

國立臺北藝術大學  
電影與新媒體學院  
新媒體藝術學系碩士班

M. F. A. Program, Department of New Media Art  
School of Film and New Media  
Taipei National University of the Arts

碩士論文

虛實共構的時空意識——符號堆疊的藝術實踐

Symbol Stacking: Temporal and Spatial Perception Between the Virtual and the Real

研究生：江育任 撰

Graduate Student：Yu-Jen Chiang

指導教授：孫士韋教授

Professor：Shih-Wei Sun

中華民國 115 年 1 月

January, 2026

國立臺北藝術大學

Taipei National University of the Arts

國立臺北藝術大學 新媒體藝術學系碩士班

碩士學位考試委員會審定書

114學年度第1學期

江育任 (CHIANG, YU-JEN) 君所提之(論文/作品連同書面報告  
/技術報告/專業實務報告)

題目：(中文) 虛實共構的時空意識——符號堆疊的藝術實踐

(英文) Symbol Stacking: Temporal and Spatial Perception  
Between the Virtual and the Real

經本委員會審定通過，特此證明。

學位考試委員會

召集人 郭令權 

委員 郭令權 

王連晟 

孫士韋 

指導教授 孫士韋 

系主任 

中華民國 115 年 1 月 8 日

國立臺北藝術大學 新媒體藝術學系碩士班  
畢業大綱審定書

江育任 CHIANG, YU-JEN 君 (創新科技組 Emerging  
Technology, 學號 211162017) 所提之作品連同書面報告

題目：(中文) 虛實共構的時空意識

(英文) Temporal and Spatial Consciousness Between the Virtual and  
the Real

經本委員會審定通過，特此證明。

召集人 郭令權

委員 郭令權

王連晟

孫士韋

指導教授 孫士韋

系主任

2024 年 11 月 27 日  
November 27, 2024

國立臺北藝術大學 新媒體藝術學系

# 112 學年度碩士班學年評鑑

## 評鑑委員審定書

創新科技組

江育任 君 (學號 211162017) 所提之評鑑作品

### 《面鏡 The Mirror》

經本委員會審議，決議

通過

不通過

評鑑委員

召集人

委員

召集人	戴磊明		
委員	林文瑞	王連成	魏德榮
	楊淑娟	孫七尊	黃若竹
	戴磊明	蕭雁宇	葉嘉心
	袁子心	林書翰	

中華民國 112 年 9 月 22 日

國立臺北藝術大學

Taipei National University of the Arts

# 虛實共構的時空意識——符號堆疊的藝術實踐

指導教授：孫士韋

研究生：江育任

## 摘要

在數位科技滲透日常的當代，現實與虛擬的邊界日益模糊，原本建立於現實世界中的穩定線性時空維度，正逐漸轉變為一個在虛實之間不斷重組與生成的經驗場域。在此脈絡下，時間、空間與主體意識將在多重媒介與符號運作中相互交織、共構而生。本研究以「虛實共構的時空意識」為核心命題，探討在當代虛實交織的環境中，各種符號與媒材如何透過「符號堆疊」的創作方法，在時間、空間與主體意識三個層次上，形塑出我們對身體感知、記憶、身份與自我認同的理解。

當代的感知將不再只是在人際間或實體環境中與符號的互動關係，科學技術與演算系統亦逐步成為人類的知覺生成的重要構成要素。本研究透過三件藝術創作實踐，分別以影像、互動裝置與線上系統為媒介，各自以不同的角度回應感知經驗在虛實交錯的時空條件下，可能被拆解、轉譯與生成的狀態。《National Freeway 5》探討運動中的靜止，透過影像堆疊技術表現身體感知與記憶時間之間的錯置經驗；《The Mirror》以雙面鏡與即時互動系統呈現觀看與被觀看的視覺結構及其伴隨的權力流動；《Self Echoes》則結合生成式人工智慧與社群資料，生成虛擬替身「數位回聲」，映照演算法時代中身份被資料化後的流動與再現。

**關鍵字：**時空意識、身體感知、符號堆疊、身份資料化、互動裝置、生成式人工智慧、人臉辨識

國立臺北藝術大學  
Taipei National University of the Arts

Symbol Stacking: Temporal and Spatial Perception Between the Virtual and the Real

Professor: Shih-Wei Sun

Graduate Student: Yu-Jen Chiang

### **Abstract**

As digital technologies increasingly permeate everyday life, the boundary between the real and the virtual has become blurred. What was once a stable and linear spatiotemporal framework is now experienced as a dynamic field continuously reorganized between virtual and physical conditions. Within this context, time, space, and subjectivity are no longer fixed structures but are co-constructed through the operation of media and symbols. Centered on the concept of spatiotemporal perception shaped through virtual–physical entanglement, this paper examines how the creative method of symbol stacking shapes perceptions of the body, memory, identity, and self-recognition.

國立臺北藝術大學  
Taipei National University of the Arts

This paper presents three works, a video, an interactive installation, and an online system, each exploring how perceptual experience is fragmented, translated, and regenerated in hybrid environments. *National Freeway 5* investigates stillness within motion through image stacking, revealing the disjunction between embodied perception and remembered time. *The Mirror* utilizes a two-way mirror and a real-time interactive system to reveal the visual structures of seeing and being seen, along with the shifting power relations inherent in the gaze. *Self Echoes* integrates generative AI with social media data to produce “digital echoes,” reflecting the fluidity and representation of identity in an algorithmic era.

**Keywords:** Spatiotemporal Perception, Embodied Perception, Symbol Stacking, Datafication of Identity, Interactive Installation, Generative AI, Face Recognition

國立臺北藝術大學  
Taipei National University of the Arts

## 致謝辭

個體的行為差異來自先天與環境的共同影響，能成就今日的自己，一路上承載了許多人的付出與成全，能夠在歷經三年半後順利取得碩士學位亦是眾多條件交織的成果。因此在感謝自己之前，首先最要感謝我的父母，給予我豐富的生命經驗，也交付予我為社會所見的模樣。介於有和沒有之間一直是我人生至今的過程寫照，足夠幸運的是，每次的回頭都能發掘混沌基於自己的意義。謝謝指導教授孫士韋老師的悉心指導，無論是從創作或從研究的角度，總在談話間讓並非工程背景的我，覺得自己像是在同個迴圈內的局內人，除了討論過程的指點迷津外，也總是相信我可以在各個層面上再多做一點，而常作為對話結尾的繼續加油，也真如其字面意義，成為支撐我再次向前的推力，另外也誠摯感謝擔任口試委員的郭令權老師與王連晟老師，這本論文可以被順利地開展與收束歸功於老師們在期間提供的建議與指引，在此以表言謝。

讀研究所這件事，既是沉潛於大學畢業之際未竟的根柢，也是在職場中醞釀了四年才順利冒出的枝桠。謝謝實踐大學媒體傳達設計學系的栽培，特別感謝蘇志昇老師在求學及考研過程中的次次開導與協助；也謝謝曾在 Impact Hub Taipei 共事過的每位夥伴，尤其感謝共同創辦人張士庭先生與陳昱築先生，溫柔地對待我的每個決定；謝謝國立臺北藝術大學新媒體藝術學系以及隔壁的文資學院提供多樣的學習養分，帶領我走入森林；謝謝路上一同披荊斬棘的碩班同學昱廷、培芸、芳宜、詠心、子晏在這座北市偏遠山坡上的相互砥礪，還要謝謝蒼茗學姐、竺勳學長、章晟學長、佩庭學姐在聯展過程中的照應；謝謝芳瑜協助論文的校對，以及謝謝好友阜鎧、宇淨、詠詠、竹芸、埤桿、語豪、之謙、品誠以及曾在人生的各個節點提供一切支持、支援與關心的每位親朋好友。最後謝謝未臻於完善的自己，在有和沒有之間，依然選擇持續努力。

江育任 謹識 於國立臺北藝術大學

民國 115 年 1 月

國立臺北藝術大學  
Taipei National University of the Arts

## 目次

摘要 .....	i
Abstract.....	iii
致謝辭 .....	v
目次 .....	vii
圖次 .....	xi
表次 .....	xv
第一章 緒論 .....	1
第一節 研究動機 .....	1
第二節 研究目的 .....	5
第三節 文章架構 .....	5
第二章 文獻探討 .....	7
第一節 問題脈絡梳理 .....	7
一、 虛實交錯 .....	7
二、 時間、空間與意識 .....	7
三、 符號堆疊 .....	8
四、 模糊性 .....	9
第二節 理論基礎 .....	9
一、 具身主體：身體感知與現象學 .....	10

二、	被觀看的主體：鏡像理論與觀看權力 .....	13
三、	碎片化主體：資訊哲學與數位自我 .....	17
第三節	藝術案例分析 .....	20
一、	時間與空間中的身體流動 .....	20
二、	觀看與被觀看的權力關係 .....	24
三、	演算法與虛擬身份的生成 .....	28
第四節	相關技術探討 .....	30
一、	OpenCV Library .....	30
二、	MediaPipe Solutions.....	32
三、	影像堆疊（frame stacking） .....	34
四、	臉部平均影像與臉部辨識 .....	38
五、	生成式人工智慧（Generative AI） .....	43
第三章	創作論述與執行 .....	53
第一節	《National Freeway 5》（兀兀騰騰寂寂蕩蕩） .....	53
一、	創作論述 .....	53
二、	創作動機 .....	54
三、	執行方法 .....	55
四、	作品呈現 .....	62
五、	限制與改進 .....	63

第二節	《The Mirror》(舉頭三尺有什麼)	64
一、	創作論述	64
二、	創作動機	66
三、	執行方法	68
四、	作品呈現	77
五、	使用者體驗回饋	81
六、	限制與改進	84
第三節	《Self Echoes》(我相·人相)	84
一、	創作論述	84
二、	創作動機	85
三、	執行方式	90
四、	作品呈現	109
五、	使用者體驗回饋	111
六、	限制與改進	114
第四章	作品概念與形式討論	117
第一節	《National Freeway 5》：從線性到綿延的堆疊	117
第二節	《The Mirror》：從鏡像到演算法凝視的堆疊	118
第三節	《Self Echoes》：從具身自我到虛擬替身的堆疊	119
第四節	時間、空間、主體意識的平行比較	120

一、	時間層次的流變 .....	120
二、	空間結構的轉換 .....	121
三、	主體意識的生成 .....	122
四、	小結 .....	122
第五章	結論 .....	123
第一節	研究貢獻 .....	123
第二節	符號堆疊下的當代感知狀態 .....	123
第三節	研究限制 .....	124
第四節	符號堆疊之後的感知想像 .....	125
參考文獻	.....	127
附錄一	AI 生成實驗完整提示語 .....	137
一、	原文 .....	137
二、	中文翻譯 .....	138
附錄二	多圖生成結果比較之完整提示語 .....	139
一、	第一組提示語 .....	139
二、	第二組提示語 .....	140
三、	第三組提示語 .....	141
四、	第四組提示語 .....	142
五、	第五組提示語 .....	144

## 圖次

圖 1、日常媒介經驗示意（圖片由本研究使用 Gemini 生成） .....	3
圖 2、《National Freeway 5》（兀兀騰騰寂寂蕩蕩）為錄像作品.....	3
圖 3、《The Mirror》（舉頭三尺有什麼）為互動裝置作品 .....	4
圖 4、《Self Echoes》（我相·人相）為線上互動作品 .....	4
圖 5、維拉斯奎茲，《宮女》，1656（圖片取自[14]） .....	15
圖 6、哈倫·法洛奇，《Eye/Machine》，2000（圖片取自[17]） .....	16
圖 7、文獻與理論時間軸 .....	19
圖 8、杉本博司，《U.A. Play House, New York》，1978（圖片取自[25]） .....	21
圖 9、袁廣鳴，《城市失格—西門町》，2001（圖片取自[26]） .....	21
圖 10、李基宏，《上學路上》（On The Road）作品畫面，2016（圖片取自[27]） .....	23
圖 11、李基宏，《在路上》（On The Road）作品畫面，2020（圖片取自[28]） .....	23
圖 12、白南準，《TV Buddha》，1974（2002）（圖片取自[29]） .....	25
圖 13、丹·格雷厄姆，《Present Continuous Past(s)》，1974（圖片取自[30]） .....	25
圖 14、薩拉文，《The Class of 1988 & The Class of 1967》，1998（圖片取自[31]） ..	27
圖 15、薩拉文，《Every Playboy Centerfold: The Decades》，2002（圖片取自[32]） ..	27
圖 16、陳順築，《男丁》，2005（圖片取自[33]） .....	27
圖 17、扎克·布拉斯，《Facial Weaponization Suite》，2012-2014（圖片取自[34]） ..	29
圖 18、華森人工智慧實驗室，《AI Portraits》範例圖片，2019（圖片取自[35]） .....	29
圖 19、OpenCV 主要功能（圖片擷取自官網[37]） .....	30
圖 20、MediaPipe 主要功能（圖片擷取自官網[39]） .....	32
圖 21、Face Mesh 可在不同角度的臉部影像中生成 3D 臉部網格（圖片取自[40]） ..	33
圖 22、邁布里奇，《The Horse in Motion》，1878（圖片取自[42]） .....	34
圖 23、馬雷，《Pelican in Flight》，約 1882（圖片取自[43]） .....	35
圖 24、合成肖像（composite portraits），1883（圖片取自[52]） .....	38

圖 25、基於常模的人臉辨識模型之二維向量呈現（圖片取自[54]） .....	39
圖 26、Philip J. Benson 與 David I. Perrett 提出的圖像方法（圖片取自[56]） .....	40
圖 27、相關技術年表 .....	52
圖 28、《National Freeway 5》影像流程 .....	55
圖 29、《National Freeway 5》於國立臺北藝術大學展出 .....	62
圖 30、《National Freeway 5》於 106 乙線眺望點進行作品記錄 .....	63
圖 31、《The Mirror》作品外觀 .....	64
圖 32、使用者只能看見自己的反射 .....	65
圖 33、使用者可以看見鏡中影像 .....	65
圖 34、俄羅斯網站《insecam》網頁截圖（圖片取自[87]） .....	66
圖 35、樓梯轉角的廣告方鏡 .....	67
圖 36、《The Mirror》作品外部構造示意 .....	68
圖 37、《The Mirror》作品使用之雙面鏡效果示意 .....	69
圖 38、《The Mirror》作品內部構造示意 .....	70
圖 39、〈Floagent〉中熱鏡與紅外線攝影機系統使用示意圖（圖片取自[88]） .....	71
圖 40、《The Mirror》系統實際執行結果 .....	74
圖 41、《The Mirror》作品互動流程圖 .....	75
圖 42、《The Mirror》於國立臺北藝術大學地下美術館展出 .....	79
圖 43、《The Mirror》於王道銀行藝廊展出 .....	80
圖 44、《The Mirror》於基隆美術館展出 .....	80
圖 45、《The Mirror》會議論文於 IEEE ICCE-TW 2025 發表 .....	81
圖 46、《The Mirror》問卷結果——「情緒反應」部分 .....	82
圖 47、《The Mirror》問卷結果——「知覺認知」部分 .....	83
圖 48、《Self Echoes》之前後端部署框架 .....	91
圖 49、《Self Echoes》使用者介面 .....	92

圖 50、《Self Echoes》系統流程圖 .....	93
圖 51、人臉分析處理流程 .....	96
圖 52、利用 DBSCAN 與 OpenCV 找出最主要的人臉樣本 .....	96
圖 53、摘要內容範例 .....	97
圖 54、圖像生成簡易流程示意 .....	97
圖 55、Draw Things 應用程式 .....	98
圖 56、各種開源模型之多圖生成結果比較（第一、二組） .....	102
圖 57、各種開源模型之多圖生成結果比較（第三、四組） .....	103
圖 58、各種開源模型之多圖生成結果比較（第五組） .....	103
圖 59、各種開源模型之多圖生成結果比較 .....	104
圖 60、《Self Echoes》結果畫面 .....	107
圖 61、《Self Echoes》結果海報 .....	107
圖 62、帳號（@self_echoes_）畫面截圖 .....	108
圖 63、《Self Echoes》作品示意圖 .....	109
圖 64、《Self Echoes》海報論文於 ISAT 2025 進行發表 .....	111
圖 65、使用者問卷「情緒感受」統計圖表 .....	112
圖 66、使用者問卷「滿意度與相似度」統計圖表 .....	112
圖 67、使用者問卷「虛擬替身認同」統計圖表 .....	114
圖 68、圖像預處理不同參數之設定結果 .....	114
圖 69、時間、空間、主體意識發展框架 .....	120

國立臺北藝術大學  
Taipei National University of the Arts

## 表次

表 1、本研究實際使用之主要 OpenCV 函式 .....	31
表 2、MediaPipe Face Mesh 與 Dlib 簡易比較 .....	33
表 3、影像堆疊概念於各時期文獻中的原始用語對照 .....	37
表 4、臉部平均影像與臉部辨識之各時期重要文獻與對照說明 .....	42
表 5、生成對抗網路模型之各時期重要文獻與對照說明 .....	45
表 6、擴散模型之各時期重要文獻與對照說明 .....	46
表 7、文字生成模型 (GPT) 之各時期文獻與對照說明 .....	48
表 8、文字生成模型 (Meta、Anthropic、Google) 之各時期文獻與對照說明 .....	50
表 9、近年可運用之大型語言模型比較 .....	51
表 10、《National Freeway 5》作品畫面 .....	56
表 11、《National Freeway 5》作品實驗過程 .....	59
表 12、《National Freeway 5》畫面品質比較 .....	60
表 13、《National Freeway 5》讀取影格數設定比較 .....	61
表 14、《The Mirror》作品互動流程 .....	73
表 15、《The Mirror》系統處理過程 .....	76
表 16、五階段平均耗時表 .....	94
表 17、各種開源模型之生成結果比較 .....	99
表 18、各種開源模型之多圖生成結果比較 .....	101
表 19、Stable Diffusion 3.5 Large 與使用 LoRA 生成結果比較表 .....	106
表 20、《Self Echoes》網頁截圖 .....	110
表 21、SD 3.5 Large 與 FLUX.1 Krea 模型之多圖生成結果比較 .....	116

國立臺北藝術大學  
Taipei National University of the Arts

# 第一章 緒論

## 第一節 研究動機

時間與空間是我們認知世界的基礎，通過記憶、身體知覺等內在機制構成了每個人獨特的經驗和認知。在今日數位科技高度滲透的時代，時間與空間除了是身體經驗與記憶的線性軸線，也是經由資料、演算法、虛擬環境不斷重組、堆疊與生成的多重場域。網路、社群媒體與人工智慧深刻改變了人們理解自我與世界的方式，我們的日常、身體感知、個人記憶乃至於社會身份，原本清晰可辨的時空、身體與自我邊界，變得越來越模糊。這種「模糊性」不僅體現在現實與虛擬、運動與靜止、主體與客體、過去與現在日益曖昧的界線，也讓個體在多層次時空中經歷身份的不確定。這些過去在藝術與哲學中不斷被追問的主題，也正在虛實交錯的時空裡被重新組構。

而在這股新興技術浪潮中，藝術創作的媒材和形式也一直不斷地迭代、創新，身為新媒體藝術創作者的我們除了關切如何運用這些新興技術外，或許也應該投入關注主體的自我意識如何在這些技術的參與之下發生流動和轉換。簡而言之，我們應該如何擁抱新科技而不迷失其中，是新媒體藝術創作者在當代藝術中一件重要的事情。若從「時間」與「空間」的維度進行討論，這些維度不僅只是一種刻度化的具象呈現（例如可以進行量測的物理空間、在空間變化刻度化下的鐘錶時間），更可以是一種抽象化的概念（例如由意識帶領發展出的空間、身體感知下構築出的時間）。虛與實的對應關係，在新媒體藝術中尤為明顯，顯示器、投影設備的應用，帶來讓我們能同時存在於實體和虛擬空間的感官體驗。然而，這種虛與實的關係並非新生事物，若沿歷史脈絡向前追溯至十六至十八世紀的歐洲，就有藝術家與哲學家開始以鏡子作為媒介開始探討實與虛、可見與不可見的議題。因此可見，時間與空間之間的虛實交錯與模糊性，一直是藝術創作中持久存在且深具意義的存在。無論是近代早期繪畫中鏡像的象

徵意涵，還是當代新媒體藝術中的觀看體驗，創作者不斷地透過這些探索來表達自我、詮釋世界，並擾動觀者對時間和空間的既有認知。

而在日常生活中，我們經常置身於大量且快速的媒介訊號之中，例如光線明暗的瞬間變化、畫面切換、他者目光的短暫停留，以及螢幕上不斷更新的影像與文字。當這些轉瞬即逝的痕跡被記錄、保存並在後續以重疊、累積或重組的方式再次回到我們面前時，原本容易被略過的過程便可能被重新感知，甚至促使我們重新辨認自身在當下的觀看位置與時間經驗，如圖 1 概念意象所示。在這個背景下，本研究以「符號堆疊」指稱此類痕跡被保留並再現的過程，並以三件創作作為本研究的主體，錄像作品《National Freeway 5》（兀兀騰騰寂寂蕩蕩）、互動裝置《The Mirror》（舉頭三尺有什麼），以及線上互動系統《Self Echoes》（我相·人相），分別從身體感知、鏡像監控與數位自我三個面向，實驗當代人在多重時空符號堆疊下的自我意識與感知經驗。

第一件作品《National Freeway 5》（兀兀騰騰寂寂蕩蕩）以向前移動的隧道錄像，疊合出一種介於靜止與運動、真實與記憶之間的時空感知，探討身體如何在現代基礎建設與單調空間中經歷感官遲滯與時間錯置（圖 2）。第二件作品《The Mirror》（舉頭三尺有什麼）則以互動裝置的形式，結合雙面鏡與臉部堆疊技術，透過鏡面反射影像與演算疊像的交錯，表現在被觀看與主動觀看之間注視權力的互換，個體身份、監視權力如何於空間中反覆堆疊與變動（圖 3）。第三件作品《Self Echoes》（我相·人相）以網站形式邀請使用者進行互動，運用演算法擷取社群媒體資料，以生成式 AI 進行資料處理與重組，讓數位自我成為可被模仿、拼貼、甚至再創造的虛擬替身，反映演算法時代社群媒體中身份被資料化後進而產生的碎片化主體（圖 4）。



圖1、日常媒介經驗示意（圖片由本研究使用Gemini生成）

國立臺北藝術大學

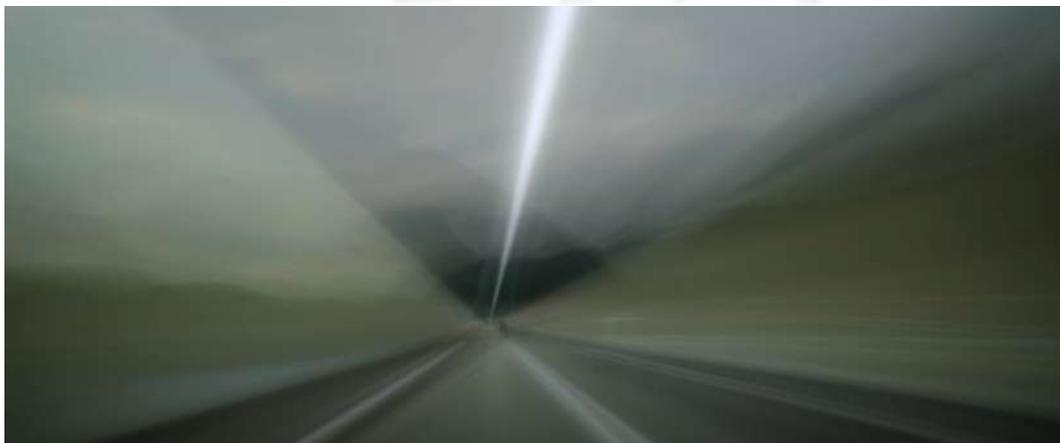


圖2、《National Freeway 5》（兀兀騰騰寂寂蕩蕩）為錄像作品



圖3、《The Mirror》（舉頭三尺有什麼）為互動裝置作品

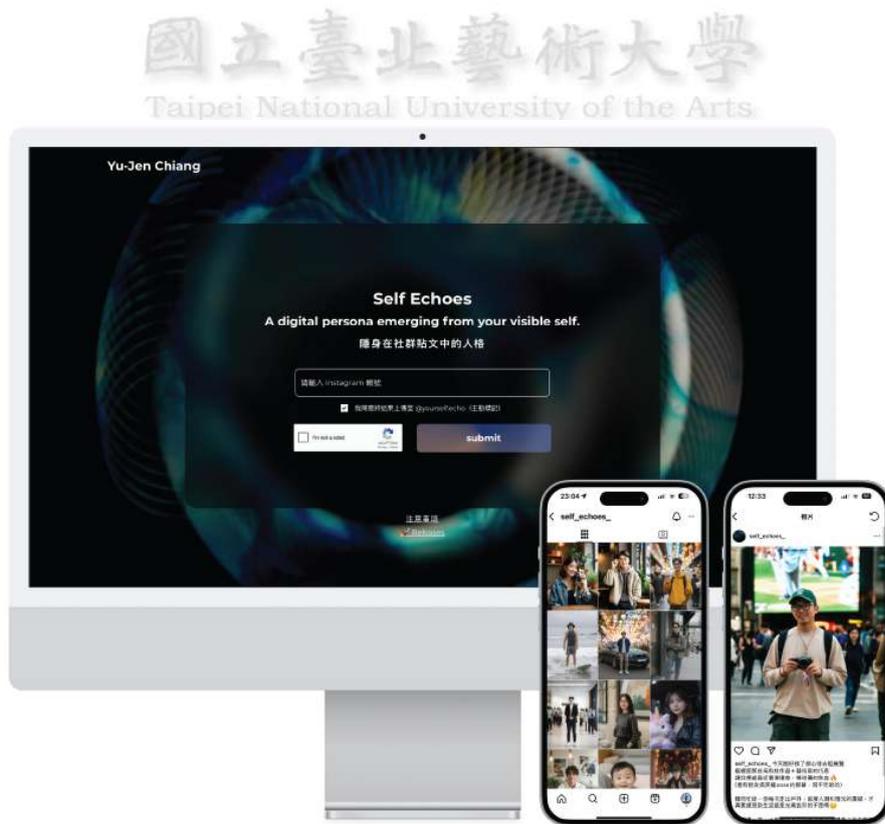


圖4、《Self Echoes》（我相·人相）為線上互動作品

三件作品在表現手法與媒材上各有差異，卻共同指向符號堆疊與多重時空意識的概念，意即我們不再只是活於直覺認知的線性時空中，我們也不再是擁有單一穩定身份的個體，而是在虛與實之間經歷了身體、記憶、身份，於不同層次上被重組、模糊與再現的資訊集合。這種符號堆疊下產生的模糊性不僅是創作技術上的選擇，更是對數位時代自我認知與世界經驗的提問：「當我們的時空經驗、記憶與自我都可以被資料化、堆疊、演算重構時，我們如何在不確定、曖昧、多重詮釋之中，辨識自身與世界的關係，開拓出新的可能？」

## 第二節 研究目的

本篇研究希望透由三件符號堆疊的創作實踐，探討數位時代虛實共構時空意識的形成與表現，具體目標包括：一、梳理二十世紀以來對時間、空間、身體與自我意識的主要哲學理論；二、分析符號堆疊如何在藝術創作中生成抽象空間概念，進一步探討意識空間如何與實體空間產生關聯，及其在虛實交錯下對抽象時空感知的展現方式；三、透過創作實踐，實驗身體、影像與數位資料在新媒體藝術中的堆疊與交錯，並分別以現象學、鏡像理論、資訊哲學等多元觀點，詮釋三件作品如何形塑個體的自我身體、記憶、身份感知與認同。

## 第三節 文章架構

本研究分為五個章節，結構安排如下。第一章：緒論，為研究動機與目的之論述；第二章梳理關於時間、空間、身體、鏡像、自我與資訊哲學的主要理論，作為後續分析基礎，並討論新媒體藝術相關案例，以及與作品實作相關的技術發展脈絡；第三章將分別論述三件作品，《National Freeway 5》（兀兀騰騰寂寂蕩蕩）聚焦身體與空間的

時空交錯；《The Mirror》(舉頭三尺有什麼)探討鏡像、監控與身份疊合；《Self Echoes》(我相·人相)結合 AI 與社群資料，呈現數位自我的生成；第四章將分析三件作品的符號堆疊呈與三件作品對於時間、空間、意識的綜合比較；第五章則將對研究成果進行結論。

國立臺北藝術大學  
Taipei National University of the Arts

## 第二章 文獻探討

正如前章所述，隨著數位科技與演算法日益深度介入藝術創作，虛擬與實體、線性與非線性的時空、個體意識與數位資料之間的界線逐漸鬆動，這種現象促使我們重新思考，探討數位時代虛實共構時空意識的形成與表現。為釐清本研究三件創作所關注的問題脈絡，本章首節將提出相關的觀察角度並梳理其脈絡。接著，第二節將根據作品中身體經驗、鏡像監控、數位自我等不同面向，分別探究現象學、鏡像理論與資訊哲學等學術觀點。而後，透過第三節介紹當代新媒體藝術的相關案例與第四節的技術工具介紹，補充本研究討論的實踐面向。

### 第一節 問題脈絡梳理

#### 一、虛實交錯

「虛實交錯」為通常指涉物理世界與數位世界的相互滲透與重疊，屬於新媒體藝術語境中常見的術語，常透過螢幕、投影或感測裝置，使觀者的經驗同時遊走於實體與虛擬的界線之間。在本研究中，「實體」不應只等同於物理空間，還應包含我們對世界的感知與認識，這樣一來，「虛實交錯」不單純只是被理解作為物理與數位的二分，而是具有時空感知、經驗認識與數位環境的三者交融。

#### 二、時間、空間與意識

時間與空間往往被理解為單一且連續的線性維度，時間是鐘錶式的均勻流逝，空間則是可量測的座標體系。然而，隨著深入身體、意識等非線性時空的討論，加上新媒體如影像技術、通訊技術與演算法的介入，時間可以被切割、加速或延展，也讓空間得以被壓縮、擴充或堆疊。在新媒體藝術中，錄像的蒙太奇手法、網路即時連線的

佈置、互動裝置的感測回饋，皆展現了這種時空重構的現象，使觀者穿梭於多重並置的時間與空間層次之中。與此同時，主體的意識也在這種多重時空條件下發生轉變。自我從原本依附於單一個體中的主體，轉為在不同平台與場域中不斷被拆解、拼貼與再造的身份存在。即時影像、鏡像技術、演算法生成乃至於虛擬替身，成為當代藝術中對這種意識流動的隱喻。自我認同因此呈現出漂移與模糊的特徵，不斷地在虛與實、主體與客體、現實與數位之間被重新定義。

因此，本研究將「時空意識」界定為外在環境與內在主體的重組，一方面，新媒體打破了線性邏輯，使時間與空間展現多重並置的可能性；另一方面，主體意識也在此條件下展現出碎片化與流動性。兩者交織，使得「時空意識」不僅是對外在世界的感知方式，更是自我在多重時空條件下生成與變異的核心場域。

### 三、符號堆疊

「符號」概念源自索緒爾（Ferdinand de Saussure, 1857-1913）的語言學理論[1]，語言是一種基於符號及意義，用於表達觀念的符號系統。所有的溝通都需要藉由使用符號來進行，而每個符號皆由「能指（signifiant）」與「所指（signifié）」構成，能指，是指用於引發人們對於特定對象、事物聯想的表現形式，通常表現為視覺影像、詞語、聲音、手勢、身體動作、物件等等；所指，則是能指所引發的概念與意義。

「堆疊（stack）」一詞，通常是指電腦科學領域中的其中一種抽象資料型別（Abstract Data Type, ADT），代表以後進先出（Last In First Out, LIFO）為原理進行運作的數學模型[2]。常與以先進先出（First In First Out, FIFO）為原理進行運作的「佇列（queue）」相提並論，兩者皆屬於有序的線性資料結構，意即規定了資料「後進先出」或「先進先出」的存放狀態。

本研究所提出之「符號堆疊」，其中「符號」引用語言學理論中一切在溝通中用於引發聯想的各式媒材；「堆疊」一詞雖在電腦科學中是具有明確定義的資料結構，但本研究將此理解為一種將符號進行持續疊加與重組過程的感知隱喻。透過對影像、文本、身體特徵或行為資料等媒材的擷取與演算，將異質性的符號元素在時間序列中相互交疊與共存，進而創造出具層次感的感知經驗。若進一步從電腦科學的運算邏輯來檢視，其運作方式更接近於一種「累積 (accumulate)」、「平均 (average)」或生成式人工智慧中「遞迴 (recursion)」概念的計算模式，將先前所累積的結果作為基礎，使狀態在時間中持續疊加、轉化與演變。

#### 四、模糊性

模糊性一般被理解為界線的鬆動與意義的多重性，既可指涉感知經驗的不確定，也可指涉符號與詮釋的多義性。在新媒體藝術中，虛實交錯、時空重構、意識流動與符號堆疊的交會，往往使得作品難以被單一化地解讀，產生真實與虛擬、運動與靜止、主體與客體之間的曖昧狀態。在本研究中，這種模糊性不僅僅是一種技術效果，也是在多層次交疊下所產生的觀看狀態，為觀者提供多重詮釋與反思的可能。

## 第二節 理論基礎

本節旨在為前述概念建立理論依據。實際上，這些提出的概念是從各個創作實踐中逐步浮現成型，由於作品創作橫跨了許多面向，前一節所提出的核心概念（虛實交錯、時空意識、符號堆疊、模糊性）即是在本研究的創作實踐中彼此交錯的現象。

為了釐清這些概念的理論來源與問題脈絡，本研究將從三個主要取向展開討論，包括現象學與身體感知、鏡像理論與觀看權力、資訊哲學與數位自我。現象學提供理解身體如何在虛實交錯與非線性時空中建立感知經驗的基礎；鏡像理論與權力理論顯

現自我在觀看與被觀看的互動中如何產生流動與模糊；資訊哲學則使我們得以理解自我在符號堆疊與演算法重組中的變異。這三個理論取向雖然分屬不同領域，卻能互相為本研究後續的作品分析提供更完整的詮釋。

## 一、具身主體：身體感知與現象學

### （一） 身體與時間意識的奠基

柏格森 (Henri Bergson, 1859-1941) 在 1889 年的《時間與自由意志》中提出綿延 (duration / la durée) 作為不可切割的時間流。他將時間區分為科學量度的「空間化時間 (spatialized time)」與內心感受的「生命時間 / 綿延 (lived Time / pure duration)」[3]。柏格森以生物學的立場提出「生命時間」是人類深層意識中的連續經驗，具有連續性且不可隨意中斷，過去的意識以及對未來的預先思考都會持續存在。過程中我們可以持續承載著這些意識，醞釀與生成出更多的意義，再開展出新的生命面貌。

