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

Kartik Sopan Phopase

kartik.phopse_24uds@sanjivani.edu.in

Under the Guidance of

Mr. Jay Rathod

ACKNOWLEDGEMENT

I want to say a big thank you to everyone who helped make this project, Image Generation using Stable Diffusion & Comfy UI, a success.

I owe a lot to my supervisor, Jay Rathod. His guidance, feedback, and support were key to my internship. His know-how and mentoring had a big impact on how this project turned out.

I'm also grateful to Microsoft's Corporate Social Responsibility (CSR) team. They gave me the chance to work on this real life project as part of the internship program. Their focus on new ideas and making a difference in society created a great space for me to grow, both as a person and in my career.

I want to give a shout-out to the people who made Stable Diffusion and Comfy UI. Their open-source tools were vital to bring this project to life.

I want to say a big thank you to my collage professors. They've always been there for me cheering me on and backing me up.

I'm also thankful to all the team members and other interns for working together so well, which made this experience fun and a great learning opportunity. This project has given me useful hands-on experience, and I'm excited to use what I've learned in my future work. Thanks again to everyone who helped make this happen.

ABSTRACT

This project, Image Generation using Stable Diffusion & Comfy UI, deals with the integration of the latest AI image generation techniques with a user-friendly interface of Comfy UI. The main aim was to make a tool that would leverage the Stable Diffusion model, a most pioneering generative model, in association with Comfy UI, intended to make the image generation process more approachable for both techies and non-technical users.

Problem Statement:

The rise of image generation models powered by artificial intelligence continues to baffle many users due to rapid advancements and changes within the industry. The lack of accessible technologies serves as the primary hurdle for those with little understanding of AI.

Objective:

Create a software application that will enable users to generate images using Stable Diffusion without the need to engage with the model's code directly.

Use Comfy UI to create an embedded user interface that is easy to use and enhances the user experience. Illustrate the potential that Stable Diffusion has in creating different types of images.

Methodology:

The project focused on the integration of Stable Diffusion into Comfy UI UI with the purpose of making a user friendly image generation platform. Picture generation UI was capable of accepting text instructions, adjusting parameters, instant showing of images and other necessary features so that every user could create images that matched their expectations.It was then thoroughly tested to enhance model performance and ensure that the interface is user-friendly to a wide range of users.

Key Results:

It produced a wide array of high-quality images based on the custom text prompts that were inputted, and the users appreciated the usability and responsiveness of the user interface, which could easily transition to generating images directly from inputting prompts.

Conclusion:

This is a project, which shows tremendous potential in unifying advanced generative models, such as Stable Diffusion, with intuitive interfaces to expand accessibility to AI-driven creative tools.

TABLE OF CONTENT

Abstract 3

Chapter 1. Introduction 1

1.1 Problem Statement ....

1.2 Motivation ....

1.3 Objectives...

1.4. Scope of the Project ....

Chapter 2. Literature Survey 3

Chapter 3. Proposed Methodology 5

Chapter 4. Implementation and Results 7

Chapter 5. Discussion and Conclusion 11

References...14

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	A Bottle containing a beautiful stars and a tree	7
Figure 2	A futuristic city skyline at dusk, glowing with neon lights.	8
Figure 3	A serene forest in early autumn, with golden leaves scattered on the ground.	8
Figure 4	A vintage 1920s car parked under a starry night sky, surrounded by mountains.	9
Figure 5	A cute cat wearing a wizard hat, casting a magical spell in a fantasy forest.	9
Figure 6	A majestic dragon flying over a medieval castle, with stormy clouds in the background	10

CHAPTER 1

Introduction

1.1Problem Statement:

The generative AI model, specifically the image generation tool, has only recently gained much popularity in making so many users go around exploring the complexities of the technical tools and models. For instance, Stable Diffusion models, having tremendous potential to produce creative and high-quality visuals from text descriptions, often require a considerable deal of technical expertise, curtailing accessibility to a broader audience. This also means that the majority of these existing image-generating platforms either are too complex or do not have enough features to generate any desired outcome. It is that difference between advancement in technology and usability that raises a significant bar to entry: preventing the potential users, whether designers or content creators, to fully utilize this tool.

