

CrashChat: A Multimodal Large Language Model for Multitask Traffic Crash Video Analysis

Kaidi Liang, Ke Li, Dr. Xianbiao Hu, Dr. Ruwen Qin*

*Transportation Information Laboratory (TIL), Stony Brook University, Department of Civil
Engineering, Stony Brook, NY 11794, USA*

08/19/2025

AGENDA

- Background and Motivation
- Problem Statement
- Methodology
- Results Comparison
- Conclusion and Future Work

BACKGROUND AND MOTIVATION

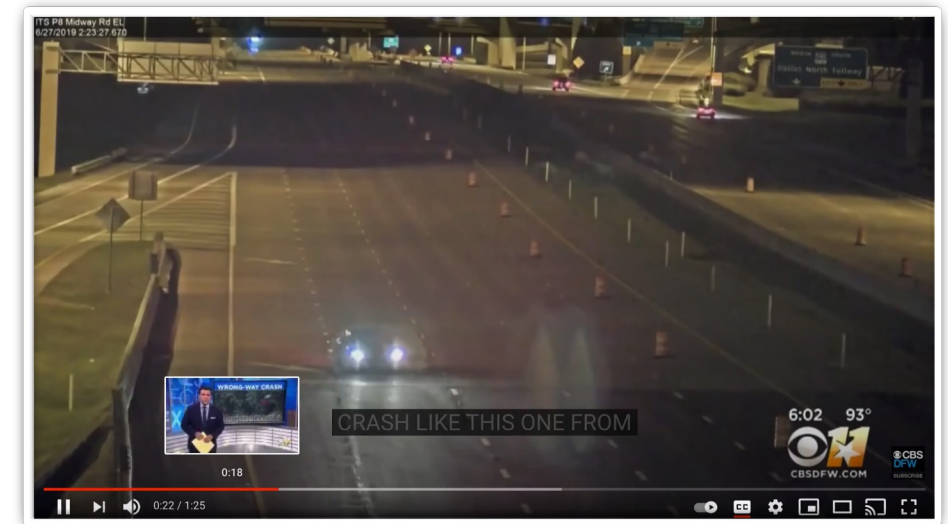
Challenge One: The Annotation Bottleneck in Large-Scale Crash Video Analysis



Dashcam Video



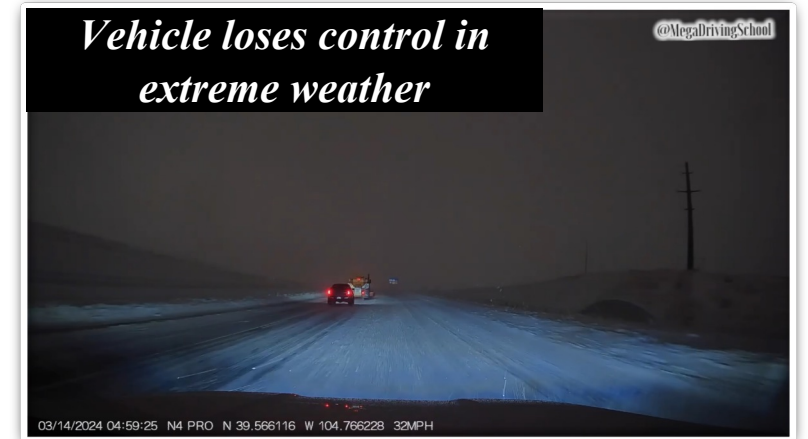
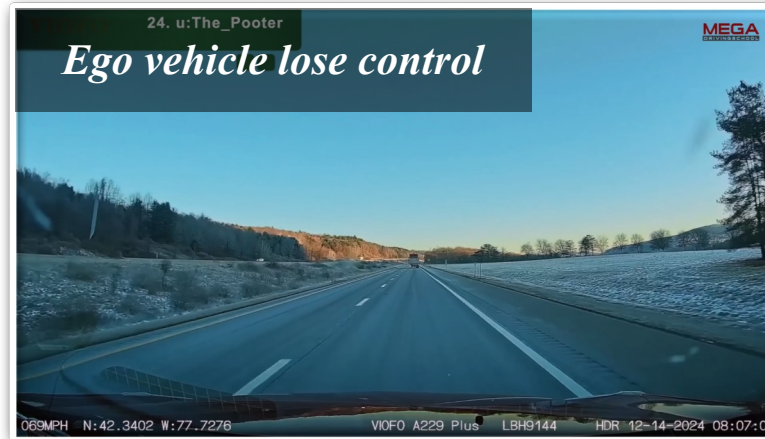
Social Media Video



