
Survey on Language Model in Intelligent Vehicles

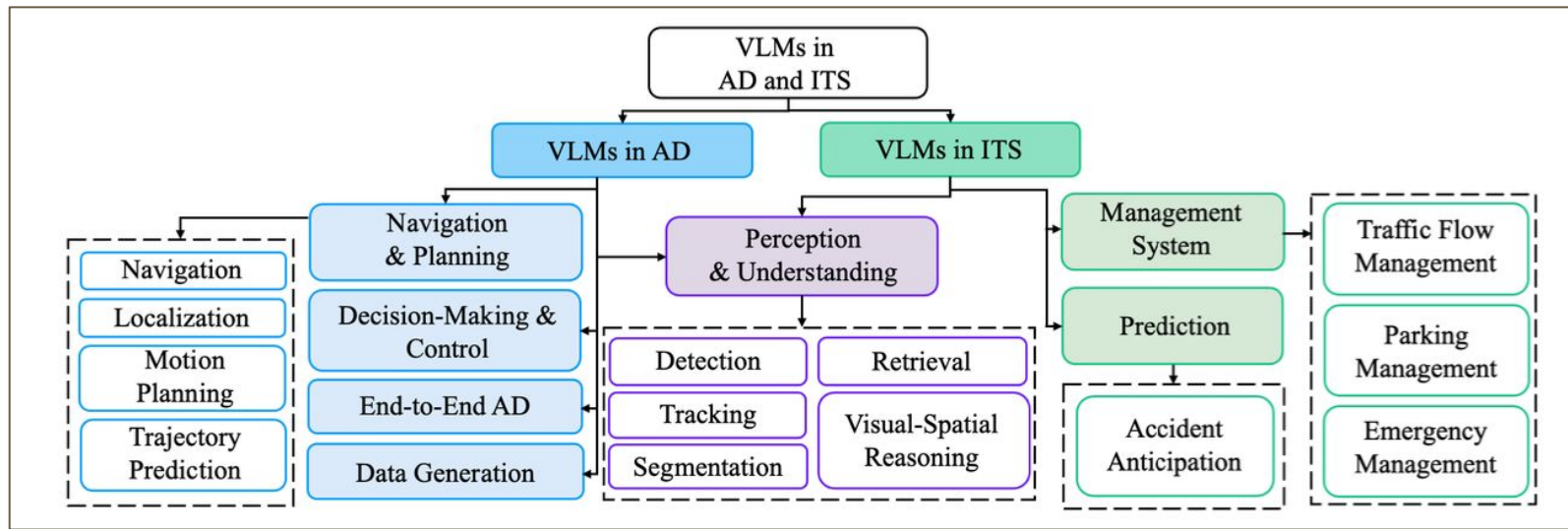
— B09901142 電機四 呂睿超 —

Introduction

- Background:
 - Language Models are critical in enhancing the understanding and interpretation capabilities in Intelligent vehicles
- Objective:
 - To explore and summarize the current state of language models in intelligent vehicles, identifying key applications, challenges, and future directions.

Literature Review (1/2)

- Zhou, X.,... (2023). Vision-Language Models in Autonomous Driving(AD) and Intelligent Transportation Systems(ITS)
 - This paper surveys existing research regarding VLM usage in AD & ITS



Literature Review (2/2)

- ZhenJie,Y. ,...(2023). LLM4Drive: A Survey of Large Language Models for Autonomous Driving
 - This paper surveys existing research regarding LLMs in autonomous driving, focusing on advancements, applications, and challenges.

