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

Nandakrishnan O,

nandakrishnan1311@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I would like to take this opportunity to express my deep sense of gratitude to all individuals who have contributed to the successful completion of this project. Their guidance, support, and encouragement have been invaluable throughout this journey.

Firstly, I extend my sincere gratitude to my supervisor, **Jay Rathod Sir**, for being a great mentor and advisor. His unwavering support, insightful feedback, and constructive criticism have played a crucial role in shaping this project. His confidence in me has been a source of motivation, and it has been a privilege to work under his guidance. His advice and encouragement not only helped me during this project but also provided valuable lessons that will shape my professional journey.

I am also grateful to **AICTE, TechSaksham, Microsoft, and SAP** for providing this incredible internship opportunity and equipping me with valuable learning resources. Their support has significantly contributed to my knowledge and skills in this domain.

Lastly, I extend my heartfelt thanks to my peers and family for their continuous encouragement and motivation. Their unwavering support has been instrumental in overcoming challenges and staying focused on my goals.

This experience has been a rewarding learning journey, and I sincerely appreciate the contributions of everyone who has been a part of it.

Nandakrishnan O

ABSTRACT

This project explores image generation using **Stable Diffusion** and **Comfy UI**, focusing on the development of AI-generated visuals through deep learning models. With the growing advancements in artificial intelligence, generative models have become powerful tools for creating high-quality, realistic images. Stable Diffusion, a state-of-the-art deep learning model, enables users to generate visually appealing and detailed images from textual prompts. This project aims to optimize the workflow of image generation using Comfy UI, a user-friendly interface designed for seamless interaction with Stable Diffusion.

The primary objective of this study is to harness the capabilities of Stable Diffusion to create high-quality images while leveraging Comfy UI's flexible workflow management. The methodology involves setting up and fine-tuning the model, experimenting with various prompts, adjusting parameters such as guidance scale, sampling methods, and inference steps, and analyzing the generated outputs for quality and consistency. Additionally, the project explores different techniques to enhance image fidelity, such as text embeddings, control nets, and inpainting methods, ensuring better precision and creativity in AI-generated visuals. By systematically evaluating different configurations, the project identifies optimal workflows that enhance efficiency and output quality.

The results demonstrate that Comfy UI significantly improves the image generation process by providing an interactive and modular environment. It simplifies complex AI workflows, making it accessible to users with varying levels of expertise. The project highlights the efficiency and versatility of Stable Diffusion in creating AI-generated media and showcases the impact of deep learning on digital content creation.

This study contributes to the growing field of AI-generated art and media, offering potential applications in digital design, entertainment, game development, and automation. By optimizing workflows and enhancing creative possibilities, this project reinforces the role of artificial intelligence in revolutionizing the way visuals are created and utilized in various industries.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1Problem Statement:

The rise of artificial intelligence in creative fields has revolutionized industries such as **digital art, game development, advertising, filmmaking, and content creation**. AI-driven tools like **Stable Diffusion** have demonstrated remarkable capabilities in generating high-quality, photorealistic images from textual prompts, offering endless possibilities for designers, artists, and developers. However, despite its potential, the adoption of AI-based image generation faces several critical challenges that hinder accessibility, efficiency, and usability.

1. Complexity of Implementation

Deploying deep learning models for image generation is a highly technical process. It requires expertise in **machine learning, Python programming, and GPU optimization**, making it difficult for non-experts to utilize these tools effectively. Many users struggle with installing dependencies, configuring models, and fine-tuning parameters to achieve optimal results.

2. High Computational Demands

AI-generated image models, especially those based on deep learning, demand **significant processing power**. Stable Diffusion, in particular, requires **high-end GPUs** or cloud-based solutions, making it inaccessible for users with limited hardware resources. Efficiently generating high-resolution images without excessive computational overhead remains a major challenge.

