Image Generation using Stable Diffusion & Comfy UI

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

Garapati Prem Kumar, premkumargarapati777@gmail.com

Under the Guidance of

Abdul Aziz Md

Master Trainer, Edunet Foundation

ACKNOWLEDGEMENT

I want to take a moment to express my deepest gratitude to everyone who made this project, "**Image Generation using Stable Diffusion & Comfy UI**," such a meaningful and enriching experience. This journey has been both challenging and rewarding, and I couldn't have done it without the support and guidance of so many wonderful people.

First and foremost, I want to thank our master trainer, **Mr. Abdul Aziz Md**, for his incredible patience, knowledge, and mentorship. Your ability to explain complex concepts in such a clear and approachable way made all the difference. Your encouragement and belief in us kept me motivated, even when things got tough. Thank you for being such an inspiring guide.

I'd also like to extend my heartfelt thanks to our program manager, **Mr. Pavan Kumar U**, for giving us this amazing opportunity and for always being there to support us. Your dedication to making this internship program a success didn't go unnoticed, and I'm so grateful for the effort you put into ensuring everything ran smoothly.

To my fellow students, thank you for being helpful to each other. Whether it was attending classes, or just helping in project making, your support meant the world to me. I couldn't have asked for a better group to learn and grow with.

This project has taught me so much—not just about technology, but also about teamwork, perseverance, and the value of collaboration. I'll always look back on this experience with gratitude and pride.

Thank you all once again for being such an important part of this journey.

ABSTRACT

This project, *"Image Generation using Stable Diffusion & Comfy UI,"* was an exciting dive into the world of AI-powered image creation. The goal was to address the challenge of generating high-quality, realistic images efficiently, which is often time-consuming and requires significant expertise using traditional methods. We wanted to explore how AI could simplify this process and make it accessible to more people, especially in creative fields like design, marketing, and art.

Our main focus was on using Stable Diffusion, a cutting-edge AI model, and combining it with Comfy UI, a user-friendly interface, to create a seamless and intuitive image generation tool. The idea was to make it easy for users to input simple prompts and get stunning, custom images without needing deep technical knowledge.

To achieve this, we set up the Stable Diffusion model, integrated it with Comfy UI, and experimented with different prompts and settings to see how they influenced the results. We also worked on optimizing the interface to ensure it was straightforward and enjoyable to use.

The results were impressive—Stable Diffusion consistently produced high-quality images that matched the input prompts, and Comfy UI made the entire process smooth and accessible. It was amazing to see how even small changes in prompts or settings could lead to entirely different outputs, showcasing the flexibility and power of the tool.

In the end, this project not only demonstrated the potential of AI in creative industries but also highlighted the importance of user-friendly design in making advanced technology accessible to everyone. It was a rewarding experience, and I'm excited to see how this work can inspire further innovation in the field.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Problem Statement:

Creating high-quality, custom images is often a time-consuming and challenging process, especially for those without technical expertise or access to expensive tools. Traditional methods require significant manual effort, making it difficult for individuals and small teams to produce visuals quickly and efficiently. This project tackles this problem by exploring how AI-driven tools like Stable Diffusion, combined with a user-friendly interface like Comfy UI, can simplify image generation. By making this process faster and more accessible, we aim to empower more people to bring their creative ideas to life.

This problem matters because visuals are everywhere—in marketing, design, entertainment, and education. Solving this challenge can open up new possibilities for creativity and innovation, making it easier for everyone to create stunning images without needing advanced skills or resources.

1.2 Motivation:

I chose this project because I've always been fascinated by the intersection of art and technology. The idea of using AI to create images felt like a perfect way to explore this intersection. Stable Diffusion is a powerful tool that can generate incredible visuals, and Comfy UI makes it accessible even to those who aren't tech-savvy. This combination has the potential to revolutionize how we create and use images in our daily lives.

The applications of this project are endless. For example:

- Artists and Designers: They can use it to quickly generate concept art or brainstorm ideas.

- Marketers: They can create custom visuals for campaigns without relying on expensive designers.

