Dual-AEB: Synergizing Rule-Based and Multimodal Large Language Models for Effective Emergency Braking

Wei Zhang[1,3,*], Pengfei Li[1,*], Junli Wang[1,4], Bingchuan Sun[2], Qihao Jin[5],
Guangjun Bao[2], Shibo Rui[2], Yang Yu[2], Wenchao Ding[5], Peng Li[1]✉ and Yilun Chen[1]✉

Abstract— **Automatic Emergency Braking (AEB) systems are a crucial component in ensuring the safety of passengers in autonomous vehicles. Conventional AEB systems primarily rely on closed-set perception modules to recognize traffic conditions and assess collision risks. To enhance the adaptability of AEB systems in open scenarios, we propose Dual-AEB, a system combines an advanced multimodal large language model (MLLM) for comprehensive scene understanding and a conventional rule-based rapid AEB to ensure quick response times. To the best of our knowledge, Dual-AEB is the first method to incorporate MLLMs within AEB systems. Through extensive experimentation, we have validated the effectiveness of our method. Codes will be publicly available at `https://github.com/ChipsICU/Dual-AEB`.**

I. Introduction

The Autonomous Emergency Braking (AEB) system is a critical safety feature in autonomous vehicles, designed to mitigate or prevent collisions by automatically activating the brakes when a potential collision is detected [1]. Numerous studies [1]–[5] have demonstrated the effectiveness of AEB systems, with reductions in rear-end collisions ranging from **25%** to **50%**.

Conventionally, AEB systems can be roughly categorized into two types: decision-making-only methods [6]–[15] and end-to-end methods [16], [17]. Decision-making-only methods use perception results of predefined perception categories (e.g., people, cars, bicycles) and apply rule-based techniques [10]–[12], [18] or deep reinforcement learning [13], [14] for braking decisions. End-to-end methods [16], [17], meanwhile, process raw sensory data directly to inform AEB decisions, allowing the system to benefit from comprehensive sensory inputs. These methods generally ensure safety in most driving scenarios.

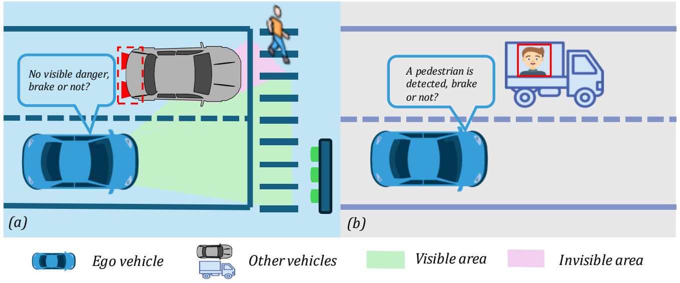

No visible danger, brake or not?
(a)
A pedestrian is detected, brake or not?
(b)
Ego vehicle
Other vehicles
Visible area
Invisible area

Fig. 1. Conventional AEB systems tend to fail in these situations: **(a)** when early detection of a pedestrian is required to brake in advance and avoid danger, and **(b)** when incorrect perception triggers AEB unnecessarily. Such scenarios are challenging for conventional AEB methods.

However, their ability to handle complex driving situations is limited due to a lack of comprehensive scene understanding. For example, in Fig. 1 (a), the scene describes a pedestrian positioned in the ego vehicle's blind spot, intending to cross at a green-light intersection. Typically, decision-making-only methods would not activate braking in this scenario due to the absence of pedestrian perception information, making it impossible to predict an impending collision. Similarly, while end-to-end methods process raw sensory data, they often lack the reasoning capacity to interpret indirect cues—such as the illuminated brake lights on the vehicle to the left of the ego vehicle—that may indicate a potential hazard ahead. In Fig. 1 (b), a truck with a facial advertisement is driving on the ego vehicle's left. Both decision-making-only and end-to-end methods may misinterpret the advertisement as a pedestrian, potentially triggering the AEB system and causing unnecessary braking. A truly effective AEB system should incorporate comprehensive scene understanding, enabling it to differentiate between real hazards and non-threatening elements, thereby ensuring appropriate braking responses.

To address these challenges, we propose the **Dual-AEB** system, which offers the following main advantages: (1) **Comprehensive Scene Understanding**: The Dual-AEB system integrates advanced Multimodal Large Language Models (MLLMs) to achieve a deep understanding of the driving environment. By processing comprehensive data—including environmental conditions, critical perception information, and ego-vehicle states—MLLMs enhance overall situational awareness while reducing the risk of false positives and missed detections. (2) **Optimized Response Time**: The Dual-AEB system leverages the strengths of both the conventional AEB module and MLLM components. The conventional AEB ensures a quick initial response to imminent threats, while the MLLM component provides detailed analyses in complex scenarios. This synergistic approach minimizes response time and maximizes accuracy during critical moments. (3) **Flexible Modular Design**: The Dual-AEB

[1] Institute for AI Industry Research (AIR), Tsinghua University, China, {lipeng, chenyilun}@air.tsinghua.edu.cn, li-pf22@mails.tsinghua.edu.cn.
[2] Lenovo Research.
[3] Harbin Institute of Technology, China, hitwizard@outlook.com.
[4] University of Chinese Academy of Sciences, China, wangjunli2022@ia.ac.cn.
[5] Academy for Engineering and Technology, Fudan University, China, qhjin24@m.fudan.edu.cn, dingwenchao@fudan.edu.cn.
[*] These authors have contributed equally to this work.

system's modular architecture facilitates seamless upgrades and component replacements as technology advances. This feature guarantees long-term efficacy and continuous improvement, preparing the system to meet future challenges.

To summarize, our contributions are as follows:

- We present Dual-AEB, the first work that integrates MLLMs to enhance conventional AEB systems by leveraging their comprehensive scene understanding to improve braking decisions.
- Our method is validated through extensive experiments on both open-loop and closed-loop benchmarks, demonstrating its effectiveness.
- Qualitative analysis on our in-house real-world scenario dataset further confirms the practicality of deploying this system.

II. Related Work

A. Autonomous Emergency Braking (AEB)

AEB systems are essential for vehicle safety, as it autonomously detects risks and activates brakes to mitigate or avoid collisions, significantly reducing traffic accident rates [1]–[5], [19]–[21]. Over time, AEB systems have evolved to utilize either decision-making-only or end-to-end methods, ensuring safety in general scenarios.

