Proceedings of the Seventh International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 15, Paper 21.3
Civil-Comp Press, Edinburgh, United Kingdom, 2026
ISSN: 2753-3239, doi: 10.4203/ccc.15.21.3

Feasibility of Vision–Language Model–Based Collision Risk Assessment as a Driver Assistance System for Tram Drivers: A Preliminary Study

S. Hovanotayan[1], W. Wang[1], K. Nakano[1], T. Takata[2], M. Ueda[3], K. Sakakidani[3] and T. Maeoka[3]

[1]Institute of Industrial Science, The University of Tokyo, Japan
[2]Kyosan Electric Manufacturing Co., Ltd., Yokohama, Japan
[3]Hiroshima Electric Railway Co., Ltd., Japan

Abstract

As trams typically share traffic spaces with cars, cyclists and pedestrians, collisions between trams and surrounding road agents have become an increasingly critical problem. Assessing potential risk is undeniably an effective measure to mitigate such safety issues. Nevertheless, comprehensively grasping tramway scenarios remains challenging due to their inherent complexity compared to other traffic environments. Recently, vision-language models (VLMs) have emerged as promising tools for visual understanding in various fields, especially in autonomous driving, because of their advanced scene analysis and reasoning abilities. In this paper, we investigate the feasibility of deploying VLMs for collision risk assessment dedicated to tram contexts. Existing VLMs, including GPT-5, Gemini 3 Pro, Sonnet 4.5, Llama 3.2 Vision, Qwen 3 VL and DeepSeek VL 2, are selected. Firstly, a dataset of tram scenarios is collected from a tram driving simulator, ranging from safe to dangerous cases. Each VLM is prompted to analyse scene features, assess collision risk and suggest driving actions in the given scenarios. In the preliminary tests, each model successfully managed to perform its designated tasks with promising performance. In future work, the VLMs' outputs will be compared with evaluations by professional tram drivers obtained through questionnaire surveys, thereby serving as ground truth data.

Keywords: tram, safety, collision risk assessment, artificial intelligence, computer vision, vision-language model, driver assistance system.

1 Introduction

Urban public transport comprises various modes, such as buses, taxis, boats, trains and trams. Among these, trams are regarded as one of the most popular urban transit networks, in which the vehicles usually drive on tramway tracks on the streets teeming with cars, cyclists and pedestrians. Owing to these mixed-traffic environments, trams are inherently susceptible to crashes with other road users, differing from trains that mainly operate on isolated and dedicated tracks. Hence, to ensure operational safety and security, collision risk assessment should be rigorously addressed in tram services. In addition, numerous tram services are manually operated by a single driver on board, for example, Hiroshima trams. Further driven by this fact, an assistive tool is vital to reduce human errors caused by fatigue, distraction or misjudgements, which can result in traffic accidents. Nowadays, there is abundant research aimed at proposing risk assessment frameworks, especially for autonomous cars. Lately, Li et al. studied a risk assessment model by using the Gaussian distribution function with physical safety metrics such as Time-to-Collision (TTC), Time-to-Stop (TTS) and Time-to-Escape (TTE) to predict risk levels in accordance with individual driving style [1]. Similarly, Noh has proposed a probabilistic method grounded in Bayesian reasoning for evaluating potential threat of a collision between cars at an intersection [2]. Whereas we can clearly witness research and design in VLMs proceeding at a rapid pace within academia and industries, the same amount of attention has not been observed in tram applications, remaining relatively unactive and limited. Considering the complicated environments of tram driving, an accurate risk assessment task necessitates high-level human-like knowledge reasoning capabilities. To achieve this, the main purpose of this study is to introduce a novel collision risk assessment model as a driver assistance system for tram drivers based on VLM methods.

In recent years, vision-language models (VLMs), an extension of large language models (LLMs), have played a significant role in a wide spectrum of computer vision tasks, such as semantic segmentation, object detection, pattern recognition, image-context understanding as well as visual question answering (VQA) etc. In comparison to Deep Neural Networks (DNNs), which generally require large-scale datasets and laborious training processes, VLMs, as being trained by data available on the Internet, such as text-image pairs and rich-linguistic knowledge, it is strongly believed that they are easy to implement and have better understanding performance along with reasoning capabilities [3]. Especially, TrafficRiskGPT model has been developed for road traffic-context risk assessment. This model is based on the LLaMA3-B model integrated with Casual Inference Chain of Thought (CI-CoT) technique to help reduce its inference time and improve model's reliability. The results indicate, in the identical scenes, the proposed model achieved lower collision rates in comparison to GPT4o-mini with faster inference speed [4]. Motivated by this work, it is highly expected that VLMs are able to offer accurate and reliable scene analysis, collision risk assessment and also appropriate driving actions to tram drivers under a certain circumstance, resulting in safer and more efficient tram services.

