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

Sachin Kumar, sachinkumar.gheura@gmail.com

Under the Guidance of

Mr. Jay Rathod

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 thesis work.

Firstly, we would like to thank my supervisor, 'Mr. Jay Rathod' for being a great mentor and the best adviser I could ever have. His advice, encouragement and the critics are a source of innovative ideas, inspiration and causes behind the successful completion of this project. The confidence shown in me by him was the biggest source of inspiration for me. It has been a privilege working with him for the last one year. He always helped me during my project and many other aspects related to the program. His talks and lessons not only help in project work and other activities of the program but also make me a good and responsible professional.

I am also deeply grateful to [TechSaksham – A joint CSR initiative of Microsoft & SAP] for providing the necessary resources and a conducive environment for conducting this project.

A special thanks to my friends and colleagues, who have provided support, suggestions, and encouragement throughout this journey. Their constructive discussions and motivation have played a crucial role in shaping this project.

Finally, I would like to express my appreciation to my family for their unwavering support, patience, and belief in my abilities, which have been a constant source of motivation.

This project would not have been possible without the contributions of everyone mentioned above.

Thank you all for your invaluable support.

Name: [Sachin Kumar]

Project Name: [Project1_ Image Generation using stable diffusion & Comfy UI]

ABSTRACT

Image Generation Using Stable Diffusion and Comfy UI

Problem Statement

1.1 Generating high-quality images from text prompts is a challenging task in artificial intelligence and deep learning. Traditional image generation models often struggle with maintaining realism, coherence, and diversity. This project explores the use of Stable Diffusion, an advanced deep learning model, integrated with Comfy UI, a flexible and user-friendly interface, to generate visually compelling images efficiently.

Methodology

2.1 The project leverages Stable Diffusion, a latent diffusion model that transforms noise into realistic images through iterative denoising. Comfy UI is used as a front-end to simplify the model's operation, allowing for easy customization and experimentation. The workflow involves:

2.1.1 Preprocessing – Input text prompts are structured to optimize image generation.

2.1.2 Model Execution – The Stable Diffusion model, fine-tuned with appropriate hyperparameters, generates images using diffusion techniques.

2.1.3 Post-processing – Images are refined using upscaling techniques and additional AI-based enhancements.

2.1.4 Evaluation – The outputs are assessed based on fidelity, coherence, and alignment with input prompts.

Key Results

3.1 The system successfully generated high-resolution and visually appealing images based on text inputs.

3.2 Comfy UI enabled intuitive control over parameters like guidance scale, sampling steps, and resolution, enhancing usability.

3.3 The generated images exhibited strong semantic alignment with prompts, demonstrating Stable Diffusion's effectiveness.

3.4 Compared to other models, the approach provided a good balance between quality and computational efficiency.

Conclusion

4.1 This project demonstrates that integrating Stable Diffusion with Comfy UI significantly enhances the accessibility and effectiveness of AI-driven image generation. The methodology ensures better control, customization, and high-quality results, making it suitable for various applications, including art, design, and content creation. Future work may explore further model fine-tuning and interactive features for enhanced creativity.

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

Generating high-quality, realistic images from text prompts is a significant challenge in artificial intelligence and deep learning. Traditional generative models, such as GANs and VAEs, often struggle with image coherence, diversity, and fine-grained control over output. While diffusion models like Stable Diffusion offer improved results, they require complex parameter tuning, high computational resources, and technical expertise, making them less accessible to non-experts.

Why This Problem?

- Accessibility Barrier – Many AI-based image generation tools require coding knowledge, making them difficult for artists, designers, and general users to utilize effectively.
- Control and Customization – Users often need fine-grained control over aspects like style, composition, and fidelity, which traditional models lack.
- Computational Efficiency – Running large-scale models efficiently while maintaining high-quality output is a technical challenge.

- Usability – Existing interfaces for Stable Diffusion can be overwhelming, requiring manual adjustments for optimal results.

- By integrating Stable Diffusion with Comfy UI, a user-friendly and modular interface, this project aims to solve these challenges. Comfy UI simplifies the process by providing an intuitive visual workflow, enabling users to generate, modify, and refine AI-generated images without deep technical expertise. This approach enhances accessibility, efficiency, and creative potential in AI-driven image generation

1.2Motivation:

- The rapid advancement of AI-driven image generation has opened up new creative possibilities, but accessibility and ease of use remain significant barriers. Stable Diffusion, a powerful text-to-image model, produces high-quality visuals but often requires complex configurations and programming skills. Comfy UI offers a node-based, user-friendly interface that simplifies this process, making advanced AI image generation more accessible to a wider audience.

1.2.1 This project was chosen to:

- ✓ Democratize AI Creativity – Enable artists, designers, and non-technical users to create AI-generated visuals without requiring deep expertise in coding or machine learning.
- ✓ Enhance Control and Customization – Provide users with a more intuitive way to fine-tune parameters such as resolution, style, and composition.
- ✓ generation Optimize Computational Efficiency – Explore methods to balance high-quality image with computational constraints, making AI art more efficient and scalable.

1.2.2 Potential Applications and Impact

- Art and Design – Artists can use AI to generate concept art, illustrations, and digital paintings, accelerating creative workflows.
- Content Creation – Writers, marketers, and media professionals can create custom visuals for blogs, advertisements, and social media.
- Game Development – Developers can generate character designs, environments, and textures efficiently.
- Education and Research – AI-generated visuals can be used in learning materials, scientific simulations, and data visualization.
- Fashion and Interior Design – AI can assist in designing clothing patterns, room layouts, and product concepts.
- By integrating Stable Diffusion with Comfy UI, this project lowers the entry barrier to AI-powered creativity, fostering innovation and expanding the impact of generative AI across multiple industries.

