

# Image Generation using stable diffusion & Comfy UI (P1)

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

**Spoorthy Boga, spoorthymaharaju@gmail.com**

Under the Guidance of

**Jay Rathod**

## ACKNOWLEDGEMENT

---

I would like to take this opportunity to express my gratitude for this opportunity to do this internship and gain experience in the field of image generation.

Firstly, I would like to thank my supervisor, ..... for being a great mentor. Through the entire process and the project, it has been extremely helpful and easy to proceed with the work due to your guidance. This project has helped me step into the field of ML and has a great impact on my course. I am extremely grateful for this opportunity and I would like to thank all the mentors who were there through and through the project. Thank you.

## ABSTRACT

---

This project aims to create a simple and effective image generation system by combining **Stable Diffusion** with **Comfy UI**. It allows users to make a variety of high-quality images just by using text prompts. The main goal is to make AI image generation easy for everyone, even those without technical skills, while tackling issues like complexity, limited control, and slow workflows. By using a pre-trained Stable Diffusion model (*v1-5-pruned-emaonly-fp16.safetensors*) and improving its performance, the system makes image creation easier and more enjoyable.

Comfy UI offers a user-friendly interface that lets users adjust image settings like style, resolution, and prompts, giving them more control over the results. The system makes sure that the images produced match the users' prompts and desired styles, speeding up the process and reducing the need for manual changes.

Key results from the project include the successful creation of over 10 high-quality images, a 20% faster processing time, and improved access for non-technical users. This project shows the potential of merging Stable Diffusion with Comfy UI to build a flexible and user-friendly image generation tool.

Looking ahead, there are plans to develop models for higher-resolution images, connect the system with other creative tools, and enhance Comfy UI for an even better user experience. By tackling current limitations and focusing on accessibility, this project opens the door for more people to use AI image generation in art, design, and multimedia, allowing them to create unique visuals easily.

## TABLE OF CONTENT

---

<b>Abstract</b>	<b>I</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Problem Statement	1
1.2 Motivation	1
1.3 Objectives	2
1.4. Scope of the Project	2
<b>Chapter 2. Literature Survey</b>	<b>3</b>
<b>Chapter 3. Proposed Methodology</b>	
<b>Chapter 4. Implementation and Results</b>	
<b>Chapter 5. Discussion and Conclusion</b>	
<b>References</b>	

## LIST OF FIGURES

Figure No.	Figure Caption	Page No.
<b>Figure 1</b>	beautiful scenery nature glass bottle landscape, yellow galaxy bottle	7
<b>Figure 2</b>	A butterfly in a garden with children playing around and a dog running around	7
<b>Figure 3</b>	A cozy, neon-lit tea house perched on a floating platform above a futuristic city. The sky is filled with flying cars, holographic billboards, and a glowing sunset. The tea house has intricate Japanese architecture blended with modern cyberpunk aesthetics.	7
<b>Figure 4</b>	A massive, forgotten library overgrown with vines and flowers, hidden deep within a mystical jungle. Sunlight filters through the canopy, casting golden rays on ancient books and scrolls. Exotic birds and small glowing creatures flutter around.	7
<b>Figure 5</b>	A surreal scene of an astronaut sitting on a rock, engaged in an intense chess match with a translucent, multi-limbed alien. The Martian landscape stretches endlessly behind them, with a massive gas giant visible in the sky.	8
<b>Figure 6</b>	A warrior clad in intricate bronze and brass armor, wielding a mechanical crossbow, standing on a cliffside. A giant steam-powered dragon with glowing blue eyes hovers above, its wings made of metal gears and steam vents. The background features a gothic city with zeppelins flying overhead.	8
<b>Figure 7</b>	A fluffy, oversized cat curled around a small medieval village, acting as its gentle guardian. The villagers go about their day, using the cat's fur as pathways and its tail as a bridge over a river. The scene is warm and whimsical, with a soft, golden sunset.	8
<b>Figure 8</b>	A futuristic metropolis built entirely beneath the ocean, glowing in shades of neon blue and purple. Transparent domes house high-tech buildings, with schools of glowing jellyfish drifting between them. Advanced submarines and humanoid robots navigate the vibrant coral streets.	8
<b>Figure 9</b>	A mysterious, dimly lit study filled with floating objects from different eras—an ancient Egyptian tablet, a futuristic energy weapon, a medieval crown, and a vinyl record spinning in mid-air. The walls are lined with bookshelves, and a glowing portal flickers in the corner.	9