Therefore, in this research, the performance of VLMs in understanding tramway environment scenarios from driver cab's perspective and assessing collision risk will

be preliminary examined. For this purpose, a group of well-known public multi-modal models and VLMs is selected, namely, GPT, Gemini, Sonnet, Llama, Qwen and DeepSeek VL. Next, we will simulate and gather diverse scenarios, encompassing many situations like safe tram driving and crashing with a turning-right car etc. Then, given the simulated scenes, we will give different kinds of prompt to the VLMs, including zero-shot, few-shot and Chain-of-Thought, to evaluate scene understanding and risk assessment. Ultimately, all outputs produced from each VLM with different kinds of prompts will later be compared and discussed with opinions and reasoning from expert tram drivers to validate VLMs' feasibility to assist drivers.

2 Methodology

In this section, the overview of the proposed collision risk assessment model based on VLMs, as depicted in Figure 1, the details of each chosen VLMs, the definitions of prompts as well as a scenario data collection method are provided in the subsequent subsections.

2.1 Proposed Collision Risk Assessment Model

We propose a collision risk assessment model primarily built upon the VLMs, which promote an end-to-end collision risk assessment solution without explicitly deriving an algorithm or predefining thinking processes.

The overall framework is illustrated in Figure 1. Since the primary purpose of this model is to help facilitate a collision risk assessment in the frontal environments of an ego-tram vehicle and decision-making of human drivers during operation, the input of the proposed model is set to be visual information from a front camera installed in a driver's cab. An image received from the camera in real time is then transmitted to VLMs. Consequently, multiple prompting strategies will be given to the VLM, asking it to generate responses to three key questions, for example, how is the collision risk, scene interpretation and driving policy in current operating conditions, under different prompting paradigms: zero-shot, few-shot and chain-of-thought (CoT), which will be further described later in another subsection. Eventually, the VLM outputs a scene description, an assessed risk level, which is either Safe, Attentive or Dangerous, and recommended suitable driving action towards human drivers.

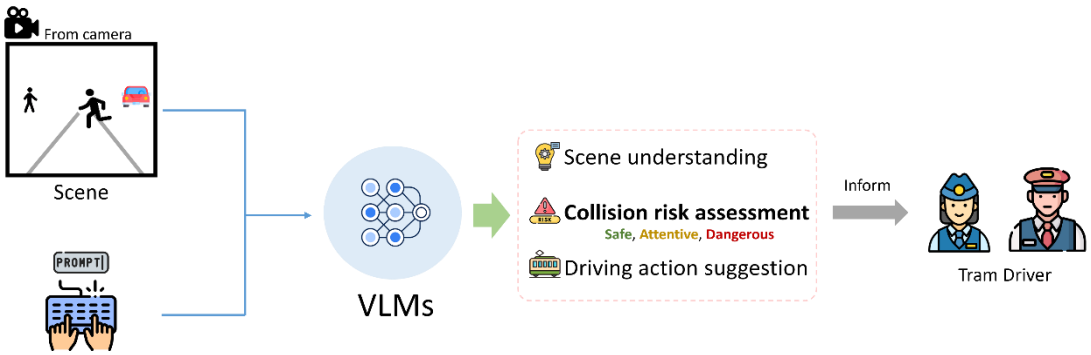

Figure 1: VLM-based collision risk assessment framework.

In this research, collision risk is categorized into three levels: “Safe”, “Attentive” and “Dangerous”. “Safe” refers to a running condition in which the likelihood of an accident is negligible. “Attentive” represents an operating situation where there is an observable possibility of a vehicle collision if insufficient caution is exercised, while “Dangerous” denotes a collision is highly likely or imminent, needing an immediate action to avoid it.

2.2 Vision-Language Models

VLMs have become breakthroughs in equipping machines with capabilities to process multimodal information in forms of images and text. In recent years, several general-purpose VLMs have been released to the public, for example, Llama and Qwen etc. In parallel, other multimodal models like GPT, Gemini and Claude have also become widely available. Although they were not originally developed as VLMs, they demonstrate effective vision-language reasoning capabilities and can function as VLMs. Table 1 shows a list of representative state-of-the-art models.

Model	Type	Developer	Release year
GPT-5	Multimodal Model	OpenAI	2025
Gemini 3 Pro		Google	2025
Claude Sonnet 4.5		Anthropic	2025
Llama 3.2 Vision	Vision-Language Model	Meta	2024
Qwen 3 VL		Alibaba Cloud	2025
DeepSeek VL2		DeepSeek AI	2024

Table 1: Existing multimodal and vision-language models.

