國立台東大學

資訊工程學系碩士班

碩士論文

智慧眼鏡結合多模態生成式AI之效能與應用特性研究：以室內設計輔助為例

A Study on Performance and Application Characteristics of Smart Glasses Combined with Multimodal Generative AI: Using Interior Design Assistance as a Case Study

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](#_Toc182699195)

[1-1 研究背景與動機 3](#_Toc182699196)

[1-3 研究流程 5](#_Toc182699197)

[1-4 研究限制與範圍 6](#_Toc182699198)

[第二章 文獻探討 6](#_Toc182699199)

[2-1 AR眼鏡 6](#_Toc182699200)

[2-2 ComfyUI 7](#_Toc182699201)

[2-3 生成式AI 7](#_Toc182699202)

[2-4 選用的ComfyUI模型 8](#_Toc182699203)

[2-4-1 TripoSR 8](#_Toc182699204)

[2-4-2 Flux 8](#_Toc182699205)

[2-4-3 SF3D 8](#_Toc182699206)

[2-4-5 SDXL 9](#_Toc182699207)

[第三章 流程設計與架構 10](#_Toc182699208)

[3-1 AR眼鏡前端與使用者介面 10](#_Toc182699209)

[3-2 ComfyUI 與WorkFlow 設置 10](#_Toc182699210)

[3-3 3D物件生成模型 10](#_Toc182699211)

[3-4 Unity API串接方式 10](#_Toc182699212)

[第四章 研究結果 11](#_Toc182699213)

[4-1 AR眼鏡選用 11](#_Toc182699214)

[4-2 3D生成模型選擇 11](#_Toc182699215)

[4-3 節點參數設置 11](#_Toc182699216)

[4-4 模型生成速度 11](#_Toc182699217)

[第五章 結論與建議 12](#_Toc182699218)

[5-1 合適的AR眼鏡款式 12](#_Toc182699219)

[5-2 選用模型的差異及生成時間突破 12](#_Toc182699220)

[5-3 參數或額外功能性的節點設置 12](#_Toc182699221)

[參考文獻 13](#_Toc182699222)

第一章 緒論

1-1 研究背景與動機

隨著資訊科技的快速發展，智慧眼鏡作為新一代的人機互動介面，正逐漸改變人們與空間互動的方式，並在遊戲、醫療保健、教育、軍事國防等領域展現出廣泛的應用前景。根據市場研究機構 Orion Market Research 的預測，全球 AR/VR 智慧眼鏡市場預計在 2023 年到 2030 年間，將以 8.1% 的複合年增長率持續成長，由於「隨看即得」的特性，讓智慧眼鏡可以在各個行業帶來革新應用。

生成式 AI 是人工智慧領域中一個快速發展的分支，其核心概念在於\*\*讓 AI 不僅僅是分析和預測現有資料，更能創造新的資料，\*\*是一種技術，它更像是一種賦予 AI 創造力的新方法，讓 AI 能夠像人類一樣進行創作。 生成式 AI 模型通常具備多模態能力，可以處理文字、圖像、聲音、程式碼等多種類型的資料，在跨媒體內容的生成和理解上擁有驚人潛力。

這個技術的發展建基於三股主要力量：

資料工程的進步: 現代化的資料工程發展出大數據治理技術，能夠收集和清理網路上大量的數據，為訓練 AI 模型提供了可靠的資源。

電腦算力的提高: 以 NVIDIA 為首的科技大廠投入加速器硬體技術研發，提供超高速的叢集運算能力，為生成式 AI 模型的訓練和推理提供必要的運算資源。

AI 演算法的突破:深度學習技術取得重大突破，例如生成對抗網路 (GAN)、擴散模型 (Diffusion Model) 和採用注意力機制的變換器 (Transformer) 網路，這些技術的發展讓 AI 能生成更逼真和充滿創意的內容，並在自然語言處理領域取得巨大成功。

