

清華電機實作專題成果展 網頁成果摘要

Website Abstract for Undergraduate Project Exhibition

Game Scene Generation Based on Diffusion Model

基於 Diffusion Model 的 2D 遊戲場景生成開發

專題領域：系統組

組 別：B423

指導教授：黃朝宗

組員姓名：郭立陽、林育安

研究期間：2024年2月19日至2024年12月31日止，共10個月

1. 報告摘要

本專題旨在利用擴散生成模型（Diffusion Model）開發 2D 遊戲場景，使用現今熱門的擴散模型 Stable Diffusion 生成，然而 Stable Diffusion 生成的場景圖片雖然真實，但生成的場景若要接近想要的遊戲風格，就必須透過微調模型參數來調整模型，因此本專題使用 LoRA（Low-Rank Adaptation）進行模型微調，將真實圖片轉換為符合不同遊戲風格的 2D 場景。專題中，我們收集了 2D 遊戲背景圖來訓練 LoRA 模型，生成符合遊戲需求的場景圖像。

研究方法包括深入研究 DDPM、DDIM 等擴散生成模型，理解了這些擴散模型在逐步去噪與圖片生成中的原理，這為後續的模型微調與生成奠定了理論基礎，並通過收集大量 2D 遊戲（楓之谷）背景圖訓練 LoRA 微調模型。

訓練 LoRA 模型的過程中，透過調整了多種訓練參數，包括 batch size、Epoch 等等，訓練出適合的微調模型。完成微調模型訓練後，我們使用 Gradio 函式庫創建了互動式用戶介面，實現了生成模型與 UI 的集成，這樣一來，就無需再依賴其他的 Stable Diffusion 介面，即可通過快速又簡化的操作流程生成符合要求的遊戲場景，提升了生成效率與可用性，並且保留多樣的生成參數控制，同時我們也使用 Godot 實現了遊戲介面，以利 demo。

本專題展示了 LoRA 結合 Stable Diffusion 在遊戲開發中的應用，並提升了對擴散模型架構和微調技術的理解。同時，也培養了自主解決問題與團隊合作的能力，為未來的學習與研究奠定了基礎。

2. 研究目的及動機

我們的專題旨在透過擴散生成模型（Diffusion Model）實現 2D 遊戲場景的自動化開發，提供一種高效且低成本的解決方案，以應對傳統遊戲背景設計耗時且效率低下的問題。專題的動機來自我們在實際遊戲體驗中的觀察：遊戲中的地圖與場景變換對玩家體驗至關重要，但這些場景通常依賴手工繪製和設計，不僅耗時且耗力。隨著遊戲規模的擴大以及玩家對場景多樣化需求的迅速提升，傳統設計方式在效率與資源分配上面臨瓶頸。因此，我們提出以 Diffusion Model 為核心的自動化生成技術，針對不同遊戲風格生成高質量背景，降低開發成本並提升效率。

為了克服現有技術的局限，我們基於 Stable Diffusion 模型進行研究，並使用 LoRA 微調技術進一步優化生成效果。雖然 Stable Diffusion 在生成擬真風景圖和照片方面表現優異，但未經 LoRA 微調時，輸出的結果難以滿足 2D 遊戲所需的視覺風格。我們透過模型微調，成功開發基於 SDXL 的 LoRA 模型，使其能夠生成符合 2D 遊戲需求的多樣化場景，甚至將真實圖片轉換為楓之谷風格。我們以校園內拍攝的照片進行測試，驗證了模型在風格轉換上的卓越表現。

研究過程中，我們學習了擴散模型的 Forward 與 Reverse Process，並以大量 2D 遊戲背景圖像進行訓練與優化，最終成功開發基於 SDXL 的 LoRA 模型。該模型不僅能透過 text-to-image 生成多樣化場景，還能將真實圖片轉換為楓之谷風格。我們以校園內拍攝的幾張照片為例，驗證了模型在風格轉換上的出色表現。

此外，我們設計了 Gradio 使用介面，簡化操作流程，讓使用者方便地生成遊戲背景。為了進一步展示專題成果，我們將生成的背景圖像應用於 Godot 遊戲引擎，開發了一款簡單的遊戲，充分展現模型應用於遊戲開發的潛力，並驗證 AI 技術在遊戲場景自動化設計上的實用性。這項研究不僅有效降低了開發時間成本，也為未來遊戲設計與場景生成提供了創新方向。

