

113-049-5521
IOT-112-IC002

我國人工智慧車聯網之 號誌控制(1/2)-都會區幹道實作 與交流道區域模式發展探討



交通部運輸研究所

中華民國 113 年 7 月

113-049-5521
IOT-112-IC002

我國人工智慧車聯網之 號誌控制(1/2)-都會區幹道實作 與交流道區域模式發展探討

著者：蕭偉政、許添本、謝秉均、董尚義、陳煒騰、黃業傑、
林小淳、郭哲維、黃祖瑜、吳東凌、周家慶

交通部運輸研究所

中華民國 113 年 7 月

國家圖書館出版品預行編目(CIP)資料

我國人工智慧車聯網之號誌控制. (1/2): 都會區幹道實
作與交流道區域模式發展探討/蕭偉政, 許添本, 謝秉
均, 董尚義, 陳煒騰, 黃業傑, 林小淳, 郭哲維, 黃祖瑜,
吳東凌, 周家慶著. -- 初版. -- 臺北市: 交通部運輸研
究所, 民 113.07

面; 公分

ISBN 978-986-531-614-3(平裝)

1.CST: 人工智慧 2.CST: 交通號誌 3.CST: 標示系統
4.CST: 交通管理

557.822

113009631

我國人工智慧車聯網之號誌控制(1/2)-
都會區幹道實作與交流道區域模式發展探討

著 者: 蕭偉政、許添本、謝秉均、董尚義、陳煒騰、黃業傑、林小淳、
郭哲維、黃祖瑜、吳東凌、周家慶

出版機關: 交通部運輸研究所

地 址: 105004 臺北市松山區敦化北路 240 號

網 址: www.iot.gov.tw (中文版>數位典藏>本所出版品)

電 話: (02)2349-6789

出版年月: 中華民國 113 年 7 月

印 刷 者: 全凱數位資訊有限公司

版(刷)次冊數: 初版一刷 57 冊

本書同時登載於交通部運輸研究所網站

定 價: 280 元

展 售 處:

交通部運輸研究所運輸科技及資訊組・電話: (02)2349-6789

五南文化廣場: 400002 臺北市中區中山路 6 號・電話: (04)2226-0330

國家書店松江門市: 104472 臺北市中山區松江路 209 號・電話: (02)2518-0207

GPN: 1011300858 ISBN: 978-986-531-614-3 (平裝)

著作財產權人: 中華民國(代表機關: 交通部運輸研究所)

本著作保留所有權利, 欲利用本著作全部或部分內容者, 須徵求交通部運輸研究所
書面授權。

交通部運輸研究所合作研究計畫出版品摘要表

出版品名稱：我國人工智慧車聯網之號誌控制(1/2)-都會區幹道實作與交流道區域模式發展探討			
國際標準書號（或叢刊號） ISBN 978-986-531-614-3(平裝)	政府出版品統一編號 1011300858	運輸研究所出版品編號 113-049-5521	計畫編號 112-IC002
本所主辦單位：運輸科技及資訊組 主管：吳東凌 計畫主持人：吳東凌 研究人員：周家慶 聯絡電話：(02)23496756 傳真號碼：(02)25450426	合作研究單位：資拓宏宇國際股份有限公司 計畫主持人：蕭偉政 研究人員：許添本、謝秉均、董尚義、陳煒騰、黃業傑、林小淳、郭哲維、黃祖瑜 地址：22041 新北市板橋區縣民大道 2 段 7 號 6 樓 聯絡電話：(02) 8969-1969		研究期間 自 112 年 4 月 至 112 年 12 月
關鍵詞：AI 號誌控制；深度強化學習；適應性號誌控制；微觀車流模擬			
摘要： 近年來人工智慧在軟硬體技術的突飛猛進與各領域應用的迅速發展，可預期未來運用人工智慧、影像辨識、資通訊(ICT)、車聯網(V2X)與 5G 等技術，特別有助於紓緩因交通號誌控制不夠智慧，導致民眾行的痛點。本計畫延續 111 年「我國人工智慧車聯網之號誌控制模式探討」之研究計畫，進行與縣市政府合作之現場實驗測試，實作測試過程中視需求持續精進人工智慧號誌控制模式，透過強化學習通用環境之探討與建置、導入多目標強化學習演算法、以及探討車流模擬測試模型於各種車流場景之泛用性，期望能提升人工智慧號誌控制模式之運作效率。此外，同時納入高速公路匝道儀控之人工智慧號誌控制課題與相關模擬模式發展探討，優先構建交流道區域號誌協控車流模擬環境、以及進行分散式 AI 號誌協控模型(多代理人機制)初步發展設計，以期逐步發展我國在人工智慧號誌控制之實力構建，期能支援我國都市智慧號誌控制在人工智慧與車聯網環境下之車流運作效率及交通安全提升。 本計畫在人工智慧強化學習號誌控制模型優化與精進上，能依據各路口各方向交通變化，即時計算下個階段的號誌時制內容；在臺南市的「台 86-19 甲」單一路口案例，即時計算下一個時相號誌綠燈秒數，在臺北市的「中山北路-德行東路」3 個路口幹道則以單一代理人控制 3 個路口，以共同號誌周期長度與不同時比方式來進行動態調整，以兼顧車流續進。實測結果顯示，臺南市及臺北市實驗場域依據不同偏好進行決策，且在各個時段中均能找到一組偏好，大部分的績效比固定時制為佳；另多任務 AI 強化學習模式模型即使將不同時段資料一起進行訓練，雖然尚無法比分開訓練的模型績效為佳，但整體表現已接近。在交流道區域的匝道儀控與平面路口強化學習號誌協控模型發展部分，遴選桃園楊梅交流道來構建模擬實證場域，導入多號誌控制代理人機制，並於績效評估上額外考量駕駛人感受等相關指標，將匝道和平面道路的輸入資料集中到一個中央化代理人(agent)的「中央化訓練(Centralized Training)」，本計畫於 113 年將進行強化學習號誌協控模型的學習訓練與實測工作。			
出版日期	頁數	定價	本出版品取得方式
113 年 7 月	234	280	凡屬機密性出版品均不對外公開。普通性出版品，公營、公益機關團體及學校可函洽本所免費贈閱；私人及私營機關團體可按定價價購。
備註：本研究之結論與建議不代表交通部之意見。			

PUBLICATION ABSTRACTS OF RESEARCH PROJECTS
INSTITUTE OF TRANSPORTATION
MINISTRY OF TRANSPORTATION AND COMMUNICATIONS

TITLE: Study of Artificial Intelligence Traffic Signal Control - Urban Main Roads and Highway Interchange Areas Implementation			
ISBN (OR ISSN) ISBN 978-986-531-614-3 (pbk.)	GOVERNMENT PUBLICATIONS NUMBER 1011300858	IOT SERIAL NUMBER 113-049-5521	PROJECT NUMBER 112-IC002
DIVISION: Transportation Technology and Information Systems Division DIVISION DIRECTOR: Tun-Ling Wu PRINCIPAL INVESTIGATOR: Tun-Ling Wu PROJECT STAFF: Ja-Ching Chou PHONE: 886-2-23496756 FAX: 886-2-25450426			PROJECT PERIOD FROM April 2023 TO December 2023
RESEARCH AGENCY: International Integrated Systems, Inc. PRINCIPAL INVESTIGATOR: Wei-Cheng Hsiao PROJECT STAFF: Tien-Pen Hsu, Ping-Chun Hsieh, Shang-Yee Tung, Wei-Teng Chen, Yeh-Chieh Huang, Hsiao Chun Lin, Che-Wei Kuo, Tsu-Yu Huang ADDRESS: 6F., No. 7, Sec. 2, Xianmin Blvd., Banqiao Dist., New Taipei City 22041, Taiwan (R.O.C.) PHONE: (02) 8969-1969			
KEY WORDS: Artificial Intelligence and Adaptive Traffic Signal Control; Deep Reinforcement Learning; Microscopic Simulation			
ABSTRACT: In recent years, the rapid development of artificial intelligence in software and hardware technologies and its widespread application across various domains has led to significant progress. We can anticipate that future utilization of technologies such as artificial intelligence, image recognition, information and communication technology (ICT), vehicle-to-everything (V2X), and 5G will play a crucial role in alleviating the pain points caused by incapable intelligent traffic signal control. Building upon the research project titled "Exploration of Artificial Intelligence Vehicle-to-Everything Signal Control Models in Taiwan" conducted in 2022, this project aims to collaborate with local governments for on-site experimental testing. During the implementation of these tests, the artificial intelligence signal control model can be continuously refined based on practical experience. We explore and establish a robust environment for reinforcement learning, introduce multi-objective reinforcement learning algorithms, and investigate the versatility of traffic flow simulation models across various scenarios. The ultimate goal is to enhance the operational efficiency of artificial intelligence signal control models. Additionally, we address the challenges related to artificial intelligence signal control for freeway ramps. We prioritize the construction of a simulated environment for interchange area signal coordination and the preliminary development of a decentralized AI signal coordination model using a multi-agent mechanism. Through these efforts, we aim to gradually strengthen Taiwan's capabilities in artificial intelligence signal control and contribute to improving traffic efficiency and safety in urban environments under the context of artificial intelligence and vehicular networking. This project focuses on optimizing and enhancing artificial intelligence (AI) reinforcement learning models for traffic signal control. It can calculate the signal timing for the next phase in real-time based on traffic changes at each intersection. For example, in real-time, at the 'Tai 86-19 Jia' intersection in Tainan City, it calculates the green light duration for the next phase. In Taipei City, at the 'Zhongshan North Road - Dexing East Road' corridor, a single agent controls three intersections, dynamically adjusting the common signal cycle length and phase ratios to ensure continuous traffic flow. Field tests in Tainan and Taipei have shown that decisions based on different preferences can be made, and for most time periods, a set of preferences can be found that performs better than fixed-time controls. Additionally, the multi-task AI reinforcement learning model may not outperform models trained separately, even when trained with data from different time periods, but its overall performance is close. In the development of ramp metering and surface intersection reinforcement learning signal coordination models for interchange areas, Yangmei Interchange in Taoyuan was selected as the testing site. A multi-signal control agent mechanism is introduced. Performance evaluation considers additional indicators related to driver perception, and the 'Centralized Training' approach centralizes input data from ramps and surface roads to a single central agent. The trained machine learning model is then applied to construct a simulated experimental field at the Yangmei Interchange in Taoyuan. This reinforcement learning-based coordination model for interchange areas in Taoyuan will be further trained and tested in 2024.			
DATE OF PUBLICATION July 2024	NUMBER OF PAGES 234	PRICE 280	
The views expressed in this publication are not necessarily those of the Ministry of Transportation and Communications.			

目錄

目錄	III
圖目錄	V
表目錄	IX
第一章 計畫概述	1
1.1 計畫背景與目的	1
1.2 計畫工作流程	2
1.3 研究內容與工作項目	2
第二章 文獻回顧與關鍵課題	5
2.1 國內外號誌控制策略探討	5
2.1.1 平面路口之號誌控制	5
2.1.2 交流道之號誌協控	10
2.2 強化學習於號誌控制之應用	16
2.2.1 強化學習於平面路口之號誌控制	16
2.2.2 強化學習應用於交流道協控	19
2.2.3 我國應用人工智慧或車聯網於號誌控制相關研究	24
2.3 人工智慧強化學習(Reinforcement learning, RL)模式	28
2.3.1 強化學習架構	28
2.3.2 強化學習演算法	30
2.3.3 小結	50
2.4 人工智慧強化學習在號誌控制應用之關鍵課題	50
第三章 強化學習號誌控制模型探討	53
3.1 實測號誌控制模型檢討	53
3.2 號誌控制模型優化與精進	59
3.2.1 AI 強化學習模型精進架構	59
3.2.2 AI 強化學習模型訓練模擬環境程式與軟硬體說明	68
3.2.3 AI 強化學習模型訓練結果	69
第四章 強化學習號誌控制實測	81
4.1 實驗場域規劃與分析	81
4.1.1 單一路口場域(臺南市「台 86-19 甲線」路口)	81
4.1.2 幹道多路口場域(臺北市「中山北路-德行東路」)	84
4.2 人工智慧強化學習號誌控制核心運作模組	87
4.3 實測計畫	92
4.3.1 車輛偵測器驗證計畫	92
4.3.2 模擬測試	96

4.3.3 實測計畫.....	135
4.3.4 實測情形.....	137
4.4 事前事後量化績效評估.....	141
4.4.1 臺南市實驗場域實測結果.....	141
4.4.2 臺北市實驗場域實測結果.....	146
4.4.3 實測分析.....	151
4.5 小結.....	152
第五章 強化學習號誌協控模擬環境構建規劃.....	155
5.1 強化學習號誌協控型態探討.....	155
5.2 強化學習號誌協控模式發展構想.....	156
5.3 實驗場域規劃與分析.....	158
5.4 實驗場域模擬模式構建規劃.....	164
5.4.1 交通參數與績效指標設計.....	164
5.4.2 交通參數資料蒐集規劃.....	165
5.4.3 模擬模式建構設計.....	167
5.4.4 模擬模式校估驗證.....	170
5.5 小結.....	176
第六章 結論與建議.....	177
6.1 結論.....	177
6.2 建議.....	178
參考文獻.....	181
附錄一 期中審查意見回覆表.....	185
附錄二 期末審查意見回覆表.....	203

圖目錄

圖 1.2-1 計畫執行流程	2
圖 2.1.1-1 設置逆向左轉車道路口以及路口時相示意圖	6
圖 2.1.1-2 單方向設置逆向左轉車道路口圖以及路口時相示意圖	7
圖 2.1.1-3 半觸動號誌邏輯	7
圖 2.1.1-4 適應性號誌控制於多路口車流推進之分段示意圖	10
圖 2.1.2-1 匝道儀控與速率控制協控系統	11
圖 2.1.2-2 主線速率控制系統設施設置說明圖	11
圖 2.1.2-3 控制邏輯架構圖	12
圖 2.1.2-4 模擬求解流程圖	13
圖 2.1.2-5 上匝道與都市間協作模式流程圖	15
圖 2.1.2-6 下匝道與都市間協作模式流程圖	15
圖 2.2.2.1-1 TTS result	20
圖 2.2.2.1-2 SD(TWT) result	20
圖 2.2.2.1-3 高速公路主線與匝道協控策略模型示意圖	21
圖 2.2.2.1-4 狀態輸入示意圖	21
圖 2.2.2.1-5 主線儀控率	22
圖 2.2.2.1-6 匝道儀控率	22
圖 2.2.2.1-7 主線速率流量關係圖	22
圖 2.2.2.2-1 CTDE 多智能體強化學習架構圖	23
圖 2.2.2.2-2 MADDPG 之 CTDE 架構示意圖	24
圖 2.3.1-1 強化學習概念示意圖	29
圖 2.3.1-2 號誌控制架構基本模型	30
圖 2.3.2-1 DDPG 運作架構	34
圖 2.3.2-2 DDPG 演算法 (https://spinningup.openai.com/en/latest/algorithms/ddpg.html)	34
圖 2.3.2-3 RL 代理人類型	36
圖 2.3.2-4 PPO 的目標式 L^{clip} 與 $r(\theta)$ 的關係圖	37
圖 2.3.2-5 PPO 運作架構圖	37
圖 2.3.2-6 PPO 演算法 (https://spinningup.openai.com/en/latest/algorithms/ppo.html)	38
圖 2.3.2-7 SAC 演算法 (https://spinningup.openai.com/en/latest/algorithms/sac.html)	39
圖 2.3.2-8 兩目標馬可夫決策過程之 Pareto front 與 CCS 示意圖	41
圖 2.3.2-9 PGMORL 之搜索概念示意圖	42
圖 2.3.2-10 Conditioned Network 架構示意圖	42
圖 2.3.2-11 Q-Pensieve 運作架構圖	44
圖 2.3.2-12 Q-Pensieve 演算法	45
圖 2.3.2-13 強化學習演算法之歸納與衍生	45
圖 2.3.2-14 單模態任務目標參數示意圖	47
圖 2.3.2-15 多模態任務目標參數示意圖	47
圖 2.3.2-16 MMAML 架構示意圖	48

圖 2.3.2-17 MMAML 訓練虛擬碼	49
圖 2.3.2-18 多任務強化學習訓練架構	50
圖 3.1-1 前期代理人 AI 強化學習模型模型結構圖	53
圖 3.1-2 臺北市實驗場域前期模型號誌化路口分布圖	53
圖 3.1-3 X2 路口現場端時制示意圖	55
圖 3.1-4 X4 路口現場端時制示意圖	55
圖 3.1-5 X6 路口現場端時制示意圖	56
圖 3.1-6 X6 路口現場端時制示意圖	56
圖 3.1-7 X2 路口臺北市交控中心修改時制示意圖	57
圖 3.1-8 X2 路口編號 CF 與 F0 時制比較圖	57
圖 3.1-9 臺北市實驗場域中山北路六段/克強路設備偵測範圍	58
圖 3.1-10 臺北市實驗場域中山北路六段/195 巷設備偵測範圍	58
圖 3.1-11 臺北市實驗場域中山北路六段/德行西路設備偵測範圍	58
圖 3.2.1-1 本期臺北市實驗場域 Policy 模型結構圖	62
圖 3.2.1-2 本期臺北市實驗場域 Critic 模型結構圖	63
圖 3.2.1-3 本期臺南市實驗場域 Policy 模型結構圖	63
圖 3.2.1-4 本期臺北市實驗場域 Critic 模型結構圖	64
圖 3.2.1-5 本期 Q-Pensieve 代理人模型訓練架構圖	66
圖 3.2.3-1 臺北市實驗場域全時段 AI 強化學習模型訓練指標	69
圖 3.2.3-2 臺北市實驗場域平日晨峰各偏好績效彙整圖	70
圖 3.2.3-3 臺北市實驗場域平日離峰各偏好績效彙整圖	71
圖 3.2.3-4 臺北市實驗場域平日昏峰各偏好績效彙整圖	71
圖 3.2.3-5 臺北假日昏峰各偏好績效彙整圖	72
圖 3.2.3-6 臺南市實驗場域全時段訓練指標	73
圖 3.2.3-7 臺南市實驗場域平日晨峰(7:00-8:00)各偏好績效彙整圖	74
圖 3.2.3-8 臺南市實驗場域平日晨峰(8:00-9:00)各偏好績效彙整圖	74
圖 3.2.3-9 臺南市實驗場域平日離峰各偏好績效彙整圖	75
圖 3.2.3-10 臺南市實驗場域平日昏峰各偏好績效彙整圖	75
圖 3.2.3-11 臺南市實驗場域假日昏峰各偏好績效彙整圖	76
圖 3.2.3-12 臺南市實驗場域多任務 AI 強化學習模型平日離峰驗證紀錄	77
圖 3.2.3-13 臺南市實驗場域多任務 AI 強化學習模型平日昏峰驗證紀錄	77
圖 3.2.3-14 臺南市實驗場域多任務 AI 強化學習模型平日離峰各偏好績效彙整圖	78
圖 3.2.3-15 臺南市實驗場域多任務 AI 強化學習模型平日昏峰各偏好績效彙整圖	78
圖 4.1.1-1 臺南市「台 86-19 甲線」路口	81
圖 4.1.2-1 臺北市「中山北路-德行東路」幹道	84
圖 4.1.2-2 德行西路、195 巷及克強路路口道路幾何示意圖	84
圖 4.2-1 號誌控制核心運作模組架構	88
圖 4.2-2 AI 智慧號誌控制系統執行流程	88
圖 4.2-3 資料串接架構規劃	89
圖 4.2-4 實測監控畫面-GIS 監控圖台	90

圖 4.2-5 實測監控畫面-系統管理.....	90
圖 4.2-6 實測監控畫面-「智慧號誌控制運行時制計畫」查詢頁面.....	91
圖 4.2-7 實測監控畫面-「車流偵測資料」查詢頁面.....	91
圖 4.2-8 實測監控畫面-「系統運作異常紀錄」查詢頁面.....	91
圖 4.3.1-1 臺南市「台 86-19 甲線」偵測設備位置.....	92
圖 4.3.1-2 臺北市「中山北路-德行東路」偵測設備位置.....	95
圖 4.3.2.1-1 VISSIM 模擬路網示意圖.....	97
圖 4.3.2.1-2 VISSIM 實際路況與模擬影片對比示意圖.....	97
圖 4.3.2.1-3 SUMO 路網編輯介面示意圖.....	98
圖 4.3.2.2-1 模擬結果評估流程示意圖.....	99
圖 4.3.2.3-1 平日晨峰 5 時相本期 AI 強化學習模型模擬各分相秒數穩定性分析.....	102
圖 4.3.2.3-2 本期 AI 強化學習模型於平日晨峰 5 時相各分相秒數合理性分析.....	102
圖 4.3.2.3-3 本期 AI 強化學習模型於平日晨峰 4 時相模擬各分相秒數穩定性分析.....	103
圖 4.3.2.3-4 本期 AI 強化學習模型於平日晨峰 4 時相模擬各分相秒數合理性分析.....	104
圖 4.3.2.3-5 本期 AI 強化學習模型於平日離峰模擬各分相秒數穩定性分析.....	105
圖 4.3.2.3-6 本期 AI 強化學習模型於平日離峰模擬各分相秒數合理性分析.....	105
圖 4.3.2.3-7 本期 AI 強化學習模型於平日昏峰模擬各分相秒數穩定性分析.....	106
圖 4.3.2.3-8 本期 AI 強化學習模型於平日昏峰模擬各分相秒數合理性分析.....	107
圖 4.3.2.3-9 本期 AI 強化學習模型於假日昏峰模擬各分相秒數穩定性分析.....	108
圖 4.3.2.3-10 本期 AI 強化學習模型於假日昏峰模擬各分相秒數合理性分析.....	109
圖 4.3.2.4-1 中山北路/克強路平日晨峰模擬合理性分析.....	112
圖 4.3.2.4-2 中山北路/195 巷平日晨峰模擬合理性分析.....	112
圖 4.3.2.4-3 中山北路/德行西路平日晨峰模擬合理性分析.....	113
圖 4.3.2.4-4 中山北路/克強路平日離峰模擬合理性分析.....	117
圖 4.3.2.4-5 中山北路/195 巷平日離峰模擬合理性分析.....	117
圖 4.3.2.4-6 中山北路/德行西路平日離峰模擬合理性分析.....	118
圖 4.3.2.4-7 中山北路/克強路平日昏峰模擬合理性分析.....	122
圖 4.3.2.4-8 中山北路/195 巷平日昏峰模擬合理性分析.....	123
圖 4.3.2.4-9 中山北路/德行西路平日昏峰模擬合理性分析.....	123
圖 4.3.2.4-10 中山北路/克強路假日昏峰模擬合理性分析.....	127
圖 4.3.2.4-11 中山北路/195 巷假日昏峰模擬合理性分析.....	128
圖 4.3.2.4-12 中山北路/德行西路假日昏峰模擬合理性分析.....	128
圖 4.3.4-1 台 86 線往國 3 方向車流狀況.....	138
圖 4.3.4-2 台 86 線往國 3 方向車流狀況.....	138
圖 4.3.4-3 中山北路北往南方向「忠誠路-德行東路」.....	139
圖 4.3.4-4 克強路右轉中山北路北往南方向車流.....	140
圖 4.3.4-5 中山北路六段/克強路口車流狀況.....	140
圖 4.4.1-1 平、假日 AI 臺南場域實測各分相綠燈秒數趨勢圖.....	142
圖 4.4.1-2 平日 11/7&11/14 AI 號誌控制與 TOD 旅行時間績效示意圖.....	145
圖 4.4.1-3 平日 11/8&11/15 AI 號誌控制與 TOD 旅行時間績效示意圖.....	145

圖 4.4.1-4 假日 AI 號誌控制與 TOD 旅行時間績效示意圖	146
圖 4.4.2-1 平、假日 AI 臺北市實驗場域實測各分相綠燈秒數趨勢圖	147
圖 4.4.2-2 臺北市實驗場域控制路口 TOD 平日昏峰綠燈續進示意圖	147
圖 4.4.2-3 臺北市實驗場域控制路口 AI 號誌控制平日昏峰綠燈續進示意圖	148
圖 4.4.2-4 平日 11/16&11/23 AI 號誌控制與 TOD 旅行時間績效示意圖	149
圖 4.4.2-5 平日 11/20&11/27 AI 號誌控制與 TOD 旅行時間績效示意圖	150
圖 4.4.2-6 假日 11/18&11/11 AI 號誌控制與 TOD 旅行時間績效示意圖	150
圖 4.4.2-7 假日 11/19&11/12 AI 號誌控制與 TOD 旅行時間績效示意圖	150
圖 4.4.3-1 實測階段遇到控制器異常情形示意	152
圖 5.2-1 號誌協控模式訓練結構圖	158
圖 5.3-1 高速公路匝道與平面道路實驗場域示意圖-以楊梅交流道為例	159
圖 5.4.2-1 路口轉向交通量示意圖-以環東路與中山北路路口為例	165
圖 5.4.2-2 本計畫採用之延滯	166
圖 5.4.2-3 GVP 旅行時間蒐集路段示意圖-以楊梅交流道為例	167
圖 5.4.3-1 楊梅交流道場域 VISSIM 建模搭配空拍底圖範例	167
圖 5.4.3-2 楊梅交流道場域 SUMO 建模範例	168
圖 5.4.3-3 SUMO 圖像化時相動線設定及文字矩陣時相動線設定	168
圖 5.4.3-4 VISSIM 中號誌時制設定範例	169
圖 5.4.3-5 SUMO 交通需求設定範例	169
圖 5.4.3-6 VISSIM 交通需求設定範例	169

表目錄

表 2.1.1-1 觸動號誌與固定時制績效比較表.....	6
表 2.1.1-2 情境績效比較表.....	8
表 2.1.1-3 控制策略及情境績效比較表.....	9
表 2.1.1-4 ACTS-T 控制策略及情境優劣比較表.....	10
表 2.1.2-1 模擬系統之系統參數設定表.....	13
表 2.1.2-2 模擬之績效比較表.....	14
表 2.1.2-3 上匝道與都市協控模型之目標式.....	14
表 2.2.2.1-1 TTS 與 SD(TWT)於各策略的結果.....	20
表 2.2.3-1 人工智慧強化學習號誌控制個環節參數.....	27
表 2.3.2-1 強化學習演算法之比較.....	45
表 3.1-1 前期 B1 模型臺北市實驗場域 AI 強化學習模型架構說明.....	54
表 3.1-2 前期 A2 模型臺南市實驗場域 AI 強化學習模型架構說明.....	54
表 3.2.1-1 臺北市實驗場域 AI 強化學習模型規劃表.....	60
表 3.2.1-2 臺北市實驗場域 AI 強化學習模型獎勵計算公式表.....	60
表 3.2.1-3 臺南市實驗場域模型規劃表.....	61
表 3.2.1-4 臺南市實驗場域 AI 強化學習模型獎勵計算公式表.....	61
表 3.2.1-5 臺北市與臺南市實驗場域模型比較表.....	64
表 3.2.1-6 臺北市實驗場域 AI 強化學習模型秒數限制表.....	65
表 3.2.1-7 臺北市實驗場域 AI 強化學習模型行人時相限制表.....	65
表 3.2.1-8 臺南市實驗場域 AI 強化學習模型秒數限制表.....	65
表 3.2.1-9 Q-Pensieve 訓練參數.....	67
表 3.2.2-1 本計畫 AI 強化學習模型訓練使用之軟硬體規格.....	69
表 3.2.2-2 國網中心使用之規格.....	69
表 3.2.3-1 臺北市實驗場域獎勵目標測試績效權重表.....	70
表 3.2.3-2 平日晨峰偏好績效彙整表.....	70
表 3.2.3-3 平日離峰偏好績效彙整表.....	71
表 3.2.3-4 平日昏峰偏好績效彙整表.....	72
表 3.2.3-5 假日昏峰偏好績效彙整表.....	72
表 3.2.3-6 臺南市實驗場域獎勵目標測試績效權重表.....	73
表 3.2.3-7 平日晨峰(7:00-8:00)時段一偏好績效彙整表.....	74
表 3.2.3-8 平日晨峰(8:00-9:00)時段二偏好績效彙整表.....	74
表 3.2.3-9 平日離峰偏好績效彙整表.....	75
表 3.2.3-10 平日昏峰偏好績效彙整表.....	76
表 3.2.3-11 假日昏峰偏好績效彙整表.....	76
表 3.2.3-12 多任務 AI 強化學習模型平日離峰偏好績效彙整表.....	78
表 3.2.3-13 多任務 AI 強化學習模型平日昏峰偏好績效彙整表.....	78
表 4.1.1-1 臺南市「台 86-19 甲線」路口道路幾何特性.....	82
表 4.1.1-2 臺南市「台 86-19 甲線」路口平、假日號誌時制計畫.....	82

表 4.1.1-3 臺南市「台 86-19 甲線」路口各時段延滯與服務水準	83
表 4.1.2-1 德行西路、195 巷及克強路路口道路幾何特性彙整表	85
表 4.1.2-2 德行西路、195 巷及克強路路口號誌時制計畫	85
表 4.1.2-3 臺北市實驗場域平日上下午尖峰延滯與服務水準	86
表 4.3.1-1 路口車流轉向辨識平日晨峰(07:45~08:00)準確度驗證	94
表 4.3.1-2 路口車流轉向辨識平日昏峰(17:45~18:00)準確度驗證	94
表 4.3.1-3 匝道車流辨識平日晨、昏峰準確度驗證	94
表 4.3.1-4 VD 平日晨峰準確度驗證	96
表 4.3.1-5 VD 平日昏峰準確度驗證	96
表 4.3.2.3-1 台南場域平均兩小時 TOD 與 AI 權重偏好績效比較彙整表	100
表 4.3.2.3-2 平日晨峰 5 時相模擬偏好延滯情境績效彙整表	101
表 4.3.2.3-3 平日晨峰 4 時相模擬偏好延滯情境績效彙整表	103
表 4.3.2.3-4 平日離峰模擬偏好延滯情境績效彙整表	104
表 4.3.2.3-5 平日昏峰模擬偏好延滯情境績效彙整表	106
表 4.3.2.3-6 假日昏峰模擬偏好延滯情境績效彙整表	108
表 4.3.2.4-1 平均兩小時 TOD 與 AI 權重偏好績效比較彙整表	110
表 4.3.2.4-2 平日晨峰 TOD 及本期 AI 強化學習模型模擬注重平均表現績效彙整表	111
表 4.3.2.4-3 中山北路/克強路平日晨峰模擬穩定性分析 1	113
表 4.3.2.4-4 中山北路/克強路平日晨峰模擬穩定性分析 2	113
表 4.3.2.4-5 中山北路/195 巷平日晨峰模擬穩定性分析	114
表 4.3.2.4-6 中山北路/德行西路平日晨峰模擬穩定性分析 1	114
表 4.3.2.4-7 中山北路/德行西路平日晨峰模擬穩定性分析 2	114
表 4.3.2.4-8 平日離峰 TOD 及本期 AI 強化學習模型模擬注重平均表現績效彙整表	116
表 4.3.2.4-9 中山北路/克強路平日離峰模擬穩定性分析 1	118
表 4.3.2.4-10 中山北路/克強路平日離峰模擬穩定性分析 2	119
表 4.3.2.4-11 中山北路/195 巷平日離峰模擬穩定性分析	119
表 4.3.2.4-12 中山北路/德行西路平日離峰模擬穩定性分析 1	120
表 4.3.2.4-13 中山北路/德行西路平日離峰模擬穩定性分析 2	120
表 4.3.2.4-14 平日昏峰模擬注重主幹道表現績效彙整表	121
表 4.3.2.4-15 中山北路/克強路平日昏峰模擬穩定性分析 1	124
表 4.3.2.4-16 中山北路/克強路平日昏峰模擬穩定性分析 2	124
表 4.3.2.4-17 中山北路/195 巷平日昏峰模擬穩定性分析	124
表 4.3.2.4-18 中山北路/德行西路平日昏峰模擬穩定性分析 1	125
表 4.3.2.4-19 中山北路/德行西路平日昏峰模擬穩定性分析 2	125
表 4.3.2.4-20 假日昏峰模擬注重通過量表現績效彙整表	126
表 4.3.2.4-21 中山北路/克強路假日昏峰模擬穩定性分析 1	129
表 4.3.2.4-22 中山北路/克強路假日昏峰模擬穩定性分析 2	129
表 4.3.2.4-23 中山北路/195 巷假日昏峰模擬穩定性分析	130
表 4.3.2.4-24 中山北路/德行西路假日昏峰模擬穩定性分析 1	130
表 4.3.2.4-25 中山北路/德行西路假日昏峰模擬穩定性分析 2	131

表 4.3.2.4-26 克強路與中山北路口 AI 強化學習模型與 TOD 的各分相平均秒數對照表	132
表 4.3.2.4-27 195 巷與中山北路口 AI 強化學習模型與 TOD 的各分相平均秒數對照表	133
表 4.3.2.4-28 德行西路與中山北路口 AI 強化學習模型與 TOD 的各分相平均秒數對照表	134
表 4.3.3-1 實測監控日程	136
表 4.3.3-2 實測例外事件一覽表	137
表 4.4.1-1 假日 AI 號誌控制路口延滯績效表	143
表 4.4.1-2 假日 TOD 路口延滯績效表	143
表 4.4.1-3 平日 11/7 AI 號誌控制路口延滯績效	143
表 4.4.1-4 平日 11/14 TOD 路口延滯績效	144
表 4.4.1-5 平日 11/8 AI 號誌控制路口延滯績效	144
表 4.4.1-6 平日 11/15 TOD 路口延滯績效	144
表 4.4.2-1 假日 AI 號誌控制&TOD 路口延滯績效表	148
表 4.4.2-2 平日 AI 號誌控制&TOD 路口延滯績效表	149
表 4.4.3-1 實測結果綜整表	151
表 5.2-1 交流道場域 AI 強化學習模式規劃表	156
表 5.3-1 實驗場域施作路口幾何特性彙整表-以楊梅交流道為例	159
表 5.3-1 實驗場域施作路口幾何特性彙整表-以楊梅交流道為例(續)	160
表 5.3-2 楊梅交流道平日時段切分與時制計畫	161
表 5.3-2 楊梅交流道平日時段切分與時制計畫(續)	162
表 5.3-3 楊梅交流道假日時段切分與時制計畫	163
表 5.3-3 楊梅交流道假日時段切分與時制計畫(續)	164
表 5.4.4-1 國內外模擬校估績效指標及範圍	172
表 5.4.4-2 國內外模擬校估績效指標及範圍	173
表 5.4.4-3 GEH 和誤差百分比的比較分析	173
表 5.4.4-4 VISSIM 楊梅場域流量 GEH 及 MAPE	174
表 5.4.4-5 SUMO 楊梅場域流量 GEH 及 MAPE	175

第一章 計畫概述

1.1 計畫背景與目的

近期人工智慧快速發展於各領域應用，在軟硬體技術上更是一大突破，可預期未來運用人工智慧、影像辨識、資通訊(ICT)、車聯網(V2X)與 5G 等技術，有助於紓緩因交通號誌控制智慧化不足，進而導致民眾行之痛點。本所於 104 年至 108 年進行一系列智慧交通服務效能與安全車聯網研究計畫，針對我國都市交通號誌控制與車聯網整合部分優先探討，使各式車輛接收即時號誌動態資訊，以提供安全、優先通行與節能之道路交通環境。交通部「淡海新市鎮智慧交通場域試驗研究計畫」以本所車聯網系列研究為基礎，進一步針對號誌控制與電信車聯網(C-V2X)整合的通訊協定需求，參酌 SAE 標準研擬「號誌控制器與車聯網路側設施間資通訊標準」。本所亦於 109 年「應用人工智慧技術進行交通數據蒐集暨號誌控制之研究」計畫發展以仿真學習(Imitation Learning, IL)與以號誌「週期」為基礎之人工智慧強化學習(RL)號誌控制，而國內亦有相關學術研究與桃園市的實作案例。同時本所於 110 年「構建 5G 智慧交通數位神經中樞(1/2)－功能架構探討與系統規劃」研究中，探討車聯網與路口號誌控制運作課題。

在 111 年度中，本所發展「我國人工智慧車聯網之號誌控制模式探討」研究計畫，除回顧彙整國內外應用車聯網，以及人工智慧強化學習於號誌控制之應用與演算法外，並發展車聯網與人工智慧強化學習之號誌控制模擬環境，藉由車聯網緊急車輛優先號誌、單一路口多時相號誌控制，及幹道連鎖號誌控制模擬之人工智慧號誌控制模式構建、學習與訓練、模擬與績效評估，為後續人工智慧號誌控制奠下研發基礎。計畫模擬實驗場域包含臺南市「台 86-19 甲」單一路口、臺北市「中山北路-德行東路」3 個路口幹道與高雄市「台 88 鳳山出口」(過埤路-鳳頂路)單一路口。因此，112 年根據 111 年所發展人工智慧號誌控制模式，進行與縣市政府合作之現場實驗測試，過程中持續優化人工智慧號誌控制模式。112 年同時納入高速公路匝道儀控之人工智慧號誌控制課題，並探討相關模擬發展模式，期待逐步發展我國對於人工智慧號誌控制之實力。113 年將延續 112 年在交流道區域匝道儀控與平面號誌導入人工智慧強化學習探討成果，發展人工智慧強化學習之號誌協控模型，並與高速公路局及縣市政府共同協作，在高速公路匝道儀控與平面道路路口動態號誌範圍內，進行號誌協控訓練與實測。本年度計畫主要目標將發展符合我國主要交通環境下，導入人工智慧號誌，藉由與實務單位場域結合，支援其智慧號誌控制與管理需求，取得加乘效果，使提高車流運作整體效率與交通安全。

1.2 計畫工作流程

本計畫執行期間為 112 年 4 月 13 日起至 112 年 12 月 31 日。工作項目可區分為「都會區幹道強化學習號誌控制模式實測驗證」、「交流道區域強化學習號誌協控模型探討與模擬環境建置」、以及「其他相關工作」，諸如研究成果投稿與成果行銷推廣活動等三大類。專案期間劃分為期中階段與期末階段，依照規劃完成各階段之目標，執行流程如圖 1.2-1 所示。

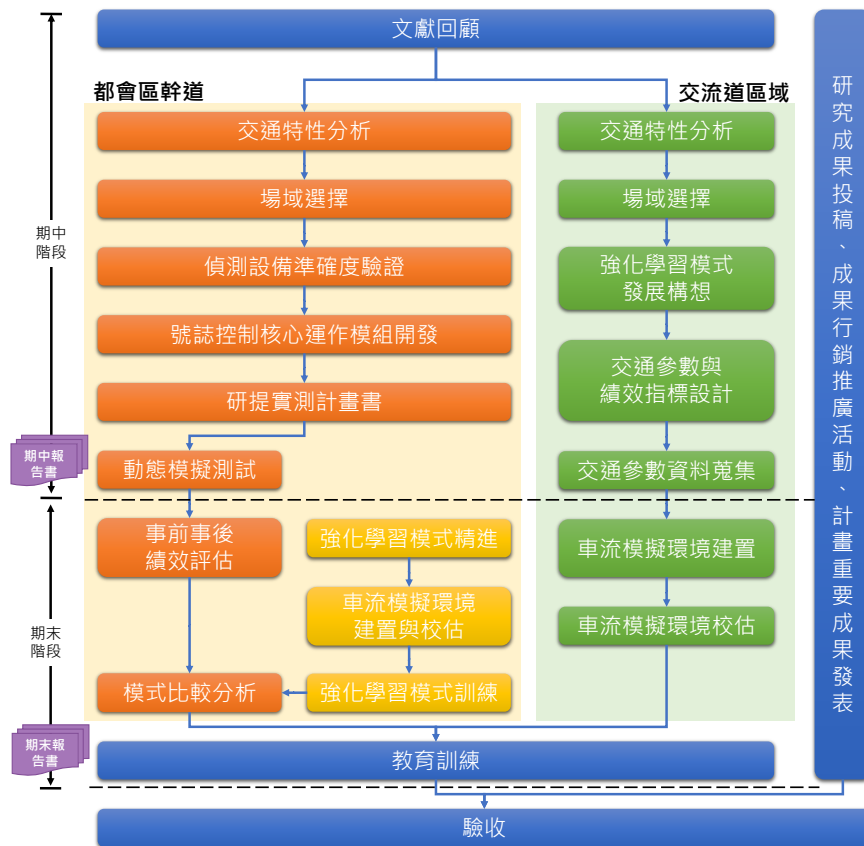


圖 1.2-1 計畫執行流程

1.3 研究內容與工作項目

一、人工智慧強化學習模式應用探討

持續盤點與彙整國內外應用人工智慧強化學習(Reinforcement learning, RL)於號誌控制案例，包括交流道區域匝道儀控與平面號誌之協控，以及演算法應用。

二、現場實測

依據前期 111 年度於臺北市「中山北路-德行東路」3 個路口幹道與臺南市「台 86-19 甲」路口等縣市(以下簡稱「合作縣市」)共同發展人工智慧強化學習號誌控制模式訓練成果，進行現場實驗測試。

- (一)對於合作縣市之車輛偵測器資料與品質進行確認與必要之準確度驗證，以及利用前期校估之 VISSIM 模擬環境，進行前期發展人工智慧強化學習號誌控制模型之精進、

效能評估、參數調校或再訓練工作。此項目執行過程得視需要，採用不同人工智慧強化學習演算法進行模型發展之嘗試與優化。

- (二)結合合作縣市現場車輛偵測器與號誌動態資料，以及經由訓練後之人工智慧強化學習號誌控制模型成果，發展核心運作模組產生下一階段號誌控制參數，該模組須具備簡易運作狀態監控、紀錄與績效分析等能力，以及應用於合作縣市之不同人工智慧強化學習模型參數與運作環境使用彈性。
- (三)研提合作縣市實驗場域測試所需之現場設備，包含號誌控制器、車輛偵測器與通訊系統等需求，以及其交控中心配合與整合本計畫人工智慧強化學習號誌控制核心運作模組與運作所需之程序或軟體需求。
- (四)協助合作縣市交控中心開發符合本計畫實測所需軟體，以及相關實測工作進行。
- (五)實測前研提實測計畫書與完成動態模擬測試流程，並與合作縣市進行現場實驗測試相關協調工作。
- (六)實測之事前事後量化績效評估。

三、模擬模式構建與號誌協控模型探討

進行高速公路交流道區域匝道儀控與平面路口動態號誌整合之模擬模式構建與人工智慧強化學習號誌協控模型探討。

- (一)依據本計畫發展人工智慧號誌控制模式需求，自六都與高速公路介接之交流道，包含前期 3 處實驗場域在內，遴選楊梅交流道作為實驗場域，場域選擇考量模擬模式與模型構建所需資料，以及後續實測現場資通訊軟硬體可配合因素，故係從都會區角度進行探討，僅考量平常日與高速公路獨立式匝道儀控。
- (二)研提高速公路交流道區域匝道儀控與平面路口號誌整合之人工智慧強化學習號誌協控模型，其發展構想與初步設計。在模型發展與設計過程中，本團隊將依循需求，評估導入多號誌控制代理人機制，於績效評估中考量駕駛人感受層面。
- (三)設計實驗場域模擬模式所需之各式交通參數與績效指標時，將多方考量人工智慧強化學習號誌控制模型常見之狀態(State)與獎勵(Reward)參數。
- (四)根據模式設計之交通參數與績效指標內容，蒐集該實驗場域資料。
- (五)於實驗場域蒐集交通參數方面，利用縣市交控系統與高速公路局交控系統即時蒐集或歷史之交通參數資料，如車隊長度、延滯等。搭配現況人工調查方式，以確認模擬模式能達成模擬績效指標值，與現場調查績效指標值誤差亦小於或等於 10%。若因實務上之限制無法達成效果時，具體提出說明後可將誤差值門檻調整為 20%

第二章 文獻回顧與關鍵課題

2.1 國內外號誌控制策略探討

2.1.1 平面路口之號誌控制

號誌控制是道路交叉口的重要環節之一，分配不同來向的車流順暢且安全的通過路口是號誌控制的重要目的之一。而其中田永平(2006)[52]提及獨立路口的時制設計可透過公式推導、數學規劃模型及模擬進行求解。幹道的號誌系統，除了獨立路口的參數，更需要考量時差及綠燈續進帶，幹道連鎖的延滯、通過量等績效指標，也是判斷號誌策略是否被妥善的運用。陳麗雯(2020)[60]亦將兩套號誌最佳化軟體，TRANSYT-7F 以及受到交通領域高度使用的號誌最佳化軟體 Synchro，對於幹道場域進行最佳化求解，並比較兩者求解結果。

根據 Jiang(2020)[53]，動態控制是目前交通控制中常見的手段，其中包括定時式、觸動式及動態式等模式。這些控制策略在國際交通運輸界受到重視，尤其是在智慧交通系統（ITS）領域。目前臺北市所建置的智慧號誌控制系統中，以觸動控制和適應性號誌策略為主。觸動號誌控制根據道路上偵測的車流資料調整綠燈時間，應用於主次要道路交通量懸殊的地點；適應性號誌控制策略旨在增強路口號誌時制的智能化，以快速應對交通狀況的變化，同時改進動態查表等交通控制方法的不足之處，提高運作績效。而近期，許多都市透過人工智慧建立號誌控制策略以嘗試解決長期壅塞的區域，主要可分為時相決策或週期決策。

一、觸動控制

(一)半觸動號誌控制模式之建構與驗證(Chang et al, 2012)[61]

為探討半觸動號誌在改善延滯的情況，Chang(2012)對於半觸動號誌進行情境分析及敏感度分析，分別在都會區具上下游路口處，以及郊區獨立路口處進行半觸動號誌設立的績效評估。其中考慮了幹道的 V/C ratio，以及幹支道流量比。

摘錄其部分結果如表 2.1.1-1 所示，假設每車道 C 值為 1200PCU，幹道為兩車道、支道一車道的 T 型路口，在速限 50KM/HR 的情境下，當幹支道流量比例懸殊，如 10:1、10:2 時，觸動控制改善之延滯績效成效較佳。當兩者流量比例接近時，觸動邏輯容易感應車流而延長綠燈，造成另一方向車流總延滯增加幅度大。在實驗場域，觀察市區廊道道路，即使幹支道流量差異大，亦不太適合施行感應性號誌。其原因為市區廊道號誌多設計為連鎖續進，該感應性路口因支道無車輛而開啟幹道綠燈後，上下游車流仍為紅燈停等狀態，造成多數情形為無效綠燈，反而因車流無法有效續進造成幹道延滯增加，與感應性號誌欲減少幹道延滯之目標相悖。

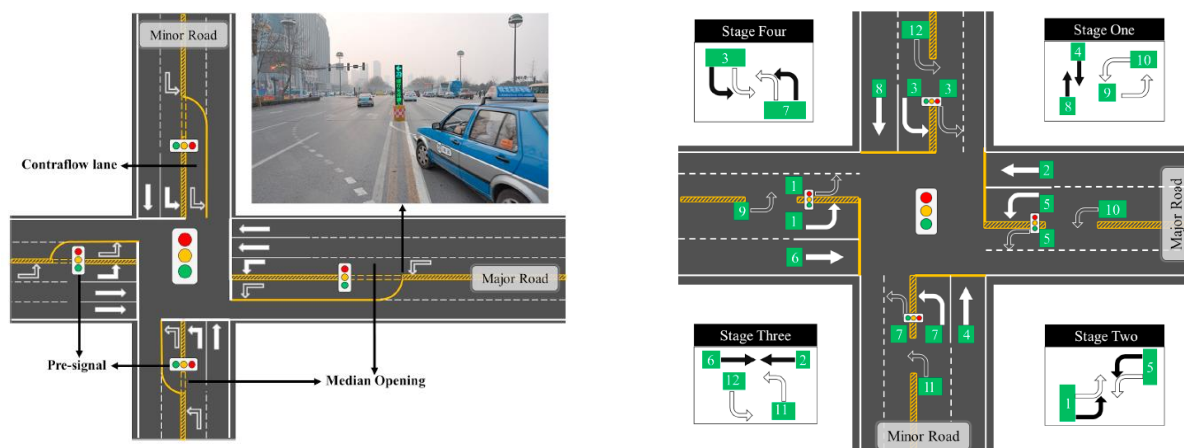
表 2.1.1-1 觸動號誌與固定時制績效比較表

V/C 值	幹支道流量比例	方向	定時號誌-平均延滯時間(秒/車)	半觸動號誌-平均延滯時間(秒/車)	績效值	定時號誌-路口平均延滯(秒/車)	半觸動號誌-路口平均延滯(秒/車)	績效值
0.8	10:1	幹道 1	5.37	3.31	38.31%	12.79	9.15	28.47%
		幹道 3	5.81	3.27	43.72%			
		支道 2	27.19	20.86	23.27%			
	9:1	幹道 1	5.13	3.56	30.54%	12.33	9.24	25.04%
		幹道 3	5.94	3.71	37.52%			
		支道 2	25.92	20.45	21.09%			
	8:1	幹道 1	5.31	4.38	17.65%	12.12	9.47	21.85%
		幹道 3	5.38	4.77	11.26%			
		支道 2	25.66	19.26	24.94%			
	10:2	幹道 1	5.79	4.53	22.06%	10.37	9.20	11.22%
		幹道 3	5.15	4.74	7.60%			
		支道 2	20.16	18.34	9.03%			
	10:3	幹道 1	5.11	5.01	2.19%	11.01	9.38	14.83%
		幹道 3	4.39	4.91	-12.36%			
		支道 2	23.54	18.22	22.61%			

資料來源：(Chang et al, 2012) [61]

(二)Developing an actuated signal control strategy to improve the operations of contraflow left-turn lane design at signalized intersections (Wu et al, 2019)[47]

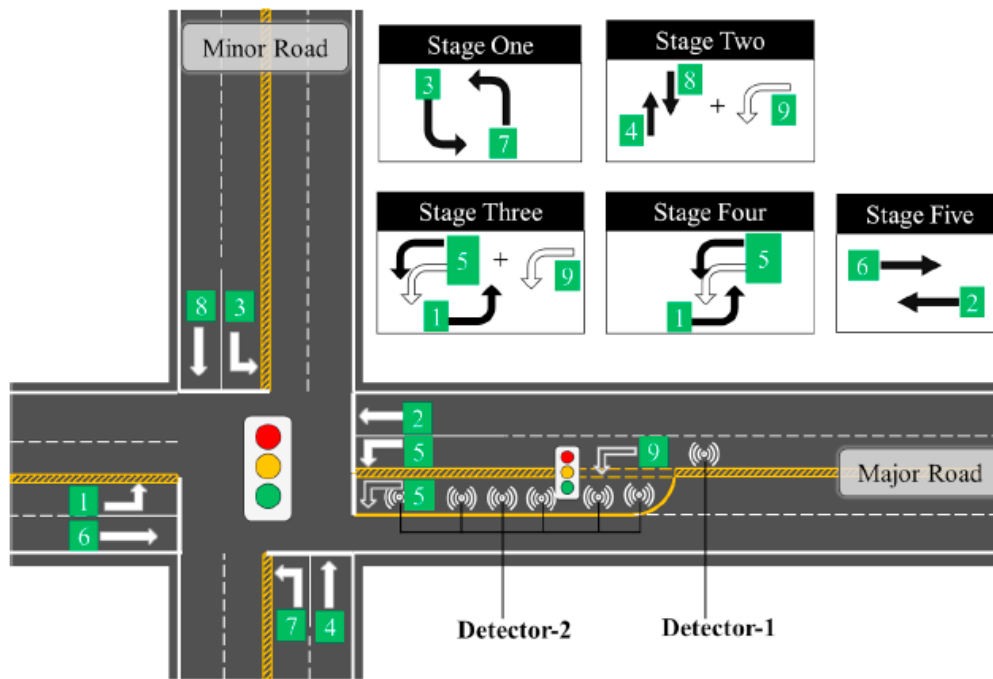
半觸動號誌除了適用於幹支道交通量差異較大之場域外，Wu(2019)提出以可改善逆向左轉車道(Contraflow Left-Turn Lane, CLL)車輛延滯時間。在中國，為了緩解信號化十字路口左轉車輛引起的交通擁堵，越來越多地採用了 CLL 設計。CLL 設計即為在分隔帶的對向直行車道上，設置一個或多個左轉車道，如圖 2.1.1-1 所示。其中，該車道設計由於會駛入對向車道，故路口近端將設置 pre signal 設備，此號誌讓車輛除了被允許的等待左轉時間外，禁止進入對向車道，以免發生逆向的車輛衝突。而釋放停等車流進入對象等候左轉，亦能夠疏解原線道左轉車隊長度。



資料來源：(Jiaming Wu, Pan Liu, Xiao Qin, Huaguo Zhou, Zhao Yang, 2019) [47]

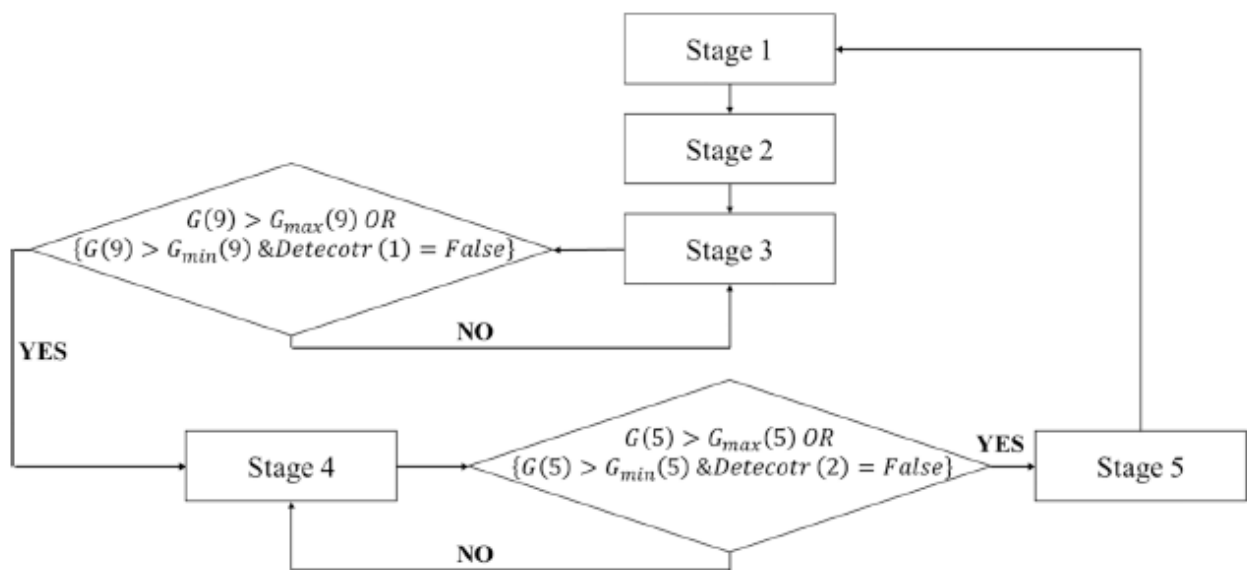
圖 2.1.1-1 設置逆向左轉車道路口以及路口時相示意圖

但 CLL 的 pre signal 設計運作，亦產生問題。由於左轉車輛只能於特定時相進入 CLL，故設計的 CLL 長度可能有無法完全有效利用的問題；另一方面，如果在左轉時相結束時 CLL 中左轉車隊無法完整紓解，亦即在後續的時相，該停等車隊將全數變成逆向車流，存在交通安全上的問題。故 Wu(2019)以簡單一方向設置 CLL 的道路幾何條件下，提出半觸動控制邏輯策略，最大綠最小綠、時制重整，以及偵測器設置位置，以期改善左轉車輛延滯的情況。



資料來源： (Jiaming Wu, Pan Liu, Xiao Qin, Huaguo Zhou, Zhao Yang, 2019) [47]

圖 2.1.1-2 單方向設置逆向左轉車道路口圖以及路口時相示意圖



資料來源： (Jiaming Wu, Pan Liu, Xiao Qin, Huaguo Zhou, Zhao Yang, 2019) [47]

圖 2.1.1-3 半觸動號誌邏輯

後續 Wu(2019)以固定時制(Fixed CLL)、無 CLL 的傳統左轉車道、以及觸動版本的 CLL 做比較，如表 2.1.1-2 所示。結論可得知觸動版本之 CLL 較使用固定時制之 CLL 佳，但績效略小於傳統的雙左轉專用車道之幾何配置。

表 2.1.1-2 情境績效比較表

Characteristics of different design concepts.

Design concepts	Control strategy	L (m)	n_{con}	n_{CLL}	Cycle length(s)
Actuated CLL	Actuated	50	1	1	–
Fixed CLL	Fixed Time	56	1	1	150
Fixed Conventional	Fixed Time	–	2	0	150

Comparison of the throughput volume associated with different design concepts.

Design Concept	Throughput Volume (veh/h)	Demand (veh/h)						
		300	350	400	450	500	550	600
Fixed-CLL	Mean	303.8	340.6	405.6	448.8	499.2	531.4	531.8
	Std. Deviations	74.6	62.7	60.2	69.9	56.2	15.9	15.4
Actuated-CLL	Mean	303.8	352.3	403.7	451.2	499.2	546.2	593.8
	Std. Deviations	81.8	93.5	83.1	98.7	109.7	87.1	67.2
Fixed-Conventional	Mean	303.8	351.4	405.1	449.3	501.6	549.6	601.0
	Std. Deviations	74.0	75.1	77.0	85.2	93.6	87.4	76.6

Comparison of delay performance associated with different design concepts.

Design Concept	Delay (s)	Demand (veh/h)						
		300	350	400	450	500	550	600
Fixed-CLL	Mean	62.1	60.7	63.1	68.9	88.6	208.1	223.3
	Std. Deviations	11.1	11.5	12.3	12.9	20.9	40.5	28.6
Actuated-CLL	Mean	60.3	56.8	57.6	55.7	61.0	68.7	130.1
	Std. Deviations	12.1	11.9	10.6	10.4	12.0	20.4	34.9
Fixed-Conventional	Mean	52.0	52.3	53.1	55.9	56.7	57.6	65.1
	Std. Deviations	12.1	12.2	10.0	9.8	8.6	11.4	11.6
P-value of t-Test ^a		0.46	0.10	0.01	< 0.01	< 0.01	< 0.01	< 0.01

資料來源：(Jiaming Wu, Pan Liu, Xiao Qin, Huaguo Zhou, Zhao Yang, 2019) [47]

二、適應性號誌控制

(一)都市適應性號誌控制原理與發展(Hu et al, 2020)[55]

適應性號誌(Adaptive Traffic Control System, ATCS) 透過偵測車流高低來調整綠燈時比或延長綠燈，而運作方式可透過群組共同考量及控制來加強其控制彈性及績效表現。適應性控制邏輯依演算差異可區分為二大類，在當前週期或一固定時間內，計算下一階段之號誌週期與時比的週期性號誌控制邏輯(Cyclic Control Logic)；以及以即時蒐集所得的車流資訊，在每一個布階中比較各種競爭方案的績效值，此為週期內號誌控制邏輯(Time-Step Control Logic)。

Hu(2020)亦匯集國內外，目前商用適應性號誌控制邏輯。國外週期適應性號誌控制有少量多次頻繁微調參數之 SCOOT；以均衡鄰近衝突路段的飽和度為目標制定時制計畫，整體作法似進階動態查表的 SCATS；經由車流預測模式推估車流需求選擇最適宜

之週期時間，再經由基因演算法求解群組最佳的時比與時差之 BALANCE。

而國外週期內適應性號誌控制有並藉過去數個週期之交通流量資料，透過水平推移法預測未來數個週期之流量型態的 OPAC；藉由路口間的 P2P 串連分享即時車流資訊，並透過 PREDICT 模式推估上游車輛或車隊到達路口之時間，藉由演算法預測下一分鐘之隨機車流到達型態之 RHODES。

國內早期於 1988 年發表適用國內道路情境之適應性號誌控制邏輯 COMDYCS 系列，其可適用於小型路網、雙時相、純車流情境。後續有版本納入轉向比估計模式，有的版本進一步發展混合車流 CTM 車流預測模式。COMDYCS-III 則是以 SAST 架構開發，屬於週期內號誌控制邏輯，其透過上下游偵測器蒐集車流資料，計算路口各路段之停等長度與總延滯時間，再判斷是否延長或結束綠燈。Hu(2020) 亦於單一路口進行實證，比較各號誌控制策略在到達率固定及變動的情境下績效為何，結果如表 2.1.1-3 所示。

表 2.1.1-3 控制策略及情境績效比較表

績效指標	適應性號誌			定時號誌			觸動式號誌		
	固定到達	變動到達	增幅(%)	固定到達	變動到達	增幅(%)	固定到達	變動到達	增幅(%)
每車延滯	32.22	33.33	3.09	32.14	83.54	159.93	35.54	38.79	9.14
停等次數	1.27	1.38	8.66	1.37	4.54	231.39	1.48	1.78	20.27
最大等候線長度	91.10	92.19	1.20	78.81	316.45	301.54	109.15	138.83	27.19

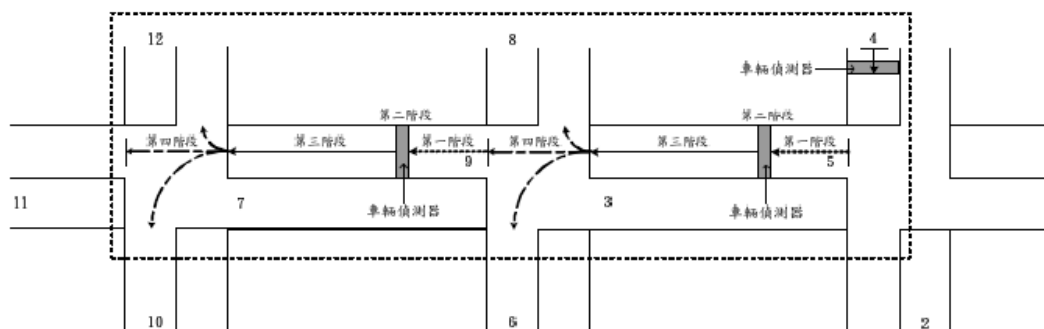
資料來源：都市適應性號誌控制原理與發展(Hu et al, 2020)[55]

(二)幹道群組適應性號誌控制模式之開發研究 (Wu, 2011) [54]

Wu(2011)針對適應性號誌策略，以系統整體旅行延滯最小化為目標，提出都市幹道號誌群組之適應性號誌控制模式，並結合路段上游之先期車流資訊，提出四階段之車流推進與預測模式，以「幹道系統整體實施連鎖控制，並以一個完整週期時制作為實施適應性控制決策之基本控制單位」為決策基本概念，考量群組內路口之週期、時比與時差等因素。而適應性啟動之運作流程每個時相皆會更新當前車隊資訊，藉以判斷該時相可完全紓解該方向車隊的綠燈時間。

Wu(2011)針對幹道群組車流，發展一幹道車流推進與預測模式，其中包含多路口車流推進模組，以及車流預測模組。多路口車流推進模組每秒進行交通量資料蒐集與車流推移，提供號誌控制模型在決策進行時即時的幹道群組交通資訊。車流預測模組則在號誌模型決策時，提供未來的流量推移預測資訊，作為預估不同決策方案下的比較依據。實測場域選擇於中華東路廊道，其實測模組分為幹道群組內適應性控制模式(ACTS-T)及幹道群組週期性適應性模組(ACTS-C)，並與 SYNCHRO 計算出之最佳化應時時制比較。結果以 ACTS-T 模式為例，Case1~Case12 為低中高流量，各流量再細分

為遞增、遞減、持平穩定及持續變化的流量狀態，其中低流量不考慮遞減，高流量不考慮遞增狀態，如表 2.1.1-4 所示。



資料來源：幹道群組適應性號誌控制模式之開發研究 (Wu, 2011) [54]

圖 2.1.1-4 適應性號誌控制於多路口車流推進之分段示意圖

表 2.1.1-4 ACTS-T 控制策略及情境優劣比較表

實驗案例	ACTS-T 控制策略 (較佳：✓；無差異：-；較差：✗)	
	較幹道群組最佳化定時	較 ACTS-C
Case1	✓	✗
Case2	✓	✗
Case3	✓	✗
Case4	✓	✓
Case5	✓	✓
Case6	✓	✓
Case7	✓	✓
Case8	✓	✓
Case9	✓	✓
Case10	✓	✓
Case11	✓	✓
Case12	✓	✓
總計	10 (100.0%)	7 (70%)

資料來源：幹道群組適應性號誌控制模式之開發研究 (Wu, 2011) [54]

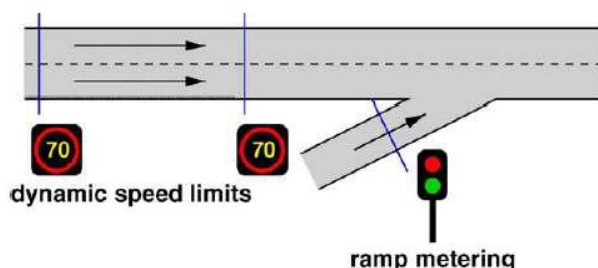
2.1.2 交流道之號誌協控

交流道系統與一般市區號誌控制的不同，是高速公路、匝道、平面道路三個目標的相互關係，高速公路主線是否壅塞及上匝道的衝擊影響，是多數研究長年致力改善的目標之一。協控即為不同系統或不同單位，為了達成地區的交通控制而進行共同控制。而匝道儀控為目前國內最常使用的高速公路交通控制手段，依據高速公路主線的壅塞情況，可以降低匝道儀控率避免主線車流密度超過臨界值，雖能減少匯入主線的車流造成的衝擊，但實施低儀控率之時段過長亦會影響平面道路的車流情況，因此可在主線不壅塞時盡量放大儀控率，保持平面道路順暢。

一、匝道與高速公路主線道的協控

(一)高速公路匝道儀控與速率控制協控模式之研究(Hsu et al, 2009) [62]

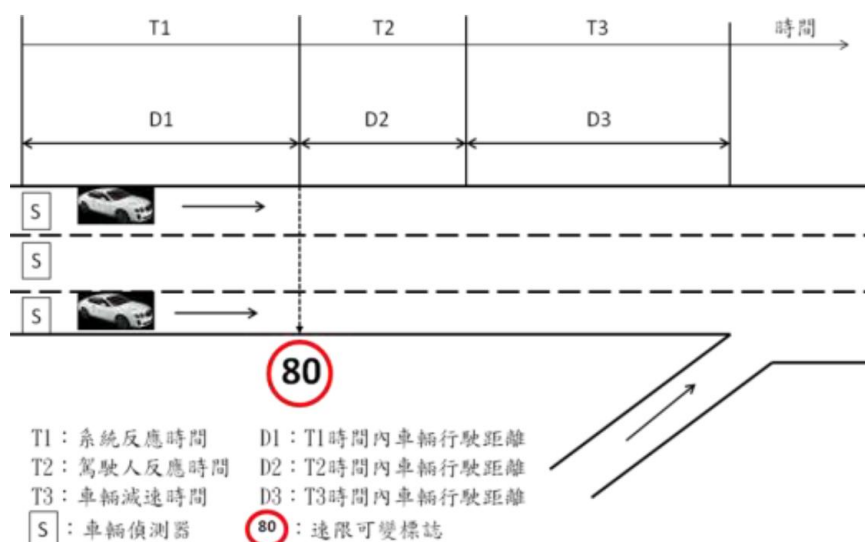
雖匝道儀控為常用手段，但當主線車輛密度過高時，匝道儀控為了避免匯入車流衝擊而不斷縮減儀控率；惟避免匝道等候車隊長度過長，到達匝道設置的最小儀控率時，在此情況下匝道儀控即無法避免壅塞情形發生(Hsu et al, 2009)。在考量高速公路主線以及匝道的壅塞情況，為提升兩者的績效，透過調整匝道儀控率，亦在高速公路上游佈設偵測器，調整主線的車輛速率，可降低車間速率變異，使車流在高密度的情況下，維持較高的速率及流量。而降低上游車輛進入瓶頸的速率，也可以延緩衝擊波向上游蔓延的情況。藉由調整匝道匯入處上游的主線速率及匝道儀控率，兩者協控以降低車輛交會時受影響的程度，幾何示意如圖 2.1.2-1 所示。



資料來源：(Tien-Pen Hsu, Chung-Yu Huang, & Yu-Jui Lin, 2009) [62]

圖 2.1.2-1 匝道儀控與速率控制協控系統

因偵測器是以每隔一固定時距方式收集交通特性資料給速率控制決策模式，故匝道匯入處上游車輛在此時距中已行駛 $D1$ 的距離，因此上游偵測器位置應至少距速限可變標誌 $D1$ 之距離，此時駕駛人發覺速限可變之速限內容，經過判斷開始採取減速措施，這段時間內，車輛保持原速率行駛了 $D2$ 的反應距離；反應結束後車輛開始減速，在速率降至速率控制之目標值時，車輛又行駛了 $D3$ 的減速距離。(Hsu et al, 2009)



資料來源：(Tien-Pen Hsu, Chung-Yu Huang, & Yu-Jui Lin, 2009) [62]

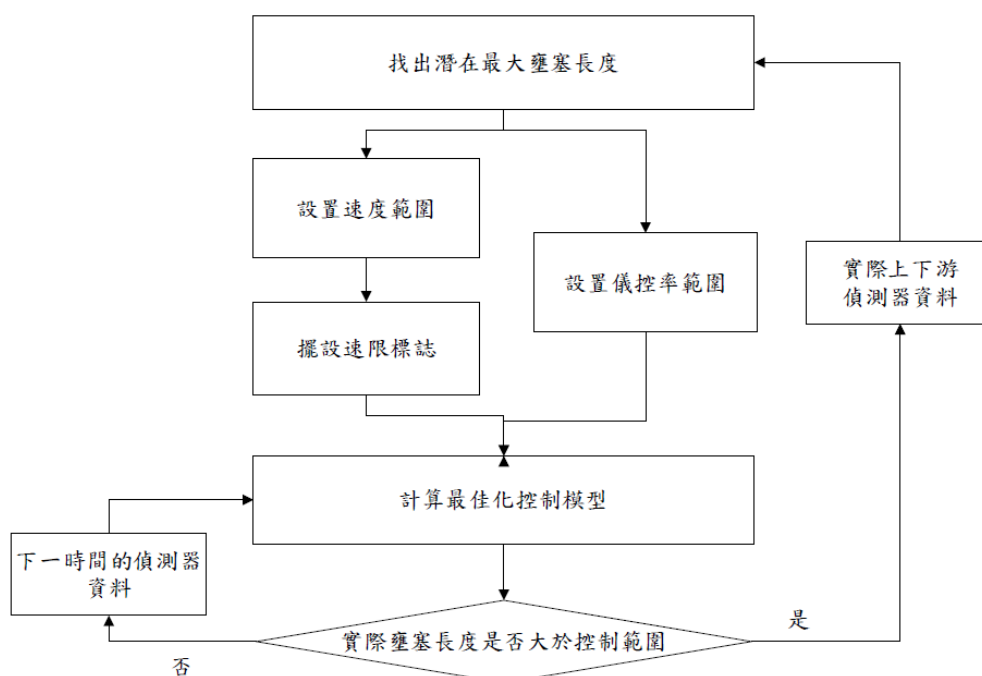
圖 2.1.2-2 主線速率控制系統設施設置說明圖

總結來說，此篇文獻提出，欲考量匝道與主線道的協控，除了匝道儀控外，主線道

需取一段交會瓶頸區的上游距離作為主線協控距離。而此距離則透過反應距離、安全減速度以及目標車輛速度幾個參數決定。

(二)高速公路速率控制與匝道儀控最佳化協控模式之研究(Hsu et al, 2011) [63]

文獻中提出透過觀測主線道的可能壅塞車隊長度，設定一最佳化協控模型，該研究控制邏輯架構圖如圖 2.1.2-3 所示。考量儀控率上限使得主線的流量密度不超過臨界值，下限需避免匝道等候車隊長度溢出至市區道路。同時計算瓶頸路段的上游速率上下限，此速率上下限是藉由宏觀車流理論計算自由流速率，以及考量駕駛安全加減速後透過牛頓運動定律而得。再經由觀測該次速限位置區間、速限上下限及儀控率上下限調整後，是否壅塞仍超出設定的範圍。如超出設定範圍，則需重新調整上述設定值區間；如壅塞情況仍在設定區間中，則持續觀測直到壅塞結束。



資料來源：(Tien-Pen Hsu & Po-Yuan Hsueh, 2011) [63]

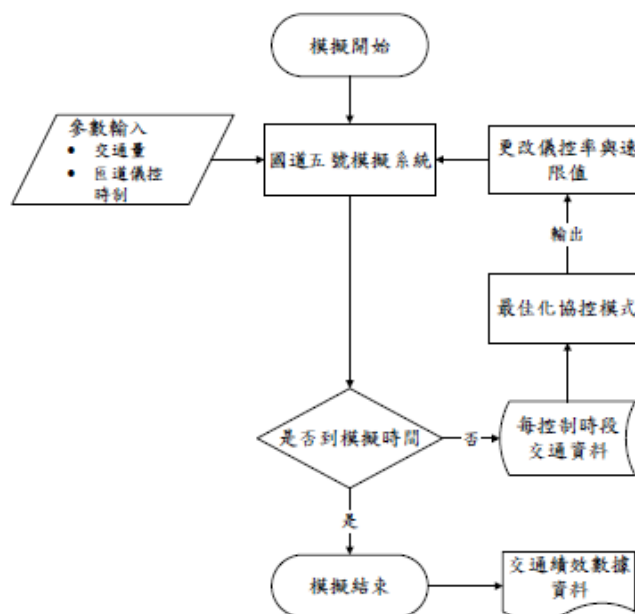
圖 2.1.2-3 控制邏輯架構圖

該控制策略以國道五號為分析案例，並透過 VSSIM 模擬軟體建模進行績效分析，模擬流程如圖 2.1.2-4。其中，該模擬模型的基本參數設定，如表 2.1.2-1 所示，而動態輸入參數為交通量及匝道儀控時制；而該模擬每個 step 所產出的輸出，如匝道等候車隊長度、上游主線流量、上游主線密度及速率等，則為最佳化模型的輸入參數。

表 2.1.2-1 模擬系統之系統參數設定表

	高速公路主線	初入隧道	隧道中段
車道數	2 車道	2 車道	2 車道
最高速限	90 公里/小時	80 公里/小時	80 公里/小時
變換車道最小車頭距	1 公尺	1 公尺	1 公尺
最小停等間距	2.82 公尺	2.82 公尺	2.82 公尺
安全間距參數	0.23 秒	4.06 秒	3.07 秒
進入跟車門檻	6 公尺	6 公尺	6 公尺

資料來源：(Tien-Pen Hsu & Po-Yuan Hsueh, 2011) [63]



資料來源：(Tien-Pen Hsu & Po-Yuan Hsueh, 2011) [63]

圖 2.1.2-4 模擬求解流程圖

(三)小結

在上述兩篇文獻中(Hsu et al, 2009[62]; Hsu et al, 2011[63])，皆有評估在模擬環境中，實施控制策略所帶來的績效。兩則文獻皆包含了以主線通過量及主線平均旅行時間做為績效評估指標。具體來說，當控制策略介入時，普遍的情境可在主線得到比未介入時好的行車績效，例如比較少的總車輛旅行時間，亦或在模擬時間區間中有比較大的總通過量。而在(Hsu et al, 2011)中，亦將匝道的停等狀況及延滯納入績效考量，整理如表 2.1.2-2。

由績效評估得知，兩篇文獻之協控策略，皆可讓控制範圍內之績效指標有不等的提升。總通過量增加，意指可讓更多車輛通過瓶頸路段區域；總旅行時間和總延遲時間下降，意為藉由此速限控制可以使得車輛塞在瓶頸路段的時間縮小；匝道等候車隊長度縮短則為上匝道的需求更能被滿足，且對地方道路的影響也降低。

表 2.1.2-2 模擬之績效比較表

	Hsu et al, 2009	Hsu et al, 2011
主線通過量	V	V
主線平均旅行時間	V	V
主線平均延滯時間	—	V
匝道平均等候車隊長度	—	V
上匝道通過量	—	V
上匝道平均旅行時間	—	V
上匝道平均延滯時間	—	V
匝道最大等候車隊長度	—	V

二、匝道與地方道路號誌控制的協控

匝道與地方道路的協控，則主要回顧運研所於 2010 年發行之「標準化都市交通管理之通訊協定研發(二)」研究報告，於該報告 4.2.1 節與 4.2.2 節分別提到上匝道及都市間的協控模式，提到下匝道與都市間的協控模式。

(一)上匝道與地方道路協控模式

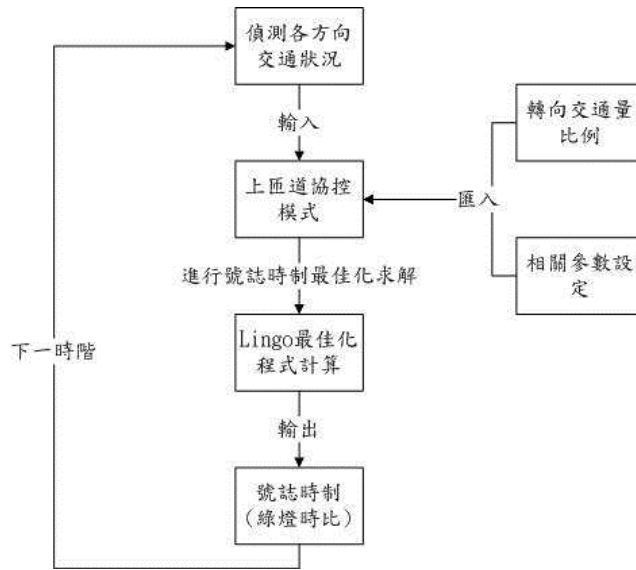
上匝道與都市間的交通問題，主因為匝道儀控為高公局控管，而地方道路號誌為地方政府交控中心控管，如兩者分別考量，則可能發生高速公路壅塞時降低儀控率，惟地方號誌仍允許車輛進入匝道導致回堵；亦或地方號誌綠燈無法搭配儀控時相等。

該報告以最佳化模型求解協控模式，主要目標為匝道壅塞長度，以壅塞程度不同作為控制目標的門檻值。其目標式考量兩種狀況：匝道已壅塞，以及匝道未壅塞。若上匝道處已壅塞，則為避免回堵至地方道路，目標式為最小化匝道等候車輛數，其意義為原排隊車輛數加上市區進入匝道車輛數減去進入主線車輛數須求解最小值。若匝道處於非壅塞情況，為提高匝道使用效率，而採取匝道通過量最大為控制目標，其意義為匝道排隊數量加上進入主線車輛數必須為最大值。並考量求解效率及時間，該文獻透過權重將多目標模型轉為單目標式之模型，整理如表 2.1.2-3。而該模式於限制式中，亦考量地方道路的車輛數及時相、匝道儀控率的上下限值、主線壅塞程度及主線壅塞臨界值等，模式求解後得到路口各時相的綠燈時比，以及匝道儀控率。主要模式運作流程圖如圖 2.1.2-5 所示。

表 2.1.2-3 上匝道與都市協控模型之目標式

情境	目標式
匝道已壅塞	$J = \min \left(x - rcA + \sum_{i=1}^n g_i c q_i \right) \quad R_{ramp} > \beta$
匝道未壅塞	$J = \max \left(\sum_{i=1}^n g_i c Q_i + rcA \right) \quad R_{ramp} \leq \beta$
合併目標式	$J = \min \left[\left(x - rcA + \sum_{i=1}^n g_i c q_i \right) \times \alpha - \left(\sum_{i=1}^n g_i c q_i + rcA \right) \times (1 - \alpha) \right]$

資料來源：交通部運輸研究所, 2010

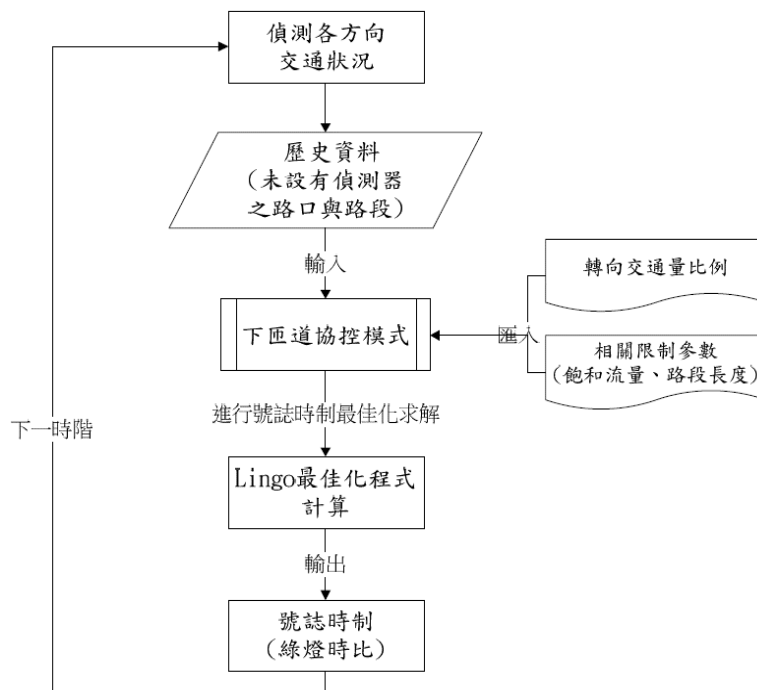


資料來源：交通部運輸研究所, 2010

圖 2.1.2-5 上匝道與都市間協作模式流程圖

(二)下匝道與都市間的協控模式

下匝道的文獻較少，現存亦無下匝道儀控設備，故多以最小延滯時間以及路段最大通過量進行設計。該報告以最小總車輛停等時間為目標式進行號誌設計，考量了每一週期進入及離開的車輛數。其中設定路段流量守恒及容量限制、各方向之車道容量及流動限制、防止下匝道車隊溢流回主線道限制以及車輛紓解率等限制式。求解將得路口各方向之綠燈時間，其主要模式運作流程圖如圖 2.1.2-6 所示。



資料來源：交通部運輸研究所, 2010

圖 2.1.2-6 下匝道與都市間協作模式流程圖

(三)小結

該報告主要基於匝道的壅塞車隊長度，建構上下匝道與都市間的地方道路協控的最佳化模型。其中上匝道口因為現存高公局的匝道儀控設備，故模型中考量了儀控率、匝道車隊長度以及地方道路的綠燈時長，目標式為以權重混和考量最小化車隊長度以及最大化疏解流量。而下匝道口之號誌歸屬於地方管轄，並無儀控率可作為模式輸入，故以最小總車輛停等時間做為目標式求解，並考量路段流量及車道流量限制、匝道車隊長度限制以及車輛於平面匝道出口的疏解率。

2.2 強化學習於號誌控制之應用

本計畫透過即時號誌動態資訊給各式車輛，以提供安全、優先通行、節能的道路交通環境，故應以適應性號誌控制(Adaptive Traffic Signal Control, ATCS)為出發點，根據路口各方向即時交通變化作動態彈性調整號誌；然過往傳統適應性號誌控制仍存在限制，因演算法本身須給定適當假設條件，並對號誌控制問題進行一定程度簡化後建立控制模型，甚至是納入預測模式的結果後，再以最佳化方法來求解號誌控制策略。此方式很大程度上取決於對交通流、道路環境之理解程度，無法學習並適應突發狀況外；當面臨規模較大之路網範圍時，其簡化問題難度、最佳化難度亦隨之急遽提升(Wang et al., 2018) [45]。

如 2.1 小節所述，若欲在複雜環境、難以透過模型建構關係的情況下求解，則強化學習(Reinforcement Learning, RL)適用於此問題；以號誌控制為例，當欲進行號誌控制時，可根據控制路口範圍內所能獲取到之車流環境狀態資訊，進行下一時相長度的決策，並透過具序列性的馬可夫決策過程，使整體控制路口範圍內的車輛在這決策周期內有著最佳的通行效率。然而市區路口環境複雜，使得狀態(state)空間與動作(action)空間存在高維度問題，使用基本 RL 方法則可能不容易找出有用特徵，故目前研究與應用大多以深度強化學習方法(Deep Reinforcement Learning, DRL)以處理高維度與複雜之控制問題，本小節以下將分別回顧國外文獻與國內過去相關研究計畫成果。

2.2.1 強化學習於平面路口之號誌控制

一、單一路口號誌控制

過往應用強化學習方法於號誌控制問題時，主要分為兩種類別(Mohammad Noaen et al., 2022) [31]：第一類為單一代理人(agent)的 DRL 應用，主要適用於單一路口或路口數較少的環境中的號誌控制問題；第二類為多個代理人(agent)的多代理人深度強化學習(Multi-Agent Deep Reinforcement Learning, MADRL)應用，此方法則適用於範圍較大的路網環境。此外，目前研究多以交通模擬軟體作為績效驗證(Qadri et al., 2020) [41]。

關於應用單一代理人之 DRL 方法至單一路口號誌控制問題，基於 DQN 的方法為最常見的方法之一，其限制在於代理人所需之狀態(state)空間與動作(action)空間皆為離散分佈；而常見離散動作可分為兩類，其一為在特定時間點決定是否切換至下一時相，可分為固定時相順

序(Abdulhai et al., 2020 [1]; Brys et al., 2014 [6]; Mannion et al., 2016 [30]; Li et al., 2016 [26])以及無順序性時相(Houli et al., 2010 [19]; Wan et al., 2019 [45]; Genders et al., 2016 [15])；其二為在每時相週期結束時針對下一週期特定時相長度進行調整(Zheng et al., 2019) [50]。Genders 等人(2016)使用 CNN 將 SUMO 車流模擬影像轉換為狀態輸入以估算 Q 函數值，證實使用 DRL 方法於號誌控制比傳統方法來的更有效。Wan 等人(2018)提出了動態計算折扣因子的方法，並應用於 DQN 方法中以設計可變的綠燈時間，證實其績效優於定時時制計畫。Shabestary 等人(2018) [36]則在 DQN 方法中將累積延滯值差作為獎勵，動作空間中則使用 8 種不同時相。相較於前者，Wei 等人(2018)則將各方向車道之車隊長度總和、停等時間總和、號誌狀態、前次動作路口通過車輛數、旅行時間總和加權為獎勵函數的計算。

相較於基於 DQN 之方法，號誌控制問題應用基於 Actor-Critic 的 DRL 方法之研究則相對較少。Mousavi 等人(2017) [34]應用 Actor-Critic 方法於單一路口的號誌控制問題，其中狀態輸入係根據 SUMO 之車流模擬影像，以 CNN 方式進行轉換以檢測各車道的車隊長度、車速、以及車輛位置，動作則設定為東西向或南北向之時相，獎勵則累積延滯差；該研究發現，相較於 DQN 方法，Actor-Critic 方法可找出更穩定的策略。Genders 等人(2018)則採取 A3C 方法應用於單一路口模擬，值得一提，該研究針對 3 種狀態輸入進行比較，分別是各車道佔有率與平均車速、各車道密度與停等車隊長度、以及車輛位置與車速，結果發現僅藉由傳統方式、及路側偵測器蒐集之數據所訓練之模型績效並未差於使用複雜偵測器作為狀態輸入的模型。

