**AI Image Generator**

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

**Vaibhav Rajaram Tatkare, vaibhavtatkare2004@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 my supervisor, 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.

I sincerely thank AICTE, Microsoft, and SAP for organizing this internship and giving me the opportunity to work on this project. I am also grateful to my instructor for guiding me throughout the process. This project helped me learn a lot about AI image generation, and I truly appreciate all the support and resources provided.

#### …..

……

….

……

#### **ABSTRACT**

The AI Image Generator project explores how artificial intelligence can create images from text descriptions. It uses deep learning, specifically the Stable Diffusion model, to generate images based on given prompts.

The objectives of this project include understanding the working principles of AI-powered image generation, implementing a model that can generate high-quality images, and evaluating the model's performance. The methodology involves leveraging pre-trained diffusion models and fine-tuning them based on specific requirements.

This report outlines the problem statement, motivation, literature survey, methodology, implementation details, results, and future work. The findings indicate that AI-generated images can significantly enhance creative processes by reducing the time and effort required for manual artwork creation. However, challenges such as image coherence, resolution, and ethical considerations remain areas for future improvement.

**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 6**

**Chapter 5. Discussion and Conclusion 9**

**References**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | U-Net Diffusion Process | **4** |
|  | CNN edges capturing | **4** |
|  | ComfyUI Interface | **6** |
|  | Ancient Indian Temple in the forest | **7** |
|  | Ancient Indian Temple in the forest (2nd attempt) | **7** |
|  | Lakeview sunset | **8** |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

Creating high-quality images by hand takes time and skill. AI-powered tools can help by automatically generating images based on text descriptions. This project aims to develop an AI system that can quickly create images from text.

* 1. **Motivation:**

The advancement of generative AI in art and design has opened new possibilities in various domains, including entertainment, advertising, and gaming. The ability to generate unique images through AI reduces dependency on manual artistry while accelerating content creation. This project explores the use of AI to enhance creativity and efficiency in digital art.

* 1. **Objective:**

Implement an AI-powered image generation model.

 Build an AI model to generate images.

 Learn how Stable Diffusion works.

 Test the quality of images created with different prompts.

 Identify the strengths and weaknesses of AI-generated images.

**1.4 Scope of the Project:**

This project focuses on leveraging AI to generate images from text-based inputs. The scope includes understanding Stable Diffusion, implementing an image-generation pipeline, and evaluating the results. However, challenges such as computational requirements and potential biases in AI-generated content are outside the immediate scope of this study.

**CHAPTER 2**

**Literature Survey**

* 1. **Review of relevant literature**

AI image generation has advanced significantly in recent years, with models like Generative Adversarial Networks (GANs) and Diffusion Models leading the way. Early AI-generated images lacked detail and consistency, but improvements in deep learning have made modern AI-generated visuals more realistic.

Stable Diffusion, introduced as an alternative to GANs, has proven effective in generating high-quality images from text prompts. Research shows that diffusion models can generate detailed and diverse images, making them useful for creative applications such as digital art, game design, and advertising [1].

Moreover, techniques for detecting patterns and features within images, which are essential for AI image generation, have been explored extensively. The face detection method by Yang et al. [1] and edge detection techniques by Canny [2] have paved the way for further refinement of feature extraction in AI models.

Furthermore, Szeliski's work on computer vision algorithms [3] contributes to the broader understanding of how image generation models process and synthesize visual information, enhancing the effectiveness and realism of AI-generated images.

* 1. **Existing Models, Techniques, and Methodologies**

Several AI-based approaches have been developed for image generation:

* Generative Adversarial Networks (GANs) – GANs use two competing neural networks (a generator and a discriminator) to create realistic images. However, they often suffer from mode collapse, where the model generates limited variations of images[1].
* Variational Autoencoders (VAEs) – VAEs learn efficient image representations but sometimes produce blurry images.
* Stable Diffusion – A state-of-the-art model that progressively refines noise into detailed images. It is more stable and scalable than GANs, making it suitable for text-to-image tasks
  1. **Gaps and Limitations in Existing Solutions**

While AI image generation has improved, some challenges remain:

* Image Coherence and Quality – Some AI-generated images may lack fine details or show distortions[1].
* High Computational Cost – Running diffusion models requires powerful hardware, limiting accessibility[1].
* Ethical Concerns – Biases in training data can affect image fairness and accuracy.