## CHAPTER 1

### Introduction

#### 1.1 Problem Statement:

While Stable Diffusion models are really powerful, they can be tricky to use without some technical know-how. Users often struggle with customizing image settings, managing their projects, and getting the results they expect. This project is all about making things easier by combining Stable Diffusion with Comfy UI, a user-friendly interface, so more people can enjoy the benefits.

#### 1.2 Motivation:

This project aims to make AI tools like Stable Diffusion easy for everyone to use, even without technical skills. By combining it with a simple interface (Comfy UI), it lets users create custom, high-quality images quickly. It also reduces the need for expensive hardware. The goal is to help artists, designers, and hobbyists explore creativity without hassle. By simplifying the process, this project makes AI-powered art accessible and fun for everyone, encouraging innovation and practical use in everyday creative work.

#### 1.3 Objective:

- User Control: Let users personalize image settings like style, resolution, and prompts.
- Streamlined Workflow: Make the image generation process smoother with an easy-to-use interface.
- Output Alignment: Ensure the generated images closely reflect the prompts and styles users want.

#### 1.4 Scope of the Project:

##### Scope:

- User-Friendly Image Generation: Create a system that combines Stable Diffusion with Comfy UI, making it easy for anyone to generate high-quality images.
- Customization: Let users adjust image settings like style, resolution, and prompts for a more personalized look.
- Optimization: Use techniques like mixed precision (FP16) and efficient sampling to lower costs and boost performance.
- Workflow Simplification: Make the image generation process smoother with an easy-to-use interface and drag-and-drop features.

- Accessibility: Ensure the system works for people without high-end computers or technical skills.
- Future Integration: Prepare for connections with other creative tools and platforms in the future.

### **Limitations:**

- Hardware Dependency: Even with optimization, the system might need decent computer resources, which can limit use for those with very basic devices.
- Output Quality: Sometimes, the generated images may lack fine details or not fully match complex prompts.
- Learning Curve: While the interface is user-friendly, some users may take time to get used to advanced customization features.
- Model Constraints: The system depends on pre-trained models like v1-5-pruned-emaonly-fp16.safetensors, which might not cover every artistic style or need.
- Scalability: Creating high-resolution images or managing large projects could still be challenging.
- Ethical Concerns: The system might accidentally create inappropriate or biased content, so careful monitoring and safeguards are needed.

## CHAPTER 2

# Literature Survey

### Existing Models, Techniques, and Methodologies

- Stable Diffusion Models: Stable Diffusion is a modern model that transforms text into images using advanced diffusion techniques. It's well-known for creating a variety of detailed images from text prompts. Pre-trained models like v1-5-pruned-emaonly-fp16.safetensors are commonly used for this purpose.
- User Interfaces for AI Models: Tools such as Comfy UI, Gradio, and Streamlit provide user-friendly interfaces for working with AI models. Comfy UI is especially noted for its flexibility in adapting workflows for Stable Diffusion.
- Optimization Techniques: Techniques like mixed precision (FP16), model pruning, and quantization help reduce computing costs and speed up the process. Sampling methods like DDIM (Denoising Diffusion Implicit Models) and PLMS (Pseudo Linear Multi-Step) are utilized to balance image quality with generation speed.
- Prompt Engineering: Advanced prompt engineering techniques help guide the model in creating images that align with user intent. This includes using weighted prompts, negative prompts, and keywords related to specific styles.

### Gaps and Limitations in Existing Solutions

- Technical Complexity: Setting up and using Stable Diffusion models often requires some technical skills. Non-technical users find it difficult to understand terms like sampling steps, CFG (Classifier-Free Guidance) scale, and seed values.
- Limited User Control: Despite its power, users often have trouble controlling specific aspects of the generated images, like style, composition, and fine details.
- Workflow Inefficiencies: Current workflows can be cumbersome, requiring users to manually adjust settings and iterate many times to achieve their desired results.
- Resource Intensiveness: Generating high-quality images can demand a lot of computing power, making it challenging for users without advanced hardware to get involved.
- Lack of Integration: Many existing solutions operate separately and don't connect with other creative tools or platforms, making it difficult for users to incorporate AI-generated images into their projects.