3. 研究方法

(1) 文獻閱讀與理論研究

在專題初期，我們首先對擴散模型的相關文獻進行了系統性閱讀，主要包括 DDPM (Denoising Diffusion Probabilistic Models) 和 DDIM (Denoising Diffusion Implicit Models) 等論文。我們深入分析了這些模型的架構，並嘗試理解背後較為複雜的數學推導過程，這些理論基礎為後續的實作提供了重要支持。

(2) Stable Diffusion WebUI 探索

在理論學習完成後，我們開始實際操作 Stable Diffusion WebUI，並嘗試使用一些有趣的功能來決定專題的具體方向。在此過程中，我們嘗試了 Text-to-Image、Image-to-Image、Inpainting、Tile 等功能，並重點探索了 ControlNet 的應用。

ControlNet 是一個功能強大的擴展工具，它通過條件控制信號來精確調整生成圖像的結果，例如邊緣檢測、姿勢檢測及深度圖等，讓生成的圖像更加符合具體的需求。基於 ControlNet 的多樣化應用，我們進一步閱讀了相關的論文，深入了解其技術細節與背後的實現機制。

(3) 模型微調與方向確定

經過前期探索後，我們發現，由於 Stable Diffusion 的模型架構過於龐大，對其進行完整的訓練和優化需要非常高的硬體要求與時間成本。考量到這些限制，我們決定採用 LoRA (Low-Rank Adaptation) 技術進行模型微調。為了獲得一些實際的訓練經驗並確定該方法的可行性，我們先利用網路上的開源圖庫進行初步的模型微調與參數調整。這不僅讓我們掌握了訓練過程中的參數調整技巧，也證實了 LoRA 技術在我們應用場景中的實際效果。由於當時主題尚未完全確定，我們選擇了現成的開源數據集來進行測試，而在最終確定專題方向後，我們自行收集了符合專題需求的數據集。

最終，經過多次測試與討論，我們確定專題的最終方向為 RPG 遊戲場景的生成。

(4) Implementation of LoRA and Parameter Optimization

在確定 RPG 遊戲場景生成作為專題方向後，我們開始深入研究 LoRA 技術，以探索如何在不改變擴散模型大框架的情況下進行有效微調。在這一過程中，我們略讀了「Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning」這篇論文。這篇論文探討了為什麼較大的模型能在較低的 Intrinsic 維度下有效運行，這解釋了為什麼像 Stable Diffusion 這樣的較大模型在微調時可以更加有效。這一理論基礎讓我更加理解 LoRA 技術的應用潛力。

此外，我們也比較了 LoRA 和 Dreambooth 兩種技術之間的差異。Dreambooth 雖然能夠更好地將視覺特徵完美融入生成圖像，但需要對 Unet 層中的所有參數進行微調，這導致模型體積大且訓練成本高。相較之下，LoRA 技術可以在不破壞原有模型的情況下，即插即用，並且不會產生額外的推理延遲，這讓我們更傾向於採用 LoRA 來進行 RPG 遊戲場景的生成。

在進行 LoRA 微調時，我們也深入閱讀了 LoRA 技術相關的文獻，儘管這些文獻主要針對大型語言模型（LLM）而非 Stable Diffusion，但其理論部分相似。這些研究為我在後續調整參數時提供了理論依據，使我們在微調時有了更明確的方向，而不是隨意地進行參數調整。

LoRA 技術不僅可以在模型中有效地切換任務，還能保持原始模型的共享權重不變，只需要替換 A 和 B 矩陣即可實現不同任務的微調。這種「即插即用」的特性大大提升了模型的靈活性，特別適合我們專題的需求。

在進行模型訓練時，我們使用了四組不同的數據集，嘗試訓練超過十種不同參數及 dataset 的 LoRA Model，最終，因為 SDXL 能夠生成更加豐富且細膩的圖像細節，因此我們選用的模型是 SDXL 的 LoRA 模型(MS_SDXL)

在進行訓練過程中，我們需要為每張圖像進行標註（tagging），此過程可以通過 Stable Diffusion 的一些插件來自動化。最初我們嘗試手動標註，但結果不甚理想，這可能是因為人工標註的準確性與 AI 自動標註方式存在差異。因此，我們最終採用了 WebUI 中的自動標註插件來生成標籤。然而，值得注意的是，並非所有自動生成的標籤都適用於模型訓練。例如，在訓練動漫角色或人物 LoRA 模型時，如果該角色的特徵包括眼鏡、白色頭髮、耳環等，這些屬性不應該直接作為模型的自動標籤，而應當通過添加 trigger word 來標註，這樣在模型訓練過程中，這些特徵能夠被準確地引導至 trigger word，從而提升模型的生成效果，並避免每次 inference(生圖)時重複手動輸入這些已知特徵。