Surveillance Video

BACKGROUND AND MOTIVATION

Challenge Two: Crash Videos Are Low-Light, Long-Tail, and Data-Scarce

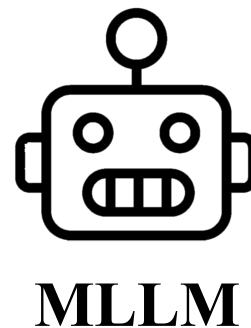
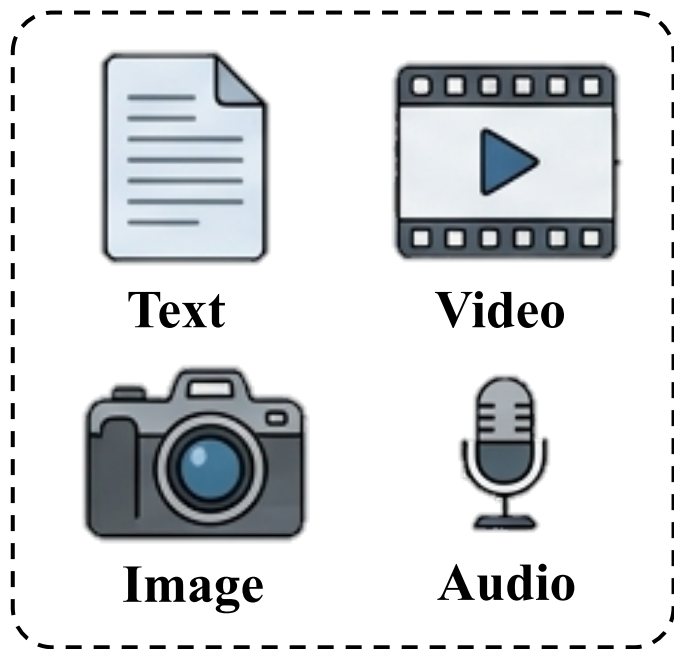


Low light condition, long-tail events and unbalanced data distribution!

