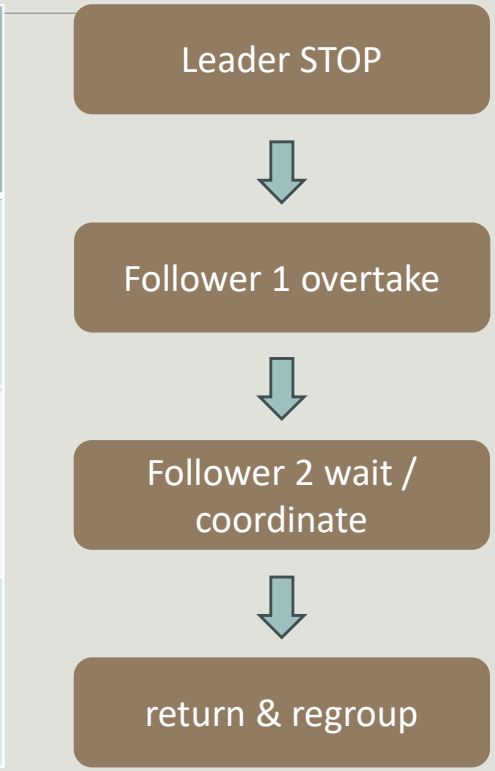
基於 CARLA 的多智能體強化學習車隊換道決策方法研究
Multi-Agent Reinforcement Learning for Cooperative Platoon Lane-Change Decision in CARLA

