Image generation using Stable Diffusion and 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

Meghadri Mukherjee,

meghadri.mm.2003@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to **Mr. Jay Rathod**, my guide and mentor, for his invaluable guidance, encouragement, and support throughout the development of this project, “**SMS Spam Detection System Using NLP**”. His expertise and insights were instrumental in shaping the direction of the project and overcoming challenges.

I am also grateful to **TechSaksham** initiative by **Microsoft & SAP** for providing me with this internship opportunity, a conducive environment, and access to resources that enabled me to successfully complete this project.

Lastly, I extend my thanks to my colleagues and team members for their collaboration and support, and to my family and friends for their constant encouragement during the internship period.

Meghadri Mukherjee

ABSTRACT

The discipline of image generation via the use of artificial intelligence has seen an unbelievable and stunning transformation, highly facilitated by the rise of cutting-edge diffusion models. Amongst these, Stable Diffusion has been a highly flexible deep learning model that has turned out to be one of the best means of generating high-quality images via textual inputs. This specific project explores the intricacies of image generation, more specifically exploring the capabilities of Stable Diffusion as well as ComfyUI, a node-based, user-friendly interface that seeks to simplify the process and render it less burdensome to handle diffusion models. Within this report, we thoroughly discuss a range of key components, ranging from the problem statement that outlines our question, the underlying motive that powers our work, the specific objectives we sought to accomplish, the advanced methodology that we used, the process of implementation we followed, the results that emerged through our experimentation, and lastly, the conclusions that may be derived from the findings.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

Conventional image generation processes are unrealistic, computationally expensive, or demand large domain knowledge. The demand for high-quality, simple-to-use, and effective AI-based image generation tools resulted in the creation of diffusion-based models like Stable Diffusion. Such models are hard to utilize and implement. This project investigates the use of Stable Diffusion with ComfyUI to create an efficient and easy method of image synthesis.

1.2 Motivation:

The inspiration for this project is the increasing need for AI-generated images across industries like digital art, game development, advertising, and content creation. Although software like Stable Diffusion is available, it usually involves programming or complex configuration. ComfyUI provides a graphical user interface that streamlines image creation and makes it available to more people.

1.3 Objective:

- To study the architecture and working principles of Stable Diffusion.
- To explore the functionalities of ComfyUI and its integration with Stable Diffusion.
- To implement an optimized workflow for generating images using a node-based approach.
- To evaluate the quality, performance, and usability of the proposed system.

1.4 Scope of the Project:

This project focuses on the following key aspects:

- Implementing image generation using Stable Diffusion with ComfyUI.
- Understanding the advantages of a node-based workflow.
- Analyzing the effectiveness of AI-generated images for various applications.
- Comparing results with alternative AI-based image generation methods.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature

The field of AI-based image generation has evolved significantly over the years. Early approaches relied on traditional computer graphics techniques such as procedural generation and rule-based systems. However, with the advent of deep learning, generative models like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models have revolutionized image synthesis.

Recent research has shown that diffusion-based models, particularly Stable Diffusion, outperform earlier techniques in terms of image quality, flexibility, and realism. Several studies have explored the applications of diffusion models in creative design, content generation, and artistic rendering. Tools like DALL·E, Midjourney, and Stable Diffusion have demonstrated the potential of AI-driven image synthesis, but they often require coding expertise or complex configurations to achieve optimal results.

2.2 Existing Models, Techniques, and Methodologies

1. Generative Adversarial Networks (GANs): GANs, such as StyleGAN and BigGAN, were among the first successful deep-learning models for image generation. These models work by training a generator and a discriminator in a competitive framework, producing high-quality images. However, GANs suffer from mode collapse, making them less reliable for diverse image generation tasks.
2. Variational Autoencoders (VAEs): VAEs encode input images into a latent space and then reconstruct them, allowing for controlled variation in the generated images. While useful for representation learning, VAEs often struggle with generating highly detailed and realistic images.
3. Diffusion Models (Stable Diffusion): Diffusion models, such as Denoising Diffusion Probabilistic Models (DDPMs) and Latent Diffusion Models (LDMs), use a stepwise noise removal process to generate images from random noise. Stable Diffusion, an advanced LDM, is particularly effective at generating high-quality images with fine details and consistency, making it one of the most widely used models for AI-driven image generation.
4. ComfyUI for Image Generation: While Stable Diffusion is powerful, its implementation often requires coding expertise and technical knowledge of machine learning pipelines. ComfyUI, a node-based interface, simplifies this process by allowing users to create, modify, and experiment with different diffusion workflows visually. Unlike command-line interfaces or complex scripts, ComfyUI enables a more accessible and flexible approach to image generation.