Decision-Making-Only Methods. These methods typically rely on a limited set of closed-set perception results, such as detecting pedestrians, vehicles, and bicycles, to determine the necessity of braking actions. These decisions are often based on metrics like Time To Collision (TTC) [10]–[12], [18] or are designed using control algorithms [22]–[25], and sometimes involve learning-based approaches [13]–[15]. While these methods are straightforward and computationally efficient, they suffer from significant limitations in complex, dynamic environments [6]–[9], [26], [27]. Relying on a predefined closed-set of objects can result in the omission of vital environmental information, potentially leading to the failure of the AEB in critical situations.

End-to-End Methods. End-to-end methods bypass traditional decision-making pipelines by directly using raw perception data for AEB decisions [14], [16], [17], [28]–[30]. They offer flexibility and can continuously improve with more data [31], enabling the detection and response to hazards that rule-based systems might miss. However, these approaches face challenges, including the need for large labeled datasets, inconsistent performance on unseen scenarios [32], [33], susceptibility to overfitting, and opaque decision-making processes [34], which hinder their reliability in critical and complex situations.

Overall, both decision-making-only and end-to-end methods face challenges in handling complex driving scenarios. Decision-making-only methods rely on predefined perception categories, limiting their ability to respond to unexpected elements. End-to-end AEB systems, while processing raw sensory data, often struggle with reasoning through complex relationships in the scene. To provide effective braking decisions in complex driving scenarios, a comprehensive understanding of the scene is required for AEB systems.

B. Multimodal Large Language Models (MLLMs)

The advent of large models such as ChatGPT [35], [36] and Gemini has brought us closer to achieving trustworthy autonomous driving [37]–[42] and robotics [43]–[50]. Works like GPT-Driver [37]–[41], [51] have demonstrated superior performance over previous methods through prompt-tuning and fine-tuning techniques in autonomous driving benchmarks. DriveVLM [42] introduces a dual system design, akin to the human brain's slow and fast thinking processes, which efficiently adapts to varying complexities in driving scenarios. This innovative approach helps end-to-end autonomous driving models address corner cases and enhances the overall system's performance ceiling, closely aligning with the focus of our work. Inspired by these works, we integrate MLLMs into AEB systems, aiming to enhance their ability for comprehensive scene understanding and, in turn, provide more effective braking decisions.

III. Dual-AEB

A. Overview

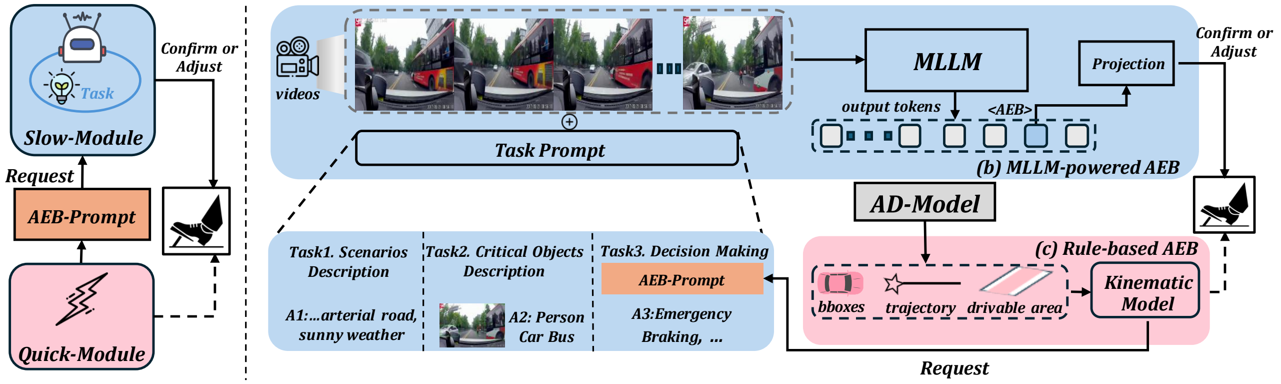
The entire workflow of the Dual-AEB system is illustrated in the Fig. 2. This system consists of two main components: the quick module, called the rule-based AEB, and the slow module, named the MLLM-powered AEB. The rule-based AEB, using conventional rule-based methods, is responsible for the initial decision. When triggered, the quick module packages this initial decision into text, named as the *AEB-Prompt*, and sends it to the slow module, the MLLM-powered AEB. The slow module, utilizing MLLM to analyze the received information, makes the final decision, either confirming or adjusting the quick module's initial decision. This framework can be seamlessly integrated with other autonomous driving algorithms, making it a flexible framework that can be easily expanded or updated.

B. Rule-Based AEB Module

The rule-based AEB module receives perception and planning results from the autonomous driving modules (AD models), including the bounding boxes of agents (B), the ego vehicle's planned trajectory (P), and the drivable area (D). This information is used to evaluate potential collisions within the trajectory horizon (H) by utilizing the kinematic bicycle model (K) [52], with a step size (Δt) for temporal progression. This module then uses Time To Collision (TTC) and Collision (C) [53] as evaluation metrics to assess the need for emergency braking [54]. The decision to initiate braking is then made by comparing the calculated trigger time (t_{trigger}) with a predefined threshold ($t_{\text{threshold}}$). The detailed steps of this process are outlined in Algorithm 1.

C. MLLM-Powered AEB Module

The MLLM-Powered AEB module processes a series of sequential front-view driving images accompanied by predefined task prompts. It consists of a trained MLLM backbone and a projection network. The MLLM backbone analyzes these inputs and generates textual responses, termed Text Generation (TG), which include an <AEB> token to



(a) Overall Architecture

Fig. 2. **Overview of our method.** Dual-AEB **(a)** includes both the quick (rule-based AEB) and slow (MLLM-powered AEB) modules. After receiving information from autonomous driving models (AD-Models), the braking signal can either be directly output by **(c)**, as indicated by dashed lines, or sent to **(b)**, as indicated by solid lines, where the MLLM-powered AEB evaluates and decides whether to confirm or adjust.

