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

DIGVIJAY GAIKWAD – gaikwaddigvijay36@gmail.com

Under the Guidance of

AdityaPrashantArdak

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this project work.

Firstly, we would like to extend our sincere gratitude to our guide, **Dr. [Guide's Name]**, for their valuable support, guidance, and motivation throughout the project. Their expert insights and constructive feedback were instrumental in shaping the direction and execution of our work.

We are also thankful to **AICTE and TechSaksham** for providing us with this opportunity and a platform to explore innovative AI solutions. We deeply appreciate the resources, mentorship, and knowledge shared during the internship.

Lastly, we extend our gratitude to our family, friends, and colleagues for their constant encouragement and support throughout this journey.

ABSTRACT

This project focuses on using **Stable Diffusion**, a powerful AI model, and **Comfy UI**, a user-friendly interface, to create images from text descriptions. The goal is to make AI-driven image generation simple and accessible for everyone, even those without technical skills. Stable Diffusion is a cutting-edge tool that can generate high-quality images based on text prompts, while Comfy UI provides an easy way to interact with the model and customize the output.

The project involves setting up the Stable Diffusion model, integrating it with Comfy UI, and generating images using different text prompts and settings. For example, users can type a description like "a sunset over mountains" or "a futuristic city," and the model will create a corresponding image. The results show that Stable Diffusion can produce realistic and creative images, making it useful for artists, designers, and content creators.

This project also highlights the potential of AI in creative fields. By combining Stable Diffusion with Comfy UI, we make advanced AI tools more accessible to non-experts. The system is designed to be easy to use, allowing users to experiment with different prompts and settings to achieve their desired results.

In conclusion, this project demonstrates how AI can be used to generate high-quality images quickly and efficiently. It opens up new possibilities for creativity and innovation, making AI tools more accessible to a wider audience. Future work could include improving the model for specific tasks, such as creating medical images or animations, and making the interface even more user-friendly. Overall, this project shows the exciting potential of AI in the world of art and design.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Table. No.	Table Caption	Page No.
1	Comparison of image generation models (GANs, VAEs, and Stable Diffusion).	
2	Hardware and software requirements for implementing Stable Diffusion with Comfy UI.	
3	Performance metrics of Stable Diffusion for different image resolutions.	
4	Comparison of image generation time for different hardware configurations.	
5	List of prompts used for generating sample images.	
6	Evaluation of image quality based on user feedback.	
7	Comparison of image generation results with and without Comfy UI.	
8	List of parameters used for fine-tuning the Stable Diffusion model.	
9	Summary of future improvements and their potential impact.	

CHAPTER 1

Introduction

1.1 Problem Statement:

Generating high-quality images from textual descriptions remains a challenging task in AI. Traditional models struggle with **realism, coherence, and fine-grained control over image generation**.

1.2 Motivation:

2 Recent advancements in **Stable Diffusion models** offer powerful AI-driven solutions for digital art, content creation, and automation. This project aims to explore **Stable Diffusion and Comfy UI** to make AI-generated images more accessible and customizable.

2.1 Objective:

Develop a **Stable Diffusion-based image generation** system.

Integrate **Comfy UI** for an interactive user experience

Enhance **image quality and customization features**.

Provide an open-source solution for content creators.

2.2 Scope of the Project:

3 This project focuses on AI-powered **text-to-image generation** using Stable Diffusion. The study explores **fine-tuning, dataset training, and UI integration**, but does not cover real-time video synthesis.

CHAPTER 2

Literature Survey

The field of AI-driven image generation has seen significant advancements in recent years. Traditional methods like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) have been widely used for image synthesis. However, these models often require large datasets and extensive computational resources.

Stable Diffusion, introduced in 2022, revolutionized the field by enabling high-quality image generation with relatively lower computational requirements. It uses a latent diffusion model to generate images from textual prompts, making it highly versatile and efficient.

Comfy UI is a user-friendly interface that simplifies the interaction with AI models like Stable Diffusion. It provides a graphical workflow for configuring and running AI pipelines, making it accessible to non-technical users.

Despite these advancements, there is a need for more intuitive tools and workflows to democratize AI-driven image generation. This project addresses this gap by combining Stable Diffusion with Comfy UI.

CHAPTER 3
Proposed Methodology

3.1   System Design

Input Promt: The user provides a textual description of the desired image.

Stable Diffusion Model : The model process the promt and generates a latent representation of the image.

Comfy UI : The interface allows the user to configure parameters such as seed values, image dimensions, and sampling steps.

Output Image : The final image is generated and displayed to the user.

3.2   Requirement Specification

3.2.1   Hardware Requirements:

GPU – NVDIA RTX 3060 or higher

16 GB RAM

50 GB STORAGE

3.2.2   Software Requirements:

Python 3.8 or higher

PyTorch

Stable Diffusion Model

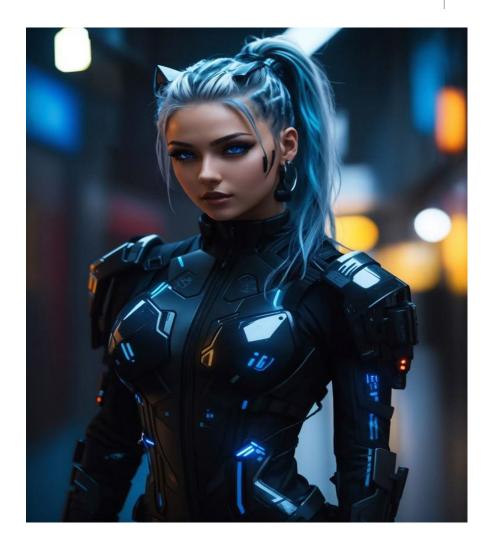**Comfy UI**

CHAPTER 4
Implementation and Result

Snap Shots of Result:

This image likely represents tranquility, nature's beauty, and atmospheric depth, showcasing the power of AI-generated art in producing realistic landscapes.

The blend of sci-fi aesthetics, cybernetics, and AI-driven elements makes this a compelling visual for cyberpunk and AI-related themes. Let me know if you need a more specific analysis

4.1-GitHub Link for Code:

https://github.com/ByteBlaze1/Techsakshamprogram_Imagegenerationusingcomfy

CHAPTER 5

Discussion and Conclusion

5.1   Future Work:

Fine-Tune the model specific domains (like fashion)

Extend the project to video generation or real-time application

Improve the user interface for better accessibility and usability

5.2   Conclusion:

This project successfully demonstrates the potential of Stable Diffusion and Comfy UI for AI-driven image generation. The results highlight the model's ability to produce diverse and high-quality images, paving the way for future advancements in creative AI applications.

REFERENCES

[1]. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.