3. Lack of Workflow Optimization

Generating AI-based images is not just about running a model—it involves **experimenting with prompts, adjusting parameters like sampling methods, noise levels, and guidance scales, and optimizing model performance**. Traditional command-line-based tools make this process cumbersome, often leading to inefficiencies and inconsistent outputs.

4. Limited Accessibility for Non-Technical Users

While AI-driven image generation has great potential, many existing tools lack a **user-friendly interface**. Without an intuitive way to experiment with different settings and workflows, non-technical users struggle to leverage these tools effectively. There is a growing need for an interface that **removes coding barriers and simplifies the AI workflow**.

Project Solution

This project aims to **bridge the gap** between AI-powered image generation and user accessibility by integrating **Stable Diffusion with Comfy UI**. Comfy UI provides a **node-based, drag-and-drop interface** that makes AI workflows more **intuitive, interactive, and customizable**. By offering a **visual representation of AI pipelines**, users can easily experiment with different configurations without deep coding knowledge.

This project will focus on:

- **Simplifying the deployment** of Stable Diffusion, making AI-generated images more accessible.
- **Optimizing workflows** using Comfy UI to improve efficiency and usability.
- **Reducing computational complexity** by exploring lightweight model configurations and cloud-based alternatives.
- **Enhancing user experience** through structured guides, tutorials, and pre-configured templates.

  By addressing these challenges, this project aims to make AI-driven image generation more **efficient, scalable, and accessible**, empowering a **wider range of users—from artists and designers to researchers and developers—to harness the power of AI in creative content generation**.

1.2Motivation:

The motivation behind this project arises from the increasing **demand for AI-driven creativity and automation** in various industries. AI-generated content has rapidly gained popularity in fields like **digital art, marketing, game development, film production, advertising, and social media content creation**. Tools like **Stable Diffusion** have enabled users to generate high-quality, detailed images from textual descriptions, opening new possibilities for both professionals and hobbyists. However, the **complexity of AI workflows** often limits the accessibility of such technologies.

1. Need for a More Accessible AI Image Generation Process

Current AI-based image generation tools require a **steep learning curve**, as users need to understand **Python, deep learning frameworks, and model fine-tuning**. This creates a barrier for non-technical users, artists, and designers who may not have a background in programming or machine learning. A more **intuitive and user-friendly approach** is essential to **democratize AI image generation** and make it more **accessible to a broader audience**.

2. Simplifying AI Workflows with Comfy UI

Traditional command-line interfaces (CLI) for AI models can be **tedious and unintuitive**, requiring users to manually configure settings and experiment through trial and error. **Comfy UI** offers a **visual, node-based approach**, where users can build AI workflows in a **modular and interactive manner**. This project aims to **leverage Comfy UI to simplify Stable Diffusion workflows**, allowing users to generate and manipulate AI-generated images **without extensive coding knowledge**.

3. Enhancing the Quality and Customization of AI-Generated Images

AI-generated content is not just about creating images—it is about generating **high-quality, customizable, and creative outputs** that meet specific artistic or professional needs. By integrating Comfy UI, this project enables users to **fine-tune model parameters**, apply **control nets**, use **text embeddings**, and explore **inpainting and outpainting techniques**, thereby expanding the **creative potential of AI**.

4. Bridging the Gap Between AI and Creative Industries

AI-powered tools have the potential to **revolutionize digital art, animation, game assets, and product design**. However, many designers and artists find AI technology **intimidating or difficult to use**. By making AI-driven image generation **more approachable**, this project aims to bridge the gap between **deep learning technology and creative professionals**, ensuring that AI becomes a **valuable tool rather than a technical hurdle**.

Project Impact

By developing an optimized workflow using **Stable Diffusion and Comfy UI**, this project:

- **Empowers non-technical users** to explore AI-generated art effortlessly.
- **Enhances productivity** in industries that require rapid content generation.
- **Encourages experimentation** by providing a structured yet flexible AI workflow.
- **Demonstrates the potential** of AI-driven creativity in real-world applications.

  This project is not just about implementing AI image generation—it is about **unlocking new possibilities, enabling creativity, and making AI accessible to everyone, from beginners to professionals**.

