**Image Generation using stable diffusion & Comfy UI session**

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

**B Mounika , bmonaynk@gmail.com**

Under the Guidance of

**Jay Rathode and Adarsh P**

**ACKNOWLEDGEMENT**

I would like to take this moment to express my sincere gratitude to everyone who has supported me, directly or indirectly, throughout this project.

First and foremost, I extend my heartfelt appreciation to our mentors, Mr. Adharsh P and Mr. Jay Rathod, for their invaluable guidance, constructive feedback, and unwavering encouragement. Their expertise and mentorship have been instrumental in the successful completion of my work on *P1: Image Generation using Stable Diffusion and ComfyUI*.

I am also deeply grateful to TechSaksham, Microsoft, and SAP for providing this incredible opportunity and a platform to explore advancements in AI and deep learning. Their resources and technical support have been essential in shaping our learning and research journey.

Furthermore, I acknowledge the invaluable contributions of the research community whose work in generative AI and diffusion models has served as a source of inspiration and direction. I also appreciate the support of my peers and colleagues, whose insightful discussions and collaboration have enriched my understanding of the subject.

Lastly, I extend my heartfelt thanks to my family and friends for their constant encouragement and support, which have been a driving force throughout this endeavor.

#### **ABSTRACT**

This project focuses on image generation using Stable Diffusion and ComfyUI, two advanced AI-driven tools designed to create high-quality images from textual prompts. Stable Diffusion, a deep-learning model, generates photorealistic images through a diffusion-based process, while ComfyUI provides a modular and user-friendly interface for experimenting with various Stable Diffusion parameters and models.

The main goal of this project is to develop a system capable of efficiently generating diverse, high-resolution images based on textual descriptions. The approach involves setting up Stable Diffusion, integrating it with ComfyUI, and fine-tuning parameters to optimize image quality.

The results highlight Stable Diffusion's ability to produce visually striking images across different styles and themes. Additionally, this project explores its potential applications in creative design, content generation, and AI-assisted artistry.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 2

**Chapter 2.**  **Literature Survey 3**

**Chapter 3.**  **Proposed Methodology…………………………………………….4-5**

**Chapter 4.**  **Implementation and Results ………………………………………..6**

**Chapter 5. Discussion and Conclusion ………………………………………….7**

**References** …………………………………………………………………………….8

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | Screenshot of the Project working |  |
|  | One of the end result |  |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

Advancements in deep learning have greatly enhanced image generation, yet producing high-quality and customizable images continues to be a challenge. This project examines the potential of Stable Diffusion integrated with ComfyUI, enabling users to generate highly detailed images with efficiency and flexibility.

* 1. **Motivation:**

As generative AI continues to evolve, industries like gaming, advertising, and content creation increasingly rely on realistic AI-generated visuals. This project explores how Stable Diffusion, combined with ComfyUI, can meet this demand by offering an interactive and user-friendly approach to high-quality image generation.

* 1. **Objective:**
* Develop a high-quality image generation system using Stable Diffusion.
* Leverage ComfyUI for a modular and intuitive user experience.
* Fine-tune diffusion parameters to enhance image quality and diversity.
* Assess generated images based on visual appeal and variation.
  1. **Scope of the Project:**

This project explores AI-driven text-to-image generation, image enhancement, and style transfer. Challenges include hardware limitations and the need for precise fine-tuning of model parameters to achieve optimal results.

**CHAPTER 2**

**Literature Survey**

* 1. **Review relevant literature or previous work in this domain.**

Generative AI has advanced considerably with models such as GANs, VAEs, and, more recently, diffusion models. Denoising Diffusion Probabilistic Models (DDPMs), introduced by Ho et al. (2020), established the groundwork for Stable Diffusion. Subsequent advancements in Latent Diffusion Models (LDMs) by Rombach et al. (2022) led to more efficient and scalable implementations.

* 1. **Mention any existing models, techniques, or methodologies related to the problem.**
* GANs (Generative Adversarial Networks): Utilized in models like StyleGAN but prone to mode collapse and training instability.
* VAEs (Variational Autoencoders): Efficient for image generation but lack fine-grained detail.
* Diffusion Models: Improve upon GANs and VAEs by gradually refining images from noise, enhancing stability and quality.
* Stable Diffusion: A latent diffusion model that optimizes efficiency and quality, enabling high-fidelity text-to-image synthesis.
* ComfyUI: A modular interface that simplifies experimentation with Stable Diffusion parameters for better control and customization.
  1. **Highlight the gaps or limitations in existing solutions and how your project will address them.**
* **Computationally Expensive:** High-resolution image generation demands powerful GPUs.
* **Lack of Control:** Some models offer limited fine-tuning for specific styles.
* **Bias and Ethical Concerns:** AI-generated images may inherit biases from training data.