(5) Gradio User Interface Integration

在訓練完 LoRA 模型後，我們使用了 Gradio 來開發一個簡易的使用者互動介面，將微調的模型集成到這個介面中，實現了便捷的圖像生成操作。Gradio 是一個開源的 Python 庫，專為創建基於機器學習模型的即時 UI 而設計。其主要技術特點在於能夠快速將複雜的模型輸出轉換為簡單直觀的前端介面，無需繁複的開發過程即可讓使用者與模型進行互動。這樣的設計不僅加快了我們模型應用的開發進度，還使得整體操作更加簡潔與高效。

具體而言，我們利用 Gradio 的簡單 API，將訓練完成的 LoRA 模型直接嵌入到自建的界面中，取代了依賴第三方的 Stable Diffusion 介面。這樣的集成大幅減少了操作複雜度，使得我們可以更靈活地測試和應用模型，同時提升了開發流程的效率。透過 Gradio 的用戶介面，我們能夠輕鬆地調整輸入參數，實時查看生成的圖像結果，並根據需求進行多次調整，這為我們的專題提供了一個高效且易於使用的解決方案。

(6) Game Demo Development with Godot

我們利用 Stable Diffusion 生成的圖片作為遊戲背景，並使用 Godot 遊戲引擎製作簡易的小遊戲 Demo。首先，我們對生成的圖片進行格式化處理，包括裁剪、調整分辨率等，使其適配遊戲背景的需求，並將其導入 Godot 的資源管理器中作為場景背景。接著，我們在 Godot 中設置遊戲場景，使用 2D 節點搭建遊戲背景，並加入角色與互動物件，例如跑動和跳躍功能，讓玩家可以操作角色進行基本的互動。除此之外，我們也加入了基礎的遊戲邏輯，例如角色的碰撞檢測和場景邊界限制，讓遊戲場景更具完整性。整體實作展示了我們生成的圖片在遊戲背景應用中的潛力，並通過這個 Demo 突出了圖像生成技術與遊戲引擎整合的可能性。

4. 研究成果

研究結果顯示，經過 LoRA 微調的 Stable Diffusion 模型能夠生成風格統一且符合預期的 2D 遊戲場景圖像。在多次調整訓練參數後，我們成功開發出一個能夠快速生成遊戲背景的模型，且生成結果在風格和品質上與訓練目標遊戲相符。這些生成的場景圖像不僅在視覺效果上達到了高品質，還具有良好的一致性和連貫性，能夠適應不同遊戲場景的需求。我們的 Gradio 用戶界面大幅簡化了操作流程，讓使用者可以輕鬆生成並即時調整參數以控制生成結果。這樣的設計不僅方便了使用者操作，還提供了更高的靈活性，讓開發者能夠針對不同的遊戲風格進行多樣化的生成嘗試，此外，我們也使用 Godot 引擎來進行 demo，使我們得以展示生成圖像在遊戲場景中的實際應用，驗證了該技術的可行性和實用性。

5. 總結

本專題成功結合擴散生成模型和 LoRA 技術開發出了生成 2D 遊戲背景的系統，有效地減少了手工繪畫遊戲場景的時間成本，並提升了生成的效率及品質。這一方法在遊戲開發中具有很大的應用潛力，為未來的遊戲背景自動生成提供了新方向。我們的研究成果顯示，經由 LoRA 微調後的 Stable Diffusion 模型能夠有效生成符合預期風格的高品質遊戲場景，透果此方法，不僅能夠降低遊戲開發的門檻，還能夠讓開發者專注於更具創意性的工作，而非繁瑣的圖像設計，本專題不僅展示了擴散模型在遊戲開發中的應用前景，也加深了我們對 Diffusion Model 的理解，並增強了團隊協作及解決問題的能力。