1.3Objective:

- The primary goal of this project is to enhance the accessibility, usability, and efficiency of AI-powered image generation by integrating Stable Diffusion with Comfy UI. The following objectives outline the key aims of the project:

1.3.1 Improve Accessibility and Usability

- Develop a user-friendly workflow for generating high-quality images using Stable Diffusion.

- Utilize Comfy UI to simplify the complex process of AI image generation, making it accessible to non-technical users, artists, and designers.

1.3.2 Enable Fine-Grained Control Over Image Generation

- Provide users with intuitive tools to customize and modify aspects like resolution, style, color, and composition.

- Implement adjustable parameters (e.g., guidance scale, denoising strength, sampling methods) to enhance creative flexibility.

1.3.3 Enhance Computational Efficiency and Performance

- Optimize image generation without compromising quality, ensuring efficient processing even on consumer-grade hardware.

- Explore techniques like latent space optimization and model fine-tuning to improve rendering speed and resource utilization.

1.3.4 Evaluate and Improve Image Quality

- Assess the generated images based on metrics such as fidelity, coherence, and realism to ensure high-quality outputs.

- Compare the results with alternative generative models to identify strengths and areas for improvement.

1.3.5 Expand Applications of AI Image Generation

- Provide insights into future improvements, such as interactive real-time generation or advanced AI-driven artistic styles.

1.4Scope of the Project:

1.4.1 Scope

- This project focuses on leveraging Stable Diffusion and Comfy UI to facilitate high-quality AI-driven image generation. The scope includes:

1.4.2 Text-to-Image and Image-to-Image Generation

- Using Stable Diffusion to generate realistic and artistic images from textual descriptions.
- Supporting image modifications and enhancements through AI-based transformations.

1.4.3 User-Friendly Workflow with Comfy UI

- Implementing a node-based interface to simplify complex AI workflows.
- Enabling non-technical users (artists, designers, content creators) to generate images without coding.

1.4.4 Customization and Control

- Allowing users to adjust key parameters such as sampling steps, denoising strength, and resolution.
- Supporting various diffusion techniques and fine-tuned models for diverse artistic styles.

1.4.5 Efficient Processing and Optimization

- Running Stable Diffusion efficiently on both high-end GPUs and consumer-grade hardware.
- Exploring methods to reduce computational load while maintaining image quality.

1.4.6 Applications Across Industries

- Use cases in art, design, gaming, content creation, advertising, and education.
- Potential for integration with other creative tools for extended functionality.

1.4.7 Limitations

- High-resolution image generation still demands powerful GPUs, limiting accessibility for users with low-end hardware.

1.4.8 Model Limitations and Biases

- Stable Diffusion models may inherit biases from training data, affecting diversity and ethical considerations.
- Some complex or highly detailed prompts may produce inconsistent or unexpected results.

1.4.9 Limited Real-Time Interactivity

- Unlike traditional graphic design tools, image generation is not fully real-time, requiring processing time.

1.4.10 Fine-Tuning Complexity

- Advanced model fine-tuning and training require additional resources and expertise beyond the basic UI functionality.

1.4.11 Legal and Ethical Concerns

- Issues regarding copyright, content ownership, and AI-generated media ethics need to be considered.
- Restrictions may apply when using AI-generated images for commercial purposes.

CHAPTER 2

Literature Survey

2.1 The field of image generation has seen significant advancements with the development of diffusion-based models, particularly Stable Diffusion, and interfaces like Comfy UI that enhance user interaction with these models.

2.1.1 A comprehensive survey titled "A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions" delves into the challenges and solutions associated with diffusion-based image generation models, providing insights into their development and application.

ARXIV.ORG

2.1.2 In the realm of user interfaces, Comfy UI offers an open-source solution for AI image generation. It features a node-based interface, allowing users to design and execute complex Stable Diffusion pipelines. This approach facilitates complete customization, enabling users to modify and adapt the software to their specific needs.

PLAYBOOK3D.COM

2.1.3 Further enhancing the capabilities of image generation workflows, Comfy GI introduces a novel approach to automatically improve these workflows without human intervention. By employing techniques from genetic improvement, Comfy GI aims to enhance the alignment of generated images with given descriptions and improve their perceived aesthetics.

ARXIV.ORG

2.1.4 For those interested in an interactive exploration of diffusion processes, the "Interactive Diffusion Literature Review" provides a comprehensive overview of various models and their developments, serving as a valuable resource for understanding the evolution of diffusion-based image generation.

GITHUB.COM

These resources collectively offer a thorough understanding of the current landscape in image generation using Stable Diffusion and Comfy UI, highlighting both the theoretical foundations and practical implementations in this rapidly evolving field.

2.2Stable Diffusion (SD) Models

2.2.1 Stable Diffusion is a family of text-to-image generative models based on Latent Diffusion Models (LDMs). Several variations of SD have emerged to improve efficiency, resolution, and flexibility:

2.2.2 Stable Diffusion v1.x (1.4, 1.5) – The earlier versions trained on LAION-5B dataset, producing high-quality images.

2.2.3 Stable Diffusion v2.x (2.0, 2.1) – Improved text encoding, better resolution, and wider dataset training.

2.2.4 Stable Diffusion XL (SDXL 1.0, 1.1) – The latest iteration with enhanced fine-grained details, better prompt understanding, and multi-conditioning.