過往應用車流模擬軟體於國內場域實作時，其中面臨最大的困難點之一為國內特有混合車流型態難以描述，故亦須進行應用 DRL 方法之國內文獻回顧。許添本等人(2020) [57]在單一路口情況下，採用 DQN 結合 Double Q-learning、競爭網路架構(Dueling Network Architecture)及優先經驗回放之技術方法，每 5 秒決定是否切換時相；該篇研究透過 CNN 將路口各方向進行格位轉換來建立 Q 網路作為環境狀態輸入，動作考量固定時相順序下時相是否切換，獎勵則考量車輛停等時間及路口通過量所組成之加權指標，最終發現基於 DQN 方法可運用於國內混合車流環境，並即時因應各種車流變化。許添本等人(2021) [58]則為克服 DRL 需大量訓練時間及有時會有訓練效率不佳的問題，採用了示範式 DQN 方法來處理單一路口號誌控制問題。該篇研究將狀態資訊以格位方式輸入，以便包含汽車與機車之位置與速度及當前時相；動作選擇係延長當前時相或切換至下一時相；獎勵函數由通過路口車輛數、停等時間與時相切換懲罰加權組成，並以最大化回合累積獎勵為目標。最終發現績效優於現有定時時制與觸動式號誌控制邏輯。

胡大瀛等人(2021) [56]則針對 DRL 應用於單一路口之路口號誌控制，其不同獎勵機制對號誌 Agent 的學習過程與訓練成果進行分析，該研究設計三種環境回饋號誌 Agent 決策的獎勵機制，分別如下：

- (一)第一種為車隊等候長度，即方向停止線前車輛數，將綠燈方向停止線前車輛數減去紅燈方向停止線前車輛數，若綠燈方向停止線前的車輛數大於紅燈方向的車輛數，則

環境獎勵號誌代理一個正值，表示此動作策略能夠提供更多用路人通行此路口，提升整體效率。

(二)第二種為路口通行量，即於號誌代理執行動作開啟設置的直右綠燈 15 秒或延長直右綠燈 5 秒內，統計進入路口的車輛數，並與最後一秒紅燈方向停止於停止線前的車輛數相減，若路口通行量大於紅燈方向停止的車輛數，則環境獎勵號誌代理一個正值，表示此動作策略能提供更多的用路人通行路口，提升整體效益。

(三)第三種為計算每一輛車於路口停止等待的時間並做加總，利用路口全部車輛累積停等時間做為設計，將上一回合動作策略所導致的車輛總累積停等時間減去此回合動作策略所導致的車輛總累積停等時間，而每輛車於該路口的累積停等時間會持續存在或增加，直到離開路口，故此回合的累積停等時間會較上一回合大而形成負的獎勵值，透過持續持續縮小兩回合間的累積停等時間，使獎勵值能夠負的愈少，提升效益。

二、多路口號誌控制

應用 DRL 於多路口號誌控制問題時，除訓練出具好的控制策略的 Agent 外，其中一個相當關鍵的議題為如何使多個 Agent 相互合作以進行決策，以最佳化大範圍路網總體績效，此類型問題即為多代理人深度強化學習(Multi-Agent Deep Reinforcement Learning, MADRL)。在處理這類型問題時，主要會面臨 2 個挑戰：第一為在多個 Agent 情況中，對某一 Agent 而言，因其主要互動環境包含其餘 Agent，故執行動作時所需觀察的環境變化不僅是該路口之狀態，亦包括其餘 Agent 的動作變化，導致該 Agent 更新策略參數時所欲考慮的複雜度更高；第二為多路口號誌控制環境下狀態空間與動作空間的高維度與複雜度，將隨著路口數量呈指數性成長，係因該問題目標為整體路網之績效最大化，故須考慮每一路口 Agent 之聯合狀態空間與聯合動作空間。為解決上述問題，過往文獻通常透過資訊溝通、協調機制、或調整價值網路及策略網路架構以克服。

過往文獻中對於應用 MADRL 至多路口號誌控制問題上時，其中一個常見的概念為將每一路口分別以獨立之 DRL 方法訓練一個 Agent，Agent 間相互溝通資訊，使各 Agent 得以利用鄰近 Agent 之訊息進行協調控制。而常見的資訊如路口下游狀態(Chu et al., 2019 [8]; Dresner et al., 2006 [11]; El-Tantawy et al., 2013 [12]; Da Silva et al., 2006 [9]; Arel et al., 2010 [4])、鄰近路口所有狀態(Arel et al., 2010 [4]; Nishi et al., 2018 [35])、鄰近路口隱藏狀態(Tan et al., 2019) [42]等。考量將各路口分別視為獨立 Agent 以進行訓練之 MADRL 方法較不容易收斂，Chu 等人(2019)提出多代理人 A2C 方法(Multi-Agent A2C)，該篇研究引入兩種技術以改善前述之問題，第一為在 Agent 狀態輸入加入鄰近 Agent 的策略資訊，使之能了解鄰近 Agent 之車流狀態及協調策略的概念；第二為在目標函數中加入空間折扣因子，按照路口間距離遠近加權鄰近 Agent 之獎勵與狀態。

另個常見的概念為調整價值網路及策略網路架構，Tan 等人(2019)將大範圍路網將大範圍路網分割成數個小範圍路網後，用一個類似於 DQN 的 Per-action DQN 來估算集中化的 Q 函數值，用 Wolpertinger-DDPG 去最佳化每個小範圍路網的策略函數。Li 等人(2021) [27]則是在此架構下設計了一個知識分享機制，將所有 Agent 從環境中得到的資訊存入一個知識容器 (Knowledge Container) 中，進而讓每個 Agent 能夠從中獲得環境資訊，並透過自己的方式去分析重組，得到一組自己特有的知識資訊，再加入本身觀測到的狀態資料一起去進行決策。

2.2.2 強化學習應用於交流道協控

2.2.2.1 強化學習於交流道協控模型

有關於強化學習如何實作於號誌控制，詳述於 2.2.1 小節。而其中提及兩個常見的應用方式，單一代理人應用於單一或少量路口，並使其間隔固定時段決策路口時相，亦或透過上一個時相判斷下一個時相的配置；多代理人大多應用於較大的路網配置，其概念為各路口配置單一代理人，再透過獎勵各代理人彼此考量績效的方式，達到路口間的連鎖關係，而常見的獎勵函數包括通過路口車輛數、車隊排隊長度(亦或停等延滯)等。

惟多代理人的運算複雜程度，隨時間及空間的增長而成非線性增加，故多代理人的實際應用仍多應用於模擬軟體階段，跨單位及跨不同公路系統的協控，具有更複雜的求解程度，其文獻數量亦不多。以下兩則文獻，一為 Lu (2017) [29] 所著，其使用 Q-learning 演算法，考量的狀態為主線車流量及匝道車流量，動作為離散儀控率，獎勵目標為最小化路網車輛數。二為 Yu(2022) [59]，使用多代理學習獨立 Q 網路，考量狀態為主線儀控上下游的速率及佔有率、匝道儀控率，動作為離散儀控率，而由於該文獻選定地點為國道五號雪山隧道口，考量地緣關係，獎勵設為最大化雪隧口流量。

一、Coordinated Ramp Metering with Equity Consideration Using Reinforcement Learning (Chao Lu et al, 2017) [29]

Lu(2017)提及過往的匝道儀控，大多犧牲了儀控率以換取高速公路主線暢通。當主線近於壅塞時，儀控策略大多為縮減儀控率，避免主線密度超過宏觀車流理論中的密度臨界值。而該篇文獻中，透過設定強化學習的 reward 為最小化主線的旅行時間以達成主線優化的目的。其特點為考量車輛數時，主線道及匝道車輛數皆考量進去，且在目標式中設定權重相等，代表著犧牲匝道儀控率的情況會降低。該篇文獻中所論述 equity，目的是不論尖峰或離峰，車輛停等的時間標準差不應太大。該文獻之強化學習模型應用場域為英國 M6 高速公路的某一區段，最後得出之改善結論如下表 2.2.2.1-1 所示，主線旅行時間下降約 18%。且加入 equity 因子後，各匝道停等車輛的等候時間變得相近，匝道上的車隊長度亦有減少，且過往尖峰等候時間峰值情況亦有改善，如圖 2.2.2.1-1~2.2.2.1-2 所示。

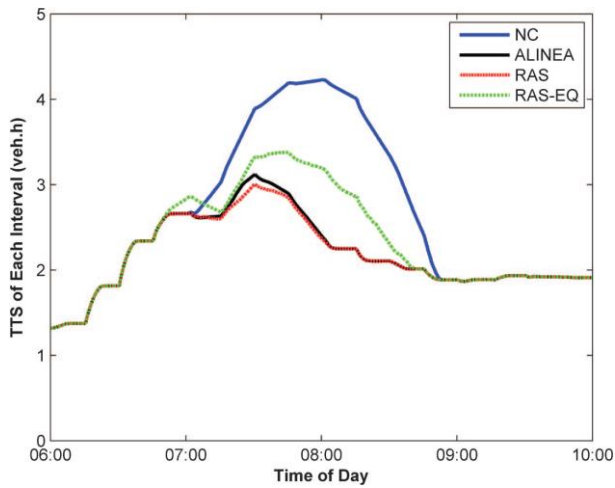


圖 2.2.2.1-1 TTS result

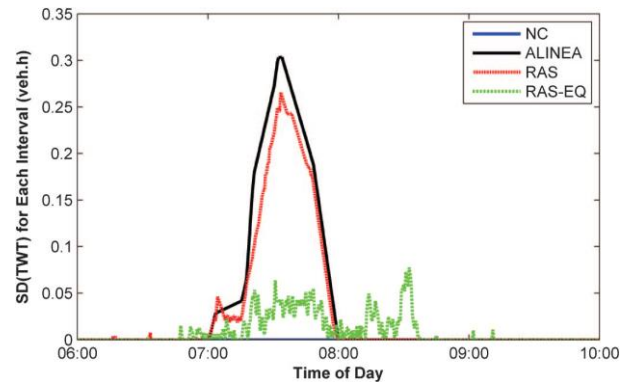


圖 2.2.2.1-2 SD(TWT) result

資料來源：(Lu, 2017) [29]

表 2.2.2.1-1 TTS 與 SD(TWT)於各策略的結果

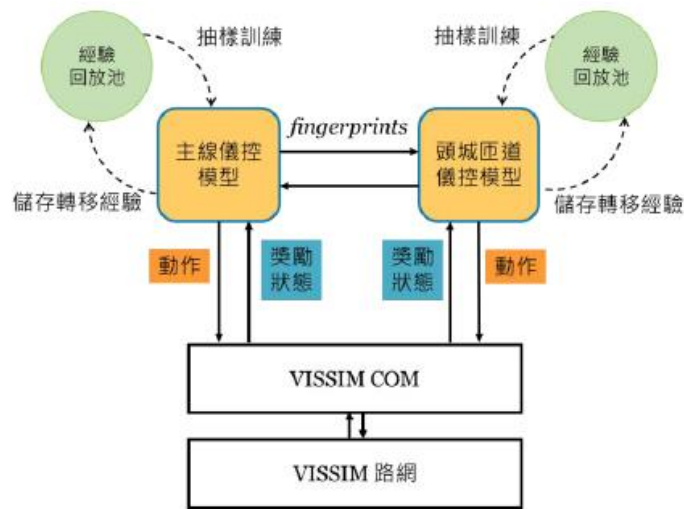
Different situations	NC	ALINEA	RAS	RAS-EQ
TTS (veh.h)	2,561.0	2,103.3	2,088.3	2,254.3
TTS reduction (%)	—	17.9	18.5	12.0
SD (TWT) (veh.h)	0.0	35.1	28.5	0.1

資料來源：(Lu, 2017)

其中 TTS 為總旅行時間(total travel time)，SD(TWT)為總停等時間(total waiting time)的標準差，NC 為原始狀況(no control)，ALINEA 為 Papageorgiou 等人於 1991 年提出的即時回饋式動態儀控策略，RAS 為一強化學習的匝道儀控策略，RAS_EQ 則為該文獻提出之平等 RAS 控制策略。

二、應用多代理深度強化學習於高速公路主線及匝道聯合儀控策略——以國道五號為例(陳又均) [59]

Yu(2022)應用強化學習，建立國道五號主線與匝道協控模型。其中考量的狀態有主線的速率及車流量數以及匝道儀控率，為避免 Q learning 可能因多目標發生維度爆炸的問題，而採用獨立 Q 學習(Independent Q learning)，一種分散式強化學習法，藉由將其他 agent 行為考量進狀態的一部份，避免因維度而使模型運作不穩定，該文獻之模型運作模型如圖 2.2.2.1-3 所示。



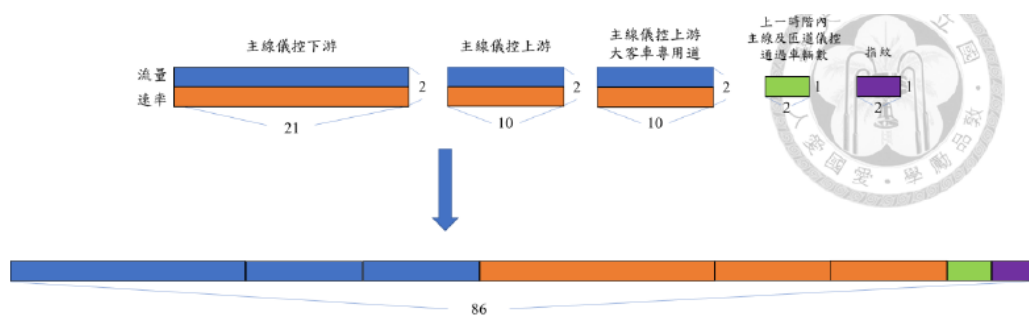
資料來源：(Yu, 2022) [59]

圖 2.2.2.1-3 高速公路主線與匝道協控策略模型示意圖

在強化學習所需要設定的狀態、動作及獎勵中，該文獻分別做以下考量：

(一)狀態

由於協控為多目標模型，故採用獨立 Q 學習來放寬維度上的限制。該文獻以流量及速率作為環境狀態，狀態範圍分為三個部分考量，分別為主線儀控上游、主線儀控下游以及上游之大客車專用道，並將上一時階內主線及匝道的通過車輛數作為狀態的一部份，如圖 2.2.2.1-4 所示。



資料來源：(Yu, 2022) [59]

圖 2.2.2.1-4 狀態輸入示意圖

(二)動作

該文獻參考過往匝道儀控策略為離散儀控，故設定為每五分鐘一次的定時定時動作，調整儀控綠與現行的國道五號匝道儀控率一致。

(三)獎勵

強化學習的獎勵針對不同應用個案，獎勵設定也會不同。過往小範圍控制策略多採用最小化旅行時間亦或最小化壅塞車輛數。該文獻針對國道五號，設定的獎勵為每一時階最大通過瓶頸流量，並以現行流量作為獎勵比較基準。該瓶頸點為國五 26.705K，

雪隧北上處。

此文獻最後輸出績效，可發現雖將匝道儀控率列為狀態的一部份作為考量，主體仍是高速公路主線道的疏解情形，故在主線的旅行速率及流量皆有提升，主線儀控率偏高；而匝道儀控率上下跳動，但整體平均儀控率仍偏低，故匝道的等候車隊長度較現況長，其輸出結果如圖 2.2.2.1-5~2.2.2.1-6 所示。

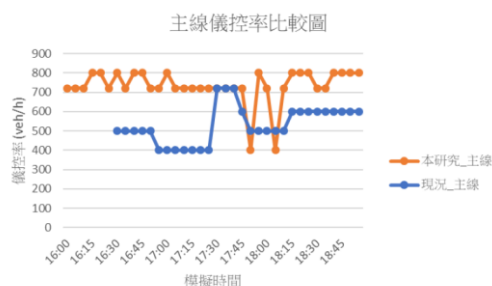


圖 2.2.2.1-5 主線儀控率

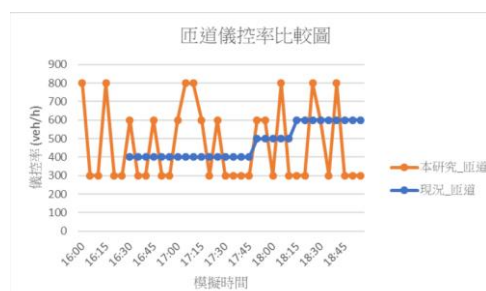
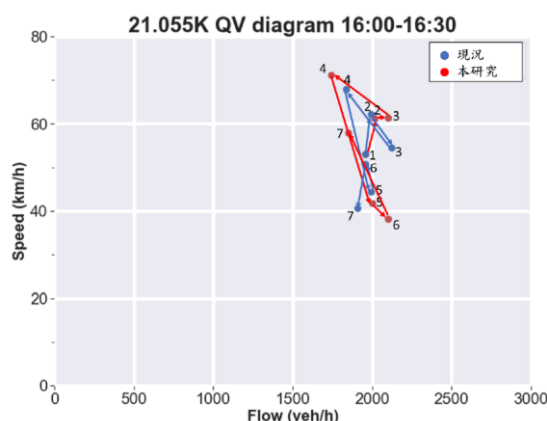


圖 2.2.2.1-6 匝道儀控率



資料來源：(Yu, 2022) [59]

圖 2.2.2.1-7 主線速率流量關係圖

2.2.2.2 強化學習於號誌協控應用之系統架構

一、人工智慧強化學習號誌協控模型發展構想

本計畫針對交流道區域，將研提高速公路交流道區域匝道儀控與平面路口號誌整合之人工智慧強化學習號誌協控模型發展構想與進行初步設計。模型發展與設計過程評估導入多號誌控制代理人機制，並於績效評估上額外考量駕駛人感受等相關指標。基於對單一平面號誌的學習模式，本計畫考慮將此模式推廣至平面號誌與匝道儀控的協同控制應用場景，對於多個智能體在合作和競爭的環境中進行學習。其整體訓練主要有以下兩種方法：

(一)集中式控制 (centralized control)

採用集中式控制方式進行匝道與平面協控，此模式僅有單一代理人者，而學習者可同時取得匝道與平面路口之狀態且能同時決定兩地之號誌。此模式透過增加狀態描述

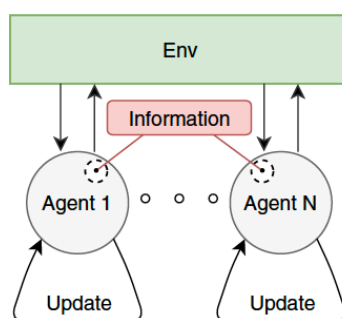
的維度與動作即可運用 SAC 與 Q-Pensieve 方法來訓練。值得一提的是，多目標方法如 Q-Pensieve 在此設定提供相當多彈性，比如若想考慮多種匝道與平面路口的偏好權衡，我們可設計多目標獎勵函數使其同時包含匝道與平面之獎勵訊號。

(二)分散式控制 (decentralized control)

透過 multi-agent RL 達成分散式控制 (decentralized control) 之匝道與平面協控，場景有兩個學習者 (分別可控制匝道儀控與平面號誌)，在測試階段各學習者僅能取得局部之狀態資訊，而無法得知全局之狀態。

在深度強化學習領域目前最主流的做法為 Centralized Training and Decentralized Execution (CTDE) 架構，其架構如圖 2.3.2.2-1，每個智能體會各自產生策略與環境互動 (因此稱為 decentralized execution)，而每個智能體在訓練期間更新其策略時會共享狀態資訊 (因此稱為 centralized training)。

當前已有許多在強化學習領域使用 CTDE 架構的研究工作。這些工作根據不同的訓練架構可分為多個類別。首先，第一個類別是參數共享，例如 Ahilan 和 Dayan 等人的研究，他們主要是通過訓練多個代理人策略，透過不同獨立訓練體之間的參數共享，以提高訓練效率[Ahilan et al., 2019[3]; Sunehag et al., 2018[40]]。另一種則為基於價值的方法，Foerster 等人提出將 deep Q-learning 透過 CTDE 的方式推廣至多智能體強化學習 [Foerster et al., 2016[13]; Foerster et al., 2017[14]]，此外 Lowe 等人提出將標準的 DDPG 透過 CTDE 的方式推廣為 Multi-Agent DDPG (MADDPG)，藉由訊息共享，Q 網路能學習到全局的預期獎勵，通過此 Q 函數，每個獨立的策略網路都能學到全局的最佳解。最後則是 Master-slave 架構，Kumar 和 Kong 等人於強化學習中導入 Master-slave 架構[Kumar et al., 2017[24]; Kong et al., 2017[23]]，中央的訓練員負責為每個從屬的訓練體提供 action。每個從屬訓練體則通過其局部觀察狀態來微調中央指導的行動，以實現合作訓練的目標。更詳細的 CTDE 文獻介紹可參考 [Gronauer and Diepold, 2022][16]。

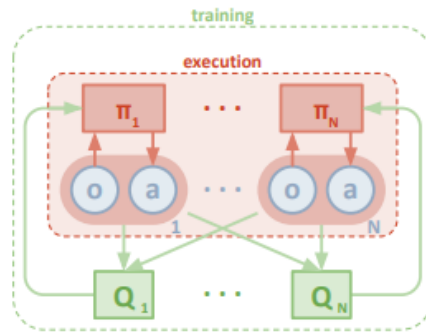


資料來源：(Gronauer and Diepold, 2022) [16]

圖 2.2.2.2-1 CTDE 多智能體強化學習架構圖

為了更具體說明 CTDE 之特點，以下進一步用 MADDPG 為例並配合圖 2.2.2.2-2 來描述 CTDE 架構：

- (一)在策略網路部分，每個學習者有各自的 policy network，而此網路僅使用局部狀態作為輸入並輸出該號誌之動作(以 $\pi(1)(s(1)), \pi(2)(s(2))$ 表示)。
- (二)在 Q 網路部分，將使用全局的狀態與兩個學習者的動作作為輸入，即可表示為 $Q(s(1), s(2), \pi(1)(s(1)), \pi(2)(s(2)))$ 。



資料來源：(Lowe et al., 2017)

圖 2.2.2.2-2 MADDPG 之 CTDE 架構示意圖

二、小結

於高速公路交流道區域匝道儀控與平面路口號誌整合中，我們需要有一個中央的 critic 能針對匝道和平面路口的協控去最佳提供訓練目標，因此我們導入了 CTDE 的訓練架構。由於 Q 網路僅在訓練階段會使用而不會參與局部測試階段，故此設計並不違背前述之設計，因此 CTDE 模式可與 SAC 和 Q-Pensieve 都相容，可以自然地推廣單目標與多目標強化學習演算法。

2.2.3 我國應用人工智慧或車聯網於號誌控制相關研究

一、應用人工智慧技術進行交通數據蒐集暨號誌控制之研究(109 年)

過往傳統適應性號誌控制策略常面臨 3 種困境，以致難以推廣應用，包括：

- (一)若欲進行適應性號誌控制策略，其偵測器之運作穩定度需達一定水準，而傳統車輛偵測器又以環路線圈式運作最穩定，惟其維修與成本均較高；此外，策略實施亦受限於所蒐集之交通參數種類與品質，致使控制成效難有明顯突破。
- (二)目前仍無通用之數學控制邏輯模式來計算處理適應性控制問題，
- (三)傳統適應性控制策略實施後，路口車流特性亦將會有相關變化，而原先規劃之參數恐已無法符合實施後之車流特性，進而使績效不佳，故實施後之參數亦應依照即時資訊加以調整。

為解決上述困境，該研究藉由模仿學習(Imitation Learning, IL)與以週期為基礎的 AI 強化學習，來選擇最適號誌時制計畫，以及下載至路口執行，以即時反應車流變化與提高號誌化路口運作效率。該研究選定前期於臺中市太平區太平匝道下游交通環境較單純之樂業路與十甲東路、樂業路與東英路兩路口作為實驗場域，並架設 AI 影像偵測設備、e-Tag 讀取器、號誌控制器及 AI IPC，並以 4G 無線通訊方式進行資料傳輸及號誌控制指令下傳。

該研究考量強化學習於尋找最適號誌控制時，需耗費大量試誤的探索時間才會收斂，致使較無實務環境之應用案例，故利用時制最佳化分析軟體 PaSO，先行針對實驗範圍產生考量幹道續進之最佳化時制計畫，以及時段參數來進行 AI 強化學習，最後將學習後之 AI 號誌控制模式，於現場進行實際運作。該研究事先透過微觀交通模擬軟體 SUMO 建構實驗場域模擬環境，進而產生 AI 號誌控制之模擬績效，最後在實測場域上進行實測。實測結果顯示，可改善實驗場域尖峰時段車流績效均達 10%以上。

二、構建 5G 智慧交通數位神經中樞(1/2)－功能架構探討與系統規劃(110 年)

該研究基於國內對運用 5G 結合 AIoT 等創新科技之需求日益茁壯，期望透過 5G 高速傳輸加速跨域即時交通資料蒐集，降低資訊傳遞落差，並針對交通管理及服務進行技術盤點，開發人工智慧(AI)及大數據管理決策核心技術，迅速掌握城市人流、車流、公共運輸、交通號誌等系統即時狀況，提供即時精確的交通管理決策。為此，該研究旨在進行我國「智慧交通數位神經中樞」之都市交通控制通訊協定檢討，並遴選實作示範場域之縣市、進行整體場域之設備佈建規劃，以及進行緊急車輛之智慧號誌控制系統規劃，並整合 C-V2X 進行路口號誌控制之概念性驗證。

該研究最終建議「智慧交通數位神經中樞系統」未來必須建立端、網、霧、雲的標準平台，在這平台下須透過介面標準進行傳輸，而這樣的選用是開放的並且兼顧資安，雲平台是共享可擴充的網路且必須是安全的。未來將可透過 5G 網路，有效處理邊緣運算，無縫接軌各式各項的運輸系統雲端系統(公有雲)及營運公司雲(私有雲)，做好各類系統混和雲服務，亦可將所有整合、預測、分析、自主反應後的所有資訊傳輸到雲平台，將各項資訊整合及訊息展示，在通訊協定，或者在端設備的整合，則應同步考量相容性和整合性，進而做出「智慧交通數位神經中樞系統」的智慧決策。

除針對都市交通控制通訊協定提出檢討成果外，該研究亦在緊急車輛之智慧號誌控制系統發展之規劃與概念性驗證中提出建議，因場域以單一路口進行測試，未來將擴展至連續兩個路口，表示 C-V2X 範圍可能涵蓋兩路口，在車輛端接受資訊需透過 GPS 確認自身位置後進行資訊過濾，確保接收正確資訊；此外，車輛端針對預估抵達路口時間需要導入較多評估與計算，例如透過 SPAT 資訊計算路口尚未紓解車輛所需秒數，此外若是採取在一時間點同時進行兩路口預估抵達時間的方式便須導入通過第一路口的時制計畫進行計算，確保 A 路口紅綠燈停等時間不會影響 B 路口的預估抵達時間。因此，整體的觸動機制與預估抵達時間息息相關，需要通盤考量設計。

三、我國人工智慧車聯網之號誌控制模式探討(111 年)

根據上述兩案累積之研究基礎，該研究探討符合我國交通運行環境之人工智慧號誌控制模型，藉由與道路管理單位整合實作，支援其號誌控制與管理需求，期望能產生實際成效，提升車流運作效率及交通安全；此外，該研究以六都場域為目標，與管理單位交流並進行實地訪查評估，選擇於適當情境進行以下三類型場域模擬人工智慧導入號誌控制之實驗測試，包括：(1)單一路口多時相、(2)連鎖幹道、(3)車聯網環境之單一路口，地點包含臺南市「台 86-19 甲」單一路口、臺北市「中山北路-德行東路」3 個路口幹道、高雄市「台 88 鳳山出口」(過埤路-鳳頂路) 單一路口。除在上述三處場域進行交通調查，蒐集場域交通特性資料以及本計畫人工智慧號誌控制設計所需之資料項目，例如：交通量、停等延滯等，並同時採用交通模擬軟體 VISSIM 與開源模擬軟體 SUMO 分別依所蒐集場域資料建模與校估參數，作為後續人工智慧決策模型參數訓練與績效模擬測試的軟體環境。

該研究著重於車聯網所能提供的即時資料，探討如何將該資料應用於 RL 的動態號誌控制層面。該研究所提出之人工智慧號誌控制方法將能依據各路口各方向交通變化即時演算號誌時制，而非在未知即時交通狀態下，以事先產生方式來給定號誌時制，因此號誌計畫之週期長度、續進寬帶應可動態彈性調整，以因應交通車流狀況的即時變異。

該研究結論可分為 6 個面向，以下分別進行說明：

(一)文獻回顧與方法論

該研究界定適應性號誌控制問題必須反應即時(real time)交通狀況的變異，此變異的問題本質主要在於「系統狀態的不確定性」與「具有時空傳導作用的車流動態性」。此外，若這些方法沒有考量「系統績效」最適化，就可能會得到經驗法則式的局部劣解(inferior solution from rule of thumb)，績效在空間上的呈現有可能產生扭曲。同時強化學習結合隨機性(stochastic)微觀車流模擬模型，以機率模型描述車流系統狀態的不確定性，並在這樣的運作情境下構建決策模型、求解決策模型，使所產出的決策模型具備優化系統績效「期望值 (expectation)」的能力，突破多數既有號誌控制方法為「確定性」模型或計算方式的限制。

(二)交通調查

傳統上在進行交通號誌改善，不論是交通影響評估或是號誌時制重整，皆會針對欲分析的各種交通情境進行完整的交通調查，以利在模擬模型中進行各方案的參數校估。而在該研究中由於考量到強化學習是學習環境中的車流模型因果關係與其趨勢，因此嘗試以調查部分具有代表性的交通情境作為模擬器參數校估的資料來源。而其他不包含在交通調查時段內的交通情境，則可透過變動已完成校估的模擬器產出資料供強化學習進行訓練與測試。

(三)交通模擬模型構建

強化學習方法論克服傳統號誌方法論上的隨機性問題。而為使該特色最大程度發揮與盡可能的貼近實際車流狀況，採用微觀車流模擬模型的作法是最佳的作法。在構建模擬模型的過程中，除按照流成進行相關參數的校估、確保模擬與實際誤差在一定程度之外，亦須確保動態號誌控制的燈號轉換指令是否可行、是否符合現況通訊協定所能滿足，以及在存取即時交通資料上是否正確。

(四)人工智慧強化學習 DDPG 與 PPO 演算法號誌控制應用

在強化學習演算法的選擇上，針對號誌控制議題應具備處理連續性決策(scalar action)的能力。該研究中所提到的兩種強化學習演算法，DDPG (Deep Deterministic Policy Gradient)與 PPO (Proximal Policy Optimization)皆為能夠應用連續性決策的演算法，改進了過去國內強化學習號誌控制常用 DQN 離散型決策。並且在此基礎之上，能將演算法採用多代理人作法，應用至多路口場域當中。測試結果顯示在模擬環境中，不論是 DDPG 或是 PPO 在號誌問題上皆可取得良好的績效改善表現，代表該研究中對於強化學習的狀態、決策、獎勵設定合理，兩者都具備持續發展研究的價值。

(五)訓練模型設計

該研究強化學習方案所探討的狀態、決策、獎勵、演算法列於表 2.2.3-1。

表 2.2.3-1 人工智慧強化學習號誌控制個環節參數

狀態	決策	獎勵	演算法
• 分車道分車種車輛數 • 轉向交通量 • 停等車隊長度 • 車輛數偵測器分布位置與數量 • 聯網車座標與速度 • 鄰近路口號誌資訊	• 逐時相決策下一時相長度 • 逐週期決策下一週期所有時相長度	• 系統總停等延滯 • 系統總通過車輛數	• DDPG • PPO

資料來源：我國人工智慧車聯網之號誌控制模式探討(民國 111 年)

過去在探討適應性控制實作時，其中一項關切的議題即是偵測器數量。偵測器數量直接關係到建置與後續養護的成本。在該研究中發現到對於強化學習方法來說，比起偵測器數量，佈設的位置似乎更為關鍵。在透過專家判斷適當佈設位置後，可以減少偵測器的數量需求同時保留一定程度的改善成效。並且在市區場域中由於控制路口與邊界路口距離相當接近，因此邊界路口的即時號誌資訊對於強化學習的控制是至關重要的。

(六)績效測試

既有號誌控制邏輯、方法或模型常以假設交通系統為確定性方式提出，因此，在建置之前實應先以模擬方式測試其績效，以確保其在具有不確定性的情境下是否能如預

期的發揮改善的作用及其堅固性。模擬績效測試應以隨機性車流模擬情境為測試基礎，至少應包含平/例假日晨昏尖離峰等 8 種情境測試，且每一情境至少應包含高變異與低變異兩種呈現不確定性的子情境。另外雖該研究中尚未達到實測階段，但仍建議實測階段的事前事後實測績效進行每日 24 小時實測，並持續至少一個月以上，採取動態控制與既有時制(已經時制重整後)隔日互換測試的方式進行，以確保事前事後測試的情境是相近的。

此外，該研究透過格位轉化作法，將個別聯網車的離散資料轉化為結構化的強化學習輸入資料來源，並採用卷積層網路(CNN)結構萃取資料特徵，在模擬環境中的實驗顯示此種作法之可行性。由於連網車資料具有某些偵測器資料較難提供的特徵，因此預期未來以連網車資料作為強化學習號誌控制資料來源仍是相當具有潛力的作法。

該研究最後提出 3 點建議如下：

- 1.該研究中人工智慧強化學習號誌控制應用是在既有時段型態與時相組態下，產生多個尖離峰時段的號誌控制強化學習模型及其訓練成果，後續可以探討如何自動依車流情境，來動態產生時相組態及其對應之最適化號誌控制機制與內容。
- 2.在號誌控制的最高最小綠燈限制上，其中最高綠燈秒數的限制，可透過將路口各方向的平均停等延滯或服務水準指標納入獎勵函數來放鬆限制，避免 RL 在只考慮系統總延滯下，因支道車輛數稀少而將該方向紅燈秒數設計超長，以此來達到減少系統總延滯的目的。而在最小綠燈秒數方面，則是可以透過在路口佈設行人偵測設備，提供欲通過路口的行人數量與類型等資訊，並將行人通過路口的績效納入獎勵函數中讓 RL 進行訓練，便有機會解決最小綠燈秒數的限制議題。
- 3.針對臺南市場域的交通情境設定上，該研究以平日與例假日做為目標。而在連接國道關廟交流道處經常實施匝道儀控，將可能發生本路口紓解率下降，致使強化學習效益打折扣。後續研究可以考量匝道儀控率的影響，並納入回堵情況的樣本供強化學習模型進行訓練。

2.3 人工智慧強化學習(Reinforcement learning, RL)模式

2.3.1 強化學習架構

強化學習是一種機器學習方法，目的在訓練學習者 (learner) 從未知且具有隨機性的動態環境 (dynamical environment) 中學習最佳順序性策略以達成特定目標，該方法通過讓學習者在環境中不斷地進行試驗和錯誤，從而學習什麼是正確的動作，進而將其與環境相關聯以實現目標。具體來說，強化學習是能讓學習者透過對環境進行一系列的動作，以使得獲得的累積獎勵 (cumulative reward) 最大化的學習方式。學習的目標為以迭代方式持續改善目前的策略 (policy)，透過不斷的訓練，期望能夠獲得最佳策略 (optimal policy)。學習方式是透過學習

者對環境執行動作 (action) 後，環境會回傳獎勵 (reward) 和新的狀態 (state)，不斷進行此程序以逐漸優化策略，如下圖 2.3.1-1。在交通號誌控制問題場景，強化學習可以透過訓練來適應不同的交通情況，例如調整紅綠燈的時間，以達到最優化的交通流暢度和效率。

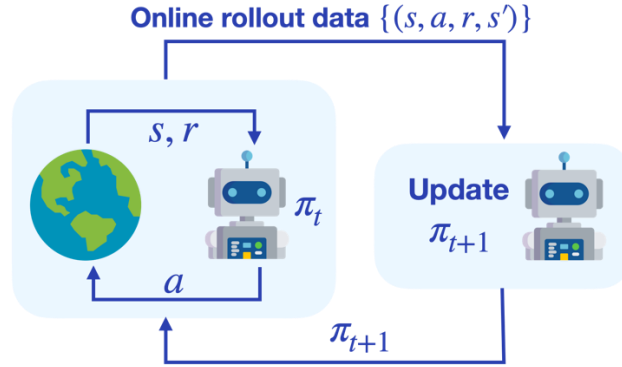


圖 2.3.1-1 強化學習概念示意圖

強化學習最常使用之動態環境模型為馬可夫決策過程 (Markov Decision Process，簡稱 MDP)，而標準的 MDP 包含以下要素(Bellman, 1957) [5]：

- 一、環境狀態集合 S ，在每個時刻 t 環境的狀態表示為 $S_t \in S$
- 二、動作集合 A ，為學習者能採取之動作，在觀察到當前狀態 S_t 後，學習者會選擇一個行動 $A_t \in A$ 。
- 三、獎勵函數 $R(\cdot, \cdot): S \times A \rightarrow R$ 其定義為 $R(s, a) = E[R_{t+1} | S_t = s, A_t = a]$ ，描述在狀態 s 採取行動 a 後預期獲得的立即獎勵期望值，此外使用 R_{t+1} 來描述在每個時刻 t 下，觀察狀態 S_t 後採取 A_t 所得到之獎勵。
- 四、狀態轉移函數 $P(\cdot, \cdot, \cdot): S \times A \times S \rightarrow [0, 1]$ ，其定義為 $P(s, a, s') = P[S_{t+1} = s' | S_t = s, A_t = a]$ ，滿足馬可夫性質 (Markov property)，亦即當下狀態的轉移機率僅與當下狀態與動作有關而與其他歷史資訊無關。
- 五、初始狀態分佈 ρ 。
- 六、 $\gamma \in [0, 1]$ 為獎勵折扣因子，旨在反映未來獎勵的「現在」價值。

為評估學得策略，強化學習常使用以下狀態價值函數 (state-value function) 與動作價值函數 (action-value function) 於演算法設計：

$$V^\pi(s) := E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{k+1} | S_0 = s]$$

$$Q^\pi(s, a) := E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{k+1} | S_0 = s, A_0 = a]$$

以上兩個價值函數皆在反映在某狀態下採取該策略之長期累積獎勵。從價值函數的角度出發，強化學習的目標也可描述為尋找能使價值函數最大化之策略，亦即尋找 $Q^*(s, a) = \max_{\pi} Q(s, a)$ 或 $V^*(s) = \max_{\pi} V(s)$ 。或以隨機梯度下降法(stochastic gradient descent)，使系統總停等延滯期望值最小化的方向更新類神經網路參數 θ 。

以上是基於強化學習(Reinforcement Learning, 後簡稱 RL)方法論之應用，而在本案中所採用之演算法在 RL 領域中的歸類屬於類似 Actor-Critic 的學習架構，因此需要明確定義狀態(State)、決策(Action)、獎勵(Reward)，方能使整套方法合理運作。上述 RL 號誌控制訓練演算基本模型如圖 2.3.1-2 所示，而前期強化學習演算法的詳細說明如 3.1 節說明。本案現階段的強化學習演算法的詳細說明則會在 3.2 節說明。

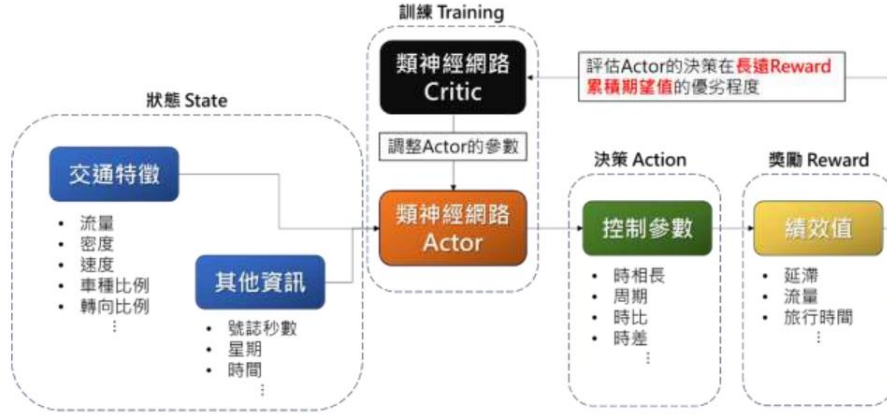


圖 2.3.1-2 號誌控制架構基本模型

資料來源：[\[5\]](#)

2.3.2 強化學習演算法

強化學習方法可根據切入角度大致可分為以下幾個主要類別：

一、基於動態模型之強化學習 (model-based RL)：

強化學習其中一個基礎且直觀的切入角度為對於 MDP 進行「基於學習之最佳控制」(learning-based control)。在標準的最佳化控制問題設定，控制器演算法的主要目標為在給定環境的動態模型資訊的情形下，透過迭代式的演算法進行規劃 (planning) 以逐步找出最佳決策，其中最常見的方式包含策略迭代法 (policy iteration) 與價值迭代法 (value iteration)，這類迭代方法皆基於前一章節介紹之 Bellman equations 之原理來設計 (Sutton & Barto, 2018) [\[39\]](#)。由於強化學習需在未知環境中做決策，基於動態模型之強化學習會透過以下兩個主要程序之交替運作來改善決策：

(一)學習與估計環境之動態模型：首先，學習者先基於與環境互動所得到之軌跡資料來估計環境之狀態轉移機率以及獎勵函數，此項任務可透過監督式學習方法來達成 (Janner, Fu, Zhang, & Levine, 2019) [\[22\]](#)，比如最大似然估計 (maximum likelihood estimation)。具體來說，若將估計之動態模型以 $\mathcal{M}_\theta$ 做參數化表示，則於 t 時間下之動態模型可估計為

$$\theta_t^{\text{MLE}} := \arg \max_{\theta} \prod_{i=1}^t p(s_i, s_{i+1}, a_i; \theta),$$

在實際應用中，此動態模型 $\mathcal{M}_\theta$ 會使用深度神經網路來表現與參數化，而上式之最

大值可透過常用的隨機梯度下降法來取得。

(二)基於估計之動態模型進行規劃：基於目前得到之 $\mathcal{M}_\theta$ ，學習者可使用任一最佳控制之規劃方法（比如價值迭代）來取得此模型對應之最佳決策。

儘管此類方法相當直觀，但基於動態模型之學習往往需要克服以下幾點挑戰。首先是「探索」問題，為了使動態模型的估計準確，必須確保學習者與環境互動取得之軌跡是足夠多元且包含環境之大部分區域，此一問題常透過使用強調探索之策略來收集軌跡或是引進樂觀估計 (optimism) 來解決，比如常使用的 Upper-Confidence RL 方法。再者，由於環境本身的隨機性，要足夠準確地估計動態模型往往相當困難，如何著重估計與最佳策略直接相關的環境部分是一大挑戰。

二、無動態模型之強化學習 (model-free RL)~價值型方法 (value-based methods)：

與基於模型的強化學習不同，價值型的強化學習並不仰賴動態模型估計，而是通過學習來逼近最佳之價值函數。比如，著名的 Q-learning 方法即為透過維持對最佳動作價值的估計來學習，具體作法為每個時刻基於所觀察得到之環境狀態與獎勵，進行以下迭代更新：

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_t(s_t, a_t) \left(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right)$$

其中 $\alpha_t(s_t, a_t)$ 為 learning rate。值得一提的是，Q-learning 方法有兩大特色：(1) 此方法不需使用動態模型，因此為 model-free，(2) 此方法可支援 off-policy 學習，亦即與環境互動之策略可與學習得到的策略不同，此特色可使強化學習在資料取得上更有彈性。在實際應用上，Q-learning 往往需配合深度神經網路來實現，此方法稱為深度 Q 網路 (Deep Q network, 或簡稱 DQN)，其具體作法為將原始的 Q-learning 更新轉換為最小化以下損失函數(Mnih et al., 2013)

[32]：

$$L_{DQN}(\phi) := E_{(s,a,s',r)} \left[\left(\left(r + \gamma \max_{a' \in A} Q(s', a'; \phi^-) \right) - Q(s, a; \phi) \right)^2 \right]$$

三、無動態模型強化學習(model-free RL)-策略最佳化方法(policy optimization methods)

與價值型方法不同，策略最佳化是將強化學習視為一個最佳化問題，亦即對總預期獎勵 $V^\pi(\rho) := E_{s \sim \rho}[V^\pi(s)]$ 進行最佳化。首先，策略最佳化方法從參數化策略開始，即 $\pi(a|s) \equiv \pi_\phi(a|s)$ ，其中 ϕ 為策略參數，在實際應用中目前最常用的策略參數化方式是使用深度神經網路 (Lillicrap, et al., 2016) [28]，常見的形式為

$$\pi_\phi(a|s) = \frac{\exp(f_\phi(s, a))}{\sum_{a' \in A} \exp(f_\phi(s, a'))}$$

其中 f_ϕ 代表深度神經網路之輸出值。至於更新策略的方式可舉一個例子來說明，目前廣泛使用的策略最佳化方法之一是隨機策略梯度演算法 (stochastic gradient ascent)，策略梯度方法

的基本理論結果是策略梯度定理 (policy gradient theorem)， $\nabla_{\phi} V_{\phi}^{\pi}(\rho)$ 具有封閉形式的表達如下：

$$\nabla_{\phi} V^{\pi_{\phi}}(\rho) = E_{\pi_{\phi}} [\nabla_{\phi} \log \pi_{\phi}(s, a) Q^{\pi_{\phi}}(s, a)]$$

為了取得或估計以上之策略梯度，我們還需要估計動作價值函數，例如可透過蒙特卡羅方法 (Monte Carlo methods) 或時間差分方法 (temporal difference methods) 來估計 (Mnih, et al., 2016)。相較於價值型方法，策略最佳化的主要優點有兩個：(1) 策略最佳化可以學習非確定性策略 (non-deterministic policy)，這是價值型方法 (如 Q-learning) 無法實現的。(2) 策略最佳化較容易處理連續動作控制 (continuous control) 問題，相較之下價值型方法之決策流程由於包含對動作價值函數取最大值而較不易處理連續動作，往往需要仰賴額外的離散化或是以抽樣方式來近似最大值。考量本計畫交通號誌調控為連續動作控制，後續的說明將著重介紹與本計畫執行相關之策略最佳化演算法。

四、DDPG (Deep Deterministic Policy Gradient, 深度確定性策略梯度法)

所謂 RL 訓練演算法，在此定義為從代理人(Actor)類神經網路初始化狀態，逐步更新其類神經網路中權重(weight)與偏差(bias)，使其達到最終能夠處理任務的一系列過程中所採用的計算方法，例如於 2.1.2 節中所介紹各種方法。在前期案中所採用的訓練演算法，是基於深度確定性策略梯度法(Deep Deterministic Policy Gradient)[10](以下簡稱 DDPG)訓練演算法，在 off-policy 的 Actor-Critic 架構下所加以修改成多代理人版本，並且亦加入了許多諸如目標網路、經驗回放等訓練技術，集結合成的獨特訓練演算法。在本節將會以既有文獻中最為接近之方法 DDPG，來做演算法的概念說明。

DDPG 是屬於 Actor-Critic 類的一種方法分支，其重點特徵在於「確定性(Deterministic)」的特性，該特性能更有效的提升傳統 Stochastic Policy Gradient 類的訓練效率。要了解這項特殊的作法之前，必須先了解原始 Actor-Critic 的基本背景架構。

$$E_p(X) = \int P_i X_i \quad \text{式(1)}$$

Actor 的目標是最大化累計獎勵的期望值，在機率分布 P 當中，某個隨機變數 X 的期望值可以用式(1)表示。其中 X_i 是所有 X 的可能性，而 P_i 則是對應該 X_i 的機率密度函數。而 Actor 的期望回報，就可用下式定義：

$$V(s) = E_{\pi}(s) [r + \gamma V(s')] \quad \text{式(2)}$$

其中 s 為當前狀態； s' 為下一個狀態； r 為這個轉移過程中所產生的立即獎勵； γ 為折減比例，介於 0~1 間； π 為當下 Actor 的 policy。式(2)即為著名的 Bellman equation 的估值函數。這時候可以將式(2)改定義一個 Q 值的函數：

$$Q(s, a) = r + \gamma V(s') \quad \text{式(3)}$$

式(3)中本質與式(2)是相同的概念，僅是將原本 policy π 的部分視為變數的一部分。為了衡量目前這個 policy π 的優劣程度，因此定義一個 $J(\pi)$ 的函數，來表示這個策略從起始狀態 s_0 開始到任務結束所能獲得的累積獎勵期望值。

$$J(\pi) = E(V(s_0)) \quad \text{式(4)}$$

梯度函數可以被表示為

$$\begin{aligned} \nabla_{\theta} J(\pi_{\theta}) &= \int_S \rho^{\pi}(s) \int_A \nabla_{\theta} \pi_{\theta}(a|s) Q^{\pi}(s, a) da ds \\ &= \mathbb{E}_{s \sim \rho^{\pi}, a \sim \pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a|s) Q^{\pi}(s, a)] \end{aligned} \quad \text{式(5) 引自 [10]}$$

式(5)[9]中， S 為狀態的集合空間；

$\rho^{\pi}(s)$ 為狀態在策略 π 下的機率分佈函數；

A 為決策的集合空間；

θ 為策略參數。

式(5)的梯度函數，是基於未來決策仍是隨機，或是說仍無法確定決策為何的狀態下，所推導出之梯度近似值。然而很直覺的，由於 Actor 的目標即是去獲取最大利益的一種策略，因此在式(5)中其實對於未來的策略未必是隨機的，而可以直接用 policy 去產出未來的決策。在此概念之下，梯度函數可以被改寫為：

$$\begin{aligned} \nabla_{\theta} J(\mu_{\theta}) &= \int_S \rho^{\mu}(s) \nabla_{\theta} \mu_{\theta}(s) \nabla_a Q^{\mu}(s, a) \Big|_{a=\mu_{\theta}(s)} ds \\ &= \mathbb{E}_{s \sim \rho^{\mu}} [\nabla_{\theta} \mu_{\theta}(s) \nabla_a Q^{\mu}(s, a) \Big|_{a=\mu_{\theta}(s)}], \end{aligned} \quad \text{式(6)}$$

推導過程可參考[9]

至此，便可以以 Actor-Critic 的架構來進行梯度下降的策略更新：

$$\delta_t = r_t + \gamma Q^{\omega}(s_{t+1}, \mu_{\theta}(s_{t+1})) - Q^{\omega}(s_t, a_t), \quad \text{式(7)}$$

$$\omega_{t+1} = \omega_t + \alpha_{\omega} \delta_t \nabla_{\omega} Q^{\omega}(s_t, a_t), \quad \text{式(8)}$$

$$\theta_{t+1} = \theta_t + \alpha_{\theta} \nabla_{\theta} \mu_{\theta}(s_t) \nabla_a Q^{\omega}(s_t, a_t) \Big|_{a=\mu_{\theta}(s_t)}, \quad \text{式(9)}$$

DDPG 的運作架構可參考圖 2.3.2-1。值得注意的是，輸入 Critic 與輸入 Actor 的 state 變數，兩者是相同的。此舉是為了確保 Actor 與 Critic 獲得的資訊量是相同的。若兩者所獲得的資訊程度不相同，例如 Critic 有額外的資訊輸入，則可能會造成 Critic 能夠判斷更新方向，但 Actor 卻無法判斷 state 情境的差異，致使 Actor 更新方向不明確的狀況發生。所以採用 Actor-Critic 訓練方法時，即使 Actor 與 Critic 的網路結構可能有巨大的差異，但輸入資訊則會盡可能的保持一致。在上述作法下，完整 DDPG 演算法可以圖 2.3.2-2 表示。

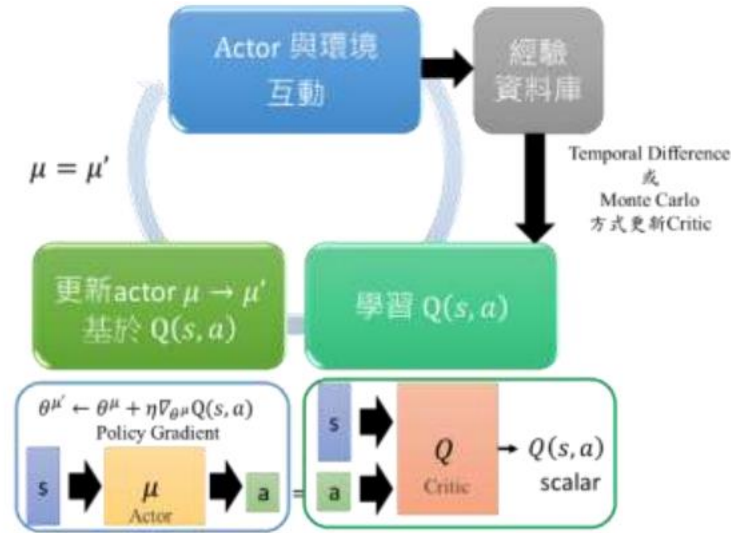


圖 2.3.2-1 DDPG 運作架構

Algorithm 1 Deep Deterministic Policy Gradient

- 1: Input: initial policy parameters θ , Q-function parameters ϕ , empty replay buffer $\mathcal{D}$
- 2: Set target parameters equal to main parameters $\theta_{\text{targ}} \leftarrow \theta$, $\phi_{\text{targ}} \leftarrow \phi$
- 3: **repeat**
- 4: Observe state s and select action $a = \text{clip}(\mu_{\theta}(s) + \epsilon, a_{\text{Low}}, a_{\text{High}})$, where $\epsilon \sim \mathcal{N}$
- 5: Execute a in the environment
- 6: Observe next state s' , reward r , and done signal d to indicate whether s' is terminal
- 7: Store (s, a, r, s', d) in replay buffer $\mathcal{D}$
- 8: If s' is terminal, reset environment state.
- 9: **if** it's time to update **then**
- 10: **for** however many updates **do**
- 11: Randomly sample a batch of transitions, $B = \{(s, a, r, s', d)\}$ from $\mathcal{D}$
- 12: Compute targets

$$y(r, s', d) = r + \gamma(1 - d)Q_{\phi_{\text{targ}}}(s', \mu_{\theta_{\text{targ}}}(s'))$$
- 13: Update Q-function by one step of gradient descent using

$$\nabla_{\phi} \frac{1}{|B|} \sum_{(s, a, r, s', d) \in B} (Q_{\phi}(s, a) - y(r, s', d))^2$$
- 14: Update policy by one step of gradient ascent using

$$\nabla_{\theta} \frac{1}{|B|} \sum_{s \in B} Q_{\phi}(s, \mu_{\theta}(s))$$
- 15: Update target networks with

$$\begin{aligned} \phi_{\text{targ}} &\leftarrow \rho \phi_{\text{targ}} + (1 - \rho) \phi \\ \theta_{\text{targ}} &\leftarrow \rho \theta_{\text{targ}} + (1 - \rho) \theta \end{aligned}$$
- 16: **end for**
- 17: **end if**
- 18: **until** convergence

圖 2.3.2-2 DDPG 演算法 (<https://spinningup.openai.com/en/latest/algorithms/ddpg.html>)

而以下說明部分訓練技巧，而這些技巧並非僅限於 DDPG 訓練演算法中適用，更像是一些通用的實用方法[10]。

(一)Target Network (目標網路)在更新 Critic 的過程中，Bellman equation 中需要納入 s' 的

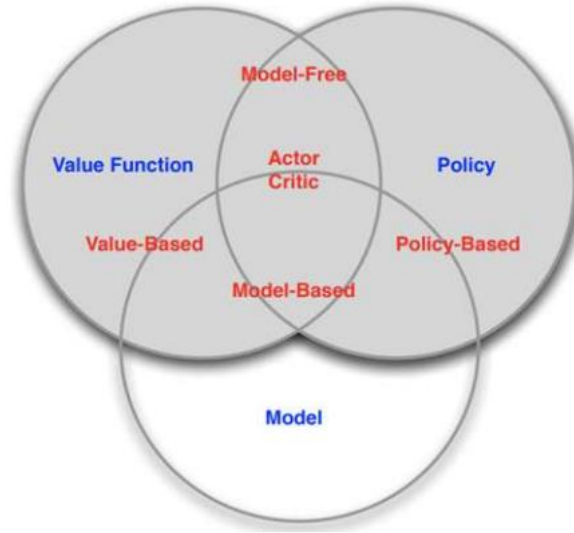
Q 估值，然而若 Critic 處於一個高頻率更新參數的狀態下，也會使 s' 的估值非常不穩定，從而導致 Critic 的訓練過程中無法收斂。Target Network 的概念即是產生另一個與 Critic 結構完全相同的類神經網路，稱為 Target Critic。在參數更新的時候僅針對 Critic 做梯度更新，而 Target Critic 則完全不改動，使 s' 的 Q 估值保持穩定。而每隔一段時間再將 Target Critic 更新成最新狀態的 Critic，重複此步驟。除了一段時間更新 Target Critic，也有另一種常見的作法稱為 soft update，即每次更新時 Critic 與 Target Critic 都接受梯度更新，但 Target Critic 更新的幅度僅為 Critic 的千分之 1 (該數值可由訓練者自訂)，限制 Target Critic 每次變動的幅度，從而達到穩定 s' 的 Q 估值的目標。

(二)Experience Replay (經驗回放)經驗回放技巧是 off policy 中常用的一種提升樣本效率的做法，其方法即為記錄下所有代理人與環境互動過程中的 (s,a,r,s') 轉移樣本，集結成一個巨大的經驗資料庫。在訓練過程中則從中隨機抽取一定數量的樣本訓練 Critic 或 Actor。經驗回放法能夠有效地重複多次使用同一筆樣本，因此能有更有效的應用蒐集來的資料。

(三)Prioritization Experience Replay (優先經驗回放) 優先經驗回放是經驗回放的改良作法，原始方法中是從資料庫中以隨機的方式抽取一定數量的樣本來做梯度更新。然而當訓練進入後期，資料庫中大多數的樣本都已經非常好的配適當前的估值函數，僅少部分的樣本仍未能有效的配適。若以隨機的方式抽取則較難抽到這些配適度較低的樣本，因此優先經驗回放方法即是去計算所有樣本的配適程度(loss 值)，並且賦予配適程度較差的樣本有較高的機率被抽中的一種機制，使訓練過程可以更快速的配適所有資料庫中的樣本。

五、PPO (Proximal Policy Optimization, 近端策略優化)

前期(111 年)研究中所採用的第二種 RL 演算法為 OpenAI 於 2017 所提出的近端策略優化(Proximal Policy Optimization, 以下簡稱 PPO)[38]。PPO 同樣是 Actor-Critic 架構的演算法，其結合價值函數(Value function)與策略函數(Policy)的技巧(如圖 2.3.2-3)，二函數均以類神經網路表示。訓練過程是藉由與環境互動過程產生的軌跡，進行二函數的參數更新。價值函數網路的參數是朝向和估計值的均方差變小的方向更新，策略函數的參數則是朝向系統期望獎勵最大化的方向作更新。



資料來源：Silver, D., Lectures on Reinforcement Learning <https://www.davidsilver.uk/teaching/>, 2015.

圖 2.3.2-3 RL 代理人類型

更新 Actor 的策略梯度(policy gradient)通常使用隨機梯度上升法(stochastic gradient ascent)，常見的梯度估計值為：

$$\hat{g} = \hat{E}_t[\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \hat{A}_t] \quad \text{式(10)}$$

其中 $\pi_{\theta}(a_t | s_t)$ 為隨機策略(stochastic policy)，亦即在 s_t 的狀態下產出 決策 a_t 的機率， θ 是此策略函數的參數。 $\hat{A}_t$ 則是步階 t 的優勢函數(advantage function)估計值。實際應用時，透過建構目標函數，軟體可自動計算出式(10)的梯度方向 $\hat{g}$ 。目標函數形式如下：

$$L^{PG}(\theta) = \hat{E}_t[\log \pi_{\theta}(a_t | s_t) \hat{A}_t] \quad \text{式(11)}$$

實務上用同個軌跡(trajjectory)對此目標式進行數次最佳化時，常導致破壞性的大幅度策略更新(destructively large policy updates)。為解決此問題，TRPO[37]提出在目標式下加入 KL-divergence 的限制式，以限縮策略更新的幅度，其數學規劃式如下。

$$\max_{\theta} \hat{E}_t \left[\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_k}(a_t | s_t)} \hat{A}_t \right] \quad \text{式(12)}$$

$$s.t. \hat{E}_t \left[KL[\pi_{\theta_k}(\cdot | s_t), \pi_{\theta}(\cdot | s_t)] \right] \leq \delta \quad \text{式(13)}$$

此處 θ_k 是更新前的策略參數。此數學規劃式可透過對目標式的線性估計及限制式的二次估計，以共軛梯度法(conjugate gradient algorithm)求取近似解。而在同樣的背景動機下，PPO 則是應用其他技巧使新舊策略不至於差異過大，且相較於 TRPO，PPO 顯著地較易應用，其績效表現至少和 TRPO 一樣好。

PPO 使用的目標式 L^{Clip} 如下：

$$\frac{1}{|D_k|T} \sum_{\tau \in D_k} \sum_{t=0}^T \min(r_t(\theta) A^{\pi_{\theta_k}(s_t, a_t)}, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) A^{\pi_{\theta_k}(s_t, a_t)}) \quad (14)$$

$$\text{其中, } r_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} \quad (15)$$

$r_t(\theta)$ 是新舊策略在給定 s_t 時產生 a_t 的機率的比率。從目標式 L^{Clip} 可看出，首項 $r_t(\theta) A^{\pi_{\theta_k}(s_t, a_t)}$ 與(12)相同，第二項則是將 $r_t(\theta)$ 限制在 $[1 - \epsilon, 1 + \epsilon]$ 的區間，最後取此二項的最小值。圖 2.3.2-4 分別對優勢函數大於零及小於零的情況，繪製 L^{Clip} 和 r 值的關係圖，紅點是 r 值的初始位置(初始時新舊策略相同，所以 r 值為 1)。PPO 目標式 L^{Clip} 的設計可讓 $r_t(\theta)$ 值在超出 $[1 - \epsilon, 1 + \epsilon]$ 時，目標式對 $r(\theta)$ 的梯度為 0，藉此產生使新舊策略差異不至於過大的效果。

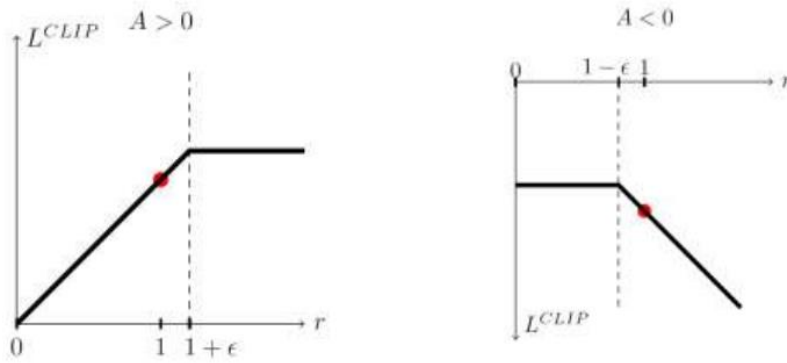


圖 2.3.2-4 PPO 的目標式 L^{Clip} 與 $r(\theta)$ 的關係圖

圖 2.3.2-4 之左圖為當 $A > 0$ 時， $r(\theta)$ 往右側移動至 $1 + \epsilon$ 後便無法再增加 L^{Clip} ，避免 $r(\theta)$ 過大。圖 2.3.2-4 之右圖為當 $A < 0$ 時， $r(\theta)$ 往左側移動至 $1 - \epsilon$ 後便無法再減少 L^{Clip} ，避免 $r(\theta)$ 過小。

PPO 也是使用 Actor-Critic 架構的方法，運作架構如圖 2.3.2-5；演算法如圖 2.3.2-6。先以隨機策略 π_{θ_k} 跟環境互動取得軌跡 τ 後，計算軌跡中各時階 t 的未來獎勵總值 R_t 以及優勢函數 A_t 。接著便可針對式(14)使用梯度上升法更新 θ ，再以梯度下降法更新 Critic 的參數 ϕ 。

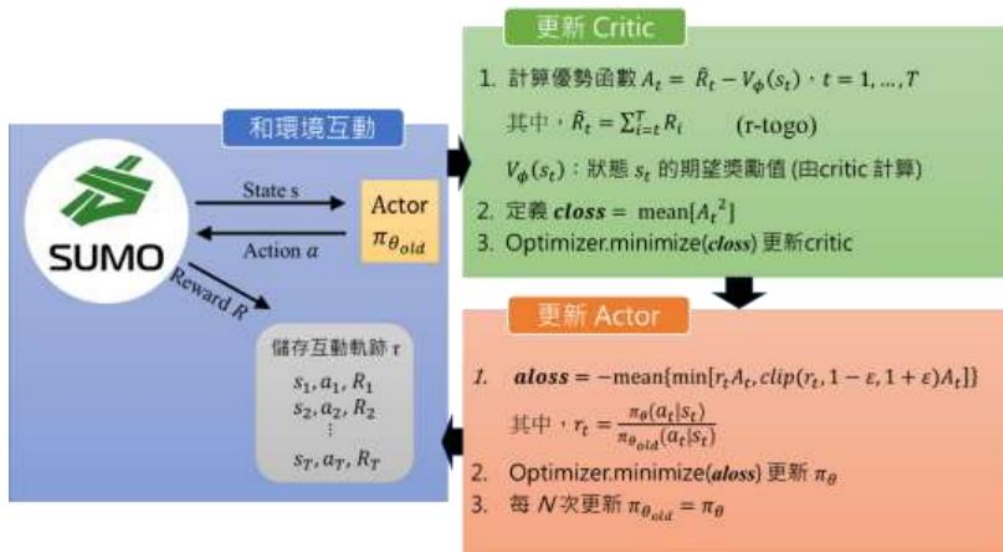


圖 2.3.2-5 PPO 運作架構圖

Algorithm 1 PPO-Clip

- 1: Input: initial policy parameters θ_0 , initial value function parameters ϕ_0
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment.
- 4: Compute rewards-to-go $\hat{R}_t$.
- 5: Compute advantage estimates, $\hat{A}_t$ (using any method of advantage estimation) based on the current value function V_{ϕ_k} .
- 6: Update the policy by maximizing the PPO-Clip objective:

$$\theta_{k+1} = \arg \max_{\theta} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \min \left(\frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A^{\pi_{\theta_k}}(s_t, a_t), g(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)) \right),$$

typically via stochastic gradient ascent with Adam.

- 7: Fit value function by regression on mean-squared error:

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T (V_{\phi}(s_t) - \hat{R}_t)^2,$$

typically via some gradient descent algorithm.

- 8: **end for**

圖 2.3.2-6 PPO 演算法 (<https://spinningup.openai.com/en/latest/algorithms/ppo.html>)

DDPG 與 PPO 算法分別由著名的 Google DeepMind 與 OpenAI 提出，DDPG 能輸出確定性決策，並改善 DQN 難以應用在連續型決策空間的缺點，且其沿用 DQN 的經驗回放等技巧，大幅提升樣本使用率。而 PPO 的 Actor 則是輸出隨機策略，透過目標式加入上下界的技巧避免策略更新變化太大，並藉 importance sampling 提升樣本使用率。OpenAI 將 PPO 當作 baseline 算法，並且著名的 OpenAI five[46]亦使用 PPO 作為核心算法。因此本研究除應用類似 DDPG 的訓練演算法外，亦嘗試使用 PPO 訓練 AI 號誌控制代理人[38], [37], [46]。

六、SAC(Soft-Actor Critic, 軟性動作-評論演算法)

軟性動作-評論演算法(Soft-Actor Critic, 以下簡稱 SAC)[18]，SAC 也是屬於 Actor-Critic 類的一種方法分支架構的演算法，而此算法的重點在於其納入了熵度量(Entropy)進回報函數(reward)以提高探索率並希望能學習到一種盡可能隨機行動的策略，也同時能夠完成任務目標。相比於 PPO 與 DDPG，其目標的學習目標很直接，就是利用策略(policy)並以式(16)使回報期望值得到最大。而最大熵則是利用式(16)並以其為基礎進行更新使式(17)要求策略的動作(action)的熵與回報期望值為最大。

$$\sum_t \mathbb{E}_{(\mathbf{s}_t, \mathbf{a}_t) \sim \rho_{\pi}} [r(\mathbf{s}_t, \mathbf{a}_t)] \quad \text{式(16)}$$

$$\pi^* = \arg \max_{\pi} \sum_t \mathbb{E}_{(\mathbf{s}_t, \mathbf{a}_t) \sim \rho_{\pi}} [r(\mathbf{s}_t, \mathbf{a}_t) + \alpha \mathcal{H}(\pi(\cdot | \mathbf{s}_t))], \quad \text{式(17)}$$

至此，便能以此掌控每次動作與策略更新時的隨機性。由於此原因，策略的決策與執行方程式與演算法也會造成改變，詳細會在以下進行說明。前期策略 Action 與執行都是以 Bellman equation 為基礎，而加入熵並承上 α 係數會計算新的 Q 值函數(式[18])。

$$Q_{soft}(s_t, a_t) = r(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1}, a_{t+1}} [Q_{soft}(s_{t+1}, a_{t+1}) - \alpha \log(\pi(a_{t+1} | s_{t+1}))] \quad \text{式(18)}$$

但要記住此算式是以舊回報值為基礎而計算的。而如果把熵回報加入係數裡，就能得出此策略的 Q 值函數更新法(式[19、20])。

$$\mathcal{T}^\pi Q(s_t, \mathbf{a}_t) \triangleq r(s_t, \mathbf{a}_t) + \gamma \mathbb{E}_{s_{t+1} \sim p} [V(s_{t+1})], \quad \text{式(19)}$$

$$V(s_t) = \mathbb{E}_{\mathbf{a}_t \sim \pi} [Q(s_t, \mathbf{a}_t) - \alpha \log \pi(\mathbf{a}_t | s_t)] \quad \text{式(20)}$$

推導過程可參考[17]。

而能使其策略能收斂，Haarnoja 等人利用了目標數值的最小化(KL Divergence)來趨近最佳的 Q 值函數 $\exp(Q_{soft}^\pi(s_t, \cdot))$ 並取得式(21)的策略更新方程式。

$$\pi_{\text{new}} = \arg \min_{\pi' \in \Pi} D_{\text{KL}} \left(\pi'(\cdot | s_t) \left\| \frac{\exp(\frac{1}{\alpha} Q^{\pi_{\text{old}}}(s_t, \cdot))}{Z^{\pi_{\text{old}}}(s_t)} \right\| \right) \quad \text{式(21)}$$

Algorithm 1 Soft Actor-Critic

```

1: Input: initial policy parameters  $\theta$ , Q-function parameters  $\phi_1, \phi_2$ , empty replay buffer  $\mathcal{D}$ 
2: Set target parameters equal to main parameters  $\phi_{\text{targ},1} \leftarrow \phi_1, \phi_{\text{targ},2} \leftarrow \phi_2$ 
3: repeat
4:   Observe state  $s$  and select action  $a \sim \pi_\theta(\cdot | s)$ 
5:   Execute  $a$  in the environment
6:   Observe next state  $s'$ , reward  $r$ , and done signal  $d$  to indicate whether  $s'$  is terminal
7:   Store  $(s, a, r, s', d)$  in replay buffer  $\mathcal{D}$ 
8:   If  $s'$  is terminal, reset environment state.
9:   if it's time to update then
10:    for  $j$  in range(however many updates) do
11:      Randomly sample a batch of transitions,  $B = \{(s, a, r, s', d)\}$  from  $\mathcal{D}$ 
12:      Compute targets for the Q functions:

$$y(r, s', d) = r + \gamma(1 - d) \left( \min_{i=1,2} Q_{\phi_{\text{targ},i}}(s', \tilde{a}') - \alpha \log \pi_\theta(\tilde{a}' | s') \right), \quad \tilde{a}' \sim \pi_\theta(\cdot | s')$$

13:      Update Q-functions by one step of gradient descent using

$$\nabla_{\phi_i} \frac{1}{|B|} \sum_{(s,a,r,s',d) \in B} (Q_{\phi_i}(s, a) - y(r, s', d))^2 \quad \text{for } i = 1, 2$$

14:      Update policy by one step of gradient ascent using

$$\nabla_\theta \frac{1}{|B|} \sum_{s \in B} \left( \min_{i=1,2} Q_{\phi_i}(s, \tilde{a}_\theta(s)) - \alpha \log \pi_\theta(\tilde{a}_\theta(s) | s) \right),$$

      where  $\tilde{a}_\theta(s)$  is a sample from  $\pi_\theta(\cdot | s)$  which is differentiable wrt  $\theta$  via the reparametrization trick.
15:      Update target networks with

$$\phi_{\text{targ},i} \leftarrow \rho \phi_{\text{targ},i} + (1 - \rho) \phi_i \quad \text{for } i = 1, 2$$

16:    end for
17:  end if
18: until convergence

```

圖 2.3.2-7 SAC 演算法 (<https://spinningup.openai.com/en/latest/algorithms/sac.html>)

在上述作法下，完整 SAC 演算法可以圖 2.3.2-7 表示。要注意的是，關於 Q 值的神經網絡(目標網絡)。雖然與 DDPG 相同，都是帶 target soft Q 網絡與參數 θ ，使 s' 的 Q 估值保持穩

定，但 DDPG 的 Target Critic 會完全不改動。而 SAC 則會在每一次利用式(21)自動更新參數。而且為了更快速穩定的訓練，Haarnoja 等人引入了兩個 Q 網路，然後每次選擇 Q 值小的一個作為 target Q 值並利用其上述的算式更新 Q 值與其策略。雖然上述所有演算法如 PPO 與 DDPG 都有在前期產出不錯的訓練成果，但其都是以單目標為基礎，而在交通的領域裡，單目標會使 AI 的訓練時間拉長而且多路口與時相的持續控制問題也會陷入問題，所以本案會引用多目標強化學習模式與其 Q-Pensieve 的演算法。詳細內容會於後續章節說明。

七、多目標強化學習(MORL)模式

(一)多目標馬可夫決策過程

強化學習亦可用於學習具有多個最佳化目標之問題場景，我們可將標準的 MDP 推廣為多目標之 MDP (multi-objective MDP，或稱 MOMDP)。與標準的 MDP 類似，MOMDP 依然具有狀態集合、動作集合、獎勵函數、狀態轉移函數、折扣因子等等要素。另一方面，MOMDP 與一般 MDP 主要的差異在於獎勵函數，MOMDP 的獎勵函數可輸出一個多維度的向量 $\mathbf{R}(s, a) = (R^{(1)}(s, a), \dots, R^{(d)}(s, a))$ ，以粗體表示。以交通場景為例，獎勵訊號可以包含多種效標如車輛排隊長度 (queue length)、車輛延遲 (delay)、路口等待時間 (waiting time)、號誌切換延遲 (switching delay) 等等。在多目標問題中，我們依然可定義價值函數如下，而唯一的差別僅在於價值函數皆為向量函數(Yang, Sun, & Narasimhan, 2019) [49]。

$$\begin{aligned} V^\pi(s) &:= E_\pi[\sum_{k=0}^{\infty} \gamma^k \mathbf{R}_{k+1} | S_0 = s] \\ Q^\pi(s, a) &:= E_\pi[\sum_{k=0}^{\infty} \gamma^k \mathbf{R}_{k+1} | S_0 = s, A_0 = a] \end{aligned}$$

相對應的總獎勵也可表示為 $\mathbf{V}^\pi(\rho) := E_{s \sim \rho}[\mathbf{V}^\pi(s)]$ ，其亦為一個 d 維度之向量。在多目標問題，我們可對於各種策略 π 所對應之總獎勵 $\mathbf{V}^\pi(\rho)$ 進行探討：以兩個目標為例 ($d = 2$)， $\mathbf{V}^\pi(\rho)$ 可表示在二維平面上 (如下圖 2.3.2-8 之各個圓點)，而從這些總獎勵向量我們可找出相對好之向量集合稱為 Pareto frontier，定義如下：首先我們用 $\pi \geq \pi'$ 來表示 $\mathbf{V}^\pi(\rho) \geq \mathbf{V}^{\pi'}(\rho)$ ，亦即 π 在各個目標都能勝過 π' ，則我們可定義 Pareto frontier 為

$$PF := \{\mathbf{V}^\pi(\rho) | \nexists \pi' s.t. \pi' \geq \pi\}$$

以下圖 2.3.2-8 為例，點 A, B, C, D, E, F, G, H, K 皆屬於 Pareto frontier。廣義來說，多目標問題的主要目標為找到 Pareto frontier 中所有點其對應之策略，這也是與單目標問題 (僅需找到一個最佳策略) 之本質差異。

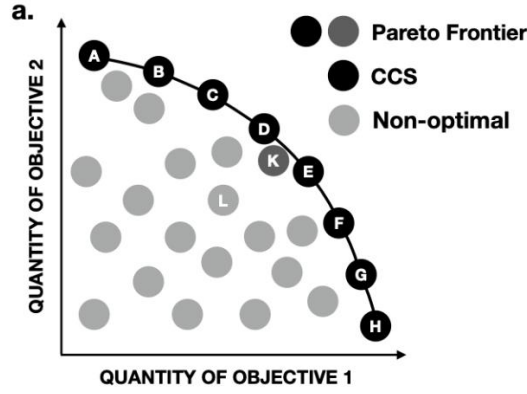


圖 2.3.2-8 兩目標馬可夫決策過程之 Pareto front 與 CCS 示意圖

(二)目標偏好與線性純量方法

在多目標問題之中，如上述定義之 Pareto front，我們可看出最佳策略會隨著對於各目標之偏好 (preference) 而改變。目前最常使用的描述偏好方式為線性純量 (linear scalarization) 方法，亦即透過一組權重 $\lambda = (\lambda_1, \dots, \lambda_d)$ 來描述偏好並可將獎勵向量轉成純量 $\lambda^\top \mathbf{R}(s, a) = \sum_{i=1}^d \lambda_i R^{(i)}(s, a)$ ，在給定一種偏好下多目標問題即可簡化為一個單目標問題。在不失一般性下，我們可專注在 $\lambda = (\lambda_1, \dots, \lambda_d)$ 滿足為機率分佈的情況下。基於此概念，找尋 Pareto frontier 可轉換為解決一群單目標問題：

對於各種非負偏好 $\lambda = (\lambda_1, \dots, \lambda_d)$ 滿足 $\sum_{i=1}^d \lambda_i = 1$ ，學習者的目標是輸出對應於各種偏好之最佳策略，即透過與環境互動同時解決一整群單目標問題。

(三)常用多目標強化學習模式介紹

基於以上的線性純量方法，現有的多目標強化學習演算法大致可分為兩類：

1.顯性搜尋(explicit search)

許多以多目標強化學習 (MORL) 為主題的先前研究，都是透過明確地搜尋獎勵空間的 Pareto frontier 來更新一個或多個策略。以近期被提出的 PGMORL 演算法 (Xu, et al., 2020) [48] 為例，此篇提出了一種演化方法來搜索 Pareto frontier，並借助預測模型來確定搜索方向，主要概念可簡要呈現如下圖 2.3.2-9，此方法主要有三個步驟：首先在 warm-up 階段此方法會初始化多個 worker，在每次迭代時各個 worker 會從目前相對較好的策略出發並使用策略梯度的來估計來搜索更接近 Pareto frontier 策略，最後會將迭代過程中的策略做適當的分群並拼湊出 Pareto frontier。

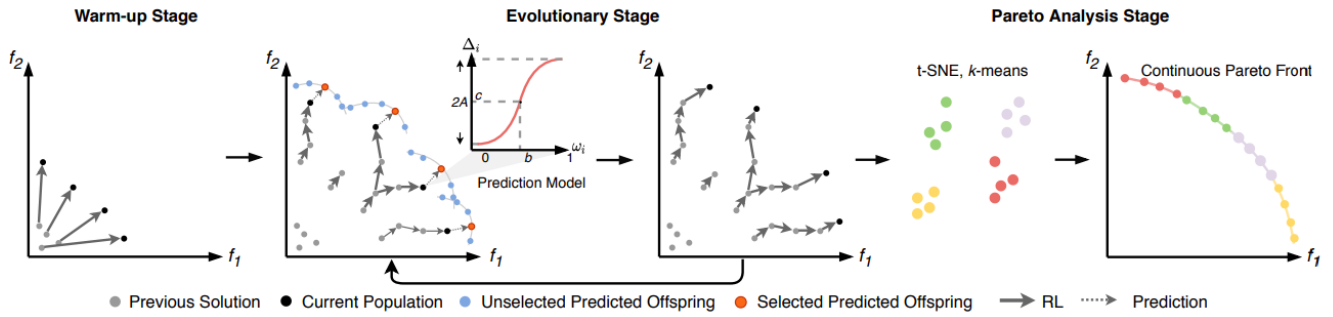


圖 2.3.2-9 PGMORL 之搜索概念示意圖

目前還有其他既有的類似顯性搜索策略，比如 (Van Moffaert & Nowe, 2014) [43] 提出了 Pareto 的 Q-learning 方法，是將即時獎勵和未來收益向量分別存儲並利用 Pareto 優勢來選擇 Q 學習中的動作。另外，(Mossalam, Assael, Roijers, & Whiteson, 2016) [33] 則是通過基於角落權重的優先順序選擇方法，使用 scalarized Q-learning 解決 MORL 問題。此外，近期的研究 (Kyriakis, Deshmukh, & Bogdan, 2022) [25] 則是通過逼近 Pareto frontier 的一階必要條件來提出一種策略梯度方法。

儘管有前述許多顯性搜索方法，但上述明確搜索算法的一大缺點是因每次搜索過程之間的知識共享非常少，而使得樣本效率較低。

2. 隱性搜尋 (implicit search)

另一類演算法旨在通過隱式搜索改進多個偏好的策略。以近期的研究 (Abels, Roijers, Lenaerts, Nowe, & Steckelmacher, 2019) [2] 為例，此篇提出了一種名為「條件網路」(Conditioned Network, 或簡稱 CN) 的方法，將標準的單目標 DQN 擴展到學習偏好相關的多目標 Q 函數，具體方法為將偏好向量作為 DQN 的輸入，而期望透過訓練使 DQN 能對於各種偏好輸出相對應之最佳 Q 函數。條件網路之架構圖如下圖 2.3.2-10。

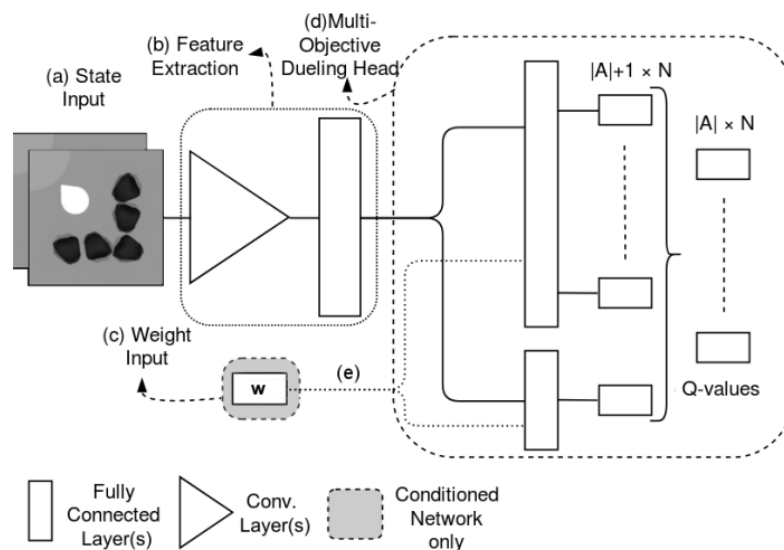


圖 2.3.2-10 Conditioned Network 架構示意圖

另一近期的研究為 (Yang, Sun, & Narasimhan, 2019) [49]，此篇提出了包絡 Q 函數 (envelope Q function) 的概念，以促進不同偏好下的 Q 函數之間的知識共享，使任何策略都可以從其他偏好的經驗中受益，從而使訓練更有效率。基於此概念，(Zhou, Chen, & Gu, 2020) 也基於包絡 Q 函數提出了基於動態模型的包絡值迭代，並給予多目標 Q 學習之收斂保證[49], [48], [43] [33], [25], [2]。

八、Q-Pensieve

Q-Pensieve[20]也是屬於 Actor-Critic 類的一種方法分支架構的演算法，可以視為 SAC 算法的衍生。與 SAC 相同都納入熵量(Entropy)為 Q 值函數裡的一種參數。但有所不同的是 Q-Pensieve 納入了多目標強化學習，使其不用多次尋找 Pareto 前線也能在共同確定政策更新方向時進行數據分享也能提高樣本效率。Q-Pensieve 的環境狀態與動作集合 $Q(S,A)$ 是利用多目標馬可夫決策過程為多維度的向量(vector: $S, A, P, r, \gamma, D, S\lambda, \Lambda, d$)(詳情於 2.1.6 小節)。此演算法重要的變數為 $d, S\lambda$, 和 Λ 。 d 代表的是多目標的數量， Λ 代表的是參數(preference)向量的系列組合，而 $S\lambda$ 代表的是 preference 參數的存量化算式。

總而言之，在單目標訓練裡的 SAC 裡，在訓練的每一步，AI 會收到 State, Action, 與獎勵的向量 r ，而策略更新會用 SAC 小節裡的算式[18:21]來取得新的 Q 值函數。而 Q-Pensieve 延展了其算式，使新的算式[22,23]融入了上述的重要變數。

$$Q^\pi(s, a) = r(s, a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s, a)} [V^\pi(s')], \quad \text{式(22)}$$

$$V^\pi(s) = \mathbb{E}_{a \sim \pi(\cdot|s)} [Q^\pi(s, a) - \alpha \log \pi(a|s) \mathbf{1}_d], \quad \text{式(23)}$$

$\mathbf{1}_d$ 代表所以目標的向量。而 $\pi(\cdot|\cdot; \lambda)$ 是用於調整並取得參數的偏好及對於 Q 值函數的好壞。而關於其演算法的更新, Hung 等人也是引用了 SAC 的策略更新方程式(SAC 小節之算式(21))。但記得此方程式為應用於單目標的，而在多目標的情況下，因資料無法共享導致其抽取的樣本效率不佳。為了處理這個問題，Hung 等人開發了新的回放資料庫與其方程式(24)，而此資料庫是為了存取此目標的 Q 值以及參數，之所以選擇這方式是因其 Q 值與參數已經存在於 action 裡了，並不需要在進行更新，只需存取於資料庫(與 replay buffer 相似)，並不會利用到多餘的資源。

$$\pi_{k+1}(\cdot|\cdot; \lambda) = \arg \min_{\pi' \in \tilde{\Pi}} D_{KL} \left(\pi'(\cdot|s; \lambda) \parallel \frac{\exp \left(\sup_{\lambda' \in W_k(\lambda), Q' \in Q_k} \frac{1}{\alpha} \lambda^\top Q'(s, \cdot; \lambda') \right)}{Z_{Q_k}(s)} \right), \quad \text{式(24)}$$

