipations rather than a better actual prediction.
3. **State-of-the-art performance**: Our models surpass existing methods by fine-tuning V-JEPA2 [1] on high-quality, ego-centric real-world data, demonstrating that combining advanced architectures with appropriate data significantly outperforms both academic methods and commercial vision-based forward collision warning systems.
4. **Real-world data importance**: We demonstrate that practical collision prediction requires training in genuine traffic situations rather than synthetic or controlled envi-

ronments. A preliminary analysis (Figure 9a) already reveals the natural long-tailed structure of real-world collisions, underscoring the value of diverse large-scale data.

The rest of this paper is organized as follows: Section 2 reviews related work in collision prediction datasets and methods and video understanding. Section 3 presents our methodology including model architecture and training protocols. Section 4 describes our experimental setup and re-annotation process. Section 5 presents a comprehensive analysis and results. We conclude in Section 6 and discuss potential directions for future research in Section 7.

2. Related Work

2.1. Traffic Accident Prediction Methods

Recent advances in traffic accident prediction have produced various approaches, though we focus our comparison on methods with open-source implementations that enable reproducible evaluation on our ego-centric test sets.

UString [7] uses RNN-based architectures with adaptive loss functions that emphasize frames near collision events, dynamically adjusting loss contribution based on temporal proximity to accidents. DSTA [8] employs transformer-based architectures with dynamic spatial-temporal attention mechanisms, allowing the model to focus on relevant spatial regions while tracking their temporal evolution. Both methods provide open-source implementations, facilitating direct comparison.

While other approaches using graph neural networks [9], reinforcement learning [10], and multi-modal fusion [11] exist in the literature, the lack of publicly available implementations prevents fair comparison on our ego-centric datasets. Therefore, our experimental comparison focuses on UString and DSTA as the current open-source state-of-the-art.

2.2. Collision Prediction Datasets and Temporal Annotations

Existing datasets differ significantly in their temporal annotation approaches. DAD [2] contains 1,750 videos and defines accidents based on abnormal behavior onset rather than impact moments. DoTA [3] provides 4,990 videos with subjective annotations marking when accidents appear "inevitable," resulting in inconsistent temporal boundaries. DADA-2000 [4] offers 1,962 videos annotating from vehicle appearance to collision, which may mark accidents as predictable too early when vehicle presence alone doesn't indicate danger.

The Nexar dataset [6] addresses these limitations through consensus-based alert times from multiple annotators, establishing both the earliest moment of recognizable danger (t_{alert}) and precise collision timing ($t_{\text{collision}}$), enabling evaluation of both prediction accuracy and temporal appropriateness.

2.3. Foundation Models for Video Understanding

Video foundation models have evolved from temporal extensions of image architectures to sophisticated self-supervised approaches. SlowFast networks [12] process videos at multiple temporal resolutions, while Video Swin Transformers [13] extend window-based attention temporally.

Self-supervised methods like VideoMAE [14] reconstruct masked spatial-temporal patches, while V-JEPA [15] predicts abstract representations rather than raw pixels. V-JEPA2 [1] further improves masking strategies and training objectives for temporal understanding, making it particularly suited for anticipatory tasks. BADAS is the first application of V-JEPA2 for collision prediction, demonstrating that fine-tuning on ego-centric data substantially outperforms task-specific architectures.

2.4. Industrial ADAS Systems

Current ADAS Forward Collision Warning (FCW) systems typically combine radar and camera sensors with physics-based models to calculate time-to-collision and trigger alerts [16]. For comparison purposes, we focus on vision-only implementations that process monocular dashcam inputs. The open-source FCW system [17] uses YOLO [18] for object detection and [19] for lane detection, raising alerts when detected objects fall within distance thresholds. This provides a baseline representing current deployed vision-based ADAS technology.

3. Methodology

3.1. Problem Formulation

Building on the Nexar Challenge [5], we reformulate collision prediction as an ego-centric task. Given a dashcam video sequence $\mathcal{V} = \{f_1, ..., f_T\}$, we predict at each timestep t the probability $p_t = P(\text{ego-collision}|f_{1:t})$ that the ego vehicle will be involved in a collision within time horizon τ.