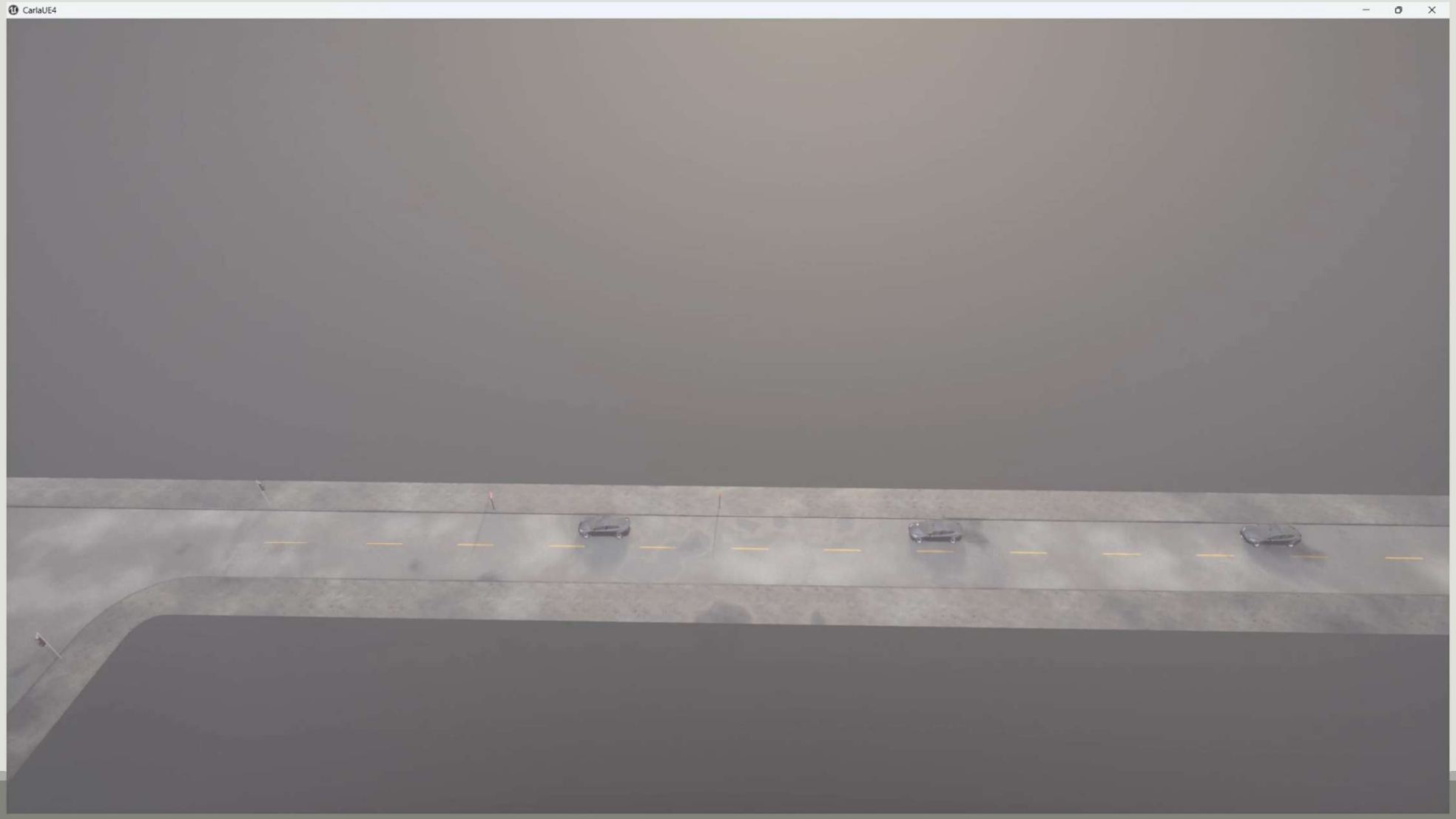
資訊系115 陳奕嘉 E94115011

研究背景、缺口、目標

區塊	內容
背景	自動駕駛不只需要單車決策，在車隊行駛中，後方車輛也必須根據前車狀態進行協同反應。
缺口	Leader 停止後的完整超車、返回、重新編隊較少被處理
目標	用 CARLA + MARL-CTDE 學習兩台 followers 的協同決策



問題情境 — Leader 突然停止造成阻塞


CarlaUE4

問題複雜度：三個動作形成指數成長的決策空間

Action space |A| = 3
0: Follow
1: Left overtake
2: Right return

$N_{sequence} = |A|^T = 3^T$

|A| = 可選 action 數量
T = decision steps

Decision steps T	Possible action sequences 3^T
50	7.18×10^{23}
100	5.15×10^{47}
200	2.66×10^{95}

本研究透過 action sequence space 說明：
即使 action space 很小，多個時間步的**連續決策(**sequential decision-making**)**仍會形成大量可能的**行為序列(**action sequences**)**。

核心方法：MARL-CTDE 車隊協同決策

	Single-agent PPO	MARL-CTDE platoon
設定	1 leader + 1 follower	1 leader + 2 followers
目標	驗證超車決策可學習	學習兩台 followers 的協同超車
協同行為	無，單車決策	follower_1 先超車，follower_2 等待安全時機
Action space	3 actions	joint action space = $3^2 = 9$
Sequence space	3^T	9^T
定位	Baseline	專題核心方法

MARL 系統架構：Shared-policy PPO with CTDE

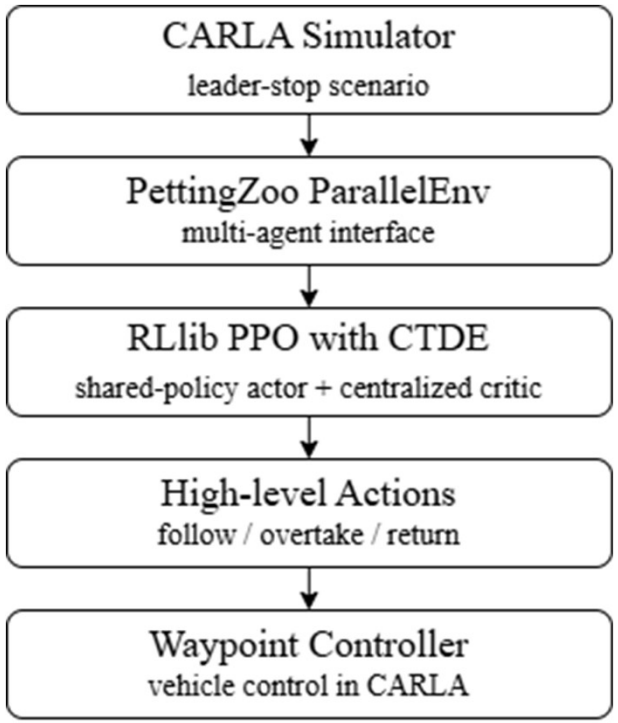


Fig. MARL Platoon Control with CTDE.