Z_{Q_k} 為 Q_k 的正量化值， Q_k 則為每一個 k 值的 Q 系列組合。

而 W_k 則為每一個 k 值的參數系列組合。

而以下將說明此訓練的三種重要更新點：

(一)Policy-Evaluation

在策略選定的過程中，Bellman Equation 會以參數(preference)為基準，並用多目

標歸一化指數函數(算式(25))而計算新的 Q 值函數。

$$(\mathcal{T}_{\text{MO}}^{\pi} \mathbf{Q})(s, a; \boldsymbol{\lambda}) = \mathbf{r}(s, a) + \gamma \mathbb{E}_{s' \sim \mathcal{P}(\cdot | s, a), a' \sim \pi(\cdot | s'; \boldsymbol{\lambda})} [\mathbf{Q}(s', a'; \boldsymbol{\lambda}) - \alpha \log \pi(a' | s'; \boldsymbol{\lambda}) \mathbf{1}_d]. \quad \text{式(25)}$$

(二)Q Replay Buffer

由此可見，Q-Pensieve 會用現在的 Q 與前面存取的 Q 進行策略更新，Q replay buffer 的用途就是用於儲存前面所有動作造成的 Q 值，並隨機抽取 Q 進行對比和 Policy Iteration。

(三)Soft-Policy Iteration

Policy Iteration 會利用 SAC 並以脫離策略學習多目標的梯度方向與計算出其目標函數(算式(5))。

$$\mathcal{L}_{\pi}(\theta; \boldsymbol{\lambda}) = \mathbb{E}_{s \sim \mu} \left[\mathbb{E}_{a \sim \pi_{\theta}} \left[\sup_{\boldsymbol{\lambda}' \in W(\boldsymbol{\lambda}), \mathbf{Q}' \in \mathcal{Q}} \{ \alpha \log (\pi_{\theta}(a | s; \boldsymbol{\lambda})) - \boldsymbol{\lambda}'^{\top} \mathbf{Q}'(s, a; \boldsymbol{\lambda}') \} \right] \right]. \quad \text{式(26)}$$

其運作架構如圖 2.3.2-11，演算法如圖 2.3.2-12。

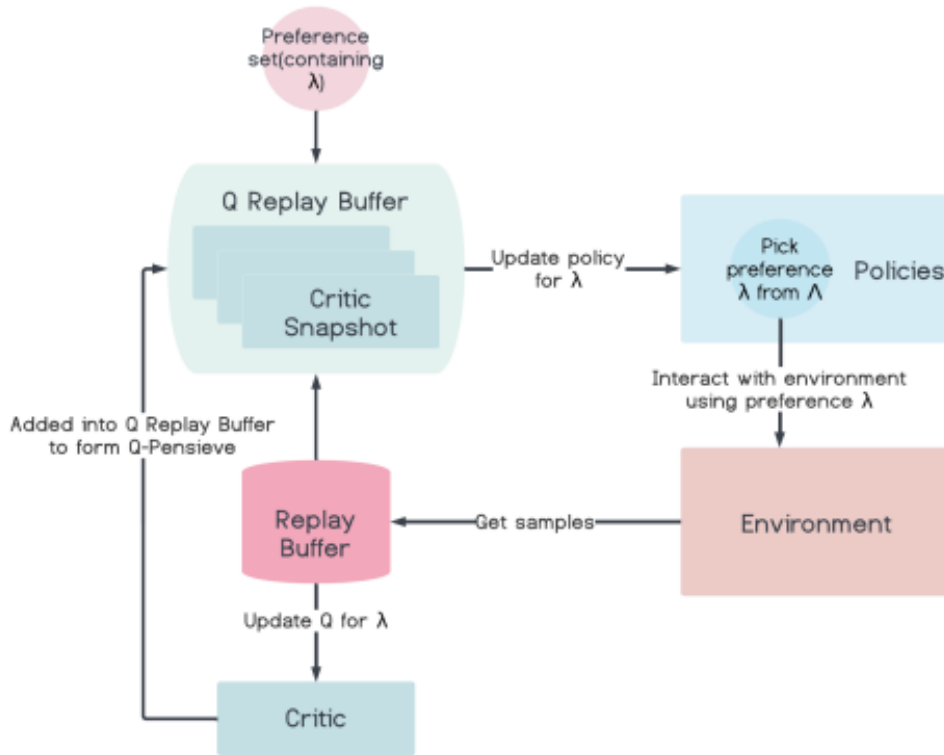


圖 2.3.2-11 Q-Pensieve 運作架構圖

Algorithm 1: Q-Pensieve

Input : ϕ_1, ϕ_2, θ , preference sampling distribution $\mathcal{P}_\lambda$, number of preference vectors N_λ , the soft update coefficient τ , actor learning rates η_π , critic learning rates η_Q

Output : ϕ_1, ϕ_2, θ

```

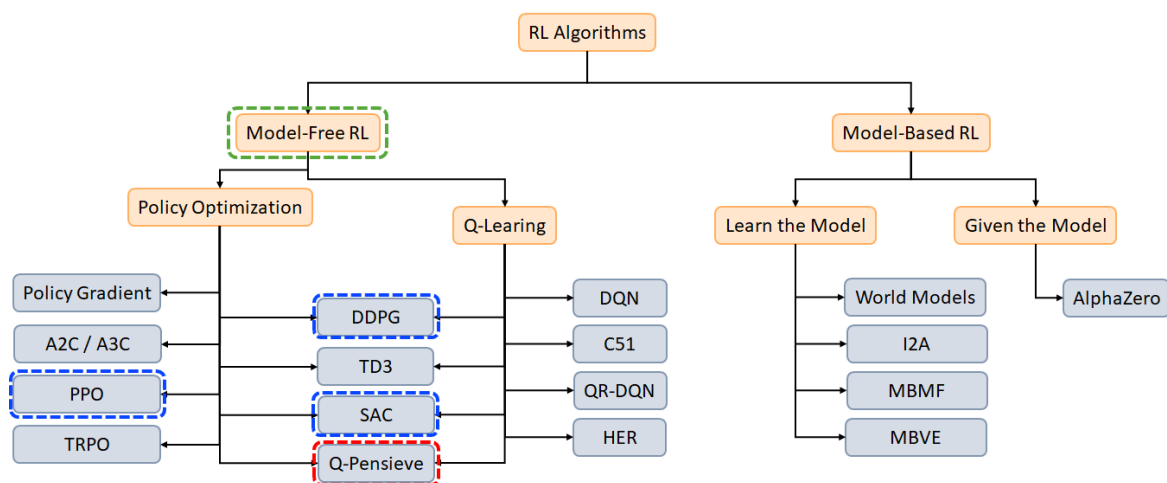
1  $\bar{\phi}_1 \leftarrow \phi_1, \bar{\phi}_2 \leftarrow \phi_2$ ;
2  $\mathcal{M} \leftarrow \emptyset$  ▷ Initialize replay buffer;
3  $\mathcal{B} \leftarrow \emptyset$ ;
4 for each iteration do
5   sample weight  $\lambda$  from  $\Lambda$  according  $\mathcal{P}_\lambda$ ;
6   for each environment step do
7      $a_t \sim \pi_\theta(\cdot | s_t; \lambda)$ ;
8      $s_{t+1} \sim P(\cdot | s_t, a_t)$ ;
9      $\mathcal{M} \leftarrow \mathcal{M} \cup \{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$ ;
10  for each gradient step do
11    sample  $N_\lambda - 1$  preferences and add them to set  $W$ ;
12     $W \leftarrow W \cup \{\lambda\}$ ;
13     $\phi_i \leftarrow \phi_i - \eta_Q \hat{\nabla}_{\phi_i} \mathcal{L}_Q(\phi_i; \lambda), \mathcal{B} \leftarrow \mathcal{B} \cup \{\phi_i; \lambda\}$  for  $i \in \{1, 2\}$ ;
14    compute  $\hat{\nabla}_\theta$  with eq. 11 with  $W$ ;
15     $\theta \leftarrow \theta - \eta_\pi \hat{\nabla}_\theta \mathcal{L}_\pi(\theta; \lambda)$ ;
16     $\bar{\phi}_i \leftarrow \tau \phi_i + (1 - \tau) \bar{\phi}_i$  for  $i \in \{1, 2\}$ ;

```

資料來源：[20]

圖 2.3.2-12 Q-Pensieve 演算法

九、強化學習演算法比較



資料來源: Part 2: Kinds of RL Algorithms — Spinning Up documentation. (n.d.).

https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html [21]

圖 2.3.2-13 強化學習演算法之歸納與衍生

表 2.3.2-1 強化學習演算法之比較

	SAC	PPO	DDPG	Q-Pensieve
策略(Policy)	隨機策略 (Stochastic)	隨機策略 (Stochastic)	確定性策略 (Deterministic)	隨機策略 (Stochastic)
樣本效率(Sample Efficiency)	高	低	高	高
離線策略學習 (Off-Policy Learning)	有	否	有	有
多目標(Multi-Objective)	否	否	否	有

(一)DDPG

DDPG 啟用脫離策略學習(off-policy learning)與確定性策略(Deterministic Policy)。優點為因確定性策略(Deterministic Policy)減少了訓練的複雜性而能更快速與有效的優化連續動作的問題應用。而 DDPG 使用了目標網路，經驗回放技巧而穩定了脫離策略學習(off-policy learning)探索的多維狀態空間與其策略的更新過程，並大幅提升樣本使用率。缺點為 DDPG 會因參數的敏感問題造成每次 critic 訓練的精準度有很大的差異，例如：學習率(learning rate)與車輛或向量的多寡，與確定性策略(Deterministic Policy)無法對於潛在較佳的 action 進行有效率的探索。

(二)PPO

PPO 啟用同策略學習(on-policy learning)與隨機策略(Stochastic Policy)。優點為因隨機策略(Stochastic Policy)解決了大規模和高維狀態空間的問題而能有效優化策略函數，並因 PPO 只使用策略函數來進行優化使其減少算法的複雜性。而 on-policy 通過限制策略更新的幅度確保了訓練過程的穩定性。而 PPO 使用的 Clipping reward 充分利用了過去的經驗數據(importance sampling)而提高樣本使用率。主要缺點為其訓練結果會受限於近期的經驗，容易陷入一個局部區域最佳性能的策略，另外一個缺點為 Policy 的模型只能表單一波峰的動作概率分佈，無法對於全局的其他次佳策略進行有效率的探索。

(三)SAC

SAC 啟用策略學習(off-policy learning)與隨機策略(Stochastic Policy)。優點為因同時使用策略學習(off-policy learning)跟隨機策略(Stochastic Policy)使 SAC 繼承了 DDPG 與 PPO 的優點。SAC 也利用了熵正則化的概念(Entropy-regularized RL)來增加策略的探索性與隨機性，使其能有效地對於全域空間進行探索。並通過使用兩個 Q 值函數 $Q(s,a)$ 來減少錯誤估計。最終使用策略函數來取得最大化的獎勵(reward)並找到對應的動作概率分佈，使模型可以去找到其他全域最佳解，而不僅僅陷入局部最優解。這件事是 DDPG 與 PPO 所無法達成的。而 SAC 的缺點也因其增加了熵係數導致參數計算資源和訓練時間花費極大。而更新參數方面需借重專家經驗或需透過相當多次反覆的超參數搜尋才能達成。這導致需利用多個神經網絡來進行優化使其訓練過程也相對較為複雜，最後也因無法實行知識共享(透過共享不同訓練週期的 Q 值函數參數與不同獎勵偏好(Preference)對應的 Q 值函數)導致 SAC 難以在多種目標中確保資料的是使用效率以達成最適權衡。

(四)Q-Pensieve

Q-Pensieve 啟用與 SAC 一樣模型。因其原因 Q-Pensieve 繼承了 SAC 的優點不只增加了策略的探索性與隨機性而也增加多目標的學習方法使 Agent 能進行知識共享

盡管目標不同。使用方法為利用 off-policy 學習並取得另外一個目標的偏好(preferences)與環境(states)和動作(actions)。並且同時利用不同目標的 Q 值函數 $Q(s,a)$ 進行 knowledge sharing 與更新原來的 Policy network。但缺點也因其龐大的 Q 值函數導致演算法本身與神經網路架構皆較為複雜。

十、Task-Aware (多任務強化學習模式)

(一)Multimodal Model-Agnostic Meta-Learning via Task-Aware Modulation (MMAML) 簡介

現有的大多數模型無關元學習器都是單模態(Unimodal)用於單一初始化(single initialization/reward)。從複雜的任務分布中隨機抽樣的不同任務可能需要截然不同的一組參數(parameter θ)，這使得很難找到單一獎勵能涵蓋所有的任務目標，使其接近所有目標參數 (如圖 2.3.2-14)。而如果任務分布是多模態的，其中模態之間是不相交且相距較遠的，可以想像使用一組獨立的元學習器，每一個元學習器涵蓋一個模態，可能更好地掌握整個分布(如圖 2.3.2-15) (Vuori, 2019b) [44]。然而，任何任務與元學習器中的一個關聯不僅需要額外的任務身份信息，而這些信息通常並不可用或在模式不明確時可能存在分歧，而且這樣做也會阻止在任務分布的不同模式之間進行知識轉移。簡而言之，如果任務分布是多模態的且模式之間相距較遠，單一初始化的模型無法很好地處理所有任務。嘗試使用多個獨立的元學習器來涵蓋每一個模式可能更為有效，但這樣做會引入額外的挑戰，並可能限制跨模式的知識轉移。

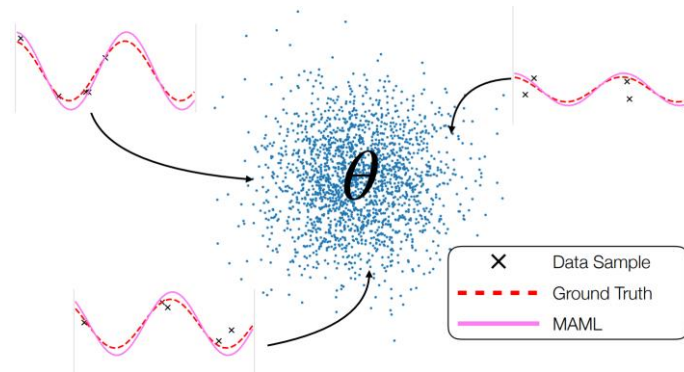


圖 2.3.2-14 單模態任務目標參數示意圖

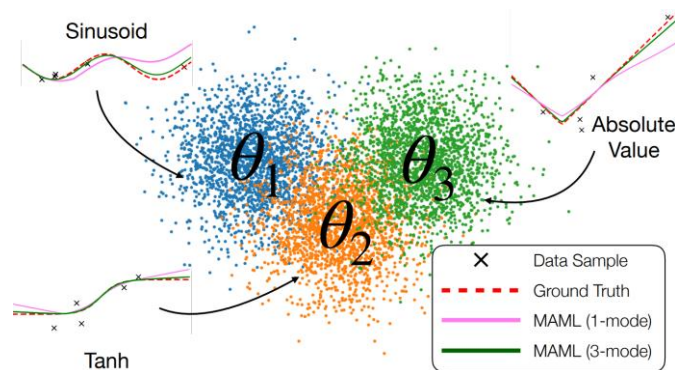


圖 2.3.2-15 多模態任務目標參數示意圖

為處理這個問題，MMAML 運用 Model-Based 與 Model-agnostic Meta-Learning (MAML) 的優點：演算法與方法(Method) (Chelsea Finn, Pieter Abbeel, Sergey Levine, 2017) [7]。演算法旨在使用過去的經驗中學習，提高在此任務上的學習能力訓練一個模型(unimodal)，使得它能夠在新的任務上以最少的訓練數據快速適應(與 SAC 跟 Q-Pensieve 相同，詳情如後 AI 強化學習模式比較分析所述)。而方法旨在利用以上多種不同的 unimodal 來處理多種數據模態的任務，使各模型能夠在不同類型數據的任務中進行學習，並進行整合與泛化，使其更加多樣化和適應各種現實場景。而在號誌控制等跨模態檢任務中(multi-task)，因號誌的週期與秒數不同，模型需在不同模態之間找到相關信息，使其原因使 MMAML 特別有用。

(二)MMAML 用法(Method)

具體來說，MMAML 希望能增強跨多模態任務分布泛化能力的方法。與學習參數空間中的單一初始化點不同，它會有兩個 Neural Network (Modular network 與 Task network)。Modular network 會先進行關於任務相關的數據檢查(datapoints)來偵測一個任務身份(task identity)並估計其參數(parameters)。在得到估計後，其估計會作為輸入(input)對目標任務(task network)進行初始化調節，並以後續不同的任務模式進行推斷與參數設定(parameter T) (Vuori, 2019b) [44]。並以這些調節的參數對目標任務進行梯度適應，逐步改善其性能(如圖 2.3.2-16)。

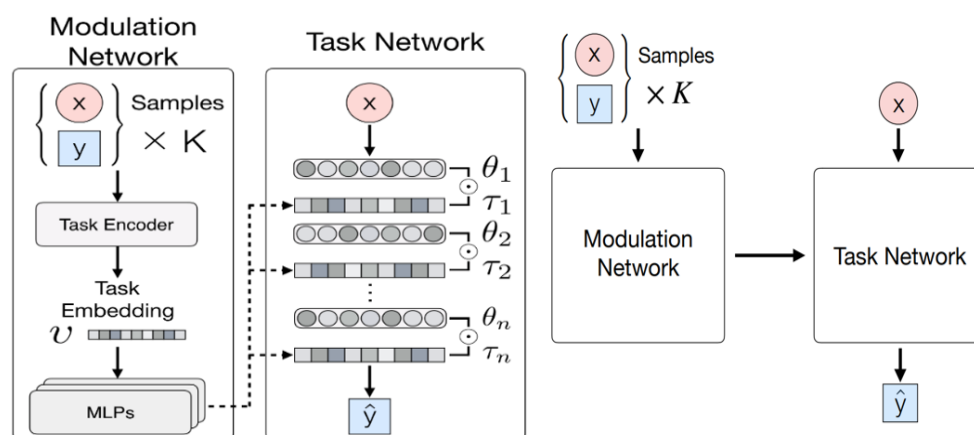


圖 2.3.2-16 MMAML 架構示意圖

Algorithm 1 MMAML META-TRAINING PROCEDURE.

```
1: Input: Task distribution  $P(\mathcal{T})$ , Hyper-parameters  $\alpha$  and  $\beta$ 
2: Randomly initialize  $\theta$  and  $\omega$ .
3: while not DONE do
4:   Sample batches of tasks  $\mathcal{T}_j \sim P(\mathcal{T})$ 
5:   for all  $j$  do
6:     Infer  $v = h(\{x, y\}_K; \omega_h)$  with K samples from  $\mathcal{D}_{\mathcal{T}_j}^{\text{train}}$ .
7:     Generate parameters  $\tau = \{g_i(v; \omega_g) \mid i = 1, \dots, N\}$ 
        to modulate each block of the task network  $f$ .
8:     Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta, \tau); \mathcal{D}_{\mathcal{T}_j}^{\text{train}})$  w.r.t the K samples
9:     Compute adapted parameter with gradient descent:
         $\theta'_{\mathcal{T}_j} = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta, \tau); \mathcal{D}_{\mathcal{T}_j}^{\text{train}})$ 
10:   end for
11:   Update  $\theta$  with  $\beta \nabla_{\theta} \sum_{\mathcal{T}_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta', \tau); \mathcal{D}_{\mathcal{T}_j}^{\text{val}})$ 
12:   Update  $\omega_g$  with  $\beta \nabla_{\omega_g} \sum_{\mathcal{T}_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta', \tau); \mathcal{D}_{\mathcal{T}_j}^{\text{val}})$ 
13:   Update  $\omega_h$  with  $\beta \nabla_{\omega_h} \sum_{\mathcal{T}_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta', \tau); \mathcal{D}_{\mathcal{T}_j}^{\text{val}})$ 
14: end while
```

圖 2.3.2-17 MMAML 訓練虛擬碼

1. Modulation Network

要產出與任務有關的參數，Modulation Network(LSTM)會先從 $\{x_k, y_k\}_{k=1, \dots, K}$ 為輸入並由 Task Encoder 產出 list of 任務向量 v (以 0,1 為表示)。

$$v = h(\{(x_k, y_k) \mid k = 1, \dots, K\}; w_k) \quad \text{式(27)}$$

$$\tau_i = g_i(v; \omega_g), \text{ where } i = 1, \dots, N; N \text{ 為 task network 的任務數量} \quad \text{式(28)}$$

利用算式 27 計算出參數 τ (task-related parameter，請看算式 28)，並用於 task network。而因 task network 具有多個任務，Modulation Network 會產出多個方塊，對應的則是每一個任務的 Input (state and parameter) 而且會進行位移和縮放為了能與其他 Output 有對應(knowledge transfer)。而有良好的任務初始值。

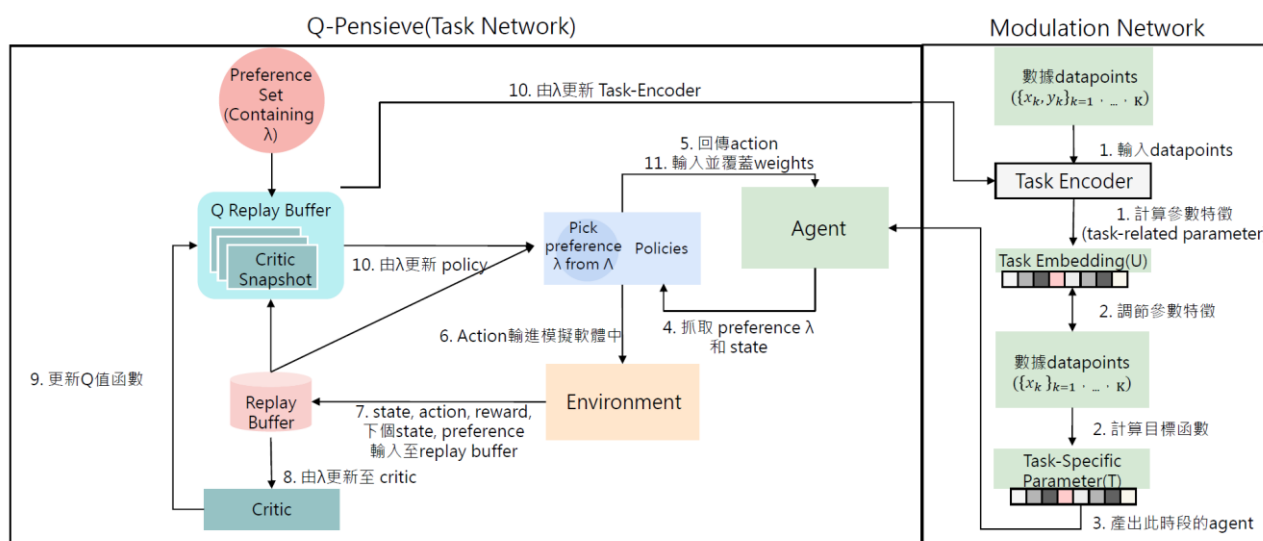
2. Task Network

Task network 具有多個任務，全部會用單一 agent (neural-network) 但會有多個 weights。而每一任務都會進行梯度適應(step action/gradient descent)，逐步改善其性能並覆蓋之前的 weights 與參數 T 。但要注意的是只有參數 T 會進行更新。因為參數 τ 代表 Task network 的初始化，只有在全部任務跑完並產出 data points(y_k)，才會進行更新(如圖 2.3.2-17)。

3. 多路口號誌控制的應用

而在多路口的號誌控制，Modulation Network (LSTM) 會以每一個路口的號誌周

期與紅綠燈秒數為資料點(data points)，並產出對應的任務目標與參數(θ, ω)，得以當作輸入給 Task Network。Task Network (Q-Pensieve)則會以每一個路口號誌(任務)進行梯度適應並逐步改善其性能。直到跑完所有路口並更新任務目標與參數(θ, ω)。



資料來源：[44], [7]

圖 2.3.2-18 多任務強化學習訓練架構

2.3.3 小結

2.3.2 節中各種強化學習方法，第一點至三點主要解釋 RL 模型的基本演算法。第四點至五點解釋前期計畫所使用 DDPG 與 PPO 演算法的運作方法與其限制，會使 AI 強化學習模型掉入區域最佳解而無法對於全局的其他次佳策略進行有效率的探索。第六點解釋 SAC 演算法在 DDPG 與 PPO 的基礎上加上最大熵(Entropy)為探索參數，並引用多 Q 網路使其更快收斂，但因前三種演算法都是單目標導致在交通的領域裡，單目標會使 AI 的訓練時間拉長而且多路口與時相的持續控制問題也會陷入問題。

因此本計畫引用 Q-Pensieve 的演算法應用於交通領域，以路口延滯、通過量與旅行時間作為多目標強化學習模式績效之一。而因 Q-Pensieve 會因時段與其輸入維度不同而有多種訓練模型，使其開發了 Task-Aware 希望能將多種時段與其資料型態一起做為決策之一。後續計畫會以 Q-Pensieve 與 Task-Aware 為主要演算法並繼續優化其性能並期望能在其他場域快速收斂與績效顯著。

2.4 人工智慧強化學習在號誌控制應用之關鍵課題

一、強化學習套用至交通環境之模型設定仍未有定論

由 2.3.1 小節可知，目前關於強化學習應用於路口號誌控制上，其環境狀態輸入、動作、獎勵設定皆未有定論，其關鍵點在於演算法之狀態空間與動作空間是否允許連續分布、以及該研究所關注的課題為何；此外，由於本計畫目的為與實務場域結合，來支援其號誌控制與管理需求，以提升車流運作效率，故資料蒐集設備限制亦納入模型設定考量之中。

二、應用多代理人深度強化學習之困境

由 2.3.1、2.3.3 小節可知，無論是國外文獻研究方向或場域實施計畫，皆以應用多代理人 (Agent) 深度強化學習於多路口號誌控制為未來趨勢，故如何妥善處理多代理人相互合作學習成為相當重要的議題，其中牽涉到 2 個困難點：第一為在多個代理人情況中，對某一代理人而言，因其主要互動環境包含其餘代理人，故執行動作時所需觀察的環境變化不僅是該路口之狀態，亦包括其餘代理人的動作變化，導致該代理人更新策略參數時所欲考慮的複雜度更高；第二為多路口號誌控制環境下狀態空間與動作空間的高維度與複雜度，將隨著路口數量呈指數性成長，係因該問題目標為整體路網之績效最大化，故須考慮每一路口代理人之聯合狀態空間與聯合動作空間。

三、實測場域既有限制

目前人工智慧強化學習號誌控制在既有時段型態與時相組合情況下，可產出數個符合尖離峰時段型態之號誌控制強化學習模式與訓練成果。為優化人工智慧強化學習模式應用於號誌控制領域，後續可以加入自動車流情境參數，動態調整時相組態，以更加符合號誌控制機制與內容。另外，既有號誌控制模式擁有最大綠燈與最小綠燈秒數限制，未來若考量路口各方向平均停等延滯、服務水準或偵測通過路口之行人數等績效指標，納入獎勵函數進行模式訓練，或有機會放寬該限制議題。

四、跨領域資訊整合

近期應用人工智慧強化學習於號誌控制系統之發展迅速，以及持續將與各縣市及高速公路局跨領域合作，故可嘗試介接與分析匝道儀控率或交流道車輛速度與流量等多元資料，將其納入模擬環境中進行模式訓練，使匝道各式資訊可以與人工智慧強化學習號誌控制邏輯結合。預期未來整合 貴所、產業、學術研究之能源，共同協助人工智慧號誌控制面之發展能量。

五、路側設備整合與調校議題

根據實作經驗，現場端整合與調校各式路側設備議題亦為重要。為滿足人工智慧強化學習號誌控制所需之偵測參數，須提高現場端車流偵測器設備準確度，降低資料利用錯誤率。另一方面，使用之號誌控制器韌體版本若能更新至最新，將可達到近似逐秒控制效能，並且為確保人工智慧強化學習號誌控制模式能即時與穩定運作，則須強化通訊設備連線率。

第三章 強化學習號誌控制模型探討

3.1 實測號誌控制模型檢討

一、前期(111 年)計畫 AI 強化學習模型架構

前期 AI 強化學習模型神經網路的架構圖如圖 3.1-1，臺北市及臺南市實驗場域均以此神經網路架構所構建，輸入為交通偵測資訊及號誌資訊，其中交通偵測資訊在臺北市實驗場域模型為停等資訊，臺南市實驗場域模型為轉向交通量資訊，模型輸出為下個時相的秒數。

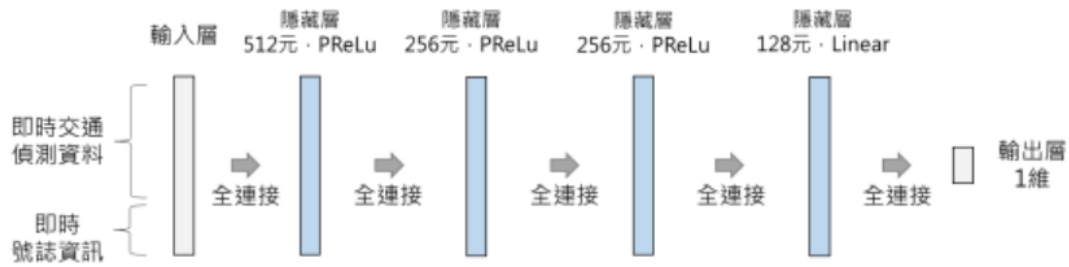


圖 3.1-1 前期代理人 AI 強化學習模型模型結構圖

二、前期(111 年)計畫 AI 強化學習模型使用的演算法

前期臺北市及臺南市實驗場域均使用 DDPG 作為訓練的演算法，並於 VISSIM 與 SUMO 兩套車流模擬軟體中的模擬環境進行訓練及測試，相關強化學習演算法的說明與比較可參考 2.3.2 節詳述。

三、AI 強化學習模型代理人說明

(一)臺北市實驗場域

本計畫臺北市實驗場域共計七個路口，前期模型使用三個 agent 分別控制三個路口，分別為：X2 中山北六段/克強路、X3 中山北六段/195 巷及 X4 中山北六段/德行西路，如下圖 3.1-2。



圖 3.1-2 臺北市實驗場域前期模型號誌化路口分布圖

前期模型以 3 個代理人分別控制 X2 中山北六段/克強路、X3 中山北六段/195 巷及 X4 中山北六段/德行西路，並依照不同偵測設備設置位置分別開發 B1、B2、B3、B4 共 4 個 AI 強化學習模型，因本期偵測設備設置位置為參考前期 B1 模型設置，以下說明 B1 模型代理人架構，如表 3.1-1，詳細偵測器布設位置可參考 4.3.1 節。3 個代理人輸入資訊均為號誌資訊及道路交通資訊，號誌資訊為代理人決策當下號誌時相指標與秒數資訊，道路交通資訊為決策當下分車道分車種(大車/小車/機車)車輛數，決策頻率為該路口時相變更時決策，代理人訓練目標以 X2 中山北六段/克強路、X3 中山北六段/195 巷及 X4 中山北六段/德行西路，此三個 AI 控制的路口系統總停等延滯最小。

表 3.1-1 前期 B1 模型臺北市實驗場域 AI 強化學習模型架構說明

定義	內容
代理人數量	3
輸入	號誌資訊：決策當下號誌時相指標與秒數資訊
	道路資訊(AI 控制路口)：決策當下分車道分車種(大車/小車/機車)車輛數
輸出	下個時相 AI 控制路口之綠燈秒數
決策頻率	AI 控制路口時相變更時決策
獎勵目標	系統總停等延滯

(二)臺南市實驗場域

前期模型為 1 個代理人控制 1 個路口，並依照不同偵測設備設置位置及偵測內容分別開發 A1、A2、A3、A4，因現場的偵測設備可偵測到轉向量資訊，故選擇前期發展的 A2 模型做為實測的模型，詳細偵測器布設位置可參考 4.3.1 節。前期 A2 模型中輸入分為號誌資訊及道路交通資訊，號誌資訊為決策當下號誌時相指標與秒數資訊，道路交通資訊則為決策瞬間，前一週期該時相轉向交通量，決策頻率為該路口時相變更時決策，訓練目標為系統總停等延滯最小。

表 3.1-2 前期 A2 模型臺南市實驗場域 AI 強化學習模型架構說明

定義	內容
代理人數量	1
輸入	號誌資訊：決策當下號誌時相指標與秒數資訊
	道路資訊：決策瞬間，前一週期該時相轉向交通量(分車種)
輸出	下個時相路口之綠燈秒數
決策頻率	時相變更時決策
獎勵目標	系統總停等延滯

四、前期模型訓練簡化議題

本期計畫於臺北市實驗場域 7 處路口經比對前期模型與當時現場時制後，確認有 4 處路口分相不一致，分別為 X2 中山北六段/克強路、X4 中山北六段/德行西路、X6 礁溪街/克強

路、X7 礁溪街/德行西路，進一步探討得知前期模型將部分分相進行簡化，以縮短訓練所需時間，以下針對簡化之四處路口進行說明。

(一)簡化東西向的號誌分相

經查 X2 中山北六段/克強路，前期模型簡化東西向號誌分相，當時現場使用編號 F0 時制計畫，如圖 3.1-3 所示，紅框處為模型簡化之分相，其中分相 3 為行人綠，分相 4 多開往西向之車流，模型中將分相 3、分相 4 進行簡化，執行固定秒數，於分相 5 時使用 AI 預測秒數排解分相 4 尚未消化之車潮。

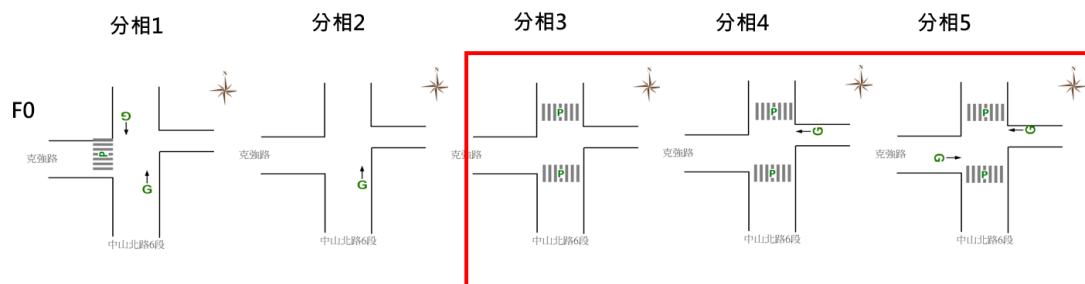


圖 3.1-3 X2 路口現場端時制示意圖

(二)簡化時制計畫

經查 X4 中山北六段/德行西路，現場端下午 16:30 至 17:00 使用編號 CF 時制計畫，其餘時段使用編號 A9 時制計畫，如下圖 3.1-4 所示，由於 CF 時制計畫僅執行半小時，考量整體訓練效率，簡化為所有時段均執行 A9 時制計畫。

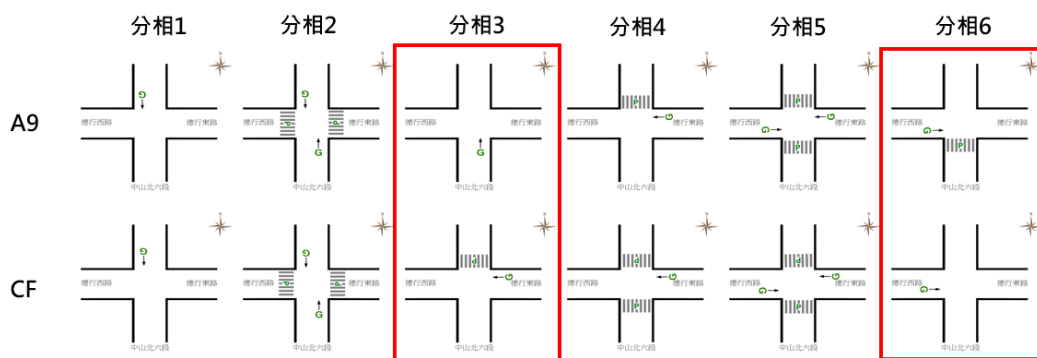


圖 3.1-4 X4 路口現場端時制示意圖

(三)簡化行人綠燈

經查 X6 礁溪街/克強路與 X7 礁溪街/德行西路，部分時段有行人綠燈的分相，前期模型省略此分相，以下分別說明：

1.X6 礁溪街/克強路

經查詢當時現場端中午 12:00 至 13:00、下午 16:00 至 17:00 使用編號 AB 時制

計畫，其餘時段使用編號 00 時制計畫，如下圖 3.1-5 所示，由圖可知 AB 時制計畫之分相 2 為行人綠，與車流無關，模型中將分相 2 進行簡化，執行固定秒數，於分相 1、分相 3 使用 AI 決策預測秒數。

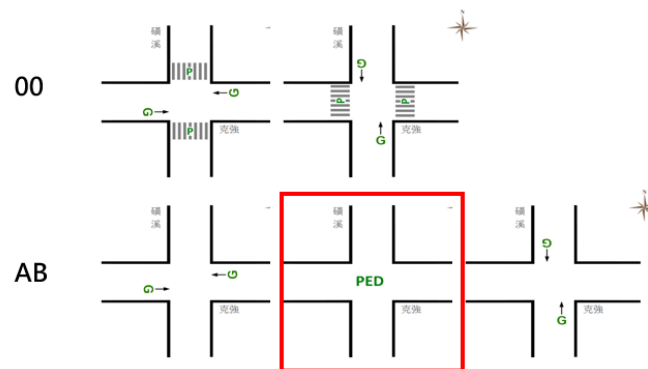


圖 3.1-5 X6 路口現場端時制示意圖

2.X7 礮溪街/德行西路

經查詢當時現場端於平日上午尖峰 7:00 至 9:00、平日下午尖峰 16:30 至 17:00，使用編號 AF 時制計畫，其餘時段使用編號 AD 時制計畫，如下圖 3.1-6 所示，由圖可知 AD 時制計畫之分相 4、AF 時制計畫之分相 5 皆為行人綠，與車流無關，模型中將 AD 時制計畫之分相 4、AF 時制計畫之分相 5 進行簡化，執行固定秒數。

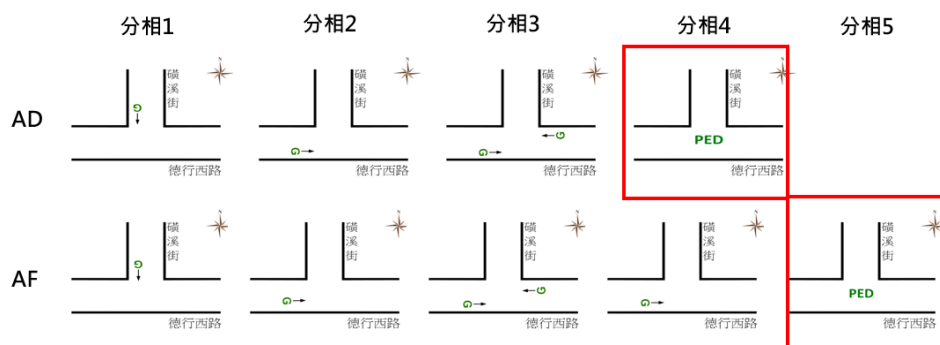


圖 3.1-6 X6 路口現場端時制示意圖

五、臺北市實驗場域時制計畫時相組態變動

前期 AI 強化學習模型依當時之分相數進行設計與訓練，111 年 12 月 9 日時於 X2 中山北六段/克強路新增一個分相，導致實際輸入的維度於 X2 路口處，多了 AI 模型一個輸入維度，如圖 3.1-7、3.1-8，模型訓練時，X2 路口輸入為時制計畫時相編號 F0。而於本期實測時，實際時制計畫時相編號改為 CF，多了一個北向車流的分相，當實測期間，現場號誌的分相走到分相 1 時，前期 AI 強化學習模型無法輸入時相編號編號 CF 的分相 1 編號及剩餘的秒數，使前期模型無法進行實測。

時制計畫歷史設定

設備編號	時制計畫	週期	時差	方向	時相	分相數	下載時間	執行方式
SZYII10	9	200	70	北	F0	5	2022-08-09 11:17:44	中心下載
SZYII10	9	200	70	北	CF	6	2022-12-09 12:07:43	中心下載
SZYII10	9	200	70	北	CF	6	2023-05-16 20:32:52	中心下載
SZYII10	9	200	70	北	CF	6	2023-05-25 18:52:17	中心下載

圖 3.1-7 X2 路口臺北市交控中心修改時制示意圖

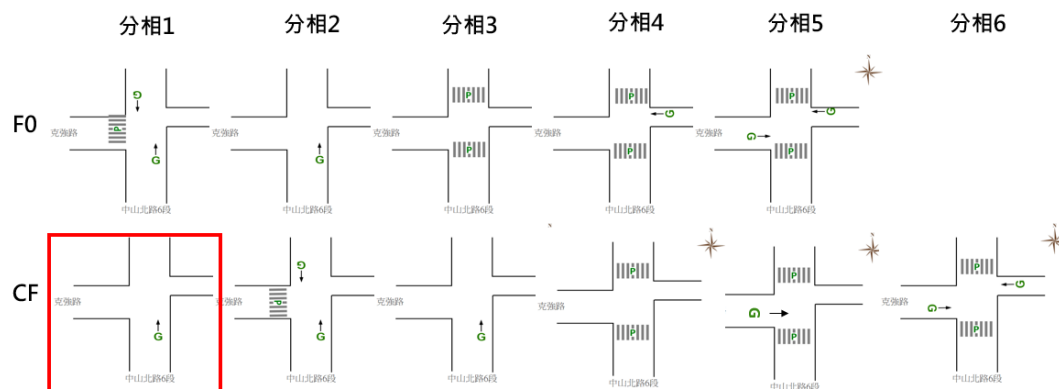


圖 3.1-8 X2 路口編號 CF 與 F0 時制比較圖

除以上原因外，前期模型 X2 中山北六段/克強路輸出為下個分相的綠燈秒數，111 年 12 月 9 日 X2 中山北六段/克強路新增一個分相，而本案於後續實測期間，現場號誌的分相走到分相 1 時，前期模型無法輸出編號 CF 的分相 1 綠燈秒數，同樣使得前期模型無法實測。

綜上所述可知，X2 中山北六段/克強路現行時制計畫改變，使得前期 AI 強化學習模型之輸入輸出無法對應至現行時制計畫，導致前期模型無法於現場進行實測，另外對於號誌組態的改變，不限於 AI 強化學習模型須重新構建，幾乎現行的動態號誌控制模型，都須重新構建。

六、本期 AI 強化學習模型優化原因及方向

(一)前期模型與現況差異

1.時制計畫改變

因臺北市實驗場域 X2 中山北六段/克強路時制計畫改變，如圖 3.1-7 及此章第五節說明，使前期 AI 強化學習模型輸入不一致，無法實測前期的模型，此部分將依據實測場域目前的時制計畫，構建新的 AI 強化學習模型。

2.車流資訊

臺北市實驗場域設備偵測停等面臨兩個模型與現況的落差，分別為起始偵測位置的不同，與偵測長度的不同，以下分別說明優化的方向：

(1)起始偵測位置的不同

前期模型訓練時，輸入的停等偵測位置皆為車輛停止線前開始計算，現場設備則是除了克強路往東的方向外，其餘所有車道的設備起算位置皆從停等線前 15 或 20 公尺才開始偵測停等，如圖 3.1-9、3.1-10 所示，後續將依照現場偵測設備的起始位置，重新訓練模型。



圖 3.1-9 臺北市實驗場域中山北路六段/克強路設備偵測範圍



圖 3.1-10 臺北市實驗場域中山北路六段/195 巷設備偵測範圍



圖 3.1-11 臺北市實驗場域中山北路六段/德行西路設備偵測範圍

(2)偵測長度的不同

前期模型訓練時，輸入的停等偵測長度部分區域高於現場偵測設備，導致實測前期模型時，AI 在判斷時會有誤差，如圖 3.1-9、3.1-10 所示，後續將依照現場偵測設備的偵測長度，重新訓練模型。

(二)本期模型優化的方向

1.多路口協控之精進

為了使臺北市實驗場域 3 個 AI 控制路口的週期一致，本期將以 1 個代理人控制此 3 個路口，並改由每週期決策一次，與前期的 3 個代理人控制 3 個路口，每時相決策一次不同。期待本期 AI 強化學習模型，在同週期的情況下，幹道車流能有好的續進。

2.多目標演算法之精進

前期使用 DDPG 的演算法作為訓練的 AI 強化學習模型，該演算法為確定性策略(Deterministic Policy)，雖然在訓練上更快速，但無法像本期使用之演算法 Q-Pensieve，找出全域之最佳解，另外 DDPG 係以單一目標的獎勵(reward)進行訓練，無法同時針對不同目標去做最佳化(詳細強化學習演算法之比較可參閱 2.3.2)。因此，本期將以 Q-Pensieve 演算法重新訓練臺北市及臺南市實驗場域 AI 強化學習模型，此模型可根據不同目標獎勵(Reward)的需要，動態地將它們以線性組合的方式形成新的最終預期目標，並同時針對所有組合進行最佳化。這將使得 AI 強化學習模型更靈活地適應各種情況，如幹道優先或支道優先等，提高訓練效率並改善性能。

3.多任務強化學習模型

前期模型須依照不同時制組態，拆分不同時段並訓練對應之模型，本期除了開發多目標強化學習模型也額外開發了多任務強化學習模型(Task-Aware)。因多目標學習也會因時制組態的維度不同而拆分並訓練對應之模型。但多任務則會將不同時制組態拆分成不同的任務(Task)，並訓練一個模型，使該模型能學習不同時制組態，不須為不同時段訓練不同模型(詳細文獻及架構可參閱 2.3.2)，此模型因計畫期程因素，僅於臺南市實驗場域之模擬環境軟體中進行測試，並於下一章呈現其結果。

3.2 號誌控制模型優化與精進

3.2.1 AI 強化學習模型精進架構

一、AI 強化學習模型規劃

(一)臺北市實驗場域

表 3.2.1-1 臺北市實驗場域 AI 強化學習模型規劃表

定義	內容
代理人數量	1
輸入	號誌資訊：上個週期路口週期秒數及時相比例
	道路資訊(AI 控制路口)：各分相結束時分車道分車種(大車/小車/機車)停等車輛數及總車輛數
輸出	下個週期 AI 控制路口之週期秒數及時相比例
決策頻率	AI 控制路口週期結束時決策
獎勵目標	1. 幹道延滯時間與固定時制的變化量 2. 支道延滯時間與固定時制的變化量 3. 幹道通過量與固定時制的變化量 4. 支道通過量與固定時制的變化量

表 3.2.1-2 臺北市實驗場域 AI 強化學習模型獎勵計算公式表

定義	計算方式
幹道延滯時間與固定時制的變化量	$\frac{\text{AI 幹道每秒平均延滯時間} - \text{固定時制幹道每秒平均延滯時間}}{\text{固定時制幹道每秒平均延滯時間}}$
支道延滯時間與固定時制的變化量	$\frac{\text{AI 支道每秒平均延滯時間} - \text{固定時制支道每秒平均延滯時間}}{\text{固定時制支道每秒平均延滯時間}}$
幹道通過量與固定時制的變化量	$\frac{\text{AI 幹道每秒平均通過量} - \text{固定時制幹道每秒平均通過量}}{\text{固定時制幹道每秒平均通過量}}$
支道通過量與固定時制的變化量	$\frac{\text{AI 支道每秒平均通過量} - \text{固定時制支道每秒平均通過量}}{\text{固定時制支道每秒平均通過量}}$

註 1：幹道為南北向的中山北路段，支道則為連接中山北路的東西向路段

臺北市實驗場域 AI 強化學習模型規劃由 1 個代理人控制 3 個路口，分別為中山北六段/克強路口(X2)、中山北六段/195 巷(X3)、中山北六段/德行西路口(X4)，因三個路口由 1 個代理人決策，決策時間點 3 個路口必須相同。因此，決策頻率為此 3 個路口週期結束時決策，此 3 個路口週期會一致，輸入為上個週期路口週期秒數、各分相之比例及 AI 控制路口分車道分車種的總車輛數及停等車輛數，其中車速低於 5km/hr 計為停等車輛，總車輛數則不考慮車速。只要該車道有偵測到車輛便計入總車輛數，輸出則為下個週期此 3 個路口的週期秒數及各個分相的比例。

為使 AI 強化學習模型能更有效率的學習，上述獎勵目標將與固定時制(TOD)於該時段的平均績效表現對應，期待 AI 強化學習模型能往高於固定時制(TOD)的績效探索與學習；另外為避免往過大與過小的時相秒數探索，獎勵目標均轉換為每秒為單位，即該時相的總通過量除以該時相秒數，為該時段內的每秒為單位的通過量。延滯亦為相同的計算方式，並代表該時段內的每秒平均延滯，期待 AI 強化學習模型避免為提高通過量而加入過大秒數，為降低延滯減至過低的秒數。詳細獎勵目標計算公式如上表 3.2.1-2；為使代理人能蒐集到週期內的資訊，輸入(state)資訊以週期內所有路口的時相為統計。

(二)臺南市實驗場域

表 3.2.1-3 臺南市實驗場域模型規劃表

定義	內容
代理人數量	1
輸入	號誌資訊：路口時相即將結束時的時相編號與剩餘秒數
	道路資訊：各分相結束時分車種(大車/小車/機車)轉向量
輸出	下個時相路口之綠燈秒數
決策頻率	路口時相結束時決策
獎勵目標	1. 台 19 甲延滯時間與固定時制的變化量 2. 台 86 延滯時間與固定時制的變化量 3. 台 19 甲通過量與固定時制的變化量 4. 台 86 通過量與固定時制的變化量

表 3.2.1-4 臺南市實驗場域 AI 強化學習模型獎勵計算公式表

定義	計算方式
台 19 甲延滯時間與固定時制變化量	$\frac{\text{AI 台 19 甲每秒平均延滯時間} - \text{固定時制每秒平均延滯時間}}{\text{固定時制每秒平均延滯時間}}$
台 86 延滯時間與固定時制的變化量	$\frac{\text{AI 台 86 每秒平均延滯時間} - \text{固定時制每秒平均延滯時間}}{\text{固定時制每秒平均延滯時間}}$
台 19 甲通過量與固定時制的變化量	$\frac{\text{AI 台 19 甲每秒平均通過量} - \text{固定時制每秒平均通過量}}{\text{固定時制每秒平均通過量}}$
台 86 通過量與固定時制的變化量	$\frac{\text{AI 台 86 每秒平均通過量} - \text{固定時制每秒平均通過量}}{\text{固定時制每秒平均通過量}}$

臺南市實驗場域 AI 強化學習模型規劃由 1 個代理人控制該路口，輸入為路口時相即將結束時的時相編號、剩餘秒數與時相剩餘秒數及分車種之轉向量，輸出則為下個時相路口之綠燈秒數，決策頻率為每個時相結束時決策。獎勵目標與臺北市實驗場域設計概念相同，為了使 AI 強化學習模型有效率的學習，與避免往過大與過小的時相秒數探索，臺南市實驗場域獎勵目標均與固定時制(TOD)的平均績效比較，通過量與延滯的設計概念亦與臺北市實驗場域相同，均轉換為該時相的單位平均績效，詳細獎勵目標計算公式 3.2.1-4；為使代理人能蒐集到時相內的資訊，獎勵目標及輸入資訊以時相為單位進行統計。

二、實測模型架構

(一)臺北市實驗場域

本期臺北市實驗場域 policy 模型神經網路的架構圖如圖 3.2.1-1，輸入為交通偵測資訊及號誌資訊，其中交通偵測資訊為 AI 號誌控制路口分相結束時，分車種的總車輛數及停等車輛數，號誌資訊則為上個週期的各路口的分相比例及週期秒數。獎勵偏好為訓練過程中，代理人自動產生一個總和為 1 的向量，獎勵偏好維度大小與目標獎勵

維度一致，輸入模型前，交通偵測資訊、號誌資訊和獎勵偏好會串接為一個輸入用資料數據一併傳入模型中，輸入的資料透過隱藏層的傳遞去擷取特徵資料，並藉由通過激活函數(Activation Function) Relu 去避免梯度消失問題，最終模型的輸出為一個平均數與變異數，用來表示模型預測的動作機率分布，透過從機率分布的抽樣，可以得到下個週期 AI 控制路口之週期秒數及各分相之比例。

本研究使用的 Critic 架構會有兩個 Q 值負責計算 policy 做出決策的品質，每次選擇 Q 值小的一個作為 target Q 值(詳見 2.3.2 第三節 SAC 說明)，其輸入基準為週期並與 policy 相同在額外加上 policy 的決策結果，同時通過兩個隱藏層，最後透過激活函數 Relu 計算獎勵績效的期望值，以供後續 policy 學習的參考避免過激的決策發生，critic 模型神經網路的架構圖如圖 3.2.1-2。

經查臺北市實驗場域中，預計由 AI 強化學習模型控制之路口中山北六段/克強路、中山北六段/195 巷有行人綠燈分相，因 AI 強化學習模型僅透過車流資訊控制號誌，行人資訊並不在 AI 強化學習模型分析的範圍內。為此在模型設計裡，AI 強化學習模型輸出週期秒數及時相比例時，如以固定時制控制 AI 強化學習模型行人綠燈秒數作為該分相輸出結果，會影響其他分相之秒數輸出結果和路口上的續進，故此兩個路口在行人綠燈的分相中，輸出為現場固定時制(TOD)行人綠燈的時相比例，AI 強化學習模型則會以此比例調整該時相比例之輸出結果。

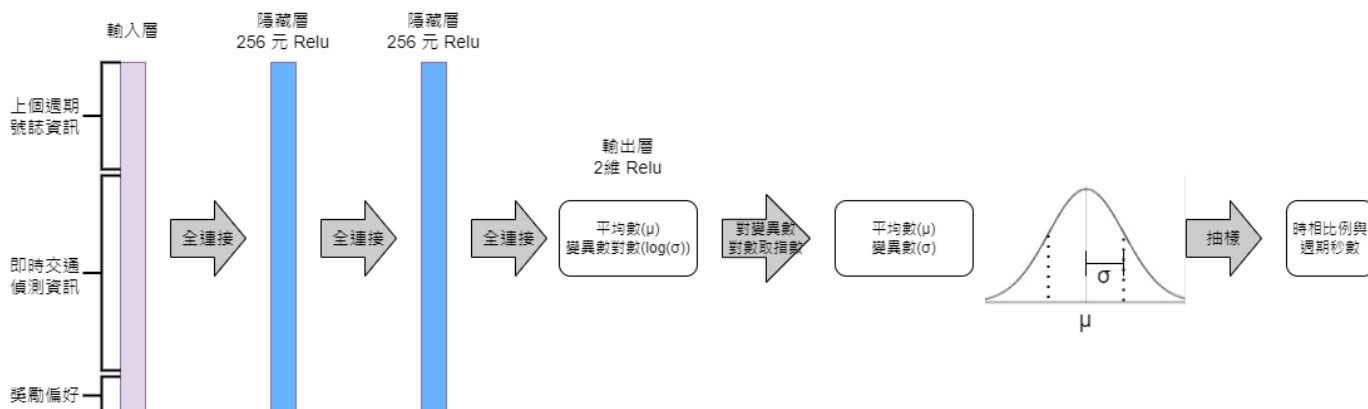


圖 3.2.1-1 本期臺北市實驗場域 Policy 模型結構圖

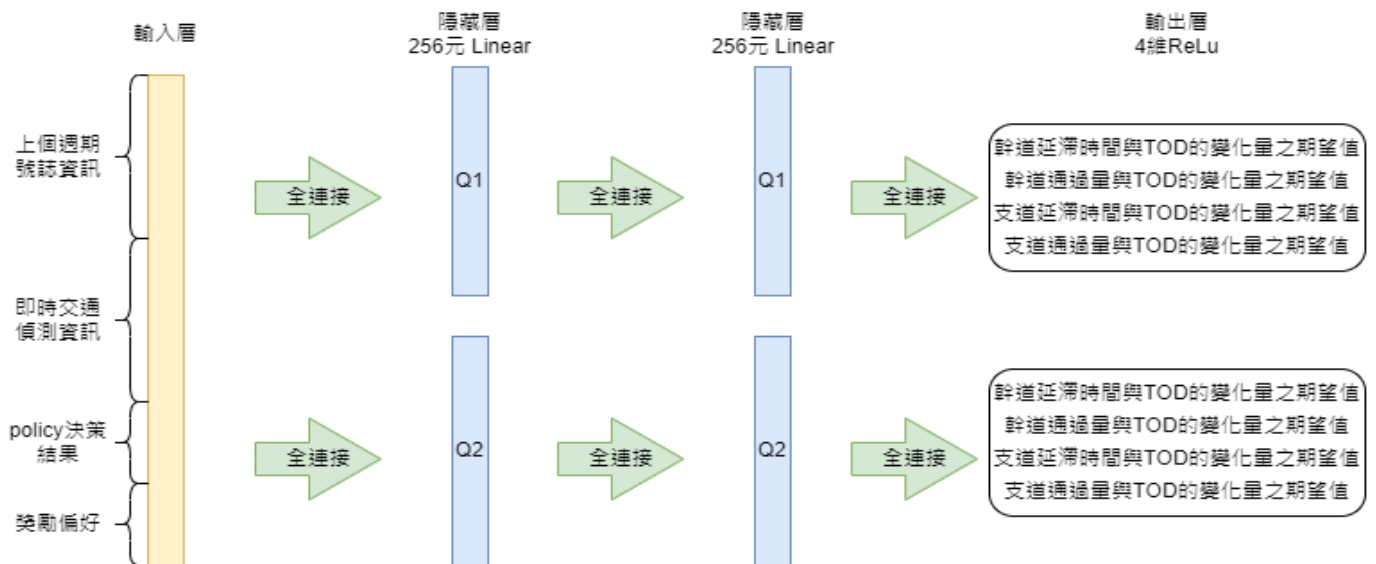


圖 3.2.1-2 本期臺北市實驗場域 Critic 模型結構圖

(二)臺南市實驗場域

本期臺南市實驗場域 AI 強化學習模型神經網路的架構圖如圖 3.2.1-3，輸入為交通的偵測資訊及號誌資訊，其中交通的偵測資訊為轉向資訊，號誌資訊則為目前處在的時相及時相剩餘秒數，獎勵偏好為訓練過程中，代理人自動產生一個總和為 1 的向量，獎勵偏好維度大小與目標獎勵維度一致，輸入模型前，交通的偵測資訊、號誌資訊和獎勵偏好會串接為一個輸入用資料數據一併傳入模型中，輸入的資料透過隱藏層的傳遞去擷取特徵資料，並藉由通過激活函數 Relu 去避免梯度消失問題，最終模型的輸出為一個平均數與變異數，用來表示模型預測的動作機率分布，透過從機率分布的抽樣，可以得到下個時相的秒數。

臺南市實驗場域 critic 架構與臺南市實驗場域的類似，惟其 policy 輸入基準不同，故臺南市實驗場域為時相控制，導致 critic 輸入為時相作為基準而非週期，後續 critic 學習與計算則與臺北市實驗場域一樣，臺南市實驗場域 critic 模型神經網路的架構圖如圖 3.2.1-4。

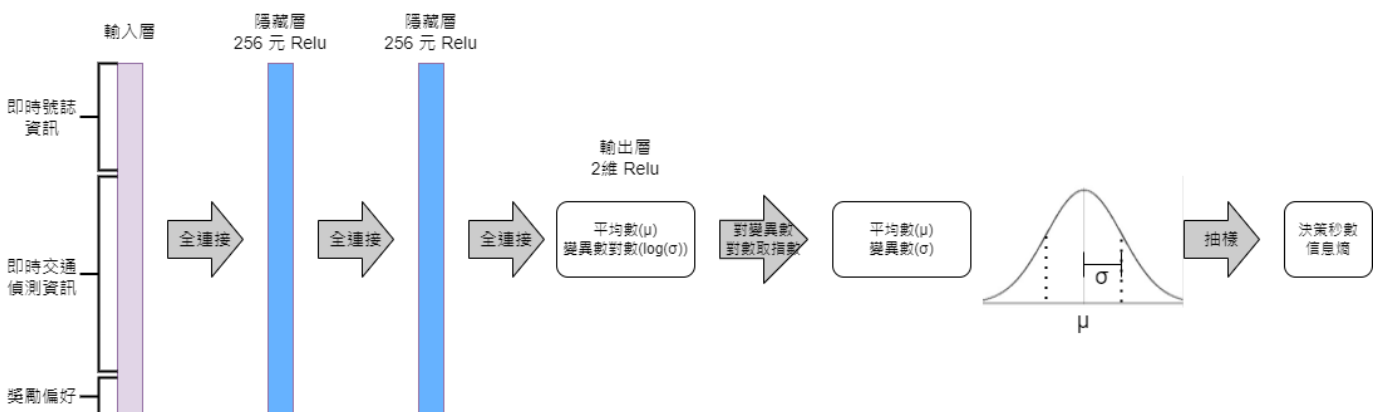


圖 3.2.1-3 本期臺南市實驗場域 Policy 模型結構圖

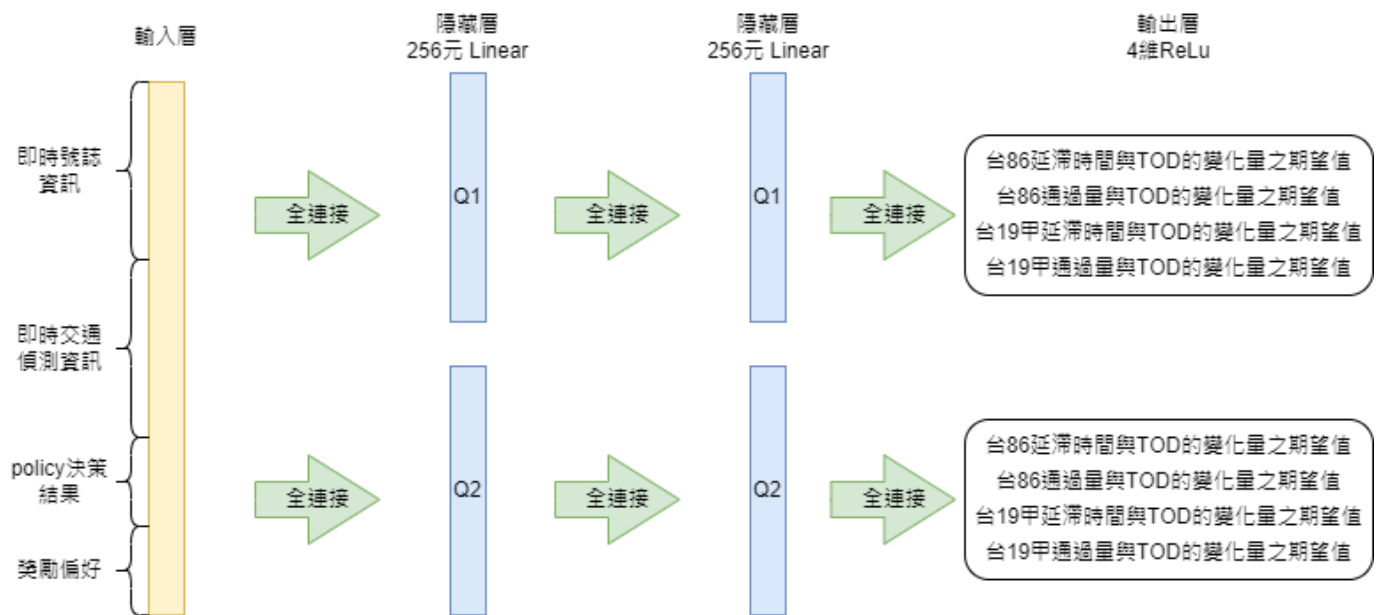


圖 3.2.1-4 本期臺北市實驗場域 Critic 模型結構圖

(三)臺北市及臺南市實驗場域模型比較

本期臺北市及臺南市實驗場域模型架構比較如表 3.2.1-5，因現場偵測設備的不同，臺北市實驗場域模型為分車道車輛數與停等作為模型輸入，臺南市實驗場域則為轉向量；由於臺北市實驗場域為多路口幹道，為維持 AI 號誌控制路口週期一致，故決策頻率為每週期決策一次，輸出週期秒數與分相比比例，臺南市實驗場域則為單一路口，決策頻率為時相決策，輸出則為分相秒數；臺北市與臺南市實驗場域之獎勵設計概念相同，均以幹支道的延滯與通過量之多目標獎勵。

當臺北市實驗場域週期結束時，將蒐集上個週期的號誌及車流資訊，使 AI 強化學習模型預測下個週期的秒數與時比；臺南市實驗場域則為時相結束時，蒐集時相內的車流與號誌資訊，使 AI 強化學習模型預測下個時相的綠燈秒數。AI 強化學習模型神經網路運算時間極短段，一秒內便能完成輸入資訊之運算並輸出資訊，因此，臺北市與臺南市實驗場域模型主要差異為輸入輸出及決策頻率。

表 3.2.1-5 臺北市與臺南市實驗場域模型比較表

場域	現場偵測設備	輸入車流資訊	決策頻率	輸入號誌資訊	輸出	獎勵
臺北市	VD 雷達偵測器偵測車道車輛數	各分相結束時分車道分車種(大車/小車/機車)停等車輛數及總車輛數	週期結束	上個週期及時比	下個週期秒數及時比	● 幹道延滯時間與固定時制的變化量 ● 支道延滯時間與固定時制的變化量 ● 幹道通過量與固定時制的變化量

						● 支道通過量與固定時制的變化量
臺南市	CCTV 影像辨識偵測轉向量	各分相結束時分車種(大車/小車/機車)轉向量	時相結束	時相編號與剩餘秒數	下個時相的分相秒數	● 台 19 甲延滯時間與固定時制的變化量 ● 台 86 延滯時間與固定時制的變化量 ● 台 19 甲通過量與固定時制的變化量 ● 台 86 通過量與固定時制的變化量

(四)臺北市及臺南市實驗場域 AI 強化學習模型秒數限制說明

實測時為避免 AI 強化學習模型輸出之秒數與用路人習慣不同，導致民眾反彈，因此增加 AI 強化學習模型輸出之綠燈秒數限制，臺北市及臺南市實驗場域之限制因場域設計與行人綠燈時相而有所不同，如表 3.2.1-6、3.2.1-7 與 3.2.1-8。

表 3.2.1-6 臺北市實驗場域 AI 強化學習模型秒數限制表

	X2 時相 2 限制比例	X2 時相 3 限制比例	X3 時相 1 限制比例	X4 時相 1 限制比例
下限	總週期的 0.55	時相 2 的 0.2	總週期的 0.4	總週期的 0.4
上限	總週期的 0.75	時相 2 的 0.3	總週期的 0.6	總週期的 0.5

表 3.2.1-7 臺北市實驗場域 AI 強化學習模型行人時相限制表

比例同固定時制	X2 時相 4 行人綠比例	X3 時相 3 行人綠比例
7:00-9:00(平日)	總週期的 0.045	總週期的 0.175
9:00-13:00(平日)	總週期的 0.06	總週期的 0.233
13:00-16:30(平日)	總週期的 0.045	總週期的 0.275
16:30-19:00(平日)	總週期的 0.045	總週期的 0.275
19:00-21:00(平日)	總週期的 0.06	總週期的 0.233
16:30-21:00(假日)	總週期的 0.045	總週期的 0.175
19:00-21:00(假日)	總週期的 0.045	總週期的 0.175

表 3.2.1-8 臺南市實驗場域 AI 強化學習模型秒數限制表

	早開時相及左轉保護時相	臺 19 甲直右時相	臺 86 直右時相
下限	15 秒	20 秒	60 秒
上限	45 秒	45 秒	120 秒

臺南市實驗場域為時相控制，故 AI 強化學習模型輸出之綠燈秒數會由 AI 強化學習模型以即時車流狀況為依據，在綠燈上下限中進行決策；臺北市實驗場域則為週期控制，時相秒數會以總週期的比例做為限制並且只會控制南北向之時相與行人綠燈時向。設定之綠燈秒數限制皆由原 TOD 秒數作為基礎進行改良。

三、Q-Pensieve 模型訓練架構說明

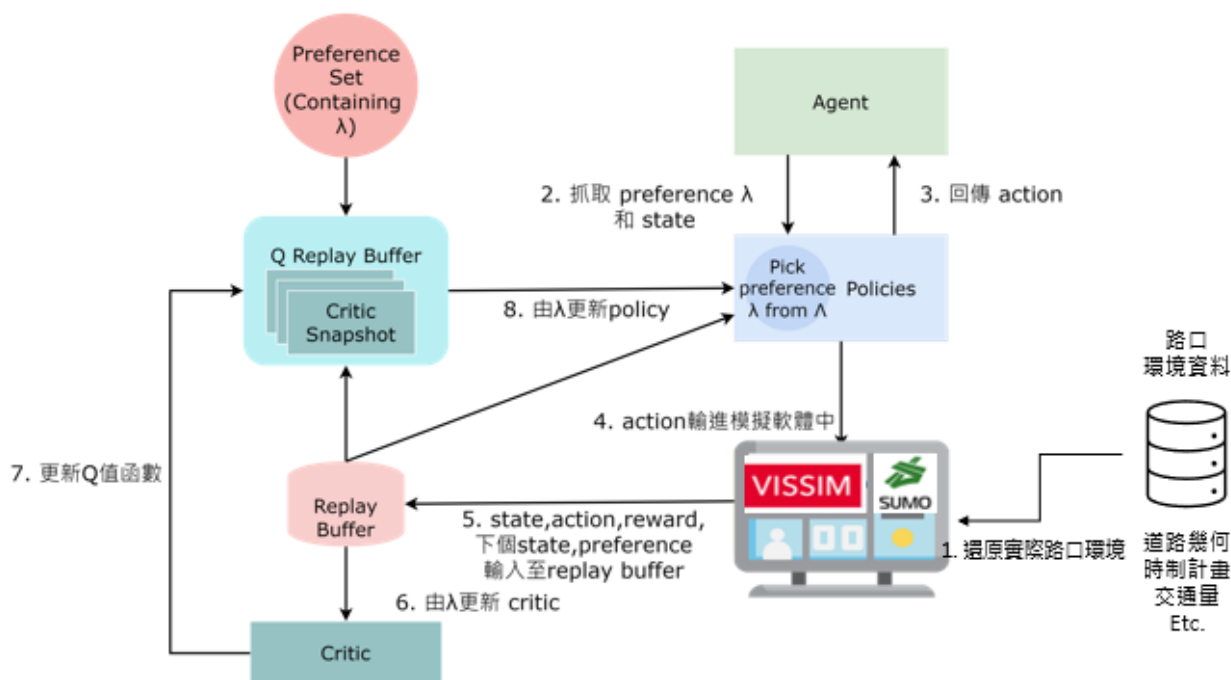


圖 3.2.1-5 本期 Q-Pensieve 代理人模型訓練架構圖

本期使用的強化學習演算法為 Q-Pensieve，相關演算法的說明與比較詳 2.3.2 節，Q-Pensieve 訓練架構如圖 3.2.1-3，並說明如下：

- (一)由人工調查或現場偵測器收集的資料，利用 VISSIM 跟 SUMO(模擬環境)還原現場的車流情況。
- (二)代理人從模擬環境抓取 State 跟從本身選取一個策略函數的獎勵偏好(preference λ)並給予 Policy Network。Policy Network 是決定在給定狀態(state)和獎勵偏好(preference)下採取哪個動作(action)的函數。其目標是最大化從環境中獲得的累積獎勵並利用持續更新與優化，以提高在不同狀態下選擇動作的能力。
- (三)Policy Network 從策略函數中選擇一種動作(action)準備與環境互動。
- (四)將此動作輸入進模擬環境裡產出下個環境(state)和獎勵(Reward)。
- (五)將這次的環境(state)、動作(action)、獎勵(reward)和下個環境(state)梳理成一個 data transition (數據資料)存進 replay buffer(資料庫)。
- (六)從資料庫裡隨機選擇一定數量的環境(state)、動作(action)、獎勵(reward)和下個環境(state)給予 Critic Network 並由它評估產出此環境(state)的策略預期價值(會有好有壞)。而後續將會持續更新和改進以更準確地估計此策略的預期價值。

(七)而這價值會以 Q 值函數 $Q(s,a)$ 為輸出(output)，並儲存到另外一種資料庫(Q replay buffer)。Q replay buffer 會存取 Critic Network 所產出的 Q 值函數並以兩種資料：
1.preference set(λ)、2.資料庫(replay buffer)的各種過去資料作為輸入。

(八)從資料庫(replay buffer)裡隨機選擇一定數量的 experience，Q-replay buffer 會利用這些 Action 使用的 λ 和 preference set 抽取出不同 λ 產出不同的 Q 值函數來更新 Policy Network。而更改過的 Policy Network 會重複步驟三並繼續執行步驟四到八直至 Policy Network 達到最優解(無法更新更好的 Q 值函數)。

四、Q-Pensieve 演算法超參數(hyperparameter)說明

Q-Pensieve 的訓練過程會有超參數需要調整，像是 Q-replay buffer 大小或是 Q-replay buffer 的更新頻率，實作上我們會使用(Hung et al, 2023.) [20] 所列出的訓練參數，使用的參數如表 3.2.1-3。Hung 等人提到，Q-replay buffer 更新頻率的選取對於訓練結果不敏感，選擇 1000~2000 做為更新頻率都是合適的。針對 Q-Replay Buffer 和 Preference Set 則是透過實作的觀察，權衡訓練時間和最終性能所選擇，兩者的大小只要設定為 4，都能在犧牲較少的訓練時間達到較高的訓練增長。

表 3.2.1-9 Q-Pensieve 訓練參數

Parameter	Value
Optimizer	Adam
Learning Rate	0.0003
Discount Factor	0.99
Replay Buffer Size	1000000
Depth of Hidden Layers	2
Number of Hidden Units per Layer	256
Number of Samples per Minibatch	256
Nonlinearity	ReLU
Target Smoothing Coefficient	0.005
Target Update Interval	1
Gradient Steps	1
Q-Replay Buffer Size	4
Preference Set Size	4
Q-Replay Buffer Update Frequency	1000

五、多任務強化學習模式(Task-Aware)

因 Q-Pensieve 演算法依舊處於單模態的學習模式導致在不同維度下的時制組態無法進行知識共享，而多任務強化學習則會將不同的時制組態根據維度給予不同的任務(task)，透過增加一個 Modulation Network，實現單一模式訓練並學習不同的時制組態(如 2.1.2 與 3.1.6 多任務強化學習之說明與下圖 3.2.1-7 所示)。後續將呈現臺南市實驗場域的兩個時段的模擬績效，此多任務強化學習模型受限於時程係，不進行實測。

(一)使用所有不同時段的交通資料通過 Modulation Network 去得到對應時段的特徵參數

(task embedding)，代理人通過特徵參數與 Policy Network 內部的參數內積去取得對應時段的 Policy Network。

(二)Policy Network 從策略函數中選擇一種行動(action)準備與環境互動。

(三)將此行動輸入進對應時段的模擬環境裡產出下個環境(state)和獎勵(reward)。

(四)將這次的偏好(preference, λ)、環境(state)、動作(action)、獎勵(reward)、下個環境(state)和特徵參數(task embedding)梳理成一個 data transition (數據資料)存進 replay buffer(資料庫)。

(五)從資料庫(replay buffer)裡隨機選擇一定數量的 data transition，Q-replay buffer 會利用這些 action 使用的偏好和 preference set 抽取出不同偏好產出不同的 Q 值函數來更新 Policy Network 和 Modulation Network。

(六)從資料庫(replay buffer)裡隨機選擇一定數量的 data transition，給予 Critic Network 並由它評估產出此環境(state)的策略預期價值(會有好有壞)，根據真實環境給予的獎勵(reward)，來更新 Critic Network。

(七)Critic Network 會以 Q 值函數 $Q(s,a)$ 為輸出(output)，並儲存到另外一種資料庫(Q-replay buffer)。Q-replay buffer 會存取 Critic Network 所產出的 Q 值函數並以兩種資料：preference set(Λ)、資料庫(replay buffer)的各種過去資料作為輸入。

(八)代理人會重複更新直至 Policy Network 達到最優解(無法更新更好的 Q 值函數)。

3.2.2 AI 強化學習模型訓練模擬環境程式與軟硬體說明

VISSIM 以 COMAPI 指令動態調整或取得模擬環境中的資訊，SUMO 則是以 TraCI API，在模擬中取得環境的狀態或是操作號誌燈號及秒數。在訓練開始前，取得模擬資料如轉向、流量與路網等資料，在開始模擬後初始化模擬期間所需記錄的參數，於暖機時間後才開始記錄環境(state)、獎勵(reward)等模型訓練所需的資訊，代理人須做決策時，將環境(state)、獎勵(reward)提供予代理人，再將代理人產出的動作(action)轉成模擬軟體中的號誌秒數，並持續執行模擬與蒐集績效；從蒐集模擬軟體中的資訊到執行代理人產生的動作(action)為一個 step，重複執行 step，直到超過模擬時間，才跳出模擬，而模擬時間為一個 episode。

另外，本研究訓練所使用的軟硬體如下表 3.2.2-1，AI 強化學習模型訓練以 pytorch 構建神經網路與強化學習架構，並以 VISSIM、SUMO 作為模擬環境，其中 VISSIM 的版本為 2022 SP09、SUMO 為 1.16.0、pytorch 為 2.0.1、Python 為 3.9。

表 3.2.2-1 本計畫 AI 強化學習模型訓練使用之軟硬體規格

項目	規格
CPU	Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz
GPU	NVIDIA GeForce RTX 2080
Python	3.9
Pytorch	2.0.1
VISSIM	2022 SP09
SUMO	1.16.0

本研究演算法需使用大量資源訓練各區域與時段的偏好模型，為了快速訓練與收斂，本案使用國家高速網路與計算中心(國網中心 TWCC)之運算資源，因國網中心提供 jupyter 開發環境與 AI 基本訓練環境與其能力能同時部署多個程式，做為 AI 強化學習模型分散式訓練中，利於訓練多個時段的模型，本研究使用其相關的資源如下表 3.2.2-2。

表 3.2.2-2 國網中心使用之規格

項目	規格
型號	cm.super
CPU	4 核
GPU	1 張
記憶體容量	60GB

3.2.3 AI 強化學習模型訓練結果

一、臺北市實驗場域訓練結果

(一)AI 強化學習模型訓練收斂過程

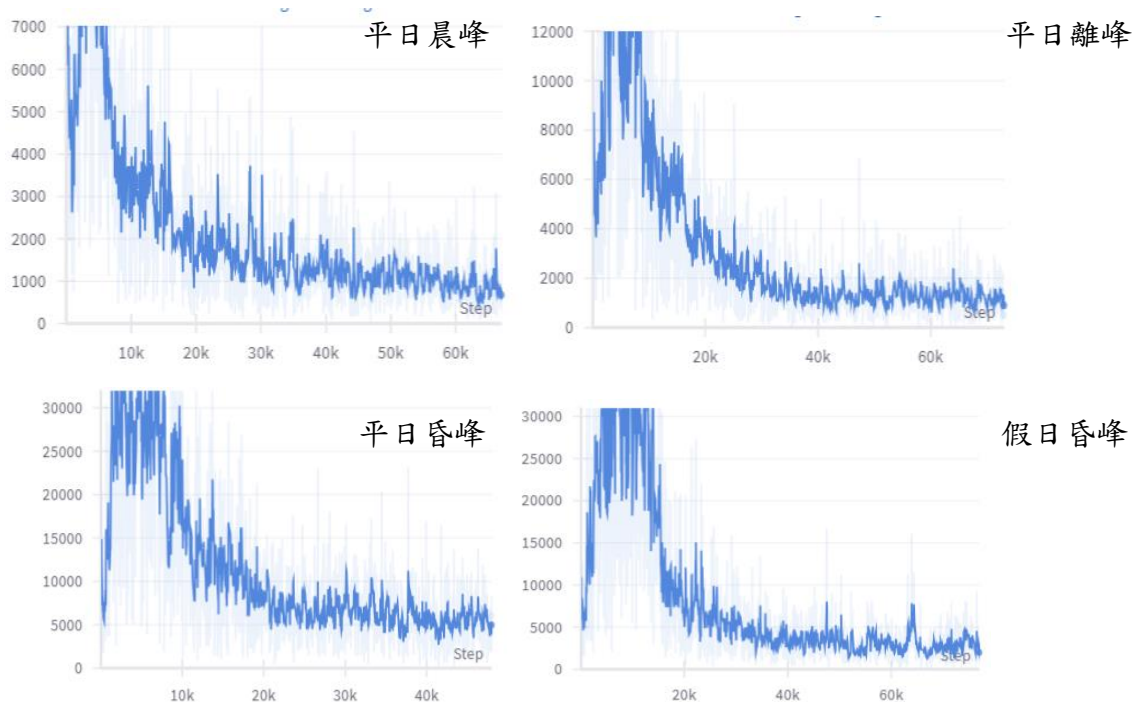


圖 3.2.3-1 臺北市實驗場域全時段 AI 強化學習模型訓練指標

訓練過程中績效將隨著類神經網路逐漸收斂而逐漸的進步，由圖 3.2.3-1 所示，y 軸代表 Loss，x 軸代表訓練的步數(step)。在臺北市實驗場域訓練大約 6 萬步與每一步為週期轉換時做更新，也因其演算法的探索特性其績效指標會有較大的浮動但有收斂的趨勢。其收斂程度會由上圖與模擬績效來決定是否使用此模型作為動態與實測用。

(二)模擬績效

依據上一小節訓練與收斂好的模型，並利用在 3.2 節所描述的 4 種偏好展示其績效。偏好值以百分比進行配比，且為了使 AI 強化學習模型對於獎勵目標的權重較為敏感，以及同時均等注重通過量及延滯時間，本研究以 8:2 為主要測試比例，例如：偏好通過量會以表 3.2.1-2 一樣的格式[0.1,0.4,0.1,0.4]對模型與流量進行配比與計算預測值，其獎勵目標的權重如下表 3.2.3-1，為了方便比較績效，以下績效呈現均以固定時制之比較百分比呈現。

表 3.2.3-1 臺北市實驗場域場域獎勵目標測試績效權重表

	幹道延滯時間與 固定時制的變化 量的權重比	支道延滯時間與 固定時制的變化 量的權重比	幹道通過量與固 定時制的變化量 的權重比	支道通過量與固 定時制的變化量 的權重比
偏好延滯	0.4	0.4	0.1	0.1
偏好通過量	0.1	0.1	0.4	0.4
偏好幹道	0.4	0.1	0.4	0.1
偏好支道	0.1	0.4	0.1	0.4
平均	0.25	0.25	0.25	0.25

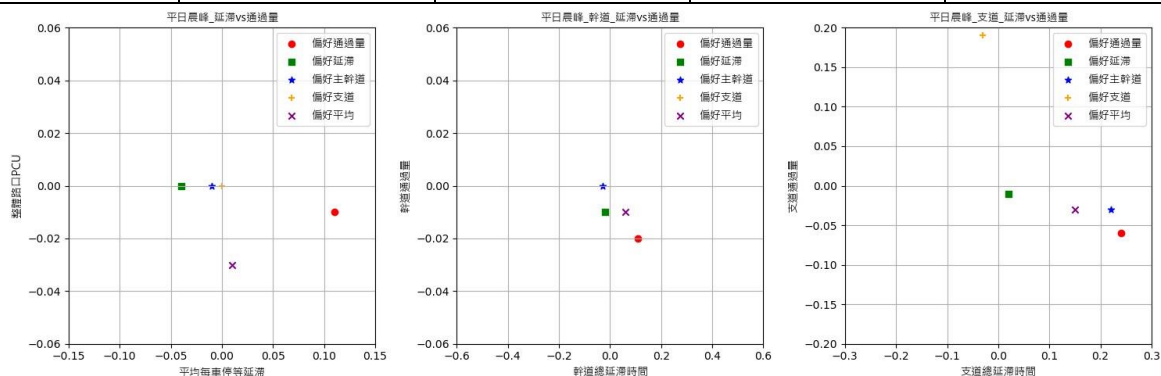


圖 3.2.3-2 臺北市實驗場域平日晨峰各偏好績效彙整圖

表 3.2.3-2 平日晨峰偏好績效彙整表

平日晨峰 (7:00-9:00)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	幹道每小 時單一車 道停等延 滯時間	幹道每小 時單一車 道通過量	支道每小 時單一車 道停等延 滯時間	支道每小 時單一車 道通過量
偏好通過量	+11%	-1%	+11%	-2%	+24%	-6%
偏好延滯	-4%	0%	-2%	-1%	+2%	-1%
偏好主幹道	-1%	0%	-3%	0%	+22%	-3%
偏好支道	0%	0%	+48%	-18%	-3%	+19%
偏好平均	1%	-3%	+6%	-1%	+15%	-3%

平日晨峰的表現為偏好延滯為最佳，如表 3.2.3-2 與圖 3.2.3-2 所示，可以發現此時段與此偏好配置除了支道停等延滯時間外，其餘績效均比固定時制好。另外，可看出偏好幹道的績效在幹道的通過量與總延滯時間最好，幹道延滯降低最多，幹道通過量雖然沒有比固定時制高，但與固定時制表現差不多；偏好支道的績效則是在支道的通過量與總延滯時間表現最好，支道延滯降低最多，支道通過量則比固定時制高了 19%。

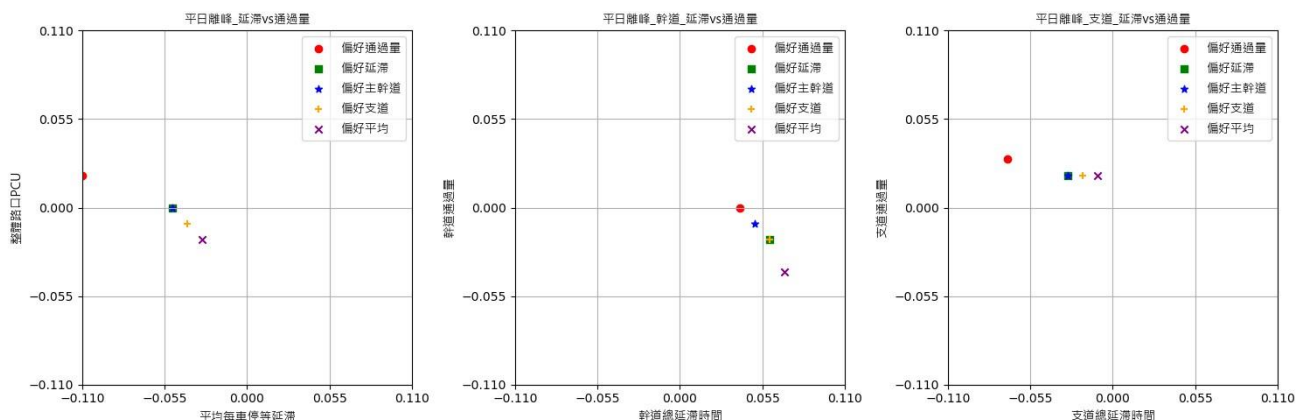


圖 3.2.3-3 臺北市實驗場域平日離峰各偏好績效彙整圖

表 3.2.3-3 平日離峰偏好績效彙整表

平日離峰 (12:30-14:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	幹道每小 時單一車 道停等延 滯時間	幹道每小 時單一車 道通過量	支道每小 時單一車 道停等延 滯時間	支道每小 時單一車 道通過量
偏好通過量	-11%	+2%	+4%	0%	-7%	+3%
偏好延滯	-5%	0%	+6%	-2%	-3%	+2%
偏好主幹道	-5%	0%	+5%	-1%	-3%	+2%
偏好支道	-4%	-1%	+6%	-2%	-2%	+2%
偏好平均	-3%	-2%	+7%	-4%	-1%	+2%

