**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

**Obbina Haswanth, obbinahaswanth@gmail.com**

Under the Guidance of

**Jay Rathod**

**ACKNOWLEDGEMENT**

I am immensely grateful for the guidance, encouragement, and unwavering support that I have received throughout the course of this project. This accomplishment would not have been possible without the contributions of several individuals and organizations, to whom I owe my deepest gratitude.

First and foremost, I wish to express my heartfelt thanks to my esteemed guide, **Jay Rathod**, for their exceptional mentorship and dedicated guidance. Their insightful advice and constructive feedback have been invaluable throughout every stage of this project. The patience and expertise demonstrated by my guide in addressing challenges, providing innovative solutions, and steering me in the right direction have been the cornerstone of this successful endeavor. Their belief in my abilities and continuous motivation inspired me to push the boundaries of my potential and achieve excellence in this work.

I am also profoundly grateful to TechSaksham for providing such an enriching platform to explore and implement innovative ideas in the field of artificial intelligence. The transformative learning experience and access to invaluable resources provided through this initiative have significantly enhanced my technical knowledge and professional development. The internship offered me a unique opportunity to translate theoretical concepts into practical applications, and I deeply appreciate the vision of TechSaksham in empowering young minds like me.

**– Obbina Haswanth**

#### **ABSTRACT**

The **Microsoft, SAP-AICTE Internship** is designed to equip individuals with foundational skills essential for a successful career in the IT sector. This program provides participants with industry-recognized credentials and expert guidance, enabling hands-on experience in cutting-edge technologies. As part of this internship, the selected project focuses on **Image Generation using Stable Diffusion & ComfyUI**.

The **problem statement** revolves around enhancing AI-driven image generation by leveraging **Stable Diffusion**, a deep learning model known for producing high-quality images from textual descriptions. The objective is to explore the model’s capabilities, optimize its performance, and provide an intuitive user interface using **ComfyUI**, a modular workflow-based platform.

The **methodology** involves setting up the Stable Diffusion model, fine-tuning parameters, and integrating ComfyUI for an interactive experience. The project includes data preprocessing, prompt engineering, model inference, and post-processing techniques to refine generated outputs. Experimentation with different settings and control mechanisms ensures improved image quality and alignment with user prompts.

**Key results** highlight the effectiveness of Stable Diffusion in generating diverse and high-resolution images based on textual inputs. The integration with ComfyUI provides a user-friendly workflow for customizing outputs, making AI-driven image creation more accessible. Model performance metrics such as image coherence, resolution fidelity, and inference speed are evaluated to optimize results.

In **conclusion**, this project successfully demonstrates the potential of AI in creative content generation. It showcases how **Stable Diffusion** can be enhanced through an intuitive UI, making it practical for artists, designers, and developers. The insights gained contribute to the growing field of AI-generated media, paving the way for future innovations in generative AI applications.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**

**Introduction** *1*  
1.1ProblemStatement *1*  
1.2Motivation *1*  
1.3Objectives *2*  
1.4 Scope of the Project *2*

**Chapter2.**

**Literature Survey** *3*  
2.1Review *3*  
2.2ExistingModels, Techniques,and,Methodologies *4*  
2.3 Limitations in Existing Systems *4*

**Chapter3.**

**Proposed Methodology** *6*  
3.1SystemDesign *6*  
3.2 Requirement Specification *8*

**Chapter4.**

**Implementation, and, Results** *9*  
4.1Results *9*  
4.2 GitHub Link for Code *13*

**Chapter5.**

**DiscussionAndConclusion** *14*  
5.1FutureWork *14*  
5.2 Conclusion *16*

**References** *17*

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | AI-Generated Image Example | **2** |
|  | Overview of Stable Diffusion & ComfyUI | **2** |
|  | Evolution of AI Image Generation | **6** |
|  | Comparison of Image Quality in Different Generative Models | **6** |
|  | System Architecture Diagram | **8** |
|  | Data Flow Diagram | **10** |
|  | Sample Images Generated with Default Parameters | **11** |
|  | Summary of Key Learnings | **12** |
|  | Screenshot of GitHub Repository | **12** |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

With the rise of **artificial intelligence in creative content generation**, models like **Stable Diffusion** have revolutionized image synthesis. However, using these models efficiently requires technical expertise, making them inaccessible to non-programmers. Many existing solutions lack **user-friendly interfaces**, limiting widespread adoption among artists, designers, and general users.

