CEDAR: CONCISE, EVIDENCE-GROUNDED DRIVING ACTION RECOMMENDATION VIA CASCADE REASONING AND POLICY REFINEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Large vision-language models have advanced autonomous driving toward interpretable scene understanding, yet existing models tend to produce verbose responses without explicitly surfacing the visual evidence behind their decisions, limiting their practicality for real-time driving assistance. We formulate the task of concise, evidence-grounded driving recommendation generation, where actionable decisions are explicitly backed by decision-relevant visual evidence and contextual reasoning. To this end, we construct **CODA-GRD** (Grounding, Reasoning & Decision), a unified Evidence-Reason-Suggestion benchmark providing structured supervision for visual grounding, reasoning, and action recommendation. Built upon it, we propose **CEDAR**, trained via a progressive pipeline: Entity Presence Grounding Pretraining aligns visual entities with structured evidence; Chain-of-Insight Cascade Reasoning (CCR) propagates context from Reason to Evidence to Suggestion under a reason-aware contrastive objective; and Driving Policy Refinement applies branch-wise Group Relative Policy Optimization with task-specific rewards. Together, these stages enable CEDAR to generate concise recommendations explicitly grounded in visual evidence and reasoning. Experiments on CODA-GRD and three external benchmarks, OmniDrive, LingoQA, and DriveCoT, show consistent improvements over strong baselines.

1 INTRODUCTION

Recent advances in large vision-language models have pushed autonomous driving beyond standalone visual perception toward multimodal scene understanding and decision reasoning, enabling unified interpretation of traffic scenes, surrounding agents, and driving contexts (Sima et al., 2024; Nie et al., 2024; Tian et al., 2025). Accordingly, generating reliable, actionable driving recommendations has become an essential capability for intelligent driving systems and garnered increasing research interest (Chen et al., 2025a; Wang et al., 2024b).

Despite these advances, existing driving vision-language models tend to produce verbose action recommendations without explicitly surfacing the supporting visual evidence. Prior works predominantly focus on holistic scene understanding, VQA, and chain-based reasoning (Sima et al., 2024; Nie et al., 2024; Tian et al., 2025; Wang et al., 2024b). For instance, DriveLM (Sima et al., 2024) graph-couples perception, prediction, and planning, whereas Reason2Drive (Nie et al., 2024) models reasoning as a multi-step chain. However, these exhaustive paradigms yield lengthy outputs ill-suited for real-time driving assistance and leave key visual evidence unexposed, hindering concise, evidence-grounded decision support.

To address this gap, we formulate the task of concise, evidence-grounded driving recommendation generation, in which final actions are backed by explicit driving rationales and decision-relevant visual evidence. To enable training under this formulation, we first construct the dataset **CODA-GRD** (**G**rounded **R**easoning & **D**ecision) to support our exploration. In particular, we reorganize the CODA-LM (Chen et al., 2025a) into a unified Evidence–Reason–Suggestion benchmark that delivers structured supervision for visual grounding, contextual reasoning, and action recommendation. To effectively address this problem, we propose a progressive three-stage training pipeline for this task. First, Entity Presence Grounding Pretraining reinforces the association between visual

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

entities and structured driving evidence. Second, we perform driving recommendation generation through cascade reasoning. Our proposed Chain-of-Insight Cascade Reasoning (CCR) propagates contextual information sequentially from Reason to Evidence and then to Suggestion; a reason-aware contrastive learning objective is incorporated to sharpen the discriminative capacity of contextual reasoning. Third, we introduce Driving Policy Refinement to optimize the quality of each individual output branch. Leveraging branch-wise Group Relative Policy Optimization, we separately optimize the Evidence, Reason, and Suggestion branches with task-specific reward functions. This staged optimization allows the model to generate concise driving recommendations that are explicitly grounded in pertinent visual evidence and contextual reasoning. With the above steps, we finally build up our CEDAR, a framework for **C**oncise **E**vidence-grounded **D**riving **A**ction **R**ecommendation Generation via Cascade Reasoning and Policy Refinement.

To systematically validate our formulation, we evaluate the performance on our constructed CODA-GRD and three additional driving benchmarks, and the results well verify the superiority of our method on concise and evidence-grounded driving action recommendation. In summary, our contributions are three-fold:

- We identify the limitations of prevailing response-centric driving generation and formalize concise, evidence-grounded driving recommendation generation. Built on this formulation, we develop a CEDAR framework that explicitly links driving rationales, decision-relevant visual evidence, and actionable recommendations.
- We propose Chain-of-Insight Cascade Reasoning (CCR) along with a progressive three-stage training pipeline. CCR enables sequential propagation of contextual information across Reason, Evidence and Suggestion. Coupled with evidence-aware pretraining, reason-aware contrastive learning and branch-wise policy refinement, it jointly boosts visual grounding, reasoning and decision generation.
- We conduct extensive experiments on our proposed benchmark CODA-GRD plus three external driving benchmarks: OmniDrive, LingoQA, and DriveCoT. Experimental results consistently verify the superiority of our framework on visual grounding, contextual reasoning, and driving recommendation generation.

2 Related Work

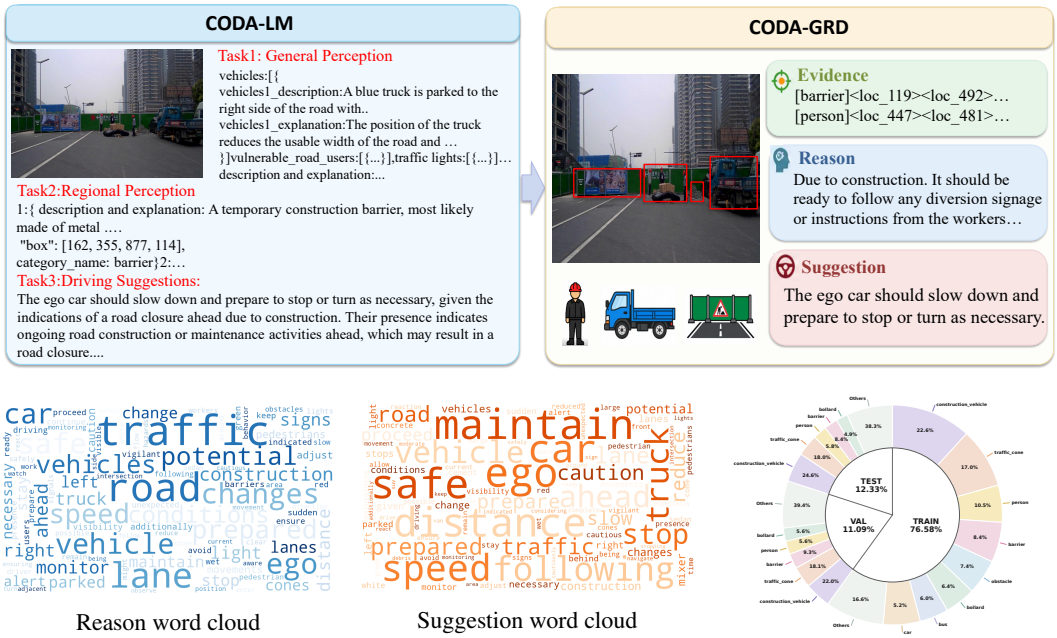
Vision-Language Models for Autonomous Driving. Recent language-aware driving methods mainly focus on scene-level reasoning (Sima et al., 2024; Nie et al., 2024) or action-oriented planning (Shao et al., 2024a; Pan et al., 2024; Tian et al., 2025), with recent efforts enhancing spatial reasoning and VLA alignment (Wang et al., 2025; Renz et al., 2025). However, these paradigms prioritize scene understanding or direct planning, often yielding verbose reasoning traces while leaving decision-relevant visual evidence under-exposed. In contrast, we explicitly structure Evidence, Reason, and Suggestion to generate concise, evidence-grounded driving recommendations.

Driving Reasoning and Policy Optimization. Supervised reasoning has empowered vision-language models for autonomous driving through structured chains-of-thought (Nie et al., 2024; Wang et al., 2024b) and spatiotemporal trajectory alignment (Kong et al., 2025; Chen et al., 2025b). However, supervised fine-tuning merely encourages imitation of annotated targets rather than directly optimizing generation quality. While recent works incorporate reinforcement learning via GRPO-based policy optimization (Jiang et al., 2025), our framework performs branch-wise policy refinement over Evidence, Reason, and Suggestion with task-specific rewards, jointly optimizing visual grounding, contextual reasoning, and concise recommendation generation.

