SitVLM2Drive: Situational Awareness Vision-Language Models to Enhance Safety and Reliability in Autonomous Driving

Supplementary Material – Dataset

1. Introduction

This supplementary material highlights the SitVLM2Drive dataset, engineered through a rigorous lifecycle process following ISO-PAS 8800, from analysis, design, development, and implementation, with detailed statistical analysis. The multifaceted analysis presented here complements the main text and provides researchers with additional insights into the dataset's design and applicability to real-world AV scenarios. In the following sections, we describe the analysis, design, development, and implementation phases, elaborate on the annotation process, and present additional visualizations that extend the main text's findings. An outline is provided below for readers interested in targeted aspects of the work.

2. Lifecycle for SitVLM2Drive

The lifecycle of SitVLM2Drive is organized into multiple phases, as illustrated in Figure 1. Each phase is designed to ensure data robustness, relevance, and compliance with AI safety.

2.1. Dual submission

The Analysis stage lays the foundation for the entire SitVLM2Drive dataset lifecycle, which consists of two layers: the Inception and Conceptual phases.

2.1.1 Inception Dataset

The Inception phase initiates the scope of SitVLM2Drive by identifying diverse, real-world driving contexts that encompass both routine and edge - case autonomous vehicle (AV) scenarios. To achieve this, we will develop a graph of objects with defined properties (six-space), represented by the tuple < position, status, relation, ranking, impacting ego safety, time > through a Directed Acyclic Graph (DAG), to analyze causal relations from a traffic system perspective and building Causal Graph (CG). This approach ensures that the curated data accurately reflects diverse traffic patterns, environmental conditions, and safety-critical events. Special emphasis is placed on hazardous elements such as pedestrian crossings, construction zones, and sudden obstacles to enhance the dataset's relevance to safety.

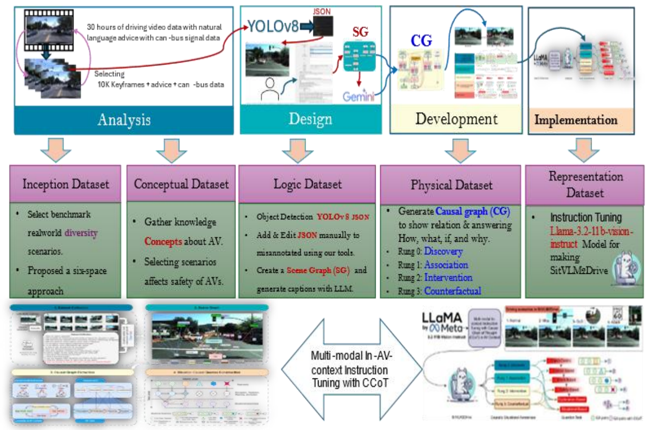


Figure 1: Overview of the SitVLM2Drive dataset lifecycle, divided into four major phases—Analysis, Design, Development, and Implementation.

2.1.2 Conceptual Dataset

Building on this groundwork, the Conceptual phase defines foundational knowledge for scene understanding. Core topics—like traffic rules, hazard perception, and agent interactions (vehicles, pedestrians, traffic signals)—are refined to ensure coverage of essential driving phenomena (see Section 3). During this stage, scenario filtering ensures the dataset highlights impactful situations, laying the conceptual bedrock for high-fidelity annotations and causal reasoning in subsequent stages.

2.2. Design

In the Design phase, the primary goal is to formalize logic within each driving scene

2.2.1 Logic Dataset

In the Logic phase, attention shifts to in-depth annotation of scene structure and relationships. Object-detection models (e.g., YOLOv8) provide initial bounding boxes for vehicles, pedestrians, and signage. Subsequently, expert annotators meticulously verify and refine these annotations, ensuring accurate representation by adding the precise position and status for all identified objects. The outcome is a set of Six Space capturing inter-object relationships—who or what interacts with whom and how these interactions evolve (See our Tools Guidelines GUI

Image Annotation PDF for more information). Large Language Models (LLMs) subsequently generate captions describing each scene's underlying logic, ensuring that crucial context is documented for later analysis.

2.3. Development

The Development stage centers on the dataset's physical aspects and causal relationships.

2.3.1 Physical Dataset

During the Physical phase, the focus pivots to causal analysis and incorporating real-world measurements such as distances, velocities, and timing. Researchers construct causal queries—addressing "who, what, if, why?"—to track the chain of events and identify factors contributing to each scenario's risk or complexity. Through thorough data management (e.g., metadata tagging, consistency checks), this stage integrates the quantifiable dimensions of driving environments, bolstering authenticity and granularity in situational understanding. Table 2 provides an overview of the dataset's core fields alongside sample values, illustrating how these components collectively capture the complexity of real-world driving situations.

2.4. Implementation

2.4.1 Representation Dataset

Finally, the Representation phase shapes SitVLM2Drive into a multimodal, instruction-tuned resource. Advanced vision–language models (such as *Llama-3.2-11b-vision-instruct*) ingest the annotated dataset to learn from in-context reasoning prompts—e.g., Our Causal Chain-of-Thought (CCoT). These enhanced capabilities enable counterfactual queries and scenario exploration, thereby supporting high-level AV research tasks and decision-making. The result is a comprehensive dataset that can be seamlessly integrated into cutting-edge autonomous vehicle development pipelines.

