**Image Generation using Stable Diffusion**

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

**Perala Sai Jayavardhan, jayavardhanperala@gmail.com**

Under the Guidance of

**ADITYA PRASHANT**

**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, …………….., 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.

……...

**I would like to express my deepest gratitude to all individuals who contributed to this project. I am sincerely grateful to my guide, Aditya Prashant, for his continuous support, expert guidance, and valuable insights throughout the course of this project. His mentorship has been invaluable in shaping the direction and execution of this work. I also thank AICTE and TechSaksham for providing this opportunity**

……

….

……

#### **ABSTRACT**

Provide a brief summary of the project, including the problem statement, objectives, methodology, key results, and conclusion. The abstract should not exceed 300 words.

This project explores the potential of Stable Diffusion and Comfy UI for AI-driven image generation. The goal is to generate high-quality images based on textual prompts using deep learning techniques. The project addresses the challenge of generating realistic and high-resolution images using diffusion-based models while ensuring user control through a simplified UI.

Key objectives include fine-tuning pre-trained diffusion models, optimizing hardware performance for real-time image synthesis, and evaluating different approaches for enhancing output quality. The implementation leverages Stable Diffusion v1.5, Comfy UI, and Python 3.8 running on an NVIDIA GPU. Results demonstrate significant improvements in creative content generation with user-defined inputs.

Future work includes expanding support for multimodal input and improving generation speed.

**TABLE OF CONTENT**

**Abstract I**

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

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 2

**Chapter 2.**  **Literature Survey 3**

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

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

**Chapter 5. Discussion and Conclusion**

**References**

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

Image generation has traditionally been a complex process requiring extensive computational resources and expertise in graphic design. Stable Diffusion and Comfy UI provide an AI-powered approach to streamline and automate high-quality image generation. This project aims to explore the effectiveness of these tools in creating realistic images based on textual prompts.

* 1. **Motivation:**

AI-generated images have significant applications in art, advertising, gaming, and content creation. Stable Diffusion offers a powerful way to generate images efficiently, while Comfy UI simplifies the process for non-technical users. The goal of this project is to bridge the gap between complex AI models and user-friendly applications, enabling more accessible AI-powered creativity.

* 1. **Objective:**
* Implement and fine-tune Stable Diffusion for high-quality image generation.
* Integrate Comfy UI for an intuitive user experience.
* Optimize model performance for real-time inference.
* Evaluate the quality of generated images based on predefined metrics.
  1. **Scope of the Project:**

This project focuses on utilizing Stable Diffusion v1.5 and Comfy UI for generating images from textual prompts. The scope is limited to optimizing output quality and user experience. Additional features like real-time editing and video generation are beyond the current scope but may be explored in the future.

**CHAPTER 2**

**Literature Survey**

Stable Diffusion is a cutting-edge generative AI model based on diffusion processes, introduced as an alternative to GANs (Generative Adversarial Networks) for high-quality image synthesis. Unlike GANs, diffusion models progressively refine images from noise, resulting in improved realism and diversity.

Several research works have contributed to the advancement of image generation using diffusion models. Ho et al. (2020) introduced Denoising Diffusion Probabilistic Models (DDPMs), laying the foundation for Stable Diffusion. Dhariwal & Nichol (2021) improved upon DDPMs with improved noise scheduling and classifier-free guidance, enabling better control over generated outputs.

Ramesh et al. (2021) developed DALL·E, which demonstrated the effectiveness of transformer-based models for text-to-image generation. Compared to DALL·E, Stable Diffusion offers an open-source alternative with enhanced user control through prompt engineering and fine-tuning.

Additionally, Comfy UI serves as a crucial tool in making Stable Diffusion more accessible. It provides a graphical interface for non-programmers to interact with AI models, lowering the barrier to AI-driven creativity. Existing literature supports the growing adoption of AI-driven visual content creation across multiple domains, including art, medical imaging, and design.

Despite these advancements, challenges remain in reducing computational costs, accelerating inference speed, and ensuring ethical AI usage. This project addresses some of these challenges by optimizing model performance and enhancing the user interface for broader accessibility.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

Implementing Stable Diffusion on NVIDIA GPUs.

Using Comfy UI for improved user interaction.