2.3 Gaps in Existing Solutions and How This Project Addresses Them

Despite the advancements in AI image generation, several limitations exist:

- High Learning Curve: Many existing tools require programming skills, making them less accessible to non-technical users.
- Limited Customization: Some AI image generators, such as Midjourney, offer fewer customization options compared to Stable Diffusion but are easier to use.
- Computational Intensity: GANs and VAEs require extensive computational resources, whereas diffusion models provide a more optimized approach but still require efficient workflows.
- Lack of Intuitive Interfaces: While Stable Diffusion is highly effective, its usability is hindered by the complexity of prompt engineering, model selection, and parameter tuning.

How This Project Addresses These Gaps

This project integrates Stable Diffusion with ComfyUI to bridge the gap between powerful image generation capabilities and user accessibility. By leveraging a visual, node-based approach, ComfyUI allows users to build custom workflows, fine-tune model parameters, and generate images efficiently without requiring extensive coding knowledge. The project aims to:

- Make AI-based image generation more user-friendly through an intuitive graphical interface.
- Provide greater customization options by enabling users to manipulate various diffusion parameters visually.
- Optimize performance by implementing efficient workflow designs for Stable Diffusion using ComfyUI.

By addressing these limitations, this project contributes to making advanced AI-based image generation more accessible to artists, designers, and researchers who may not have a deep background in machine learning.

CHAPTER 3

Proposed Methodology

3.1 System Design

Getting Started with ComfyUI

Setting Up ComfyUI with Stable Diffusion

Setting up ComfyUI with Stable Diffusion is surprisingly straightforward.

Initial Setup:

```
# Clone the ComfyUI repository
git clone https://github.com/comfyanonymous/ComfyUI.git
cd ComfyUI
```

Figure 1: ComfyUI set up via GitHub

Model Integration

Downloaded the v1-5-pruned-emaonly-fp16 model from HuggingFace

Placed it in ComfyUI_windows_portable\ComfyUI\models\checkpoints

This specific path is crucial as ComfyUI automatically searches here during startup

Understanding the Connection

Now that we understand both the theory (diffusion models and U-Net) and the practical implementation (ComfyUI), we can see how everything connects:

1. Text → CLIP

Your prompt is processed by CLIP

Creates embeddings that guide the diffusion

2. Diffusion Process

KSampler implements the reverse diffusion

Uses U-Net architecture internally

Gradually denoises based on your settings

3. Final Output

VAE decoder converts latent space to image

Produces the final generated image

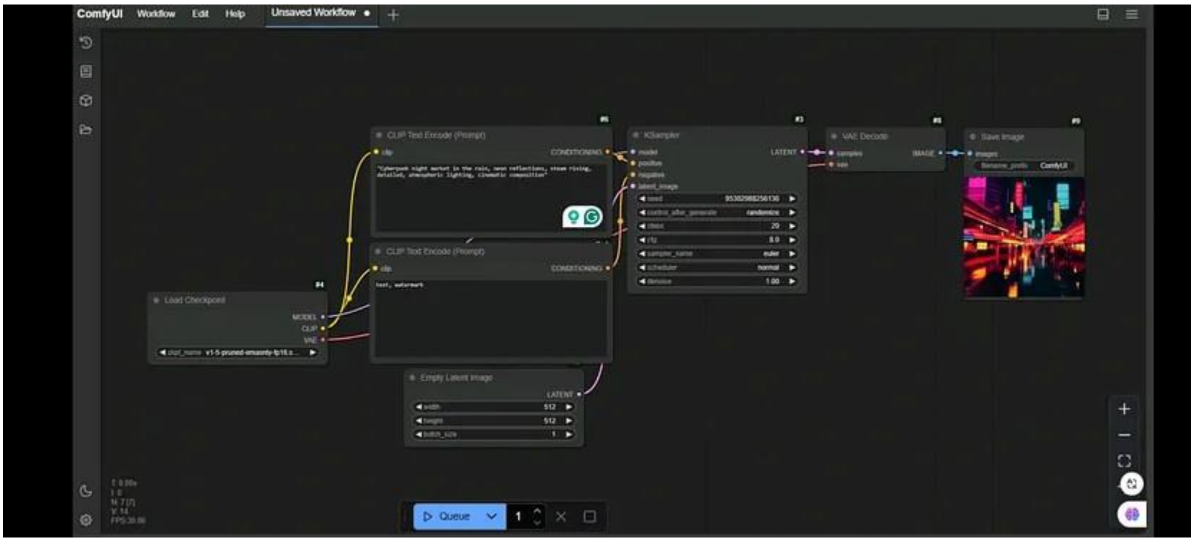

Figure 2: Workflow of SD 1.5

3.2 Requirement Specification

- ComfyUI
- SD1.5
- SD 3.5 Medium

3.2.1 Hardware Requirements:

- GPU: At least 4 GB of VRAM,. NVIDIA GPUs are recommended
- CPU: A modern processor like Intel Xeon E5 or i5 or Ryzen 5 or higher.
- RAM: At least 8 GB of system memory is recommended, but 16 GB or more is better.
- Operating system: Windows 10 or 11, or Linux.
- Storage: At least 40 GB of hard disk space is recommended. An SSD is recommended to speed up loading and running of model files.

3.2.2 Software Requirements:

- Python environment is required, and Python libraries like torch and transformers need to be installed. Some plugins may require Git for installation.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