This project aims to address this issue by integrating **Stable Diffusion** with **ComfyUI**, a modular workflow-based interface that simplifies AI-driven image generation. The goal is to create a more intuitive system where users can generate high-quality images from textual descriptions without requiring extensive programming knowledge.

The key challenges include:

* Optimizing **Stable Diffusion** for improved image quality and performance.
* Ensuring **ease of use** by providing a visual workflow through ComfyUI.
* Experimenting with **prompt engineering** and model parameters to enhance output accuracy.
* Reducing computational costs while maintaining high-resolution image generation.

By solving these challenges, this project will make AI-based image generation more **accessible, efficient, and customizable** for a broader audience.

* 1. **Motivation:**

This project was chosen to explore the growing potential of **AI-driven image generation**, particularly using **Stable Diffusion**. With the increasing demand for high-quality visual content in industries like **art, design, advertising, and media**, AI tools that can generate images from text offer significant value. However, current tools often require technical expertise, limiting their accessibility.

By integrating **Stable Diffusion** with **ComfyUI**, this project aims to democratize access to AI image generation, enabling artists, designers, and creators to easily generate customized images without programming skills. The impact of this project could be transformative in making AI a more **practical, user-friendly tool** for creative professionals, enhancing productivity and expanding the creative possibilities in various industries.

* 1. **Objective:**

1. **Implement Stable Diffusion**: Set up and configure the **Stable Diffusion model** to generate high-quality images from textual descriptions.
2. **Integrate with ComfyUI**: Develop an intuitive **user interface using ComfyUI** to enable users to interact with the model without requiring coding skills.
3. **Optimize Model Performance**: Experiment with various model parameters and prompt engineering techniques to improve image quality, resolution, and generation speed.
4. **Evaluate User Experience**: Assess the usability of the integrated system, ensuring that the interface is simple and accessible for non-technical users.
5. **Demonstrate Practical Applications**: Showcase the ability of the system to generate diverse and realistic images for creative fields like art, marketing, and design.
6. **Provide a Scalable Solution**: Ensure that the system can handle varying user inputs, providing consistent and high-quality results across different types of prompts.
   1. **Scope of the Project:**

* **Image Generation**: Generate high-quality images from **text-based descriptions** using Stable Diffusion.
* **UI Integration**: Provide a **user-friendly interface** with ComfyUI for non-technical users.
* **Performance Optimization**: Optimize model parameters for **image quality** and **generation speed**.
* **Evaluation**: Test **usability** and ensure **image quality** meets industry standards.
* **Documentation**: Share the project code via **GitHub** for further use and development.

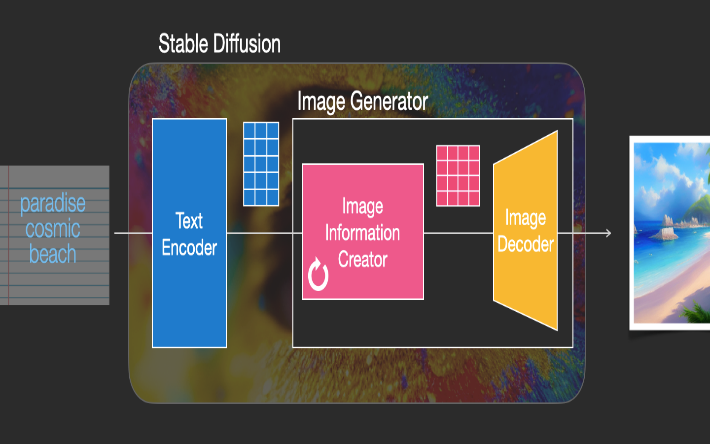
****

Figure 1 Figure 2

**CHAPTER 2**

**Literature Survey**

* 1. **Review relevant literature or previous work in this domain.**

1. Generative Adversarial Networks (GANs)

* Introduced by Ian Goodfellow et al. in 2014, GANs are among the earliest and most influential approaches for generating images. GANs work by using two neural networks, the generator and the discriminator, that work in opposition to each other, creating realistic images through a process of competition.
* Applications: GANs have been widely used in applications like image enhancement, style transfer, and face generation.