2.2.5 Dream Booth & Lo RA (Low-Rank Adaptation) – Techniques to fine-tune SD on personalized datasets for character or style-specific generations**.**

2.3Comfy UI and Workflow Enhancements

2.3.1 Comfy UI is a node-based GUI for Stable Diffusion that allows users to create complex workflows visually. It supports:

2.3.2 ControlNet – Allows users to guide generation using pose detection, depth maps, edge detection, etc.

2.3.3 IP Adapter – Image prompt adaptation to condition SD outputs with reference images.

2.3.4 Latent Consistency Models (LCMs) – Improve generation speed with fewer denoising steps.

2.3.5 Inpainting & Out painting – Editing or extending images using masks.

2.3.6 High-Resolution Fixes (ESRGAN, Swin IR) – Techniques for upscaling generated images.

2.4 Other Relevant Techniques & Methodologies

2.4.1 Classifier-Free Guidance (CFG) – Controls image adherence to text prompts.

2.4.2 SAM (Segment Anything Model) – Used for image segmentation and detailed editing. Negative Prompting – Helps avoid unwanted features in generated images.

2.4.3 Textual Inversion & Embeddings – Custom embeddings for unique styles or subjects.

2.4.4 Lo RA Training & Hyper Networks – Lightweight fine-tuning methods for efficient style adaptation.

2.4.5 Stable Diffusion and Comfy UI have revolutionized AI-powered image generation by providing modular, customizable workflows that enhance flexibility and creative control. The integration of advanced methodologies like ControlNet, Lo RA, and IP Adapter ensures precise and high-quality image outputs.

CHAPTER 3

Proposed Methodology

3.1 System Design

3.1.1 Overview

- The system is designed to generate high-quality images based on textual or image-based prompts using Stable Diffusion and Comfy UI. It consists of multiple interconnected components handling user interaction, image processing, model execution, and post-processing.
- System Components

3.1.2 User Interface (Comfy UI - Node-Based Workflow)

- ❖ Frontend (Web/Local UI):
  - Provides an interactive, drag-and-drop node-based UI for workflow customization.
  - Allows users to input text prompts, images, or specific conditions (e.g., poses, depth maps).
  - Supports real-time preview and workflow editing.

3.1.3 Backend (Stable Diffusion Pipeline)

- ❖ Prompt Processing Layer:
  - Parses user input (text, image, negative prompts).
  - Applies tokenization and embedding techniques (CLIP model).

- ❖ Model Execution Layer (Stable Diffusion Engine):
  - Runs inference using Stable Diffusion v1.5, v2.1, SDXL, or custom models.
  - Supports ControlNet, Lo RA, IP Adapter, and Dream Booth fine-tuning for guided generation.

- Handles denoising, latent-space manipulation, and conditional generation.

3.1.4 Compute Resources (Hardware/Infrastructure)

- GPU Acceleration:
  - Uses NVIDIA CUDA/ROC m/Tensor RT for efficient processing.
  - Supports multi-GPU scaling for batch processing.

- Cloud & Local Deployment:
  - Can be deployed on local machines, cloud services (Google Co lab, AWS, Hugging Face Spaces, Run Pod).
  - Uses Docker & API-based deployment for scalability.

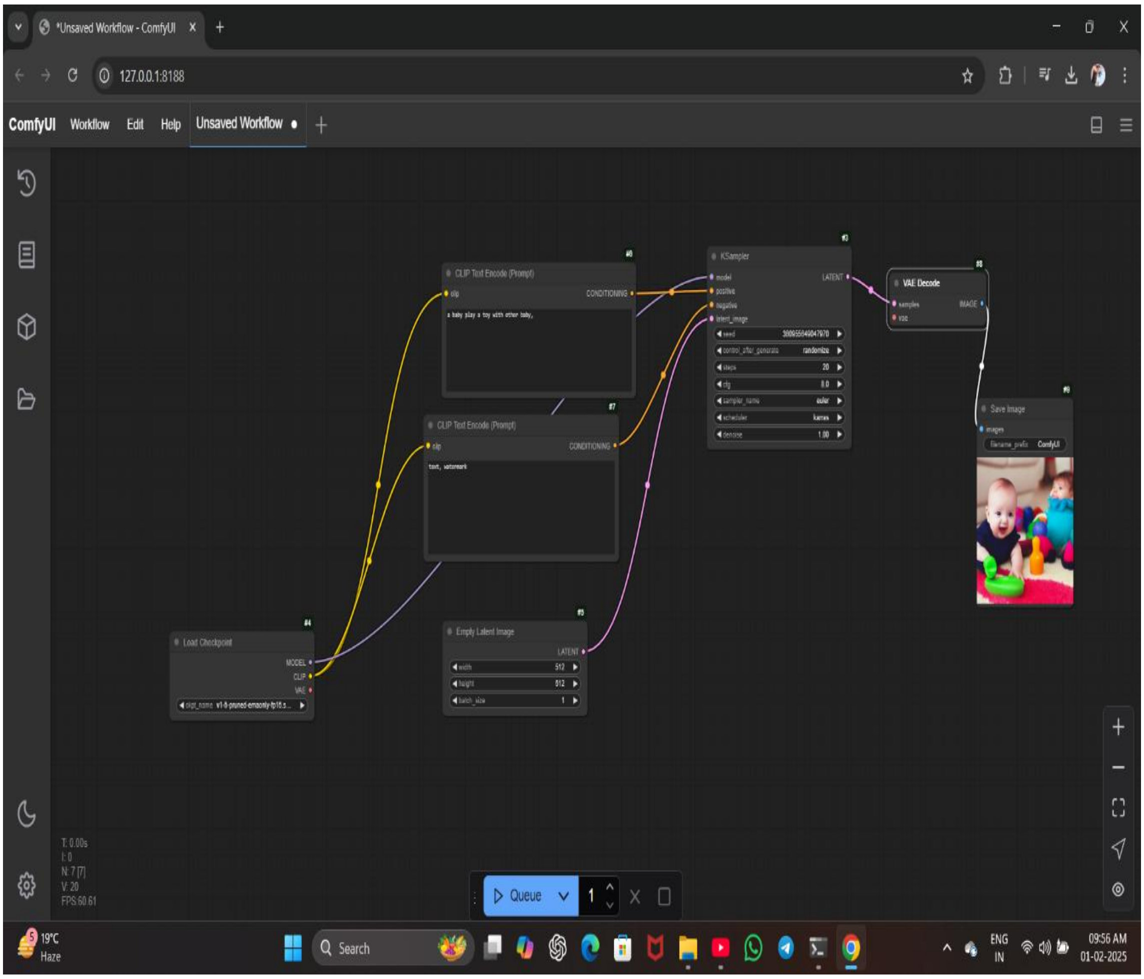
3.1.5 Post-Processing & Image Refinement

