國立台東大學

資訊工程學系碩士班

碩士論文

AI輔助學習：基於Unity和ComfyUI的互動式創意生成教育工具

AI-Assisted Learning: An Interactive Creativity Generation Educational Tool Based on Unity and ComfyUI

研 究 生 : 胡修銘

指導教授 : 賴盈勳 博士

狄旻

目錄

[第一章 緒論 3](#_Toc183983167)

[1-1 研究背景與動機 3](#_Toc183983168)

[1-2 研究目的 3](#_Toc183983169)

[1-3 研究流程 4](#_Toc183983170)

[1-4 研究限制與範圍 4](#_Toc183983171)

[第二章 文獻探討 5](#_Toc183983172)

[2-1 AR智慧眼鏡 5](#_Toc183983173)

[2-2 ComfyUI 6](#_Toc183983174)

[2-3 生成式AI 6](#_Toc183983175)

[2-4 選用的ComfyUI模型 7](#_Toc183983176)

[2-4-1 TripoSR 7](#_Toc183983177)

[2-4-2 Flux 7](#_Toc183983178)

[2-4-3 SF3D 8](#_Toc183983179)

[2-4-5 SDXL 8](#_Toc183983180)

[第三章 流程設計與架構 10](#_Toc183983181)

[3-1 AR眼鏡前端與使用者介面 10](#_Toc183983182)

[3-2 ComfyUI 與WorkFlow 設置 10](#_Toc183983183)

[3-3 3D物件生成模型 10](#_Toc183983184)

[3-4 Unity API串接方式 10](#_Toc183983185)

[第四章 研究結果 11](#_Toc183983186)

[4-1 AR眼鏡選用 11](#_Toc183983187)

[4-2 3D生成模型選擇 11](#_Toc183983188)

[4-3 節點參數設置 11](#_Toc183983189)

[4-4 模型生成速度 11](#_Toc183983190)

[第五章 結論與建議 12](#_Toc183983191)

[5-1 合適的AR眼鏡款式 12](#_Toc183983192)

[5-2 選用模型的差異及生成時間突破 12](#_Toc183983193)

[5-3 參數或額外功能性的節點設置 12](#_Toc183983194)

[參考文獻資料 13](#_Toc183983195)

第一章 緒論

1-1 研究背景與動機

近年來人工智慧技術快速發展，特別是生成式AI的興起，如ChatGPT等大型語言模型的出現，不僅改變了人們的生活方式，也為教育領域帶來重大影響。在此背景下，培養學習者的AI素養已成為現代教育的重要課題。AI素養不僅包含對AI技術的基本認知與應用能力，更強調培養學習者具備AI倫理意識、批判性思考能力，以及在AI時代中所需的問題解決能力。

而智慧眼鏡作為一種新興的穿戴式科技，具備擴增實境(AR)功能，能夠將虛擬資訊無縫地整合到現實環境中，為學習者提供即時的視覺回饋與互動體驗。結合生成式AI與智慧眼鏡的創新教學模式，不僅能讓學習更具互動性和沉浸感，更能通過實時的AI輔助，幫助學習者更直觀地理解AI技術的應用場景與影響。

本研究計畫將致力於建立一個整合生成式AI與智慧眼鏡的創新學習環境。研究將聚焦於設計適合學生的AI素養的應用工具，透過智慧眼鏡提供的混合實境體驗，讓學習者能夠：

1. 直接體驗並操作生成式AI的各種應用場景
2. 理解Gen AI生成過程與其對參數對結果的影響
3. 培養對AI技術的正確認知與相關工具使用能力
4. 發展利用AI工具解決實際問題的能力

本研究將採用混合研究方法，結合量化與質性數據分析，評估此創新教學方式對提升學生AI素養的成效。研究成果預期能為科技教育領域提供新的教學模式參考，同時為培養具備AI素養的新世代人才提供有效的教育方案。此研究主題高度符合當前教育部推動的數位轉型與科技創新教育政策，也呼應了培養具備未來關鍵能力人才的重要目標。

通過此研究，我們期望能建立一個可推廣的AI素養教育模式，培養學生在AI時代所需的核心素養，包括：AI技術認知能力、AI倫理思辨能力、AI應用創新能力，以及跨領域整合能力。這不僅有助於提升學生的科技素養，更能為未來AI時代培育具備關鍵競爭力的人才。