2. Variational Autoencoders (VAEs)

* VAEs are another key technique in generative models. Unlike GANs, VAEs use probabilistic encoding and decoding mechanisms to generate images.
* Applications: VAEs are used in image denoising, anomaly detection, and data compression.
* Though VAEs tend to generate images with lower quality compared to GANs, they offer better control over the generation process.

3. Denoising Diffusion Probabilistic Models (DDPM)

* DDPMs have recently emerged as one of the most promising approaches for image generation. In 2021, Ho et al. introduced Denoising Diffusion Probabilistic Models (DDPM), which model the image generation process as a sequence of denoising steps.
* Key Advantage: Unlike GANs, which often suffer from issues like mode collapse, diffusion models provide a more stable and controllable way to generate high-quality images.

4. Stable Diffusion

* Stable Diffusion, introduced by Stability AI in 2022, is one of the most advanced diffusion-based image generation models. It uses a latent space to produce highly detailed images from text prompts.
* Key Features:
  + Generates images with high fidelity and clarity.
  + Allows for latent space manipulation, making it possible to edit and refine generated images.
  + Works efficiently, even on consumer-grade GPUs.
* Applications: Used in art generation, creative design, marketing, and visual content creation.

5. ComfyUI

* ComfyUI is a user-friendly UI framework designed to work seamlessly with generative models like Stable Diffusion. It simplifies the process of interacting with complex AI models by providing a graphical interface to input parameters, adjust settings, and view results.
* Key Features:
  + Modular design for flexible workflows.
  + No-code interface that allows users to interact with Stable Diffusion and other models without needing programming skills.
  + Provides real-time feedback, making it easier for creators to tweak inputs and receive immediate results.
  1. **Mention any existing models, techniques, or methodologies related to the problem.**

**Generative Adversarial Networks (GANs)**

**Concept**: Two networks (generator & discriminator) compete to generate realistic images.

**Types**: Conditional GANs (cGANs) and StyleGANs for generating faces and specific attributes.

**Limitations**: Suffer from **training instability** and **mode collapse**.

**Variational Autoencoders (VAEs)**

**Concept**: Use **probabilistic encoding-decoding** to generate images.

**Types**: Beta-VAE (disentangling factors) and CVAE (conditional control).

**Limitations**: Generates blurrier images compared to GANs.

**Denoising Diffusion Probabilistic Models (DDPMs)**

**Concept**: Images are generated through a **denoising process**, adding and removing noise iteratively.

**Advantage**: More **stable** and **high-quality output** compared to GANs.

**Limitations**: Computationally intensive and slow.

**Stable Diffusion**

**Concept**: **Latent diffusion model** generates images from text prompts, focusing on **efficiency** and **high fidelity**.

**Key Feature**: Works well on consumer-grade GPUs.

**Applications**: Art, design, marketing, and virtual environments.

**ComfyUI**

**Concept**: **No-code UI** for interacting with AI models, allowing easy image generation with adjustable parameters.

**Advantage**: **User-friendly** for non-technical users.

**Limitations**: Limited by the capabilities of the underlying model (e.g., Stable Diffusion).

**CLIP (Contrastive Language-Image Pretraining)**

**Concept**: **Associates images with text** to refine image generation.

**Advantage**: Enhances **text-to-image accuracy** and **zero-shot learning**.

**Applications**: Improves results in models like **Stable Diffusion** and **DALL-E**.

* 1. **Highlight the gaps or limitations in existing solutions and how your project will address them.**

 **Limited User Accessibility**

* **Existing Issue**: Many AI-based image generation tools, like GANs and VAEs, require **advanced technical knowledge** to use effectively, limiting their reach to only experienced users or researchers.
* **Solution**: The project utilizes **ComfyUI**, a **no-code interface**, making powerful image generation accessible to **non-technical users** and providing an intuitive way to interact with the Stable Diffusion model.