PROBLEM STATEMENT

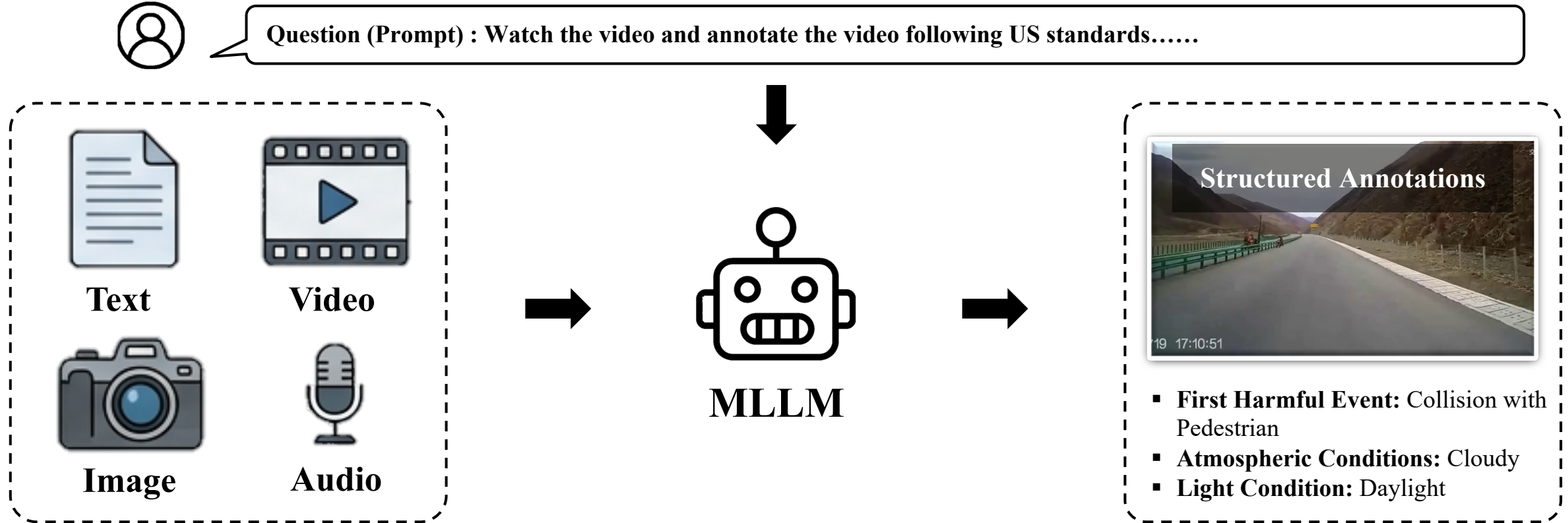


Question (Prompt) : Watch the video and annotate the video following US standards.....



- **First Harmful Event:** Collision with Pedestrian
- **Atmospheric Conditions:** Cloudy
- **Light Condition:** Daylight