- Educators: They can use it to create engaging visual aids for teaching.

- Hobbyists: Anyone can use it to create personalized art or visuals for fun.

This project has the potential to make image generation more inclusive and accessible, which is why I was so excited to work on it.

1.3 Objective:

The main goals of this project are:

1. To understand how Stable Diffusion works and use it to generate high-quality images.

2. To integrate Stable Diffusion with Comfy UI, making the tool easy and intuitive to use.

3. To experiment with different prompts and settings to see how they affect the results.

4. To show how this tool can be used in real-world scenarios, like marketing, design, or education.

5. To make advanced image generation tools accessible to everyone, not just experts.

1.4 Scope of the Project:

This project focuses on using Stable Diffusion and Comfy UI to generate static images. It involves setting up the model, experimenting with prompts, and creating a user-friendly workflow. I also explored how this tool can be applied in different fields, like marketing and design, to show its real-world potential.

Limitations:

- The project uses pre-trained models, so I didn't train any new models from scratch.

- The quality of the images depends on the prompts and settings, which can take some trial and error to get right.

- It doesn't cover real-time or video-based image generation, as it's limited to static images.

- Hardware limitations, like needing a good GPU, can affect performance.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature

Over the past few years, AI-driven image generation has made incredible progress. Early methods like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) paved the way for creating realistic images using AI. GANs, for example, introduced the idea of two neural networks competing to generate and refine images, but they were often tricky to train and sometimes produced inconsistent results. VAEs offered a more stable approach but sometimes resulted in images that lacked detail or looked blurry.

More recently, diffusion models have taken center stage. These models, including Stable Diffusion, work by gradually adding noise to an image and then learning how to reverse the process. This approach has led to highly detailed and realistic outputs, making it a game-changer in the field. Open-source models like DALL·E, MidJourney, and Stable Diffusion have made AI-generated art more accessible, allowing anyone to create stunning visuals from simple text prompts.

2.2 Existing Models, Techniques, and Methodologies

Several models and techniques have been developed to tackle image generation:

2.2.1 Generative Adversarial Networks (GANs): Models like StyleGAN and BigGAN are known for creating high-resolution images, but they require a lot of computational power and can be difficult to train.

2.2.2. Variational Autoencoders (VAEs): VAEs are more stable than GANs but often produce less detailed images.

2.2.3. Diffusion Models: Models like DALL·E 2, Imagen, and Stable Diffusion have set new standards for image generation. Stable Diffusion, in particular, stands out because it's open-source and doesn't require as much computational power as some other models.

2.2.4. Text-to-Image Models: Tools like CLIP (Contrastive Language–Image Pretraining) allow users to generate images from text descriptions, making it easier to create visuals based on specific ideas.

2.3 Gaps and Limitations in Existing Solutions

Despite the progress, there are still some challenges:

1. Complexity and Accessibility: Many advanced models are hard to set up and use, especially for people without a technical background.

2. Hardware Requirements: Generating high-quality images often requires powerful GPUs, which can be expensive and out of reach for many users.

3. Lack of User-Friendly Interfaces: Most models are command-line based or require coding knowledge, which can be intimidating for creative professionals or hobbyists.

4. Limited Customization: Some models struggle to generate highly specific or niche images based on detailed prompts.

How This Project Addresses the Gaps

This project tackles these challenges by:

1. Simplifying the Process: By combining Stable Diffusion with Comfy UI, the project creates a user-friendly interface that makes advanced image generation accessible to everyone, even those without technical expertise.

2. Optimizing for Accessibility: The project explores ways to make the tool work efficiently on a wider range of hardware, so more people can use it without needing expensive equipment.

3. Enhancing Customization: By experimenting with different prompts and settings, the project shows how users can create highly specific and tailored images.

