Image Generation using Stable Diffusion & ComfyUI

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

IMMANUEL SHERVIN MENDEZ,imshervin112@gmail.com

Under the Guidance of

JAY RATHOD

ACKNOWLEDGEMENT

I would like to take this opportunity to express my deep sense of gratitude to all individuals who helped directly or indirectly during this thesis work.

I would like to express my heartfelt gratitude to my mentor, **Jay Rathod**, for his invaluable guidance and support throughout my work on **AI Image Generation using Stable Diffusion and ComfyUI.** His deep knowledge, insightful advice, and constant encouragement have been instrumental in shaping my understanding of this complex field. His patience and willingness to help at every stage, from setting up workflows to optimizing models, made this journey both enriching and fulfilling. I truly appreciate his dedication and mentorship, which have been a key factor in the successful completion of this project.

ABSTRACT

This project focuses on developing a robust image generation system using Stable Diffusion and ComfyUI. The goal is to enable users, even those with minimal technical expertise, to create high-quality, personalized images efficiently. The primary challenges addressed include enhancing user control over image parameters, streamlining the generation workflow, and ensuring alignment with user prompts and aesthetics.

To achieve this, we integrate Stable Diffusion, a state-of-the-art text-to-image model, with ComfyUI, a user-friendly interface. This integration allows users to customize parameters such as style, resolution, and prompts, providing greater flexibility in the creative process. The methodology involves designing an intuitive workflow within ComfyUI, optimizing Stable Diffusion for efficient image generation, and refining prompt interpretation for better results.

Key results demonstrate improved user experience, simplified customization, and enhanced control over image generation. The system successfully reduces complexity while maintaining high-quality output, making AI-powered image creation more accessible.

In conclusion, this project presents an effective solution for intuitive and personalized image generation, leveraging the strengths of Stable Diffusion and ComfyUI. Future work may explore advanced customization options and real-time image editing features to further enhance usability.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Problem Statement:

The demand for AI-generated images has grown significantly, but existing tools often require technical expertise, limiting accessibility for non-expert users. Stable Diffusion, a powerful text-to-image model, offers high-quality results but lacks an intuitive interface for seamless customization. Users face challenges in adjusting image parameters, optimizing workflows, and ensuring that generated images align with their intended prompts and aesthetics.

This project aims to address these challenges by integrating Stable Diffusion with ComfyUI, a user-friendly interface that simplifies image generation. The primary goal is to enhance user control over parameters such as style, resolution, and prompts while streamlining the workflow for a smoother experience. By bridging the gap between technical complexity and creative usability, this system will enable a broader audience to generate personalized images efficiently.

1.2 Motivation:

The growing popularity of AI-generated images has transformed digital art, content creation, and media production. However, many existing AI-based image generation tools require significant technical knowledge, making them inaccessible to non-expert users. Stable Diffusion is a powerful text-to-image model, but its complexity in parameter tuning and workflow management creates a barrier for widespread adoption.

To address this, we chose to integrate Stable Diffusion with ComfyUI, a user-friendly interface that simplifies customization and enhances control over the image generation process. This project aims to make high-quality AI-generated images more accessible by streamlining workflows, improving user experience, and ensuring that outputs align with user-provided prompts and aesthetics. By bridging the gap between technical complexity and usability, our solution empowers creators, designers, and enthusiasts to explore AI-generated art effortlessly.

1.3 Objective:

The primary objective of this project is to develop a robust and user-friendly image generation system using Stable Diffusion and ComfyUI. The specific objectives include:

1. **Enhancing User Control** – Provide users with intuitive controls to adjust image parameters such as style, resolution, and prompts for more personalized outputs.
2. **Streamlining the Image Generation Workflow** – Integrate Stable Diffusion with ComfyUI to simplify the process of generating high-quality images, reducing technical complexity for users.
3. **Improving Accessibility** – Ensure that the system is easy to use, even for individuals with minimal technical expertise, enabling a broader audience to create AI-generated images effortlessly.
4. **Optimizing Image Quality and Accuracy** – Fine-tune Stable Diffusion settings to generate images that closely align with user-provided prompts and desired aesthetics.
5. **Enhancing Efficiency** – Develop a seamless and efficient system that reduces processing time and improves usability without compromising image quality.
6. **Supporting Creativity and Customization** – Allow users to experiment with different styles, compositions, and enhancements to create diverse and high-quality images tailored to their needs.