3 Methodology

3.1 Data Preparation

To support evidence-grounded reasoning and concise driving action recommendation, we reorganize the annotations of CODA-LM (Chen et al., 2025a) into three complementary supervision signals: Evidence, Reason, and Suggestion, as shown in Figure 1.



Reason word cloud

Suggestion word cloud

Figure 1: Overview of CODA-GRD data preparation and statistics. The top panel illustrates how a CODA-LM sample is converted into CODA-GRD, while the three bottom panels show the Reason and Suggestion word clouds and the Evidence category distribution.

Data Source. CODA-LM (Chen et al., 2025a) provides hierarchical multimodal annotations for real-world road corner cases, covering three tasks: General Perception, Regional Perception, and Driving Suggestions. General Perception describes critical traffic entities and their potential impact on the ego vehicle, while Regional Perception provides object-level annotations with category names, bounding boxes, and driving-related descriptions. Driving Suggestions provide action-oriented guidance with contextual and causal explanations.

Suggestion Extraction. To obtain concise and actionable decision supervision, we separate the direct action-oriented content from the contextual and causal information in the original Driving Suggestions and retain the former as concise suggestion.

Reason Construction. After extracting the concise Suggestion, the remaining contextual and causal information in the original Driving Suggestions is further organized as Reason. Unlike Suggestion, which directly specifies the expected driving behavior, Reason preserves the explanatory information associated with the current traffic context and describes why the corresponding driving action is required, thereby providing reasoning supervision for driving decision learning.

Evidence Construction. To identify the essential elements required to verify the driving suggestion and the reasoning behind it, we construct structured Evidence from the object-level Regional Perception annotations. For each regional object, we use its annotated semantic category c_i and bounding box $b_i = [x_i, y_i, w_i, h_i]$. The box is converted into corner coordinates as

$$(x_1^i, y_1^i, x_2^i, y_2^i) = (x_i, y_i, x_i + w_i, y_i + h_i), \tag{1}$$

and further normalized and discretized according to the image resolution to obtain location tokens.

Some decision-relevant pedestrian entities described in General Perception do not have corresponding spatial annotations in Regional Perception. To improve their spatial coverage, we employ Qwen2.5-VL-3B-Instruct (Bai et al., 2025b) to supplement missing bounding boxes and encode them using the same semantic-spatial representation. The i-th entity is represented as:

$$e_i = [c_i]\langle loc_{\hat{x}_1^i}\rangle\langle loc_{\hat{y}_1^i}\rangle\langle loc_{\hat{x}_2^i}\rangle\langle loc_{\hat{y}_2^i}\rangle, \tag{2}$$

where $\hat{x}_1^i, \hat{y}_1^i, \hat{x}_2^i, \hat{y}_2^i$ denote the discretized corner coordinates. All entity representations are concatenated to form the final Evidence sequence.

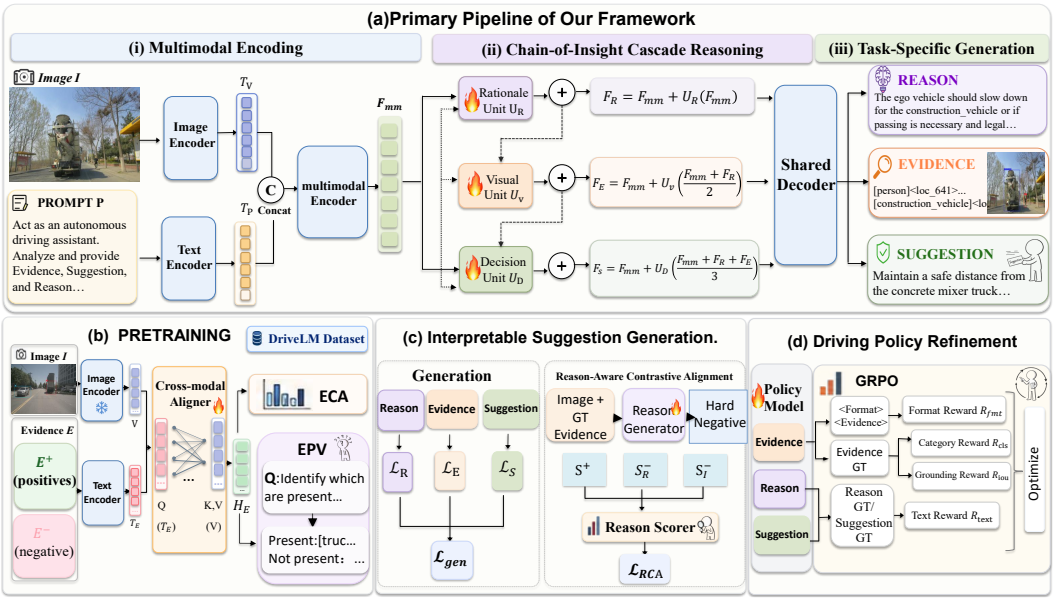


Figure 2: Overview of our framework and three-stage training paradigm. (a) CEDAR employs CCR for latent Reason–Evidence–Suggestion refinement. (b) Entity Presence Grounding Pretraining improves Evidence–vision alignment via ECA and EPV. (c) Cascade Reasoning performs Reason–Evidence–Suggestion generation with RCA. (d) Driving Policy Refinement uses branch-wise GRPO with task-specific rewards.

3.2 CEDAR Framework

Overview. As illustrated in Fig. 2, our CEDAR framework adopts a multimodal encoder-language-decoder architecture with Chain-of-Insight Cascade Reasoning. Given a driving image I and task prompt P, the Vision Encoder extracts visual tokens, which are combined with the prompt embeddings and processed by the Multimodal Encoder to obtain the shared representation F_{mm}. F_{mm} is subsequently refined by the Chain-of-Insight Cascade Reasoning module, and the resulting task-specific representations are decoded by the shared Language Decoder into Reason, Evidence, and Suggestion. The framework is optimized in three stages: Entity Presence Grounding Pretraining, Driving Suggestion Generation via Cascade Reasoning, and Driving Policy Refinement.

3.2.1 Entity Presence Grounding Pretraining

Before delving into the evidence-grounded suggestion generation, we first strengthen the semantic-spatial grounding between structured Evidence and visual observations. Sequence-level supervision alone could encourage the model to fit output correlations without explicitly establishing whether the corresponding traffic entities and locations are supported by the visual scene. We therefore introduce a pretraining stage to establish reliable Evidence-vision correspondence before higher-level reasoning and decision learning. For this stage, we construct additional training samples from the front-view images and object annotations of DriveLM (Sima et al., 2024).

During this stage, the Vision Encoder remains frozen, and we optimize the LoRA parameters within both the Multimodal Encoder and Language Decoder. To establish Evidence–vision correspondence, we introduce a lightweight auxiliary cross-modal alignment module for pretraining. Given an image I and an Evidence sequence E, the frozen Vision Encoder extracts visual features V, whereas E is embedded and processed by the Multimodal Encoder to yield the Evidence representation T_E. The cross-modal representation is computed as

$$H_E = \mathcal{A}_{\text{cm}}(T_E, V, V), \tag{3}$$

where $\mathcal{A}_{\text{cm}}$ denotes the auxiliary cross-modal aligner.

Evidence–Image Contrastive Alignment. To strengthen the correspondence between Evidence and visual content, we align the Evidence representation T_E with the cross-modal representation H_E in a shared embedding space. Specifically, T_E and H_E are mean-pooled and mapped by two

independent projection heads, $\mathcal{P}_t$ and $\mathcal{P}_v$, respectively. Their similarity is then optimized using a sigmoid-based contrastive objective:

$$\mathcal{L}_{\text{ECA}} = -\frac{1}{|\mathcal{Q}|} \sum_{(E,I,y)\in\mathcal{Q}} \log \sigma \left[y \left(\alpha \cos \left(\mathcal{P}_t(\text{Pool}(T_E)), \mathcal{P}_v(\text{Pool}(H_E)) \right) + b \right) \right], \tag{4}$$

where $\mathcal{Q}$ is the set of image-evidence pairs with match label $y \in \{+1, -1\}$, $\mathcal{P}_t$ and $\mathcal{P}_v$ are the textual and cross-modal projection heads, and α, b are learnable scale and bias parameters. Note that the cross-modal aligner and projection heads are only active during evidence-aware pretraining, whereas the learned LoRA parameters are retained for subsequent stages.