PROBLEM STATEMENT

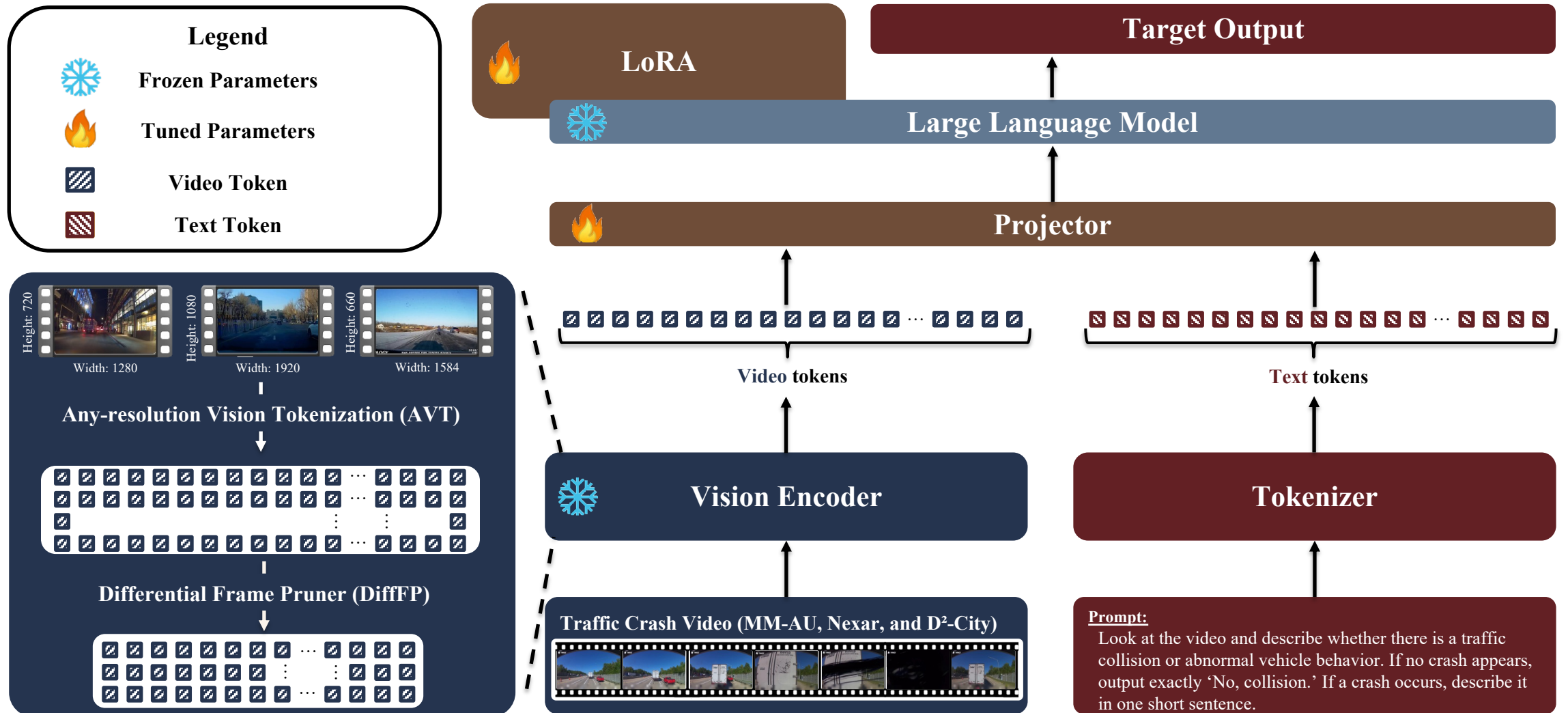


Problem Statement

To address above challenges, we investigate Multimodal Large Language Models (MLLMs) as a scalable solution for automated annotation, semantic reasoning, and fine-grained understanding of rural crash videos.