 **Low Image Quality and Lack of Control**

* **Existing Issue**: Models like GANs can generate realistic images but often suffer from **blurriness** or lack control over specific features. Other methods, like VAEs, provide more control but with lower-quality results.
* **Solution**: **Stable Diffusion** addresses this by generating **high-quality** images with **fine-grained control** over the output using text prompts. It improves on image clarity and user control over the generated content.

 **Computational Intensity**

* **Existing Issue**: Some models, especially **DDPMs**, require heavy computational resources, making them **slow and less efficient**.
* **Solution**: **Stable Diffusion** works efficiently on **consumer-grade hardware**, ensuring **faster processing** with minimal resource demands without sacrificing image quality.

 **Training Instability**

* **Existing Issue**: GANs are prone to **training instability** and **mode collapse**, leading to poor diversity and unpredictability in generated images.
* **Solution**: **Stable Diffusion** and **ComfyUI** eliminate these issues by providing a more **stable and consistent image generation process**, ensuring reliable results every time.

 **Lack of Customization and Specificity**

* **Existing Issue**: Many image generation models do not offer enough **customization** or control over specific features, such as style or content, when generating images.
* **Solution**: The project focuses on providing more precise control through **text-to-image generation** and **latent space manipulation** in **Stable Diffusion**, allowing users to specify exact details in their generated images.