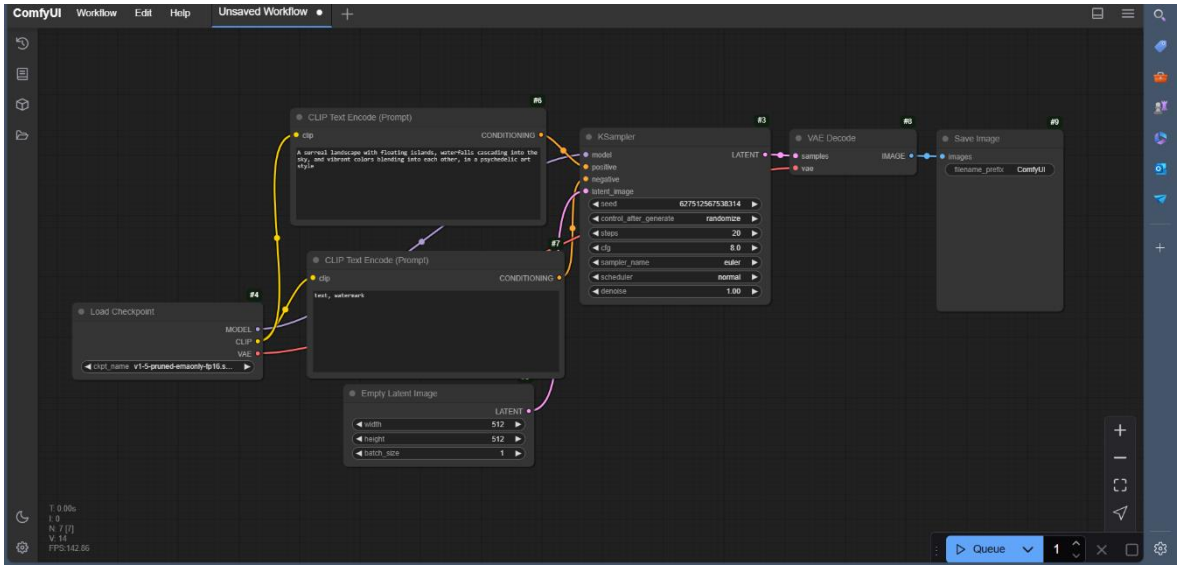
4. Bridging the Gap: The project demonstrates practical applications of AI-driven image generation in fields like marketing, design, and education, showing how it can solve real-world problems.

By addressing these gaps, this project builds on existing advancements and makes them more practical and accessible for a broader audience. It’s about empowering more people to create and innovate, regardless of their technical background.

CHAPTER 3

Proposed Methodology

3.1 System Design



The above image shows a ComfyUI workflow, which is a node-based interface for Stable Diffusion which is used for Image Generation in our project. Here's a breakdown of each phase in the process:

3.1.1. Loading the Checkpoint (Model Selection)

- Node: Load Checkpoint (#4)
- This node loads a Stable Diffusion model (v1-5-pruned-emaonly in this case).
- It provides outputs for Model, CLIP, and VAE, which are required for text-to-image generation.

3.1.2. Text Encoding (Generating Prompt Embeddings)

- Nodes: CLIP Text Encode (Prompt) (#6 & #7)
- First CLIP Text Encode (#6): Encodes the main prompt
    - Example prompt: *"A surreal landscape with floating islands, waterfalls cascading into the sky, and vibrant colours blending into each other, in a psychedelic art style."*

- Second CLIP Text Encode (#7): Encodes additional text (like "watermark") for fine-tuning or negative prompting.

3.1.3. Latent Space Initialization

- Node: Empty Latent Image (#5)
- Defines the latent image dimensions (512x512) and batch size (1).

3.1.4. Sampling Process (Generating the Image)

- Node: KSampler (#3)
- Uses a diffusion sampler (Euler) to iteratively refine the image in the latent space.
- Parameters:
  - Steps: 20 (more steps improve quality but increase generation time).
  - CFG (Classifier-Free Guidance): 8.0 (controls prompt adherence).
  - Denoise: 1.0 (full noise removal for final output).
  - Latent Image Input: Takes the empty latent image and applies transformations.

3.1.5. Decoding Latent Space to an Image

- Node: VAE Decode (#8)
- Converts the processed latent image into a visible image.