平日離峰的表現為偏好通過量為最佳，如表 3.2.3-3 與圖 3.2.3-3 所示，可以發現此時段與此偏好配置除了幹道停等延滯時間外，其餘績效均比固定時制好。另外，偏好通過量在幹支道與 PCU 的通過量均較其他偏好來的高。

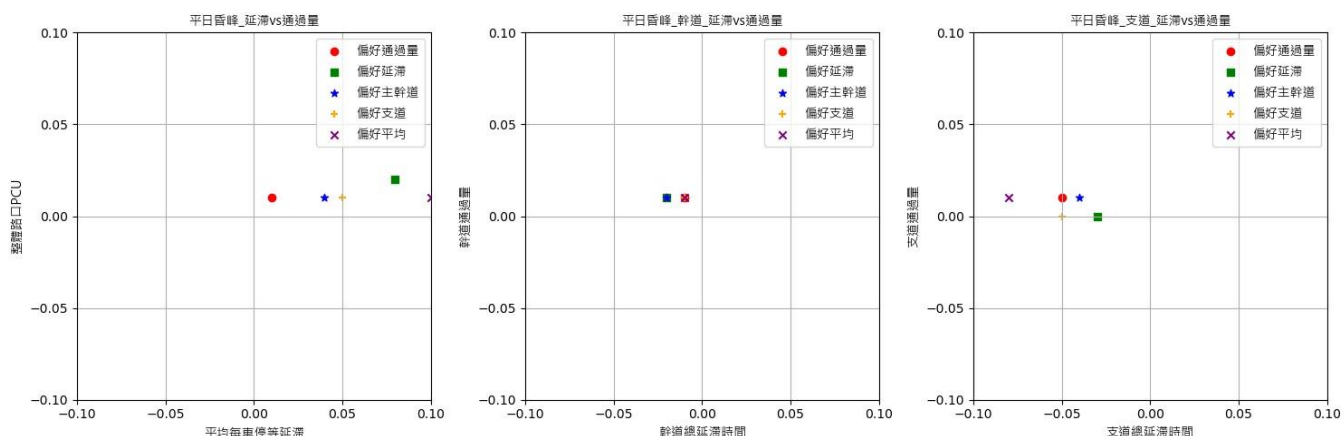


圖 3.2.3-4 臺北市實驗場域平日昏峰各偏好績效彙整圖

表 3.2.3-4 平日昏峰偏好績效彙整表

平日昏峰 (16:30-18:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	幹道每小 時單一車 道停等延 滯時間	幹道每小 時單一車 道通過量	支道每小 時單一車 道停等延 滯時間	支道每小 時單一車 道通過量
偏好通過量	+1%	+1%	-1%	+1%	-5%	+1%
偏好延滯	+8%	+2%	-2%	+1%	-3%	0%
偏好主幹道	+4%	+1%	-2%	+1%	-4%	+1%
偏好支道	+5%	+1%	-1%	+1%	-5%	0%
偏好平均	+10%	+1%	-1%	+1%	-8%	+1%

平日昏峰的表現為偏好通過量為最佳，如表 3.2.3-4 與圖 3.2.3-4 所示，可以發現此時段與此偏好配置除了平均延滯外，其餘績效均比固定時制好。

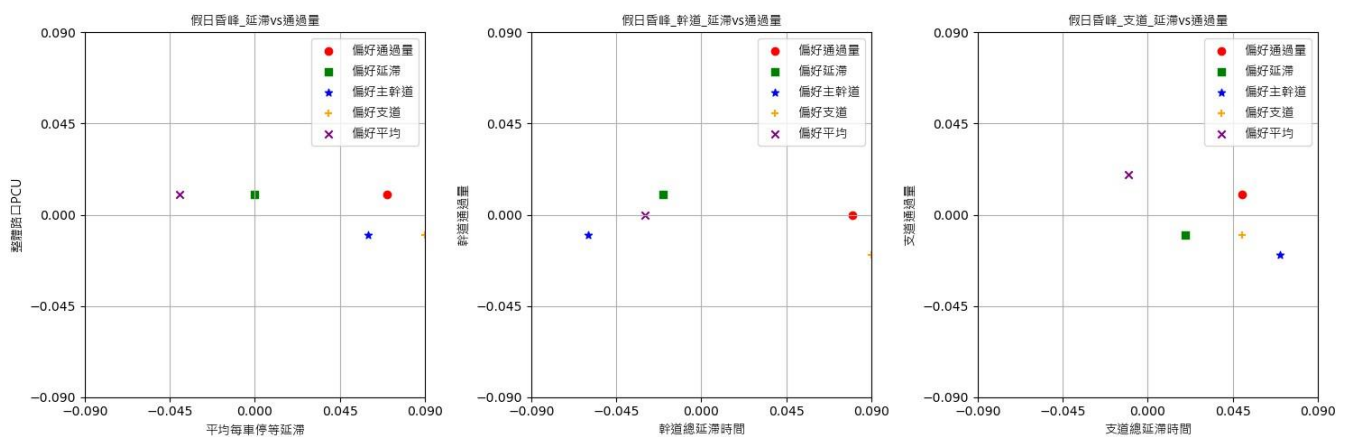


圖 3.2.3-5 臺北假日昏峰各偏好績效彙整圖

表 3.2.3-5 假日昏峰偏好績效彙整表

假日昏峰 (16:30-18:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	幹道每小 時單一車 道停等延 滯時間	幹道每小 時單一車 道通過量	支道每小 時單一車 道停等延 滯時間	支道每小 時單一車 道通過量
偏好通過量	+7%	+1%	+8%	0%	+5%	+1%
偏好延滯	0%	+1%	-2%	+1%	+2%	-1%
偏好主幹道	+6%	-1%	-6%	-1%	+7%	-2%
偏好支道	+9%	-1%	+9%	-2%	+5%	-1%
偏好平均	-4%	+1%	-3%	0%	-1%	+2%

假日昏峰的表現為偏好平均為最佳，如表 3.2.3-5 與圖 3.2.3-5 所示，可以發現此時段與此偏好配置除幹道通過量外，其餘績效均比固定時制好。另外，偏好幹道在幹道的通過量與幹道的延滯均較其他偏好來的高。

二、臺南市實驗場域訓練結果

(一)AI 強化學習模型訓練收斂過程

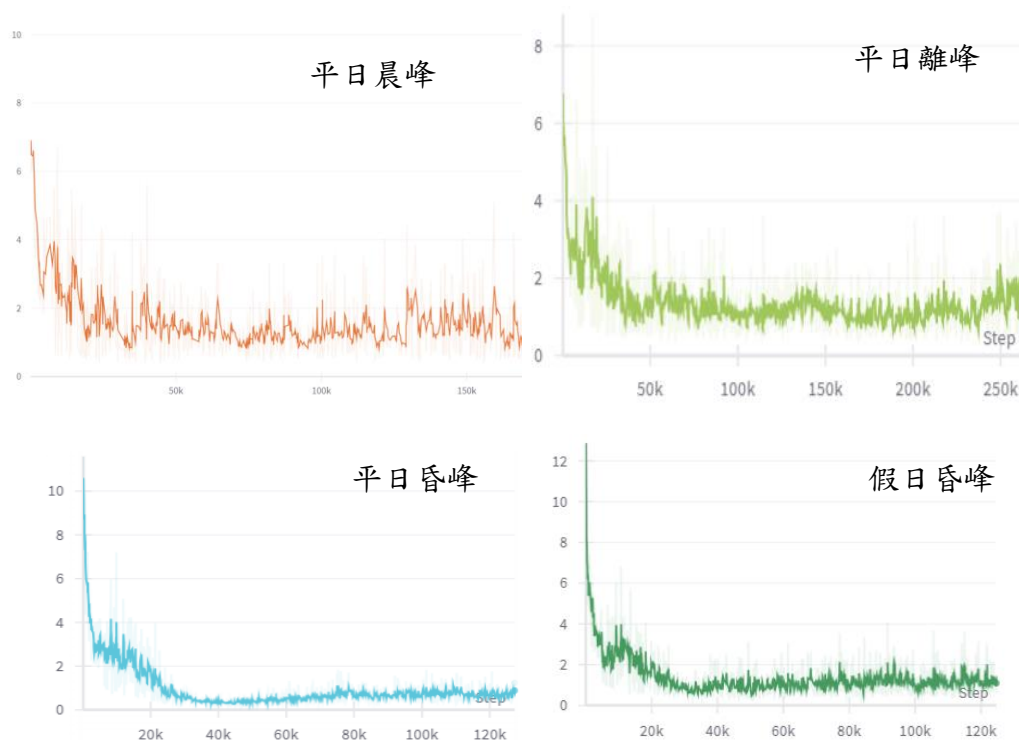


圖 3.2.3-6 臺南市實驗場域全時段訓練指標

臺南市實驗場域與臺北市實驗場域相同都有收斂的趨勢，如圖 3.2.3-6 所示，y 軸代表 Loss，x 軸代表訓練的步數。但不同之處在於臺南市實驗場域訓練是以時相轉換時做為更新步數收斂步數大約為 200 萬步。而模型的使用方針與臺北市實驗場域相同，利用上圖與模擬績效決定是否使用該模型。

(二)模擬績效

依據上一小節訓練與收斂好的模型，並利用在 3.2 節所描述的四種偏好展示其績效。偏好值會以百分比進行配比，且會以 8:2 為主要測試比例，例如偏好通過量會以表 3.2.1-3 一樣的格式[0.1,0.4,0.1,0.4]對模型與流量進行配比與計算預測值，其獎勵目標的權重如下表 3.2.3-6，因平日晨峰時制組態分成兩套，故晨峰模擬績效以兩個時段呈現。

表 3.2.3-6 臺南市實驗場域獎勵目標測試績效權重表

	台 86 延滯時間 與固定時制的變 化量的權重比	台 19 甲延滯時 間與固定時制的 變化量的權重比	台 86 通過量與 固定時制的變化 量的權重比	台 19 甲通過量 與固定時制的變 化量的權重比
偏好延滯	0.4	0.4	0.1	0.1
偏好通過量	0.1	0.1	0.4	0.4
偏好幹道	0.4	0.1	0.4	0.1
偏好支道	0.1	0.4	0.1	0.4
平均	0.25	0.25	0.25	0.25

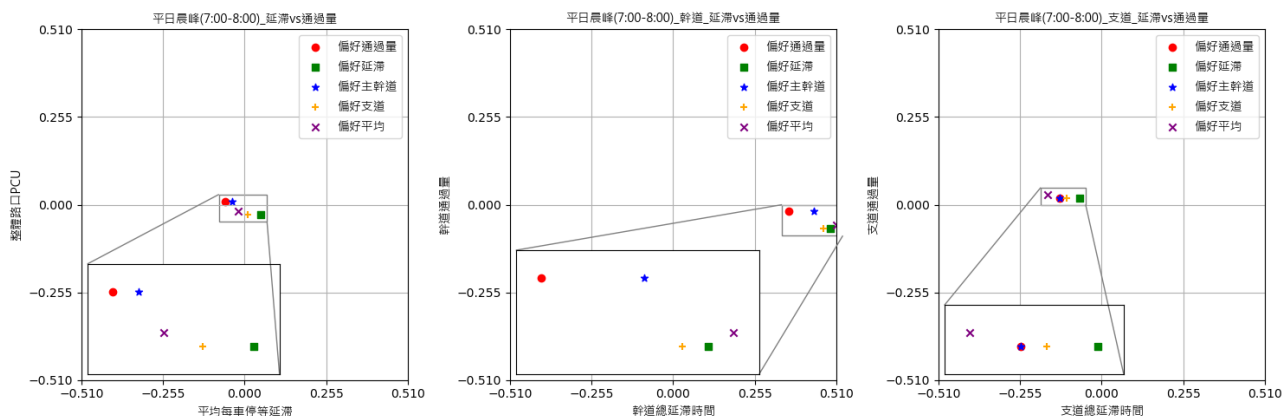


圖 3.2.3-7 臺南市實驗場域平日晨峰(7:00-8:00)各偏好績效彙整圖

表 3.2.3-7 平日晨峰(7:00-8:00)時段一偏好績效彙整表

平日晨峰 (7:00-8:00)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	-6%	+1%	+36%	-2%	-13%	+2%
偏好延滯	+5%	-3%	+49%	-7%	-7%	+2%
偏好主幹道	-4%	+1%	+44%	-2%	-13%	+2%
偏好支道	+1%	-3%	+47%	-7%	-11%	+2%
偏好平均	-2%	-2%	+51%	-6%	-17%	+3%

平日晨峰時段一的表現為偏好通過量為最佳，如表 3.2.3-7 與圖 3.2.3-7 所示，可以發現此偏好使整體路口的停等延滯降低 6%與拉高通過量 1%，除了臺 86 的延滯績效較差外，其他的績效均較固定時制好。

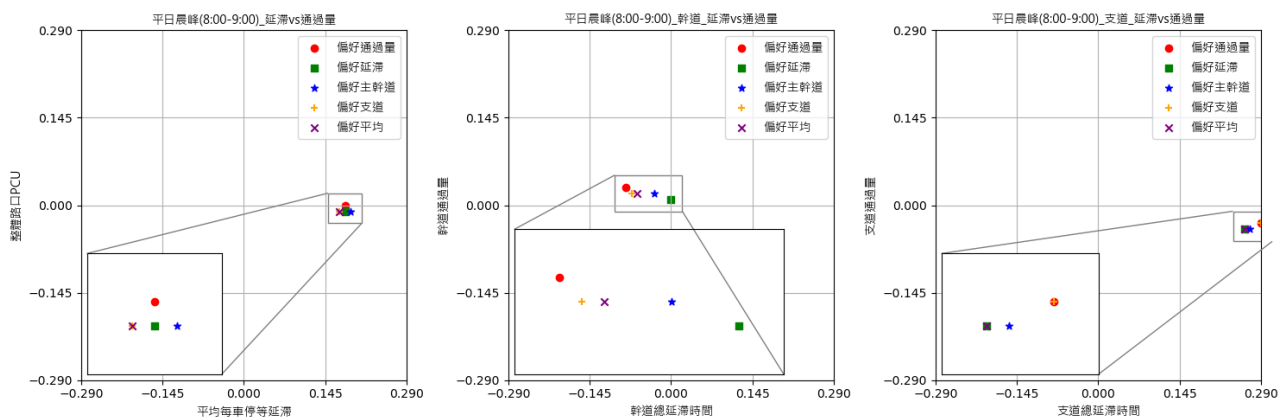


圖 3.2.3-8 臺南市實驗場域平日晨峰(8:00-9:00)各偏好績效彙整圖

表 3.2.3-8 平日晨峰(8:00-9:00)時段二偏好績效彙整表

平日晨峰 (8:00-9:00)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	+18%	0%	-8%	+3%	+29%	-3%
偏好延滯	+18%	-1%	0%	+1%	+26%	-4%
偏好主幹道	+19%	-1%	-3%	+2%	+27%	-4%

偏好支道	+17%	-1%	-7%	+2%	+29%	-3%
偏好平均	+17%	-1%	-6%	+2%	+26%	-4%

平日晨峰時段二的表現為偏好通過量為最佳，如表 3.2.3-8 與圖 3.2.3-8 所示，雖然此偏好使整體路口的停等延滯拉高 18%與通過量沒變化。但台 86(幹道)的延滯時間較固定時制低，同時通過量較固定時制高。

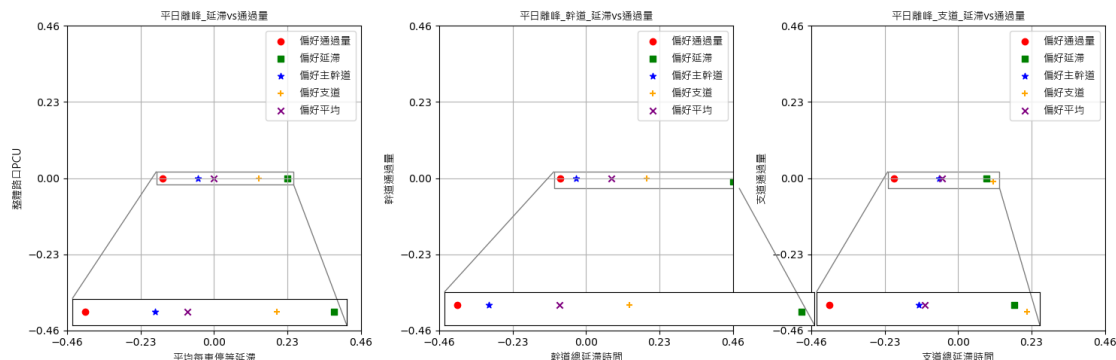


圖 3.2.3-9 臺南市實驗場域平日離峰各偏好績效彙整圖

表 3.2.3-9 平日離峰偏好績效彙整表

平日離峰 (12:30-14:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	-16%	0%	-8%	0%	-20%	0%
偏好延滯	+23%	0%	+46%	-1%	+9%	0%
偏好主幹道	-5%	0%	-3%	0%	-6%	0%
偏好支道	+14%	0%	+19%	0%	+11%	-1%
偏好平均	0%	0%	+8%	0%	-5%	0%

平日離峰的表現為偏好通過量為最佳，如表 3.2.3-9 與圖 3.2.3-9 所示，整體路口的平均停等延滯相較於固定時制降低了 16%，同時台 86 與台 19 甲的停等延滯時間都較固定時制低。偏好延滯與支道的延滯時間均遜於固定時制，推測原因為離峰車流量較低，且本模型為避免實測時出現過大、過小的綠燈秒數，加上時相之上下限，此部分亦會影響延滯的表現，未來如以績效為目標，可嘗試將上下限去除進行訓練，再測試其績效。

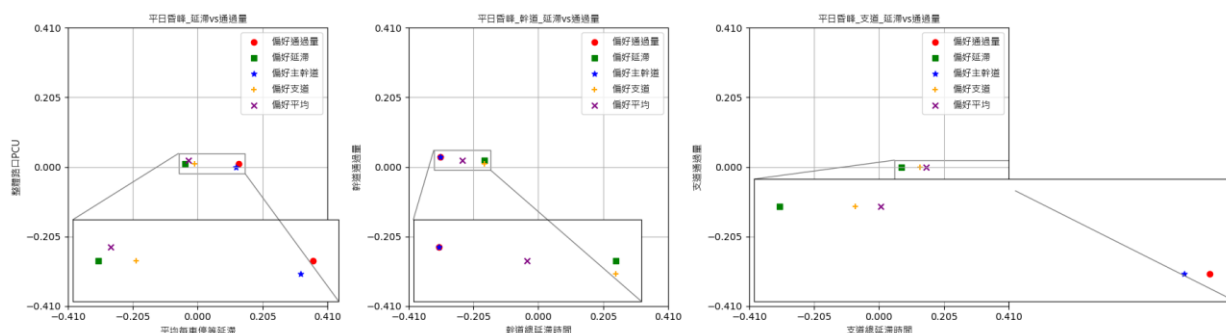


圖 3.2.3-10 臺南市實驗場域平日昏峰各偏好績效彙整圖

表 3.2.3-10 平日昏峰偏好績效彙整表

平日昏峰 (16:30-18:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	+13%	+1%	-31%	+3%	+41%	-5%
偏好延滯	-4%	+1%	-17%	+2%	+7%	0%
偏好主幹道	+12%	0%	-31%	+3%	+39%	-5%
偏好支道	-1%	+1%	-17%	+1%	+13%	0%
偏好平均	-3%	+2%	-24%	+2%	+15%	0%

平日昏峰的表現為偏好延滯為最佳，如表 3.2.3-10 與圖 3.2.3-10 所示，除了台 19 甲停等延滯時間較固定時制差外，其他的績效均比固定時制好。另外，偏好幹道的績效，在幹道的停等延滯與通過量表現最好。

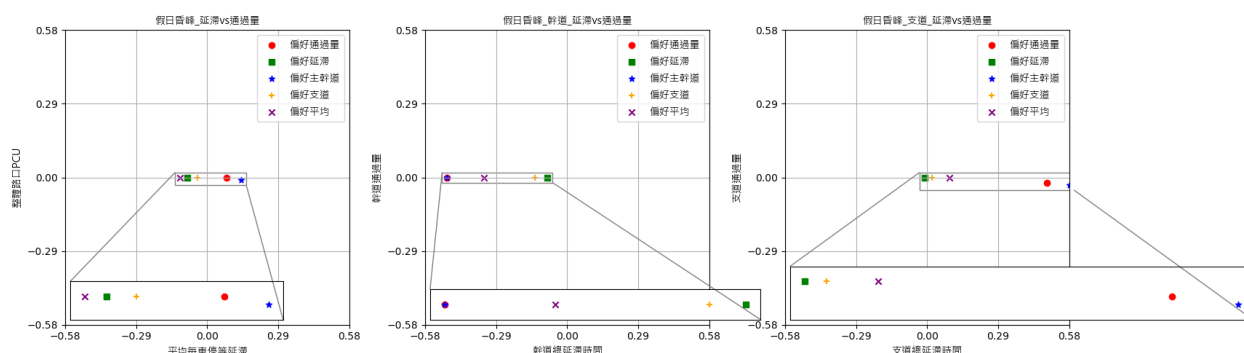


圖 3.2.3-11 臺南市實驗場域假日昏峰各偏好績效彙整圖

表 3.2.3-11 假日昏峰偏好績效彙整表

假日昏峰 (16:30-18:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	+8%	0%	-49%	0%	+49%	-2%
偏好延滯	-8%	0%	-8%	0%	-1%	0%
偏好主幹道	+14%	-1%	-49%	0%	+58%	-3%
偏好支道	-4%	0%	-13%	0%	+2%	0%
偏好平均	-11%	0%	-34%	0%	+9%	0%

假日昏峰的表現為偏好延滯為最佳，如表 3.2.3-11 與圖 3.2.3-11 所示，雖然通過量與固定時制差不多，但延滯績效均比固定時制好。另外，偏好幹道的績效，在幹道的停等延滯與通過量表現最好；偏好支道的績效在支道的停等延滯與通過量表現上，除了略遜於偏好延滯外，均優於其他偏好。

三、多任務強化學習模型

多任務強化學習模型的架構已在 3.2.1 節第六點說明，此節將呈現其平日昏峰與離峰之模擬績效。

(一)AI 強化學習模型訓練過程



圖 3.2.3-12 臺南市實驗場域多任務 AI 強化學習模型平日離峰驗證紀錄

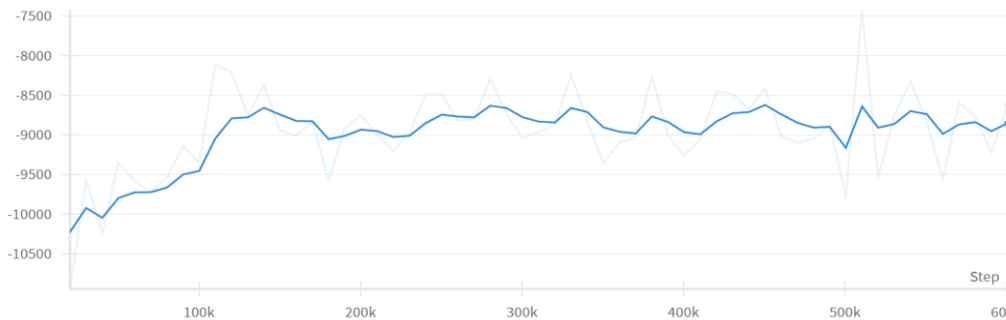


圖 3.2.3-13 臺南市實驗場域多任務 AI 強化學習模型平日昏峰驗證紀錄

由於各個 Loss 都受到同一個 Embedding Network 的影響，導致 Loss 收斂效果不夠顯著。因此，本研究選擇利用訓練過程中的驗證記錄來呈現 Task-Aware 的訓練過程，見圖 3.2.3-12 和圖 3.2.3-13。在圖中，y 軸表示訓練過程中綜合了獎勵目標與偏好權重的加權計算值，而 x 軸代表訓練步數。這裡所謂的加權值是指每次訓練中 Critic 更新的 Q 值，會根據我們所訂定的績效權重表(見表 3.2.3-1)進行評估，以確定該 Q 值在各種獎勵偏好下的好壞程度(當值為正且 y 軸持續增加時表示效果較佳，反之則較差)。然後，利用這些評估結果來更新策略。換句話說，這裡的加權值反映了 Q 值在各種獎勵偏好下的好壞程度。而圖 3.2.3-12 與 3.2.3-13 則是為此提供的驗證紀錄表。由此可見，驗證曲線在每次訓練後(X 值增加)，y 軸的驗證值有隨著訓練過程中逐步進步，負值越來越接近零，離峰時段有達到正值，負值代表此驗證值還未超越固定時制績效，正值則代表有超越固定時制績效。

(二)模擬績效

依據上一小節訓練與收斂好的模型，並利用在 3.2 節所描述的四種偏好展示其績效。偏好配比與前一節相同，以下呈現其結果。

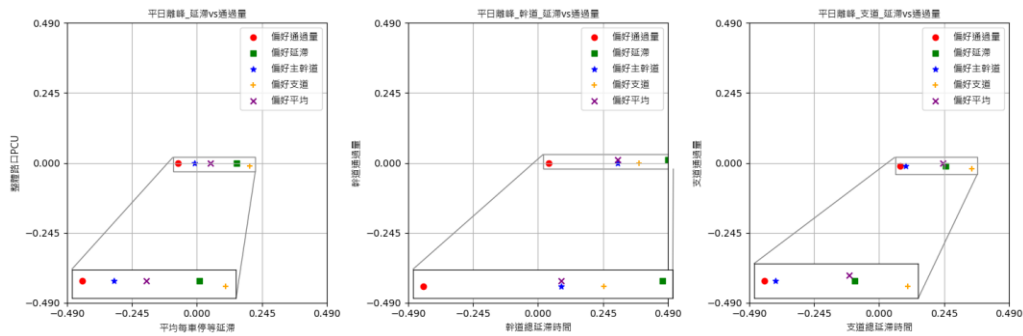


圖 3.2.3-14 臺南市實驗場域多任務 AI 強化學習模型平日離峰各偏好績效彙整圖

表 3.2.3-12 多任務 AI 強化學習模型平日離峰偏好績效彙整表

平日離峰 (12:30-14:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	-7%	0%	+4%	0%	+8%	-1%
偏好延滯	+15%	0%	+49%	+1%	+25%	-1%
偏好主幹道	-1%	0%	+30%	0%	+10%	-1%
偏好支道	+20%	-1%	+38%	0%	+35%	-2%
偏好平均	+5%	0%	+30%	+1%	+24%	0%

在平日離峰的環境中，可以發現多任務 AI 強化學習模型可以做到與固定時制相近的通過量，但在台 86 與台 19 甲的停等延滯時間上還有不少的改良空間，推測是模型為了使平日昏峰的表現上升做出的妥協，不過整體的表現基本上與只有訓練平日離峰單一模型的績效相近。

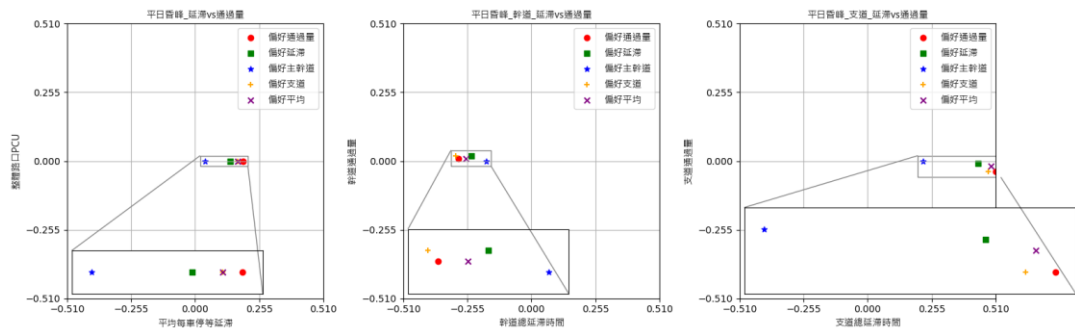


圖 3.2.3-15 臺南市實驗場域多任務 AI 強化學習模型平日昏峰各偏好績效彙整圖

表 3.2.3-13 多任務 AI 強化學習模型平日昏峰偏好績效彙整表

平日昏峰 (16:30-18:30)	單一方向 平均每車 停等延滯	每小時整 體路口 PCU	台 86 每小 時單一車 道停等延 滯時間	台 86 每小 時單一車 道通過量	台 19 甲每 小時單一 車道停等 延滯時間	台 19 甲每 小時單一 車道通過 量
偏好通過量	+19%	0%	-29%	+1%	+51%	-4%
偏好延滯	+14%	0%	-24%	+2%	+44%	-1%
偏好主幹道	+4%	0%	-18%	+0%	+22%	0%
偏好支道	+17%	0%	-30%	+2%	+48%	-4%
偏好平均	+17%	0%	-26%	+1%	+49%	-2%

多任務 AI 強化學習模型在平日昏峰的表現也與平日昏峰單一模型的績效接近，除了在台 19 甲的停等延滯時間較不理想以外，其他的表現基本上都是與固定時制相近，甚至於優於固定時制的績效的。

四、小結

如表 3.2.3-1 到 3.2.3-11 所示，可以發現臺南市實驗場域及臺北市實驗場域 AI 強化學習模型依據不同偏好進行決策，各個時段中均能找到一組偏好在大部分的績效比固定時制佳；另多任務 AI 強化學習模型在不同時段與時制組態下一起訓練的情形中具有一定的績效，雖然無法比分開訓練的模型績效好，但整體表現已接近。後續將挑選模擬績效最佳的 AI 強化學習模型於現場實測。關於其他績效分析的說明，在後續章節實測模擬進行說明。

第四章 強化學習號誌控制實測

4.1 實驗場域規劃與分析

4.1.1 單一路口場域(臺南市「台 86-19 甲線」路口)

本計畫前期計畫(111 年「我國人工智慧車聯網之號誌控制模型探討」計畫)回顧與彙整國內外應用車聯網與人工智慧強化學習在號誌控制之應用與演算法，發展車聯網與人工智慧強化學習之號誌控制模擬環境，藉由車聯網緊急車輛優先號誌、單一路口多時相號誌控制與幹道連鎖號誌控制模擬之人工智慧號誌控制模型構建、學習與訓練、模擬與績效評估，來為後續人工智慧號誌控制奠下研發基礎。

由於前期計畫完成訓練後的 RL 號誌控制模型在不同場域間無法移轉應用，成果僅適用於原設計之實驗場域，臺南市「台 86-19 甲線」單一路口、臺北市「中山北路-德行東路」3 個路口幹道、高雄市「台 88 鳳山出口」(過埤路-鳳頂路) 單一路口。而本計畫考慮計畫時程，以及對路口的掌握程度等因素，選擇以臺南市「台 86-19 甲線」單一路口、臺北市「中山北路-德行東路」作為前期計畫研究成果之實驗場域，以下分別介紹該兩個實驗場域。



圖 4.1.1-1 臺南市「台 86-19 甲線」路口

前期計畫於 111 年 5 月至 6 月期間進行交通調查，調查項目包括：現況道路特性、現行時制計畫、轉向交通量、停等延滯四個項目，以下說明調查之分析結果。

一、道路特性

調查範圍道路包含台 86、19 甲、關廟交流道。其道路幾何特性資料彙整如表 4.1.1-1，包含了車道數、車道寬、分隔方式等。

表 4.1.1-1 臺南市「台 86-19 甲線」路口道路幾何特性

路名	道路類型	車道數	車道寬	分隔方式	車種管制
19 甲 北向	一般 道路	2+左轉袋+ 右轉槽化	4.0(直外)、3.5(直內)、3.3(左)、5.5(右)	實體分隔	內側禁行 機車
19 甲 南向	一般 道路	2+左轉袋+ 右轉槽化	4.4(直外)、3.5(直內)、3.3(左)、5.5(右)	實體分隔	內側禁行 機車
台 86	一般 道路	2+2 左轉袋+ 右轉槽化	3.3(直外)、3.3(直內)、3.3(左外)、5.5(右)	實體分隔	內側禁行 機車
關廟 交流道	快速 道路	2+2 左轉袋+ 右轉槽化	3.0(直外)、3.0(直內)、3.1(左外)、 3.1(左內)、5.0(右)	實體分隔	內側禁行 機車

資料來源：我國人工智慧車聯網之號誌控制模式探討(民國 111 年)

二、號誌時制計畫

台 86-19 甲線路口時相組態與時制計畫內容詳表 4.1.1-2 所示。

表 4.1.1-2 臺南市「台 86-19 甲線」路口平、假日號誌時制計畫

時制	週期	時相1			時相2			時相3			時相4			時相5		
		G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR
0710 0800																
		16	6	0	86	5	3	15	5	3	30	5	3	35	5	3
21	220	16	6	0	86	5	3	15	5	3	30	5	3	35	5	3
時制	週期															
0800~0900 、 1230~1430		16	6	0	86	5	3	15	5	3	30	5	3			
01	160	66	5	3	22	5	3	25	5	3	15	5	3			
時制	週期															
1600 1850		10	6	0	106	5	3	18	5	3	30	5	3	38	5	3
06	240	10	6	0	106	5	3	18	5	3	30	5	3	38	5	3

資料來源：本計畫整理

三、尖峰小時轉向交通量

平日上午尖峰小時發生在 07:00~08:00，主要流量為台 86 線往東；平日下午尖峰小時則發生於 17:00~18:00，主要流量為台 86 線往西，十分具有方向性；平日離峰的尖峰小時則發生於 14:00~15:00；假日尖峰小時則發生於 16:45~17:45。基於平日上下午尖峰之方向性明顯特性，目前號誌時制計畫在上午尖峰時，針對台 86 線往東有設計一個左直右全開的時相 1；而在下午尖峰時，對台 86 線往西有設計一個左直右全開的時相 1。而從平日離峰跟假日尖峰的

小時轉向量資料來看，幹道台 86 線的流量都明顯比支道台 19 甲線來得大，幹支道流量差異較平日上下午尖峰顯著。

四、尖峰小時延滯

觀察臺南市「台 86-19 甲線」路口的延滯與服務水準資料，發現幹道台 86 線的服務水準皆較台 19 甲線來得好，尤其在平日上下午尖峰與假日尖峰時段，幹道服務水準皆在 B~C 級左右，但支道的服務水準只有 E~F 級。

表 4.1.1-3 臺南市「台 86-19 甲線」路口各時段延滯與服務水準

時段	方向		各方向延滯	服務水準	
平日 晨峰	A	國 3 往西	40.3	C	平均延滯
	B	台 19 甲往北	80.4	F	44.7
	C	台 86 往東	27.3	B	服務水準
	D	台 19 甲往南	69.5	E	C
平日 離峰	A	國 3 往西	37.5	C	平均延滯
	B	台 19 甲往北	80.1	F	45.7
	C	台 86 往東	30.9	C	服務水準
	D	台 19 甲往南	83.8	F	D
平日 昏峰	A	國 3 往西	25.8	B	平均延滯
	B	台 19 甲往北	84.5	F	57.5
	C	台 86 往東	43.7	C	服務水準
	D	台 19 甲往南	160.0	F	D
假日 昏峰	A	國 3 往西	23.4	B	平均延滯
	B	台 19 甲往北	91.8	F	47.4
	C	台 86 往東	43.3	C	服務水準
	D	台 19 甲往南	107.4	F	D

資料來源：我國人工智慧車聯網之號誌控制模式探討(民國 111 年)

4.1.2 幹道多路口場域(臺北市「中山北路-德行東路」)

前期計畫於 111 年 6 月至 7 月期間進行交通調查，調查項目包括：現況道路特性、現行時制計畫、轉向交通量、停等延滯四個項目，以下說明調查之分析結果。

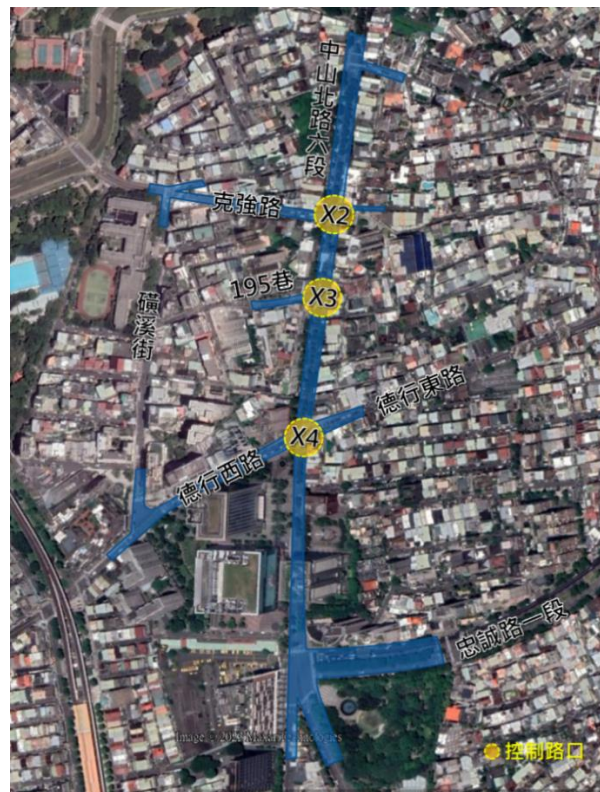
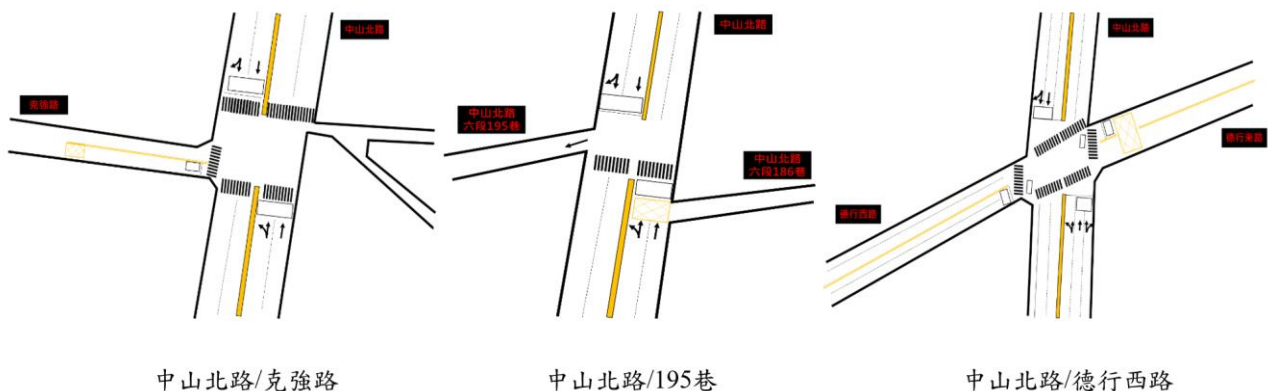


圖 4.1.2-1 臺北市「中山北路-德行東路」幹道

一、道路特性

調查範圍道路包含中山北路、德行東西路、195 巷及克強路，其道路幾何特性資料彙整如表 4.1.2-1(包含車道數、車道寬、分隔方式等)。



資料來源：我國人工智慧車聯網之號誌控制模式探討(民國 111 年)

圖 4.1.2-2 德行西路、195 巷及克強路路口道路幾何示意圖

表 4.1.2-1 德行西路、195 巷及克強路路口道路幾何特性彙整表

路名	道路類型	車道數	車道寬	分隔方式	車種管制
路口 2 中山北路北向	一般道路	2	5.0(直外)、2.8(直內)	實體分隔	混合車道
路口 2 中山北路南向	一般道路	2	4.0(直外)、2.8(直內)	實體分隔	混合車道
路口 2 克強路	一般道路	1	3.2	雙黃線	混合車道
路口 3 中山北路北向	一般道路	2	5.4(直外)、2.8(直內)	實體分隔	混合車道
路口 3 中山北路南向	一般道路	2	4.2(直外)、2.8(直內)	實體分隔	混合車道
路口 4 中山北路北向	一般道路	3	3.0(直右)、3.0(直) 、3.0(直左)	實體分隔	混合車道
路口 4 中山北路南向	一般道路	3	3.0(直右)、3.0(直) 、3.0(直左)	實體分隔	混合車道
路口 4 德行東路	一般道路	1	4.5	雙黃線	混合車道
路口 4 德行西路	一般道路	2	3.0(直左)、3.5(直右)	雙黃線	混合車道

資料來源：我國人工智慧車聯網之號誌控制模式探討(民國 111 年)

二、號誌時制計畫

整理臺北市中山北路六段/克強路、中山北路六段/195 巷、中山北路六段/德行西路等路口時制計畫與時相組態，如表 4.1.2-2 所示。

表 4.1.2-2 德行西路、195 巷及克強路路口號誌時制計畫

路口		時相1			時相2			時相3			時相4			時相5			時相6		
		G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR
中山北路 /克強路																			
平日晨峰	200	9	0	0	104	3	2	5	3	2	9	0	0	14	0	0	44	3	2
平日離峰	150	9	0	0	66	3	2	5	3	2	9	0	0	14	0	0	32	3	2
平日昏峰	200	8	0	0	96	3	2	9	3	2	9	0	0	14	0	0	49	3	2
假日昏峰	200	9	0	0	104	3	2	5	3	2	9	0	0	14	0	0	44	3	2
中山北路 /195巷																			
平日晨峰	200	135	3	2	20	3	2	30	3	2									
平日離峰	150	95	3	2	10	3	2	30	3	2									
平日昏峰	200	115	3	2	20	3	2	50	3	2									
假日昏峰	200	135	3	2	20	3	2	30	3	2									
中山北路 /德行西路																			
平日晨峰	200	10	0	0	85	3	2	5	3	2	23	0	0	40	3	2	17	3	2
平日離峰	150	15	0	0	57	3	2	5	3	2	14	0	0	29	3	2	10	3	2
平日昏峰	200	10	0	0	84	3	2	15	0	0	7	0	0	50	3	2	19	3	2
假日昏峰	200	10	0	0	87	3	2	5	3	2	15	0	0	45	3	2	18	3	2

資料來源：本計畫整理

三、尖峰小時轉向交通量

3 處路口在平日尖峰小時發生時段略有不同。平日上午尖峰小時克強路、195 巷及德行西路路口分別發生在 07:15~08:15、08:30~9:30、07:30~08:30；平日下午尖峰小時則分別發生於 17:00~18:00、16:45~17:45、17:00~18:00。而在平日離峰跟假日尖峰時段，其尖峰小時的時段較為相近。平日離峰尖峰小時則皆發生在 13:45~14:45；假日尖峰則分別發生於 17:00~18:00、17:15~18:15、16:45~17:45。臺北場域不論平日尖峰、平日離峰或假日尖峰等時段，其南北向中山北路幹道直行比例皆占了 80%以上。支道克強路主要是左右轉向中山北路幹道方向；195 巷進入單行道；德行東、西路則是直行比例有 60~75%上下，為其較特別之處。

四、尖峰小時延滯

從臺北市「中山北路-德行東路」的延滯與服務水準資料來看，在所有時段下，幹道中山北路南北雙向的服務水準皆有 A~B 級；但在支道方向則大部分只有 D~E 級，少數時段甚至只有 F 級。由此可知目前定時時制計畫內容，對於幹道續進的時差設計效果相當不錯。

表 4.1.2-3 臺北市實驗場域平日上下午尖峰延滯與服務水準

時段	路口	方向		各方向延滯	服務水準	
平日晨峰	X2	B	中山北路往北	9.6	A	14.5 服務水準 A
		C	克強路往東	73.0	F	
		D	中山北路往南	4.1	A	
	X3	B	中山北路往北	2.4	A	2.3 服務水準 A
		D	中山北路往南	2.2	A	
	X4	A	德行東路往西	49.0	D	25.7 服務水準 B
		B	中山北路往北	7.1	A	
		C	德行東路往東	39.2	C	
		D	中山北路往南	23.4	B	
平日離峰	X2	B	中山北路往北	12.5	A	15.1 服務水準 B
		C	克強路往東	46.3	D	
		D	中山北路往南	9.9	A	
	X3	B	中山北路往北	1.4	A	1.8 服務水準 A
		D	中山北路往南	2.1	A	
	X4	A	德行東路往西	86.6	F	38.5 服務水準 C
		B	中山北路往北	22.8	B	

時段	路口	方向		各方向延滯	服務水準	
平日 昏峰		C	德行東路往東	61.8	E	
		D	中山北路往南	22.8	B	
	X2	B	中山北路往北	14.8	A	28.8 服務水準 B
		C	克強路往東	110.5	F	
		D	中山北路往南	11.7	A	
	X3	B	中山北路往北	6.3	A	8.6 服務水準 A
		D	中山北路往南	10.4	A	
	X4	A	德行東路往西	69.7	E	37.0 服務水準 C
		B	中山北路往北	14.2	B	
		C	德行東路往東	66.9	E	
		D	中山北路往南	29.1	B	
假日 昏峰	X2	B	中山北路往北	13.9	A	23.4 服務水準 B
		C	克強路往東	78.5	E	
		D	中山北路往南	15.6	B	
	X3	B	中山北路往北	3.9	A	3.0 服務水準 A
		D	中山北路往南	2.2	A	
	X4	A	德行東路往西	99.1	F	41.0 服務水準 C
		B	中山北路往北	8.9	A	
		C	德行東路往東	98.4	F	
		D	中山北路往南	15.0	B	

資料來源：我國人工智慧車聯網之號誌控制模式探討(民國 111 年)

4.2 人工智慧強化學習號誌控制核心運作模組

本計畫開發人工智慧強化學習號誌控制核心運作模組，連結縣市交通控制中心之現場設備(車輛偵測器(VD)與號誌控制器(TC))與 AI 強化學習模型，將運作績效儲存於資料庫中，提供運作狀態監控畫面，呈現即時路口績效。

一、整體架構

人工智慧強化學習號誌控制系統整體架構如圖 4.2-1 所示，首先取得路口相關設備(如 TC、VD)運作數據，將收集到的資料傳送給 AI 強化學習模型做運算，AI 強化學習模型會即時將計算結果回傳，然後再依照結果調整時制計畫，最後傳送給號誌控制器，即時運作狀態與路口績效可透過實測監控畫面掌握。線下 AI 模組則會定時做訓練，將最新的訓練結果傳送給線上

AI 號誌決策模組使用。

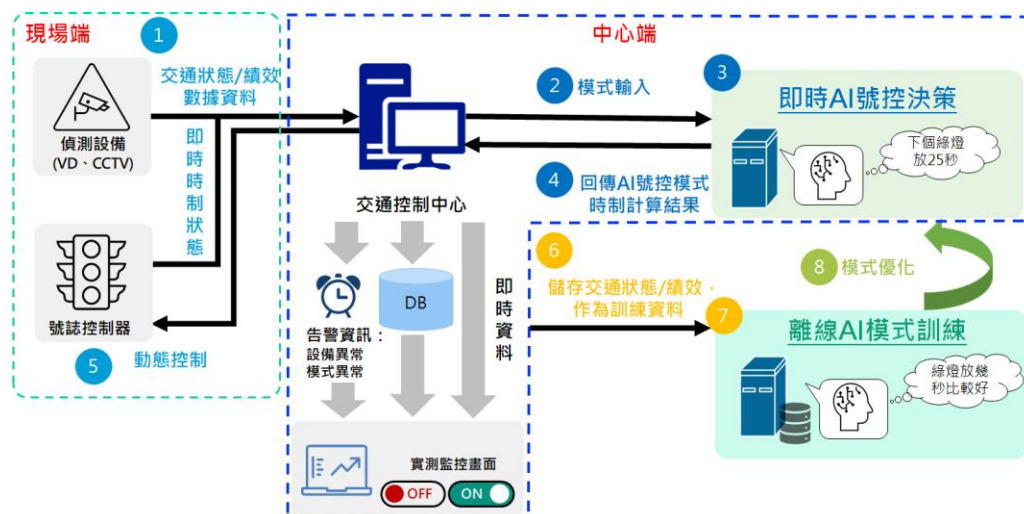


圖 4.2-1 號誌控制核心運作模組架構

系統的執行流程如圖 4.2-2 所示，首先會定時取得設備資料，如車輛偵測器分車道、車種與每秒之車輛數等，或是號誌控制器目前執行的時制計畫、時相及剩餘秒數。下一步驗證資料有無缺漏，當資料有所缺漏時會自動帶入原時制計畫秒數，避免現場因降階運轉之補償機制導致交通壅塞。若資料無缺漏，系統將再確認群組內控制器是否正常連線，緊接查看目前號制分相是否進入清道時間，若進入清道時間則呼叫線上 AI 決策模組，計算下個週期或分相秒數，最後將秒數傳送至控制器執行。

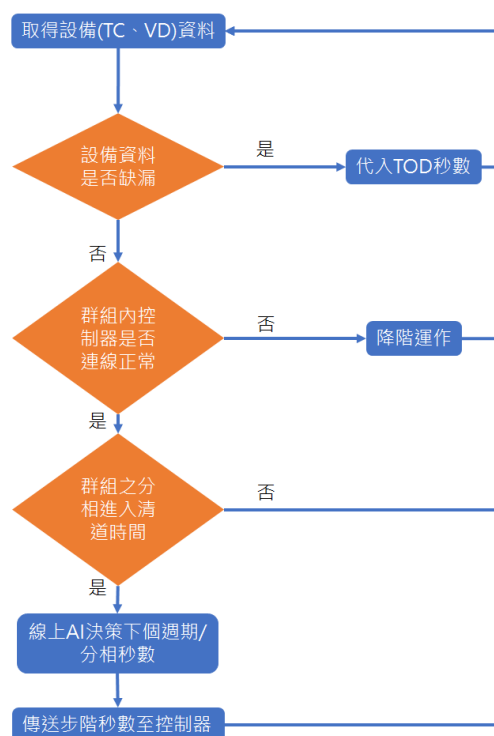


圖 4.2-2 AI 智慧號誌控制系統執行流程

二、軟體功能

設備透過即時串流服務，將資料依照頻率傳送至中心號誌控制系統，而號誌控制系統依照功能將資料傳送至 AI 強化學習模型、資料庫及監控畫面，AI 號誌控制模型計算之最適結果則下達現場號誌控制器，進行動態人工智慧強化學習號誌控制，並可查詢即時燈態與步階資訊，其資料串接架構規劃如下圖 4.2-3 所示。

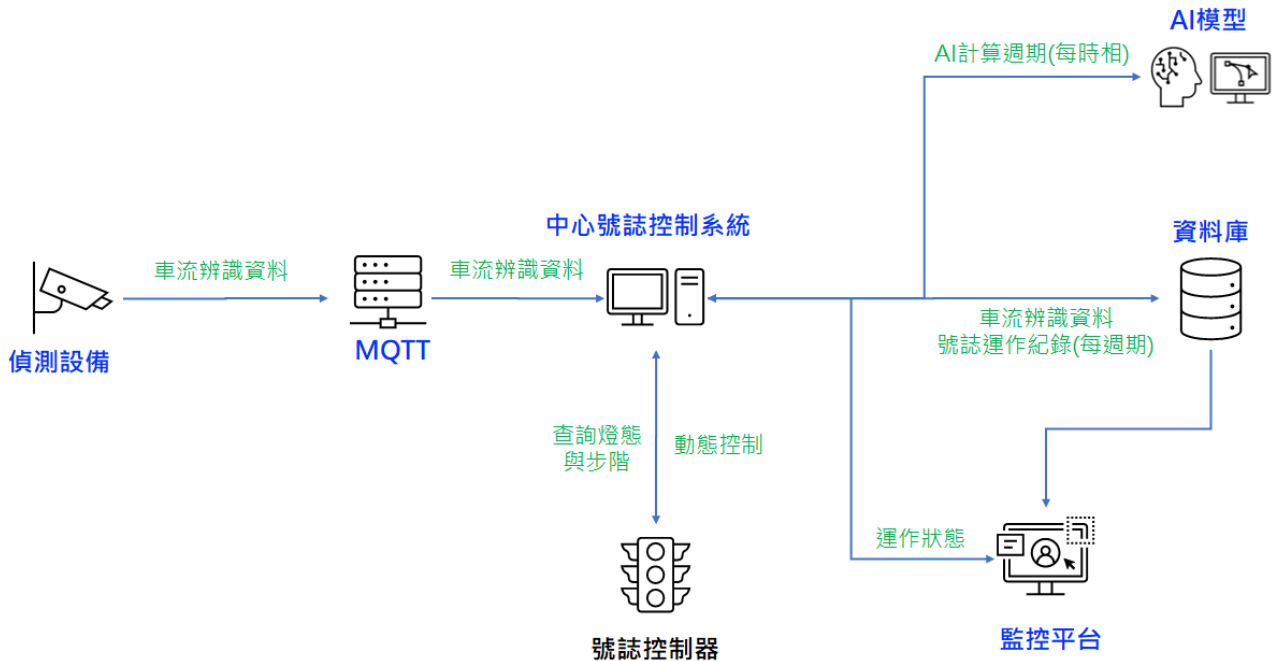


圖 4.2-3 資料串接架構規劃

三、實測監控畫面

實測監控畫面提供 GIS 監控圖台、系統管理、運作績效三大功能，協助使用者掌握實測情形。

(一)GIS 監控圖台

監控圖台分為畫面左側路口運作績效、路口狀態與畫面右側之 GIS 地圖，系統畫面如圖 4.2-4 所示。路口運作績效、路口狀態依車流偵測設備可提供之資訊不同，呈現即時路口延滯、各分相綠燈秒數比較表、綠燈續進帶、近 1 小時路口流量趨勢、轉向量交通量圖資料，使用者可藉由即時資料即時路口流量，並了解 AI 強化學習模型運算之秒數與 TOD 之差異。

GIS 地圖呈現實測路口周邊之號誌控制器、車流偵測設備、CCTV 即時影像點位與路口即時燈號，右下角則顯示設備運作狀態，可一覽有無設備斷線。點選號誌控制器圖示可知路口當下時制計畫、分相圖與現場綠燈秒數，實測中可由現場綠燈秒數得知 AI 計算秒數是否有下達號誌控制器；點選臺北市實驗場域之車流偵測設備圖示可知即

時停等車輛數，臺南市實驗場域由於使用即時影像 AI 辨識技術，故呈現 CCTV 即時影像畫面。

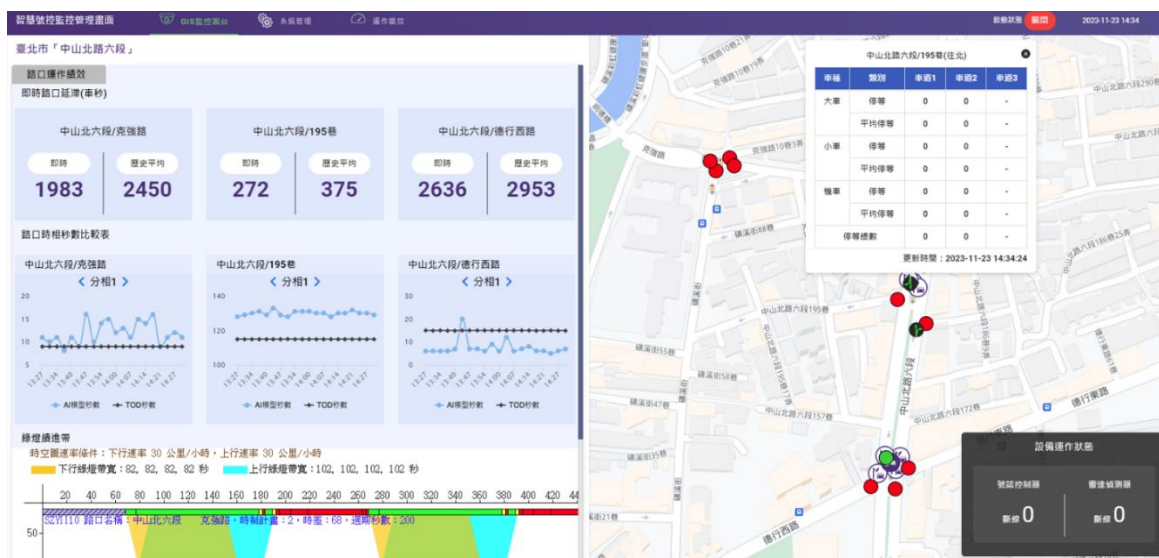


圖 4.2-4 實測監控畫面-GIS 監控圖台

(二)系統管理

使用者可於系統管理頁面進行 AI 號誌控制之排程設定，可依需求單日或每日、每週之循環排程，不須於每次實測安排人員手動開啟、關閉 AI 號誌控制。若實測過程發生意外事件，系統無自動回復 TOD 控制，亦可於此頁面手動操作。

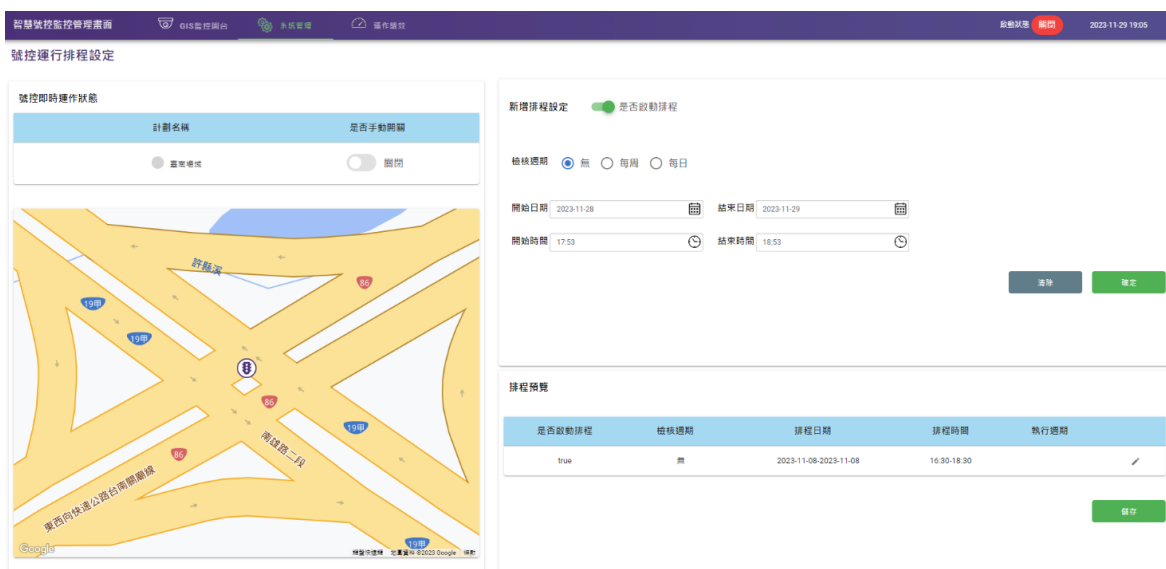


圖 4.2-5 實測監控畫面-系統管理

(三)運作績效

提供智慧號誌控制運行時制計畫、車流偵測資料(停等車流量/影像辨識車流量)、系統運作異常紀錄之資料查詢、下載功能。智慧號誌控制運行時制計畫可查詢號誌控制器現場之週期、分相開始時間、時相編號、分相編號與號誌秒數；車流偵測資料可查

詢路口各方向之車流量；系統運作異常紀錄可查詢號誌控制器暫停原因與暫停之起始、結束時間。

智慧號誌監控管理畫面

智慧號誌運行時制計畫

開始日期: 2023-11-29 19:15 結束日期: 2023-11-29 19:15

路口編號	分相開始時間	總相開始時間	路相編號	分相編號	行綠	行黃	行紅	黃燈	全紅
S281601	2023-11-29 19:14:20	2023-11-29 19:12:03	52	4	14	0	0	5	3
S281601	2023-11-29 19:13:47	2023-11-29 19:12:03	52	3	24	0	0	5	3
S281601	2023-11-29 19:13:16	2023-11-29 19:12:03	52	2	21	0	0	5	3
S281601	2023-11-29 19:12:03	2023-11-29 19:12:03	52	1	65	0	0	5	3
S281601	2023-11-29 19:11:39	2023-11-29 19:09:23	52	4	15	0	0	5	3
S281601	2023-11-29 19:11:07	2023-11-29 19:09:23	52	3	25	0	0	5	3
S281601	2023-11-29 19:10:37	2023-11-29 19:09:23	52	2	21	0	0	5	3
S281601	2023-11-29 19:09:23	2023-11-29 19:09:23	52	1	65	0	0	5	3
S281601	2023-11-29 19:08:59	2023-11-29 19:06:43	52	4	15	0	0	5	3
S281601	2023-11-29 19:08:27	2023-11-29 19:06:43	52	3	24	0	0	5	3

圖 4.2-6 實測監控畫面-「智慧號誌控制運行時制計畫」查詢頁面

智慧號誌監控管理畫面

影像辨識車流量

開始日期: 2023-11-29 19:16 結束日期: 2023-11-29 19:16

資料時間	路口編號	路口方向	大車車輛加總	小車車輛加總	機車車輛加總
2023-11-29 19:14:00	S281601	往東	0	3	1
2023-11-29 19:14:00	S281601	往西	0	11	1
2023-11-29 19:14:00	S281601	往北	0	5	0
2023-11-29 19:14:00	S281601	往東	1	4	0
2023-11-29 19:13:00	S281601	往東	0	10	6
2023-11-29 19:13:00	S281601	往西	2	2	0
2023-11-29 19:13:00	S281601	往北	1	8	0
2023-11-29 19:13:00	S281601	往東	1	6	0
2023-11-29 19:12:00	S281601	往東	0	1	0
2023-11-29 19:12:00	S281601	往西	2	23	0

圖 4.2-7 實測監控畫面-「車流偵測資料」查詢頁面

智慧號誌監控管理畫面

系統運作異常紀錄

開始日期: 2023-11-03 19:17 結束日期: 2023-11-29 19:17

設備編號	暫停類型	暫停運作原因	暫停開始時間	暫停結束時間
S281601	控制器異常	S281601 TC外部控制，恢復正常。	2023-11-15 10:18:55	
S281601	控制器異常	S281601 因通訊線路暫停運作	2023-11-08 08:00:29	2023-11-08 08:13:24
S281601	控制器異常	S281601 因通訊線路暫停運作	2023-11-08 07:10:04	2023-11-08 07:10:05
S281601	控制器異常	S281601 TC外部控制，恢復正常。	2023-11-07 15:10:17	2023-11-07 15:49:42
S281601	控制器異常	S281601 TC外部控制，恢復正常。	2023-11-07 14:26:28	2023-11-07 14:36:19
S281601	控制器異常	S281601 TC與PLC步調不一致，TC異常結束恢復23秒。	2023-11-07 07:51:44	2023-11-07 07:59:15
S281601	控制器異常	S281601 TC與PLC步調不一致，TC異常結束恢復17秒。	2023-11-07 07:40:18	2023-11-07 07:47:47
S281601	控制器異常	S281601 TC與PLC步調不一致，TC異常結束恢復28秒。	2023-11-07 07:28:51	2023-11-07 07:36:19
S281601	控制器異常	S281601 因通訊線路暫停運作	2023-11-07 07:12:04	2023-11-07 07:12:05
S281601	控制器異常	S281601 TC與PLC步調不一致，TC異常結束恢復25秒。	2023-11-05 18:28:09	2023-11-05 18:27:41

圖 4.2-8 實測監控畫面-「系統運作異常紀錄」查詢頁面

4.3 實測計畫

4.3.1 車輛偵測器驗證計畫

考量本計畫所建置之 AI 號誌控制系統係依據於各路口或各車道所佈建之智慧型偵測器即時偵測資料，經由 AI 運算決策即時調整與控制現場號誌運作，因此偵測設備是否可有效辨識物件並提供精確之偵測值極為重要。以下針對臺南市及臺北市兩個實驗場域使用之車輛偵測器準確度驗證規劃說明如下。

一、單一路口場域(臺南市「台 86-19 甲線」路口)

臺南市「台 86-19 甲線」路口偵測設備如圖 4.3.1-1 所示，路口共架設 6 台即時影像車輛偵測設備，經由 AI 影像辨識取得路口轉向量資料，經確認後，設備位置與偵測範圍皆符合本案需求。



圖 4.3.1-1 臺南市「台 86-19 甲線」偵測設備位置

(一)路口車流轉向辨識

1.驗證說明：以各路口為一計算單位，於平日晨峰(07:00~09:00)與昏峰(16:30~19:30)共 2 時段各錄製 15 分鐘影像(二時段共 30 分鐘)，計算影片時間內車流轉向辨識之準確率。

2.計算方式： $(1 - |((\text{各物件人工計數} - \text{各物件 AI 計數})) / \text{各物件人工計數}|) * 100\%$

3.準確率門檻：90%

4.測試方式：

(1)現場車況利用本計畫偵測路口之攝影機擷取影像，錄影收集車況數據，需提供各路段車流於路口轉向之數據。錄放影系統需具有時間紀錄，至少以

秒為單位，且於播放時可單張播放。

(2)每路段皆可由以上記錄人工計算求得總路口轉向量與各路段轉向量等測試值，並與後端影像辨識系統(iTraView)之實際偵測結果進行準確度驗證。

A.各路段轉向量：利用放影機可步階控制單張播放的特性，以目視觀察物件之轉向量，以判斷為左轉、右轉與直行之轉向。

B.總路口轉向量：上項的加總，並依車流轉向進行分類。

5.應用方式：作為 AI 號誌控制模型輸入參數，以及後續監控畫面使用。

(二)匝道車流辨識

1.驗證說明：以各匝道為一計算單位，於平日晨峰(06:30-08:30)與昏峰(16:30-19:30)共 2 時段各收集 10 分鐘 eTag、VD 資料(二時段共 20 分鐘)，計算時間內各車種物件數之準確率。

2.計算方式： $(1 - ((\text{各車種 VD 偵測量} - \text{各車種 eTag 偵測量}) / \text{各車種 eTag 偵測量})) * 100\%$

3.準確率門檻：90%

4.測試方式：

(1)現場車況利用偵測匝道之 eTag 門架收集車況數據，須提供小型車、大型車、聯結車等物件之數據。數據需具有時間紀錄，至少以秒為單位。

(2)每部車皆可由以上紀錄求得分車種數量之測試值，並與 VD 實際偵測結果進行準確度驗證。

5.應用方式：作為 AI 號誌控制模型輸入參數，以及作為號誌控制績效運算與監控畫面使用。

(三)準確率驗證

1.路口車流轉向辨識

依上述驗證方法，取 7/13 平日晨峰(07:45~08:00)與昏峰(17:45~18:00)2 時段，驗證其 15 分鐘內各車種轉向辨識準確率，整理如表 4.3.1-1~4.3.1-2 所示。晨峰準確率達 97.5%，昏峰準確率達 98.5%，符合準確率門檻。

表 4.3.1-1 路口車流轉向辨識平日晨峰(07:45~08:00)準確度驗證

測試時間	人工計算					AI辨識					準確率				
	大車	小車	機車	大車+小車	大車+小車+機車	大車	小車	機車	大車+小車	大車+小車+機車	大車	小車	機車	大車+小車	大車+小車+機車
每5分鐘統計															
07:45:00 ~ 07:50:00	45	339	137	384	521	42	332	133	374	507	92.4%	97.9%	97.1%	97.3%	97.2%
07:50:00 ~ 07:55:00	41	295	108	336	444	38	288	107	326	433	92.7%	97.6%	99.1%	97.0%	97.5%
07:55:00 ~ 08:00:00	36	292	91	328	419	38	283	89	321	410	94.7%	96.9%	97.8%	97.9%	97.9%
每15分鐘統計															
07:45:00 ~ 08:00:00	122	926	336	1048	1384	118	903	329	1021	1350	96.4%	97.5%	97.9%	97.4%	97.5%

表 4.3.1-2 路口車流轉向辨識平日昏峰(17:45~18:00)準確度驗證

測試時間	人工計算					AI辨識					準確率				
	大車	小車	機車	大車+小車	大車+小車+機車	大車	小車	機車	大車+小車	大車+小車+機車	大車	小車	機車	大車+小車	大車+小車+機車
每5分鐘統計															
17:45:00 ~ 17:50:00	23	340	73	363	436	23	339	65	362	427	100.0%	99.7%	89.0%	99.7%	97.9%
17:50:00 ~ 17:55:00	20	322	110	342	452	24	315	107	339	446	83.3%	97.8%	97.3%	99.1%	98.7%
17:55:00 ~ 18:00:00	18	261	44	279	323	20	259	40	279	319	90.0%	99.2%	90.9%	100.0%	98.8%
每15分鐘統計															
17:45:00 ~ 18:00:00	61	923	227	984	1211	67	913	212	980	1192	91.0%	98.9%	93.4%	99.6%	98.4%

2. 匝道車流辨識

依上述驗證方法，取 5/16 平日晨峰(06:30~07:00)與昏峰(17:00~17:30)2 時段，驗證其 30 分鐘內各車種轉向辨識準確率，整理如表 4.3.1-3 所示。晨峰北上、南下準確率分別為 95.8%及 96.6%，昏峰北上、南下準確率分別為 91.7%及 93.8%，符合準確率門檻。

表 4.3.1-3 匝道車流辨識平日晨、昏峰準確度驗證

測試時間			方向	VD	ETAG	準確率
每30分鐘統計				大車+小車	大車+小車	大車+小車
06:30:00	~	07:00:00	北上	270	259	95.8%
			南下	304	294	96.6%
17:00:00	~	17:30:00	北上	273	252	91.7%
			南下	224	211	93.8%

二、幹道多路口場域(臺北市「中山北路-德行東路」)

臺北市「中山北路-德行東路」偵測設備位置如圖 4.3.1-2 所示，於路段號誌桿上架設車輛偵測器，共計架設 9 台，經執行團隊確認，設備位置與偵測範圍皆符合本案需求。

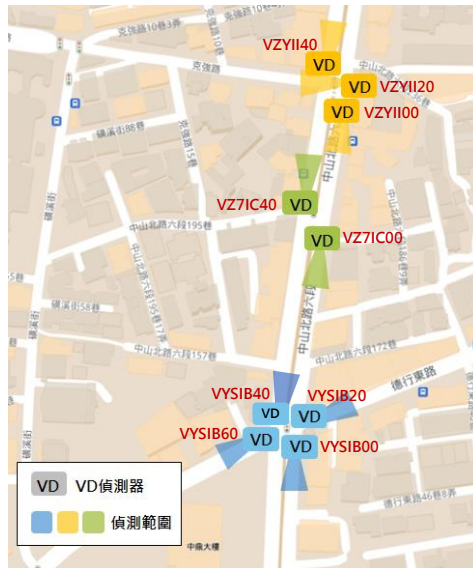


圖 4.3.1-2 臺北市「中山北路-德行東路」偵測設備位置

(一)物件雷達辨識

1.驗證說明：以各路口為一計算單位，於平日晨峰(07:00~10:00)與昏峰(16:30~19:30)共 2 時段各錄製 15 分鐘影像(二時段共 30 分鐘)，計算影片時間內之大車物件數、小車物件數與機車物件數之準確率。

2.計算方式： $(1 - |((\text{各物件人工計數} - \text{各物件雷達偵測量}) / \text{各物件人工計數})|) * 100\%$

3.準確率門檻：90%

4.測試方式：

(1)現場車況利用人工錄影收集偵測路段之車況數據，需提供大車、小車、機車等物件之數據。錄放影系統需具有時間紀錄，至少以秒為單位，且於播放時可單張播放。

(2)每部車皆可由以上記錄求分車種數量測試值，並與 VD 實際偵測結果進行準確度驗證。

5.分車種數量：利用放影機可步階控制單張播放的特性，以目視計算呈現在螢幕上的車輛長度，以判定為大車、小車或機車。

6.應用方式：作為 AI 號誌控制模型輸入參數，以及作為號誌控制績效運算與監控畫面使用。

(二)準確率驗證

依上述驗證方法，取 9/19 平日晨峰與昏峰 2 時段，驗證其 15 分鐘內抽樣 5 分鐘各車種轉向辨識準確率，整理如表 4.3.1-4~4.3.1-5 所示。晨峰準確率達 97.8%，昏峰準確

率達 99.8%，符合準確率門檻。

表 4.3.1-4 VD 平日晨峰準確度驗證

路口	設備編號	方向	計數時段	VD 變化量	人工計數流量	人工_小	人工_中	人工_大	(人工-VD)/人工
中山北路六段 /克強路口	VZYI120	東向	0803-0808	63	58	21	35	2	-9.0%
	VZYI140	南向	0835-0840	194	181	87	93	1	-7.0%
	VZYI100	北向	0815-0820	160	165	60	99	6	3.0%
中山北路六段 /195巷口	VZ71C00	北向	0906-0910	137	136	56	73	7	-1.0%
	VZ71C40	南向	0856-0900	117	115	33	76	6	-2.0%
中山北路六段 /德行西路口	VYSIB20	東向	1015-1020	84	82	43	33	6	-3.0%
	VYSIB40	南向	1035-1040	150	153	39	106	8	2.0%
	VYSIB60	西向	0940-0945	66	67	47	18	2	2.0%
	VYSIB00	北向	0957-1002	232	221	85	134	2	-5.0%
									-2.2%

表 4.3.1-5 VD 平日昏峰準確度驗證

路口	設備編號	方向	計數時段	VD 變化量	人工計數流量	人工_小	人工_中	人工_大	(VD-人工)/VD
中山北路六段 /克強路口	VZYI120	東向	1715-1720	74	72	35	37	0	-3.0%
	VZYI140	南向	1653-1658	112	110	39	66	5	-2.0%
	VZYI100	北向	1730-1735	110	116	46	69	1	5.0%
中山北路六段 /195巷口	VZ71C00	北向	1750-1755	153	153	52	98	3	0.0%
	VZ71C40	南向	無影片						
中山北路六段 /德行西路口	VYSIB20	東向	無影片						
	VYSIB40	南向	1825-1830	110	115	56	56	3	4.0%
	VYSIB60	西向	無資料						
	VYSIB00	北向	1905-1910	197	189	97	89	3	-4.0%
									-0.2%

4.3.2 模擬測試

4.3.2.1 車流模擬軟體

本計畫預計蒐集長時間的交通調查資料，以 4.2 節之 RL 模型進行動態模擬，號誌控制代理人透過與環境的互動，從經驗中學習出最優化的控制策略，惟因現實上不可能允許代理人直接在實際場域探索各種可能的控制策略，因此仍必須建置一套在不影響現實車流下，讓代理人嘗試下控制策略取得績效的環境，而 VISSIM 及 SUMO 兩種交通模擬軟體恰為符合需求的最佳選項。兩車流模擬軟體操作簡單說明如下：

一、VISSIM

為德國 PTV 公司所開發之 VISSIM 專業交通車流模擬器。其特點在於大量可依據不同情境調整之客制化參數，足以相當程度的模擬出各式不同地區的駕駛行為特性。其中也包含我國機車駕駛行為模擬，例如同車道超車(鑽行)、兩段式左轉、停等區等等。模擬與實際畫面對比可參考圖 4.3.2.1-1 及 4.3.2.1-2 所示。

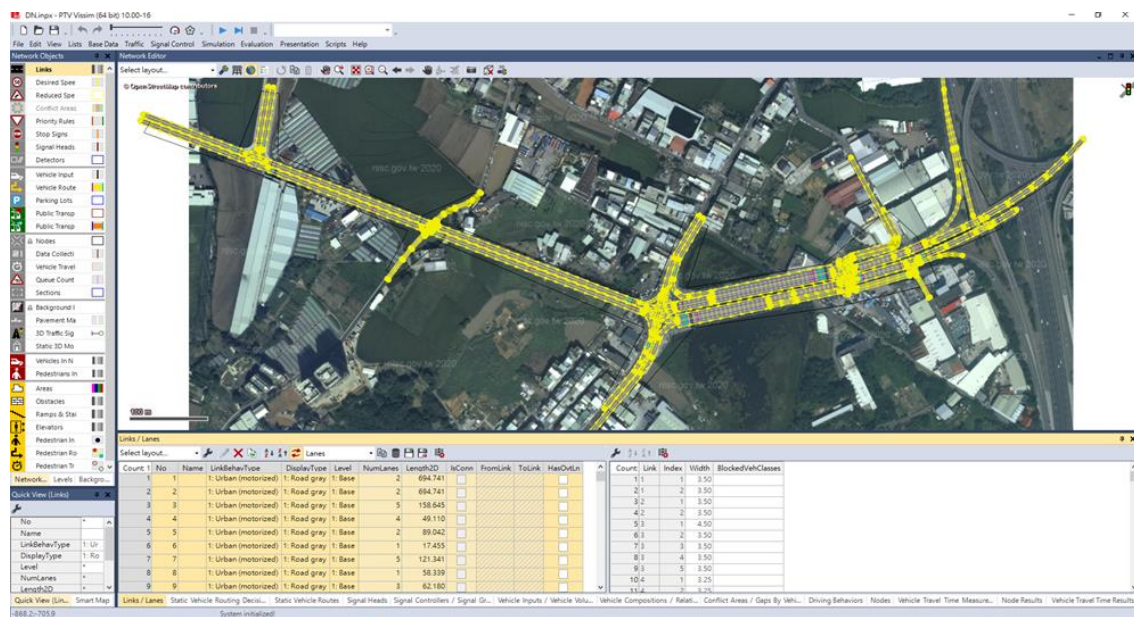


圖 4.3.2.1-1 VISSIM 模擬路網示意圖



圖 4.3.2.1-2 VISSIM 實際路況與模擬影片對比示意圖

VISSIM 車流模擬軟體是目前廣受國際交通業界好評與應用的交通模擬器，除前述豐富的軟體功能與客制化能力之外，對於本案的強化學習號誌控制而言相當重要的一點在於微觀車流軟體相較於定性(Deterministic)的巨觀車流模擬，能夠提供一個隨機性(Stochastic)的交通環境給代理人去做學習。這點將影響代理人的績效表現，因為實際的交通環境必然存在一定程度的隨機性，因此其必須要在類似的環境下探索，才能學到最佳化績效期望值的策略。倘若使用定性的巨觀車流模型讓代理人進行學習，則學出的結果將會是服從定性巨觀車流下的解析最佳解，與實際狀況就會有所差距，可能導致代理人失去原本能因應各種不同車流狀態並即時調整的最大強項。

VISSIM 作為一款微觀車流模擬器，一般的使用情境為所謂「靜態分析」，意即當模擬模型相關參數設定完畢之後，模型便會根據該組參數搭配選定之亂數種子進行模擬分析，在整套模擬過程當中將不再會有任何外部的影響因素干擾模擬過程。然而在動態控制的架構下，靜態分析並無法滿足動態控制的需求。因此 VISSIM 提供了一套能訪問或控制模擬器內所有物件的外部接口，可將數據與指令透過程式語言的指令集實現各式動態控制的模擬，稱為 VISSIM-COM。VISSIM-COM 接口定義了一個階層式物件模型，其中最初由 GUI 提供的模擬器的功能和參數可以通過編譯程式來訪問或操作各項物件，例如可以透過程式自由的控制

紅綠燈的秒數，或是讀取任何定義車道的車輛數等資訊。VISSIM-COM 能夠支援大多數主流的程式語言(例如 C++、Visual Basic、Java 等)。通過 VISSIM-COM，使用者能夠動態地操作大多數內部對象的屬性。

本研究透過 Python script 在 VISSIM API 環境，作為強化學習代理人訓練模組以及 VISSIM 模擬環境間的接口，可在 VISSIM 模擬過程中讀取各樣交通車流狀態以及號誌狀態，並以代理人計算出的時制/相秒數來動態地控制 VISSIM 號誌控制內容。其中亦將常用功能參數化，如最大最小綠燈長度、黃燈秒數、全紅秒數、時相順序等等，以利在擴充路網或調整參數時能夠順利調整。同時，API 環境亦預先定義了數種常用的 KPI 指標，如停等延滯、通過量、旅行時間等等，方便 AI 訓練時有不同的績效可供參考，而在計畫執行的過程，亦可視業主需求新增更多種類的 KPI。

二、SUMO

SUMO(Simulation of Urban MObility)是一個微觀、多運具、空間連續和時間離散的車流模擬平台。其應用最早始於 2001 年，2002 年首次發布開源版本。主要部分的開發工作由德國航空航天中心(Deutsches Zentrum für Luft- und Raumfahrt, DLR)的交通系統研究所承擔。並透過外部團隊支持不同模擬套件的擴展。在過去的十年中，SUMO 已經發展成為一套功能齊全的車流模型套件。其路網導入工具能夠讀取不同來源的道路資料，可支援 VISUM、VISSIM、Shapefiles、OSM、RoboCup、MATsim、OpenDRIVE 和 XML 等格式。需求產生工具可支援不同方式作輸入，如起訖矩陣、路徑需求、流量計數、動態指派等。另外其高效的運算可用於整個城市的車流模擬，以 1GHz 的主機運算一秒可更新 10 萬台車的狀態。自 2001 年以來，SUMO 已在多個國家和國際研究項目中使用，包括交通號誌評估、路徑指引評估、交通調查方法評估、車輛通訊模擬、交通預測等領域的應用。圖 4.3.2.1-3 為 SUMO 路網編輯介面。

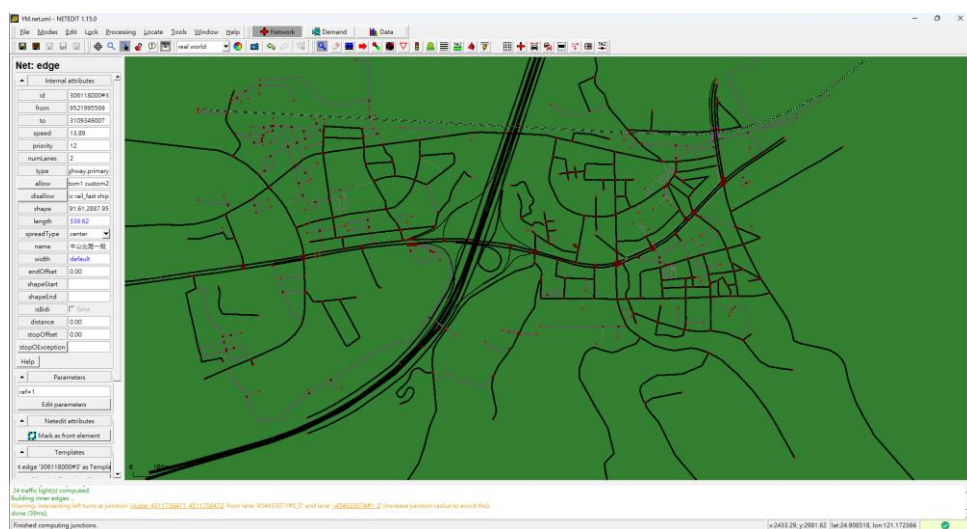


圖 4.3.2.1-3 SUMO 路網編輯介面示意圖

SUMO 提供 TraCI(Traffic Control Interface)工具包，這個套件可以實現訪問者(例如 Python script)對正在運作的模擬即時的存取操作其中的各項設定與行為。因此可用於取得號誌、車輛

狀態、偵測器數值等，亦可使用在動態控制號誌燈號，或是改變特定車輛的參數或行駛路徑等。在 SUMO 模擬過程中，可以透過 TraCI 套件即時動態改動或取得資訊。TraCI 是 Traffic Control Interface 的簡稱。

4.3.2.2 模擬評估流程及結果

透過 RL 模型與車流模擬環境 VISSIM 及 SUMO 互動，模擬運算時相綠燈秒數，再透過人工檢視模型運算結果的合理性以及穩定度，以此循環過程優化模型，流程詳圖 4.3.2.2-1 所示。並依此合理性及穩定性概念，分別檢視臺南市實驗場域及臺北市實驗場域以現場偵測設備收集之資料，輸入模型後針對現場偵測資料所產出之模擬分相綠燈秒數、週期及模擬績效(如通過量及延滯等)。力求其於實測前，RL 模型判定綠燈時長具一定效度，多方確保其可行性。

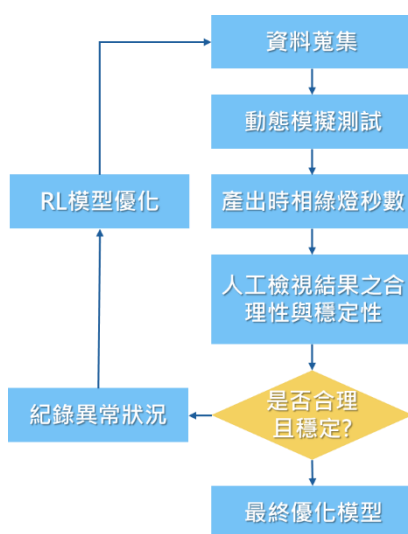


圖 4.3.2.2-1 模擬結果評估流程示意圖

一、合理性

為維護行車安全及促進道路使用效率，號誌設計須符合交通流向、流量及路況，亦即某行駛方向交通量多，其分相綠燈秒數運算結果應足夠安全通過且延滯最小，不合理之情況如某轉向交通量很多，但其分相綠燈秒數運算結果很短；反之同樣非為合理狀況。判定標準以 TOD 為參考基準，單位時間(秒)通過車流差異比應為正值且介於 1~4 秒之間，亦即分相配置之秒數至少能通過適當比例車流，並符合 TOD 最大週期及最小綠 10 秒之限制。其公式計算如下：

$$\chi = \frac{V}{T}$$

V：流量 / T：分相秒數 / χ ：單位時間(秒)通過車流差異比

二、穩定度

同天同時段內之穩定度與不同天同時段之穩定度確認，當出現各時相綠燈秒數差異很大之情況，就需特別注意該天是否有突發狀況，選擇無突發狀況之其他天的交通調查資料進行

運算確認穩定度。判定標準以 AI 強化學習模型產出之分相秒數及該分相通過之流量皆應介於 $\mu \pm 2\sigma$ ，亦即 95%信賴區間內，其公式計算如下：

$$\mu_{V,T} - 2\sigma_{V,T} \leq \chi \leq \mu_{V,T} + 2\sigma_{V,T}$$

μ_v ：AI 強化學習模型分相秒數通過流量平均值

σ_v ：AI 強化學習模型分相秒數通過流量標準差

μ_T ：AI 強化學習模型分相秒數平均值

σ_T ：AI 強化學習模型分相秒數標準差

$\chi_{V,T}$ ：AI 強化學習模型產出之該分相通過之流量或分相秒數

4.3.2.3 單一路口場域(臺南市「台 86-19 甲線」路口)模擬測試分析

依偏好停等延滯為最低之獎勵目標，分析其進行 10 次不同亂數種子的 training episode，藉此產出對應之平日晨峰 5 時相(07：10~08：00)、平日晨峰 4 時相(08：00~09：00)、平日離峰(12：30~14：30)、平日昏峰(16：30~18：30)及假日昏峰(16：30~18：30)時段之分相秒數與績效結果。將績效分為幹道及支道總延滯時間與通過量，以對應 AI 強化學習模型訓練之獎勵目標，以及整體路口的平均每車停等延滯及 PCU，同時其輸出秒數亦須符合 4.3.2.2 節所定義之合理性及穩定性。不同權重的獎勵目標所產生的績效整理如下：