胡塞爾 (Edmund Husserl, 1859-1938) 自 1905 年的演講所發展出的《內時間意識現象學》<sup>1</sup> 提出，時間不應該只是鐘錶式物理時間，而是一種內在於意識的結構，「內在時間意識 (internal time-consciousness)」由「當下意識 (intention)」、「回顧意識 (retention)」、「前瞻意識 (protention)」三個面向構成[4]。我們之所以能感受到時間的持續流動，是因為意識同時具有回顧與前瞻的能力，能保留過去，也能期待未來。這樣的內在時間觀，強調了時間經驗的生成與流動。

雖然柏格森與胡塞爾提出的理論分屬不同的背景脈絡，前者立基於直覺與生命哲學，後者則以現象學方法建構意識描述，但兩者都共同挑戰了「時間等同於外在物理刻度」的理解，並強調時間經驗的內在性。梅洛龐蒂 (Maurice Merleau-Ponty, 1908-

---

<sup>1</sup> 胡塞爾有關「時間意識」的第一篇論文發表於 1905 年的演講，並於 1928 年編輯成著作出版。

1961) 則進一步將「時間的內在性」與「身體的在場性」結合。在 1945 年的《知覺現象學》中，他批判笛卡兒式的身心二元論，認為身體不是被動的工具，而是我們與世界互動的根本途徑[5]。透過身體的運動與多感官的協作，我們不僅佔據物理空間，更主動生成「身體空間」。人的身體與所在的世界密切交織，身體的感知與運動並不僅是對外界的反應，也是生成時間與空間經驗的核心。梅洛龐蒂將兩者的時間觀進一步延伸到具身的感知論述之中，讓時間存在於身體的感知與展現的運動之中，而空間也由身體的在場性與意識的持續性共同構築。

## (二) 當代數位危機與具身人文主義

進入二十一世紀後，數位科技普及於生活之中，當代學者持續地在思考過往以身體為基礎的時空意識，在被介面技術與演算法中介時，身體如何持續的經驗世界。2021 年，德國現象學與精神病學家湯瑪斯·福克斯 (Thomas Fuchs, 1958-) 在《In Defense of the Human Being》中接續了具身現象學的基礎，以「具身人文主義 (humanism of embodiment)」批判與反對將心靈與意識簡化為可被數位流程在任何載體上模擬的資訊或演算法，如功能主義 (functionalism)<sup>2</sup>、化約論自然主義 (reductionist naturalism)<sup>3</sup>、大腦中心主義 (cerebrocentrism)<sup>4</sup> 等等，指出此類觀點忽視了身體作為意識、生命力與自我決定能力之基礎，挑戰當代科學、技術和文化中將人類還原為數據與算法的趨勢，並且警告數位化生活將導致人類面臨日益嚴重的「去具身化」[6]。同一時期，

---

<sup>2</sup> 功能主義 (functionalism)，主張心靈狀態是由其在因果關係中所扮演的角色來定義，而不是由其構成的物理材料來定義。這意味著只要一個實體 (無論是血肉之軀的大腦還是電腦) 能扮演相同的功能角色，它就可以被視為擁有「心靈」。

<sup>3</sup> 化約論自然主義 (reductionist naturalism) 結合兩個概念，化約主義 (reductionism)，或譯作簡化論、還原主義，認為複雜的系統可以用其組成部分來解釋；自然主義 (naturalism) 認為科學方法是認識世界最可靠的方式，且無法透過超自然力量來解釋或理解。

<sup>4</sup> 大腦中心主義 (cerebrocentrism) 目前尚無標準中文翻譯，根據其含義譯為「大腦中心主義」，認為大腦是所有心智活動、意識、個性和經驗的唯一或最核心的中心。

美國哲學家尚恩·蓋勒 (Shaun Gallagher, 1948-) 在 2020 年的《Action and Interaction》等著作中指出，儘管面臨去具身化的威脅，身體感知並未消失，而是在數位介面中持續以行動主義 (enactivism) 的方式運作，感知並非對外界資訊的被動接收，而是源於身體與環境的行動與互動，即使身體的感知被技術中介，身體仍然是主體行動與意識生成的場所[7]。

### (三) 社會加速與異化的批判

另一方面，德國社會學家哈特穆特·羅薩 (Hartmut Rosa, 1965-) 分析了技術化社會如何以效率、可控性與加速為核心邏輯，壓縮並均質化生活時間，他將此現象稱之為社會加速 (social acceleration) [8]。他以「共鳴 (resonance)」作為回應現代人被迫以更快節奏的生活，時間變得線性、可計算、可預測，最終失去回應性 (responsiveness)，如此的時間結構也使主體與世界的關係變得空洞與異化 (alienation)，唯有重新建立不可控、具回應性的世界關係，主體才能恢復感知的活力。雖然羅薩並非現象學者，但其對現代時間結構的批判與現象學對內在時間的論述上具有高度的一致性。

### (四) 小結

主體對於時間與空間的感知與意識必須建基於具身經驗，此理論脈絡也為本研究分析《National Freeway 5》(兀兀騰騰寂寂蕩蕩) 提供了對應的框架，作品將國道五號雪山隧道所象徵的線性、高效率與可控性的線性時間，轉化為身體性的、非線性的感知時間，使被壓縮的高速通行經驗重新展現為一種具有回應性的時間流動，從而在現代性的封閉空間中重新喚回身體與世界的關係。

## 二、被觀看的主體：鏡像理論與觀看權力

身體作為感知的原點，自我認同透過觀看他者而形塑，後在進入群體互動的過程，身體也不可避免地存在於社會與環境的凝視之中，而觀看也成為了權力機制投射之下最直接的行為表現。當代，觀看也不再僅只於群體間個體的直接互動行為，更在技術的輔助下，成為更牢固的權力結構，身體則在技術的觀看下被轉譯為可被計算與辨識的符號。這種轉譯的過程，或許可以追溯至人類最早透過鏡子這一中介物來認識自我的歷史。

### （一）人體鏡像與隱喻

鏡子作為物件，自古便是多重時間與空間的橋樑象徵，在《如何觀看世界》，尼古拉斯·米爾佐夫（Nicholas Mirzoeff, 1962-）指出，無論中世紀的歐洲，或是前哥倫布時期的美洲，鏡子最初以黑曜石作為材料，因其特殊性而具有占卜的用途[9]。而鏡像作為一種人類慾望投射與想像的載體，也有著他者視線的隱喻，不僅涉及個體對自我的認同，也牽連社會權力、文化倫理與數位環境中的觀看結構。在個體層次，法國精神分析學家拉岡（Jacques Lacan, 1901-1981）的「鏡像階段」理論首次在1936年提出，他認為人類對自我的原初認識，是約莫在6至18個月大的嬰兒時期，藉著外在的影像認識到自己作為一個整體的存在[10]。然而，這種認知並非對真實身體的直接掌握，其「自我認同」的整合過程為，從認為鏡像為真實，到發現鏡像為虛幻，再到發覺鏡中虛幻的影像就是自己的身體，對鏡中影像產生認同。因此，自我認同自一開始便帶有「他者」的存在，我們對自我的想像與認識，必然依賴於外部影像的映照與回饋。

另一方面，東方傳統中亦存在以鏡子作為倫理隱喻的思維。唐太宗李世民（598-649）曾言：「以銅為鏡，可以正衣冠；以古為鏡，可以知興替；以人為鏡，可以明得失。」此言凸顯了鏡子除了是映照工具，也是作為對歷史與他者修正自我方式的隱喻。

## (二) 觀看權力與規訓視線

在社會層次，傅柯 (Michel Foucault, 1926-1984) 在《詞與物》中分析了維拉斯奎茲 (Diego Velazquez, 1599-1660) 的作品《宮女》(The Maids of Honour)，如圖 5 所示，其構圖透過畫面中間的鏡子混合了空間的可見與不可見[11]。畫家本人於畫面中的視點與我們視線相連接，而鏡中又映照出國王皇后的畫像，使主體與客體不停地相互交換角色，讓國王皇后、畫家、觀者三者共享畫中角色的目光，讓觀看本身成為一種權力機制。另外，傅柯也在 1975 年出版的《規訓與懲罰》中以環形監獄(panopticon) 作為現代監控社會的隱喻，並深化出「全景敞視主義 (panopticism)」，當個體身處在隨時可能被看見的情境下，他們將會不斷地進行自我內化監視並加以自我規訓[12]。觀看因此成為一種治理技術，滲透於身體與日常之中。同樣與觀看權力有關，約翰·伯格 (John Berger, 1926-2017) 在《觀看的方式》中則批判女性如何在父權社會的觀看下被形塑為「景觀」，成為男性慾望的投射對象，被固定在被觀看的客體位置。「鏡子的真正功能是讓女性成為共犯，和男人一樣，首先把自己當成一種景觀」，鏡子此時意味著一種自我審視，女性看鏡子更多的是看別人如何看自己[13]。

從這個討論來看，自我意識的生成始終處於觀看與被觀看的動態過程，觀看與被觀看也一直都與社會的權力有著深刻地連結，從嬰兒期的心理認同，到社會化的自我規訓，觀看不只是個體內在的心理經驗，也是社會權力結構下自我持續被塑造的過程。然而在當代處處為技術所中介的條件下，若人類不再是最主要進行觀看的主體，其運作邏輯或許也將會有相當大的差異。



圖5、維拉斯奎茲，《宮女》，1656（圖片取自[14]）

國立臺北藝術大學  
Taipei National University of the Arts

### （三） 技術中介下的觀看

美國藝術史學家強納森·柯拉瑞（Jonathan Crary, 1951-）在1990年出版《觀察者的技術》中提到，現代的觀看已非全然是以肉眼觀察光線反射的自然成像，縱然此模式會繼續存在，但其視覺機能將會被整合進各類光學儀器、電腦設備的構造中，並且開始被影像製造技術所支配。至此，觀看將成為一種被技術與體制所共同形塑的行為，而非純粹的感官經驗[15]。隨著電腦視覺的出現與普及，進行觀看的主體似乎除了人體，也可能是正在進行運算的機器。「操作用影像（operational image）」的概念由德國電影導演哈倫·法洛奇（Harun Farocki, 1944-2014）在2000年提出的作品《Eye/Machine》中出現（圖6），用於指稱一種當代大量存在，卻並非用於提供人類進行審美或認知，而是為了讓機器進行運算、分析、控制而存在，應用於軍事、工業等自動化機械儀器中用於執行任務的功能性影像，而此類影像已然屬於一種「機器的

視覺」[16]。隨後的許多學者，例如特雷弗·佩格藍（Trevor Paglen, 1974-），將法洛奇的觀點視為理解整個機器視覺體制的基礎，2016年提出「不可見的影像」的概念，並且也在2019年正式發表的《Invisible Images: Your Pictures Are Looking at You》提到，許多當代的數位影像並非主要被人觀看，而是被演算法用於分類、比對與控制[18]。

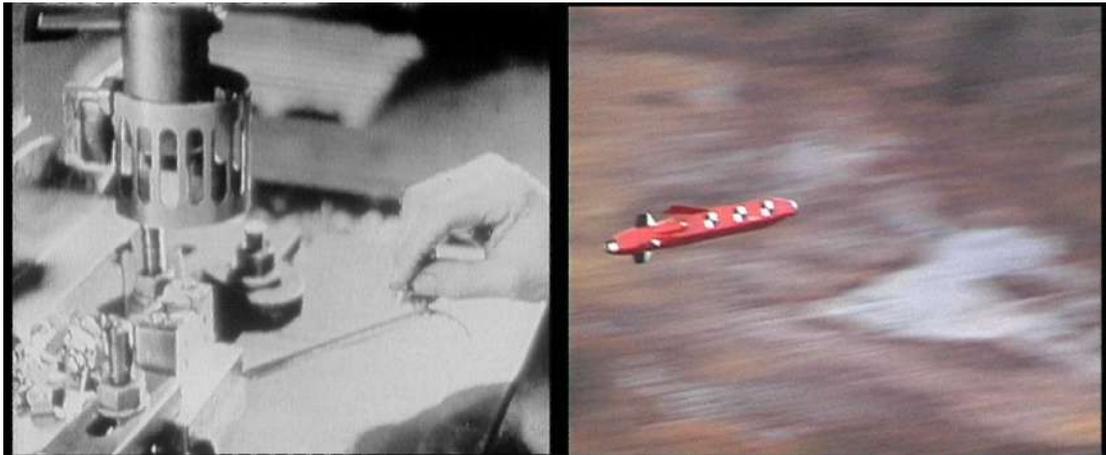


圖6、哈倫·法洛奇，《Eye/Machine》，2000（圖片取自[17]）

#### （四） 小結

上述理論的發展脈絡為《The Mirror》（舉頭三尺有什麼）提供了相當程度的論述依據，作品從人類作為被觀看主體的經驗開始，藉以鏡子這一現實生活中常見且本身即具有自我映照與他者視線隱喻的物件，引導使用者透過互動思考觀看機制中的權力關係。當鏡中的影像系統將拍攝到的人臉影像進行運算並產出結果時，觀看本身也將從個體層次的自我認同與社會層次的自我規訓中，進一步轉變為一種在機器觀看條件之下，身體在系統中被拆解為由數據座標與特徵點構成的可計算符號，如作品所用的人臉網格技術。

### 三、碎片化主體：資訊哲學與數位自我

在前兩部分的討論中可以知道，主體的時空意識一方面源於身體的感知與行動，另一方面也在他者與社會環境的視線中持續被塑造。然而，若機器視覺使「觀看」成為人與技術共同的行為，那麼當「主體」也同時以身體與演算法、資料結構等技術條件共同構成新的存在方式時，自我如何在資訊環境中被表述、再現？

#### （一）社群媒體中的自我揭露與敘事危機

當代社群媒體中的個人檔案、照片與動態更新，已成為塑造理想自我的鏡像實踐。2012年，韓裔德國哲學家韓炳哲（Byung-Chul Han, 1959-）在《透明社會》中指出，當代的監控模式已由傅柯筆下的他律監控，轉為透過自我揭露而運作的自律監控[19]，人們主動在社群媒體上分享生活細節，將部分自我展現於他人的觀看之下，這種經過美化形象自我展示成為在數位空間中的鏡像，同時也持續接受來自他者的「喜愛」與「留言」等回饋，讓自我在他者的視線下持續修正，形成一種動態且可計算的主體。2023年，韓炳哲進一步在《敘事的危機》中延續批判，在資訊統治的條件下，面臨了敘事危機（crisis of narration），主體日益被扁平化為離散而破碎的資料點，自我游移於演算法分類的各種片段之間，喪失了時間的連續性與深度，難以形成一個具有連貫的生命敘事。

#### （二）多重身份的碎片化

對於多重身份與多重自我的呈現，或許可以追溯至1959年美國社會學家厄文·高夫曼（Erving Goffman, 1922-1982）提出的「前台／後台」的隱喻。在《日常生活中的自我呈現》中，他指出在日常生活中，人們於「前台」展演符合社會期待的角色，而在「後台」則展現未必為公眾所見的面貌，自我即是一種透過與社會互動而持續被建構的表演性產物[20, 21]。數位環境則延伸了這一結構，使前台呈現更加持續、可編

輯、可計算，使自我表演不再僅止於社會情境之中。1995年，美國社會學家雪莉·特克（Sherry Turkle, 1948-）在《虛擬化身》中指出，人在不同平台上建立多重身份，甚至同時經營多個自我[22, 23]。她認為自我的流動性與碎片化讓一個人可能同時在社群平台展現專業形象、私人情感、甚至匿名的遊戲身份，這些身份既分裂又未必能整合，數位媒介不僅改變了我們與世界互動的方式，也深刻重塑了我們理解自我的方式。

### （三） 資訊模式與具身虛擬性

然而，如果這些平台上碎裂的多重身份都構成主體的一部分，那麼主體如何在資訊條件下維持其時空意識？在本節的第一部分中曾提過湯瑪斯·福克斯批判了當代科學、技術和文化中將人類還原為數據與算法的觀點。然而向前推至1999年，美國後現代文學評論家凱瑟琳·海爾斯（N. Katherine Hayles, 1943-）即在《後人類時代》中追溯了控制論後人類主義（cybernetic posthumanism）時代以來形成的「去身體化後人類主體」觀念。此一觀點認為，自我本質上是一種可被抽離、運算，並且能在不同介質之間轉譯的資訊模式，而身體僅是資訊的暫時容器，這些論述最終導向了「意識可上傳」、「身體可被替代」等去物質化想像。然而，海爾斯批判「資訊可以脫離物質而存在」掩蓋了身體作為認知與經驗之基礎的不可替代性。她主張的「具身虛擬性（embodied virtuality）」，即是在強調資訊永遠必須在某種物質媒介中具體化，因此無法與身體完全分離。真正的主體性，是由身體的限制條件（具身性）與資訊模式的可塑性（虛擬性）共同構成[24]。

### （四） 小結

在此意義下，資訊性自我並不完全取代具身自我，而是一種在虛實共構的環境中，以演算法邏輯為基礎所生成的第二種自我形式。它與具身自我並存，並在不同場域中運作，具身自我展現生命的時間綿延與敘事深度，而資料化主體則體現資訊模式下的可塑性。此一脈絡也將為《Self Echoes》（我相·人相）作品提供分析基礎，透過使用

者因身體而生的貼文與影像，蒐集語言與視覺風格，運算生成一個數位替身，旨在探討這個數位替身如何作為身體之外的另一種資料化主體，我們所面對的究竟是不是「我」（時間軸整理如圖7）。

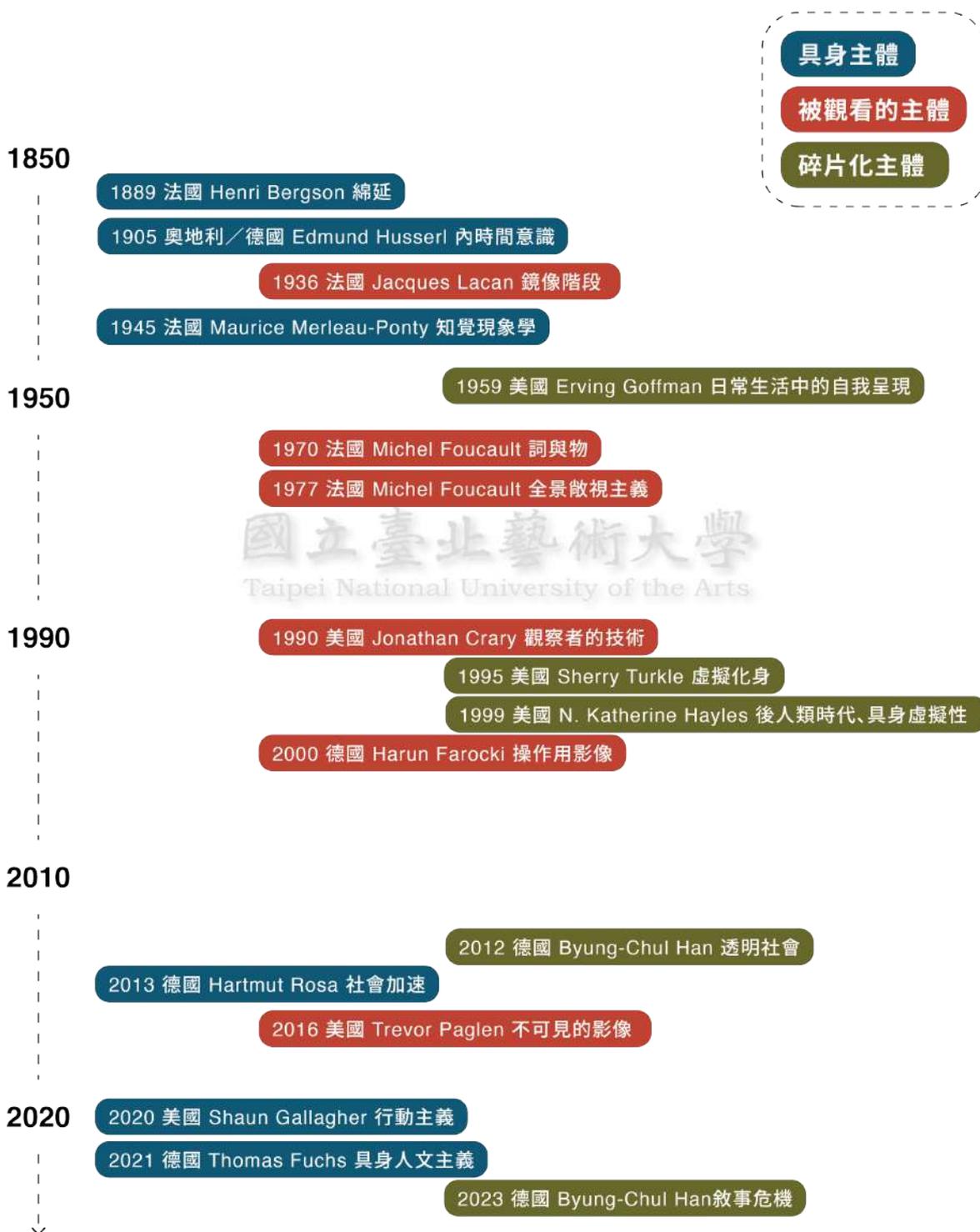


圖7、文獻與理論時間軸

### 第三節 藝術案例分析

新媒體的發展使得物理現實和數位虛擬之間形成了多重交錯的場域，這些場域又在在地影響了我們在不同層次的時間、空間和自我的認識，也為身體與意識帶來了新的存在方式。因此，本節將透過一些對當代藝術案例的分析，觀察不同創作者如何以錄像、裝置或攝影等形式，使虛實之間的模糊與流動具體化。

#### 一、時間與空間中的身體流動

##### (一) 杉本博司《Theaters》

日本攝影師杉本博司（Hiroshi Sugimoto, 1948-）的早期系列作品《Theaters》透過攝影展現了他對時間和空間的思考，圖 8 為其系列中的一件作品。他將攝影中通常用來凍結瞬間、切割時間的快門，轉化為一種儲存整個時間過程的載具。在這些拍攝了劇院和電影院大銀幕的作品中，他的快門在電影開始時開啟，結束時關閉，曝光時間是一部電影的全長，在整個播放過程中，流動的影像不斷地曝光在底片上，直至電影結束，最終在畫面中的銀幕上留下的是一片空白光場。我們在這些作品中看見的空白銀幕並非虛無，而實際上是整部電影的畫面總和，將所有流動的影像納進抽象的時間容器。透過這種攝影手法，探索了時間的流逝和空間的模糊性，使觀者在觀看時意識到時間本身作為影像材料的存在，並且體驗到時間和空間交織所構成的獨特詮釋。



圖8、杉本博司，《U.A. Play House, New York》，1978（圖片取自[25]）



圖9、袁廣鳴，《城市失格—西門町》，2001（圖片取自[26]）

## (二) 袁廣鳴《城市失格—西門町》

圖 9 為袁廣鳴 (1965-) 的 2001 年的作品《城市失格—西門町》，在前後兩個月的時間，於西門町的相同定點，將拍攝出來的三百多張照片用電腦合成進行拆解重組，重新拼湊出一個無人無車的城市影像。「城市失格」系列的表現手法讓時間在此被濃縮，三百多次時間的定格瞬間，被壓縮進單一畫面中，在現實中重構出另一個不曾在某個時間點上存在，卻又依附在實體空間上的意識空間。當我們進入的這個空間之後，重新喚起了記憶中的連續經驗，對現實空間（西門町）的回顧意識與知覺會持續與虛擬空間（城市失格）相互連結產生對話，尤其在我們經歷 2021 年新冠疫情三級警戒期間過後，對於城市失格提供的虛擬空間，對照當時因為防疫措施造成的無人外出的現實空間景觀，多重時空的回顧意識相互對應，或將綿延出更多的詮釋空間。

## (三) 李基宏《上學路上》與《在路上》

李基宏 (1972-) 於 2016 年的錄像作品《上學路上》(圖 10)，記錄他從家裡送小孩走到學校共計一個學期的影像；2020 年的社宅公共藝術作品《在路上》(圖 11) 則是將紀錄了 31 組居民以家門為起(終)點的行走軌跡重疊，將現實空間影像堆疊，生成出一個存在堆疊之中的意識空間。他日積月累的記錄了因身體運動而創造出的空間變化，讓身體成為了度量時間的標準，在將自身經歷的空間重疊在一起的同時，也重疊了那個當下的內在時間意識，創造出一個不存在且具有模糊性，依內在時間意識交疊流動的抽象空間，空間中的每個角度看起來都是模糊不清的，但根據每個觀者在過去的不同空間經驗與時間記憶，又可從中意識到自我經驗與他者經驗的交錯。



圖10、李基宏，《上學路上》(On The Road) 作品畫面，2016 (圖片取自[27])



圖11、李基宏，《在路上》(On The Road) 作品畫面，2020 (圖片取自[28])

## 二、觀看與被觀看的權力關係

### (一) 白南準《TV Buddha》

韓裔美國藝術家白南準(Nam June Paik, 1932-2006)1974年的作品《TV Buddha》為一件將佛像與錄影機、電視機相對放置的裝置作品，佛像實體與佛像在螢幕上的影像同時存在，兩者互為參照(如圖12所示)，佛像凝視著螢幕裡自己即時被攝錄的影像，形成一個不斷循環的迴路。此時，現實與虛擬的時空被連動，佛像既是觀看的主體，同時又是被觀看的客體，現實存在的佛像既凝視了虛擬的影像，又被虛擬的影像凝視。而，身為第三方的觀者也被捲入整個觀看系統之中，從而意識到虛與實、主體與客體、宗教凝視與科技再現，以及觀看本身的權力結構。

### (二) 丹·格雷厄姆《Present Continuous Past(s)》

美國藝術家丹·格雷厄姆(Dan Graham, 1942-2022)1974年的作品《Present Continuous Past(s)》也是一件由鏡子、攝影機與延遲錄影系統組成的空間裝置。在圖13中可見，觀眾進入空間後，會同時看到鏡中映照的自己，以及經由攝影機捕捉、八秒鐘延遲後出現在螢幕上的影像，讓觀眾能立刻感受到虛實影像交錯的體驗，真實身體與其影像並存於同一場域，而時間上不同步。當下與8秒前、16秒、24秒前等過去的影像並置，現在與剛過去的瞬間互相干擾，時間被分層與堆疊，而不是線性流動。觀者在其中既是當下的主體，也是螢幕上剛過去的客體；既是觀看者，也是被觀看者。這種自我與影像的交錯，使作品與觀者的自我意識持續在動態的時空中流動。



圖12、白南準，《TV Buddha》，1974（2002）（圖片取自[29]）



圖13、丹·格雷厄姆，《Present Continuous Past(s)》，1974（圖片取自[30]）

### (三) 傑森·薩拉文《Amalgamations》

美國藝術家傑森·薩拉文(Jason Salavon, 1970-)1998年開始的《Amalgamations》系列是以程式自動化影像疊合技術，將多張照片平均化合成「集合肖像」的系列影像作品。其中，《The Class of 1988 & The Class of 1967》(圖14)分別取自藝術家本人與其母親在德州沃斯堡(Fort Worth, Texas)的畢業紀念冊照片，經由平均運算合成兩個世代的「班級集體臉孔」。於2002年創作的《Every Playboy Centerfold: The Decades》(圖15)則是他將1953年至1999年間《花花公子》(Playboy)雜誌中，每一期的「Centerfold」(中頁女郎照片，為正中央以3頁連在一起的摺頁特寫)進行數位疊合與平均化處理，生成五張代表不同年代的「平均女性」影像。這些不再呈現具體個體又模糊、近乎抽象的圖像，除了是以現實空間堆疊出抽象空間的手法，也是使個體身體在數據堆疊下轉化為群體意識的象徵。

### (四) 陳順築《男丁》

圖16為陳順築(1963-2014)2005年的作品《男丁》，他利用一張父親的照片，重疊他與哥哥在三個不同階段中的各一張照片，創造出三個不同也不在現實存在的父子、兄弟肖像。這些肖像符號將自己視為客體存在，在結構上創造了一個因為他者的凝視而存在虛擬空間，在具有模糊性的影像中觀看自己在不同時期與父親、兄長的面貌，除了將自我與家族記憶投射於影像中，也在呈現上以符號堆疊作為實踐的方法。



圖14、薩拉文，《The Class of 1988 & The Class of 1967》，1998（圖片取自[31]）

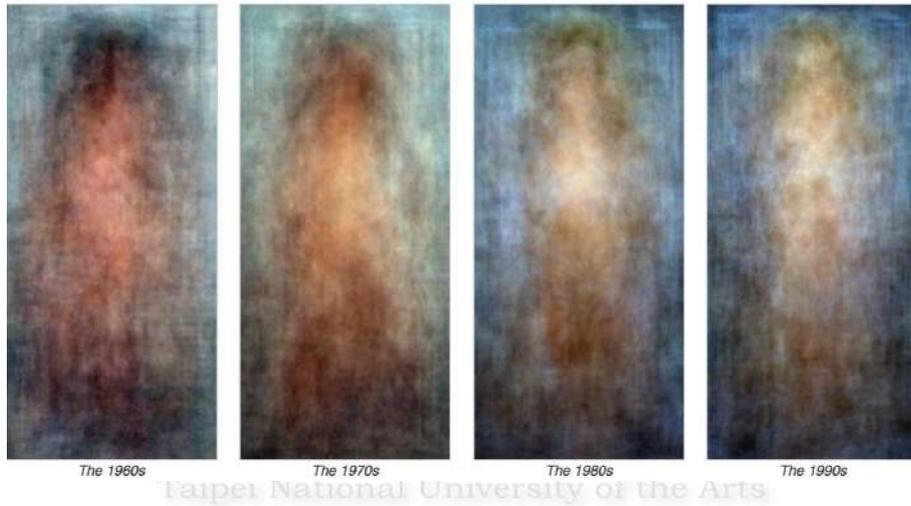


圖15、薩拉文，《Every Playboy Centerfold: The Decades》，2002（圖片取自[32]）



圖16、陳順築，《男丁》，2005（圖片取自[33]）

### 三、演算法與虛擬身份的生成

#### (一) 扎克·布拉斯《Facial Weaponization Suite》

美國新媒體藝術家扎克·布拉斯 (Zach Blas, 1981-) 於 2011 至 2014 年間進行的《Facial Weaponization Suite》則將臉部資料的堆疊轉化為一種批判行動。藝術家邀請參與者提供臉部影像資料，透過電腦演算將多張臉孔平均化與疊合，最終生成一個「集體臉孔」，如圖 17。這些演算結果不再是一個清晰的肖像，取而代之的是一個抽象的形體，之後再被製作成各種單一顏色的面具，供參與者在錄像或公共遊行中佩戴。這些面具的臉孔資料原先來自真實時空中的人臉，在進行虛擬化後堆疊生成出形狀扭曲、難以被人臉辨識系統正確判讀的影像，從而意圖作為對監控技術的抵抗。此作品不僅是對數位監控社會的批判，也在身份與臉部資料的堆疊中，在數位環境中生成具有政治性的自我意識。

國立臺北藝術大學

#### (二) 華森人工智慧實驗室《AI Portraits》

由 IBM 與美國 MIT 麻省理工學院聯手成立的華森人工智慧實驗室 (MIT-IBM Watson AI Lab) 於 2019 年推出一個線上專案《AI Portraits》，讓使用者可以透過網頁在上傳照片之後，獲得一張由 AI 資料庫生成的古典肖像，如圖 18，古典肖像與當代自拍文化因此被串連在一起，使用者在跨越歷史的圖像轉換中因而體驗到過去與現在的混合。



Taipei National University of the Arts

圖17、扎克·布拉斯，《Facial Weaponization Suite》，2012-2014（圖片取自[34]）



圖18、華森人工智慧實驗室，《AI Portraits》範例圖片，2019（圖片取自[35]）

## 第四節 相關技術探討

### 一、 OpenCV Library

OpenCV (Open Source Computer Vision Library) 是由 Gary Bradski 於 1999 年在 Intel 開發，並於 2000 年公開發行的一套開源跨平台電腦視覺函式庫，可以在商業與研究領域自由使用[36]。它提供豐富的影像處理與分析工具，涵蓋即時影像處理、物件辨識、特徵提取與運動追蹤等任務，如圖 19 所示。由於其開源特性，OpenCV 已成為電腦視覺開發中最為廣泛採用的工具之一，也被廣泛應用於電腦視覺、車載影像、智慧監控等領域。