1.3Objective:

The primary objective of this project is to **develop an efficient, user-friendly, and accessible AI-driven image generation system** using **Stable Diffusion** and **Comfy UI**. This project aims to bridge the gap between **complex deep learning models and intuitive creative workflows**, making AI-generated content more **usable, customizable, and practical** for a wide range of users.

The key objectives of this project are as follows:

1. Implement Stable Diffusion for AI-Driven Image Generation

- Deploy **Stable Diffusion**, a state-of-the-art deep learning model for text-to-image generation.
- Enable the creation of **high-quality, realistic, and detailed AI-generated images** based on user-defined text prompts.
- Explore different **model versions, checkpoints, and embeddings** to enhance output quality and versatility.

  2. Develop an Optimized Workflow Using Comfy UI

- Utilize **Comfy UI's node-based interface** to design an **interactive and structured AI image generation pipeline**.
- Simplify the workflow by eliminating the need for complex coding, making AI-generated image creation **more accessible to artists, designers, and non-technical users**.

- Experiment with different **processing pipelines, noise schedules, and model parameters** to optimize results.

3. Evaluate and Fine-Tune Models for Improved Image Quality

- Conduct extensive **testing and experimentation** to improve image clarity, resolution, and detail.
- Optimize key parameters like **sampling methods, guidance scale, and denoising strength** to enhance output precision.
- Implement **post-processing techniques** such as **image upscaling, inpainting, and outpainting** to refine AI-generated visuals.

4. Provide a Detailed Guide for Users to Replicate and Customize the Workflow

- Document the entire setup and configuration process, including installation guides for **Stable Diffusion and Comfy UI**.
- Create step-by-step tutorials explaining **how to generate, modify, and fine-tune AI-generated images** using the toolset.
- Offer recommendations on **best practices, troubleshooting methods, and advanced techniques** to help users maximize their creative potential.

5. Explore Future Enhancements and AI Integration Possibilities

- Investigate **additional AI tools and techniques** such as ControlNet, LoRA (Low-Rank Adaptation), and custom-trained models to enhance output flexibility.
- Explore potential **integration with other AI applications** for **real-time generation, automation, or interactive experiences**.
- Analyze possible use cases in **digital marketing, content creation, gaming, virtual reality, and product design**.

Project Significance

By achieving these objectives, this project will **demonstrate the power of AI-driven creativity, improve accessibility to deep learning models, and contribute to the growing field of AI-generated media**. It aims to **empower individuals and businesses** by providing a structured, efficient, and innovative approach to **AI-based image generation**.

1.4Scope of the Project:

This project focuses on the **implementation, optimization, and usability enhancement** of AI-driven image generation using **Stable Diffusion** and **Comfy UI**. It explores various techniques to simplify the workflow, improve image quality, and make AI-generated art more **accessible and practical** for different users, including artists, designers, developers, and researchers.

The scope of this project includes the following key areas:

1. Setting Up Stable Diffusion and Comfy UI

- Installation and configuration of **Stable Diffusion** for **text-to-image generation**.
- Implementation of **Comfy UI** to provide a **node-based, drag-and-drop interface** for enhanced usability.
- Exploring different **model architectures, pre-trained weights, and GPU optimization techniques** for efficient image processing.

2. Exploring Various Image Generation Techniques

- Experimenting with **different prompts, model variations, and control parameters** to understand their impact on image quality.
- Utilizing **custom embeddings, hypernetwork models, and text encoders** for enhanced AI creativity.
- Applying **ControlNet, LoRA (Low-Rank Adaptation), and fine-tuning** to customize image outputs based on specific artistic requirements.

3. Analyzing the Performance and Quality of Generated Images

- Evaluating key factors such as **image resolution, texture details, color accuracy, and realism**.
- Comparing different **sampling methods (Euler, DDIM, LMS, etc.)** to determine their effect on output efficiency.
- Optimizing settings such as **guidance scale, denoising strength, and inference speed** for better performance.

