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

Thiruveedhi Shanmukha Sai,

shanmukhasai432@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to **Mr. Jay Rathod**, my guide and mentor, for his invaluable guidance, encouragement, and support throughout the development of this project, "**Image generation using stable diffusion & Comfy UI**". His expertise and insights were instrumental in shaping the direction of the project and overcoming challenges.

I am also grateful to **TechSaksham** initiative by **Microsoft & SAP** for providing me with this internship opportunity, a conducive environment, and access to resources that enabled. me to successfully complete this project.

Lastly, I extend my thanks to my colleagues and team members for their collaboration and support, and to my family and friends for their constant encouragement during the internship period.

Thiruveedhi Shanmukha Sai

ABSTRACT

The **Image generation using stable diffusion & comfy UI** is designed to generate A User-Friendly Image Generation Platform using Denoising diffusion model. Recent advancements in deep learning have led to the development of powerful image generation models like Stable Diffusion. However, these models often require technical expertise and cumbersome interfaces, limiting their accessibility to non-experts.

This project presents a novel approach to image generation using Stable Diffusion, a state-of-the-art denoising diffusion model, combined with Comfy UI, an intuitive user interface library. Our system enables users to generate high-quality images from text prompts, leveraging the power of deep learning and diffusion-based synthesis. Comfy UI provides an accessible and interactive way to adjust parameters, explore different styles, and refine the generated images. We demonstrate the effectiveness of our approach through a range of experiments, showcasing the ability to generate diverse, realistic, and contextually relevant images. Our project has the potential to democratize image generation, enabling artists, designers, and hobbyists to unlock new creative possibilities.

The platform will be built using a combination of Python, Stable Diffusion, and Comfy UI. Users will be able to input text prompts, select from various style and theme options, and adjust parameters to control the image generation process. The platform will utilize GPU acceleration to ensure efficient and rapid image generation.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

The rapid advancement of deep learning techniques has led to significant breakthroughs in image generation. However, existing image generation models often require extensive technical expertise, cumbersome interfaces, and substantial computational resources, limiting their accessibility and usability.

Despite the potential of image generation technology, the current state-of-the-art models are not user-friendly, and their complexity creates a barrier to entry for non-technical users. This limitation hinders the widespread adoption of image generation technology, restricting its benefits to a niche audience.

Create a user-friendly and intuitive image generation platform that leverages the power of Stable Diffusion, a state-of-the-art denoising diffusion model, and Comfy UI, a flexible and interactive user interface library. The platform should enable users to generate high-quality images from text prompts, explore different styles, and refine the generated images without requiring extensive technical expertise.

Key Challenges:

- **Complexity:** Existing image generation models require extensive technical expertise, making them inaccessible to non-technical users.
- **User Experience:** Current interfaces for image generation models are often cumbersome and difficult to navigate, leading to a poor user experience.
- **Computational Resources**: Image generation models require substantial computational resources, making them difficult to deploy and use, especially for users with limited resources.

1.2Motivation:

The motivation behind this project is to democratize image generation technology, making it accessible and usable for a broader audience, including artists, designers, hobbyists, and non-technical users.

Potential Applications: -

- **Art and Design:** Enable artists and designers to generate new ideas, explore different styles, and automate repetitive tasks.
- **Advertising and Marketing**: Generate high-quality images for advertisements, product showcases, and social media campaigns.

- **Education and Training:** Create interactive and engaging educational materials, such as virtual labs, simulations, and interactive diagrams.
- **Entertainment and Media:** Generate images for movies, video games, and virtual reality experiences.
- **Healthcare and Medical Imaging:** Generate synthetic medical images for training and research purposes.
- **Fashion and Apparel**: Generate images of clothing and accessories for e-commerce and fashion design.
- **Architecture and Real Estate:** Generate images of buildings and interior designs for architectural visualization and real estate marketing.

Impacts of the project: -

- **Increased Creativity**: Enable users to generate new and innovative ideas, leading to increased creativity and productivity.
- **Improved Efficiency**: Automate repetitive tasks and reduce the time required to generate high-quality images.
- **Enhanced User Experience**: Provide users with an intuitive and user-friendly interface, making it easy to generate high-quality images.
- **Democratization of Image Generation**: Make image generation technology accessible to a broader audience, regardless of their technical expertise.
- **New Business Opportunities**: Enable new business opportunities in industries such as advertising, education, and entertainment.
- **Advancements in Research and Development**: Generate synthetic images for research and development purposes, leading to advancements in fields such as healthcare and materials science.
- **Increased Accessibility**: Make image generation technology accessible to people with disabilities, enabling them to participate in creative activities.