Entity-Level Presence Verification. To impose fine-grained supervision over traffic entities, we further introduce an entity-level presence verification task. For each driving image, we combine annotated Evidence entities with randomly sampled negative traffic candidates and shuffle the combined pool to build a verification query Q:

"Identify which of the following are present in the driving scene: [candidates]."

The model is trained to generate responses following the template:

"Present: [positive entities]. Not present: [negative entities]."

where each positive entity retains its full category and location-token representation. We optimize the model with the language modeling loss

$$\mathcal{L}_{\text{EPV}} = -\frac{1}{L} \sum_{t=1}^{L} \log p_\theta(a_t \mid a_{<t}, I, Q), \tag{5}$$

where $A = \{a_t\}_{t=1}^{L}$ denotes the target response sequence.

Combining the two pretraining tasks, our overall pretraining objective reads

$$\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{ECA}} + \mathcal{L}_{\text{EPV}}. \tag{6}$$

3.2.2 Driving Suggestion Generation via Cascade Reasoning

Chain-of-Insight Cascade Reasoning. Driving reasoning, evidence identification, and action generation are inherently interdependent: identifying relevant traffic elements relies on the current driving context, while reliable driving decisions should be supported by both reasoning and scene evidence. To model such dependencies, we introduce Chain-of-Insight Cascade Reasoning (CCR), which comprises three functional units: the Rationale Unit $\mathcal{U}_R$ for reasoning clue learning, the Visual Unit $\mathcal{U}_V$ to capture regional clues, and the Decision Unit $\mathcal{U}_D$ for suggestion-oriented feature aggregation. All three units are implemented via lightweight Transformer adapters.

To establish the contextual prior required for evidence modeling, the Rationale Unit first adapts the shared multimodal representation F_{mm} to the current driving context, producing the reasoning-aware representation F_R. The Visual Unit integrates F_{mm} with the reasoning-aware context F_R and further refines the fused features to obtain the Evidence representation F_E. Finally, driving suggestion generation requires both contextual reasoning and corresponding traffic evidence. The Decision Unit therefore integrates F_{mm}, F_R, and F_E and further refines them into the decision-oriented representation F_S:

$$\begin{aligned} F_R &= F_{mm} + \mathcal{U}_R(F_{mm}), \qquad F_E = F_{mm} + \mathcal{U}_V\left(\frac{F_{mm} + F_R}{2}\right), \\ F_S &= F_{mm} + \mathcal{U}_D\left(\frac{F_{mm} + F_R + F_E}{3}\right). \end{aligned} \tag{7}$$

The resulting F_R, F_E, and F_S are separately fed into the shared Language Decoder to independently generate Reason, Evidence, and Suggestion, respectively. Therefore, the CCR progression represents progressive refinement within the latent space rather than sequential textual generation.

Interpretable Suggestion Generation. To explicitly align each unit with its targeted semantic context, we introduce task-specific supervision during training. Specifically, Evidence, Reason, and Suggestion are optimized with their respective language modeling losses $\mathcal{L}_E$, $\mathcal{L}_R$, and $\mathcal{L}_S$. The joint generation objective is

$$\mathcal{L}_{\text{gen}} = \mathcal{L}_E + \mathcal{L}_R + \mathcal{L}_S. \tag{8}$$

Reason-Aware Contrastive Alignment. To complement autoregressive supervision for Reason generation, we introduce reason-aware contrastive alignment with generated hard negatives. Specifically, an auxiliary reason generator $\mathcal{G}_R$ is trained on driving images and ground-truth evidence to synthesize hard-negative reasons. We then generate candidate reasons and select the most informative hard negative using a similarity-based strategy. For each sample i, the generator produces a candidate set $\tilde{\mathcal{R}}_i$. We filter candidates whose similarity ρ to the ground truth falls outside a pre-defined interval, yielding a refined set $\mathcal{H}_i$. If $\mathcal{H}_i$ is non-empty, a hard negative R_i^- is randomly drawn from it; otherwise, a reason from another sample is selected. Formally, the similarity is formulated as:

$$\rho(\tilde{R}, R^{gt}) = 0.4 \times \text{BLEU}(\tilde{R}, R^{gt}) + 0.6 \times \text{ROUGE-L}(\tilde{R}, R^{gt}). \tag{9}$$

For each sample i, the hard-negative construction is formulated as

$$\begin{cases} \tilde{\mathcal{R}}_i = \mathcal{G}_R(I_i, E_i^{gt}), \\ \mathcal{H}_i = \left\{ \tilde{R}_{ij} \in \tilde{\mathcal{R}}_i \mid 0.30 \le \rho(\tilde{R}_{ij}, R_i^{gt}) \le 0.65 \right\}, \\ R_i^- = \begin{cases} \text{Rand}(\mathcal{H}_i), & |\mathcal{H}_i| > 0, \\ \text{Rand}\left(\{R_j^{gt}\}_{j \neq i}\right), & |\mathcal{H}_i| = 0, \end{cases} \end{cases} \tag{10}$$

where $\text{Rand}(\cdot)$ denotes random sampling operation, $\tilde{R}_{ij}$ is the j-th candidate in $\tilde{\mathcal{R}}_i$.

A lightweight Reason scorer computes cosine similarity between visual and Reason representations. We denote the matched-pair similarity as s^+, the current-image–negative-Reason similarity as s_R^-, and the mismatched-image–Reason similarity as s_I^-. The contrastive objective is

$$\mathcal{L}_{\text{RCA}} = \max\{m - s^+ + s_R^-, 0\} + \max\{m - s^+ + s_I^-, 0\}, \tag{11}$$

where m denotes the ranking margin. The overall training objective is

$$\mathcal{L}_{\text{ISG}} = \mathcal{L}_{\text{gen}} + \lambda_R \mathcal{L}_{\text{RCA}}, \tag{12}$$

where λ_R balances generation and Reason-aware contrastive supervision.

3.2.3 Driving Policy Refinement

To mitigate SFT exposure bias and encourage policy exploration, we adopt Group Relative Policy Optimization (GRPO) (Shao et al., 2024b) with a four-component reward function tailored to our task:

Format Reward (R_{fmt}) enforces adherence to the structured Evidence format defined in Sec. 3.1. Let $\hat{E}$ denote the generated Evidence, $\mathcal{S}_E$ the set of valid Evidence patterns, and $\mathbb{I}(\cdot)$ the indicator function: $R_{\text{fmt}} = \mathbb{I}(\hat{E} \in \mathcal{S}_E)$.

Category Reward (R_{cls}) measures the category agreement between generated and ground-truth Evidence while accounting for the number of instances in each category. Let $\hat{n}_c$ and n_c denote the predicted and ground-truth numbers of instances for category c, respectively. The category reward is defined as the F1 score computed from these category-wise instance counts:

$$R_{\text{cls}} = \text{F1}(\hat{\mathbf{n}}, \mathbf{n}). \tag{13}$$

Grounding Reward (R_{iou}) evaluates spatial precision through category-consistent one-to-one IoU matching. Let $\mathcal{M}$ denote the matched box pairs and N_{gt} the number of ground-truth entities:

$$R_{\text{iou}} = \frac{1}{N_{\text{gt}}} \sum_{(\hat{b}, b) \in \mathcal{M}} \text{IoU}(\hat{b}, b). \tag{14}$$

Table 1: Comparison of zero-shot MLLMs and fine-tuned LVLMs on CODA-GRD. Evidence grounding is evaluated by F1 and IoU, while Suggestion and Reason are evaluated using BLEU-1 (B-1), BLEU-4 (B-4), METEOR (M), ROUGE-L (R-L), and CIDEr (C). AVG denotes the average across all 12 metrics. The best results are in bold, and the second-best are underlined.