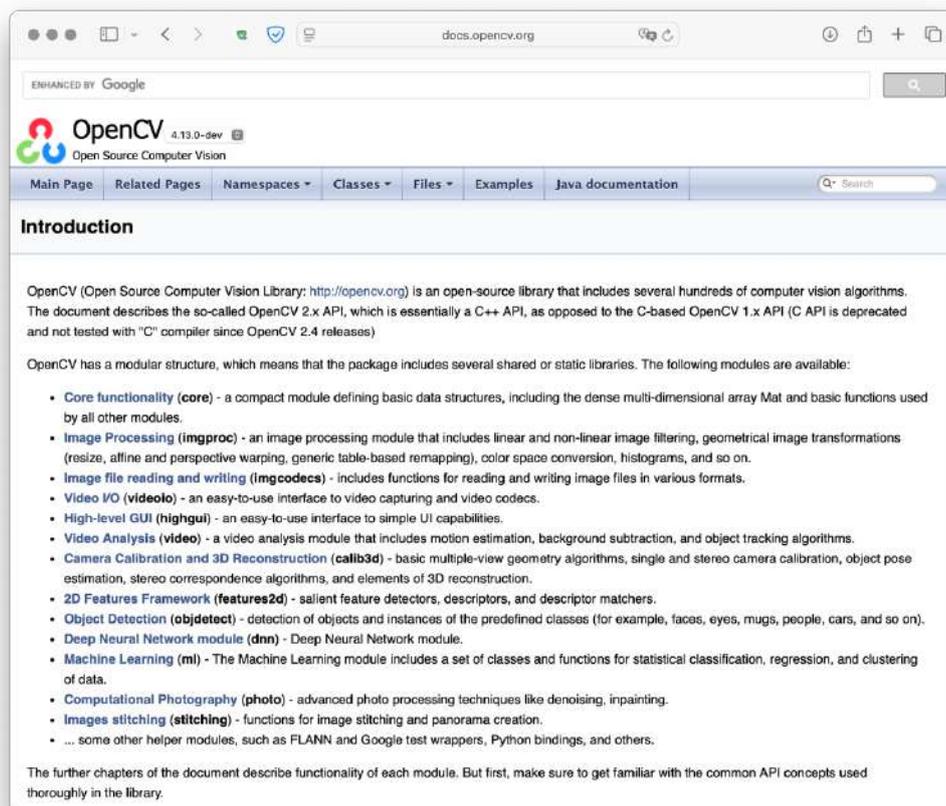


圖19、OpenCV主要功能（圖片擷取自官網[37]）

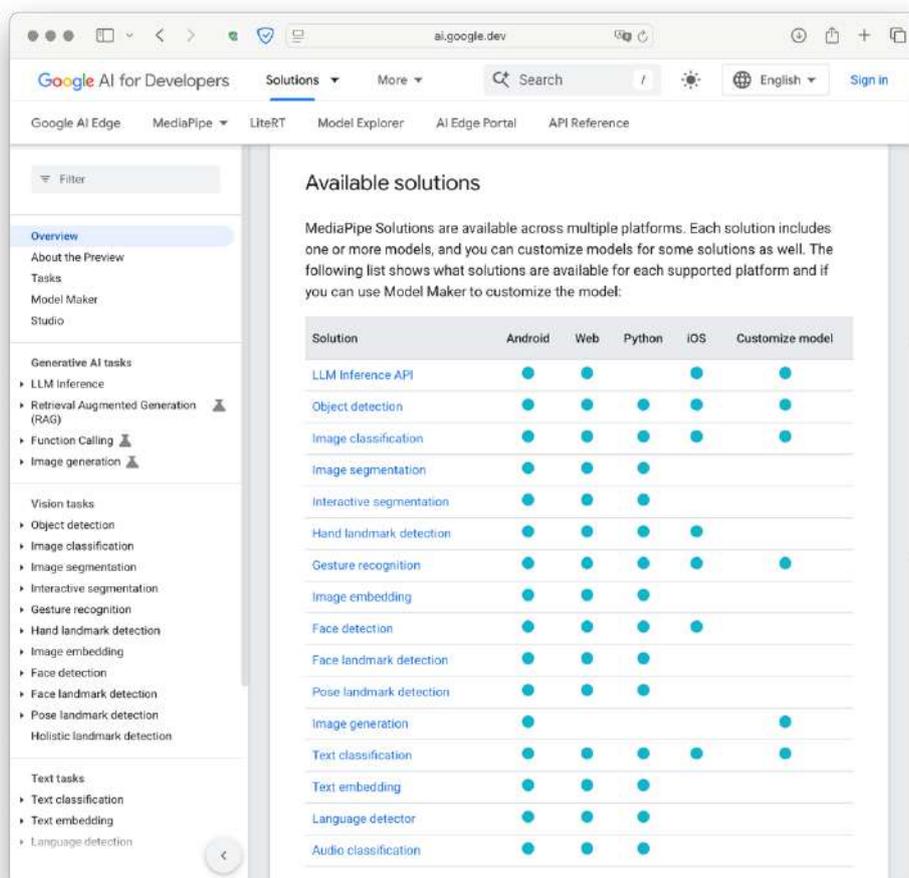
在本研究中，OpenCV 主要用於影像輸入與輸出、影像運算與處理以及影像繪製與標註等作業，表 1 整理了本研究所提出三件作品實際使用之主要函式與其功能說明，例如使用 cv2.VideoCapture 函式可讀取影片或即時攝影機畫面，搭配 cv2.VideoWriter 函式可以輸出經處理後的影像；而透過 cv2.addWeighted 函式可以進行多個影像的加權平均，藉由設定透明度（ $\alpha$  值）控制各個影像間的合成比例，產生堆疊與混合的視覺效果。

表1、本研究實際使用之主要OpenCV函式

類別	函式名稱	主要功能
影像輸入與輸出	cv2.VideoCapture()	讀取影片與攝影機影像輸入
	cv2.VideoWriter()	儲存影片結果
	cv2.imread()	讀取靜態圖片
	cv2.imwrite()	儲存圖片
影像運算與處理	cv2.addWeighted()	進行影像疊合加權與透明度控制
	cv2.resize()	調整影像尺寸
	cv2.cvtColor()	色彩空間轉換
	cv2.Laplacian()	清晰度運算分析
影像繪製與標註	cv2.rectangle()	標註臉部區域
	cv2.putText()	標示文字
	cv2.imshow()	顯示影像視窗

## 二、MediaPipe Solutions

MediaPipe 是一個由 Google 於 2019 年提出的開源跨平台機器學習架構[38, 39]，專為處理複雜的視覺辨識和動態追蹤任務而設計，尤其適合應用於即時互動系統與多媒體創作中。MediaPipe 以模組化的方式提供了多樣的解決方案 (solutions)，如人臉偵測、手勢辨識、人體骨架偵測與動態追蹤等等，如圖 20 所示。MediaPipe 整合了機器學習模型與即時視覺處理技術，可透過與 OpenCV 等其他工具的結合，實現更進階的影像處理與更豐富的互動功能，因此也被廣泛應用於電腦視覺研究、行為識別、動態捕捉及新媒體藝術創作等領域。



Solution	Android	Web	Python	iOS	Customize model
LLM Inference API	●	●		●	●
Object detection	●	●	●	●	●
Image classification	●	●	●	●	●
Image segmentation	●	●	●		
Interactive segmentation	●	●	●		
Hand landmark detection	●	●	●	●	
Gesture recognition	●	●	●	●	●
Image embedding	●	●	●		
Face detection	●	●	●	●	
Face landmark detection	●	●	●		
Pose landmark detection	●	●	●		
Image generation	●				●
Text classification	●	●	●	●	●
Text embedding	●	●	●		
Language detector	●	●	●		
Audio classification	●	●	●		

圖 20、MediaPipe 主要功能 (圖片擷取自官網[39])

在本研究所提出的《The Mirror》中，即是使用 Face Mesh Solution<sup>5</sup>進行人臉偵測與關鍵點追蹤[40]，能在單一或即時串流的影像中，精準捕捉使用者臉部的 468 個特徵關鍵點，並且使用機器學習推斷 3D 臉部網格，如圖 21 所示。相較於傳統的人臉偵測方法（如 Dlib 等），Face Mesh 擁有更高的效能與穩定性，能以較低的運算成本實現較準確且有效率的臉部追蹤功能，兩者之簡易比較請見表 2。

表2、MediaPipe Face Mesh與Dlib簡易比較

模型	關鍵點	網格	運算速度
Dlib	68	2D	較慢
MediaPipe Face Mesh	468	3D	較快

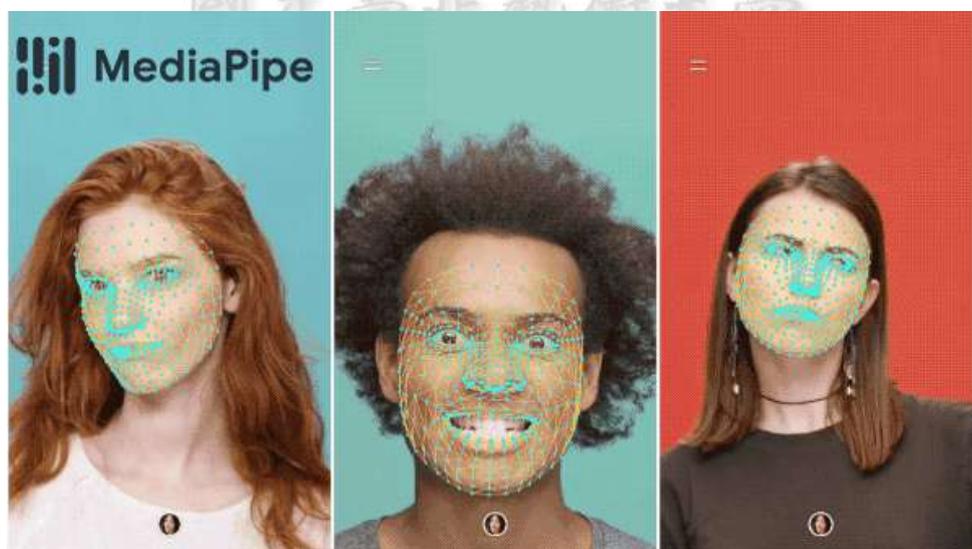


圖21、Face Mesh可在不同角度的臉部影像中生成3D臉部網格（圖片取自[40]）

<sup>5</sup> Google 於 2023 年 5 月對 MediaPipe 進行改版，原先的 Face Mesh 模組已更新為 Face Landmark Detection，本研究使用之 Face Mesh 因而被歸入 MediaPipe Legacy Solutions 之中。

### 三、影像堆疊 (frame stacking)

影像堆疊 (frame stacking) 是一種將多張影像或連續影格依照時間序列排列，透過堆疊使時間 (temporal) 與運動 (motion) 資訊在單一影像中被視覺化的運算方法。其概念源頭可向前追溯至十九世紀的運動影像實驗，英國攝影師邁布里奇 (Eadweard Muybridge, 1830-1904) 使用多部相機拍攝運動中的動物[41]，如《The Horse in motion》(圖 22)，這些分格式影像形成了時間的切片，並藉助放映機還原連續性的時間。而後法國生理學家馬雷 (Étienne-Jules Marey, 1830-1904) 發明了馬雷攝影槍 (Marey's chronophotographic gun)，將一個運動物體的分離影像重複感光在一張底片上，用以傳達運動的連續關係，如《Pelican in Flight》(圖 23)，而這種技術被稱為「時間攝影 (chronophotography)」。



圖22、邁布里奇，《The Horse in Motion》，1878 (圖片取自[42])

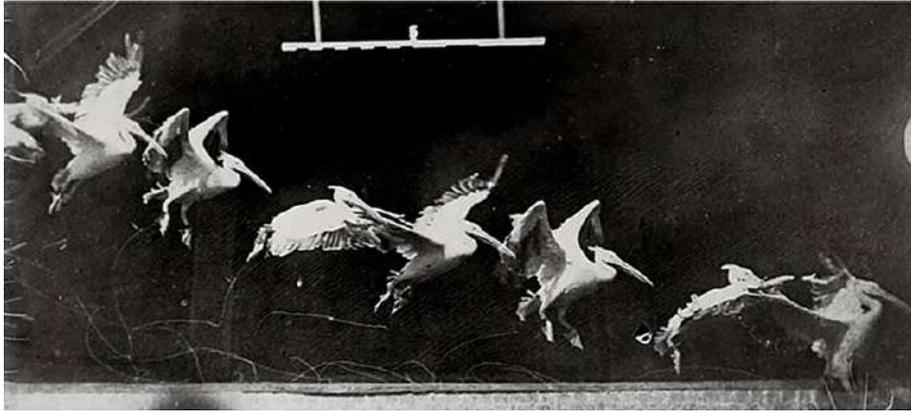


圖23、馬雷，《Pelican in Flight》，約1882（圖片取自[43]）

1963年任職於美國 Itek Corporation<sup>6</sup>的 R. J. Kohler 與 Hutson K. Howell 發表了〈Photographic Image Enhancement by Superimposition of Multiple Images〉[44]，此技術報告在多處文獻中以相同卷期與頁碼被引用，由此可知，「光學重疊(superimposition of multiple images)」作為一種可以提升航太與天文影像品質的方法，成為進入數位影像處理之前，技術討論上的早期技術來源。隨著電腦與數位影像處理的發展，美國羅森菲爾德 (Azriel Rosenfeld, 1931-2004) 教授在 1969 年於《Picture Processing by Computer》中提到了一種電腦運算概念，稱作「多份副本的平均(averaging of multiple copies)」，將光學的概念轉化為數位影像處理中可以降低雜訊的一項基礎電腦演算法[45]。而後在 1970 至 1990 年代間，影像堆疊技術常被用於去除影像雜訊與長曝光模擬。1977 年 Rafael C. Gonzalez 與 Paul. A. Wintz 在《Digital Image Processing》<sup>7</sup>中將

---

<sup>6</sup> Itek Corporation 是一家美國國防承包商，專門從事間諜衛星的相機系統及其他各種軍事偵察系統[105]。Kohler 與 Howell 任職於 Itek 的線索可從其發表的文件得知。《Photographic Considerations for Aerospace》有三種版本。根據多處文獻比對，1961 年版本為編號 IL-9026-8 的技術報告[106]，作者為四位 Itek 工程師 (Brock,G.C., Harvey,D. I., Kohler,R.J., Myskowski,E.P.)；1965 年版本為正式出版第一版[107]、1966 年版本為增訂第二版[108]，由 Hall, H. J.與 Howell, H. K.擔任編者，四位工程師則為作者，並由此得知 H. K. Howell 全名為 Hutson Koger Howell (1924-2002)。

<sup>7</sup> 2002 年版為目前可取得之最早版本，書中已明確使用「Image Averaging」一詞。由於該書延續自 Gonzalez 與 Wintz 於 1977 與 1987 年出版的早期版本，因此可推知此概念最早應可追溯至 1977 年

「影像平均 (image averaging)」視為影像增強的基礎方法之一[46]；1989 年美國的 Kevin M. Kavanaugh 等人則在醫學影像分析中提出「視訊影格平均 (Video Frame Averaging)」的應用，以連續影格平均降低雜訊，提升影像品質[47]；而任職於美國松下先進電視與影像實驗室 (Panasonic Advanced TV-Video Laboratories, Inc.) 的 Jill M. Boyce 也在 1992 年中發表「自適應運動補償影格平均 (adaptive motion compensated frame averaging)」，在影像處理與電視技術領域中，以運動補償演算法來精確對齊連續影格，實現更高品質的視訊降噪[48]。之後，影像堆疊的概念從降低雜訊擴展到資訊融合，美國的 Paul E. Debevec 與 Jitendra Malik 在 1997 年提出了高動態範圍 (High Dynamic Range, HDR) 的原理，將堆疊概念應用於不同曝光影像的融合，重建出具有更多光度資訊的影像[49]。

進入二十一世紀後，影像堆疊的概念被延伸至人工智慧與深度學習領域，影像堆疊可以被延伸為一種狀態表示方法。英國的 Volodymyr Mnih 等人於 2015 年發表，將「影格堆疊作為輸入 (frame stacking as input)」運算概念引入深度強化學習 (DQN) 模型[50]，原文指出：「The function ... applies this preprocessing to the m most recent frames and stacks them to produce the input ...」，透過數個連續影格的堆疊為神經網路的輸入，能夠有效地將時間 (temporal) 與運動 (motion) 資訊轉換為神經網路可處理學習的空間特徵 (spatial)，以提升模型對動態情境的理解能力。表 3 整理各時期文獻中提及「影像堆疊」概念的應用與技術脈絡，以及文獻中原始用語的對照，可助於理解此概念如何從光學實驗、數位影像處理，逐步延伸成為當代人工智慧中的時間性資料表現方法。

---

的版本。《Digital Image Processing》最初作者為 Gonzalez 與 Wintz (1977、1987)，其後由 Gonzalez 與 Woods 接續撰寫，並於 1992、2002、2018 年間多次改版。

表3、影像堆疊概念於各時期文獻中的原始用語對照

時期	原始用語	文獻	作者	應用與技術脈絡
1880	chronophotography	《Pelican in Flight》	Étienne-Jules Marey	將分離影像重複感光於單一底片以呈現運動連續性。
1963	superimposition of multiple images	〈Photographic Image Enhancement by Superimposition of Multiple Images〉	R. J. Kohler、Hutson K. Howell	以多重曝光重疊影像降低顆粒雜訊。
1969	averaging of multiple copies	〈Processing by Computer〉	Azriel Rosenfeld	將多張同場景影像堆疊平均以降低雜訊。
1977	image averaging	《Digital Image Processing》	Rafael C. Gonzalez、Paul. A. Wintz	系統性整理影像處理方法，包括影像平均。
1989	video frame averaging	〈Effects of Video Frame Averaging, ...〉	Kevin M. Kavanaugh <i>et al.</i>	影格平均應用於醫學影像降噪
1992	adaptive motion compensated frame averaging	〈Noise reduction of image sequences using adaptive motion compensated frame averaging〉	Jill M. Boyce	以運動補償對齊幀序以提升視訊降噪。
1997	high dynamic range	〈Recovering High Dynamic Range Radiance Maps from Photographs〉	Paul E. Debevec、Jitendra Malik	跨曝光堆疊的資訊融合。
2015	...stacks them to produce the input to ...	〈Human-level control through deep reinforcement learning〉	Volodymyr Mnih <i>et al.</i>	將幾個最近影格堆疊作為輸入，學習時序資訊。

#### 四、臉部平均影像與臉部辨識

臉部平均影像的生成邏輯延續了前一節所述的「影像堆疊」概念，皆以多重影像的重疊與平均作為運算核心。若前者著重於時間序列的堆疊，臉部平均影像則將此概念轉化為空間上的整合，透過多張臉部影像的對齊與平均，使群體特徵在單一影像中被視覺化。其技術起源可追溯至十九世紀的人類學研究，英國科學家高爾頓（Francis Galton, 1822-1911）於 1878 年發表「合成肖像（composite portraits）」的實驗，透過多重曝光將不同個體的面孔重疊感光於同一底片，以呈現所謂的「典型面貌」[51]。這種以光學重疊達成平均結果的技術實驗，也奠定了後續臉部平均技術的觀念基礎。

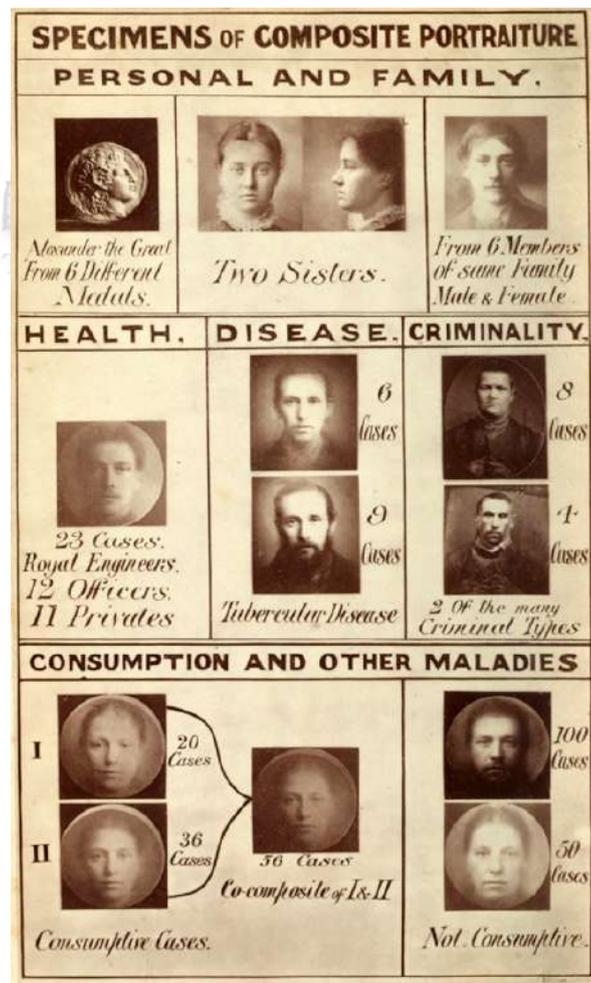


圖24、合成肖像（composite portraits），1883（圖片取自[52]）

二十世紀末，心理學與視覺認知研究延續了此一脈絡。美國的 Judith H. Langlois 與 Lori A. Roggman 於 1990 年在〈Attractive Faces Are Only Average〉一文中指出，人們傾向認為經平均化處理的臉孔是更具吸引力的，臉部平均影像因此與「典型性 (prototypicality)」擁有密切的關聯[53]。英國的 Tim Valentine 則於 1991 年提出一個「人臉多維空間框架 (multidimensional space framework)」<sup>8</sup>來解釋人腦如何識別臉孔 [54]，而臉部平均即是其框架下其中一個抽象模型的空間原點 (如圖 25)，所有個體的臉部特徵皆以此作為距離基準。

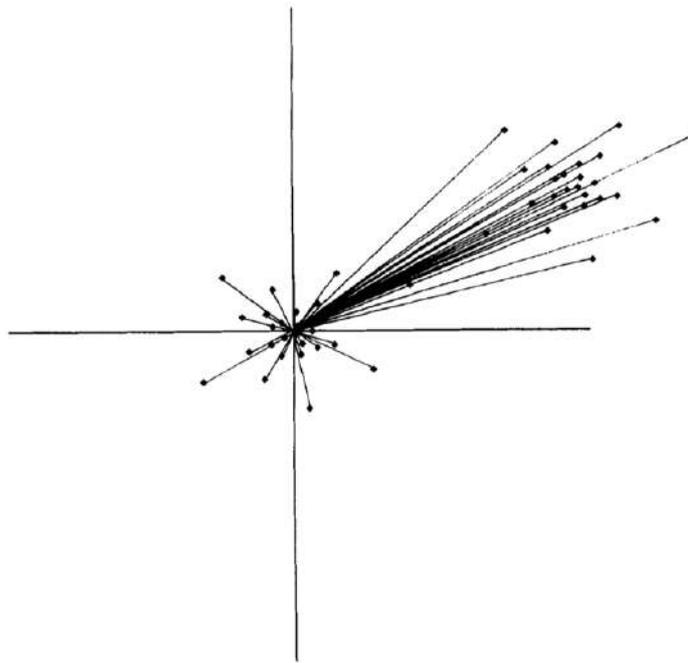


圖25、基於常模的人臉辨識模型之二維向量呈現 (圖片取自[54])

同一年，在電腦視覺領域中，美國 MIT Media Lab 的 Matthew Turk 與 Alex Pentland 也提出了「電腦如何識別臉孔」的「特徵臉演算法 (eigenfaces for recognition)」，建

---

<sup>8</sup> 這個框架下有兩種模型：Norm-based model：臉以相對於常模 (Norm) 或原型 (Prototype) 的向量表示；Exemplar-based model：臉以樣本形式儲存，透過相似度比較進行辨識，沒有抽象的常模。論文中的實驗無法區分兩種模型的優劣，且框架中也透過結合「距離常模的遠近」和「範例密度」來解釋所有現象。

立了一個由「特徵臉 (eigenfaces)」組成的低維度子空間 (face space)，用來描述臉部圖像的變化，其中成為臉部辨識核心基準的訓練樣本平均向量，即是「平均臉 (average face)」 [55]。1993 年，英國的 Philip J. Benson 與 David I. Perrett 提出一種電腦圖像方法，將高爾頓「合成肖像」的技術結果以電腦運算實現並改良 [56]。他們將同一個人的多張照片數位化後手動標上 186 個點，再透過形變與座標平均來大幅降低對齊時產生的模糊現象，因而能自動運算出臉部平均影像，即「臉部原型 (facial prototype)」。他們以多張 Margaret Thatcher 的照片為範例，如圖 26 所示，圖上放兩列為六張原始影像，第三列 (a) 為直接疊合知結果；(b) 為高爾頓技術之數位化處理結果；(c) 為經過形變與座標平均處理後之結果。



圖26、Philip J. Benson與David I. Perrett提出的圖像方法（圖片取自[56]）

此後二十年間，臉部平均影像的研究開始結合不同新興技術與方法，從統計分析，逐漸發展至 3D 模型、深度學習等。德國的 Volker Blanz 與 Thomas Vetter 於 1999 年發表的研究將臉部識別與合成帶入 3D 領域，並以高精度統計模型驗證了臉部特徵在多維參數空間中進行運算的可行性，也成功地將心理學上的抽象概念，轉化為在參數空間中的具體向量操作[57]。而隨著深度學習與電腦視覺的進步，臉部特徵在的技術處理工具也更為精確。包括前面曾提到由 Google 在 2019 年推出的 MediaPipe[38]，就是利用機器學習來推斷深度資訊，讓人臉關鍵點 (facial landmark) 與 3D 網格 (face mesh) 可以即時被偵測。2020 年土耳其的 Sefik Serengil 與 Alper Özpınar 開發的 DeepFace 封裝框架[58-60]，以多種深度模型 (face recognition models) 作為辨識的運算後端<sup>9</sup>，使臉部特徵能以多維向量 (multi-dimensional vectors) 在統一的空間 (embedding space) 中進行比對，此框架也封裝多種偵測與對齊工具 (face detection and alignment)<sup>10</sup>，讓研究者能在相同介面下測試、比較多種深度臉部辨識模型與偵測工具，作為實務應用與創作實驗的基礎工具。

國立臺北藝術大學  
Taipei National University of the Arts

表 4 為整理與臉部平均影像、臉部辨識相關之各時期文獻，可助於理解從光學實驗、2D 像素數位影像處理到 3D 模型、深度學習的技術脈絡。在本研究提出的作品中，除了《The Mirror》使用 MediaPipe 的 Face Mesh Solution 之外，在《Self Echoes》也使用 DeepFace 模型進行人臉影像偵測與辨識等相關的預處理。

---

<sup>9</sup> VGG-Face、FaceNet、OpenFace、DeepFace、DeepID、ArcFace、Dlib、SFace、GhostFaceNet 與 Buffalo\_L。

<sup>10</sup> OpenCV、Dlib、MtCnn、Faster MtCnn、RetinaFace、MediaPipe、Yolo、YuNet 與 CenterFace。

表4、臉部平均影像與臉部辨識之各時期重要文獻與對照說明

時期	類別	文獻	作者	說明
1878	光學合成	〈Composite Portraits, Made by ...〉	Francis Galton	以多重曝光重疊不同個體臉部的「合成肖像」。
1990	心理研究	〈Attractive Faces Are Only Average〉	Judith Langlois、Lori Roggman	實證經平均處理臉孔更具有吸引力。
1991	認知模型	〈A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition〉	Tim Valentine	提出「多維臉部空間模型」，以平均臉作為辨識基準。
1991	電腦視覺	〈Eigenfaces for Recognition〉	Matthew Turk、Alex Pentland	建立臉部特徵子空間，將平均臉作為臉部辨識的平均向量基準。
1993	電腦圖像處理	〈Extracting Prototypical Facial Images from Exemplars〉	Philip J. Benson、David I. Perrett	以電腦生成經過形變與座標平均處理後之「臉部原型」。
1999	3D統計模型	〈A Morphable Model for the Synthesis of 3D Faces〉	Volker Blanz、Thomas Vetter	以多張臉部資料建立模型，奠定臉部對齊與生成技術基礎。
2019	深度學習模型	MediaPipe Face Mesh	Google	以機器學習推斷臉部深度資訊，即時偵測468個關鍵點並建立3D網格。
2020	開源框架整合	DeepFace	Sefik I. Serengil、Alper Özpınar	Python開源框架，整合多種辨識模型與偵測工具，提供統一模組化的人臉辨識流程。

## 五、生成式人工智慧 (Generative AI)

人工智慧 (Artificial Intelligence, AI) 是指用電腦程式來模擬和呈現人類智慧的技術總稱。隨著硬體運算能力的升級, AI 的發展逐漸聚焦於機器學習 (Machine Learning, ML), 一種讓電腦從資料中自動分析獲得規律, 並利用規律對未知資料進行預測的演算法。其中, 深度學習 (Deep Learning, DL) 則是機器學習的一個方法路徑, 核心在於多層人工神經網路的建構, 能有效處理大規模影像、語音與文本等複雜數據。生成式人工智慧 (Generative Artificial Intelligence, Generative AI) 則屬於深度學習的應用分支, 除了分類或預測, 還能夠根據輸入的資料自動生成文字、影像、聲音或其他多媒體內容的人工智慧技術。目前主流的生成式人工智慧模型可分為「文字生成文字 (text-to-text)」與「文字生成影像 (text-to-image)」兩大類。除此之外, 尚包含「影像生成影像 (image-to-image)」、「文字生成語音 (text-to-speech)」、「文字生成影片 (text-to-video)」等多元應用。

國立臺北藝術大學  
Taipei National University of the Arts

在影像生成領域, 從 2014 年到 2024 年十年間, 生成式人工智慧的主要技術從「生成對抗網路 (GAN)」轉移至「擴散模型 (diffusion model)」, 並且發展出許多可應用的模型工具。生成對抗網路 (GAN) 最早由美國資訊科學家 Ian Goodfellow 等人於 2014 年所發表的〈Generative Adversarial Nets〉中提出[61], 其架構包含生成器 (generator) 與鑑別器 (discriminator) 兩個相互對抗的神經網路, 透過對抗式學習逐步提升生成影像的真實度, 此後衍生出的 DCGAN、StyleGAN 等, 分別強化了影像細節與人臉合成的品質。2015 年, 美國的 Alec Radford 將卷積網路 (CNN) 結合 GAN, 提出〈Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks〉, 大幅的提升了生成影像品質, 也奠定了影像基礎架構[62]。2018 年任職於美國輝達研究團隊 (NVIDIA Research) 的 Tero Karras 等人在〈Progressive Growing of GANs for Improved Quality, Stability, and Variation〉提出 ProGAN 模型與「漸進式增長」訓練方法[63], 從低解析度逐步擴展網路層數以生成  $1024 \times 1024$  的影

像，並且於 2019 年在〈A Style-Based Generator Architecture for Generative Adversarial Networks〉發表了 StyleGAN 模型[64]，加入「風格嵌入 (style-based)」架構，能精確控制臉部表情與紋理，可用於生成高解析度臉部表情的影像，其系列模型自發表至 2020 年間被廣泛應用，文獻整理如表 5。

以 GAN 為主的生成模型雖能生成逼真影像卻缺乏語意控制。近年來，擴散模型 (diffusion model) 在影像品質和生成多樣性上逐漸超越 GAN，並且成為主流的生成架構，此類模型文獻整理於表 6 中。Jonathan Ho 等人於 2020 年發表〈Denoising Diffusion Probabilistic Models〉，提出「去噪擴散概率模型 (Denoising Diffusion Probabilistic Model, DDPM)」[65]，透過模擬影像逐步去除雜訊的反向過程，以機率方式生成高品質影像，這一模型的穩定性與生成品質遠超過 GAN，被視為生成模型領域的重大里程碑。2021 年 OpenAI 推出 DALL·E 與 CLIP Text Encoder [66]，將文字描述嵌入影像生成流程，成為了文生圖的基礎架構。2022 年，DALL·E 2 以擴散模型提升解析度與語意對應準度。同年 7 月，Midjourney 以雲端服務形式推出，強調藝術風格化構圖，廣受創作者社群討論。OpenAI 則於 2023 年 10 月推出後續版本 DALL·E 3，並與 GPT 進行整合，能直接在 ChatGPT 介面中生成影像[67]。德國 CompVis<sup>11</sup> 聯合創始人 Robin Rombach 等人也於 2022 年在〈High-Resolution Image Synthesis with Latent Diffusion Models〉提出潛在擴散模型 (LDMs)，將擴散過程轉移至潛在空間 (latent space) 進行[68]，這個空間可以被用來進行數據壓縮、生成新數據，或在不同數據類型之間建立關聯，例如，透過整合 CLIP Text Encoder 作為文字嵌入條件，可以將圖像和文字映射到同一個潛在空間，而因此能使生成內容在語意與構圖上被精確控制，同時也大幅降低運算成本。Latent Diffusion 之後在同年 8 月與英國 Stability AI 以及美國 Runway AI 合作發展為 Stable Diffusion 1.1-1.4，並於 10 月發表 Stable Diffusion

---

<sup>11</sup> Computer Vision and Learning research group at Ludwig Maximilian University of Munich，慕尼黑大學電腦視覺與學習研究小組（原海德堡大學電腦視覺小組）

1.5，其開源性促使全球社群迅速產生大量衍生模組，如 LoRA (low-rank adaptation)、ControlNet 等，成為圖像生成式 AI 領域中最廣為人知的工具之一。而後 RunwayML 及 CompVis 相繼退出，部分原開發成員隨後在德國成立 Black Forest Labs，於 2024 年推出 FLUX.1[69]，在光影、材質與真實感表現上更接近照片級水準。而 Stability AI 則持續發展出 2023 年發表的 Stable Diffusion XL (SDXL)[70]與 2024 年發表的 Stable Diffusion 3 以及 3.5[71]，也更進一步提升語意一致性與影像真實感，在本研究提出的《Self Echoes》中即是使用 Stable Diffusion3.5 模型進行圖像生成。

表5、生成對抗網路模型之各時期重要文獻與對照說明

年份	模型	文獻	作者 / 單位	說明
2014	GANs	〈Generative Adversarial Nets〉	Ian Goodfellow <i>et al.</i>	首次提出 GAN 框架，以生成器、鑑別器進行對抗學習。
2015	DCGAN	〈Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks〉	Alec Radford <i>et al.</i>	將 CNN 與 GAN 結合，提升生成穩定性與品質。
2018	ProGAN	〈Progressive Growing of GANs for Improved Quality, Stability, and Variation〉	Tero Karras <i>et al.</i> (NVIDIA)	漸進式增長訓練，生成 1024×1024 影像，為 StyleGAN 的基礎。
2019	StyleGAN	〈A Style-Based Generator Architecture for Generative Adversarial Networks〉	Tero Karras <i>et al.</i> (NVIDIA)	為人臉生成的里程碑，引入風格嵌入，能精確控制臉部表情與紋理等等。

表6、擴散模型之各時期重要文獻與對照說明

年份	模型	文獻	作者 / 單位	說明
2020	DDPM	〈Denoising Diffusion Probabilistic Models〉	Jonathan Ho <i>et al.</i>	去噪式擴散生成，高品質且多樣性佳。
2021 / 1	DALL·E	〈Zero-Shot Text-to-Image Generation〉	Aditya Ramesh <i>et al.</i> (OpenAI)	DALL·E 初版研究論文，描述文字至影像生成概念與 CLIP 結構。
2022 / 4	DALL·E 2	〈Hierarchical Text-Conditional Image Generation with CLIP Latents〉	Aditya Ramesh <i>et al.</i> (OpenAI)	DALL·E 2 的正式論文，說明 CLIP latent 與 Diffusion 架構。
2022 / 7	LDM	〈High-Resolution Image Synthesis with Latent Diffusion Models〉	Robin Rombach <i>et al.</i>	Stable Diffusion 的基礎研究，在潛在空間進行擴散過程，大幅降低運算成本
2022 / 7	Midjourney	—	Midjourney Inc.	雲端影像生成平台，以風格與構圖美感著稱。
2022 / 8	SD 1.1-1.4	—	Stability AI, Runway ML, CompVis	採 LDM 架構的開源模型，開啟社群生態。
2022 / 10	SD 1.5	—	Stability AI, Runway ML, CompVis	改良版本，為最廣泛使用之開源模型。
2023 / 7	SD XL	〈SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis〉	Dustin Podell <i>et al.</i> (Stability AI)	支援 1024×1024 解析度，畫質明顯提升。

年份	模型	文獻	作者 / 單位	說明
2023 / 10	DALL·E 3	〈Improving Image Generation with Better Captions〉	James Betker <i>et al.</i> (OpenAI)	與 GPT-4 整合，可於 ChatGPT 介面中生成影像。
2024 / 2	SD 3	〈Scaling Rectified Flow Transformers for High-Resolution Image Synthesis〉	Patrick Esser <i>et al.</i> (Stability AI)	新一代擴散架構，提升語意一致性與運算效率。
2024 / 4	FLUX.1	—	Black Forest Labs	強調光影與材質真實感。
2024 / 6	SD 3.5	—	Stability AI	改良版本，進一步提升生成速度與畫面一致性。