3. Detailed Scenario and Infrastructure Annotations

We categorize our dataset into two primary groups: object-level and scene-level annotations. At the object level, we established various grouping contexts tailored to the nature of the objects, aimed at enhancing the dataset's applicability in real-world scenarios, such as Traffic Infrastructure, Road Users, Vehicles, and Ego Vehicles. Conversely, the scene-level annotations encompass a range of elements, including broader contextual information (e.g., weather, traffic density) and higher-level semantics (e.g., captions, causes, goals), all of which consider the diverse scene and environmental contexts.

3.1. Scenario-Level

Scene-level annotations can also include:

- Caption
- Maneuver
- Cause
- Goal-Oriented
- Safe
- Action Suggestions
- Traffic Regulations Suggestions

These higher-level descriptors allow you to incorporate narrative and decision-making hints.

3.2. Object-Level

Object-Level annotations: Focus on identifying and classifying individual vehicles, pedestrians, ego, and infrastructure components:

3.2.1 Traffic Infrastructure

Annotations under the Traffic Infrastructure document all static elements of the road environment (See Table 3):

- Static Elements: Traffic signals, road signs, crosswalks, parking areas, and environmental hazards.
- Special Zones: Specific areas such as construction zones, hazardous zones, parking spaces, bus stops, gas stations, and designated zones (e.g., Smartphone Zombie Ribbon).
- Additional Elements: Sidewalks, street furniture, parking meters, intersections (e.g., Four-Way, T-Intersections, Y-junctions, N-Intersections), roundabouts, over/underpasses, and medians (islands).
- Road Types and Conditions: Highways, city streets, rural roads, railway lines, bus lines, and bike lanes, along with road conditions such as slippery or uneven surfaces.
- Road Markings and Objects: Lane lines, crosswalks, arrows, stop lines, text markings, and zone colors, as well as road objects like debris, potholes, guardrails, barriers, cones, and bollards (see Section 4).
- Traffic Signals: Detailed specifications, including walk lights (with audio signals), non-flashing and flashing signals (Red, Amber, Green), countdown

timers, green arrows, bus/tram signals, and animated signals for pedestrians and cyclists.

3.2.2 Road Users and Vehicles

The dataset also categorizes entities that interact with or navigate the road. Road Users are Non-vehicle entities such as pedestrians, riders (e.g., cyclists, skateboarders), and animals.

3.2.3 Vehicles

This category encompasses both motorized and non-motorized modes of transport. The categories include standard vehicles, trucks, motorcycles (or motorized scooters), buses, construction vehicles, trains, airplanes, and emergency vehicles (e.g., ambulances, fire trucks, and police vehicles). Notably, the Ego Vehicle (the autonomous vehicle itself) is distinctly annotated in a separate grouping from other vehicles because it is more significant for us to categorize it alone.

4. Action Tasks

The dataset includes standardized definitions for various driving actions to support dynamic decision-making. These annotations provide temporal and contextual details for each maneuver:
Action List: [Moving Forward, Stopping, Parking, Turning Left, Turning Right, U-Turning, Breaking, Accelerating, Decelerating, Reversing, and Overtaking]
These action tasks facilitate modeling both routine and emergency maneuvers, ensuring that AV systems can generate appropriate responses under diverse conditions.

5. Road and Lane Annotations

The dataset provides fine-grained annotations of road and lane configurations to support advanced spatial reasoning:
Lane Numbering: Identification of lanes based on the ego vehicle's direction (e.g., "just ego direction").

- Lane Descriptions: Detailed labels such as "lane 1," "lane 2," progressing from the passing lane to the lane adjacent to the sidewalk, including additional types:
  - Turn Lane Added
  - Left Turn Only Lane
  - Right Turn Only Lane
  - Temporary Lane Shift
  - Lane Reduction Ahead
  - Lane Marking Faded
- Median Types: Classification into "Median-Raised," "Median-Depressed," "Median-Closed," "Median-Flush," or "Median-None."
  - For example, "Median Flush Lane Solid Line White, "Median Flush Lane Dashed Line Yellow," etc.
- Inter-lane Markings: Annotation of lane markings between lanes (e.g., "Lane_Solid_Line_White_1_2," "Lane_Dashed_Line_Yellow_1_2").
- Road Type and Intersection Conditions: Specification of the type of road (e.g., highway, city street) and a detailed description of road sections within intersections. For instance, based on the ego vehicle's position, road sections are numbered from the right (Ego Section 1) onward.

6. Statistical Analysis and Visualization

A thorough statistical analysis of SitVLM2Drive underscores its utility for autonomous vehicle (AV) research. The dataset comprises various data modalities, including video frames, annotated textual captions, QA pairs, and object relationships. In what follows, we detail the key statistics and visual analyses.