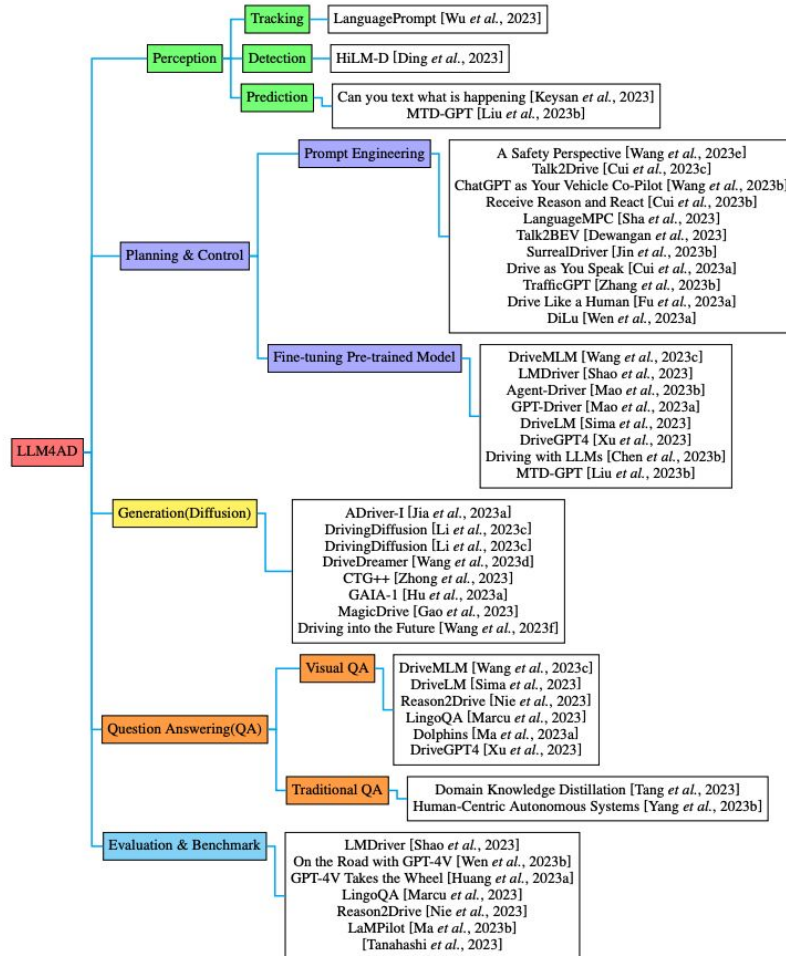


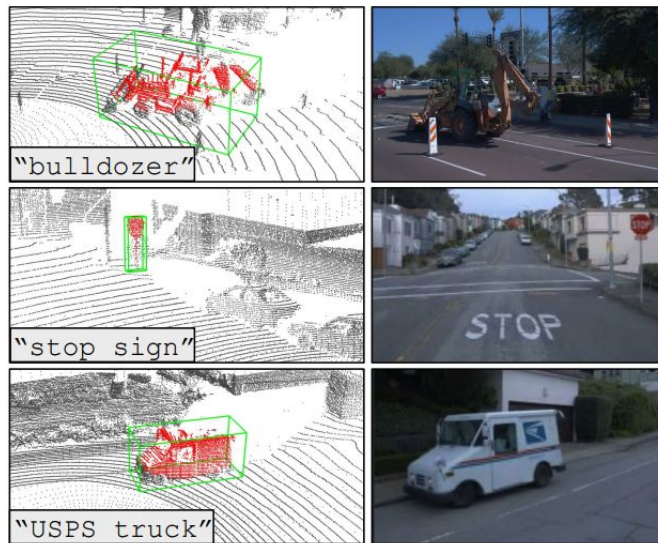
Figure 3: Large Language Models for Autonomous Driving Research Tree

How to conduct the survey?

- Research on some papers in each section below
 - Perception and Understanding
 - Planning and Control
 - Dataset
 - Evaluation
- And compare with papers that have tried to solve end to end autonomous driving
 - e.g. DriveGPT4: Interpretable End-to-end Autonomous Driving via Large Language Model