在文字生成領域，文字生成文字 (Text-to-Text) 模型又稱大型語言模型 (Large Language Models, LLMs)，文字生成技術的核心在於「自回歸式語言模型 (Autoregressive Language Model)」，其原理是根據先前的文字序列預測下一個字元或詞彙。2017 年，任職於 Google 的 Ashish Vaswani 等人提出〈Attention Is All You Need〉[72]，以「Transformer」架構取代傳統的循環神經網路 (RNN)，大幅提升長上下文語意理解與推理能力，成為後續生成模型的基礎。

OpenAI 也在 Alec Radford 加入後，於 2018 年發表 GPT-1 論文〈Improving Language Understanding by Generative Pre-Training〉[73]，提出「預訓練 (pre-training)」與「微調 (fine-tuning)」的雙階段訓練流程。其後 2019 年推出的 GPT-2 擁有 15 億參數[74]；2020 年推出的 GPT-3 更具有 1,750 億參數，使自然語言生成達到人類可讀等級[75]。隨著硬體與資料規模的進步，此類模型也已不斷迭代，進一步朝向多模態化發展 (Multimodal LLMs)。2023 年發表的 GPT-4 支援影像與文字輸入，並整合於 ChatGPT

之中[76]。2024 年，GPT-4o 提升了推理、即時語音與長文本處理能力[77]。此外，若要執行更複雜的任務，也能夠串接 API 選擇特定模型進行開發，例如執行指令方面更加精準的 GPT-4.1[78, 79]。相關文獻整理如表 7。

表7、文字生成模型（GPT）之各時期文獻與對照說明

時期	模型	文獻	作者 / 單位	說明
2018 / 6	GPT-1	〈Improving Language Understanding by Generative Pre-Training〉	Alec Radford <i>et al.</i> (OpenAI)	首採預訓練與微調架構，約 1.17 億參數。
2019 / 2	GPT-2	〈Language Models Are Unsupervised Multitask Learners〉	Alec Radford <i>et al.</i> (OpenAI)	15 億參數，擁有長文本生成與多任務能力。
2020 / 6	GPT-3	〈Language Models Are Few-Shot Learners〉	Tom B. Brown <i>et al.</i> (OpenAI)	1,750 億參數，首次達到人類等級語言生成。
2023 / 3	GPT-4	〈GPT-4 Technical Report〉	Josh Achiam <i>et al.</i> (OpenAI)	參數量未知 <sup>12</sup> ，支援文字與影像輸入，整合於 ChatGPT。
2024 / 5	GPT-4o	〈GPT-4o system card〉	Aaron Hurst <i>et al.</i> (OpenAI)	原生多模態模型，可即時處理文字、影像與語音。
2025 / 4	GPT-4.1	〈Introducing GPT-4.1 in the API〉	OpenAI	API 版本，強化指令執行與長文本推理。

<sup>12</sup> 自 GPT-4 後，官方文件就沒有列出實際參數量，經推估其參數量約為 1.76 兆。

除 OpenAI 外，其他主要系統也都自 2023 年起快速發展，使文字生成的能力不僅限於語言，還可跨越多種資料型態，推動了生成式人工智慧的發展，Meta 於 2023 至 2024 年間以開源形式推出 LLaMA 與其迭代的系列模型[80-82]；Anthropic 在同年推出 Claude 系列模型，並且在 2024 年已迭代至 Claude 3.5，強調了長文本生成、倫理與安全性約束的能力[83]；Google Gemini Team 也在 2023 年 12 月推出了 Gemini 系列模型，支援文字、圖像、影片、音訊等多模態輸入與推理[84]，並在 2024 至 2025 年間推出的 Gemini 1.5、2.0、2.5，迭代強化了長上下文記憶與影片理解能力[85, 86]，相關文獻整理如表 8。

在開發選擇上，Meta 的 LLaMA 模型為開源形式，然而此模型需要部署於本機中，對運算資源要求較高，因此使用上擁有較高門檻，而其他模型則多數提供付費 API 供開發者開發使用，模型之間也有各自的優勢與限制，表 9 為近年可開發使用之大型語言模型比較。在本研究提出的《Self Echoes》中即是使用 Open AI 的 GPT4.1 模型進行文字生成。（此節相關技術年表整理如圖 27）

表8、文字生成模型（Meta、Anthropic、Google）之各時期文獻與對照說明

時期	模型	文獻	作者 / 單位	說明
2023 / 3	LLaMA	〈Llama: Open and efficient foundation language models〉	Hugo Touvron <i>et al.</i> (Meta)	開源模型，含 70-650 億參數四版本。
2023 / 7	LLaMA 2	〈Llama 2: Open foundation and fine-tuned chat models〉	Hugo Touvron <i>et al.</i> (Meta)	開源模型，含 70-700 億參數三版本。
2023 / 12	Gemini	〈Gemini: A Family of Highly Capable Multimodal Models〉	Google Gemini Team	原生多模態模型，整合文字、影像、影片與音訊輸入。
2024 / 3	Claude 3	〈The Claude 3 Model Family: Opus, Sonnet, Haiku〉	Anthropic	強化精度、速度與即時性，支援長文本與邏輯推理。
2024 / 3	Gemini 1.5	〈Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context〉	Google Gemini Team	強化長上下文記憶與影片理解能力。
2024 / 4	LLaMA 3	〈The Llama 3 Herd of Models〉	Aaron Grattafiori <i>et al.</i> (Meta)	改進推理品質，含 80、700 億參數二版本。
2024 / 6	Claude 3.5 Sonnet	〈Claude 3.5 Sonnet Model Card Addendum〉	Anthropic	增強程式生成與多語言能力。
2024 / 9	Gemini 2.0 / 2.5	〈Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities〉	Gheorghe Comanici <i>et al.</i> (Google)	多步推理，表現提升，支援影片分析。

表9、近年可運用之大型語言模型比較

年份	模型名稱	開發單位	優勢	限制	說明
2025	GPT-4.1 系列 (API)	OpenAI	支援文字與影像輸入，長文推理能力強。	上下文與多模態功能依方案不同。	開發者導向的後端模型，本研究採用之主要工具。
2024	Gemini 2.5 系列 (API)	Google	原生多模態，支援影片與多步推理。	生成錯誤碼機率仍高，穩定性待改進。	強化程式與影片處理，低延遲。
2024	Claude 3.5 系列 (API)	Anthropic	語意一致性高，適合長篇敘事。	不支援影像輸入，風格保守。	注重倫理與安全性，長文本結構穩定。
2024	LLaMA 3	Meta	開源可自由微調，適合學術研究與實驗。	需部署於本機，運算資源門檻高。	適用於學術研究與自建生成實驗。

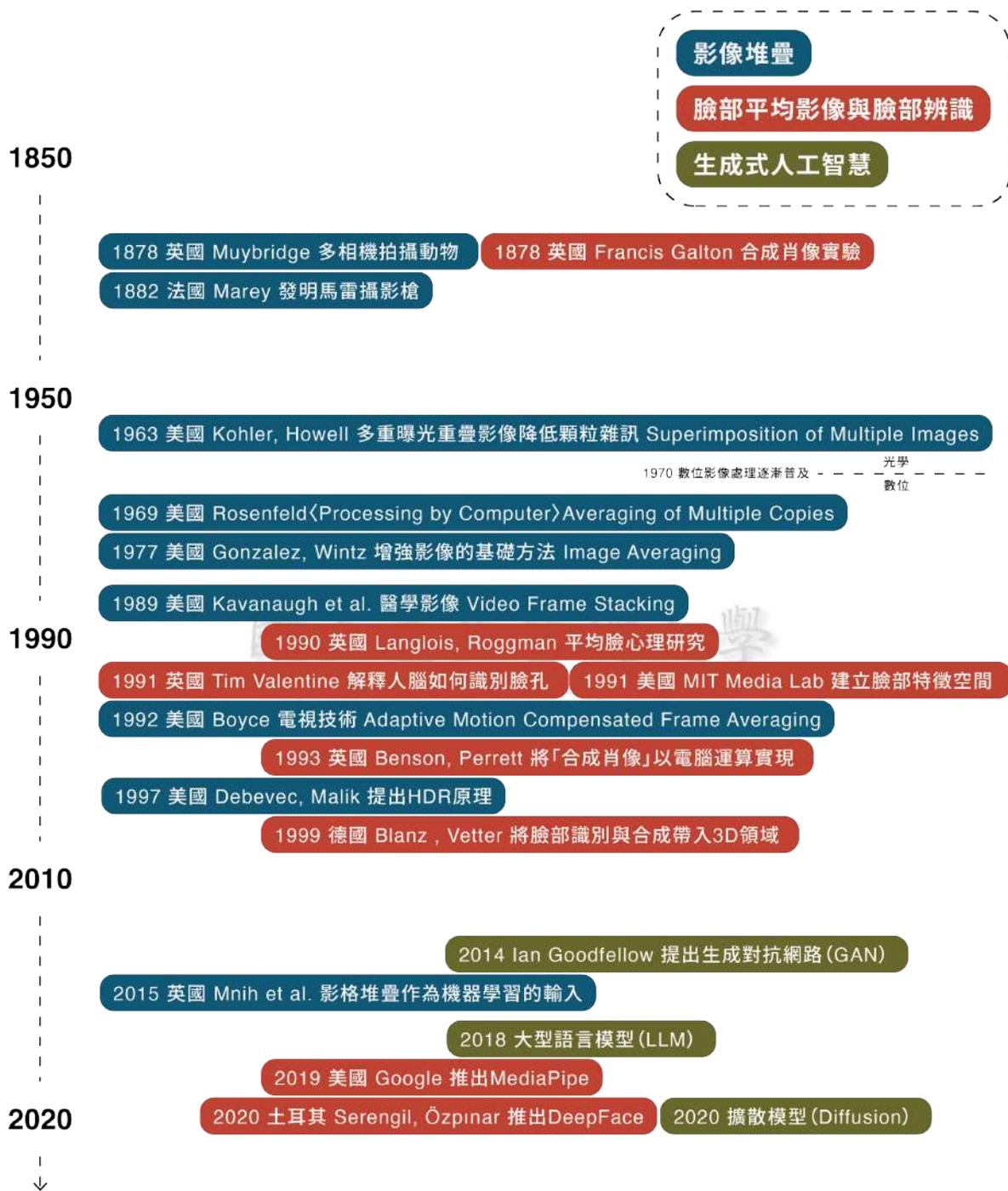


圖27、相關技術年表

### 第三章 創作論述與執行

本章節將論述三件作品的創作意圖與執行細節。第一件錄像作品《National Freeway 5》(兀兀騰騰寂寂蕩蕩)討論了身處在快速移動的實體空間中，所帶來如靜止不動的身體空間意識與記憶時間知覺交錯的精神意識；第二件互動裝置作品《The Mirror》(舉頭三尺有什麼)圍繞虛實空間之間的互動關係，將實體空間中的物件作為與虛擬空間中的對象建立關係的媒介，以數位監控與隱私為題，展開對自我身份等精神意識的討論；至於第三件作品《Self Echoes》(我相·人相)，則是以一個線上網頁的呈現形式，討論現代人生活在社群等虛擬空間中所形塑出的數位自我與身分意識。

#### 第一節 《National Freeway 5》(兀兀騰騰寂寂蕩蕩)

##### 一、創作論述

兀兀，如如不動；騰騰，自在無拘；寂寂，寂靜清閒；蕩蕩，寬廣無邊。<sup>13</sup>國道五號上的雪山隧道，作為現代多數人來往台灣東半部的共同記憶，長達十三公里耗時十五年建成，成為台灣最具代表性的公路長隧道。自 2006 年開通至今已近二十年，以約十分鐘的路程徹底改變了過去台 9 線北宜公路九彎十八拐的用路經驗，成為橫跨東西的重要通道。然而，長隧道內部幾乎可視為封閉環境，快速移動的過程中，駕駛者不得不將視線專注並固定在透視法畫面中間消失的盡頭，其封閉的駕駛環境讓快速的移動或許反而成了不再移動，此時對於時間推移與空間變換的身體感知與記憶產生

---

<sup>13</sup> 參考出處有二。一為唐·六祖惠能大師(638-713)圓寂遷化之前開示學人的最後一首偈語，「兀兀不修善，騰騰不造惡，寂寂斷見聞，蕩蕩心無著。」《六祖壇經·付囑品第十》；二為明·唐寅(1470-1523)〈自笑〉詩：「兀兀騰騰自笑癡，科名如鬢髮如絲。」形容昏昏沉沉的樣子。

了交錯。排風設施、汽車引擎、輪胎摩擦地面等低頻聲響的迴響與震動透過聽覺與觸覺重複在身體裡低沉地迴蕩，在介於動與不動之間的狀態中體驗感受向前行的速度，眼前不動視線似乎超越隧道圓拱型邊界的畫面，構築出一片超越現實，如動不動，狹窄卻寬廣無邊的移動風景。

## 二、創作動機

在童年記憶的家庭旅遊經驗中，有著「九彎十八拐」之稱的北宜公路為宜蘭人外出旅程中的必經之路，在進入翻山越嶺的蜿蜒道路前總會擔心暈車引起的不適感。約莫在我十歲的那年，國道五號蔣渭水高速公路頭城至蘇澳路段先行通車，當時在高架的高速公路上，對於遠方高聳的山腰上兩個半圓形的山洞入口充滿許多的好奇與期待。在隧道通車後的前期，十多公里長的封閉空間、覆蓋調頻的廣宣電台、不時出現的隧道廣播系統，還有其他如速限、車距管制等等，它在當時帶給用路人幾近全新的用路體驗。而後將近二十年間，雪隧成為了我在家庭旅遊、求學返鄉、差旅往來、背井還鄉時的道路風景，甚至成為溝通上的重要的定位分界線。對我而言，國道五號與雪山隧道的意義早已超越單純的交通功能，它不僅是宜蘭與外地之間的橋樑，更是一個分散在記憶中，串連情感的通道，這段路承載了許多生命的轉折，也見證了宜蘭人在時代變遷下的集體經驗。

近年，從過往的乘客身份轉變為駕駛人的角色，讓我以一種不同以往的身體感受重新經驗這段路程，不同於以往快速穿梭的印象，在駕駛過程中看著隧道前方的消失點，兩側的水泥壁持續往左右移動、頭頂照明設備規律的排列、排風設備低沉的運作聲、輪胎摩擦柏油鋪面的震動，所有身體能感知的元素都以各自的頻率不斷重複地的發生，彷彿是整個動態空間被壓縮進了靜態平面，空間感被重疊在一起，時間感被放慢了。這種恍惚的感受促使我思考，當身體已在相對受限的車室中，而周遭環境又在快速移動的狹窄空間中，無法大範圍活動的身體是否反而進入了一種近似靜止，與此

同時意識卻又延展無邊的狀態？這樣的身體經驗與時間感知的交錯，正是此作品想要討論的核心動機。

### 三、執行方法

《National Freeway 5》為單頻道形式的錄像作品，錄像長度為 5 分 06 秒，影像比例為 21:9。作品影像流程請見圖 28，影像畫面請見表 10 內容，在畫面安排上，一開始的畫面為行進雪山隧道入口的畫面；進入隧道後，在一再重複的風景中，隨著時間角色的淡出造成畫面漸漸相互堆疊而模糊，隨推移而隱約浮現許多隧道外時間被錯置的山景畫面；聲音則是在拍攝過程中同步進行錄製，道路設備、風扇運轉聲、胎噪等聲響皆被真實記錄，中段堆疊上戶外的蟬鳴聲以及用路壅塞時段會播報的隧道廣播內容。畫面與聲音的搭配，意圖再現動與不動的身體感知，與寬廣無邊的意識狀態。最後作品讓物理空間與鐘錶時間再次加入，使畫面逐漸清晰後離開隧道而結束。

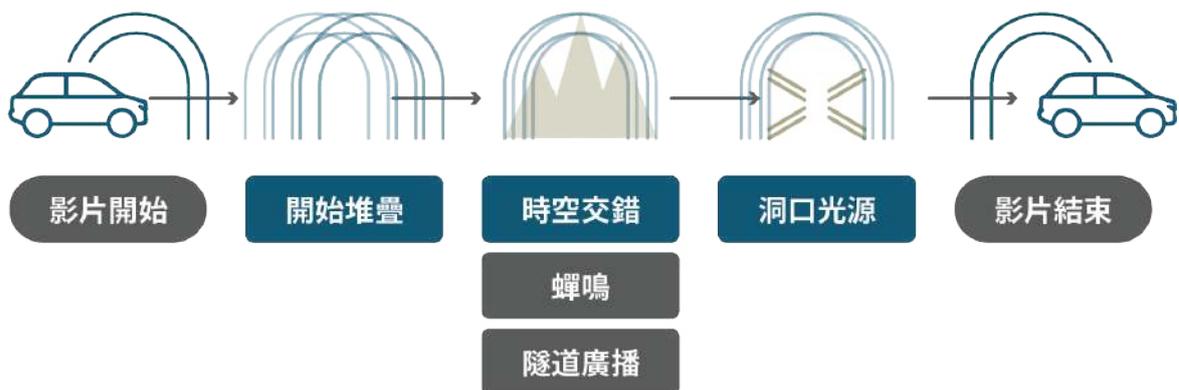


圖 28、《National Freeway 5》影像流程

表10、《National Freeway 5》作品畫面

時間	畫面
0:05	
01:33	
02:40	
04:45	

## （一） 影像來源

此作品在概念正式落地前的技術測試階段，素材來源主要提取自行車記錄器所拍攝之行車影像，曾經以「台北市萬華區至北投區關渡一帶」以及「萬華區至宜蘭縣羅東鎮」的路段作為畫面起迄點。期間實驗涵蓋了位於隧道、高架道路、市區道路，以及日間、夜間、晴天、雨天、陰天等不同環境條件下的行車影像片段，測試結果如表 11，結果顯示，天候明暗、沿途地景、車輛多寡及停等時間長短皆大幅影響畫面組成。然而，測試也發現行車記錄器因為受限於主要用途、長時間錄影、體積、儲存容量等軟硬體設計，其對比度、銳利度等畫面品質並不如預期，因此在後期正式選定以「國道五號南下路段」作為作品的拍攝場景後，拍攝素材包含國道上封閉的隧道空間與戶外的道路路段，選定於天氣狀況良好、車流較少的清晨時段，將攝影器材固定於車輛前檔玻璃上進行正式拍攝，顯著的提升了影像品質，畫質比較如表 12。

## （二） 堆疊方式與透明度設定

上述來源所擷取的影像，在本作品中主要使用 Python 結合 OpenCV 函式庫進行影像處理。在獲取原始影像片段之基礎資訊（如影格長寬、總影格數）之後，逐格將每一影格之透明度調整為「總影格數」分之一。最後將所有經過透明度處理的影格進行線性堆疊運算，並將此運算過程儲存為新的影像檔案。

本作品的主要核心視覺是在播放約一分半鐘後，漸漸出現由多層影像疊而成模糊影像，除了隧道內的影像在經過運算處理後產生了特殊的視覺模糊性，暗示逐漸昏沈的身體感，作品中段（約二分半鐘）再多疊加了戶外路段經過運算後的影像，隱喻無邊無際的意識狀態。當中原本透過攝影機拍攝、在現實中擁有相當距離的「隧道內」與「戶外」等多個不同實體空間畫面被接連在一起，藉此達成時間與空間在視覺上，透過壓縮而重構了實體存在的距離概念，使影像在視覺上同時存在封閉與開闊、陰暗到光亮的多重空間，形成一種連續卻跳躍的動態視覺經驗。而堆疊方式的設定亦是依

據實驗過程中的結果而定，從表 13 可以發現，「逐格讀取」的方法相較於「每十格讀取一格」，隧道上方的燈具與側邊的數字燈箱在畫面的呈現上擁有較高的連續性，因此採用「逐格讀取」作為本作品的堆疊方法。

### （三） 聲響設計

聲音部分採現場同步錄製，真實記錄了隧道內部的環境聲場，包括道路設備運作聲、低頻運轉及車輛行駛之胎噪。為呼應視覺上多時空交錯的畫面以及戶外景觀的出現，聲響設計在中段也疊加了另外錄製的戶外自然環境中的蟬鳴聲，以及僅在用路高峰時段出現的隧道交通廣播。透過視覺與聲音符號的跨時空堆疊，試圖也在聽覺層次上表現內部與外部空間交織的感知狀態。

表11、《National Freeway 5》作品實驗過程

1、市區道路、日間、陰天	2、市區道路、夜間、雨天
	
3、高架道路、日間、晴天	4、隧道、日間、晴天
	
5、隧道、高架道路、日間、陰天	6、高架道路、市區道路、日間、晴天
	

表12、《National Freeway 5》畫面品質比較

以行車紀錄記拍攝之畫面



國立臺灣藝術大學  
Taipei National University of the Arts



表13、《National Freeway 5》讀取影格數設定比較

<p>每十格讀取一格</p>

<p>逐格讀取</p>


#### 四、作品呈現

此作品曾於國立臺北藝術大學行政大樓前廣場公開展出（如圖 29），當時的呈現方式為將錄像裝置於汽車內的多媒體主機前，透過車內音響輸出聲音，而錄像後方為朝向東南方的景象，在地理位置上正好呼應台北、雪山隧道、宜蘭三點一線的關係。此次的展出，試圖為觀眾營造一種內外呼應與虛實共構的觀看體驗，在觀看錄像的同時也能感受得到車室周圍的環境、聲音的變化以及汽車引擎發動中的些微震動。而後，為了想讓影像與現實空間的關聯能更緊密，遂找尋到一處位於新北市坪林區 106 乙線上的眺望點進行作品記錄（如圖 30），此處雖無法直接看見隧道入口，卻也能夠遠眺國道五號進入雪山隧道前的路段及坪林方向的山景，讓影像的內外空間能夠更緊密地相互呼應。



圖29、《National Freeway 5》於國立臺北藝術大學展出



圖30、《National Freeway 5》於106乙線眺望點進行作品記錄

## 五、限制與改進

國立臺北藝術大學  
Taipei National University of the Arts

以目前的錄像裝置呈現方式而言，雖然作品藉由車室空間、窗外景象與外部聲響的互動，形成了獨特的觀看體驗，但整體仍存在可改進之處，儘管影片已達4K畫質，但在僅有12吋的顯示器上，畫面尺寸過小使得影像細節難以被完整感受。此外，也有車室空間同一批次至多只能容納五位觀眾的限制問題，此現象促使我重新思考作品的展示邏輯，若將此作品回歸至黑盒子空間，採用大尺寸投影形式，是否能使影像的時間延展性與空間沉浸感更為完整地被經驗？抑或是外部具有隨機性的角色參與，其實也擁有存在的必要性？這樣的想法或許也提供了我下一階段的測試與實驗方向。

## 第二節 《The Mirror》（舉頭三尺有什麼）

### 一、創作論述

舉頭三尺的地方有什麼？我們的一思一念、一舉一動，是否正被關注著？「萬事勸人休瞞昧，舉頭三尺有神明。」這句俗諺出自明代萬曆年間（1573- 1620）的《增廣賢文》，代表的是精神上的自我道德規範，勸誡凡事應三思而後行。《The Mirror》（舉頭三尺有什麼）是一個雙面鏡影像互動裝置作品，外觀如圖 31 所示，形式上引用傅柯把環形監獄作為現代紀律社會的隱喻，以日常物件鏡子為媒介，在實體空間中創造一個想像的虛擬空間，一面會監視人臉的鏡子，裝置內部的監視鏡頭紀錄著每個曾出現在鏡中的人，每個人的過去時間的影像也都會因此而殘留。



圖31、《The Mirror》作品外觀

透過雙面鏡與光線的控制，在使用者的鏡像反射與虛擬的生成影像之間產生不同的觀看方式與權力關係。首先，觀看的能力被賦予在攝影機一方，當攝影機開始監看與紀錄時，使用者此時只能看見自己的反射，如圖 32；而當光線漸弱時，觀看的權力

由裝置內部轉移至外部的使用者，使用者將得以看見鏡中出現一張疊合著自己與眾多  
人臉的生成影像，如圖 33，雙面鏡讓他們能夠窺探那些曾被攝影機看到的人（包括一  
部份的自己）。然而，如果觀看者想在鏡子前持續地窺視，隨著互動循環週期的增加，  
其內部生成的影像會越來越趨近於觀看者自身的鏡像，進一步的隱喻觀看權力的喪失。



Taipei National University of the Arts

圖32、使用者只能看見自己的反射

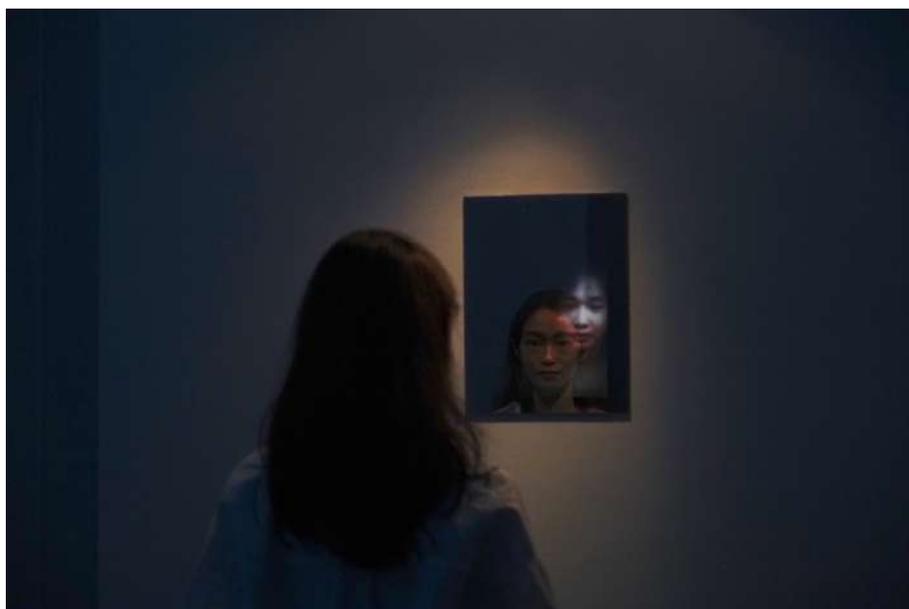


圖33、使用者可以看見鏡中影像

在現代社會中，「不為人知」或許只是我們的自以為是。各類型監視設備讓我們得以觀看身體所不能及的區域，許多的行為和狀態皆可能不經意地留下記錄，在這之中，我們是否有辦法想像自己在每一次的呼吸之間又留下了哪些新的痕跡？想像原本應屬於我們的個人隱私，似乎在不知不覺的大家都看得到了？

## 二、創作動機

對於個資與隱私問題的關注，始自於過去任職於社群溝通與行銷企劃的工作經驗，隨著演算法與數據分析技術的進展，現代科技使得內容辨識與行為追蹤等應用層面的功能升級，讓受眾輪廓能更容易精準地捕捉分析，然而，即使有著相關法律的規範，個資網路外洩的事件仍時有耳聞，也確實帶來分隱私層面的信任問題。除此之外，作為現代科技的日常使用者，我們大多都有過被追蹤並投放廣告的親身經歷，似乎我們的行為甚至是對話都進入了廣告商的資料庫，成為一筆筆能夠判斷我們喜好的依據。其實隱私問題的意識遠遠不只有上述的這些，舉例來說，俄羅斯網站《insecam》就以網站的方式，公開了全球所有使用預設密碼的 IP 網路監視器，只要稍有不留意，包含我們在內的資料、影像，就有可能變成公開資訊。(如圖 34)

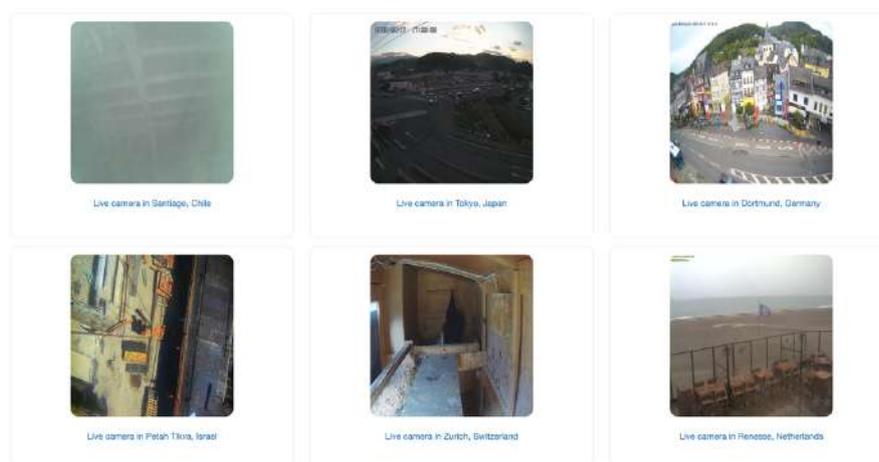


圖34、俄羅斯網站《insecam》網頁截圖（圖片取自[87]）

另一方面，在實體空間中對於鏡面的想像，則源自於自身近年起居位在台北公寓的租屋處，在樓梯轉角牆面上掛著的一面徵信社廣告方鏡，如圖 35 所示。雖不知這些鏡子的來由，卻在許多類似的公寓空間中或鄉間廟宇間，都能看見類似的存在。它既是日常空間裡最尋常不過的物件，卻由於鏡面的主要功能以及徵信社廣告相連結，帶來一種被窺視的引申聯想，兩者指向了一種觀看與被觀看的權力關係，使得站在原本鏡子前的日常又多了一絲地的不確定。



圖35、樓梯轉角的廣告方鏡

由此延伸的想像是，如果實體的鏡子如同網路環境一般，其內部存在著一直凝視著我們的觀看者，而我們卻無法看見對方？這樣的想像在公寓與日常空間裡顯得不切實際，但不肖人士透過鏡面進行非法窺視的案例也並非不曾發生。這種「可疑卻無從確定」的狀態，恰巧對應了數位環境中的隱私焦慮，我們明知行為軌跡、對話內容可能被演算法追蹤、被廣告商收錄，但卻無法看見背後真正的「觀看者」是誰、以何種方式凝視著我們。因此，鏡子不再只是單純反射自我的物件，它或許成為一個權力的隱喻，一種模糊的監視機制。就像網路平台上的監控方式，讓人時刻處於「可能被看見」的狀態，而這種不確定感才是最讓人不安的地方。基於這樣的心理與經驗思考，作品嘗試將日常鏡像與數位監控的概念並置，把看與被看的關係具象化為一種可見卻又模糊、虛實交錯的空間。

### 三、執行方法

#### (一) 硬體設計

##### 1. 外部構造

裝置外觀為一道懸掛雙面鏡的白牆，上方裝有投射燈具，構造如圖 36 所示。懸掛鏡面尺寸以公寓樓梯間牆上常見規格作為設計參考，營造日常中特有的觀看場域。

雙面鏡（或稱單向玻璃、雙向鏡）是一種特殊的半透鏡，其中一面具有低濃度金屬塗層，作用原理是玻璃兩側的光線強度不同時，光線弱的一方能像透明玻璃那樣看見光線強的一方，而光線強的一方只能看到自己的影像，使其看起來像普通的鏡子，觀看方式請參考圖 37。差異在於雙面鏡通常呈透黑色，後方物體因此在觀看上亮度會稍微降低。

國立臺北藝術大學  
Taipei National University of the Arts



圖36、《The Mirror》作品外部構造示意

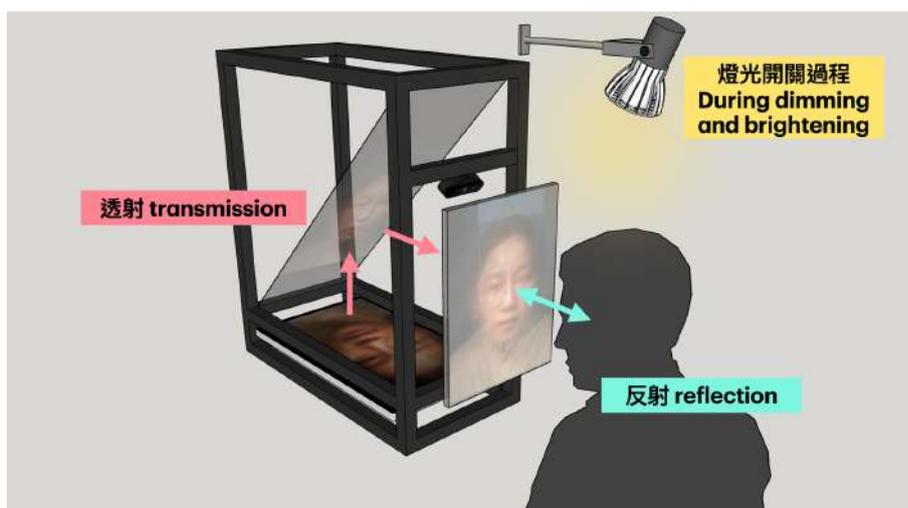
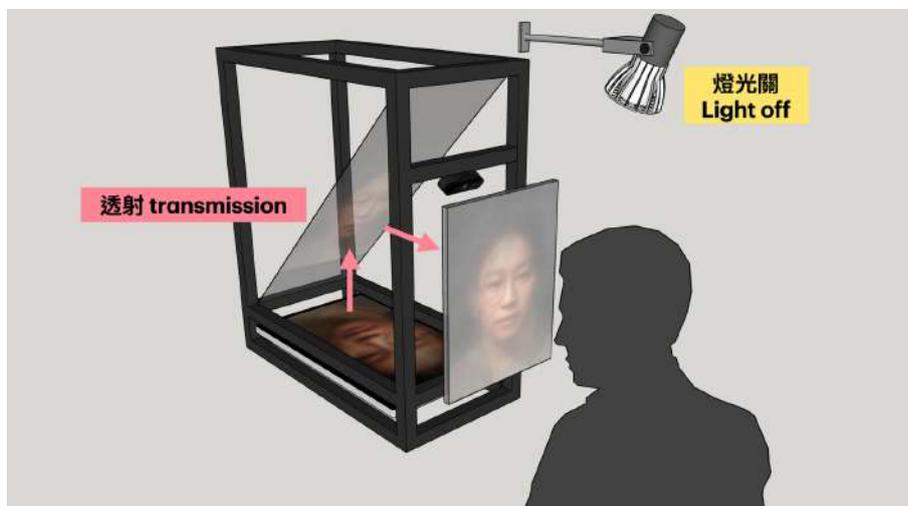