表 4.3.2.3-1 台南場域平均兩小時 TOD 與 AI 權重偏好績效比較彙整表

時段		TOD		AI_通過量		AI_平均		AI_延滯		AI_主幹道		AI_支道	
		平均每車停等延滯	每小時整體路口 PCU	平均每車停等延滯	每小時整體路口 PCU	平均每車停等延滯	每小時整體路口 PCU	平均每車停等延滯	每小時整體路口 PCU	平均每車停等延滯	每小時整體路口 PCU	平均每車停等延滯	每小時整體路口 PCU
平日晨峰 5 相	μ	37.20	4,974	39.71	4,896	39.46	4,841	37.96	5,079	40.06	4,851	39.27	4,862
	δ	-	-	2.51	-78	2.26	-133	0.76	105	2.86	-123	2.07	-112
平日晨峰 4 相	μ	37.20	4,974	27.60	4,801	28.80	4,684	29.27	4,860	28.23	4,790	29.70	4,630
	δ	-	-	-9.60	-173	-8.40	-291	-7.93	-114	-8.97	-183	-7.50	-344
平日離峰	μ	38.50	6,899	32.84	6,562	39.12	6,557	45.79	6,874	37.32	6,548	44.86	6,527
	δ	-	-	-5.66	-337	0.62	-342	7.29	-25	-1.18	-351	6.36	-372
平日昏峰	μ	65.30	9,181	73.73	8,759	63.26	8,846	59.77	9,265	73.37	8,742	65.09	8,775
	δ	-	-	8.43	-422	-2.04	-335	-5.53	84	8.07	-439	-0.21	-406
假日昏峰	μ	58.60	8,670	63.44	8,201	51.94	8,259	51.60	8,674	66.75	8,180	56.32	8,234
	δ	-	-	4.84	-469	-6.66	-411	-7.00	4	8.15	-491	-2.28	-436

註 1： μ =10 個兩小時 episode，總 20 小時之績效平均值， δ (差值)=時段績效表現與 TOD 差值

註 2：停等延滯單位為秒，通過量單位為 PCU/hr。

註 3：欄位偏好名稱對應表 1-2 獎勵目標偏好權重表之偏好名稱

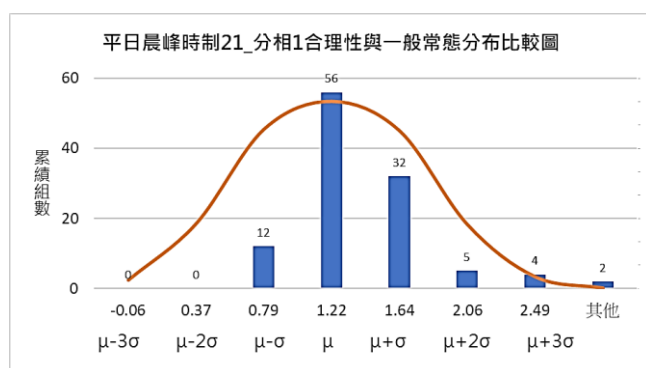
(一)平日晨峰 5 時相(07：10~08：00)

本期 AI 強化學習模型模擬產出之通過量約 5,079PCU，平均停等延滯為 38.0 秒；其各分相秒數之單位時間(秒)通過車流差異比皆為正值，各分相差異均值分別為分相一的 1.22 PCU/秒、分相二的 0.28 PCU/秒、分相三的 0.09 PCU/秒、分相四的 0.64 PCU/秒及分相五的 0.12 PCU/秒；週期及各分相趨勢圖顯示其週期秒數約呈現 281 秒上下集中趨勢，各分相秒數趨勢分別為分相一的 45 秒、分相二的 100 秒、分相三的 43 秒、分相四的 42 秒及分相五的 51 秒，整體皆較 TOD 的 220 秒號誌周期高。

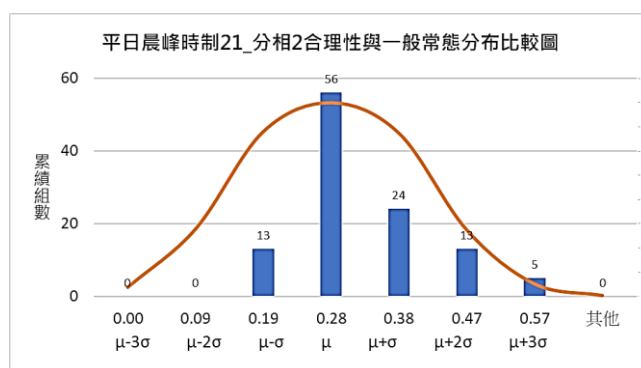
表 4.3.2.3-2 平日晨峰 5 時相模擬偏好延滯情境績效彙整表

	A (幹道)			B (支道)			C (幹道)			D (支道)			通過量 (PCU)	平均停等延滯 (秒)
	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉		
TOD	287	632	320	189	462	469	269	1144	124	244	263	423	4974	37.2
本期 AI	386	595	361	294	416	503	388	1113	140	382	226	469	5079	38.0

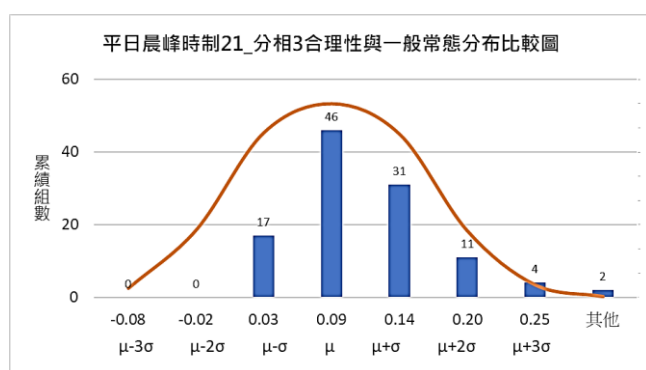
註：本表為 10 個 episode 之績效平均值



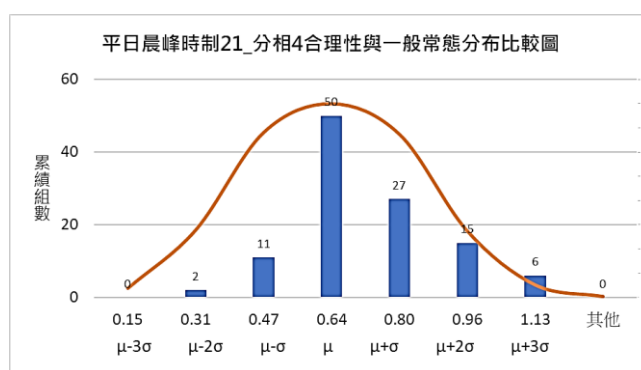
分相 1



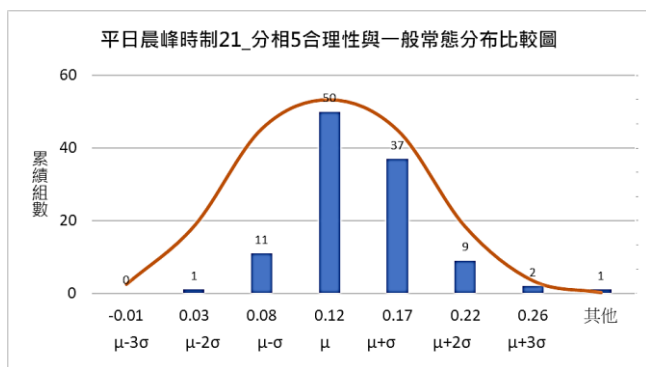
分相 2



分相 3

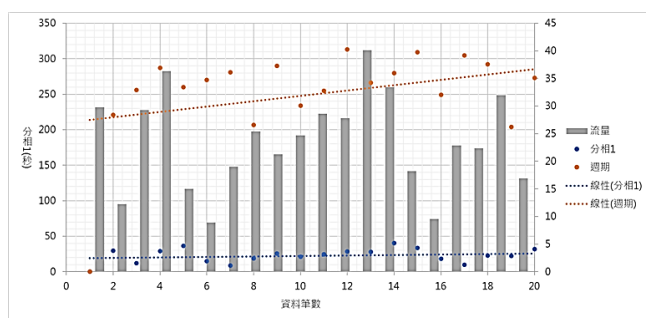


分相 4

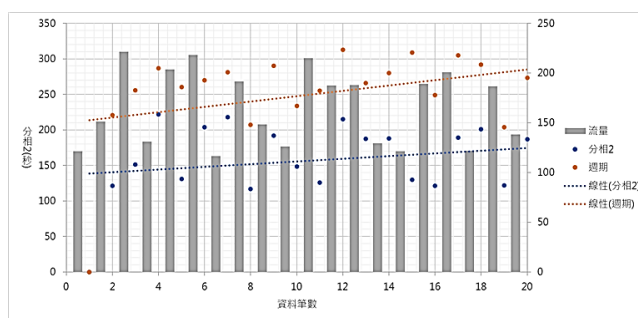


分相 5

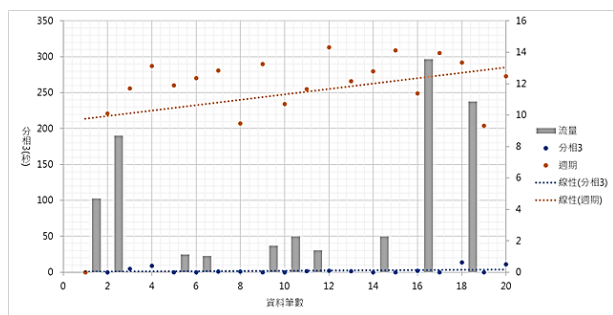
圖 4.3.2.3-1 平日晨峰 5 時相本期 AI 強化學習模型模擬各分相秒數穩定性分析



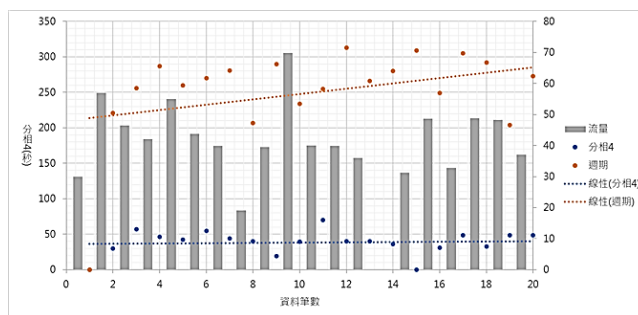
分相 1



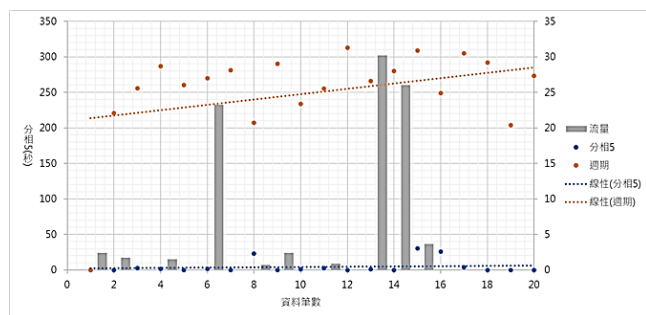
分相 2



分相 3



分相 4



分相 5

圖 4.3.2.3-2 本期 AI 強化學習模型於平日晨峰 5 時相各分相秒數合理性分析

(二)平日晨峰 4 時相(08：00~09：00)

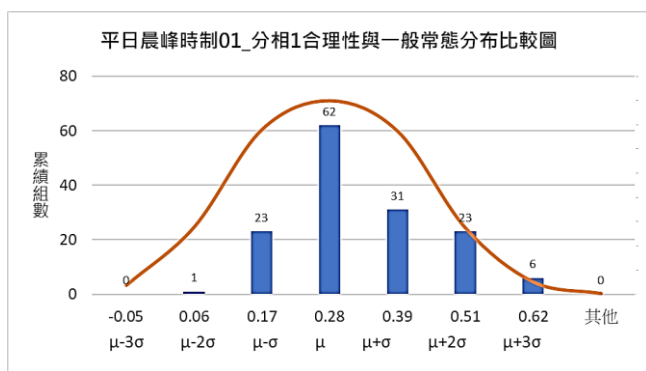
本期 AI 強化學習模型模擬產出之通過量約 4,863PCU，平均停等延滯為 29.3 秒；其各分相秒數之單位時間(秒)通過車流差異比皆為正值，各分相差異均值分別為分相一

的 0.28 PCU/秒、分相二的 0.15 PCU/秒、分相三的 0.37 PCU/秒及分相四的 0.12 PCU/秒；週期及各分相趨勢圖顯示其週期秒數約呈現 223 秒上下集中趨勢，各分相秒數趨勢分別為分相一的 87 秒、分相二的 43 秒、分相三的 41 秒及分相四的 51 秒，整體皆較 TOD 的 160 秒號誌周期高。

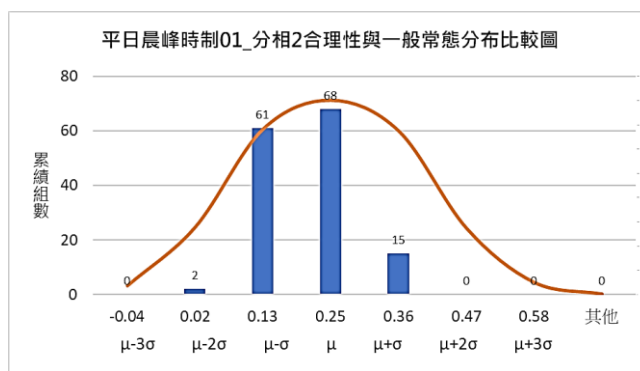
表 4.3.2.3-3 平日晨峰 4 時相模擬偏好延滯情境績效彙整表

	A (幹道)			B (支道)			C (幹道)			D (支道)			通過量 (PCU)	平均停等延滯 (秒)
	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉		
TOD	287	632	320	189	462	469	269	1144	124	244	263	423	4974	37.2
本期 AI	475	641	313	193	386	482	317	924	112	366	251	361	4863	29.3

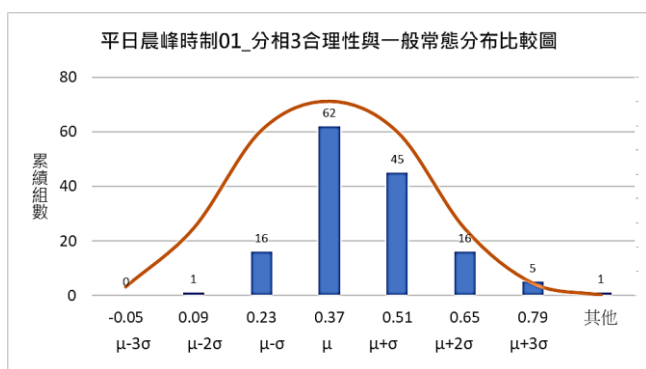
註：本表為 10 個 episode 之績效平均值



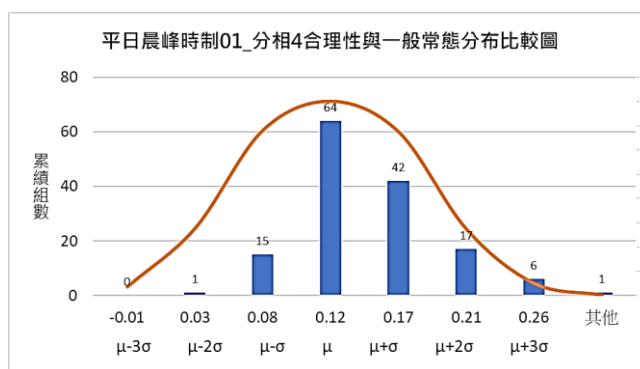
分相 1



分相 2

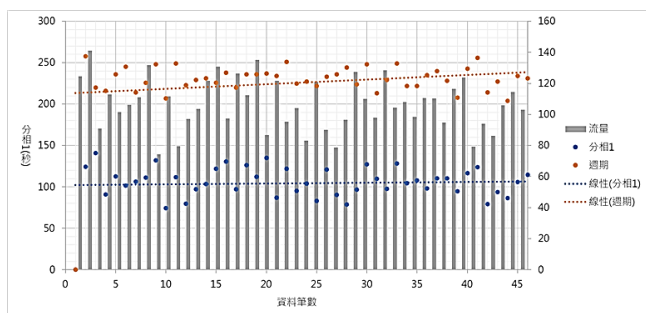


分相 3

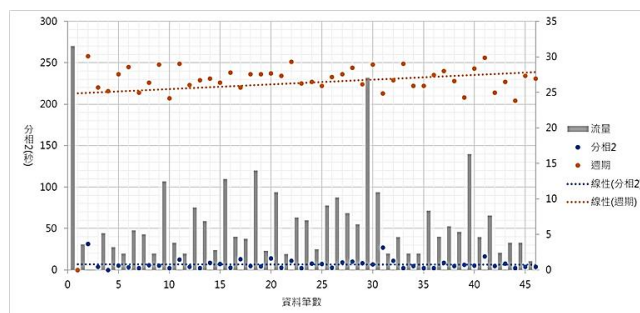


分相 4

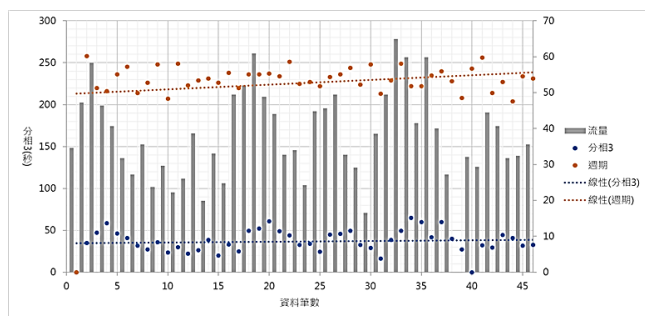
圖 4.3.2.3-3 本期 AI 強化學習模型於平日晨峰 4 時相模擬各分相秒數穩定性分析



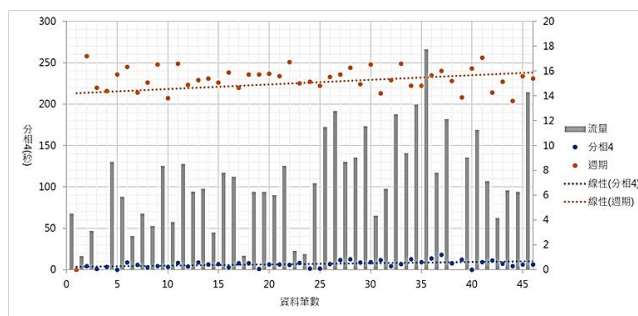
分相 1



分相 2



分相 3



分相 4

圖 4.3.2.3-4 本期 AI 強化學習模型於平日晨峰 4 時相模擬各分相秒數合理性分析

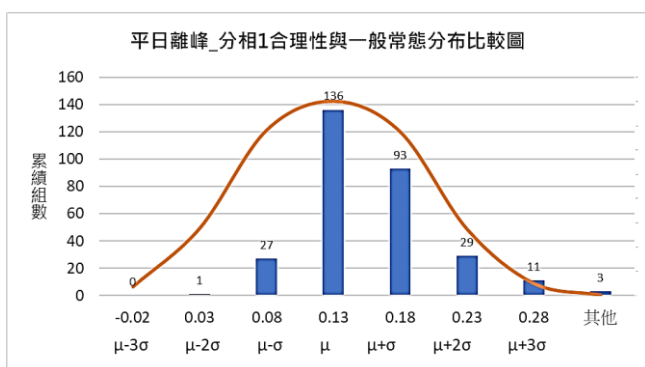
(三)平日離峰(12：30~14：30)

本期 AI 強化學習模型模擬產出之通過量約 6,882 PCU，平均停等延滯為 44.5 秒；其各分相秒數之單位時間(秒)通過車流差異比皆為正值，各分相差異均值分別為分相一的 0.13 PCU/秒、分相二的 0.13 PCU/秒、分相三的 0.22 PCU/秒及分相四的 0.14 PCU/秒；週期及各分相趨勢圖顯示其週期秒數約呈現 231 秒上下集中趨勢，各分相秒數趨勢分別為分相一的 102 秒、分相二的 43 秒、分相三的 45 秒及分相四的 41 秒，整體皆較 TOD 的 160 秒號誌周期高。

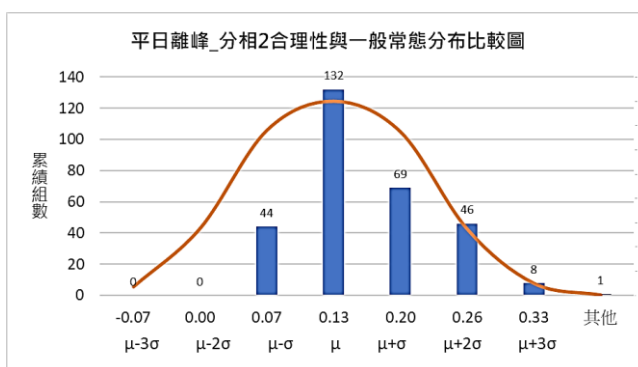
表 4.3.2.3-4 平日離峰模擬偏好延滯情境績效彙整表

	A (幹道)			B (支道)			C (幹道)			D (支道)			通過量 (PCU)	平均停等延滯 (秒)
	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉		
TOD	486	1073	529	309	370	433	518	1246	211	505	388	519	6899	38.5
本期 AI	485	1059	528	309	368	434	518	1231	210	524	389	521	6875	45.8

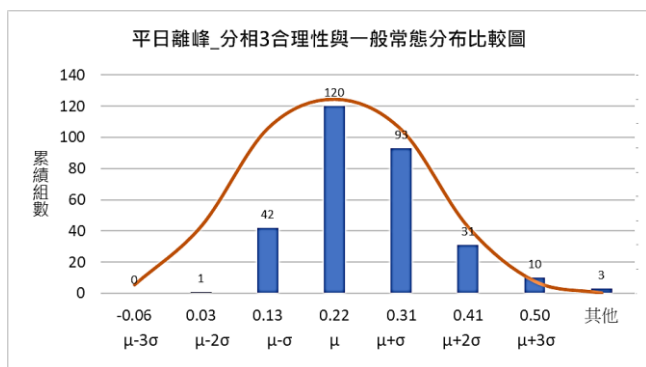
註：本表為 10 個 episode 之績效平均值



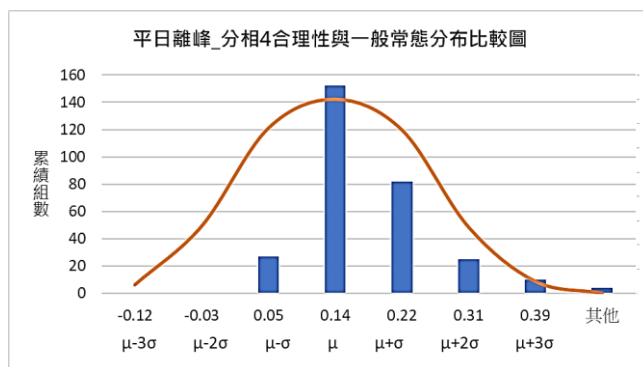
分相 1



分相 2

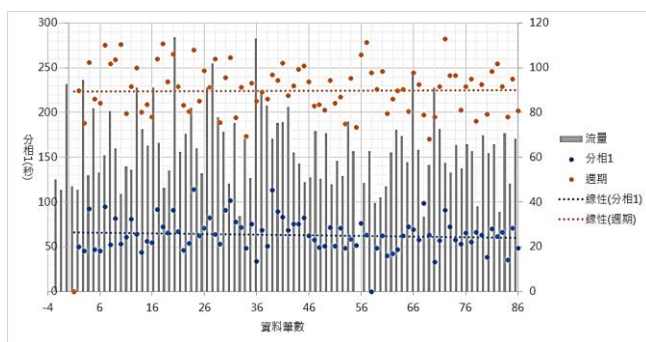


分相 3

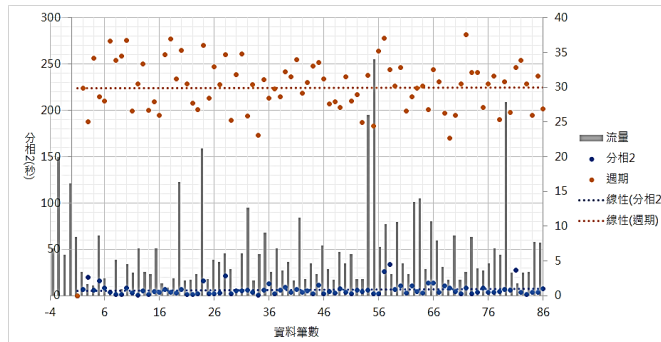


分相 4

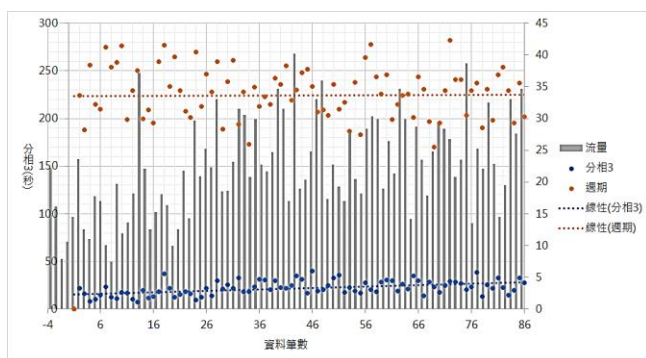
圖 4.3.2.3-5 本期 AI 強化學習模型於平日離峰模擬各分相秒數穩定性分析



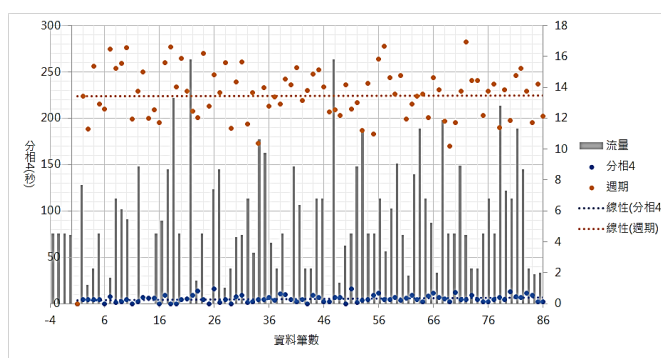
分相 1



分相 2



分相 3



分相 4

圖 4.3.2.3-6 本期 AI 強化學習模型於平日離峰模擬各分相秒數合理性分析

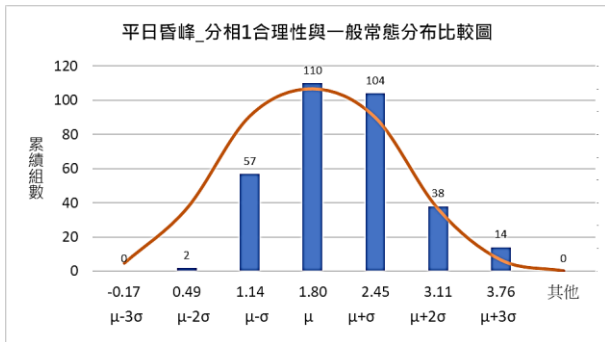
(四)平日昏峰(16：30~18：30)

本期 AI 強化學習模型模擬產出之通過量約 9,438 PCU，平均停等延滯為 56.2 秒；其各分相秒數之單位時間(秒)通過車流差異比皆為正值，各分相差異均值分別為分相一的 1.80 PCU/秒，分相二的 0.20 PCU/秒、分相三的 0.19 PCU/秒、分相四的 0.57 PCU/秒及分相五的 0.10 PCU/秒；週期及各分相趨勢圖顯示其週期秒數約呈現 211 秒上下集中趨勢，各分相秒數趨勢分別為分相一的 26 秒、分相二的 88 秒、分相三的 26 秒、分相四的 31 秒及分相五的 40 秒，模擬分相二秒數平均較 TOD 分相二秒數低 26 秒，週期亦較 TOD 的號誌週期 240 秒低。

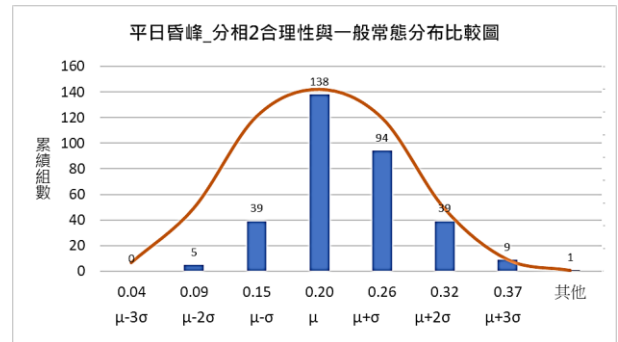
表 4.3.2.3-5 平日昏峰模擬偏好延滯情境績效彙整表

	A (幹道)			B (支道)			C (幹道)			D (支道)			通過量 (PCU)	平均停等延滯 (秒)
	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉		
TOD	695	2136	612	326	581	541	452	1570	244	494	688	662	9181	65.3
本期 AI	1016	2084	664	446	524	573	881	1510	288	676	596	689	9266	59.8

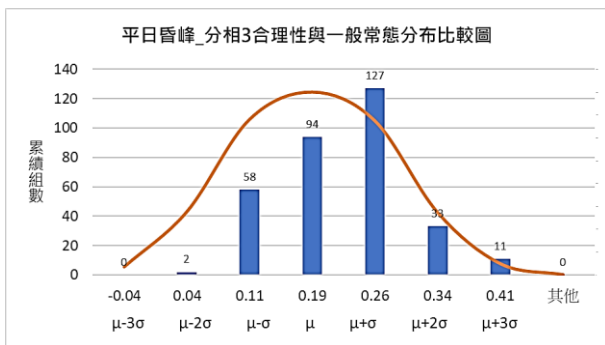
註：本表為 10 個 episode 之績效平均值



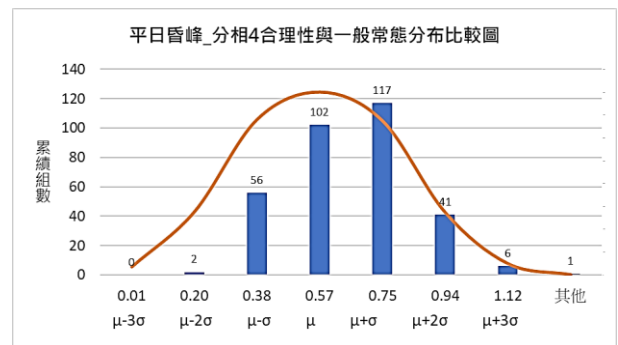
分相 1



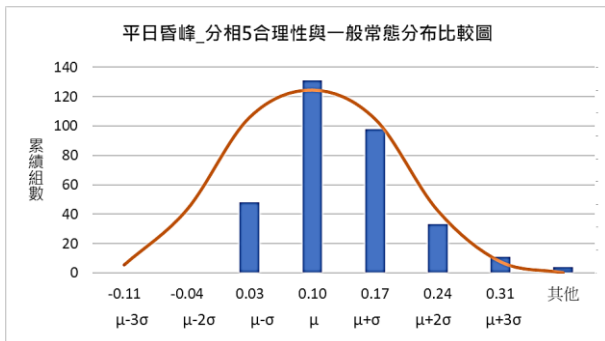
分相 2



分相 3

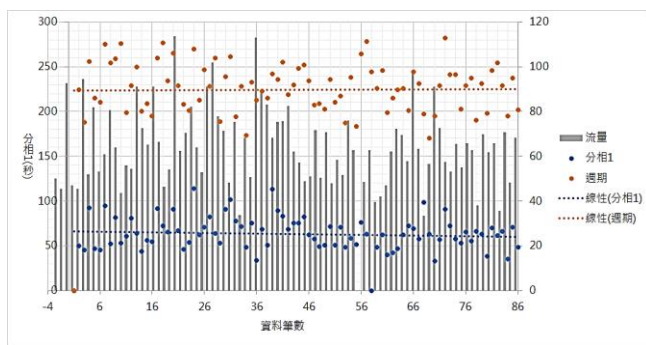


分相 4

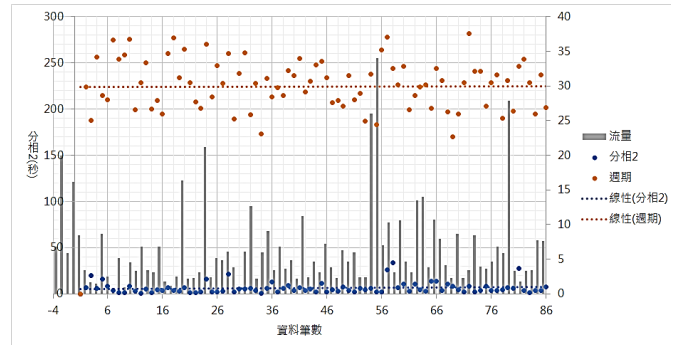


分相 5

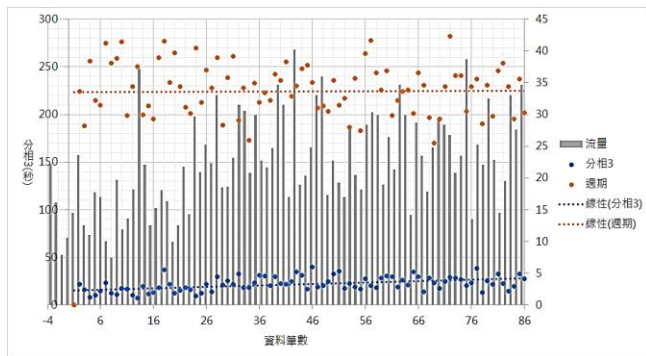
圖 4.3.2.3-7 本期 AI 強化學習模型於平日昏峰模擬各分相秒數穩定性分析



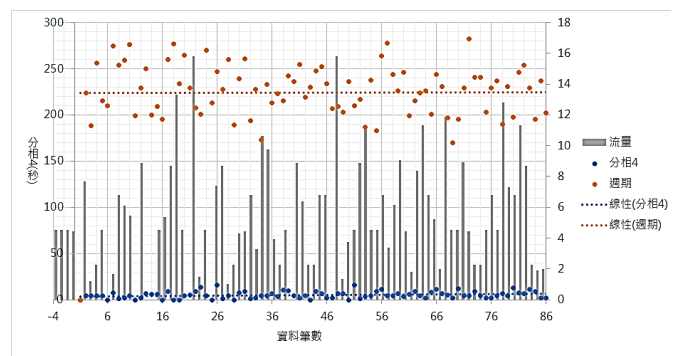
分相 1



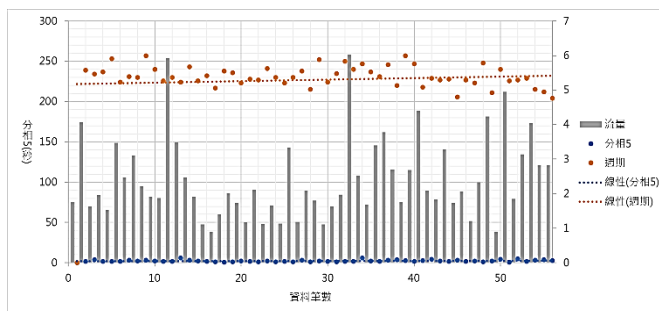
分相 2



分相 3



分相 4



分相 5

圖 4.3.2.3-8 本期 AI 強化學習模型於平日昏峰模擬各分相秒數合理性分析

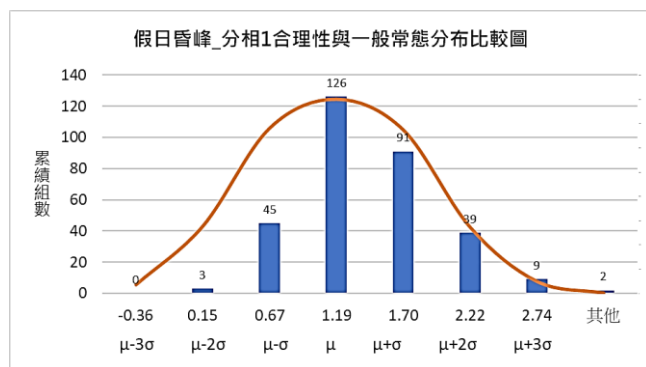
(五) 假日昏峰(16:30~18:30)

本期 AI 強化學習模型模擬產出之通過量約 8,570 PCU，平均停等延滯為 54.8 秒；其各分相秒數之單位時間(秒)通過車流差異比皆為正值，各分相差異均值分別為分相一的 1.19 PCU/秒、分相二的 0.25 PCU/秒、分相三的 0.09 PCU/秒、分相四的 0.54 PCU/秒及分相五的 0.06 PCU/秒；週期及各分相趨勢圖顯示其週期秒數約呈現 219 秒上下集中趨勢，各分相秒數趨勢分別為分相一的 31 秒、分相二的 73 秒、分相三的 27 秒、分相四的 41 秒及分相五的 47 秒，模擬分相二秒數平均較 TOD 分相二秒數低 41 秒，週期亦較 TOD 的號誌週期 240 秒略低。

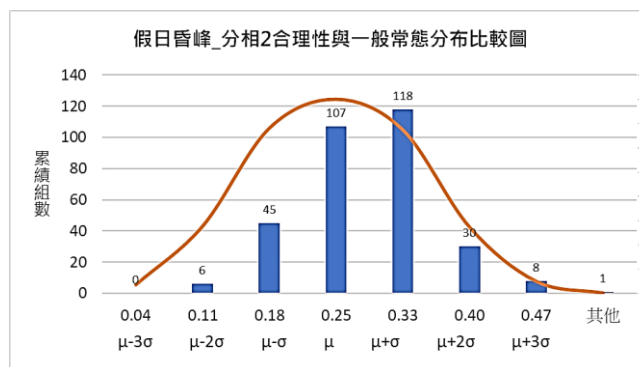
表 4.3.2.3-6 假日昏峰模擬偏好延滯情境績效彙整表

	A (幹道)			B (支道)			C (幹道)			D (支道)			通過量 (PCU)	平均停 等延滯 (秒)
	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉	左轉	直行	右轉		
TOD	719	1822	666	399	569	524	437	1595	272	511	524	584	8670	58.6
本期 AI	1027	1782	768	556	536	584	606	1443	306	694	485	641	8675	51.6

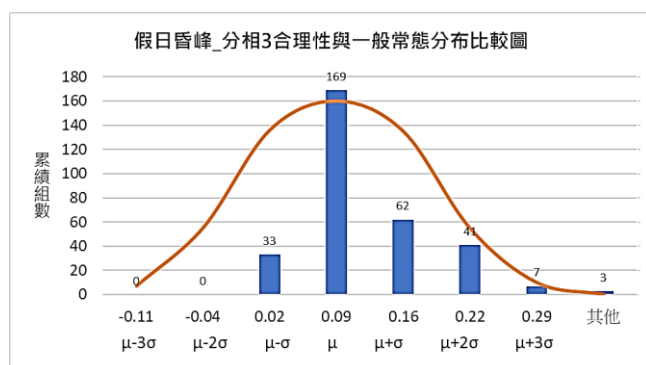
註：本表為 10 個 episode 之績效平均值



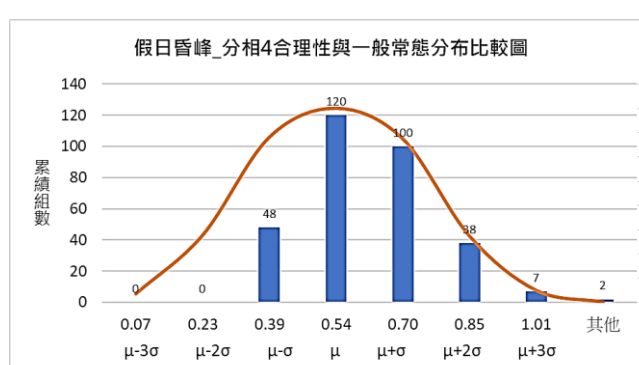
分相 1



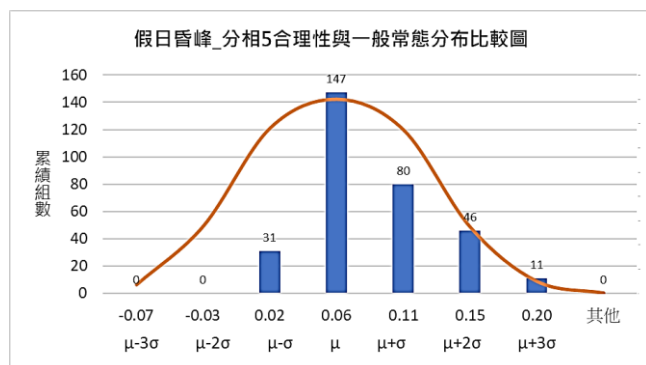
分相 2



分相 3

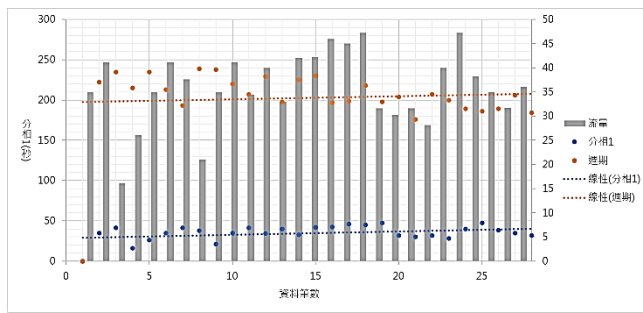


分相 4

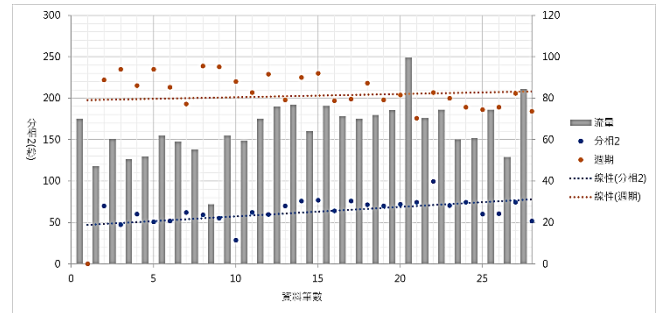


分相 5

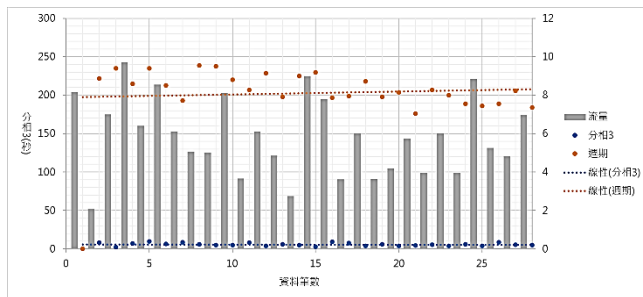
圖 4.3.2.3-9 本期 AI 強化學習模型於假日昏峰模擬各分相秒數穩定性分析



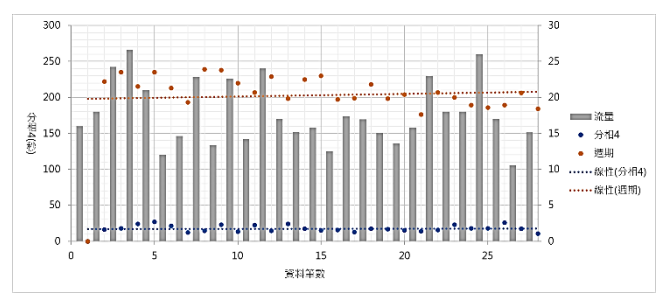
分相 1



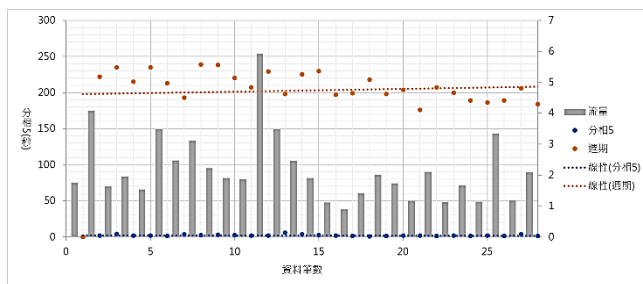
分相 2



分相 3



分相 4



分相 5

圖 4.3.2.3-10 本期 AI 強化學習模型於假日昏峰模擬各分相秒數合理性分析

4.3.2.4 幹道多路口場域(臺北市「中山北路-德行東路」)模擬測試分析

模擬環境測試以平日晨峰(07:00~09:00)、平日離峰(中午 12:30~14:30)、平日昏峰(16:30~18:30)及假日昏峰(16:30~18:30)4 個時段為主，將依照不同的獎勵目標，給予不同的權重，分別為注重延滯表現、通過量表現、主線表現、支線表現及平均表現等權重。將績效分為幹道及支道總延滯時間與通過量，以對應 AI 強化學習模型訓練之獎勵目標，以及整體路口的平均每車停等延滯及 PCU，同時其輸出秒數亦須符合 4.3.2.2 節所定義之合理性及穩定性。透過分析不同權重的獎勵目標所產生的績效，進而提供未來應用的參考，TOD 與各權重績效表現整理如表 4.3.2.4-1 所示，並說明如下。

表 4.3.2.4-1 平均兩小時 TOD 與 AI 權重偏好績效比較彙整表

時段		TOD		AI_通過量		AI_平均		AI_延滯		AI_主幹道		AI_支道	
		單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU
平日 晨峰	μ	41.8	17046	31.1	17028	30.3	17119	30.3	17020	30.6	16968	30.6	16967
	δ	—	—	-10.6	-19	-11.5	73	-11.5	-27	-11.2	-78	-11.1	-79
平日 離峰	μ	33.8	15651	33.0	15828	32.4	15861	33.8	15736	33.4	15772	33.0	15755
	δ	—	—	-0.8	177	-1.4	209	0.0	85	-0.4	121	-0.8	103
平日 昏峰	μ	27.15	17972	31.29	18016	31.29	18016	33.79	18043	30.96	18234	29.84	18053
	δ	—	—	4.1	44	4.1	44	6.6	71	3.8	262	2.7	82
假日 昏峰	μ	23.59	17177	27.2	17385	27	17290	27.03	17019	27.45	17319	26.56	17294
	δ	—	—	3.6	208	3.4	113	3.4	-158	3.9	143	3.0	117

註 1： μ (平均值)=為 10 個兩小時 episode，總 20 小時之績效平均值， δ (差值)=時段績效表現與 TOD 差值

註 2：停等延滯單位為秒，通過量單位為 PCU/hr。

註 3：欄位偏好名稱對應表 1-2 獎勵目標偏好權重表之偏好名稱

(一)平日晨峰(07：00~09：00)

1.秒數分布

AI 模擬週期秒數趨勢集中約 209 秒，略高於 TOD 的 200 秒。

(1)中山北路/克強路路口(X2)

各分相秒數平均趨勢分別為分相一的 9 秒、分相二的 105 秒、分相三的 27 秒、分相四的 8 秒、分相五的 36 秒及分相六的 24 秒。

(2)中山北路/195 巷路口(X3)

各分相秒數平均趨勢分別為分相一的 121 秒、分相二的 42 秒及分相三的 46 秒。

(3)中山北路/德行西路路口(X4)

各分相秒數趨勢分別為分相一的 18 秒、分相二的 73 秒、分相三的 16 秒、分相四的 37 秒、分相五的 40 秒及分相六的 25 秒。

2.績效

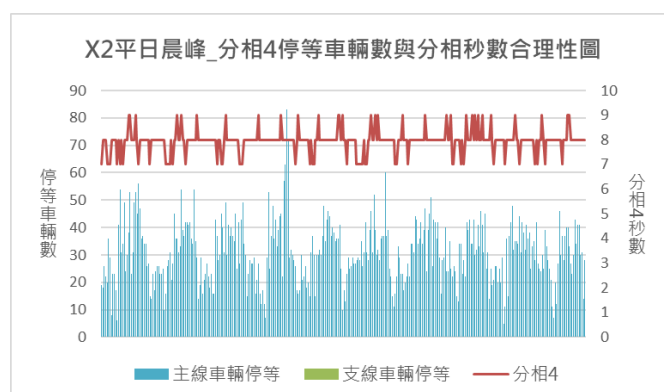
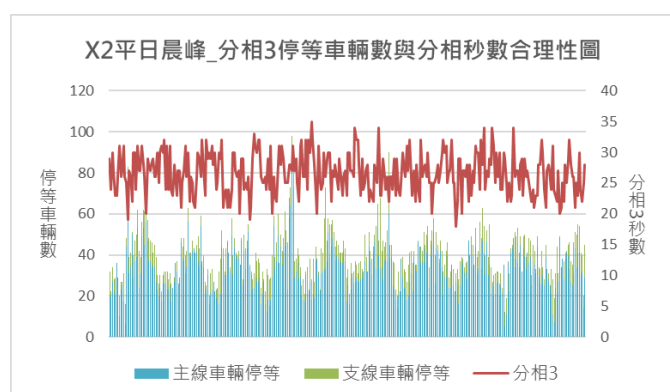
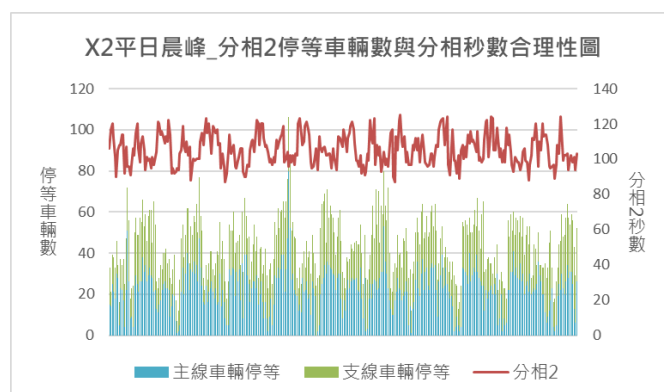
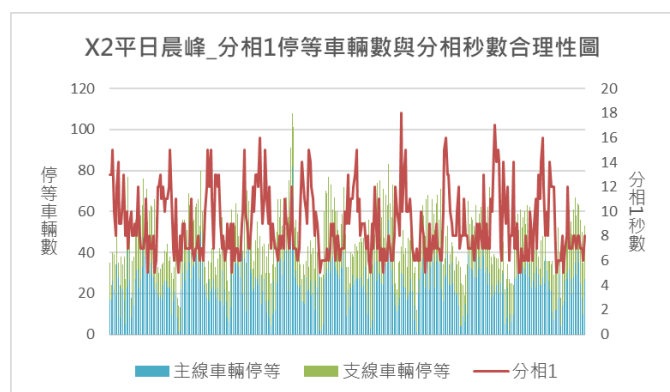
在注重平均表現偏好下，單一方平均每車停等延滯減少 11.5 秒，每小時整體路口增加 73PCU。

表 4.3.2.4-2 平日晨峰 TOD 及本期 AI 強化學習模型模擬注重平均表現績效彙整表

次數	TOD						AI 強化學習模型					
	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	幹道每小時 單一車道		支道每小時 單一車道		單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	幹道每小時 單一車道		支道每小時 單一車道	
			停等 延滯	通過量	停等 延滯	通過量			停等 延滯	通過量	停等 延滯	通過量
1	43.1	17208	10342	644	22794	518	29.0	16984	11317.4	615	24410.1	511
2	39.1	17331	9993	638	21855	510	27.7	17252	11217.9	637	24181.9	504
3	45.1	17096	10104	625	22341	500	32.0	16918	11432.3	611	23655.4	498
4	47.0	17326	10372	629	23112	523	33.0	17211	11464.2	629	25237.4	512
5	64.5	17341	10336	638	22602	514	29.5	17054	12753.3	617	25805.4	499
6	32.1	17194	10190	634	21441	526	28.9	17012	11081.5	626	27003.6	507
7	36.4	17340	10302	635	19833	528	28.6	17073	11118.8	630	22613.9	517
8	39.5	15449	10259	624	19745	511	31.0	17196	11559.7	631	25264.9	507
9	36.9	17144	10070	610	20520	517	28.2	17452	11375.1	643	23794.2	517
10	34.1	17035	10094	618	20745	522	35.0	17039	11079.4	619	24161.2	509
μ	41.8	17046	10206	630	21499	517	30.3	17119	11440.0	626	24612.8	508

註 1：停等延滯單位為秒，通過量單位為 PCU/hr，每一行為該時段兩小時模擬平均值， μ (平均值)總 20 小時之績效平均值。

註 2：本表為各 episode 之績效平均值



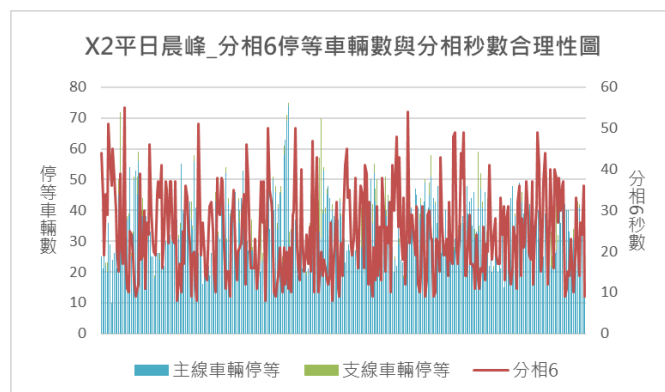
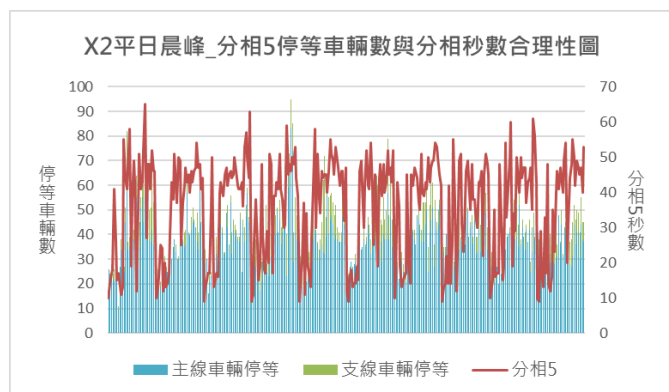


圖 4.3.2.4-1 中山北路/克強路平日晨峰模擬合理性分析

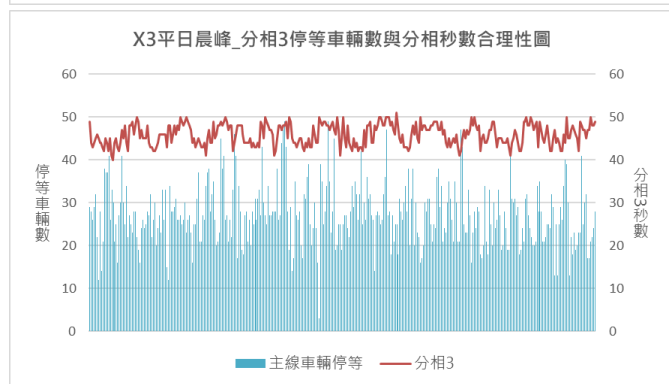
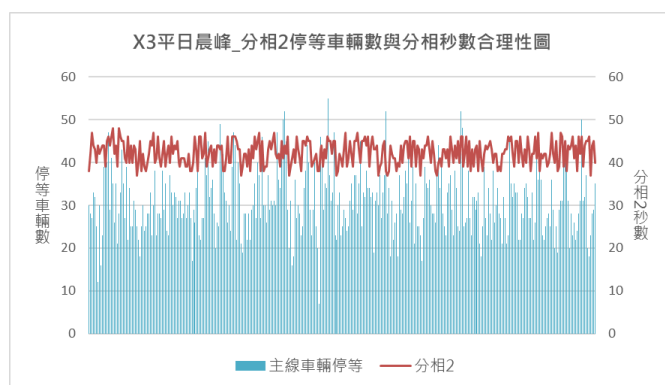
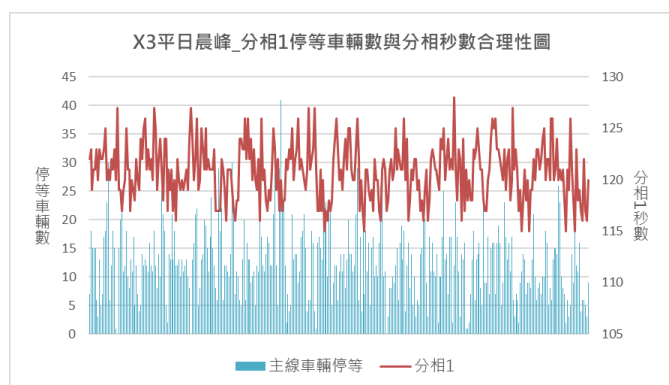
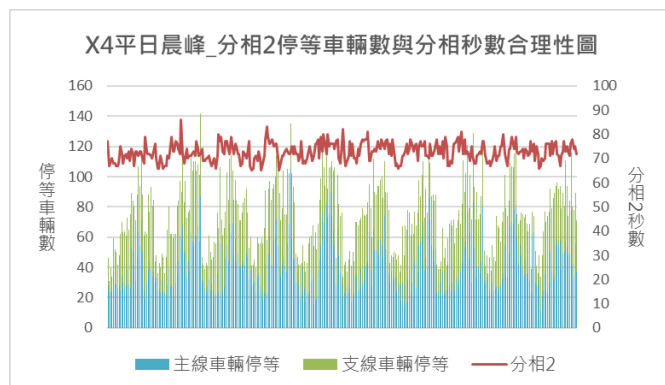
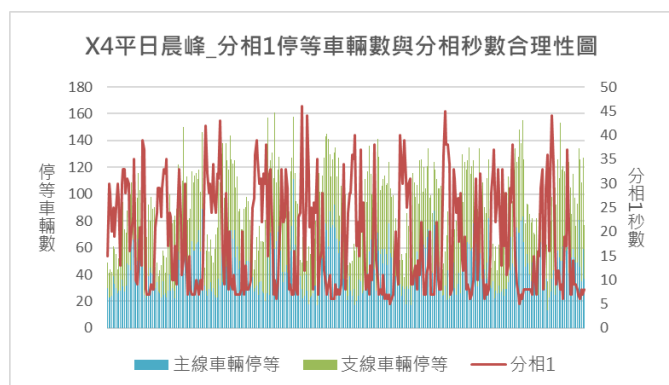


圖 4.3.2.4-2 中山北路/195 巷平日晨峰模擬合理性分析



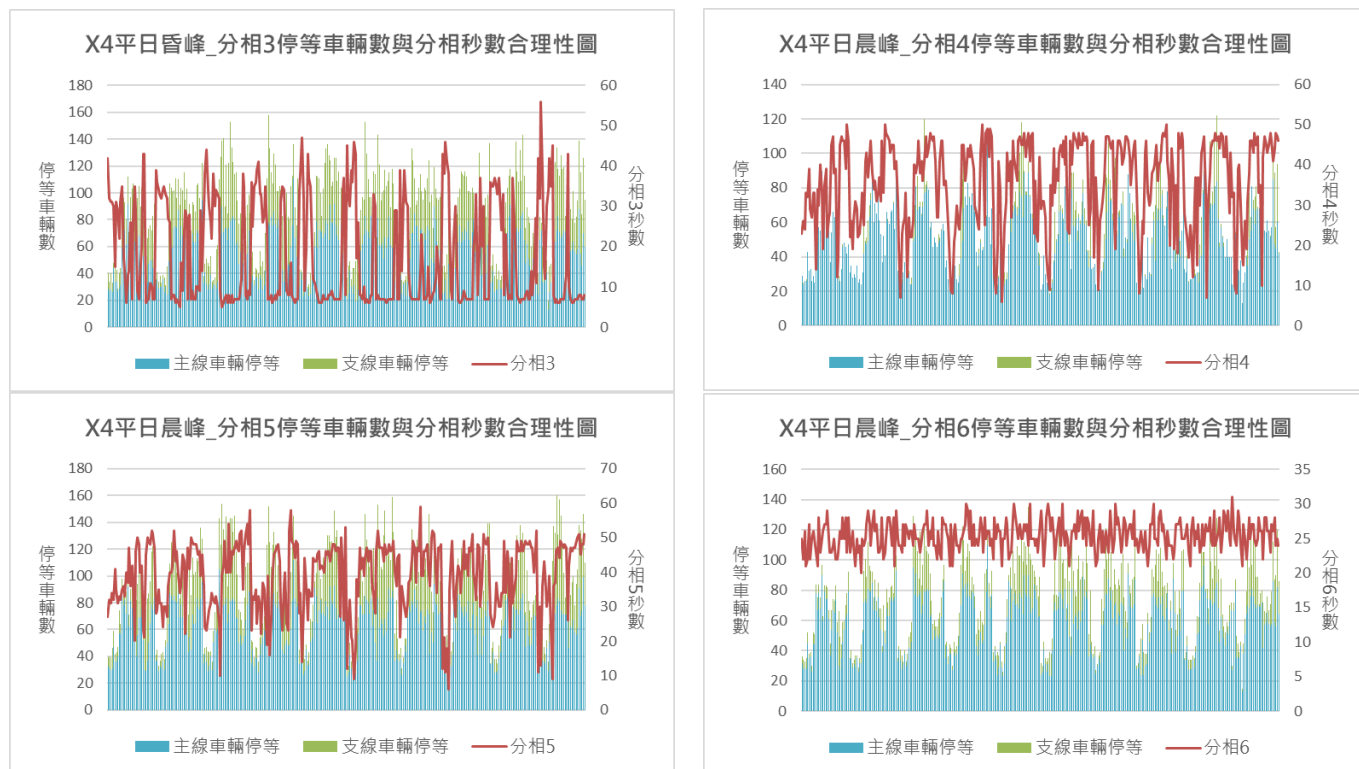


圖 4.3.2.4-3 中山北路/德行西路平日晨峰模擬合理性分析

表 4.3.2.4-3 中山北路/克強路平日晨峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.42	2	0.47	2	0.49	2	0.47	2	0.45
3	0.28	3	0.32	3	0.38	3	0.36	3	0.53
4	0.17	4	0.30	4	0.45	4	0.43	4	0.25
5	0.28	5	0.30	5	0.43	5	0.43	5	0.55
6	0.40	6	0.32	6	0.19	6	0.17	6	0.28
7	0.40	7	0.34	7	0.25	7	0.28	7	0.28
8	0.34	8	0.30	8	0.36	8	0.25	8	0.32

註：分相 4 為行人專用時相，配給固定秒數

表 4.3.2.4-4 中山北路/克強路平日晨峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	92	1.99	8.53	0.97
	1	50	1.55	9.34	0.98
分相 2	0	60	6.32	104.90	0.95
	1	82	7.53	113.04	0.95
分相 3	0	61	2.96	26.16	0.95
	1	81	3.25	28.75	0.93
分相 5	0	90	2.33	10.54	0.96
	1	52	3.79	12.65	0.96
分相 6	0	48	10.06	41.38	0.90
	1	25	7.38	46.48	0.96
	2	38	10.21	38.16	0.97
	3	11	5.01	49.64	0.91
	4	20	6.56145	49	1

表 4.3.2.4-5 中山北路/195 巷平日晨峰模擬穩定性分析

分相 1		分相 2	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.47	2	0.42
3	0.47	3	0.43
4	0.45	4	0.43
5	0.26	5	0.40
6	0.19	6	0.25
7	0.23	7	0.32
8	0.26	8	0.30

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	75	2.55	119.91	0.97
	1	67	2.66	121.85	1.00
分相 2	0	47	2.92	42.53	1.00
	1	33	2.82	42.03	1.00
	2	55	3.05	42.71	1.00
	3	7	2.54	41.86	1.00

註：分相 3 為行人專用時相，配給固定秒數

表 4.3.2.4-6 中山北路/德行西路平日晨峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 4		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.49	2	0.26	2	0.13	2	0.55	2	0.38	2	0.62
3	0.40	3	0.45	3	0.36	3	0.43	3	0.42	3	0.57
4	0.49	4	0.30	4	0.26	4	0.30	4	0.43	4	0.55
5	0.40	5	0.26	5	0.21	5	0.30	5	0.26	5	0.55
6	0.28	6	0.26	6	0.34	6	0.13	6	0.36	6	0.55
7	0.19	7	0.38	7	0.36	7	0.17	7	0.09	7	0.55
8	0.21	8	0.45	8	0.47	8	0.32	8	0.32	8	0.55

表 4.3.2.4-7 中山北路/德行西路平日晨峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	47	8.17	21.09	0.96
	1	95	7.46	17.35	0.98
分相 2	0	40	4.07	75.35	0.95
	1	44	2.95	72.34	0.95
	2	58	3.66	74.02	0.97
分相 3	0	11	5.92	31.91	0.91
	1	29	6.33	22.45	1.00
	2	29	8.63	25.28	1.00
	3	23	7.85	26.91	0.96
	4	16	7.59	27.56	1.00
	5	6	10.19	24.67	1.00
	6	7	5.65	29.57	1.00
	7	21	10.90	27.14	0.95
分相 4	0	80	6.54	34.30	0.94
	1	62	4.45	37.21	0.97
分相 5	0	33	8.49	22.00	0.97
	1	71	10.11	29.79	0.96
	2	7	9.71	22.71	1.00
	3	31	11.74	35.45	0.94
分相 6	0	107	2.3	25.3	1.0
	1	35	2.5	26.5	1.0

(二)平日離峰(12：30~14：30)

1.秒數分布

平日離峰 09：00~13：00 運行時制計畫 04 週期 150 秒；13：00~16：30 運行時制計畫 02 週期 200 秒，AI 模擬週期秒數趨勢集中約 209 秒，皆略高於 TOD 150 秒及 200 秒。

離峰時段 AI 仍給予幹道 209 秒週期，經評估與事前調查的結果相近。該路段於交通調查時，平日尖峰與離峰的交通量及分布特性並無非常顯著之差異，故雖與離峰使用半周期或短周期之常用作法相異，應可將 AI 輸出結果視為合理。

(1)中山北路/克強路路口(X2)

各分相秒數平均趨勢分別為分相一的 9 秒、分相二的 89 秒、分相三的 22 秒、分相四的 8 秒、分相五的 37 秒及分相六的 43 秒。

(2)中山北路/195 巷路口(X3)

各分相秒數平均趨勢分別為分相一的 108 秒、分相二的 38 秒及分相三的 63 秒。

(3)中山北路/德行西路路口(X4)

各分相秒數趨勢分別為分相一的 17 秒、分相二的 77 秒、分相三的 22 秒、分相四的 29 秒、分相五的 37 秒及分相六的 27 秒。

2.績效

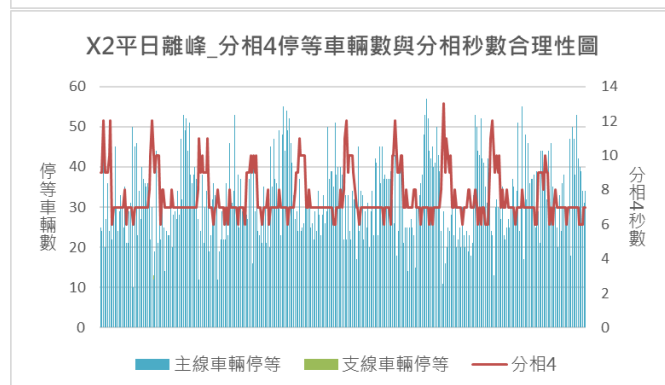
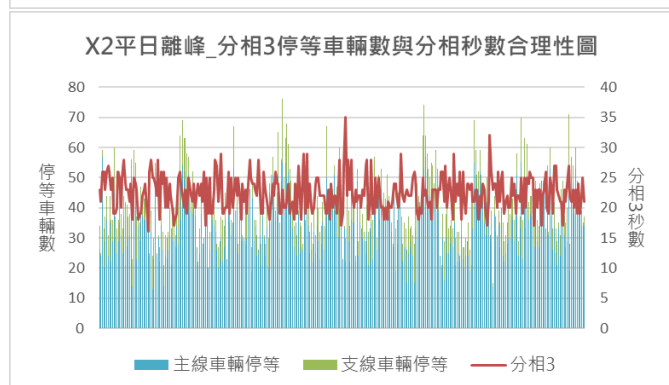
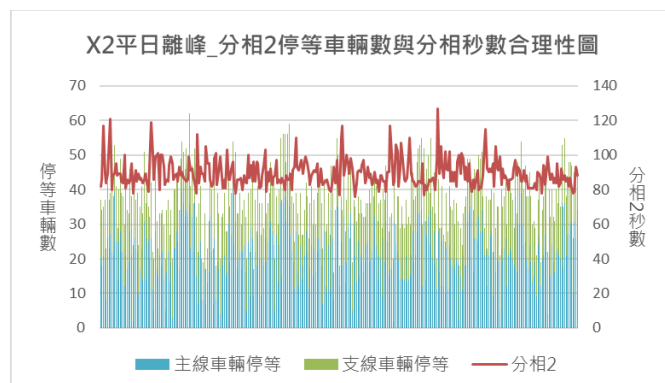
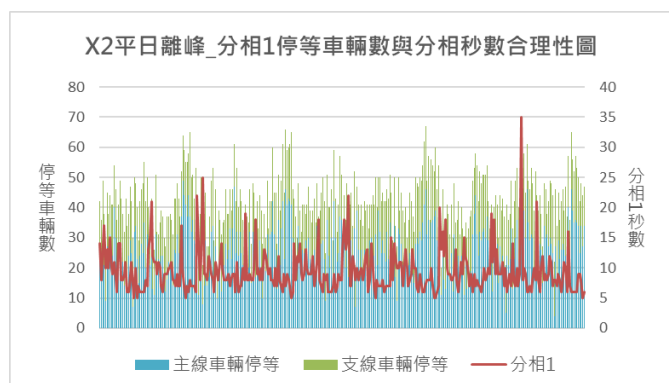
在注重平均表現偏好下，單一方向平均每車停等延滯減少 1.4 秒，每小時整體路口增加 209 PCU。

表 4.3.2.4-8 平日離峰 TOD 及本期 AI 強化學習模型模擬注重平均表現績效彙整表

次數	TOD						AI 強化學習模型					
	單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	幹道每小時 單一車道		支道每小時 單一車道		單一方 向平均 每車停 等延滯	每小時 整體路 口 PCU	幹道每小時 單一車道		支道每小時 單一車道	
			停等 延滯	通過量	停等 延滯	通過量			停等 延滯	通過量	停等 延滯	通過量
1	34.5	15502	9515.8	517	17303.1	397	32.2	15714	10424.8	519	17733.3	394
2	35.6	15317	8732.6	515	23211.7	382	33.7	15518	9882.6	512	21960.9	393
3	34.9	15550	8895.4	520	20402.5	386	32.0	15803	9613.3	535	22085.1	379
4	33.5	15865	9089.6	543	18619.3	399	31.7	16182	9986.7	539	17516.2	400
5	31.6	15698	9084.4	534	18866.6	395	31.3	15659	9569.5	530	18148.5	398
6	31.0	15827	8927.9	537	16686.1	403	33.1	15924	9874.7	528	17079.2	397
7	31.6	15654	9041.3	519	16577.6	397	29.9	16079	9958.5	526	15630.6	396
8	32.6	15440	9425.0	521	17203.4	392	31.6	15535	9985.9	520	16377.6	389
9	38.6	16070	9351.1	539	16697.3	388	34.8	16388	10283.6	544	17290.2	386
10	34.5	15590	9293.1	521	17996.5	389	33.5	15805	10353.3	522	17025.6	396
μ	33.8	15651	9135.6	527	18356.4	393	32.4	15861	9993.3	528	18084.7	393

註 1：停等延滯單位為秒，通過量單位為 PCU/hr，每一行為該時段兩小時模擬平均值， μ (平均值)總 20 小時之績效平均值。

註 2：本表為各 episode 之績效平均值



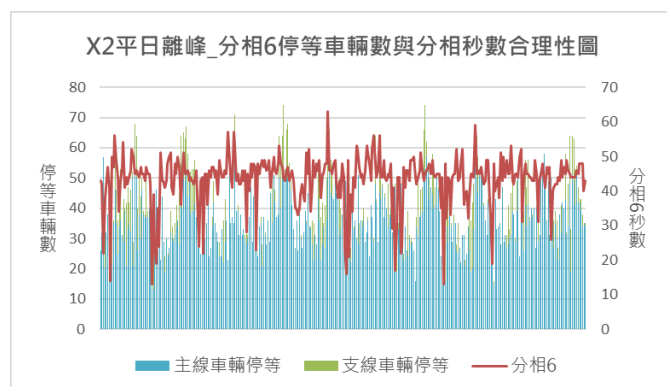
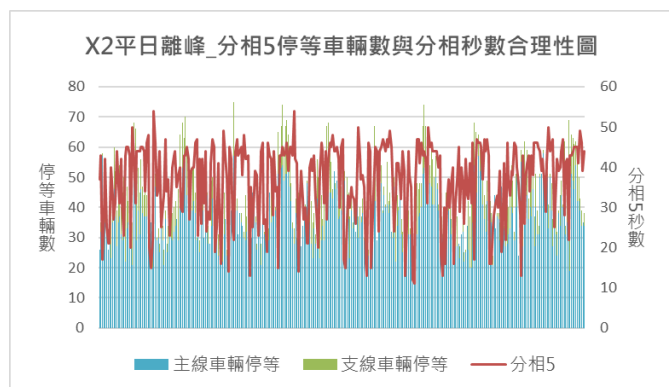


圖 4.3.2.4-4 中山北路/克強路平日離峰模擬合理性分析

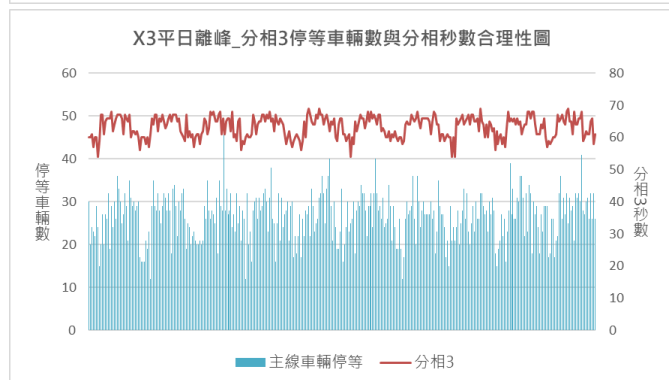
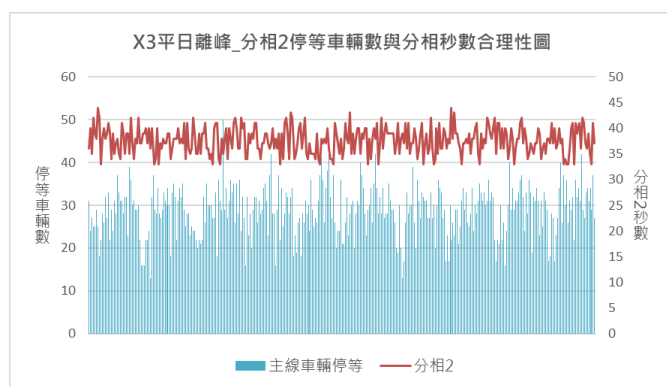
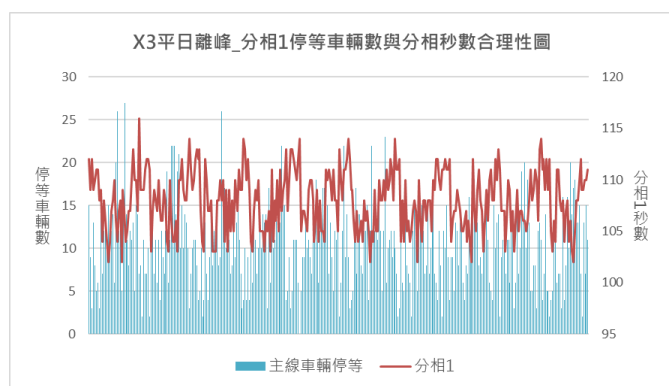


圖 4.3.2.4-5 中山北路/195 巷平日離峰模擬合理性分析

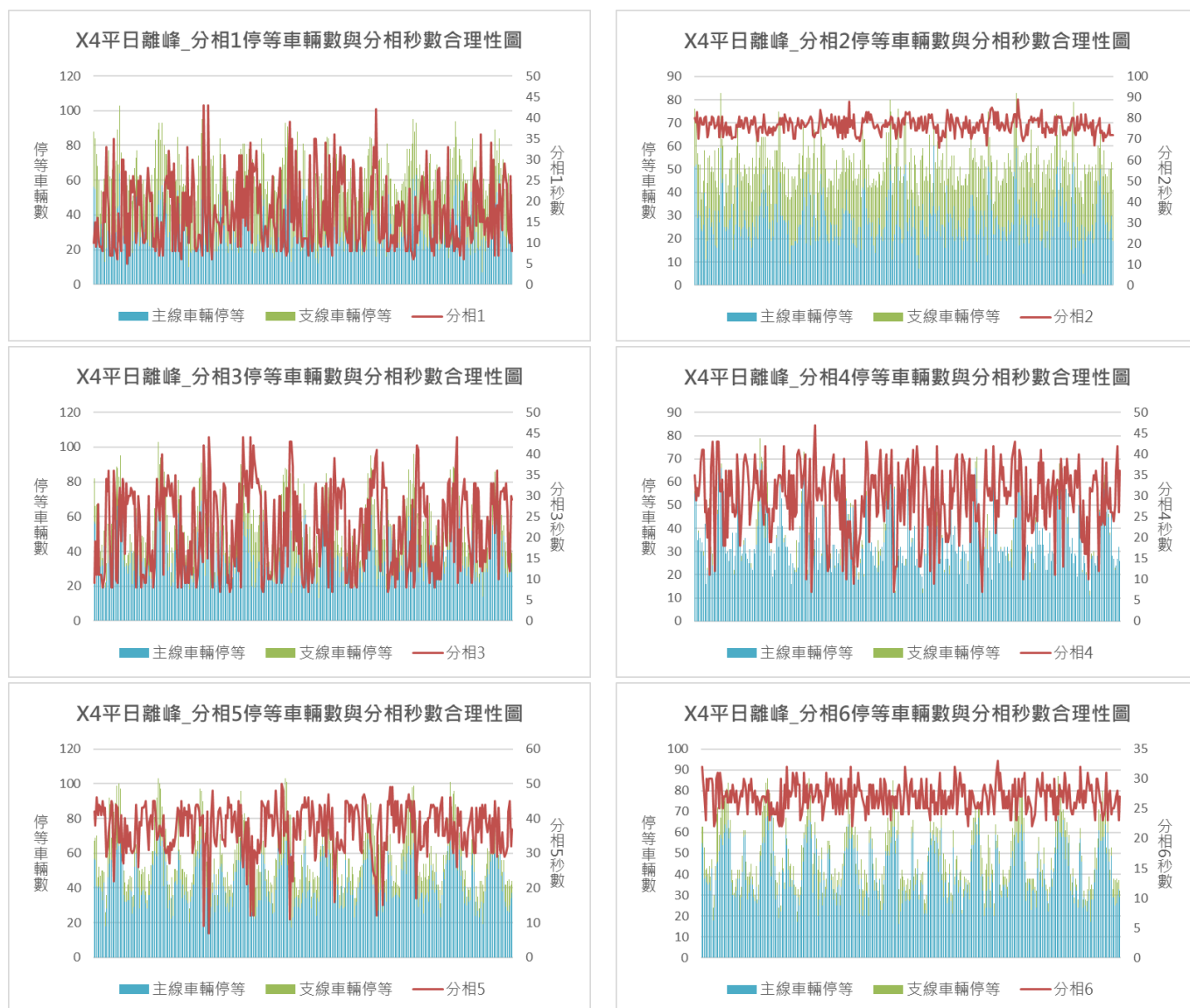


圖 4.3.2.4-6 中山北路/德行西路平日離峰模擬合理性分析

表 4.3.2.4-9 中山北路/克強路平日離峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.42	2	0.23	2	0.15	2	0.40	2	0.42
3	0.30	3	0.25	3	0.30	3	0.26	3	0.45
4	0.15	4	0.23	4	0.26	4	0.30	4	0.30
5	0.11	5	0.26	5	0.26	5	0.30	5	0.36
6	0.15	6	0.32	6	0.26	6	0.42	6	0.45
7	0.53	7	0.26	7	0.21	7	0.43	7	0.38
8	0.23	8	0.23	8	0.23	8	0.36	8	0.42

註：分相 4 為行人專用時相，配給固定秒數

表 4.3.2.4-10 中山北路/克強路平日離峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	38	10.18	12.68	0.95
	1	20	3.79	11.65	0.95
	2	24	3.04	9.75	0.92
	3	15	4.69	11.60	0.87
	4	17	5.00	10.71	0.94
	5	16	7.49	13.69	0.94
	6	25	4.54	12.44	0.92
分相 2	0	33	7.34	85.55	0.97
	1	26	8.14	90.04	0.96
	2	33	9.78	90.76	1.00
	3	25	6.29	89.76	1.00
	4	15	6.83	92.13	1.00
	5	23	9.40	90.22	0.91
分相 3	0	51	3.26	22.73	0.98
	1	62	2.98	22.31	0.95
	2	42	3.39	22.48	0.95
分相 5	0	19	9.74	33.21	1.00
	1	24	7.41	39.58	0.96
	2	14	8.48	38.57	1.00
	3	31	9.90	36.16	0.97
	4	13	10.27	40.31	1.00
	5	27	9.50	32.78	1.00
	6	27	10.37	33.67	1.00
分相 6	0	65	10.59	42.89	0.94
	1	55	8.81	40.35	0.95
	2	35	10.89	37.91	0.94

表 4.3.2.4-11 中山北路/195 巷平日離峰模擬穩定性分析

分相 1		分相 2	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.30	2	0.36
3	0.38	3	0.60
4	0.34	4	0.70
5	0.25	5	0.55
6	0.36	6	0.64
7	0.34	7	0.64
8	0.34	8	0.60

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	50	2.79	105.08	0.94
	1	40	2.74	107.98	0.98
	2	65	3.05	106.42	0.97
分相 2	0	21	2.92	38.86	0.90
	1	15	2.81	38.20	1.00
	2	67	2.64	36.82	1.00
	3	52	2.47	37.71	0.96

註：分相 3 為行人專用時相，配給固定秒數

表 4.3.2.4-12 中山北路/德行西路平日離峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 4		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.53	2	0.32	2	0.19	2	0.30	2	0.36	2	0.77
3	0.38	3	0.40	3	0.28	3	0.28	3	0.36	3	0.64
4	0.34	4	0.32	4	0.30	4	0.23	4	0.21	4	0.45
5	0.11	5	0.25	5	0.26	5	0.42	5	0.28	5	0.38
6	0.32	6	0.38	6	0.28	6	0.26	6	0.17	6	0.81
7	0.30	7	0.25	7	0.34	7	0.38	7	0.42	7	0.58
8	0.15	8	0.26	8	0.23	8	0.25	8	0.32	8	0.55

表 4.3.2.4-13 中山北路/德行西路平日離峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	74	3.26	7.93	0.95
	1	81	3.57	8.02	0.95
分相 2	0	41	2.43	79.54	0.93
	1	72	2.67	78.56	0.94
	2	42	2.17	79.24	0.95
分相 3	0	24	2.59	8.08	0
	1	24	4.67	10.38	1
	2	18	1.38	7.56	2
	3	27	2.93	9.15	3
	4	24	3.19	8.92	4
	5	18	5.33	10.06	5
	6	20	5.97	10.45	6
分相 4	0	28	5.77	41.39	0.93
	1	25	8.23	38.80	0.96
	2	43	5.19	40.49	0.93
	3	20	4.86	40.45	0.95
	4	39	4.51	41.05	0.95
分相 5	0	25	1.64	44.44	0.92
	1	13	3.20	43.38	0.92
	2	25	4.36	42.64	0.92
	3	29	1.74	43.34	1.00
	4	18	1.67	43.72	0.89
	5	23	1.91	45.13	0.96
	6	22	2.97	43.41	0.95
分相 6	0	38	2.28	27.71	1.00
	1	32	2.04	28.66	1.00
	2	15	2.25	26.73	1.00
	3	28	2.13	27.57	0.96
	4	13	2.06	28.62	1.00
	5	29	1.91	28.17	1.00

(三)平日昏峰(16：30~18：30)

1.秒數分布

AI 模擬週期秒數趨勢集中約 239 秒，高於 TOD 的 200 秒。

(1)中山北路/克強路路口(X2)

各分相秒數平均趨勢分別為分相一的 10 秒、分相二的 103 秒、分相三的 20 秒、分相四的 8 秒、分相五的 39 秒及分相六的 58 秒。

(2)中山北路/195 巷路口(X3)

各分相秒數平均趨勢分別為分相一的 100 秒、分相二的 82 秒及分相三的 58 秒。

(3)中山北路/德行西路路口(X4)

各分相秒數趨勢分別為分相一的 10 秒、分相二的 103 秒、分相三的 20 秒、分相四的 8 秒、分相五的 39 秒及分相六的 58 秒。

2.績效

在注重主幹道表現之偏好下，單一方向平均每車停等延滯增加 3.8 秒，每小時整體路口增加 262 PCU。

表 4.3.2.4-14 平日昏峰模擬注重主幹道表現績效彙整表

次數	TOD						AI 強化學習模型					
	單一方向平均每車停等延滯	每小時整體路口 PCU	幹道每小時單一車道		支道每小時單一車道		單一方向平均每車停等延滯	每小時整體路口 PCU	幹道每小時單一車道		支道每小時單一車道	
			停等延滯	通過量	停等延滯	通過量			停等延滯	通過量	停等延滯	通過量
1	25.8	18260	12788.1	680	14713.8	581	33.4	18533	12514.8	702	16092.8	573
2	25.6	18056	12216.5	655	15057.9	571	39.6	18139	12140.3	650	15979.9	572
3	27.8	18263	12351.9	652	15590.6	580	24.4	18579	11670.3	677	17383.9	573
4	29.9	17854	12453.1	658	17796.7	577	33.0	18041	12503.0	662	17828.6	579
5	30.9	18253	12548.5	650	24276.0	575	30.6	18322	12255.0	654	23209.8	537
6	25.2	17816	12197.5	674	13381.4	574	26.4	18215	11953.2	690	13106.6	574
7	25.9	17769	12396.7	662	14761.9	579	35.4	17786	12719.5	653	16070.4	575
8	25.8	18053	12164.8	657	15844.4	584	27.2	18432	12186.0	680	17695.9	588
9	27.6	17924	12359.4	677	16160.2	593	31.9	18352	12386.3	692	16160.9	589
10	27.0	17467	12526.4	656	14633.8	576	27.7	17938	12412.2	672	16309.1	564
μ	27.2	17972	12400.3	662	16221.7	579	31.0	18234	12274.0	673	16983.8	572

