Image Generation using Stable Diffusion

And 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

Kailash Chandra Mannem

kailash.mannem@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

We would like to take this moment to sincerely thank everyone who supported us throughout this project. Their guidance, encouragement, and help made this journey smoother and more insightful.

First and foremost, we are deeply grateful to our mentor, **Jay Rathod**, for always being there to guide us. His support, valuable suggestions, and belief in our abilities kept us motivated and helped us overcome challenges. His insights not only improved our understanding of **AI-based image generation** but also inspired us to think creatively and work more efficiently.

We also want to thank **AICTE** for giving us this amazing internship opportunity. It allowed us to explore **Stable Diffusion and ComfyUI**, learn new skills, and gain hands-on experience in a field we are truly passionate about.

Lastly, a big thank you to our friends and family for their constant encouragement and support. Their patience and motivation kept us going, even during tough times.

ABSTRACT

AI-generated images are becoming more popular, but many existing tools are either too complex or don't offer enough control over the final output. The **problem** users face is a user-friendly and comfortable interface that allows them to generate and customize images easily. Many tools require technical knowledge, making it difficult for beginners to use them effectively.

The main **objective** of this project is to develop a simple and user-friendly system for AI-powered image generation. By integrating **Stable Diffusion** with **ComfyUI**, the goal is to give users better control over the customization process while improving image quality and processing speed. This ensures that anyone, regardless of experience, can generate high-quality images with minimal effort.

For the **methodology**, **Stable Diffusion** was used as the core AI model, while **ComfyUI** provided an interactive drag-and-drop interface for adjusting settings. Various experiments were conducted by testing different prompts and optimizing workflows to enhance both the output quality and efficiency of the system.

The **results** showed that the combination of **Stable Diffusion** and **ComfyUI** made the image-generation process significantly easier and more efficient. The system was able to produce detailed, high-quality images while maintaining flexibility for customization.

In **conclusion**, this project successfully enhances AI-powered image generation by making it more accessible and user-friendly. With further improvements such as **real-time editing, multi-modal inputs, and advanced customization options**, the system can be refined to provide even better results in the future.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

The **problem** being addressed in this project is the lack of a simple and intuitive way for users to generate and customize AI-generated images. Many existing tools require technical or coding knowledge, having an uphill learning curve, or offer limited control over the final output. This makes it difficult for beginners, designers, and content creators to fully leverage AI for creative tasks. The **significance** of this issue lies in the growing demand for high-quality AI-generated visuals in fields like digital art, marketing, and content creation. By simplifying the process and providing a user-friendly interface, this project makes AI-powered image generation more accessible, allowing a wider range of users to create customized images with ease and efficiency.

1.2Motivation:

This project was chosen to bridge the gap between **AI-powered image generation and user accessibility**, making it easier for anyone to create high-quality images without requiring advanced technical nor coding knowledge. With the growing demand for AI-generated visuals in **graphic design, content creation, advertising, gaming, and digital art**, providing a simple yet powerful tool can significantly enhance creativity and productivity. By integrating **Stable Diffusion** with **ComfyUI**, users gain more control over the generation process, allowing for customization while maintaining ease of use. The potential impact extends to **artists, designers, marketers, and developers**, enabling them to generate visuals quickly and efficiently. As AI-generated content continues to evolve, this project helps in making such technology more **accessible, practical, and widely adopted** across various industries.

1.3Objective:

The main objective of this project is to develop a **user-friendly AI image-generation system** by integrating **Stable Diffusion** with **ComfyUI** to simplify the creative process while providing better control and customization. The project aims to make AI-generated images more **accessible, efficient, and flexible** for users of all skill levels. Key goals include enhancing **image quality**, optimizing **processing speed**, and ensuring an **intuitive interface** that allows users to fine-tune outputs effortlessly. By achieving these objectives, the project seeks to empower **artists, designers, and content creators** with a powerful yet easy-to-use tool for generating high-quality visuals.

1.4Scope of the Project:

The **scope** of this project includes developing an AI-powered image-generation system using **Stable Diffusion** and **ComfyUI**, providing users with an intuitive interface to create high-quality images with customizable parameters. The system is designed for **graphic designers, artists, content creators, and AI enthusiasts** who want a simple yet flexible tool for generating visuals. It allows users to modify prompts, adjust settings, and experiment with different styles while maintaining ease of use. However, there are **limitations**, such as the dependency on **hardware resources**, which may affect processing speed on lower-end systems. Additionally, while the tool offers customization, it is still limited by the **capabilities of the Stable Diffusion model**, meaning complex artistic requirements may require manual refinements. Future improvements could focus on **real-time editing, enhanced processing efficiency, and better integration with external design tools**.

CHAPTER 2

Literature Survey

AI-powered image generation has gained significant attention in recent years, with models like Stable Diffusion, DALL-E, and MidJourney revolutionizing the way digital content is created. These models leverage deep learning, particularly latent diffusion models (LDMs), to generate high-quality images from text prompts. Research in this domain has explored techniques such as GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders) before transitioning to diffusion models, which offer better detail, structure, and flexibility. Studies have shown that Stable Diffusion performs exceptionally well in producing diverse and high-resolution images while maintaining creative control through text-to-image prompting.