Algorithm 1 Rule-Based AEB Module

```
Input: $B_{ego,t_0}, B_{other,t_0}, P, D, K, \Delta t, H$
Initialize: $trigger\_times = []$
for $t = \Delta t$ to $H \times \Delta t$ step $\Delta t$ do
    $B_{ego,t} = UpdateEgoBBox(B_{ego,t-\Delta t}, P[t/\Delta t], K)$
    $B_{other,t} = UpdateOtherBBox(B_{other,t-\Delta t}, K)$
    $TTC_t = CalculateTTC(B_{ego,t}, B_{other,t}, D)$
    $C_t = CalculateCollision(B_{ego,t}, B_{other,t}, D)$
    if $TTC_t < t_{threshold}$ or $C_t$ then
        Append $t$ to $trigger\_times$
    end if
end for
Decision:
if $trigger\_times$ is not empty then
    return Brake
end if
```

indicate when emergency braking is needed. The projection network then utilizes the hidden state associated with the <AEB> token, applying a linear layer followed by a sigmoid activation to output a braking signal—referred to as Braking Signal Generation (BSG)—in the range of 0 to 1.

D. Training Data Construction

The data required for Dual-AEB training includes front-view video inputs, question-answer pairs, and braking signals. We generated this data for the MLLM-powered AEB by utilizing the MM-AU [55] dataset, which captures real-world accident scenarios, and the Bench2Drive [56] dataset, featuring complex simulation scenarios from CARLA [57]. Videos are directly available from both datasets, and Bench2Drive also provides braking signals. For MM-AU, braking signals are derived by analyzing the chronological sequence of traffic accidents. For instance, in video sequences where a collision is imminent, emergency braking is required, and in such cases, the ground truth for the braking signal is set to 1.

For the question-answer pairs, we designed three sub-tasks focused on driving scenarios, critical objects, and decision-making processes, structuring the reasoning process in a step-by-step manner [58], [59]. Using GPT-4, we generated diverse question-answer templates and filled them with ground truth. Details are outlined below.

Scenarios Description. The driving environment significantly influences the complexity of driving tasks [42]. For instance, rural roads are more likely to encounter sudden appearances of animals. Annotations include details on weather variations, visibility, road traction, time of day, and road types. An example of an annotated scenario description question-answer pair is, **Q**: “*What are the environmental details captured in this driving video?*” **A**: “*The ego vehicle navigates an arterial roadway under clear, sunny conditions in an urban environment during daylight.*”

Critical Objects Description. Annotations for critical objects include details such as 2D bounding boxes in $(x\_min, y\_min, x\_max, y\_max)$ format, distance to the ego vehicle, traffic signals, and the intentions of surrounding agents. For example, “*A black vehicle with a blinking left turn signal, located at [(197, 474), (295, 667)], is 14.26 meters away from the ego vehicle, suggesting it is preparing to make a left turn.*” These details provide critical information for the system to determine the precise position of the object within the scene. With this information, Dual-AEB can forecast the object’s future movements, resulting in more accurate and informed decision-making in various scenarios.

Decision Making. In the final sub-task, questions are supplemented with *AEB-Prompt*, which represent initial decisions generated by the rule-based AEB module. For example, a *AEB-Prompt* might be, “*Initial decision: A collision with the black vehicle on the left is expected in 1.2 seconds, and I decide to brake.*” To prevent the MLLM-powered AEB from becoming overly reliant on *AEB-Prompt*, 50% of the training data includes incorrect initial decisions. We categorize AEB actions into three meta-actions: *Normal*, *Early Warning*, and

Emergency Braking [1]. A possible answer might be, "*Early Warning. The presence of a truck and another black car ahead requires heightened awareness. <AEB>*"

E. Training and Inference

During training, AD models are trained on Bench2Drive [56], and the MLLM-powered AEB employs a composite loss function that integrates both TG and BSG to ensure cohesive semantic learning. The overall loss $\mathcal{L}$ is defined as:

$$\begin{aligned} \mathcal{L} = &- \sum_{n=1}^{N} \sum_{i=1}^{V} y_{n,i} \log(\hat{y}_{n,i}) \\ &- \left[z \log(\hat{z}) + (1-z) \log(1-\hat{z}) \right], \end{aligned} \tag{1}$$

where $\hat{y}_{n,i}$ and $y_{n,i}$ represent the predicted and ground truth probabilities of word i at position n in the text sequence, with V being the vocabulary size. The variables z and $\hat{z}$ denote the ground truth and predicted braking signals, respectively.

During inference, the rule-based AEB receives perception and planning results from the pre-trained AD model. It can either directly output its braking signal or initiate an interaction with the MLLM-powered AEB via the *AEB-Prompt*. If interaction is triggered, the MLLM-powered AEB engages in a multi-round Q&A process to address the three sub-tasks. Ultimately, during the decision-making task, it determines whether to confirm or adjust the initial decision made by the rule-based AEB. The final braking signal is then output after passing through the projection network.

IV. Experiments

A. Datasets

To provide a more comprehensive evaluation of our method, we assess it on two datasets. We perform open-loop evaluations using both MM-AU [55] and Bench2Drive [56], and conduct closed-loop evaluations using the Bench2Drive [56] benchmark. We carefully construct 120,113 samples from MM-AU and 132,922 samples from Bench2Drive, with the proportions of *Normal*, *Early Warning*, and *Emergency Braking* approximately at 1:1:1. Using a 9:1 ratio, we split this data into training and test sets.

B. Metrics

In the open-loop evaluation, we focus on the accuracy of the model's predicted braking signals and the quality of the generated text. For the Braking Signal Generation (BSG), we utilize standard AEB task metrics [17], namely *Precision* and *Recall*. Positive cases refer to situations where AEB activation is necessary, and negative cases to those where it is not needed. True Positives (TP) are instances of correctly triggered *Emergency Braking*, while True Negatives (TN) are correct non-activations (*Early Warning/Normal*). Conversely, False Positives (FP) represent erroneous activations, and False Negatives (FN) denote instances where necessary AEB actions are missed. The formulas for *Precision* and *Recall* are provided as follows:

$$\text{Precision} = \frac{TP}{TP+FP}, \ \text{Recall} = \frac{TP}{TP+FN}, \tag{2}$$

for the Text Generation (TG) aspect, we evaluate the quality of the generated text using BLEU4 [60], METEOR [61], and ROUGE-L [62] metrics.