註 1：停等延滯單位為秒，通過量單位為 PCU/hr，每一行為該時段兩小時模擬平均值，μ(平均值)總 20 小時之績效平均值。

註 2：本表為各 episode 之績效平均值

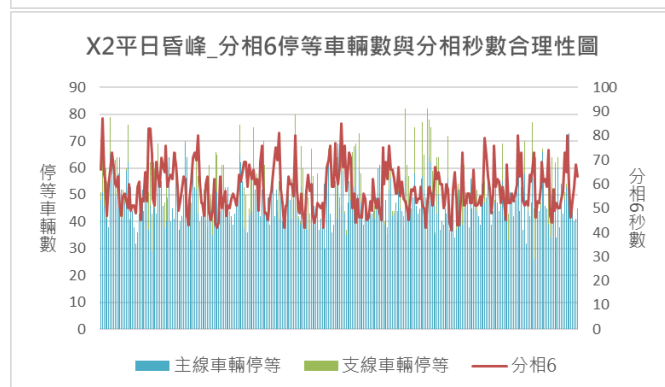
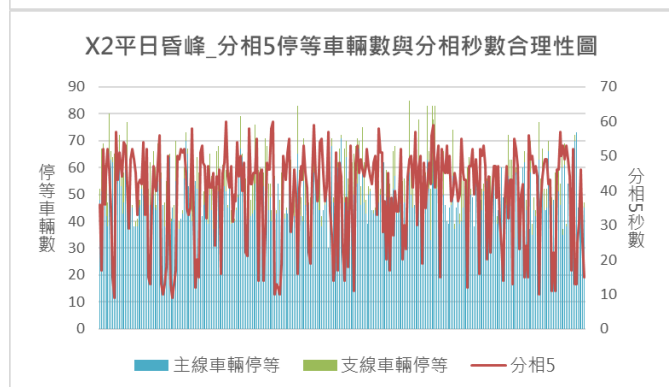
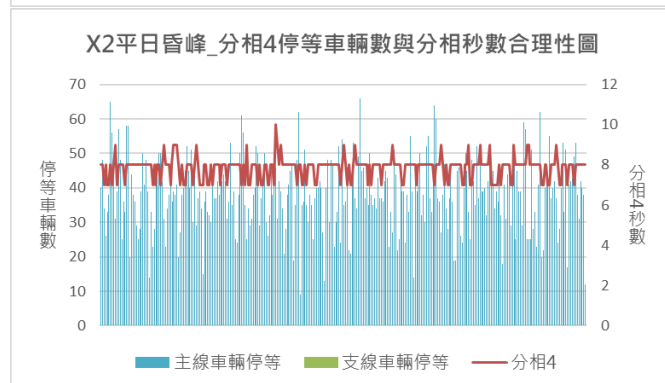
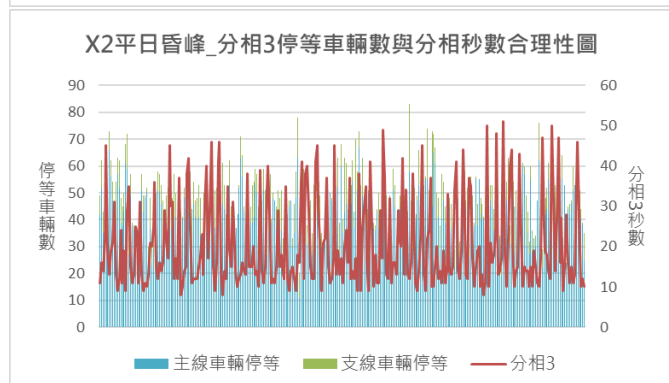
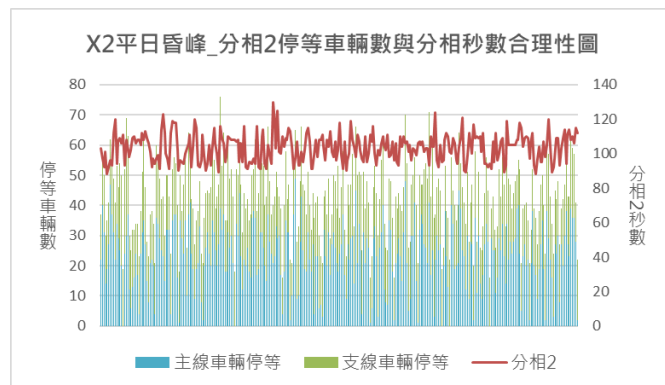
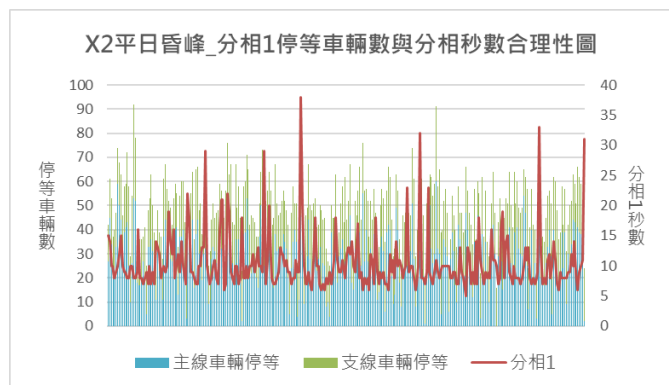
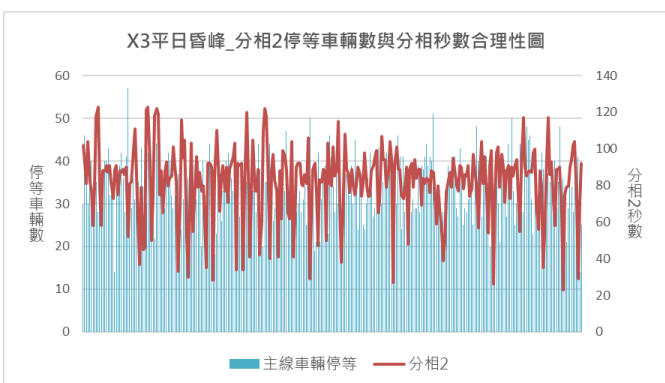
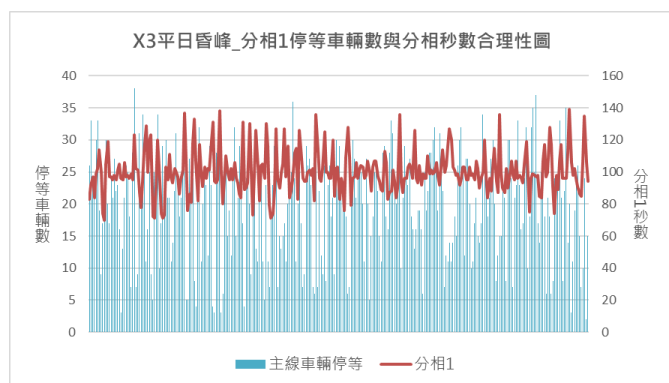


圖 4.3.2.4-7 中山北路/克強路平日昏峰模擬合理性分析



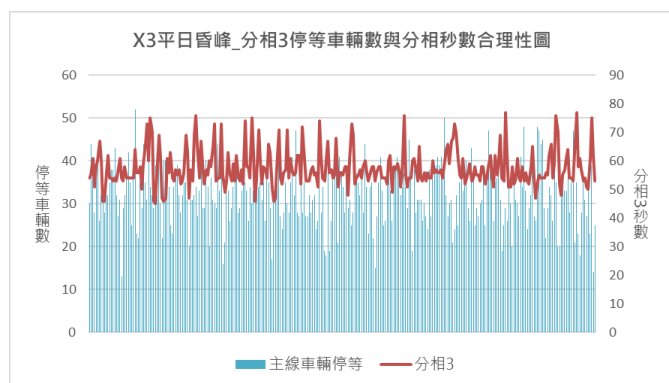


圖 4.3.2.4-8 中山北路/195 巷平日昏峰模擬合理性分析

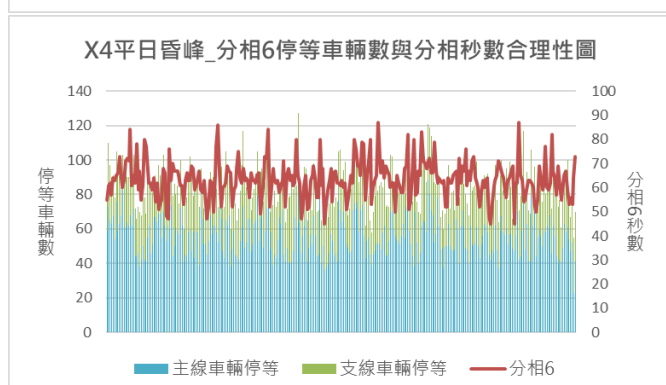
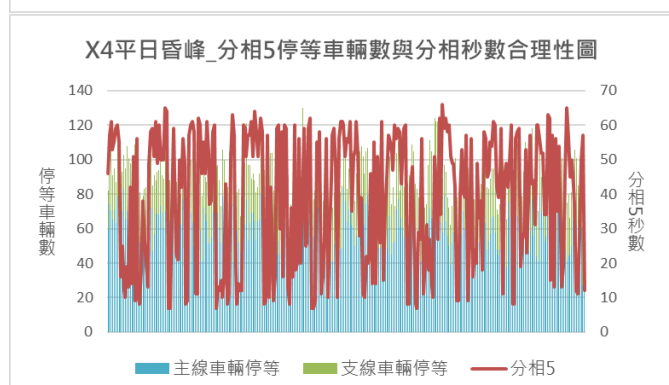
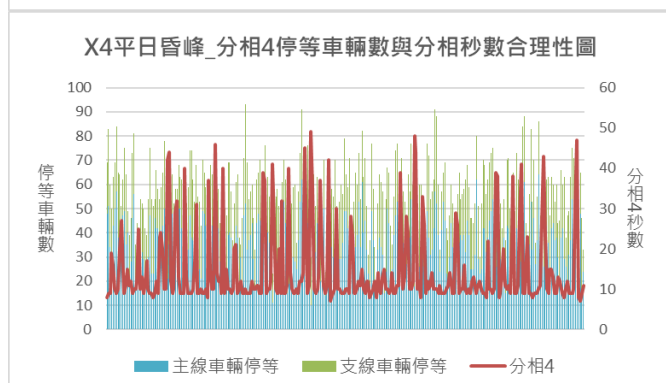
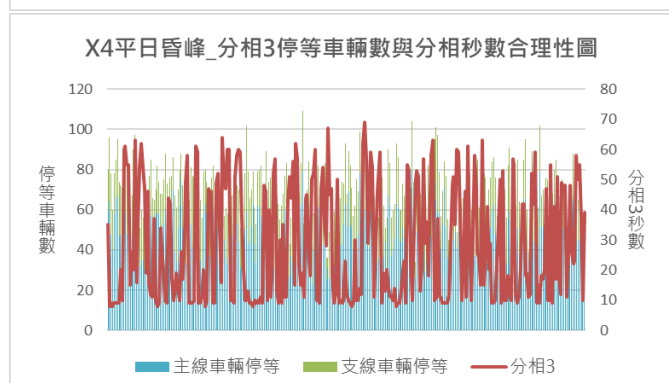
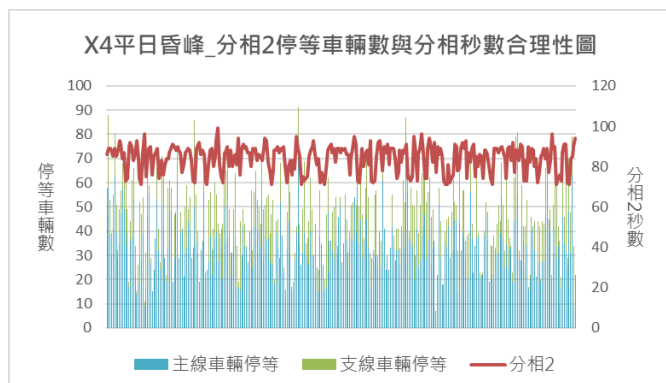
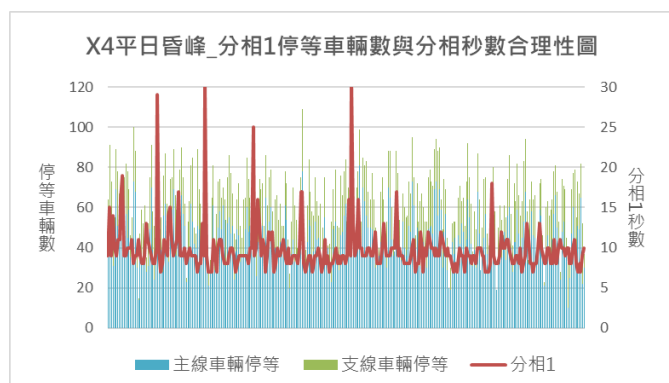


圖 4.3.2.4-9 中山北路/德行西路平日昏峰模擬合理性分析

表 4.3.2.4-15 中山北路/克強路平日昏峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.28	2	0.45	2	0.36	2	0.25	2	0.19
3	0.17	3	0.28	3	0.43	3	0.25	3	0.25
4	0.45	4	0.21	4	0.36	4	0.34	4	0.30
5	0.28	5	0.25	5	0.38	5	0.26	5	0.28
6	0.32	6	0.36	6	0.38	6	0.34	6	0.30
7	0.30	7	0.34	7	0.42	7	0.25	7	0.36
8	0.25	8	0.40	8	0.40	8	0.23	8	0.34

註：分相 4 為行人專用時相，配給固定秒數

表 4.3.2.4-16 中山北路/克強路平日昏峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	18	1.81	8.33	0.89
	1	25	1.27	9.04	0.96
	2	32	2.06	7.63	0.97
	3	30	1.16	8.37	0.93
分相 2	0	55	7.83	84.58	0.96
	1	50	12.19	80.42	0.92
分相 3	0	51	3.26	22.73	0.98
	1	62	2.98	22.31	0.95
	2	42	3.39	22.48	0.95
分相 5	0	11	12.81	31.82	1.00
	1	37	13.11	27.41	1.00
	2	32	11.58	36.47	0.94
	3	25	11.67	27.72	0.96
分相 6	0	13	5.42	54.77	1.00
	1	14	8.64	49.79	0.93
	2	17	8.61	52.59	0.94
	3	26	5.53	54.12	0.92
	4	9	10.32	49.22	1.00
	5	15	5.57	53.20	1.00
	6	11	9.98	51.09	0.91

表 4.3.2.4-17 中山北路/195 巷平日昏峰模擬穩定性分析

分相 1		分相 2	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.42	2	0.26
3	0.36	3	0.30
4	0.21	4	0.28
5	0.36	5	0.49
6	0.36	6	0.30
7	0.40	7	0.23
8	0.21	8	0.34

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	51	13.86	97.76	0.92
	1	54	7.18	103.30	0.98
分相 2	0	22	4.90	34.09	0.95
	1	19	5.19	35.05	0.95
	2	24	2.28	35.21	1.00
	3	8	2.59	36.88	1.00
	4	32	5.39	34.56	0.94

註：分相 3 為行人專用時相，配給固定秒數

表 4.3.2.4-18 中山北路/德行西路平日昏峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 4		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.32	2	0.34	2	0.45	2	0.42	2	0.28	2	0.28
3	0.26	3	0.26	3	0.36	3	0.53	3	0.34	3	0.23
4	0.15	4	0.43	4	0.55	4	0.19	4	0.23	4	0.32
5	0.23	5	0.25	5	0.25	5	0.34	5	0.21	5	0.36
6	0.30	6	0.26	6	0.17	6	0.36	6	0.17	6	0.38
7	0.32	7	0.17	7	0.36	7	0.36	7	0.26	7	0.17
8	0.30	8	0.09	8	0.23	8	0.42	8	0.32	8	0.19

表 4.3.2.4-19 中山北路/德行西路平日昏峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	48	0.85	8.85	0.96
	1	57	1.07	8.18	0.93
分相 2	0	8	2.12	89.25	0.88
	1	31	15.77	80.35	0.90
	2	38	6.19	86.58	0.97
	3	28	9.48	86.39	0.96
分相 3	0	10	0.48	9.30	1.00
	1	26	1.26	7.92	0.88
	2	40	0.79	8.70	0.95
	3	29	0.86	8.66	0.93
分相 4	0	49	7.44	9.80	0.96
	1	47	4.83	9.72	0.98
	2	9	0.44	9.22	1.00
分相 5	0	21	6.73	59.05	0.90
	1	37	3.22	61.35	0.97
	2	47	7.17	59.02	0.91
分相 6	0	18	2.23	30.83	1.00
	1	32	3.97	29.88	0.97
	2	8	6.32	27.00	0.88
	3	11	6.54	29.18	0.91
	4	13	2.69	30.62	1.00
	5	23	4.66	29.39	0.91

(四)假日昏峰(16:30~18:30)

1.秒數分布

AI 模擬週期秒數趨勢集中約 209 秒，高於 TOD 的 200 秒。

(1)中山北路/克強路路口(X2)

各分相秒數平均趨勢分別為分相一的 9 秒、分相二的 105 秒、分相三的 27 秒、分相四的 8 秒、分相五 36 的秒及分相六的 24 秒。

(2)中山北路/195 巷路口(X3)

各分相秒數平均趨勢分別為分相一的 122 秒、分相二的 43 秒及分相三的

44 秒。

(3)中山北路/德行西路路口(X4)

各分相秒數趨勢分別為分相一的 16 秒、分相二的 76 秒、分相三的 19 秒、分相四的 35 秒、分相五的 37 秒及分相六的 27 秒。

2.績效

在注重通過量表現之偏好下，單一方向平均每車停等延滯略增加 3.6 秒，每小時整體路口增加 208 PCU。

表 4.3.2.4-20 假日昏峰模擬注重通過量表現績效彙整表

次數	TOD						AI 強化學習模型					
	單一方向平均每車停等延滯	每小時整體路口 PCU	幹道每小時單一車道		支道每小時單一車道		單一方向平均每車停等延滯	每小時整體路口 PCU	幹道每小時單一車道		支道每小時單一車道	
			停等延滯	通過量	停等延滯	通過量			停等延滯	通過量	停等延滯	通過量
1	21.1	17151	8868.3	634	12981.6	494	24.7	17194	9355.5	640	16950.3	463
2	21.3	17030	8825.2	624	12684.5	481	24.8	17189	9017.6	629	16923.3	454
3	25.8	17419	9521.0	620	17222.3	479	29.3	17568	10485.5	626	18554.8	468
4	23.3	17439	9574.5	638	15702.3	475	28.6	17496	9979.4	638	22104.6	425
5	26.0	17221	9700.5	614	17940.8	476	31.9	17560	9734.3	637	23958.8	452
6	21.7	16745	8668.4	630	12918.1	491	22.9	17216	9040.7	637	14330.9	480
7	22.1	17250	9168.4	631	13593.1	486	28.7	17454	9195.1	642	19079.4	451
8	27.2	17047	9137.9	622	22653.0	466	25.5	17365	9430.3	631	18858.4	464
9	24.2	17139	9533.0	622	13864.1	507	29.7	17373	9701.5	630	21231.2	469
10	23.1	17327	8860.9	631	13741.4	492	25.9	17434	9761.2	631	16967.1	473
μ	23.6	17177	9185.8	627	15330.1	485	27.2	17385	9570.1	634	18895.9	460

註 1：停等延滯單位為秒，通過量單位為 PCU/hr，為該時段兩小時模擬平均值。

註 2：本表為各 episode 之績效平均值

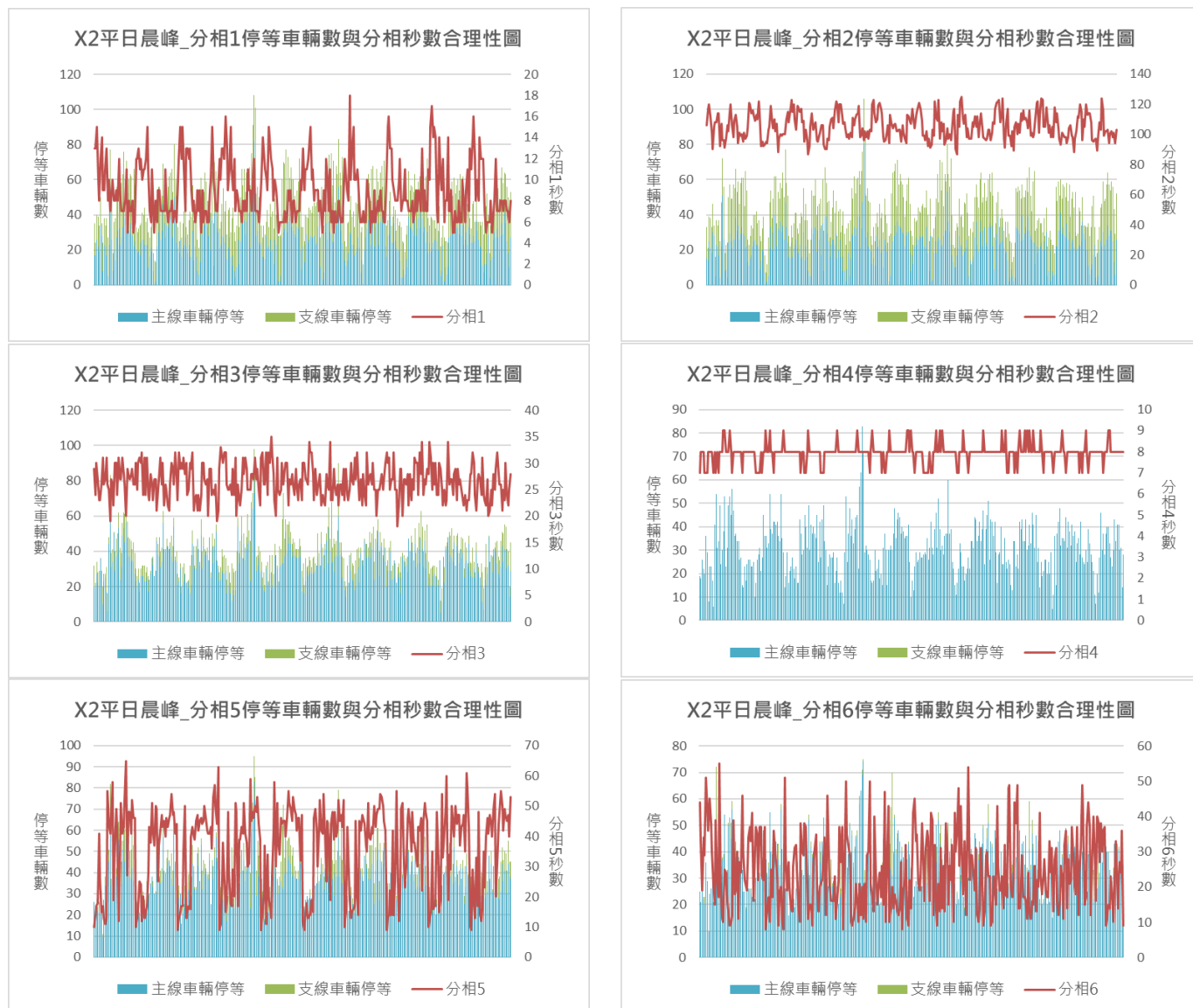
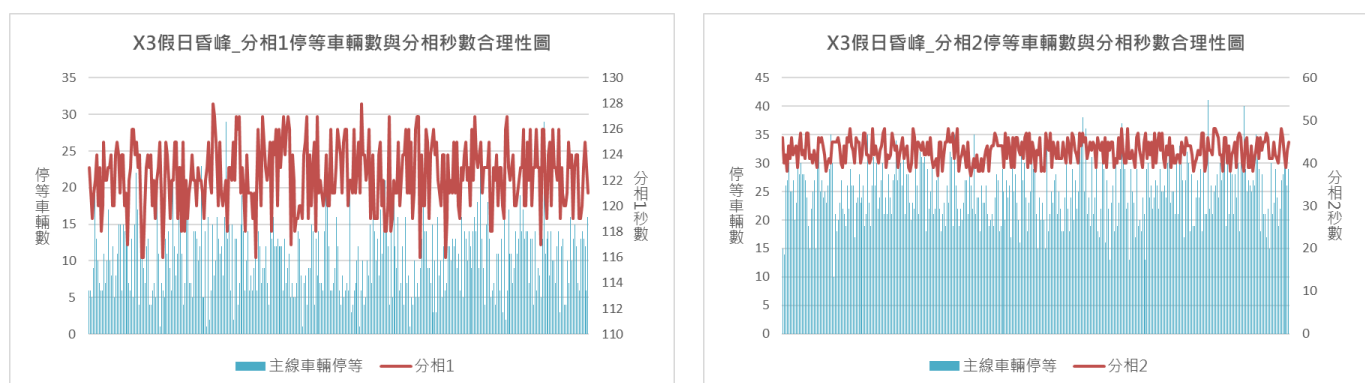


圖 4.3.2.4-10 中山北路/克強路假日昏峰模擬合理性分析



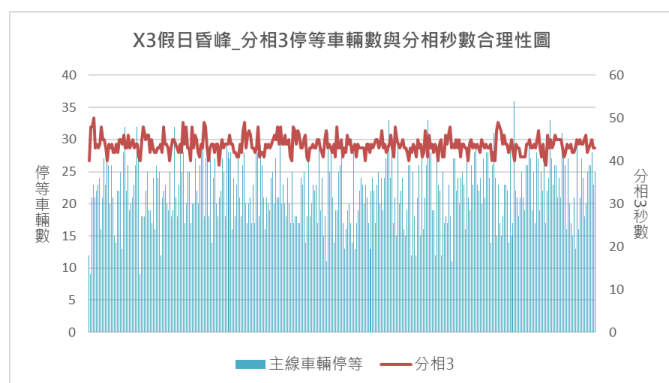


圖 4.3.2.4-11 中山北路/195 巷假日昏峰模擬合理性分析

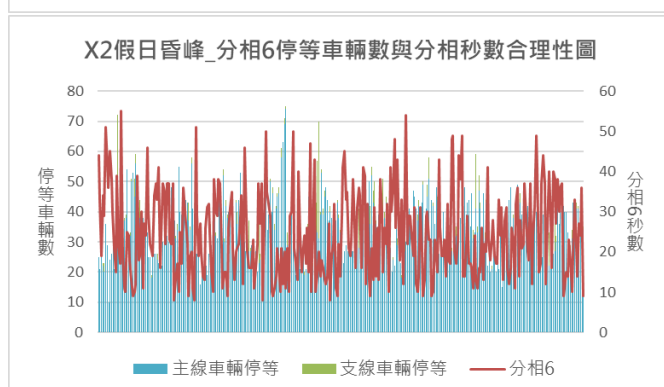
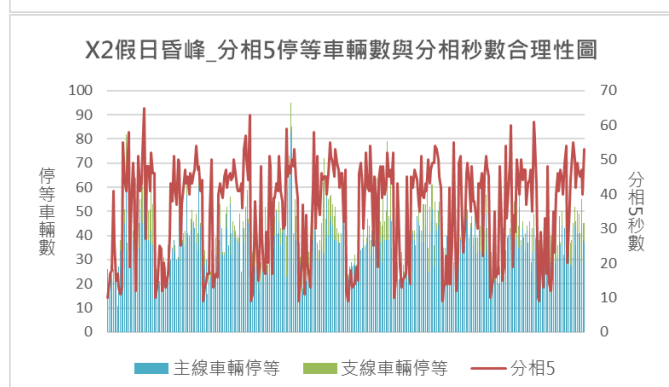
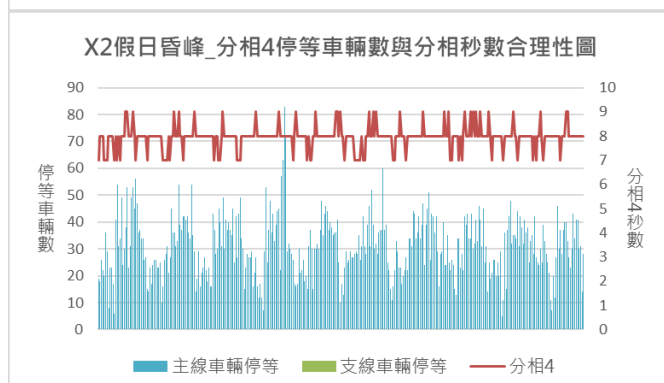
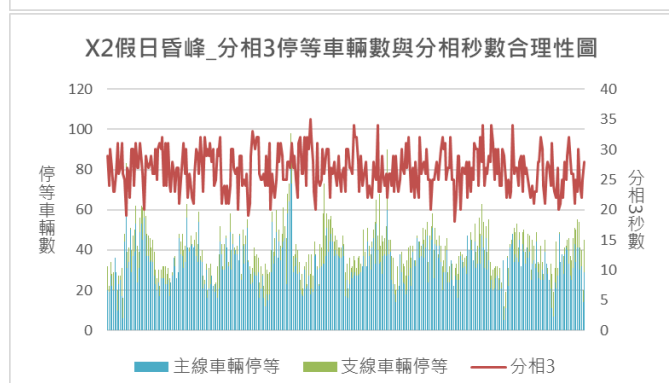
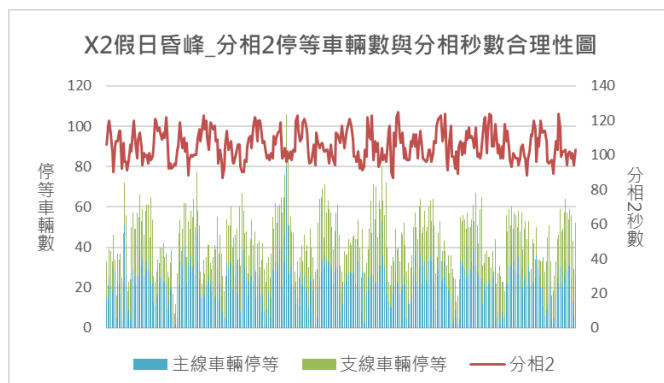
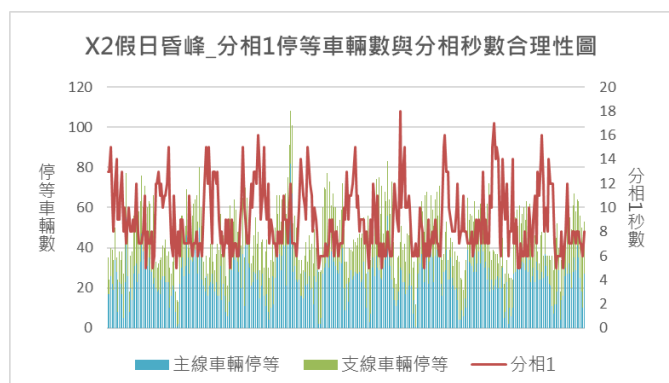


圖 4.3.2.4-12 中山北路/德行西路假日昏峰模擬合理性分析

表 4.3.2.4-21 中山北路/克強路假日昏峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 5		分相 6	
群集 編號	變異數 分析具 有顯著 水準特 徵比	群集 編號	變異數 分析具 有顯著 水準特 徵比	群集 編號	變異數 分析具 有顯著 水準特 徵比	群集 編號	變異數 分析具 有顯著 水準特 徵比	群集 編號	變異數 分析具 有顯著 水準特 徵比
2	0.23	2	0.15	2	0.21	2	0.21	2	0.28
3	0.23	3	0.17	3	0.13	3	0.28	3	0.15
4	0.26	4	0.19	4	0.15	4	0.15	4	0.23
5	0.15	5	0.17	5	0.25	5	0.21	5	0.11
6	0.25	6	0.19	6	0.13	6	0.21	6	0.30
7	0.26	7	0.21	7	0.25	7	0.25	7	0.13
8	0.23	8	0.21	8	0.15	8	0.19	8	0.13

註：分相 4 為行人專用時相，配給固定秒數

表 4.3.2.4-22 中山北路/克強路假日昏峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	7	23.58	24.57	1.00
	1	6	17.06	30.67	1.00
	2	17	10.01	17.00	1.00
	3	6	23.22	20.83	1.00
分相 2	0	5	9.53	104.40	1.00
	1	9	7.90	102.11	1.00
	2	9	5.31	99.22	0.89
	3	6	5.71	101.17	1.00
	4	1		92.00	1.00
	5	4	6.68	110.00	1.00
	6	2	0.71	99.50	1.00
分相 3	0	2	2.83	27.00	1.00
	1	5	3.39	25.00	1.00
	2	13	3.75	26.38	1.00
	3	6	4.68	25.50	1.00
	4	2	2.12	23.50	1.00
	5	3	2.08	26.33	1.00
	6	5	3.51	23.60	1.00
分相 5	0	6	1.21	12.67	1.00
	1	3	1.15	12.33	1.00
	2	3	0.00	12.00	1.00
	3	5	0.45	12.80	1.00
	4	7	0.82	13.00	1.00
	5	2	0.00	12.00	1.00
	6	10	1.17	12.60	0.90
分相 6	0	7	14.58	46.29	1.00
	1	15	18.59	38.40	1.00
	2	3	20.88	36.00	1.00
	3	3	20.13	38.33	1.00
	4	5	18.31	46.60	1.00
	5	3	15.87	21.00	1.00

表 4.3.2.4-23 中山北路/195 巷假日昏峰模擬穩定性分析

分相 1		分相 2	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.13	2	0.21
3	0.17	3	0.28
4	0.15	4	0.23
5	0.19	5	0.15
6	0.06	6	0.11
7	0.21	7	0.25
8	0.13	8	0.23

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	4	1.00	136.50	1.00
	1	7	2.06	135.71	1.00
	2	11	1.19	135.73	1.00
	3	5	0.89	134.40	1.00
	4	5	1.10	135.20	1.00
	5	3	0.58	135.67	1.00
	6	1	-	136	1
分相 2	0	12	1.51	24.42	1.00
	1	11	1.54	23.82	1.00
	2	13	1.70	23.31	1.00

註：分相 3 為行人專用時相，配給固定秒數

表 4.3.2.4-24 中山北路/德行西路假日昏峰模擬穩定性分析 1

分相 1		分相 2		分相 3		分相 4		分相 5		分相 6	
群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比	群集編號	變異數分析具有顯著水準特徵比
2	0.21	2	0.13	2	0.19	2	0.17	2	0.25	2	0.23
3	0.30	3	0.11	3	0.13	3	0.15	3	0.11	3	0.17
4	0.15	4	0.15	4	0.13	4	0.15	4	0.19	4	0.25
5	0.21	5	0.08	5	0.25	5	0.11	5	0.30	5	0.13
6	0.17	6	0.13	6	0.06	6	0.25	6	0.26	6	0.26
7	0.26	7	0.06	7	0.09	7	0.17	7	0.13	7	0.23
8	0.11	8	0.15	8	0.15	8	0.17	8	0.38	8	0.19

表 4.3.2.4-25 中山北路/德行西路假日昏峰模擬穩定性分析 2

	集群編號	資料數量	標準差	平均數	$\mu \pm 2\sigma$
分相 1	0	13	3.36	6.46	0.92
	1	13	9.83	10.15	0.85
	2	10	0.84	6.40	1.00
分相 2	0	10	3.37	68.70	1.00
	1	13	4.07	68.23	0.92
	2	6	2.23	66.83	1.00
	3	7	5.09	69.57	1.00
分相 3	0	6	5.05	37.67	1.00
	1	9	6.04	36.44	1.00
	2	2	2.12	32.50	1.00
	3	15	7.15	31.13	0.93
	4	4	6.48	39.00	1.00
分相 4	0	5	7.80	31.40	1.00
	1	7	11.05	24.14	1.00
	2	9	4.18	30.33	1.00
	3	2	7.07	28.00	1.00
	4	5	10.26	25.80	1.00
	5	8	8.50	23.00	1.00
分相 5	0	2	6.36	46.50	1.00
	1	2	3.54	48.50	1.00
	2	9	3.49	45.22	1.00
	3	5	3.58	46.60	1.00
	4	5	2.24	50.00	1.00
	5	2	2.12	45.50	1.00
	6	4	7.07	52.00	1.00
	7	7	5.32	47.43	1.00
分相 6	0	10	2.39	24.20	1.00
	1	9	1.94	24.44	1.00
	2	5	1.79	23.20	1.00
	3	6	1.90	24.00	1.00
	4	2	2.12	24.50	1.00
	5	4	2.06	22.25	1.00

本計畫訓練之 AI 強化學習模型使用車流模擬軟體測試，經上述合理性及穩定性公式檢視，各時段產出結果趨勢近乎一致，因此具參考價值，可施行實作。以表 4.3.2.4-26 至 4.3.2.4-28 為各路口 AI 號誌控制模型平均秒數與 TOD 各分相內容對照表

表 4.3.2.4-26 克強路與中山北路口 AI 強化學習模型與 TOD 的各分相平均秒數對照表

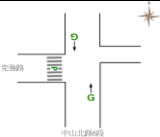
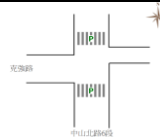
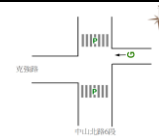
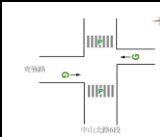
週期 秒數	分相 1	分相 2	分相 3	分相 4	分相 5	分相 6
						
200 秒	9 秒	107 秒	12 秒	9 秒	14 秒	49 秒
209 秒	9 秒	105 秒	27 秒	8 秒	36 秒	24 秒
200 秒	9 秒	106 秒	12 秒	9 秒	14 秒	50 秒
209 秒	9 秒	89 秒	22 秒	8 秒	37 秒	43 秒
200 秒	9 秒	99 秒	16 秒	9 秒	14 秒	54 秒
238 秒	10 秒	103 秒	20 秒	8 秒	39 秒	58 秒
200 秒	9 秒	107 秒	12 秒	9 秒	14 秒	49 秒
209 秒	9 秒	105 秒	27 秒	8 秒	36 秒	24 秒

表 4.3.2.4-27 195 巷與中山北路口 AI 強化學習模型與 TOD 的各分相平均秒數對照表

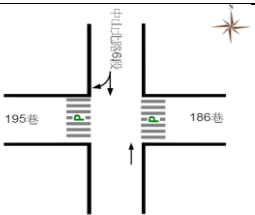
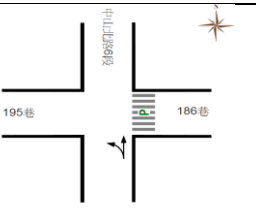
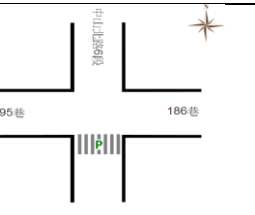
	週期秒數	分相 1	分相 2	分相 3
分相圖				
平日晨峰 TOD	200 秒	140 秒	25 秒	35 秒
平日晨峰 AI	209 秒	121 秒	42 秒	46 秒
平日離峰 TOD	200 秒	120 秒	25 秒	55 秒
平日離峰 AI	209 秒	108 秒	38 秒	63 秒
平日昏峰 TOD	200 秒	120 秒	25 秒	35 秒
平日昏峰 AI	239 秒	100 秒	82 秒	58 秒
假日昏峰 TOD	200 秒	140 秒	25 秒	35 秒
假日昏峰 AI	209 秒	122 秒	43 秒	44 秒

表 4.3.2.4-28 德行西路與中山北路口 AI 強化學習模型與 TOD 的各分相平均秒數對照表

週 期 秒 數	分相 1	分相 2	分相 3	分相 4	分相 5	分相 6
200	10 秒	89 秒	10 秒	23 秒	45 秒	23 秒
209	18 秒	73 秒	16 秒	37 秒	40 秒	25 秒
200	15 秒	92 秒	15 秒	15 秒	45 秒	18 秒
209	17 秒	77 秒	22 秒	29 秒	37 秒	27 秒
200	10 秒	88 秒	15 秒	7 秒	55 秒	25 秒
238	10 秒	103 秒	20 秒	8 秒	39 秒	58 秒
200	10 秒	91 秒	10 秒	15 秒	50 秒	24 秒
210	16 秒	76 秒	19 秒	35 秒	37 秒	27 秒

4.3.3 實測計畫

一、臺南市實驗場域計畫

(一)實測日期

平假日各兩天，平日為 11 月 7 日(二)、11 月 8 日(三)，假日為 11 月 4 日(六)、11 月 5 日(日)。

(二)實測時段

- 1.平日上午尖峰：07：00 至 09：00。
- 2.平日下午離峰：12：30 至 14：30。
- 3.平日下午尖峰：16：30 至 18：30。
- 4.假日尖峰：16：30 至 18：30。

(三)實測地點：臺南市關廟區「台 86-台 19 甲線」路口。

(四)實測指標

- 1.通過量：使用 AI 影像辨識取得通過量資料。
- 2.停等延滯：使用 GVP 資料進行推估。

二、臺北市實驗場域計畫

(一)實測日期

平假日各兩天，11 月 16 日(四)、11 月 18 日(六)、11 月 19 日(日)、11 月 20 日(一)。

(二)實測時段

- 1.平日上午尖峰：07：00 至 09：00。
- 2.平日下午離峰：11 月 16 日(四)12：30 至 14：30，11 月 20 日(一)13：00 至 15：00。
- 3.平日下午尖峰：16：30 至 18：30。
- 4.假日尖峰：16：30 至 18：30。

(三)實測地點：臺北市「中山北路/克強路」、「中山北路/195 巷」、「中山北路/德行東路」3 個路口幹道。

(四)實測指標

- 1.通過量：使用車輛偵測器蒐集資料。
- 2.停等延滯：使用車輛偵測器蒐集資料。
- 3.旅行時間：使用 GVP 資料進行推估。

三、實測流程

操作方式如以下設定內容：

(一)設定人工智慧號誌控制排程

- 1.進入智慧號誌控制監控管理畫面，點選系統管理。
- 2.設定排程開始與結束之日期、時間、週期後點選確認。
- 3.於排程預覽確認欲設定之排程，確認無誤點選儲存。

(二)手動開啟/關閉

- 1.進入智慧號誌控制監控管理畫面，點選系統管理。
- 2.於號誌控制即時運作狀態點選「是否手動開關」，即可開啟、關閉人工智慧號誌控制。

四、監控方式

為掌握系統運作狀況，實測初期將安排人力於現場觀察，確保號誌運作正常且不造成路況異常或惡化，同時安排人員監控管理畫面，監控內容與時程安排如下：

- (一)監看現場設備運作狀態，確認是否有設備斷線。
- (二)確認系統是否正常運作，若人工智慧號誌控制運作異常則回復原 TOD 控制。
- (三)確認人工智慧號誌控制指令是否有下達至現場號誌。

表 4.3.3-1 實測監控日程

日期	監控內容
第一天	1.現場監控：安排 1-2 人於實測時段全程在現場觀察號誌運作與車流狀況，即時回報現場情形，如遇例外事件或路況異常則手動操作回復 TOD 控制。 2.監控畫面：人工進行畫面監測，如遇上述監控內容異常則手動恢復 TOD 控制。 3.系統監測：系統自動監測例外事件，事件觸發則自動恢復 TOD 控制。
第二天	1.監控畫面：人工進行畫面監測，如遇上述監控內容異常則手動恢復 TOD 控制。 2.系統監測：系統自動監測例外事件，事件觸發則自動恢復 TOD 控制。
第三天	1.監控畫面：人工進行畫面監測，如遇上述監控內容異常則手動恢復 TOD 控制。 2.系統監測：系統自動監測例外事件，事件觸發則自動恢復 TOD 控制。
第四天	1.監控畫面：人工進行畫面監測，如遇上述監控內容異常則手動恢復 TOD 控制。 2.系統監測：系統自動監測例外事件，事件觸發則自動恢復 TOD 控制。

五、例外事件

號誌控制模組於實測期間持續偵測下表 4.3.3-2 之例外事件，當例外事件發生時將自動降階恢復原固定時制計畫，避免因號誌控制導致現場交通狀況混亂之情形。

表 4.3.3-2 實測例外事件一覽表

事件類型	狀態描述	因應方式
現場號誌閃光	現場號誌為綠衝突或時制閃光。	恢復原固定時制計畫，待號誌恢復，由人工智慧號誌控制重新取得控制權。
現場控燈	現場有人員直接手動進行號誌控制。	恢復原固定時制計畫，待現場控燈結束後，人工智慧號誌控制重新取得控制權。
車流偵測設備斷線	無法回傳資料，導致人工智慧無法運算輸出秒數。	恢復原固定時制計畫，待設備斷線排除，再由人工智慧號誌控制取得控制權。
號誌控制器斷線	號誌控制器斷線，導致人工智慧號誌控制指令無法下達。	恢復原固定時制計畫，待號誌控制器恢復連線，再由人工智慧號誌控制取得控制權。
號誌控制器異常	號誌控制器異常，無法正常執行步階秒數。	恢復原固定時制計畫，待異常排除，再由人工智慧號誌控制取得控制權。

(一)自動降階執行流程

- 1.暫停人工智慧號誌控制系統。
- 2.恢復原固定時制計畫。
- 3.3 個週期（8-12 分）後轉換至原固定時制計畫，轉換期間由號誌控制器依當下狀況自行運算進行秒數補償。

(二)人工智慧號誌控制重新取得控制權流程

- 1.系統偵測例外事件結束。
- 2.恢復人工智慧號誌控制。
- 3.轉換需 2 個週期（5-8 分），期間現場號誌依固定時制計畫運作。

4.3.4 實測情形

一、臺南市實驗場域

(一)平日實測

1.11 月 7 日(週二)

(1)晨峰

實測期間號誌控制器降階 3 次，經查原因應為網路問題導致 AI 號誌控制指令來不及於原 TOD 時相秒數內下達，設備無接收到指令而回復 TOD。觀察現場台 86 線往國 3 車隊較長，但可於 1 週期內順利紓解。



圖 4.3.4-1 台 86 線往國 3 方向車流狀況

(2)離峰：觀察現場車流狀況順暢，無特殊情形。

(3)昏峰：觀察現場台 86 線雙向明顯車多，但可於 1 週期內順利紓解。

2.11 月 8 日(週三)

(1)晨峰

07：48 時臺南市交控中心通知台 86 線出現回堵情形，觀察 AI 號誌控制計算秒數有因應台 86 線車流量而增加綠燈秒數，後續車流順利紓解。

(2)離峰：觀察現場車流狀況順暢，無特殊情形。

(3)昏峰：觀察各方向車流量皆湧現車潮，台 19 甲線之車流量明顯較前幾日多，約在 1-2 週期內順利紓解。

(二)假日實測

1.11 月 4 日(週六) 昏峰

觀察現場台 86 線往國 3、台 19 甲往台 86 線車多，下午五點後台 86 線車流增多，路口仍可於 1 週期紓解完畢。



圖 4.3.4-2 台 86 線往國 3 方向車流狀況

2.11 月 5 日(週日) 昏峰

實測期間號誌控制器降階 6 次，經查原因應為網路問題導致 AI 號誌控制指令來不及於原 TOD 時相秒數內下達，設備無接收到指令而回復 TOD。觀察實測開始時觀察國 3 下交流道明顯車多，車隊已回堵至下匝道處，17:00 開始台 86 線雙向車多，17:30 後車流開始減少。

二、臺北市實驗場域

(一)平日實測

1.11 月 15 日(週三) 晨峰

實測時段開始時發現車流偵測設備自凌晨開始斷線，確認程式端與設備端後，得知斷線原因為前日 CCTV 即時影像廠商拆機時不慎破壞車流偵測設備電路，直至當天下午 5 點修復，故當天實測暫停，調整平日實測日期為 11 月 16 日、11 月 20 日。

2.11 月 16 日(週四)

(1)晨峰

車流偵測設備於兩小時中斷線 4 次，為避免車流偵測設備頻繁斷線導致號誌控制器進入補償，影響現場車流，請工程師於昏峰前調整程式，若偵測設備斷線則直接給予 TOD 秒數之指令。

觀察早上 7 點時德行東路東往西方向車流、中山北路南往北方向左轉克強路需兩週期紓解，每一週期中山北路北往南方向皆會自忠誠路壅塞至德行東路口，8 點半開始車流減少。



圖 4.3.4-3 中山北路北往南方向「忠誠路-德行東路」

(2)離峰

離峰時段車流偵測設備與號誌控制器各斷線一次，觀察實測範圍車流順

暢，惟德行東路東往西方向車多，但一週期可紓解。

(3) 昏峰

實測前觀察實測範圍已車多壅塞，實測開始時號誌控制器斷線，直至 16:49 才順利啟動 AI 號誌控制，至結束前號誌控制器共斷線 5 次。觀察現場中山北路北往南方向車流明顯多於南往北方向，克強路右轉中山北路往南方向車多，約 2-3 個週期會出現車多使克強路行車無法順利右轉情形，判斷可能與行人早開時相有關。各路口車流約 1-2 個週期便可順利壅擠，實測至下午 6 點時實測場域與上下游路口皆呈現車流順暢狀態。



圖 4.3.4-4 克強路右轉中山北路北往南方向車流

3.11 月 20 日(週一)

(1) 晨峰

實測期間號誌控制器斷線 7 次，經了解得知臺北市交控中心通訊伺服器於凌晨出現異常，導致這幾日實測期間號誌控制器頻繁斷線。德行東西路於 07:30 時車隊尾端需 2 次停等，東往西車流約至上游一個路口，西往東超過一個路口，克強路於 07:30 時因路口不淨空無法右轉往南，造成回堵到上游橋上，直到中山北車流減少才會有較明顯的有效綠燈，8 點開始實測範圍整體車流減少。



圖 4.3.4-5 中山北路六段/克強路口車流狀況

(2)離峰

實測期間號誌控制器斷線 9 次，14：48 時第 9 次斷線，由監控人員手動關閉結束此時段。現場僅中山北路幹道車流較多，但仍屬順暢。

(3)昏峰

由於今日號誌連線狀況不佳，討論觀察前半小時，若仍頻繁斷線則終止該時段實測，自下午 5 點開始 AI 號誌控制執行約 10-20 分鐘即斷線 1 次，故於 18：15 由監控人員手動關閉結束實測。觀察車流較 11 月 15 日少，約 1 週期可紓解完成，17：40 開始車流減少。

(二)假日實測

1.11 月 18 日(週六) 昏峰

實測期間號誌控制器斷線 3 次。現場車流德行東路西往東方向、克強路西往東方向與中山北路北往南方向皆出現車多狀況，觀察一個週期皆可順利紓解。

2.11 月 19 日(週日) 昏峰

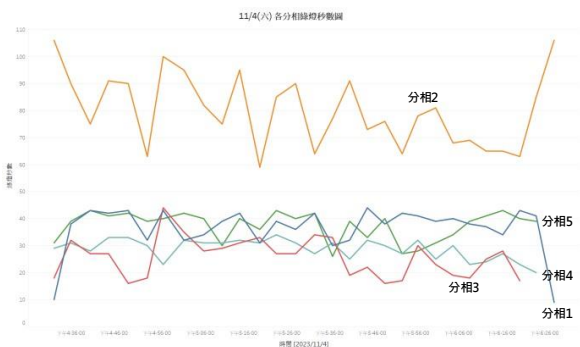
實測期間號誌控制器斷線 6 次。現場車流德行東路西往東方向車多，但仍可於一週期紓解完成，五點過後中山北路北往南方向車多壅塞，導致克強路部分週期車流會壅塞至明德橋前，需 1-2 個週期紓解。

4.4 事前事後量化績效評估

4.4.1 臺南市實驗場域實測結果

一、實測路口分相秒數變化

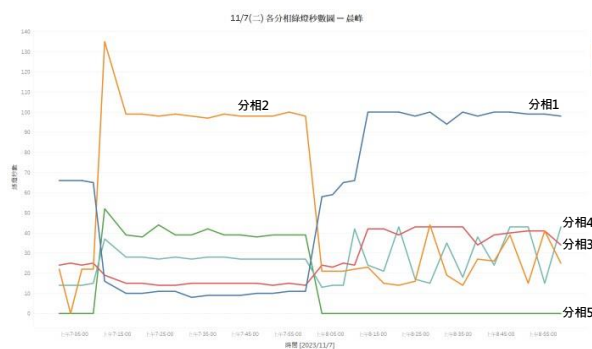
為逐時相決策分相綠燈秒數，考量當下運行時相編號與該時相剩餘秒數，以及各分相結束時各車種轉向量，判斷下一個分相 AI 控制路口之綠燈秒數。



11/4(六)假日昏峰各分相綠燈秒數圖



11/5(日)假日昏峰各分相綠燈秒數圖



11/7(二)平日晨峰各分相綠燈秒數圖



11/7(二)平日離峰各分相綠燈秒數圖



11/7(二)平日昏峰各分相綠燈秒數圖



11/8(三)平日晨峰各分相綠燈秒數圖



11/8(三)平日離峰各分相綠燈秒數圖



11/8(三)平日昏峰各分相綠燈秒數圖

圖 4.4.1-1 平、假日 AI 臺南場域實測各分相綠燈秒數趨勢圖

二、路口平均延滯績效

本場域蒐集各臨近路口各方向路段之 GVP 資料，進而推估各臨近路口之路段於 AI 號誌控制前後之延滯表現，具體作法為當車輛駛離上游路口進入目標路段時，其行駛速率會有一段距離不受路口號誌及停等車隊影響，僅受車流影響，且若維持此速率將可順利通過停止線；而實際受號誌影響後通過停止線的時間落差即為號誌帶給車流的整體影響(包含減速、停等、再啟動加速之延滯等)。

又本場域為單一號誌路口，且各方向距上、下游號誌路口達至少 500m 以上，不受號誌及停等車隊影響之車流速率資料取樣較具參考價值，因此採用此法。實測路口延滯績效表現說明如下。

(一)假日昏峰績效表現

1.11 月 4 日(六)較 11 月 11 日(六)路口整體延滯增加 16.1 秒，服務水準自 B 級降到

C 級。

2.11 月 5 日(日)較 11 月 12 日(日)路口整體延滯增加 8.3 秒，服務水準自 B 級降到 C 級，兩天假日 AI 表現較 TOD 表現略差。

表 4.4.1-1 假日 AI 號誌控制路口延滯績效表

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/4	台86/ 台19甲	假日昏峰 1630~1830	關廟交流道往台86	↙	1895	1232	51.3	D	37.7
			南雄路二段往關新路一段	↘	725	1877	78.2	E	
			台86往關廟交流道	↗	1522	2557	106.6	F	
			關新路一段往南雄路二段	↖	711	1965	81.9	F	C

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/5	台86/ 台19甲	假日昏峰 1630~1830	關廟交流道往台86	↙	2429	1937	80.7	F	31.7
			南雄路二段往關新路一段	↘	800	1936	80.7	F	
			台86往關廟交流道	↗	1507	1510	62.9	E	
			關新路一段往南雄路二段	↖	727	1832	76.3	E	C

表 4.4.1-2 假日 TOD 路口延滯績效表

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/11	台86/ 台19甲	假日昏峰 1630~1830	關廟交流道往台86	↙	2867	1236	51.5	D	21.6
			南雄路二段往關新路一段	↘	776	1834	76.4	E	
			台86往關廟交流道	↗	1960	930	38.7	C	
			關新路一段往南雄路二段	↖	960	1901	79.2	E	B

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/12	台86/ 台19甲	假日昏峰 1630~1830	關廟交流道往台86	↙	3272	2406	100.3	F	23.4
			南雄路二段往關新路一段	↘	889	1717	71.5	E	
			台86往關廟交流道	↗	2248	1061	44.2	C	
			關新路一段往南雄路二段	↖	953	1984	82.7	F	B

(二)11 月 7 日(二)&11 月 14 日(二)平日晨峰、離峰及昏峰績效表現

1.晨峰：路口整體延滯減少 5.4 秒，服務水準自 C 級升到 B 級

2.離峰：路口整體延滯增加 0.7 秒，服務水準維持 B 級

3.昏峰：路口整體延滯減少 7.3 秒，服務水準自 C 級升到 B 級

表 4.4.1-3 平日 11/7 AI 號誌控制路口延滯績效

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/7	台86/ 台19甲	平日晨峰 0700~0900	關廟交流道往台86	↙	1285	1148	47.8	D	26.9
			南雄路二段往關新路一段	↘	827	1558	64.9	E	
			台86往關廟交流道	↗	1555	690	28.8	B	
			關新路一段往南雄路二段	↖	825	1637	68.2	E	B
	台86/ 台19甲	平日離峰 1230~1430	關廟交流道往台86	↙	1002	380	47.5	D	16.6
			南雄路二段往關新路一段	↘	352	605	75.6	E	
			台86往關廟交流道	↗	858	266	33.2	C	
			關新路一段往南雄路二段	↖	462	601	75.2	E	B
	台86/ 台19甲	平日昏峰 1630~1830	關廟交流道往台86	↙	1820	670	27.9	B	29.8
			南雄路二段往關新路一段	↘	492	1745	72.7	E	
			台86往關廟交流道	↗	1297	686	28.6	B	
			關新路一段往南雄路二段	↖	716	2263	94.3	F	B

表 4.4.1-4 平日 11/14 TOD 路口延滯績效

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/14	台86/ 台19甲	平日晨峰 0700~0900	關廟交流道往台86	↙	1226	1221	50.9	D	32.3
			南雄路二段往關新路一段	↘	677	1496	62.3	E	
			台86往關廟交流道	↗	1356	626	26.1	B	
			關新路一段往南雄路二段	↖	782	2087	87.0	D	
	台86/ 台19甲	平日離峰 1230~1430	關廟交流道往台86	↙	969	409	51.1	D	15.9
			南雄路二段往關新路一段	↘	362	464	58.0	D	
			台86往關廟交流道	↗	842	339	42.3	C	
			關新路一段往南雄路二段	↖	469	540	67.5	E	
	台86/ 台19甲	平日昏峰 1630~1830	關廟交流道往台86	↙	1875	1387	57.8	D	34.3
			南雄路二段往關新路一段	↘	475	1702	70.9	E	
			台86往關廟交流道	↗	1269	1332	55.5	D	
			關新路一段往南雄路二段	↖	830	1928	80.3	F	

(三)11月8日(三)&11月15日(三)平日晨峰、離峰及昏峰

- 1.晨峰：路口整體延滯增加 7.1 秒，服務水準維持 C 級
- 2.離峰：路口整體延滯減少 2.1 秒，服務水準自 B 級升到 A 級
- 3.昏峰：路口整體延滯減少 2.5 秒，服務水準自 C 級升到 B 級

表 4.4.1-5 平日 11/8 AI 號誌控制路口延滯績效

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/8	台86/ 台19甲	平日晨峰 0700~0900	關廟交流道往台86	↙	1213	1601	66.7	E	41.8
			南雄路二段往關新路一段	↘	671	1853	77.2	E	
			台86往關廟交流道	↗	1540	1940	80.8	F	
			關新路一段往南雄路二段	↖	751	1884	78.5	E	
	台86/ 台19甲	平日離峰 1230~1430	關廟交流道往台86	↙	1034	375	46.8	D	14.2
			南雄路二段往關新路一段	↘	395	478	59.8	D	
			台86往關廟交流道	↗	889	266	33.2	C	
			關新路一段往南雄路二段	↖	471	529	66.2	E	
	台86/ 台19甲	平日昏峰 1630~1830	關廟交流道往台86	↙	1939	579	24.1	B	29.1
			南雄路二段往關新路一段	↘	507	1678	69.9	E	
			台86往關廟交流道	↗	1307	764	31.8	C	
			關新路一段往南雄路二段	↖	801	2502	104.2	F	

表 4.4.1-6 平日 11/15 TOD 路口延滯績效

	路口	時間	鄰近路口方向	方向	通過車輛數 (輛/小時)	總延滯時間 (秒)	平均延滯 (秒)	服務水準	路口整體
11/15	台86/ 台19甲	平日晨峰 0700~0900	關廟交流道往台86	↙	1154	1205	50.2	D	34.7
			南雄路二段往關新路一段	↘	657	1595	66.5	E	
			台86往關廟交流道	↗	1377	1047	43.6	C	
			關新路一段往南雄路二段	↖	755	1856	77.3	E	
	台86/ 台19甲	平日離峰 1230~1430	關廟交流道往台86	↙	1026	383	47.8	D	16.3
			南雄路二段往關新路一段	↘	364	516	64.5	E	
			台86往關廟交流道	↗	886	353	44.1	C	
			關新路一段往南雄路二段	↖	468	609	76.2	E	
	台86/ 台19甲	平日昏峰 1630~1830	關廟交流道往台86	↙	1971	1517	63.2	E	31.7
			南雄路二段往關新路一段	↘	505	1593	66.4	E	
			台86往關廟交流道	↗	1408	1206	50.3	D	
			關新路一段往南雄路二段	↖	843	1926	80.2	F	

二、旅行時間績效

傳統之旅行速率調查係採用浮動車法，調查車輛於車隊中保持車隊之均速行駛。然在號誌

化幹道中車輛不同的出發時間點其可續進的路口數不同，即使依交通工程手冊至少蒐集 6 趟資料，亦可能無法真正反應整體車流績效。

因此本計畫透過蒐集 GVP 資料瞭解事前事後路段平均旅行時間，預計事前及事後將各至少蒐集 1 週之資料。由於短時間內車流隨號誌而有週期性變化，為能全面掌握路況 GVP 資料蒐集之頻率將避開號誌週期，如：號誌週期為 150 秒則不能每 150 秒或每 300 秒蒐集一次資料。

平日 11 月 7 日(二)AI 號誌控制晨峰及昏峰表現都略優於 11 月 14 日，隔一天 11 月 8 日(三)AI 號誌控制晨峰出現兩處峰期，且昏峰表現也較 11 月 15 日略差；假日 AI 號誌控制旅行時間表現皆略較 TOD 表現為差。此情形亦對應在延滯表現上，唯路口服務水準尚維持在 C 級以上，車流運行尚算通順。

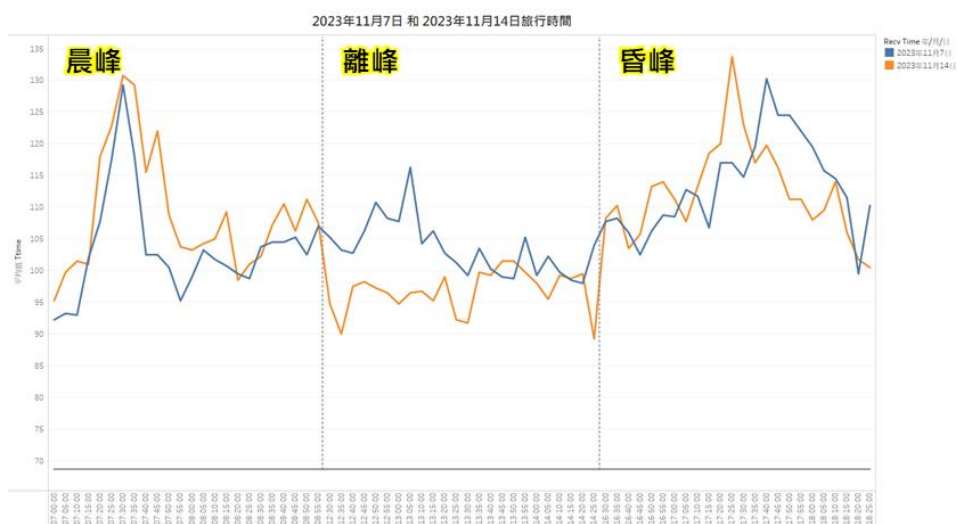


圖 4.4.1-2 平日 11/7&11/14 AI 號誌控制與 TOD 旅行時間績效示意圖

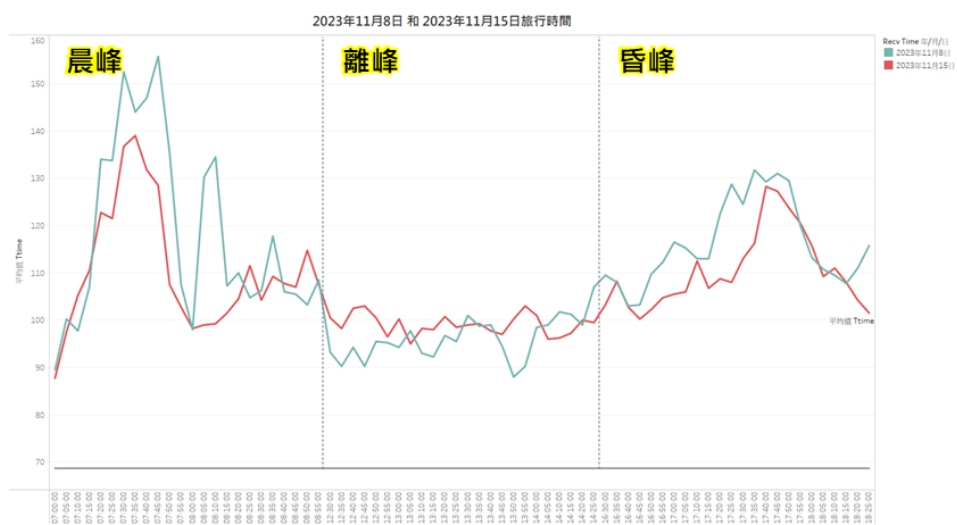


圖 4.4.1-3 平日 11/8&11/15 AI 號誌控制與 TOD 旅行時間績效示意圖

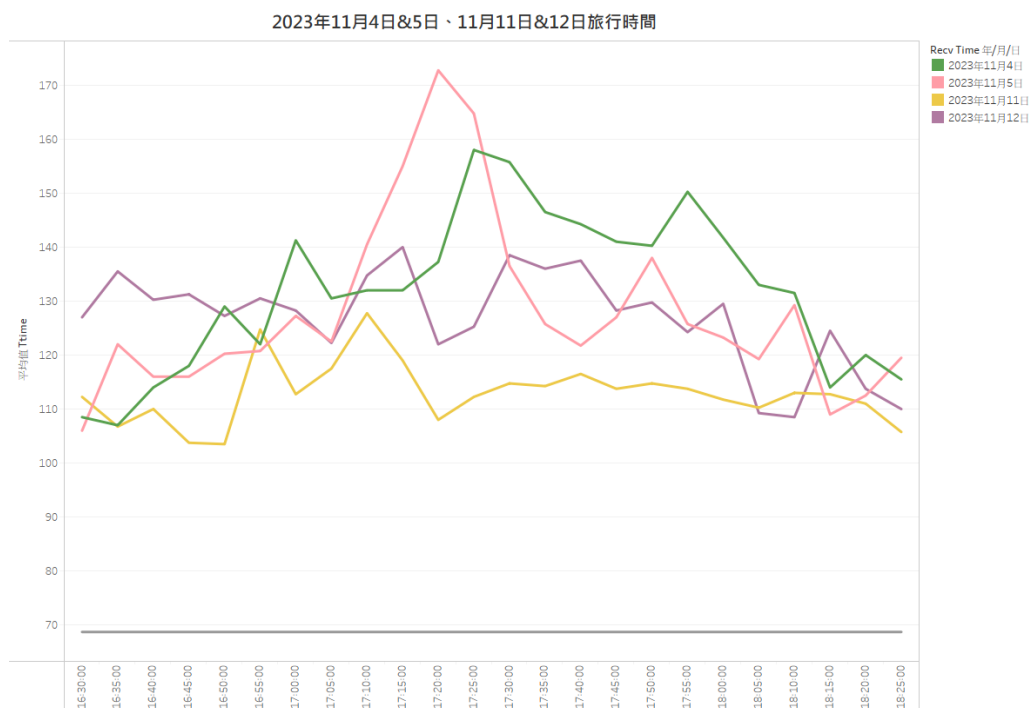
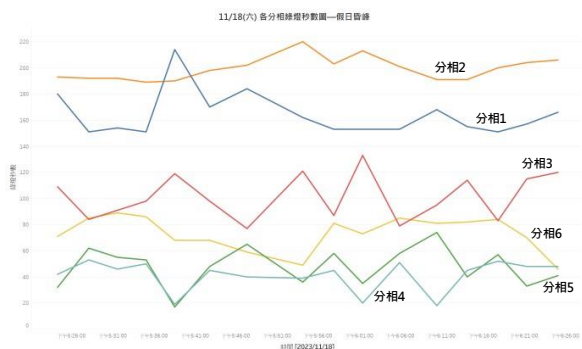


圖 4.4.1-4 假日 AI 號誌控制與 TOD 旅行時間績效示意圖

4.4.2 臺北市實驗場域實測結果

一、實測路口分相秒數變化

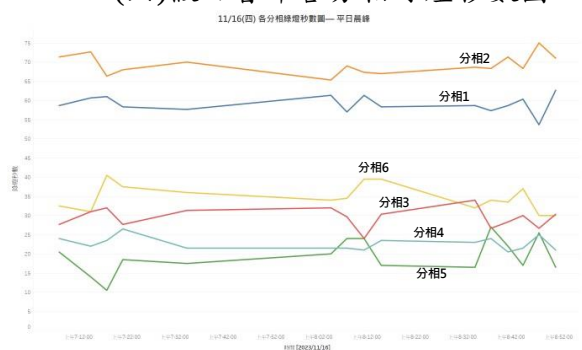
臺北市實驗場域為當周期該分相停等車輛數決策後續分相綠燈秒數。



11/18(六)假日昏峰各分相綠燈秒數圖



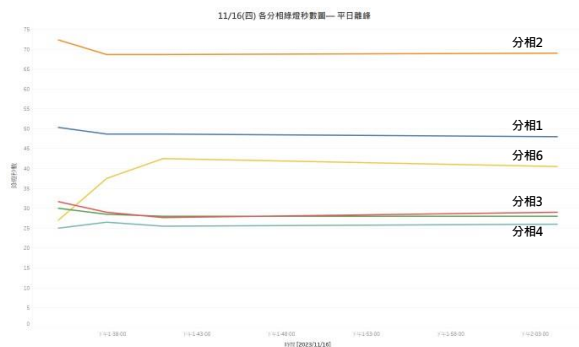
11/19(日)假日昏峰各分相綠燈秒數圖



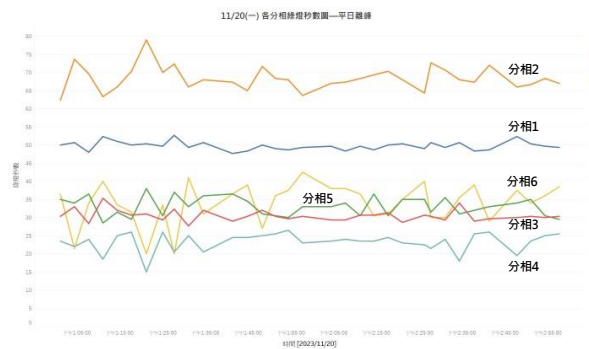
11/16(四)平日晨峰各分相綠燈秒數圖



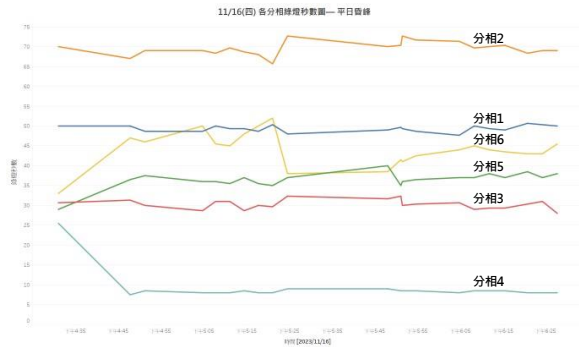
11/20(一)平日晨峰各分相綠燈秒數圖



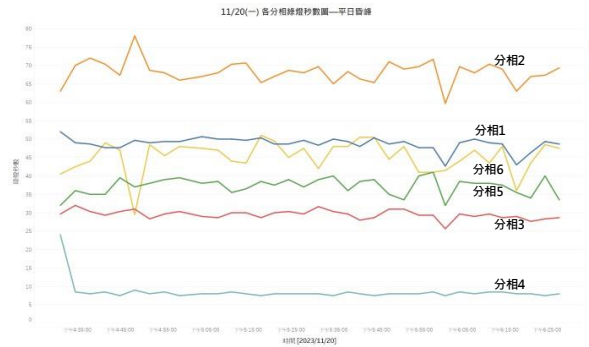
11/16(四)平日離峰各分相綠燈秒數圖



11/20(一)平日離峰各分相綠燈秒數圖



11/16(四)平日昏峰各分相綠燈秒數圖



11/20(一)平日昏峰各分相綠燈秒數圖

圖 4.4.2-1 平、假日 AI 臺北市實驗場域實測各分相綠燈秒數趨勢圖

二、綠燈續進情形

紀錄 TOD 號誌控制時之三處控制路口綠燈續進情形，其時段起點為平日昏峰 16:30，上行約 100 秒，下行約 80 秒，詳圖 4.4.2-2 所示；AI 號誌控制時之三處控制路口，因各分相綠燈秒數會隨偵測之停等車輛變化，綠燈續進情形上行平均約 93 秒，下行平均約 83 秒，詳圖 4.4.2-3 所示。其與 TOD 號誌控制連鎖差異較明顯的地方為德行西路口，對比兩張圖可發現 TOD 號誌控制之上行、下行秒數每個周期皆不變，而 AI 號誌控制於該路口秒數配置依時間序演進而變化增加，除因應路口車輛停等狀態外，應還有對應昏峰下班車流增加之情形。

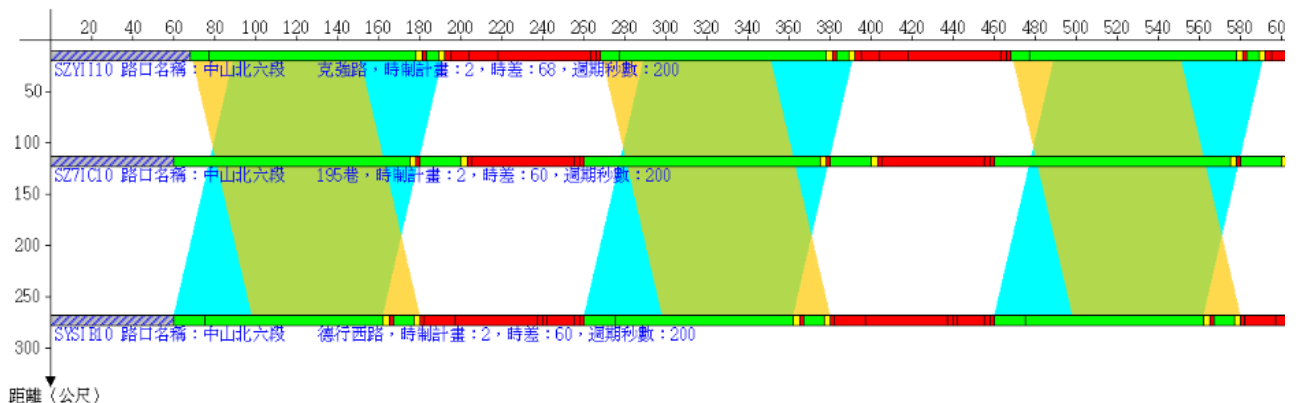


圖 4.4.2-2 臺北市實驗場域控制路口 TOD 平日昏峰綠燈續進示意圖

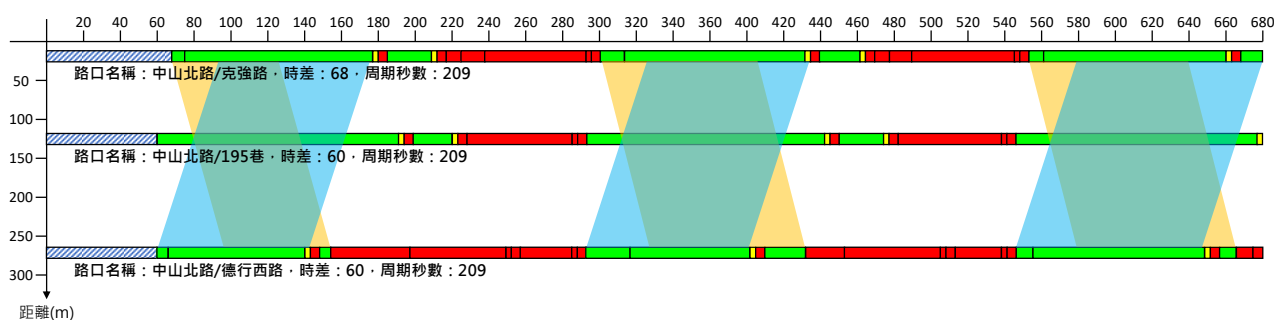


圖 4.4.2-3 臺北市實驗場域控制路口 AI 號誌控制平日昏峰綠燈續進示意圖

三、路口平均延滯績效

傳統停等延滯調查係人員站在停止線附近，每 15 秒計數停等車輛數，本場域融合現代科技，採架設之路口偵測各車道停等車輛數之雷達設備資料，依路口方向分類，整理時間序列，加總同一方向全部車道之停等車輛，每 15 秒計數一次停等車輛數，偵測距離可達 100 公尺，唯可能受限於道路兩側行道樹遮蔽或現場網路不穩或斷線問題，造成事後資料蒐集分析績效誤差。此法改善傳統人工判讀問題及人力耗費，實測路口延滯績效表現說明如下。

(一)假日昏峰 AI 號誌控制平均延滯皆較 TOD 略差，但服務水準大多維持不變，僅 11 月 19 日(日)195 巷及德行西路服務水準分別從 A 降到 B、C 降到 D。

(二)平日 11 月 16 日(四)AI 號誌控制服務水準皆與 11 月 23 日維持一樣的水準，唯平日 11 月 20 日(一)195 巷及德行西路晨峰服務水準分別自 A 級降至 B 級，B 級降至 C 級，但昏峰德行西路服務水準自 D 級提升至 C 級，路口延滯減少 9.5 秒。

表 4.4.2-1 假日 AI 號誌控制&TOD 路口延滯績效表

		11月18日				11月11日			
		總延滯	車輛數	平均延滯	等級	總延滯	車輛數	平均延滯	等級
中山北路/克強路	下午尖峰總延滯	10896	3701	44.2	C	9440	4418	32.1	C
中山北路/195巷		3712	3165	17.6	B	4180	3329	18.8	B
中山北路/德行西路		16548	4440	55.9	D	16490	4682	52.8	D

		11月19日				11月12日			
		總延滯	車輛數	平均延滯	等級	總延滯	車輛數	平均延滯	等級
中山北路/克強路	下午尖峰總延滯	9912	4037	36.8	C	8212	4098	30.1	C
中山北路/195巷		3332	2945	17.0	B	1320	3126	6.3	A
中山北路/德行西路		14634	4118	53.3	D	12474	4357	42.9	C

表 4.4.2-2 平日 AI 號誌控制&TOD 路口延滯績效表

		11月16日				11月23日			
		總延滯	車輛數	平均延滯	等級	總延滯	車輛數	平均延滯	等級
中山北路/克強路	上午尖峰總延滯	7104	4009	26.6	B	4488	3729	18.1	B
	離峰總延滯	7612	3440	33.2	C	7358	3068	36.0	C
	下午尖峰總延滯	7936	3391	35.1	C	9010	3799	35.6	C
中山北路/195巷	上午尖峰總延滯	2038	3239	9.4	A	2368	2847	12.5	A
	離峰總延滯	688	3090	3.3	A	544	3122	2.6	A
	下午尖峰總延滯	3000	3263	13.8	A	3080	3507	13.2	A
中山北路/德行西路	上午尖峰總延滯	9558	5425	26.4	B	8970	5382	25.0	B
	離峰總延滯	13312	4774	41.8	C	12834	4971	38.7	C
	下午尖峰總延滯	13498	4081	49.6	D	14626	4613	47.6	D

		11月20日				11月27日			
		總延滯	車輛數	平均延滯	等級	總延滯	車輛數	平均延滯	等級
中山北路/克強路	上午尖峰總延滯	7210	3767	28.7	B	6604	4014	24.7	B
	離峰總延滯	7968	3612	33.1	C	7622	3738	30.6	C
	下午尖峰總延滯	7936	3726	31.9	C	9440	4418	32.1	C
中山北路/195巷	上午尖峰總延滯	3534	3042	17.4	B	1686	3283	7.7	A
	離峰總延滯	1496	3410	6.6	A	848	3564	3.6	A
	下午尖峰總延滯	3000	4169	10.8	A	3072	4423	10.4	A
中山北路/德行西路	上午尖峰總延滯	12678	5385	35.3	C	10034	5666	26.6	B
	離峰總延滯	15064	5105	44.3	C	14430	5397	40.1	C
	下午尖峰總延滯	13498	4815	42.0	C	16586	4828	51.5	D

四、旅行時間績效

本計畫透過蒐集 GVP 資料瞭解事前事後路段平均旅行時間，平日 11 月 16 日(四)AI 號誌控制除晨峰因系統、設備多次斷線導致旅行時間較 TOD 高外，離峰及昏峰表現雖略差，但都貼合 TOD 走勢；平日 11 月 20 日(一)晨峰及離峰旅行時間表現亦多貼合 TOD 走勢，惟昏峰表現明顯優於 TOD 表現，對應路口延滯應為克強路平均延滯較 TOD 減少 0.2 秒又德行西路平均延滯減少 9.5 秒所致。假日兩天旅行時間績效 AI 號誌控制皆略差於 TOD。

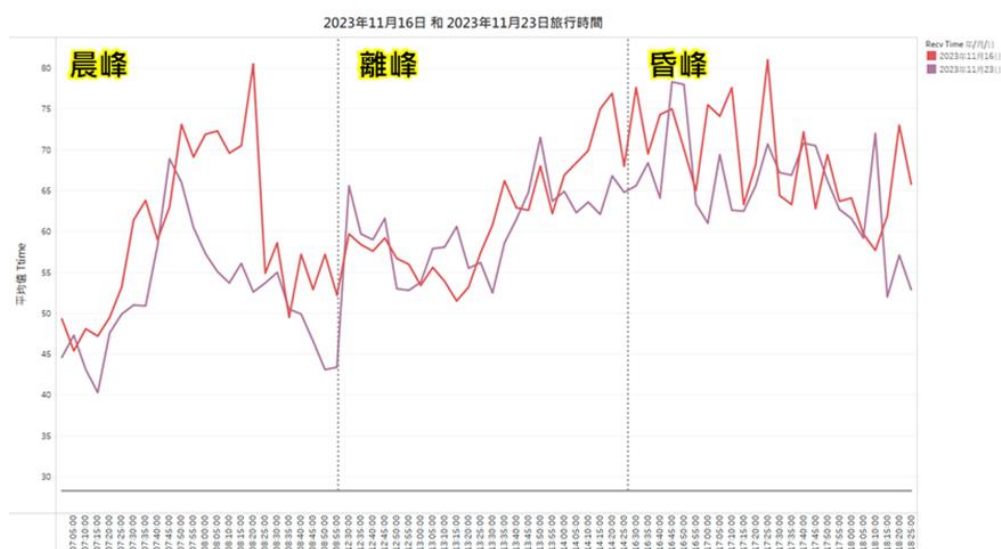


圖 4.4.2-4 平日 11/16&11/23 AI 號誌控制與 TOD 旅行時間績效示意圖

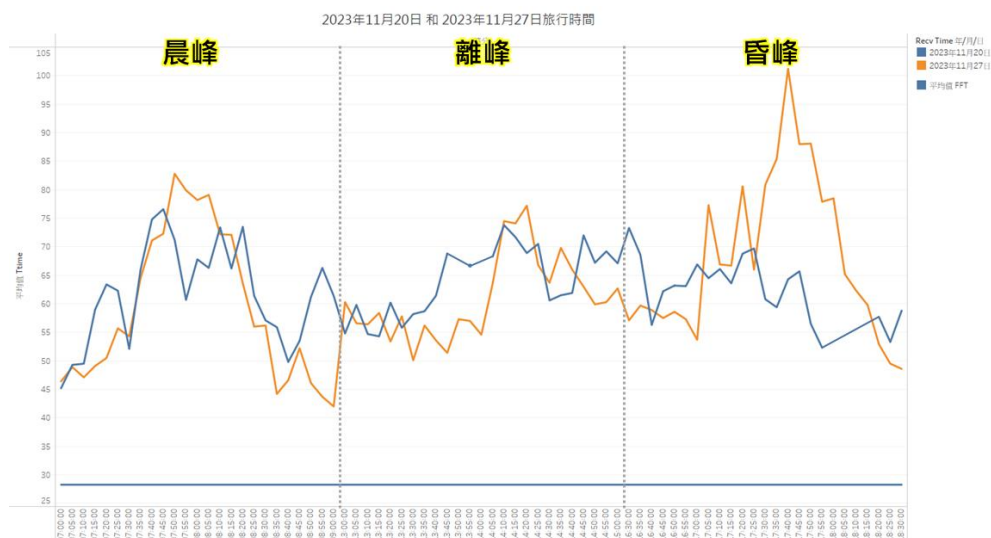


圖 4.4.2-5 平日 11/20&11/27 AI 號誌控制與 TOD 旅行時間績效示意圖

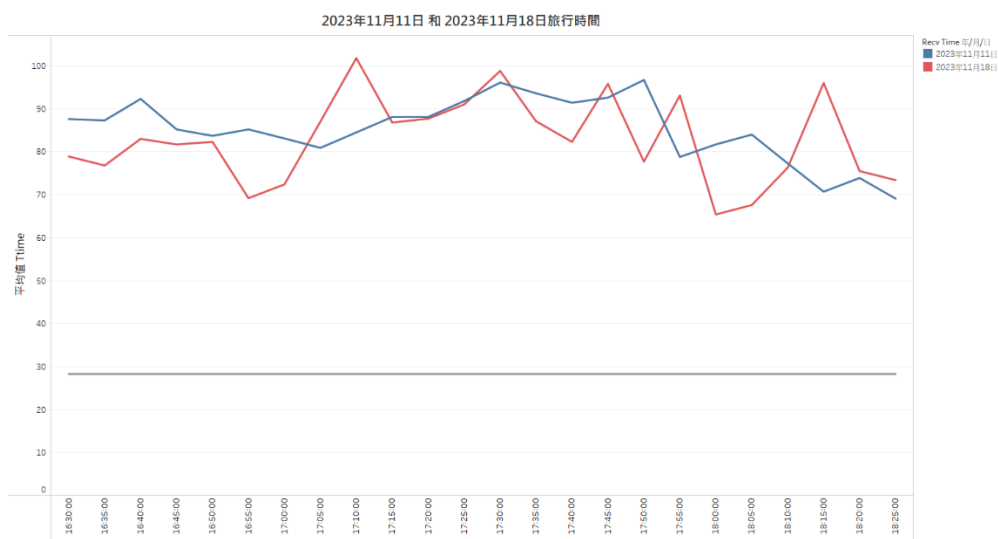


圖 4.4.2-6 假日 11/18&11/11 AI 號誌控制與 TOD 旅行時間績效示意圖

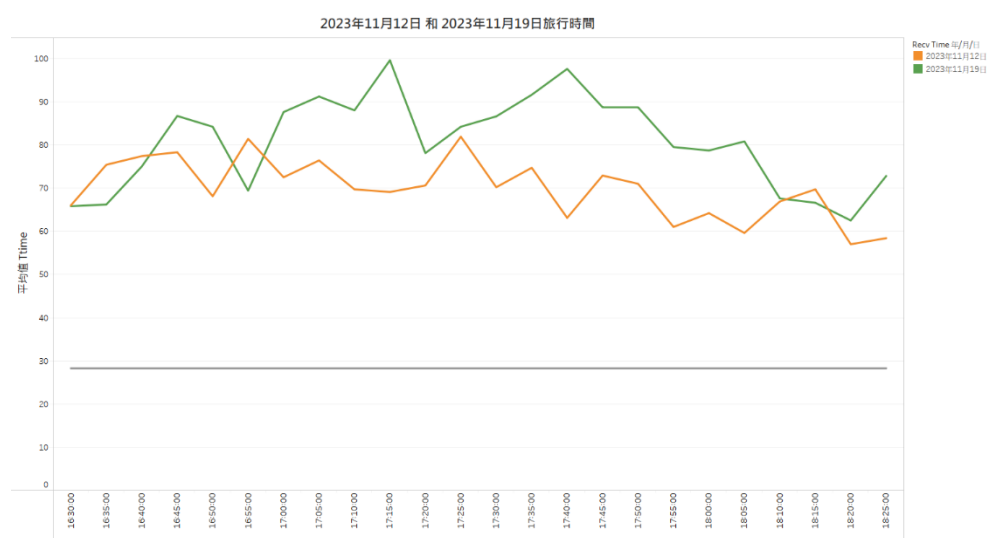


圖 4.4.2-7 假日 11/19&11/12AI 號誌控制與 TOD 旅行時間績效示意圖

4.4.3 實測分析

一、實測結果

本次兩場域實測結果，依設計之績效指標，多數情況與 TOD 大致持平。在實測時段，假日 AI 較沒有改善績效；平日 AI 強化學習模型於昏峰時多數有改善績效，而在晨峰時 AI 控制效果則較不明顯。時段資料彙整如表 4.4.3-1，相關實測結果詳參 4.4.1、4.4.2 節。

