Supplementary Material

EchoTraffic: Enhancing Traffic Anomaly Understanding with Audio-Visual Insights

Zhenghao Xing¹*, Hao Chen¹*, Binzhu Xie¹, Jiaqi Xu¹, Ziyu Guo¹, Xuemiao Xu², Jianye Hao³, Chi-Wing Fu¹, Xiaowei Hu^{4†}, and Pheng-Ann Heng¹ ¹ The Chinese University of Hong Kong ² South China University of Technology ³ Tianjin University ⁴ Shanghai Artificial Intelligence Laboratory

^{*}Joint first authors.

[†]Corresponding author (huxiaowei@pjlab.org.cn).

There are four parts in this supplementary material.

Part 1 - Qualitative Results: Additional results illustrating EchoTraffic's performance in diverse traffic scenarios.

Part 2 - Quantitative Results: Additional evaluation of fine-tuning Video-LLaMA2 under EchoTraffic's settings and comparisons with Gemini 2.0.

Part 3 - Dataset and Experiment Details: Details of the AV-TAU dataset and the configuration of the GPT-Eval metric.

Part 4 - Limitations and Future Work: Discussion on current limitations and potential future research.

Part 1 - Qualitative Results

This section showcases EchoTraffic's performance across various traffic scenarios, comparing its robustness and accuracy with VideoLLaMA2 [2] and Holmes-VAD [4].

(a) Description	n: What anomaly is presented in the video?
VideoLLaMA2:	The video shows a truck driving on the highway with traffic behind it. The unexpected event in the video scene is a car accident on the highway. The truck driver is seen stopping and helping the drivers involved in the accident.
Holmes-VAD:	No, the video shows a blue truck driving down a highway, passing by other vehicles, and the camera capturing the view from inside the truck, which is a normal and common scene of everyday traffic.
EchoTraffic:	This video was recorded by a dash cam. In this video, the car in the video rear-ended with a blue truck in front of it while driving on the highway. You can hear the sound of the video car colliding with the truck in front and the alarm sound.
(b) Causation:	Why did the unusual event occur in the video?
VideoLLaMA2:	The video shows a truck driving on the highway with traffic. The anomaly in the video is the sudden appearance of a car next to the truck. This unexpected event caused the driver of the truck to react quickly and brake suddenly, resulting in a dangerous situation.
Holmes-VAD:	The video shows a blue truck driving down a highway, passing by other vehicles, and then suddenly veering off the road and crashing into a guardrail, which is abnormal because it deviates from the expected safe and controlled driving behavior.
EchoTraffic:	The cause of the accident was that the driver of the video car failed to maintain a safe distance from the truck in front, and the truck driver took over the road but failed to notice the movements of the video car behind in time.
(c) Prevention	: How could the anomaly have been prevented in the video?
VideoLLaMA2:	The video shows a truck driving on the highway with traffic behind it. The incident in the video could have been avoided if the truck driver had maintained a safe distance from the car behind it. By keeping a safe following distance, the truck driver would have had more time to react to any sudden changes in traffic or road conditions, reducing the likelihood of a collision.
Holmes-VAD:	The video shows a blue truck driving down a highway, passing by other vehicles, and then suddenly veering off the road and crashing into a guardrail, which is abnormal because it deviates from the expected safe and controlled driving behavior.
EchoTraffic:	To avoid this accident, the driver of the video car should maintain a safe distance from surrounding vehicles, slow down and ensure sufficient reaction time. The driver of the truck should carefully observe the driving conditions of vehicles on the road.
(d) Response:	How should the event be managed after the anomaly?
VideoLLaMA2:	The video shows a truck driving on the highway with traffic behind it. The anomaly in the video is that the truck suddenly stops, and the cars behind it start honking. This unexpected event creates a tense atmosphere on the road. Participants should be cautious and react quickly to avoid any accidents.
Holmes-VAD:	The video shows a blue truck driving down a highway, passing by other vehicles, and the camera capturing its movement from different angles, but suddenly, the truck veers off the road and crashes into a guardrail, which is an abnormal event. Participants should react by stopping the video and discussing the anomaly, which is the truck's sudden loss of control and subsequent crash, and they should consider the factors that may have contributed to this event, such as the driver's actions, road conditions, or vehicle malfunction.
EchoTraffic:	After the accident occurs, the video car should turn on the double flashes to warn the vehicles behind, then record the situation and damage details at the scene of the accident as evidence. After determining the responsibility, the vehicle should be moved to a safe location immediately to prevent traffic jams and further dangers.