Several models and techniques have been developed for AI-driven image generation. DALL-E 2 by OpenAI and MidJourney offer commercial AI art generation with user-friendly interfaces but often come with restrictions on customization and usage rights. Stable Diffusion, an open-source model, provides more flexibility as it can be fine-tuned and run locally. Methodologies like ControlNet, LoRA (Low-Rank Adaptation), and DreamBooth allow for enhanced control and fine-tuning of AI-generated images. ComfyUI is an advanced workflow-based interface that simplifies interactions with Stable Diffusion, offering node-based processing for better customization.

While current AI image-generation models produce high-quality visuals, they often have limitations in terms of accessibility, customization, and ease of use. Many platforms, like DALL-E and MidJourney, operate on a paid basis, restricting free access. Even Stable Diffusion, though open-source, lacks an intuitive interface, making it difficult for beginners to navigate without technical expertise. ComfyUI, while powerful, can still be overwhelming due to its node-based workflow. This project bridges these gaps by providing a user-friendly system that integrates Stable Diffusion with ComfyUI to make AI-powered image generation easier, more accessible, and highly customizable. The focus is on improving the workflow, speed, and control while ensuring that both beginners and professionals can utilize the tool effectively

CHAPTER 3

Proposed Methodology

3.1 System Design

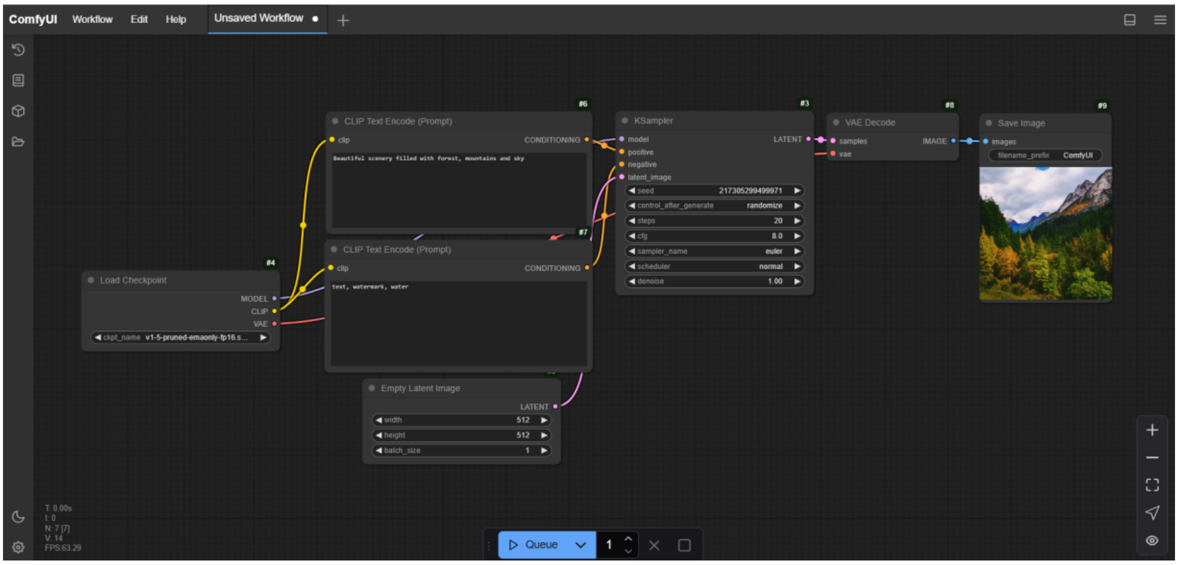

Figure 1: Comfy UI Workflow

The above figure is a workflow of Comfy UI, a graphical interface for creating and running workflows using the ComfyDiffusion library. It consists of different nodes performing a specific task. The different Nodes are:

1. Load Checkpoint (#4)

- This node loads a pre-trained model checkpoint. The specific checkpoint being used here is a version of Stable Diffusion model optimized for performance.
- It provides three outputs: **MODEL** (the diffusion model), **CLIP** (the text encoder), and **VAE** (the Variational Autoencoder).

2. CLIP Text Encode (Prompt) Nodes (#6 and #7)

- These nodes encode text prompts into embeddings that can be understood by the model.

- Node #6 encodes the positive prompt while Node #7 encodes the negative prompt which should not be present in the generated image.

3. Empty Latent Image (#8)

- This node creates an empty latent image with specified dimensions (**width** and **height**) and a batch size.

4. KSampler (#3)

- The KSampler node is responsible for sampling from the diffusion model to generate the latent image.
- It takes several inputs:
  - **model**: The diffusion model from the Load Checkpoint node.
  - **positive**: The positive prompt embedding from CLIP Text Encode node #6.
  - **negative**: The negative prompt embedding from CLIP Text Encode node #7.
  - **latent_image**: The empty latent image from the Empty Latent Image node.
- It also has parameters like **seed** (for reproducibility), **steps** (number of diffusion steps), **cfg** (classifier-free guidance scale), **sampler_name** (sampling algorithm), **scheduler** (step scheduler), and **denoise** (denoising strength).