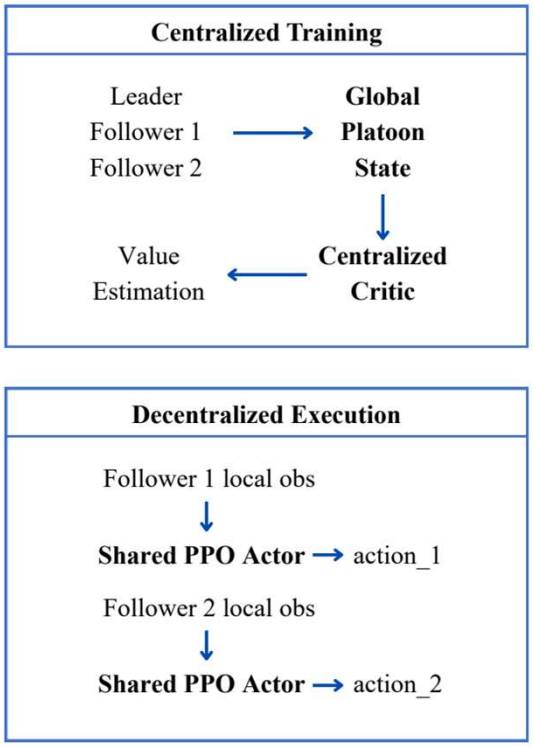


Fig. CTDE design

訓練看全局，執行靠各車局部決策。

Baseline：Single-agent PPO 驗證核心超車決策

系統架構

CARLA 模擬環境
↓
Gymnasium Env
↓
PPO Agent
↓
High-level Action
↓
Rule-based Waypoint Controller
↓
CARLA Vehicle Control