Setting	Method	Evidence		Suggestion					Reason					AVG
		F1	IoU	B-1	B-4	M	R-L	C	B-1	B-4	M	R-L	C	
Zero-Shot	Qwen2.5-VL-72B (Bai et al., 2025b)	10.27	74.79	12.28	3.42	15.35	18.96	0.13	7.85	0.63	10.45	14.18	0.01	14.03
	Qwen3-VL-235B (Bai et al., 2025a)	11.33	82.92	18.43	3.20	18.56	21.16	0.15	17.33	1.16	15.91	15.08	0.02	17.10
	LLaVA-v1.6-13B (Liu et al., 2024)	0.14	62.78	2.28	0.35	2.49	2.82	0.01	2.24	0.15	2.50	2.83	0.00	6.55
	ERNIE-4.5-VL (Baidu ERNIE Team, 2025)	3.26	67.15	20.65	5.31	22.66	23.92	0.18	18.11	1.50	16.90	15.53	0.03	16.27
	Gemini-2.5-Flash (Comanici et al., 2025)	0.93	64.40	16.48	3.48	18.83	19.95	0.13	11.60	0.93	12.76	13.92	0.01	13.62
	Qwen3-VL-Plus (Bai et al., 2025a)	0.42	69.82	20.89	3.25	20.54	21.64	0.16	18.99	1.27	16.77	14.82	0.03	15.72
	Qwen2.5-VL-Max (Bai et al., 2025b)	10.77	75.83	19.75	5.56	21.85	23.62	0.19	14.91	1.10	14.37	14.99	0.02	16.91
	GPT-4o (Hurst et al., 2024)	0.53	59.34	15.41	4.19	17.48	20.95	0.13	9.74	0.80	11.84	13.35	0.01	12.81
Fine-Tuned	Qwen2.5-VL-3B (Bai et al., 2025b)	55.31	86.43	28.22	10.28	34.23	30.34	0.28	17.64	2.79	18.35	19.58	0.04	25.29
	Qwen2-VL-2B (Wang et al., 2024a)	56.67	84.84	26.41	9.95	33.04	28.86	0.32	19.00	3.20	19.28	19.69	0.05	25.11
	Qwen2-VL-7B (Wang et al., 2024a)	58.24	86.20	28.84	10.76	**36.68**	30.13	0.25	15.30	2.46	17.15	18.59	0.03	25.39
	InternVL2.5-4B (Chen et al., 2024)	38.81	78.81	28.03	10.56	35.28	30.28	0.28	16.93	2.69	17.84	19.67	0.04	23.27
	MiniCPM-V-2.6 (Yao et al., 2024)	38.23	77.40	27.33	11.10	33.24	30.82	0.37	23.38	4.10	22.90	21.39	0.06	24.19
	LLaVA-OneVision-0.5B (Li et al., 2025)	24.48	69.85	27.76	10.40	33.69	29.11	0.30	17.12	2.41	17.94	17.88	0.04	20.92
	Phi-3.5-Vision (Abdin et al., 2024)	24.33	75.60	27.37	11.44	32.57	31.33	0.39	12.51	2.16	22.32	14.82	0.01	21.24
	Baseline	55.56	89.71	27.00	10.90	32.91	31.02	0.38	21.60	3.57	21.30	20.99	**0.07**	26.25
	CEDAR-0.89B	**61.63**	**89.86**	**30.63**	**12.05**	35.85	**34.06**	**0.45**	**30.86**	**7.14**	**31.78**	**23.49**	0.06	**29.82**

Text Reward (R_{text}) evaluates the generated Reason and Suggestion using ROUGE-L. Given a generated text $\hat{Y}$ and its ground-truth annotation Y^{gt}: $R_{\text{text}}(\hat{Y}, Y^{gt}) = \text{ROUGE-L}(\hat{Y}, Y^{gt})$, where $Y \in \{R, S\}$.

Optimization Objective. For each input prompt x and branch $k \in \{E, R, S\}$, we sample a group of G outputs $\{o_1^k, o_2^k, \ldots, o_G^k\}$ from the current policy π_θ. The reward for the i-th output is defined according to the corresponding branch:

$$R_k^{(i)} = \begin{cases} \alpha_{\text{fmt}} R_{\text{fmt}}^{(i)} + \alpha_{\text{cls}} R_{\text{cls}}^{(i)} + \alpha_{\text{iou}} R_{\text{iou}}^{(i)}, & k = E, \\ R_{\text{text}}(o_i^k, Y_k^{gt}), & k \in \{R, S\}. \end{cases} \tag{15}$$

To compute the advantage A_i^k, we normalize the rewards independently within each branch group. Let$r_i^k = \pi_\theta(o_i^k \mid x, F_k)/\pi_{\text{old}}(o_i^k \mid x, F_k)$ denote the probability ratio. The surrogate objective for the i-th sample is formulated as

$$J_{\text{clip}}^{k,i} = \min\left(r_i^k A_i^k,\ \text{clip}(r_i^k, 1-\epsilon, 1+\epsilon) A_i^k\right). \tag{16}$$

The final GRPO objective averages the surrogate objectives while applying a KL-divergence penalty to constrain policy deviation:

$$\mathcal{L}_{\text{RL}} = -\sum_{k \in \{E,R,S\}} \omega_k \mathbb{E}\left[\frac{1}{G}\sum_{i=1}^{G} J_{\text{clip}}^{k,i} - \beta D_{\text{KL}}\left(\pi_\theta(\cdot \mid x, F_k) \| \pi_{\text{ref}}(\cdot \mid x, F_k)\right)\right]. \tag{17}$$

4 Experiments

Implementation details, training settings, and evaluation metrics are provided in the Appendix.

4.1 Quantitative Results

Performance on **CODA-GRD** is shown in Tab. 1. CEDAR achieves the highest AVG of 29.82 among the methods, surpassing the baseline (26.25) by 13.60%. It also outperforms Qwen2-VL-7B (25.39) and Qwen2.5-VL-3B (25.29). The margin is more pronounced over zero-shot MLLMs, among which Qwen3-VL-235B achieves the best AVG of 17.10. With only 0.89B parameters, our model attains the best overall AVG while maintaining competitive performance. Performance on

Table 2: Comparison of zero-shot MLLMs and fine-tuned LVLMs on LingoQA.

Setting	Method	Suggestion					Reason					AVG
		B-1	B-4	M	R-L	C	B-1	B-4	M	R-L	C	
Zero-Shot	Qwen2.5-VL-72B (Bai et al., 2025b)	4.72	0.86	3.28	5.96	0.03	19.92	2.50	19.88	20.89	0.10	7.81
	Qwen3-VL-235B (Bai et al., 2025a)	7.59	1.26	7.19	10.41	0.04	12.88	1.42	21.11	16.41	0.00	7.83
	ERNIE-4.5-VL (Baidu ERNIE Team, 2025)	7.46	1.22	6.61	9.81	0.04	15.40	2.05	24.92	18.93	0.01	8.64
	Gemini-2.5-Flash (Comanici et al., 2025)	7.03	1.20	7.02	11.14	0.04	19.80	2.74	23.77	21.50	0.06	9.43
	Qwen3-VL-Plus (Bai et al., 2025a)	8.33	1.16	9.12	10.44	0.03	11.90	1.32	22.38	15.62	0.00	8.03
	Qwen2.5-VL-Max (Bai et al., 2025b)	8.78	1.48	8.31	12.10	0.04	14.35	1.62	21.32	17.35	0.01	8.54
	GPT-4o (Hurst et al., 2024)	7.91	1.38	6.50	10.34	0.03	19.73	2.27	21.44	21.00	0.08	9.07
Fine-Tuned	Qwen2.5-VL-3B (Bai et al., 2025b)	75.71	62.48	78.54	79.96	5.21	48.63	33.56	50.06	49.89	2.56	48.66
	Qwen2-VL-2B (Wang et al., 2024a)	76.11	63.06	79.10	80.18	5.29	50.08	34.01	51.85	50.89	2.49	49.31
	Qwen2-VL-7B (Wang et al., 2024a)	72.43	57.70	74.79	77.42	4.74	47.46	32.64	48.54	50.23	2.52	46.85
	InternVL2.5-4B (Chen et al., 2024)	75.26	61.58	78.48	79.18	5.07	41.77	24.63	43.20	43.04	1.61	45.38
	MiniCPM-V-2.6 (Yao et al., 2024)	72.75	57.68	75.92	77.23	4.61	49.33	33.71	51.04	51.09	2.55	47.59
	LLaVA-OneVision-0.5B (Li et al., 2025)	75.66	62.47	79.73	79.70	5.23	43.54	25.75	45.21	44.79	1.73	46.38
	Phi-3.5-Vision (Abdin et al., 2024)	60.23	47.74	65.77	69.13	3.62	37.17	20.73	41.27	40.12	1.07	38.68
	Baseline	72.93	55.77	76.33	77.76	4.13	48.57	31.50	50.55	49.47	2.28	46.93
	CEDAR-0.89B	**76.85**	**64.40**	**80.89**	**80.68**	**5.53**	**52.98**	**37.57**	**55.65**	**54.68**	**2.84**	**51.21**

Table 3: Comparison of zero-shot MLLMs and fine-tuned LVLMs on DriveCoT.