Figure 1. A highway scenario captured by a dash camera, illustrating the challenges of detecting collisions with minimal visual context.



(a) Descriptior	n: What anomaly is presented in the video?		
VideoLLaMA2:	The video shows a car driving down the street at night, with the headlights illuminating the road ahead. The sound of the engin and the hum of the tires can be heard in the background. Suddenly, the car's windshield wipers start moving erratically, as they are trying to clear the windshield of something. This is an unusual event that takes place in the video.		
Holmes-VAD:	Yes, the video shows a car driving down a street at night with its headlights on, but suddenly, the car veers off the road a crashes into a tree, which is an abnormal event because it deviates from the expected safe and controlled driving behavior.		
EchoTraffic:	The video shows a car driving on a rainy road at night, shot from a dashcam in the car. The car in the video collides with a black car that is turning left. In this video, you can clearly hear the sound of the video car colliding with the black car.		
(b) Causation:	Why did the unusual event occur in the video?		
VideoLLaMA2:	The video shows a car driving down the street at night, with the lights on. The anomaly in the video is the red light that appears on the car's dashboard. This could indicate that there is something wrong with the car's brakes or other systems.		
Holmes-VAD:	The video shows a car driving down a street at night with its headlights on, but suddenly, the car veers off the road and crashes into a tree, which is an abnormal event because it deviates from the expected safe and controlled driving behavior.		
EchoTraffic:	The cause of the accident is that the driver of the video car is distracted and fails to notice the dynamics of the black car in front in time. Additionally, the car is driving too fast, reducing the reaction time and braking distance.		
(c) Prevention	: How could the anomaly have been prevented in the video?		
VideoLLaMA2:	This type of anomaly could be mitigated in the future by implementing stricter safety measures and regulations for drivers . Additionally, advancements in technology such as self-driving cars and improved vehicle sensors could help prevent accidents from occurring. It is also important for drivers to remain vigilant and attentive on the road , especially during adverse weather conditions like rain or fog.		
Holmes-VAD:	To mitigate the anomaly of a car driving down a street at night with its headlights on, which is abnormal because it's unusual for a car to have its headlights on in a well-lit area, the following measures could be taken:\n\n1. Implement stricter traffic regulations, such as fines or penalties for drivers who leave their headlights on in well-lit areas, to encourage safe driving practices.\n2. Increase public awareness campaigns about the importance of using headlights in low-light conditions to improve road safety.\n3. Invest in better street lighting and traffic management systems to ensure that all areas are adequately lit, reducing the need for drivers to rely on their headlights.		
EchoTraffic:	To avoid this accident, the driver of the video car should maintain a safe distance from surrounding vehicles, slow down and ensure sufficient reaction time. The driver of the black car should make careful observations before turning and signal intentions clearly to other drivers.		
(d) Response:	How should the event be managed after the anomaly?		
VideoLLaMA2:	The video shows a car driving down the street at night, with the lights on. The anomaly in the video is that there are no other cars or people visible on the street, despite the car's headlights illuminating the road ahead. This could be due to various factors, such as poor lighting conditions or a malfunctioning car sensor. Participants in the video should exercise caution and proceed with caution, as there may be hidden dangers or obstacles on the road.		
Holmes-VAD:	The video shows a car driving down a street at night with its headlights on, passing by other cars and pedestrians, and the camera capturing the view from inside the car. This is a normal video because it depicts a common and everyday scene of a car driving at night, with no unusual or abnormal events occurring.		
EchoTraffic:	After the accident occurs, the driver of the vehicle should stop immediately and turn on the hazard lights to warn the vehicles behind that an accident has occurred. Then get off the vehicle to check the injuries, call the emergency number immediately and provide necessary first aid measures. Contact the traffic police and relevant departments to handle the matter .		