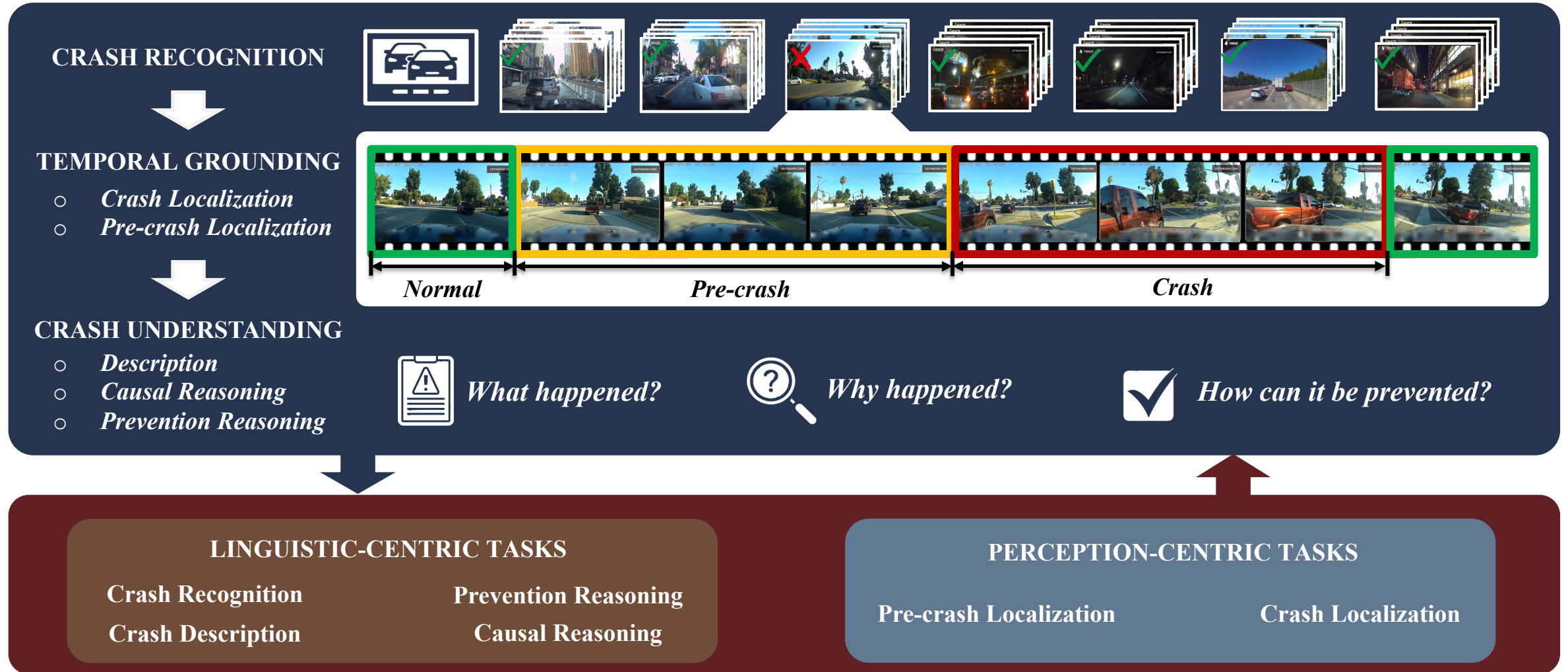
METHODOLOGY

Classic Multi-modal Large Language Model Framework