In this study, six state-of-the-art models composed of GPT-5, Gemini 3 Pro, Claude Sonnet 4.5, Llama 3.2 Vision, Qwen 3 VL as well as DeepSeek VL 2 are selected without additional training or parameter tuning to accomplish the research objectives.

2.3 Prompting

Like LLMs that execute their specific intended task solely based on given prompts to them, such as task explanation and brief or detailed samples in text formats, VLMs also process by prompting. Prompting, as a primary interface connecting between users and VLMs, involves designing and structuring users’ requirements to the model. A well-defined prompt is helpful in precisely and elaborately instructing VLMs to perform either simple or even greatly sophisticated jobs effectively. In addition, varied prompt templates also have significant impacts on the VLMs’ performance to harness their abilities in understanding and inference speed, ultimately affecting the desired quality level of outcomes.

In the context of VLMs, prompting can be classified into three categories; Zero-shot, Few-shot and Chain-of-Thought. They will be meticulously verified regarding how they influence scene and risk analysis under tram environments. The detailed explanations of these three prompts will be provided in the next subsections.

2.3.1 Zero-shot prompt

Zero-show prompting is a prompting strategy guiding VLMs to carry out their tasks by a direct instruction without specific samples or instructions. As implied by the term of zero-shot, VLMs must entirely rely on the given prompt of task description and their own pre-trained knowledge to figure out how to successfully finish the tasks by themselves. This method is considered to be the simplest form of prompting.

2.3.2 Few-shot prompt

Rather than supplying no examples as in zero-shot, few-shot prompting enriches the previous technique by proving a few samples alongside the task guidelines, usually at least two representative examples to the models. With this prompting approach, it will elevate the model's understanding, resulting in more precise and preferred responses, since the model can learn more from both the pre-trained insight and the underlying patterns observed in the provided samples.

2.3.3 Chain-of-Thought

CoT aims at complicated tasks because it generates a series of intermediate logical steps, stimulating VLMs to learn better. Particularly, one of the main advantages of this prompting is that it helps decompose any complicated jobs, which comes with multiple steps into intermediate steps. This means the models will process reasoning step-by-step. By reasoning through small steps, CoT is able to handle complex tasks better than zero-shot and few-shot prompting, returning better reasoning accuracy and logical consistency. Additionally, it offers reasoning transparency.

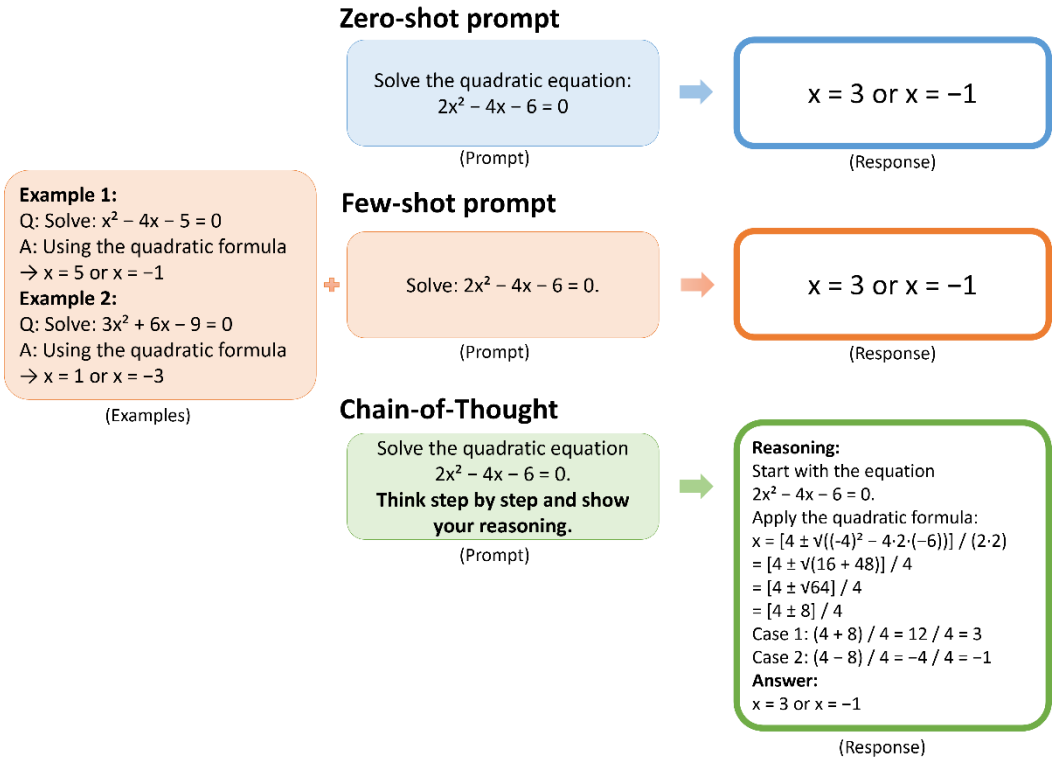

Figure 2: Examples of zero-shot, few-shot and CoT prompts.

