**IMAGE GENERATION USING STABLE DIFFUSION AND COMFYUI**

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

**Bhavana S,**

**bhuvanashivakumar1404@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.

**I sincerely express my gratitude to my guides, Jay Rathod, Adarsh p for their invaluable support and mentorship throughout this project. My heartfelt appreciation goes to TechSaksham, Microsoft, and SAP for offering this learning opportunity. I also extend my thanks to my peers and family for their continuous encouragement and motivation throughout this journey**

#### **ABSTRACT**

his project explores **image generation using Stable Diffusion and ComfyUI**, focusing on creating AI-generated images from text prompts through a user-friendly workflow. The implementation involves setting up ComfyUI, integrating the **Hugging Face Table Diffusion model**, and configuring the checkpoint files. The process includes running the application on **CPU or GPU**, which opens the **ComfyUI browser interface** for image generation. The workflow is designed to streamline **image creation** by adjusting prompts and model parameters efficiently. The key contributions include a structured workflow, effective module integration, and performance optimization for high-quality outputs.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 1

1.4. Scope of the Project 1

**Chapter 2.**  **Literature Survey 2**

**Chapter 3.**  **Proposed Methodology 4**

**Chapter 4.**  **Implementation and Results 5**

**Chapter 5. Discussion and Conclusion 6**

**References** 7

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

AI-generated images are becoming increasingly useful in various industries such as art, design, and advertising. However, efficient workflows for generating high-quality images with minimal effort are still evolving.

* 1. **Motivation:**

The project was chosen to explore **ComfyUI**, an advanced tool for handling **Stable Diffusion workflows**, making AI image generation more accessible and structured.

* 1. **Objective:**
* To set up Stable Diffusion with ComfyUI.
* To create high-quality AI-generated images.
* To develop structured workflows for better image control.
  1. **Scope of the Project:**

This project focuses on using **Stable Diffusion and ComfyUI** to generate images based on text prompts, improving creativity and efficiency in AI art generation.

**CHAPTER 2**

**Literature Survey**

**2.1 Review of Relevant Literature  
AI-based image generation has undergone significant advancements, evolving from early neural network models to more sophisticated approaches like Generative Adversarial Networks (GANs) and Diffusion Models. GANs, first introduced by Ian Goodfellow in 2014, have played a crucial role in image synthesis, allowing the generation of realistic images through adversarial training. StyleGAN and BigGAN further improved the quality and resolution of generated images. However, GANs suffer from instability and mode collapse, limiting their scalability.**

**Diffusion models, such as Denoising Diffusion Probabilistic Models (DDPMs) and Stable Diffusion, emerged as alternatives that overcome these issues. Diffusion models work by gradually refining images from noise, leading to more stable and diverse image outputs. Stable Diffusion, in particular, has gained popularity due to its ability to generate high-quality images from text inputs efficiently.**

**2.2 Existing Models and Techniques  
Several approaches have been developed to improve AI-driven image synthesis:**

**BigGAN & StyleGAN: Provide high-resolution and detailed image synthesis.**

**OpenAI’s DALL·E: Generates diverse and complex images from textual descriptions.**

**Latent Diffusion Models (LDMs): Enhance stability and efficiency in image generation.**

**ComfyUI: A graphical interface designed to streamline Stable Diffusion workflows, allowing better control and accessibility for users.**

**2.3 Gaps and Limitations in Existing Solutions  
Despite these advancements, current methods still face limitations:**

**Computational Complexity: GANs and diffusion models require significant computational power, making them resource-intensive.**

**User Accessibility: Many AI-based image generation tools lack an intuitive interface, making them difficult for non-experts to use.**

**Fine Control: Generating specific image outputs requires extensive tweaking and prompt engineering.**

**How This Project Addresses the Gaps**

**By integrating Stable Diffusion with ComfyUI, this project simplifies the image generation process through a structured and visual workflow.**

**It enhances user accessibility by providing a node-based interface that allows users to easily customize model parameters.**

**It optimizes computational efficiency by leveraging pre-trained diffusion models, reducing the need for extensive training.**

**The project contributes to the field by improving workflow efficiency and usability for AI-generated content creation.**

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

The proposed system follows a structured workflow to generate images from text prompts efficiently. The process involves:

* **CLIP Text Encoding**: Converts user-provided text prompts into numerical embeddings using the **Contrastive Language-Image Pretraining (CLIP) model**.
* **Stable Diffusion Model**: Utilizes a latent diffusion model to generate high-quality images based on the encoded prompts.
* **ComfyUI Interface**: Provides a node-based graphical interface to manage the generation process seamlessly.

1. **Requirement Specification**

Mention the tools and technologies required to implement the solution.

**Hardware Requirements:**

* GPU: NVIDIA RTX 3060 or higher for optimal performance
* RAM: Minimum 16GB for smooth execution
* Storage: 50GB+ for model files and generated images

**Software Requirements:**

* ComfyUI
* Hugging Face model repository

**CHAPTER 4**

**Implementation and Result**

**Steps to Set Up the Project**

1. Download ComfyUI from the official repository:  
   [ComfyUI GitHub Repository](https://github.com/comfyanonymous/ComfyUI?tab=readme-ov-file#direct-link-to-download)
2. Extract and navigate to the ComfyUI directory.
3. Download the Stable Diffusion model checkpoint from Hugging Face:  
   [Hugging Face Model](https://huggingface.co/Comfy-Org/stable-diffusion-v1-5-archive/blob/main/v1-5-pruned-emaonly-fp16.safetensors)
4. Place the downloaded model inside the ComfyUI/models/checkpoints/ directory.
5. Start ComfyUI by running the CPU file (start\_cpu.bat or python main.py).
6. Open the ComfyUI browser interface, load the workflow, and set the prompt parameters.
7. Generate AI images by queuing the task and reviewing the outputs.
   1. **Snap Shots of Result:**
   2. **GitHub Link for Code:**

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**
* Improving model fine-tuning for better image generation.
* Integrating ControlNet for more precise outputs.
* Enhancing user experience by adding predefined workflows**.**
  1. **Conclusion:**

This project successfully demonstrates AI-based image generation using Stable Diffusion and ComfyUI. It highlights the importance of structured workflows in producing high-quality images efficiently.

**REFERENCES**

* 1. 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),* 10684-10695.
  2. **Dhariwal, P., & Nichol, A. (2021).** "Diffusion Models Beat GANs on Image Synthesis." *Advances in Neural Information Processing Systems (NeurIPS).*
  3. **CompVis Group (2022).** "Stable Diffusion: Latent Text-to-Image Diffusion Model.”