Ego-vehicle incidents: Scenarios where the ego vehicle experiences collision or executes emergency maneuvers, including direct collisions, near-misses requiring evasive action, and situations prevented only by emergency intervention.

Non-ego incidents: Accidents visible but irrelevant to ego-vehicle safety, including adjacent lane collisions and distant events.

Near-collisions (emergency maneuvers that prevented accidents) are treated as positive training examples, providing rich supervisory signals from successfully-resolved dangerous situations.

3.2. BADAS Architecture

BADAS uses V-JEPA2 [1] backbone with patch aggregation and classification head:

V-JEPA2 Backbone: Encoder patchifies videos using $2 \times 16 \times 16$ tubelets and passes tokens through ViT-L transformer. We process 16 frames of size 256×256, yielding 2048 latent patches with dimension $D = 1024$.

Patch Aggregation: We aggregate patches using attentive probe [20]. Given patches $X \in \mathbb{R}^{P \times D}$, learned queries $Q \in \mathbb{R}^{M \times D}$ compute attention scores $A = \text{softmax}(QX^\top/\sqrt{D})$. These aggregate patches via weight matrix $W \in \mathbb{R}^{D \times d}$ to produce features $AXW \in \mathbb{R}^{M \times d}$, concatenated to final vector of size Md (with $M = 12$, $d = 64$).

Prediction Head: Three-layer MLP with GELU activation, layer normalization, and 0.1 dropout probability maps aggregated features to collision probability. Hidden dimension: 768.

3.3. Training Protocol

We fine-tune V-JEPA2 end-to-end using AdamW optimizer with learning rate $\eta = 1 \times 10^{-5}$, weight decay 1×10^{-4}, and cosine annealing schedule. Binary cross-entropy loss with mixed precision training and gradient clipping (norm 5.0) ensures stability. Early stopping on validation set prevents overfitting.

3.4. Dataset Re-annotation Protocol

We systematically re-annotated DAD, DoTA, and DADA-2000 for ego-centric evaluation:

Ego-Involvement Classification: We manually reviewed all accidents, categorizing them as ego-involved - recording vehicle is directly involved in the accident, or non-ego involved - accidents in adjacent lanes, perpendicular traffic or visible but non-threatening.

Human Alert Time: 10 annotators with defensive driving certification marked when they would initiate defensive action. Consensus time (median) represents realistic human response timing.

Figure 3 shows human reaction patterns across 726 events: median 1.70s, mean 1.81s (SD=0.82s), with 90% of alerts between 0.70-3.47s before collision. Alerts beyond 95th percentile (3.47s) likely respond to normal variations rather than genuine threats.

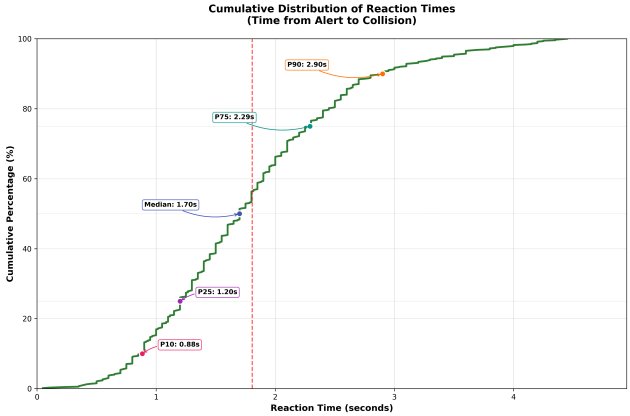

Figure 3. Cumulative distribution of human reaction times across 726 ego-involved collisions. Median: 1.70s, 90% range: 0.70-3.47s before impact.

Synthetic Negatives: For datasets lacking negative samples, we extract first 4 seconds from videos where human alert occurs 4.5s or later after video start, ensuring realistic driving without collision precursors.

Table 2. Dataset composition after ego-centric re-annotation

Dataset	Real Neg.	Real Pos.	Synth. Neg.	Total
DADA-2000	0	75	38	113
DAD	301	13	0	314
DoTA	0	327	40	367

Our consensus-based protocol provides unified temporal reference across datasets, addressing the existing inconsistencies and enabling fair comparison. The re-annotations are publicly available [1].

3.5. Model Variants

BADAS-Open: Trained solely on Nexar's public dataset (1,500 videos) with balanced collision/normal driving representation. Achieves state-of-the-art on major benchmarks. Model and code are publicly available [2] under the APSv2 license.