6.1. Overview of Dataset Statistics

We performed an extensive statistical analysis of the SitVLM2Drive dataset to understand its composition and demonstrate its utility for autonomous vehicle (AV) research. Table 1 provides a detailed dataset summary, including the number of videos, frames, question-answer (QA) pairs, object types, and their distribution across dynamic and static categories. The dataset also includes various reasoning questions, such as discovery, association, intervention, and counterfactual, enabling a robust evaluation of AV models across multiple reasoning tasks. Notably, 40% of the dataset is categorized as safety-critical, highlighting its emphasis on scenarios that directly impact driving decisions.

Table 1 Dataset

Category	Count	Percentage
Total Videos	495	
Total Frames	10250	
Total QA Pairs	2083050	
Total Objects	158600	
Unique Object Types	76	

6.2.

6.3. Caption and Maneuver Analysis

We analyzed the captions accompanying the videos to understand the diversity and richness of the scene annotations. Figure 2 illustrates the caption word cloud, while Figure 3 provides a Maneuver word cloud highlighting frequently used terms. The captions exhibit significant variability in length and vocabulary, reflecting the complexity of the annotated scenarios.

Figure 1 Word Cloud for Captions

Figure 2 Word Cloud for Maneuver

6.4. Question-Answer Distribution

The QA pairs in SitVLM2Drive span various reasoning tasks. Figures 4 and 5 present word clouds for questions and answers, respectively, highlighting the diversity of topics and responses covered in the dataset.

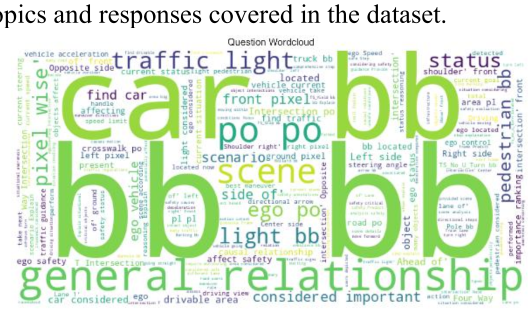


Figure 3 Word Cloud for Questions

6.5. Object and Importance Ranking Analysis

The dataset includes various objects and relationships critical for AV perception and decision-making. Figure 6 shows the most common objects, while Figure 7 illustrates the distribution of object importance ranking via objects class,.

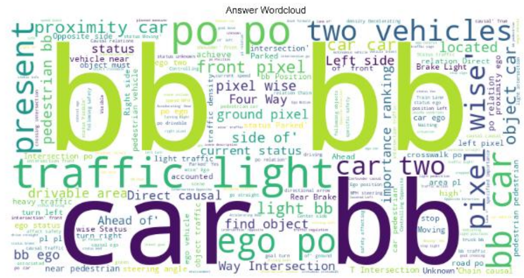


Figure 4 Word Cloud for Answers

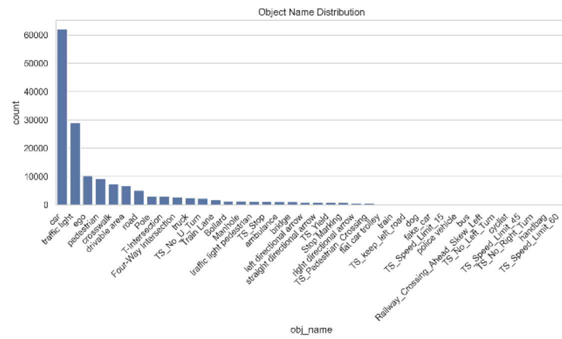


Figure 5 Top 40 Object Names

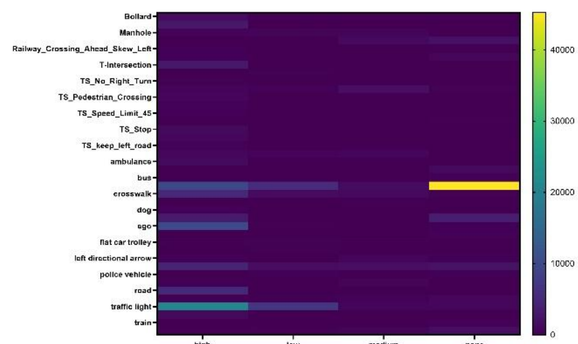


Figure 6 Object Importance Ranking via Object Classes

6.6. Action and Regulation Analysis

Figures # and # explore the dataset’s focus on action suggestions and traffic regulation guidance. These aspects are critical for modeling safe and compliant AV behavior.

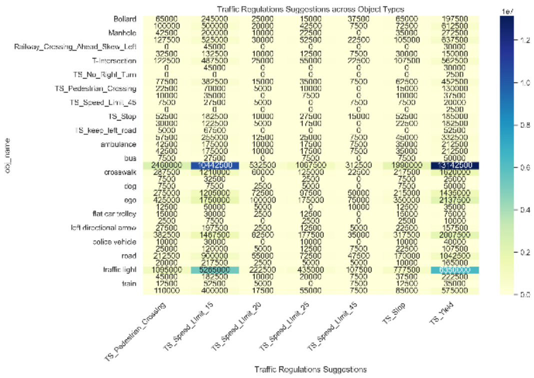


Figure 7 Traffic Regulation Suggestions

Figure 8 Goal-Oriented for Ego

7. Causally Situational Awareness-Based Question Framework Overview

In our approach, we employ a layered question-and-answer framework to enhance situational awareness in autonomous vehicles by systematically classifying Q&A pairs across four "rungs" of complexity (Discovery, Association, Intervention, and Counterfactual), various autonomous vehicle tasks (perception, prediction, plan, and action), and distinct question styles (Object-Centric, Context-Based, Event(Temporal)-Based, Safety-Based, Explanation-Based, and Planning-Based). Each Q&A pair includes a question template (e.g., "How many [objects] are there?") and an answer template (e.g., "There are [count] [objects] in the scene."), along with metadata tags (Type, AV_Task, question_task) indicating which system function and rung of complexity the question addresses (Table 4).