How This Project Addresses These Gaps:

* Improve image quality by experimenting with different prompt structures and model settings.
* Optimize processing to make AI-generated images more efficient.
* Raise awareness of ethical concerns and explore ways to minimize bias in AI-generated content.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

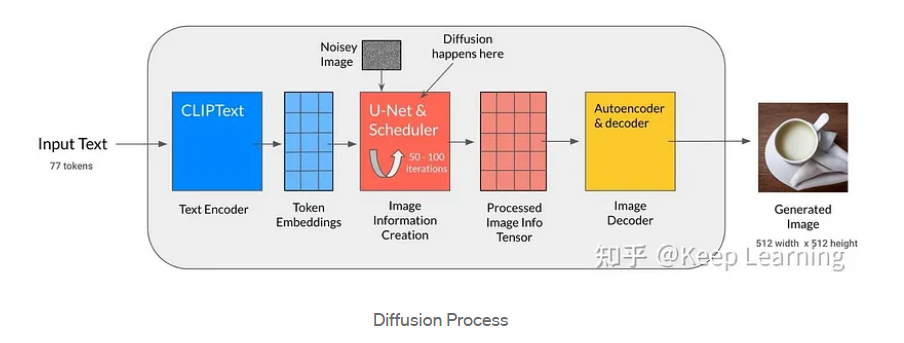


Figure 1 – U-Net Diffusion Process

There are 4 major processes involved in diffusion namely CNN, Pooling, Encoding and Decoding. Let’s explore them in detail:

**Convolution Neural Networks**

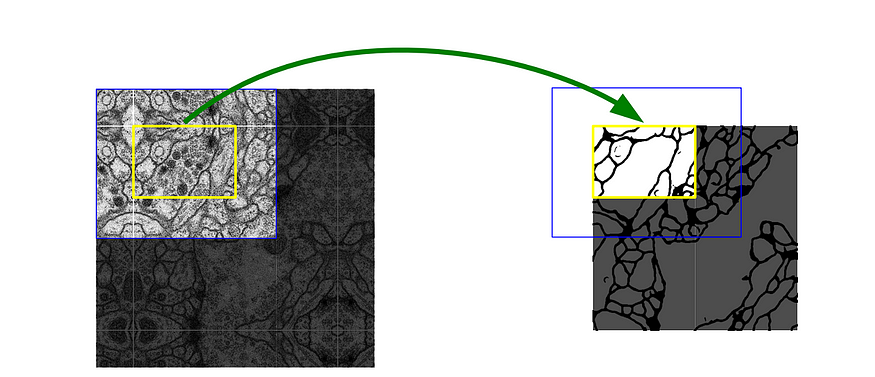


Figure 2 – CNN edges capturing

**Feature Extraction by CNN**

U-Net is a kind of neural network that helps turn text into images by learning patterns in pictures. CNN looks at small parts of an image to find important details using something called convolution. It saves these details to understand what’s in the picture.

**Pooling**

Max Pooling is like making a drawing smaller while keeping only the important parts, so the model doesn’t get confused by extra details.

**Encoding and Decoding**

1. **Encoding** the model starts with random noise and slowly learns important parts by breaking the image down.
2. **Decoding** takes what it learned and builds the image back, adding missing details and making it look clear. It’s like first drawing a rough sketch and then filling in the colours and small details to make it look real.
   1. **Requirement Specification**
      1. **Hardware Requirements:**

 CPU: Intel i5/i7 or AMD equivalent

 RAM: At least 8GB (16GB preferred)

