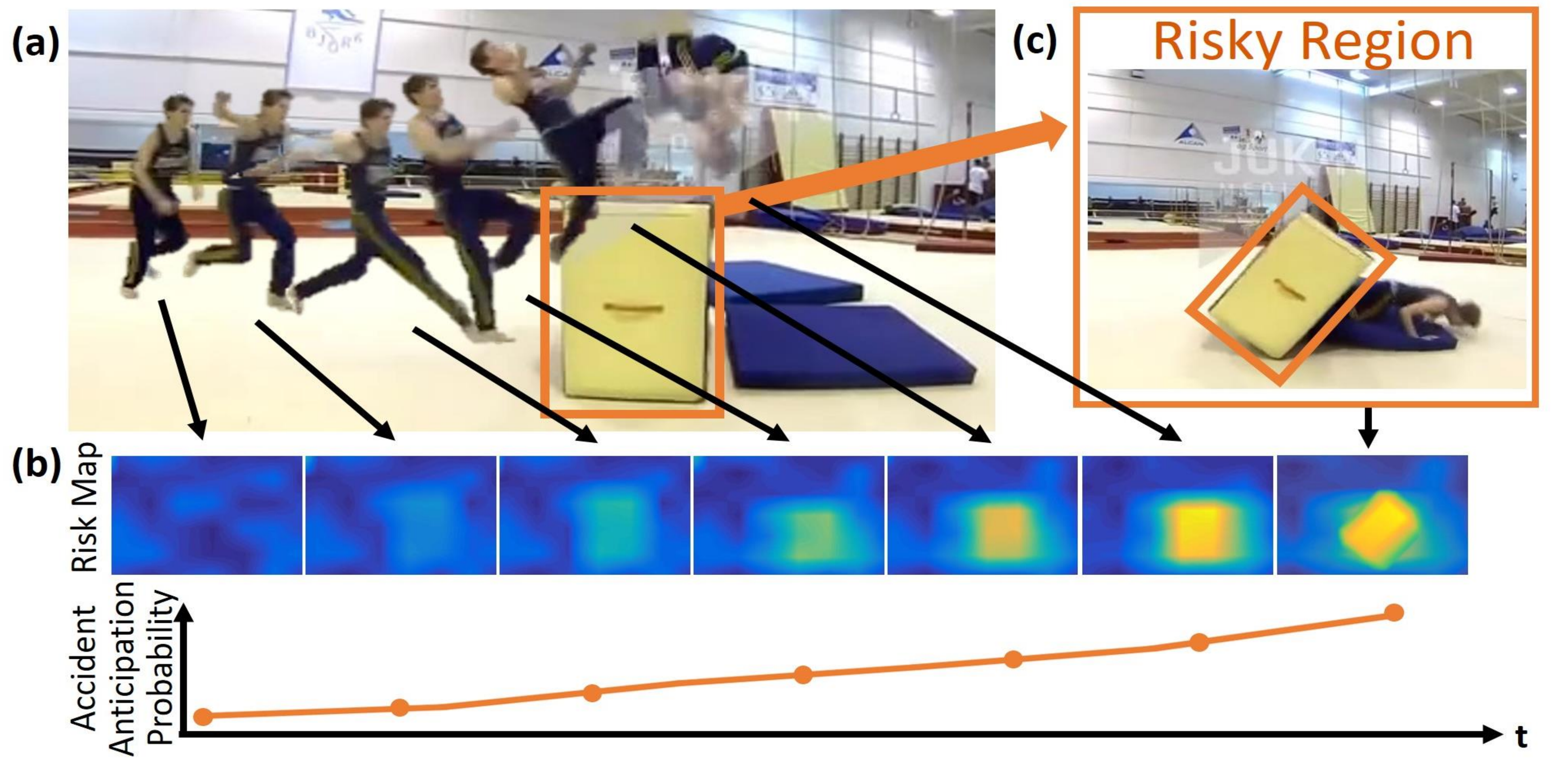
Agent-Centric Risk Assessment: Accident Anticipation and Risky Region Localization