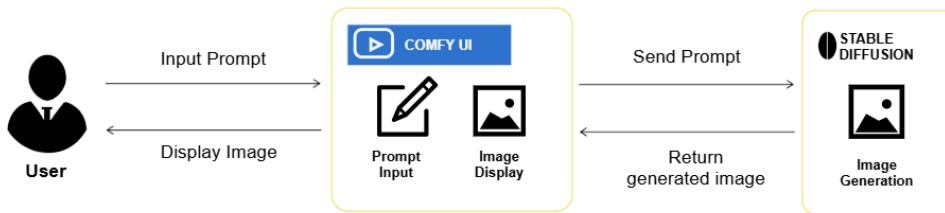
## How This Project Addresses the Gaps

- Simplified User Experience: By merging Stable Diffusion with Comfy UI, this project provides a user-friendly interface that eliminates technical barriers, allowing non-technical users to generate images easily.
- Enhanced User Control: The system allows users to adjust image parameters like style, resolution, and prompts through a straightforward interface, giving them more control over the outcomes.
- Streamlined Workflow: Comfy UI's drag-and-drop features simplify the image generation process, reducing the need for manual tweaks and iterations.
- Optimized Performance: The project employs optimization techniques like mixed precision (FP16) and effective sampling methods to decrease computing costs and enhance performance, making the system more accessible.
- Future Integration Potential: The project lays the groundwork for connecting the system with other creative tools and platforms, facilitating smoother workflows for artists and designers.
- Focus on Accessibility: By addressing resource demands and technical challenges, the project makes it easier for a broader audience, including non-technical users and creators, to access AI-powered image generation.

## CHAPTER 3

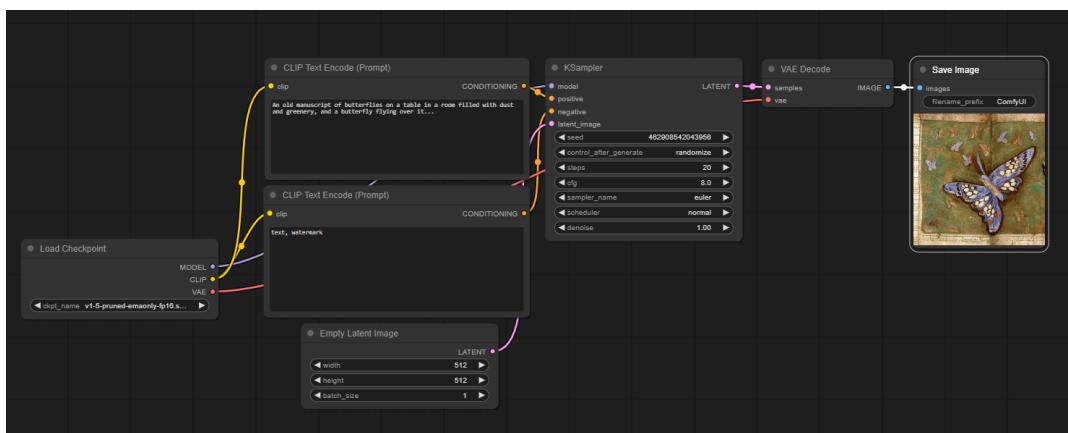
### Proposed Methodology

#### 3.1 System Design



- The user starts by giving a text prompt that describes the image they want.
- In ComfyUI, this prompt is entered into a section where users can adjust their descriptions. After refining the prompt, the user submits it, and ComfyUI sends it to Stable Diffusion for image creation.
- Stable Diffusion takes the prompt and uses AI techniques to create an image based on the text. Once the image is ready, it is sent back to ComfyUI.
- ComfyUI then displays the generated image for the user to see. The user can look at the final image or change their prompt to get a different result.