3.1.6. Saving the Generated Image

- Node: Save Image (#9)
- Saves the final image with the filename prefix "ComfyUI".

Summary

1. Load Checkpoint → Loads Stable Diffusion model.
2. CLIP Text Encode → Converts text prompt into embeddings.
3. Initialize Latent Space → Defines image size.
4. KSampler → Generates an image using a diffusion model.
5. VAE Decode → Converts latent image to a visible format.
6. Save Image → Stores the final output.

This workflow efficiently processes a text-to-image AI generation task using Stable Diffusion in ComfyUI.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

1. GPU: A decent NVIDIA GPU with at least **8 GB of VRAM** (e.g., RTX 2060 or higher).
2. CPU: A multi-core processor (e.g., Intel i5 or AMD Ryzen 5).
3. RAM: At least **16 GB** for smooth operation.
4. Storage: An SSD with **20 GB of free space** for storing models and images.

3.2.2 Software Requirements:

1. Operating System: Windows 10/11, Linux (Ubuntu recommended), or macOS.
2. Python: Version 3.8 or higher.
3. CUDA (for NVIDIA GPUs): Version 11.7 or higher.
4. Stable Diffusion Model: Pre-trained model (e.g., Stable Diffusion 1.4/1.5 or 2.0).
5. Comfy UI: For a user-friendly interface.
6. Key Libraries: PyTorch, Transformers, NumPy, Pillow, and Torchvision.
7. Web Browser: Chrome or Firefox to access Comfy UI

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Figure 1: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 2: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 3: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 4: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 5: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 6: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 7: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 8: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

Figure 9: A futuristic cityscape at sunset, with towering skyscrapers, flying cars, and neon lights reflecting on a calm river, in a cyberpunk art style.

4.2GitHub Link for Code:

https://github.com/Search-Prem/Image-Generation-using-stable-diffusion-Comfy-UI.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

While this project has shown how powerful Stable Diffusion and Comfy UI can be for creating images, there's still a lot of room for improvement and exciting possibilities to explore in the future:

5.1.1 Better Customization:

- Fine-tuning the model to work better for specific needs, like creating medical images, architectural designs, or even personalized art styles.

5.1.2 Real-Time Image Generation:

- Making the tool faster so it can generate images in real-time, which could be amazing for live art or video editing.

5.1.3 Easier to Run:

- Optimizing the model to work on less powerful computers, so more people can use it without needing expensive hardware.

5.1.4 More User-Friendly Features:

- Adding new features to Comfy UI, like batch processing (generating multiple images at once) or tools for editing images directly in the interface.

5.1.5 Ethical Safeguards:

- Addressing concerns about AI-generated content, like copyright issues or misuse, by adding guidelines or tools to prevent abuse.

5.1.6 Video and Animation:

- Expanding the model to create short videos or animations, which could open up new creative possibilities for storytelling and content creation.

By working on these areas, we can make AI tools even more powerful, accessible, and useful for everyone.

5.2 Conclusion

This project, **"Image Generation using Stable Diffusion & Comfy UI",** has been an exciting journey into the world of AI and creativity. By combining the advanced capabilities of Stable Diffusion with the simplicity of Comfy UI, we've created a tool that makes it easy for anyone to generate stunning, custom images with just a few clicks.

The project shows how AI can be a game-changer for creative industries, education, marketing, and even personal projects. It's all about making advanced technology accessible to everyone, not just experts.

In the end, this project is a small step toward a future where AI helps us bring our ideas to life in ways we never thought possible. It's been a rewarding experience, and I'm excited to see how this work inspires others to explore and innovate in this field.

REFERENCES

1. Comfy UI Documentation. (2023). User Guide for Comfy UI. Retrieved from [Comfy UI GitHub Repository](https://github.com/comfyui/comfyui).

2. www.huggingface.co - for Stable Diffusion download

3. Reddit, ChatGPT for queries.

These references were the foundation for this project, providing the knowledge and tools needed to bring it to life.