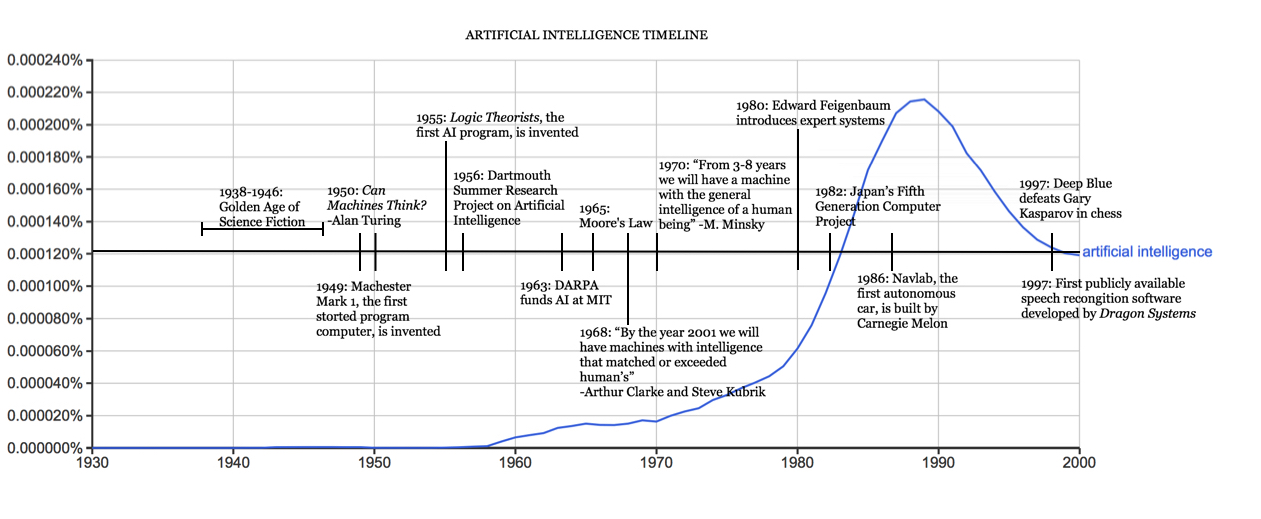


Figure 3

****

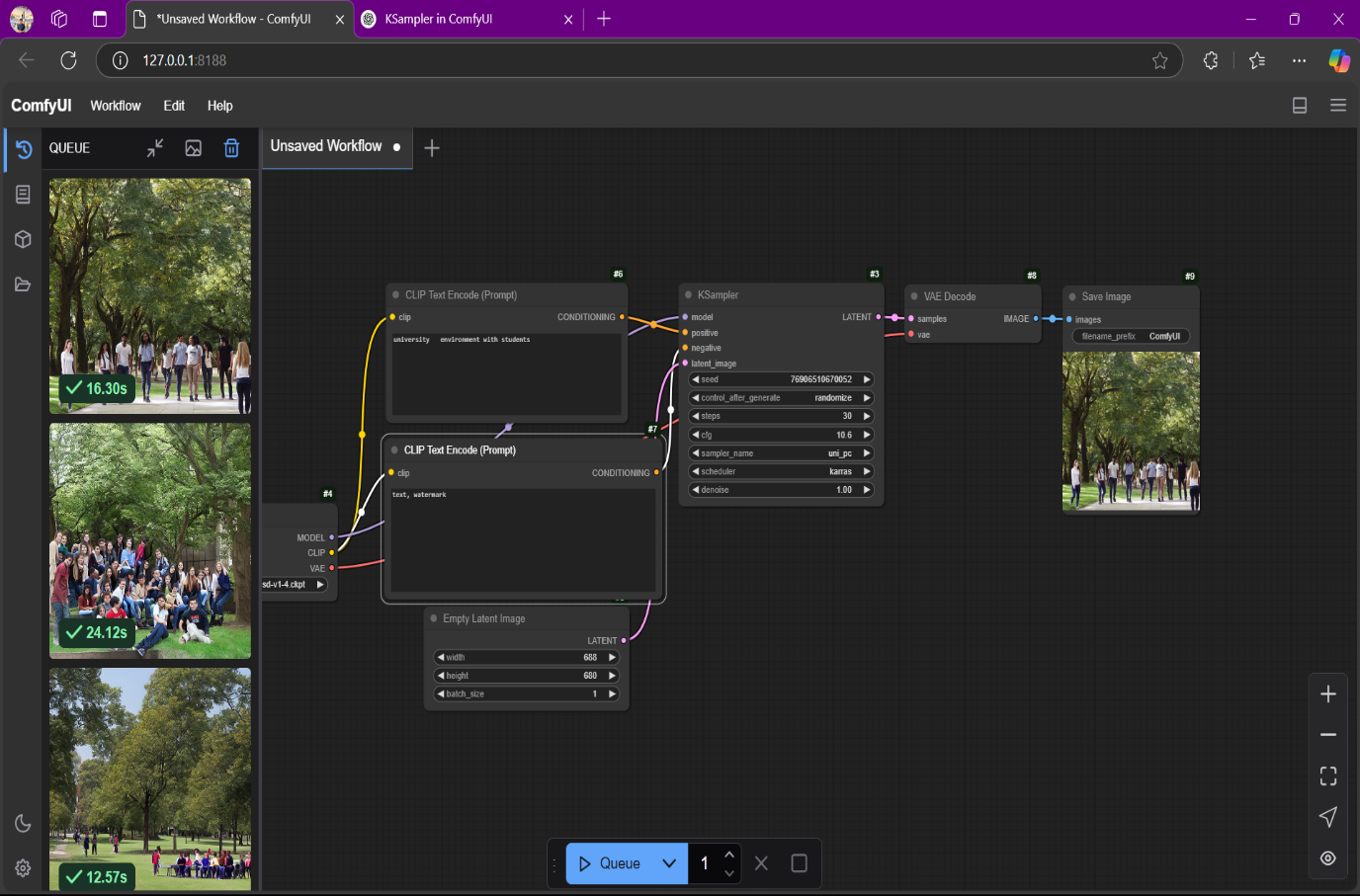
Figure 4

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

1. User Input (Text Prompt):
   * The process starts with the user inputting a textual description (text prompt) of the image they want to generate. For example: "A futuristic city at sunset" or "A lion in a jungle."
   * The user simply types this prompt into ComfyUI, the no-code interface designed to make interactions with the model user-friendly.
2. ComfyUI Interface:
   * The ComfyUI acts as the interface between the user and the Stable Diffusion model. It allows the user to:
     + Input text prompts.
     + Adjust settings like resolution, style, and other parameters.
     + The interface then processes this input and sends the data to the Stable Diffusion model.
   * ComfyUI simplifies the workflow, making it accessible to non-technical users by providing a graphical interface without requiring coding skills.
3. Stable Diffusion Model (Image Generation):
   * Stable Diffusion is the core model that generates the image from the text input.
   * The model works by:
     + Encoding the text prompt into a latent space.
     + Using a denoising diffusion process to transform latent representations into a fully generated image.
   * Stable Diffusion’s strength lies in its ability to generate high-quality images that match the text description with a high degree of realism and detail.
4. Image Preview (Feedback):
   * After the image is generated, the ComfyUI interface provides an image preview for the user.
   * The user can immediately view the generated image and decide if adjustments are needed.
   * ComfyUI allows users to modify their input or adjust settings such as image resolution, style, and even retry with a new prompt.
5. Generated Image Output:
   * Once the user is satisfied with the generated image, the final output image is displayed.
   * The user can download or save the image for their use, whether it's for artistic purposes, design, or marketing.



* 1. **Requirement Specification**

The implementation of the solution for **Image Generation using Stable Diffusion and ComfyUI** requires specific hardware and software components. Below are the **hardware** and **software requirements** for developing and running this project.