1.2Motivation:

This project was made to close the gap between the complexity of AI models such as Stable Diffusion and the need for a very intuitive user interface. By integrating Stable Diffusion with Comfy UI, we take the power of AI-driven image generation from complex computers to users that are not even significantly techy. The potential fields of application of this project include digital art, marketing, content creation, education, and game development. High-quality, customized images generated rapidly based on text prompts can greatly impact creativity, productivity, and innovation across all industries while democratizing access to advanced AI technologies.

1.3Objective:

1. Development of a user-friendly interface for Stable Diffusion for high-quality image generation.

2. To make the process of image generation more accessible and user-friendly by allowing users to generate images from text descriptions without needing deep technical knowledge.

1.4Scope of the Project:

This project is restricted to the use of Stable Diffusion in Comfy UI where the focus is on developing a user-friendly approach towards image generation. The scope excludes model development or training, rather it aims towards increasing the accessibility and usability of an existing model. Some of the notable functionalities include image generation from text input, customizable prompt inputs, and controls over the output (image resolution, aspect ratio, etc.). Some of the limitations are as follows:

Scope of Created Imagery: The number and varying style of images is constrained by the ability of Stable Diffusion. Even though the images generated are of high standard, they do not fulfill all creatives needs.

Interface Functionality: In line with the major goal, the project intends to maintain a high level of simplicity hence some of the more advanced features and extensive cutomizations found on other platforms will not be included.

Available Resources: The scope is also limited by the available computational power for Stable Diffusion, as it is expected that image generation will consume a significant amount of resources, irrespective of whether it is done locally or on the cloud

CHAPTER 2
Literature Survey

Over the past several years, there has been tremendous progress made into generative models, particularly within the scope of image synthesis. The most notable models belonging to this category include GANs, VAEs, and Diffusion Models. Among these, Stable Diffusion is one of the most widely used and powerful generative models due to its ability to produce high quality images from textual input. Below is an overview of the relevant models and methodologies in this space:

1. Generative Adversarial Networks

GANs have been one of the most impactful techniques in generative modeling introduced by Ian Goodfellow back in 2014. They comprise two networks: a generator and a discriminator that model a two player game. While GANs have been successful in image generation tasks like realistic generation of humans, objects, and landscapes, they are difficult to train and suffer from several problems like mode collapse when the generator does not produce varied outputs. Other famous models include StyleGAN and BigGAN, which have excelled in face and object generation. However style, Big, and other types of GANs have had major challenges, such as the lack of efficient computation and extensive work required to enhance low resolution images, while moderately low resolution images are easy and fast to produce.

2. Variational Autoencoders (VAEs)

VAEs, proposed by Kingma and Welling in 2013, are another alternative to generative modeling, learning a representation of the data in the form of a latent variable. Although VAEs are good at capturing the data distribution, they are less detailed and blurrier compared to GANs and diffusion models. Despite that, their simplicity and effectiveness in generating images from a continuously distributed latent space makes them relevant in certain fields like anomaly detection and image interpolation. ”

3. Diffusion Models (For example, Stable Diffusion)

Stable Diffusion, as proposed by Rombach et al. in 2022, is an example of a diffusion models that uses random noise as a seed and turns it into a coherent image. Like Rombach, most researchers are using diffusion models in their recent papers with the expectation that these models outperform former methods and do not require additional work.”

Unlike GANs, mode collapse is generally not a challenge for diffusion models, and these tend to yield more stable, high-quality output. These models have received much attention for the variety and high resolution of images produced from text-based descriptions.

However, despite the impressive capabilities of Stable Diffusion, the typical user will often find it difficult to use due to matters of complexity in deployment and usage as it would require a good understanding of machine learning models and their interaction techniques. Computational Resources: Diffusion models for generating images are very computationally expensive and require powerful GPUs or cloud-based platforms. This may become a barrier to non-technical users or to those with less access to these resources. Customization: Stable Diffusion does allow for customizing image generation but mostly through manual intervention (e.g., adjusting parameters in code or settings), which is not very intuitive for non-technical users.

4. Comfy UI

Comfy UI is a pretty new innovation which tries to give an interface easy to work with generative models like Stable Diffusion. Offering a visual interface with drag-and-drop functionalities makes it easy to input text prompts, change settings, and generate images.