2.4 Scenario Collection

In this section, the collection of a tram scene dataset is presented. The data is prepared in video format. The scenario videos are gathered from the tram driving simulator of Hiroshima Electric Railway company. This simulator is mainly leveraged to train new tram drivers, enabling them to experience the virtual world of Hiroshima trams before engaging in actual operations. It delivers a wide spectrum of tram scenarios from the front-facing perspective, such as driving in urban areas of Hiroshima city where it is full of cars, as well as in sparsely populated regions.

In term of risky scenes, this simulator has a variety of scene setting options to train drivers to handle cautious and hazardous situations. Each scenario will be recorded for about 10 s to acquire scenario features, covering the moments before an accident takes place until its occurrence. This dataset consists of rich scenario information that can make a fair comprehension and evaluation of VLMs for the risk assessment task. The list of test scenarios for the preliminary experiments will be provided in the next section.

Figure 3: Tram driving simulator at Hiroshima tram company.

Figure 4: Sample scenarios from the simulator.

3 Experiment and Results

3.1 Test Scenarios

In the test scenario setup, twelve different scenarios were generated in the tram driving simulator instead of real-world operations to validate the effectiveness of the VLMs considering safety concerns. The scenarios illustrate Hiroshima tram environments from the driver's view with the ego vehicle running on the left-hand pair of rail tracks with the standard gauge of 1.435 m. Similarly, the simulator provides the scenarios that replicate the most common risky scenes of Hiroshima trams, including no-risk, near-miss circumstances and accidental cases interacting with surrounding cars, bikes and pedestrians. All the details are summarized in Table 2