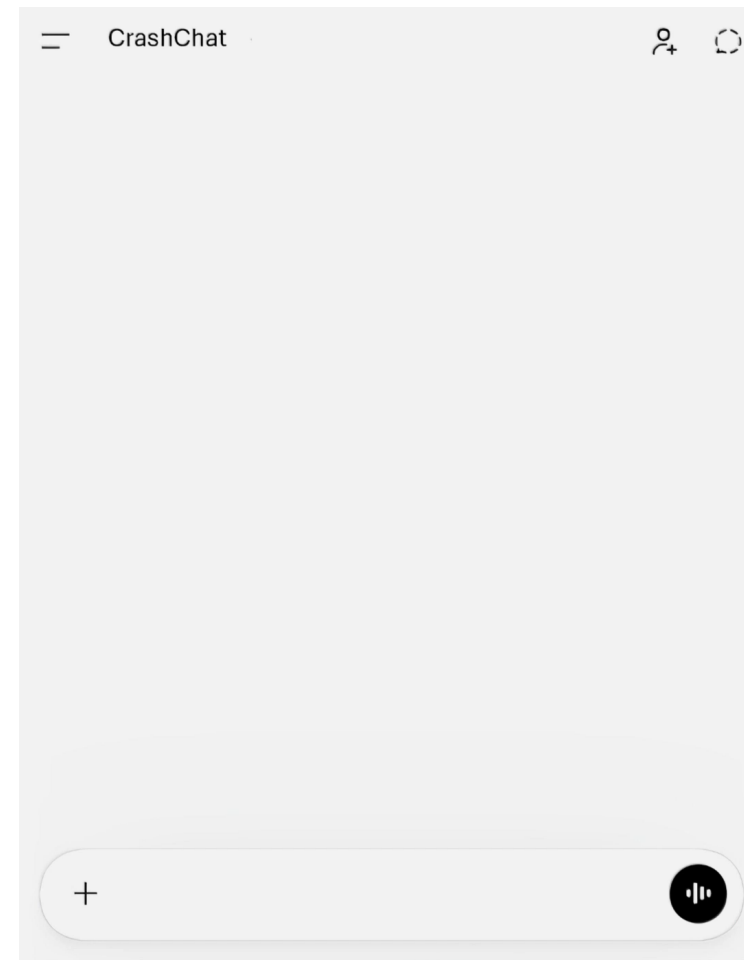
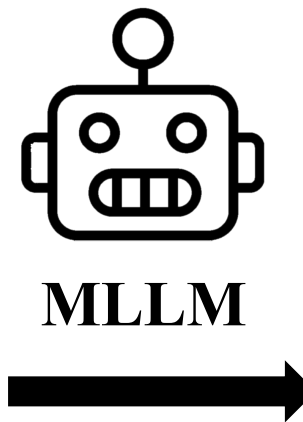
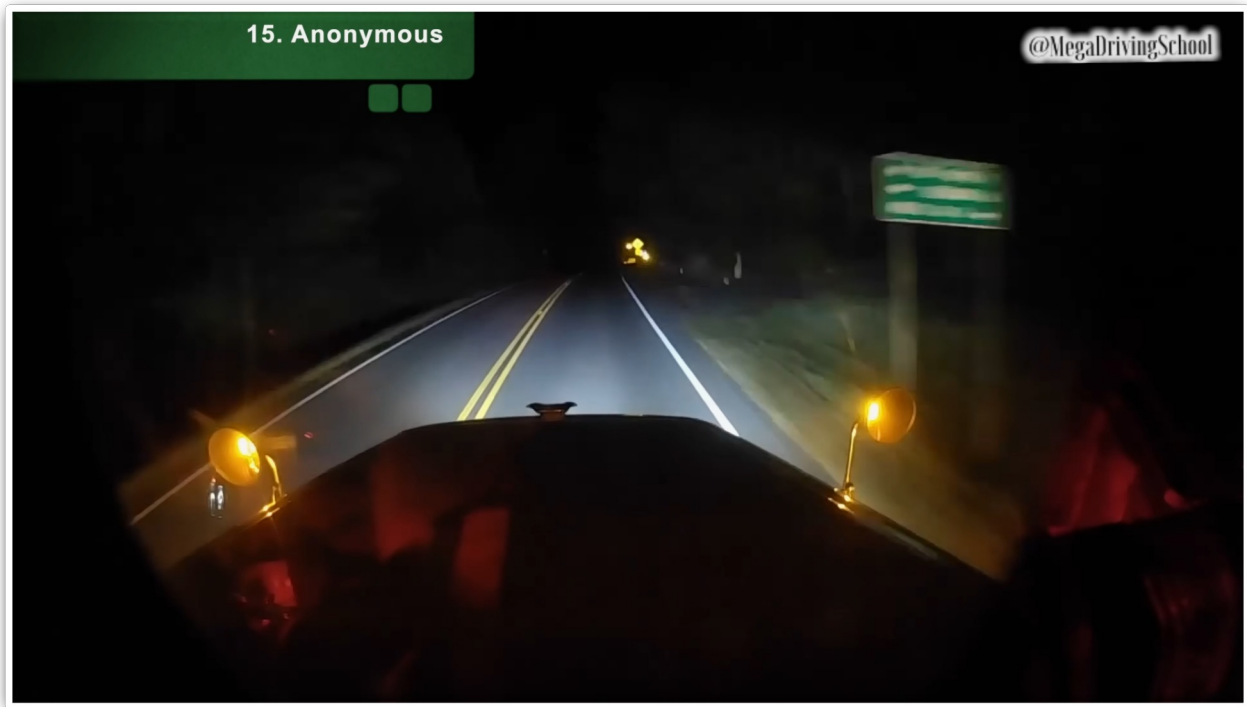
METHODOLOGY