CHAPTER 3

Proposed Methodology

System Design

Stable Diffusion with Comfy UI to create a more user-friendly experience for generating images from text prompts. The system is made up of several modules that work together to process user inputs, generate images, and display the results in an easy-to-use format.

User Interface (Comfy UI):
This is the front-end interface where users engage with the system. It offers a straightforward way for users to enter text prompts (descriptions of the images they wish to create) and a space to showcase the generated image.
The UI enables users to adjust basic settings such as image resolution, style, and aspect ratio, giving them control over the output without needing to understand the technical details.
Request Handler / API Layer:

After the user submits the text prompt and settings, the Request Handler (API layer) processes the input and forwards the request to the backend (Stable Diffusion model).
This layer handles the communication between the UI and the backend, making sure that the data is properly exchanged and the user gets the feedback (for example, an image) once the model generates it.
Stable Diffusion Model (Backend):

The actual core of the system is the stable diffusion model. Given a text input from the user, it will produce good-quality images. In a nutshell, the idea to use such a process is that it is the diffusion process whereby noise transforms itself into a highly detailed image and guided by a textual prompt by the user.
The model used in the back end is pre-trained and capable of inferring images based on given prompts.
Image Generation and Post-Processing:

After generating the image, it may undergo post-processing to ensure it meets acceptable quality standards and aligns with the user's expectations. This could involve increasing the resolution or enhancing the colors. The final post-processed image is then prepared for presentation, with the aim of meeting the desired output quality at this stage.

Image Display Area:

The generated image is shown in the image rendering area of the Comfy UI. Users are able to view, save, or even edit the image here. The UI is responsive so that users can interact with it seamlessly.

3.1 Requirement Specification

GPU: A high performance GPU is important to run the Stable Diffusion efficiently. NVIDIA RTX 3060 and RTX 3080, at least.

CPU: For handling the back-end logic and API requests besides non-GPU tasks, one needs an Intel Core i7 or similar CPU.

RAM: 16 GB of RAM would be enough to generate images smoothen and handle large weights of the models.

Storage: Provide an SSD storage with available space of at least 50 GB to store a pre-trained model, images, and temporary files.

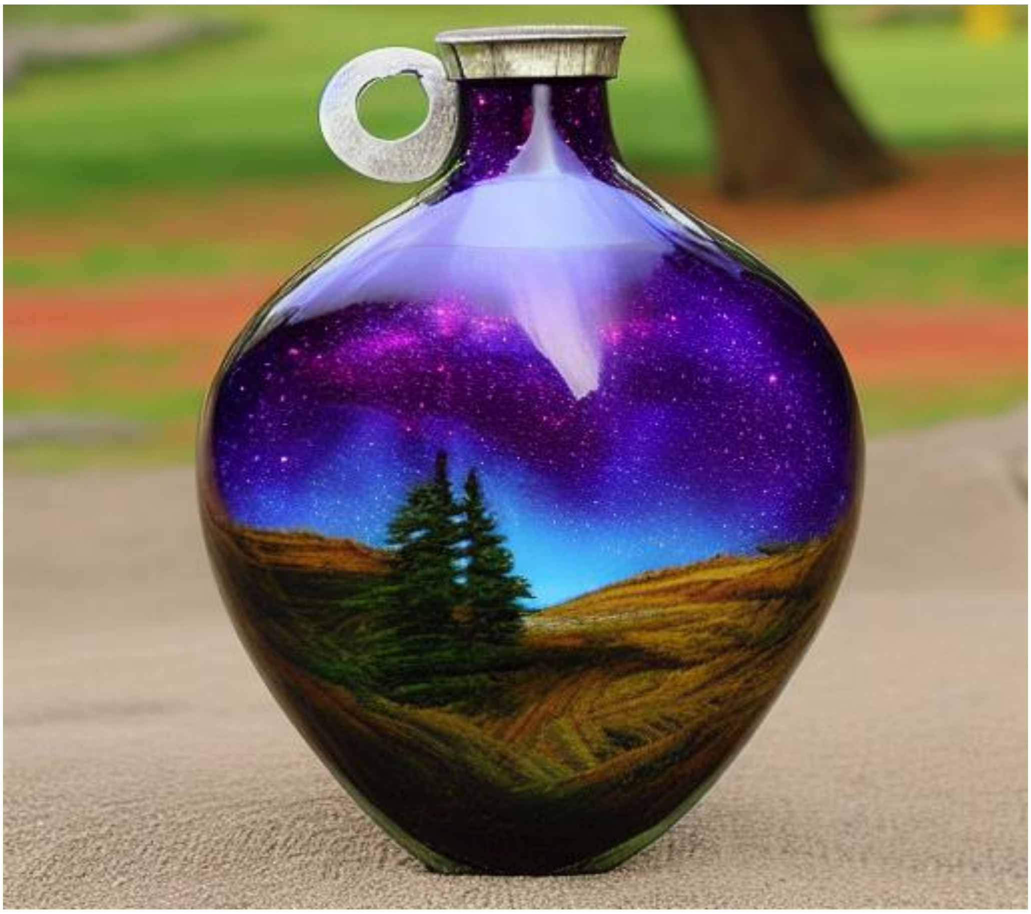
Cloud/Remote Compute Resources For large-scale jobs or those who have limited access to local GPUs