1.3Objective:

The objective of the project is: -

- **Develop a User-Friendly Interface**: Design and develop an intuitive and user-friendly interface using Comfy UI that enables users to easily generate high-quality images using Stable Diffusion.
- **Enable High-Quality Image Generation**: Utilize Stable Diffusion to generate high-quality images that are comparable to state-of-the-art image generation models.
- **Simplify the Image Generation Process**: Streamline the image generation process, making it easier for users to generate high-quality images without requiring extensive technical expertise.

- 4.**User Experience**: Enhance the overall user experience by providing a responsive, interactive, and engaging interface.
- **Foster Creativity and Innovation**: Encourage users to explore new ideas, styles, and themes, and provide a platform for creative expression and innovation.
- **Optimize Performance**: Optimize the performance of the image generation platform, ensuring efficient use of computational resources and fast image generation times.
- **Establish a community**: Foster a community of users, developers, and researchers who can contribute to the platform, share knowledge, and collaborate on new projects.
- **Continuously Improve the Platform**: Regularly update and improve the platform, incorporating new features, models, and technologies to stay at the forefront of image generation research.
- **Explore New Applications**: Investigate new applications and use cases for the image generation platform, such as education, healthcare, and entertainment.

1.4Scope of the Project:

The scope of this project includes:

- **Image Generation**: Developing a platform that utilizes Stable Diffusion to generate high-quality images from text prompts.
- **User Interface**: Designing and developing an intuitive and user-friendly interface using Comfy UI that enables users to easily interact with the image generation platform.
- **Customization Options**: Providing users with customization options, such as adjusting parameters, selecting styles, and refining generated images.
- **Performance Optimization**: Optimizing the performance of the platform to ensure efficient use of computational resources and fast image generation times.

The limitations of this project include:

- **Technical Expertise**: While the platform aims to be user-friendly, some technical expertise may still be required to fully utilize the customization options and optimize performance.
- **Computational Resources**: The platform requires significant computational resources to generate high-quality images, which may limit its accessibility to users with lower-end hardware.
- **Data Quality**: The quality of the images generated is dependent on the quality of the training data, which may be limited by the availability and diversity of the data.

- **Style and Content**: The platform may not be able to generate images that are outside the scope of the training data, which may limit its ability to generate certain styles or content.
- **Ethical Considerations**: The platform may be used to generate images that are misleading, offensive, or unethical, which highlights the need for responsible use and development of technology.

CHAPTER 2

Literature Survey

2.1 Review relevant literature

The domain of image generation using Stable Diffusion and Comfy UI has seen significant developments in recent times. Researchers have explored the potential of Stable Diffusion in generating high-quality images from text prompts.

One notable study presented a system model that utilizes Stable Diffusion for text-to-image synthesis, leveraging the power of generative adversarial networks (GANs) and natural language processing (NLP). This approach aims to bridge the semantic gap between textual descriptions and visual content.

Another relevant work discusses the implementation of Artificial Intelligence-based image creation technology for conceptual ideas in 3D visual modeling. This research highlights the potential of AI-generated images in various applications, including education, healthcare, and entertainment.

In terms of user interface, Comfy UI has been explored as a tool for controlled image generation with Stable Diffusion. This interface enables users to refine their generated images and provides a user-friendly experience.

Key concepts in this domain include:

- **Latent Space:** A condensed representation of an image that highlights its key elements.
- **Text Encoder and Tokenizer:** Used to encode user-specific text prompts for image generation.
- **Denoising Process:** A critical step in Stable Diffusion that refines the generated image through iterative steps.

2.2 Existing Models, Techniques, or Methodologies:

Several Models, Techniques and Methodologies have been used for Image generation using stable diffusion & comfy UI

Existing Models used for Image generation using stable diffusion & comfy UI:

- **Stable Diffusion Model:** A type of generative model that uses a process called diffusion-based image synthesis to generate images.

- 2. **DALL-E Model:** A text-to-image synthesis model that uses a combination of natural language processing and computer vision techniques.
- 3. **Latent Diffusion Model:** A type of generative model that represents images as a sequence of latent variables.

