Appendix:
CycleCrash: A Dataset of Bicycle Collision Videos for Collision Prediction and Analysis

Nishq Poorav Desai   Ali Etemad   Michael Greenspan
Queen's University, Canada
{n.desai, ali.etemad, michael.greenspan}@queensu.ca

A. Dataset Availability

The dataset is publicly available and can be accessed at https://github.com/DeSinister/CycleCrash/. The initial release of the dataset comprises:

- List of links for all the videos in the dataset, with their start and end timestamps.
- All 13 annotations described in the paper, for all the video clips.
- PyTorch-based implementation for the pre-processing used in the paper, and data loader codes for efficiency and consistency.
- PyTorch-based implementation for VidNeXt and its ablation variants.

B. Keyword Search for dataset creation

For creating the dataset, the following keywords were used to search and collect the videos in CycleCrash: Bicycle accident dashcam; Bicycle crash; Bicycle crash dashcam; Bicyclist car accident; Bicyclist car collision dashcam; Bicyclist collision; Bicyclist collision compilation; Bicyclist crash; Bike accident; Bike crash; Bike crash compilation; Bike hit car dashcam; Bike hit dashcam; Bike hit and run; Car hit cyclist; Cyclist accident; Cyclist accident dashcam; Cyclist being hit; Cyclist being hit compilation; Cyclist collision dashcam; Cyclist crash dashcam; Cyclist fall down; Cyclist fault dashcam; Cyclist hit and run; Cyclist hit dashcam; Cyclist hit pedestrian; Cyclist hits bus; Cyclist hits pole; Cyclist hitting car; Cyclist hits road sign; Cyclist hits train; Cyclist near miss; Dashcam cycle accident; Dashcam cycle accident compilation; Hit cyclist.

For the videos exhibiting cyclists navigating safely in an urban traffic scenario, without depicting risky behaviour, we use the clips from driving tour videos collected using the following search keywords: City Driving tour; Downtown Driving tour; Urban Driving tour; Urban City Driving tour.

C. Multi-tasking

We performed a multi-task version of our model VidNeXT and present the results below in Tab. A1. The multitasking model slightly underperforms possibly due to task interference.

D. Future Research Directions

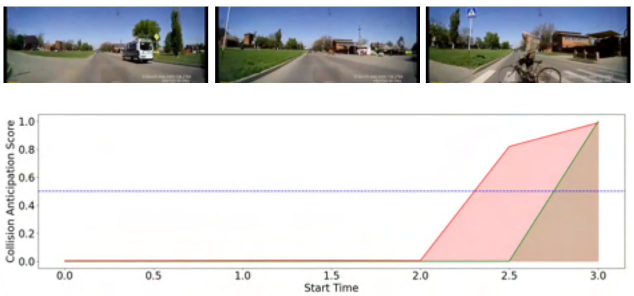
We visualize a few examples of the failure cases of our method for the Collision Anticipation task, in Fig. A1. We observe that in these instances, the model predicts an accident earlier than it actually occurs. Investigating the underlying reasons for this and designing effective solutions are an interesting line of future inquiry. Additionally to enhance overall training and performance, specialized losses, long-term recognition methods, or spatiotemporal augmentation techniques could be employed to upsample the minority classes. Moreover, incorporating scene-related labels such as 'type of object involved,' 'camera position,' and 'ego-vehicle involved' could provide additional context to improve model performance. Finally, we have introduced tasks such as 'Right-of-way Prediction' and 'Severity Classification', which have the potential to be extended to include interactions among autonomous vehicles and other vehicles (cars, motorcycles, buses, etc).

E. Additional Data Statistics

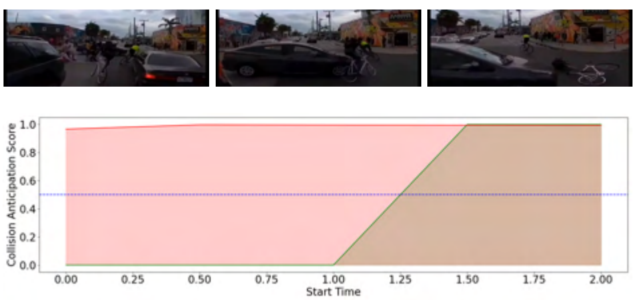
We present more statistics for our dataset in Fig. A2 and Fig. A3. Here, Fig. A2 (*i*) indicates if the ego-vehicle (the

Method	Risk		Right-of-way		Collision		Time-to-coll.	Severity		Fault		Age		Direction		Object	
	Acc.↑	F1↑	Acc.↑	F1↑	Acc.↑	F1↑	MSE↓	Acc.↑	F1↑	Acc.↑	F1↑	Acc.↑	F1↑	Acc.↑	F1↑	Acc.↑	F1↑
ViViT [1] (single-task)	65.12	39.06	52.84	53.74	57.01	<u>69.92</u>	**1.33**	47.51	24.47	53.37	50.42	93.56	66.34	36.29	27.99	**46.30**	26.34
X3D-M [2] (single-task)	64.76	38.75	59.83	<u>57.57</u>	63.72	61.08	1.44	54.45	24.70	52.16	52.19	94.34	53.78	<u>47.82</u>	31.85	42.72	23.79
VidNeXt (single-task)	66.20	**41.96**	**64.28**	57.51	64.84	**70.84**	<u>1.38</u>	**59.66**	**31.78**	**65.16**	52.51	**94.57**	**67.88**	**47.94**	31.20	42.31	<u>28.37</u>
VidNeXt (multi-task)	51.86	31.34	59.09	55.38	62.86	62.87	2.55	45.17	29.62	57.16	51.43	93.56	65.66	44.85	31.04	40.4	27.64

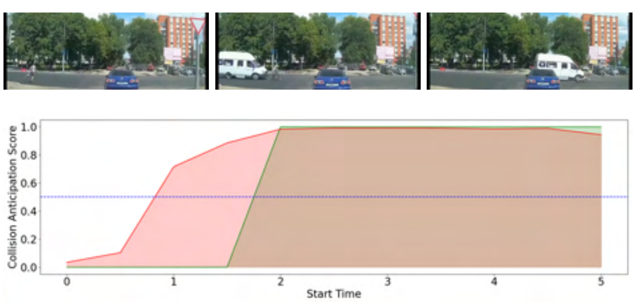
Table A1. Multi-task version of VidNeXT.