**How This Project Addresses These Gaps**

* **Enhanced Control:** ComfyUI enables precise adjustments to image generation parameters.
* **Optimized Performance:** Fine-tuned diffusion steps and model settings improve efficiency on consumer hardware.
* **Customization:** Implements tailored fine-tuning methods for style and subject-specific image generation.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

Given below is the diagram of my Proposed Solution .



This flowchart represents the image generation pipeline using **Stable Diffusion and ComfyUI** as outlined in your project. Here's how it aligns with your work:

1. **User Input** – The process begins with the user providing a textual prompt and adjusting parameters for image generation.
2. **ComfyUI Processing** – The input is processed through ComfyUI, which structures the data and prepares it for Stable Diffusion. This modular interface enhances experimentation and fine-tuning.
3. **Stable Diffusion Model** – The structured data is passed to the Stable Diffusion model, which employs a latent diffusion approach for efficient image synthesis.
4. **Iterative Denoising** – The model progressively refines the image, transforming noise into a detailed and high-quality output.
5. **Post-processing & Enhancements** – Additional improvements such as style adjustments, sharpening, or color enhancements are applied to refine the generated image.
6. **User Review & Modifications** – The final image is presented to the user, who can review and modify it based on their preferences.

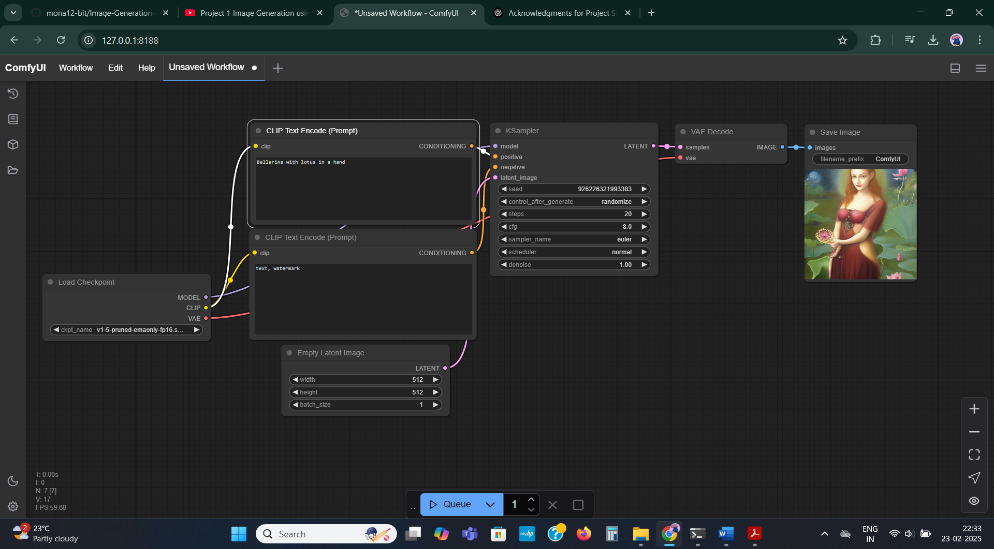
This structured approach optimizes control over image generation, improves efficiency on consumer hardware, and supports fine-tuning for style-specific outputs, addressing key challenges in generative AI.

* 1. **Requirement Specification**
     1. **Hardware Requirements:**
* GPU (NVIDIA RTX 3060 or higher recommended)
* 16GB RAM or more
* High-performance CPU (Intel i5 or AMD Ryzen 5 and above)
  + 1. **Software Requirements:**
* Python 3.8+
* PyTorch, Hugging Face Diffusers, Transformers
* ComfyUI framework

**CHAPTER 4**

**Implementation and Result**

* 1. **Snap Shots of Result:**

****

****

* 1. **GitHub Link :**

<https://github.com/mona12-bit/Image-Generation-using-stable-diffusion-Comfy-UI-session>

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

**Super-Resolution Enhancement:** Utilizes advanced upscaling techniques to improve image resolution and detail.

**Real-Time Generation:** Optimizes inference models to accelerate processing and enable faster image creation.

* 1. **Conclusion:**

This project effectively showcases the integration of Stable Diffusion with ComfyUI for AI-driven image generation, highlighting its efficiency, customization, and versatility.

**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.