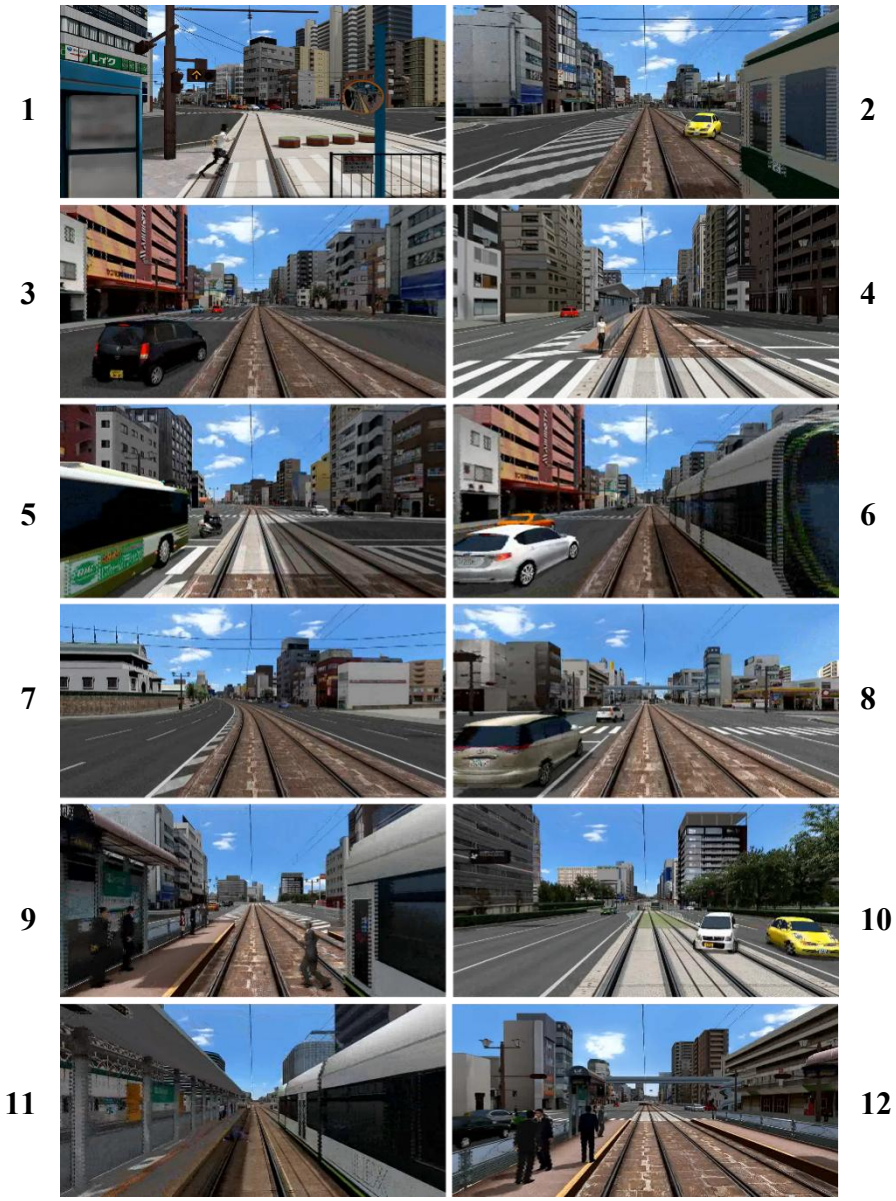

Figure 5: Sample scenarios used for the experiments.

Scene No.	Scene description	Collision
1	**Pedestrian sudden crossing when departing from the station** When the tram is parked at the station and the tram signal turns green for departure, a pedestrian suddenly crosses the track from nowhere	No
2	**Blind-spot pedestrian/car due to an oncoming tram** While running with no surrounding agents except an oncoming tram, the oncoming tram blocks the driver's perspective and creates a blind spot. And as the ego-tram approaches the oncoming tram, a car suddenly crosses the tram track from behind it.	Yes
3	**Car turning right at the intersection (No oncoming tram)** While running forward with cars in the same direction, a car at the intersection suddenly makes a right turn in front of the tram.	Yes
4	**Pedestrian waiting close to the track** A pedestrian is standing very close to the tram track, waiting to cross.	No (Near Miss)
5	**Motorbike suddenly turning right** At an intersection with a green light, a motorbike is waiting to turn right. As the tram gets closer to the intersection, the motorbike suddenly turns right in front of the tram.	No (Near Miss)
6	**Car turning right at the intersection (No oncoming tram)** While running forward with cars in the same direction, a car at the intersection suddenly makes a right turn in front of the tram.	Yes
7	**Clear tram path** The track is clear, with no obstructions and no nearby traffic participants	No
8	**Two cars on the left; second car suddenly turns right** Two cars are on the left side of the tram (same lane, same direction). The first car is waiting to turn right, but the second car suddenly turns right, cutting across the tram's path.	Yes
9	**Passenger crossing behind a parked tram at the station** While approaching a station with another tram parked on the opposite platform for boarding, a passenger suddenly crosses the track from behind the parked tram.	Yes
10	**Car suddenly crossing into a parking lot** While running forward (not at an intersection), a car suddenly crosses the tram track in front of the tram to enter a parking lot on the left side (on the tram's lane).	Yes
11	**Passenger falling onto the track** When approaching a station, multiple passengers are on the platform. One passenger suddenly falls onto the tram track.	Yes
12	**Passenger stopping at the yellow line** When approaching a station, several passengers are waiting on the platform. One passenger walks forward and stops at the yellow safety line (within the vehicle limit) without paying attention to the approaching ego-tram	Yes

Table 2: Scene description.

3.2 Input Prompt Design

We applied prompt engineering to design the input prompts for our experiments. The creation of prompt words is associated with the core of three prompting technique, as previously explained; zero-shot, few-shot and CoT. In general, the content and structure of the word prompts rely on the tasks and the complexity of the models. In this study, all prompts are meticulously formulated to ensure the models can correctly understand and captures all relevant and coherent information in the running tram scenes, further enhancing the interpretability and performance of VLMs to accomplish the tasks.

Based on the research objectives of investigating the performance of VLMs in
- capturing related scene features and traffic element and giving precise scene analysis
- accurately assessing collision risk
- providing appropriate, useful and logical driving actions

the input prompts utilized in the experiments are derived accordingly, as shown in the following table.

<table>
<tr><th>Type</th><th colspan="2">Input Prompt</th></tr>
<tr><td>Zero-shot</td><td colspan="2">“You are a tram assistant system”
Look at the given video and provide three things:
1. Provide scene description
2. Assess collision risk level (Safe, Attentive, or Dangerous)
3. Recommend appropriate action for the tram driver</td></tr>
<tr><td rowspan="2">Few-shot</td><td>Video example 1:

Scene: A stationary oncoming tram obstructs the ego tram driver’s field of view, increasing the risk of passengers emerging unexpectedly and potentially leading to a collision.
Risk: Attentive
Action: Proceed with caution</td><td>Video example 2:

Scene: A motorbike is suddenly crossing ahead of the tram at the intersection, while the ego tram is approaching.