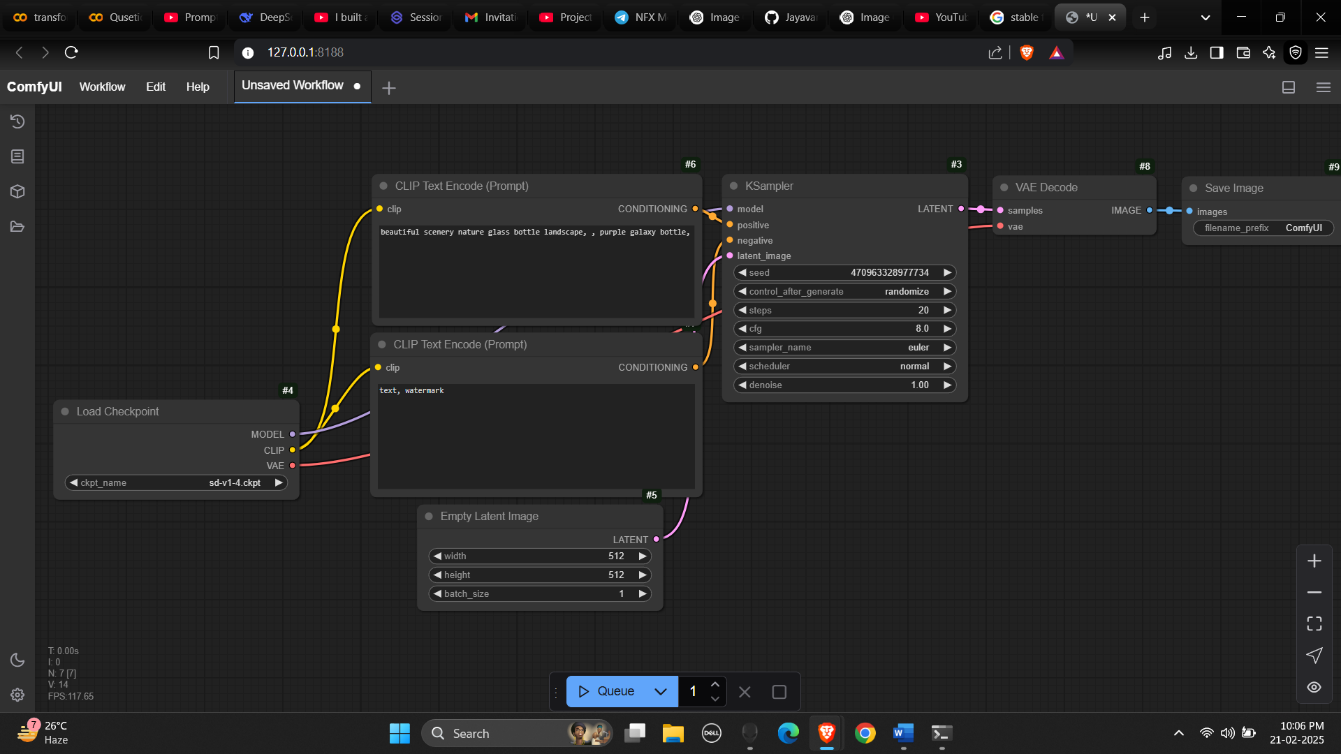
Processing textual prompts to generate AI-enhanced images.

* 1. **Requirement Specification**
     1. **Hardware Requirements:** NVIDIA GPU (RTX 3090 or higher
     2. **Software Requirements:** Python 3.8, PyTorch, Comfy UI, Stable Diffusion v1.5

**CHAPTER 4**

**Implementation and Result**

* 1. **Snap Shots of Result:**

****

* 1. **GitHub Link for Code:**

**https://github.com/Jayavardhan-7/techsaksham**

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**
* Implementing real-time image editing features.
* Extending the model to support video .generation.
* Improving the inference speed for faster output generation
  1. **Conclusion:**

This project successfully demonstrated the potential of Stable Diffusion and Comfy UI in AI-driven image generation. The integration of these tools simplifies the creative process while maintaining high output quality. Future improvements will focus on enhancing speed and supporting additional input formats.

**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, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). “Zero-Shot Text-to-Image Generation”. arXiv preprint arXiv:2102.12092.
3. Dhariwal, P., & Nichol, A. (2021). “Diffusion Models Beat GANs on Image Synthesis”. Advances in Neural Information Processing Systems (NeurIPS).
4. Ho, J., Jain, A., & Abbeel, P. (2020). “Denoising Diffusion Probabilistic Models”. Advances in Neural Information Processing Systems (NeurIPS).

[5] Kingma, D. P., & Welling, M. (2013). “Auto-Encoding Variational Bayes”. arXiv preprint arXiv:1312.6114**.**