In the closed-loop evaluation, our primary focus is on the model's overall driving performance. We utilize the *Driving Score* and *Success Rate* metrics from Bench2Drive [56]. The *Driving Score* aggregates multiple driving metrics from Bench2Drive into a weighted sum, while the *Success Rate* measures the proportion of successfully completed scenarios out of the total number of scenarios. Additionally, we introduced the *Collision Rate*, defined as the average number of collisions per scenario, to better assess the model's ability to avoid collisions.

C. Implementation Details

For the rule-based AEB module, we instantiate two widely used models, UniAD [63] and VAD [64], to generate perception and planning results, thereby validating the flexibility of our framework. We adopt a step size of 0.2 seconds and a 3-second horizon, aligning with the 3-second future trajectory provided by the Bench2Drive benchmark planner. For the MLLM-powered AEB module, we employ the state-of-the-art LLaVA-OneVision [65] model as the backbone and perform full fine-tuning of all its components. Following the recommendations in [65], a learning rate of 2×10^{-6} is applied to the vision encoder, while a learning rate of 1×10^{-5} is used for the other components.

For closed-loop evaluation, the rule-based AEB module is provided with information from VAD or UniAD and interacts with the MLLM-powered AEB module every 2.5 seconds, and the MLLM-powered AEB is trained on the Bench2Drive dataset before being evaluated in CARLA.

All experiments are conducted on a server equipped with 8 NVIDIA A100 80G GPUs. To ensure the deployability of the framework in real-world scenarios, inference time consumption tests are performed on consumer-grade devices; here, we use an NVIDIA Jetson Orin.

D. Closed-Loop Experimental Results

We leverage Qwen-0.5B as the foundational language model for LLaVA-OneVision to ensure faster response times within the Bench2Drive closed-loop simulation benchmark. Detailed results are provided in Table I. Specifically, with Dual-AEB, VAD's *Driving Score* increased by **14.74%**, while the *Success Rate* remained unchanged. For UniAD-Base, the *Driving Score* improved by **6.84%**, and the *Success Rate* increased from 9.54 to 10.00. Among all evaluated models [63], [64], [66]–[69], VAD with Dual-AEB achieved the highest performance in terms of *Driving Score*.

Dual-AEB helps improve the closed-loop performance of these models by providing effective braking decisions. **However, since it does not alter the autonomous driving model's output trajectory, the improvement on** ***Success Rate*** **is limited.** Despite this, the overall driving performance of these end-to-end models is enhanced, demonstrating the potential for Dual-AEB to be integrated into other autonomous driving systems.

TABLE I

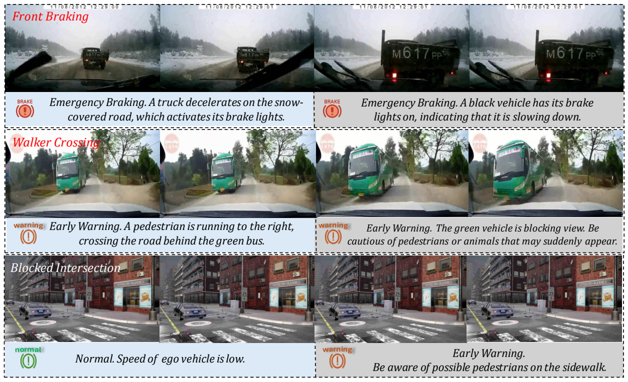
CLOSED-LOOP RESULTS IN BENCH2DRIVE. * REPRESENTS EXPERT FEATURE DISTILLATION. THE METRICS INCLUDE DRIVING SCORE (**D**), SUCCESS RATE (**S**) AND THE IMPROVEMENT OF SUCCESS RATE (ΔS).

Method	**Input**	**Metrics ↑**		
		D	**S**	ΔS
TCP* [67]	Ego State + Front Cameras	23.63	7.72	-
TCP-ctrl*	Ego State + Front Cameras	18.63	5.45	-
TCP-traj*	Ego State + Front Cameras	36.78	26.82	-
ThinkTwice* [68]	Ego State + 6 Cameras	39.88	28.14	-
DriveAdapter* [69]	Ego State + 6 Cameras	42.91	30.71	-
AD-MLP [66]	Ego State	9.14	0.00	-
UniAD-Tiny [63]	Ego State + 6 Cameras	32.00	9.54	-
UniAD-Base [63]	Ego State + 6 Cameras	37.72	9.54	-
VAD [64]	Ego State + 6 Cameras	39.42	10.00	-
UniAD-Base+Dual-AEB	Ego State + 6 Cameras	40.32	10.00	**0.06**
VAD+Dual-AEB	Ego State + 6 Cameras	**45.23**	10.00	0.00

TABLE II

IMPACT OF DIFFERENT METHODS IN OPEN-LOOP EVALUATIONS (**P**: PRECISION, **RECALL**, **B4**: BLEU-4, **M**: METEOR, **R**: ROUGE-L). THE MODELS EVALUATED INCLUDE QWEN-0.5B (Q0.5B), QWEN-7B (Q7B), AND YOLO, ACROSS TWO DATASETS: MM-AU AND B2D (BENCH2DRIVE).

Dataset	**Method**	**Metrics ↑**				
		P	**Recall**	**B4**	**M**	**R**
MM-AU	YOLO	26.62	38.99	-	-	-
	Q0.5B	72.90	73.80	13.18	30.86	33.69
	Q 7B	**73.60**	**75.20**	**20.12**	**36.30**	**37.95**
B2D	YOLO	56.27	67.34	-	-	-
	Q0.5B	65.25	**94.35**	21.74	42.74	40.04
	Q 7B	**67.36**	93.14	**28.09**	**46.85**	**45.07**



Emergency Braking. A truck decelerates on the snow-covered road, which activates its brake lights.

Emergency Braking. A black vehicle has its brake lights on, indicating that it is slowing down.

Early Warning. A pedestrian is running to the right, crossing the road behind the green bus.

Early Warning. The green vehicle is blocking view. Be cautious of pedestrians or animals that may suddenly appear.

Normal. Speed of ego vehicle is low.

Early Warning.
Be aware of possible pedestrians on the sidewalk.

Fig. 3. Qualitative analysis: MLLM-powered AEB provided reasonable descriptions for different meta actions. The left side (🟦) represents the ground truth results, while the right side (⬜) represents the predicted results.