1-2 研究目的

本研究基於生成式AI與智慧眼鏡的結合潛力，設定以下研究目的：

探索智慧眼鏡在室內設計領域的創新應用模式 透過結合生成式AI的capabilities，開發適合智慧眼鏡運行的生成式AI使用流程，本研究以室內設計輔助系統為應用情境。

研究重點著重在ComfyUI的生成式流程以及眼鏡上呈現3D物件效果的比較，為室內設計工作帶來效率提升。

評估ComfyUI框架在智慧眼鏡平台的整合效能 研究ComfyUI節點式架構與智慧眼鏡的整合方式，探討如何優化工作流程，以及在資源受限的智慧眼鏡平台上實現高效能的AI模型運算。同時評估不同模型在實際應用場景中的表現，為未來相關研究提供參考基準。

建立智慧眼鏡結合生成式AI的應用設計準則透過實際開發經驗，歸納出適合智慧眼鏡的生成式AI應用設計原則，包含使用者體驗考量、效能優化策略，以及如何善用ComfyUI的優勢特性。這些準則將有助於未來相關應用的開發。

1-3 研究流程

本研究依循下列步驟進行探討：

前置研究階段

1. 蒐集並分析智慧眼鏡、生成式AI及ComfyUI的相關文獻與技術資料
2. 評估現有智慧眼鏡硬體規格與運算能力
3. 研究ComfyUI框架的功能特性與擴充性

系統開發階段

1. 設計適合智慧眼鏡的使用者介面
2. 建立ComfyUI工作流程
3. 開發API整合介面
4. 進行效能優化與測試

實驗驗證階段

1. 執行系統功能測試
2. 進行效能評估
3. 收集使用者回饋
4. 歸納研究發現與建議

1-4 研究限制與範圍

 硬體限制

本研究僅針對特定型號的智慧眼鏡(Jorjin plus)進行測試，受限於目前智慧眼鏡的運算能力、電池續航力與散熱性能。同時，網路連線品質可能影響系統運作效能。

 軟體限制

研究主要基於ComfyUI框架進行開發，使用其支援的生成式AI模型。在功能方面，著重於2D/3D物件的生成與視覺化呈現，不包含複雜的物理模擬或即時渲染功能。

 應用範圍

本研究以室內設計輔助為主要應用情境，探討智慧眼鏡結合生成式AI的可行性。研究成果雖可作為其他領域的參考，但可能需要因應不同場景需求進行調整。

 技術限制

目前生成式AI模型的運算需求較高，在智慧眼鏡等資源受限設備上的效能表現可能受到影響。同時，模型的生成品質與速度也需要在實際應用中取得平衡。

第二章 文獻探討

2-1 AR智慧眼鏡

AR (Augmented Reality) 眼鏡是一種穿戴式裝置，能將虛擬資訊疊加在使用者的真實視野中。近年來，AR眼鏡技術發展迅速，主要可分為以下幾個面向：

1. 光學顯示技術

* 波導光學 (Waveguide Optics)：使用特殊設計的光學元件引導光線，實現輕薄的顯示方案
* 反射光學 (Reflective Optics)：通過反射鏡面將影像投射到使用者眼前
* 全像光學 (Holographic Optics)：運用全像技術實現更自然的深度顯示

1. 感測器整合

* 空間定位感測器：實現精確的空間追踪
* 手勢識別感測器：提供自然的人機互動介面
* 環境光感測器：自動調整顯示亮度

1. 運算平台

* 整合式處理器：針對AR應用優化的系統單晶片(SoC)
* 神經網路加速器：支援AI模型運算
* 邊緣運算能力：降低對雲端運算的依賴

1. 目前市場主流產品

* Microsoft HoloLens 2：企業級AR眼鏡的代表作
* Magic Leap 2：提供廣視角的沉浸式體驗
* Nreal Light：消費級AR眼鏡的新選擇

2-2 ComfyUI

ComfyUI 是一款功能強大、操作簡便的 AI 生成插件，其基於節點式的介面設計，讓使用者可以自由地組合各種功能模組和節點，創建出複雜的 AI 繪圖流程。支援 Stable Diffusion 1.0、2.0、XL 等模型，並且具有高度的擴展性和相容性，可以整合各種插件和模型，滿足使用者不同的需求。(羅禎俋，2024)