Techniques used for Image generation using stable diffusion & comfy UI:

- **Diffusion-Based Image Synthesis**: A technique used in Stable Diffusion to generate images by iteratively refining a noise signal.
- **Text Encoder and Tokenizer:** Techniques used to encode user-specific text prompts for image generation.
- **Denoising Process:** A critical step in Stable Diffusion that refines the generated image through iterative steps.

Methodologies used for Image generation using stable Diffusion & comfy UI:

- **Generative Adversarial Networks (GANs):** A methodology used to train generative models, including Stable Diffusion.
- **Natural Language Processing (NLP):** A methodology used to analyze and understand user-specific text prompts.

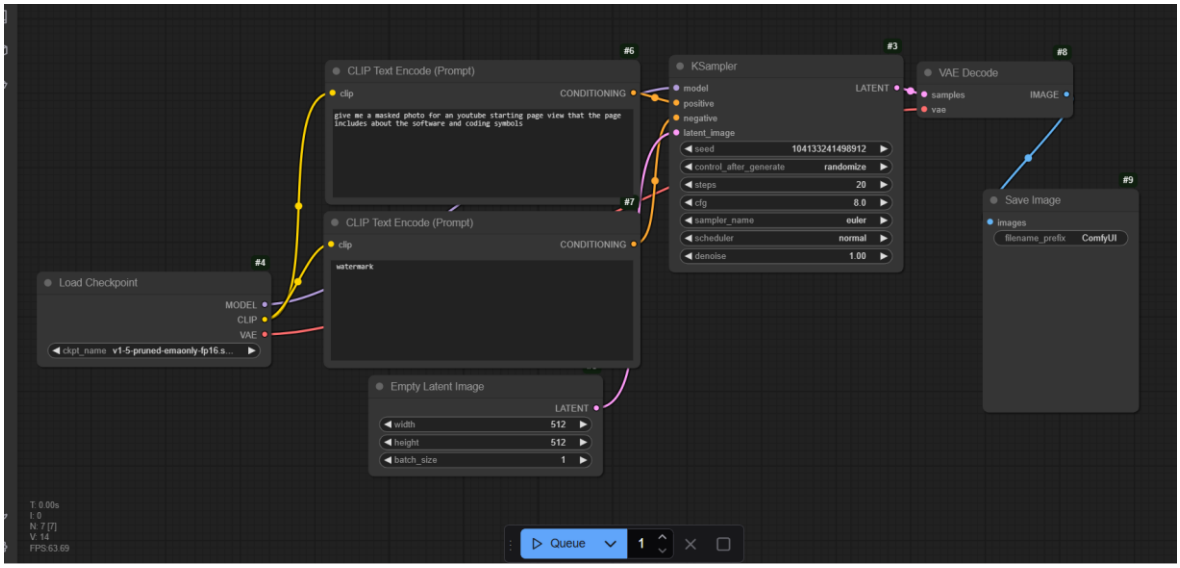
2.3 Gaps and limitations in Existing Solutions

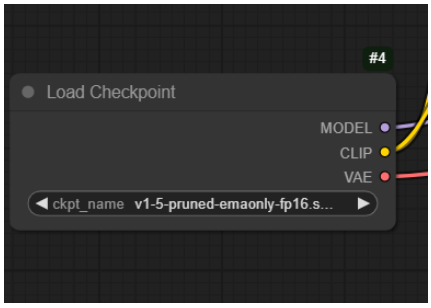
- **Limited User Control**: Existing solutions often lack intuitive user interfaces, making it difficult for non-technical users to control the image generation process.
- **Quality and Consistency**: Existing solutions may struggle to generate high-quality images consistently, particularly for complex or abstract prompts.
- **Lack of Customization**: Existing solutions may not provide sufficient customization options, limiting users' ability to refine generated images.
- **Computational Resources:** Existing solutions may require significant computational resources, making them inaccessible to users with limited hardware.
- **Ethical Concerns:** Existing solutions may raise ethical concerns, such as the potential for generating misleading or offensive images.
- **Mode Collapse**: Existing models, such as GANs, may suffer from mode collapse, resulting in limited diversity in generated images.
- **Unrealistic Images**: Existing models may generate unrealistic or unnatural images, particularly for complex or abstract prompts.
- **Lack of Control**: Existing models may not provide sufficient control over the image generation process, limiting users' ability to refine generated images.