E. Open-Loop Experimental Results

In this part, we aim to evaluate the comprehensive scene understanding capabilities of MLLM-powered AEB in real-world scenarios. We compare the *Precision*, *Recall*, and text generation quality of the trained LLaVA-OneVision model based on Qwen-0.5B and Qwen-7B, taking YOLO-AEB [30] as the baseline. Since the YOLO model processes single images, we take the last frame from each video split as input. We present the quantification results in Table II, and provide qualitative examples in Fig. 3.

Compared to YOLO-AEB, MLLM-powered AEB achieves higher *Precision* and *Recall*, thanks to the powerful and comprehensive scene understanding capabilities of MLLMs, which enable more effective braking decisions. Additionally, the increase in model scale further contributes to the observed performance improvements.

On the Bench2Drive benchmark, both YOLO and the MLLM-powered AEB demonstrate improved performance compared to their results on MM-AU. This enhancement is likely due to the simpler and more uniform objects in the simulated scenarios, which facilitate the models in learning effective braking behaviors. However, these improvements do not extend to *Precision*. Specifically, the MLLM's performance is somewhat hindered because Bench2Drive's simulation data is generated by an expert model [70] that consistently avoids danger by braking. As a result, some scenarios appear inherently safe, reducing the model's tendency to initiate braking and leading to lower *Precision*.

F. Ablation Study

Impact of Different AEB Modules. To assess the effectiveness of various Dual-AEB modules in providing accurate braking decisions, we conducte an evaluation. The specific results are presented in Table III.

After integrating the rule-based AEB module into VAD and UniAD, both models exhibit a significant reduction in *Collision Rate*, resulting in an improved *Driving Score*. Replacing the rule-based AEB with an MLLM-powered AEB, which possesses advanced scene understanding capabilities, further decreases the *Collision Rate*, highlighting the MLLM's stronger ability to perceive potential dangers. Finally, with the implementation of the full Dual-AEB, the *Driving Score* improves once more. This demonstrates that the rule-based AEB can provide rapid responses in dangerous scenarios when the MLLM is not invoked—since it operates intermittently—thereby compensating for the limitations of the MLLM.

Impact of Different Task Modules. In previous experiments, we design three sequential tasks to guide the model's TG process. This section compares the performance of different task combinations on the MM-AU and Bench2Drive open-loop datasets (Table IV). The results indicate that incorporating additional tasks enhances the inference performance of the MLLM-powered AEB module, with the critical objects description task providing the most substantial improvement. This task allows the model to better understand the intentions of other objects, resulting in more effective braking.

Impact of Different Trigger Intervals of the MLLM-powered AEB. In previous experiments, we test the MLLM-powered AEB module in Bench2Drive with a 2.5-second trigger interval. Here, we evaluate the impact of different intervals: 0.5, 2.5, and 5 seconds. As shown in Table V,

TABLE III
IMPACT OF DIFFERENT AEB MODULES IN CLOSE-LOOP EVALUATION (**R**: RULE-BASED AEB, **M**: MLLM-POWERED AEB). THE METRICS INCLUDE DRIVING SCORE (**D**), SUCCESS RATE (**S**), AND COLLISION RATE (**C**).

Models	Modules		Metrics		
	R	**M**	**D ↑**	**S ↑**	**C ↓**
VAD	✗	✗	39.42	10.00	70.19
	✔	✗	43.22	10.00	56.43
	✗	✔	44.75	10.00	53.42
	✔	✔	**45.23**	**10.00**	**50.86**
UniAD	✗	✗	37.72	9.54	87.06
	✔	✗	38.47	9.54	74.53
	✗	✔	39.46	10.00	71.68
	✔	✔	**40.32**	**10.00**	**69.02**

TABLE IV
IMPACT OF DIFFERENT TASK MODULES (**S**: SCENARIOS DESCRIPTION, **O**: CRITICAL OBJECTS DESCRIPTION, **D**: DECISION MAKING).

Datasets	**S**	**O**	**D**	**Precision**	**Recall**
MM-AU	✗	✗	✗	56.80	57.62
	✔	✗	✗	58.68	62.54
	✔	✔	✗	71.44	68.60
	✔	✔	✔	**72.90**	**73.80**
Bench2Drive	✗	✗	✗	57.44	68.32
	✔	✗	✗	59.14	72.69
	✔	✔	✗	62.53	88.96
	✔	✔	✔	**65.25**	**94.35**

TABLE V
IMPACT OF DIFFERENT TRIGGER INTERVALS ON BENCH2DRIVE (**D**: DRIVING SCORE, **C**: COLLISION RATE, **T**: AVERAGE INFERENCE TIME).

Models	**Trigger Time**	**D ↑**	**C ↓**	**T (s) ↓**
VAD	0.5	**45.92**	**49.98**	8.10
	2.5	45.23	50.86	1.55
	5.0	42.14	54.23	**0.80**
UniAD	0.5	**40.56**	**68.74**	8.25
	2.5	40.32	69.02	1.60
	5.0	38.87	73.54	**0.80**

TABLE VI
COMPARISON OF AVERAGE INFERENCE TIME CONSUMPTION RESULTS ON DIFFERENT DEVICES USING TENSORRT.

Modules	TensorRT	Devices	
		A100	**Jetson Orin**
Dual-AEB-F	✗	1.55s	11.44s
	✔	0.96s	4.86s
Dual-AEB-S	✗	0.68s	3.87s
	✔	0.31s	1.21s

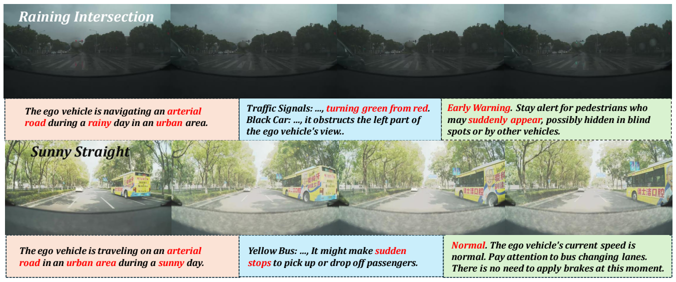
shorter intervals improve *Driving Score*, but the 0.5-second interval provides only marginal gains and significantly increases inference time. Thus, a 2.5-second interval offers a better balance between performance and efficiency.

