電機資訊學院 2025 作 BRAIN PLUS HAND 實作專題競賽

Game Scene Generation Based on Diffusion Model

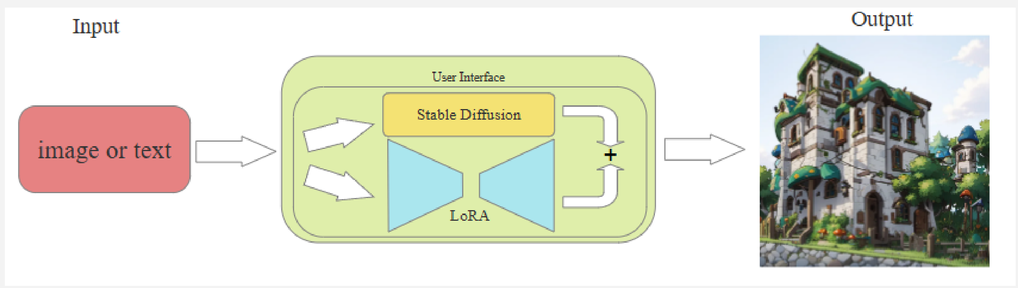
基於 Diffusion Model 的 2D 遊戲場景生成開發

組別 : EECS03　　組員 : 林育安、郭立陽

Abstract

This project aims to use diffusion model for developing 2D game scenes, providing an efficient and cost-effective solution to the time-consuming nature of traditional scene design. Our motivation comes from observing that scene transitions are crucial to player experience but often rely on manual, labor-intensive work. As games scale and player demands for diverse scenes increase, traditional methods struggle with efficiency and resource allocation. We propose an automated generation approach based on diffusion model to produce high-quality backgrounds for various game styles, reducing development costs and improving efficiency.

System Design



Input can be image or text and finally generate output game scene.

Training Model

Version	MS_v7, v8	MS_v9	MS_v10	MS_SDXL
Train batch size	1	2	2	2
Epoch	6	10	10	10
LR Scheduler	Cosine with restarts	Cosine with restarts	Cosine with restarts	Cosine with restarts
Text Encoder learning rate	1E-5	5E-5	5E-5	5E-5
Unet learning rate	1E-4	1E-4	1E-4	1E-4
Network Rank (Dimension)	64	16	16	4
Network Alpha	32	8	8	1
Max resolution	768, 768	512, 512	768, 768	512, 512
dataset	Dataset 3: 25 photos		Dataset 4: 28 photos	

Some versions of training models and their key parameters

During the model training phase, the initial training results were poor and the quality of the generated images was unsatisfactory due to a lack of prior experience. Even when using the same parameters, each training run exhibited random variations, leading to unstable outcomes.

Solution: We conducted an in-depth review of relevant literature and online resources to optimize the model parameters and adjust the training strategies, gradually improving the generation results. Additionally, by repeatedly training the model and testing different parameter settings, we ultimately identified a stable configuration. This significantly enhanced the model's stability and consistency.

We experimented with training more than ten different LoRA models with various parameters and datasets. Ultimately, we selected the SDXL LoRA model (MS_SDXL) over SD1.5 because of its superior ability to generate richer and more detailed image textures.

Experimental Results

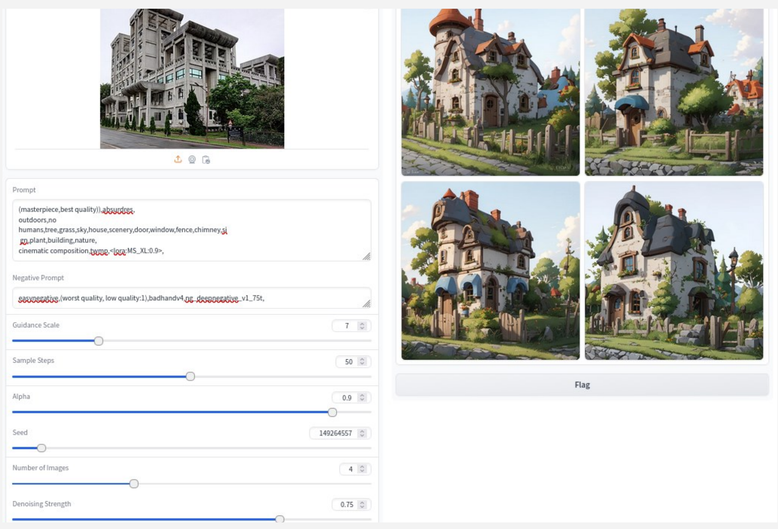
Game scene generated by NTHU Physics building

Game scene generated by NTHU Delta building

These experimental results demonstrate the effectiveness of using the our model to transform real-world image into stylized game scenes. The generated images highlight the model's capability in preserving architectural details while adding unique artistic elements.

User Interface

We utilized the Python Gradio library to develop an user interface for our model, allowing users to adjust parameters and input data .



User Interface

Conclusion

We successfully used the Diffusion Model to generate 2D game scenes, providing a new, automated tool for game background development. By adjusting model parameters and fine-tuning, our model performs well in generating images with specific game styles. The integration of Gradio greatly simplified the user interface operations, enabling developers to easily generate scene images that meet their needs, further enhancing the convenience and efficiency of the entire development process .

Reference

[1] Jonathan Ho, Ajay Jain, Pieter Abbeel. Denoising Diffusion Probabilistic Models .arXiv:2006.11239

[2] Hu et al. “LoRA: Low-Rank Adaptation of Large Language Models, ” In ICLR 2021.

[3] Jiaming Song, Chenlin Meng, Stefano Ermon, Denoising Diffusion Implicit ModelsarXiv:2010.02502

[4] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, Robin Rombach . SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis .arXiv:2307.01952