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

Sagar Chavan

sagarchavan142003@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I sincerely thank my mentors, **Jay Rathod Sir** and **Aadarsh Sir**, for their invaluable guidance and support throughout this project.

Their insights, encouragement, and constructive feedback have greatly enhanced my understanding and progress. Their confidence in my abilities has been a significant source of motivation, pushing me to think critically and strive for excellence.

Despite the challenges, their mentorship has been incredibly enriching and impactful. I have gained not only technical knowledge but also a deeper appreciation for professionalism and responsibility.

My sincere thanks also go to **Pavan Kumar Sir** for their support and efforts in ensuring the smooth execution of the project.

Finally, I am thankful for the opportunity provided by **TechSaksham Initiative by Microsoft and SAP** for facilitating this learning experience, which has been instrumental in my growth.

ABSTRACT

This project leverages the power of **Stable Diffusion**, an advanced AI model for generating high-quality images from text prompts, through the user-friendly interface of **Comfy UI**. Stable Diffusion has gained prominence for its ability to create detailed and realistic images, and Comfy UI enhances its accessibility by providing an intuitive, modular environment that requires no coding expertise.

The primary aim of this project is to enable users, regardless of technical background, to generate creative and visually appealing images with ease. By combining a sophisticated AI model with a straightforward interface, the project demonstrates how cutting-edge technology can be democratized for broader use cases, including art, design, and content creation.

The essential steps involved in the project include downloading and setting up the **Comfy UI** framework, integrating the Stable Diffusion model into the system by placing it in the designated `models/checkpoints` directory, and using Comfy UI's interface to input prompts and generate images. This process eliminates the need for programming, focusing instead on simplicity and efficiency.

The results showcase the system's ability to generate high-resolution, contextually accurate images based on user prompts, highlighting the practical effectiveness of Stable Diffusion when paired with Comfy UI. The project not only demonstrates the versatility of AI in creative domains but also emphasizes accessibility for non-technical users.

In the future, this project can be expanded by integrating additional models, experimenting with advanced prompt engineering for more refined outputs, and exploring new customization features within Comfy UI to enhance user experience and creative possibilities.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

The process of generating high-quality, realistic images using AI models like Stable Diffusion often requires significant technical expertise, including programming knowledge and familiarity with complex frameworks. This creates a barrier for non-technical users, artists, and creators who wish to utilize these powerful tools for artistic expression, content creation, and design.

The lack of accessible and user-friendly interfaces limits the widespread adoption of AI-driven image generation technologies, despite their immense potential. There is a need for a solution that bridges the gap between advanced AI capabilities and ease of use, enabling users with little to no technical background to harness the power of AI for generating creative and visually appealing images.

This project addresses this problem by leveraging Comfy UI, a modular and intuitive interface for Stable Diffusion, to simplify the image generation process and make it accessible to a broader audience.

1.2 Motivation:

The rapid advancements in AI, particularly in generative models like Stable Diffusion, have revolutionized the fields of art, design, and content creation. However, the potential of these technologies remains largely untapped by non-technical users due to the steep learning curve associated with their implementation.

This project was motivated by the desire to democratize AI-powered image generation, making it accessible to individuals with little to no coding experience. By using Comfy UI, a user-friendly interface, the project seeks to empower artists, creators, and enthusiasts to explore their creativity without being hindered by technical complexities.

Additionally, the motivation stems from the increasing demand for tools that are both powerful and easy to use, enabling people from diverse backgrounds to integrate AI into their workflows. This project not only aims to bridge the gap between advanced technology and usability but also inspires a broader adoption of AI in creative and professional domains.

1.3 Objective:

- Simplify AI-powered image generation using Stable Diffusion through the Comfy UI interface.
- Enable non-technical users to generate high-quality, creative images without requiring coding expertise.
- Provide a user-friendly and modular platform for experimenting with AI-based image generation.
- Promote accessibility and encourage wider adoption of generative AI technologies in art and content creation.
- Explore the potential of Stable Diffusion for various applications, including design, storytelling, and digital art.

1.4 Scope of the Project:

- Provides an accessible AI-powered image generation platform using Stable Diffusion and Comfy UI.
- Enables non-technical users to create high-quality images without coding knowledge.
- Supports various applications, including digital art, design, content creation, and storytelling.
- Allows customization through different AI models and prompt engineering for diverse creative outputs.
- Can be expanded by integrating additional features such as style transfer, upscaling, and multi-modal inputs.
- Offers potential for future improvements in user experience, automation, and integration with other creative tools.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature

The field of AI-driven image generation has seen significant advancements with the development of deep learning models like **Stable Diffusion, DALL·E, and Midjourney**. These models have demonstrated remarkable capabilities in generating high-quality images from text prompts. Stable Diffusion, introduced by **Rombach et al. (2022)**, has been widely adopted due to its efficiency and ability to run on consumer-grade GPUs. Various interfaces, such as **Dream Studio, Automatic1111, and Comfy UI**, have been developed to simplify interaction with these models, enabling broader accessibility.