Comparison of Inference Time Consumption on Different Devices. After conducting frequency analysis, we perform average inference time tests on the NVIDIA Jetson Orin and optimize the models using TensorRT. The results are presented in Table VI. To achieve faster response times, we develop a version of Dual-AEB that executes only the decision-making task (Dual-AEB-S), based on the original Dual-AEB that executes all tasks (Dual-AEB-F). By combining trigger time with accelerated inference, we further reduce the average inference time.

G. Cases Study

MM-AU comes from network dashcam recordings, and Bench2Drive originates from simulations. There are still some differences between them and real-world driving data. To further explore the capabilities of our model in real-world driving scenarios, we conduct tests on our in-house dataset, which is collected from a vehicle equipped with a complete autonomous driving hardware and software system. Fig. 4 shows several case results, which are similar to the scenarios mentioned in Fig. 1.

In the first scenario, conventional AEB often proceed without providing an early warning to the driver, which may result in insufficient reaction time when a pedestrian suddenly appears. In the second scenario, conventional AEB may mistakenly identify an advertisement on a bus as a pedestrian, triggering unnecessary emergency braking. In contrast, Dual-AEB offers early warnings and filters out false positives, thereby enhancing both safety and the overall effectiveness of the system.



The ego vehicle is navigating an arterial road during a rainy day in an urban area.

Traffic Signals: ..., turning green from red. Black Car: ..., it obstructs the left part of the ego vehicle's view..

Early Warning. Stay alert for pedestrians who may suddenly appear, possibly hidden in blind spots or by other vehicles.

The ego vehicle is traveling on an arterial road in an urban area during a sunny day.

Yellow Bus: ..., It might make sudden stops to pick up or drop off passengers.

Normal. The ego vehicle's current speed is normal. Pay attention to bus changing lanes. There is no need to apply brakes at this moment.

Fig. 4. Qualitative analysis: Dual-AEB provides reasonable descriptions in our in-house dataset. 🟧 represents the scenarios description task, 🟦 represents the objects description task, and 🟩 represents the decision making task.

V. CONCLUSION

In this study, we present Dual-AEB, an innovative method that integrates conventional rule-based modules with advanced Multimodal Large Language Models (MLLMs) to enhance autonomous emergency braking (AEB). The conventional component ensures rapid initial responses, while the MLLM component enhances decision accuracy by processing complex environmental data and ego-vehicle states. Evaluations on open-loop and closed-loop benchmarks demonstrate the effectiveness of this system in providing robust braking strategies. Further testing on real-world scenarios confirms the practical applicability and strong performance of the Dual-AEB system. Enhancing precision in real-world applications is the focus of our future work.

References

[1] L. Yang, Y. Yang, G. Wu, X. Zhao, S. Fang, X. Liao, R. Wang, and M. Zhang, “A systematic review of autonomous emergency braking system: Impact factor, technology, and performance evaluation,” *JOURNAL OF ADVANCED TRANSPORTATION*, vol. 2022, APR 18 2022.

[2] J. B. Cicchino, “Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates,” *Accident Analysis & Prevention*, vol. 99, pp. 142–152, 2017.

[3] E. Jeong and C. Oh, “Methodology for estimating safety benefits of advanced driver assistant systems,” *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 12, no. 3, pp. 65–77, 2013.

[4] B. Fildes, M. Keall, N. Bos, A. Lie, Y. Page, C. Pastor, L. Pennisi, M. Rizzi, P. Thomas, and C. Tingvall, “Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes,” *Accident Analysis & Prevention*, vol. 81, pp. 24–29, 2015.

[5] I. Isaksson-Hellman and M. Lindman, “Evaluation of the crash mitigation effect of low-speed automated emergency braking systems based on insurance claims data,” *Traffic injury prevention*, vol. 17, no. sup1, pp. 42–47, 2016.

[6] H. Jeppsson and N. Lubbe, “Simulating automated emergency braking with and without torricelli vacuum emergency braking for cyclists: Effect of brake deceleration and sensor field-of-view on accidents, injuries and fatalities,” *ACCIDENT ANALYSIS AND PREVENTION*, vol. 142, JUL 2020.

[7] H. Chajmowicz, J. Saade, and S. Cuny, “Prospective assessment of the effectiveness of autonomous emergency braking in car-to-cyclist accidents in france,” *TRAFFIC INJURY PREVENTION*, vol. 20, no. 2, SI, pp. S20–S25, NOV 25 2019.

[8] G. Savino, J. Brown, M. Rizzi, M. Pierini, and M. Fitzharris, “Triggering algorithm based on inevitable collision states for autonomous emergency braking (AEB) in motorcycle-to-car crashes,” in *2015 IEEE INTELLIGENT VEHICLES SYMPOSIUM (IV)*, ser. IEEE Intelligent Vehicles Symposium, 2015, pp. 1195–1200, iEEE Intelligent Vehicles Symposium, Seoul, SOUTH KOREA, JUN 28-JUL 01, 2015.

[9] G. Savino, M. Pierini, J. Thompson, M. Fitzharris, and M. G. Lenne, “Exploratory field trial of motorcycle autonomous emergency braking (MAEB): Considerations on the acceptability of unexpected automatic decelerations,” *TRAFFIC INJURY PREVENTION*, vol. 17, no. 8, pp. 855–862, 2016.

[10] M. Minderhoud and P. Bovy, “Extended time-to-collision measures for road traffic safety assessment,” *ACCIDENT ANALYSIS AND PREVENTION*, vol. 33, no. 1, pp. 89–97, JAN 2001.

[11] M. Wessels, C. Zaehme, and D. Oberfeld, “Auditory information improves time-to-collision estimation for accelerating vehicles,” *CURRENT PSYCHOLOGY*, vol. 42, no. 27, pp. 23 195–23 205, OCT 2023.

[12] K. Vogel, “A comparison of headway and time to collision as safety indicators,” *ACCIDENT ANALYSIS AND PREVENTION*, vol. 35, no. 3, pp. 427–433, MAY 2003.

[13] H. Chae, C. M. Kang, B. Kim, J. Kim, C. C. Chung, and J. W. Choi, “Autonomous braking system via deep reinforcement learning,” in *2017 IEEE 20TH INTERNATIONAL CONFERENCE ON INTELLIGENT TRANSPORTATION SYSTEMS (ITSC)*, ser. IEEE International Conference on Intelligent Transportation Systems-ITSC. IEEE, 2017, 20th IEEE International Conference on Intelligent Transportation Systems (ITSC), Yokohama, JAPAN, OCT 16-19, 2017.