圖37、《The Mirror》作品使用之雙面鏡效果示意

## 2. 內部構造

在鏡面的後方則是連通裝置的內部，如圖 38，其內部構造包括電腦主機、燈光控制系統、浮空投影裝置與攝影機。在燈光控制系統部分，使用的投射燈具為使用交流電鹵素燈泡，其優點為可以使用旋鈕式調光器直接以電流控制光線的明暗，無需額外轉換電壓。然而缺點也顯而易見，相比新型 LED 燈，會出現明顯地發熱之外，也更為耗能。硬體上使用 Arduino 作為控制器，連接電腦主機與伺服馬達，當電腦主機發出操作訊號後，以伺服馬達直接驅動調光器進行即時的光線控制。

在浮空投影裝置部分，透過一個斜置的透明壓克力板反射下方橫置的液晶螢幕顯示器，利用影像的反射，產生影像浮現於空中的視覺效果，比起直接將螢幕貼合鏡面的裝置形式，在觀看上擁有更多的空間感。但由於因為雙面鏡與浮空投影裝置均會削弱的顯示內容的亮度，於是在裝置過程中改良了浮空投影裝置的構造，在透明壓克力板後方間隔 2 公分處再加上一面透黑色鏡面壓克力，雖然空間感不如典型浮空投影裝置，卻能作為空間感與顯示亮度的折衷方案，且正因為雙面鏡的濾光作用，顯示器的反射邊緣也因此而模糊。結合燈光控制，更進一步地創造了環境光線的明暗差異，讓鏡內的影像在互動的一開始能被隱藏。



圖38、《The Mirror》作品內部構造示意

攝影機的配置，原先預計參考日本電氣通信大学安藤将平（Shohei Ando）與小泉直也（Naoya Koizumi）發表的技術研究案例〈Floagent: Interaction with Mid-Air Image via Hidden Sensors〉[88]，計畫使用熱鏡（hot mirror）與紅外線攝影機（IR camera）達到隱藏攝影機的同時，也取得使用者的平行視角影像，如圖 39 所示。但因為材料取得不易，加上紅外線攝影機只能輸出的黑白影像，在希望保留影像色彩的前提下進行了一項技術實驗，保留研究案例中的硬體配置架構，將紅外線攝影機改為普通全彩攝影機，並以雙面鏡兩側明暗運作原理為基礎，將原本的熱鏡也替換成雙面鏡。

然而實驗發現，此方法在外部光線強度能讓使用者接受的的狀態下，攝影機會同時捕捉到雙面鏡的反射影像；且若將原本的雙面鏡也安裝上去，攝影機拍攝的畫面將經過兩道減光濾鏡，因為外部光線的不足，也將導致捕捉影像亮度的不足。因此經過調整後，最終方案將使攝影機從反射鏡的正後方，移動至反射鏡前，並固定於上方，配合顯示器、光線、影像觀看角度的調整下，除了影像視角調整為些微的俯視視角外，基本上也已達成了最初想要隱藏攝影機的目的。

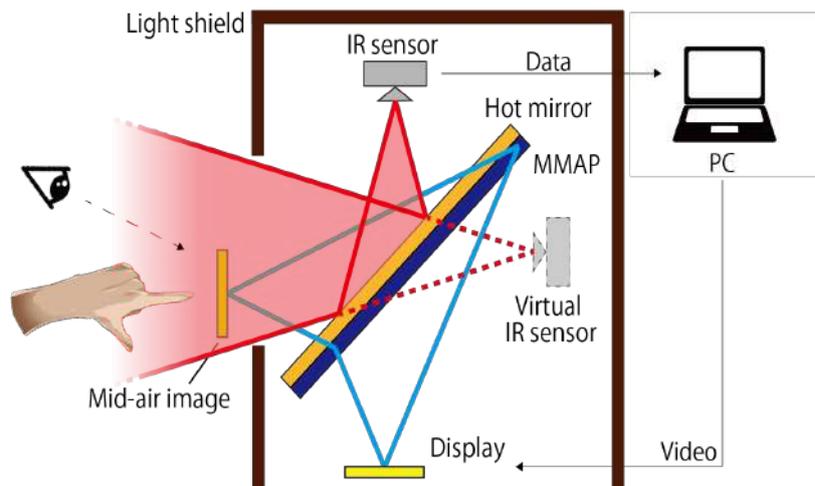


圖39、〈Floagent〉中熱鏡與紅外線攝影機系統使用示意圖（圖片取自[88]）

## (二) 互動設計

整體互動流程請參閱表 14。在初始狀態中，光線為開啟狀態，當觀看者接近鏡子時，使用者只能看到鏡子上自己的鏡像，就如同平常的鏡面觀看體驗一樣，而此時隱藏於內部的攝影機若偵測到觀看者的臉部影像，將會觸發整個互動體驗流程。系統將在背景中開始啟動搜集運算，在這個階段中，使用者仍然只會看見自己，此階段時間長度為五秒。在接下來的第六秒，攝影機停止偵測與紀錄，裝置上方的投射光線也會開始逐漸變暗<sup>14</sup>，使用者也在此時才能夠看見鏡子中生成出來的虛擬影像逐漸浮現，此過程時間長度為三秒。之後再經過約十五秒的時間，光線將會開始重新亮起，三秒後，虛擬影像也將再次消失在鏡子中，回到只能看見使用者反射的狀態，一個互動循環在此結束，新的互動循環將會在五秒之後重新開始。

整體流程的時間配置經過調整與測試，互動時間設定為大約三十秒左右，包括偵測五秒、燈漸弱三秒、維持約十五秒、燈漸強三秒、間隔五秒，試圖透過較短時間的互動循環，一方面避免等待過長造成中途離開，另一方面讓同一使用者願意駐足並再一次地參與體驗。由於疊加影像會在多次循環後逐漸趨近於使用者面孔，在約二至三個週期後，呈現出的影像將會與使用者的面孔越來越相似。這也試圖讓使用者去做下一步的思考：「究竟剛剛看見的影像是什麼？」以及「它為什麼會越來越像我？」。

---

<sup>14</sup> 在最初版本的設計邏輯是，如果光線太暗，相機就會停止錄影。在因應不同場地環境而進行改良後，現行版本改以秒數作為判斷基準。

表14、《The Mirror》作品互動流程

示意圖	時間 (秒)	說明
	1	初始狀態，燈光完全開啟，使用者走進感應區。
	2-5	燈光完全開啟，使用者看見外部反射。
	6-8	燈光漸弱，使用者同時看見外部反射與內部透射。
	9-20	燈光幾近關閉，使用者看見內部透射。
	21-24	燈光漸強，使用者同時看見外部反射與內部透射。
	25-29	燈光完全開啟，使用者看見外部反射，體驗結束。

### (三) 系統設計

本系統以即時運算為核心，以主機運算並且連動燈光、攝影機、顯示器等控制系統，在影像處理流程上，共包含四個主要步驟。在表 15 中，以八張不同使用者經過 AI 處理的照片作為範例，說明整體運作過程。在實際展示中，系統設定以約 8.8% 的透明度在迴圈中進行權重處理，即約有 50 張臉部影像能對最終影像產生影響，形成一個即時更新的「臉孔疊影」。因此觀眾於鏡面中所見的是由前 25 位左右的參與互動的使用者（每人兩張影像）所累積合成的影像，影像會隨著新使用者的加入而動態更新，既保存過往參與者的痕跡，也不斷地融入新的面孔。圖 40 為此系統實際執行下，因不同使用者加入觀看的互動過程，而造成圖像不斷改變之結果。此設計讓鏡面中的「臉」可以保持在持續變動的狀態，既反映當下的參與者，也同時疊合過往的觀看痕跡，透過動態影像的堆疊，營造出一種介於個體與群體、自我與他者之間的面貌。



圖40、《The Mirror》系統實際執行結果

圖 41 以流程架構圖方式來簡單示意整體使用者的互動過程，最初從燈光開啟的狀態開始，當用戶觸發隱藏攝影機時，系統會儲存和處理他們的影像，將新的影像依照權重堆疊至原本的合成影像上面。當光線變暗時，新的合成影像在鏡中被顯露出來，而後，當燈再次亮起，合成影像漸漸淡去，使用者也會再次看到他們的反射。

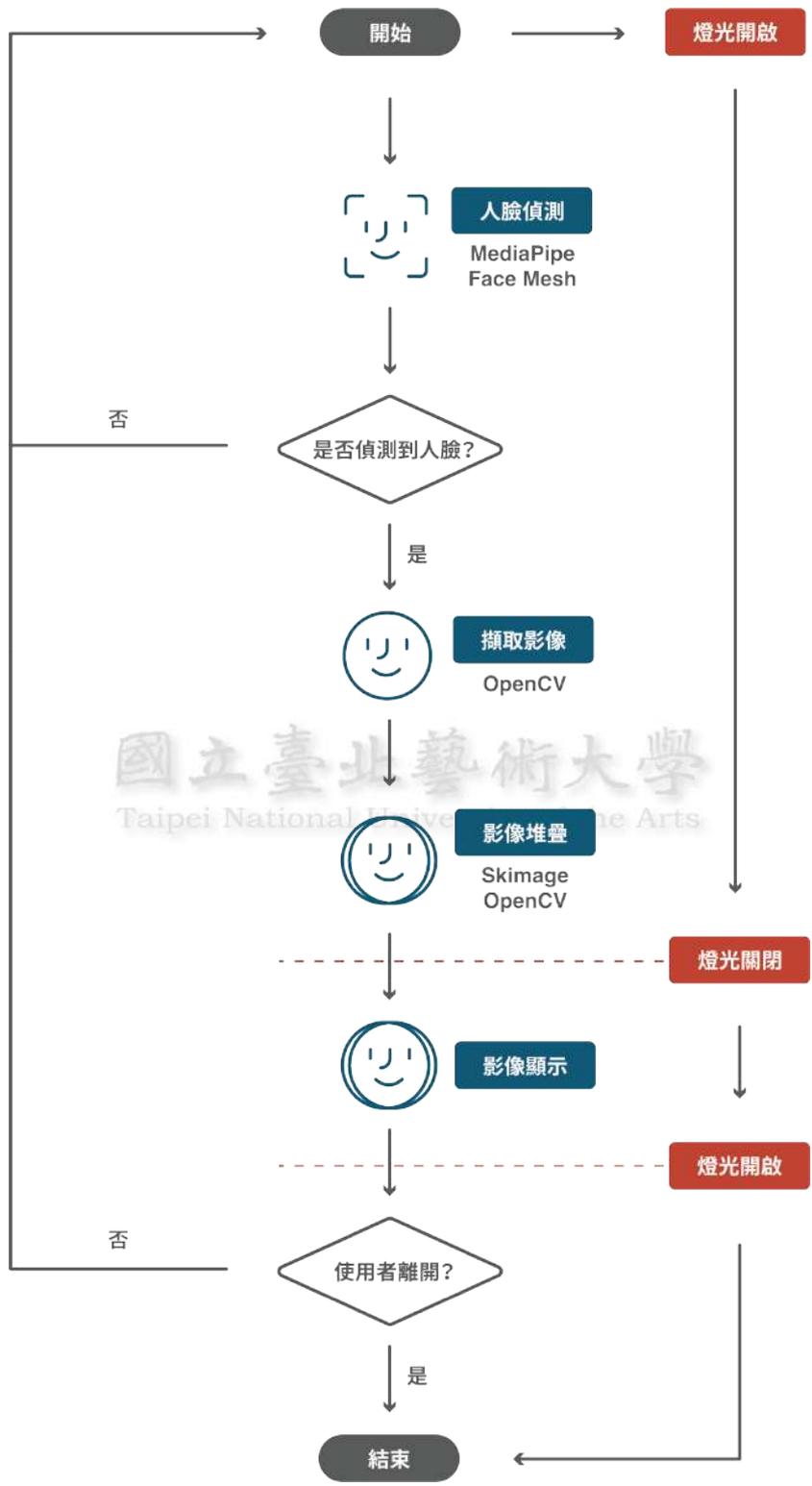
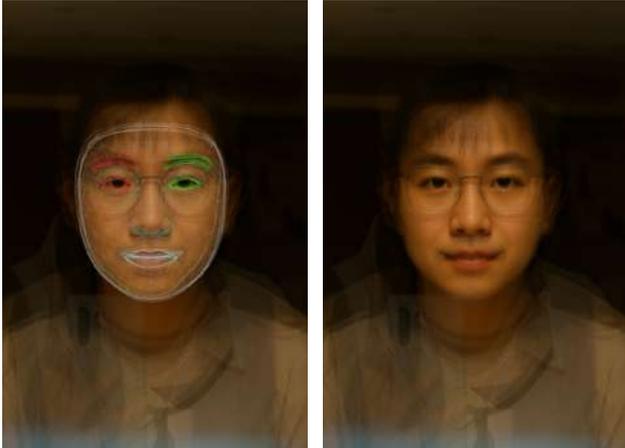


圖41、《The Mirror》作品互動流程圖

表15、《The Mirror》系統處理過程

示意圖（影像經 AI 進行去識別處理）	步驟說明
	<p>1. 人臉影像擷取：</p> <p>攝影機即時偵測使用者並擷取使用者的臉部影像，作為輸入。</p>
	<p>2. Google MediaPipe 特徵標記：</p> <p>調用 Face Mesh API，在偵測到臉部影像後，為每張臉部影像建立 468 個特徵點，完整涵蓋五官輪廓與臉部細節。</p>
	<p>3. Skimage 影像對齊：</p> <p>引用 SimilarityTransform 函式進行最小平方法對齊（least squares alignment），將所有臉部影像在位置、比例與旋轉角度上依照參考點調整至統一規格。</p>

示意圖（影像經 AI 進行去識別處理）	步驟說明
	<p>4. OpenCV 影像疊合：</p> <p>最後使用 <code>cv2.addWeighted</code> 函式進行透明度調整與影像疊加，將臉部影像融合最終生成的畫面呈現於顯示器並反射在鏡面上。</p>

#### 四、作品呈現

### 國立臺北藝術大學

本作品至今擁有三次實體公開展出經驗，以及一次的國際研討會論文發表。實體展出分別於 2023 年《Make Sense》112 學年度新媒體藝術研究所學年評鑑展於國立臺北藝術大學地下美術館（圖 42）、2024《科技賦格——數位環境的永續及想像》展於王道銀行藝廊（圖 43）以及《2024 基隆美展》（攝影與新媒體藝術類）於基隆美術館三樓公開展出（圖 44）。此外，本作品曾另外以研討會論文形式，投稿 2025 IEEE 台灣消費電子國際研討會（IEEE ICCE-TW 2025），並於 2025 年 7 月 18 日在高雄展覽館進行論文發表（圖 45）。

在第一次的展出中將作品裝置於一樓樓梯前方的轉角處，裝置高度為 240 公分，鏡面畫心高度設計為 165 公分，導致部分想參與互動的使用者無法順利啟動臉部偵測。在程式設計上，也並無置入影像與對齊參考點的存檔讀取機制，需要依靠每次的對位校正來重新啟動，也無法讓原本生成的影像在開關機後持續使用。

上述問題在第二次展出前進行了相當程度的改善，除了將畫心高度降低至約 145 公分，讓大部分的使用者都有辦法參與互動機制，還加入了資料存檔與讀取機制，也設計了自動執行等開關機機制，讓整體的開關機程序更加容易操作外，也不會因為程式中斷而造成資料遺失。在環境因素上，藝廊的環境光線的可控性並不如第一次展出的場館可以依靠鏡頭和光線的物理條件來判斷互動條件，因此為了確保整體互動機制和流程順利，重新調整了程式上的邏輯，同時也改善了原本效能不足造成卡頓的問題。然而此次展出上仍有改善空間，在這次經驗中，作品為了想要與樓梯空間進行對應，而將作品裝置於藝廊空間的樓梯下方空間，因為裝置位置內縮至樓梯下而導致稍微隱蔽，空間呈現在頭頂空間也略為侷促，另外，通道側邊的大面積採光玻璃也造成鏡中虛擬影像觀看上有些許影響。

在第三次於基隆美術館三樓的展出中，汲取前兩次的展出經驗，終於能在此階段以相當完整的展佈方法進行展出，大致保留了第二次展出所做的調整，將裝置整體高度因應空間高度而重新調整回 240 公分，同時也使用更為顯眼的體驗位置指引，讓使用者在初次接觸時，就能在最佳位置參與作品體驗。



圖42、《The Mirror》於國立臺北藝術大學地下美術館展出



圖43、《The Mirror》於王道銀行藝廊展出

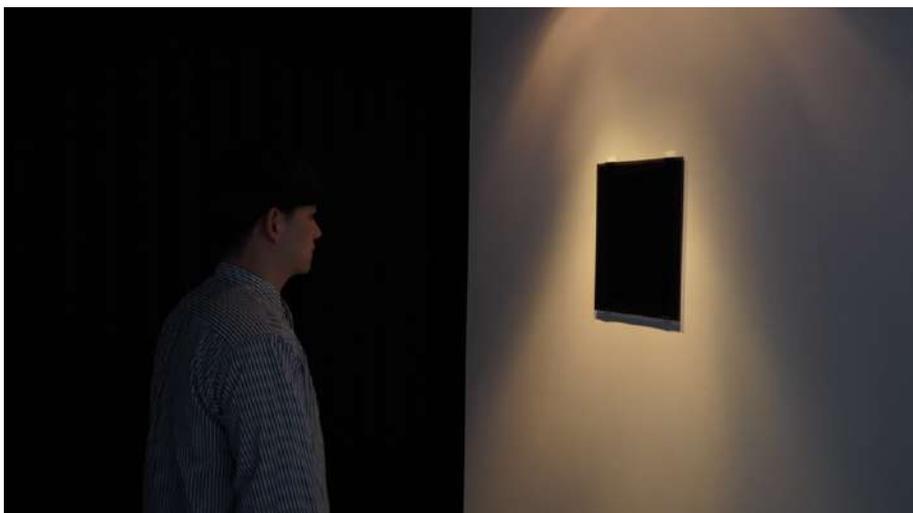


圖44、《The Mirror》於基隆美術館展出



圖45、《The Mirror》會議論文於IEEE ICCE-TW 2025發表

## 五、使用者體驗回饋

為評估互動作品的觀展體驗與概念傳達效果，本研究針對作品進行了使用者問卷調查，並輔以開放式意見蒐集與觀察記錄。問卷依據李克特量表（Likert Scale）設計 [89]，涵蓋情緒反應、知覺認知、自我辨識與虛擬替身認同等面向，藉以分析觀眾在過程中的感知差異與心理變化。

本調查於王道銀行藝廊及基隆美術館的展出期間進行，本次調查共回收 15 份有效樣本，受試者年齡介於 20 至 40 歲之間，20 至 30 歲 11 人、30 至 40 歲 4 人，性別分布生理女性 10 人、生理男性 5 人，職業背景涵蓋學生、金融業、服務業、文化藝術相關從業者等。問卷分為情緒反應與知覺認知兩大部分。情緒反應部分包含十個形容詞（新奇、有趣、愉悅、平靜、不安、沉重、焦慮、孤獨、懸疑、詭異），以 1 至 5 分的區間評估受試者的感受。知覺認知部分則透過四個問題，檢視受試者是否能理解作品、辨識鏡中臉部影像的變化，以及是否意識到潛在的監控情境。多數受試者的興趣或專業領域與藝術、設計、人文研究相關，部分涉及科技、環境與社區營造，推測受測群體對本作品所探討的「鏡像、監控」主題可能具有一定的敏感度。

圖 46 為調查情緒反應之結果，受試者在「新奇的」(平均值 4.13 分)、「有趣的」(平均值 4.20 分)的評分最高；「愉悅的」(平均值 3.07 分)與「平靜的」(平均值 3.07 分)次之；有一小部分受試者回饋其感受到「不安的」(平均值 2.60 分)、「懸疑的」(平均值 2.53 分)與「詭異的」(平均值 2.67 分)等略為不安的氛圍。「沉重的」(平均值 2.00 分)、「焦慮的」(平均值 2.00 分)、「孤獨的」(平均值 1.93 分)的評分則偏低，顯示作品並未引起過度負面情緒，此結果的分數分佈大致與預期相符。

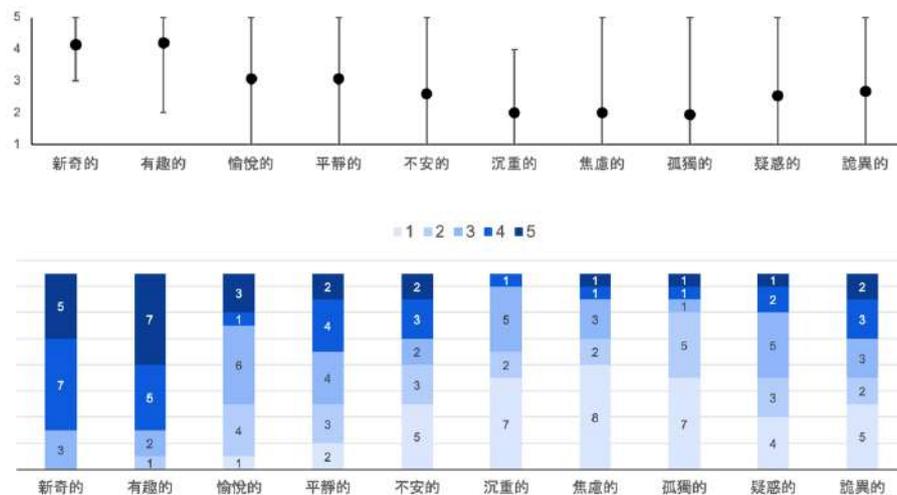


圖46、《The Mirror》問卷結果——「情緒反應」部分

在知覺認知的部分，問題設計如下：

Q1：您覺得您對此作品在體驗後有足夠的了解嗎？

Q2：鏡子內存在一個臉的影像？

Q3：鏡子內的臉逐漸產生一些變化並且越來越像自己？

Q4：鏡子的內外關係讓我意識到我可能正在被監看？

Q5：鏡子的內外關係讓我意識到現代社會的監控現象？

根據圖 47 結果顯示，受試者普遍能理解作品核心 (Q1 平均值 4.07)，幾乎所有人都注意到鏡中臉部影像 (Q2 平均值 4.93)，約有三分之二的人發現影像逐漸變得

「像自己」(Q3 平均值 3.87), 並意識到自身處於可能被監看的情境(Q4 平均值 3.93), 也有超過半數的受試者更能領會到作品欲傳達的現代社會監控隱喻(Q5 平均值 3.63)。

Q2 作為最基礎的條件, 受試者們確實將鏡內影像當作一個可辨識的他者, 而在 Q3 中, 受試者進入了自我認同被觸發的狀態; 而 Q4 的結果也顯示, 受試者的感知並不止停留於自我辨識, 而是進一步地意識到自身處於在可能被看見的情境之中, 這種自我警覺正是傅柯在環形監獄隱喻中內在歸訓的運作機制。最後在 Q5 的調查結果, 部分受試者能將此體驗轉化為對現代監控現象的理解, 說明作品所呈現的並非只是單純的鏡像和影像, 而是技術中介下可進行運算並在系統中被轉譯為可運算的符號。

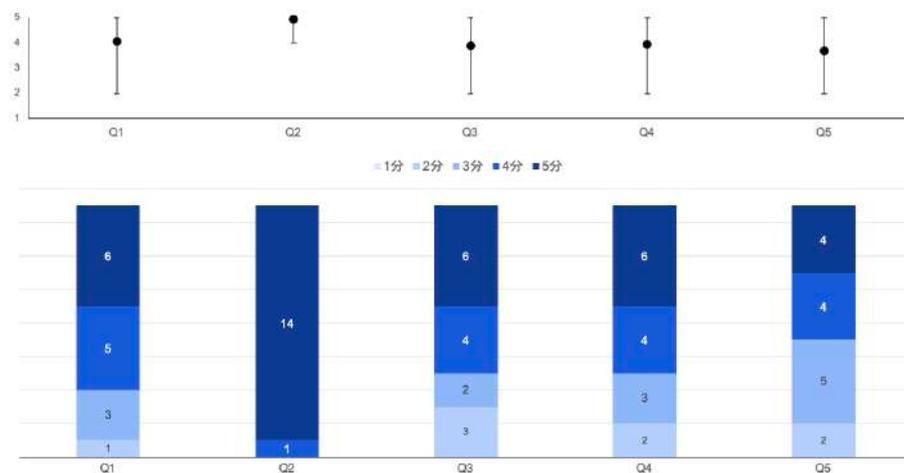


圖47、《The Mirror》問卷結果——「知覺認知」部分

在開放式問題的回饋中, 有受試者提到互動性帶來的驚喜, 「很酷, 裝置會隨不同觀眾互動後留下不同面孔。」也有人分享: 「和很多人同時看鏡子很有趣」, 甚至有人將體驗比喻為「與鏡中自己的對話」。少數受試者則提出了一些改善建議, 如「視覺效果不足」、「互動流程不夠流暢」、「部分技術問題仍待改善」。另有受試者提及一則反思, 「眾人臉部的總和是否是一種新時代的黃金比例?」顯示作品除了引導對於隱私與監控的聯想, 也引發了對於自我與群體形象的思考。綜合而言, 在量化數據與

質性回饋中，《The Mirror》雖然在技術與呈現上仍有改善空間，但仍在提供互動體驗的同時，達成了階段性的設計目標。

## 六、限制與改進

《The Mirror》的即時互動機制，使觀眾在與鏡像的互視過程中同時成為觀看者與被觀看者。大部分觀眾在體驗中能意識到監控隱喻並感受到詭異的張力，說明即時互動某種程度上也成功喚起了觀眾對「被觀看」的自我覺察。但在現階段的呈現中，為了取得使用者體驗與系統效能的平衡，鏡中虛擬影像的位置參考點是經過預設的，並且在運算過程中是固定不變的，這也導致不同身高的使用者在觀看影像時將會需要有意識的去移動身體，讓鏡中反射與影像相互對應。若在未來的改進中，能夠解決目前虛擬影像與反射影像的即時對位卡頓，也將能讓「符號堆疊」的意義更加完整。

### 第三節 《Self Echoes》（我相·人相）

#### 一、創作論述

《Self Echoes》（我相·人相）是一件網頁形式的線上互動藝術作品，探討社群媒體所生成的數位人格，以及它與真實自我之間的落差。作品名稱取材自《金剛經》「四相」中的「我相」與「人相」<sup>15</sup>。佛典中所說的「相」是一種對自我的概念化執取，其中也包括外在的形貌，在今日社群媒體的脈絡下，呈現出強烈的呼應，我們透過貼

---

<sup>15</sup> 「四相」出自佛教經典《金剛經》，指的是我相、人相、眾生相、壽者相。其中，「我相」意指恆常不變的本質我（atman）；「人相」為自我或人格（pudgala）的概念；「眾生相」是對一切有情生命（sattva）的執取；「壽者相」則指靈魂或生命體的延續（jīva）。經文強調修行者若執著於這四種觀念，即不能真正體悟菩薩之道。其核心在於揭示「無我」：不論是本質我、自我、群體或靈魂，都只是概念上的假名，並非恆常不變的實體。

文展示、編輯並塑造自我形象，卻往往執取於一個並非恆常的自我投射。當個人的時空經驗被轉化為網路貼文，它們進入一個去物質化的虛擬世界，成為可儲存、觀看與重新排列的數位資訊。而這些貼文之間的時間空隙中，或許，一個與現實身份有別的「虛擬分身」、一個由貼文、影像與演算法所構成的「數位回聲」已逐步形成。這個虛擬分身既非真實的存在，也並非全然的虛構，而是介於兩者之間的流動身份。

作品的運作方式為：系統讀取使用者公開 Instagram 帳號，分析近期貼文中的圖像與文字，分別萃取出人臉特徵與語言風格，並轉化為新的影像提示與模擬語氣的文字。在此基礎上，生成模型產生對應圖像，最後形成一則圖文貼文，成為使用者的「數位回聲」。本研究回顧使用者的自身的社群歷史，在經運算並重新生成演繹後，試圖引導使用者思考：當我們在社群媒體上展示自我時，作為一個在數位環境中被觀看的客體，與真實的主體間產生了哪個與真實自我不盡相同的數位身份？這個代替我們被觀看的替身是誰？而我們是否能從這些數位回聲中，藉機再次觀看與認識自己？

《Self Echoes》延續前兩件作品對「虛實共構的時空意識」之探討脈絡，進一步將這種時間與觀看問題轉化至社群媒體與 AI 生成的場域之中。作品將使用者公開的社群資料作為素材，透過符號堆疊生成演算法觀看下產生的數位自我。在這個意義上，《Self Echoes》既是前兩件作品的延伸，也是對虛擬空間中「自我、時間與意識」關係的實驗。

## 二、創作動機

### （一）社群媒體與演算法

在將近三十年的成長經驗中，各類型的社群平台在近二十年間成為附加於現實生活中十分習以為常的一部分，也因此見證了許多不同社群平台的消長。在早期以無名小站等部落格、網誌為主的社群使用經驗中，影像與文字等內容與個別使用者高度綁

定，發布時間為主要排序依據，觀看行為可被理解為一種近似線性的時間回顧。然而，自 Facebook 與 Instagram 分別於 2008 年、2015 年進入台灣市場，同時伴隨著智慧型手機的快速普及，Facebook 取代了網頁部落格，成為的多數人隨手分享生活感想、日常紀錄，並與朋友們進行互動的網路空間。而後 Instagram 的興起，讓影像成為社群中的主要媒材，限時動態功能的加入，也讓這些快拍、即時分享的短篇圖文取代過去的長篇書寫成為主流，而短影音在各大平台的出現，更是成功的吸引了使用者的注意力，成為近年社群媒體發展的重要趨勢。根據台灣網路資訊中心（TWNIC）於 2024 年度發布的《台灣網路報告》[90]，Facebook 與 Instagram 在上網者中合計擁有超過七成的市佔率，特別是形式上以圖片、影片、限時動態為主，文字為輔的 Instagram，僅次於 Facebook，穩居台灣民眾最常使用之社群媒體第二名，有 21.12% 民眾以此做為最常使用的媒體，也較 2023 調查結果 17.61% 呈上升趨勢。

與早年最具差異之處，在於演算法在平台中的應用，系統機制開始介入影像與文字的辨識、審閱與分類，並且依據使用者的觀看與互動習慣，進行選擇性的推播與排序，使相同內容可能以在不同時間點反覆出現在觀看者眼前。使用者的觀看不再只是依循時間順序展開，而是受到平台機制而被安排著。近年新型社群平台（如同樣隸屬於 Meta 旗下的 Threads）之興起，其演算法的設計即被視為可以迅速累積使用者的重要因素之一，進一步凸顯演算法在當代社群媒體生態中的關鍵角色。在此觀看經驗的轉變中，自我於社群媒體中的呈現逐漸不再僅是個人主動記錄的結果，而是同時受到平台運算邏輯的重組與再現，進而反覆成為當下觀看的一部分，時間感也因此呈現出非線性、跳動的狀態。

## （二） 社群平台成為展演場域

Instagram 的社群形式，使其成為使用者自我展演的重要場域。影像中的色彩、場景與人物姿態，文字中的語氣、節奏與詞彙，皆可由使用者自行決定後上傳，構成一組自身欲展現形象的圖文集合，既是一場個人的日常情緒展演，也在平台規則與演算

法邏輯的作用下成為可供他人瀏覽的資訊流。當我們透過這些圖文集合形塑我們的自身形象，使其他使用者成為這場自我展演中的觀眾，然而，這樣的認識也與真實的我們存在相當的差距。社群媒體上的圖文，往往是經過篩選與編輯後的片段，它既反映了真實，亦不代表全然的真實，圖文因而逐漸構築出一種可被觀看、卻未必能對應個體實際狀態的自我形象。

### （三） 中國哲學「氣」範疇的認識

另一方面，在接觸家族發展脈絡與研究所的修業過程中，逐漸建構出了我對「天、地、人」三者平衡的認知，而「氣」成為理解與貫通其中的關鍵概念。「气，雲气也，象形」，在《說文解字》中，气（氣的初文）本為象形字，原始意為自然界的雲氣，在人類的實踐和認識過程中逐步擴展意涵。在早期華夏文明的初民世界中，時間、空間不是認知的框架，而是生命的框架，生命起源與風相關的論述依照現今的邏輯看待，這應當是身處季風亞洲天文地理底下的產物，然而對風的神聖化與神秘化即是初民對於存在風背後深層無形力量的探究，反映出一種融合時空架構於主體的人性構造。[91]

春秋時期，《左傳》提出「天生六氣」<sup>16</sup>——陰、陽、風、雨、晦、明，為自然氣候變化的六種現象，後世思想家將此氣論進一步與五行（木、火、土、金、水）、八卦（天、地、水、火、山、澤、雷、風）系統相互涵攝。五行作為自然意象，相應五味、五色、五聲等關係，形成一種將天地運行與人類感官相連結的世界觀。戰國時期，《莊子》言「人之生，氣之聚也」<sup>17</sup>，把生命理解為氣的聚合與散逸；醫學經典《黃帝內經》承襲並深化先秦的氣論思想，認為氣是構成人體與天地的基本要素，五行之氣分布於臟腑系統，其運動變化決定了身體的生理功能與病理現象；在風水理論中，

---

<sup>16</sup> 來自《左傳·昭公元年》，「天有六氣，降生五味，發為五色，徵為五聲，淫生六疾，六氣曰陰，陽，風，雨，晦明也，分為四時，序為五節，...」

<sup>17</sup> 《莊子·知北遊》，「人之生，氣之聚也。聚則為生，散則為死。」

「氣」則構成地理與人居關係的核心。《葬書》提到「氣乘風則散，界水則止」，風水以氣的聚散流動判斷地勢的吉凶，山川河流被視為氣脈的通道，建築與聚落的選址必須順應、引導、蓄積天地之氣，以確保人事和諧興盛。漢代董仲舒則透過「天人感應」論，雖主要論述政治秩序，但也延伸至人性與氣的關係；《白虎通》更指出，人類稟受陰陽五行之氣而生，其性情、道德、精神皆因此而有差異<sup>18</sup>。到了宋明理學，朱熹提出「氣質之性」的觀念，認為個體因稟受之氣不同而形成各自的人格傾向，修養的過程便是調和、涵養氣質。

簡單來說，氣，乃至於五行，常被理解為萬物構成的根源。五物之所以成為五行，是因「行」為流動之物；隨著思想傳承，它不僅指向自然界中流動不定的生成力量，也成為理解生命、身體、人格與心性差異的重要依據，其核心特質並非穩定的實體，而是一種介於可感與不可見之間、以流動與聚散為主要特性的存在形式。如果從近代西方科學的脈絡來看，十九世紀物理學中的「以太（aether）」，曾被假設為充滿宇宙、承載光與電磁波的不可見介質；電氣（electricity）的流動則被視為推動萬物運動的能量形式；而質點（particles）模型則以最小單位的粒子解釋世界的生成與互動。這些概念都呈現了某一種「無形而能量化」的存在，其流動、傳導，構成人與世界的關係。