2.2 Existing Models, Techniques, and Methodologies

- **Stable Diffusion**: A latent diffusion model that generates high-quality images through text prompts, balancing performance and computational efficiency.
- **DALL·E**: Developed by OpenAI, it uses transformer-based architectures to generate creative images but requires API-based access.
- **Midjourney**: A proprietary AI-powered image generation tool that provides artistic and stylized results but lacks open-source accessibility.
- **Comfy UI**: A modular and node-based interface for Stable Diffusion that allows users to visually configure and experiment with model settings without requiring coding skills.

2.3 Gaps and Limitations in Existing Solutions

- **Technical Complexity**: Many AI image generation models require advanced knowledge of Python, ML frameworks, and GPU configurations, making them inaccessible to non-technical users.
- **Limited Customization**: Some existing platforms do not provide full flexibility in model tuning and prompt refinement.
- **Computational Requirements**: Many solutions demand high-end GPUs, restricting accessibility for users with limited hardware resources.

2.4 How This Project Addresses These Gaps

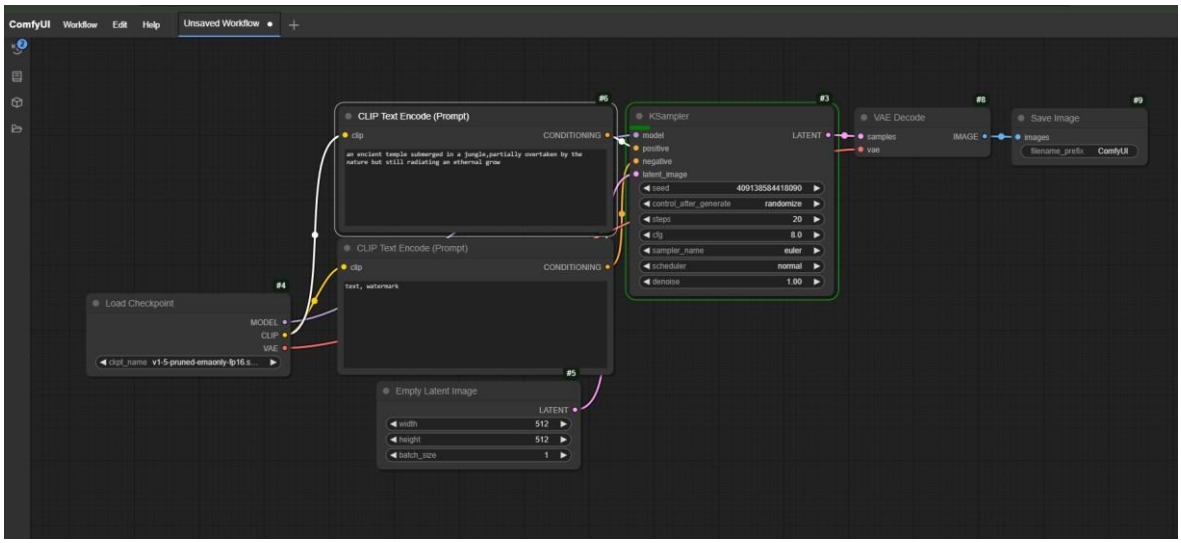
- **User-Friendly Interface**: By leveraging **Comfy UI**, this project eliminates the need for coding and simplifies model interactions.
- **Open-Source & Customizable**: Users can experiment with different models and settings without restrictions.
- **Efficient Resource Utilization**: Comfy UI allows Stable Diffusion to run on consumer GPUs, making AI image generation more accessible.

This project aims to bridge the gap between **technical complexity and usability**, empowering users to generate high-quality images effortlessly.

CHAPTER 3

Proposed Methodology

3.1 System Design



This diagram represents a **Comfy UI workflow** for generating an image using **Stable Diffusion**. Each block (node) has a specific role in the process. Here's a breakdown of each block:

1. Load Checkpoint

- **Purpose:** Loads the pre-trained Stable Diffusion model (v1-5-pruned-emaonly-fp16.safetensors).
- **Inputs:** None.
- **Outputs:**
  - MODEL: The Stable Diffusion model.
  - CLIP: The text encoder (for prompts).
  - VAE: The variational autoencoder (for decoding images).

2. CLIP Text Encode (Prompt)

- **Purpose:** Encodes the input text prompt (an ancient temple submerged in a jungle...) into embeddings that the model can use for image generation.
- **Input:** The CLIP encoder from the Load Checkpoint block.
- **Output:** CONDITIONING data used by the model.