Risk: Dangerous
Action: Brake immediately.</td></tr>
<tr><td colspan="2">Prompt
“You are a tram assistant system” and look at the examples and corresponding outputs. Then, provide three things: scene, risk (Safe, Attentive, Dangerous) and action, by following the given examples.</td></tr>
<tr><td>CoT</td><td colspan="2">“You are a tram assistant system” Analyse the video step by step.

1. Describe everything relevant and identify all dynamic actors (pedestrians, cars, bikes, cyclists) and their approximate movement in the scene.

2. Evaluate collision risks based on a 3-level scale (Safe, Attentive, Dangerous)

3. Recommend appropriate driving action.

Show your reasoning clearly before giving the final answer.”</td></tr>
</table>

Table 3: Examples of the prompt used in the experiments.

3.3 Preliminary Test Results

To validate the VLMs-based predictive risk assessment models, scenario videos were truncated to exclude collision scenes, each ending approximately two seconds before the collision, ensuring that the model assesses risk solely from pre-incident visual cues. All VLMs were prompted to analyse and identify traffic agents in the given scenes, to evaluate collision risk based on a three-level risk scale and to suggest driving actions. According to the preliminary test results, each VLM is capable of comprehending the assigned tasks under zero-shot, few-shot or CoT prompts. However, there are obvious differences among their responses. Specifically, when it comes to scene analysis, the zero-shot prompts return answers fastest, but produce short analyses that sometimes describe traffic elements irrelevant to the scenarios or inaccurately, making the risk assessment and suggested driving actions slightly unreliable, whereas the few-shot ones permit the VLMs to generate more accurate, detailed, valid and well-organized responses, which help us acquire more reliable and robust risk evaluation and driving actions despite having longer processing time. In the cases of CoT prompting, the VLMs present their thinking process in a clear, step-by-step manner, and as expected, have a noticeable longest inference time. In addition, all models seem to correctly identify all important features embedded in the driving scenes consistently, leading to precise collision risk assessment and suitable driving actions to tram drivers. For a fair evaluation on their performances, the entire results will be further compared with evaluations by professional drivers and analysed, as depicted in Figure 6. Figure 7-8 illustrates samples results from the preliminary tests.