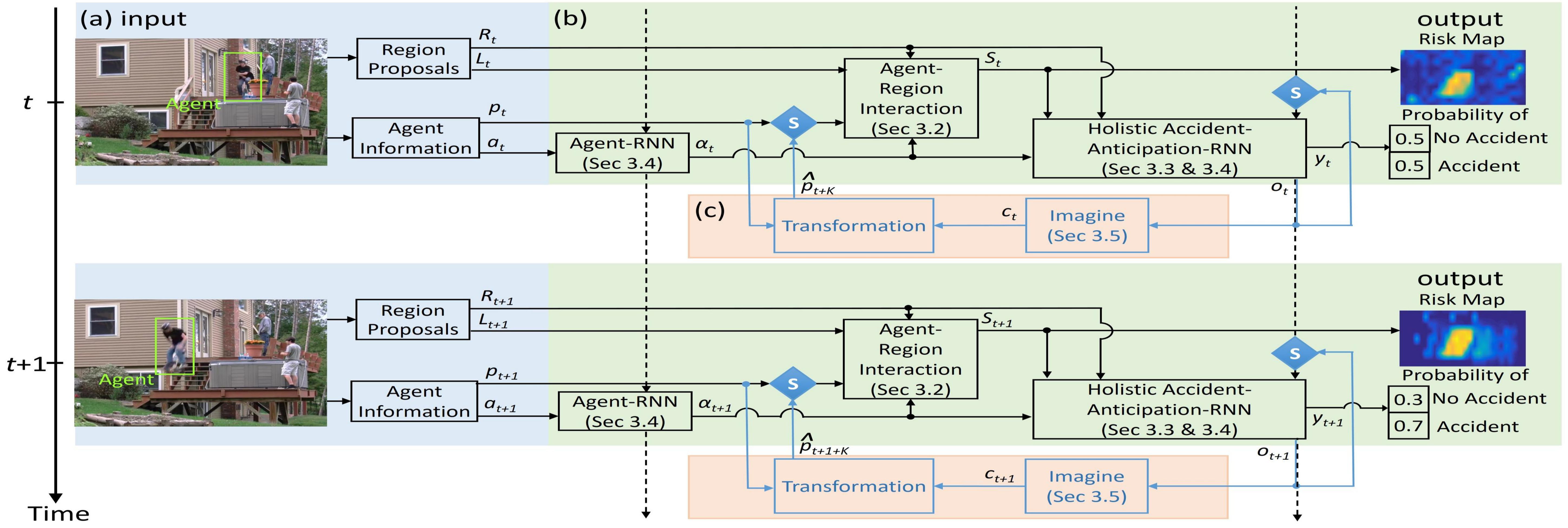
Kuo-Hao Zeng*†, Shih-Han Chou*, Fu-Hsiang Chan*, Juan Carlos Niebles†, Min Sun*
†Stanford University *National Tsing Hua University

Risk Assessment



- Agent-centric: We define our problem as risk assessment with respect to the agent appearing at each time step.
- Risk assessment contains two domains, (1) in the temporal domain (accident anticipation) and (2) in spatial domain (risky region localization).
- Accident anticipation is to predict an accident before it occurs.
- Risky region localization is to spatially localize the regions in the scene that might be involved in a future accident.

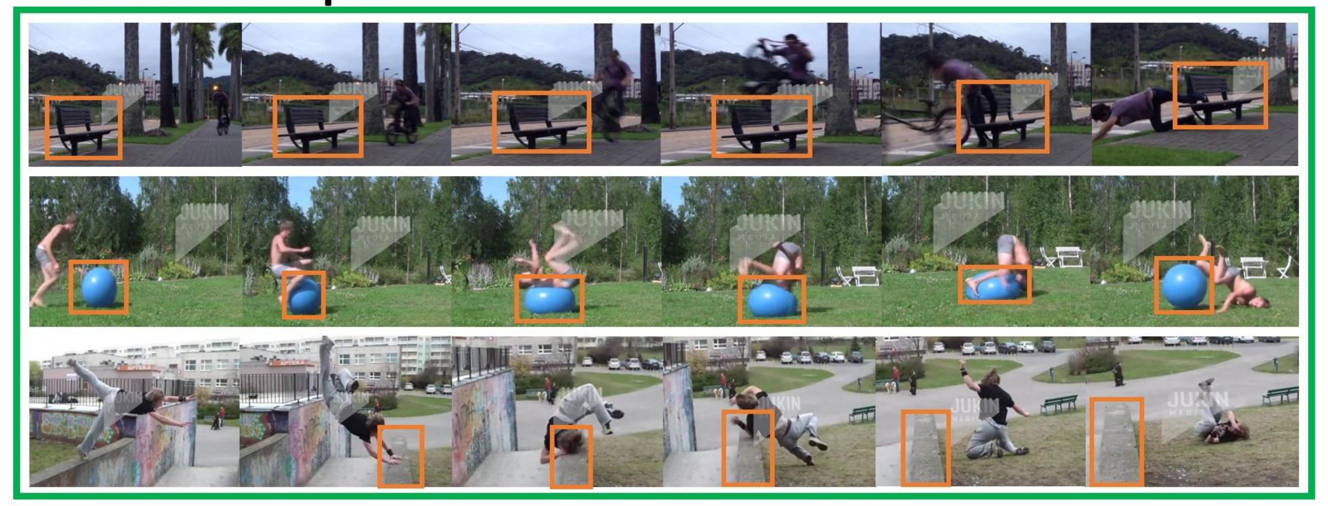
Contribution – Our Proposed Model



1. We utilize the dynamic parameter layer to efficiently model the relative spatial relation and coupled appearance between agent and region (Panel (b)).
2. We use the generative property of RNN to self-train it to encode the behavior of the agent as well as generate (i.e., imagine) its future trajectory (Panel (c)).
3. The imagined future trajectory becomes new inputs to our model to assess risk in a longer term (Panel (c) to Panel (b)).