Setting	Method	Suggestion					Reason					AVG
		B-1	B-4	M	R-L	C	B-1	B-4	M	R-L	C	
Zero-Shot	Qwen2.5-VL-72B (Bai et al., 2025b)	10.12	1.54	7.45	11.30	0.06	11.65	1.20	13.43	14.16	0.01	7.09
	Qwen3-VL-235B (Bai et al., 2025a)	9.08	1.24	8.49	11.17	0.04	7.51	0.60	15.58	11.09	0.00	6.48
	ERNIE-4.5-VL (Baidu ERNIE Team, 2025)	9.77	1.46	7.61	11.57	0.04	9.33	0.85	18.67	13.28	0.00	7.26
	Gemini-2.5-Flash (Comanici et al., 2025)	9.72	1.43	9.58	13.45	0.05	11.38	1.05	15.86	14.17	0.01	7.67
	Qwen3-VL-Plus (Bai et al., 2025a)	7.78	1.02	8.35	9.86	0.03	7.01	0.53	15.01	10.05	0.00	5.96
	Qwen2.5-VL-Max (Bai et al., 2025b)	9.02	1.31	7.85	10.71	0.05	8.31	0.74	16.26	12.42	0.00	6.67
	GPT-4o (Hurst et al., 2024)	9.63	1.42	7.79	11.29	0.06	12.72	1.14	15.65	14.46	0.02	7.42
Fine-Tuned	Qwen2.5-VL-3B (Bai et al., 2025b)	55.46	33.05	55.14	60.29	1.94	51.66	29.88	55.36	56.87	2.63	40.23
	Qwen2-VL-2B (Wang et al., 2024a)	55.37	34.37	53.99	59.81	2.11	52.89	31.45	54.30	57.39	2.78	40.45
	Qwen2-VL-7B (Wang et al., 2024a)	57.27	34.01	57.43	61.31	2.11	49.70	26.75	52.75	55.94	2.44	39.97
	InternVL2.5-4B (Chen et al., 2024)	52.33	31.38	52.95	58.95	1.77	46.18	26.33	46.63	51.09	2.29	36.99
	MiniCPM-V-2.6 (Yao et al., 2024)	56.30	34.76	57.08	61.30	**2.20**	50.55	29.16	51.64	55.59	2.63	40.12
	LLaVA-OneVision-0.5B (Li et al., 2025)	57.40	34.74	56.29	61.22	2.19	52.50	30.85	53.01	57.35	2.66	40.82
	Phi-3.5-Vision (Abdin et al., 2024)	42.39	25.22	42.45	49.21	1.38	31.91	14.43	38.03	40.27	0.94	28.62
	Baseline	54.96	32.58	53.55	59.17	1.97	53.22	30.91	55.19	59.24	2.76	40.36
	CEDAR-0.89B	**58.40**	**35.63**	**59.73**	**63.90**	2.18	**58.13**	**34.02**	**56.94**	**61.80**	**3.08**	**43.38**

LingoQA (Marcu et al., 2024) is compared in Tab. 2, which shows that zero-shot MLLMs achieve an AVG of at most 9.43, while the best fine-tuned LVLM, Qwen2-VL-2B, reaches 49.31. Our method further improves the AVG to 51.21, surpassing Qwen2-VL-2B and the baseline by 3.85% and 9.12%, respectively, with consistent gains across Suggestion and Reason metrics. On **DriveCoT** (Wang et al., 2024b) (Tab. 3), Gemini-2.5-Flash and LLaVA-OneVision-0.5B achieve the best zero-shot (7.67) and fine-tuned (40.82) AVG, respectively. Our method reaches 43.38, improving over LLaVA-OneVision-0.5B and the baseline by 6.27% and 7.48%, with the best Reason results.

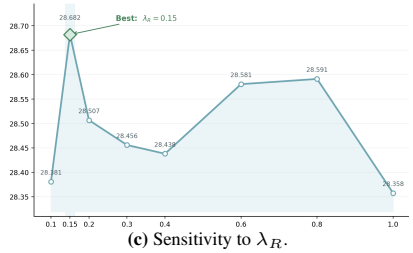
4.2 Ablation Studies and Discussion

We conduct comprehensive ablation studies on CODA-GRD, as summarized in Tab. 4. **Impact of Component Modules.** Tab. 4(a) evaluates each component. CCR improves Overall from 26.25 to 28.22 and Reason from 13.51 to 17.03. RCA and Entity Presence Grounding Pretraining further increase Overall to 28.68, while RL-based policy refinement achieves the best Overall of 29.82 with an Evidence score of 75.75. **Impact of CCR Depth.** Tab. 4(b) shows that the $1/2/2$ configuration achieves the best Overall of 28.68, whereas $2/2/2$ reduces it to 28.00, validating the effectiveness of asymmetric depth allocation. **Sensitivity to the Contrastive Weight.** Tab. 4(c) evaluates λ_R. Overall remains stable within 28.36–28.68, with the best result at $\lambda_R = 0.15$. **Impact of Cascade Architecture.** Tab. 4(d) shows that cascade improves Overall from 27.63 to 28.22, mainly benefiting Reason (15.74→17.03). **Impact of Entity Presence Grounding Pretraining.** Tab. 4(e) verifies

Table 4: Ablation of components (a), CCR depth (b), λ_R sensitivity (c), cascade architecture (d), and entity presence grounding pretraining (e) on CODA-GRD. Evidence, Suggestion, and Reason are task-wise averages, while Overall averages all 12 metrics.

Components				Task-wise Average			
CCR	RCA	Pre.	RL	Evidence	Suggestion	Reason	Overall
				72.64	20.44	13.51	26.25
✓				73.44	21.32	17.03	28.22
✓	✓			73.43	21.16	17.24	28.24
✓	✓	✓		72.42	21.54	18.33	28.68
✓	✓	✓	✓	**75.75**	**22.61**	**18.67**	**29.82**

(a) Component ablation.



(c) Sensitivity to λ_R.

Depth			Task-wise Average			
E	R	S	Evidence	Suggestion	Reason	Overall
1	1	1	72.67	20.89	18.21	28.40
1	1	2	73.20	21.30	18.12	28.62
1	2	1	**73.24**	21.13	17.41	28.26
1	2	2	72.42	**21.54**	**18.33**	**28.68**
1	2	3	73.16	21.14	18.16	28.57
2	2	2	72.90	20.94	17.10	28.00

(b) CCR depth configuration.

Design	Evidence	Suggestion	Reason	Overall
Parallel	73.08	**21.33**	15.74	27.63
Cascade	**73.44**	21.32	**17.03**	**28.22**

(d) Parallel vs. cascade.

Setting	Evidence	Suggestion	Reason	Overall
SFT	**73.43**	21.16	17.24	28.24
SFT + ECA	73.03	**21.65**	17.08	28.30
SFT + ECA + EPV	72.42	21.54	**18.33**	**28.68**

(e) Entity presence grounding pretraining.

Table 5: DriveLM-based pretraining (a) and latent refinement order (b).

Method	DriveLM Pre.	Evidence	Suggestion	Reason	Overall
Qwen2-VL-7B	✓	71.90	21.01	11.86	25.68
Qwen2.5-VL-3B	✓	**72.52**	20.11	12.04	25.48
InternVL2.5-4B	✓	57.81	20.74	11.77	23.18
CEDAR (w/o RL)	✓	72.42	**21.54**	**18.33**	**28.68**

(a) DriveLM pretraining.

Order	Evidence	Suggestion	Reason	Overall
E-R-S	72.84	21.14	**17.35**	28.18
E-S-R	72.89	14.48	7.40	21.27
S-E-R	**73.81**	21.30	16.82	28.19
R-E-S	73.44	**21.32**	17.03	**28.22**

(b) Latent refinement order.

ECA and EPV. ECA improves Overall from 28.24 to 28.30, and adding EPV further raises it to 28.68 with Reason up from 17.24 to 18.33.

Impact of DriveLM Pretraining. Tab. 5(a) examines the influence of additional DriveLM supervision. With the same additional data, the pretrained LVLMs achieve AVG scores ranging from 23.18 to 25.68, whereas CEDAR without RL reaches 28.68, suggesting that the improvement cannot be attributed solely to additional training data. **Impact of Latent Refinement Order.** Tab. 5(b) evaluates different latent refinement orders. The proposed R-E-S order achieves the highest AVG of 28.22, marginally surpassing the other competitive orders. **Efficiency and Capability Analysis.** Fig. 3 compares efficiency and task performance across representative LVLMs. CEDAR achieves the highest Overall AVG of 29.82 with only 893.6M parameters, 3.63T FLOPs, and 2.45 GiB peak GPU memory, demonstrating a favorable performance–efficiency trade-off.

