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](#_Toc181787650)

[1-1 研究背景與動機 3](#_Toc181787651)

[1-2 研究目的 5](#_Toc181787652)

[1-3 研究流程 5](#_Toc181787653)

[1-4 研究限制與範圍 5](#_Toc181787654)

[第二章 文獻探討 5](#_Toc181787655)

[2-1 AR眼鏡 5](#_Toc181787656)

[2-2 ComfyUI 5](#_Toc181787657)

[2-3 生成式AI 5](#_Toc181787658)

[2-4 選用的ComfyUI模型 5](#_Toc181787659)

[2-4-1 TripoSR 5](#_Toc181787660)

[2-4-2 Flux 5](#_Toc181787661)

[2-4-3 SF3D 5](#_Toc181787662)

[2-4-5 SDXL 5](#_Toc181787663)

[第三章 流程設計與架構 6](#_Toc181787664)

[3-1 AR眼鏡前端與使用者介面 6](#_Toc181787665)

[3-2 ComfyUI 與WorkFlow 設置 6](#_Toc181787666)

[3-3 3D物件生成模型 6](#_Toc181787667)

[3-4 Unity API串接方式 6](#_Toc181787668)

[第四章 研究結果 7](#_Toc181787669)

[4-1 AR眼鏡選用 7](#_Toc181787670)

[4-2 3D生成模型選擇 7](#_Toc181787671)

[4-3 節點參數設置 7](#_Toc181787672)

[4-4 模型生成速度 7](#_Toc181787673)

[第五章 結論與建議 8](#_Toc181787674)

[5-1 合適的AR眼鏡款式 8](#_Toc181787675)

[5-2 選用模型的差異及生成時間突破 8](#_Toc181787676)

[5-3 參數或額外功能性的節點設置 8](#_Toc181787677)

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，ComfyUI 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室內設計輔助系統。研究重點包含使用者介面設計、互動模式規劃，以及如何善用智慧眼鏡的「隨看即得」特性，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是一個專注於3D物件重建的AI模型，具有以下特點：

* 高質量幾何重建能力
* 材質與紋理還原
* 低計算資源需求
* 快速推理速度

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