若將這樣的觀點延伸到當代數位環境中，網際網路的流動邏輯在結構上與中國哲學中「氣」的概念呈現出些許相似之處。然而，網際網路雖看似抽象存在，但其實際上是依附於電流的流動而得以運作。若以此為出發點，在當代數位環境中，網際網路或可被視為「氣」顯化後所形成的「風」，一種可被感知卻難以被直接指認其本體的流動現象；而電流，則可被理解為一種可與「氣」進行類比的能量，透過持續流動，構築出虛擬世界的運作基礎，並使資訊、影像與情緒得以被傳遞、連結與擴散。而差異在於，當演算法出現於社群媒體的運作之中，演算法決定了內容的可見性，形塑系

---

<sup>18</sup> 《白虎通·性情》，「人生於五行，性隨之偏，或仁或義，或禮或智。」

統中觀看與互動的節奏，使原本連續的能量流動被轉化為離散、可運算、可重組的計算單元，進而以碎片化的方式重新分配主體位置與自我的呈現方式，並在潛移默化中重構主體對「自我」的理解。

#### （四） 創作過程與轉向

《Self Echoes》在初期的創作計畫中，原先期望以氣與五行的框架來分析虛擬社群網路空間中的角色意識之流動型態，並進一步為其生成對應的 3D 模型，透過擴增實境（AR）技術將抽象幾何的角色意識模型投射於實體空間中。此一構想試圖延續傳統哲學中以氣與五行貫通天地、人與身體的思維範疇，將其轉譯為當代數位環境中的感知模型。然而，隨著研究與實作的推進，逐漸發覺此方法在技術發展與應用上的結構性限制。在技術嘗試中，本研究曾引用大型語言模型（LLM），在提示語（prompt）中將色彩、物件、語意等內容映射至木、火、土、金、水的分類範疇，期望以五行元素來描述使用者的意識流動。然而，五行本身作為一種範疇龐雜且高度抽象的符號分類體系，運作方式也與當今西方主流科學認知體系存在相當程度的差異，而人工智慧的訓練資料雖然幾乎包覆了整個網際網路，但其中仍然以西方科學體系為優勢主導，若直接映射至使用者圖文資料，其分析結果往往流於概括而難以中國早年的氣論與五行框架精準投射出使用者在數位環境中被拆解與再組織的主體狀態。基於此觀察，作品最終選擇回到讓語言模型直接對圖像與文字集合進行分析，並以肖像作為生成數位替身的主要依據。相較於抽象的五行映射，圖像與文字本身即是社群平台中主體被觀看、被理解與被回饋的主要形式，亦是演算法實際進行運算與分類的對象。透過此一轉向，生成結果不再試圖去重建一個整體性的意識模型，而是如實呈現主體在數位氣場中被離散與重組的狀態

因此，《Self Echoes》所關注的是，當使用者的圖像與文字被上傳至社群平台後，經歷分析、再生成與重新展示的過程中，所形成的數位自我與現實中具身主體之間存在的差距。每一次生成，皆可視為對原始資料的再書寫，使虛實之間的界線在生成式

人工智慧的遞迴邏輯概念下，層層疊加、持續推移，最終形構出一個由多重時間性與資料層次共構而成的數位主體。

### 三、執行方式

《Self Echoes》被設計為一套以網頁為操作介面的線上互動系統，使用者能在實際網路環境中，以自身的電腦、行動裝置參與作品互動，整體系統流程如 **Error! Reference source not found.**所示。在使用者輸入自己公開狀態的 Instagram 帳號後，系統將抓取該帳號近期 12 則貼文中的圖像與文字。所有影像會經過人臉偵測與分析，並透過模糊篩選與辨識機制找出主角人臉；文字部分則由語言模型分析語氣風格，萃取出語言節奏與情感基調。在整合所有圖文資訊後，生成兩項核心輸出：影像生成提示語 (prompt)、模仿語氣撰寫的文字 (caption)。接著透過文字生成影像模型產出對應圖像，最後作為使用者所見的「數位回聲」圖文將會在網頁上呈現，並依照使用者最初的選擇上傳至專屬帳號中 (@self\_echoes\_)。

為確保各模組之間能穩定運作，本作品初期先將五階段功能（圖文擷取、內容分析、文字生成、圖像生成、上傳）逐項實現後，再合併串接成一個自動化流程。中期導入簡易 HTML 介面直連本地端運行 FastAPI 的主機[92]，並用 Celery[93]與 Redis[94]將流程模組化與排程佇列。驗證流程可以依序順利運作後，後期加入狀態回報與預估剩餘時間 (ETA)，用以監控各步驟的執行節奏。穩定後以 Ngrok 將本地服務映射為 HTTPS 端點並整合 Wix Studio 前端介面[95]作為公開入口[96]。系統部署於本地端 Mac mini (M2, 16GB RAM)、Python 3.10，核心套件含 GPT、FastAPI、DeepFace、Celery、Redis、OpenCV 等，並連動使用 Draw Things 應用程式[97]進行 Stable Diffusion[98]生成。經過多次測試與修正後，系統在各階段任務間能穩定執行並保持運算流程的連貫性，完成整體互動架構的建立。

## (一) 使用者介面

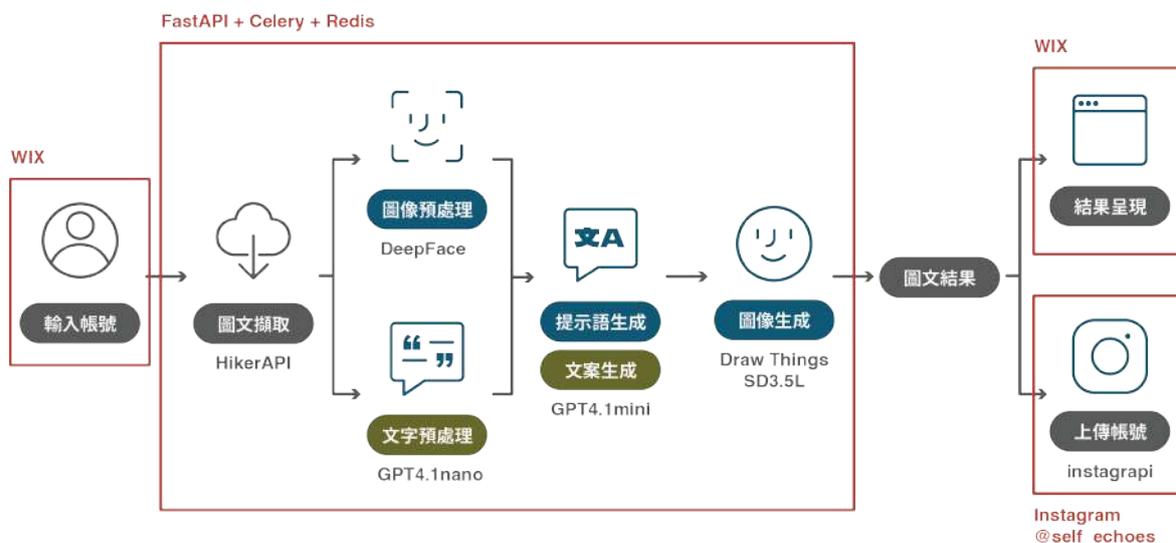


圖48、《Self Echoes》之前後端部署框架

### 1. 後端主機

如圖 48 所示，本系統採用 FastAPI 服務建立伺服器框架，並結合 Celery 與 Redis 進行任務排程與狀態追蹤，整合分析系統中的資料擷取、圖像處理、人臉偵測與圖文生成、資料上傳等模組，建構出整體互動流程。系統的每一階段皆設計成為模組化的獨立任務，透過 Celery 排入工作佇列，在理想狀態下可以讓多個用戶同時操作而不互相干擾<sup>19</sup>。任務鏈以 Celery chain 串接，依序執行擷取、分析、文字、圖像與上傳，並對同一帳號設置冷卻時間以防重複觸發造成壅塞。Redis 則作為任務快取，也用於儲存進度資訊。系統的預估等待時間 (ETA) 則是以多個歷史耗時進行平均計算。完成後端設計後，透過 Ngrok 服務將本地端的程式系統映射至網際網路中，建立成可公開存取的網址，方便使用者於網路環境中直接操作。

<sup>19</sup> 實際操作中，因為有連接至本地端的 Draw Things 應用程式進行圖像生成，因此在生成的階段中，仍需要採取單線任務佇列。

## 2. 前端介面

本系統的前端初期以 HTML 與 JavaScript 製作測試介面，用於直接與 FastAPI 後端溝通、驗證系統的各項功能，在輸入 Instagram 帳號後，系統會以 HTTP 請求方式連線至 FastAPI 的後端伺服器，啟動貼文擷取與分析流程。頁面同時會定期向伺服器查詢任務進度，並即時顯示目前階段（如「下載中」、「分析中」、「生成中」）與預估等待時間，此版本主要用於系統開發與除錯，方便在開發過程中觀察每個任務階段的執行順序與運算狀態。在系統穩定後，為了在展示中提供更友善且美觀的操作體驗，後期則將整體流程整合至 Wix Studio 平台中，以圖形化介面作為公開線上展示的正式入口，如圖 49。運用其內建的 `$w.httpFunctions.fetch` 函式與 FastAPI 伺服器 API 連線，讓觀眾能在這個圖形界面中啟動同樣的流程。使用者只需輸入帳號、選擇是否開啟自動上傳功能後，即可觸發整個生成過程。當任務完成後，系統會自動顯示生成的圖像與文字，若使用者同意上傳，則也會同時上傳至 Instagram 帳號中 (@self\_echoes\_)。

國立臺北藝術大學  
Taipei National University of the Arts

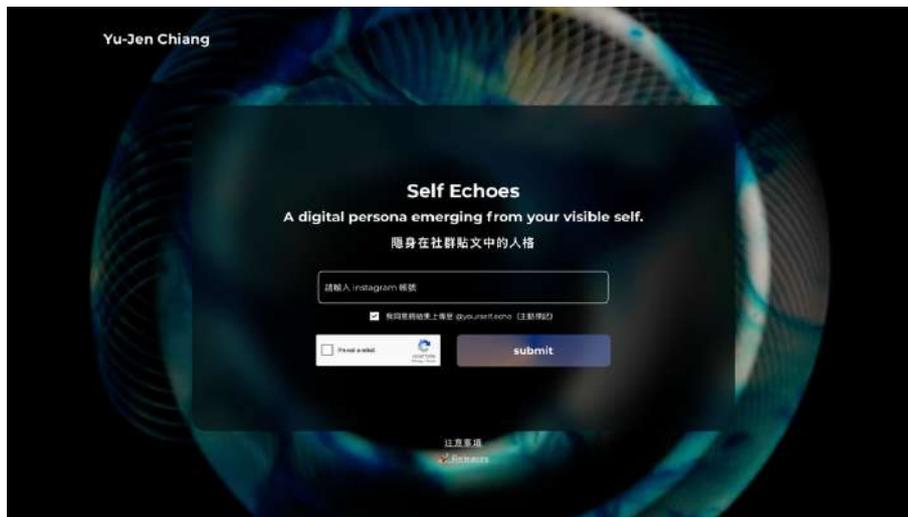


圖49、《Self Echoes》使用者介面

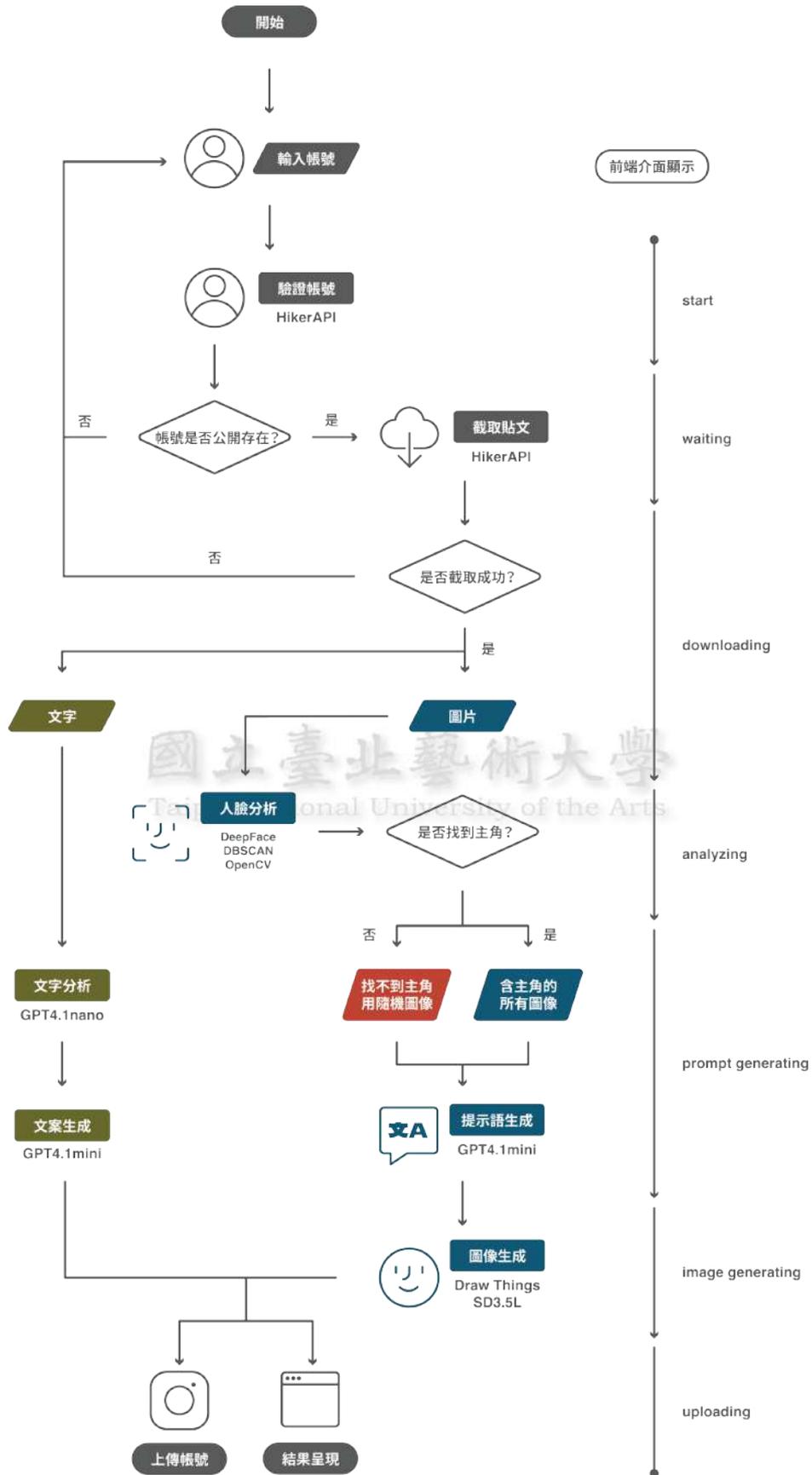


圖50、《Self Echoes》系統流程圖

## (二) AI 分析系統

分析系統以 Python 為主要開發語言，整合多個 AI 模組與影像處理工具，形成一套能自動化執行「圖文擷取→圖像預處理→文字預處理→內容分析與生成→上傳」的流程。系統整體執行時間在 10 至 15 分鐘間，也會因為不同使用者自身帳號內容多寡而有不同，表 16 為實際執行狀況下所記錄的平均耗時。

表16、五階段平均耗時表

階段	平均耗時	實測範圍
1. 圖文擷取	176 秒	108 – 203 秒
2. 圖像預處理	352 秒	340 – 364 秒
3. 文字預處理	28 秒	14 – 45 秒
4. 內容分析與生成	152 秒	147 – 155 秒
5. 上傳	27 秒	22 – 31 秒

### 1. 圖文擷取

當使用者於前端輸入 Instagram 帳號後，伺服器首先啟動第一階段任務，透過 HikerAPI[99]檢查帳號是否為有效且公開狀態。若驗證成功，系統會擷取其近期 12 則貼文中的影像與文字。所有資料均會儲存為時間序結構化資料夾，並生成中繼檔(json)記錄貼文時間、圖像 URL、文字內容等資訊，作為後續的輸入依據。

## 2. 圖像預處理（人臉分析）

第二階段負責進行影像層面的分析與特徵萃取。流程如圖 51 所示，系統會先以 DeepFace 進行人臉偵測與特徵比對的預處理，目前以 RetinaFace 作為人臉偵測器<sup>20</sup>，再以 ArcFace 對偵測到的臉提取 embedding 向量<sup>21</sup>，依據 DeepFace 於 GitHub 所公布的實驗結果，RetinaFace + ArcFace 組合的辨識準確率約為 85.2% [100]。在辨識出每張臉的性別與年齡後，利用 DBSCAN 聚類演算法<sup>22</sup>透過 embedding 向量，從每張照片出現的臉孔中找出出現頻率最高的人臉樣本 [101]，並將系統判定之主要人物標示紅色框，判定為不同於主要人物的臉孔則為綠框，再以 OpenCV 計算臉部區塊的面積以及使用 Laplacian 變異數法評估臉部影像的清晰度，篩選掉解析度不足或過於模糊的影像，如圖 52 所示，由左至右為由系統判斷之主要臉孔、主要 + 其他臉孔、其他臉孔、無臉孔。最後將所有符合標準的臉孔整合，平均所有性別與年齡資料，輸出一份性別與年齡判斷資訊。



---

<sup>20</sup> DeepFace 在人臉偵測預設使用 OpenCV，實驗中發現常常錯誤偵測，因而改用 RetinaFace。

<sup>21</sup> DeepFace 在人臉辨識預設使用 VGG-Face，實驗中發現不易控制，且考量與 RetinaFace 的相容性，最後使用 ArcFace。

<sup>22</sup> DBSCAN (Density-Based Spatial Clustering of Applications with Noise)，是一種以資料密度為基礎的分群方法。它會從資料中找出「點比較密集」的區域當成群集，而把太孤立的點視為雜訊，不需要事先設定群集數量，也能找出形狀不規則的群集。

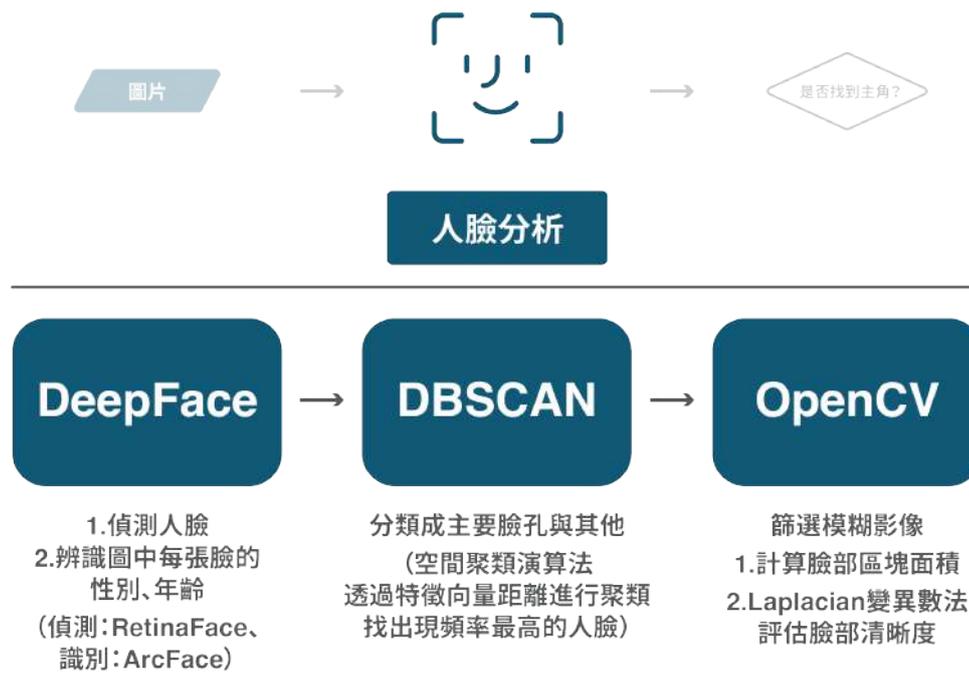


圖51、人臉分析處理流程



圖52、利用DBSCAN與OpenCV找出最主要的人臉樣本

### 3. 文字預處理

在語言部分，系統串接 OpenAI 的 GPT-4.1-nano 模型[102]分析貼文的文字，萃取出用戶在句型結構、文字長度、語氣節奏、詞彙選擇與表情符號上的習慣特徵，建立出一份針對該用戶的語言風格摘要，如圖 53 所示。此摘要將作為後續生成任務的語言基調，使 AI 能生成語句時維持與使用者貼文風格較為一致的語感。

這位使用者的貼文風格自然親切，語氣輕鬆愉快，帶有濃厚的個人色彩與生活點滴的真實感。他常用日常語句，像是「難得今天終於出門」或「在家這幾天有夠固定的作息」，還會用括弧輕描淡寫地補充細節，讓讀者感受到貼文的即時感與貼近生活的親密感，偶爾還會加入emoji來點綴，使文字更具情感色彩和趣味性。整體情緒基調偏向積極、溫暖與感性，經常流露出一種對生活細節的熱愛，以分享個人或朋友的活動、展覽、演唱會經驗來營造親近感，並透過語言展現對美好事物的細膩感受與珍惜之情，表達方式多以短句組合為主，喜歡用簡單直白的句型，讓整篇貼文看起來輕鬆又自然。关键词包括生活點滴、活動回顧、文化藝術、音樂、旅行與美食，對於展現日常細膩情感與樂趣非常擅長，也善於用貼近心靈的語言來連結讀者，使貼文既有共鳴又帶著溫馨的人情味。  
 貼文長度統計：平均 226.4 字元，最短 25 字元，最長 1241 字元。

這位使用者的貼文風格偏向自然親切、誠實真誠，語氣帶有一點幽默感，常用語助詞如「啦」、「呢」、「啊」來增添親密感和輕鬆氛圍，句型多是一段話內接續描述，較少使用複雜句式，偏向口語化的陳述。Emoji的使用適量，並且多偏向於增添情感和氛圍，例如笑臉、雨滴、動物或與內容氛圍相 matching 的表情，呈現出貼近真實生活的親和力。整體情緒基調微帶感性，略帶幽默、積極且真誠，喜歡談論旅行、生活感悟、日常點滴與自我反思，偶而有自嘲或感性句段，展現出一種輕鬆又實在的氛圍。貼文主題則以旅行、生活經驗、個人成長、心情分享為主，內容多元但始終圍繞於日常生活的點滴，透出一份對生活的熱愛與感激。  
 貼文長度統計：平均 287.8 字元，最短 43 字元，最長 518 字元。

圖53、摘要內容範例

#### 4. 內容分析與生成

此階段則由串接 GPT-4.1 mini 模型整合上述的圖文特徵資料，同時生成符合語言風格的「說明文字 (caption)」，並撰寫具體的「影像提示語 (prompt)」，輸入內容包含用戶的貼文內容(文字與篩選過的圖像)、DeepFace 影像分析結果、語言風格摘要。影像提示語會根據這些輸入內容生成包括性別、年齡、表情、姿勢、光線氛圍與拍攝語境等等資訊，運作方式如圖 54 所示，此操作意味著所有圖像會在 AI 理解後，以文字形式進行內容生成，再以此文字提示於下一步進行圖像生成，即是透過 HTTP API 功能連接安裝於本地端的 Draw Things 應用程式(圖 55)，以 Stable Diffusion 3.5 Large 模型根據設定好的參數與 prompt 生成對應的影像。

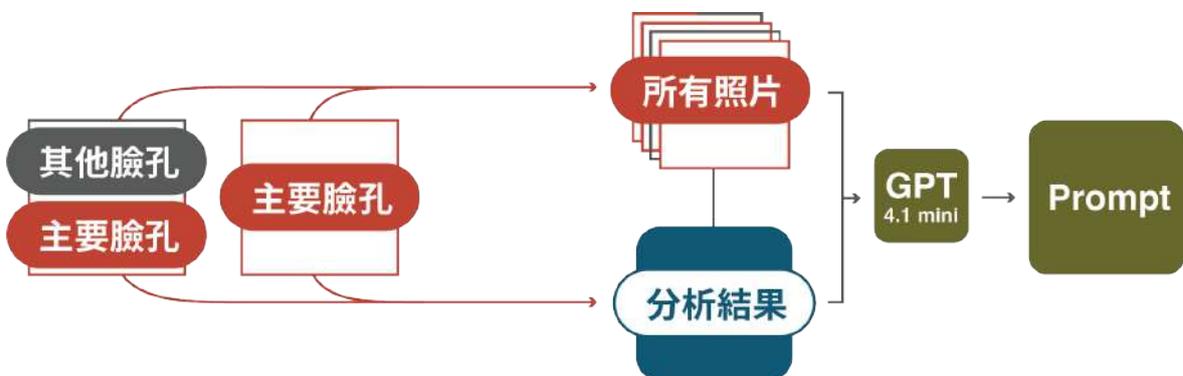


圖54、圖像生成簡易流程示意

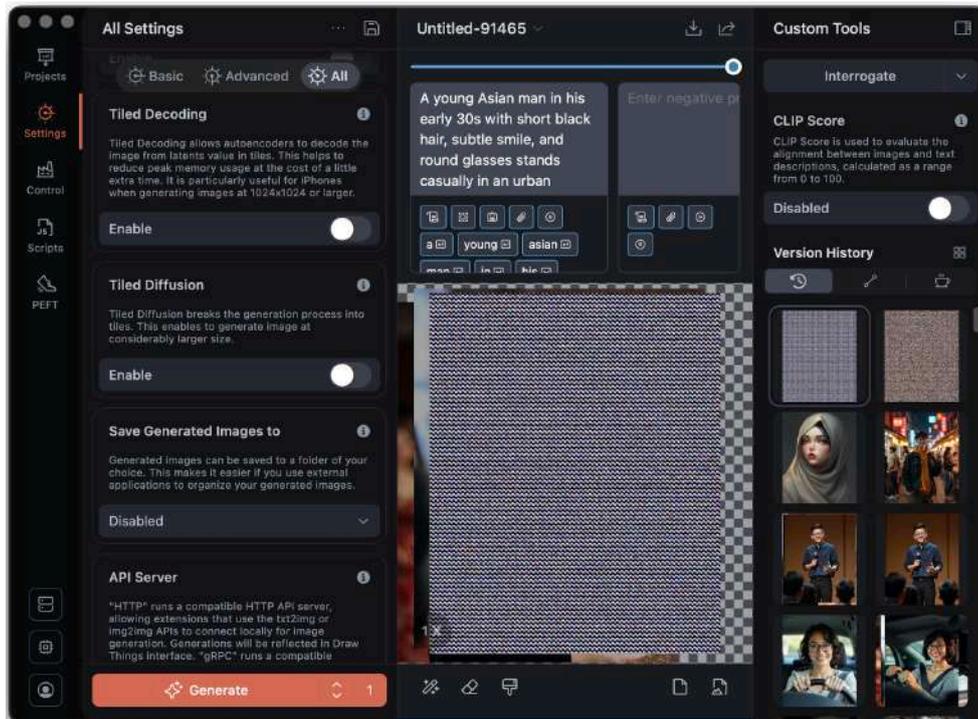


圖55、Draw Things應用程式

最初的技術開發中，以擁有豐富資源的開源模型 Stable Diffusion 1.5 進行測試，其在語意提示與構圖之間的控制相當穩定，但模型可所用的 CLIP 只能容納 77 個 token，prompt 無法詳盡描述畫面就容易被截斷，導致生成結果缺乏完整的畫面細節。於是之後因為正值 FLUX.1 與 Stable Diffusion 3.5 發表初期，陸續又測試使用了許多更大型的生成模型。然而又因為在自動化流程裡無法如同一般使用情境，能從多次的生成中挑選後細修，這裡相當仰賴有一組固定的設定參數，並在 prompt 輸入後產生單一的圖像。為了選定最適合生成模型，開發階段比較了多種開源模型與版本之輸出結果。

表 17 之測試內容以相同 prompt、相同生採樣器 (euler a trailing)、相同解析度 (1024×1280)，並以官方公布各個模型最合適的參數 (步數、文本引導 CFG) 對單一圖像生成結果進行實驗對照，生成模型之提示語 (prompt) 內容詳見附錄一。若以單次執行的生成結果來看，FLUX.1 dev 的生成表現最為優秀，而 Stable Diffusion 3.5 Medium 與 Stable Diffusion 3.5 Medium Turbo 則可能出現較為明顯的缺陷。

表17、各種開源模型之生成結果比較

模型	步數	引導	生成結果	花費時間	說明
Stable Diffusion 3.5 Large	28	3.5		14:21	<ol style="list-style-type: none"> <li>1、 臉部特徵：臉型自然圓潤且符合描述。</li> <li>2、 姿勢：自然但動作不完全符合描述。</li> <li>3、 解剖學：身型比例大致正常，但手指些微異常。</li> <li>4、 背景設定：構圖合理，符合綠樹環繞、一群人、狗。</li> <li>5、 光線氛圍：柔和但過於明亮，偏向晨光非昏光。</li> </ol>
Stable Diffusion 3.5 Large Turbo	4	0		02:12	<ol style="list-style-type: none"> <li>1、 臉部特徵：臉型略窄但大致符合描述。</li> <li>2、 姿勢：合理但動作不完全符合描述。</li> <li>3、 解剖學：身型比例大致正常，但手指些微異常。</li> <li>4、 背景設定：符合一群人、狗，但不符合綠樹環繞，構圖不合理。</li> <li>5、 光線氛圍：偏向晨光非昏光，且畫面較扁平。</li> </ol>
Stable Diffusion 3.5 Medium	40	4.5		07:29	<ol style="list-style-type: none"> <li>1、 臉部特徵：臉型略窄但大致符合描述，皮膚質感佳。</li> <li>2、 姿勢：動作符合描述，但手機處理異常。</li> <li>3、 解剖學：上下身、肩部、手指、坐姿異常。</li> <li>4、 背景設定：符合綠樹環繞、一群人，但狗沒有出現。</li> <li>5、 光線氛圍：光線處理佳。</li> </ol>

<p>Stable Diffusion 3.5 Medium Turbo</p>	<p>8</p>	<p>1.5</p>		<p>02:26</p>	<ol style="list-style-type: none"> <li>1、 臉部特徵：臉型略窄、皮膚質感佳，但眼神較呆滯。</li> <li>2、 姿勢：雙手皆持物，但兩隻手機處理異常。</li> <li>3、 解剖學：身型比例大致正常，但手指異常。</li> <li>4、 背景設定：構圖合理，符合綠樹環繞、一群人，狗，但人物異常。</li> <li>5、 光線氛圍：光線處理佳。</li> </ol>
<p>FLUX.1 dev</p>	<p>50</p>	<p>3.5</p>		<p>22:12</p>	<ol style="list-style-type: none"> <li>1、 臉部特徵：臉型過窄但大致符合描述。</li> <li>2、 姿勢：合理且完全符合描述。</li> <li>3、 解剖學：身型比例正常，手指正常。</li> <li>4、 背景設定：構圖合理，符合綠樹環繞、一群人，狗。</li> <li>5、 光線氛圍：天色、燈光處理佳，但畫面質感過於銳利。</li> </ol>
<p>FLUX.1 Schnell</p>	<p>4</p>	<p>0</p>		<p>02:29</p>	<ol style="list-style-type: none"> <li>1、 臉部特徵：臉型自然，但髮型較不符合描述。</li> <li>2、 姿勢：合理且完全符合描述。</li> <li>3、 解剖學：上下身比例異常，手指右手正常、左手異常。</li> <li>4、 背景設定：符合綠樹環繞，但只出現兩個人，卻有兩隻狗。</li> <li>5、 光線氛圍：天色、燈光處理佳，但畫面質感過於銳利。</li> </ol>

表18、各種開源模型之多圖生成結果比較

指令	SD 3.5 Medium Turbo	SD 3.5 Medium	FLUX.1 Schnell	SD 3.5 Large Turbo	FLUX.1 dev	SD 3.5 Large
一						
二						
三						
四						
五						

然而本研究最終選擇 Stable Diffusion 3.5 Large 作為最終採用模型，則是因為過程中又進一步比較了多組 prompt 生成下的綜合表現，結果如表 18，完整提示語詳見附錄二。前兩組 prompt（第一、二組）的生成結果如圖 56 所示，六個模型皆有良好的表現，能夠發現 Flux.1 系列的 Schnell 和 dev 模型，其風格與其他 Stable Diffusion 系列模型有明顯的差異。而在這組結果中值得一提的是，Flux.1 Schnell 是唯一有將第二組 prompt 中提到「一隻手拿著手機」內容完整生成的模型（圖中灰色圓圈）。



圖56、各種開源模型之多圖生成結果比較（第一、二組）

而在下兩組（第三、四組）的生成結果就能發現各個模型的明顯差異，如圖 57 所示。在第三組中，只有 Flux.1 dev 和 Flux.1 Schnell 有將 prompt 內容提到「手持麥克風」完整生成（圖中灰色圓圈），其他模型則無完整生成，然而也發現 Flux.1 Schnell 有時會額外生成一些非 prompt 內容指定的圖像元素，如第二組人像中的眼鏡（圖中黃色圓圈）；Stable Diffusion 系列則無論是哪一種模型，在手指的生成上皆有異常，尤其以 Medium、Medium Turbo 與 Large Turbo 最為嚴重（圖中紅色圓圈），Large 的異常程度則較為輕微（圖中綠色圓圈）。由此推測得知，Flux.1 系列模型不只對於人體解剖學擁有較好的掌握，對於 prompt（尤其中後段內容）也有較完整的理解。



圖57、各種開源模型之多圖生成結果比較（第三、四組）

第五組的生成結果如圖 58 所示，一樣只有 Flux.1 dev 和 Flux.1 Schnell 有將 prompt 內容完整生成（圖中灰色圓圈），其他模型則都沒有將「水缸」生成出來，但在這一組結果中「微笑的墨西哥鈍口螈」並沒有被正確理解，反而是以其他生物形象出現在畫面中。在前面比較中表現十分優異的 Flux.1 dev 在這組測試中異常的出現了非寫實畫風（圖中紫色圓圈）；對於解剖學有良好掌握的 Flux.1 Schnell 在腳趾的生成上也出現了異常；Stable Diffusion Medium 和 Medium Turbo 兩個模型各自在臉型和手指上出現異常。相較之下，Stable Diffusion Large 和 Large Turbo 兩個模型除了沒有將水缸完整生成外，則沒有出現其他大問題。



圖58、各種開源模型之多圖生成結果比較（第五組）

指令	SD 3.5 Medium Turbo	SD 3.5 Medium	FLUX.1 Schnell	SD 3.5 Large Turbo	FLUX.1 dev	SD 3.5 Large
一						
二						
三						
	出錯機率高且嚴重		臉型雷同	出錯機率高 構圖雷同	臉型雷同	會出錯， 但相對穩定
四						
五						