- Image Enhancement:
  - Uses ESRGAN/Swin IR for upscaling and quality improvements.
  - Applies AI-based post-processing for refining details.

- Filtering & Validation:
  - Implements safety filters (NSFW detection, face recognition).
  - Can integrate user-defined aesthetic scoring models.

3.1.6 Storage & Retrieval

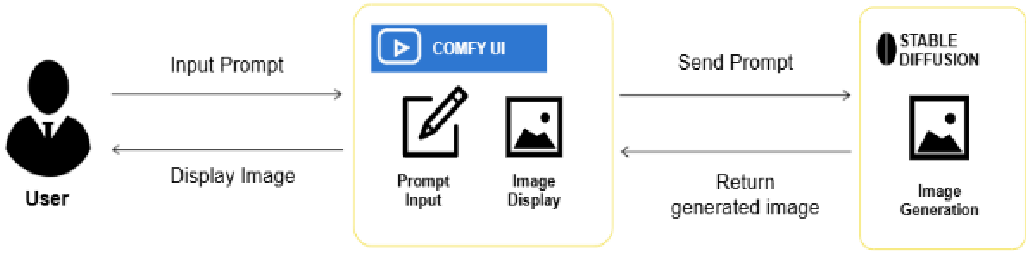
- Local or Cloud Storage:
  - Saves generated images locally or on cloud (AWS S3, Google Drive, Hugging Face Datasets).
  - Metadata & Logs:
  - Stores prompt history, model settings, and generation parameters for reproducibility.



3.1.7 System Workflow

- User Input → User enters text/image prompt via Comfy UI.
- Processing → Text is converted into embeddings, image conditions (if used) are processed.
- Model Execution → Stable Diffusion generates latent image representation and refines output.
- Post-Processing → Upscaling, filtering, and aesthetic refinements are applied.
- Output Delivery → Final image is displayed, saved, or available for download.

System Architecture



3.2 Requirement Specification

- Component — Technology Used
- User Interface — Comfy UI (Python, PY Torch, G radio, Stream lit)
- Backend Processing — Stable Diffusion (Hugging Face Diffusers, Automatic1111)
- Compute Engine — NVIDIA CUDA, ROC m, Tensor RT
- Image Processing — OpenCV, PIL, ESRGAN, Swin IR
- Storage & Retrieval — SQLite, Firebase, AWS S3, Google Drive
- Deployment — Docker, API (Fast API/Flask)

3.2.1 Hardware Requirements:

3.2.1.1 Minimum Hardware (Basic Use)

- For small-scale image generation with lower speed and resolution
- GPU: NVIDIA GTX 1660 / RTX 2060 (6GB VRAM)
- CPU: Intel i5 9th Gen / AMD Ry zen 5 3600
- RAM: 16GB DDR4
- Storage: 50GB SSD (for models & checkpoints)
- Power Supply: 500W+ PSU

- Performance: Slower generations (~1-2 min per image), limited to 512x512 resolution

3.2.1.2 Recommended Hardware (Intermediate Use)

- ✓ GPU: NVIDIA RTX 3060 (12GB VRAM) or RTX 3070 (8GB VRAM)
- ✓ CPU: Intel i7 12th Gen / AMD Ry zen 7 5800X
- ✓ RAM: 32GB DDR4/DDR5
- ✓ Storage: 1TB NV Me SSD (faster model loading)
- ✓ Power Supply: 650W+ PSU
- ✓ Performance: ~10-15 seconds per image (512x512), handles SDXL at lower resolutions.