Figure 2. A rainy-night scenario captured fy a dash camera, demonstrating the challenges of detecting collisions under low-visibility conditions.

D skn or 4	
-11/11-	
(a) Description	n: What anomaly is presented in the video?
VideoLLaMA2:	The video shows a snowmobile driving down a snowy road, and suddenly, it flips over on its side. The man standing next to the snowmobile seems surprised by the unexpected incident.
Holmes-VAD:	Yes, the video shows a group of people engaging in a snowy activity, but suddenly , one person falls down and another person rushes to help them up , which is an abnormal event because it deviates from the expected norm of people continuing with the activity without any interruptions or assistance.
EchoTraffic:	The video shows a car driving on a snow-covered road. The car loses control and crashes into a roadside obstacle. In the video, you can hear the sound of the vehicle's engine, the sound of the vehicle hitting the obstacle, and then the sound of a crowd screaming.
(b) Causation:	Why did the unusual event occur in the video?
VideoLLaMA2:	The video shows a car driving down a snowy road with people watching. The main source of sound in the video is the sound of the car's engine and the people cheering. The anomaly in the video is the car flipping over on the snowy road, which is likely due to the driver losing control of the vehicle.
Holmes-VAD:	The video shows a group of people engaging in a snowy activity, with some individuals falling down and others rushing to help them up, which is abnormal because it deviates from the expected norm of people enjoying the activity without any accidents or injuries.
EchoTraffic:	The cause of the accident is that the road surface is slippery due to snow, which reduces the grip of the tires and makes it difficult for the driver of the blue car to maintain stability. The car is driving too fast, making it impossible to respond in time when encountering obstacles.
(c) Prevention	: How could the anomaly have been prevented in the video?
VideoLLaMA2:	One way to mitigate this type of anomaly in the future is by implementing stricter safety measures and regulations for extreme winter sports activities. This could include mandatory training for drivers and participants, as well as the use of specialized equipment such as helmets and protective gear.
Holmes-VAD:	To mitigate the anomaly of a person falling off a sled and being hit by another person, the following measures can be taken: \n\n1. Ensure that the sled is designed with safety features, such as a protective barrier or padding, to prevent collisions and injuries.\n2. Implement a system of communication and signaling between the sled riders and the person controlling the sled, to coordinate their movements and avoid unexpected collisions.\n3. Train and educate participants on safe sledding practices, including proper sled control, stopping techniques, and awareness of surroundings.\n4. Establish designated areas for sledding activities, with clear boundaries and supervision, to minimize the risk of accidents and injuries.
EchoTraffic:	To avoid this accident, the driver of the blue car should slow down and ensure sufficient reaction time; stay focused and avoid distracted driving.
(d) Response:	How should the event be managed after the anomaly?
VideoLLaMA2:	The video shows a car flipping on a snowy road, creating a dangerous situation for the driver and passengers. The unexpected event of the car flipping should be managed by ensuring that everyone involved is safe and receiving medical attention if necessary. Additionally, the car should be towed to a safe location to prevent further accidents or injuries.
Holmes-VAD:	The video shows a person riding a snowmobile at a high speed, losing control, and crashing into a tree, causing a loud noise and a cloud of snow to be kicked up. This event is abnormal because it involves a sudden loss of control and a severe crash, which is a hazardous and potentially harmful situation. To manage this situation, the person should be immediately stopped and checked for injuries, and the snowmobile should be inspected for any damage that may have occurred.
EchoTraffic:	After an accident, the driver should first ensure the safety of himself and the passengers in the car, quickly assess the injuries and remain calm. If the vehicle rolls into a ditch and cannot be freed by itself, stay in the car with the seat belt fastened until it is safe to leave. If it is safe to evacuate, you should immediately stay away from the vehicle to avoid secondary dangers such as fuel leakage. At the same time, call the emergency rescue number as soon as possible, report the location and accident situation, and seek professional rescue.