PPO 決策設計

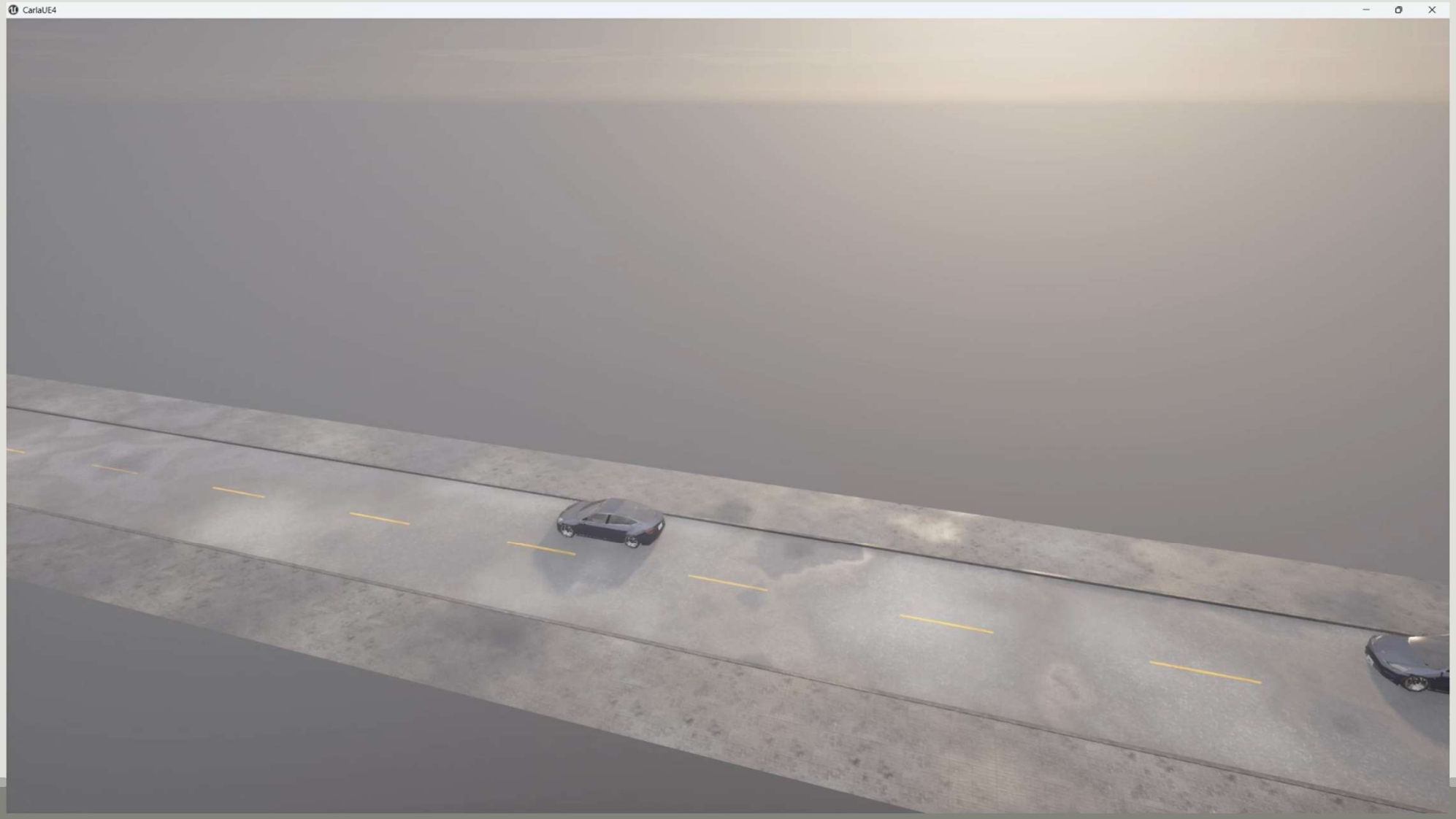
Observation
- distance to leader
- follower / leader speed
- leader stopped time
- lane availability / clearance
- lane position

Action space |A| = 3
0：Follow
1：Left overtake
2：Right return

Baseline 的意義

Purpose
- 驗證 stopped-leader overtaking 是否可學習
- PPO 學習「何時超車、何時返回」
- 作為後續 MARL platoon 的 baseline

Single-agent PPO 先驗證核心決策可學習；MARL 再處理兩台 followers 的協同問題。


CarlaUE4

Baseline Result：Single-Agent PPO 驗證任務可學習

Training summary:

245 training episodes
Overall success rate: 82.4%
Final 100 episodes: 100.0%
Collision rate: 2.0%

模型在訓練後期已趨於穩定。

Evaluation summary:

Test condition	Episodes	Success rate	Collision rate	Mean reward	Mean length
Fixed stop step = 40, speed = 15	10	100%	0%	133.69	203
Random stop timing	20	100%	0%	130.66 ± 1.62	291.9
Fixed stop step = 40, speed = 12	10	100%	0%	133.68	207