工作流節點中也有許多在 AI繪圖常見的正、反提詞 (Prompt)、採樣方法 (Sampling Methods) 及採樣步驟 (Steps)、重繪值(Classifier-free guidance，簡稱 CFG) 與 Controlnet等插件運用 (徐志溢，2023)。

不僅可以在具有 VRAM 的 GPU 設備上進行高效運算，也可以在沒有 GPU 的情況下進行運算，這為使用者提供了更大的靈活性。提供了豐富的功能，例如文生圖、圖生圖、圖片微調、局部調整、遮罩、細節修正、影片製作等等，可以滿足使用者從簡單到複雜的各種 AI 繪圖需求，並且可離線使用，使用者只需要將套件與模型下載後即可在本地端運行，不用透過對外網路 (羅禎俋，2024)。

2-3 生成式AI

生成式 AI，也稱為生成式人工智慧（generative artificial intelligence），是一種透過機器學習和深度學習技術，依託數位化圖像數據庫，運用演算法生成圖像的技術。 生成式 AI 可應用於生成文本、圖像、影音、音樂等內容，這些內容也被稱為 AIGC（AI generated content），即人工智慧生成內容 (羅禎俋，2024)，在高齡化社會和醫護人員不足的背景下，科技創新被視為改善醫療體系效能和提升照護品質的關鍵。 其中，生成式 AI 被認為是具有潛力的技術之一，可以用於輔助高齡照護，例如提供情感支持、安全監控和娛樂等 (李奉爵，2024)。

生成式 AI也能應用於教育領域，為教學帶來創新和突破，例如客製化學習體驗、互動式學習環境、自動化評量和回饋、提升教師效率、激發創造力和想像力等 (羅禎俋，2024)

2-4 選用的ComfyUI模型

2-4-1 TripoSR

TripoSR 是一個基於 Transformer 架構的 3D 重建模型，能夠從單一影像快速生成 3D 網格。 該模型在 NVIDIA A100 GPU 上的推論時間不到 0.5 秒。 TripoSR 建立在大型重建模型 (LRM) 的基礎上，並在資料處理、模型設計和訓練技術方面進行了多項改進 (Dmitry Tochilkin1, 2024)。

TripoSR 的核心組件包括一個影像編碼器、一個影像到三平面解碼器和一個基於三平面的神經輻射場 (NeRF)。 首先，輸入的 RGB 影像會被編碼成特徵向量。 然後，解碼器將這些特徵向量轉換成三平面特徵表示，用於描述 3D 物體的形狀和外觀。 最後，NeRF 模組根據三平面特徵渲染出新的視圖，並用於訓練模型 (Dmitry Tochilkin1, 2024)。

為了提高模型的性能，TripoSR 在資料收集方面進行了兩項改進：資料策展和一致的渲染設定。 資料策展的目標是從 Objaverse 資料集中選擇一個精心策展的高品質子集，以減少低品質資料對模型訓練的影響。 一致的渲染設定則確保所有訓練影像都使用相同的相機參數和光照條件生成，從而提高資料的一致性。TripoSR 的模型和訓練方面也進行了一些改進，包括使用感知損失函數 (LPIPS) 來提升重建品質。 LPIPS 是一種用於衡量兩張影像之間感知相似度的指標，能夠更好地捕捉人類視覺系統對影像差異的感知 (Dmitry Tochilkin1, 2024)。在公開資料集上的評估結果顯示，TripoSR 在 3D 重建品質和速度方面都優於其他開源模型。 在 GSO 和 OmniObject3D 資料集上進行的量化比較表明，TripoSR 在 Chamfer Distance 和 F-score 等指標上都取得了最佳性能。

TripoSR 也具有較高的計算效率，能夠在 0.5 秒內從單一影像生成 3D 網格。 與其他快速 3D 重建方法相比，TripoSR 在重建品質和速度之間取得了良好的平衡 (Dmitry Tochilkin1, 2024)。然而，TripoSR 也存在一些缺點。 與使用基於網格訓練的 SF3D 模型相比，TripoSR 的重建精度較低。 此外，在處理低動態範圍 (LDR) 輸入時，TripoSR 輸出的顏色可能與輸入影像不完全匹配。 TripoSR 模型的另一個限制是它沒有輸入姿態和內在條件，因此模型會被鼓勵去猜測物體的尺度，這可能導致尺度預測不準確 (Hanwen Jiang,2024)。

