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

Rashmi Ranjan Malik, rashmiranjan00702@gmail.com

Under the Guidance of

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 our supervisor, Jay Rathod, for being a great mentor and the best adviser we could ever have. His advice, encouragement, and constructive criticism have been a source of innovative ideas and inspiration, leading to the successful completion of this project.

ABSTRACT

This project explores the capabilities of Stable Diffusion and Comfy UI for image generation, leveraging advanced deep learning techniques. The objective is to develop an efficient image generation pipeline capable of producing high-quality visuals from textual prompts. The study involves an in-depth analysis of diffusion models, implementation of Comfy UI, and evaluation of generated images based on various parameters. The results demonstrate the potential of these models in creative industries, content creation, and AI-driven design workflows.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

With the rise of AI-generated content, there is a growing demand for high-quality image generation that balances realism and creativity. Traditional methods struggle with fine-tuned control over image attributes. Stable Diffusion, a deep learning model, presents an effective solution by generating images based on textual input while ensuring flexibility and precision.

1.2Motivation:

The motivation for this project arises from the increasing use of AI-generated imagery in fields like digital art, game design, and marketing. Stable Diffusion with Comfy UI offers an accessible and efficient framework for non-technical users to explore AI-based image synthesis.

1.3Objective:

- Implement Stable Diffusion using Comfy UI.
- Enhance control over image generation parameters.
- Analyze the quality of generated images.
- Explore real-world applications of AI-generated content.

1.4Scope of the Project:

This project focuses on leveraging Stable Diffusion and Comfy UI to develop a user-friendly image generation pipeline. The scope includes:

- Model selection and training.
- Integration of Comfy UI for ease of use.
- Performance evaluation and fine-tuning.
- Future scope for improvement.

CHAPTER 2

Literature Survey

A literature review was conducted based on multiple sources, including Introductory Notes on Stable Diffusion and Stable Video Diffusion, various research papers, and implementations of diffusion models.

2.1 Review of Relevant Literature

Diffusion models have evolved significantly, with early work focusing on denoising autoencoders and variational inference. Stable Diffusion is a breakthrough in this domain, leveraging latent diffusion models (LDMs) to generate high-resolution images efficiently. Studies on diffusion models show their superior generative ability compared to GANs, making them more robust for tasks like image synthesis, inpainting, and super-resolution.

2.2 Existing Models, Techniques, and Methodologies

Several models and techniques have contributed to the development of AI-driven image generation:

- Denoising Diffusion Probabilistic Models (DDPMs) - Basis of modern diffusion models, iteratively refining images from Gaussian noise.
- Latent Diffusion Models (LDMs) - Used in Stable Diffusion to generate images using compressed latent spaces, reducing computational costs.
- GANs (Generative Adversarial Networks) - Previously dominant in image generation, but struggled with mode collapse and training stability.

- StyleGAN & BigGAN - Advanced GAN models capable of high-fidelity image generation but with high training costs and lower flexibility compared to diffusion models.

2.3 Gaps and Limitations in Existing Solutions

While GANs and other generative models have shown promise, they exhibit issues like:

- Mode Collapse - GANs often generate limited variations of images.
- High Training Costs - Requires extensive GPU power and dataset curation.
- Lack of Interpretability - Difficult to control specific features in generated images.

Stable Diffusion and Comfy UI address these challenges by providing better control, reduced computational needs, and flexibility in image generation. Our project leverages these advantages to build an optimized image-generation pipeline accessible to non-experts.