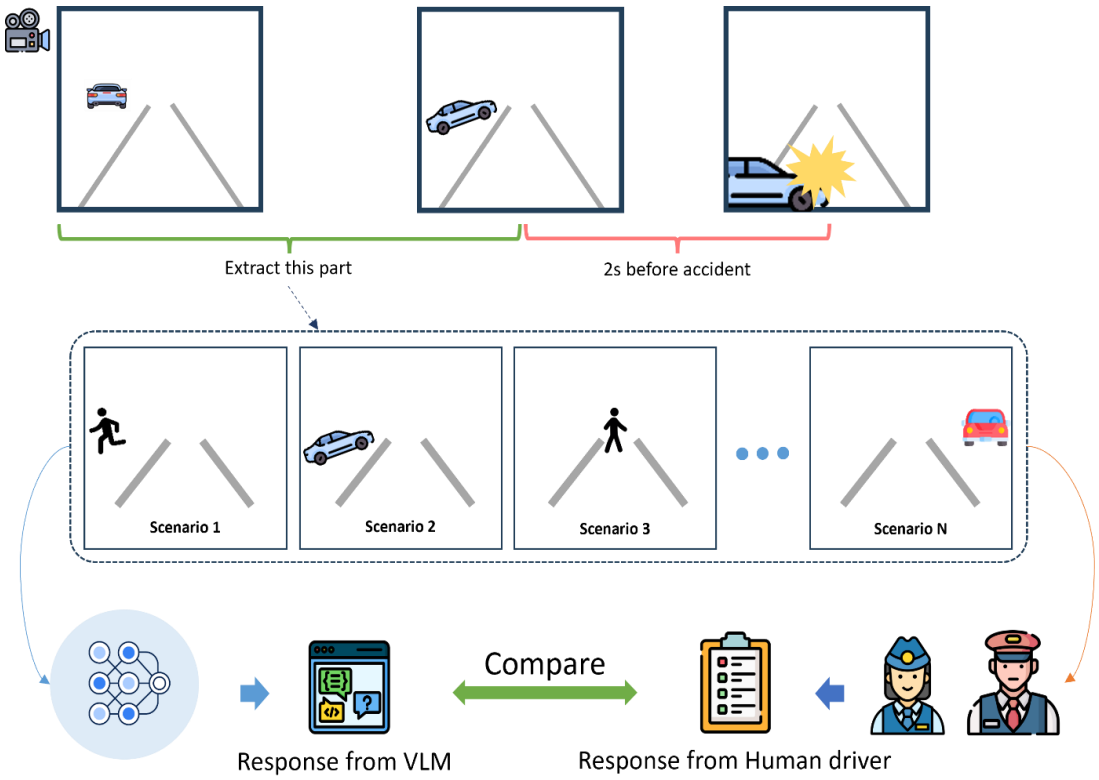

Figure 6: Scenario preparation, experiment and evaluation process overview.

GPT-5

1. Scene Description & Dynamic Actor Identification

Environment

- **Urban street with central tram tracks**
- **Signalized intersection ahead**
- **Clear weather, good visibility**
- **Road traffic lanes on both sides of the tracks**

Dynamic actors identified

(A) Ego tram

- **Location: Center tram track**
- **Motion: Moving forward at constant speed**
- **Path: Straight, fixed rail trajectory**

(B) Oncoming tram

- **Location: Same tram track, opposite direction**
- **Motion: Approaching steadily**
- **Behavior: Properly aligned on rail**

(C) Cars

Parallel cars

- **Location: Adjacent road lanes**
- **Motion: Moving parallel to tram**

Crossing car at intersection

- **Location: Intersection ahead**
- **Motion: Crossing laterally (right to left)**
- **Behavior: Does not enter tram track but approaches clearance boundary**

(D) Pedestrians / cyclists

None visible in the scene

2. Collision Risk Evaluation (3-level scale)

Actor	Risk Level	Justification
Oncoming tram	**Safe**	Fixed rail separation
Parallel cars	**Safe**	Physically separated from tram track
Crossing car	**Attentive**	Lateral movement near intersection
Overall scene	**Attentive**	Intersection introduces uncertainty

3. Recommended Safest Driving Action

⚠️ **Proceed with caution**

Maintain current speed or apply mild deceleration

Increase attention toward:

Intersection area

Lateral vehicle behavior near tracks

Prepare for braking if:

A vehicle unexpectedly enters the tram clearance zone

Summary (Final Answer)

Dynamic actors: Oncoming tram, multiple cars (parallel + crossing)

Highest risk level: Attentive

Safest action:

👉 ***Proceed forward with increased attention; be prepared to brake near the intersection***

This scene represents a **low-risk but non-trivial urban interaction**, typical of tram operation near intersections.

Gemini 3 Pro

1. Scene Description & Dynamic Actor Identification

Environment:

- **Ego Vehicle: Tram moving forward on dedicated tracks.**
- **Scene: Urban street intersection or merging area. Daylight conditions with good visibility.**

Dynamic Actors:

- **Primary Actor (Black Sedan): Located directly ahead in the ego tram's lane.**
  - **Position: The car is straddling the tram tracks, occupying the immediate path of travel.**
  - **Orientation: The vehicle is angled slightly to the left, suggesting it is either attempting to turn across the tracks or merge into traffic.**
  - **Movement Status: Based on the position and angle, the car appears to be moving slowly or is momentarily stationary while attempting a maneuver. It is an active obstruction.**
- **Secondary Actors: There are no immediate pedestrians or cyclists visible in the critical danger zone (the tracks directly ahead).**

2. Collision Risk Evaluation

Risk Level: DANGEROUS

Reasoning:

- **Direct Obstruction: The most critical factor is that the black car is physically blocking the tram's fixed path. Unlike a car, the tram cannot steer around the obstacle.**
- **Proximity: The distance to the vehicle appears to be short. Trams have significantly longer braking distances than cars due to their weight and steel-on-steel contact.**
- **Uncertain Intent: The car is in a vulnerable position (broadside/angled). We cannot guarantee the driver will clear the tracks in time. The driver may be waiting for other traffic, leaving them stuck in the tram's path.**

3. Recommended Safest Driving Action

Action: APPLY EMERGENCY BRAKE (STOP IMMEDIATELY)

Reasoning:

- The potential for a collision is imminent.
- Even if the car is moving, the tram driver cannot rely on the car clearing the path.
- The only variable the tram controls is its own speed. Reducing speed to zero immediately is the only way to guarantee safety or, at worst, minimize the force of impact.
- **Secondary Action: Sound the warning bell/horn continuously to alert the driver of the black car to the immediate danger.**

Figure 7: Sample results from GPT-5 and Gemini 3 Pro for Scene No. 6 (CoT).