表 4.4.3-1 實測結果綜整表

	日期		使用偏好		績效(服務水準)	績效說明
臺南市	平日	112/11/7 112/11/8	晨峰 5 時相	延滯	持平	路口平均延滯(秒)：整體路口的平均每車延滯，直接影響服務水準
			晨峰 4 時相	延滯	持平	
			下離峰	延滯	持平	
			昏峰	延滯	稍優	
	假日	112/11/4 112/11/5	昏峰	延滯	稍劣	通過量(輛/小時)：路口通過車輛數
臺北市	平日	112/11/16 112/11/20	晨峰	平均	稍劣	路口平均停等延滯(秒)：整體路口、分方向的平均每車因號誌的停等延滯，直接影響服務水準
			離峰	平均	持平	
			昏峰	主幹道	稍優	
	假日	112/11/18 112/11/19	昏峰	通過量	稍劣	通過量(輛/小時)：路口通過車輛數

二、實測執行問題

本次臺南市及臺北市實驗場域實測時遇到以下幾種執行面的問題，都將導致實測出現不可預期之狀況，甚至影響實測結果，說明如下：

(一)控制器步階控制異常，時制補償影響車流加劇

當原固定時制之時相秒數較短，人工智慧號誌控制核心運作模組來不及於原時相秒數結束前下達下一步階指令時，沒有接受到指令之號誌控制器將恢復原固定時制控制，進入補償，可能導致現場車流遭受影響。

(二)網路不穩、偵測設備與號誌控制器斷線，導致車流狀況無法即時掌控

實測期間臺北市實驗場域、臺南市實驗場域皆發生設備斷線情形，偵測設備一旦斷線，AI 強化學習模型將無法掌握現場即時車流狀況輸出秒數，而號誌控制器斷線，則 AI 強化學習模型運算之秒數無法下達至現場，且控制器將恢復固定時制計畫，進入補償狀況，影響現場車流。

臺北市實驗場域實測期間交通控制中心機房通訊伺服器故障，導致號誌控制器頻繁斷線，直至實測結束仍未修復，此外車流偵測設備於原訂實測第一天早晨發現因電路被破壞無法連線，緊急搶修後亦偶發斷線情形，後續請設備廠商協助更換零件後便無再度發生。

臺南市實驗場域因位置較偏遠且周邊無住家，因而其網路訊號不優，故實測時間時有號誌控制器斷線情形發生。

(三)流量型態變化，較疫情期間均有增加

以臺北為例，7月13日(四)及7月15日(六)進行交通調查，並與前期調查結果(111年6月14日)比對發現，今年交通量調查結果較去年成長近3倍，判斷可能當時仍受疫情嚴重影響，又進一步與疫情前(109年)臺北市交通管制工程處車流調查比較，今年交通量調查結果為當時之0.8倍，顯示交通量有漸漸恢復之趨勢。

設備編號	設備類型	暫停運作原因	暫停開始時間	暫停結束時間
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:30秒。	2023-11-04 18:28:22	2023-11-05 16:03:26
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:25秒。	2023-11-05 16:36:04	2023-11-05 16:44:00
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:33秒。	2023-11-05 17:23:01	2023-11-05 17:31:28
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:33秒。	2023-11-05 17:52:34	2023-11-05 18:00:17
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:28秒。	2023-11-05 18:04:22	2023-11-05 18:12:07
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:29秒。	2023-11-05 18:16:10	2023-11-05 18:24:05
S281601	控制器異常	S281601 TC與IPC步階不一致。TC提早結束綠燈:25秒。	2023-11-05 18:28:09	2023-11-06 16:27:41
S281601	時制轉換	S281601 因時相轉換暫停運作	2023-11-08 07:10:04	2023-11-08 07:10:06
S281601	時制轉換	S281601 因時相轉換暫停運作	2023-11-08 08:00:29	2023-11-08 08:13:24
S281601	控制器異常	S281601 TC外部控制，恢復定時。	2023-11-15 10:18:55	

圖 4.4.3-1 實測階段遇到控制器異常情形示意

4.5 小結

前期使用 DDPG 演算法進行臺北市及臺南市實驗場域 AI 模型訓練，該演算法為確定性策略(Deterministic Policy)，雖然在訓練上更快速，但無法找出全域之最佳解，另外 DDPG 係以單一目標獎勵進行訓練。前期各個路口皆配置一個代理人於每時相結束前決策並控制之。本期優化調整採用 Q-Pensieve 多目標演算法，可根據不同目標獎勵需要進行訓練，同時求解所有組合之最佳化。本期各個實驗場域配置一個代理人進行控制，臺北市實驗場域調整決策頻率為每週期決策，一個代理人同時控制三個路口，使得三個路口週期一致，提供更好的續進效果。

觀察本期兩場域之實測結果，大多數與 TOD 之績效持平，AI 模型於平日昏峰之績效多有改善；於假日、平日晨峰則較不明顯。回顧實測期間所發生之問題，包含：號誌控制器較晚接收到步階控制指令、網路不穩導致偵測設備與號誌控制器斷線，以上因素皆會導致號誌控制器恢復固定時制控制，進入時制補償影響現場車流，除此之外流量型態較疫情期間之變化亦可能影響實測結果。

總結本期都會區幹道 AI 模型精進與實測經驗，可知現場偵測設備與號誌控制器之調校、整合為重要課題，後續可考量進行優化，包含：車流偵測器準確度、設備通訊穩定性與連線率、號誌控制器的近似逐秒控制能力、AI 模型泛用型等。由於強化學習號誌控制模型之限制，需針對不同號誌時段與時相組態各自建模、訓練，為探討 AI 模型之實務應用，本期嘗試多任務強化學習模式(Task-aware)，將不同的時制組態根據維度給予不同的任務(task)，使各模型能夠在不同類型數據的任務中進行學習，並進行整合與泛化，使其更加多樣化和適應各種現實

場景，實現單一模式訓練並學習不同的時制組態，建議後續可持續多任務 AI 強化學習模型探討，以期未來能發展為泛用型(全時段)的 AI 強化學習號誌控制模型。

第五章 強化學習號誌協控模擬環境構建規劃

5.1 強化學習號誌協控型態探討

延續前述章節，為利發展我國於車聯網與人工智慧強化學習之號誌控制模擬環境，提升車流運作效率及交通安全提升，除與縣市政府合作之現場實驗測試，將同時納入高速公路匝道儀控之人工智慧號誌控制課題與相關模擬模式發展探討，以期逐步發展我國在人工智慧號誌控制與實力構建。本計畫在交流道區域導入人工智慧強化學習在匝道儀控與平面號誌協控，以及於後續進行人工智慧強化學習之號誌協控模型訓練與實測。

建置交流道區域匝道儀控與平面道路路口動態號誌策略模式(以下簡稱區域協控動態號誌策略)，需即時掌握交流道周邊之整體車流運作狀況，並據以執行下匝道號誌協控功能及平面道路壅塞監控反應功能，另透過上匝道儀控及壅塞訊號監測，於必要時發出平面道路壅塞告警事件，經由跨單位已建立之事件通報與反應機制將告警事件傳達予高速公路局相關區域養護工程分局人員，得以即時掌握跨區域交通壅塞狀況與需求，並據以評估匝道儀控管理作為，減少國道交流道上匝道停等車隊回堵進而影響平面幹道運行順暢之情形。

一、區域協控動態號誌策略運作原則

與交通部高速公路局進行協控議題訪談、協調會議與專案成果審查，針對跨域協調溝通介面、作業原則、交換資訊等方面反覆進行研議與修正，已獲致多項具體的執行共識可供本計畫延續據以執行。

二、區域協控動態號誌策略

(一)藉由平面及高速公路區域的車流特性偵測，判斷協控區域之交通需求

本計畫所發展區域協控動態號誌策略系統應能透過多元交通資訊來源，確實掌握實施區域的交通需求，車流監控涵蓋區域包含：國道主線路段、入口匝道末端(上游)路段以及平面交通要道(運輸路廊及橫交要道)。

(二)紓緩出口匝道車流回堵至國道主線的狀況，維持國道主線順暢

1.主要將透過監控出口匝道上游交通壅塞訊號，於產生交通需求時透過動態調整平面號誌群組時制計畫，達到紓解壅塞車流之目的。

2.號誌時制計畫產製依據現場重要路段或路徑之旅行時間、車流量、轉向比等即時車流資料，配適出平面道路群組之建議時制計畫，並下載至路口號誌控制器執行，以求迅速紓解出口匝道壅塞狀況。

(三)當出口匝道車流之競爭號誌時相路段產生壅塞訊號時，協控策略應能夠維持整體路網的均衡控制

- 1.針對控制區域之交通狀況應做整體性的監控，執行之策略亦應配合不同的需求做出適當的因應措施。
- 2.當出口匝道及平面道路均未發出壅塞訊號，或均已發生壅塞的狀況下，系統將以維持控制區域之整體交通狀況進行時制計畫配置。
- 3.當僅有出口匝道發出壅塞訊號時，系統將優先執行紓解匝道車流之號誌時制計畫；反之，若僅有平面道路發出壅塞訊號，以紓解平面壅塞為主。

(四)針對入口匝道車流若判斷因匝道儀控管制導致平面道路嚴重回堵時，在國道主線服務水準容許範圍內，系統將自動發出跨平台事件告警

- 1.透過介接國道匝道儀控號誌訊號及國道主線路況資訊，當系統偵得入口匝道停等車隊回堵至平面道路之情形，在確認匝道儀控管制已啟用且國道主線服務水準尚佳的情況下，可透過 Line notify 功能，自動傳達平面道路壅塞告警事件至縣市政府交通局與高公局交控中心組成之跨單位溝通群組。
- 2.當平面道路壅塞告警事件建立與發佈後，系統則立即重新由時制資料庫中遴選出最適於當前交通需求之時制計畫，以便提升整體道路運作效率。

5.2 強化學習號誌協控模式發展構想

一、交流道場域 AI 強化學習模式發展規劃

交流道場域將分為平面號誌與匝道儀控，為使 AI 能蒐集到平面與匝道的資訊，輸入分為匝道、平面號誌資訊與道路資訊，因匝道與平面週期不會一致，決策的頻率則為各自的週期結束時決策，決策的結果同時輸出下個匝道綠燈秒數、平面號誌週期秒數及分項比例，現場號誌則視何種號誌控制週期結束而選擇 AI 強化學習模式輸出的結果，例如：現場平面號誌週期結束時，則取 AI 強化學習模式輸出中平面號誌的結果，反之則取 AI 強化學習模式輸出中匝道儀控的結果。

表 5.2-1 交流道場域 AI 強化學習模式規劃表

定義	內容
代理人 數量	1
輸入	匝道號誌資訊：上個週期匝道綠燈秒數
	平面號誌資訊：上個週期平面道路週期秒數及分相之比例
	道路資訊：平面及高速公路車流資訊
輸出	下個週期匝道綠燈秒數、平面號誌週期秒數及分項比例
決策頻率	匝道綠燈秒數結束時或平面週期結束時決策

二、交流道場域 AI 強化學習模式訓練規劃

為增加訓練的效率，平面與匝道以同個 policy network 決策，分別為平面與匝道的週期秒

數及時相比比例，因匝道與平面的週期不會一致，訓練週期會以秒為單位，並根據當下平面與匝道的決策剩餘秒數是否為零，決定是否再提供下一輪的時相控制，並輸入對應的決策結果至 environment；state 輸入則為整個場域，平面及匝道的車流狀況與當下號誌剩餘秒數；Critic network 分為平面及匝道，並於每秒間將環境(state)及各自的動作(action)與獎勵(reward)加入 replay buffer 中。

三、Centralized Training Decentralized Execution (CTDE)

在強化學習模型設計上，為使模型能蒐集到平面與匝道的資訊，輸入資訊包括匝道與平面道路號誌資訊，以及交流道周邊的平面道路各路口各分相結束時的分車道分車種(大車/小車/機車)停等車輛數及總車輛數、高速公路各 ETC 偵測區域的總車流量與旅行時間；輸出參數為下個週期匝道綠燈秒數、平面號誌週期秒數及分相比比例；獎勵目標為號誌協控範圍高速公路與平面道路之旅行時間、停等延滯、通過量等。在號誌協控強化學習模型訓練上，導入中央化訓練與分散化執行(Centralized Training Decentralized Execution, CTDE)架構，CTDE 通常適用於多代理人；在設計上，高速公路與平面道路各有一個代理人，所有輸入資訊均集中到一個中央化的 Critic Network，在號誌協控模式訓練過程，都可以針對各群組(高速公路與平面道路)提供策略函數，並將動作(下個週期號誌控制內容)分散式的下載到模擬環境。

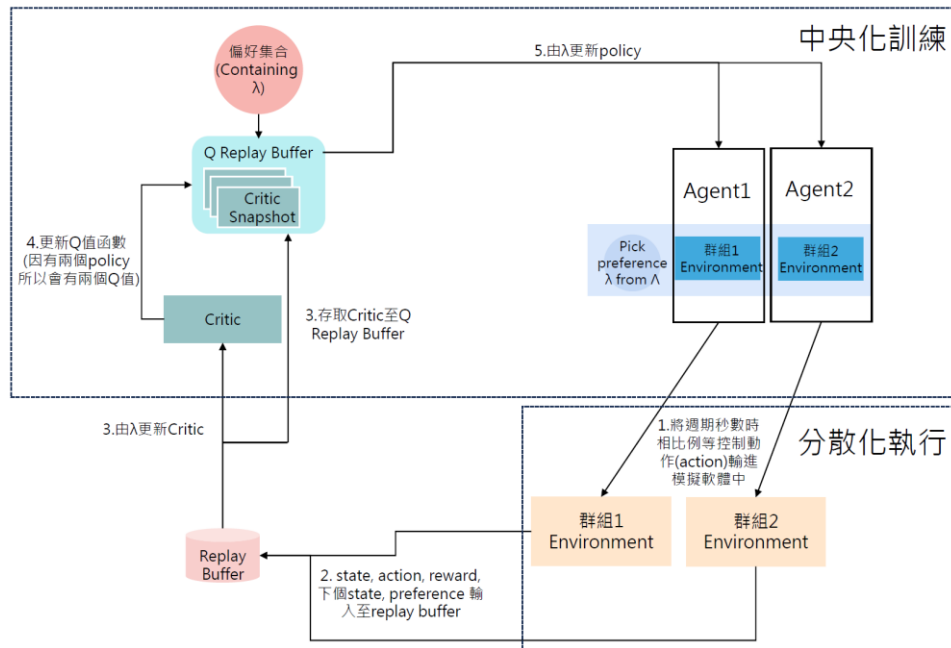
CTDE 架構通常應用於大規模機器學習系統，特別是涉及同時訓練多個策略。這個架構主要分為兩個階段：

(一)中央化訓練 (Centralized Training)

在這個階段，匝道和平面道路的數據集被集中到一個中央化的 agent。然後，這個中央化的 agent 用這個大型數據集來訓練整個交通模型。這樣做的好處是可以在大量數據上進行訓練，提高模型的性能和準確性。

(二)分散化執行 (Decentralized Execution)

一旦機器學習模型訓練完成，就可以針對匝道與平面道路提供策略函數，並使用其對於給予的環境(state)做動作(action)。



5.3 實驗場域規劃與分析

根據發展人工智慧號誌控制模式需求，自六都與高速公路介接的交流道(含前期計畫3處實驗場域)中遴選一處上匝道與平面道路作為實驗場域，執行區域協控計畫，藉由以交通運輸模型為基礎之智慧號誌控制系統，改善交流道周邊區域的交通。

實驗場域之選擇須考量模擬模式與模型構建所需資料，以及後續實測現場資通訊軟硬體可配合因素。本計畫係從都會區角度進行探討，因此僅考量平常日與高速公路獨立式匝道儀控，以下以楊梅交流道實驗場域為例說明，實際實驗場域需選定後待與機關同意而定。

一、建置緣由：選定場域之緣由及必要性

因楊梅市區欠缺外環道，民眾欲北上或南下大多須經楊梅交流道，導致車流量集中，尖峰時段沿線易壅塞。

二、偵測範圍：實驗場域偵測路口及範圍

台 1 線中山北路一段雙向(梅獅路至環東路間)為進出國道 1 號楊梅交流道即前往埔心及楊梅方向之重要道路，劃定號誌化路口 1 至路口 9(路口 8 為往楊梅方向下匝道出口，非屬號誌化路口且永久閃黃)，如圖 5.3-1 所示。

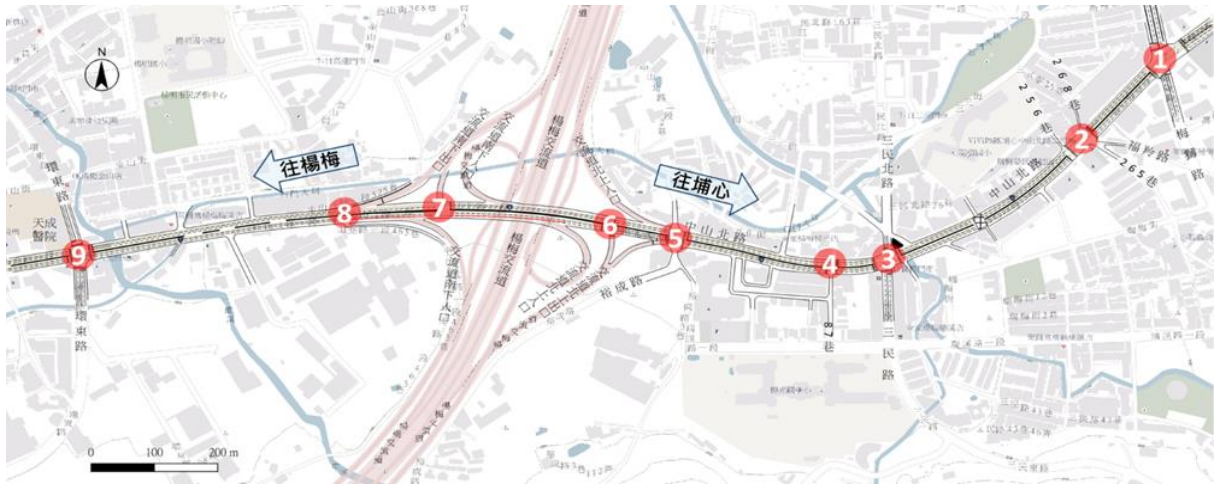


圖 5.3-1 高速公路匝道與平面道路實驗場域示意圖-以楊梅交流道為例

三、交通現況調查：於實驗場域作路口幾何及時制計畫調查以有效建構模擬環境

(一)現況道路特性

針對調查路口進行該路口各個方向之道路幾何及交通管制措施調查，調查項目包含：車道數、分隔型式、車道寬度、車道配置、路口轉向管制(如禁左或單行道)等。計畫施作路口幾何特性於現地調查後將彙整如表 5.3-1 所示。

表 5.3-1 實驗場域施作路口幾何特性彙整表-以楊梅交流道為例

路口 1	方向	A	B	C	D	E
	車道數	2 快 1 最外側	1 快 1 最外側	2 快 1 最外側	1 快	1 快 1 最外側
	分隔型態	實體分隔	標線分隔	實體分隔	單向道	實體分隔
	車道 1	左	左直	左	左直右	左
	車道 2	直	直右	直	-	直右
	車道 3	直右	-	直右	-	-
路口 2	方向	A	B	C	D	E
	車道數	2 快 1 最外側	1 快	2 快 1 最外側	1 快	1 快
	分隔型態	實體分隔	無分隔	實體分隔	無分隔	無分隔
	車道 1	直	直	直	直	直
	車道 2	直	-	直	-	-
	車道 3	直	-	直	-	-
路口 3	方向	A	B	C	D	E
	車道數	2 快 1 最外側	1 快 1 最外側	2 快 1 最外側	1 快 1 最外側	-
	分隔型態	實體分隔	標線分隔	實體分隔	標線分隔	-
	車道 1	左	直	左	直	-
	車道 2	直	直	直	直	-
	車道 3	直右	-	直右	-	-
路口 4	方向	A	B	C	D	E

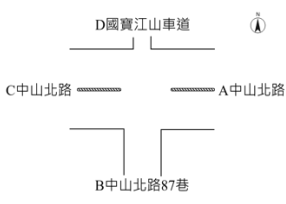
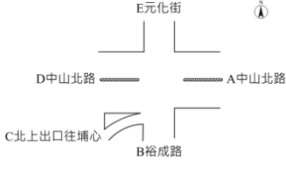
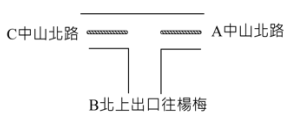
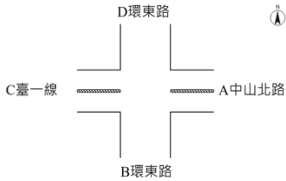
	車道數	2 快 1 最外側	1 快	2 快 1 最外側	-	-
	分隔型態	實體分隔	無分隔	實體分隔	-	-
	車道 1	直	直	直	-	-
	車道 2	直	-	直	-	-
	車道 3	直	-	直	-	-
路口 5	方向	A	B	C	D	E
	車道數	2 快 1 最外側	1 快	1 快	2 快 1 最外側	1 快
	分隔型態	實體分隔	標線分隔	單行道	實體分隔	無分隔
	車道 1	左直	直	直	左直	直
	車道 2	直	-	-	直	-
	車道 3	直右	-	-	右	-

表 5.3-1 實驗場域施作路口幾何特性彙整表-以楊梅交流道為例(續)

路口 6	方向	A	B	C	D	E
	車道數	2 快 1 最外側	1 快	2 快 1 最外側	-	-
	分隔型態	實體分隔	單行道	實體分隔	-	-
	車道 1	直	直	直	-	-
	車道 2	直	-	直	-	-
	車道 3	直右	-	直	-	-
路口 7	方向	A	B	C	D	E
	車道數	3 快	-	2 快 1 最外側	2 快	-
	分隔型態	實體分隔	-	實體分隔	單行道	-
	車道 1	直	-	直	左	-
	車道 2	直	-	直	左	-
	車道 3	直	-	直	-	-
路口 9	方向	A	B	C	D	E
	車道數	2 快 1 最外側	1 快 1 最外側	2 快 1 最外側	1 快 1 最外側	-
	分隔型態	實體分隔	標線分隔	實體分隔	標線分隔	-
	車道 1	左	左直	左	左直	-
	車道 2	直	右	直	直右	-
	車道 3	直右	-	直右	-	-

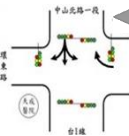
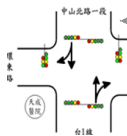
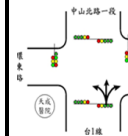
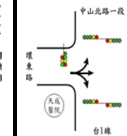
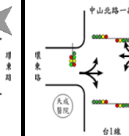
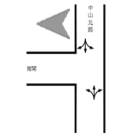
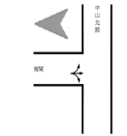
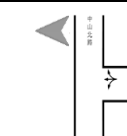
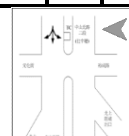
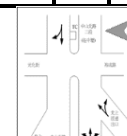
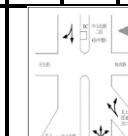
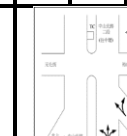
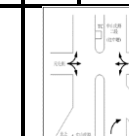
(二)現行時制計畫

針對控制範圍內路口進行號誌時制計畫調查，包括週期、時相、綠燈、紅燈、全紅、以及各路口間之時差等資料。

楊梅交流道於本計畫範圍周邊平面路口現況時段切分及號誌週期於各路口略有不同，平日及假日的時段切分型態亦有差值；其中平日晨峰、昏峰時段週期為 200 秒或 180 秒(分別約介於 06:30~09:00、16:30~19:00)、平日離峰時段週期為 150 秒(約介於

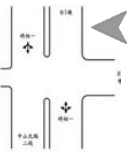
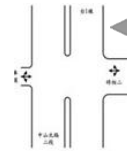
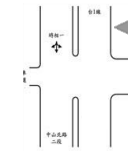
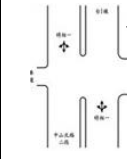
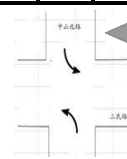
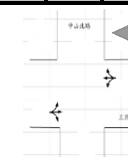
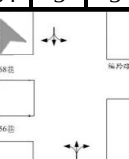
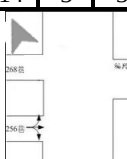
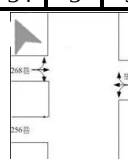
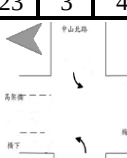
08:30~16:30、19:00~23:00)、平日夜間時段週期為 120 秒(約介於 23:00~06:00)；假日尖峰時段週期為 200 秒或 180 秒(約介於 10:00~19:00)、假日離峰時段週期為 150 秒(分別約介於 06:00~10:00、19:00~23:00)、假日夜間時段週期為 120 秒(約介於 23:00~06:00)。整理如表 5.3-2 及表 5.3-3 所示。

表 5.3-2 楊梅交流道平日時段切分與時制計畫

路口	時段	週期	時差	時制計畫	時相1			時相2			時相3			時相4			時相5		
					G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR
路口9 中山北環東路口																			
	06:00 ~ 08:30	200	0	1	49	3	3	53	4	3	18	4	3	14	3	3	34	3	3
	08:30 ~ 16:30	180	0	2	49	3	3	39	4	3	18	4	3	9	3	3	33	3	3
	16:30 ~ 19:00	200	0	3	44	3	3	58	4	3	18	4	3	14	3	3	34	3	3
	19:00 ~ 23:00	180	0	2	49	3	3	39	4	3	18	4	3	9	3	3	33	3	3
	23:00 ~ 06:00	120	0	4	14	3	3	38	4	3	18	4	3	0	0	0	24	3	3
路口7 交流道南匝																			
	06:30 ~ 09:00	200	0	33	128	4	3	58	4	3									
	09:00 ~ 16:30	150	0	32	88	4	3	48	4	3									
	16:30 ~ 19:00	200	0	33	128	4	3	58	4	3									
	19:00 ~ 23:00	150	0	32	88	4	3	48	4	3									
	23:00 ~ 06:30	120	0	31	70	4	3	36	4	3									
路口6 交流道北匝																			
	06:30 ~ 09:00	200	0	33	128	4	3	58	4	3									
	09:00 ~ 16:30	150	0	32	88	4	3	48	4	3									
	16:30 ~ 19:00	200	0	33	128	4	3	58	4	3									
	19:00 ~ 23:00	150	0	32	88	4	3	48	4	3									
	23:00 ~ 06:30	120	0	31	70	4	3	36	4	3									
路口5 中山北裕成路口																			
	06:30 ~ 09:00	200	0	33	25	0	0	46	0	0	36	3	3	15	3	3	60	3	3
	09:00 ~ 16:30	150	0	32	20	0	0	31	0	0	30	3	3	11	3	3	40	3	3
	16:30 ~ 19:00	200	0	33	25	0	0	46	0	0	36	3	3	15	3	3	60	3	3
	19:00 ~ 23:00	150	0	32	20	0	0	31	0	0	30	3	3	11	3	3	40	3	3
	23:00 ~ 06:30	120	0	31	15	0	0	26	0	0	20	3	3	11	3	3	30	3	3

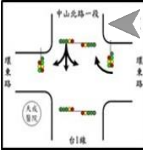
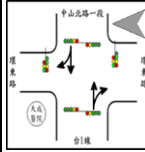
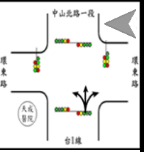
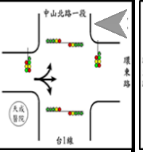
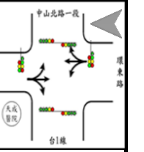
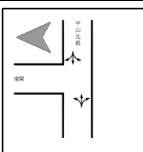
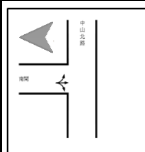
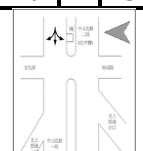
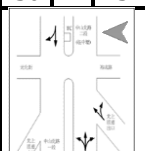
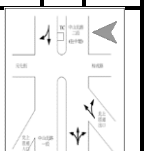
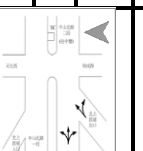
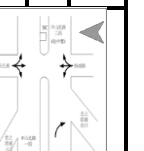
資料來源:桃園市政府交通局，本計畫調查

表 5.3-2 楊梅交流道平日時段切分與時制計畫(續)

路口	時段	週期	時差	時制計畫	時相1			時相2			時相3			時相4			時相5		
					G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR
路口4 中山北 87巷路 口																			
	06:30 ~ 09:00	200	0	5	89	3	3	25	3	2	65	0	0	10	0	0			
	09:00 ~ 16:30	150	0	7	74	3	3	26	3	2	29	0	0	10	0	0			
	16:30 ~ 19:00	200	0	6	98	3	3	25	3	2	56	0	0	10	0	0			
	19:00 ~ 23:00	150	0	7	74	3	3	26	3	2	29	0	0	10	0	0			
	23:00 ~ 06:30	120	0	3	54	3	3	26	3	2	0	0	0	29	0	0			
路口3 中山北 三民路 口																			
	06:30 ~ 09:00	200	0	2	94	3	3	14	3	3	74	3	3						
	09:00 ~ 16:30	150	0	1	74	3	3	14	3	3	44	3	3						
	16:30 ~ 19:00	200	0	4	98	3	3	19	3	3	65	3	3						
	19:00 ~ 23:00	150	0	1	74	3	3	14	3	3	44	3	3						
	23:00 ~ 06:30	120	0	3	54	3	3	14	3	3	34	3	3						
路口2 中山北 福羚路 口																			
	06:30 ~ 09:00	200	0	2	123	4	3	25	3	4	31	3	4						
	09:00 ~ 16:30	150	0	1	77	4	3	23	3	4	29	3	4						
	16:30 ~ 19:00	200	0	2	123	4	3	25	3	4	31	3	4						
	19:00 ~ 23:00	150	0	1	77	4	3	23	3	4	29	3	4						
	23:00 ~ 06:30	120	0	3	52	4	3	23	3	4	24	3	4						
路口1 中山北 梅獅路 口																			
	06:30 ~ 09:00	200	0	2	63	4	3	19	3	3	10	3	2	44	3	3	34	3	3
	09:00 ~ 16:30	150	0	1	53	4	3	9	3	3	0	0	0	34	3	3	29	3	3
	16:30 ~ 19:00	200	0	4	63	4	3	19	3	3	10	3	2	44	3	3	34	3	3
	19:00 ~ 23:00	150	0	1	53	4	3	9	3	3	0	0	0	34	3	3	29	3	3
	23:00 ~ 06:30	120	0	3	43	4	3	9	3	3	0	0	0	24	3	3	19	3	3

資料來源:桃園市政府交通局，本計畫調查

表 5.3-3 楊梅交流道假日時段切分與時制計畫

路口	時段	週期	時差	時制計畫	時相1			時相2			時相3			時相4			時相5		
					G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR
路口9 中山北環 東路口																			
	06:00 ~ 10:00	150	0	5	24	3	3	38	4	3	18	4	3	9	3	3	29	3	3
	10:00 ~ 14:00	180	0	2	49	3	3	39	4	3	18	4	3	9	3	3	33	3	3
	14:00 ~ 19:00	180	0	2	49	3	3	39	4	3	18	4	3	9	3	3	33	3	3
	19:00 ~ 23:00	150	0	5	24	3	3	38	4	3	18	4	3	9	3	3	29	3	3
	23:00 ~ 06:00	120	0	4	14	3	3	38	4	3	18	4	3	0	0	0	24	3	3
路口7 交流道南 匝																			
	06:00 ~ 10:00	150	0	32	88	4	3	48	4	3									
	10:00 ~ 19:00	200	0	33	128	4	3	58	4	3									
	19:00 ~ 23:00	150	0	32	88	4	3	48	4	3									
	23:00 ~ 06:00	120	0	31	70	4	3	36	4	3									
路口6 交流道北 匝																			
	06:00 ~ 10:00	150	0	32	88	4	3	48	4	3									
	10:00 ~ 19:00	200	0	33	128	4	3	58	4	3									
	19:00 ~ 23:00	150	0	32	88	4	3	48	4	3									
	23:00 ~ 06:00	120	0	31	70	4	3	36	4	3									
路口5 中山北裕 成路口																			
	06:00 ~ 10:00	150	0	32	20	0	0	31	0	0	30	3	3	11	3	3	40	3	3
	10:00 ~ 19:00	200	0	33	25	0	0	46	0	0	36	3	3	15	3	3	60	3	3
	19:00 ~ 23:00	150	0	32	20	0	0	31	0	0	30	3	3	11	3	3	40	3	3
	23:00 ~ 06:00	120	0	31	15	0	0	26	0	0	20	3	3	11	3	3	30	3	3

資料來源:桃園市政府交通局，本計畫調查

表 5.3-3 楊梅交流道假日時段切分與時制計畫(續)

路口	時段	週期	時差	時制計畫	時相1			時相2			時相3			時相4			時相5		
					G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR	G	Y	AR
路口4 中山北87巷路口																			
	06:00 ~ 10:00	150	0	1	74	3	3	26	3	2	0	0	0	39	0	0			
	10:00 ~ 19:00	200	0	6	74	3	3	26	3	2	29	0	0	10	0	0			
	19:00 ~ 23:00	150	0	1	74	3	3	26	3	2	0	0	0	39	0	0			
	23:00 ~ 06:00	120	0	3	54	3	3	26	3	2	0	0	0	29	0	0			
路口3 中山北三民路口																			
	06:00 ~ 10:00	150	0	1	74	3	3	14	3	3	44	3	3						
	10:00 ~ 19:00	200	0	4	98	3	3	19	3	3	65	3	3						
	19:00 ~ 23:00	150	0	1	74	3	3	14	3	3	44	3	3						
	23:00 ~ 06:00	120	0	3	54	3	3	14	3	3	34	3	3						
路口2 中山北福羚路口																			
	06:00 ~ 10:00	150	0	1	77	4	3	23	3	4	29	3	4						
	10:00 ~ 19:00	200	0	2	123	4	3	25	3	4	31	3	4						
	19:00 ~ 23:00	150	0	1	77	4	3	23	3	4	29	3	4						
	23:00 ~ 06:00	120	0	3	52	4	3	23	3	4	24	3	4						
路口1 中山北梅獅路口																			
	06:00 ~ 10:00	150	0	1	53	4	3	9	3	3	0	0	0	34	3	3	29	3	3
	10:00 ~ 19:00	200	0	4	63	4	3	19	3	3	10	3	2	44	3	3	34	3	3
	19:00 ~ 23:00	150	0	1	53	4	3	9	3	3	0	0	0	34	3	3	29	3	3
	23:00 ~ 06:00	120	0	3	43	4	3	9	3	3	0	0	0	24	3	3	19	3	3

資料來源:桃園市政府交通局，本計畫調查

5.4 實驗場域模擬模式構建規劃

分別使用專業交通運輸模擬軟體 VISSIM 與開源交通運輸模擬軟體 SUMO，構建實驗場域的模擬模式。並以交通調查資料進行模擬環境參數校估，使模擬器的行為儘量符合實際車流的運作狀況，以供強化學習號誌控制代理人作為探索與學習最佳控制策略的環境。

5.4.1 交通參數與績效指標設計

進行實驗場域模擬模式所需各式交通參數與績效指標設計時應考量人工智慧強化學習號誌控制模型常見之「環境(State)」與「獎勵(Reward)」參數，以下分別說明可設計之交通參數

與績效指標，將依實際執行策略與方向訂定。

一、交通參數

(一)流量：道路或區域中的交通量，通常以小客車當量數(PCU)表示，亦可體現車道通行能力。

(二)速度：交通流中的車輛行駛速度，通常以平均速度或速度分佈表示道路通行狀態。

二、績效指標

(一)旅行時間：用於評估交通流的通行效率。

(二)停等延滯：在模擬中因交通擁堵而停滯，用於評估交通流的流動性。

(三)環境指標：行車油耗之減少量換算二氧化碳減少量之環境改善指標。

5.4.2 交通參數資料蒐集規劃

針對實驗場域除現況道路特性及時制計畫以外，必要之交通資料蒐集說明如下：

一、路口轉向交通量

分車種(大型車、小型車、機踏車)、流向(左轉、直進、右轉)統計調查路口所有通過停止線之車輛數，並以每 5 分鐘紀錄 1 次。

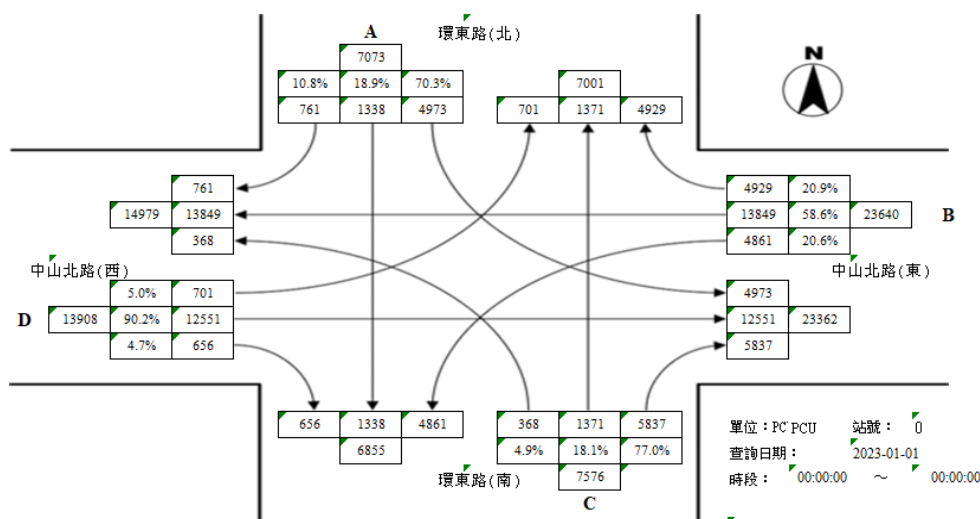
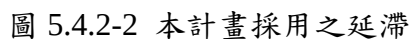


圖 5.4.2-1 路口轉向交通量示意圖-以環東路與中山北路路口為例

二、路口平均延滯

傳統停等延滯調查係人員站在停止線附近，每 15 秒計數停等車輛數，若等候車隊過長將無法於 15 秒內精確計算完所有的停等車輛數，導致在壅塞時停等延滯調查資料淪於調查員之自由心證。而若採用攝影機記錄，亦受限於攝影機架設高度，遠方車輛可能無法於畫面上清

為清楚定義路段延滯範圍，以下透過車輛之軌跡時間圖進行說明(如圖 5.4.2-2)，其中深綠色粗實線為車輛受紅燈影響減速加入停等車隊停等再加速過通路口之軌跡，過去一般停等延滯調查即調查圖中之 AB 段之停等時間。然而整個街廓的 GVP 資料無法判斷車輛停止及行駛各多少時間，因此延滯評估不採用傳統停等延滯。即當車輛駛離上游路口進入目標路段時，其行駛速率會有一段距離不受路口號誌及停等車隊影響，而僅受車流量影響，若維持此速率將可於時間點 C 通過停止線；而實際受號誌影響後通過停止線的時間點為 D，C 點及 D 點之間的時間落差即為號誌帶給車流的整個影響範圍，其中包含車輛減速、停等、再加速產生的延滯，故延滯評估範圍即為 C、D 兩點間之時間間隔。



(一)當街廓大於 250 公尺：以該臨近路口停止線上游 200 公尺至街廓邊界之 GVP 資料推估之。

(二)當街廓小於等於 250 公尺：採用附近街廓之資料，若為幹道則取時比大的街廓資料，而支道則往上游找道路幾何相同且不受號誌影響之路段。

傳統之旅行速率調查係採用浮動車法，調查車輛於車隊中保持車隊之均速行駛。然在號誌化幹道中車輛不同的出發時間點其可續進的路口數不同，即使依交通工程手冊至少蒐集 6 趟資料，亦可能無法真正反應整體車流績效。

166

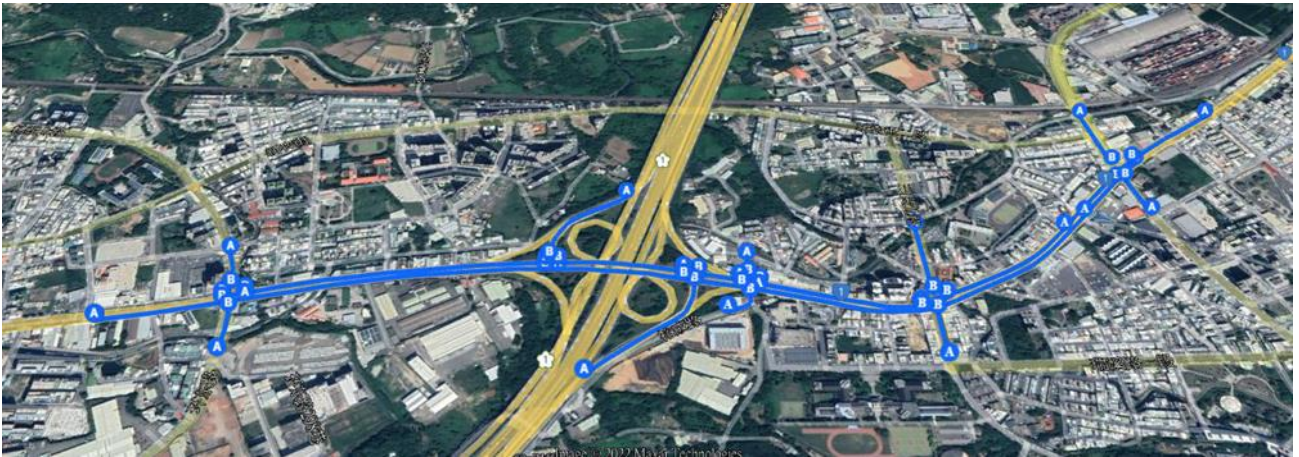


圖 5.4.2-3 GVP 旅行時間蒐集路段示意圖-以楊梅交流道為例

5.4.3 模擬模式建構設計

本案交通模擬模型分別以 VISSIM 與 SUMO 建立模型，模擬模型建置如圖 5.3-1 及 5.4.3-2。依照人工交通調查數據輸入相關交通需求、轉向比例、車種比例及運行之號誌時制計畫，並進行校估工作。模擬模型建置過程包含以場域衛星空照圖做為底圖做為背景，搭配比例尺確認路網距離與實際一致。

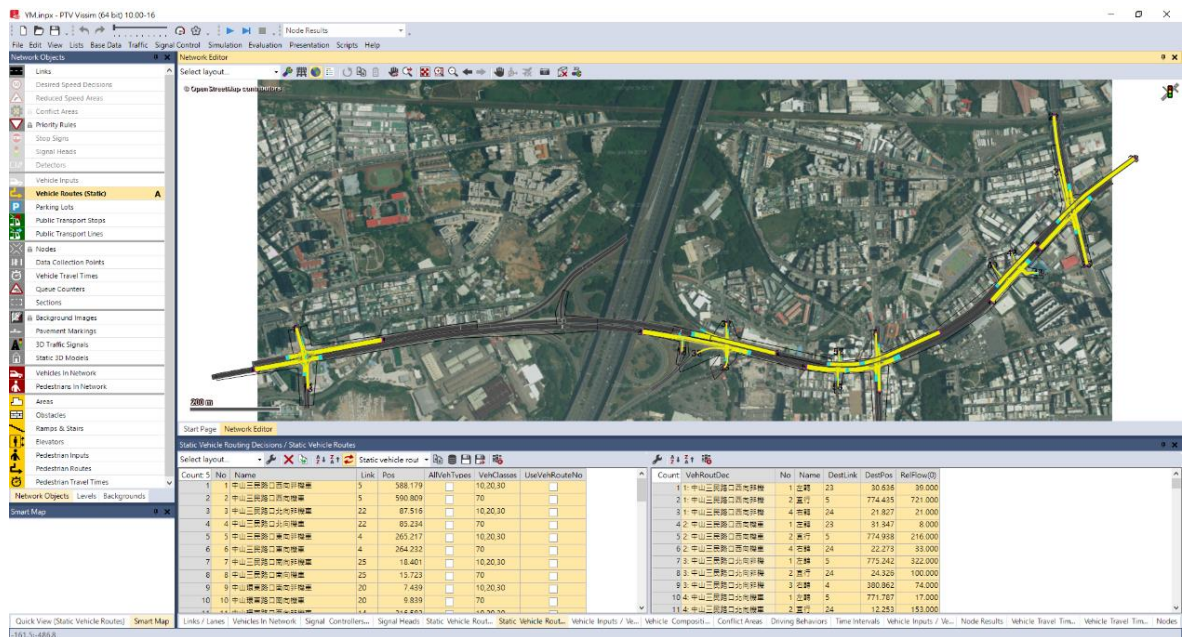


圖 5.4.3-1 楊梅交流道場域 VISSIM 建模搭配空拍底圖範例

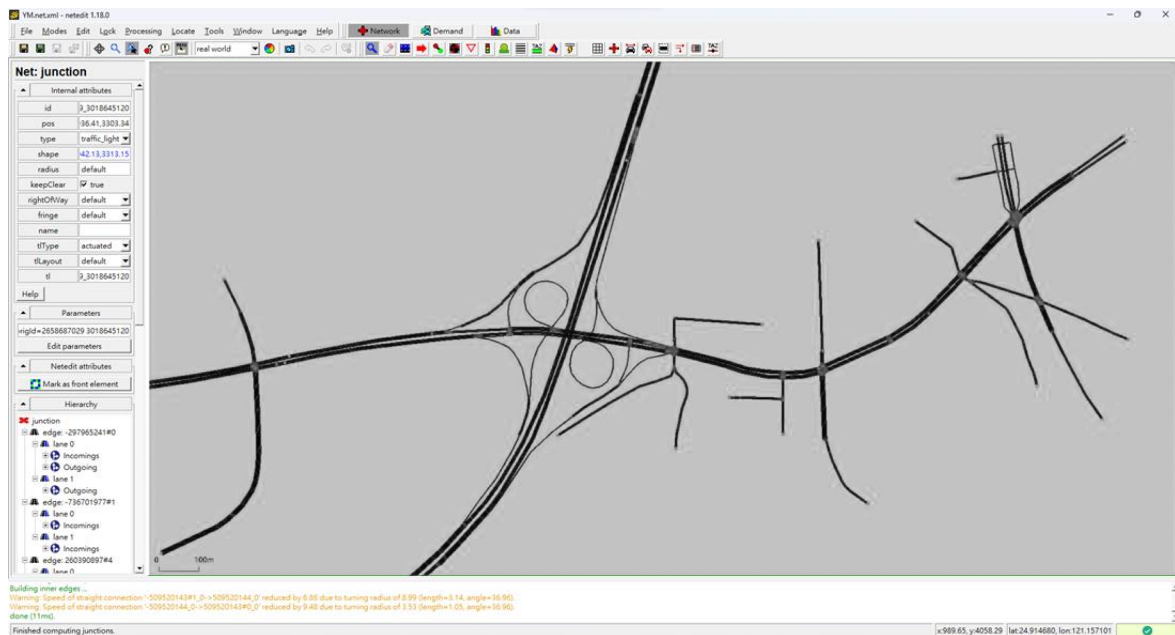


圖 5.4.3-2 楊梅交流道場域 SUMO 建模範例

透過空拍底圖可以清楚的看到道路連結的幾何線型，再搭配實地場勘、調查影片或是街景圖等方式可以更詳細的了解場域的車道佈設、號誌運作方式等等。通過以上資料可以將場域路網結構繪製建模。

在號誌運作方面，根據號誌時制計畫設定各號誌時相動線與秒數以及分時段執行不同號誌時制計畫(TOD)。號誌時制動線設定在 SUMO 中可以圖像化方式設定路口節點中各時相動線的連接方式，也可以採用文字矩陣方式設定各燈頭的燈號、秒數等等，如圖 5.4.3-2 所示。而在 VISSIM 中則是類似燈頭設定的方式，以圖像化的方式設定號誌控制器中所有號誌燈頭在號誌週期中的燈號與秒數，如圖 5.4.3-3 所示。

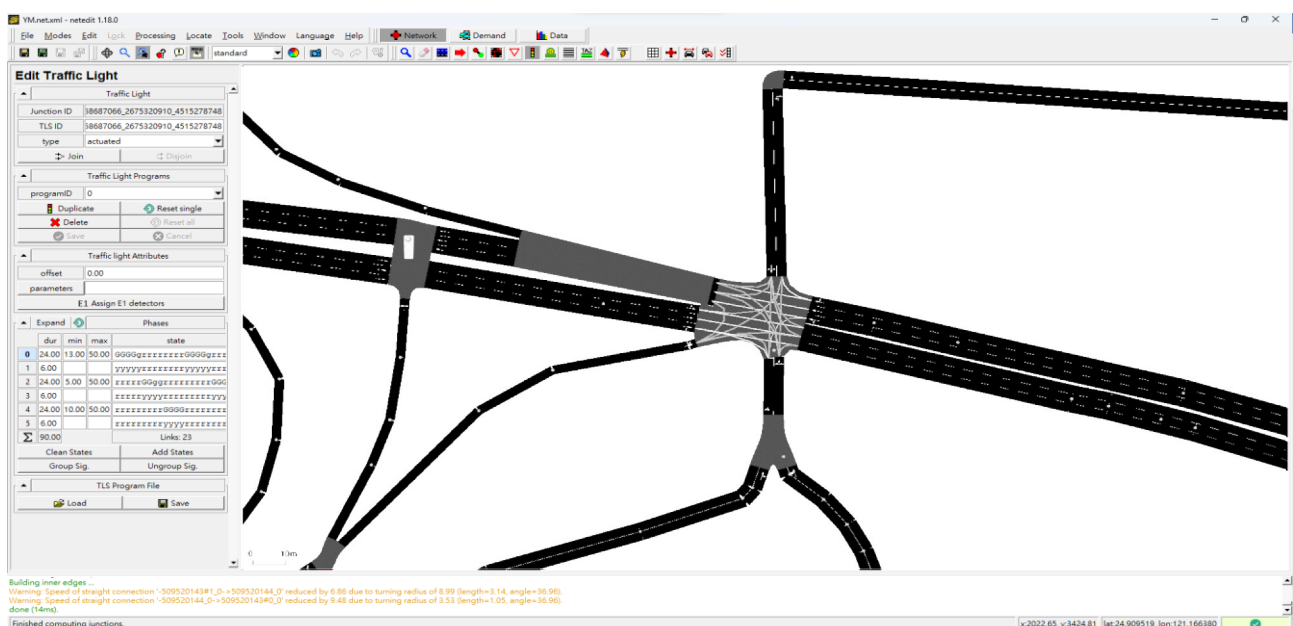


圖 5.4.3-3 SUMO 圖像化時相動線設定及文字矩陣時相動線設定

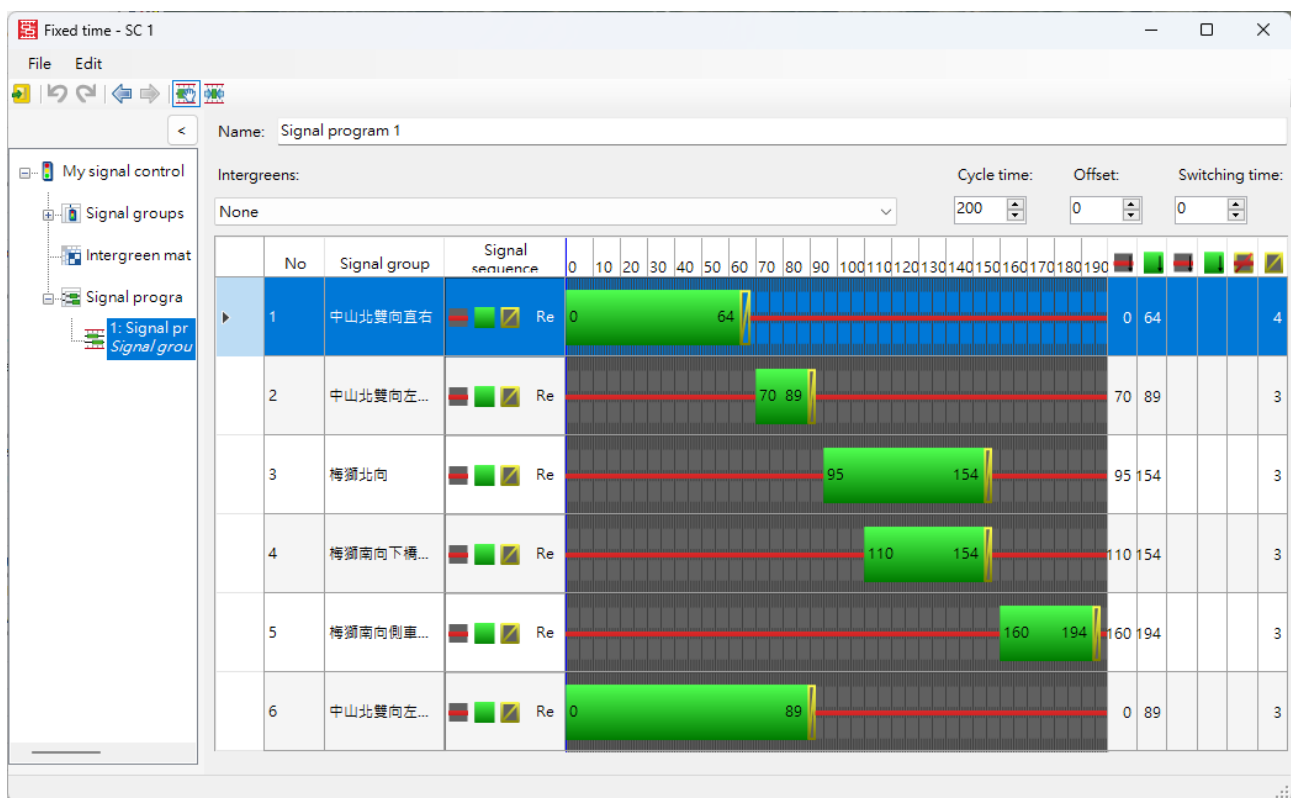


圖 5.4.3-4 VISSIM 中號誌時制設定範例

在完成基本的模擬模型結構與號誌運作設定後，接著須設定交通需求相關的設定。在 SUMO 中須透過計算車輛在發車點出現的機率值完成交通需求量的設定，而該機率值則是通過交通調查資料換算所得。而在 VISSIM 中則較為直觀，僅需設定在時段內預期出現的車輛數量，軟體即會以 Poisson 分配轉換發車機率值讓車輛到達。



圖 5.4.3-5 SUMO 交通需求設定範例

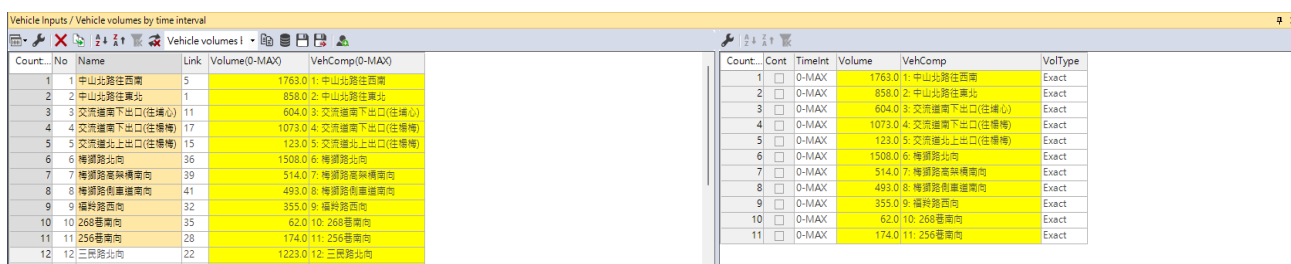


圖 5.4.3-6 VISSIM 交通需求設定範例

完成車輛需求設定後，最後需要設定各路口的轉向比例，在這方面兩套模擬軟體作法基本相同，皆為設定各轉向的機率數值，而在模擬過程中車輛即會隨機的決定在通過路口時的轉向動線，轉向比例同樣是從交通調查資料得到。至此，模擬模型基本的結構與運作方式大致底定。接下來須通過誤差模擬驗證的過程微調駕駛人行為參數，以確保最終的模擬行為與現況場域交通表現一致。

5.4.4 模擬模式校估驗證

交通模擬方法是評估交通影響衝擊與改善方案的有效做法，在有限的資源下盡可能的獲得量化的方案評估結果。然而為了確保評估方案與現實交通環境能有相同的表現行為，模擬器的校準過程相當重要。以下參考國內外採用模擬方法之建議手冊與案例，將校估流程按步驟說明。

一、模擬場域資料蒐集

路網建置需蒐集相關交通資料，一般可分為靜態資料與動態資料，說明如下。

(一)靜態資料：路網構建階段輸入之資料多為靜態道路設施資料，不隨模擬過程變動，可由道路幾何配置調查、道路工程圖面獲得，項目如下。

- 1.道路寬度、車道數、車道寬度
- 2.轉向、車道縮減、或車道增加
- 3.大眾運輸場站位置與長度
- 4.停止線、號誌燈頭的位置
- 5.以及與號誌控制時相之關聯
- 6.「停」、「讓」標誌之位置
- 7.機車停等區和機車待轉區尺寸和位置

(二)動態資料：完成路網建構後，須進行車流特性等相關動態資料設定，動態資料會隨模擬執行過程變化，主要為輸入交通組成與交通量。交通組成係定義輸入交通量中，各類型運具所佔比例；交通量則依賴路網端點之上游路口轉向交通量，模擬系統根據兩項資料產生路網中相對應之各車種車輛數。相關動態資料可能包括：

- 1.路徑決策點的位置，以及時間與車種之決策情形
- 2.非號誌化路口或路口允許左轉之優先控制
- 3.大眾運輸之路徑、班次、班距與停靠站時間

4.進入路網之交通量與車種組成

二、模擬路網建置

路網之構建除道路幾何之劃設之外，還包括路口管制設定、交通量之輸入及基本參數之設定。

(一)道路構建與路口管制方式設定

- 1.構建道路路網基本幾何線型：以背景圖為底，於圖上構建研究所需範圍之道路線型及車道數。
- 2.輸入車道寬度與各車道之行駛車輛限制
- 3.設定各路段之連結路徑
- 4.設定路口控制方式(號誌時制計畫)
- 5.設定績效指標偵測器

(二)交通量輸入

- 1.設定車種組成
- 2.輸入路網各路段之端點交通量
- 3.設定路口轉向流量比例

(三)參數設定

- 1.車輛幾何(車長、車寬、車重)
- 2.車輛性能(最大加/減速率、期望加/減速率、期望速率分配)
- 3.車輛種類(小汽車、公車、大貨車、機車、腳踏車)
- 4.駕駛行為參數(跟車參數、變換車道參數、側向參數)

三、參數校估

(一)校估指標選取

AI 強化學習模式的訓練過程是透過大量與環境互動的經驗累積所得出的決策優化過程。而此一與環境的互動在實務上不宜直接讓 AI 強化學習模式在真實的交通環境下進行訓練，因此如何創建一個與真實交通狀態近似的互動環境對於 AI 強化學習模式的訓練可謂至關重要的。透過設定模擬軟體(VISSIM 與 SUMO)中甚多參數可供使用者根據經驗或校估結果輸入，如車流期望速度、車輛最大加減速度、接受間距等。各項參

數皆有軟體提供之預設值，然而為了使模擬更貼近實際狀況，如何正確的輸入這些參數方能反映真實車流情況，使模擬績效指標值與現場調查績效指標值誤差能在接受範圍內就極為重要。

驗證的項目主要指標為轉向交通量(流量)，計算方式為計算各方向、各轉向於 2 小時內的通過 pcu 數量與實際調查值以平均絕對百分比誤差(MAPE)做為量化誤差指標。而由於文獻回顧中較少採用此項指標進行驗證工作，因此在本案中將平均每車延滯做為輔助驗證指標。

本計畫亦參考了近年來國內使用 VISSIM 的相關經驗做為初始參數設定，以加速模擬校估的工作進度。為了確保模擬結果能夠反應實際交通行為，參考國內外相關案例後整理如表 5.4.4-1 所示，設定本計畫參數組必須滿足以下門檻：

- 1.轉向交通量之 MAPE≤10%
- 2.總停等延滯 MAPE≤20%
- 3.旅行時間 MAPE≤15%

從模擬校估相關文獻回顧當中可以發現到，絕大部分的微觀模擬仍是以交通流量作為主要的校估指標，在延滯部分則是受到實務上調查困難，調查自身數據存在一定程度誤差，故較少文獻使用。

表 5.4.4-1 國內外模擬校估績效指標及範圍

國內外案例	採用校估績效指標		
	平均流量	平均旅行時間 (or 平均速率)	平均 停等延滯
107 年，交通改善措施微觀車流模擬成效評估案	< 10%	--	--
102 年，交通技術研發與人才培育規劃研究(4/4)	< 20%	--	< 30%
96 年，公路工程設計輔助工具引進與應用之研究	< 10%	< 15%	--
105 年，台 9 線蘇花改通車後對蘇澳地區交通衝擊影響評估分析	< 20%	< 20%	--
107 年，大新竹運輸走廊整合道路交通與電信資訊應用	< 20%	--	--
Calibration of VISSIM for Indian Heterogeneous Traffic Conditions, 2013	< 10%	--	--
Core of Rosslyn Transportation Study Existing Conditions Report, 2018	< 15%	< 15%	--
Calibration of VISSIM and QuickZone programs for work zone traffic impact analysis, 2015	--	< 15%	< 15%

資料來源：本計畫整理

而採用 MAPE 作為驗證統計量經常受到的批評為當部分轉向或路段的調查值本身數值較低時，容易很大程度的影響 MAPE 的數值，導致其評估結果有失客觀；反之，當調查數值極大時，MAPE 也容易低估，造成配適良好的假象。在文獻回顧中，為避

免此種狀況，亦常見到以 GEH 統計量(Geoffrey E. Havers)來做為衡量模型是否為良好適配的指標。GEH 的計算方式如下：

$$GEH = \sqrt{\frac{2(M - C)^2}{M + C}}$$

其中，M 為模型輸出數值；C 為實際觀測數值。

由於 GEH 並非如 MAPE 可對於不同項目直接進行平均得到單一衡量數值，因此參考美國內華達州運輸部門所建議的做法，至少 75%的轉向交通量之 GEH 需小於 5。

表 5.4.4-2 國內外模擬校估績效指標及範圍

GEH 接受門檻(兩項皆需滿足)
GEH<5，對於至少 85%的路段
GEH<10，對於至少 100%的路段
GEH<5，對於至少 75%的轉向量
GEH<10，對於至少 95%的轉向量

而 GEH 統計量則可以一定程度的避免由於數值偏小造成的偏誤狀況。GEH 統計量在極大數值與極小數值的狀況下，可以有效的避免偏誤的狀況發生，表 5.4.4-3 中說明即使誤差百分比都在 10%的狀況，絕對誤差量仍可能非常高；反之在誤差量低的情況則也可能有百分比極高的狀況發生。而表中 GEH 統計量相較於誤差百分比則顯得更為客觀評估模擬與實際的誤差狀況。

表 5.4.4-3 GEH 和誤差百分比的比較分析

M	C	GEH	%Difference
10000	9000	10.3	10%
100	900	3.2	10%
100	90	1.0	10%
10000	9520	4.9	5%
100	850	4.9	18%
100	57	4.9	75%

本計畫已完成平日晨峰資料校估之模擬路網比對交通調查數值，其確各路口模擬誤差驗證如表 5.4.4-4 所示。

表 5.4.4-4 VISSIM 楊梅場域流量 GEH 及 MAPE

路口	臨近路口方向	方向	調查資料	模擬資料	MAPE 指標	GEH 指標
			通過車輛數 (輛/小時)	通過車輛數 (輛/小時)		
中山北路/ 梅獅路	中山北路往西南	↙	1,700	1,573	7.47%	3.14
	梅獅路往北	↑	1,441	1,300	9.78%	3.81
	中山北路往東北	↗	1,395	1,023	26.67%	10.70
	陸橋往南	↓	607	498	17.96%	4.64
	陸橋側車道往南	↓	506	486	3.95%	0.90
中山北路/ 三民路	中山北路往西南	↙	1,710	1,587	7.19%	3.03
	三民路往北	↑	1,323	1,283	3.02%	1.11
	中山北路往東北	↗	1,756	1,712	2.51%	1.06
	三民路往南	↓	707	769	8.80%	2.29
中山北路/ 裕成路	中山北路往西南	←	1,585	1,583	0.13%	0.05
	裕成路北向	↑	264	281	6.44%	1.03
	中山北路往東北	→	1,507	1,449	3.85%	1.51
	交流道北上出口	↗	302	272	9.93%	1.77
	裕成路南向	↓	407	355	12.78%	2.66
中山北路/ 交流道北閘	中山北路往西南	←	1,781	1,533	13.92%	6.09
	交流道北上出口	↑	147	125	14.97%	1.89
	中山北路往東北	→	1,690	1,563	7.51%	3.15
中山北路/ 交流道南閘	中山北路往西南	←	1,894	1,745	7.87%	3.49
	中山北路往東北	→	1,755	1,563	10.94%	4.71
	交流道南下出口	↓	642	547	14.80%	3.90
中山北路/ 環東路	中山北路往西南	↙	2,037	2,355	15.61%	6.79
	環東路北向	↑	531	523	1.51%	0.35
	中山北路往東北	↗	859	855	0.47%	0.14
	環東路南向	↓	501	456	8.98%	2.06
平均					9.0%	

表 5.4.4-5 SUMO 楊梅場域流量 GEH 及 MAPE

路口	臨近路口方向	方向	調查輸入	模擬輸出	MAPE 指標	GEH 指標
			通過車輛數 (輛/小時)	通過車輛數 (輛/小時)		
中山北路/ 梅獅路	中山北路往西南	↙	1,700	1,641	3.47%	1.44
	梅獅路往北	↑	1,441	1,304	9.51%	3.70
	中山北路往東北	↗	1,395	1,013	27.38%	11.01
	陸橋往南	↓	607	518	14.66%	3.75
	陸橋側車道往南	↓	506	509	0.59%	0.13
中山北路/ 三民路	中山北路往西南	↙	1,710	1,563	8.60%	3.63
	三民路往北	↑	1,323	1,158	12.47%	4.68
	中山北路往東北	↗	1,756	1,672	4.78%	2.03
	三民路往南	↓	707	819	15.87%	4.06
中山北路/ 裕成路	中山北路往西南	←	1,585	1,453	8.33%	3.39
	裕成路北向	↑	264	251	4.92%	0.81
	中山北路往東北	→	1,507	1,461	3.05%	1.19
	交流道北上出口	↗	302	273	9.60%	1.71
	裕成路南向	↓	407	368	9.58%	1.98
中山北路/ 交流道北開	中山北路往西南	←	1,781	1,598	10.28%	4.45
	交流道北上出口	↑	147	135	8.16%	1.01
	中山北路往東北	→	1,690	1,543	8.70%	3.66
中山北路/ 交流道南開	中山北路往西南	←	1,894	1,604	15.31%	6.93
	中山北路往東北	→	1,755	1,333	24.05%	10.74
	交流道南下出口	↓	642	563	12.31%	3.22
中山北路/ 環東路	中山北路往西南	↙	2,037	2,221	9.03%	3.99
	環東路北向	↑	531	498	6.21%	1.45
	中山北路往東北	↗	859	866	0.81%	0.24
	環東路南向	↓	501	513	2.40%	0.53
平均					9.6%	

(二)參數調校

參數調整時並無一定程序，但應一次調整單一參數，藉以了解各參數之影響力，調整後若問題還是無法解決，再以組合方式調整兩項以上之參數，但原則須確定其相互作用所產生之影響必須是要對符合實際車流情況有所貢獻。重複以上程序直到誤差小於門檻為止。

(三)路網微調

此步驟可於參數調校後仍無法令模擬車流與實際車流的誤差小於所設定門檻時進

行。例如普遍性的違規(如跨越雙白線)等，雖於交通規則並不允許，然為忠實反映實際狀況，仍應將此一情形透過路網微調步驟修正。

5.5 小結

綜觀以強化學習決策號誌控制策略相關研究，多數模型選址為交流道場域，其因高速公路與平面道路兩系統之車流，在交流道區域互相牽制，如能改善平面道路及匝道儀控之協控車流問題，可有效優化交流道瓶頸區域之疏解效能。實驗場域之選址，則根據發展人工智慧號誌控制模式需求，自六都中遴選一處交流道作為實驗場域，需考量模擬模式與模型構建所需資料，以及後續實測現場資通訊軟硬體可配合等因素。楊梅交流道場域由國道一號與臺一線組成，該區域無外環道路，亦無鄰近都會高架快速道路系統，故此交流道為民眾進出楊梅區域的重要路段；車流量集中易壅塞，評估此協控系統具優化空間。

交流道場域具有兩交通系統，由於交控中心隨車流情況變化儀控率，且儀控率非固定周期，異於平面號控系統，故 AI 模型對兩者之決策頻率也不相同。決策的頻率則為各自的週期結束時決策，決策的結果同時輸出下個匝道綠燈秒數、平面號誌週期秒數及分相比例；現場號誌則視何種號控週期結束而選擇 AI 輸出的結果，例如：現場平面號誌週期結束時，則取 AI 輸出中平面號誌的結果，反之則取 AI 輸出中匝道儀控的結果。

不同於本期都會道路(臺北及臺南場域)RL 模型訓練，下一階段交流道場域訓練模型規劃為共同訓練、分別執行之 AI 模型架構(Centralized Training Decentralized Execution, CTDE)，主要為針對匝道及平面道路訓練不同策略，亦可共用同一個 state 資訊。中央化訓練(CT)使得平面號誌與匝道儀控使用同一個 policy network，增加了訓練效率，並且共同閱讀平面及主線之道路資訊，提高模型訓練之準確性；分別執行(DE)則因應了協控中不同系統需在不同頻率下決策。

訓練模型之過程，交通模擬模型是不可或缺之重要角色，給予現場實測之前可能的交通流表現，提前設定觀察點及擬定因應策略。本次場域使用 VISSIM 及 SUMO 兩建模軟體進行建模，並以交通調查資料進行模擬環境參數校估，使模擬器儘量符合實際車流的運作狀況。在考量強化學習模型部分，在交通模型中將蒐集路段車流量及速度作為狀態(state)，路段旅行時間及停等延滯為交通績效評估(performance)。針對交通模型的仿真度校估，則以 MAPE 及 GEH 兩項指標，並使用路口調查流量為指標進行校估。兩項指標校估結果皆顯示不論是 VISSIM 亦或 SUMO 模型，皆符合國內外建模誤差接受規範，為可使用之模擬模型。

下一階段強化學習模型預計在 CTDE 架構下進行訓練，以楊梅交流道為場域，透過模擬軟體建構環境，將路段車流量、速度、號誌燈態作為狀態，並以路段通過量、停等延滯為獎勵，policy 將會輸出動作，為下個周期之匝道綠燈秒數以及平面號誌之週期秒數及分相比。CTDE 以中央化訓練提高效率，分別執行以因應不同決策頻率。透過狀態轉移、獲得的回饋，模型不斷優化號誌控制策略，以期提升交通疏解量能。

第六章 結論與建議

本計畫延續去(111)年「我國人工智慧車聯網之號誌控制模式探討」研究計畫成果，擴充導入多目標的人工智慧強化學習號誌控制模型發展，並於臺北市的「中山北路-德行東路」3個路口幹道與臺南市的「台 86-19 甲」單一路口進行實測；計畫同時納入高速公路與平面道路號誌協控課題，遴選桃園市楊梅交流道為實驗場域，並發展模擬模式與探討人工智慧強化學習號誌協控模型。都會區幹道號誌控制部份，本計畫採用之人工智慧號誌控制方法，延續前期以強化學習(Reinforcement Learning, RL)為方法論，導入多目標 Q-Pensieve 演算法進行強化學習號誌控制模型優化，優化內容包括多目標獎勵(Reward)，以及不同的幹道與支道優先權重策略，並於臺北市的「中山北路-德行東路」3個路口幹道與臺南市的「台 86-19 甲」單一路口進行實測。計畫同時納入高速公路與平面道路號誌協控課題，遴選桃園市楊梅交流道為實驗場域，並發展模擬模式與探討人工智慧強化學習號誌協控模型，優先構建交流道區域號誌協控車流模擬環境、以及進行分散式 AI 強化學習模式號誌協控(多代理人機制)初步發展設計。

6.1 結論

- 1.在人工智慧強化學習(Reinforcement learning, RL)號誌控制應用上，本計畫採用在連續動作控制領域中的多目標強化學習方法 Q-Pensieve，該方法可視為多目標強化 SAC 演算法，並針對 Q 值估算進行加強，啟用與 SAC 一樣的模型，並繼承 SAC 的優點，除了增加策略的探索性與隨機性，確保策略成長性和支持離線策略學習，也透過知識共享，有效率解決多目標強化學習效能議題。
- 2.本計畫所提人工智慧強化學習號誌控制模型優化與精進上，能依據各路口各方向交通變化，即時計算下個階段的號誌時制內容；在臺南市的「台 86-19 甲」單一路口案例，即時計算下一個時相號誌綠燈秒數，在臺北市的「中山北路-德行東路」3個路口幹道則以單一代理人控制 3 個路口，以共同號誌周期長度與不同時比方式來進行動態調整，以兼顧車流續進。實測結果顯示，臺南市及臺北市實驗場域依據不同偏好進行決策，且在各個時段中均能找到一組偏好，大部分的績效比固定時制為佳；另多任務 AI 強化學習模式模型即使將不同時段資料一起進行訓練，雖然尚無法比分開訓練的模型績效為佳，但整體表現已接近。
- 3.在臺南市與臺北市實測的事前事後績效分析方面，
 - (1)臺南市實驗場域平日時段，晨昏峰路口整體延滯減少 5.4 秒~7.3 秒，服務水準自 C 級升到 B 級，離峰時段，路口整體延滯略微增加 0.7 秒，但服務水準仍然維持 B 級水準。假日昏峰 11 月 4 日(六)較 11 月 11 日(六)路口整體延滯增加 16.1 秒，服務水準自 B 級降到 C 級；11 月 5 日(日)較 11 月 12 日(日)路口整體延滯增加 8.3 秒，服務水準自 B 級降到 C 級，兩天假日 AI 強化學習模型表現較 TOD 表現略差。平日 11 月 7 日(二)AI 號誌控制晨峰及昏峰表現都略優於 11 月 14 日，隔一天 11 月 8 日(三)AI 號

誌控制晨峰出現兩處峰期，且昏峰表現也較 11 月 15 日略差；假日 AI 號誌控制旅行時間表現皆略較 TOD 表現為差。此情形亦對應在延滯表現上，唯路口服務水準尚維持在 C 級以上，車流運行尚算通順。

- (2)臺北市實驗場域於平日 11 月 16 日(四)AI 號誌控制服務水準皆與 11 月 23 日維持一樣的水準，惟平日 11 月 20 日(一) 195 巷及德行西路晨峰服務水準分別自 A 級降至 B 級，B 級降至 C 級，但昏峰德行西路服務水準自 D 級提升至 C 級，路口延滯減少 9.5 秒。假日昏峰 AI 號誌控制平均延滯皆較 TOD 略差，但服務水準大多維持不變，僅 11 月 19 日(日)195 巷及德行西路服務水準分別從 A 降到 B、C 降到 D。平日 11 月 16 日(四)AI 號誌控制除晨峰因系統、設備多次斷線導致旅行時間較 TOD 高外，離峰及昏峰表現雖略差，但都貼合 TOD 趨勢。假日兩天旅行時間績效 AI 號誌控制皆略差於 TOD。

4.針對交流道區域的匝道儀控與平面路口強化學習號誌協控模型發展部分，

- (1)本計畫導入多號誌控制代理人機制，並於績效評估上額外考量駕駛人感受等相關指標，考慮將匝道和平面道路的輸入資料集中到一個中央化代理人(agent)的「中央化訓練(Centralized Training)」，以及將完成訓練的機器學習模型，針對匝道與平面道路提供策略函數，並使用其對於給予的環境(state)做動作(action)的「分散式控制(Decentralized Execution)」方式進行。
- (2)在號誌協控實驗場域選擇上，本計畫遴選桃園楊梅交流道，主要考量為楊梅市區欠缺外環道，民眾欲北上或南下大多須經楊梅交流道，導致車流量集中，尖峰時段沿線易壅塞，且北側鄰近幼獅工業區、南側鄰近竹科龍潭園區，因此南北向大車多，而東西向台 1 線銜接中壢市區及大興工業區甚至新竹工業區，為重要交通匯集路口，因此可以就協控模型進行探討與測試。
- (3)在模擬環境建置與校估上，本計畫採用 VISSIM 與 SUMO 兩套交通模擬軟體來建立桃園楊梅交流道周邊路網模型，包含台 1 線中山北路一段雙向(梅獅路至環東路間)為進出國道 1 號楊梅交流道即前往埔心及楊梅方向之重要道路，劃定號誌化路口 9 處，以及高速公路楊梅交流道鄰近主線區域。而為了確保模擬環境與現實交通環境能有相同的表現行為，參考國內外採用模擬方法之建議手冊與案例，以流量 MAPE 與 GEH 進行模擬環境校估。結果顯示以 VISSIM 建置模擬環境其流量 MAPE 為 9.0%，GEH 小於 5 為 88%，GEH 小於 10 為 96%；以 SUMO 建置模擬環境其流量 MAPE 為 9.6%，GEH 小於 5 為 88%，GEH 小於 10 為 92%。符合前述建議之至少 75%的轉向交通量之 GEH 需小於 5 之目標。

6.2 建議

- 1.在都會區強化學習號誌控制模型的演算法、輸入狀態、輸出決策探討上，建議增加時制計畫

改變的因應方式，並納入調整時相順序、與鄰近群組連鎖、時差續進之因素，以使人工智慧強化學習在號誌控制應用上更具彈性與全面性。

2. 目前強化學習號誌控制模型受限於不同號誌時段與時相組態，須各自建模與訓練學習，本計畫所嘗試使用可結合不同時段狀態，同時進行訓練的多任務強化學習模式，在模擬環境下，雖然尚無法比個別訓練的強化學習模型績效為佳，建議持續進行此多任務 AI 強化學習模式的實務應用探討，以期未來能發展為泛用型(全時段)的 AI 強化學習號誌控制模型。
3. 根據實作經驗，現場端整合與調校各式路側設備議題亦為重要。為滿足人工智慧強化學習號誌控制所需之偵測參數，須提高現場端車流偵測器設備準確度，降低資料利用錯誤率。另一方面，使用之號誌控制器韌體版本若能更新至最新，將可達到近似逐秒控制效能，並且為確保人工智慧強化學習號誌控制模式能即時與穩定運作，則須強化通訊設備連線率。並可根據實測經驗，調整程式運作邏輯，預先提出設備斷線或資料缺漏處理機制，以增加程式容錯能力與穩定性。建議後續可透過協調持續進行本年度實驗場域之實測，強化實測控制邏輯、減少設備影響程度、延長實測時間，已取得 AI 強化學習模型較長期且穩定的績效評估。
4. 國內行人交通安全議題日益受到重視，然目前各式號誌控制邏輯均僅考量車流因素，後續建議可嘗試將常用之行人專用時相、行人早開、行人對同向右轉車流的影響等納入 AI 強化學習號誌控制模型的探討。
5. 探討 AI 強化學習號誌控制在模型與學習訓練之維護機制，並探討設備通訊與偵測穩定度、設備偵測方式差異、偵測器類型與數量簡化對於號誌控制績效之影響，期能移植成熟模式部分至其他需求場域，藉以提高都市路網運作效率及交通安全。
6. 在交流道區域強化學習號誌協控方面，
 - (1) 針對人工智慧強化學習號誌協控模型進行構建，並探討單一代理人、多代理人之差異，比較集中控制與 CTDE (集中訓練與分散控制)不同 AI 強化學習模型之優缺點與適用情境。
 - (2) 實作人工智慧強化學習號誌協控模型，並透過實測進行分析相關量化效益，且依據本計畫執行經驗，建議延長實測時段，以獲得較為精確之 AI 強化學習模式評估結果。

參考文獻

1. Abdulhai, B., Pringle, R., & Karakoulas, G. J. (2003). Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering*, 129(3), 278-285.
2. Abels, B., Roijers, D. M., Lenaerts, T., Nowe, A., & Steckelmacher, D. (2019). Analyzing the impact of opponent models in zero-sum games. *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*, 1773-1775.
3. Ahilan S, Dayan P (2019) Feudal multi-agent hierarchies for cooperative reinforcement learning.
4. Arel, I., Liu, C., Urbanik, T., & Kohls, A. G. (2010). Reinforcement learning based multi-agent system for network traffic signal control. *IET Intelligent Transport Systems*, 4(2), 128-135.
5. BELLMAN, R. (1957). A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6(5), 679–684. <http://www.jstor.org/stable/24900506>
6. Brys, T., Pham, T. T., & Taylor, M. E. (2014). Distributed learning and multiobjectivity in traffic light control. *Connection Science*, 26(1), 65-83.
7. Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *International Conference on Machine Learning*, 2017. 1, 2, 3, 5, 6, 8, 9.
8. Chu, T., Wang, J., Codecà, L., & Li, Z. (2019). Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, 21(3), 1086-1095.
9. Da Silva, B. C., Basso, E. W., Perotto, F. S., C. Bazzan, A. L., & Engel, P. M. (2006). Improving reinforcement learning with context detection. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems* (pp. 810-812).
10. David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International Conference on Machine Learning*, 2014.
11. Dresner, K., & Stone, P. (2006). Multiagent traffic management: Opportunities for multiagent learning. In *International Workshop on Learning and Adaption in Multi-Agent Systems* (pp. 129-138). Springer, Berlin, Heidelberg.
12. El-Tantawy, S., Abdulhai, B., & Abdelgawad, H. (2013). Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto. *IEEE Transactions on Intelligent Transportation Systems*, 14(3), 1140-1150.
13. Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 29.
14. Foerster, J., Nardelli, N., Farquhar, G., Afouras, T., Torr, P. H., Kohli, P., & Whiteson, S. (2017, July). Stabilising experience replay for deep multi-agent reinforcement learning. In *International Conference on Machine Learning* (pp. 1146-1155). PMLR.