BADAS1.0: Commercial variant trained on 40k proprietary sequences (20k collisions/near-misses, 20k normal driving). Uses identical architecture and training protocols, but 25× more data, enabling data scaling effect evaluation and superior edge-case handling.

[1] https://github.com/getnexar/BADAS-Open/tree/main

[2] https://huggingface.co/nexar-ai/BADAS-Open

Table 3. Ego-centric collision prediction performance across benchmarks.

	DAD (n=116)			DADA (n=113)			DoTA (n=367)			Nexar (n=1344)		
Method	AP	AUC	mTTA	AP	AUC	mTTA	AP	AUC	mTTA	AP	AUC	mTTA
DSTA	0.06	0.59	2.9	0.74	0.60	6.5	0.92	0.55	4.8	0.53	0.54	9.8
UString	0.06	0.61	2.4	0.69	0.57	5.3	0.90	0.54	4.3	0.48	0.48	9.1
FCW ADAS	0.11	0.69	2.0	0.77	0.78	3.8	0.94	0.69	3.0	0.58	0.64	8.7
BADAS-Open	0.66	0.87	2.7	0.87	0.77	4.3	0.94	0.70	4.0	0.86	0.88	4.9
BADAS1.0	**0.94**	**0.99**	2.7	**0.90**	**0.87**	4.6	**0.95**	**0.72**	4.0	**0.91**	**0.91**	3.9

4. Experiments

4.1. Experimental Setup

We evaluate BADAS against state-of-the-art collision prediction methods and industrial ADAS systems. Our baseline methods include UString [7], DSTA [8] and FCW ADAS [17], a vision-only forward collision warning system using YOLOv11 [18] for object detection and TuSimple ResNet-34 [19] for lane detection.

Evaluation is performed on three benchmarks re-annotated for ego-centric assessment: DAD [2], DADA [4], DoTA [3], and the Nexar test set. For datasets originally containing only positive samples (DADA, DoTA), we generate synthetic negatives as described in Section 3. Performance is measured using Average Precision (AP) for ranking quality, Area Under Curve (AUC) for discrimination capability, and mean Time-To-Accident (mTTA) for temporal accuracy.

5. Results

5.1. Quantitative Performance

Table 3 presents a comprehensive summary of our evaluation results. BADAS models achieve substantial improvements over existing methods, with BADAS1.0 reaching 0.94 AP on DAD compared to 0.06 for baseline methods. The performance gap is particularly striking on the Nexar dataset—the largest and most comprehensive benchmark—where BADAS-Open achieves 0.86 AP versus 0.48-0.53 for academic baselines and 0.58 for the industrial FCW system.

Notably, BADAS models maintain consistent performance across diverse datasets, indicating robust generalization. The mTTA values reveal a critical distinction: baseline methods report physically implausible predictions (9-10 seconds before collision on Nexar), while BADAS maintains more realistic 3-5 second windows better aligned with human prediction capabilities. The industrial FCW system, despite using state-of-the-art detection models, underperformed due to its rule-based logic generating excessive false positives in dense traffic—highlighting the advantages of end-to-end learning from real collision data.

5.2. Qualitative Analysis

Figure 4 illustrates prediction scores over time for representative test samples. BADAS-Open (red) demonstrates stable, confident predictions that rise sharply as collisions approach, while baseline methods exhibit erratic patterns with premature or inconsistent alerting. This stability translates to practical deployment benefits leading to fewer false alarms and more reliable threat assessment.

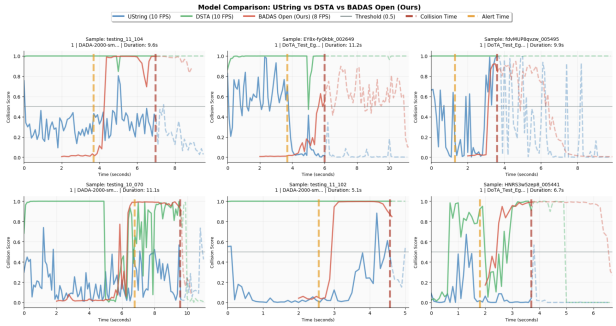

Figure 4. Collision predictions scores over time for 6 samples: UString (blue), DSTA (green) and BADAS-Open (red). Vertical dashed lines indicate collision time, horizontal line shows detection threshold.