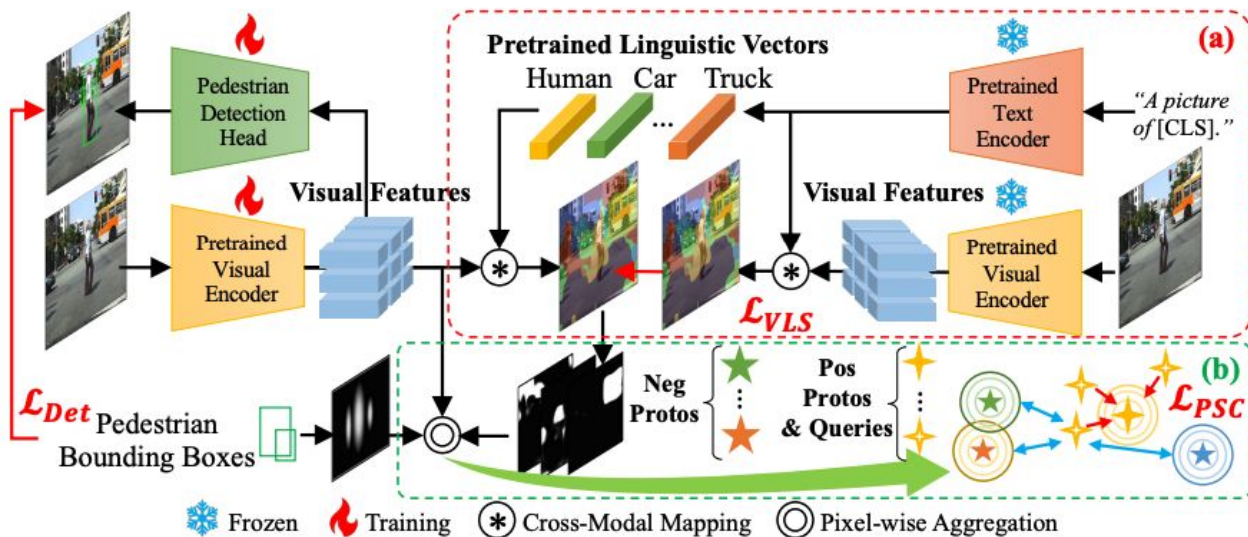
Perception and Understanding(1/2)

- Object detection : Utilizing VLM, we can conduct open-vocabulary detection and segmentation - even on unseen objects
 - e.g. Unsupervised 3D Perception with 2D Vision-Language Distillation for Autonomous Driving



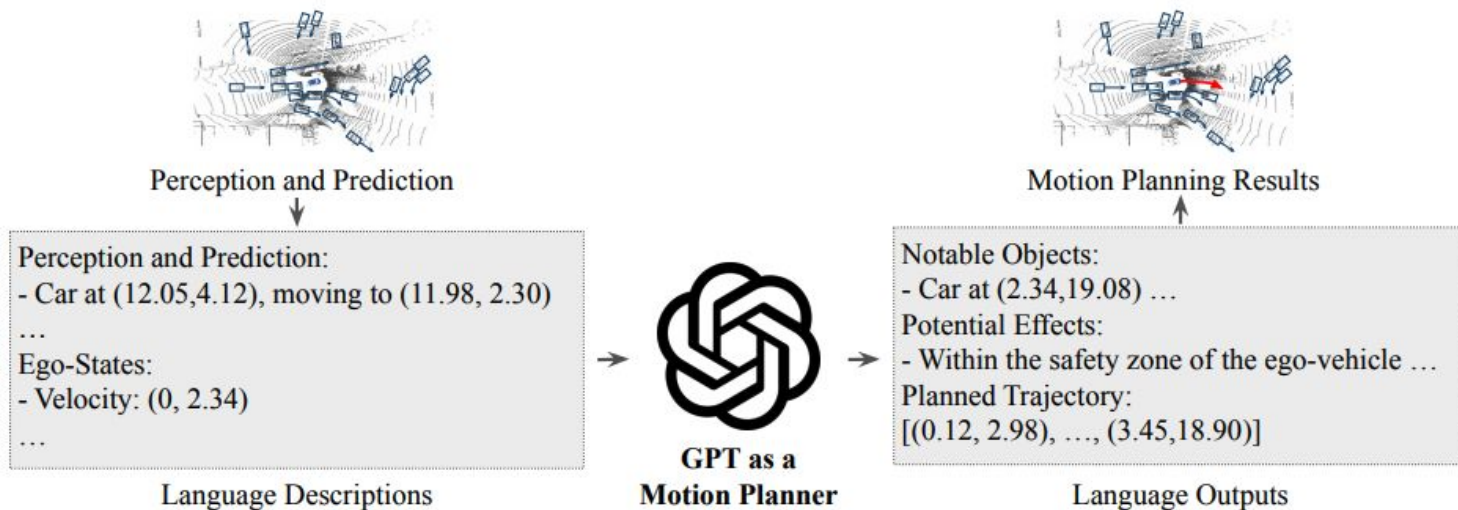
Perception and Understanding (2/2)