圖59、各種開源模型之多圖生成結果比較

綜合比較至目前，出錯機率高且嚴重異常的 Stable Diffusion Medium 和 Medium Turbo 是首先需要被排除考慮的選項；Flux.1 dev 模型則是除了偶發地出現一次非寫實畫風外，其他生成結果中一直都有相當好的生成品質，其和 Flux.1 Schnell 兩個模型在身體架構、手指的處理上也是其他模型無法比擬的，然而，如果進行多圖比較則會發現，或許是訓練內容的不足導致其生成出的人像皆有著過於相似的問題，雖然幾乎所有模型都存在類似問題，但在 Flux.1 上此問題更為明顯。相對而言，Stable Diffusion 3.5 Large 與 Stable Diffusion 3.5 Large Turbo 則擁有較高的臉部特徵多樣性，較少出現相同人臉的問題。然而，除了身體結構、手指等解剖學的異常，Stable Diffusion 3.5 Large Turbo 生成出的圖通常會以正面平視的視角呈現，有構圖過於雷同的問題。依上述情況綜合而言，這些模型在經過幾次的測試比較後，Stable Diffusion 3.5 Large 似乎是現階段較為合適的模型（圖 59）。

另外，若再將生成的時間納入考量，在現有的硬體限制下，Stable Diffusion 3.5 Large 平均單圖生成時間為 15 分鐘，而使用 Stable Diffusion 3.5 Large Turbo 雖然只需平均 3 分鐘的時間，卻又面臨構圖單調的問題。在測試了 Stable Diffusion 3.5 Large 原版模型加上 Tensorart 發布的 TensorArt-TurboX-SD3.5Large-8steps LoRA 模型[103]之後，從結果（表 19）可以發現，雖然生成品質和細節雖然皆不如高步數的原版模型，但在節省了超過三分之二的時間下，仍相當程度的保留了原始構圖，能在人物情緒、光線氛圍以及生成時間取得相對折衷的平衡，因此此方案被採用為系統中主要的圖像生成引擎。

表19、Stable Diffusion 3.5 Large與使用LoRA生成結果比較表

模型	LoRA	步數	CFG	生成結果	花費時間
Stable Diffusion 3.5 Large	無	28	3.5		14:21
Stable Diffusion 3.5 Large	Tensorart TurboX sd3.5L 8steps	8	1		04:30

## 5. 上傳與展示

最後生成的圖文結果會傳送至前端網頁上顯示，如圖 60，同時，此結果也會依照設定好的版面位置製作成海報並提供檔案的下載連結，如圖 61。此外，系統會另根據使用者最一開始的設定，決定是否將生成的圖文內容上傳至 Instagram 帳號 (@self\_echoes\_)，並由 instagrapi API[104]負責上傳任務，上傳結果於帳號中的呈現如圖 62。



圖60、《Self Echoes》結果畫面



圖61、《Self Echoes》結果海報



圖62、帳號 (@self\_echoes\_) 畫面截圖

### (三) 例外處理

為提升穩定，系統設計例外處理的機制，如遇私密帳號則提示檢查或更換帳號；未偵測人臉時則取消預處理，回退使用所有原始資料；外部 API 失效時，回報並記錄錯誤，供後續維護修正。

#### 四、作品呈現

如同前面章節所述，《Self Echoes》是一個不具有實體形式的線上互動系統，使用者在能夠連接網路的條件下，可以以自身的電腦、行動裝置參與作品（圖 63）。其在呈現上，作品設計一套網站與自動化系統，讓每個使用者在網頁上可以進行分析和生成，介面截圖如表 20 所示，生成結果除了能在網頁上顯示，也同步匯集在新創建的 Instagram 帳號中作為公開展示。此系統上線期間為 2025 年 4 月 21 起三個月，此期間網頁造訪人次為 252 人次，系統執行次數為 148 次，實際參與資料上傳有 45 位使用者。此外，本作品曾另以研討會海報論文形式，投稿第八屆科技藝術國際學術研討會（ISAT 2025），並於 2025 年 7 月 13 日在國立臺灣大學進行發表（圖 64）。

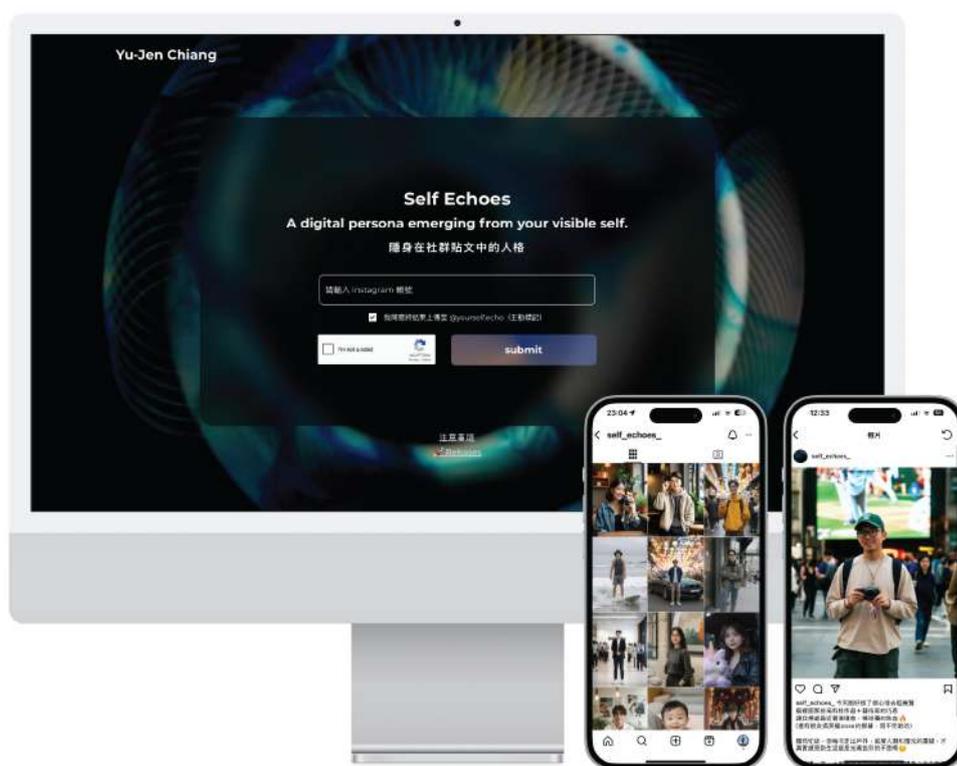


圖 63、《Self Echoes》作品示意圖

表20、《Self Echoes》網頁截圖

<p>初始畫面：注意事項</p>	<p>輸入畫面</p>
	
<p>(1/5) 下載中</p>	<p>(2/5) 分析中</p>
	
<p>(3/5) prompt 生成中</p>	<p>(4/5) 圖像生成中</p>
	
<p>(5/5) 上傳中</p>	<p>結果頁面</p>
	



圖64、《Self Echoes》海報論文於ISAT 2025進行發表

## 五、使用者體驗回饋

在《Self Echoes》的使用者問卷中，共回收共計 17 份有效樣本，受試者年齡以 25–34 歲為主 (15/17 人)，生理女性 9 位、生理男性 7 位、1 位不願透露。問卷涵蓋社群媒體使用習慣、情緒感受、自我辨識、虛擬替身認同與系統操作五個面向。絕大多數受試者表示自己「很常使用 Instagram」(14/17 人)，少部分受試者則是「偶爾使用」(3/17 人)，因此可理解此受試群體對自我呈現與社群平台文化具有高度熟悉度。

在「情緒感受」上，結果如圖 65。多數使用者對生成的貼文表現出「新奇的」(平均值 4.00 分)、「有趣的」(平均值 4.06 分)、「喜歡的」(平均值 3.53 分)與「接受的」(平均值 3.59 分)等正向感受，而負面情緒如「疑惑的」(平均值 1.88 分)、「不安的」(平均值 1.29 分)、「詭異的」(平均值 1.29 分)、「沉重的」(平均值 1.18 分)、「焦慮的」(平均值 1.12 分)、「討厭的」(平均值 1.12 分)等反應則明顯偏低，因此生成內容整體上並未引起排斥或恐懼。

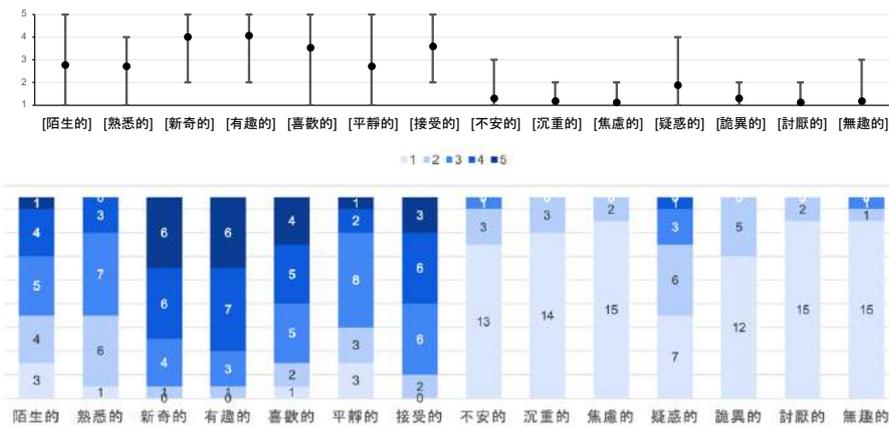


圖65、使用者問卷「情緒感受」統計圖表

針對圖像與文字的內容滿意度，結果如圖 66。使用者對「衣著」與「場景」整體評價較高，平均值分別為 3.65 分與 3.53 分，而人物本身的樣貌平均滿意度為 3 分。文字部分方面，「語調」（平均值 3.18 分）與「用詞」（平均值 3.24 分）。進一步觀察「自我辨識」的面向，受試者認為生成圖像與自己上傳習慣相似度之評分為 2.82 分，文字語氣為 2.88 分、文字內容為 3 分，整體相似度自評分數皆居中。此一落差代表生成結果雖能在風格與語言層次引發辨識，卻仍與使用者對自身的完整認知存在距離。

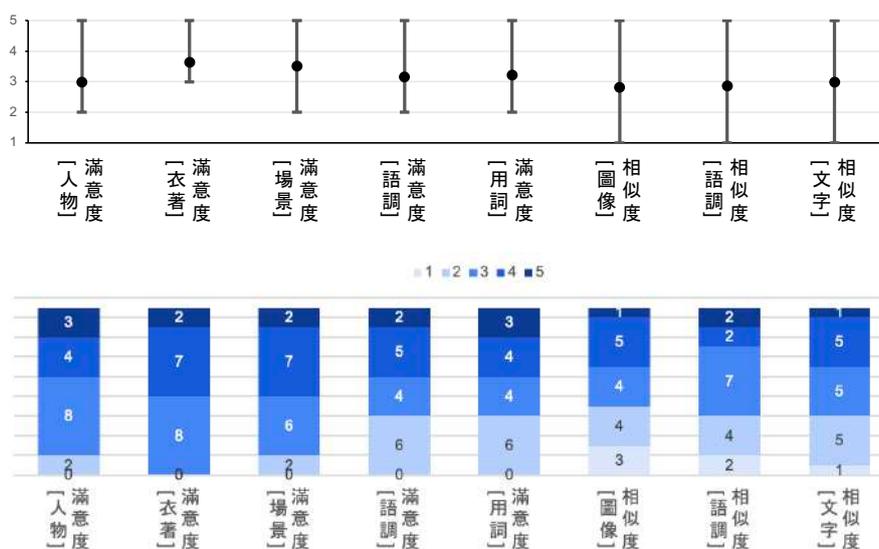


圖66、使用者問卷「滿意度與相似度」統計圖表

在「虛擬替身認同」方面，結果如圖 67。使用者多認為目前表現出來的替身人格特質偏向「一個虛構的我」(10 人)，其次為「期待別人看到的我」(6 人)，值得注意的是，「真實生活中的我」、「學我風格的陌生人」、「無法辨認的陌生人」各自只獲 1 人選擇，無人選擇「網路上的我」作為當下的替身認同；多數人對未來替身的期待為「網路上的我」(9 人)、「期待別人看到的我」(6 人)、「一個虛構的我」(3 人)與「真實生活中的我」(2 人)，由此可推斷，現階段的替身傾向被理解為一種具有展示性、想像性的自我延伸。

此外，使用者在自我反思的相關題目中，對於「展現了自己平常較少展現的性格」、「讓我對自己有新的發現」、「是否願意讓替身幫忙代言」結果皆呈現相當兩極化的分布，顯示部分使用者在自我呈現與行動上仍具較保留的態度。綜合上述使用者經驗的結果，《Self Echoes》所生成的數位替身現階段被多數人視為「一個虛構的我」與「期待別人看到的我」，並未被視為對具身自我的取代，而是一種與之並存的資料化主體形式。此結果呼應凱瑟琳·海爾斯在《後人類時代》中所提出資訊模式仍需依附於具體媒介與身體經驗的「具身虛擬性」觀點，無法脫離具身條件而獨立存在。

因此這也讓我思考，或許《Self Echoes》的後續發展不應再繼續試圖生成「更真實的我」，而是轉向如何讓使用者在當代主體在虛實共構條件下，感知、辨識自身的資料化版本，進而反思「我」在當代資訊環境中，究竟是如何透過資料化、演算與觀看機制被持續生成與理解。

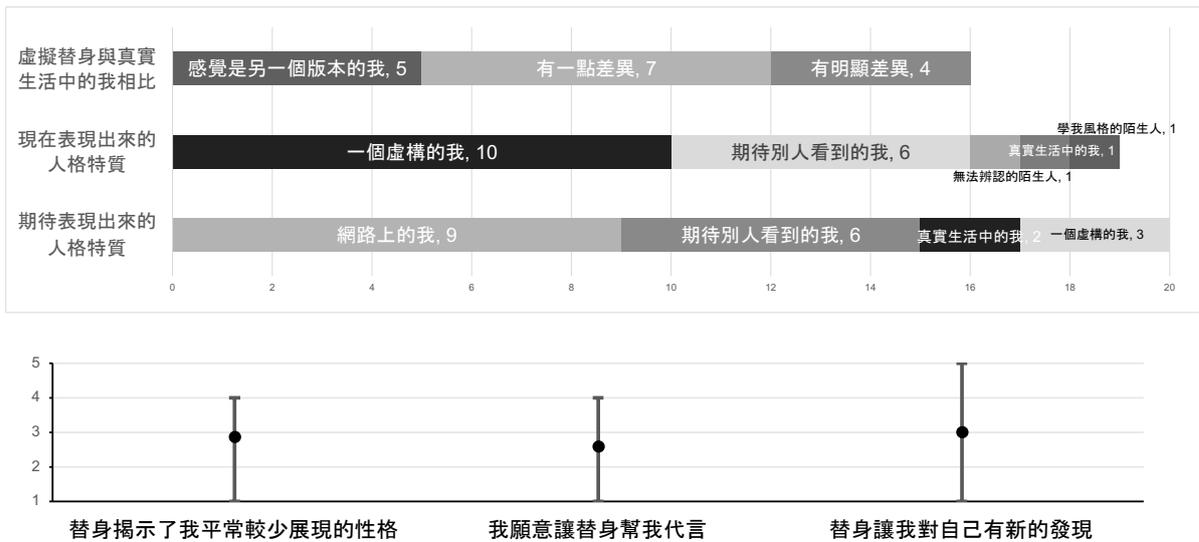


圖67、使用者問卷「虛擬替身認同」統計圖表

## 六、限制與改進

在圖像預處理的 DBSCAN 演算法設定中，鄰域半徑 ( $\epsilon$ ,  $\epsilon$ ) 為 3.5，最小出現次數為 2，結果如圖 68，左圖為  $\epsilon < 3.5$ ，中圖為  $\epsilon = 3.5$ ，右圖為  $\epsilon > 3.5$ ，然而在不同的照片組中最準確的  $\epsilon$  也會有些微不同，因此在現階段的設定中，實際能找出正確使用者人臉的準確度僅為 65%，這也造成系統後續進行年齡判讀時，可能因為誤讀其他臉部資訊而造成誤差。實際結果誤差約為 5 歲，略高於 DeepFace 公布年齡誤差的  $\pm 4.65$  MAE。



圖68、圖像預處理不同參數之設定結果

而且，在目前的版本中，系統針對單一帳號的完整流程約需要 10 至 15 分鐘，由於圖像生成階段需連接至本地端的 Draw Things 應用程式執行，系統雖然以多進程模型為目標進行開發設計，但現階段仍需採取單線任務佇列運作。因此當多人同時操作時，等待時間往往會隨著佇列長度而成倍數增加。雖然在開發後期新增了進度查詢的功能來回應等待過久的問題，但運算效率仍舊是個無法被忽視的瓶頸。未來可考慮以線上伺服器的方式進行改善，確保多用戶操作下的整體流暢度。

在生成模型的應用上，現行版本所採用的影像生成模型仍存在一定侷限，上一部分的使用者研究中也顯示出此限制。正好，2025 年 8 月 Black Forest Labs 與 Krea AI 合作發布了 FLUX.1 Krea [dev] 模型，在生成的真實感與人物表情姿態上皆展現了更高的可用性。我以現有分析模組重新對個人帳號中的貼文進行再生成測試，發現該模型在語意理解與圖像表現上能更準確地反映提示內容，測試結果請參閱表 21。若能將此模型整合進系統流程，並且在花費時間上取得平衡，將有助於提升生成內容的精緻度。

在作品呈現方式上，《Self Echoes》目前以網頁系統的形式運作，目的是讓使用者能夠直接參與。然而此種形式使作品在結構與介面上更接近一個功能性的應用系統，其作為藝術創作的詮釋性將會是未來的改進方向，思考如何透過文本與呈現方式的重新設計，使作品的焦點從分析與生成的演算方法轉向人工智慧介入社群媒體的詮釋，使《Self Echoes》成為更深入探討「數位自我」與「演算法觀看」的創作實踐。

表21、SD 3.5 Large與FLUX.1 Krea模型之多圖生成結果比較

模型	SD 3.5 Large	FLUX.1 Krea
一		
二		
三		
四		
五		

## 第四章 作品概念與形式討論

### 第一節 《National Freeway 5》：從線性到綿延的堆疊

《National Freeway 5》雖以在國道上高速行駛的行車畫面為主要創作內容，高速公路在現代社會中是追求效率與物理時空壓縮的展現，目的在於使人以可行的最短時間穿越空間，使身體被置於標準化與高度線性的時間框架之中。但此作品最主要是想傳達的除了直觀上線性物理時空被現代建設所壓縮之外，更想傳遞身體在生命經驗中的主觀感知時間，其中包括身體載具中進行移動時，意識介於清醒與出神之間，形成具有延展性的模糊時間狀態。

在環境景觀高度重複且單調的行駛過程中，反覆的隧道結構、均質的光線、穩定的速度，使得意識不再精準追蹤物理時間的流逝，此情境也讓我們能夠喚起身體最直覺對時間流逝的知覺。作品為了呈現這種主觀的感受時間，以堆疊的方式將感知視覺化，引用柏格森對於綿延的定義，生命時間是具有連續性的意識經驗，當影像在不斷地進行堆疊時，我們所見的不再只是單一且侷限於當下的隧道景觀，而是許多過去存在的影像在當下畫面中共同顯現的狀態，使觀者得以在單一畫面中同時感知過去殘留、當下瞬間與未來趨向。若以胡賽爾提出內在於意識時間的三分結構作為對應，在作品中段，那些看似模糊的影像殘影正是過去瞬間在意識中的滯留，而畫面末段隧道逐漸凝聚的遠方光點，則象徵著駛離隧道的前瞻意識。影像堆疊使這三種意識相互交織，讓時間經驗共存於在同一視覺平面中。

從具身性的角度來看，作品也試圖重新聚焦到身體在經驗時間時的重要性。雖然高速移動通常會讓我們的感知變得遲鈍，但作品所呈現出模糊卻可辨識的光線、隧道輪廓，以及因速度而被拉伸的視覺節奏，都是想讓觀者意識到自己是帶著身體在經過這段路，而非只是單純地在觀看一段具功能性的行車畫面。

若再從羅薩索分析的社會加速角度來理解，作品或許也反映了現代生活中時間被急速壓縮的共同經驗。現代社會在交通、科技與生活節奏上皆不斷追求「更快速」的狀態，高速公路或許就是這種加速邏輯之下的其中一個產物。在加速的生活裡，人們逐漸難以在日常經驗中找到停頓或感受時間厚度的時刻，時間似乎只剩下功能性的計算與行程安排。而此作品讓高速公路這個功能性空間從「需要快速穿越的場域」變成「值得被感受的場域」，從高度線性的時間中抽離，重新找回身體感知的節奏，使行車影像從單純的交通導向，成為重新審視時間、身體與世界關係的契機。

## 第二節 《The Mirror》：從鏡像到演算法凝視的堆疊

《The Mirror》以雙面鏡、光線切換與即時影像讓使用者在同一面鏡子前面經歷兩種不同的觀看方式，分別為觀看光學反射的鏡像，以及由即時系統運算出的臉部集合影像，並用此兩種交替的觀看模式討論監控與隱私之間觀看與被觀看的關係。反射鏡像可藉拉岡的鏡像理論作為對應，在此理論中主體會藉由鏡子所見的完整形象來辨認自我並建立主體性，作品以此作為個體從鏡像中觀看自我的隱喻，也讓作品的使用者以看見「我」的影像作為互動起點。然而在光線切換後出現的透視影像，是一張不屬於單一個體，而是由演算法運算多張臉部碎片的平均面孔，這使使用者在鏡中面對的是一個混合著他者與自我的模糊影像，進而從鏡中辨認影像裡某一個部分的「我」。

若以觀看的權力關係切入，傅柯提出的全景敞視概念也能作為另一種解讀，個體在可能被觀看的情境下會進入自我的規訓。在《The Mirror》的互動機制中，使用者會在觀察到鏡中影像有著自己留下的痕跡後，才會意識到自己可能正在被記錄、分析，進而進入上述自我規訓的情境。使用者也會在這之後，使用者也會不再把這面鏡子單純地視為日常的反射裝置，而是將其視為一種監視設備的象徵存在。這也呈現出當代監控中監視者不再可見的特徵，使用者在鏡前看不見他人，也看不見攝影機，即便如此，仍會對鏡中的影像賦予超越一般「影像」的意義。系統的運算邏輯以及會記錄使

用者的可能性讓鏡子因此成為一個擁有隱性權力的介面。另一層面，使用者在鏡中看見的也並非只是單純的他者視線，而是一種在自動化運算邏輯下的複合產物，此影像也可以對應理解為法洛奇所提出，當代許多用於分類、比對、控制，不為人類觀看而生的「機器視覺」。在如此的延伸下，觀看與被觀看的權力不僅持續存在，也甚至因為演算法的介入而變得更加無所不在。

### 第三節 《Self Echoes》：從具身自我到虛擬替身的堆疊

《Self Echoes》將符號堆疊的實驗場景，從時空感知轉移到主體意識的重構。作品透過擷取使用者在社群平台上原始具備身體性的貼文、照片、語言風格與臉部特徵，並將其生成式 AI 系統中轉換為演算法可處理的資料，再造生成為一組具有堆疊意義的綜合性的虛擬替身。生成式 AI 的運作與遞迴邏輯概念相關，即當前狀態依賴先前狀態；而在此生成機制中，堆疊可被視為遞迴運作所外顯的可見現象，表現為先前狀態被保留下來並層層累積。本研究以「資料化主體」討論這個在後人類語境之下，主體如何在虛實交織的環境中被重新定義，作品所使用的資料源自於社群平台，與韓炳哲所描述的「自我揭露」現象密切相關。在社群媒體環境中，個人透過照片、語言與形象管理，展演一個更接近理想化，接近如高夫曼所說的「前台」形象，只不過在資訊時代，前台極可能又被碎片化成其他各種大大小小的前台，也就是如特克所指稱的模樣。這些自我呈現成為塑造不同數位身份的重要機制，除了構成了他者的社會凝視，也同時表現出演算法觀看下的機器視覺。

若以海爾斯的框架來理解，她認為真正的主體性是由具身性與資訊模式共同構成，而《Self Echoes》的運作方式也仰賴一種從具身性到資訊模式的轉換，從使用者回饋中的結果，此作或可視作是一種「具身虛擬性」為何重要的實例，也由於內容的生成純粹由生成式人工智慧依照其分析出的資訊模式堆疊而成，因此不可避免地帶有福克斯所批判的去具身化傾向。使用者的真實身體在轉換後，其身體所留下的影像、語氣、

文字節奏都被抽取為可運算的模式，再被堆疊組構為一個脫離身體、卻由身體資料所生成的數位形象，但作品的產出在其生成的事實與意義上仍依附於身體曾留下的媒介記錄，也因此具有某種與真實自我相連的存在方式。在此脈絡下，作品所呈現的數位自我並無法完整代表真實自我的再現，而更像是在資訊環境中生成、並被演算法強化的一種新的前台自我。虛擬替身是使用者的部分延伸，但同時也是演算法在資料堆疊下重新定義出的另一種可能的「我」。

#### 第四節 時間、空間、主體意識的平行比較

本節將從形式分析的角度，對三件作品置於同一組分析框架中進行比較與對照（圖 69），也藉此說明虛實共構之下的時間流變、空間轉換與意識生成機制。



圖69、時間、空間、主體意識發展框架

##### 一、時間層次的流變

三件作品在時間的構成上形成一條由「感知當下」到「疊加當下」，再到「再造當下」的發展軸線。在《National Freeway 5》中，時間呈現為一種身體與感官共同構成的感知當下，作品以影像堆疊讓物理時間被壓縮，將線性流動轉化為一種主觀的、

身體性的時間經驗。當影格在同一畫面中累積顯現，當下不再是單點瞬間，而成為包含過去殘留與未來趨向的延展狀態；《The Mirror》則將時間推向一種由反射與演算共同生成的疊加當下。鏡中的影像累積了過往使用者臉部殘影的演算結果。當新的面孔出現在鏡前時，系統在背景中將使用者自我的當下（被捕捉的影像）將同時與他者的當下（過往使用者）堆疊，在光線調動的過程中，再與現在（鏡面反射）一同出現，這種時間結構使鏡子成為交會點，每一位使用者的臉既屬於自己的當下，也被納入他人的時間中；《Self Echoes》則進一步使時間成為可被資料化處理與生成的再造當下。作品透過人工智慧將使用者過去分散於不同時間點的圖像、文字內容轉化為可運算的資訊素材，再經由生成系統重組為新的貼文，使時間成為可被「再造」的資訊流。這些生成的圖文內容並不屬於某一個真實確定的瞬間，而是由過去的數據片段所重構（演算、壓縮、重組）出既源於真實又呈現虛擬的現在。

## 二、空間結構的轉換



三件作品在空間上分別透過不同媒介建構出從實體到虛擬、從可觸及到抽象的空間變化。《National Freeway 5》的空間起於具體物理環境，但在作品中被轉化為一種感知、記憶的經驗場域，影像空間的壓縮堆疊使原本封閉的車道被抽象化為光與速度的流動軌跡，觀眾從觀看具體的地景，轉變為經驗一個由影像運動與聲音迴盪構成的具身感知空間；《The Mirror》的空間結構以鏡面作為核心媒介，在虛實之間，鏡面既屬於反射現場的裝置物件，也是顯示數位影像的載體。反射現實中的使用者的同時，也疊合了使用者與過往使用者留下的影像，使鏡面成為一個同時包含現場身體、自我影像與他者影像的反射介面，一個不斷變動更新、多重視線交會的空間。而當使用者意識到影像痕跡的累積與被分析的可能性，鏡前的場域便由反射空間轉化為具有權力關係的監控場域；至於《Self Echoes》則由社群平台展開空間資料化的轉換。使用者的貼文原本嵌在日常時空中，但當作品擷取這些內容進行分析時，原始情境被抽離並

被轉寫為資料結構；最終生成結果回到平台呈現時，觀看者面對的已是資料被再構成內容的場域。

### 三、主體意識的生成

主體意識在虛實之間的轉化，使「我」能夠同時是感知者，也同時作為被系統觀看、記錄的角色，甚至再推向可被資料化生成的替身存在。在《National Freeway 5》中，主體主要仍是具身感知主體。觀者在觀看中被引導意識我正在以身體穿越，並且進入一種由光影與聲音構成的感官流動；《The Mirror》中的主體意識則同時賦予具身主體成為被凝視的主體。使用者在鏡前成為影像的一部分，其臉部即時被捕捉、分析、混合於他人影像之中。此時的「我」從一個進行觀看對象的主體，變成一個被「我」觀看的對象。使用者既能在鏡中看見自己，又能意識到有個不是「我」的影像，在系統機制設計中，使用者又或將發現不是自己的影像，其實有著自己的存在，這個存在使使用者成為了一種被感知的主體；在《Self Echoes》中，主體意識更進一步轉化為資料化生成的替身主體。使用者過去的貼文與影像資料被抽取成資訊模式，再被 AI 生成為新的圖文內容，此時的「我」跳脫出了單純進行感知或被感知對象的主被動關係，而是以替身形式成為被構造出來的被生成者。他既來自真實的行為痕跡，又脫離原本的身體與意識，成為介於自我與他者、真實與虛擬之間的模糊存在。使用者在觀看生成貼文時，同時經驗到那既像是自己，又被演算法重組而不再是自己的陌生感。

### 四、小結

綜觀三件作品可見，時間、空間與主體三者交會出虛實共構的時空意識。從《National Freeway 5》的感知經驗，到《The Mirror》的互動觀看，再至《Self Echoes》的生成回聲，作品中的虛實轉換不僅重塑了觀看或參與作品的方式，也各自定義了自我的存在狀態。

## 第五章 結論

本研究以「虛實共構的時空意識」作為整體命題，研究中提出的三件藝術創作實踐，雖各自擁有不同的探討面向與創作媒材，卻都圍繞在四個相互交織的共同核心概念——「虛實交錯」、「時空意識」、「符號堆疊」與「模糊性」。這四個概念構成了本研究的主要架構，「虛實交錯」為我們經驗的「時空意識」構成了現象的場域；「符號堆疊」成為了創作的方法或路徑；「模糊性」則作為對感知經驗的詮釋。也就是如同緒論中所說，本研究主要探討的是，在當代虛實交織的環境中，各種符號與媒材如何在時間、空間與主體意識三個層次上，堆疊出我們對自我身體、記憶、身份感知與認同的理解。

### 第一節 研究貢獻

本研究的主要貢獻在於將語言學中用於溝通目的的「符號」與電腦科學或資料處理的「堆疊」（嚴格意義之下，三件作品的操作更接近平均、積累、遞迴，而本研究以視覺感知的角度，將此種重疊狀態稱之為堆疊），轉化為一種具有時空深度，可被實踐、可被感知、可被反思的經驗結構，並嘗試使「符號堆疊」成為讓觀者與使用者進行時空感知的創作方法。

### 第二節 符號堆疊下的當代感知狀態

第二章所提出的文獻背景脈絡，為三件創作提供了重要的理論基礎，現象學的身體感知論述使本研究可以在《National Freeway 5》中討論時間的延展性與意識的流動；鏡像理論與觀看權力的論述則提供本研究對《The Mirror》為自我與他者的互視結構進行詮釋；資訊哲學、資料化主體與後人類觀點則為本研究的《Self Echoes》提供身

體在資訊流與演算法中如何發生主體性改變的理解。這些理論不僅協助發掘與理解作品的概念架構，也支持本研究提出「符號堆疊」作為理解虛實共構時空意識的創作方法，也使「符號堆疊」可以作為我們如何重新理解「自身」與「世界」關係的回應，其所指的並不只是外在環境的變化，也同時是一種關於我們如何感受到時間、如何在空間中定義自己存在，以及如何在世界中感知自我的各種深刻經驗。

三件作品的創作時序，正也呈現出從內而外，自下而上的發展脈絡。《National Freeway 5》以具身主體的視角觀看與感受周圍的現實環境帶來的時空回饋；《The Mirror》讓主體透過實體物件，使他者影像介入互動，從而在鏡像與監視器構成的視覺結構中，意識自己作為被觀看主體的角色；《Self Echoes》則進入虛擬網路世界，讓使用者的影像與文字被系統與演算法轉化為可運算的符號，經由重新排列、生成，使自我呈現成為一種不斷被非實體他者塑形的過程。由最初單一影像的時間延展，逐步轉向多的影像之間的時空關係，最終進入虛擬世界由系統觀看下的影像結果，如此也顯示，當代的感知不再只是在人際間、在實體環境中與各類符號的互動關係，科學技術與演算系統也將比以往更多地參與人類的知覺系統，呼應海爾斯所提出，具身虛擬性中具身性與虛擬性的共伴關係。

### 第三節 研究限制

然而目前本研究也存在一些限制。其一，作品中目前使用的人工智慧與演算法技術在其應用上具有一定的不透明性，資料處理與影像生成過程可能受到模型偏差的影響，無法完全掌控其輸出結果。其二，本研究以三件創作為案例，雖呈現了虛實共構經驗的多個面向，仍然難以全面涵蓋虛實共構經驗的所有範圍，未來可延伸至更多技術媒介或不同互動形式的探討。其三，研究對於使用者經驗研究所獲得的研究樣本數有限，後續研究可加入更多參與式觀察或較為深入的訪談回饋，以擴充使用者在虛實交錯中的主體經驗的分析與理解。

#### 第四節 符號堆疊之後的感知想像

若從本研究三件創作實踐的發展脈絡與其所呈現的趨勢來看，可以將符號堆疊理解為一種在當代文化與科技條件下逐漸浮現的感知狀態。隨著數位技術與運算條件持續進步與發展，過往可能隨時間消散的行為痕跡、影像紀錄與語言表述等等，愈來愈容易以資料的形式被保存、累積並且長期留存。當這些資料在未來持續存在，並被生成式人工智慧或各類演算法系統根據個體所留下的資料痕跡加以分析、重組與生成，進而於不同時間與空間條件下回饋至主體經驗之中時，堆疊將可能不再只是畫面效果，而是主體經驗的生成機制。若個體的感知與表述被轉寫為可運算的符號單元，再以新的排列方式回到觀看者面前，這些影像、文本或其他轉譯形式，則可能同時具有「曾經經驗過」與「未曾親身經歷」的特性，主體所經驗到的究竟是自身記憶的延伸還是他者觀看的回饋，是真實還是虛構，或許其邊界也會比我們現今的認知還要更加模糊。

在這樣的感知想像中，生成式人工智慧與演算法系統或可被理解為不再只是進行書寫、分析、創作或對話的輔助工具，而可能成為不斷進行保存、轉譯、回饋的結構，甚至逐漸與當代影像、聲音等五感知覺體驗技術結合，成為承載感知殘留的裝置，抑或是在生理或心理層面融入我們的身體或心靈。這種經由外部系統保存、回饋的感知結果，或可被類比為如同《哈利波特》中用於保存與回顧知覺的「儲思盆」(Pensieve)，成為能夠承載過往經驗、並在他時他地被再次調用的媒介。然而，呼應本研究中《The Mirror》提出的議題，如果這些感知經驗的提取與回饋並非完全受個體意志所掌控，而更多可能受到商業邏輯或演算法系統條件的影響時，各種符號堆疊出的形貌或認知或將間接成為一種在當代感知經驗中潛在的不確定因素。

國立臺北藝術大學  
Taipei National University of the Arts

## 參考文獻

- [1] Ferdinand de Saussure, Trans. by George Wolf. *Saussure's First Course of Lectures on General Linguistics (1907): From the Notebooks of Albert Riedlinger*. (1st ed) Oxford [England]: Pergamon, 1996 Available: <https://worldcat.org/title/34473717>.
- [2] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to algorithms," 2009. Available: <https://worldcat.org/title/676697295>  
<http://www.books24x7.com/marc.asp?bookid=49924>.
- [3] Henri Bergson and Frank Lubecki Pogson, *Time and Free Will: An Essay on the Immediate Data of Consciousness*. George Allen & Unwin, 1910.
- [4] E. Husserl, Trans. by James S. Churchill. *The Phenomenology of Internal Time-Consciousness*. Bloomington: Indiana University Press, 1964.
- [5] M. Merleau-Ponty, *Phenomenology of Perception*. New York: Routledge, 1945.
- [6] T. Fuchs, *In Defence of the Human being: Foundational Questions of an Embodied Anthropology*. OUP Oxford, 2021 Available: <https://books.Google Scholar.com.tw/books?id=fS5CEAAAQBAJ>.
- [7] S. Gallagher, *Action and Interaction*. Oxford University Press, 2020 Available: <https://books.Google Scholar.com.tw/books?id=h8jXDwAAQBAJ>.
- [8] H. Rosa, *Social Acceleration: A New Theory of Modernity*. Columbia University Press, 2013.
- [9] Nicholas Mirzoeff, Trans. by 徐達艷. *如何觀看世界*. (初版) 中國上海: 上海文藝出版社, 2017 Available: <https://worldcat.org/title/1050523989>.
- [10] J. Lacan, *Écrits: A Selection*. (1st) London: Routledge, 2001.
- [11] M. Foucault, *The Order of Things: An Archaeology of the Human Sciences*. (1st American ed.) Tavistock Publications, 1970.
- [12] M. Foucault, *Discipline and Punish : The Birth of the Prison*. (1st American ed) New York: Pantheon Books, 1977 Available: <https://worldcat.org/title/3328401>.