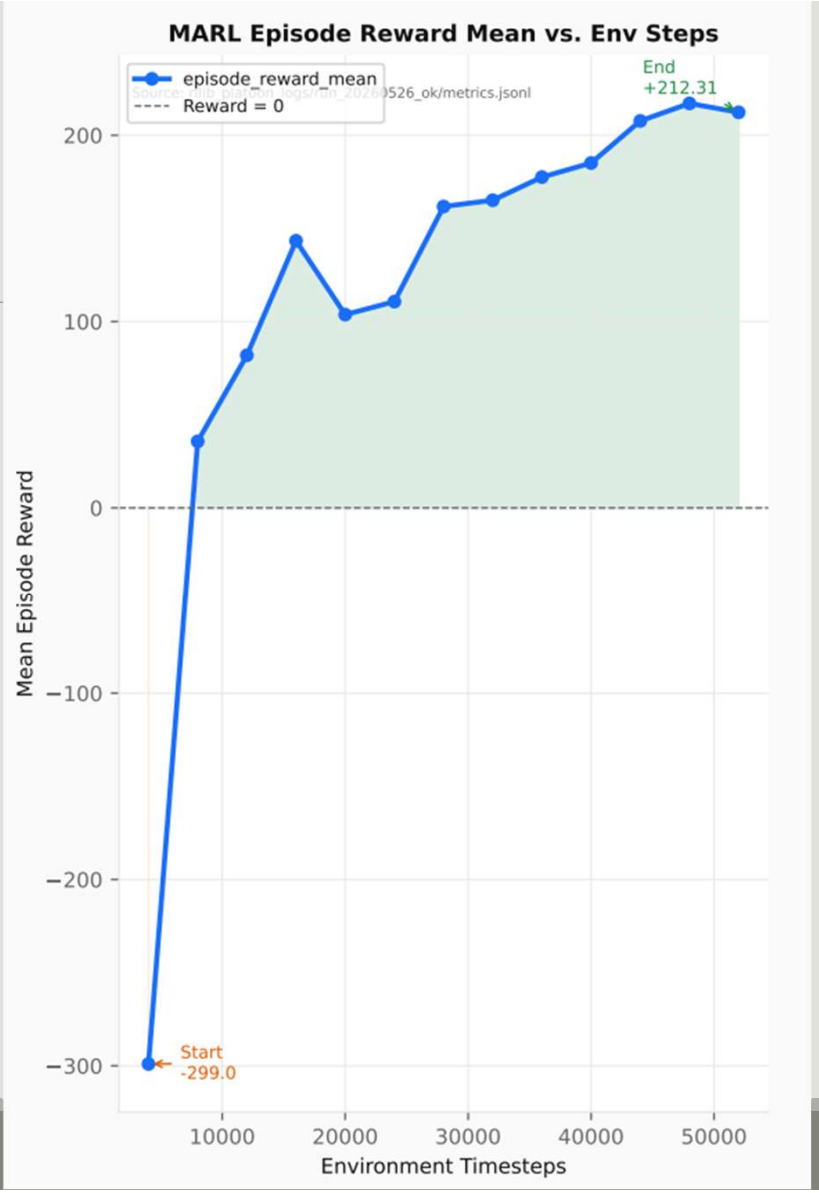
結果亮點：
模型在固定與隨機停止時間下皆能成功，表示其學到的是根據狀態決策，
而非記憶固定停止時機。

MARL Training Result：兩台 RL Followers 的協同超車學習

Setup
- 1 leader + 2 RL followers
- Shared-policy PPO with centralized critic
- 52k environment steps, 13 PPO iterations

Main Result
- Mean episode reward: −299.03 → +212.31
- Coordination conflicts: 0 recorded

Key Interpretation
- Reward 由負轉正，表示模型從失敗/逾時行為轉向有效任務完成
- 兩台 followers 已初步學到依序超車與避免衝突



結論與未來期望

結論

- 本專題建立了一個基於 CARLA 的 stopped-leader overtaking 任務，用於研究自動駕駛車隊在前車停止時的超車決策。
- Single-agent PPO baseline 驗證了核心超車決策是可以被學習的。
- MARL-CTDE 架構進一步將任務延伸到車隊協同決策，並結合 shared-policy PPO 與 centralized critic，使兩台 followers 能學習協同超車行為。

未來期望

- 提升模型在不同 CARLA 道路、出生點與 waypoint 拓樸下的穩健性。
- 改善 MARL 的重新編隊行為。
- 將系統從 2 台 followers 擴展到 n 台 followers。

感謝聆聽