Our Epic Fail Dataset

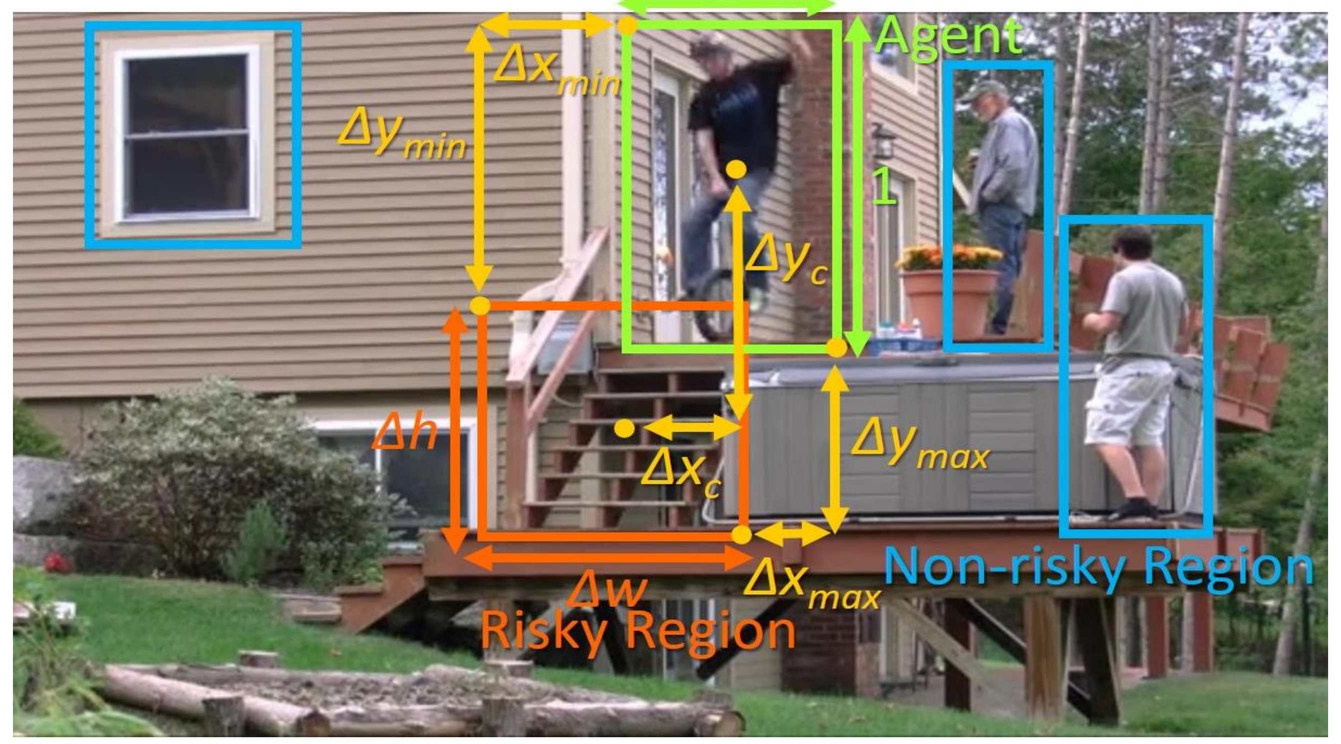
Positive examples



Negative example

- We ask annotator to annotate the region ***causing*** the failure event.
- The agents and the risky regions are annotated by 2D bounding boxes.
- We manually identify the time when accident occurs in a subset of raw videos and sample short videos of 3-4 seconds from the subset.

Relative Configuration



- Risk assessment of all regions are illustrated with respect to the agent (green box).
- In our agent-centric perspective, the orange box indicates a risky region and the blue boxes indicate non-risky regions.
- We normalize the horizontal and vertical axes of the agent separately to unit one.