- pedestrian detection: enhance the human-like object confusion and achieve unsupervised/ annotation-free method
 - e.g. VLPD: Context-Aware Pedestrian Detection via Vision-Language Semantic Self-Supervision



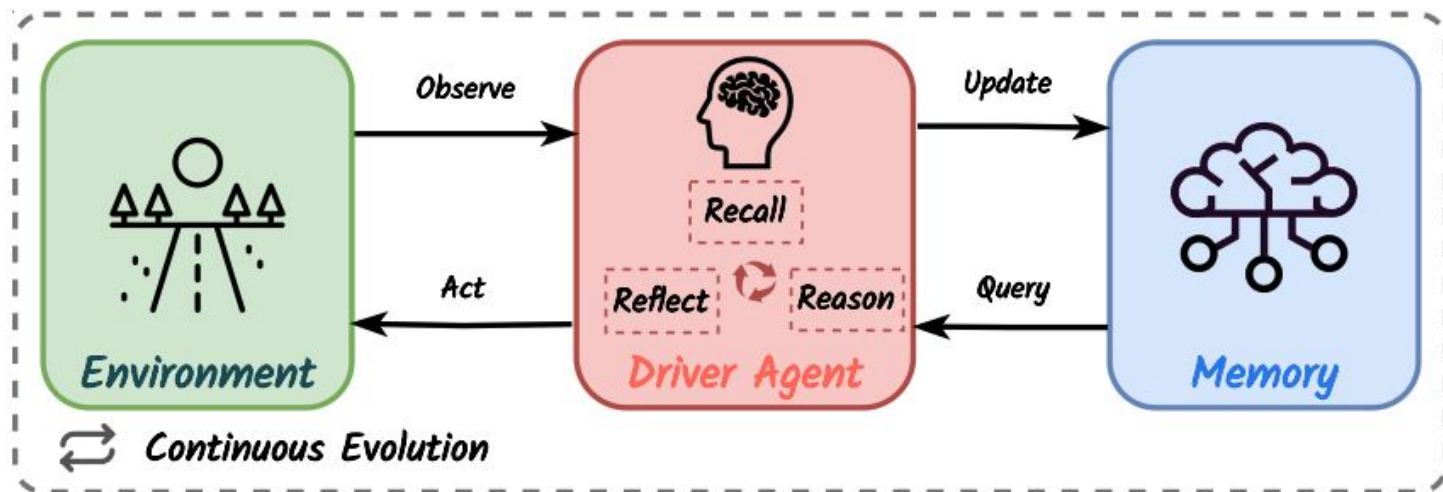
Planning and Control (1/2)

- Motion planning → Language model problem
 - e.g. GPT-Driver: Learning to Drive with GPT



Planning and Control (2/2)

- Memory module to record experience, to leverage LLMs to facilitate reasoning and re-flection
 - e.g. DiLU: A Knowledge-Driven Approach to Autonomous Driving with Large Language Models



Dataset

- Vision Question Answering (VQA): These datasets enhances the understanding of different driving scenarios
 - NuScense-QA: Visual Question Answering for autonomous driving scenarios.
 - DRAMA: Joint risk localization with visual reasoning of driving risks in a free-form language description.
- Object Detecting Dataset
 - KITTI: Benchmark for various autonomous driving tasks.
 - Cityscapes: Annotated urban street scenes for object detection and segmentation.

Evaluation

- Try to use VLMs such as GPT-4V to conduct comprehensive evaluation
 - On the Road with GPT-4V
 - LingoQA

Prompt:



Describe the condition of all the traffic lights in the image, if there are countdowns, please describe them together.

GPT-4V:

The image has two red traffic lights, both without any visible countdowns.

Prompt:

How about the yellow lights in the image?

GPT-4V:

In addition to the red traffic lights, there are two yellow streetlights illuminating the road, but they are not traffic signals.

On the Road with GPT-4V

Action Justification



Q: What is your next action and why?

A: Slowing down, pedestrians crossing the road.

LingoQA

What other to do in the survey

- Challenge and research gap
- Comparison
- Survey of most recent papers