5. VAE Decode (#9)

- This node decodes the latent image generated by the KSampler into a full-resolution RGB image.
- It takes the **vae** from the Load Checkpoint node and the **samples** from the KSampler node as inputs.

6. Save Image (#10)

- This node saves the final generated image to disk.
- It takes the **images** output from the VAE Decode node and allows specifying a **filename_prefix**.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

- **Processor:** A multi-core CPU (Intel i3/i5 or AMD Ryzen 5/7) or higher.
- **GPU:** NVIDIA GPU with at least **4GB VRAM** (RTX 3060 or higher recommended) for faster model inference. Works without GPU but slow processing time will be faced.
- **RAM:** Minimum **8GB RAM** (32GB recommended for handling large models)
- **Storage:** At least **40GB of free SSD space** to store models and generated images

3.2.2 Software Requirements:

- **Operating System:** Windows 10/11, Linux (Ubuntu recommended), or macOS
- **Programming Language:** Python 3.8 or higher
- **AI Framework: Stable Diffusion (Hugging Face)**
- **Interface & Workflow Engine: ComfyUI**
- **Deep Learning Libraries:** TensorFlow / PyTorch
- **Dependencies & Packages:** NumPy, OpenCV, Matplotlib, PIL, and Hugging Face Diffusers
- **Environment Management:** Conda or Virtualenv for managing dependencies
- **Optional Enhancements:** ControlNet, LoRA, DreamBooth for additional fine-tuning capabilities

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

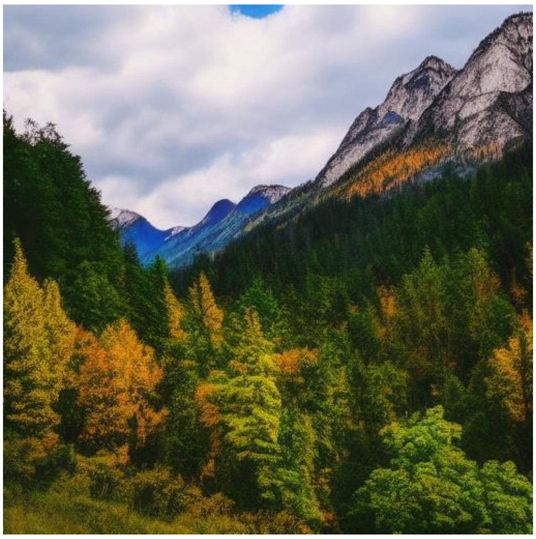

Figure 2: Generated Image 1

Positive Prompt: Beautiful scenery filled with forest, mountains and sky

Negative Prompt: text, watermark, water

Figure 3: Generated Image 2

Positive Prompt: Beautiful Forest Temple surrounded by trees

Negative Prompt: People

Figure 4: Generated Image 3

Positive Prompt: City center packed with people and blinding lights at night

Negative Prompt: Trees, Water

Figure 5: Generated Image 4

Positive Prompt: Aerial view of city at night filled with colourful lights

Negative Prompt: ---

Figure 6: Generated Image 5

Positive Prompt: Cluster of galaxies revolving around a black hole

Negative Prompt: Satellites

Figure 7: Generated Image 6

Positive Prompt: A futuristic cyberpunk female warrior, glowing neon lights, high-tech cityscape

Negative Prompt: Blurry, artifacts, watermark

4.2 GitHub Link for Code:

Link: https://github.com/kailashmannem/Image-Generation

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Future improvements for this project include optimizing performance for **low-end hardware** by developing **lighter model versions** or enabling **cloud-based processing** for users with limited GPU power. Enhancing the system with **real-time editing and interactive feedback** can allow users to modify images dynamically without restarting the generation process. Expanding capabilities to support **multi-modal inputs** like text, sketches, or reference images would provide greater creative flexibility. Additionally, **fine-tuning with custom datasets** and incorporating advanced techniques like **LoRA and DreamBooth** can improve the accuracy and personalization of generated images. Improving **user experience (UX)** in **ComfyUI** with **preset templates, guided workflows, and AI-powered prompt suggestions** will make the tool more beginner-friendly. Lastly, integrating **APIs with design tools like Photoshop or Figma** can extend its usability for professionals, making AI-powered image generation **more efficient, accessible, and versatile**.

5.2 Conclusion:

This project significantly enhances **AI-powered image generation** by making it more **accessible, user-friendly, and customizable**. By integrating **Stable Diffusion with ComfyUI**, it simplifies the complex process of AI-driven creativity, enabling both beginners and professionals to generate high-quality images effortlessly. The project's impact extends to **graphic design, digital art, content creation, and marketing**, where users can create visuals quickly and efficiently without deep technical expertise. Additionally, it addresses key limitations in existing tools by offering **greater control, flexibility, and a more intuitive interface**.

REFERENCES

[1]. Hugging Face Stable Diffusion Model: https://huggingface.co/stabilityai/stable-diffusion-3.5-large

[2]. ComfyUI GitHub – https://github.com/comfyanonymous/ComfyUI

[3]. OpenAI. DALL-E: Creating Images from Text – https://openai.com/dall-e