4. Enhancing Workflow Efficiency and Usability

- Streamlining the **image generation process** to make it accessible for both technical and non-technical users.
- Developing a **user guide** to assist users in customizing their workflows with minimal coding knowledge.
- Exploring **automated image enhancement techniques**, such as **upscaling, inpainting, and background removal**.

5. Potential Enhancements and Future Integrations

- Investigating the **integration of Stable Diffusion with other AI tools**, such as **chatbots, automation scripts, and VR applications**.
- Exploring potential applications in **content creation, marketing, digital art, gaming, and product design**.
- Analyzing the feasibility of **real-time AI-assisted image generation** for dynamic and interactive use cases.

Out of Scope

- The project does not focus on **developing a new AI model from scratch** but instead utilizes **pre-trained Stable Diffusion models**.
- Ethical considerations such as **AI bias, deepfake generation, and misuse of AI-generated images** are acknowledged but not the primary focus of this study.
- The project does not cover **hardware development** but instead explores software-based optimizations for existing GPU and cloud computing infrastructures.

Project Significance

By defining this scope, the project ensures a **focused and structured approach** to AI-generated image creation. It aims to **bridge the gap between deep learning models and practical applications**, providing users with an **efficient, accessible, and high-quality AI image generation workflow**.

Limitations of the Project

While this project successfully demonstrates AI-driven image generation using **Stable Diffusion** and **Comfy UI**, several limitations must be acknowledged. These limitations arise due to technical constraints, hardware requirements, and the inherent challenges of deep learning models.

1. High Computational Requirements

- Stable Diffusion models require **high-performance GPUs with significant VRAM** for smooth processing.
- Running on lower-end hardware may result in **slower image generation, increased processing time, or limited model capabilities**.
- Cloud-based solutions can be costly, limiting accessibility for users without dedicated hardware.

2. Dependence on Pre-Trained Models

- The project does not involve training a new AI model but instead **relies on existing pre-trained Stable Diffusion models**.
- The quality and style of generated images are **limited by the capabilities of these models**.
- Custom model fine-tuning is possible but requires **additional computational resources and expertise**.

3. Bias and Ethical Considerations

- AI models, including Stable Diffusion, may **inherit biases from their training data**, affecting the diversity and representation of generated images.
- Certain **prompt restrictions** may prevent the generation of specific content due to ethical and policy constraints.
- The potential for **misuse in deepfake generation or unethical AI applications** remains a concern.

4. Complexity in Achieving Desired Outputs

- The effectiveness of image generation depends on **prompt engineering**—users may need **trial and error** to refine prompts for optimal results.
- Some images may **lack consistency, coherence, or fine details**, requiring additional post-processing.
- Advanced features like **ControlNet, LoRA, or DreamBooth fine-tuning** require deeper technical knowledge, making them less accessible to beginners.

5. Limited Real-Time Capabilities

- While the project focuses on batch image generation, it **does not support real-time AI image synthesis** for interactive applications like video generation or live rendering.
- Integration with **real-time AI workflows** may require additional tools and optimizations beyond the scope of this project.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature

Image generation using deep learning has evolved significantly, with advancements in Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and most recently, diffusion models. GANs, introduced by Goodfellow et al. (2014), revolutionized the field by enabling high-quality synthetic image generation. VAEs (Kingma & Welling, 2013) provided a probabilistic approach to image synthesis, though they often suffered from blurry outputs. However, both methods had inherent challenges, such as mode collapse in GANs and blurry outputs in VAEs.

Diffusion models, particularly Stable Diffusion, introduced by Rombach et al. (2022), have set new benchmarks in high-resolution, text-to-image generation. Unlike GANs and VAEs, diffusion models use a denoising process to iteratively refine an image from a random noise distribution, resulting in photorealistic and detailed outputs. Stable Diffusion, an open-source latent diffusion model, has gained widespread popularity due to its ability to generate high-quality images with relatively lower computational cost compared to models like DALL·E (Ramesh et al., 2022).