生成式 AI 的興起，為各行各業帶來了新的突破和應用場景，例如：

**智慧 AOI 檢測**: 生成式 AI 可以生成各種可能的缺陷圖像，擴增訓練數據集，提高模型的泛化能力和檢測準確率，特別是在需要大量影像資料的智慧檢測辨識上，能有效縮減模型產生的時程。

**藝術與設計**: 生成式 AI 能快速產出圖像、轉換藝術風格或提供配色建議，提升藝術家或設計師的工作效率，甚至改變傳統的設計流程。

**影視娛樂**: 生成式 AI 被應用於影視特效、遊戲開發等領域，透過深偽 (Deepfake) 或虛擬人技術，創造出逼真的視覺效果和互動體驗。

**教育**: 生成式 AI 可輔助教學，例如批改作業、製作學習歷程檔案或多媒體教材等。

**醫療保健**: 生成式 AI 被用於醫學影像處理、疾病診斷、新藥研發等任務。

在人工智慧發展的潮流中，ComfyUI 作為一個創新的開源框架， 是一款開源的人工智慧工作流程工具，最大特色就是採用節點式的視覺化介面。讓使用者可以用拖拉的方式，把不同的 AI 功能組合在一起，特別適合用來生成或修改圖片。

功能支援方面，ComfyUI 展現出優異的擴充性與相容性。系統支援多種主流擴散模型，包含 Stable Diffusion 1.x、2.x 系列及 SDXL 等，同時擴展至影片生成（Stable Video Diffusion）和音訊處理（Stable Audio）等多媒體領域，在工作流程管理層面，ComfyUI 提供了完整的工作流程儲存與讀取機制，並具備模型和檢查點管理功能。系統整合的即時效能監控與除錯工具，有效協助研究者追蹤與優化實驗流程。這些功能的整合不僅提升了研究效率，更確保了實驗過程的可追溯性與重現性

本研究旨在探討如何在智慧眼鏡等資源受限設備上有效運行生成式 AI 模型，ComfyUI 展現出在學術研究與實務應用上的諸多優勢。其高度客製化的使用者介面允許研究者根據個別需求進行調整，完善的除錯機制則有助於實驗過程中的問題排除。基於上述特性，選用ConfyUI做為生成式AI開發工具，並探索其在領域的應用。相信隨著技術的進步和應用生態的完善，生成式 AI 與智慧眼鏡的結合將為人們帶來更智能、更便捷、更富創造力的未來生活體驗。

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(Magic Leap 2 , 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，這些都是本研究需要克服的限制。

 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是一個創新的開源AI工作流程框架，具有以下特點：

1. 架構特性

* 節點式視覺化介面：直觀的拖拉操作
* 模組化設計：靈活組合不同功能
* 工作流程管理：支援儲存與分享

1. 核心功能

* 多模型支援：相容各種Stable Diffusion模型
* 參數優化：細緻的參數調整介面
* 批次處理：高效率的任務處理機制

1. 擴充性

* 自定義節點：支援開發新功能
* 模型整合：易於導入新模型
* API介面：方便與其他系統整合

1. 效能優化

* GPU記憶體管理
* 運算資源調度
* 快取機制設計

2-3 生成式AI

生成式AI技術近年來取得重大突破，主要發展方向包含：

1. 模型架構演進

* GAN (Generative Adversarial Networks)
* Diffusion Models
* Transformer-based Models

1. 多模態能力

* 文字生成圖像
* 圖像轉換與編輯
* 3D物件生成

1. 應用領域

* 藝術創作
* 產品設計
* 建築視覺化

1. 效能優化技術

* 模型壓縮
* 量化技術
* 推理加速

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

Flux是一個專注於2D圖像生成的模型:

* 能夠產生高質量的圖像
* 支持多樣化的風格和內容
* 推理速度較快
* 可集成於ComfyUI的工作流程

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

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