 Storage: Minimum 20GB free space

* + 1. **Software Requirements:**

 Python

 Comfy UI

 Stable Diffusion Model

**CHAPTER 4**

**Implementation and Result**

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

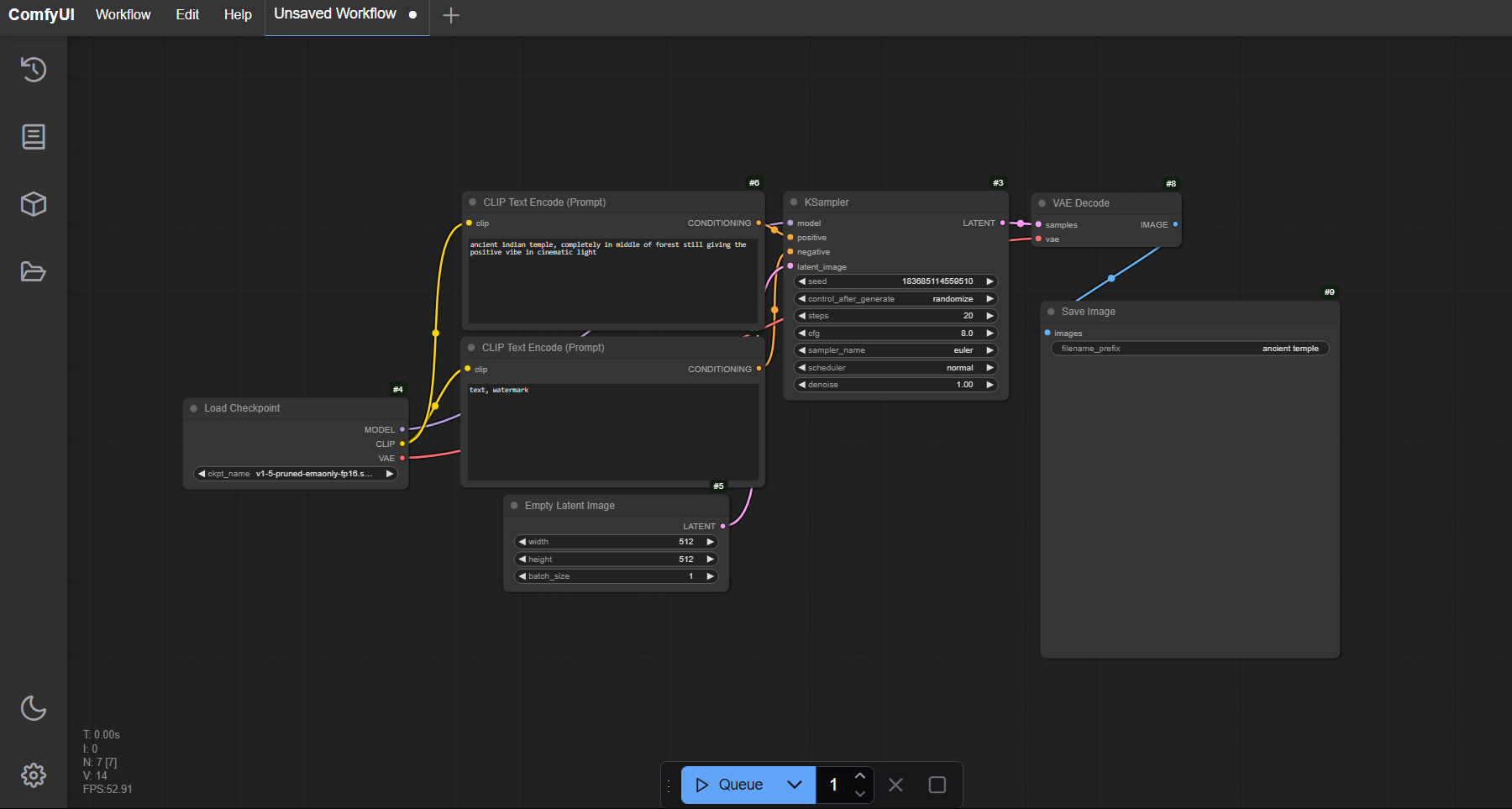


Figure 3 – ComfyUI Interface

**Explanation :**

1. Load Checkpoint: Loads the Stable Diffusion model (v1-5-pruned-emaonly).
2. CLIP Text Encode (Prompt): Encodes the main prompt (ancient Indian temple, completely in middle of forest).
3. CLIP Text Encode (Negative Prompt): Encodes a negative prompt (watermark to avoid unwanted elements).
4. Empty Latent Image: Defines the image size (512x512).
5. KSampler: Generates an image using 20 steps, Euler sampling, and a CFG scale of 8.
6. VAE Decode: Converts the latent image into a final image.
7. Save Image: Saves the output with the filename prefix ancient temple.

Output Image :

****

Figure 4 – Ancient Indian Temple in the forest

Image generated for, an ancient Indian temple, hidden in the heart of a forest, radiates a calming, positive energy, bathed in soft cinematic light.



Figure 5 – Ancient Indian Temple in the forest (2nd attempt)



Figure 6 – Lakeview sunset

A serene sunset over a calm lake, with vibrant orange and purple hues reflecting on the water, surrounded by lush green mountains

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

https://github.com/VaibhavT04/AI-Image-Generator-Diffusion-Model-and-ComfyUI.git

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**
* **Improving Image Resolution and Coherence:** Enhance image quality and consistency through model optimization and advanced techniques like super-resolution.
* **Enhancing Computational Efficiency:** Optimize the model for faster processing and lower computational requirements, possibly through algorithm improvements or parallel processing.
* **Exploring Ethical Considerations and AI Biases:** Address potential biases in AI-generated artwork and its impact on traditional art, ensuring fairness and inclusivity.
  1. **Conclusion:**

The AI Image Generator project successfully demonstrates the potential of artificial intelligence in creating digital artwork that is both innovative and visually appealing. While AI-generated images offer numerous advantages, such as increased efficiency and creative possibilities, challenges like data biases and the high computational resources required remain significant hurdles. As the field continues to evolve, future work will aim to enhance the model’s performance by optimizing its resolution, computational efficiency, and addressing ethical considerations, making it more accessible and impactful for diverse creative applications.

**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. John Canny, “A Computational Approach to Edge Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 8, No. 6, 1986.
3. Richard Szeliski, "Computer Vision: Algorithms and Applications," Springer, 2010.