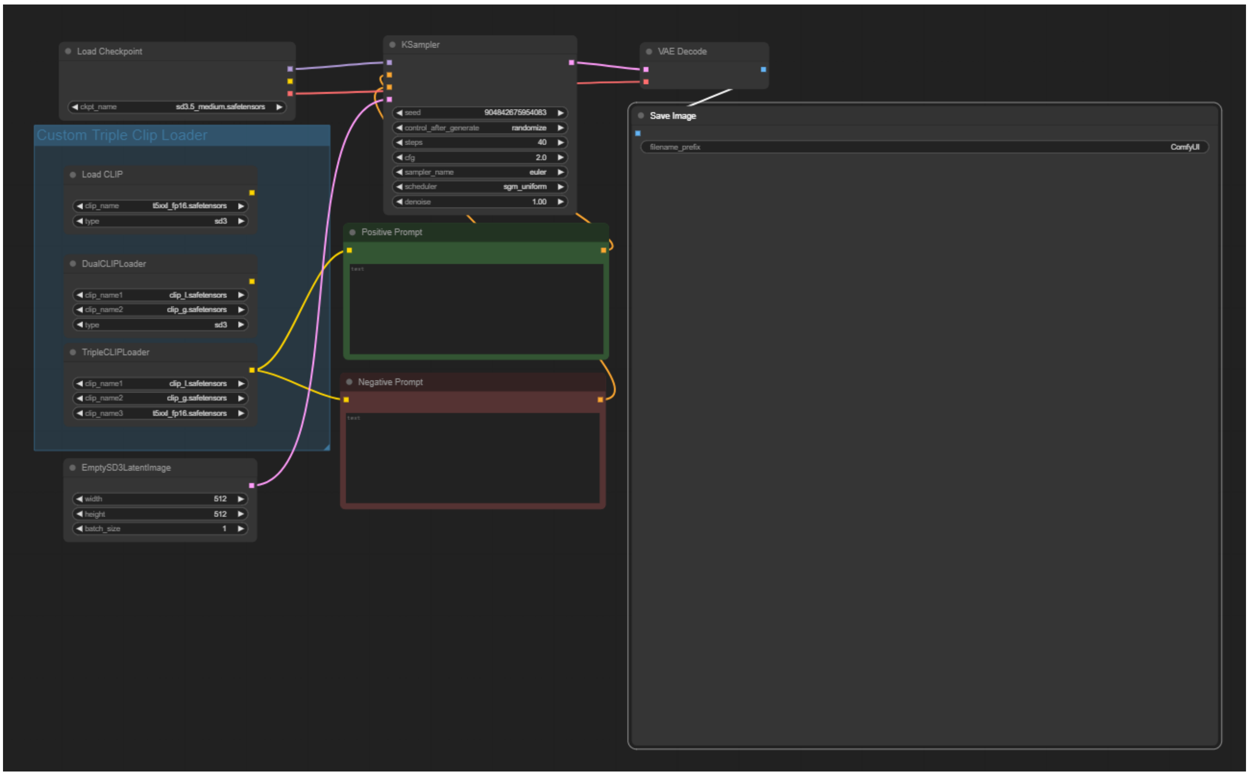

Figure 3: WorkFlow of SD 3.5 Medium model

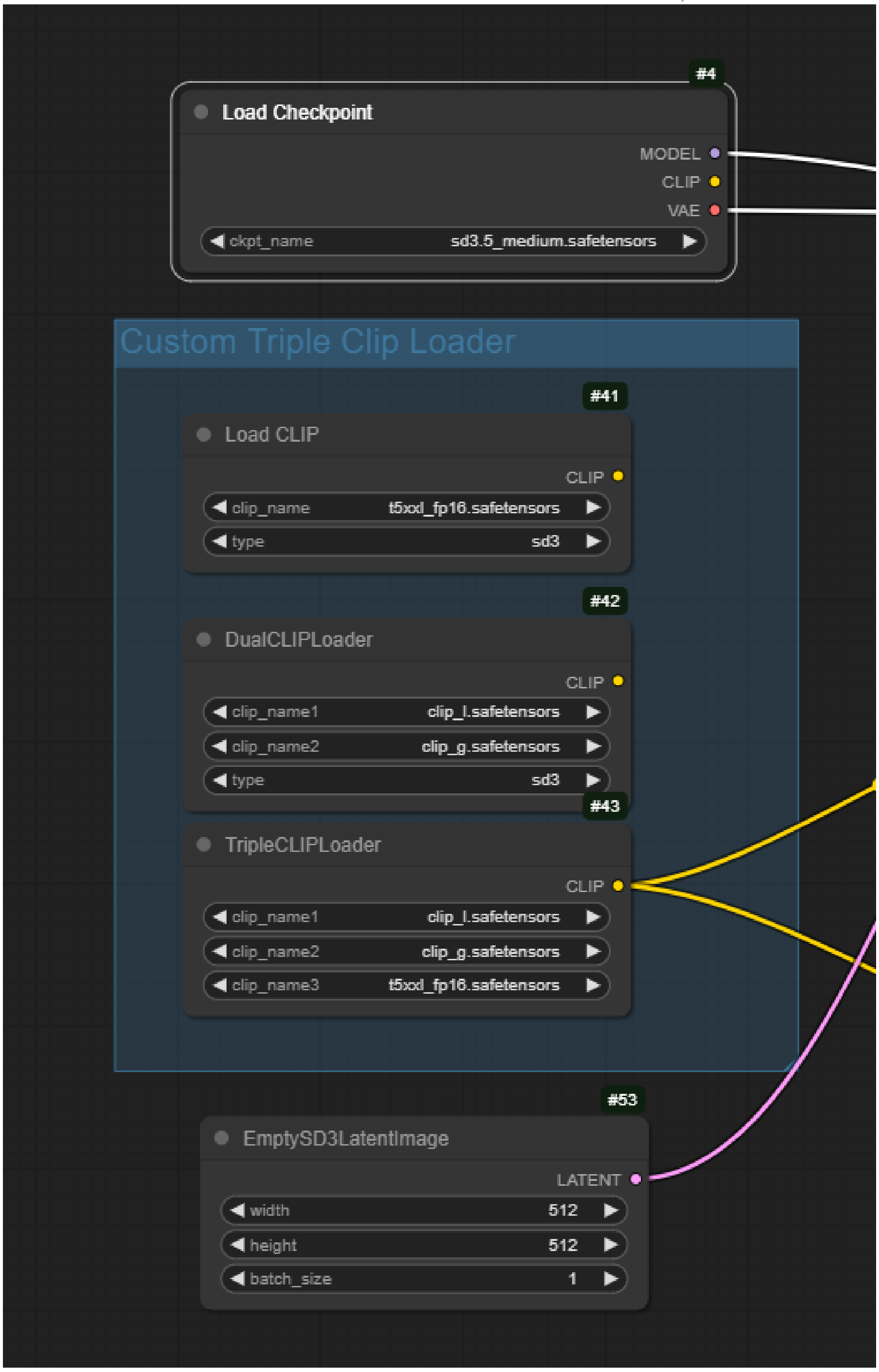

Figure 4: Checkpoint with Triple clip loader

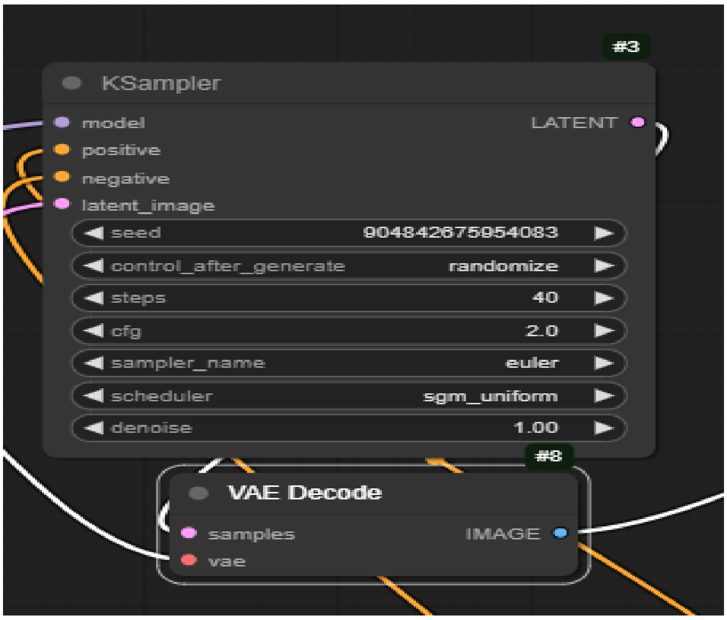

Figure 5: Ksampler

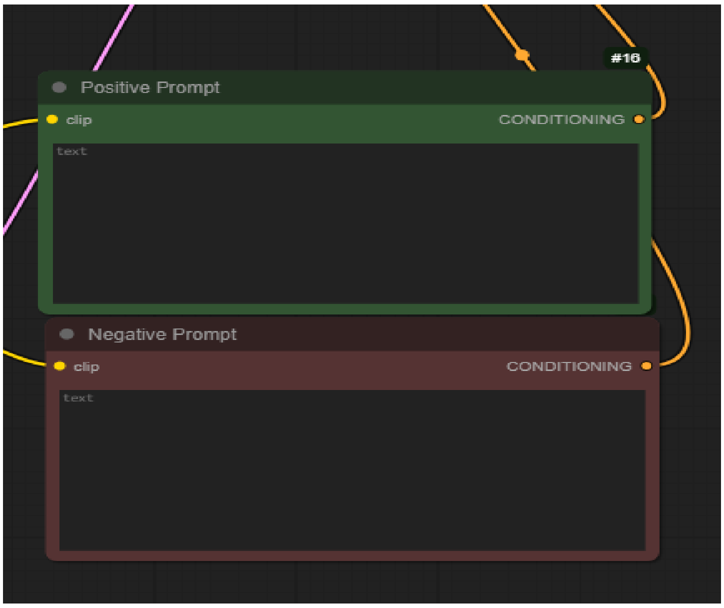

Figure 6: Positive prompt and Negative prompt input boxes

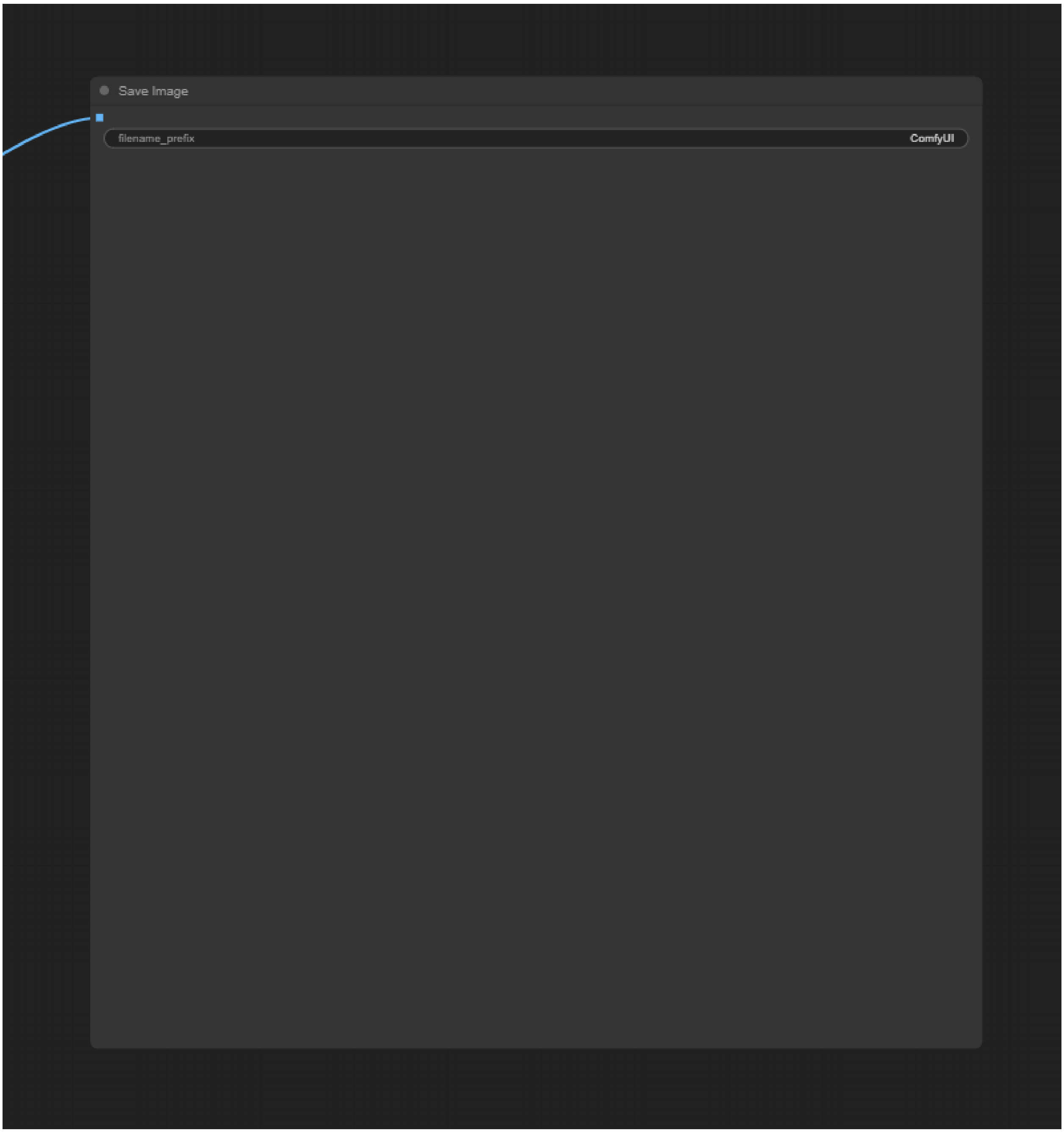

Figure 7: Output box

4.2 GitHub Link for Code:

https://github.com/Flex4Meghadri/SD-3.5.git

CHAPTER 5
Discussion and Conclusion

5.1 Future Work:

Needs a lot of improvement for human related image generation. Planning on making a desktop application using stable diffusion integrating via comfyUi

5.2 Conclusion:

This project successfully demonstrates the potential of integrating Stable Diffusion with ComfyUI for AI-based image generation. The results indicate that a graphical, node-based UI can lower the entry barrier for non-programmers while maintaining high-quality outputs. Future work can explore further enhancements, including model fine-tuning and real-time interactive generation.

REFERENCES

[1] *Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models."* Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. *2022*.

[2] *https://pub.towardsai.net/introductory-notes-on-stable-diffusion-and-stable-video-diffusion-f9fc8ad10c39*

[3] *https://huggingface.co/models?other=stable-diffusion*

[4] *https://github.com/Stability-AI/sd3.5*

[5] *https://stability.ai/news/introducing-stable-diffusion-3-5*

[6] *https://www.comfy.org/*

[7] *https://chatgpt.com/gpts*