GPU Instances In the cloud, offers services such as AWS, Google Cloud, and Microsoft Azure.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:



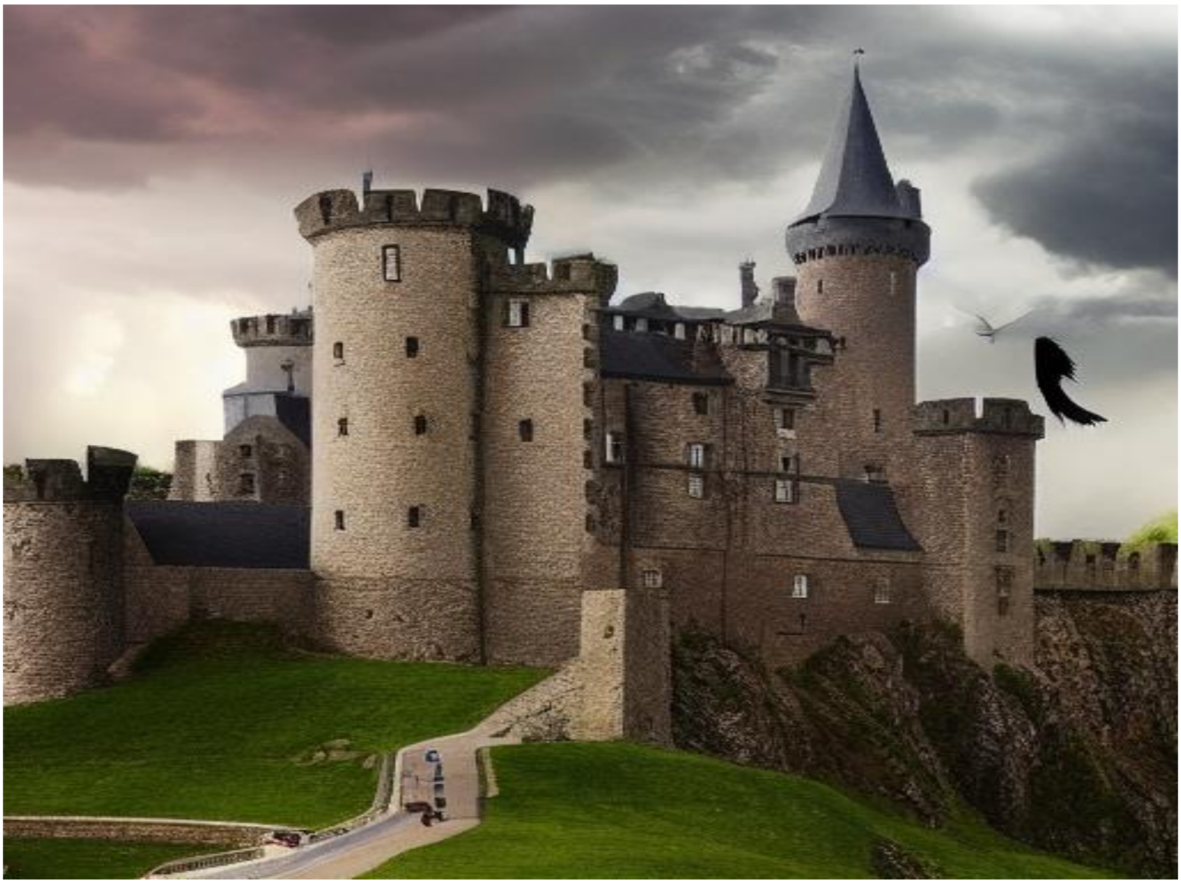
A Bottle containing a beautiful stars and a tree

A futuristic city skyline at dusk, glowing with neon lights.