2.2 Existing Models, Techniques, and Methodologies

Several models and frameworks exist for text-to-image generation:

- **DALL·E:** Developed by OpenAI, DALL·E and its successor, DALL·E 2, use transformer-based models for generating images from textual descriptions. While powerful, they are not open-source and require significant computational resources.
- **Stable Diffusion:** Unlike DALL·E, Stable Diffusion is open-source and allows local execution on consumer-grade GPUs, making it accessible to a broader audience. It uses latent diffusion models (LDMs) to generate images efficiently.
- **MidJourney:** A commercial AI-driven text-to-image generator, producing artistic images with high aesthetic appeal.

- **ComfyUI:** A modular and node-based graphical interface for Stable Diffusion, allowing users to experiment with various diffusion settings, model parameters, and workflows without requiring extensive programming knowledge.

2.3 Gaps and Limitations in Existing Solutions

Despite the advancements in diffusion models, several challenges remain:

- **High Computational Requirements:** Although Stable Diffusion is more efficient than its predecessors, generating high-quality images still demands considerable GPU resources.
- **Limited Control Over Outputs:** While textual prompts guide image generation, fine-grained control over specific attributes remains a challenge. ComfyUI aims to improve this by providing a visual workflow for enhanced control.
- **Ethical and Bias Concerns**: AI-generated images can sometimes reflect biases present in training data, leading to ethical concerns about fairness and misuse.

Proposed Contributions

- By leveraging **ComfyUI**, our project aims to enhance the usability and accessibility of Stable Diffusion, making high-quality image generation more intuitive.
- We will explore optimizations that reduce computational overhead while maintaining output quality.
- Our project will experiment with **custom models and fine-tuning techniques** to improve control over image generation and mitigate biases.

This work seeks to bridge the gap between powerful AI-driven image generation and user-friendly accessibility through an intuitive UI and optimized workflows.