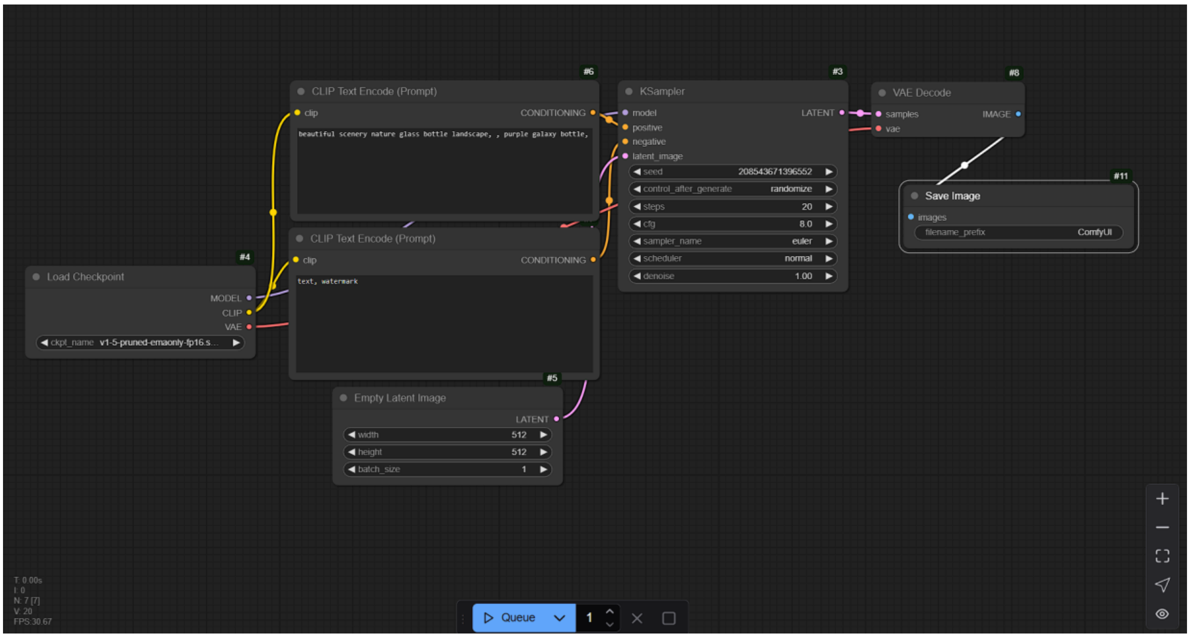
CHAPTER 3

Proposed Methodology

3.1 System Design

The proposed system involves:

1. Text-to-Image Generation using Stable Diffusion.
2. Comfy UI Interface for Enhanced Control.
3. Fine-Tuning the Model for Improved Output Quality.



3.1.1. Load Checkpoint

- ➢ Loads the Stable Diffusion model (v1-5-pruned-emaonly-fp16.safetensors).
- ➢ Outputs:
  - Model (sent to KSampler).
  - CLIP (Text Encoder) (sent to CLIP Text Encode nodes).
  - VAE (Variational Autoencoder) (sent to VAE Decode).

3.1.2. CLIP Text Encode (Prompt)

- Encodes the positive prompt:
    - "beautiful scenery nature glass bottle landscape, purple galaxy bottle"
- The encoded output is passed to KSampler as positive conditioning.

3.1.3. CLIP Text Encode (Prompt)

- Encodes the negative prompt:
    - "text, watermark" (used to avoid unwanted text or watermarks in the generated image).
- The encoded output is sent to KSampler as negative conditioning.

3.1.4. Empty Latent Image

- Creates a blank latent tensor (512x512).
- This acts as the starting point for diffusion in KSampler.

3.1.5. KSampler

- Performs the denoising and sampling process using the given model and prompts.
- Inputs:
    - Model (from Load Checkpoint).
    - Positive prompt encoding.
    - Negative prompt encoding.
    - Latent image (512x512).
- Parameters:
    - Seed: 208543671396552
    - Steps: 20
    - CFG Scale: 8.0
    - Sampler: Euler
    - Scheduler: Normal
    - Denoise Strength: 1.00
- Outputs:
    - A processed latent image, which is sent to VAE Decode.

3.1.6. VAE Decode

- Decodes the latent image into a visible image.
- Takes the VAE model (from Load Checkpoint) and the latent image (from KSampler).
- Outputs the final RGB image.

3.1.7. Save Image

- Saves the generated image with the prefix "ComfyUI".

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

I. GPU: NVIDIA RTX 3060 or higher
II. RAM: 16GB+
III. Storage: SSD (500GB+)

3.2.2 Software Requirements:

I. Python 3.x
II. Comfy UI Framework
III. Stable Diffusion Model
IV. CUDA & PyTorch

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

- Beautiful scenery nature glass bottle landscape, purple galaxy bottle

- A beautiful hidden waterfall in a tropical jungle, exotic plants, glowing fireflies, fantasy-style, highly detailed, HDR

- Futuristic cyberpunk city at night, neon lights, rain-soaked streets, holographic advertisements, flying cars, ultra-detailed, cinematic composition, inspired by Blade Runner

- A floating island in the sky, waterfalls flowing into the clouds, glowing magical ruins, ultra-detailed, dreamlike aesthetics

- A powerful sorcerer casting a spell, swirling blue energy around their hands, wearing a flowing robe with intricate patterns, fantasy-inspired artwork

- A friendly dragon with big expressive eyes, sitting on a pile of gold, cartoon-style, adorable and colorful

- A haunted mansion in the middle of a dark forest, eerie fog, glowing red windows, unsettling atmosphere, horror concept art

- A shadowy figure standing under a flickering streetlight, deep shadows, unsettling mood, cinematic horror scene

- A tiny kitten sleeping on a giant fluffy cloud, warm pastel colors, dreamy and soft aesthetic, ultra-cute

4.2 GitHub Link for Code:

https://github.com/Ray9hub/AICTE-Project-snaps

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- Integration with video diffusion models.
- Optimization for faster image generation.
- Real-time implementation in creative tools.

5.2 Conclusion:

This project successfully demonstrates the power of Stable Diffusion and Comfy UI in AI-generated imagery. The flexibility of these tools enables users to generate high-quality images with greater control, bridging the gap between AI and artistic 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.