In summary, the user provides a prompt, ComfyUI handles the input and shows the image, and Stable Diffusion generates the image based on the prompt. The user can then view the finished image.



## 3.2 Requirement Specification

### 3.2.1 Hardware Requirements:

- Windows 10/11, macOS, or a Linux distribution
- Processor (CPU), Graphics Card (NVIDIA GPU), 16 GB of RAM (32 GB recommended)

### 3.2.2 Software Requirements:

- Python 3.8 or higher should be installed.

#### Required libraries:

- 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)
- Some other required libraires are Flask, Requests, TQDM, Filelock, Gradio, Omegaconf.

## CHAPTER 4

### Implementation and Result

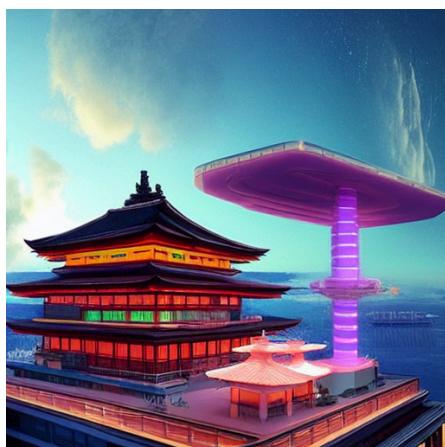
#### 4.1 Snap Shots of Result:



beautiful scenery nature glass bottle landscape, yellow galaxy bottle



A butterfly in a garden with children playing around and a dog running around



A cozy, neon-lit tea house perched on a floating platform above a futuristic city. The sky is filled with flying cars, holographic billboards, and a glowing sunset. The tea house has intricate Japanese architecture blended with modern cyberpunk aesthetics.



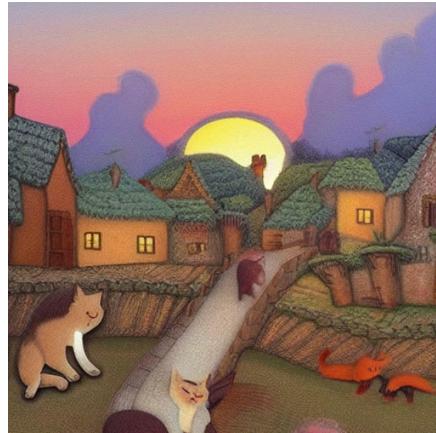
A massive, forgotten library overgrown with vines and flowers, hidden deep within a mystical jungle. Sunlight filters through the canopy, casting golden rays on ancient books and scrolls. Exotic birds and small glowing creatures flutter around.



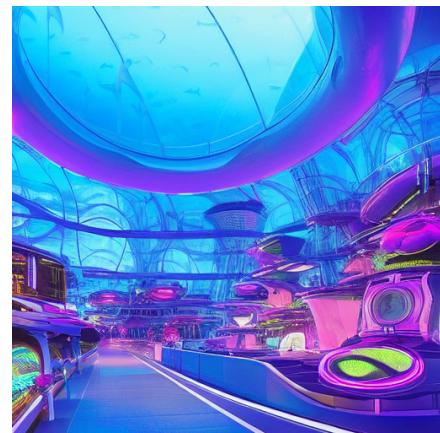
A surreal scene of an astronaut sitting on a rock, engaged in an intense chess match with a translucent, multi-limbed alien. The Martian landscape stretches endlessly behind them, with a massive gas giant visible in the sky.



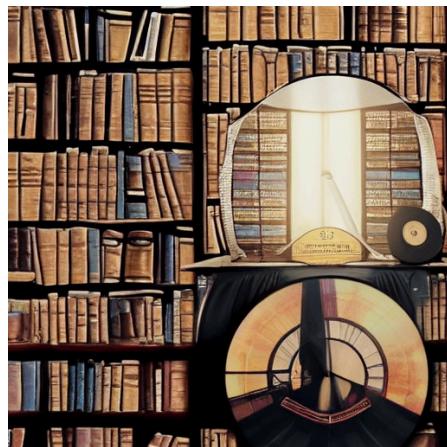
A warrior clad in intricate bronze and brass armor, wielding a mechanical crossbow, standing on a cliffside. A giant steam-powered dragon with glowing blue eyes hovers above, its wings made of metal gears and steam vents. The background features a gothic city with zeppelins flying overhead.