Qwen 3 VL

Scene: An oncoming tram is stopped directly ahead, blocking the driver's line of sight. This creates a potential hazard as pedestrians may appear suddenly from behind the tram, raising the risk of collision.
Risk: Attentive
Action: Move forward cautiously and remain alert for unexpected movement.

Claude: Sonnet 4.5

Scene: An oncoming tram is traveling on the parallel track in the opposite direction. The tram is approaching and passing by the ego tram's position. The tracks appear clear with proper separation between the two tramlines, and both vehicles are maintaining their respective lanes on what appears to be a straight section of track in an urban area.
Risk: Attentive
Action: Maintain speed and stay alert for any unexpected movements or passengers near the tracks between the two trams.

Llama 3.2 Vision

Scene: A tram is approaching the platform where the ego tram is likely to stop. The oncoming tram is visible on the adjacent track, indicating that it is not an immediate obstruction but is relevant to the scene as it shares the same stopping area. The presence of the oncoming tram suggests that passengers might be waiting or moving around the area, which could potentially lead to unexpected movements into the ego tram's path.
Risk: Attentive
Action: Proceed with caution

DeepSeek VL2

Scene: The tram is approaching a station where passengers are waiting to board.
Risk: Attentive (Potential for collisions or accidents due to passengers stepping onto the tracks)
Action: Slow down and prepare to stop if necessary.

Figure 8: Sample results for Scene No. 9 (Few-shot).

4 Conclusions and Contributions

In this study, the authors have taken a pioneering step into an unexplored research area, particularly a topic on the potential application of VLMs in assessing collision risk in the context of tramways. The widely recognized multi-modal models and VLMs, GPT-5, Gemini 3 Pro, Claude Sonnet 4.5, Llama 3.2 Vision, Qwen 3 VL and DeepSeek 2 VL, were preliminarily tested using zero-shot, few-shot and CoT prompts with the scene videos containing diverse running situations from the simulator.

Based on the preliminary tests, the findings indicate that these VLMs demonstrate promising potential for collision risk assessment, semantic scene understanding and recommending reasonable driving actions across varied tram driving conditions. More specifically, the models exhibit the capability to capture salient contextual features and components: passengers, surrounding cars, traffic lights and oncoming trams etc., in the given scenarios to some extents depending on the models and the prompting strategy. Overall, it could be briefly concluded that VLMs may serve as a supportive driver assistance system in practical applications, potentially contributing to improved safety in tram operations. Additionally, this work is anticipated to encourage further research and development not only in vision-language models for safety-critical tram applications, but also in multimodal intelligent transportation innovations.

For future work, full validation of each VLM's performance on the intended tasks is planned. Expert tram drivers will be recruited to take part in the questionnaire-based survey. The questionnaires are structured to elicit expert judgments on risk level, salient traffic elements and appropriate driving responses for each individual test scenario. Answers, opinions and reasoning processes from the expert drivers will be leveraged as ground truth data. After completion of the survey, the ground truth data will be compared with the outputs of VLMs to gain comprehensive insight into their strengths and limitations. Similarly, incorporating a broader spectrum of tram driving scenes will be helpful to verify model generalizability. Furthermore, to better facilitate real-world operations, the next effort will focus on the enhancement of the VLMs themselves. With respect to scene understanding capabilities, training and tuning the models using larger-scale tram-specific scene datasets are expected to be beneficial. Moreover, if such tram-traffic datasets, covering scene images and traffic information, become available in the future, we could further propose a novel VLM tailored to collision risk assessment and traffic management in tram domains, as well as related computer vision and automation across various transportation modes.

References

[1] Guofa Li, Yifan Yang, Tingru Zhang, Xingda Qu, Dongpu Cao, Bo Cheng and Keqiang Li, "Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios", Transportation Research Part C: Emerging Technologies, Volume 122, 2021.

[2] Samyeul Noh, " Decision-Making Framework for Autonomous Driving at Road Intersections: Safeguarding Against Collision, Overly Conservative Behavior, and Violation Vehicles", IEEE Transactions on Industrial Electronics, Volume 66, No. 4, 2019.

[3] Jingyi Zhang, Jiaxing Huang, Sheng Jin and Shijian Lu, "Vision-Language Models for Vision Tasks: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 46, No. 8, United Kingdom, 2024.
[4] Wuchang Zhong, Jinglin Huang, Maoqiang Wu, Weinan Luo and Rong Yu, "Large language model based system with causal inference and Chain-of-Thoughts reasoning for traffic scene risk assessment", Knowledge-Based Systems, Volume 319, 2025.