Figure 3. A snowy countryside road captured by the camera, highlighting the difficulty of collision detection in adverse weather conditions.

D	
(a) Descriptio	n: What anomaly is presented in the video?
VideoLLaMA2:	The video shows a truck driving down the road with a car behind it, and then a car drives down the road with a truck in the background. The truck is seen driving down the road with a car on the side, and then a truck is parked on the side of the road. The truck is also seen driving down the road with a truck in the background.
Holmes-VAD:	The video shows a white car driving down a road, passing by a red truck, and then turning left, which is a normal and everyday traffic scene.
EchoTraffic:	This video was recorded by a dash cam. In this video, a truck collided with a black car and hit the guardrail. In this video, you can clearly hear the sound of a truck colliding with the black car, as well as the sound of the truck's brakes.
(b) Causation:	Why did the unusual event occur in the video?
VideoLLaMA2:	The video shows a truck driving down the road with a car behind it, and then a car drives down the road with a truck behind it. The anomaly in the video is the sudden appearance of a red circle on the screen, which seems to be a glitch or error in the video feed.
Holmes-VAD:	The video shows a white car driving down a road and suddenly veering off the road and crashing into a tree, which is abnormal because it depicts a severe car accident that deviates from the normal flow of traffic.
EchoTraffic:	The cause of the accident was that the truck driver was driving too fast and did not pay attention to the road conditions ahead.
(c) Prevention	: How could the anomaly have been prevented in the video?
VideoLLaMA2:	The video shows a truck driving down the road with a car behind it, and then a car drives down the road with a truck in the background. The truck and car are involved in a collision on the road. To reduce the likelihood of this event happening, both the truck driver and the car driver should follow traffic rules and regulations, maintain a safe distance from each other, and drive at a safe speed.
Holmes-VAD:	The video shows a white car driving down a road and a truck driving in the opposite direction, with the truck suddenly veering off the road and crashing into a tree, which is an abnormal event. \n\nTo reduce the likelihood of this event happening, the truck driver should be more cautious and maintain a safe distance from the white car, and the car driver should also be vigilant and drive at a safe speed, especially when approaching a truck that is driving in the opposite direction.
EchoTraffic:	To avoid this accident, the driver of the red truck should maintain a safe distance from surrounding vehicles, slow down and ensure sufficient reaction time; stay focused and avoid distracted driving.
(d) Response:	How should the event be managed after the anomaly?
VideoLLaMA2:	In response to the anomaly, both the truck driver and the car driver should take immediate action. The truck driver should assess the damage caused by the collision and determine if any further action is necessary, such as reporting the incident to the authorities or seeking medical attention for any injured passengers . The car driver should also assess the damage to their vehicle and consider reporting the incident to the authorities if it was a significant collision.
Holmes-VAD:	The video shows a white car driving down a road and suddenly veering off the road and crashing into a tree, which is an abnormal event. The driver of the white car should be immediately checked for injuries and the vehicle should be towed away, while the police should be called to investigate the cause of the accident.
EchoTraffic:	The driver of the accident vehicle should pay attention to the vehicles behind, turn on the warning lights to remind other drivers to pay attention to safety and avoid further traffic accidents, and check whether all people are injured. If there are people injured, call the emergency number immediately and provide necessary first aid measures. Then contact the traffic police or relevant departments to report the accident, record the situation at the scene of the accident and the details of the damage, take photos as evidence, and exchange contact information and insurance information with the other driver for subsequent claims and processing.