CHAPTER 3

Proposed Methodology

3.1 System Design

Proposed Solution

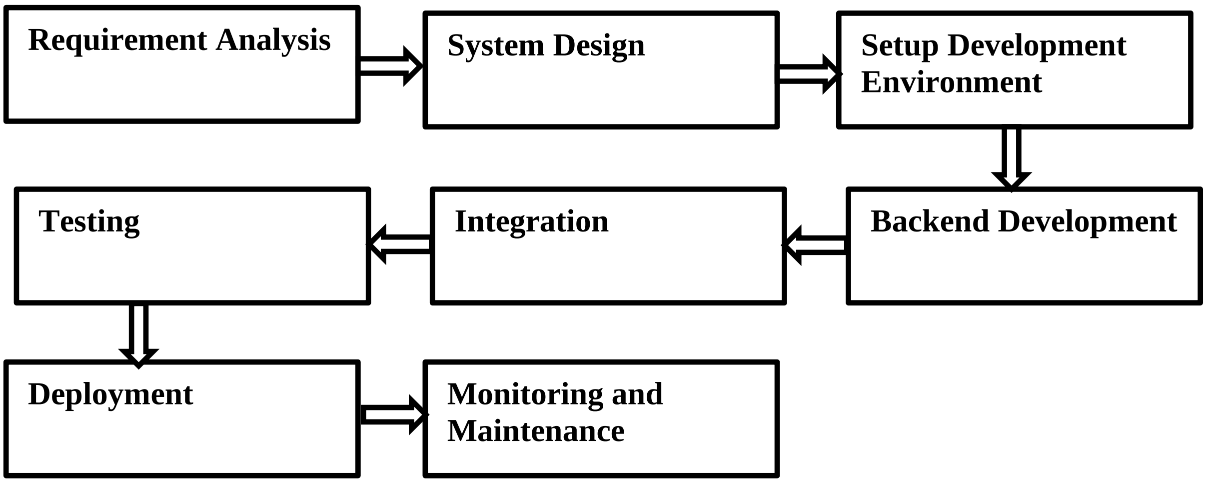

Figure 1:Block diagram of system design

Explanation of the Diagram

1. **Requirement Analysis**: Identify project needs, define objectives, and analyze available resources.
2. **System Design**: Plan the architecture, choose tools, and create data flow diagrams.
3. **Setup Development Environment**: Install dependencies, configure Stable Diffusion and ComfyUI.
4. **Backend Development**: Implement APIs and integrate Stable Diffusion for image generation.
5. **Integration**: Connect all components, ensuring seamless data flow and usability.
6. **Testing**: Conduct unit tests, performance tests, and bug fixes.
7. **Deployment**: Deploy the model and UI on cloud platforms or local servers.
8. **Monitoring and Maintenance**: Track performance, fix bugs, and update models for optimization.

This structured approach ensures efficient development, deployment, and maintenance of the project.

System Architecture

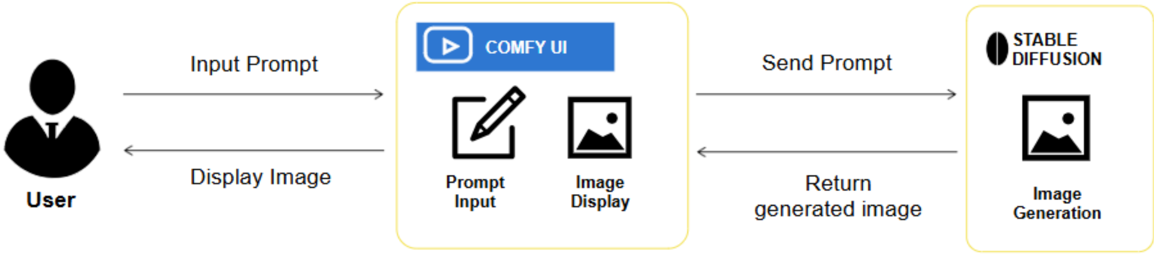

Figure 2:System Architecture

Explanation of the Diagram

The provided diagram represents the workflow of the image generation system using **Stable Diffusion** and **ComfyUI**. It consists of three main components:

1. **User**:
   - The user provides an **input prompt** (text describing the desired image).
   - Once the image is generated, it is displayed back to the user.
2. **ComfyUI** (Graphical Interface):
   - ComfyUI acts as the intermediary between the user and Stable Diffusion.
   - It receives the input prompt from the user.
   - The UI forwards the prompt to Stable Diffusion for image generation.
   - After receiving the generated image, it displays it to the user.
3. **Stable Diffusion** (Image Generation Model):
   - Stable Diffusion processes the input prompt and generates a corresponding image.
   - The generated image is returned to ComfyUI for display.

Workflow:

1. **User** enters an input prompt into ComfyUI.
2. **ComfyUI** forwards the prompt to the Stable Diffusion model.
3. **Stable Diffusion** processes the prompt using deep learning and generates an image.
4. The generated image is sent back to **ComfyUI**.
5. **ComfyUI** displays the generated image to the **User**.

This system ensures an interactive and user-friendly way to generate AI-based images efficiently.

Features and Functionalities

Feature	Description
Text-to-Image Generation	Generates high-quality images based on text input.
Customizable Parameters	Allows users to modify model settings such as resolution, steps, and style.
Batch Processing	Supports generating multiple images at once.
Pre-Trained Models	Uses Stable Diffusion pre-trained models for efficient generation.
User-Friendly UI	ComfyUI provides an easy-to-use graphical interface for interaction.
Cloud-Based Execution	Can run on Google Colab for GPU-accelerated processing.