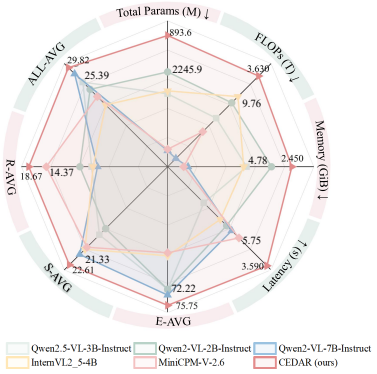


Figure 3: Efficiency–performance. Lower-is-better efficiency metrics are inverted and normalized for visualization.

5 Conclusion

We formulate the task of concise, evidence-grounded driving recommendation generation, connecting decisions with contextual reasoning and decision-relevant visual evidence. We introduce CODA-GRD, a unified Evidence–Reason–Suggestion benchmark for grounding, reasoning, and action recommendation, and propose CEDAR, integrating Entity Presence Grounding Pretraining, Chain-of-Insight Cascade Reasoning with Reason-Aware Contrastive Alignment, and branch-wise Driving Policy Refinement to progressively improve all three capabilities. Experiments on CODA-GRD and three external driving benchmarks show consistent gains over the task-specific baselines, demonstrating CEDAR's effectiveness for concise, evidence-grounded driving decision support.

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish the writing, including grammar correction, sentence refinement, and improving clarity and readability. We have reviewed all AI-assisted content and take full responsibility for the final manuscript.

REFERENCES

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, et al. Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone, 2024. URL https://arxiv.org/abs/2404.14219.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-VL Technical Report, 2025a. URL https://arxiv.org/abs/2511.21631.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-VL Technical Report, 2025b. URL https://arxiv.org/abs/2502.13923.

Baidu ERNIE Team. ERNIE 4.5 Technical Report, 2025. URL https://ernie.baidu.com/blog/publication/ERNIE_Technical_Report.pdf.

Satanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Jade Goldstein, Alon Lavie, Chin-Yew Lin, and Clare Voss (eds.), *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pp. 65–72, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics. URL https://aclanthology.org/W05-0909/.

Kai Chen, Yanze Li, Wenhua Zhang, Yanxin Liu, Pengxiang Li, Ruiyuan Gao, Lanqing Hong, Meng Tian, Xinhai Zhao, Zhenguo Li, Dit-Yan Yeung, Huchuan Lu, and Xu Jia. Automated Evaluation of Large Vision-Language Models on Self-Driving Corner Cases. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pp. 7806–7815, 2025a.

Xuesong Chen, Linjiang Huang, Tao Ma, Rongyao Fang, Shaoshuai Shi, and Hongsheng Li. SOLVE: Synergy of Language-Vision and End-to-End Networks for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12068–12077, 2025b.

Zhe Chen, Weiyun Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Erfei Cui, Jinguo Zhu, Shenglong Ye, Hao Tian, Zhaoyang Liu, et al. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*, 2024.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities, 2025. URL https://arxiv.org/abs/2507.06261.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. GPT-4o System Card. *arXiv preprint arXiv:2410.21276*, 2024.

Bo Jiang, Shaoyu Chen, Qian Zhang, Wenyu Liu, and Xinggang Wang. AlphaDrive: Unleashing the Power of VLMs in Autonomous Driving via Reinforcement Learning and Reasoning, 2025. URL https://arxiv.org/abs/2503.07608.

Fanjie Kong, Yitong Li, Weihuang Chen, Chen Min, Yizhe Li, Zhiqiang Gao, Haoyang Li, Zhongyu Guo, and Hongbin Sun. VLR-Driver: Large Vision-Language-Reasoning Models for Embodied Autonomous Driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26966–26976, 2025.

Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, and Chunyuan Li. LLaVA-OneVision: Easy Visual Task Transfer. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=zKv8qULV6n.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://aclanthology.org/W04-1013/.

Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. Llava-next: Improved reasoning, ocr, and world knowledge, January 2024. URL https://llava-vl.github.io/blog/2024-01-30-llava-next/.

Ana-Maria Marcu, Long Chen, Jan Hünermann, Alice Karnsund, Benoit Hanotte, Prajwal Chidananda, Saurabh Nair, Vijay Badrinarayanan, Alex Kendall, Jamie Shotton, Elahe Arani, and Oleg Sinavski. LingoQA: Video Question Answering for Autonomous Driving. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. doi: 10.1007/978-3-031-72980-5_15. URL https://mlanthology.org/eccv/2024/marcu2024eccv-lingoqa/.

Ming Nie, Renyuan Peng, Chunwei Wang, Xinyue Cai, Jianhua Han, Hang Xu, and Li Zhang. Reason2Drive: Towards Interpretable and Chain-Based Reasoning for Autonomous Driving. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. doi: 10.1007/978-3-031-73347-5_17. URL https://mlanthology.org/eccv/2024/nie2024eccv-reason2drive/.

Chenbin Pan, Burhaneddin Yaman, Tommaso Nesti, Abhirup Mallik, Alessandro G Allievi, Senem Velipasalar, and Liu Ren. VLP: Vision Language Planning for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14760–14769, 2024.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin (eds.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. URL https://aclanthology.org/P02-1040/.

Katrin Renz, Long Chen, Elahe Arani, and Oleg Sinavski. SimLingo: Vision-Only Closed-Loop Autonomous Driving with Language-Action Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11993–12003, 2025.

Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L. Waslander, Yu Liu, and Hongsheng Li. LMDrive: Closed-Loop End-to-End Driving with Large Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15120–15130, 2024a.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models, 2024b. URL https://arxiv.org/abs/2402.03300.

Chonghao Sima, Katrin Renz, Kashyap Chitta, Li Chen, Hanxue Zhang, Chengen Xie, Jens Beißwenger, Ping Luo, Andreas Geiger, and Hongyang Li. Drivelm: Driving with graph visual question answering. In *European conference on computer vision*, pp. 256–274. Springer, 2024.

Xiaoyu Tian, Junru Gu, Bailin Li, Yicheng Liu, Yang Wang, Zhiyong Zhao, Kun Zhan, Peng Jia, XianPeng Lang, and Hang Zhao. DriveVLM: The Convergence of Autonomous Driving and Large Vision-Language Models. In *Proceedings of the 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 4698–4726. PMLR, 2025. URL https://proceedings.mlr.press/v270/tian25c.html.

Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, et al. Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution, 2024a. URL https://arxiv.org/abs/2409.12191.

Shihao Wang, Zhiding Yu, Xiaohui Jiang, Shiyi Lan, Min Shi, Nadine Chang, Jan Kautz, Ying Li, and Jose M. Alvarez. OmniDrive: A Holistic Vision-Language Dataset for Autonomous Driving with Counterfactual Reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22442–22452, 2025.

Tianqi Wang, Enze Xie, Ruihang Chu, Zhenguo Li, and Ping Luo. DriveCoT: Integrating Chain-of-Thought Reasoning with End-to-End Driving. *arXiv preprint arXiv:2403.16996*, 2024b.

Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, and Lu Yuan. Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4818–4829, 2024.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. MiniCPM-V: A GPT-4V Level MLLM on Your Phone, 2024. URL https://arxiv.org/abs/2408.01800.

6 APPENDIX

6.1 EXPERIMENTAL SETUP

Implementation Details. All experiments are implemented in PyTorch and conducted on NVIDIA GeForce RTX 4090 GPUs. Entity Presence Grounding Pretraining and Driving Suggestion Generation via Cascade Reasoning are performed on a single GPU, while Driving Policy Refinement is conducted on four GPUs using Distributed Data Parallel (DDP). CEDAR is built upon the Florence-2-large-ft vision-language model (Xiao et al., 2024), whose Vision Encoder, Multimodal Encoder, and Language Decoder constitute the shared backbone.

For Evidence–Image Contrastive Alignment (ECA), the auxiliary cross-modal aligner comprises two multi-head attention layers with eight attention heads and a dropout rate of 0.1. The textual and cross-modal projection heads each consist of a linear projection followed by LayerNorm, with an output dimension of 256. Entity-Level Presence Verification (EPV) is implemented as an auxiliary generation task using the shared backbone.