Figure 4. A highway scenario captured by a surveillance camera, with part of the accident occurring outside the camera's view.

Part 2 - Quantitative Results

Table 1. Performance of VideoLLaMA2 [2] and EchoTraffic across different tasks and metrics. Fine-tuning was performed on VideoLLaMA2 using EchoTraffic's settings. EchoTraffic consistently outperforms VideoLLaMA2 by a significant margin.

Task	Metric	VideoLLaMA2	EchoTraffic
	BLEU	0.2113	0.2541
Description	Rouge	0.3860	0.4380
	MoverScore	0.6037	0.6205
-	BERTScore	0.8911	0.9102
	GPT-Eval	0.6526	0.6843
	BLEU	0.1939	0.2464
	Rouge	0.3203	0.3509
Causation	MoverScore	0.5598	0.6077
	BERTScore	0.8876	0.9017
	GPT-Eval	0.7840	0.8301
	BLEU	0.1559	0.1902
	Rouge	0.3790	0.4095
Prevention	MoverScore	0.5812	0.6120
	BERTScore	0.8801	0.8974
	GPT-Eval	0.7912	0.8462
	BLEU	0.2777	0.3171
	Rouge	0.3401	0.3994
Response	MoverScore	0.5992	0.6310
	BERTScore	0.8698	0.9071
	GPT-Eval	0.7776	0.8678
	BLEU	0.2097	0.2520
	Rouge	0.3564	0.3994
Overall	MoverScore	0.5860	0.6178
	BERTScore	0.8822	0.9041
	GPT-Eval	0.7514	0.8103

Table 2. Performance comparison across different tasks and metrics for Gemini 2.0 (gemini-2.0-flash-exp) [3] and EchoTraffic. Due to access limitations on Gemini 2.0, we evaluated both models on the same subset of 2,611 AV-TAU test Q&A pairs. Results indicate that EchoTraffic outperforms Gemini 2.0, with GPT-Eval (AVG) representing the average score from GPT-40 [1] and Gemini 2.0.

Task	Metric	Gemini 2.0	EchoTraffic
	BLEU	0.0637	0.2518
Description	Rouge	0.3275	0.4329
	MoverScore	0.5815	0.6199
	BERTScore	0.9050	0.9101
	GPT-Eval (AVG)	0.6539	0.7057
	BLEU	0.0857	0.2432
	Rouge	0.2146	0.3495
Causation	MoverScore	0.5737	0.6071
	BERTScore	0.8857	0.9021
	GPT-Eval (AVG)	0.7743	0.8610
	BLEU	0.0767	0.2011
	Rouge	0.2505	0.4238
Prevention	MoverScore	0.5794	0.6151
	BERTScore	0.8752	0.8985
	GPT-Eval (AVG)	0.7487	0.8611
	BLEU	0.0832	0.3093
	Rouge	0.2746	0.3940
Response	MoverScore	0.5521	0.6276
	BERTScore	0.8479	0.9077
	GPT-Eval (AVG)	0.7220	0.8658
	BLEU	0.0773	0.2514
	Rouge	0.2668	0.4000
Overall	MoverScore	0.5717	0.6174
	BERTScore	0.8785	0.9046
	GPT-Eval (AVG)	0.7247	0.8234

Part 3 - Dataset and Experiment Details

Dataset Details and Statistics

Our dataset captures a diverse range of traffic scenarios, categorized by observation view, weather conditions, and time of day, as detailed in Table 3. Notably, during the data collection, we observe that dashcam videos are more likely to include audio data, whereas most surveillance-view videos lack accompanying audio. Additionally, our dataset reveals that traffic anomaly events are often accompanied by distinct audio types, including collision sounds, braking, honking, shouting, tire sounds, and alarms, as revealed in Fig. 6. Furthermore, Fig. 5 illustrates the subtasks through four word cloud visualizations, providing a visual representation of the dataset's focus areas.