Table 1: Features and Functionalities

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

- **Processor:** Minimum Intel Core i5 (Recommended: i7 or Ryzen 7)
- **RAM:** Minimum 16GB (Recommended: 32GB for better performance)
- **GPU:** NVIDIA GPU with at least 6GB VRAM (Recommended: 12GB or more for faster inference)
- **Storage:** Minimum 50GB free space (Recommended: SSD with 100GB free space)
- **Cloud Computing (Alternative):** Google Colab Pro+ for access to high-performance GPUs

3.2.2 Software Requirements:

- **Operating System**: Windows 10/11, Linux (Ubuntu 20.04+), or macOS
- **Python**: Version 3.8 or later

- **CUDA & cuDNN**: Required for GPU acceleration (compatible with installed GPU driver)
- **PyTorch**: Latest stable version with GPU support
- **ComfyUI**: Installed and configured with Stable Diffusion models
- **Git & GitHub**: For version control and collaboration
- **NumPy** - For numerical data operations and array manipulation
- **PyTorch** - For building and deploying machine learning and deep learning models.
- **Transformers** - Architecture for the Stable diffusion model to generate the image
- **Pillow** - Image handling library.
- **Flask** - Flask is a lightweight and flexible Python web framework.
- **Requests** - Requests is a popular Python library for making HTTP requests (API Call)
- **TQDM**
- **Filelock**
- **Gradio**
- **Omegaconf**

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

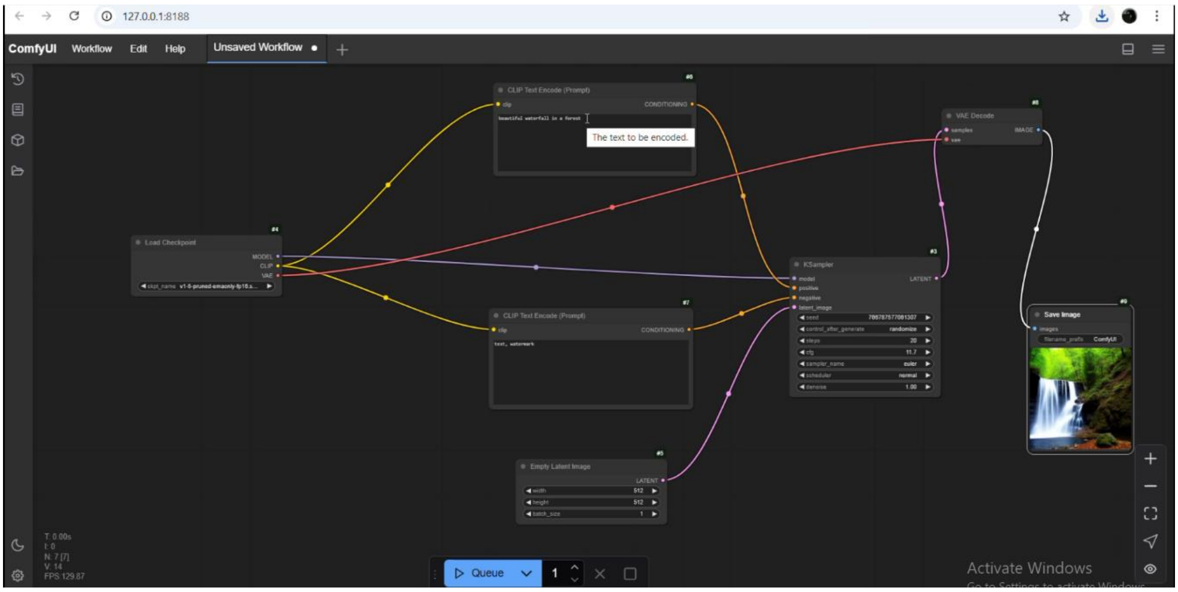

Figure 3:Output generating

Figure 4:Generated output of bottle which Filled color water.