[14] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and Y. Zhang, “A decision-making strategy for vehicle autonomous braking in emergency via deep reinforcement learning,” *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 69, no. 6, pp. 5876–5888, JUN 2020.

[15] X. Teng, S. Xu, D. Guo, Y. Guo, W. Meng, and X. Zhang, “DTTCNet: Time-to-collision estimation with autonomous emergency braking using multi-scale transformer network,” *IEEE Transactions on Mobile Computing*, pp. 1–13, 2024.

[16] R. Zhang and A. Pourkand, “Emergency-braking distance prediction using deep learning,” 2021.

[17] R. Itu and R. Danescu, “Fully convolutional neural network for vehicle speed and emergency-brake prediction,” *SENSORS*, vol. 24, no. 1, JAN 2024.

[18] S. P. Sharan, F. Pittaluga, V. K. B. G, and M. Chandraker, “LLM-Assist: Enhancing closed-loop planning with language-based reasoning,” 2023.

[19] J. B. Cicchino, “Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates,” *ACCIDENT ANALYSIS AND PREVENTION*, vol. 99, no. A, pp. 142–152, FEB 2017.

[20] B. Fildes, M. Keall, N. Bos, A. Lie, Y. Page, C. Pastor, L. Pennisi, M. Rizzi, P. Thomas, and C. Tingvall, “Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes,” *ACCIDENT ANALYSIS AND PREVENTION*, vol. 81, pp. 24–29, AUG 2015, 2nd International Conference on Transportation Information and Safety (ICTIS), Wuhan, PEOPLES R CHINA, JUN 29-JUL 02, 2013.

[21] I. Isaksson-Hellman and M. Lindman, “Evaluation of the crash mitigation effect of low-speed automated emergency braking systems based on insurance claims data,” *TRAFFIC INJURY PREVENTION*, vol. 17, no. 1, SI, pp. 42–47, 2016.

[22] K. Gounis and N. Bassiliades, “Intelligent momentary assisted control for autonomous emergency braking,” *SIMULATION MODELLING PRACTICE AND THEORY*, vol. 115, FEB 2022.

[23] T. Zhou, W. Liu, H. Li, S. Xu, and C. Sun, “Improvement of an autonomous emergency-braking decision-making system for commercial vehicles based on unsafe control strategy analysis,” *Journal of Tsinghua University (Science and Technology)*, pp. 1415–27, 2023 2023.

[24] K. Lee and H. Peng, “Evaluation of automotive forward collision warning and collision avoidance algorithms,” *VEHICLE SYSTEM DYNAMICS*, vol. 43, no. 10, pp. 735–751, OCT 2005.

[25] H. Mu, Z. Li, G. Sun, L. Li, Y. Zhao, and M. Mei, “An autonomous emergency braking strategy based on non-linear model predictive deceleration control,” *IET INTELLIGENT TRANSPORT SYSTEMS*, vol. 17, no. 3, pp. 562–574, MAR 2023.

[26] S. Nguyen, Z. Rahman, and B. T. Morris, “Pedestrian emergency braking in ten weeks,” in *2022 IEEE INTERNATIONAL CONFERENCE ON VEHICULAR ELECTRONICS AND SAFETY (ICVES)*. Compania internat integra; Flypass; Transmilenio S.A; Mobil Secret Manizalez; Pilot univ Colombia; Eafit univ; Its Colombia, 2022, iEEE International Conference on Vehicular Electronics and Safety (ICVES), Bogota, COLOMBIA, NOV 14-16, 2022.

[27] W. Yang, X. Zhang, Q. Lei, and X. Cheng, “Research on longitudinal active collision avoidance of autonomous emergency braking pedestrian system (AEB-P),” *SENSORS*, vol. 19, no. 21, NOV 2019.

[28] F. Cai and X. Koutsoukos, “Real-time out-of-distribution detection in cyber-physical systems with learning-enabled components,” *IET CYBER-PHYSICAL SYSTEMS: THEORY & APPLICATIONS*, vol. 7, no. 4, pp. 212–234, DEC 2022.

[29] M. H. Hwang, G. S. Lee, E. Kim, H. W. Kim, S. Yoon, T. Talluri, and H. R. Cha, “Regenerative braking control strategy based on ai algorithm to improve driving comfort of autonomous vehicles,” *APPLIED SCIENCES-BASEL*, vol. 13, no. 2, JAN 2023.

[30] T.-H. Wu, T.-W. Wang, and Y.-Q. Liu, “Real-time vehicle and distance detection based on improved YOLO v5 network,” in *2021 3rd World Symposium on Artificial Intelligence (WSAI)*. IEEE, 2021, pp. 24–28.

[31] S. Rajendar and V. K. Kaliappan, “Recent advancements in autonomous emergency braking: A survey,” in *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2021, pp. 1027–1032.

[32] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, “End-to-end autonomous driving: Challenges and frontiers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2024.

[33] D. Xin, L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran, “Accelerating human-in-the-loop machine learning: Challenges and opportunities,” in *Proceedings of the second workshop on data management for end-to-end machine learning*, 2018, pp. 1–4.

[34] E. Duede, “Deep learning opacity in scientific discovery,” *Philosophy of Science*, vol. 90, no. 5, pp. 1089–1099, 2023.

[35] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” 2020.

[36] Z. Yang, L. Li, K. Lin, J. Wang, C.-C. Lin, Z. Liu, and L. Wang, “The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision),” 2023. [Online]. Available: https://arxiv.org/abs/2309.17421

[37] J. Mao, Y. Qian, J. Ye, H. Zhao, and Y. Wang, “GPT-Driver: Learning to drive with GPT,” 2023.

[38] H. Shao, Y. Hu, L. Wang, S. L. Waslander, Y. Liu, and H. Li, “LM-

Drive: Closed-loop end-to-end driving with large language models," 2023.

[39] H. Sha, Y. Mu, Y. Jiang, L. Chen, C. Xu, P. Luo, S. E. Li, M. Tomizuka, W. Zhan, and M. Ding, "LanguageMPC: Large language models as decision makers for autonomous driving," 2023.

[40] D. Fu, X. Li, L. Wen, M. Dou, P. Cai, B. Shi, and Y. Qiao, "Drive like a human: Rethinking autonomous driving with large language models," 2023.