Category	Subcategory	Count
	Dashcam	26,400
Observed on Vitere	Surveillance	378
Observation view	Subcategory Dashcam Surveillance Camera Phone Sunny Cloudy Rainy Foggy Snowy Day Night	1,631
	Phone	1,456
	Sunny	23,007
	Cloudy	3,409
Weather Condition	Subcategory Dashcam Surveillance Camera Phone Sunny Cloudy Rainy Foggy Snowy Day Night	2,589
	Foggy	224
	Snowy	636
Thus of Door	Day	24,874
Time of Day	Night	4,991

Table 3. Statistics of the collected dataset categorized by observation views, weather conditions, and time of day.



Figure 5. These "word cloud" visualizations highlight the key terms in the dataset.



Figure 6. Frequency distribution of sound-related terms in the dataset.

Question Formulation

The following set of question formulations is designed to assess and annotate anomalies within the dataset. These questions address various aspects of anomaly understanding, such as *<WHAT - MULTIMODAL UNDERSTANDING>*, *<WHY - CAUSAL INFERENCE>*, *<WHEN - ANOMALY TIMING>*, *<HOW - PREVENTION STRATEGY>*, and *<HOW - EVENT RESPONSE>*. During each training iteration for a given task, a question is randomly selected to ensure diverse and comprehensive coverage.

WHAT - MULTIMODAL UNDERSTANDING:

- What unusual incident occurred in the video?
- Describe the anomaly observed in the video.
- What unexpected behavior is shown in the video?
- What anomaly is present in the video?
- What is the unexpected event in the video scene?
- In the video, what unusual event takes place?
- What anomaly can be observed in the video?
- What unexpected behavior is observed in the video?
- What irregular event occurs in the video?
- What odd occurrence is captured in the video?
- Describe the unexpected event in the video.
- What unusual scene is depicted in the video?
- What unexpected incident happens in the video?
- What irregularity is captured in the video?
- What strange event unfolds in the video?

WHY - CAUSAL INFERENCE:

• What caused the anomaly in the video?

- Why did this anomaly happen in the video?
- What led to the anomaly in the video?
- Why did the unusual event occur in the video?
- What was the reason behind the anomaly in the video?
- Why did the incident unfold this way in the video?
- What prompted the anomaly in the video?
- Why did the unusual event take place in this manner in the video?
- What was the cause of the unusual event in the video?
- What set off the anomaly in the video?
- Why did the anomaly happen as it did in the video?
- Why did the anomaly develop in the video?
- What factor caused the anomaly in the video?
- Why was there an anomaly in the video?
- What made the anomaly occur in the video?

WHEN - ANOMALY TIMING:

- When does the anomaly start and end in the video?
- At what time in the video does the anomaly begin and end?
- When does the anomaly start and when does it end in the video?
- At what points in time do the anomaly's start and end occur in the video?
- When does the anomaly begin and conclude in the video?
- At what moments do the anomaly's start and end take place in the video?
- When in the video does the anomaly begin and when does it come to an end?
- What times mark the start and end of the anomaly in the video?
- In the video, at what times do the anomaly start and finish?
- When does the anomaly start and stop in the video?
- At what timestamps does the anomaly start and end in the video?
- When do the start and end of the anomaly occur within the video?
- What times in the video show the start and end of the anomaly?
- When is the starting and ending point of the anomaly in the video?
- At what points in the video do we see the anomaly start and end?