Table 4. Performance for different label windowing strategies.

Window	Precision	Recall	AP	AUC	mTTA
1.0s	**0.792**	0.896	0.930	0.930	3.112
1.5s	0.785	0.939	**0.931**	**0.935**	3.880
2.0s	0.741	**0.964**	0.925	0.931	**4.646**

Frame-by-frame analysis in Figure 5 further demonstrates that in urban intersection and residential scenarios, BADAS variants correctly identify ego-relevant threats while maintaining low confidence during safe driving periods.

5.3. Data Scaling Effects

Figure 6 illustrates the effect of training data scale, showing a logarithmic improvement in Nexar validation AP as the dataset size increases from 1.5k to 40k videos. This consistent scaling trend supports our hypothesis that leveraging large-scale real-world data continues to yield performance gains, even at the largest scales tested.

5.4. Ablation Studies

We investigate key design choices through systematic ablations. Table 4 examines label window selection, showing that $T = 1.5$ seconds optimally balances precision (0.785) and recall (0.939). Shorter windows miss developing collision patterns, while longer windows include too many non-indicative frames.

Data augmentation and oversampling strategies (Table 5) show significant benefits, with the combination of 2× oversampling for positive samples and augmentations improv-

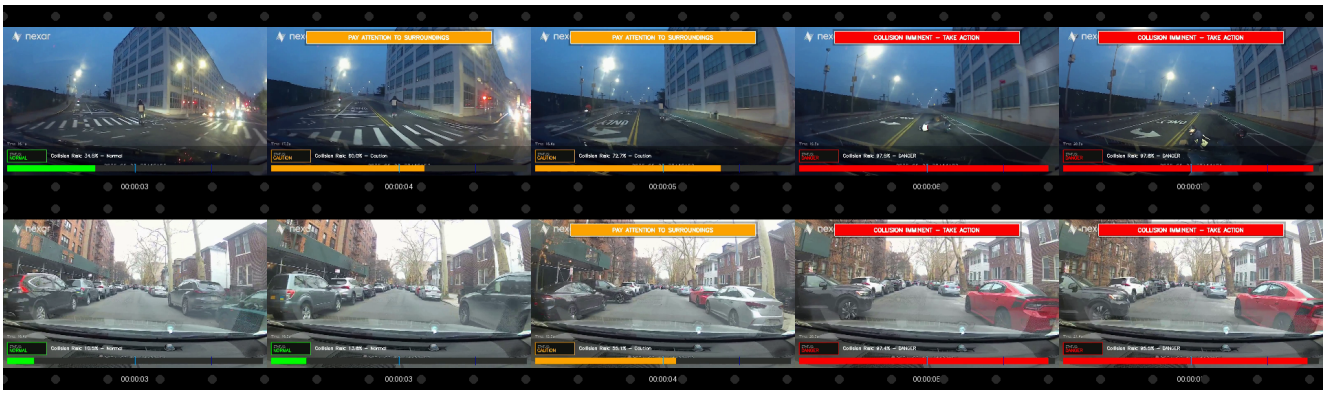

Figure 5. Frame-by-frame comparison on Nexar test footage. Top: urban intersection collision. Bottom: residential street scenario. Alert indicators show prediction confidence (green: safe, orange: caution, red: imminent collision). BADAS variants demonstrate accurate ego-centric threat assessment.

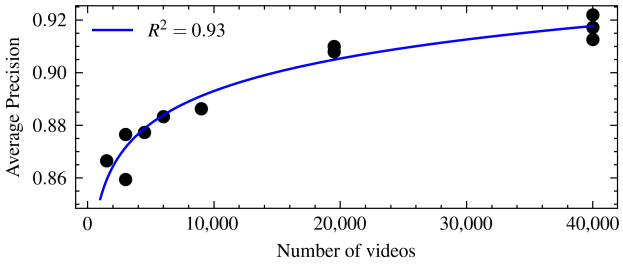

Figure 6. Validation AP versus training data size, showing logarithmic performance improvement with data scaling.

```python
import albumentations as A

A.Compose([
    A.RandomBrightnessContrast(p=0.5),
    A.HueSaturationValue(p=0.5),
    A.GaussianBlur(blur_limit=(3, 7), p=0.3),
    A.CLAHE(p=0.3),
    A.RandomGamma(p=0.3),
    A.RGBShift(p=0.3),
    A.MotionBlur(blur_limit=7, p=0.3),
    A.RandomRain(p=0.2),
    A.RandomSnow(p=0.2),
    A.RandomShadow(p=0.2),
])
```

Figure 7. Data augmentations for weather and lighting robustness.