[41] L. Chen, O. Sinavski, J. Hünermann, A. Karnsund, A. J. Willmott, D. Birch, D. Maund, and J. Shotton, "Driving with LLMs: Fusing Object-Level Vector Modality for Explainable Autonomous Driving," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 14 093–14 100.

[42] X. Tian, J. Gu, B. Li, Y. Liu, Y. Wang, Z. Zhao, K. Zhan, P. Jia, X. Lang, and H. Zhao, "DriveVLM: The convergence of autonomous driving and large vision-language models," 2024. [Online]. Available: https://arxiv.org/abs/2402.12289

[43] Z. Mandi, S. Jain, and S. Song, "RoCo: Dialectic multi-robot collaboration with large language models," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 286–299.

[44] G. Tziafas and H. Kasaei, "Lifelong robot library learning: Bootstrapping composable and generalizable skills for embodied control with language models," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 515–522.

[45] M. A. Graule and V. Isler, "GG-LLM: Geometrically grounding large language models for zero-shot human activity forecasting in human-aware task planning," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 568–574.

[46] W. Xia, D. Wang, X. Pang, Z. Wang, B. Zhao, D. Hu, and X. Li, "Kinematic-aware prompting for generalizable articulated object manipulation with llms," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 2073–2080.

[47] Z. Zhou, J. Song, K. Yao, Z. Shu, and L. Ma, "ISR-LLM: Iterative self-refined large language model for long-horizon sequential task planning," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 2081–2088.

[48] Y. Chen, J. Arkin, Y. Zhang, N. Roy, and C. Fan, "Scalable multi-robot collaboration with large language models: Centralized or decentralized systems?" in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 4311–4317.

[49] D. Rivkin, N. Kakodkar, F. Hogan, B. H. Baghi, and G. Dudek, "CARTIER: Cartographic language reasoning targeted at instruction execution for robots," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 5615–5621.

[50] J. Zhang, L. Tang, Y. Song, Q. Meng, H. Qian, J. Shao, W. Song, S. Zhu, and J. Gu, "FLTRNN: Faithful long-horizon task planning for robotics with large language models," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 6680–6686.

[51] Y. Zheng, Z. Xing, Q. Zhang, B. Jin, P. Li, Y. Zheng, Z. Xia, K. Zhan, X. Lang, Y. Chen, and D. Zhao, "PlanAgent: A multi-modal large language agent for closed-loop vehicle motion planning," 2024. [Online]. Available: https://arxiv.org/abs/2406.01587

[52] W. Xu, Q. Wang, and J. M. Dolan, "Autonomous vehicle motion planning via recurrent spline optimization," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 7730–7736.

[53] D. Dauner, M. Hallgarten, T. Li, X. Weng, Z. Huang, Z. Yang, H. Li, I. Gilitschenski, B. Ivanovic, M. Pavone, A. Geiger, and K. Chitta, "NAVSIM: Data-driven non-reactive autonomous vehicle simulation and benchmarking," *arXiv*, vol. 2406.15349, 2024.

[54] D. Dauner, M. Hallgarten, A. Geiger, and K. Chitta, "Parting with misconceptions about learning-based vehicle motion planning," in *Conference on Robot Learning (CoRL)*, 2023.

[55] J. Fang, L.-l. Li, J. Zhou, J. Xiao, H. Yu, C. Lv, J. Xue, and T.-S. Chua, "Abductive ego-view accident video understanding for safe driving perception," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 22 030–22 040.

[56] X. Jia, Z. Yang, Q. Li, Z. Zhang, and J. Yan, "Bench2Drive: Towards multi-ability benchmarking of closed-loop end-to-end autonomous driving," *arXiv preprint arXiv:2406.03877*, 2024.

[57] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," 2017.

[58] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," 2023.

[59] B. Jin, X. Liu, Y. Zheng, P. Li, H. Zhao, T. Zhang, Y. Zheng, G. Zhou, and J. Liu, "ADAPT: Action-aware driving caption transformer," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 7554–7561.

[60] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.

[61] S. Banerjee and A. Lavie, "METEOR: An automatic metric for mt evaluation with improved correlation with human judgments," in *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, 2005, pp. 65–72.

[62] C.-Y. Lin and F. J. Och, "Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics," in *Proceedings of the 42nd annual meeting of the association for computational linguistics (ACL-04)*, 2004, pp. 605–612.

[63] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, L. Lu, X. Jia, Q. Liu, J. Dai, Y. Qiao, and H. Li, "Planning-oriented autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[64] B. Jiang, S. Chen, Q. Xu, B. Liao, J. Chen, H. Zhou, Q. Zhang, W. Liu, C. Huang, and X. Wang, "VAD: Vectorized scene representation for efficient autonomous driving," *ICCV*, 2023.

[65] B. Li, Y. Zhang, D. Guo, R. Zhang, F. Li, H. Zhang, K. Zhang, Y. Li, Z. Liu, and C. Li, "LLaVA-OneVision: Easy visual task transfer," 2024. [Online]. Available: https://arxiv.org/abs/2408.03326

[66] J.-T. Zhai, Z. Feng, J. Du, Y. Mao, J.-J. Liu, Z. Tan, Y. Zhang, X. Ye, and J. Wang, "Rethinking the open-loop evaluation of end-to-end autonomous driving in nuscenes," 2023. [Online]. Available: https://arxiv.org/abs/2305.10430

[67] P. Wu, X. Jia, L. Chen, J. Yan, H. Li, and Y. Qiao, "Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline," *Advances in Neural Information Processing Systems*, vol. 35, pp. 6119–6132, 2022.

[68] X. Jia, P. Wu, L. Chen, J. Xie, C. He, J. Yan, and H. Li, "Think twice before driving: Towards scalable decoders for end-to-end autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 21 983–21 994.

[69] X. Jia, Y. Gao, L. Chen, J. Yan, P. L. Liu, and H. Li, "DriveAdapter: Breaking the coupling barrier of perception and planning in end-to-end autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 7953–7963.

[70] Q. Li, X. Jia, S. Wang, and J. Yan, "Think2Drive: Efficient reinforcement learning by thinking in latent world model for quasi-realistic autonomous driving (in carla-v2)," in *ECCV*, 2024.