This design labels basic inquiries, such as identifying or locating objects, as "Discovery" with AV_Task "perception." In contrast, more advanced or hypothetical inquiries, such as explaining maneuver decisions or testing counterfactual conditions, are tagged as "Association," "Intervention," or "Counterfactual," with corresponding tasks like "prediction" or "plan." By structuring questions and answers this way, our approach can automatically generate comprehensive, human-readable outputs (e.g., "Is the vehicle exceeding the speed limit?" -> "Yes, the current speed is [speed] MPH, exceeding [limit].") and support detailed Causal-chain-of-thought (CCoT) explanations for various driving scenarios, including normal operation, cases with missed detections, situations where relevant objects are not in the scene, and adversarial (attack) scenarios. In what follows, we describe how each scenario is handled (Figure ## for Scenarios).

7.1. Normal Driving Scenario

In the normal driving scenario, the system processes each frame by extracting objects, identifying relationships, and generating Q&A pairs that reflect straightforward driving conditions, i.e., all relevant objects are detected accurately, and no adversarial attack is in play. Here, questions at the Discovery rung focus on identifying and counting objects (e.g., "How many vehicles are there?"), While the Association rung questions that address object relationships and context (e.g., "How do object A and object B interact?"). If safety-related conditions arise—such as speed compliance—the system may shift into Intervention rung inquiries (e.g., "Is the vehicle exceeding the speed limit?").

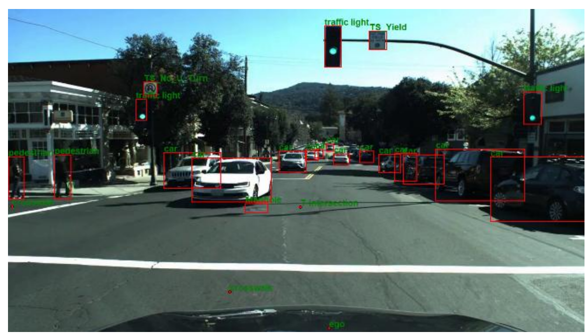


Figure 9 Normal Driving Scenario

7.2. Miss-Detected Objects

In the miss-detected scenario, particular objects may fail to register properly—possibly due to visual obstructions or partial occlusions. The system still employs the same question types (Discovery, Association, etc.) but tailors the answers to account for uncertain or missing data (e.g., "Is there a [stop sign] in the scene?" → "Uncertain: partial detection suggests [stop sign], but confidence is low."). Intervention rung questions can highlight the potential safety impact of unrecognized hazards in such cases.

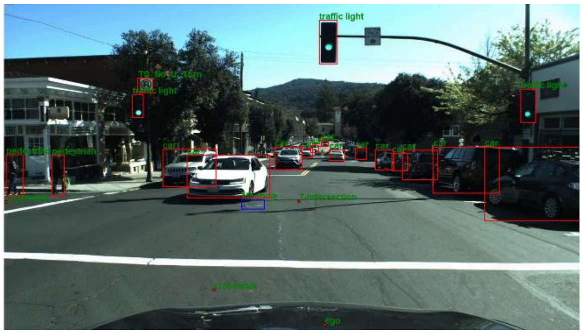


Figure 10 Example of a driving scenario involving a missed object: The yield sign is not detected by the model.

7.3. Out of Distribution

When particular objects or features are supposed to be in the scene, but the system detects none, it falls under the OoD. The Q&A framework can generate clarifications, such as “Is this an object affecting the safety of the ego → ? “No, this object is not in a drivable area for the ego,” prompting either a fallback action. This scenario essentially tests how the framework handles an unexpected element in the scene.

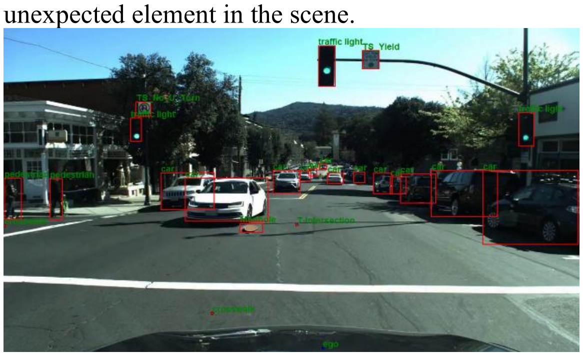


Figure 11 Example of an OoD driving scenario: Manhole status - opened.