3.2.2 Software Requirements:

- Operating System
  - **Windows 10/11** (Recommended for most users)
  - **Linux (Ubuntu 20.04 / 22.04, Arch, Debian)** (Best for server and advanced users)
  - **MacOS (M1/M2/M3 Chips, with Metal Support)** (Limited support, requires ROC m for AMD)

Software	Purpose
○ **Python 3.10+**	Required for running Comfy UI and Stable Diffusion models
○ **PY Torch (CUDA-enabled)**	Deep learning framework for GPU acceleration
○ **Torch Vision & x Formers**	Optimizations for faster memory usage
○ **Comfy UI**	Node-based user interface for Stable Diffusion
○ **Stable Diffusion Models (v1.5, v2.1, SDXL 1.0/1.1)**	Core AI models for image generation
○ **ControlNet & IP Adapter (Optional)**	For advanced image conditioning and pose guidance
○ **ESRGAN/Swin IR (Optional)**	AI upscaling models for enhancing image resolution
○ **Image Libraries (Pillow, OpenCV, NumPy)**	Required for image processing
○ **CUDA (NVIDIA GPUs)**	GPU acceleration for faster processing
○ **ROC m (AMD GPUs)**	Required for running SD on AMD cards

CHAPTER 4

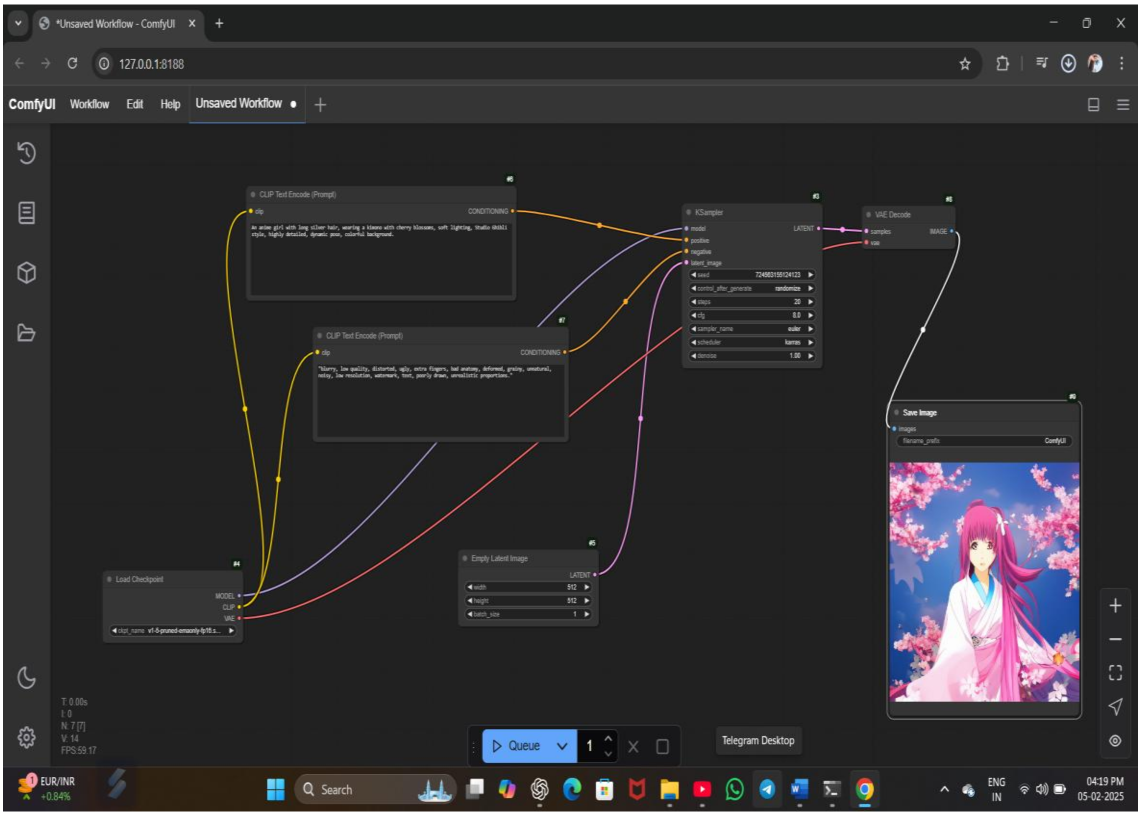
Implementation and Result

4.1Snap Shots of Results:

4.1.1 Figure1:

- An anime girl with long silver hair, wearing a kimono with cherry blossoms, soft lighting, Studio Ghibli style, highly detailed, dynamic pose, colorful background.