總體而言，TripoSR 是一種快速且高效的單一影像 3D 重建模型，它在 3D 生成領域取得了顯著進展。

2-4-2 Flux

Flux模型專注於即時3D生成：

* 動態場景生成
* 即時渲染優化
* 光影效果模擬
* 物件互動處理

2-4-3 SF3D

SF3D 是一種從單張圖像快速重建高品質帶紋理和 UV 展開的 3D 模型的方法，並具有去光照的材質屬性。 與大多數現有方法不同，SF3D 是專門為網格生成而訓練的，它採用一種快速的 UV 展開技術，能夠快速生成紋理，而不是依賴頂點顏色。 該方法還可以學習預測材質參數和法線貼圖，以提高重建的 3D 網格的視覺品質。 此外，SF3D 整合了一個去光照步驟，可以有效地去除低頻照明效果，確保重建的網格可以在新的照明條件下輕鬆使用 (Mark Boss，2024)。

SF3D 的主要目標是解決現有快速 3D 重建模型中的一些關鍵問題，同時保持快速生成速度（在 H100 GPU 上只需 0.5 秒），這些關鍵問題包括 :

光照烘焙 : 現有技術通常將陰影烘培到材質中，SF3D 透過整合明確的照明和使用球面高斯函數 (SG) 的可微分著色模型，提出了分解照明和反射屬性的方法。

頂點著色 : 大多數 3D 生成模型使用頂點著色來表示物體紋理，這會導致生成的 3D 資源在遊戲等應用中使用效率低下。SF3D 提出一種高度可並行化的快速基於盒子投影的 UV 展開技術，以實現 0.5 秒的生成時間。

步進立方體偽影 : 前饋網路通常會創建立體表示，例如 Triplane-NeRFs，然後使用步進立方體 (MC) 演算法將其轉換為網格。 MC 會導致「步進」偽影，可以透過增加體積解析度來減少，但這會帶來巨大的計算開銷。相比之下，SF3D 使用更高效的架構來處理更高解析度的三平面圖，並使用 DMTet 生成網格，並具有學習到的頂點位移和法線貼圖，從而產生更平滑的網格表面。

缺乏材質屬性 : 以前的前饋技術生成的模型在使用不同的照明渲染時，通常看起來很暗淡。 這主要是由於輸出生成中缺乏明確的材質屬性，這會影響光的反射。 為了應對這一問題，SF3D 預測非空間變化的材質屬性。

透過這些改進，SF3D 可以從單張圖像生成高品質的 3D 網格，並具有平滑形狀、分離的材質屬性、可在0.5秒內生成及較小的物件資源(小於1MB)(Mark Boss，2024)。

2-4-5 SDXL

SDXL (Stable Diffusion XL)模型具備：

* 高解析度圖像生成
* 改進的細節表現
* 更好的構圖能力
* 多樣化的風格控制

這些模型的選擇基於以下考量：

* 運行效能要求需符合AR眼鏡的硬體限制
* 生成效果需滿足室內設計應用需求
* 推理速度需達到實時互動的要求
* 模型整合難度需考慮開發資源限制

各模型特性比較：

1. 運算效能

TripoSR：★★★★☆ (較低資源需求)

Flux：★★★☆☆ (中等資源需求)

SF3D：★★☆☆☆ (較高資源需求)

SDXL：★★★☆☆ (中等資源需求)

1. 生成質量

TripoSR：★★★★☆ (優秀的3D重建質量)

Flux：★★★★☆ (良好的即時渲染效果)

SF3D：★★★★★ (最佳的結構重建能力)

SDXL：★★★★★ (頂級的圖像生成質量)

1. 推理速度

TripoSR：★★★★★ (最快)

Flux：★★★★☆ (較快)

SF3D：★★★☆☆ (中等)

SDXL：★★★☆☆ (中等)