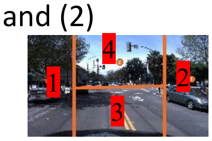
7.4. Adversarial / Attack Scenarios

Finally, in adversarial or attack scenarios, the system identifies that the image may contain Gaussian noise, Sit-Patch, or similar manipulations. In these conditions, the Q&A pairs handle both basic safety checks (e.g., verifying sensor integrity or normal detection operations) and more advanced Counterfactual rung queries, such as “Explain step-by-step how the ego vehicle should decide when to [goal] under a [Noise/Patch] attack.” Answers may include extra cautionary steps and sensor-filtering measures to mitigate adversarial effects. By assigning these advanced inquiries to the Association, Intervention, or Counterfactual rungs (with AV_Task often labeled “prediction” or “plan”), the system highlights its layered approach to adversarial robustness.

Figure 12 Example of an attack driving scenario: A pedestrian is holding a traffic sign to confront the ego car.

Table 2 Key Annotation Fields and Their Descriptions

Key	Description for Value	Example
image_id	A unique identifier (often the filename) for the image resource	frame_0000.jpg
Caption	Scene description from the driver's view, enhanced using LLM	cars are coming in the opposite lane, and while are waiting to turn left ... then the stop light is green cars are coming I turn left
Maneuver	The driver’s actions in the scene are enhanced using LLM	yield to oncoming traffic and wait to turn left when traffic in the oncoming lane is clear and, make the left turn, then continue down the block then stop at an intersection and turn left, and the stop

		light is green, I am stopped at the light I am waiting to turn left I turn left and after all cars in oncoming traffic either turn or keep straight
speed	A numeric value representing the vehicle's speed	X (e.g., [0<X] means moving, [X=0] means stopped)
steering	A numeric value for the steering command	X (e.g., [X<0] for Turn Left, [0<X] for Turn Right, and [X=0] for Straight)
Action Suggestions	Recommended driving action	Stop(Yield) see Action List
Traffic Regulations Suggestions	Traffic law or regulation advice	TS_Yield
goal-oriented	Driver's intended objective through the scene	Turn left
safe	Evaluation of the scene's safety relative to the driver's goal	It is not safe for reason cars have the right-of-way more than ego
Causes	Identifies the reasons behind certain traffic events or conditions (e.g., Traffic Congestion, Lane Addition Ahead, Sun Glare).	[Intersection Congestion….]
Object name	Name or identifier of the object detected point, Bbox, polyline	"ego", "car", …
object type	Category of the object	Ego-Infrastructure, Ego-Vehicle, Ego-Ego, Ego-Road Users
Object Position	The spatial relationship of the ego car to other objects or the location of objects relative to other elements, Geometry Position from pixel-wise (H shape) four areas, Dynamic Positioning: Include relative velocities or trajectories, and 3D Spatial Information: For environments with multi-level roads or overpasses, specify vertical positions.	Spatial Relations➔ ahead of obj X, left side... Geometry Position * ➔ left-pixel-wise… (auto) Dynamic Position➔ Decelerating before Stop Sign 3D Spatial Position➔ "Above" and "Below". *Type H image splitter. (1) and (2) represent the left and right area, (3) represent the ground area, and (4) represent the front area 

Object Status	Describes the current state or behavior of various objects (e.g., ego car, car, pedestrians, traffic lights).	"Moving"
importance ranking	Priority level or significance of the object or event	"high"
Causal cause	Indicates how the initiating event (cause) is connected to the resulting event (effect)	"Sudden braking "
Causal Effect	Describe how the outcome influences later events or conditions in the traffic scene.	"triggers traffic congestion."
Object safety	A list of safety-related categories that describe how the object affects safety. Options include: "Affects Safety," "Potentially Affect Safety," "Does Not Affect Safety," and "Requires Monitoring."	"Affects Safety"
QA	Structured Q&A pairs for scene details	"QA":[{"Q": "Question", "A": "Answer", "Type": "e.g. Discovery", "question_task": "e.g., object Centric", "AV_Task": "e.g., perception", "scene_scenario": "Normal, Miss, Attack…",}]
graph	"nodes": Representation of scene entities in a graph format	"ego <po>x, y</po>", ➔ obj ID { "object_name": "ego", "object_type": "Ego-Ego", "object_Causal": "cause or effect", "boxes": [x1,y1, x2,y2], or "point": [x, y], or "polyline": [[x1, y1],[x2, y2]... [xn, yn]], "importance_ranking": "high", "status": [], "object_safety": ["Does Not Affect Safety"], "position": [] }
	"edges": Representation of relationships between graph nodes	obj1{coord}, or obj_cause{coord}, obj2{coord}, or obj_effect{coord}", {"relation": "Relationship" or "causal relation" }
Checked	Verification status flag	A Boolean value (T/F)
Confirmed	Confirmation status flag	A Boolean value (T/F)