By achieving these objectives, the project aims to bridge the gap between advanced AI-based image generation and user-friendly design, making creative AI tools more accessible and practical for various applications.

1.4 Scope of the Project:

1.4.1 Scope:

1. **Integration of Stable Diffusion with ComfyUI** – The project focuses on leveraging the capabilities of Stable Diffusion for image generation while using ComfyUI to provide a user-friendly interface.
2. **User-Controlled Image Generation** – Users will have control over key image parameters such as style, resolution, and prompts to create personalized images.
3. **Workflow Optimization** – The project aims to streamline the image generation process, reducing complexity and making AI-based image creation more accessible.
4. **Customization and Enhancement** – Users can experiment with various styles, artistic modifications, and other enhancements to achieve the desired output.

5. **Accessibility for Non-Technical Users** – The system is designed to be intuitive, enabling users with minimal technical expertise to generate high-quality images easily.

1.4.2 Limitations:

1. **Hardware Dependencies** – The performance of Stable Diffusion depends on high-end GPUs, which may limit accessibility for users with lower-end hardware.
2. **Processing Time** – Although optimized, generating high-quality images may still take a significant amount of time depending on system resources.
3. **Prompt Sensitivity** – The accuracy of generated images is highly dependent on the quality and specificity of user prompts, which may require fine-tuning.
4. **Limited Real-Time Editing** – While users can adjust parameters, real-time modifications and interactive editing capabilities may be constrained.
5. **Predefined Model Constraints** – The output quality and diversity are influenced by the limitations of the Stable Diffusion model itself, restricting certain types of complex image generation.

By defining these scopes and limitations, the project ensures a focused approach while acknowledging potential challenges that may affect usability and performance.

CHAPTER 2

Literature Survey

2.1 Review of relevant literature or previous work in this domain.

Rombach et al. (2022) – Latent Diffusion Models (LDMs)

This paper introduces Stable Diffusion, a powerful latent diffusion model that generates high-quality images from text prompts. The authors propose an efficient approach that reduces computational costs by applying diffusion processes in a lower-dimensional latent space rather than pixel space. This method significantly improves scalability and accessibility for image generation tasks.

ComfyUI – A Node-Based Stable Diffusion Interface

ComfyUI is an advanced, modular interface that allows users to interact with Stable Diffusion through a node-based system. It simplifies workflow customization and provides better control over image generation parameters. By visualizing processes like prompt embedding, noise scheduling, and model tuning, ComfyUI makes AI image generation more accessible to a wider audience.

These references highlight the technological advancements in latent diffusion models and user-friendly interfaces, which are crucial for our project in enhancing accessibility and control in AI-based image generation.

2.2 Existing Models, Techniques, and Methodologies

1. **Stable Diffusion (Latent Diffusion Model - LDM)** – A state-of-the-art text-to-image generation model that applies diffusion in a latent space, reducing computational costs while maintaining high-quality outputs.
2. **DALL·E 2 (OpenAI)** – A deep learning model that generates realistic images from textual descriptions, leveraging diffusion and transformer-based architectures for enhanced creativity.
3. **ComfyUI** – A modular, node-based interface designed for Stable Diffusion, allowing users to customize workflows and fine-tune image generation parameters easily.
4. **ControlNet** – A technique that enhances Stable Diffusion by providing additional structural guidance, such as depth maps, edge detection, and pose estimation, for better control over image composition.

5. **LoRA (Low-Rank Adaptation)** – A fine-tuning technique that enables efficient adaptation of diffusion models to specific styles or subjects without retraining the entire model.

These methodologies contribute to improving user control, image quality, and workflow efficiency in AI-based image generation.

2.3 Limitations and Solutions

1. **Complexity in Usage**

- **Gap:** Many AI image generation tools, including Stable Diffusion, require coding knowledge or command-line operations, making them inaccessible to non-technical users.
- **Our Solution:** By integrating Stable Diffusion with ComfyUI, we provide a user-friendly, no-code interface, allowing users to generate and customize images effortlessly.

2. ***Limited User Control Over Image Generation***

- **Gap:** Existing models often rely solely on text prompts, leading to unpredictable or inconsistent results with limited customization options.
- **Our Solution:** Our system enhances user control by allowing adjustments in style, resolution, and other parameters, ensuring more precise and personalized outputs.