(a) Failure Case 1



(b) Failure Case 2



(c) Failure Case 3

Figure A1. A few examples of failure cases of VidNeXt for the Collision Anticipation task. The red line represents the collision anticipation scores predicted by the model, while the green line shows the ground truth labels. The blue line marks the threshold at a score of 0.5, which is used to determine whether a collision is anticipated or not.

vehicle carrying the dashcam) is involved in the collision in the video. We observe that most of the videos do not involve the ego-vehicle in the accident. Furthermore, we

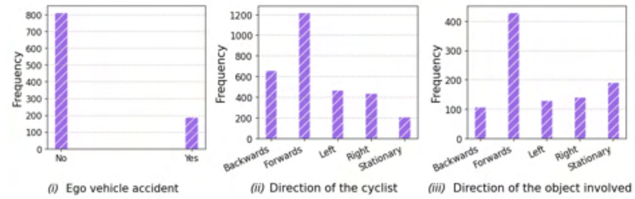


Figure A2. Distribution of CycleCrash data for (*i*) ego vehicle accidents, (*ii*) direction of cyclists, (*iii*) direction of other objects involved.

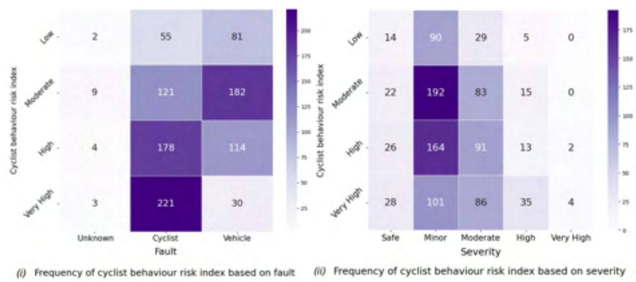


Figure A3. Relationship between (*i*) cyclist behaviour risk index and fault, (*ii*) cyclist behaviour risk index and severity.

present the distributions of the directions of the cyclists in Fig. A2 (*ii*), and the directions of the objects involved in Fig. A2 (*iii*). We observe that the most frequently occurring direction was forward in both cyclists and other objects involved. Fig. A3 (*i*) displays the frequency of fault based on the cyclist behaviour risk index. It is observed that as the cyclist behaviour risk index increases, the number of cyclists being at fault also rises, and vice-versa. Moreover, the relationship between cyclist behaviour risk index and severity is presented in Fig. A3 (*ii*). We notice an increasing number of collisions in each severity group as the cyclist behaviour risk index increases. The highest number of collisions was recorded with minor collision severity and low cyclist behaviour risk indexes.

F. Additional Visualizations

We present additional samples for different video streams, showing diversity in interactions in the CycleCrash dataset in Fig. A4 to Fig. A7.

Figure A4. Additional visualizations of cyclist interactions with cars from the CycleCrash dataset.

Figure A5. Additional visualizations of cyclists falling on their own from the CycleCrash dataset.

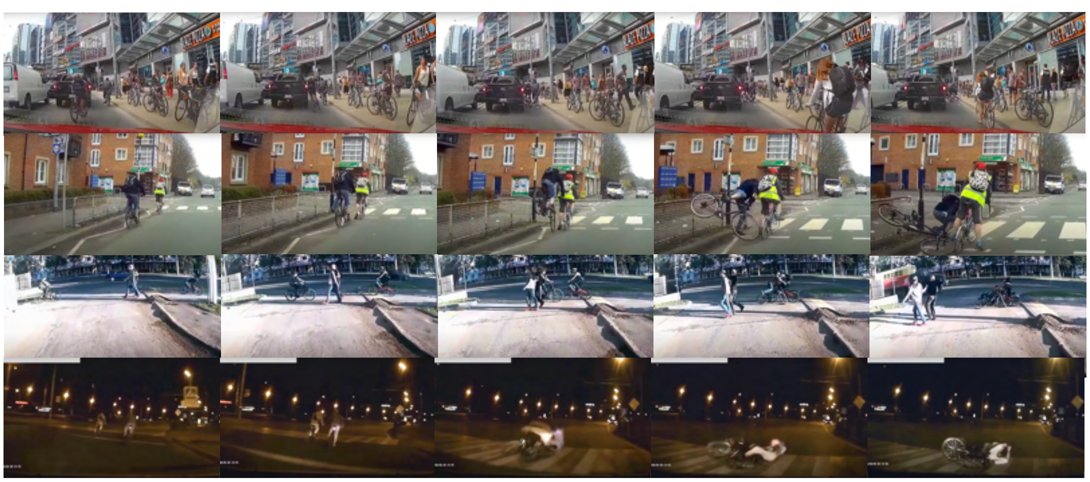


Figure A6. Additional visualizations of cyclist interactions with other cyclists from the CycleCrash dataset.

Figure A7. Additional visualizations of cyclist interactions with other motor vehicles from the CycleCrash dataset.

References

[1] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A video vision transformer. In *IEEE/CVF International Conference on Computer Vision*, 2021.

[2] Christoph Feichtenhofer. X3d: Expanding architectures for efficient video recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.