A serene forest in early autumn, with golden leaves scattered on the ground.

A vintage 1920s car parked under a starry night sky, surrounded by mountains.

A cute cat wearing a wizard hat, casting a magical spell in a fantasy forest.



A majestic dragon flying over a medieval castle, with stormy clouds in the background

4.2 GitHub Link for Code:

https://github.com/kartikphopase/imagegeneration

CHAPTER 5

Discussion and Conclusion

Future Work

Despite the current success of the implementation of the Image Generation using Stable Diffusion & Comfy UI project, allowing users to easily generate high-quality images with a simple interface, several aspects will need to be further developed and expanded:

Future Improvement of the UI and UX

The present UI can also be further enhanced with more refined options for user personalization such as:

Art Style: Allow selection of preferred style or referenced image for generation of the produced image.

Online Previews: Provide online real-time preview with dynamic adaptation to images where changing resolution or color or effects and styles apply.

Interactive Toolkit: The users interact in the tool with drag-n-drop to layup composition, which can be in editing content on generated output. Fine tuning of a model: Training the system specifically to perform some particular job within a different field

The Stable Diffusion can also be fine-tuned for other domains or applications such as fashion, architecture, medical illustration, etc. for improvement of the generated images quality and accuracy particularly to niche-industry applications.

Training of a custom model: Future Work

Specialized models trained over specific datasets are developed to give better-tailored outputs for example, highly detailed character design or product mockups.

Superior Image Resolution and Quality

Although Stable Diffusion generates high-quality images, the resolution of the images sometimes does not meet the desired level. Future versions can include:

Super-Resolution Methods: Tools for increasing the image resolution after generating it without reducing the details.

Noise reduction and image enhancement: Features that should automatically remove the noise and give more realism to the generated image.

Performance Improvements:

The generation using Stable Diffusion might be computationally intensive and require large GPU resources. Optimization of the model for faster times to generate an image or using lightweight versions of the models like distillation techniques could be used to improve efficiency.

Server-Side Optimizations: The system will enhance the performance, wait time, and may even do the batch processing/parallelization where multiple images would be produced together.

Other Integrations

Future work will include integrating this solution with other creative tools or platforms, for example, Adobe Photoshop or Blender, to support seamless workflows by allowing users to import and directly manipulate AI-generated images within their existing design environments. Adding Ethical Guidelines and Safety Filters

Ethics in relation to bias, misuse, and inappropriate content generation would, therefore become very critical in generative models such as Stable Diffusion.

Safety Filters and Moderation: The filters deployed will have functionalities for identifying and blocking dangerous, deceptive, or biased content in order to avoid the model's generation of inappropriate or unethical images.

Conclusion

The project of Image Generation using Stable Diffusion & Comfy UI has indeed filled the gap between complex generative models and everyday users. With the power of integrating State-of-the-art generative AI, Stable Diffusion, in the interface with Comfy UI, this solution empowers both designers, content creators, as well as teachers to easily produce high-quality images from textual descriptions.

This project has a couple of impacts:

Increased Accessibility: It makes AI-driven image generation accessible to non-technical users, allowing anyone to harness its power, making it accessible and democratizing advanced creative tools.

Empowering Creativity: This provides the means for image generation features that are customizable, unlocking new avenues of creative expression, marketing, educational content, and much more.

Scalability: The modular approach of Stable Diffusion with Comfy UI can be scaled up and enhanced easily, making sure that the system can be evolved as more advancements in AI technology come.

The project is successful in directly addressing the challenge of making sophisticated AI models more accessible and user-friendly for non-technical audiences. The future work mentioned here will make enhancements to the system's capabilities to cover a greater range of users and use cases while improving its performance, customization options, and ethical considerations.

In summary, this project shows the possibility of generative AI in creative applications and paves the way for broader adoption across industries to open up new opportunities for innovation and creativity.

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.