Table 3 Manually Designed For The Road Traffic Domain

Scenario	Scene-level	diverse scenes and environmental contexts	weather	lighting	Traffic Density Levels: Categorize traffic density (e.g., "Light," "Moderate," "Heavy," etc.)				
	Traffic Infrastructure	includes all static elements of the road environment, such as traffic signals, signs, road types, crosswalks, parking areas, and environmental hazards.	Especially Zone *	Construction-Zone	Hazard-Zone	Parking-Space	Bus-Stop	Gas Station	Smartphone Zombie ribbon
			Sidewalk	Street-Furniture	Parking-Meter				
			Intersection	At grade	4-Way-Intersection	T-Intersection	Y-junction	N-Intersection	
					Roundabout				
				Grade separated	Over/Underpasses	Trumpet	Diamond	Cloverleaf	Diretional
			Medians (islands)						
			Road Type	Highway	City street	Rural road	Railway line	Bus line	Bike Line
			Road Condition	Slippery	Uneven				
			Road Markings	Lane Lines	Crosswalks	Arrows	Stop Lines	Text Marking	Zone Color *
			Road Objects	Debris	Pothole	Guardrail	Barrier	Cone	Bollard
			Sign	Regulatory Sign					
				Warning Sign					
				Informational Sign					
				other					
			Traffic Light	walk light (audio signals)	Non-flashing (Red, Amber) and Flashing (Red, Amber, Green)	Countdown lights	Green arrows	Bus and tram signals	Traffic light animation (pedestrians, cyclists, and traffic)
	Road users	primarily include non-vehicle entities interacting with the road environment, such as pedestrians, animals, and portable items.	Pedestrians						
			Rider (Cyclists, skateboard, ...)						
			Animal						
	Vehicle	Vehicles include all types of motorized and non-motorized transport means.	Standard Vehicles	Truck	Motorcyclists (motorized scooters and similar devices)	Bus	Construction Vehicles		
			train	airplane					
			Emergency Vehicles	Ambulance	Fire Truck	Police Vehicles			
		Ego Vehicle	the autonomous vehicle itself						

Table 4 QA Template

Q Template	A Template	Type	Task	question_task	AV_Task
1. Infrastructure Q&A					
What infrastructure is present in the scene?	Infrastructure objects found: {list_of_infrastructure_objects}	Association	Context-Based Questions	Scene Classification	perception
Is this a driving situation involving a 4-way intersection, T-intersection, or roundabout?	The scene includes: {intersection_type}	Association	Context-Based Questions	Scene Classification	perception
What objects are located at {pos}?	{obj_name} is located at {pos}.	Discovery	Object-Centric Questions	Position Identification	perception
Where can I find {oid}?	You can find object {oid} ({obj_name}) at {pos}.	Discovery	Object-Centric Questions	Position Identification	perception
What is the status of {oid} on {pos}?	The status of object {oid} is: {status_str}.	Discovery	Object-Centric Questions	State Identification	perception
2. Traffic Signs Q&A (from infrastructure/detect_traffic_signs)					
Which traffic signs are present in the scene, and where are they located?	Traffic signs found: {sign_names} at positions {sign_positions}.	Discovery	Object-Centric Questions	object direction	perception
How many traffic signs are present in the scene?	Number of traffic signs: {count_signs}	Discovery	Object-Centric Questions	Counting	perception
Where is the {sign_name} located, and what is its status? (Position)	{sign_name} is located at {sign_position} <bb>{boxes}</bb>.	Discovery	Object-Centric Questions	Position Identification	perception
Where is the {sign_name} located, and what is its status? (Status)	{sign_name} status: {status_str}	Discovery	Object-Centric Questions	State Identification	perception
Where is the {sign_name} located, and what is its status? (Combined)	{sign_name} is located at {sign_position} <bb>{boxes}</bb>, with status: {status_str}	Discovery	Object-Centric Questions	Position & State Identification	perception
What rule does the {sign_name} at position {sign_position} impose?	*E.g. {sign_name} <bb>{boxes}</bb>: Stop means the ego vehicle must come to a complete stop.* (varies per sign type)	Association	Context-Based Questions	Scene Classification	plan
3. Traffic Lights Q&A (from infrastructure/detect_traffic_lights)					
How many traffic lights are present in the scene?	There are {num_lights} traffic light(s) detected.	Discovery	Object-Centric Questions	Counting	perception

Where is the {obj_name} <bb>{boxes}</bb> located?	The {obj_name} <bb>{boxes}</bb> is located at {position}.	Discovery	Object-Centric Questions	object direction	perception
What is the current status of the {obj_name} <bb>{boxes}</bb>?	The current status of the {obj_name} <bb>{boxes}</bb> is {light_status}.	Discovery	Object-Centric Questions	State Identification	perception
What is the importance ranking of the {obj_name} <bb>{boxes}</bb>?	The {obj_name} <bb>{boxes}</bb> has a {importance_ranking} importance ranking.	Association	Context-Based Questions	Relationship	perception
What is affecting the ego's safety from this {obj_name} <bb>{boxes}</bb>?	The {obj_name} is associated with the following safety: {safety_objects}.	Association	Context-Based Questions	Relationship	perception
4. Traffic Markings Q&A (from infrastructure/detect_markings)					
Which traffic road markings are present in the scene?	Road markings are present: {list_of_road_markings}.	Discovery	Object-Centric Questions	object direction	perception
How many traffic road markings are present in the scene?	The number of road markings is: {marking_count}.	Discovery	Object-Centric Questions	Counting	perception
Where is the {name} located?	{name} is located at {position}.	Discovery	Object-Centric Questions	Position Identification	perception
Where is the {name} status?	{name} is status: {status_str}.	Discovery	State Identification	State Identification	perception
Where is the {name} located, and what is its status?	{name} is located at {position}, with status: {status_str}.	Discovery	Object-Centric Questions	Position & State Identification	perception
5. Vehicle Q&A					
Is this a driving situation involving a police vehicle, ambulance, construction vehicle or fire truck?	Yes, includes: {list_of_emergency_vehicles}	Association	Object-Centric Questions	existence	perception
What types of vehicles are present in the scene, and how many are there?	There are {count_vehicles} vehicles. Types: {unique_types}	Discovery	Object-Centric Questions	Counting	perception
Where can I find {oid}? (vehicle)	The object {obj_name} is at position {pos}.	Discovery	Object-Centric Questions	Position Identification	perception
What is the status of {oid} on {pos}? (vehicle)	The status of object {oid} is: {status_str}	Discovery	Object-Centric Questions	State Identification	perception