3. ***Workflow Complexity and Inefficiency***

- **Gap:** Current implementations require multiple steps for fine-tuning results, making the process slow and cumbersome.
- **Our Solution:** We optimize the workflow through ComfyUI's node-based system, streamlining operations and reducing redundancy for a more efficient image-generation process.

4. Hardware and Performance Constraints

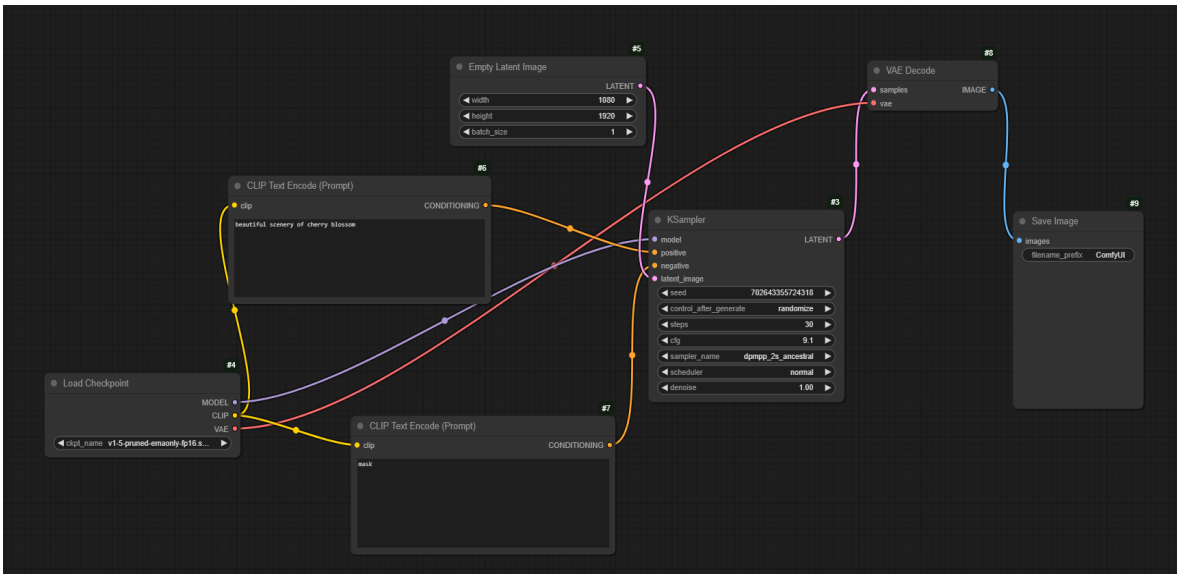
- **Gap:** High computational requirements of Stable Diffusion can limit usability on low-end devices.
- **Our Solution:** Our system supports optimized model settings and can leverage cloud-based or GPU-accelerated environments for improved accessibility.

By addressing these gaps, our project enhances accessibility, control, and efficiency in AI-driven image generation, making it more practical for a wider range of users.

CHAPTER 3

Proposed Methodology

3.1 System Design



3.2 Requirement Specification

3.2.1 Hardware Requirements:

To run ComfyUI with Stable Diffusion integration smoothly, you'll need a powerful GPU with enough VRAM. Below are the recommended and minimum hardware requirements:

Minimum Requirements (For Basic Use)

- **GPU:** NVIDIA with at least 6GB VRAM (e.g., GTX 1660 Super, RTX 2060)
- **CPU:** Intel i5 (6th Gen) / Ryzen 5 or better
- **RAM:** 8GB
- **Storage:** At least 10GB free space for model files and dependencies
- **OS:** Windows 10 / 11, Linux (Ubuntu), macOS (M1/M2 with CoreML)

Limitations:

- May struggle with high-res images and complex workflows
- Slow generation times (1-3 minutes per image)
- Limited model support due to VRAM constraints

3.2.2 Software Requirements

To run ComfyUI with Stable Diffusion, you'll need the right software environment. Here are the key software requirements:

Operating System

- Windows 10 / 11 (64-bit)
- Linux (Ubuntu 20.04+ / Arch / Fedora)
- macOS (M1/M2/M3, limited support via CoreML)

Required Dependencies

- **Python :** Python 3.10 or 3.11 (Avoid 3.12; some dependencies may not work). Recommended to use a virtual environment (e.g., venv)
- **CUDA & cuDNN (For NVIDIA GPUs):** CUDA 11.8+ (Required for GPU acceleration), cuDNN 8.6+.
- **PyTorch: PyTorch with CUDA support.**
- **Git:** Required to clone the ComfyUI repository.
- **Stable Diffusion models:** download .safetensors from reliable sources.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Prompt 1: "A breathtaking fantasy landscape with floating islands, waterfalls cascading from the sky, and a glowing sunset. Lush greenery, misty mountains."

