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

REDDY RANJITH KUMAR, 22981a04j7@raghuenggcollege.in

Under the Guidance of

Mr. Jay Rathod

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

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

ABSTRACT

This project focuses on AI-based image generation using Stable Diffusion and ComfyUI. The project explores how deep learning models can create realistic images from textual descriptions. The methodology involves using the Stable Diffusion model with ComfyUI for enhanced customization. Results demonstrate the ability of AI to generate high-quality visuals. Future work includes improving model efficiency and expanding dataset diversity.

.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

AI image generation is revolutionizing various industries, including gaming, advertising, and design. This project explores how Stable Diffusion and ComfyUI can be used to generate high-quality images efficiently

1.2Motivation:

The motivation behind this project is to explore how AI can generate realistic images and improve creative workflows. AI-based image generation has wide applications in content creation, art, and digital media.

1.3Objective:

1. Implement Stable Diffusion for AI-generated images.
2. Integrate ComfyUI for user-friendly customization.
3. Evaluate model performance and output quality.

1.4Scope of the Project:

This project focuses on implementing a deep learning-based image generation system with Stable Diffusion and ComfyUI. The project scope includes experimentation with text-to-image generation and optimizing AI model performance.

CHAPTER 2

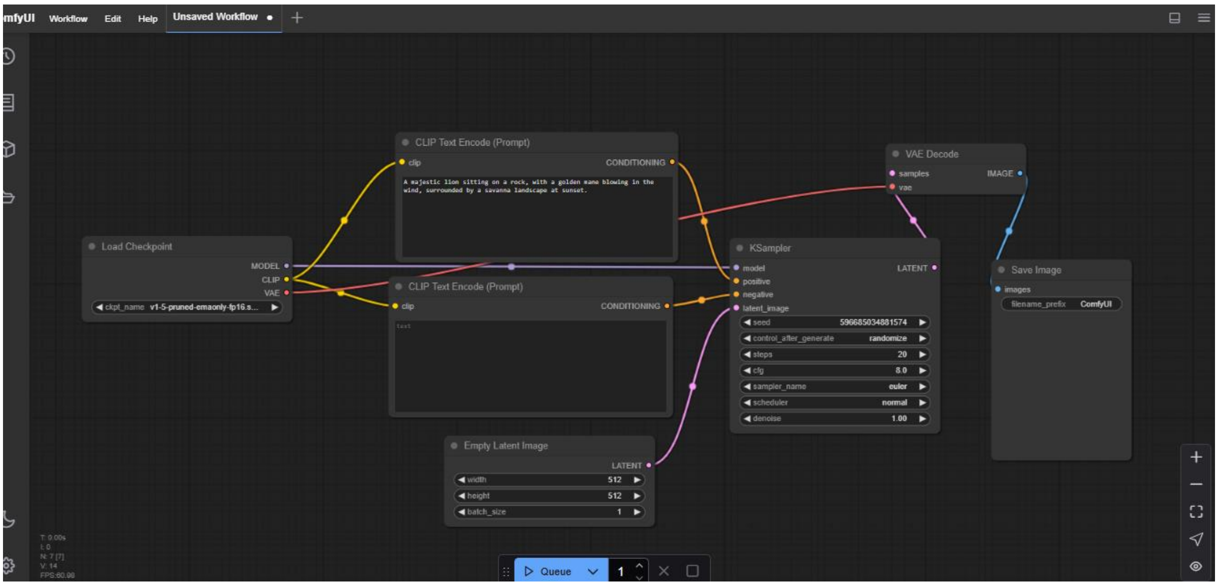
Literature Survey

Several deep learning models have been developed for AI image generation, including GANs, VAEs, and diffusion models. Stable Diffusion, introduced in 2022, has shown superior results in generating high-quality images. ComfyUI enhances user interaction with AI models, allowing greater customization.

CHAPTER 3

Proposed Methodology

3.1 System Design



1. Load Checkpoint (Model)

- Loads a pre-trained **Stable Diffusion model** (v1-5-pruned-emaonly).
- Provides **CLIP** (for text encoding) and **VAE** (for decoding images).

2. CLIP Text Encode (Prompt)

- Encodes the **positive prompt**:
  "A majestic lion sitting on a rock, with a golden mane blowing in the wind, surrounded by a savanna landscape at sunset."
- This guides the AI model to generate an image based on the description.

3. CLIP Text Encode (Negative Prompt)

- Used to encode a **negative prompt** (but it is empty in this case).
- Negative prompts help refine the output by telling the model what **not** to include.

4. Empty Latent Image

- Defines the **image resolution** (512x512).
- Initializes a blank latent image for the model to work on.

5. KSampler

- The core component that generates the image from the text prompt.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

CPU

16GB RAM

3.2.2 Software Requirements:

4 - Python 3.9+
- PyTorch
- Stable Diffusion
- ComfyUI

CHAPTER 4

Implementation and Result

4.1 Snapshots of result

Figure-1:

A bustling cyberpunk city at night, filled with neon lights, flying cars, and towering skyscrapers, with a futuristic market scene.

4.2

Figure2:

A majestic lion sitting on a rock, with a golden mane blowing in the wind, surrounded by a savanna landscape at sunset

GitHub Link for Code:

https://github.com/508522-tech/AICTE.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Future work includes improving model efficiency, enhancing customization options, and expanding datasets for better training.

5.2 Conclusion:

This project successfully implemented AI-based image generation using Stable Diffusion and ComfyUI. The results show that AI models can generate high-quality images based on text prompts, opening new possibilities in creative design and automation.

.

REFERENCES

1. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.
2. Ramesh, Aditya, et al. 'Hierarchical Text-Conditional Image Generation with CLIP Latents.' OpenAI, 2022.
3. Kingma, Diederik P., and Max Welling. 'Auto-Encoding Variational Bayes.' arXiv preprint arXiv:1312.6114 (2013).