3. CLIP Text Encode (Negative Prompt)

- **Purpose:** Encodes negative prompts (text, watermark) to guide the model in avoiding undesired elements (e.g., watermarks, text overlays) in the generated image.
- **Input:** The CLIP encoder from the Load Checkpoint block.
- **Output:** CONDITIONING data for suppressing undesired features.

4. Empty Latent Image

- **Purpose:** Defines the initial latent space where the image will be generated.
- **Parameters:**
  - width: 512 pixels.
  - height: 512 pixels.
  - batch size: Number of images to generate (1).
- **Output:** LATENT space for image generation.

5. KSampler

- **Purpose:** Generates the image in latent space using the Stable Diffusion model and encoded prompts.
- **Inputs:**
  - model: From Load Checkpoint.
  - positive: From the positive prompt (CLIP Text Encode).
  - negative: From the negative prompt (CLIP Text Encode).
  - latent image: From Empty Latent Image.
- **Parameters:**
  - steps: Number of denoising steps (20).
  - cfg: Guidance scale (8.0).
  - scheduler: Sampling method (Euler).
  - denoise: Strength of denoising (1.00).
- **Output:** LATENT image.

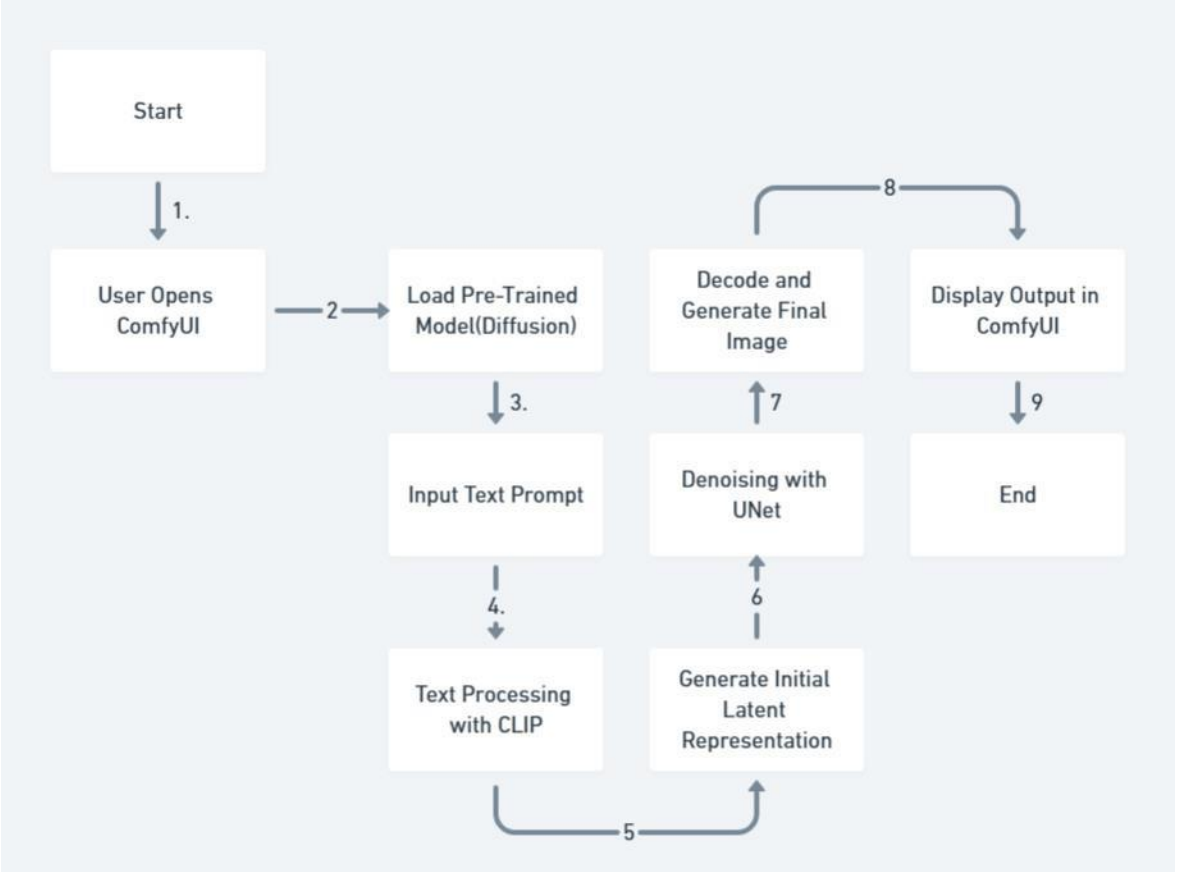
6. VAE Decode

- **Purpose:** Decodes the generated latent image into a final visible image using the VAE.
- **Inputs:**
  - samples: From the KSampler block.
  - vae: From Load Checkpoint.
- **Output:** IMAGE in standard pixel format.