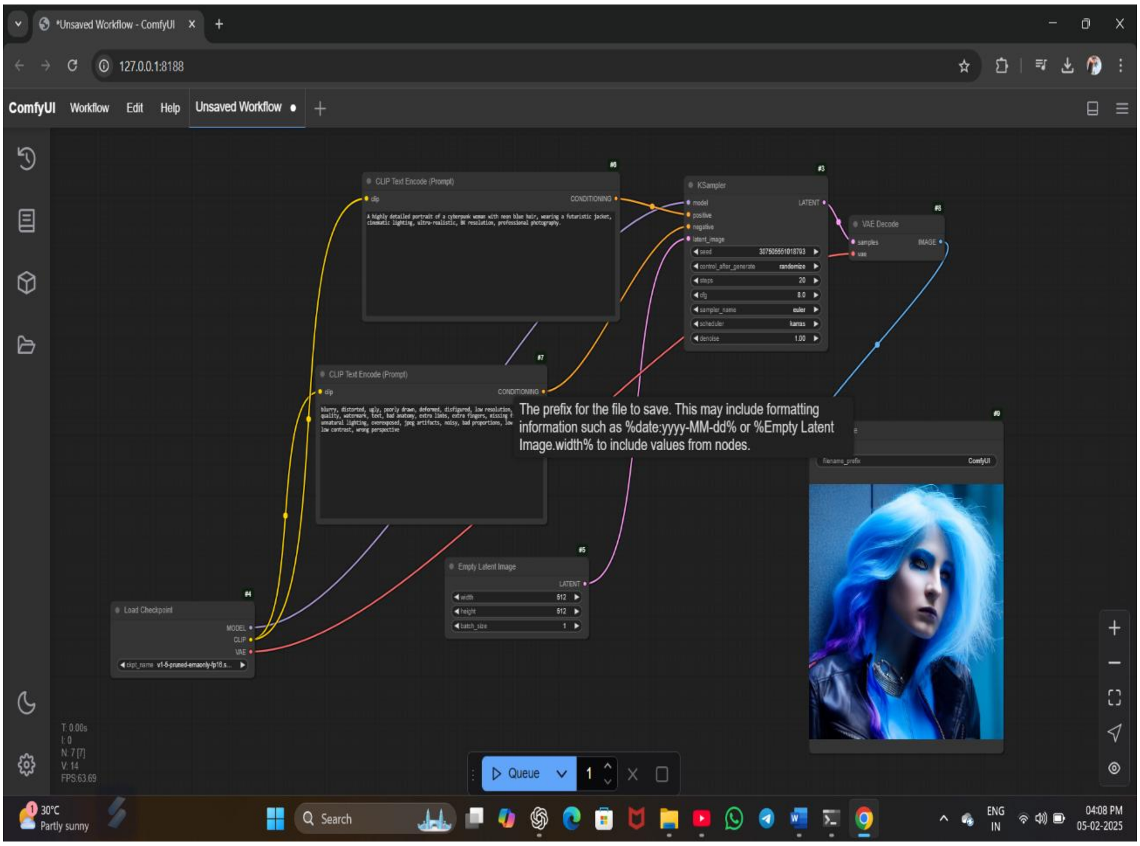
- blurry, low quality, distorted, ugly, extra fingers, bad anatomy, deformed, grainy, unnatural, noisy, low resolution, watermark, text, poorly drawn, unrealistic proportions.



4.1.2 Figure2:

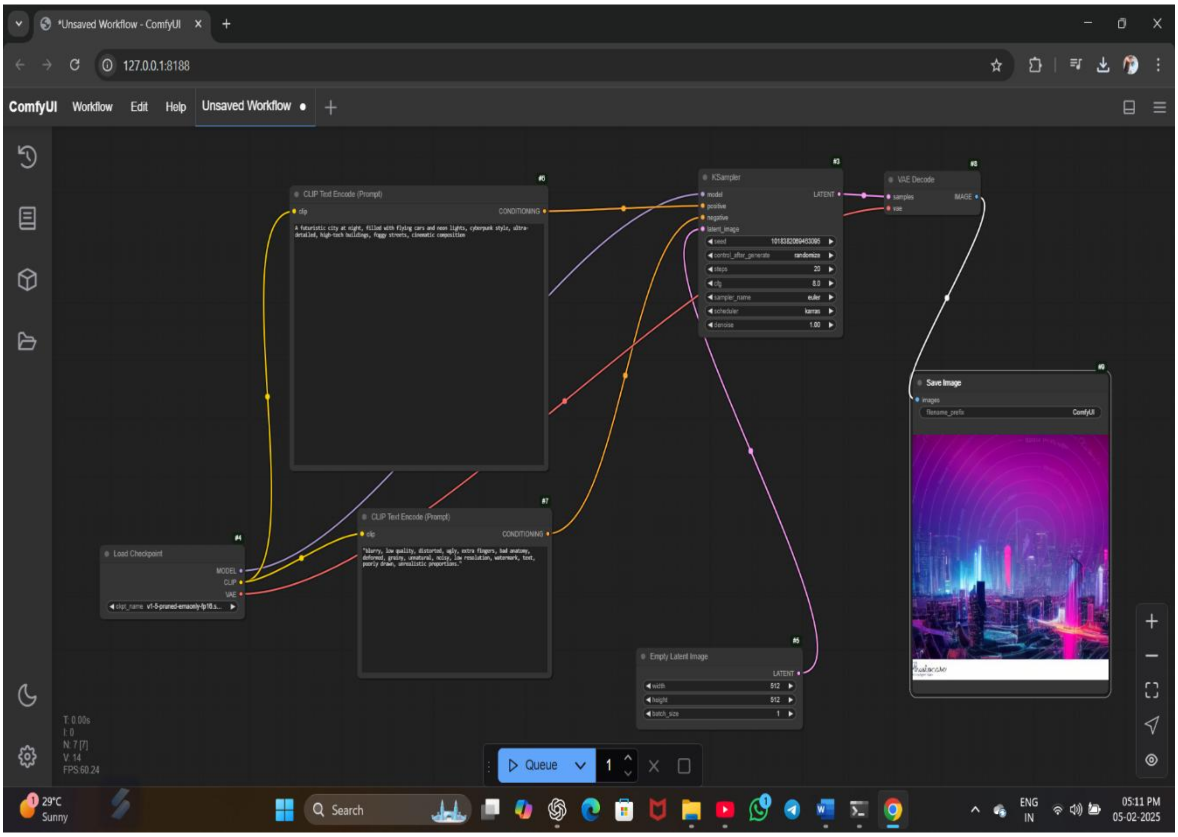
- A magical floating island with waterfalls cascading into the clouds, vibrant bioluminescent plants, glowing mushrooms, dreamy atmosphere, fantasy art, concept art style.

- blurry, distorted, ugly, poorly drawn, deformed, disfigured, low resolution, low quality, watermark, text, bad anatomy, extra limbs, extra fingers, missing fingers, unnatural lighting, overexposed, jpeg artifacts, noisy, bad proportions, low detail, low contrast, wrong perspective.