What are the spatial details and emergency status of {obj_name} {oid}?	The {obj_name} is located at {pos}. Its emergency status includes: {emergency_status_str}.	Association	Object-Centric Questions	State Identification	perception
6. Road Users Q&A					
What are the road users depicted in the scene?	The road users in this scene include: {list_of_road_users}.	Association	Context-Based Questions	Scene Classification	perception
How many road users are present in the scene?	There are {count_road_users} road user(s) detected.	Discovery	Object-Centric Questions	Counting	perception
Where is the {oid} located? (road user)	The {oid} is located at {pos}.	Discovery	Object-Centric Questions	Position Identification	perception
What is the status of the {oid}? (road user)	The current status of the {oid} is: {status_str}.	Discovery	Object-Centric Questions	State Identification	perception
What is the importance ranking of the {oid}? (road user)	The {oid} has an importance ranking of: {importance}.	Association	Context-Based Questions	Relationship	perception
What safety annotations are associated with the {oid}? (road user)	The {oid} has the following safety annotations: {safety_str}.	Intervention	Safety-Based Questions	Risk & Anomaly Detection	perception
Is the {oid} exiting the scene?	The exiting status of the {oid} is: {exiting_str}.	Discovery	Object-Centric Questions	existence	perception
7. Ego Q&A					
Where is the ego located now?	The ego is located at {position}.	Discovery	Object-Centric Questions	Object-Centric Questions	perception
What is the status of the ego? (Speed and steering)	Speed: {speed} and Steering: {steering}.	Discovery	Object-Centric Questions	Object-Centric Questions	perception
Explain the current situation for this scene.	The ego is navigating the environment with {caption}.	Association	Object-Centric Questions	Context-Based Questions	perception
Explain the maneuver that can be performed in this situation.	{maneuver}	Intervention	Object-Centric Questions	Planning-Based Questions	plan
8. Drivable Area Q&A					
Is there a drivable area for the ego?	Yes, a drivable area is present.	Discovery / Ass.	Context-Based Questions	Scene Classification	plan
Is the drivable area big or small?	The drivable area is {size_category}.	Discovery / Ass.	Context-Based Questions	Scene Classification	plan

Explain situational awareness concerning the ego in relation to the drivable area.	1) The scene includes a drivable area... 2) It defines safe navigable space... 3) It is {size_category}, affecting strategy.	Association	Context-Based Questions	Scene Classification	plan
9. Global Q&A					
What is the ego vehicle's current goal in this situation?	{goal}	Association	Object-Centric Questions	Context-Based Questions	plan
What is the best maneuver the ego vehicle can perform in this scene according to {goal}?	{maneuver}	Association	Planning-Based Questions	Planning	plan
What action should the vehicle take next to achieve {goal}?	Current speed: {speed} MPh, Steering: {steering}. Recommended Action: {action_suggestions}	Intervention	Event(Temporal)-Based Questions	Predictive	action
What traffic regulations should the ego vehicle consider to achieve {goal}?	The ego vehicle should consider: {traffic_regulations}.	Intervention	Planning-Based Questions	Planning	plan
Is the current situation considered safe?	{safety_status}. Reasoning: {safety_reasoning}	Intervention	Safety-Based Questions	Safety-Based Questions	action
Is there a {obj_id} in the scene?	Yes, a {obj_name} is present with status {status}.	Discovery	Object-Centric Questions	existence	perception
Is there not an {obj_name} in the scene?	No, a {obj_name} is present with status {status}.	Discovery	Object-Centric Questions	existence	perception
What are the causes of the current safety issue?	{cause_list}	Association	Object-Centric Questions	Scene Classification	prediction
How many objects are detected in total from the driving view?	There are {count_of_objects} objects detected in the scene.	Discovery	Object-Centric Questions	Counting	perception
Are there any causal relationships between objects in the scene?	Causal relationships detected: {causal_edges}	Association	Planning-Based Questions	Relationship	plan
10. Object Safety Categorization Q&A					
Which objects {category_phrase}? (e.g. *affect safety*, *potentially affect safety*)	*E.g.* "The objects that affect safety are: {obj_list}." or "The following objects require monitoring: {obj_list}."	Intervention	Safety-Based Questions	Risk & Anomaly Detection	plan
11. Relation Q&A					