Table 5. Impact of sampling strategies and augmentations.

Oversampling Rate	Augmentations	AP	AUC
1×	No	0.922	0.927
2×	No	0.924	0.922
1×	Yes	0.931	0.935
2×	Yes	**0.939**	**0.941**

ing AP from 0.922 to 0.939. The augmentations (Figure 7) simulate diverse weather and lighting conditions, improving robustness to real-world variations.

Architecture ablations (Table 6) show the contribution of each component. We train models with the base V-JEPA2 backbone with: only a linear head, the attentive probe with a linear head, and the attentive probe with the MLP head. We test training with and without the base model frozen. Fine-tuning the V-JEPA2 backbone provides the largest gain (0.707→0.928 AP), while the attentive probe and MLP head add incremental improvements, reaching 0.939 AP with the complete architecture.

Table 6. Architecture component contributions.

Configuration	AP	AUC
Frozen base	0.707	0.744
Frozen base + probe	0.782	0.811
Frozen base + probe + MLP	0.771	0.809
Fine-tuned base	0.928	0.935
Fine-tuned base + probe	0.929	0.936
Fine-tuned base + probe + MLP	**0.939**	**0.941**

5.5. Temporal Accuracy Analysis

Figure 8 compares Time-To-Accident distributions at 80% confidence against human consensus annotations. BADAS-Open's median TTA of 3.0 seconds closely aligns with practical requirements—providing sufficient warning while avoiding premature alerts. In contrast, UString and DSTA show median TTAs of 6.2s and 7.5s respectively, with high variance extending to implausible 14-second predictions. This unrealistic early detection often manifests as false pos-

itives in deployment, posing challenges when integrating these methods into real-world ADAS.

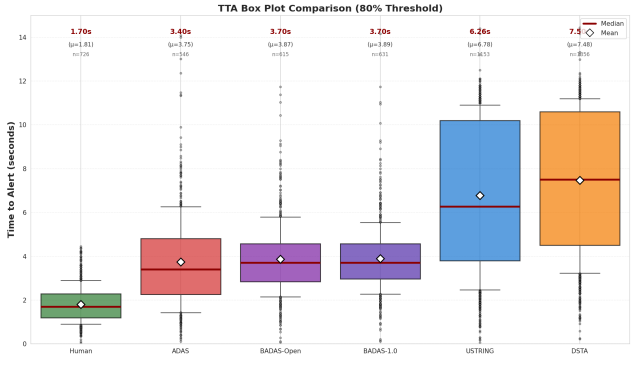

Figure 8. TTA distributions at 80% confidence. Human consensus (green) shows expert annotations. BADAS-Open demonstrates realistic timing while baselines exhibit implausibly early predictions.

5.6. Long-Tail Performance

Beyond common vehicle-to-vehicle interactions, real-world collision prediction must also recognize rare but safety-critical scenarios. To assess BADAS-Open robustness under such conditions, we measure recall across various third-party long-tail categories for confidence threshold of 0.85 relative to the recall observed on the dominant "vehicle" class.

As shown in Figure 9, the model achieves the highest recall on vehicle-related events, with only a modest drop in larger vehicles (trucks and busses). Vulnerable Road Users (VRUs) categories - pedestrians, cyclists, motorcyclists - show a notable decline in recall, and performance decreases sharply further on animal-related incidents.
This degradation highlights the challenges of generalizing to rare, visually diverse, and low-frequency event types. While this behavior is expected given the absence of such cases in the training distribution of the BADAS-Open model, these results underscore the importance of explicitly evaluating the long-tail performance in collision prediction.

6. Conclusion

This work introduces BADAS, a new approach to collision prediction that focuses on ego-vehicle safety through ego-centric problem formulation. Building on insights from the Nexar Dashcam Collision Prediction Challenge, we demonstrate that focusing exclusively on ego-vehicle threats—rather than general accident detection—dramatically improves real-world performance.