HOW - PREVENTION STRATEGY:

- What could have been done to prevent the anomaly in the video?
- How could this type of incident have been avoided in the video?
- What changes could prevent this anomaly in future scenarios?
- How might the anomaly have been averted in the video?
- What preventive measures could have stopped the anomaly in the video?
- What could reduce the likelihood of this event happening?
- How could the unusual scenario have been prevented in the video?
- How can similar anomalies be avoided in the future?
- What could be implemented to prevent such events?
- What preventive actions could be taken to avoid this situation?
- How might the risk of this anomaly be minimized?
- What safety improvements could help prevent the anomaly?
- How could similar incidents be avoided?
- What could have helped to prevent this anomaly in the video?
- How could this type of anomaly be mitigated in the future?

HOW - EVENT RESPONSE:

- What steps should be taken to handle the anomaly in the video?
- How should those involved respond to the anomaly in the video?
- What is the recommended response for the anomaly in the video?
- How should this unexpected event be managed in the video?

- What actions should be taken to address the anomaly in the video?
- What is the appropriate response to this anomaly?
- How should participants react to the anomaly in the video?
- What should be done to respond effectively to the anomaly?
- How should those involved in the video handle the anomaly?
- What steps are necessary to manage the anomaly?
- How should the situation be handled by those involved?
- What should the response be to this unexpected event?
- How should the event be managed after the anomaly?
- What's the recommended action for handling this anomaly?
- What actions should those involved take in response to the anomaly?

Details in the GPT-Eval Metric

In the GPT-Eval metric, we use GPT-40 [1] as an assistive tool to evaluate the responses generated by EchoTraffic and the comparison methods. When GPT-40 evaluates the responses from different models, we specify three key criteria for scoring: *Reasonability, Detail,* and *Consistency*. To study whether GPT-40 clearly understands its task, we first configure the system prompt as follows:

For a specific question-answer pair, we propose to provide GPT-40 with the question <QUESTION>, the reference ground truth <REFERENCE>, and the model's output <ANSWER>. GPT-40 will compare the model's output with the reference and assign a score. The user prompt is structured as follows:

Part 4 - Limitations and Future Work

Limitations

Hallucination. Multimodal LLMs often rely on processing a limited subset of video frames, which can lead to hallucinations arising from incomplete contextual information. Integrating audio signals and employing an audio-insight frame selector helps mitigate these issues. However, restricted frame selection can still cause hallucinations, such as incorrectly inferring a collision between the ego vehicle and a bicycle in front instead of accurately identifying it as a rear-end collision involving another vehicle, as illustrated in Fig. 7.

Long Video Analysis. Analyzing long-term videos presents challenges due to the large volume of data and the computational resources required. Besides, effectively processing these videos is challenging, as it requires maintaining temporal coherence across extended sequences and accurately extracting relevant spatiotemporal features, which remain areas for further refinement and optimization.

Future Work

We aim to address the limitations mentioned above by developing methods that enable our models to seamlessly analyze complete video sequences or bird's-eye view videos. Additionally, we plan to explore anomaly anticipation by integrating audio-visual cues with contextual data, with the goal of enhancing real-time monitoring and improving model resilience across diverse traffic scenarios.



Figure 7. Failure case. EchoTraffic misinterprets a rear-end collision as a crash between the ego vehicle and a bicycle ahead.

References

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. arXiv preprint arXiv:2303.08774, 2023. 7, 11
- [2] Zesen Cheng, Sicong Leng, Hang Zhang, Yifei Xin, Xin Li, Guanzheng Chen, Yongxin Zhu, Wenqi Zhang, Ziyang Luo, Deli Zhao, et al. VideoLLaMA 2: Advancing spatial-temporal modeling and audio understanding in Video-LLMs. arXiv preprint arXiv:2406.07476, 2024. 3, 7
- [3] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023. 7
- [4] Huaxin Zhang, Xiaohao Xu, Xiang Wang, Jialong Zuo, Chuchu Han, Xiaonan Huang, Changxin Gao, Yuehuan Wang, and Nong Sang. Holmes-VAD: Towards unbiased and explainable video anomaly detection via multi-modal LLM. arXiv preprint arXiv:2406.12235, 2024. 3