What is the {relation_type} relationship between {source_id} and {target_id}?	{source_id} is {relation_text} {target_id}.	Association	Context-Based Questions	Relationship	plan
12. CCoT Reasoning Q&A					
Step-by-step, explain how the ego vehicle should handle the current situation...	(7-step chain-of-thought) 1) Scene Safety... 2) Cause... 3) Ego Motion... etc.	CCot	Planning-Based Questions	Planning	plan
Provide a comprehensive, step-by-step explanation of how the ego vehicle should act...	(7-step chain-of-thought) 1) High-Ranking Objects... 2) Causes... 3) Ego Motion... etc.	CCot	Planning-Based Questions	Planning	plan
Using the provided scene details and directional steps, explain how the ego vehicle should decide...	(Detailed CCoT for normal scenario) *E.g. "... proceed if safe conditions are met..."*	CCot	Planning-Based Questions	Planning	plan
Detail the step-by-step decision-making process for the ego vehicle to {goal} in an Attack scenario...	(Detailed CCoT for Attack scenario) *E.g. "... adopt extreme caution if an attacking object is identified..."*	counterfactual	Planning-Based Questions	Planning	plan
Describe the comprehensive decision-making process for the ego vehicle to {goal} under OoD objects...	(Detailed CCoT for OoD scenario) *E.g. "... re-assess environment with heightened vigilance..."*	counterfactual	Planning-Based Questions	Planning	plan
Explain the step-by-step reasoning for the ego vehicle to {goal} in a miss-detection scenario...	(Detailed CCoT for miss-detected scenario) *E.g. "... remain alert to unseen hazards..."*	counterfactual	Planning-Based Questions	Planning	plan
13. Speed & Temporal Q&A					
What is the ego vehicle's current speed?	The current speed is {speed} MPh.	Intervention	Event(Temporal)-Based Questions	Predictive	perception
Predict the ego vehicle's acceleration or deceleration compared to the previous frame.	The vehicle {accelerated/decelerated} by {abs(speed_diff)} MPh compared to the previous frame.	Intervention	Event(Temporal)-Based Questions	Predictive	perception

Predict the ego vehicle's acceleration or deceleration in the next frame.	The vehicle is expected to {accelerate/decelerate} by {abs(speed_diff)} MPh in the next frame.	Intervention	Event(Temporal)-Based Questions	Predictive	prediction
Is the vehicle exceeding the speed limit?	Yes, the current speed of {speed} MPh exceeds the limit {limit} MPh.	Intervention	Event(Temporal)-Based Questions	Predictive	perception
Is the vehicle moving within the speed limit?	Yes, the current speed of {speed} MPh is within the limit {limit} MPh.	Intervention	Event(Temporal)-Based Questions	Predictive	perception
What is the applicable speed limit in this area?	The speed limit is {speed_limit} MPh.	Discovery/Inter.	Event(Temporal)-Based Questions	Predictive	perception
14. Turning & Steering Q&A					
Is the ego vehicle attempting to turn {intended_turn} where it is prohibited?	Yes, the ego vehicle is attempting to turn {intended_turn}, but a 'No {intended_turn}' sign is present.	Intervention	Event(Temporal)-Based Questions	Predictive	perception
Should the ego vehicle proceed with the {intended_turn} turn?	No, it should adjust its steering to comply with regulations.	Intervention	Event(Temporal)-Based Questions	Predictive	perception
Is the ego vehicle allowed to turn {intended_turn} in this area? (Allowed)	Yes, the ego vehicle is turning {intended_turn}, which is permitted by directional/arrow signs.	Discovery	Planning-Based Questions	Object-Centric Questions	perception
Is the ego vehicle allowed to turn {intended_turn} in this area? (Not allowed)	No, the ego vehicle is not allowed to turn {intended_turn} as only specific directions are permitted.	Intervention	Planning-Based Questions	Safety-Based Questions	plan
What is the ego vehicle's current steering angle?	The steering angle is {steering}° indicating it is {direction_description}.	Discovery	Planning-Based Questions	Object-Centric Questions	perception
Is the vehicle steering to the left?	Yes, it is steering to the left.	Discovery	Planning-Based Questions	Object-Centric Questions	controlling

Is the vehicle steering to the right?	Yes, it is steering to the right.	Discovery	Planning-Based Questions	Object-Centric Questions	controlling
Is the vehicle going straight?	Yes, it is moving straight with minimal steering input.	Discovery	Planning-Based Questions	Object-Centric Questions	controlling
Is the vehicle's intended direction clear?	The intended direction is unknown.	Discovery	Planning-Based Questions	Object-Centric Questions	perception
15. Importance Reasoning Q&A					
Why is the {obj_name} considered important in this scenario?	*Because {reasoning_explanation} (location, status, link to the goal, mention of attack, etc.)*	Association	Planning-Based Questions	Explanation-Based Questions	plan
16. Safety Reasoning Q&A					
Explain the safety status reasoning.	*Full explanation:* The scene is {safe/unsafe} due to {ego_status, speed, steering, environment reasons}...	Association	planning	Explanation-Based Questions	plan