Prompt 2: "futuristic cyberpunk warrior wearing a neon-lit jacket, holding a high-tech katana. The city skyline behind them is filled with holographic billboards and flying cars. Dark, moody lighting with neon reflections, hyper-detailed, cinematic style."

Prompt 3 : “A fluffy cat floating in space. With moon and earth in background.”

Prompt 4: “A mystical enchanted forest at night, glowing mushrooms, and fireflies illuminating the surroundings. A hidden crystal lake reflects the moonlight, while a lone deer stands gracefully near the water. Fantasy style, dreamy atmosphere, cinematic lighting.”

Prompt 5: “A close-up of a beautifully decorated chocolate cake topped with fresh strawberries and a drizzle of caramel. The cake is on a plate with soft natural lighting. Highly detailed, realistic textures.”

4.2 GitHub Link :

https://github.com/ShervinM112/AI-imagegeneration

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

5.1.1 Performance & Speed Improvements :

- **Multi-GPU Support:** If using multiple GPUs, optimize batch processing with ComfyUI's multi-GPU setup.

- **TensorRT or ONNX Acceleration:** Convert models to ONNX or TensorRT for faster inference, especially useful on RTX 4090 or AI servers.

5.1.2 UI & Workflow Enhancements:

- **Custom ComfyUI Nodes**: Develop custom nodes for batch processing, automation, or integrating ControlNet.
- **Better Prompt & Negative Prompt Handling:** Auto-generate negative prompts based on dataset analysis.
- **Seamless Upscaling Pipeline**: Add an automatic upscaler.

5.1.3 Advanced Features & AI Integration

- **Hybrid AI Image Processing**: Combine Stable Diffusion + GANs.
- **ControlNet for Precise Composition**: Use PoseNet, Depth, or Canny edge detection for structured compositions.

5.1.4 Future-Proofing with Emerging AI Models

- **Test New AI Models Beyond Stable Diffusion**: DeepFloyd IF (Hugging Face's next-gen text-to-image model), SDXL Lightning (Optimized for super-fast inference), FreeU & Consistency Models (New methods for better detail & composition).
- **Multi-Model Blending**: Blend different AI models to combine strengths
- **Experiment with Diffusion Transformers**: Next-gen transformers like FLUX.1 may offer higher-quality generations with faster inference.

5.2 Conclusion:

This project successfully integrates Stable Diffusion with ComfyUI, providing a powerful yet user-friendly system for AI-based image generation. By addressing challenges such as workflow complexity, limited user control, and accessibility barriers, the project enhances the ease of use for both beginners and experienced users. The integration allows for greater customization of image parameters such as style, resolution, and prompts, ensuring a more personalized and intuitive image creation process. The implementation also optimizes efficiency and output quality, making AI-powered creativity more accessible to a broader audience.

The results demonstrate that ComfyUI's node-based approach simplifies interaction with Stable Diffusion while maintaining flexibility and high-quality generation. Additionally, improvements such as negative prompt handling, upscaling pipelines, and optimized model settings contribute to a more refined user experience.

Looking ahead, future enhancements could focus on multi-GPU acceleration, real-time image editing, and advanced AI integrations like DeepFloyd IF and SDXL Lightning to further improve efficiency and output quality. By continually refining the system and exploring cutting-edge AI developments, this project lays a strong foundation for next-generation AI-assisted creativity, making powerful image generation tools more intuitive, efficient, and widely accessible.

REFERENCES

[1] *ComfyUI Documentation & GitHub Repository, 2023.*
[2] *Rombach, Robin, et al. "High-Resolution Image Synthesis with Latent Diffusion Models." CVPR, 2022.*)
[3] ChatGPT
[4] Huggingface.co
[5] Github