The Chain-of-Insight Cascade Reasoning (CCR) module contains three lightweight Transformer units. The Rationale, Visual, and Decision Units comprise 2, 1, and 2 Transformer layers, respectively, with eight attention heads per layer. For Reason-Aware Contrastive Alignment (RCA), the Reason scorer employs two independent projection heads for visual and Reason representations, each consisting of a linear projection followed by LayerNorm with an output dimension of 256.

Parameter Settings. Input images are resized to 768×768 using the default Florence-2 image processor. For Entity Presence Grounding Pretraining and Driving Suggestion Generation via Cascade Reasoning, we set the per-GPU batch size to 2 with 8 gradient accumulation steps and train for up to 30 and 50 epochs, respectively.

For Entity Presence Grounding Pretraining, optimization is performed using AdamW with learning rates of 1×10^{-5} for LoRA parameters and 2×10^{-4} for the cross-modal aligner and projection heads. A weight decay of 0.01 is applied to these parameter groups. The LoRA rank, scaling factor, and dropout rate are set to 16, 32, and 0.1, respectively.

For Interpretable Suggestion Generation (ISG), the LoRA weights learned during Entity Presence Grounding Pretraining are first merged into the Florence-2 backbone before introducing new LoRA modules. We adopt AdamW with learning rates of 5×10^{-5} for LoRA parameters and 2×10^{-4} for the CCR units and Reason scorer, with a weight decay of 0.01. The LoRA rank, scaling factor, and dropout rate are set to 32, 64, and 0.1, respectively. For RCA, the loss weight λ_R and ranking margin m are set to 0.15 and 0.5.

The auxiliary Reason generator is trained for 10 epochs with a per-GPU batch size of 4, 4 gradient accumulation steps, a learning rate of 5×10^{-5}, a weight decay of 0.01, and 500 warm-up steps. For hard-negative construction, three candidate Reasons are sampled at temperatures of 0.7, 1.0, and 1.3 using top-p sampling with $p = 0.95$. Candidates with weighted BLEU/ROUGE-L scores within $[0.30, 0.65]$ are retained.

For Driving Policy Refinement, the LoRA weights learned during the preceding training phases are successively merged into the backbone before introducing new LoRA modules with a rank of 32, a scaling factor of 64, and a dropout rate of 0.1. The CCR units are initialized from Interpretable Suggestion Generation, while the inherited Reason scorer is kept frozen. The model is trained for 2 epochs with a per-GPU batch size of 1 and a learning rate of 1×10^{-5}. For the branch-wise GRPO objective, we set the generation group size to $G = 4$, the KL coefficient to $\beta = 0.02$, and the clipping parameter to $\epsilon = 0.2$. The branch weights for Evidence, Reason, and Suggestion are set to 0.7, 0.1, and 0.2, respectively, while the format, category, and IoU reward weights for Evidence are set to 0.10, 0.25, and 0.65.

We use BF16 precision throughout training. During evaluation, beam-search decoding is adopted with 8 beams for Evidence and 10 beams for both Reason and Suggestion.

Florence-2 Baseline. We adopt Florence-2-large-ft as the task-specific baseline, using the same training data and supervision format as CEDAR while excluding Entity Presence Grounding Pretraining, CCR, RCA, and Driving Policy Refinement. This setting isolates the gains introduced by CEDAR from those of the underlying backbone.

Table 6: Comparison of zero-shot MLLMs and fine-tuned LVLMs on OmniDrive.

Setting	Method	Suggestion					Reason					AVG
		B-1	B-4	M	R-L	C	B-1	B-4	M	R-L	C	
Zero-Shot	Qwen2.5-VL-72B (Bai et al., 2025b)	15.26	1.84	12.57	13.89	0.05	6.36	0.46	9.93	14.33	0.00	7.47
	Qwen3-VL-235B (Bai et al., 2025a)	20.76	2.17	18.07	17.50	0.08	19.32	1.21	15.38	15.47	0.02	11.00
	ERNIE-4.5-VL (Baidu ERNIE Team, 2025)	21.68	2.43	19.70	18.47	0.09	21.59	1.66	17.31	16.92	0.02	11.99
	Gemini-2.5-Flash (Comanici et al., 2025)	19.50	2.39	19.17	18.31	0.09	11.73	0.88	12.79	14.99	0.00	9.99
	Qwen3-VL-Plus (Bai et al., 2025a)	22.84	2.14	20.51	18.19	0.07	23.91	1.39	17.29	15.98	0.03	12.23
	Qwen2.5-VL-Max (Bai et al., 2025b)	20.42	2.22	18.44	17.44	0.08	17.39	1.19	14.76	15.98	0.01	10.79
	GPT-4o (Hurst et al., 2024)	18.62	2.08	16.31	17.71	0.06	8.83	0.60	11.13	13.77	0.00	8.91
Fine-Tuned	Qwen2.5-VL-3B (Bai et al., 2025b)	36.56	12.58	39.51	36.48	0.63	35.47	4.66	28.31	21.49	0.11	21.58
	Qwen2-VL-2B (Wang et al., 2024a)	**38.02**	**14.41**	**41.11**	**39.02**	**0.78**	35.14	5.65	28.97	22.00	0.11	22.52
	Qwen2-VL-7B (Wang et al., 2024a)	34.00	11.38	36.33	34.16	0.66	34.22	4.60	27.27	21.74	0.10	20.45
	InternVL2.5-4B (Chen et al., 2024)	31.61	9.29	32.19	31.23	0.42	34.24	4.80	26.94	22.05	0.11	19.29
	MiniCPM-V-2.6 (Yao et al., 2024)	37.38	13.43	40.85	37.48	0.68	34.86	5.26	28.02	22.05	0.12	22.01
	LLaVA-OneVision-0.5B (Li et al., 2025)	31.37	7.44	32.25	30.36	0.44	34.68	4.33	27.89	20.72	0.11	18.96
	Phi-3.5-Vision (Abdin et al., 2024)	35.46	11.73	36.81	34.74	0.54	28.05	4.44	**31.54**	20.00	0.04	20.34
	Baseline	36.39	13.90	39.62	36.85	0.72	35.08	5.56	28.21	22.54	0.13	21.90
	CEDAR-0.89B	37.77	12.67	39.58	36.66	0.59	**35.96**	**7.35**	29.32	**25.56**	**0.15**	**22.56**

We will release the complete training and evaluation code as well as checkpoints to ensure the reproduction of all results reported in this paper.

6.2 Evaluation Metrics

To comprehensively evaluate our proposed CEDAR, we adopt task-specific metrics for Evidence grounding, Suggestion generation, and Reason generation. The detailed evaluation setup is organized as follows:

- **Evidence Grounding.** To evaluate the accuracy of generated Evidence, we employ F1 and Intersection over Union (IoU). F1 measures the correctness of grounded traffic entities by jointly considering precision and recall, while IoU quantifies the spatial overlap between predicted and ground-truth bounding boxes. These metrics jointly reflect the semantic and spatial grounding capability of the model.
- **Suggestion Generation.** For driving suggestion generation, we employ BLEU-1 and BLEU-4 (Papineni et al., 2002), METEOR (Banerjee & Lavie, 2005), ROUGE-L (Lin, 2004), and CIDEr (Vedantam et al., 2015). These metrics measure the correspondence between generated suggestions and reference annotations from different perspectives, providing a comprehensive evaluation of driving recommendation quality.
- **Reason Generation.** For driving reason generation, we adopt the same text-generation metrics, including BLEU-1, BLEU-4, METEOR, ROUGE-L, and CIDEr. These metrics evaluate the agreement between generated Reasons and ground-truth rationales, reflecting the quality of contextual driving reasoning.
- **Overall Performance.** To provide an overall assessment of the model, we additionally report AVG, which computes the arithmetic mean of task metrics. For CODA-GRD, AVG averages all 12 metrics across Evidence, Suggestion, and Reason; for external benchmarks without Evidence grounding annotations, it averages the 10 text-generation metrics of Suggestion and Reason.

This evaluation framework ensures a systematic and comprehensive assessment of CEDAR across visual grounding, contextual reasoning, and driving recommendation generation.

6.3 Additional Experiments

Evaluation on OmniDrive.

Tab. 6 reports the performance on **OmniDrive** (Wang et al., 2025). Our method achieves the highest AVG of 22.56, outperforming Qwen2-VL-2B (22.52), MiniCPM-V-2.6 (22.01), and the baseline

Table 7: Sensitivity analysis of RES reward weights. E, S, and R denote the weights of Evidence, Suggestion, and Reason rewards, respectively.