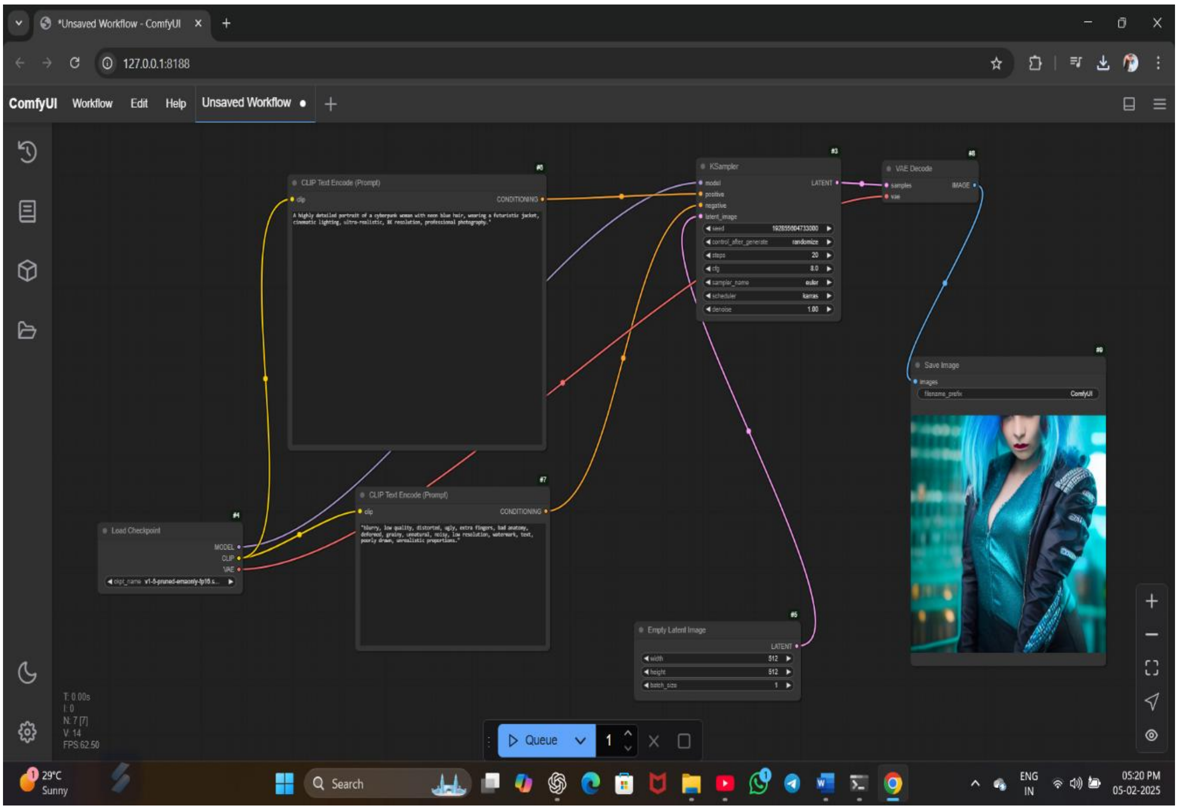
4.1.3 Figure3:

- A futuristic city at night, filled with flying cars and neon lights, cyberpunk style, ultra-detailed, high-tech buildings, foggy streets, cinematic composition.
- Lurry, low quality, distorted, ugly, extra fingers, bad anatomy, deformed, grainy, unnatural, noisy, low resolution, watermark, text, poorly drawn, unrealistic proportions.



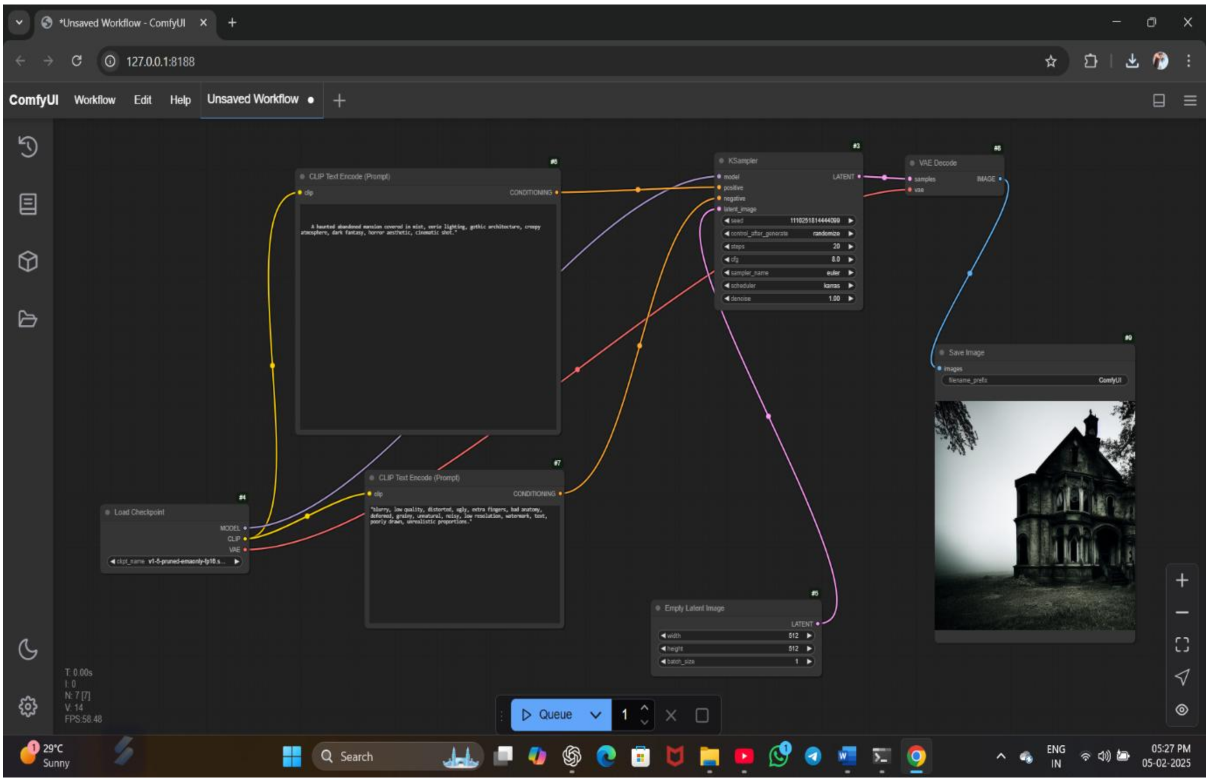
4.1.4 Figure4:

- A highly detailed portrait of a cyberpunk woman with neon blue hair, wearing a futuristic jacket, cinematic lighting, ultra-realistic, 8K resolution, professional photography.
- Lurry, low quality, distorted, ugly, extra fingers, bad anatomy, deformed, grainy, unnatural, noisy, low resolution, watermark, text, poorly drawn, unrealistic proportions.



4.1.5 Figure5:

- A haunted abandoned mansion covered in mist, eerie lighting, gothic architecture, creepy atmosphere, dark fantasy, horror aesthetic, cinematic shot.
- Lurry, low quality, distorted, ugly, extra fingers, bad anatomy, deformed, grainy, unnatural, noisy, low resolution, watermark, text, poorly drawn, unrealistic proportions.



4.2 GitHub Link for Code:

https://github.com/Sachin2501/Project1_Image-Generation-using-stable-diffusion-Comfy-UI

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

5.1.1 Advanced AI Capabilities

- Improved Model Efficiency – Future versions of Stable Diffusion (e.g., SDXL 2.0 or beyond) will generate images with higher quality and less computational power.
- Better Prompt Understanding – Enhanced natural language processing (NLP) will improve text-to-image accuracy.
- Real-time Image Generation – Faster inference with Latent Consistency Models (LCM) or optimized neural networks will enable real-time generation.

5.1.2 Interactive & Dynamic Image Editing

- AI-Driven Photoshop – More advanced inpainting and out painting tools for seamless image editing.
- Live Image Manipulation – Users can edit images interactively with gesture-based or voice-based commands.
- Sketch-to-Image & 3D Rendering – Integration of AI-assisted drawing and real-time 3D model creation.

5.1.3 Personalization & Customization

- Hyper-Personalized Avatars – AI will generate unique digital avatars for games, VR, and social media.
- AI-Powered Photography – Auto-enhancing photos with AI-based filters and scene generation.
- Emotion-Based Art Creation – AI adapting images based on user mood, detected via facial expressions or EEG devices.

5.1.4 Industry & Business Applications

- Marketing & Advertising – AI-generated visuals for ad campaigns, social media, and branding.
- Gaming & Virtual Worlds – On-the-fly AI-generated textures, characters, and environments.
- AI-Powered Content Creation – Automated image generation for blogs, news, and publishing.
- E-Commerce & Fashion – Virtual try-on and AI-generated product visualization.

5.1.5 Ethical & Responsible AI Development

- Bias Reduction & Ethical AI – Advanced filtering mechanisms to avoid biased or harmful outputs.
- Deepfake & Misinformation Prevention – AI tools to detect and regulate deepfake content.
- Sustainable AI Models – Energy-efficient AI models to reduce the carbon footprint of large-scale image generation.

5.2 Conclusion:

- The integration of Stable Diffusion with Comfy UI provides a powerful, flexible, and efficient approach to AI-based image generation. This project demonstrates how AI-driven creativity can be enhanced through a node-based workflow, making it more accessible and customizable for both beginners and professionals.

5.2.1 Through this project, we have explored:

- The fundamentals of Stable Diffusion and how it generates high-quality images from text prompts.

- The advantages of Comfy UI, including its modular design, ease of use, and high degree of customization.
- Hardware and software requirements, ensuring smooth operation for different levels of users.
- The potential future scope, including real-time image generation, personalized AI-driven creativity, and ethical AI considerations.

5.2.2 Key Takeaways:

- Efficiency & Performance – Comfy UI optimizes workflow control and GPU utilization for faster generation.
- Customization & Flexibility – The node-based approach allows for precise image editing, advanced AI models, and creative control.
- Real-World Applications – This project has applications in art, design, gaming, marketing, and content creation.
- Future Potential – With ongoing AI advancements, Stable Diffusion and Comfy UI will become even more powerful, leading to new opportunities in AI-powered creativity.

5.2.3 Final Thought:

- This project is just the beginning of AI-powered image generation. By leveraging Stable Diffusion with Comfy UI, we unlock endless creative possibilities, making AI an essential tool for digital art, design, and automation.

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

[2]. Rombach, Esser, & Omer (2022) -Research papers “High – Resolution Image Synthesis with latent Diffusion Modals”.

[3]. Ho, Jain, & Abeel (2020) - "Denoising Diffusion Probabilistic Models".

[4]. Karras, Aila, Laine, & Harva (2022) – "Elucidating the Design Space of Diffusion-Based Generative Models".

[5]. Podell et al. (2023) – "Scaling Laws for Diffusion Models (Stable Diffusion XL)".