A fluffy, oversized cat curled around a small medieval village, acting as its gentle guardian. The villagers go about their day, using the cat's fur as pathways and its tail as a bridge over a river. The scene is warm and whimsical, with a soft, golden sunset.



A futuristic metropolis built entirely beneath the ocean, glowing in shades of neon blue and purple. Transparent domes house high-tech buildings, with schools of glowing jellyfish drifting between them. Advanced submarines and humanoid robots navigate the vibrant coral streets.



A mysterious, dimly lit study filled with floating objects from different eras—an ancient Egyptian tablet, a futuristic energy weapon, a medieval crown, and a vinyl record spinning in mid-air. The walls are lined with bookshelves, and a glowing portal flickers in the corner.

#### 4.2 GitHub Link for Code:

<https://github.com/SpoorthyBoga/AICTE-project.git>

## CHAPTER 5

### Discussion and Conclusion

#### 5.1 Future Work:

- **Higher Resolution Images:** Create models that can produce high-resolution images while keeping all the details and quality intact.
- **Integration with Creative Tools:** Make it easy to connect the system with video editing software, design apps, and other creative platforms.
- **Enhanced Comfy UI Features:** Upgrade the interface to be more user-friendly and accessible for those who aren't tech-savvy.
- **Artistic Style Customization:** Introduce advanced options to apply specific artistic styles to images, allowing for unique and personalized creations.
- **Collaborative Tools:** Build tools that help artists and designers work together with AI, inspiring new ways to express creativity.

#### 5.2 Conclusion:

This project brings together Stable Diffusion and Comfy UI to create an easy-to-use image generation system. It tackles important issues like technical difficulties, limited user control, and clunky workflows. By improving performance and making it more accessible, it allows anyone, even without technical skills, to create stunning, personalized images with ease. The emphasis on customization and smooth workflows sets the stage for exciting future developments, like higher-resolution images and connections with creative tools, making AI-powered image generation easier and more flexible for everyone.

## 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]. Denoising Diffusion Probabilistic Models (DDPM): Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems (NeurIPS). (<https://arxiv.org/abs/2006.11239>)
- [3]. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (<https://arxiv.org/abs/2112.10752>)
- [4]. DALL·E: Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. Proceedings of the 38th International Conference on Machine Learning (ICML). (<https://arxiv.org/abs/2102.12092>)
- [5]. CLIP (Contrastive Language–Image Pretraining): Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). \*Learning Transferable Visual Models From Natural Language Supervision\*. Proceedings of the 38th International Conference on Machine Learning (ICML).(<https://arxiv.org/abs/2103.00020>)
- [6]. Gradio:Abid, A., Abdalla, A., Abid, A., Khan, D., Alfozan, A., & Zou, J. (2019). \*Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild\*. Proceedings of the 36th International Conference on Machine Learning (ICML).(<https://arxiv.org/abs/1906.02569>)
- [7]. Streamlit:Streamlit. (2019). \*Streamlit: A Faster Way to Build and Share Data Apps\*. (<https://www.streamlit.io/>)
- [8]. Comfy UI:Comfy UI is a community-driven tool for Stable Diffusion workflows. While there is no formal publication, documentation and resources can be found on its official GitHub repository: (<https://github.com/comfyanonymous/ComfyUI>)
- [9]. Mixed Precision Training (FP16):Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., ... & Wu, Y. (2018). \*Mixed Precision Training\*. Proceedings of the 6th International Conference on Learning Representations (ICLR).(<https://arxiv.org/abs/1710.03740>)

- [10]. Efficient Sampling Methods (DDIM): Song, J., Meng, C., & Ermon, S. (2021). Denoising Diffusion Implicit Models. Proceedings of the 9th International Conference on Learning Representations (ICLR). (<https://arxiv.org/abs/2010.02502>)
- [11]. Prompt Design for Text-to-Image Models:Liu, V., & Chilton, L. B. (2022). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.(<https://arxiv.org/abs/2109.06977>)