- [13] John Berger, Trans. by 吳莉君. *觀看的方式*. (二版) 臺北市: 麥田出版: 家庭傳媒發行, 2010 Available: <https://worldcat.org/title/1338376917>.
- [14] 維拉斯奎茲《宮女》原來是一幅自畫像 *非池中藝術網*. [Online]. Available: <https://artemperor.tw/focus/1527>. [Accessed: Sep 16].
- [15] J. Crary, *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. MIT Press, 1992 Available: <https://books.GoogleScholar.com.tw/books?id=Ea6i11KVGLkC>.
- [16] Harun Farocki, "Harun farocki: Eye / machine," [Online]. Available: <https://www.harunfarocki.de/installations/2000s/2000/eye-machine.html>. [Accessed: Dec 5].
- [17] Harun farocki. eye/machine I-III, auge/maschine I-III. 2001-2003 | MoMAMoMA. [Online]. Available: <https://www.moma.org/collection/works/147861>. .
- [18] T. Paglen, "Invisible images: Your pictures are looking at you," *Architectural Design*, vol. 89, (1), pp. 22 - 27, 2019. .
- [19] 韓炳哲, Trans. by 管中琪. *透明社會*. (初版) 台北市: 大塊文化出版股份有限公司, 2019 Available: <https://worldcat.org/title/1140423809>.
- [20] E. Goffman, *The Presentation of Self in Everyday Life*. (First Anchor books edition) New York: Anchor Books, 1959 Available: <https://worldcat.org/title/733073728>.
- [21] E. Goffman and 黃煜文., *日常生活中的自我呈現 / 厄文.高夫曼著 ; 黃煜文譯 = the Presentation of Self in Everyday Life / Erving Goffman*. (初版) 臺北市: 商周出版, 2023.
- [22] S. Turkle, *Life on the Screen : Identity in the Age of the Internet*. New York: Simon & Schuster, 1995 Available: <https://worldcat.org/title/33078560>.
- [23] Sherry Turkle, Trans. by 譚天 and 吳佳真. *虛擬化身 : 網路世代的身分認同*. (初版) 臺北市: 遠流, 1998.
- [24] N. K. Hayles, *How we Became Posthuman : Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: The University of Chicago Press, 1999 Available: <https://worldcat.org/title/39539341>.

- [25] Hiroshi Sugimoto, "U.A. playhouse, great neck, new york," *SFMOMA*. [Online]. Available: <https://www.sfmoma.org/artwork/79.221/>. [Accessed: Sep 16].
- [26] 城市失格－西門町 city disqualified - ximen district | 在瘟疫的天空下：病毒的隱喻及其他 [Online]. Available: <https://artogo.co/zh-TW/exhibition/underskyofplague/work/4218f804035c>. [Accessed: Sep 16].
- [27] 【臺北市社宅公共藝術計畫－藝居】完整是一種模糊的概念《在路上》非池中藝術網. [Online]. Available: <https://artemperor.tw/focus/3354>. [Accessed: Sep 16].
- [28] 觀看時間的刻度－李基宏《在路上》臺北市社宅公共藝術計畫非池中藝術網. [Online]. Available: <https://artemperor.tw/focus/3561>. [Accessed: Sep 16].
- [29] Nam June Paik, "TV buddha," *Nam June Paik Art Center*. [Online]. Available: <https://njpart.ggcf.kr/collections/205>. [Accessed: Sep 16].
- [30] Ursula Anna Frohne, "Figure 04 dan graham, present continuous past(s), 1974," *Research Gate*. [Online]. Available: [https://www.researchgate.net/figure/Dan-Graham-Present-Continuous-Pasts-1974\\_fig3\\_321549058](https://www.researchgate.net/figure/Dan-Graham-Present-Continuous-Pasts-1974_fig3_321549058). [Accessed: Sep 16].
- [31] Jason Salavon, "Jason salavon | the class of 1988 & the class of 1967," [Online]. Available: <http://salavon.com/work/Class/>. [Accessed: Nov 5].
- [32] Jason Salavon, "Jason salavon | every playboy centerfold, the decades (normalized)," [Online]. Available: <http://salavon.com/work/EveryPlayboyCenterfoldDecades/>. [Accessed: Nov 5].
- [33] 陳順築, 李維菁, 張世倫, 游崑, 周郁齡 and 余思穎, *家族盒子：陳順築*. (初版) 台灣: 臺北市立美術館、田園城市文化事業有限公司, 2014.
- [34] Zach Blas, "Facial weaponization suite - zach blas," *Zach Blas*. [Online]. Available: <https://zachblas.info/works/facial-weaponization-suite/>. [Accessed: Sep 16].
- [35] Charlotte Jearchive, "Turn selfies into classical portraits with the AI that fuels deepfakes," *MIT Technology Review*. [Online]. Available: <https://www.technologyreview.com/2019/07/23/102857/turn-selfies-into-classical-portraits-with-the-ai-generative-that-fuels-deep-fakes-gans/>. [Accessed: Sep 16].
- [36] G. Bradski, "The Opencv Library," *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, (11), pp. 120 - 123, 2000. .

- [37] OpenCV Team, "OpenCV - open computer vision library," [Online]. Available: <https://opencv.org>. [Accessed: Nov 12].
- [38] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, W. Chang, W. Hua, M. Georg and M. Grundmann, "MediaPipe: A Framework for Building Perception Pipelines," 2019. Available: <http://arxiv.org/abs/1906.08172>.
- [39] Google AI Edge, "MediaPipe solutions guide," [Online]. Available: <https://ai.google.dev/edge/mediapipe/solutions/guide>. [Accessed: Nov 12].
- [40] Google AI Edge, "MediaPipe face mesh," [Online]. Available: [https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face\\_mesh.md](https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face_mesh.md). [Accessed: Nov 12].
- [41] J. MUYBRIDGE, "The Horse in Motion," *Nature*, vol. 25, (652), pp. 605, 1882. Available: <https://doi.org/10.1038/025605b0>.
- [42] E. Muybridge, "The Horse in motion. "Sallie Gardner," owned by Leland Stanford; running at a 1:40 gait over the Palo Alto track, 19th June," 1878. Available: <https://www.loc.gov/item/97502309/>.
- [43] Étienne-Jules Marey, "Pelican in Flight," 1880. Available: [https://ocw.mit.edu/courses/21g-049-french-photography-spring-2017/resources/7-research\\_pelican/](https://ocw.mit.edu/courses/21g-049-french-photography-spring-2017/resources/7-research_pelican/).
- [44] R. Kohler and H. Howell, "Photographic image enhancement by superimposition of multiple images," *Photographic Science and Engineering*, vol. 7, pp. 241 - 245, 1963. .
- [45] A. Rosenfeld, "Picture Processing by Computer," vol. 1, (3), pp. 147 - 176, 1969. Available: <https://doi.org/10.1145/356551.356554>.
- [46] R. C. Gonzalez and P. A. Wintz, *Digital Image Processing*. Addison-Wesley Publishing Company, Advanced Book Program, 1977 Available: <https://books.GoogleScholar.com.tw/books?id=UQhEAQAIAAJ>.
- [47] K. M. Kavanaugh, I. M. F. Pinto, M. J. McGillem, S. F. DeBoe and G. B. J. Mancini, "Effects of video frame averaging, smoothing and edge enhancement on the accuracy and precision of quantitative coronary arteriography," *The International*

*Journal of Cardiac Imaging*, vol. 5, (4), pp. 233 – 239, 1990. Available:  
<https://doi.org/10.1007/BF01797840>.

- [48] J. M. Boyce, "Noise reduction of image sequences using adaptive motion compensated frame averaging," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992, pp. 461 – 464 vol.3.
- [49] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, 1997, pp. 369 – 378 Available:  
<https://doi.org/10.1145/258734.258884>.
- [50] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, (7540), pp. 529 – 533, 2015. Available: <https://doi.org/10.1038/nature14236>.
- [51] F. Galton, "Composite Portraits, Made by Combining Those of Many Different Persons Into a Single Resultant Figure," *The Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 8, pp. 132 – 144, 1879. Available:  
<http://www.jstor.org/stable/2841021>.
- [52] Francis Galton, "Composite Portrait," 1883. Available:  
[https://en.wikipedia.org/w/index.php?title=Composite\\_portrait&oldid=1300956112](https://en.wikipedia.org/w/index.php?title=Composite_portrait&oldid=1300956112).
- [53] J. H. Langlois and L. A. Roggman, "Attractive Faces Are Only Average," *Psychol. Sci.*, vol. 1, (2), pp. 115 – 121, 1990. Available: <https://doi.org/10.1111/j.1467-9280.1990.tb00079.x>.
- [54] T. Valentine, "A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition," *The Quarterly Journal of Experimental Psychology Section A*, vol. 43, (2), pp. 161 – 204, 1991. Available:  
<https://doi.org/10.1080/14640749108400966>.
- [55] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cogn. Neurosci.*, vol. 3, (1), pp. 71 – 86, 1991. Available: <https://doi.org/10.1162/jocn.1991.3.1.71>.

- [56] J. B. Philip and I. P. David, "Extracting Prototypical Facial Images from Exemplars," *Perception*, vol. 22, (3), pp. 257 – 262, 1993. Available: <https://doi.org/10.1068/p220257>.
- [57] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, 1999, pp. 187 – 194 Available: <https://doi.org/10.1145/311535.311556>.
- [58] S. Serengil and A. Özpınar, "HyperExtended LightFace: A facial attribute analysis framework," in *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, 2021, pp. 1 – 4.
- [59] S. Serengil and A. Özpınar, "LightFace: A hybrid deep face recognition framework," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2020, pp. 1 – 5.
- [60] S. Serengil and A. Özpınar, "A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules," *Bilişim Teknolojileri Dergisi*, vol. 17, (2), pp. 95 – 107, 2024. Available: <https://doi.org/10.17671/gazibtd.1399077>  
<https://dergipark.org.tr/en/download/article-file/3573195>.
- [61] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014. .
- [62] A. Radford, L. Metz and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv Preprint arXiv:1511.06434*, 2015. .
- [63] T. Karras, T. Aila, S. Laine and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv Preprint arXiv:1710.10196*, 2017. .
- [64] T. Karras, S. Laine and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401 – 4410.
- [65] J. Ho, A. Jain and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840 – 6851, 2020. .

- [66] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin and J. Clark, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, 2021, pp. 8748 – 8763.
- [67] J. Betker, G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee and Y. Guo, "Improving image generation with better captions," *Computer Science*.<https://cdn.openai.com/papers/Dall-E-3.pdf>, vol. 2, (3), pp. 8, 2023. .
- [68] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10684 – 10695.
- [69] Black Forest Labs, "FLUX," 2024. Available: <https://github.com/black-forest-labs/flux>.
- [70] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv Preprint arXiv:2307.01952*, 2023. .
- [71] P. Esser, S. Kulal, A. Blattmann, R. Entezari, J. Müller, H. Saini, Y. Levi, D. Lorenz, A. Sauer and F. Boesel, "Scaling rectified flow transformers for high-resolution image synthesis," in *Forty-First International Conference on Machine Learning*, 2024, .
- [72] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł Kaiser and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017. .
- [73] A. Radford, K. Narasimhan, T. Salimans and I. Sutskever, "Improving language understanding by generative pre-training," 2018. .
- [74] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, (8), pp. 9, 2019. .
- [75] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry and A. Askell, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877 – 1901, 2020. .

- [76] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Alentschmidt, S. Altman and S. Anadkat, "Gpt-4 technical report," *arXiv Preprint arXiv:2303.08774*, 2023. .
- [77] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. J. Ostrow, A. Welihinda, A. Hayes and A. Radford, "Gpt-4o system card," *arXiv Preprint arXiv:2410.21276*, 2024. .
- [78] OpenAI, "GPT-4.1 prompting guide | OpenAI cookbook," [Online]. Available: [https://cookbook.openai.com/examples/gpt4-1\\_prompting\\_guide](https://cookbook.openai.com/examples/gpt4-1_prompting_guide). [Accessed: Nov 5].
- [79] OpenAI, "Introducing GPT-4.1 in the API," [Online]. Available: <https://openai.com/index/gpt-4-1/>. [Accessed: Nov 5].
- [80] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro and F. Azhar, "Llama: Open and efficient foundation language models," *arXiv Preprint arXiv:2302.13971*, 2023. .
- [81] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava and S. Bhosale, "Llama 2: Open foundation and fine-tuned chat models," *arXiv Preprint arXiv:2307.09288*, 2023. .
- [82] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten and A. Vaughan, "The llama 3 herd of models," *arXiv Preprint arXiv:2407.21783*, 2024. .
- [83] Anthropic, "The Claude 3 Model Family: Opus, Sonnet, Haiku," 2024. Available: [https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model\\_Card\\_Claude\\_3.pdf](https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf).
- [84] G. Team, R. Anil, S. Borgeaud, J. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth and K. Millican, "Gemini: a family of highly capable multimodal models," *arXiv Preprint arXiv:2312.11805*, 2023. .
- [85] G. Team, P. Georgiev, V. I. Lei, R. Burnell, L. Bai, A. Gulati, G. Tanzer, D. Vincent, Z. Pan and S. Wang, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," *arXiv Preprint arXiv:2403.05530*, 2024. .

- [86] G. Comanici, E. Bieber, M. Schaekermann, I. Pasupat, N. Sachdeva, I. Dhillon, M. Blistein, O. Ram, D. Zhang and E. Rosen, "Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities," *arXiv Preprint arXiv:2507.06261*, 2025. .
- [87] Insecam - world biggest online cameras directory [Online]. Available: <http://www.insecam.org/>. [Accessed: Oct 22].
- [88] S. Ando and N. Koizumi, "Floagent: Interaction with mid-air image via hidden sensors," in *SIGGRAPH Asia 2022 Emerging Technologies*, Daegu, Republic of Korea, 2022, Available: <https://doi.org/10.1145/3550471.3558398>.
- [89] W. M. K. Trochim, "Likert scaling," [Online]. Available: <https://conjointly.com/kb/likert-scaling/>. [Accessed: Jun 24].
- [90] 財團法人台灣網路資訊中心, "2024 年台灣網路報告," Oct, 2024. Available: [https://report.twnic.tw/2024/assets/download/TWNIC\\_TaiwanInternetReport\\_2024\\_CH\\_summary.pdf](https://report.twnic.tw/2024/assets/download/TWNIC_TaiwanInternetReport_2024_CH_summary.pdf).
- [91] 楊儒賓., a. 楊儒賓, Y. Rubin and R. Yang, *五行原論：先秦思想的太初存有論*. (Di 1 ban) Taipei Shi: 聯經出版社, 2018 Available: <https://worldcat.org/title/1035372993>.
- [92] Sebastián Ramírez, "FastAPI," [Online]. Available: <https://fastapi.tiangolo.com/>. [Accessed: Jun 24].
- [93] Ask Solem, "Celery - distributed task queue — celery 5.5.3 documentation," [Online]. Available: <https://docs.celeryq.dev/en/stable/>. [Accessed: Jun 24].
- [94] Redis - the real-time data platform [Online]. Available: <https://redis.io/>. [Accessed: Jun 24].
- [95] Wix studio | the web platform built for agencies and enterprises [Online]. Available: <https://www.wix.com/studio>. [Accessed: Oct 22].
- [96] Ngrok | API gateway, kubernetes ingress, webhook gateway [Online]. Available: <https://ngrok.com/>. [Accessed: Jun 24].
- [97] liuliu, "Draw things: AI-assisted image generation," [Online]. Available: <https://drawthings.ai/>. [Accessed: Jun 24].

- [98] Stability AI, "Stable diffusion 3.5," [Online]. Available: <https://stability.ai/news/introducing-stable-diffusion-3-5>. [Accessed: Jun 24].
- [99] HikerAPI SaaS | home [Online]. Available: <https://hikerapi.com/>. [Accessed: Jun 24].
- [100] Serengil, "DeepFace/benchmarks," [Online]. Available: <https://github.com/serengil/deepface/tree/master/benchmarks>. .
- [101] M. Ester, H. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996, pp. 226 – 231.
- [102] OpenAI, "API platform," [Online]. Available: <https://openai.com/api/>. [Accessed: Jun 24].
- [103] TA Studios, "TensorArt-TurboX-SD3.5Large - 8steps | stable diffusion model - LORA," [Online]. Available: <https://tensor.art/models/834787530596537105>. [Accessed: Oct 22].
- [104] subzeroid, "Instagrapi," 2025. [Online]. Available: <https://subzeroid.github.io/instagrapi/>. [Accessed: Oct 22].
- [105] Anonymous "Elements of reconnaissance," *Aviation Week & Space Technology*, vol. 78, pp. 7, March 11. 1963. Available: [https://archive.org/details/Aviation\\_Week\\_1963-03-11/page/n3/mode/1up](https://archive.org/details/Aviation_Week_1963-03-11/page/n3/mode/1up).
- [106] G. C. Brock, D. I. Harvey, R. J. Kohler and E. P. Myskowski, "Photographic considerations for aerospace," Itek Corporation, Lexington, Massachusetts, Tech. Rep. IL-9026-8, 1961.
- [107] G. C. Brock, H. J. Hall and H. K. Howell, *Photographic Considerations for Aerospace*. (1st ed) Lexington, Mass.: Itek Corp., 1965.
- [108] G. C. Brock, H. K. Howell, D. I. Harvey, R. J. Kohler, E. P. Myskowski and H. J. Hall, *Photographic Considerations for Aerospace*. (Second edition) Lexington, Massachusetts: Itek Corporation, 1966.

## 附錄一 AI 生成實驗完整提示語

### 一、原文

A realistic portrait of an East Asian man in his early 30s, with a soft oval face shape, **slightly round cheeks**, expressive almond-shaped eyes behind thin, **round metal frame glasses**, and a smooth low-bridge nose with a subtly rounded tip. His lips are naturally shaped with a balanced softness, and his brows are neat with a soft arch. His hair is black, cut in a modern short style with **a slight undercut at the sides and a casual side-parted textured fringe**, showing natural East Asian hair density. He wears a relaxed casual outfit: a muted **olive-green or gray oversized t-shirt** with a subtle embroidered patch, paired with loose, comfortable dark pants, and white classic sneakers. **He holds a smartphone with one hand while adjusting his glasses or helmet with the other in a casual, informal posture.** The setting is a **cozy outdoor campsite at dusk with soft string lights casting warm glows over a tarp-covered seating area surrounded by green trees and a faint fog rolling in.** Nearby, **a group of friends and a dog are casually lounging**, some capturing the moment on their phones. The mood is warm, intimate, and relaxed, with gentle natural light fading into dusk and soft artificial lights enhancing the atmosphere. Slight sparks of excitement and quiet contemplation blend as the man enjoys a peaceful moment amid nature and friendship, embodying a sensitive and reflective yet approachable lifestyle that values cultural events, music, art, and simple joys. The image captures a harmonious balance of technology and natural life, with subtle reflections of recent concert and exhibition experiences visible in photography gear nearby.

## 二、中文翻譯

一幅 30 歲出頭的東亞男子的逼真肖像，他有一個柔軟的橢圓形臉型，略圓潤的臉頰，薄而圓的金屬鏡框眼鏡後面有表情的杏仁形眼睛，光滑的低橋鼻子，尖端微妙圓潤。他的嘴唇形狀自然，平衡柔軟，眉毛整齊，拱形柔軟。他的頭髮是黑色的，剪成現代短髮，兩側略微剪，側面有休閒的紋理流蘇，顯示出天然的東亞頭髮密度。他穿著休閒裝：一件柔和的橄欖綠或灰色寬大 T 恤，上面有微妙的刺繡補丁，搭配寬鬆舒適的深色褲子和白色經典運動鞋。他一隻手拿著智慧手機，另一隻手以隨意、非正式的姿勢調整眼鏡或頭盔。背景是一個舒適的戶外露營地，黃昏時分，柔和的燈串在被綠樹環繞的休息區投射出溫暖的光芒，周圍環繞著綠樹和微弱的霧氣。在附近，一群朋友和一隻狗隨意地閒逛，有些人用手機捕捉這一刻。氣氛溫暖、親密和放鬆，柔和的自然光淡入黃昏，柔和的人工光增強了氣氛。當男人在自然和友誼中享受和平的時刻時，輕微的興奮和靜的沉思混合在一起，體現了一種敏感、反思但平易近人的生活方式，重視文化活動、音樂、藝術和簡單的快樂。該影象捕捉到了技術和自然生活的和諧平衡，在附近的攝影器材中可以微妙地反映最近的音樂會和展覽經歷。

## 附錄二 多圖生成結果比較之完整提示語

### 一、第一組提示語

#### (一) 原文

A 36-year-old East Asian man with a soft oval face, moderately wide-set almond eyes, smooth nasal contour with a slightly rounded tip, natural balanced lips, and a relaxed but confident expression stands on a winding stone staircase of a traditional Chinese temple complex with intricate white stone railings engraved in dragon and flower motifs. His short black hair is neatly combed, with subtle natural texture. He wears a classic dark brown bomber jacket over a white shirt tucked into straight-cut beige pants, complemented by white sneakers with subtle red and blue stripes, and black aviator sunglasses reflecting soft daylight. His posture is relaxed yet upright with one hand in his pocket, subtly surveying the temple grounds and distant green mountains under an overcast sky. Gentle diffused natural light highlights his calm demeanor and the detailed temple stonework, creating a serene, reflective mood. The background shows lush green hills fading into mist, adding depth and a quiet majesty to the contemplative outdoor scene, expressing his thoughtful personality and connection to culture and nature.

#### (二) 中文翻譯

一位 36 歲的東亞男子，擁有柔和的橢圓形臉龐、中等寬度的杏仁眼、鼻樑輪廓平滑且略帶圓潤的鼻尖、自然平衡的嘴唇，以及輕鬆卻自信的表情，他站在一座傳統中國寺廟群的蜿蜒石階上，樓梯上有雕刻著龍花圖案的精緻白色石欄杆。他短短的黑髮梳得整齊，帶有細膩的自然質感。他穿著經典的深棕色飛行員夾克，外面是白色襯衫，襯衫塞進直筒米色褲子裡，搭配帶有細微紅藍條紋的白色運動鞋，以及反射柔和

日光的黑色飛行員太陽眼鏡。他的姿態放鬆卻挺直，一手插在口袋裡，悄悄地掃視著寺廟的庭院和陰沉天空下的遠方綠山。柔和的自然光線襯托出他平靜的神情與精細的寺廟石工，營造出寧靜且沉思的氛圍。背景描繪茂密的綠丘漸漸消散於霧氣中，為沉思的戶外場景增添層次與寧靜莊嚴，展現他深思熟慮的個性，以及與文化與自然的連結。

## 二、第二組提示語

### (一) 原文

A realistic portrait of an East Asian man around 32 years old, with a soft oval face shape, subtle rounded cheeks, and balanced, natural lips with a gentle curvature. His eyes are expressive, slightly almond-shaped with double eyelids and a moderate distance between them, framed by neat, slightly angled brows. He wears thin round metal glasses that rest lightly on a low nasal bridge with a smooth contour and slightly rounded tip. His hair is dark brown to black, short with natural dense texture, styled casually with a textured fringe and subtle loose waves, showing an East Asian male typical hair density. The man is dressed casually in a lightweight beige jacket worn open over a grey hoodie, paired with relaxed dark pants, and carries a small black crossbody bag with the Adidas logo. He stands in a reflective, slightly moody subway station with faint golden reflections in the tiles behind him and muted, cool artificial lighting that highlights the soft sheen on his skin. His posture is relaxed yet upright, one hand casually holding a phone, face thoughtful with a slight, sincere smile, as if about to share a candid personal story. The atmosphere feels quiet, intimate, and contemplative—a moment capturing a thoughtful, warm, and slightly introspective young adult man navigating the complexity of urban life with subtle humor and inner calm.

## (二) 中文翻譯

一幅寫實的東亞男子肖像，約 32 歲，擁有柔和的橢圓形臉型、細膩圓潤的雙頰，以及平衡自然且曲線柔和的雙唇。他的眼睛表現豐富，略呈杏仁形，雙眼瞼間距適中，眉毛整齊且略微傾斜。他戴著薄而圓的金屬眼鏡，輕放在低矮的鼻樑上，鼻樑線條平滑，鼻尖略呈圓潤。他的頭髮從深棕到黑色，短而自然濃密，造型隨意，有層次的瀏海和細緻的鬆波浪，展現出東亞男性典型的髮質密度。男子穿著輕薄米色外套，外面是灰色連帽衫，搭配寬鬆的深色長褲，背著一個印有 Adidas 標誌的小黑色斜挎包。他站在一個反光、略帶情緒的地鐵站，身後瓷磚映出淡淡的金色光芒，柔和冷冽的人造燈光突顯出他皮膚上的柔和光澤。他的姿態放鬆卻挺拔，一隻手隨意握著手機，臉上帶著若有所思的微笑，彷彿要分享一個坦率的個人故事。氛圍安靜、親密且沉思——捕捉一位深思、溫暖且略帶內省的年輕男子，以細膩幽默與內心平靜，駕馭都市生活的複雜性。

國立臺北藝術大學  
Taipei National University of the Arts

### 三、第三組提示語

#### (一) 原文

A slender Taiwanese man around 28 years old, with a soft oval face and neatly styled black hair parted slightly to the side, dark expressive almond-shaped eyes with double eyelids, a smooth low nasal bridge with a softly rounded tip, natural lips with a gentle curvature, subtle cheekbones, and straight thin brows. He wears a minimalist, semi-formal outfit: a black high-neck top paired with loose black trousers and polished black leather boots. His posture is relaxed yet attentive, seated in a modern wood and fabric armchair on a small stage in a cozy contemporary gallery space with a muted concrete wall and soft natural light filtering through a large window. On the wall behind him is a large projection screen, dimly lit to highlight abstract video art related to nature and memory. Nearby, small water bottles

and note cards rest on a simple wooden side table. The overall mood is contemplative and intimate, with a calm, diffused light casting soft shadows that emphasize his reflective expression as he gently holds a microphone mid-discussion. The environment conveys a blend of artistic discourse and personal introspection, surrounded by subtle touches of nature-inspired artworks hinting at his deep connections to forests, history, and the human body as a vessel of memory.

## (二) 中文翻譯

一位約 28 歲的苗條台灣男子，臉型柔和橢圓，黑髮整齊地側分，深邃且富有表情的杏仁形眼睛，雙眼瞼，鼻樑低垂且柔和圓潤，自然的嘴唇帶有柔和曲線，顴骨細緻，眉毛筆直且細長。他穿著極簡、半正式的服裝：黑色高領上衣搭配寬鬆的黑色長褲和拋光的黑色皮靴。他的姿態放鬆卻專注，坐在一張現代木質與布料的扶手椅上，舞台上是一個舒適的當代畫廊空間，牆面柔和的混凝土牆，柔和的自然光透過大窗戶灑入。他身後牆上掛著一塊大型投影螢幕，燈光昏暗，用以突顯與自然與記憶相關的抽象錄像藝術。附近，小水瓶和便條卡放在簡單的木製邊桌上。整體氛圍沉思且親密，柔和的燈光投下柔和的陰影，強調他沉思的表情，他輕輕地手持麥克風，正中途交談。這個環境融合了藝術論述與個人內省，周圍細膩地點綴著自然啟發的藝術作品，暗示他與森林、歷史及作為記憶載體的人體有著深厚的連結。

## 四、第四組提示語

### (一) 原文

A cheerful East Asian woman in her 25 years old with a soft oval face, natural straight black hair falling just below her shoulders, and expressive almond-shaped eyes behind thin, transparent glasses. Her nose has a smooth low bridge with a slightly rounded tip, soft lips

in a natural balanced shape, gentle low cheekbones, and natural slightly thick brows with a soft arch. She wears a comfortable beige knitted sleeveless top paired with light blue jeans and a pinkish-orange crossbody bag shaped like a cartoon animal. She stands relaxed and smiling warmly with arms raised in a joyful open gesture in front of an iconic fairy-tale castle draped in pastel pink and blue banners celebrating an anniversary, on a slightly overcast day that softens the sunlight and gives a cozy, intimate vibe. The busy theme park scene behind her subtly blurs out tourists wandering, animatronic characters, and colorful themed elements, suggesting a fun family outing or travel experience full of happy moments. The lighting is cool and diffused, creating an inviting and genuine atmosphere with a sense of cherished memories and simple joy.

## (二) 中文翻譯

一位 25 歲、開朗的東亞女性，柔和的橢圓形臉龐，自然直黑髮垂至肩膀以下，薄薄透明眼鏡後有著富有表情的杏仁眼。她的鼻樑平滑且低，鼻尖略帶圓潤，嘴唇柔和且自然平衡，低顴骨柔和，眉毛自然略濃，眉弓柔和。她穿著舒適的米色針織無袖上衣，搭配淺藍色牛仔褲，以及一個粉橘色斜挎包，包型是卡通動物造型。她站在一座標誌性童話城堡前，輕鬆地微笑，雙臂高舉，展現出歡愉的開放姿態，城堡掛滿粉紅與藍色的旗幟，慶祝週年紀念，天氣微微陰沉，柔和陽光柔和，營造出溫馨親密的氛圍。她身後繁忙的主題樂園場景巧妙地模糊了遊客、機械人偶和色彩繽紛的主題元素，營造出充滿歡樂時光的家庭出遊或旅遊體驗。燈光涼爽且柔和，營造出溫馨且真誠的氛圍，充滿珍貴回憶與簡單的喜悅。

## 五、第五組提示語

### (一) 原文

A warm, intimate indoor scene of a young East Asian woman in her late twenties, reflecting a sensitive and affectionate personality deeply connected with her beloved pets. She has a soft oval face shape with slightly wide cheeks, almond-shaped expressive eyes with double eyelids and a mildly downturned shape, a low-bridge nose with a smooth contour and gently rounded tip, natural lips with a balanced upper and full lower lip, subtle low and soft cheekbones, and neat thick brows with a soft arch. Her black hair is medium-short, slightly tousled with natural density and a casual side part. She wears a cozy oversized beige knit sweater paired with soft earth-tone comfortable pants, sitting cross-legged on a beige couch in a relaxed, reflective posture gently stroking a sleek black cat. Nearby, two playful tortoiseshell and black kittens lounge on a textured cat tree bathed in warm morning sunlight filtering through a window behind them, showing a view of urban buildings and leafy trees. The soft lighting bathes the room in a gentle glow, emphasizing a calm, contemplative atmosphere. A glass aquarium with a smiling axolotl resting on a natural substrate adds a quiet, whimsical contrast. The overall mood is one of deep care, quiet gratitude, and the tender intimacy of daily life shared with beloved animals and warm sunlight in a simple, inviting home environment.

### (二) 中文翻譯

一幅溫暖親密的室內場景，描繪一位二十多歲的東亞年輕女性，展現出與心愛寵物深厚連結的敏感且深情的個性。她擁有柔和的橢圓形臉型，臉頰略寬，杏仁形富有表情的雙眼皮，微微下垂的眼皮，鼻樑低，鼻尖柔和圓潤，自然的嘴唇，上唇與豐盈的上唇與豐盈的雙唇，低矮柔和的顴骨，以及整齊濃密的眉毛與柔和的弧線。她的黑

髮中短，略帶蓬亂，自然濃密，側分隨意。她穿著一件舒適的米色寬大針織毛衣，搭配柔軟的土色調舒適長褲，盤腿坐在米色沙發上，姿態放鬆且充滿反光，輕輕撫摸著一隻光滑的黑貓。附近，兩隻活潑的玳瑁貓和黑色小貓正懶散地躺在一棵有紋理的貓爬架上，陽光透過窗戶灑進來，映入眼簾的是都市建築和茂密的樹木。柔和的燈光為房間披上一層柔和的光暈，強調出寧靜沉思的氛圍。一個玻璃水族箱裡有一隻微笑的墨西哥鈍口螈，靜謐且充滿奇幻感。整體氛圍充滿深切關懷、靜默感激，以及與心愛動物共享的溫馨親密，與心愛的動物共享溫暖陽光，營造一個簡單且誘人的家。

國立臺北藝術大學  
Taipei National University of the Arts