第三章 流程設計與架構

3-1 AR眼鏡前端與使用者介面

3-2 ComfyUI 與WorkFlow 設置

3-3 3D物件生成模型

3-4 Unity API串接方式

第四章 研究結果

4-1 AR眼鏡選用

4-2 3D生成模型選擇

4-3 節點參數設置

4-4 模型生成速度

第五章 結論與建議

5-1 合適的AR眼鏡款式

5-2 選用模型的差異及生成時間突破

5-3 參數或額外功能性的節點設置

參考文獻資料

[聚焦AR/VR產品走向與應用趨勢 | 電腦與通訊](https://ictjournal.itri.org.tw/xcdoc/cont?sid=0M257582562602272358&xsmsid=0M208578644085020215)

[AR智慧眼鏡帶動Micro LED需求大爆發， 2026年產值達4,100萬美元|數位時代 BusinessNext](https://www.bnext.com.tw/article/70331/-micro-led-ar-smart-glasses--explode0629)

Rinon Gal NVIDIA, Tel Aviv University、Adi Haviv Tel Aviv University、Yuval Alaluf Tel Aviv University、Amit H. Bermano Tel Aviv University、Daniel Cohen-Or Tel Aviv University、Gal Chechik NVIDIA ( 2024 ). COMFYGEN: PROMPT-ADAPTIVE WORKFLOWS FOR TEXT-TO-IMAGE GENERATION

Yijun Zhao∗、Yiming Cheng†、Shiying Ding、Yan Fang、Wei Cao、Jiacheng Cao‡、Ke Liu ( 2024 ). Magic Camera

Xiangyuan Xue, Zeyu Lu, Di Huang, Zidong Wang, Wanli Ouyang, Lei Bai\* Shanghai Artificial Intelligence Laboratory ( 2024 ). ComfyBench: Benchmarking LLM-based Agents in ComfyUI for Autonomously Designing Collaborative AI Systems

徐志溢（2023）。 基於 Stable Diffusion 模型和 LoRA 優化訓練生成人臉表情

羅禎俋（2024）。 AI工具ComfyUI輔助2D人型角色動畫應用於各景別之研究

李奉爵、黃輝揚 (2024). 基於角色情感互動與主動式照護的生成式 AI 模型能力研究Role-Based Emotional Interaction and Proactive Care Capabilities

of Generative AI Models

Huan-Yi Chen、Jinn-Tsong Tsai, Ph.D. (2023). 生成式AI在生成演算法與資料結構文本之相似度與超參數性能研究Research on the Similarity and Hyperparameter

Performance of Generative AI in Generating Algorithm and Data Structure Texts

Li-Heng Chen、Ding-Ming Wang (2024). 運用生成式 AI 工具對研究生學術研究能力影響之實證研究An Empirical Study of Generative AI as a Research Tool for Graduate Students

Dmitry Tochilkin1、David Pankratz1、Zexiang Liu2、Zixuan Huang1、Adam Letts1、Yangguang Li2、Ding Liang2、Christian Laforte1、Varun Jampani1∗、 Yan-Pei Cao2∗(2024). TripoSR: Fast 3D Object Reconstruction from a Single Image

Hanwen Jiang、Qixing Huang、Georgios Pavlakos (2024). Real3D: Scaling Up Large Reconstruction Models with Real-World Images

Xiao Yu† Baolin Peng‡∗ Michel Galley‡ Jianfeng Gao‡ Zhou Yu† †Columbia University ‡Microsoft Research (2024). Teaching Language Models to Self-Improve through Interactive Demonstrations

Ke Ma, Jeanhun Chung\* (2024). A Research on AI Generated 2D Image to 3D Modeling Technology

Jie Fu †、Shun Fu、Mick Grierson (2024). Coral Model Generation from Single Images for Virtual Reality Applications ∗

Jiale Xu1,2 Weihao Cheng1 Yiming Gao1 Xintao Wang1\*† Shenghua Gao2\* Ying Shan1 1ARC Lab, Tencent PCG 2ShanghaiTech University (2024). InstantMesh: Efficient 3D Mesh Generation from a Single Image with Sparse-view Large Reconstruction Models

Mark Boss1 Zixuan Huang1,2 † Aaryaman Vasishta1 Varun Jampani1

1Stability AI 2UIUC (2024). SF3D: Stable Fast 3D Mesh Reconstruction with UV-unwrapping and Illumination Disentanglement