Our systematic re-annotation of major benchmarks reveals fundamental issues with existing datasets: a significant portion of annotated accidents do not involve the ego-

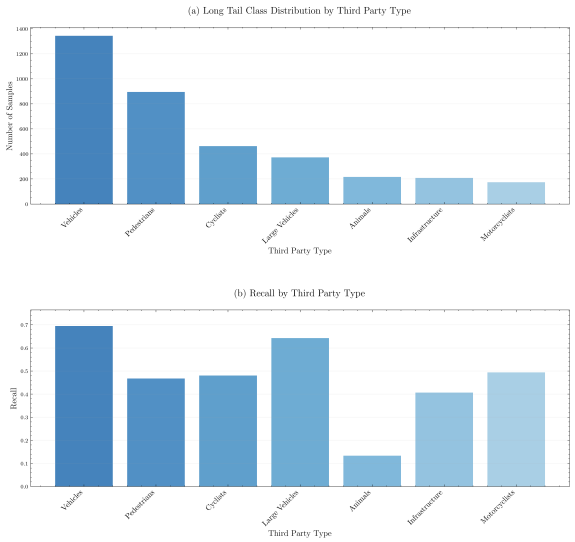

Figure 9. (a) Distribution of third-party categories, representing the interacting agents. Static infrastructure include buildings, fences, poles and traffic signs. The vehicle class corresponds to the Nexar [5] test set used throughout this work, while the remaining categories originate from additional annotated data. (b) BADAS-Open recall across these categories at a confidence threshold of 0.85

vehicle, leading models to learn patterns irrelevant to ego-vehicle safety. By filtering for ego-relevance and establishing human baseline reaction times, we create evaluation protocols that better reflect real-world deployment requirements. Our synthetic negative sampling method further improves the balance between positive and negative samples and relaxes the biased AP and AUC measurements.

We further highlight the necessity of a coherent definition and annotation scheme for alert time, to serve as reference to the predicted mTTA values. Our findings show varying levels of early prediction in all methods. This is especially important for practical systems as these early predictions will be manifested as false alerts when deployed in real ADAS or AV frameworks.

We present two model variants addressing different deployment needs: BADAS-Open, trained exclusively on 1.5k public Nexar videos, and BADAS1.0, leveraging 40k videos from Nexar's proprietary dataset. Both models achieve state-of-the-art performance when compared to leading research methods and FWC systems. The significant performance gain observed with increased data volume suggests that the potential of data scaling has not yet been fully saturated. The BADAS-Open model and code are released to the research community.

While our model outperforms existing state-of-the-art results, we also highlight the long-tail nature of collision and near-collision distributions, showing that BADAS-

Open performance significantly deteriorates on minority classes. This result is expected, as any model trained on an imbalanced dataset naturally focuses on majority classes (e.g., vehicle-to-vehicle accidents). However, edge cases must also be taken into account — first by explicitly evaluating current model performance on them, and later by developing dedicated strategies to improve their prediction.

7. Future Work

While this study provides encouraging evidence for the effectiveness of context-aware architectures in collision prediction, several open challenges remain.

Future research directions include expanding the dataset to further enhance generalization, improving mean time-to-alert (mTTA) to reduce false alerts in real-world systems, and addressing long-tail categories to better evaluate and predict diverse and rare driving scenarios. Our model ability to recognize complex and risky situations even before collisions occurred as illustrated in Figure 5 suggests the potential to extend collision prediction models beyond a binary formulation, toward a three-level taxonomy: *normal, warning, and alert*. Such an approach could be particularly beneficial for autonomous driving systems, enabling adaptive decision-making based on momentary risk levels. We refer the readers to our project page for full length examples [3].

Ultimately, advancing reliable and context-aware collision prediction can contribute significantly to the broader goal of safer, more anticipatory driver assistance systems, and may play a key role in bridging the gap between current ADAS technologies and fully autonomous driving.