CHAPTER 3

Proposed Methodology

3.1 System Design





Load Check point:

Most used Stability.ai Models

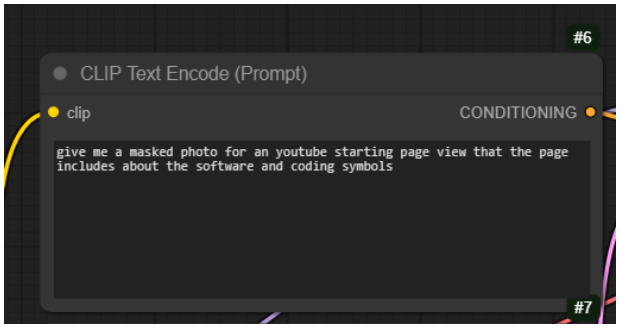
- SD 1.5
- SDXL

Fine-tuned Models:

- Specialized: style, subject

This Load Checkpoint includes:

Character, Style, Celebrity, Concept, Clothing, Base Model, Poses, Background, Tool, Buildings, Vehicle, Object, Animal, Assets, Action

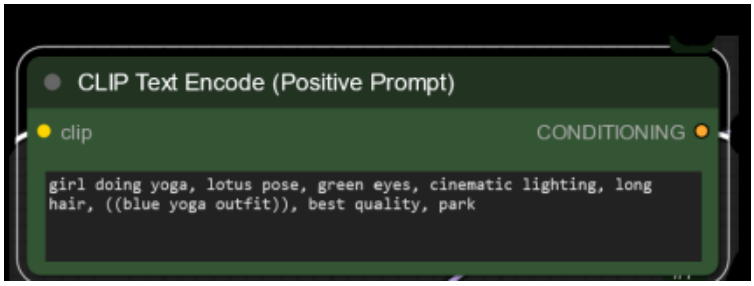


Clip Text Encode (Prompt):

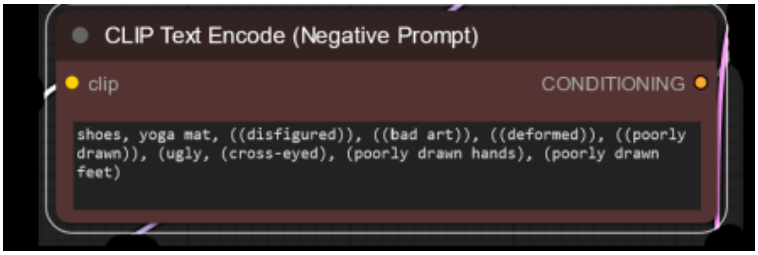
- This Clip Text Encode in simple words known as Prompts
- The prompts are two types:

1. Positive prompt
2. Negative prompt

Positive prompt:



Negative Prompt:



Empty Latent Image:

Latent Space:

1.Abstract, compressed representation of the image

2.Handles encoded features such as shapes, colors, textures and general structure

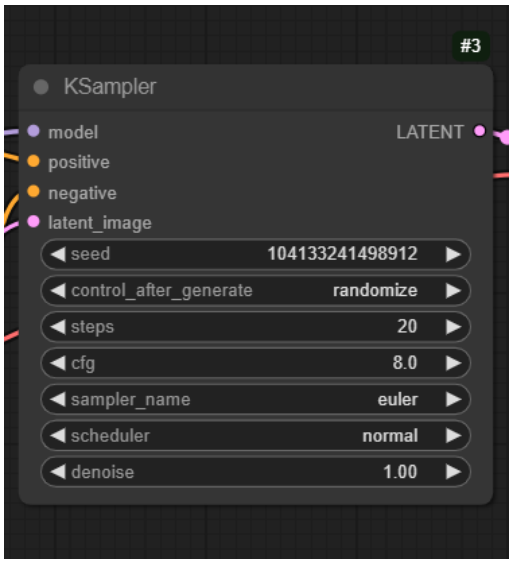
3. Manipulation of embedding vectors

Iterative and refining generation:

1.Random noise is introduced into the latent space

2.At each step the model adjusts the features to match the prompt

KSampler:



Seed:

1.Random seed used to create initial noise

2.Fixing allows you to see impact of other parameters

Samplers:

1.Algorithms guiding the iterative image generation

2.Differ in Speed and Quality

Schedulers:

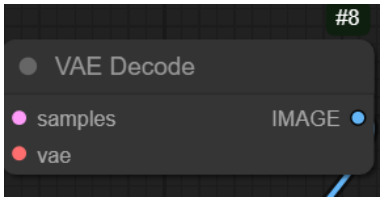
1.Control how noise is removed at each step

2.Also impact Speed and Quality, Karras is well balanced

Other Parameters:

#steps, CFG (adherence to prompt), %denoising

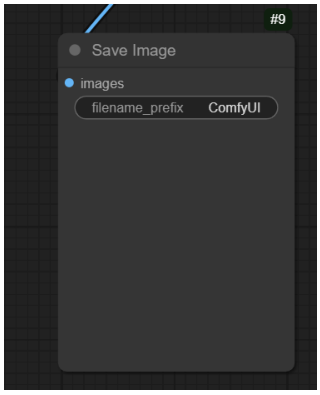
VAE Decode:



VAE (Variational Autoencoder)

• Convert Image Pixels → Latent Space

Save Image:



- Here in save image we can download the generated image

3.2 Requirement Specification

.

3.2.1 Hardware Requirements:

1. CPU: Intel Core i7 or AMD Ryzen 7 (or higher)
2. GPU: NVIDIA GeForce RTX 3080 or AMD Radeon RX 6800 XT (or higher)
3. RAM: 16 GB or more
4. Storage: 512 GB or more of SSD storage
5. Display: 1080p or higher resolution display

3.2.2 Software Requirements:

1.Comfy UI Version: Comfy UI 1.0 or later

2.Stable Diffusion Model: Stable Diffusion model (available on GitHub or Hugging Face Model Hub)

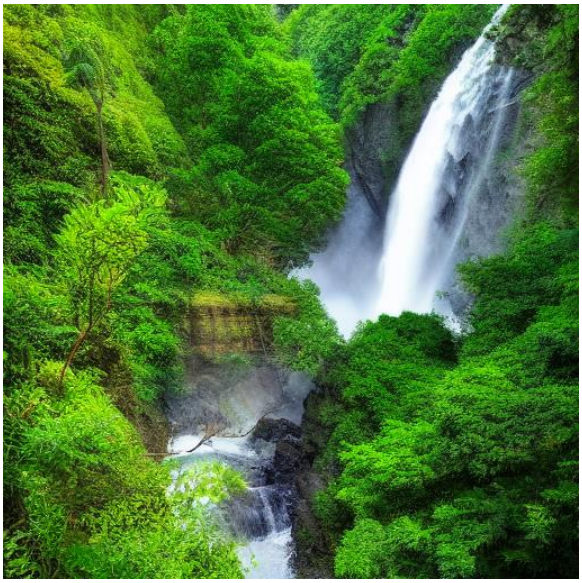
CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Prompt: -An astronaut riding a horse.

Prompt: -A dramatic waterfall cascading into a lush valley



Prompt: - Batman eating pizza for dinner

Prompt: - A massive spaceship exploring the depths of space

4.2 GitHub Link for Code:

Link: https://github.com/shanmukhasai9494/Image-generation-using-stable-diffusion-and-comfy-ui

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Incorporating other advanced techniques in image processing and analysis to improve accuracy and reliability of the generated images. Expanding the scope of the project to include other types of criminal investigations, such as generating images of missing persons or suspects in crimes other than facial recognition. Improving the speed and efficiency of the system to generate images in real-time or near real-time to support investigations that require urgent action. Integrating the system with other tools and technologies used in law enforcement, such as facial recognition software or databases of criminal records, to provide more comprehensive and accurate result.

5.2 Conclusion:

The proposed system is designed to generate facial images of suspects based on witness descriptions using modern cloud technologies and modern software engineering principles. The system consists of a user-friendly interface, a Stable Diffusion AI Model, and a secure database to store witness descriptions and generated facial images. RESTful APIs are used to communicate between components, allowing for easy integration with other systems and Applications. The Stable Diffusion AI Model used in the proposed system has been shown to be accurate in generating high-quality and diverse facial images of suspects based on witness descriptions. The model's accuracy report demonstrates its ability to generate facial images that closely resemble the actual suspects.

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.