E	S	R	Evidence	Suggestion	Reason	Overall
0.1	0.45	0.45	73.36	22.09	16.59	28.34
0.2	0.4	0.4	74.49	21.23	13.91	27.06
0.25	0.25	0.5	**75.85**	21.18	13.63	27.15
0.4	0.4	0.2	75.05	21.78	17.87	29.03
0.5	0.25	0.25	75.38	<u>22.54</u>	<u>18.43</u>	<u>29.63</u>
1.0	0	0	75.06	21.36	18.4	29.07
0.7	**0.2**	**0.1**	<u>75.75</u>	**22.61**	**18.67**	**29.82**

(21.90). Compared with the best zero-shot result of 12.23 from Qwen3-VL-Plus, CEDAR remains effective when transferred to OmniDrive.

Sensitivity to RES Reward Weights.

Tab. 7 presents the sensitivity analysis over the weights of Evidence, Suggestion and Reason rewards in GRPO policy refinement. Different reward weights lead to different trade-offs among Evidence, Suggestion, and Reason. The weighting scheme E:S:R = 0.7:0.2:0.1 achieves the highest Overall score of 29.82 among the evaluated configurations.

Reliability Assessment of Qwen-Assisted Annotations. To assess the reliability of the bounding boxes supplemented by Qwen2.5-VL-3B-Instruct, we conduct a manual evaluation along two dimensions. Each dimension is rated on a three-point scale (0–2), with the total score normalized by the maximum possible score.

Localization Reliability. We manually evaluate the generated bounding boxes in 300 images. A score of 0 indicates incomplete target coverage or an excessively large bounding box; 1 denotes complete but loose coverage without including unrelated objects; and 2 indicates accurate and tight localization. The annotations achieve a normalized assessment score of 96.00% (576/600).

Criterion	Normalized
Localization Reliability	96.00%
Reason Consistency	95.75%

Table 8: Normalized human assessment scores.

Reason Consistency. We further examine 200 images to assess the relevance of annotated entities to their corresponding driving rationales. Scores of 0, 1, and 2 indicate no relevance, partial relevance without constituting key evidence, and direct relevance to the driving rationale, respectively. The normalized assessment score reaches 95.75% (383/400).

6.4 Qualitative Examples of CODA-GRD

Fig. 4 and Fig. 5 present representative annotation examples from CODA-GRD.

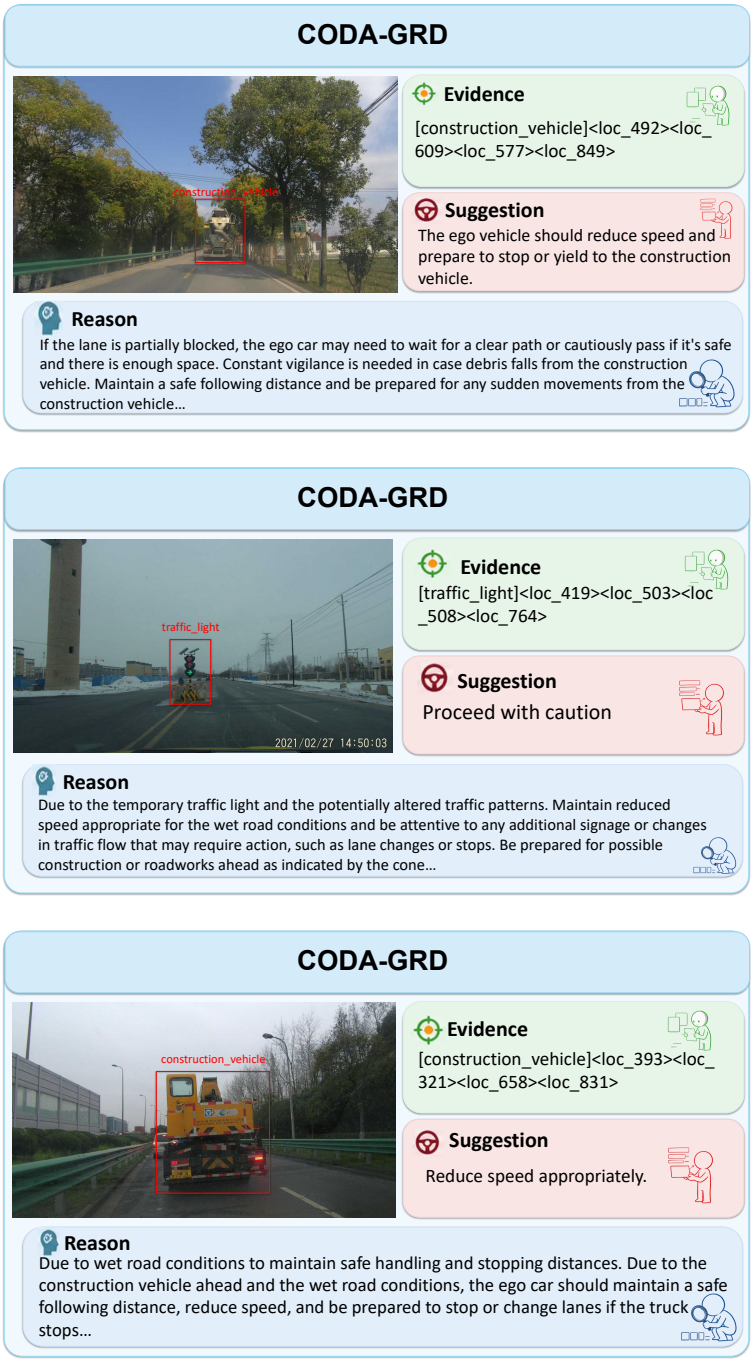


Figure 4: Representative CODA-GRD annotation examples, including Evidence, Suggestion, and Reason.

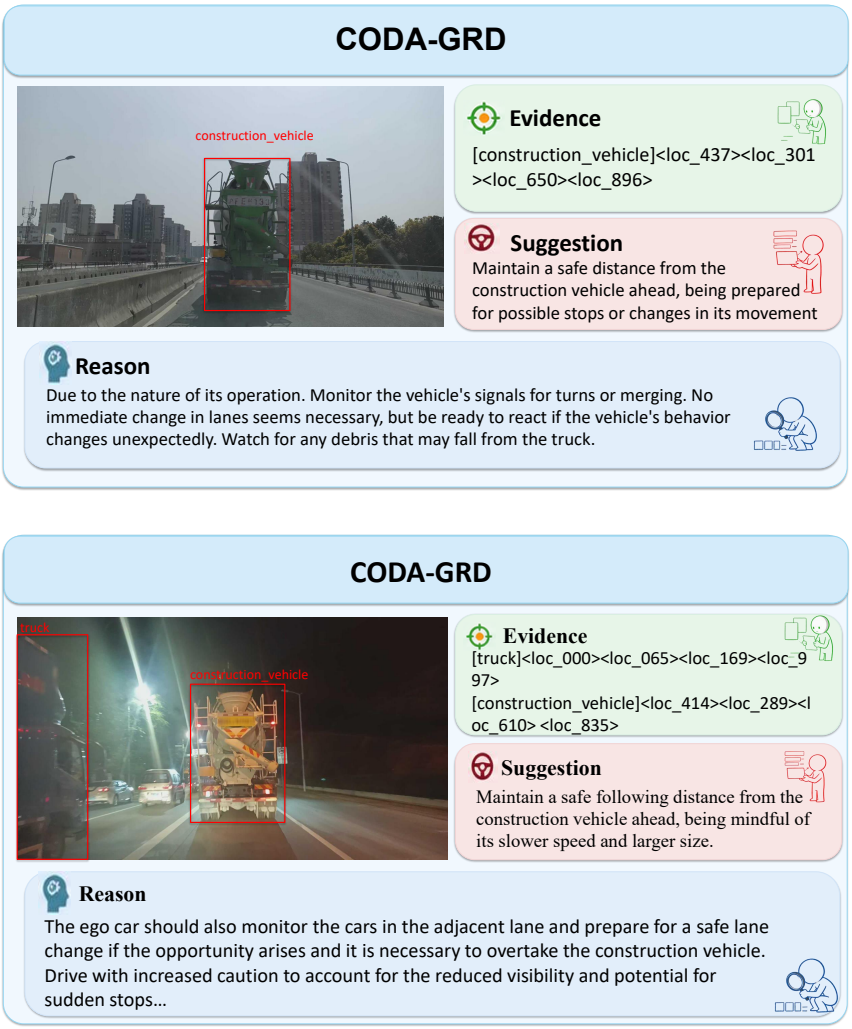


Figure 5: Additional CODA-GRD annotation examples.