7. Save Image

- **Purpose:** Saves the decoded image to disk.
- **Inputs:**
  - images: From the VAE Decode block.

3.1.1 Flowchart for the Proposed System



1. User inputs text prompt in Comfy UI.
2. CLIP processes text and generates embeddings.
3. Stable Diffusion generates an initial noisy image.
4. U Net refines the image by removing noise step by step.
5. The final image is displayed in Comfy UI for preview or saving.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

Processor (CPU): Intel Core i5 (8th Gen) / AMD Ryzen 5 or equivalent
Graphics Card (GPU): NVIDIA GTX 1660 (6GB VRAM) or AMD equivalent
RAM:16GB DDR4
Storage: 50GB free space (for models and generated images)
Operating System: Windows 10/11, Linux, macOS (with MPS support)

3.2.2 Software Requirements:

Download Comfy UI → Extract it in any folder
Download Stable Diffusion Model (e.g., v1.5, v2.1, SDXL)
Move the Model File (. safetensors) into → Comfy UI\models\checkpoints
Run Comfy UI by double-clicking run_nvidia_gpu.bat (for Windows) or running. /run (Linux/Mac)

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

4.1.1 A futuristic cyberpunk city at night, glowing neon signs reflecting off the rain-soaked streets, flying cars soaring between towering skyscrapers, a lone hooded figure standing under a holographic billboard.

4.1.2 A grand medieval castle on top of a snowy mountain, with huge banners fluttering in the wind, dragons flying in the distance, and a brave knight in shining armor standing at the gates, ready for an epic battle.

4.1.3 An ancient samurai temple hidden in the misty mountains, surrounded by cherry blossoms in full bloom, with a lone monk performing a tea ceremony as the first light of dawn touches the wooden pagoda.

4.1.4 A post-apocalyptic wasteland where nature has reclaimed skyscrapers, vines covering rusted cars, and a lone survivor wearing an exoskeleton suit, exploring the ruins with a robotic dog companion.

4.1.5 A hidden elven village inside an ancient tree, with glowing lanterns hanging from its branches, elegant treehouses connected by wooden bridges, and magical creatures fluttering around under the twilight sky.

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- Enhanced Workflow Features

  Implement additional conditioning models like ControlNet for more precise image generation.
  Add support for style transfer or fine-tuning of outputs to match specific artistic styles.

- Dynamic User Inputs

  Allow users to upload custom models and checkpoints directly via the UI.
  Integrate a text-based prompt helper or AI-powered prompt generator to assist users in crafting creative prompts .

- Real-time Image Editing

  Develop tools to edit or refine generated images in real time, such as cropping, color correction, or inpainting.
  Integrate advanced editing features, like blending multiple generated images or adding overlays.

- Batch Image Generation

  Enable support for batch processing of multiple prompts, generating a gallery of images in one go.
  Automate saving with custom naming conventions or directories.

- Integration with Other Models

  Explore combining Stable Diffusion with generative models for audio, video, or 3D object generation.
  Add compatibility with advanced text-to-image models like DALL·E 3 or Imagen (as they become publicly available).

- Web-Based Deployment

  Build a web-based interface to allow remote access to the system.
  Add cloud support for running the workflow without requiring powerful local hardware.

5.2 Conclusion:

This project demonstrates the powerful capabilities of **Comfy UI** and **Stable Diffusion** for creating stunning AI-generated images using intuitive workflows. By combining advanced models like CLIP and custom samplers, users can effortlessly produce high-quality, creative outputs with minimal effort. The modular design of Comfy UI ensures flexibility, making it suitable for both beginners and advanced users.

Through this tool, we aim to make AI-powered image generation more accessible, interactive, and user-friendly. As we continue to enhance the project with future improvements and community contributions, it has the potential to become a valuable resource for artists, designers, and enthusiasts in the field of AI-driven creativity.

This is just the beginning of what can be achieved. With ongoing innovation and collaboration, the possibilities for pushing the boundaries of AI art are endless.

REFERENCES

[1]https://www.slideshare.net/slideshow/image-generation-with-comfyui-and-stable-diffusion/27392287

[2]https://medium.com/@techlatest.net/setup-and-installation-of-comfy-ui-stable-diffusion-ai-image-generation-made-simple-on-gcp-cf94aa85b9cc

[3] https://scholar.harvard.edu/files/binxuw/files/stable_diffusion_a_tutorial.pdf

[4] https://stability.ai/news/stable-diffusion-3-research-paper