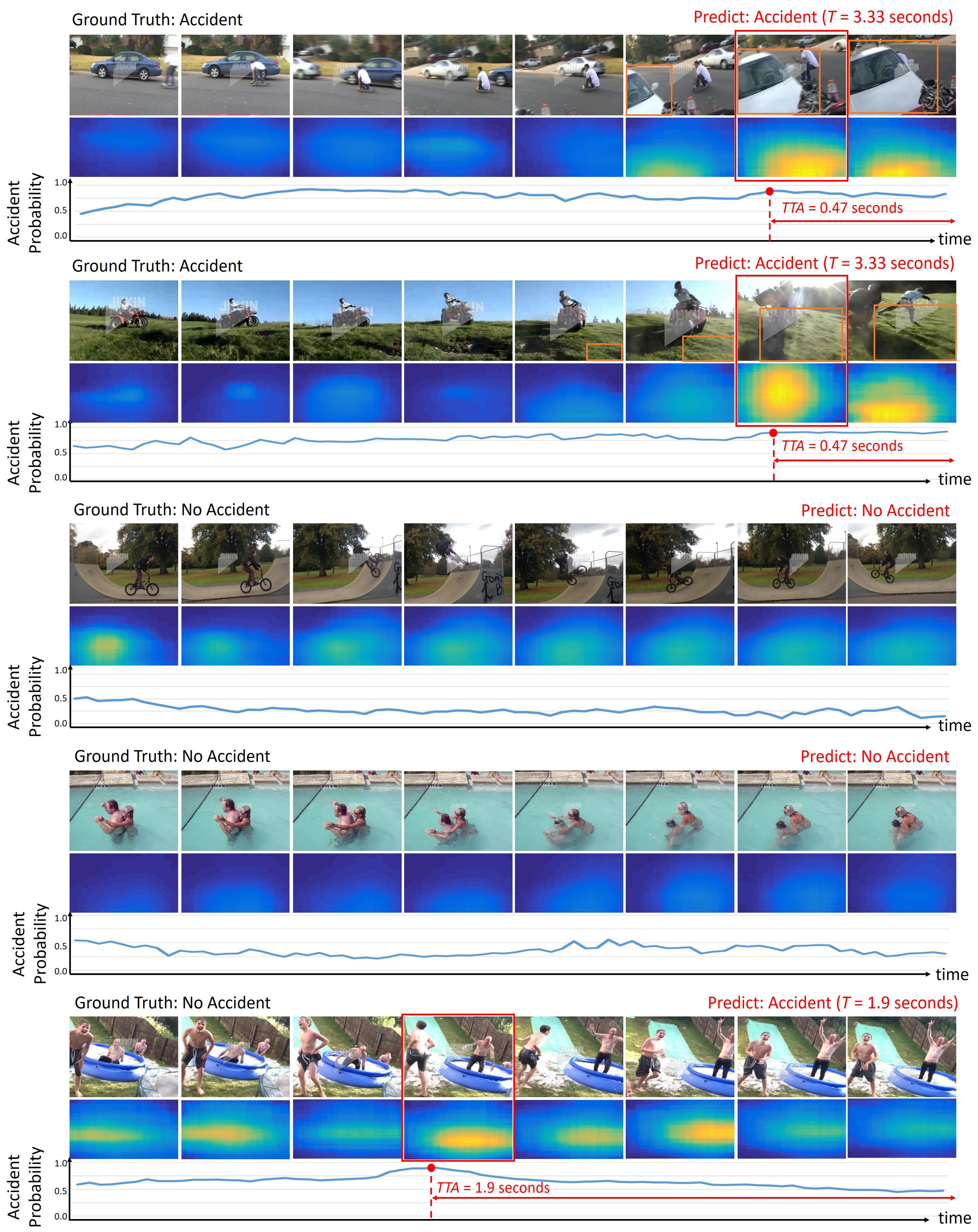
Quantitative Results

Dataset	EF		SA	
w/o memory	mAP (%)	ATTA (s)	mAP (%)	ATTA (s)
R*CNN[1]	68.6	2.47	40.7	2.64
RA	72.2	2.10	47.8	2.55
RAI	72.4	2.13	48.8	2.62
w memory	mAP (%)	ATTA (s)	mAP (%)	ATTA (s)
DSA[2]	45.7	1.16	48.1	1.34
SP[3]	40.5	0.88	47.3	1.66
L-R*CNN	69.6	2.54	37.4	3.13
L-RA	74.2	1.84	49.1	**3.04**
L-RAI	**75.1**	**2.23**	**51.4**	3.01

Dataset	EF	SA
w/o memory	mAP (%)	mAP (%)
R*CNN	3.47	34.7
RA	12.3	40.1
RAI	14.1	43.1
w memory	mAP (%)	mAP (%)
L-R*CNN	3.5	35.6
L-RA	14.0	43.8
L-RAI	**15.1**	**45.4**
Oracle	75.7	92.8

- ATTA: It's the average version of Time-to-accident (TTA) and it's used to evaluate how early the model is able to predict an accident.

Qualitative Results



Acknowledgement

科技部 Ministry of Science and Technology, MEDIATEK, NTU IOX center, NOVATEK, Panasonic

QrCode

[1] G. Gkioxari, R. Girshick, and J. Malik. Contextual action recognition with R*CNN. In CVPR, 2016.
[2] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun. Anticipating accidents in dashcam videos. In ACCV, 2016.
[3] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In CVPR, 2016.