15. Genders, W., & Razavi, S. (2016). Using a deep reinforcement learning agent for traffic signal control. arXiv preprint arXiv:1611.01142.
16. Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 55(2), 895-943.
17. Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., ... & Levine, S. (2018). Soft actor-critic algorithms and applications. arXiv preprint arXiv:1812.05905.
18. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018, July). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning* (pp. 1861-1870). PMLR.
19. Houli, D., Zhiheng, L., & Yi, Z. (2010). Multiobjective reinforcement learning for traffic signal control using vehicular ad hoc network. *EURASIP journal on advances in signal processing*, 2010, 1-7.
20. Hung, W., Huang, B. K., Hsieh, P. C., & Liu, X. (2022). Q-Pensieve: Boosting Sample Efficiency of Multi-Objective RL Through Memory Sharing of Q-Snapshots. arXiv preprint arXiv:2212.03117.
21. <https://spinningup.openai.com/en/latest/algorithms/ppo.html>
22. Janner, M., Fu, J., Zhang, M., & Levine, S. (2019). When to trust your model: Model-based policy optimization. *Proceedings of the 36th International Conference on Machine Learning*, 3158-3167.
23. Kong X, Xin B, Liu F, Wang Y (2017) Revisiting the master-slave architecture in multi-agent deep reinforcement learning.
24. Kumar S, Shah P, Hakkani-Tür D, Heck LP (2017) Federated control with hierarchical multi-agent deep reinforcement learning.
25. Kyriakis, P., Deshmukh, J., & Bogdan, P. (2022;). Pareto policy adaptation. In *International Conference on Learning Representations(ICLR)*.
26. Li, L., Lv, Y., & Wang, F. Y. (2016). Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica*, 3(3), 247-254.
27. Li, Z., Yu, H., Zhang, G., Dong, S., & Xu, C. Z. (2021). Network-wide traffic signal control optimization using a multi-agent deep reinforcement learning. *Transportation Research Part C: Emerging Technologies*, 125, 103059.
28. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. *Proceedings of the 4th International Conference on Learning Representations (ICLR)*.
29. Lu, C., Huang, J., Deng, L., & Gong, J. (2017). Coordinated ramp metering with equity consideration using reinforcement learning. *Journal of Transportation Engineering, Part A: Systems*, 143(7), 04017028.
30. Mannion, P., Duggan, J., & Howley, E. (2016). An experimental review of reinforcement learning algorithms for adaptive traffic signal control. *Autonomic road transport support systems*, 47-66.
31. Mohammad Noaeen, Atharva Naik, Liana Goodman, Jared Crebo, Taimoor Abrar, Zahra

- Shakeri Hossein Abad, Ana L.C. Bazzan, Behrouz Far (2022). Reinforcement learning in urban network traffic signal control: A systematic literature review. *Expert Systems with Applications*.
32. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
 33. Mossalam, S., Assael, Y. M., Roijers, D. M., & Whiteson, S. (2016). Multi-agent common knowledge reinforcement learning. *Proceedings of the 2016 Conference on Advances in Neural Information Processing Systems*, 3849-3857.
 34. Mousavi, S. S., Schukat, M., & Howley, E. (2017). Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems*, 11(7), 417-423.
 35. Nishi, T., Otaki, K., Hayakawa, K., & Yoshimura, T. (2018). Traffic signal control based on reinforcement learning with graph convolutional neural nets. In *2018 21st International conference on intelligent transportation systems (ITSC)* (pp. 877-883). IEEE.
 36. Shabestary, S. M. A., & Abdulhai, B. (2018). Deep learning vs. discrete reinforcement learning for adaptive traffic signal control. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp.286-293). IEEE.
 37. Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015). Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897.
 38. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*
 39. Sutton, R.S. and Barto, A.G. (1998) *Reinforcement Learning: An Introduction*, Vol. 1. MIT Press, Cambridge.
 40. Sunehag P, Lever G, Gruslys A, Czarnecki WM, Zambaldi V, Jaderberg M, Lanctot M, Sonnerat N, Leibo JZ, Tuyls K, Graepel T (2018) Value-decomposition networks for cooperative multi-agent learning based on team reward.
 41. Syed Shah Sultan Mohiuddin Qadri, Mahmut Ali Gökçe & Erdinç Öner (2020). State-of-art review of traffic signal control methods: challenges and opportunities. *European Transport Research Review*.
 42. Tan, T., Bao, F., Deng, Y., Jin, A., Dai, Q., & Wang, J. (2019). Cooperative deep reinforcement learning for large-scale traffic grid signal control. *IEEE transactions on cybernetics*, 50(6), 2687-2700.
 43. Van Moffaert, K., & Nowé, A. (2014). Decentralized partially observable Markov decision processes with Monte Carlo tree search. *Proceedings of the 2014 IEEE Conference on Computational Intelligence and Games*, 1-8.
 44. Vuorio, R., Sun, S. H., Hu, H., & Lim, J. J. (2019). Multimodal model-agnostic meta-learning via task-aware modulation. *Advances in neural information processing systems*, 32.
 45. Wan, C. H., & Hwang, M. C. (2018). Value-based deep reinforcement learning for adaptive isolated intersection signal control. *IET Intelligent Transport Systems*, 12(9), 1005-1010.
 46. Wan, C. H., & Hwang, M. C. (2019). Adaptive traffic signal control methods based on deep reinforcement learning. In *Intelligent Transport Systems for Everyone's Mobility* (pp. 195-

- 209). Springer, Singapore.
47. Wu, J., Liu, P., Qin, X., Zhou, H., & Yang, Z. (2019). Developing an actuated signal control strategy to improve the operations of contraflow left-turn lane design at signalized intersections. *Transportation research part C: emerging technologies*, 104, 53-65.
48. Xu, L., Zhang, Y., Liu, Z., Wu, J., & Zha, H. (2020). Meta reinforcement learning with adaptive hyperparameters. *Proceedings of the 37th International Conference on Machine Learning*, 10391-10402.
49. Yang, G., Sun, Y., & Narasimhan, S. G. (2019). Critic-Aware Multi-Modal Planning for Heterogeneous Multi-Robot Systems. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 5837-5843.
50. Zheng, G., Xiong, Y., Zang, X., Feng, J., Wei, H., Zhang, H., & Li, Z. (2019). Learning phase competition for traffic signal control. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1963-1972).
51. He, W., Gao, H., Zhou, C., Yang, C., & Li, Z. (2020). Reinforcement learning control of a flexible two-link manipulator: An experimental investigation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(12), 7326-7336
52. 田永平、卓訓榮、林貴璽(2005)，號誌路口時制轉換時段長度之模擬研究 (Doctoral dissertation).
53. 江伊嵐、陳奕廷、葉梓銓(2020)，智慧型動態交通號誌控制系統-應用於台北市重點壅塞區域. *土木水利*, 47(4), 20-27.
54. 吳悅慈(2011)，幹道群組適應性號誌控制模式之開發研究.
55. 胡守任、葉志韋、林定憲、劉瀚聰(2020)，都市適應性號誌控制原理與發展. *土木水利*, 47(4), 28-39.
56. 胡大瀛、李卓育(2021)，深度強化學習下號誌控制設計獎勵機制之探討，2021 年中華民國運輸學會學術論文研討會.
57. 許添本、程楷祐(2020)，以深度強化學習方式建構混合車流之 AI 最佳化號誌時制計畫，2020 年中華民國運輸學會學術論文研討會.
58. 許添本、蔡沐軒(2021)，示範式深度強化學習應用於號誌時制最佳化之研究，2021 年中華民國運輸學會學術論文研討會.
59. 陳又均(2022)，應用多代理深度強化學習於高速公路主線及匝道聯合儀控策略——以國道 5 號為例.
60. 陳麗雯、胡大瀛(2020)，都市幹道連鎖時制設計之研究與實例分析. *土木水利*, 47(4), 40-50.
61. 張佑鴻、林良泰(2012)，半觸動號誌控制模式之建構與驗證 model construction and validation for semi actuated traffic signal control
62. 黃崇宇(2009)，高速公路匝道儀控與速率控制協控模式之研究.
63. 薛博元(2011)，高速公路速率控制與匝道儀控最佳化協控模式之研究.

附錄一 期中審查意見回覆表

項次	審查意見	審查意見回覆	主辦單位意見
一、 范俊海委員			
1	報告書第二章為相關文獻回顧，為加強對於系統架構認知，建議將 P.74~P.75 的演算法之歸納與比較分析，整併於「2.1.2 強化學習演算法」該小節內。	感謝委員意見，已整併節 3.3 於節 2.3.2 之第六點強化學習演算法比較。	同意
2	報告書「2.3.3 國內過往智慧交通研究/計畫成果」小節內標題層級有誤，請重新檢視與修改。此外，表 2.3.3-1 標題為「本研究強化學習要素設定列表」是指前期(111 年)研究設定或本期(112 年)研究設定？若前期與本期研究設定不同，建議補充說明研究設定差異處或僅演算法不同。	謝謝指教，已修正： (一)節 2.3.3 調整為節 2.2.3。 (二)修改節 2.2.3 標題層級。 (三)表 2.2.3-1 標題，修正為「前期研究強化學習要素設定列表」 (四)有關前、後期研究設定差異，詳報告書後面章節所述。	同意
3	報告書「5.4.1 交通參數與績效指標設計」小節，請補充說明績效指標的計算方式(例如：如何分辨延遲時間[delay]與停滯時間[stop delay])與預計如何取得績效資料。此外，績效指標(Performance Index)應是表示獎勵(Reward)的指標，常見會有正向與負向指標，請說明為何報告書內均列出負向指標。	(一)Delay 有很多種評估方式，例如自由流 GVP 與實際 GVP 的差值，停等延滯為其中一種評估方法。而停等延滯(Stop Time Delay)定義為車輛因號誌化路口停止後所停等之時間。 (二)節 5.4.1 為模擬環境，故將透過模擬軟體之偵測器功能進行資料蒐集。惟績效與獎勵代表之意義不盡相同，其中通過量與停等延滯分別為正、負向指標。	同意
4	報告書「4.3.1 偵測器驗證計畫」小節關於高速公路 VD 與 eTag 設備偵測誤差是否顯著，可考慮進行敏感度分析，若偵測設備(雷	本計畫於節 4.3.1 中，即對於偵測器進行該設備蒐集之資料的驗證，其中亦有透過影片人工計數確認其準確率。 RL 是基於 Reward 之偏好設定以及輸入資料進行 Action 推估，故資料失準必然會影響演算法	同意

	達偵測設備、AI 影像辨識偵測設備、高公局 VD 偵測設備、高公局 eTag 偵測設備)誤差顯著，請說明是否會影響演算法運作或導致策略失準。	的輸出結果，但不影響演算法本身。	
5	針對多目標 Reward 進行事前事後績效分析時，請問該以何種偏好進行績效分析，或針對不同偏好進行決策結果的敏感度分析。	實測進行事前事後績效分析時，會依照時段特性選擇特定的偏好，若測試時間足夠時，可於同時段測試多種偏好，並分析不同偏好對於績效之影響程度。	同意
6	請補充說明動態模擬軟體測試與實測、強化學習演算法的關聯。	當 AI 模型訓練完成，會先經由動態模擬測試確認 AI 模型輸出之秒數是否合理且穩定，以避免實測時給予不符合現場車流狀況之秒數，造成交通壅塞、引發行車人不滿，若 AI 模型於動態模擬測試時給予之秒數不具合理性與穩定性，將記錄異常狀況並加以優化 AI 模型。	同意
7	請說明本案 AI 模型是否具備可移轉性或相關移轉流程。	若 AI 模型欲嘗試移轉至不同路口，初步評估需要兩個路口之車流狀況相似，且時制計畫與時相內容完全一致，同時偵測設備裝在路口的相同對應位置後，方可直接使用，其餘狀況則可能需要重新訓練 AI 模型。	同意
8	簡報 P.8 所提考量動態調整時相組態可行性，依照駕駛者慣有習性，比較適合既定的時相組態組成，建議不要異動時相組態、順序等設定。	感謝委員意見，目前 AI 模型於號誌控制的表現多為時相時比的控制，且駕駛者對於時相順序有既定的印象與習慣，評估若欲使 AI 模型輸出調整時相組態與順序理論與實務均較為困難，未來可思考並嘗試讓 AI 模型減少人工介入的比例，提高機器學習的佔比。	同意
9	請詳細說明簡報 P.18 所提及之多目標 Reward 計算方式，例如：平均停等延滯採用半小時為單位，可補充如何取得停等延滯，以及為何採用半小時為單位。	Reward 為 AI 模型訓練時使用之績效，Reward 計算的頻率為 AI 決策後產出，以臺南場域為例，Reward 計算為每個時相結束後產出；停等延滯係由模擬軟體中的車輛偵測器中取得，偵測車速低於 5 km/hr 則記為停等；平均停等延滯為測試 AI 模型績效時使用的，採用半小時為單位計算測試績效，係為 SUMO 模軟體中的流量為半小時為單位改變一次。	同意
二、 吳健生委員			
1	請說明本計畫使用 Q-Pensieve 演算法的原因，以及未來如何推廣使用本演	現行已有許多將強化學習用於交通號誌控制的文獻與案例，而具體做法皆是基於單目標強化學習，亦即在 reward 設計上會將多種指標 (如	同意

	算法。	停等延滯、通過量、號誌切換次數等) 直接做線性疊加，而這個做法主要的缺點有兩個： (一)各種指標的單位不同，如何給予適當的權重不易決定。 (二)只能學習單一種權重下對應的最佳策略，缺乏對於各種權重的自適應能力。 有鑑於此，本計劃提出使用多目標強化學習 (Multi-objective RL, or MORL) 來解決上述兩個挑戰，而使用 Q-Pensieve 的主要考量有兩個： (一)Q-Pensieve 在多種 RL benchmark 能達到迄今所有演算法的最佳表現。 (二)Q-Pensieve 是由本計劃的成員謝秉均教授及其團隊所開發，已經有較成熟的 codebase，在開發時程上預期能比使用其他 MORL 方法更佳。 在演算法的推廣部分，預期進行以下三件工作： (一)近期會將 Q-Pensieve 與在交通上的實驗結果投稿至頂尖的機器學習或交通領域期刊，以獲得較高能見度。 (二)近期會將相關成果投稿至台灣運輸相關學會報告。 (三)將會提供此演算法之開源或部分開源的程式碼，期望未來能於台灣各縣市做號誌控制使用。	
2	請說明多目標 AI 模型中不同的偏好會導致決策結果有何差異，若多目標內 Reward 有互斥情形，或多目標(通過量、平均停等延滯時間)單位不同如何比較績效差異。	(一)訓練時會依照 Reward 偏好的權重訓練，如果此時的權重為通過量最高，此時通過量就為主要的訓練目標。 (二)不同的偏好反應了對每個目標的相對重要性。我們可以通過調整獎勵函數中不同目標的權重來表現這些偏好，AI 模型會根據這個獎勵函數來訓練，以最大化整體獎勵。然而，如果多個目標存在互斥情況，模型可能需要在不同目標之間進行權衡，因為優化一個目標可能會降低其他目標的性能，這樣才能找到最佳的解決方案。 (三)在 AI 模型的訓練過程中，我們需要處理多目標和不同偏好的情況。為了實現這一目標，我們首先通常會將各個目標的數值範圍標準	同意

		化，以確保它們在相似的數量級範圍內。本計畫後續將修正 Reward，使用與現場固定時制的比較並以秒為單位，如：總延滯時間與固定時制的變化量、通過量與固定時制的變化量，詳細說明於 3.2.1 AI 模型精進架構中的獎勵目標。	
3	請說明 AI 模型基本限制條件與使用情境，包含為何使用單一 Agent 或多 Agent、控制路口數量不同等因素。	AI 模型主要依據現場環境需求進行 AI 控制架構之設計，並不受 Agent 數量的限制，本計畫臺北市實驗場域期待 AI 模型控制的三個路口週期具一致性，故使用單一 Agent 控制三個路口。	同意
4	高速公路協控常面臨高速公路與平面道路目標互相衝突或目標有主從層級，請補充說明 AI 模型是否可發揮其效益，達成協控目的。	高速公路協控遇到高速公路與平面道路目標衝突或主從層級時，AI 模型可使用權重調整高速公路與平面道路的比重分配，以符合主從層級之設定。	同意
5	請說明報告書中 AI 模型只訓練尖峰時段之原因，此外，請說明 AI 模型輸出是否改變週期或採用其他決策參數。	(一)本計畫將訓練上下午尖峰及下午離峰時段之 AI 模型，惟期中審查時僅完成臺南市實驗場域上午尖峰時段訓練，故僅呈現此成果，完整成果請詳期末報告第三章。 (二)臺南市實驗場域 AI 模型輸出為時相秒數；臺北市實驗場域 AI 模型輸出為週期秒數及時相比例，均為變動週期。請參閱期中報告節 3.2.1。	同意
6	AI 模型是否具備可移轉性，若用於其他路口使用是否一定需要重新訓練。	感謝委員意見，請參考范俊海委員之第 7 條意見回覆。	同意
7	請說明 AI 模型在訓練與實測時運算效能、演算花費時間、所需硬體設備規格、以及對於號誌控制器是否有限制。	(一)AI 模型訓練一天於臺南市實驗場域可進行十萬步(AI 決策秒數時會增加一步)；臺北市實驗場域可進行七萬步；實測時 AI 模型所需決策時間不超過一秒。 (二)本計畫 AI 模型訓練使用之軟硬體規格，請參閱期中報告書節 3.2.2。 (三)因本計畫使用步階控制，故號誌控制器需要可支援都市交通控制通訊協定 3.0 版 5F1C 指令中的時相步階轉換控制。	同意
三、 王穆衡委員			
1	建議應由計畫目標與預期效益/貢獻，在檢視計畫之	謝謝委員意見，本計畫目標利用 AI 模型支援號誌控制與管理需求，提升車流運作效率及交	同意

	設計與工作安排。	通安全。本計畫 AI 模型可依照現場狀況輸出合適之綠燈秒數，提升車流效率並減少管理人員負擔，後續將持續檢視計畫預期效益並規畫明年度工作安排。	
2	本計畫在國內具有重要價值，建議研議實作場域應有之標準設備、設施配置等規範，如有需要可向交通部申請經費，於特定地區建構標準實驗場域。	本計畫 AI 模型輸入資料為分車道、車種之車輛數(停等/轉向)，使用 CCTV 影像辨識與 VD 雷達偵測器，已於第四章補充說明實驗場域之路口幾何與車流偵測器，以做未來參考。	同意
3	請補充說明後續實證之情境分析與實驗設計。	本計畫於 112 年 11 月辦理實測，避開暑假期間、連續假期、補班等狀況。細節詳期末報告節 4.3 所述。	同意
4	建議可藉由本計畫了解 AI 模型適用性與可適應何種類型之場域，並請說明 AI 模型輸入是否只有包含控制路口資訊，應可擴大資料輸入範圍，將周圍路口納入考量。	本期臺北市實驗場域 AI 模型並無輸入周邊路口資訊，僅在 Reward 考量周邊路口狀況，如需納入周圍路口資訊，周圍路口號誌組態改變會需要重構模型。	同意
5	高速公路協控常因高速公路主線流量遠大於平面道路流量，若維持使用通過量做為績效可能導致平面道路壅塞無法紓解，請審慎選擇績效評估指標。	敬悉，交流道模擬模型將以旅行時間、延滯等作為績效指標，並會一併考量高速公路及平面道路主線車流狀態，以高速公路主線及平面道路主幹道旅行時間、延滯等最小為主。	同意
6	請補充說明平假日、尖離峰與夜間各時段，其 AI 模型運算績效表現。	期中報告時著重於上午尖峰時段進行訓練，其他時段之運算績效表現請詳期末報告第三章。	同意
7	請考量未來發表論文之主題與貢獻為何。	本計畫為 AI 強化學習與交通領域之結合，預計於機器學習和交通領域進行發表或研討會參與，期能引起更多之討論與實務研究。	同意
四、 李明聰委員			
1	請補充說明簡報 P.16 中，臺北市實驗場域 Reward 之計算方式，以及是否為訓練目標，另請補充各路口平均延滯之定義，以及如何權衡各路口權重偏好。	(一)臺北市實驗場域的 AI 模型為每週期決策一次，故 Reward 計算為週期內累積的績效，其 Reward 為幹支道的總延滯時間與通過量(詳見節 3.2.1)。 (二)平均延滯為測試 AI 模型績效時使用，計算公式為總延滯時間除以通過量。 (三)訓練過程中會隨機抽取不同路口的權重偏	同意

		好用於模型的訓練。Q-Pensieve 演算法會根據這些抽取的權重，學習各種不同權重的策略。在實際應用階段，我們可以根據當前的需求，賦予特定路口期望的權重，以引導 AI 模型生成一個最符合該權重偏好的策略。 舉例來說，如果我們希望增加 A 路口的交通流量，我們可以給予 A 路口較高的權重，這樣模型就會更傾向於生成一個偏好 A 路口流通的策略，以實現所需的流量優化目標。	
2	請說明未來於實測時預計使用之績效指標。	謝謝指教。 臺南市實驗場域將以停等延滯及系統通過車輛數做為績效指標；臺北市實驗場域將以幹道平均旅行時間、系統通過車輛數及停等延滯做為績效指標，詳節 4.3.3 所述。	同意
3	建議台北市實驗場域之績效指標可納入綠燈寬帶，以了解使用 AI 模型綠燈帶寬是否有顯著變化。	感謝委員意見，臺北市實驗場域之監控畫面納入綠燈帶寬資訊，實測期間可即時參照。	同意
4	簡報 P.47 中高速公路協控涉及多層級道路討論，由於高速公路與平面道路流量相差許多，應考量如何進行多目標與多層級績效權衡，進而達到協控之目的。	感謝委員意見，請參考吳健生委員之第 4 條意見回覆。	同意
五、 許佳興委員			
1	報告書內甘特圖建議應明確寫出預計完成之時間點。甘特圖中提及七月份會完成實測計畫書，目前期中報告中有部分篇幅說明實測計畫，請補充說明是否會提出實測計畫書。	已於期中報告中說明實測計畫內容，實測前提出完整之實測計畫書，包含實測日期與 AI 模型產出績效，供計畫場域縣市政府交通局評估。	同意
2	報告書第二章文獻回顧部分，請依照文獻首次出現順序編號、公式格式與引用格式請統一、所引用之參考圖文務必請重新繪製以避免版權問題。	謝謝委員指教，已修正。	同意
3	報告書第三章之後，若有	謝謝委員指教，已修正。	同意

	引用或參照相關文獻，請於內文明確說明與文獻之關聯性。		
4	報告書提及 AI 模型訓練簡化之議題，建議應分別介紹前期與本期模型架構與決策效果，另外針對不同實驗場域應分別說明設備感測器、資料格式、AI 模型等項目。	遵照辦理，不同實驗場域 AI 模型說明請詳節 3.2.1，使用之車流偵測器、AI 模型等資訊詳請第四章。	同意
5	因應時程緊湊，建議於報告書或後續工作會議上提出實測監測畫面或儀表板規劃設計，以利確認可行性後，進行後續建置開發。	感謝委員意見，已於後續工作會議上提出實測監控畫面設計，確認可行性。	同意
六、 劉瑞麟委員			
1	報告書請考量採用「貴所」之尊稱或「本所」自稱。	謝謝委員指教，已調整報告書用語。	同意
2	報告書 P.53-P.54 臺北場域輸入決策當下資料，輸出下個時相秒數；臺南場域輸入決策當下前一周期該時相轉向交通量，輸出下一個時相秒數。兩者在使用文字上不一樣，意義是否相同？	兩個實驗場域輸入的資料與資料的時間區間皆不相同，臺北市實驗場域輸入為決策當下的分車道及車種的車輛數；臺南市實驗場域輸入為累積轉向量，如要決策時相 3 的綠燈秒數，此時的轉向量輸入為時相 2 開始至結束時的累積轉向量。	同意
3	報告書 P.54-P.56 臺北場域「礁」溪街，應為「礮」溪街之誤植。	謝謝委員指教，已修正。	同意
4	報告書 P.56-P.59 臺北場域有較多簡化之說明，臺南場域則均無，惟於(二)模型演算法後續修正內，僅用「訓練過程效率低」及無「不同目標最佳化」等因素改採 Q-Pensieve 演算法重新訓練。可否有數據或相關結果來加強說明修正	前期 AI 模型訓練使用 DDPG 演算法，而 DDPG 最主要的缺點主要有兩個：(1) DDPG 使用無隨機性的策略，而容易收斂至 sub-optimal 的動作；(2) DDPG 是單目標強化學習，必須人為決定多種獎勵的權重，而這需要大量的超參數調整以及大量訓練資料的配合。為了解決問題 (1)，目前強化學習領域最常見的做法是直接使用 SAC 演算法，具有多模態 (multi-modal) 且有隨機性的策略，可較好地避	同意

	的依據。	免收斂至 sub-optimal 的動作，而在各種 RL benchmark 的比較中 SAC 普遍都優於 DDPG。可參考以下連結： https://spinningup.openai.com/en/latest/spinningup/bench.html https://github.com/DLR-RM/rl-baselines3-zoo/blob/master/logs/benchmark/benchmark.md 為同時解決以上(1)和(2)兩個問題，本計畫提出使用 Q-Pensieve 多目標強化學習方法，可同時學習各種權重對應的策略，同時由於 Q-Pensieve 可視為推廣版的多目標 SAC 演算法，因此 Q-Pensieve 也繼承 SAC 的多模態隨機性策略的優點。至於 Q-Pensieve 相對於單目標 SAC 的優勢，可參考 Q-Pensieve 原論文的多種實驗數據： https://openreview.net/forum?id=AwWaBXLlJE	
5	報告書圖 3.2.3-2 與圖 3.2.3-3，請補充說明平均延滯為負值其比較基礎為何。	為呈現績效結果為正斜率，本計畫將平均延滯計算結果加上負號，後續將修正為平均延滯不加負號，呈現其結果。	同意
6	報告書圖 3.2.3-2 前期 AI 績效過於分散，是否有探討其可能原因為何？	可能因前期模型訓練為單一目標，與本期多目標模型不同，使其績效延滯及通過量的目標無本期集中。	同意
7	報告書圖 3.2.3-3 Q-Pensieve 以通過量權重較高之目標值，結果集中於通過量相較偏低之結果，請補充說明其原因。	研判這個現象的可能性是績效計算的加權以及兩邊用來計算通過量的方式不同所導致，主要的不同點為訓練時使用報告中提到的計算方式，而績效則是直接在路口設置計數器，也因此兩邊的數值會有 2 至 3 倍的差異。在模型訓練上(期中報告書圖 3.2.3-2)，我們可以很明顯的感受到注重通過量的 preference 有很好的提升此項的結果。	同意
8	報告書 P71 圖 3.2.3-1 收斂曲線是否可表示其收斂，若繼續訓練(增加 STEP)是否會維持收斂結果。	從期中報告書圖 3.2.3-1 中，我們可以看到模型的性能持續提升，但它尚未完全收斂。這種情況表明模型仍然有繼續成長的潛力和空間。	同意
9	報告書 P72 圖 3.2.3-2~4，請說明為何其 STEP 高達 10^6 (或 8)，並請確認是否有收斂，其相關績效(總通過量、延滯績效)是否亦有同步變好。	感謝委員的指出，期中報告書圖 3.2.3-2~圖 3.2.3-4 的總訓練 step 數都為 10^4 才是正確的，從這幾張圖我們可以觀察到訓練 Reward 都有所提升，並且有持續成長的跡象，同時參考表 3.2.3-1 可以說明其相關績效有確實同步變好。	同意

10	報告書 P.81-91 針對北市德行/中山沿線交通現況調查分析，建議除流量調查交通數據外，是否有相關旅行時間數據，另現場仍有許多交通狀況如路側行為、轉向車與行人延滯及變換車道行為非能於交通數據得知，建議可補充相關調查分析資料，以應用至後續號誌控制。	謝謝指教。 本計畫後續進行績效評估時，已規劃蒐集旅行時間進行評估，相關說明詳述於節 4.3、4.4。 另本計畫為 AI 模型應用於都會區及交流道的發展探討，故以評估號誌對於車流造成的影響為主，其他用路人行為所造成之交通影響，可納為往後精進工項。	
11	報告書 P.101-103 用人工判斷合理性與穩定度，請說明合理性標準定義，如表 4.3.2.2-2，刪除時相 2 極低值，但下一周期明顯有高補償，應可檢討原因。未來應用時，請說明該如何操作合理性判斷。	(一)本計畫合理性標準定義以 TOD 為參考基準，單位時間(秒)通過車流差異比應為正值且介於 1~4 之間，亦即分相配置之秒數至少能通過適當比例車流，並符合 TOD 最大週期及最小綠 10 秒之限制。 (二)有關期中報告書表 4.3.2.2-2，該表格為台南場域，該處 RI 模型決策頻率為每個時相，故不同週期之同一時相彼此獨立。	同意
12	報告書 P.117 績效指標採用旅行時間、延遲時間及停滯時間，後續可檢討何種指標較符合實際情況，另台北市調整號誌亦會考慮前後路口連鎖、綠燈帶及車輛停等長度或 2 次停等，建議可納入績效評估使用。	驗證模型績效之指標分為模擬環境與實測使用兩大類，前者自模擬環境中取得，實測使用則是於現場路側設備偵測、人工調查或其他多元資料取得，故可分別設定模擬環境與實測使用之績效指標，後續將就不同績效指標權重比較考量。	同意
13	報告書 P.123 交流道模擬驗證將採轉向交通量、總停等延滯、旅行時間等 3 個參數之 MAPE 都符合個別門檻，之前臺北、臺南場域之模擬環境是否比照辦理校估驗證。	謝謝指教。 MAPE 統計量指標於前期模型皆有評估，而本計畫訓練模型沿用前期計畫，惟本計畫使用多目標模型訓練，為不同之處。	同意
14	採用 VISSIM 及 SUMO 進行 AI 模擬訓練，如何避免模擬環境及真實環境差異導致訓練結果不佳。	實務及學術上，多以 GEH 指標評估兩者差異，相關敘述詳章節 5.4.4。	同意
15	本案是否使用多元資料如	本計畫除了以通過量作為 AI 訓練的 reward，	同意

	車流量及旅行時間資料作為控制依據，如僅使用車流量資料，請補充說明 AI 號控是否可分辨塞車時及車流順暢車流量皆低之情形。	同時也使用總延滯時間作為訓練時的 reward，因此 AI 模型可透過這兩種績效判斷車流是否暢通。	
16	建議後續有 AI 模擬結果之時制計畫內容及模擬結果影片，提供台北市交控中心確認，另正式上線前應有相關執行計畫說明，避免執行後導致現場交通混亂。	謝謝委員意見，實測前皆提交實測計畫書予縣市政府交控中心並召開會議確認執行內容。	同意
17	簡報 P.32、P.35 臺北場域資料調查於 6-7 月，7 月為暑假平日交通趨勢，建議應考量是否符合一般月份。	本計畫蒐集 109~111 年不同月份之交通資料，顯見交通趨勢並無顯著差異，詳期中報告書表 4.3.3-2 及 4.3.3-3。	同意
七、卓明君委員(書面審查意見)			
1	請釐清號誌協控策略及模型開發時，是否納入道路等級之權重條件，如國道及地方道路有不同權重，而國道等級及最高速限等條件較高，故權重較高，始符合比例原則，而非以整體總停等延滯最少為佳。	感謝委員意見，請參考吳健生委員之第 4 條意見回覆。	同意
2	高速公路上、下匝道之車輛回堵，對地方道路之外側車道影響較鉅，直行車輛則受回堵車隊干擾影響，故計算時是否針對地方道路之不同車道採不同權重？	敬悉，後續交流道模型環境建立將納入考量。	同意
3	高速公路上匝道之儀控，除考量個別交流道進、出及直行交通量外，亦可能需配合下 1 處或下 2 處交流道之交通狀況予以調	敬悉，考量高速公路除本身所在交流道進出匝道之車流交織影響，亦會受上下游交流道車流衝擊波影響，爰此交流道模型環境建立將納入涵蓋上下游交流道旅行時間資料參考。	同意

	整，建議協控測策略及模式開發時，亦將不同路段特性需求納入考量。		
4	國道例假日及連續假日尖峰時段於 11 處路段實施大範圍之精進式匝道儀控管制，以維持國道合理服務水準，與地方道路之協控機制策略目標不同： (1) 國 1：楊梅至新竹南向、台北至大華系統北向、苗栗至豐原南向、西螺至埔鹽系統北向、彰化系統至埔鹽系統南向、三義至頭屋北向、彰化系統至后里北向。 (2) 國 3：土城至大溪南向、後龍至西濱北向、快官至霧峰南向。 (3) 國 5：國 3 汐止系統至南港系統南向、國 3 三鶯至南港系統北向、國 5 南港系統至坪林南向。	謝謝指教。本計畫之交流道區域設定以平常上班日為主，視執行狀況納入一般假日(週休二日)，並不考量連續假期之情形，針對例假日部分後續交流道場域模擬將與高公局協調討論相關協控策略，以納入模型建置參考。	同意
5	協控策略及模型開發時，是否將高速公路或市區幹道之施工、事故...等即時路況事件納入並透過橫向聯繫機制提出建議改道資訊等。	目前文獻回顧尚未有 AI 號誌協控模型納入即時路況事件之探討，未來可探討透過 AI 模型進行橫向聯繫機制與改道建議之可行性。	同意
八、交通部高速公路局			
1	本計畫預計控制上匝道儀控、下匝道與平面道路號誌，兩者都會影響國道主線狀況，建議實測前可先與高速公路局召開會議確認可行性。	謝謝指教。本計畫今年度僅提出協控架構，明年度實測前將先與高速公路局聯繫召開會議，確認 AI 模型可行性。	同意
2	建議補充報告書 P.34 表 2.2-3 代數意涵說明。	謝謝指教，已補述於該章節中。	同意
3	報告書 P.110 中路口 5 之楊	謝謝指教，已確認修正，詳表 5.3-2 及 5.3-3 所	同意

	梅交流道有新增號誌控制，請確認與修正。	示。	
4	報告書 P.118 由於本計畫涉及高速公路匝道，建議考量高速公路相關車流資訊作為績效評估指標。	遵照辦理，將納入高速公路主線旅行時間及速率等車流資訊作訓練及績效依據。	同意
5	高速公路局於匝道儀控時會與地方政府設定回堵點，當車流壅塞至平面道路回堵點時，高速公路局會放寬儀控以緩解地方交通，建議本計畫可考慮納入匝道回堵長度。	感謝指教，後續將確認 SUMO、VISSIM 車流模擬軟體是否能設定回堵點並了解模擬狀況，評估明年現場環境是否可配合設立回堵點之偵測。	同意
九、 新北市政府交通局(書面審查意見)			
1	本計畫建構之高速公路匝道儀控與地區道路號控模式對於解決本市上下匝道周邊道路壅塞問題確有助益，碩建構應考量便於移轉使用，以供後續地方配合改善匝道周邊交通問題。	謝謝指教，AI 模型移轉性請參考范俊海委員之第 7 條意見回覆。	同意
2	請補充說明地區道路號控運作邏輯，且地區道路因匝道壅塞產生之影響屬於路網型，請說明設置偵測點之原則？	謝謝指教。 交流道場域擇以楊梅交流道及其周邊路口為例，說明地區道路號控運作邏輯。又因昏峰車流狀態壅塞較為嚴重，雖須維持高速公路主線順暢，惟過長時間之低匝道儀控率會造成車流回堵至地方道路。 因此須設置偵測點，以匝道銜接道路之路口往上游至少推一個號誌路口為原則，避免排隊等候車隊長度過長。	同意
十、 桃園市政府交通局			
1	P.119 不受號誌影響之車輛速率街廓大小以 250 公尺為基準，以楊梅交流道為例，7 號路口至 6 號路口間街廓約為 270 公尺，且北上匝道車輛不受號誌管制，如採 GVP 是否影響其參數及績效指標？	路口平均延滯考量當車輛一進入短街廓時即受到路口號誌影響，以楊梅交流道路口 7 及路口 6 為例，路口 7 南下匝道至路口 6 北出匝道街廓大於 250 公尺，尚能以該鄰近路口停止線上游 200 公尺至街廓邊界 GVP 資料推估，詳報告書圖 5.4.2-3 所示。	同意
2	P.120 模擬模型以 VISSIM 與	AI 模型訓練所需時間應依照模型收斂狀況判	同意

	SUMO建立模型，請說明各階段實作過程(如AI模式訓練所需時間或下載至現場預計日期，從CCTV蒐集到資料轉成轉向量所需時間等作業內容。	斷，並於收斂後校估調整相關參數。待模型完成後，透過與模擬軟體環境互動確認其合理性及穩定性，再下載至現場。預計期程須依前述各階段完成狀況而定。	
3	目前AI所能執行的項目、具體學習內容、運作限制等與傳統時制重整間之差異為何?	傳統時制重整係透過短期交通調查與現況分析，透過人工檢視與時制最佳化軟體，據以研提路口群組劃分、時相選擇、時段切分、以及時制計畫產製，後續在進行時制計畫下載與微調。 而本期AI模型著重於時制計畫的調整，在既定的路口群組劃分、時相選擇、時段切分限制下，根據即時車流狀況，產出適當之時制計畫(各分相秒數)，以期減少交通壅塞發生。 未來可將人工檢視或設定的部分，調整為AI機器學習方式，朝向更具人工智慧之號誌控制模型發展。	同意
4	該計畫改善僅時比，尚無調整號誌時制週期的情況下，是否考慮或建議其他交控改善策略?	本期基於前期構建及訓練完成之模型進行實測，均採用變動週期與變動時比之號誌控制方式，後續研究可納入調整秒差(Offset)、變動時相組態或順序等不同之號誌控制模式。	同意
5	P.117績效指標中，停滯時間是否僅為停等延滯?是否包含減速或加速延滯等延滯時間?	謝謝委員意見，5.4.1節績效指標中停滯時間僅為停等延滯，後續參考委員意見調整績效指標停滯時間、延滯時間為停等延滯。	同意
6	AI場域如遇閃光號誌及行人觸動號誌時，如何運作?	當執行閃光號誌時，AI模型無法在此情形做號誌控制；行人觸動號誌與AI模型號誌控制為不同控制邏輯，亦無法同時進行號誌控制	同意
十一、臺中市政府交通局			
1	請問偵測設備資料回傳的頻率與秒數，資料回傳時間加上AI模型運算時間，是否可即時下達現場號誌控制指令。	臺北市實驗場域偵測設備為每秒回傳資料，臺南市實驗場域偵測設備為每筆回傳，兩者皆可即時回傳，AI模型運算時間不到一秒，因此現場號誌控制指令約可於1-2秒完成下達。	同意
十二、臺南市政府交通局			
1	檢視報告書 P.103 顯示目前AI模型輸出的秒數似不具合理性，依過往經驗應會被當地民眾反彈，建議應	遵照辦理，此部分訓練時，將最長與最短綠燈納入AI模型的訓練。	同意

	參考臺南市交通局之實務意見，設定最長/最短綠燈秒數限制。		
十三、交通工程技師公會			
1	AI 模型選擇影響績效甚大，建議應審慎評估使用何種模型。	謝謝指教，本計畫使用可同時學習各種權重對應的策略、具有多模態隨機性策略的 Q-Pensieve 演算法建構 AI 模型。	同意
2	建議說明使用 AI 模型所需之偵測設備，以及 AI 模型輸出設備為何。	謝謝委員意見，AI 模型需要使用車流偵測設備掌握現場交通狀況，如本案採用 AI 影像辨識車流量與 VD 車流偵測器，輸入資料後 AI 模型將輸出秒數下達至號誌控制器。	同意
3	模型構建應明確定義目標，是否有多層級或多目標，此外，目前目標定義偏向效率面，較無考量安全全面績效，可探討是否透過 AI 模型控制增加行人安全(例如：藉由 AI 模型減少車流與行人的交織，進一步提升行人安全)。	近來著重行人專用時相與行人號誌控制，未來可嘗試於 AI 模型中納入行人資訊，以同步處理行人時相控制，進而增加行人安全，於模擬環境部分，目前 VISSIM 之公司有付費軟體 VISWALK 可提供模擬之行人資訊，而 SUMO 中的行人設定則是作為車流資訊的基礎設定之一，並無法直接模擬人流狀況，於現場偵測部分，可利用 AI 影像辨識設備做行人偵測。	同意
十四、本所運輸資訊組(書面審查意見)			
1	建議將 3.3 節內容融入 2.1.2 節強化學習演算法最後部分，以小結方式呈現。	遵照辦理，期末報告已修正彙整 3.3 節至 2.3.2 節。	同意
2	P.52 前期臺南市實驗場域 AI 強化學習模型與臺北市實驗場域相同，均為 DDPG，請修正。	謝謝指教，已修正。	同意
3	請將 P.57 之「(二)模型演算法後續修正」與「(三)後續修正方向」內容整合，調整為「六、模型演算法後續修正」，分別就時制計畫與車流偵測資訊兩方面進行探討。另在「(二)模型演算法後續修正」中對於 DDPG, PPO, Q-Pensieve 等演算法在效率、更新處理、多目標等論述，請補	關於 DDPG、PPO、SAC 演算法的比較，期中報告書表 3.3-1 主要是基於 ICML 2018 的 SAC 論文 Section 2 之描述與 Soft Q-learning 的官方介紹整理而得。 SAC 論文： https://arxiv.org/abs/1801.01290 Soft Q-learning 官方介紹： https://bair.berkeley.edu/blog/2017/10/06/soft-q-learning/ 另外關於 Q-Pensieve，其最主要特色為相較於單目標的 SAC 或其他多目標強化學習方法能有更好的樣本效率，參考來源為 Q-Pensieve 論	同意

	充說明所參考或依據的文獻來源。	文： https://openreview.net/forum?id=AwWaBXLlJE	
4	P.58 對於臺北市實驗場域車輛偵測器起始偵測位置不同之修正方式，規劃以定值方式進行補償，請補充說明此補償方式之合理性與可能缺點。	若使用前期模型進行優化，則需使用定值方式進行補償，由於本期之臺北市實驗場域從頭開始建構本期 AI 模型，因此可直接依據現場偵測器偵測位置及距離，於模擬軟體中重新訓練模型，使其實際部屬上能與現場偵測資料對應。	同意
5	請補充說明圖 3.2.1-1(P.62)與圖 3.2.1-2(P.63)中自輸出層至決策秒數信息 entropy 的詳細流程，以及於「輸入的資料透過隱藏層的傳遞去擷取特徵資料」中之「特徵資料」所指為何？	本期的實作演算法 Q-Pensieve 使用 SAC 基底的隨機策略模式，模型輸出的決策會是一個 distribution，agent 會於輸出層後輸出一個 distribution，最後藉由隨機採樣和線性轉換來決定決策秒數，信息 entropy 則代表這個 distribution 的離散程度。 「特徵資料」用於描述輸入數據的重要信息或屬性。以本應用為例子，輸入資料包含各個路口的號誌資訊與路口偵測資訊，模型學習擷取的特徵資料就會是分析各個路口的轉向量，找出各個路口對於預測決策秒數的重要程度。	同意
6	請利用 P.67 圖 3.2.2-1 來詳細補充說明本案如何在模擬環境進行 AI 模型訓練。	遵照辦理，已於 3.2.2 節補充說明。	同意
7	3.2.3 節之圖編號有誤，請檢視與修正。	謝謝指教，已修正。	同意
8	請再補充說明 3.2.2 節 AI 模型訓練結果的表 3.2.3-1、表 3.2.3-2、圖 3.2.3-2(目前)、圖 3.2.3-3(目前)等詳細內容，例如：表 3.2.3-2 為 10 組亂數種子的整體路口的通過量，以(0.9, 0.1)偏好通過量情況下，何以其通過量為最低？而以(0.1, 0.9)偏好平均停等延滯下之通過量反而為最高，另平均停等延滯時間為負值之交通意義為何？	(1)這個現象的可能性為績效計算的加權以及兩邊用來計算通過量的方式不同導致的，這邊主要的不同點是訓練時是用報告中提到的計算方式，而績效則是直接在路口設置計數器的方式，也因此兩邊的數值會有 2 至 3 倍的差異。在模型訓練上(期中報告書圖 3.2.3-2)，我們可以很明顯的感受到注重通過量的 preference 有很好的提升此項的結果。 (2)因為強化學習的目標為持續與環境互動，並且在其中得到最大的 Reward，而我們的目標是最小化平均停等延滯，因此會需要在其前方加上負號調整最佳化方向。	同意
9	請再詳細補充說明P.71與P.72圖3.2.3-1(目前)至圖	(1)主要的差異可以從同一種 Reward 不同 Preference 的 y 軸數據比較出來。在使用主要偏好於通過量的 Preference 時，通常會收斂到	同意

	3.2.3-4(目前)等內容，例如： (1) 圖 3.2.3-1(目前)分別為總停等延滯時間與通過量之曲線，看不出「不同偏好」(0.8, 0.2)、(0.5, 0.5)、(0.2, 0.8)之差異。 (2) 先前偏好值採用 (0.9, 0.1)，而此處改為 (0.8, 0.2)的考量為何？ (3) -1.35×10^5總停等延滯時間與 1.34×10^5通過量的計算基準與交通上的意義為何？ (4) 何謂「通過量的結果數量會為績效的兩倍以上」？ (5) Q-Pensieve 所訓練出來的策略內容為何？	通過量較高結果，這點可以在期中報告書圖 3.2.3-2~圖 3.2.3-4 中比較出來。 (2)由於訓練時的 preference 是隨機產生的，為了能呈現整個訓練過程中，不同權重的表現，我們會根據每個 episode 使用的 preference 分為靠近(0.8, 0.2)、靠近(0.5, 0.5)和靠近(0.2, 0.8)的三組訓練結果，選擇以上三種 preference 我們可以獲得較為平均的獲得各 preference，以作為收斂曲線的呈現。 (3)總停等延滯時間主要是計算測試時間，所有車輛停滯等待的時間和通過輛。由於後面提到的重複計算問題，通過量會較真實模擬數據高兩倍以上，因此於交通意義上主要就會為 3600 秒內總共通過約 4 萬台車，並且總共有 4 萬多次車輛停等次數。 (4)這是在說明計算流通量的方式於訓練和績效上的不同，訓練目標 Reward 的計算方式為，一台通過路口的車子，會由於模擬程式的計算方式而導致被重複計數。在測試績效的程式上，我們針對重複計算車輛的問題進行修正，因此報告中提出結果才會為績效的兩倍以上。 (5) 由於目前兩個目標的數量級差距仍然相當大，因此 Q-Pensieve 學習到的策略為生成決策秒數時傾向於選擇能達到較小的平均延滯時間的綠燈秒數。	
10	P.75 表 3.3-1 各強化學習演算法比較，請補充說明參考來源或依據，另根據文獻，Q-Pensieve 特色之一為樣本效率高，請再確認本表內容之正確性。	(一)請參考本所運輸研究所的第三條意見回覆。 (二)關於 DDPG、PPO、SAC 演算法的比較，期中報告書表 3.3-1 主要是基於 ICML 2018 的 SAC 論文 Section 2 之描述與 Soft Q-learning 的官方介紹整理而得。 SAC 論文：https://arxiv.org/abs/1801.01290 Soft Q-learning 官方介紹： https://bair.berkeley.edu/blog/2017/10/06/soft-q-learning/ (三)另外關於 Q-Pensieve，其最主要特色為相較於單目標的 SAC 或其他多目標強化學習方法能有更好的樣本效率，參考來源為 Q-Pensieve 論文： https://openreview.net/forum?id=AwWaBXLlJE	同意

11	P.78 表 4.1.1-2 與 P.79 表 4.1.1-3，似遺漏時制計畫編號「04」、「08」、「22」，請再加以確認與補充。	本計畫訓練之時段僅擇平日上午尖峰 07:00~09:00、平日下午尖峰 17:00~18:00、平日離峰 14:00~15:00 及假日尖峰 17:00~18:00 四個時段，其對應之時制計畫為 01、21、06，故圖 4.1.1-2 僅列此 3 種時制計畫。	
12	P.102 請補充說明何謂「由交通量產出整理表」？另此處之 RL 模型為前期計畫 DDPG 模型，還是為今年 Q-Pensieve 模型，請再加以確認。	謝謝委員意見，應為「將模型產出之交通量整理可得表 4.3.2.2-3」。另，此處模型為 Q-Pensieve 模型。	同意
13	P.103 表 4.3.2-1 與表 4.3.2-2 中之「AI」與「AI*」所指為何，請再加以說明。	AI 係指前期 AI 模型；AI*為調整時制計畫後之前期 AI 模型，期末報告書已更改敘述為前期 AI、本期 AI。	同意
14	P.109請再加強說明圖5.2-1號誌協控模式訓練結構內容。	謝謝指教，已將訓練的流程與損失函數做具體描述並彙整至節 5.2。	
15	P.112 表 5.3-1 似遺漏路口 8 之幾何特性，請再加以補充。	路口 8 為南下匝道與台一線中山北路交會處，惟該處號誌為經常性閃黃，故視為非號誌化路口。	同意
16	部分圖表內容不清晰(例如：P.120 圖 5.4.3-2 與 P.121 圖 5.4.3-3)，以及若非本案產生之圖表請加註資料來源。	遵照辦理。	同意
17	有關報告初稿中之錯漏字、語句不適當與資料來源修正部分，請於會後洽本所。	遵照辦理。	同意
十五、主席結論			
1	審查會議各委員及與會單位研提之口頭及書面意見，請資拓宏宇國際股份有限公司整理「審查意見處理情形表」，且逐項說明回應辦理情形，並充分納入報告之修正。	遵照辦理。	同意
2	本計畫經徵詢審查委員意見，期中報告初稿審查通	遵照辦理。	同意

	過，請資拓宏宇國際股份有限公司後續依本所出版品印製相關規定撰寫報告，並納入每月工作會議查核事項進行追蹤。		
--	--	--	--

附錄二 期末審查意見回覆表

項次	審查意見	審查意見回覆	主辦單位意見
一、 范俊海委員			
1	建議績效分析時，AI 號誌控制應與適應性號誌控制或其他動態號誌控制比較，而非僅與固定時制相比。	感謝委員意見，本計畫實驗場域過往並無使用動態號誌控制，故僅可取得固定時制績效作為比較。	同意
2	報告書文獻回顧引用格式需完整。	謝謝委員指教，已於文獻引用此格式：作者（年代），並在其後放上文獻編碼。	同意
3	實測影響因素提及流量變化而導致績效受到影響，可能是 AI 號誌控制模型訓練不足，而其他非模型訓練因素，例如：網路不穩，是否有其他方式可以克服。	本次實測遭遇三個主要影響因素：控制器步階控制異常、設備斷線、流量型態變化，具體影響補充於 4.4.3 小節。控制器步階異常因素可藉由網路速率提升、程式精簡號誌指令方式克服。車流偵測設備斷線可經由程式直接下達原固定時制秒數，避免號誌補償影響績效；號誌控制器斷線則無法下達號誌指令，故無法克服。	同意
4	第二章報告書內容錯誤意見： 1.文獻回顧 P.9 之表 2.1.1-1 提及幹道與支道，因無相關資料可供參考，故無法清楚了解相關內涵與表達意義，建議應補充幹道與支道相關定義。	謝謝委員指教 1.已補充該文獻說明於本文 2.1.1 節中。	同意
	2.文獻回顧 P.11~13 之資料來源與內容對應不清楚，應調整引用格式與表達方式，此外表 2.1.1-4 所指實驗案例 Case 1~12 亦應補充說明其代表意義。	2.已將表來源及用途加註於本文中。如委員所舉例之 Case 1~12 為該研究依據流量特性分類，已補充說明於報告書章節 2.1.1 中。	同意
	3.文獻回顧 P.23 之資料來源與內容對應不清楚，均放置於整段文末，建議調整引用格式，明確表達引用範圍。	3.報告書 P.23 參考資料來源編號與節 2.2.1 文中 P.21、P.22 之資料來源為相同，惟參考資料眾多，故將參考資料來源彙整於 P.23。	同意
	4.文獻回顧 P.24 之 2.2.2 節標題為「強化學習於交流	4.報告書 2.2.2 節為交流道協控之文獻回顧章節，並且已修改原 2.2.2.2 標題，改為「強化學	同意

	道協控之號誌控制」，內容非專門介紹交流道協控，建議調整標題或內容，以使兩者相符。	習於號誌協控之應用架構」。	
	5.文獻回顧 P.38~39 建議應增加次標題，相鄰段落使用同一層級編號，卻又重新編號，容易造成閱讀不易與文意混淆。	5.已將 2.3.2 節之編號重新順編。	同意
5	第三章報告書內容錯誤意見： 1.實測模型探討部分 P.61~63，內文僅使用編號(一)，而無其他列點，建議整併內容並移除編號。	謝謝委員指教，已修正： 1.修正 3.1 之四、前期模型訓練簡化議題內文編號。	同意
	2.實測模型探討部分 P.70 與 P.71 之架構圖是否一樣，是否僅有輸入項目差異，請補充說明。	2.已補充說明於 3.2.1 之二、實測模型架構，(三)臺北與臺南模型比較，與表 3.2.1-5。	同意
6	第四章報告書內容錯誤意見： 1.實測成果 P.88 之尖峰小時轉向量項目編號錯誤，請修正。此外，4.1.2 小節 P.90~92 一樣有編號層級錯誤的問題，請修正。	1.謝謝指教，標題二、尖峰小時轉向交通量已修正為三、尖峰小時轉向交通量，並一併檢視順編後續層級。	同意
	2.實測動態模擬分析 P.121、127、131 之穩定性分析定義為應落於平均數加上兩倍標準差內範圍，請探討部分資料落於範圍外之因素，是否與取樣數不足有關聯。	2.以修訂版報告書 P.123 表 4.3.2.4-3 至表 4.3.2.4-7 為例說明之。因 AI 訓練輸入資料維度大，無法人工判斷其彼此間特徵關係強度，以 K-means 分類群集結果，顯著特徵比越大，表示其各群具關連性的比例越小，又根據最大特徵比之群集檢視其穩定性，多能落在兩倍標準差內，少數未達者因其群集資料量相對較少，但仍多有 9 成以上集中水準。	同意
	3.偏好績效分析 P.123、128、133、146、147、148 中顯示部分時段通過量 AI 號誌控制模型較固定時制佳，請評估是否差異顯著，但同時段	3.臺北市場域模擬環境測試成果，係依權重偏好各方向平均每車停等延滯及每小時整體路口 PCU 獎勵目標所產生之績效，用於實測偏好選擇，因此以修訂版報告書之表 4.3.2.4-8 平日離峰、表 4.3.2.4-14 平日昏峰及表 4.3.2.4-20 假日	同意

	績效亦可能發生停等延滯 AI 號誌控制模型較固定時制差，請補充說明整體績效表現。	昏峰其 AI 強化學習模型模擬表現皆是各偏好下相對較佳且顯著。 表 4.4.1-1、4.4.1-2 為臺南市實驗場域 AI 號誌控制與固定時制路口平日延滯績效比較，實測結果影響因素較多，如設備穩定度、連線狀態及模型訓練輸入之交通量樣態等，所以實測前兩天之假日 AI 績效表現皆較固定時制略差，而平日實測時，設備通訊較穩定，實測 AI 績效表現則多較固定時制為佳。	
7	第五章報告書內容錯誤意見： 1.報告書P.167績效分析指標使用旅行時間推估停等延滯，請補充說明與現行方式差異，並評估此推估方式是否具備合理性與可用性。	1.傳統停等延滯係人工於停止線附近作調查，若等候車隊過長無法於 15 秒內精確計算完所有的停等車輛數，導致在壅塞時停等延滯調查資料淪於調查員自由心證。而若採用攝影機記錄，亦受限於攝影機架設高度，遠方車輛可能無法於畫面上清楚呈現，造成事後判讀誤差。 本團隊提出以兩種 GVP 資料推估延滯係依據蔡輝昇老師「交通控制理論與實務」，其計算方式說明如下：蒐集當車輛駛離上游路口進入目標路段時，其行駛速率不受路口號誌及停等車隊影響之 GVP 資料，以及實際受號誌影響後通過停止線之 GVP 資料，其差值即為路口號誌帶給該路段車流的整個範圍影響，其中包含車輛減速、停等、再加速產生的延滯，可改善傳統人工判讀問題，且可減省人力成本。	同意
	2.請補充說明P.174、P.175模擬環境校估內容，確認驗證結果是否合乎可用標準。	2.有關報告書表 5.4.4-4、5.4.4-5 之 VISSIM 及 SUMO 楊梅場域流量模擬參數校估結果，符合前述三、參數校估之 MAPE 及 GEH 指標，詳報告書表 5.4.4-1 至 5.4.4-3 所示。	同意
二、吳健生委員			
1	報告書中部分圖表字體、圖例不清楚、與圖表分類錯誤，請檢視與調整內容。	感謝委員指教，已再次檢視報告書之圖表分類，並重新調整內容。	同意
2	請確認本案目標與方向，是否以發展泛用型(全時段)AI 號誌控制模型演算法與架構為主，目前模擬績效顯示分時段模型績效優於全時	因不同時段使用之時制計畫不同、時相組態有所差異，因此 AI 強化學習模型號誌控制需分時段進行模型訓練，然實務操作上若可有泛用型(全時段)之 AI 強化學習模型號誌控制將可節省訓練時間，且使用時無須抽換模型，故本期嘗試	同意

	段模型，請探討泛用型(全時段)AI 強化學習模型之適用性與預計達成之目標。	初探泛用型(全時段)號誌控制模型之可行性，若績效可與分時段模型相比擬，則未來可嘗試以泛用型(全時段)模型為主進行發展。	
3	報告書僅說明個別績效差異，建議可深入探討綜合性績效與偏好對於績效之影響。	謝謝委員指教，本團隊建置之 AI 強化學習模型初始以綜合性績效對比固定時制績效，惟效果並不顯著，因此設置不同偏好權重，不同的偏好設定下其相對應之績效表現亦較為顯著，選擇最好的偏好績效表現作為實測模型依據。	同意
4	請補充場域選擇因素、各場域既有限制、與前後期模型差異，因實測時可能有因現場路側設備問題、網路問題、地方縣市政府因實務考量而提出週期秒數、最小綠等限制，以上因素可能會影響 AI 號誌控制模型績效，建議應加強說明。	本計畫延續前期研究選擇臺北市與臺南市實驗場域進行實測，為避免 AI 強化學習模型輸出之秒不符合現場實務需求，設定之相關週期秒數與最小綠限制補充於 3.2.1 小節。	同意
5	請補充並舉例說明 AI 號誌控制模型運作方式、輸出/輸入參數、決策頻率、與運算時間。	謝謝委員指教，已補充說明於 3.2.1 之二、實測模型架構，(三)臺北與臺南模型比較，與表 3.2.1-5。	同意
6	請說明當實測遭遇設備異常、通訊問題時，是否有方式可以克服。	本次實測遭遇三個主要影響因素：控制器步階控制異常、設備斷線、流量型態變化，具體影響補充於 4.4.3 小節。號誌控制器步階異常因素可藉由網路速率提升、程式精簡方式克服。車流偵測設備斷線可經由程式直接下達原 TOD 秒數，避免號誌補償影響績效；號誌控制器斷線則無法下達號誌指令，故無法克服。	同意
三、 李明聰委員			
1	請補充說明簡報 P.18 臺北市、臺南市實驗場域模型與輸出資訊不同之原因。	謝謝委員指教，已補充說明於 3.2.1 之二、實測模型架構，(三)臺北與臺南模型比較，與表 3.2.1-5。	同意
2	簡報 P.80 顯示實測結果有好有壞，是否與模型訓練績效(簡報 P.22)有關，如簡報 P.22 偏好平均沒有位於預期趨勢中，簡報 P.26 頁偏好延滯但延滯績效並非最佳，請補充說明模型訓練績效圖形之趨勢與各偏好績效	謝謝委員指教，已補充部分績效表現與偏好對應不明顯之推測原因於 3.2.3 AI 強化學習模型訓練結果。	同意

	差異之原因。		
3	請補充說明簡報 P.26 多任務模型架構與應用情境。	謝謝委員指教，已補充說明於 3.2.1 第五項「多任務強化學習模式」中。 相關應用情境主要係不同時段使用之時制計畫或時相組態有所差異，減少因時段切分或時制計畫改變而增加 AI 強化學習模型訓練次數，期望提出一泛用型(全時段)AI 強化學習模型，除可節省訓練次數與時間外，亦可能增加 AI 強化學習模型對於時制計畫因應彈性，降低因時制計畫改變造成之影響。	同意
四、許佳興委員			
1	請確保文件品質，例如圖表不清晰、資料數據被截斷、參照不正確(頁 90)、以及公式應獨立清晰表示等等，此外，報告書內容使用「本所」用語，建議修正為「本團隊」之計畫用語。	謝謝委員指教，已重新檢視報告書內容並修正。報告書中有關團隊執行之工作內容修正為「本團隊」，而運研所之相關內容則使用「本所」。	同意
2	第二章參考文獻提出許多演算法、公式、與架構，建議可於小節最後總結說明與本計畫之關聯，以及後續發展應用。	謝謝委員指教，已於 2.3.3 小節納入總解說明關聯與後續發展應用。	同意
3	實測所遭遇之問題應可透過拉長實測時間解決，建議後續可以增加實測時間，以驗證模型績效。	本次實測遭遇之設備異常、通訊異常問題除偶發性斷線外，尚有因中心機房通訊伺服器故障導致，若故障問題未經排解，則異常發生頻率將逐漸增高，此時無法經由延長實測時間排解，若故障問題解決，則可延長實測時間扣除先前發生異常問題之時段，以取得 AI 強化學習模型號誌控制之績效，然因本計畫時程緊迫，故無法延長實測時間，未來規劃計畫期程將全面考量實測可能發生之情形進行安排。	同意
4	請補充說明簡報 41 頁，AI 號誌控制模型與固定時制績效分析是否差異顯著。	本次 AI 強化學習模型績效評估，係以數值絕對差距為比較基準，並無設立比較區間或比較門檻值。	同意
五、劉瑞麟委員			
1	臺北市場域使用「各分相結束時分車道分車種(大車/小車/機車)停等車輛數及總車		

	輛數」當作 AI 訓練模型之輸入值：		
	1. 停等車輛數及總車輛數之定義差異，應具體說明。	1. 停等車輛數之定義為，在偵測路口停等車輛時，如該車輛行駛方向對應路口號誌為紅燈，則為停等車輛；偵測時不分號誌狀態則為總車輛數。	同意
	2. 模擬時是否有考慮偵測設備偵測停範圍之限制，如壅塞程度已超出偵測範圍，無法取得準確之前述輸入值，請說明對模型訓練、績效之影響及使用此輸入值之原因為何。	2. 在模擬環境中，雖然我們能夠觀測整個路網的情況，但為求還原，我們設置在模擬中的偵測器範圍與真實偵測器的偵測範圍相同。所以模擬中 AI 強化學習模型的輸入資料，亦為實際現場偵測器能夠給予 AI 強化學習模型的輸入資料。	同意
2	模擬環境是否可考慮行人對轉向車影響、公車停靠及臨時停車等行為，請於報告書內補充說明。	模擬環境在 SUMO 環境中，可能較難還原公車停靠以及行人的影響，主要是因其功能及參數偏向車流；VISSIM 模擬環境中，具有 Viswalk 外掛程式可模擬行人，亦可還原公車停靠情況。惟本次臺南市實驗場域並無行人影響，臺北市實驗場域亦有行人專用時相，故行人影響應較小；而公車停靠的部分，臺南市實驗場域亦無此影響，臺北市實驗場域在靠近克強路北側，已接近本次實測場域末端，並經現地觀測，評估影響應在可接受範圍。	同意
3	圖 3.2.3-1，X 座標為 60K，文字說明 60 萬步，請確認是否為誤植。	謝謝委員指教，此部分為誤植，已修正圖 3.2.3-1 文字說明。	同意
4	請補充說明使用幹/支道延滯時間與固定時制的變化量及幹/支通過量與固定時制的變化量，作為獎勵目標原因？另前述獎例目標是否可於現實環境蒐集取得，以確保套用於現實狀況有一致之評估指標。	為使 AI 強化學習模型更有效率的學習，因而使用幹/支道延滯時間與固定時制的變化量及幹/支通過量與固定時制的變化量，作為獎勵目標，期待 AI 強化學習模型能往高於固定時制的績效探索與學習。由於因現場偵測設備有限，獎勵績效無法由現場取得，但本研究使用的 AI 強化學習模型可以離線訓練，因此於模擬軟體中可取得獎勵績效，模型將有機會優化。	同意
5	獎勵目標權重，僅於 P.80 敘述 8:2 與 [0.1, 0.4, 0.1, 0.4]，可否建立偏好與權重的表格，以利閱讀比較。	謝謝委員指教，已於 3.2.3 AI 強化學習模型訓練結果中，新增表 3.2.3-1 與 3.2.3-6 說明獎勵目標測試績效權重。	同意
6	對於實證結果之資料整理及說明建議應清楚，建議以	謝謝委員指教。已補述於 4.4.3 節中。	同意

	一張表綜整說明各種指標衡量之方式、指標計算方式、指標衡量日期、衡量結果等。		
7	臺北市場域通過量及延滯有改善，惟旅行時間指標卻無改善，請說明為何有此差異。	臺北市實驗場域實測結果(表 4.4.2-1 及表 4.4.2-2)假日 AI 號誌控制平均延滯皆較固定時制略差，但服務水準幾乎維持同一等級；平日第一天 AI 號誌控制平均延滯亦較固定時制略差但服務水準不變，惟第二天昏峰克強路平均延滯較固定時制減少 0.2 秒、德行西路平均延滯減少 9.5 秒，對應旅行時間績效(圖 4.4.2-4~7)，亦有相同趨勢表現。	同意
8	建議 AI 號誌控制實測調整之每周期號誌時制資料及連鎖續進圖應補充於報告書，並可就 AI 號誌控制連鎖執行與固定時制執行差異進行探討。	有關臺北市實驗場域之 AI 號誌控制實測連鎖續進執行與固定時制之差異詳 4.4.2 節之二、綠燈連鎖情形所示。	同意
9	實測時發生硬體設備異常、通訊異常等問題，是否能利用較長時間的實證，來扣除這些異常的干擾。	本次實測遭遇之設備異常、通訊異常問題除偶發性斷線外，尚有因中心機房通訊伺服器故障導致，若故障問題未經排解，則異常發生頻率將逐漸增高，此時無法經由延長實測時間排解，若故障問題解決，則可延長實測時間扣除先前發生異常問題之時段，以取得 AI 強化學習模型號誌控制之績效。	同意
10	圖 4.4.2-1 圖示不清楚，建議修正。報告圖說曲線圖用顏色表示，建議改加標示，以利黑白報告閱讀，且圖不宜太小。	謝謝委員指教，已修正圖 4.4.2-1 之圖示。	同意
11	簡報 P.22、P.23 之獎勵目標值在與臺北市、臺南市的績效差異型態迥異，建議討論其原因。	謝謝委員指教，已補充部分績效表現與偏好對應不明顯之推測原因於 3.2.3 之 AI 強化學習模型訓練結果。	同意
六、卓明君委員(書面審查意見)			
1	目前高公局與地方政府於交流道區號誌協控方式，係於入口匝道啟動儀控，造成車輛回堵至地方道路至可接受的觀察點時，透過橫向	感謝委員提醒，於後續交流道區域協控模式開發上，將多方考量可能影響地方道路及國道主線車流之干擾變數。	同意

	連繫協請放寬儀控率，以紓解回堵至地方道路之車流；出口匝道車輛如受路口號誌管制致回堵至主線造成嚴重壅塞時，亦透過橫向連繫協請增加下匝道左轉綠燈秒數等，以兼顧尖峰時段兩種不同等級道路之服務水準並紓解壅塞車流。另國道常受交流道下游事故或故障車輛影響致主線服務水準驟降情形，故不得不緊縮入口匝道儀控管制，爰本計畫在開發交流道區域協控模式時，建議納入可能影響國道主線車流之干擾變數。		
2	楊梅交流道屬於國1楊梅至新竹南向路段實施精進式匝道儀控的範圍，該路段精進式匝儀係於一般週休六日及連續假期上午開始實施，故本計畫在假日的協控模式部分，請洽本局北區交控中心研議運作方式。	感謝委員意見，因本計畫交流道儀控係從都會區角度進行探討，因此僅考量平常日與高速公路獨立式匝道儀控。	同意
七、交通部高速公路局			
1	報告書 P.20 之 2.1.2 節目前匝道銜接地地方道路處皆設有號誌，現由地方政府控制，請協助確認是否屬於下匝道儀控設備。	謝謝委員指教。經查目前於下匝道銜接處皆有地方政府維運之號誌，惟 2.1.2 節之(二)下匝道、(三)小結文意為強調下匝道處不若上匝道處有儀控率作為模式輸入，已修改本文以避免混淆。	同意
2	報告書 P.155-P.156 之 5.1、5.2 節多以順暢、壅塞等車流速率用詞作為時制計畫調整依據，但(四)則是利用道路服務水準的道路流量、容量作為判斷，希望可敘明未來時制調整之判斷依據。	謝謝委員指教。未來該場域主要仍以 2011 公路容量手冊之服務水準為判斷，5.1、5.2 節之順暢、壅塞亦由該服務水準為判斷標準。	同意
3	本計畫僅有平面壅塞，道路服務水準尚佳時有告警機	本計畫平面道路壅塞告警事件，係經由跨單位已建立之事件通報與反應機制將告警事件傳達	同意

	制，但是相反情況卻沒有，請補充說明。	予高速公路局相關區域養護工程分局人員，得以即時掌握跨區域交通壅塞狀況與需求，並據以評估匝道儀控管理作為，減少國道交流道上匝道停等車隊回堵進而影響平面幹道運行順暢之情形。 國道主線壅塞告警事件計畫蒐集 VD 或 eTag 資料，計算即時國道主線速率或旅行時間，作為判斷壅塞依據，控制平面路口號誌，減少車流持續匯入加重國道主線壅塞情形。	
4	5.1 節(三)、(四)當出入口匝道與地方平面道路皆產生壅塞訊號時候，目標對於高速公路及地方道路的時相分配是否依道路等級有權重配比差異。	謝謝委員指教。目前預計輸入模式的參數，高速公路主線及平面車流資訊皆納入其中，而後續待模型建置完成後，可再行評估不同道路權重對於模型獎勵目標的影響顯著性為何。	同意
5	P.165 之 5.4.1 績效指標設計對應高速公路績效評估定義，請再說明如何計算上、下匝道旅行時間、停等時間，以及如何評估事前事後差異。	謝謝委員指教。5.4.1 節之指標設計皆為初步評估，雖皆為交通常用之績效，惟最後定案仍待評估後決定。而目前預計於實測前後皆收取 GVP 資料作為旅行時間前事後差異之依據。	同意
八、臺北市交通管制工程處			
1	AI 強化學習模型是否可以使用多個偏好，或偏好是否可以有優先順序排列。	可於一個時段內有不同的偏好選擇，但偏好使用的順序需事先決定權重，應用時，AI 強化學習模型會依照此偏好權重，產出號誌控制最佳決策。	同意
2	本計畫使用通過量與停等延滯，然週期長短會影響此兩個績效，建議可使用其他參數(例如：旅行時間)做為績效判斷。	謝謝委員指教。規劃於明年度之計畫的交流道協控場域，將旅行時間納為績效指標之一。	同意
3	車流量調查為暑假期間，請問此車流量調查是否有納入模擬環境中進行 AI 強化學習模型訓練，是否會影響模型訓練結果。	車流量調查目的為進行模擬環境之校估，以及提供 AI 強化學習模型訓練使用，故相關車流量將影響 AI 強化學習模型訓練結果，後續可考慮納入非寒暑假之調查資料。	同意
九、新北市政府交通局(書面審查意見)			
1	實測臺南市與臺北市實驗場域績效分析均顯示假日昏峰時段 AI 表現不及	謝謝委員指教。目前已歸納出幾點原因，如車流型態(轉向比等)分布不同，會造成 AI 強化學習模型只針對某特定車流型態做訓練。當現況改	同意

	TOD，建議可初步歸納原因作為後續使用參考。	變過大，模型可能會失準，本次模型在臺南市實驗場域即遭遇此情況。 後續的交流道場域亦會以此經驗作為參考，可能會不同於以往做交通評估，只抽樣一天作為事前的作法，需要多蒐集不同時間之訓練資料。	
十、臺中市政府交通局			
1	4.3.2 節 AI 強化學習模型所計算之週期秒數與 TOD 不同，可能會造成號誌遞亮，在模擬環境中是否有將車隊無法續進造成綠燈使用率下降之因素納入考量。	謝謝委員指教。模擬環境中，雖無法直接紀錄綠燈使用率，但可以記錄車輛的行駛資訊，所以綠燈使用率低之情境，可能會以較長的總旅行時間或其他方式呈現出來。	同意
2	當動態號誌路口與上下游路口因週期秒數不一致，所產生可能之遞亮情形時，系統如何計算各分相可能尚未抵達之車輛數量而分配下一分相的號誌秒數，另外不連鎖情形是否可能造成用路人的反應。	1.目前實測運作上，係偵測近路口停車車流資訊或路口轉向量資訊，偵測設備有限之狀況下，無法偵測尚未抵達之車輛數量，故分相號誌秒數實際上取決於路口偵測到之車流狀況。 2.因本計畫 AI 強化學習模型均為變動週期，臺南市實驗場域為獨立路口距離上下游路口均超過 500 公尺以上，故較無不連鎖之影響；而臺北市實驗場域三個控制路口為同一群組，群組內保持連鎖，雖與上下游距離近，但模型有考量臨近路口週期資訊，雖有週期變動，但不影響續進，因此實測時均無接獲用路人反應。	同意
3	本計算挑選之路口尖峰週期皆為 200 秒以上，是否表示 AI 號誌控制僅適用於較長週期之路口。	AI強化學習模型號誌控制可依需求於時相或週期進行決策，且各長短週期路口皆可適用。	同意
4	觀察臺南市實驗場域所分配之週期秒數，發現前後週期震盪較大，是否會造成用路人在特定週期不耐久候，認為號誌故障而闖紅燈的情形。	若現場並無發生綠燈道路車流紓解完畢，但綠燈秒數仍未使用完畢之情形，應不會造成用路人誤判號誌故障。為妥善運用綠燈，臺南市實驗場域之 AI 強化學習模型可即時反應現場車流，於每一時相進行運算輸出下一時相秒數。	同意
十一、臺南市政府交通局(書面審查意見)			
1	4.3.2.3節單一路口場域(臺南市台86-16甲路口)模擬成果部分(P.107)，「模擬情境績效彙整表」建議可再分別表列幹、支道之通過量及平均停等延滯的模擬	謝謝委員指教。4.3.2.3 節中，模擬情境績效彙整表中之通過量，已補充說明何者為幹支道。惟平均停等延滯，團隊在產製資料時並無分幹支道，故本次結果仍以路口平均停等延滯為主。	同意

	結果。		
2	4.4.1 節臺南市實驗場域實測結果部分(P.146~P.149)，依據路口實測結果績效表，可大致看出路口整體績效不分，假日績效 AI 號誌控制較 TOD 惡化，而平日績效則較 TOD 為佳，請問團隊是否能簡單說明其兩者反差的原因？另依據實測結果，後續又將朝哪種方向進行 AI 號誌控制模式的調整與優化？	謝謝委員指教。目前假日的績效較平日差，團隊評估車流型態改變可能為主因。團隊蒐集 5 月中的車流資料時，國三往臺 86 之左轉車流轉向比較高，故 AI 強化學習模型的訓練結果亦傾向給予第一時相較大的基本秒數。惟實測日經現場觀測，左轉車流不如訓練資料大，故模型縮減第一時相秒數的效果有限。 未來優化的方向，針對流量的部分，可更新更多組現場的流量資訊，使 AI 能應付更多的車流變化；AI 訓練的部分，可嘗試調整模型調整超參數、觀測數據的標準化或是調整獎勵的設計。	同意
3	4.4.1 節臺南市實驗場域實測結果部分(P.146~P.148)，依路口平日之車流型態，其昏峰與晨峰具有較明顯的、互為相反的方向性重現性壅塞，惟實測結果晨峰績效惡化，昏峰改善，請問團隊是否能簡單說明其兩者反差的原因？	謝謝委員指教。目前推測晨峰績效較差的原因為晨峰有兩個不同時段的時制計畫。而在模擬及實測中發現，AI 強化學習模型在處理五時相時之績效皆不如四時相，可能為晨峰績效較差之原因。	同意
十二、高雄市政府交通局(書面審查意見)			
1	未來計畫場域有無於高雄轄內之幹道或交流道周邊研究之機會。	感謝委員意見，明年度計畫平面幹道號誌控制將延續本期計畫之場域，而交流道協控場域，亦規劃以本期建置完成之楊梅交流道模擬環境進行後續實測。	同意
十三、交通工程技師公會			
1	建議補充模型、偵測設備之比較表，幫助讀者掌握差異。	謝謝委員指教，已補充說明於 3.2.1 之二、實測模型架構，(三)臺北市與臺南市模型比較，與表 3.2.1-5。	同意
2	簡報 P.22 之圖表應越靠近原點之結果越好，建議可調整 XY 軸方向。	謝謝委員指教，已嘗試過原點績效圖，因各績效系均與 TOD 績效比較，績效會有正有負，如以原點績效呈現，會導致各筆資料間距更集中，不易閱讀。	同意
3	簡報 P.22、P.23 之圖表使用單位不一致，可能因此導致臺北市、臺南市實驗場域圖表呈現趨勢與結果不一。	謝謝委員指教，因場域不同與控制基準不同，呈現績效量級也會不一致。若調整量級一致，可能會導致差異比較更加不明顯。	同意
4	績效分析時應排除實測發	謝謝指教，因本次臺北市、臺南市實驗場域皆有	同意

	生問題之時段，才可表現出模型實際之績效。	發生控制器控制異常、網路不穩、偵測設備斷線等問題，扣除發生問題時間，部分時段實際運作不足 1 小時，為衡量實際 AI 強化學習模型績效，亦考慮於後續計畫延長實測時間。	
5	Q-Pensieve 使用隨機決策，若已知輸出多少秒數可取得較佳績效，為何還建議使用隨機決策輸出秒數，請補充說明。	此隨機性為訓練時，使 AI 強化學習模型具備有探索能力，實測運用上，於 AI 強化學習模型產出後，採用平均值作為決策定值輸出，已移除隨機性之影響。	同意
6	模型實測績效之差異是否可能與現實環境中沒有設備取得真實停等延滯，僅能使用預估結果做為績效有關。	臺南市實驗場域實測績效係以旅行時間延滯計算，而臺北市實驗場域則是以停等延滯計算，旅行時間延滯因資料蒐集時間尺度較停等延滯為大，可能有實測結果相對較不精確之情形。 惟旅行時間延滯之計算較傳統人工調查停等延滯便利，亦可改善傳統人工判讀問題、減省人力成本，尚具使用意義。	同意
7	AI 強化學習模型是否可以於路口實作過程同時進行訓練，或必須離線訓練完成才可於路口使用。	AI 強化學習模型藉由對於運算之輸出進行績效評分，以習得較佳的策略，現實環境較難取得用以評分之績效，故 AI 強化學習模型需在模擬環境中進行訓練，無法於實作過程同時訓練。然實作過程所取得之資料皆可於事後進行 AI 強化學習模型離線訓練，以更新 AI 強化學習模型輸出。	同意
8	交流道之模擬環境需注意與市區道路之差異：包含高速公路旅行時間、匝道停等或排隊長度。	謝謝委員提醒，本案後續交流道區域，目前預計考量之績效指標包含委員提及之旅行時間及停等延滯，在建置模擬環境及運行模型時，將會注意兩不同道路系統帶來之影響。	同意
十四、本所運輸資訊組(書面審查意見)			
1	P.68 與 P.69 在說明臺北市實驗場域與臺南市實驗場域 AI 強化學習模型規劃的獎勵設計時均指出「獎勵目標與臺北場域設計概念相同，為了使 AI 強化學習模型有效率的學習，與避免往過大與過小的時相秒數探索」，請補充說明在學習過程設計出「避免往過大與過小的時相秒數探索」的機制。	謝謝委員指教，已補充說明於 3.2.1 之一、AI 強化學習模型規劃，臺北市、臺南市實驗場域之模型說明中。	同意
2	P.72 請補充說明圖 3.2.1-3	謝謝委員指教，已補充說明於 3.2.1 之一、AI 強	同意

	架構圖的「critic」模型結構圖。	化學習模型規劃，臺北市、臺南市實驗場域之模型說明與圖 3.2.1-2、3.2.1-4。	
3	P.75 請補充說明本案運用國網中心運算資源扮演角色，以及運用方式。	謝謝委員指教，已補充說明於 3.2.2 AI 強化學習模型訓練模擬環境程式與軟硬體說明中，相關使用規格補充於表 3.2.2-2。	同意
4	P.83 請補充說明多任務強化學習模式(Task-Aware)圖 3.2.3-12 與圖 3.2.3-13 驗證紀錄中，「獎勵目標與偏好權重的加權計算值」的定義，以及強化對於此 2 個驗證紀錄圖的內容說明。	謝謝委員指教，已補充說明於 3.2.3 節之多任務強化學習模型訓練過程中。	同意
5	P.76 至 P.85 臺南市實驗場域績效評估的圖文說明，觀察各圖內容說明，似乎在平日離峰的台 19 甲支道延滯，以及台 86 幹道與台 19 甲支道的模擬績效均較定時號誌控制(TOD)大幅增加，另多任務強化學習模式僅於平日昏峰幹道延滯有所改善，其餘均不如 TOD，請對此現象進行探討，並提出後續改善建議。 各項模擬績效顯示多任務強化學習模式尚無法比分開訓練的模型績效好，請補充說明後續在分開個別訓練與多任務強化學習模式的採用考量與適用時機。	1.謝謝委員指教，已補充 3.2.3 「AI 強化學習模型訓練結果」其績效評估的圖文說明，並推測部分績效表現與偏好對應不明顯之可能因素。 2.多任務強化學習的目的在於希望利用一個模型去解決各個時段的任務，使訓練出來的模型不侷限在特定任務上，同時也希望利用任務間的相似性來最大化資料採樣的利用率。關於績效不理想的部分，我們認為是僅有使用單一 Embedding Network 導致模型中的 Actor Network 與 Critic Network 互相拉扯導致的，因此我們認為可以改善優化方向為於 Actor 和 Critic Network 都獨立去訓練 Embedding Network，使多任務可以更準確地分辨不同任務，達到提升訓練績效的目的。 3.相關應用情境主要係不同時段使用之時制計畫或時相組態有所差異，減少因時段切分或時制計畫改變而增加 AI 強化學習模型訓練次數，期望提出一泛用型(全時段)AI 強化學習模型，除可節省訓練次數與時間外，亦可能增加 AI 強化學習模型對於時制計畫因應彈性，降低因時制計畫改變造成之影響。雖目前多任務強化學習模式尚無法比分開訓練的模型績效好，但仍希望透過加強訓練，以獲得泛用型(全時段)AI 強化學習模型。	同意
6	4.3.2 節「模擬測試」部分，		
	1.請將 4.3.2.1 節「車流模擬軟體」內容移至 3.2.2 節	1.已於 3.2 節加入補述文字，避免敘述不清。(加在 3.2.2 最後：較詳細的模擬軟體說明，詳參	同意

「AI 強化學習模型訓練模擬環境程式與軟硬體說明」並加以彙整。	4.3.2 小節)惟 3.2 節旨在說明號誌控制模型，3.2.2 僅為說明訓練環境投入模型之參數；4.3 節主軸為實測計畫，4.3.2 節旨在說明實測前之模擬環境，故該段落有存在於 4.3 節之必要。	
2.此處應該是在將強化學習號誌控制模式輸出結果正式下載到路口號誌控制器前的動態模擬測試，以評估模型輸出之合理性與穩定度，而非再利用模擬環境來讓代理人任意探索各種策略組合可能性的模型學習，請再加強說明以避免誤解。	2.已修改描述。	同意
3.另本節內容與 3.2.3 節的「模擬績效」角色不同，亦請於報告中加強說明。	3.已補述於 4.3.2.2 節。	同意
4. 4.3.2.2 節 P.106 之圖 4.3.2.2-1 示意圖中包括「RL 模型優化」，惟未見說明「RL 模型優化」內容，另此請補充說明此處「RL 模型優化」與 P.72 第三章圖 3.2.1-3 本期 Agent 模型架構圖之關聯。	4. 4.3.2.2 節之優化，其為觀察輸入了現場設備資料(測試資料)後，其輸出結果是否具合理及穩定性。如輸出結果不具穩定或合理性，例如總週期長超過 10 分鐘，則在模型輸出結果前對其設定些許限制如綠燈時長上下限、合理週期長等。圖 3.2.1-3 則為訓練模型時，神經網路與資料(訓練資料)互動之成果。	同意
5.請加強說明 4.3.2.2 節對於本案強化學習號誌控制模式輸出合理性與穩定度的評估設計，並於 4.3.2.3 節與 4.3.2.4 節中依此設計，針對各案例說明其合理性與穩定度的評估結果。	5.已修改 4.3.2.2 節說明，以及 4.3.2.3 節及 4.3.2.4 節之說明。	同意
6. 4.3.2.3 節所有合理性分析相關的圖均不清晰，請加以改善。	6.已重整圖片並加強其清晰度。	同意
7.P.70 臺北市與 P.71 臺南市的 Agent 模型結構圖都有獎勵偏好，但在 4.3.2.4 節臺北市模型有探討不同	7.已在 4.3.2.3 節補充偏好說明表格及其比較。	同意

	AI 權重偏好績效比較，而在 4.3.2.3 節臺南市模型未見探討，請說明緣由與進行報告調整。		
7	4.3.3 節「實測計畫」部分， 1.請補充說明 P.142 於臺南市實驗場域使用 GVP 資料推估停等延滯，以及於臺南市實驗場域使用車輛偵測器推估通過量與停等延滯等的作法。	1.臺南市實驗場域使用 GVP 資料推估延滯，作法補充詳報告書 4.4.1 節之二、路口平均延滯績效；臺北市實驗場域使用車輛偵測器蒐集推估延滯，補充詳報告書 4.4.2 節之三、路口平均延滯績效。	同意
	2.請將附錄二之臺北市與臺南市實測計畫書重點摘錄於本文，同時亦將實測情形全部置於本文中。	2.已將實測計畫書重點摘要於 4.3.3 節，實測情形整理於 4.3.4 節。	同意
	3.本案 AI 多目標 RL 號誌控制模型採用通過量與路口平均停等延滯兩個目標，請補充說明實測過程所選用多目標權重的比例，以及說明實測結果是否符合原設計期望是否能反應？	3.謝謝委員指教，已於 3.2.3 AI 強化學習模型訓練結果中，新增表 3.2.3-1 與 3.2.3-6 獎勵目標測試績效權重表。另，本團隊建初始以平均偏好權重，惟效果並不顯著，因此設置不同的偏好設定，其相對應之績效表現亦較為顯著。	同意
	4.請補充於臺北市與臺南市實驗場域於實測過程 AI 強化學習模型所產生各階段的號誌周期長度與各路口時比、TOD 時制計畫內容。	4.臺北市與臺南市實驗場域於實測過程 AI 強化學習模型所產生號誌週期長度與各路口時比皆與模擬時相差無幾，詳報告書 4.3.2.3 及 4.3.2.4 節所示。TOD 時制計畫內容詳報告書表 4.1.1-2 及表 4.1.2-2 所示。	同意

	5.對於本案 AI 強化學習 (RL)模式實測所得路口延滯績效多數均不如現況 TOD 運作，除目前在現場設備或通訊異常因數外，請從 AIRL 模式與相關參數設計、模擬模式、車流特性與號誌控制參數等面向進行後續再優化具體建議。	5.於 Q-Pensieve 的訓練，有許多訓練參數可以做調整，像是 Q Replay Buffer 的容量大小或是 Learning Rate，目前本計畫使用的超參數主要是採用 Q-Pensieve 論文的建議數據，論文與本計畫的實驗環境性質較為不同，State 與 Action 的維度與數量級有些差異，因此調整超參數和觀測數據的標準化會是可以嘗試優化的部分，並且與本期的實驗中有觀察到，我們將同樣的績效目標轉換成不同的訓練目標(Ex:線性轉換，使用和基準的差異作為訓練目標)，都會對表現有所影響，因此，Reward 的設計，也會是優化的考慮方向。	同意
	6. 4.3.3 節的小結內容過於薄弱，請再加強與補充。	6.謝謝委員指教，已補充 4.3.3 節小結內容。	同意
8	第五章「強化學習號誌協控模擬環境構建規劃」部分， 1.對於 P.174 表 5.4.4-4 與 P.175 表 5.4.4-5 中雖然整體差異比例符合+-10%，對於差異比例大於+-10%均應探究其原因與進行可能之再精進。	謝謝委員指教。如 5.4.4 節中所提及，在模擬環境中有眾多因素影響還原性。表 5.4.4-4、5.4.4-5 中，部分路段之 MAPE 及 GEH 表現不佳，未來可針對該路段先行評估可能原因。惟整體平均符合規範，國外文獻對於 GEH 亦為整體路網綜合考量，針對性調整可能會造成其他路段表現改變，故仍須評估模擬路網之可用性，再進行模擬路網調整。	同意
9	請補充說明桃園市楊梅交流道實驗場域目前相關交控設施與團隊規劃 AIRL 模型參數於實測時資料來源等。	本計畫規劃交流道協控實測場域其資料來源分為平面幹道與高速公路兩部分，平面幹道規劃以 AI 影像辨識偵測路口轉向量與號誌資訊作為模型輸入；而高速公路部分則規劃自交通部高速公路局「交通資料庫」取得即時資料作為模型輸入。	同意
10	第七章建議內容過於薄弱，應針對本案實測結果提出後續精進與優化建議，對於桃園市楊梅交流道號誌協控於 113 年進行 AIRL 模型訓練學習與實測等應提出具體建議。	遵照辦理，已補充後續精進與優化建議與交流道協控實測具體建議。	同意
11	有關報告初稿中之錯漏字、語句不適當與資料來源修正部分，請於會後洽本所。	遵照辦理。	同意

十五、主席結論			
1	審查會議各委員及與會單位研提之口頭及書面意見，請資拓宏宇國際股份有限公司整理「審查意見處理情形表」，逐項回應辦理情形，納入修正之期末報告，於112年12月20日提交。	遵照辦理。	同意
2	本計畫經徵詢審查委員意見，期末報告初稿審查通過，請資拓宏宇國際股份有限公司後續依本所出版品印製相關規定撰寫報告。	遵照辦理。	同意