Six Core Crash Video Understanding Tasks



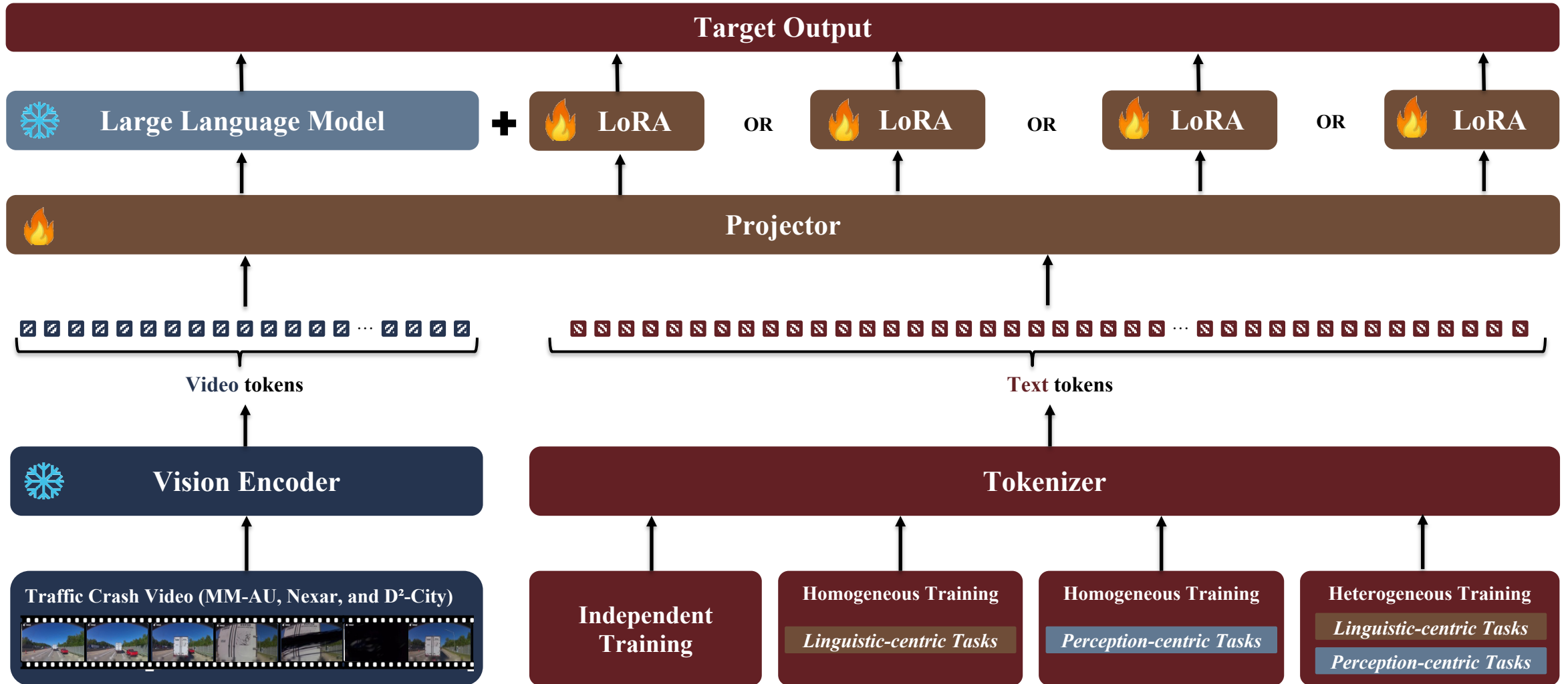
METHODOLOGY

Demo

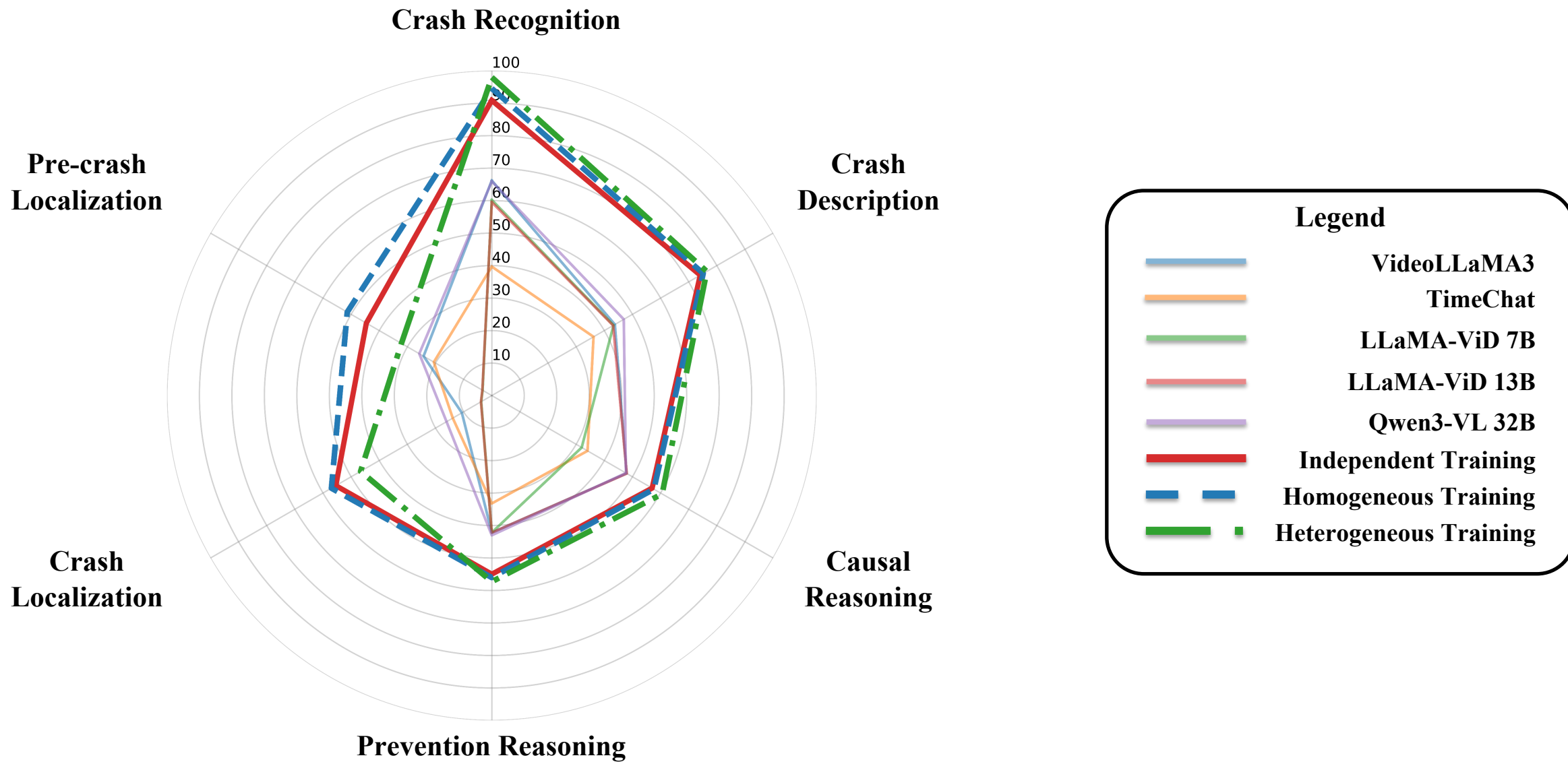


METHODOLOGY

Coupling and Decoupling Training Strategy



RESULTS COMPARISON



CONCLUSION AND FUTURE WORK

Conclusion

Fine-tuning enables MLLMs to better adapt to crash video analysis. By introducing *coupling and decoupling training strategies*, the model achieves improved performance under the same training conditions. The fine-tuned MLLM shows solid performance across all *six core tasks*, demonstrating its strong potential for crash video understanding.

Future Work

- **Dataset Constructure:** The existing dataset annotations do not follow the US standards and cover US region.
- **More efficient framework:** coupling and decoupling can improve the model performance, but cannot eliminating interference from different tasks.
- **Applications:** The model can be applied to a wide range of practical applications, such as road hazard monitoring, crash report generation, and anticipation.

Thank You!