6. 參考圖片

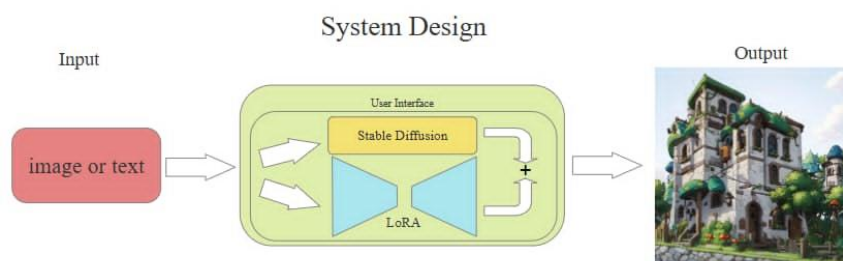


Figure 1: System Design



Figure 2: 清大物理館(左圖為輸入 右圖為遊戲背景輸出)



Figure 3: 清大台達館(左圖為輸入 右圖為遊戲背景輸出)



Figure 4: LoRA 模型各式參數 xy-plot

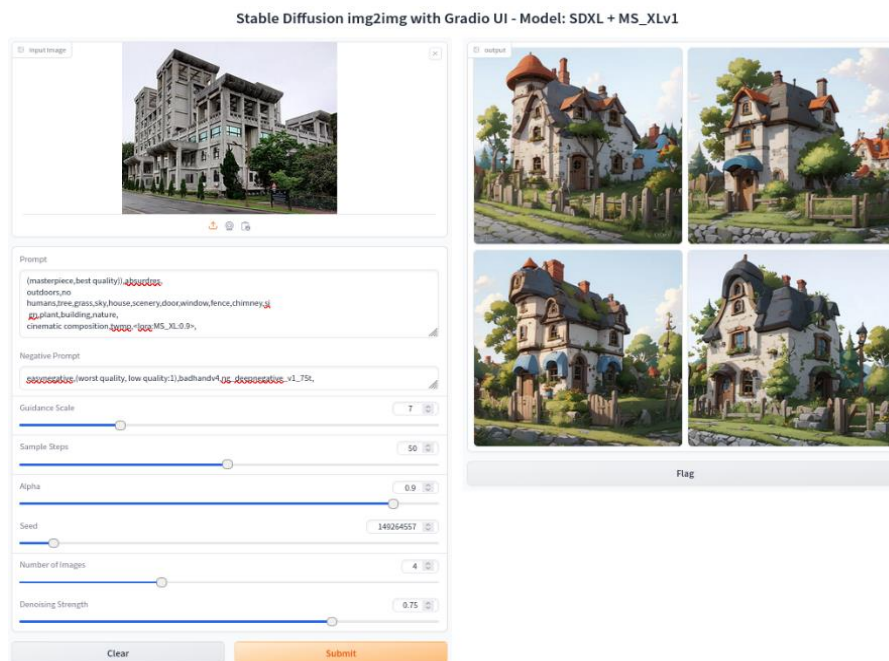


Figure 5: UI 使用者介面成果

7. 心得感想

在這個實作專題中我們成功透過 Diffusion Model 實現了 2D 遊戲場景的生成，為遊戲背景開發提供了一種全新且自動化的工具。通過調整模型參數與進行微調，我們的模型在生成具有特定遊戲風格的圖像上表現優秀。Gradio 的集成大幅簡化了使用介面的操作，使得開發者可以輕鬆生成符合需求的場景圖片，進一步提升了整個開發流程的便捷性與效率。

此專題過程中，我們培養了自我學習與團隊合作的能力，同時，實作專題最精華所在就在學會如何獨自解雜問題，在專題進行的過程中，我們也必須勇於在團隊中提出自己的想法，並學習和互相分工合作、溝通協調，才能在各種可能性中尋找出最合適的解決方案。因此我們認為實作專題的學習經歷是十分可貴的，也令我們收穫良多。

在專題的整個實作過程中，我們培養了自我學習與團隊合作的能力，並且通過解決一系列技術挑戰，深入理解了擴散模型及其在遊戲開發中的應用。這項專題最具挑戰性的部分在於如何解決未知的問題，我們經常需要面對模型訓練中的各種 bug，這些挑戰驅使我們嘗試多種方法及查閱許多網路資料，並根據實驗結果不斷改進我們的方案。

除了技術層面的挑戰，專題中也需要我們具備良好的團隊協作能力。我們在過程中都發揮了不可或缺的作用，無論是資料集的收集、模型訓練、參數調整還是介面開發，我們都在彼此交流中成長。學會與他人協作、分工以及有效溝通，是我們在這次專題中所學到的寶貴經驗，這些學習經歷對我們未來的發展至關重要，為我們未來的專業發展奠定了重要基礎。