References

[1] M. Assran, A. Bardes, D. Fan, Q. Garrido, R. Howes, M. Komeili, M. Muckley, A. Rizvi, C. Roberts, K. Sinha, A. Zholus, S. Arnaud, A. Gejji, A. Martin, F. R. Hogan, D. Dugas, P. Bojanowski, V. Khalidov, P. Labatut, F. Massa, M. Szafraniec, K. Krishnakumar, Y. Li, X. Ma, S. Chandar, F. Meier, Y. LeCun, M. Rabbat, and N. Ballas, "V-jepa 2: Self-supervised video models enable understanding, prediction and planning," 2025.

[2] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun, "Anticipating accidents in dashcam videos," in *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part IV 13*. Springer, 2017, pp. 136–153.

[3] Y. Yao, X. Wang, M. Xu, Z. Pu, Y. Wang, E. Atkins, and D. J. Crandall, "Dota: Unsupervised detection of traffic anomaly in driving videos," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 444–459, 2022.

[4] J. Fang, D. Yan, J. Qiao, J. Xue, and H. Yu, "Dada: Driver attention prediction in driving accident scenarios," *IEEE transactions on intelligent transportation systems*, vol. 23, no. 6, pp. 4959–4971, 2021.

[5] D. C. Moura, S. Zhu, and O. Zvitia, "Nexar dashcam crash prediction challenge," https://kaggle.com/competitions/nexar-collision-prediction, 2025, kaggle.

[6] D. Moura, S. Zhu, and O. Zvitia, "Nexar dashcam collision prediction dataset and challenge," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 2583–2591.

[7] T. Suzuki, H. Kataoka, Y. Aoki, and Y. Satoh, "Anticipating traffic accidents with adaptive loss and large-scale incident db," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3521–3529.

[8] M. M. Karim, Y. Li, R. Qin, and Z. Yin, "A dynamic spatial-temporal attention network for early anticipation of traffic accidents," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9590–9600, 2022.

[9] A. V. Malawade, S.-Y. Yu, B. Hsu, D. Muthirayan, P. P. Khargonekar, and M. A. Al Faruque, "Spatiotemporal scene-graph embedding for autonomous vehicle collision prediction," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9379–9388, 2022.

[10] W. Bao, Q. Yu, and Y. Kong, "Drive: Deep reinforced accident anticipation with visual explanation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7619–7628.

[11] J. Fang, L.-l. Li, J. Zhou, J. Xiao, H. Yu, C. Lv, J. Xue, and T.-S. Chua, "Abductive ego-view accident video understanding for safe driving perception," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 22 030–22 040.

[12] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "Slowfast networks for video recognition," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6202–6211.

[13] Z. Liu, J. Ning, Y. Cao, Y. Wei, Z. Zhang, S. Lin, and H. Hu, "Video swin transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 3202–3211.

[14] Z. Tong, Y. Song, J. Wang, and L. Wang, "Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training," *Advances in neural information processing systems*, vol. 35, pp. 10 078–10 093, 2022.

[15] A. Bardes, Q. Garrido, J. Ponce, X. Chen, M. Rabbat, Y. LeCun, M. Assran, and N. Ballas, "Revisiting feature prediction for learning visual representations from video," *arXiv preprint arXiv:2404.08471*, 2024.

[16] "Euro ncap assessment protocol – collision avoidance," European New Car Assessment Programme (Euro NCAP), Tech. Rep., 2020, version 1.0.4.1, October 2020. Defines Forward Collision Warning (FCW) and Autonomous Emergency Braking (AEB) test procedures. [Online]. Available: https://www.euroncap.com/media/80154/euro-ncap-assessment-protocol-sa-collision-avoidance-v1041.pdf

[17] J. Li, "Vehicle-cv-adas," https://github.com/jason-li-831202/Vehicle-CV-ADAS, 2023, accessed: 2025-10-02.

[3] https://www.nexar-ai.com/badas

[18] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint*, 2024, arXiv:2410.17725.

[19] Z. Qin, P. Zhang, and X. Li, "Ultra fast deep lane detection with hybrid anchor driven ordinal classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2022.

[20] B. Psomas, D. Christopoulos, E. Baltzi, I. Kakogeorgiou, T. Aravanis, N. Komodakis, K. Karantzalos, Y. Avrithis, and G. Tolias, "Attention, please! revisiting attentive probing for masked image modeling," 2025. [Online]. Available: https://arxiv.org/abs/2506.10178