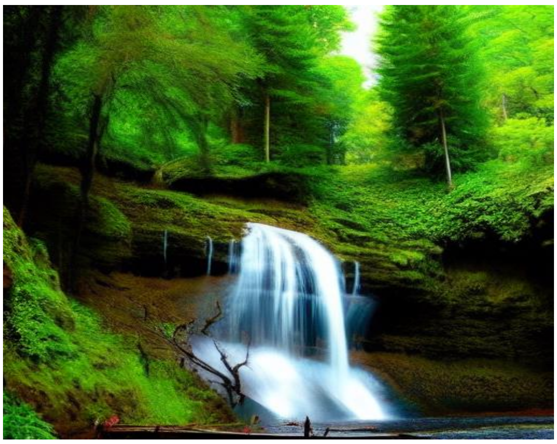

Figure 5 Waterfall inside forest(output)

Figure 6:Fish in clear water (output) **Figure 7:House in forest(output)**

4.2 GitHub Link for Code:

https://github.com/nandakrishnan1311/techsaksham

4.3Demo URL:

https://www.youtube.com/watch?v=Y-Nbj4TJ3ZA

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Future research and development can focus on several key areas to further improve the capabilities and usability of the model:

Enhanced Prompt Engineering: Improving methods for refining textual inputs to achieve more accurate and contextually relevant image generation. Implementing prompt enhancement techniques using natural language processing (NLP) can lead to more controlled and fine-tuned results.

Fine-tuning for Specific Domains: Training custom Stable Diffusion models tailored for particular industries such as medical imaging, fashion design, gaming, and advertising can lead to more specialized and accurate image outputs.

Reducing Computational Overhead: Optimizing the inference process through model pruning, quantization, or leveraging more efficient neural network architectures can reduce resource consumption and make high-quality image generation feasible on lower-end hardware, including mobile devices.

Interactive Real-time Modifications: Developing real-time interaction features where users can modify generated images dynamically, adjusting parameters such as colors, composition, lighting, and details, without having to regenerate images from scratch.

Addressing Bias and Ethical Concerns: Establishing bias-detection mechanisms, ensuring datasets are diverse, and creating ethical guidelines for AI-generated content will be essential in mitigating fairness-related challenges.

Integration with Other AI Technologies: Combining Stable Diffusion with other AI models, such as large language models (LLMs) and multimodal AI, can lead to more advanced applications, including interactive storytelling, AI-assisted design, and content generation for virtual and augmented reality.

Improved User Interfaces and Accessibility: Enhancing platforms like ComfyUI to support better user experience with drag-and-drop functionalities, guided tutorials, and integration with cloud-based services for users without high-end hardware.

5.2 Conclusion:

This project explores the advancements in text-to-image generation using Stable Diffusion and the enhanced usability provided by ComfyUI. By integrating a user-friendly interface with powerful image synthesis capabilities, the project contributes to the democratization of AI-driven image generation. The ability to fine-tune models and control the generation process enhances both creativity and practical applications in various fields.

Despite existing challenges such as high computational requirements and ethical concerns, the project sets the foundation for future improvements. The proposed enhancements in prompt engineering, computational efficiency, and ethical considerations will help refine the model further. Ultimately, this work bridges the gap between advanced AI-driven image generation and intuitive user accessibility, paving the way for broader adoption and innovation in the field.

REFERENCES

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Nets", Advances in Neural Information Processing Systems, 2014.

[2] Diederik P. Kingma, Max Welling, "Auto-Encoding Variational Bayes", International Conference on Learning Representations (ICLR), 2013.

[3] Jonathan Ho, Ajay Jain, Pieter Abbeel, "Denoising Diffusion Probabilistic Models", Advances in Neural Information Processing Systems, 2020.

[4] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[5] Prafulla Dhariwal, Alexander Nichol, "Diffusion Models Beat GANs on Image Synthesis", Advances in Neural Information Processing Systems, 2021.

[6] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever, "Zero-Shot Text-to-Image Generation", International Conference on Machine Learning (ICML), 2021.

[7] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, "Learning Transferable Visual Models From Natural Language Supervision", International Conference on Machine Learning (ICML), 2021.

[8] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Jamie Brew, "Transformers: State-of-the-Art Natural Language Processing", Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

[9] Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[10] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, Ben Poole, "Score-Based Generative Modeling through

Stochastic Differential Equations", International Conference on Learning Representations (ICLR), 2021.