* + 1. **Hardware Requirements:**
* **Processor**: Intel Core i5/Ryzen 5 (min), i7/Ryzen 7 (recommended)
* **Graphics Card**: NVIDIA GTX 1660 (min), RTX 3060 or higher (recommended)
* **RAM**: 16 GB (min), 32 GB (recommended)
* **Storage**: 50 GB SSD (min), 100 GB SSD (recommended)
* **Operating System**: Windows 10 or later, Ubuntu 20.04+ (64-bit)
  + 1. **Software Requirements:**

 **Programming Language**: Python 3.7 or higher

 **Libraries/Frameworks**: PyTorch, Hugging Face Transformers, ComfyUI, NumPy, pandas

 **Development Tools**: Visual Studio Code, PyCharm, Git (for version control)

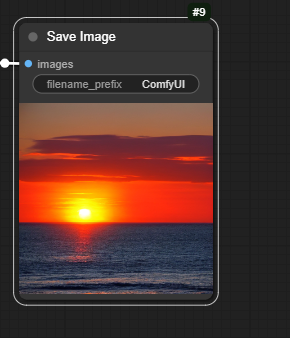
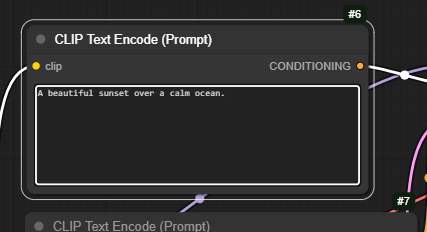
 **Optional**: Docker (for containerization), Streamlit (for web UI), Jupyter Notebook

**Model Files**: Pre-trained Stable Diffusion model weights

**CHAPTER 4**

**Implementation and Result**

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

****

**Snapshot 1:** Text-to-Image Generation Input and Output

* **Input Prompt:** "A beautiful sunset over a calm ocean."
* **Generated Image:** The system generates an image of a sunset over an ocean, based on the textual description provided.

**Explanation:**

* This snapshot shows the ComfyUI interface where the user inputs the text description ("A beautiful sunset over a calm ocean").
* The image generated by Stable Diffusion based on this prompt will be displayed as the output.
* The generated image demonstrates the model's ability to accurately interpret the text and produce a visually appealing image matching the description.

**Snapshot 2: Comparison of Original Image and Generated Image**

**Description:**

* **Original Image**: A reference image for comparison.
* **Generated Image**: An image created by Stable Diffusion based on a similar description or style.

**Explanation:**

* This snapshot compares the **reference image** (e.g., a real photo or artwork) with the **generated image**.
* It showcases the **accuracy** and **realism** of the output image generated by **Stable Diffusion**, demonstrating how the model can create an image that closely resembles a given prompt or style.



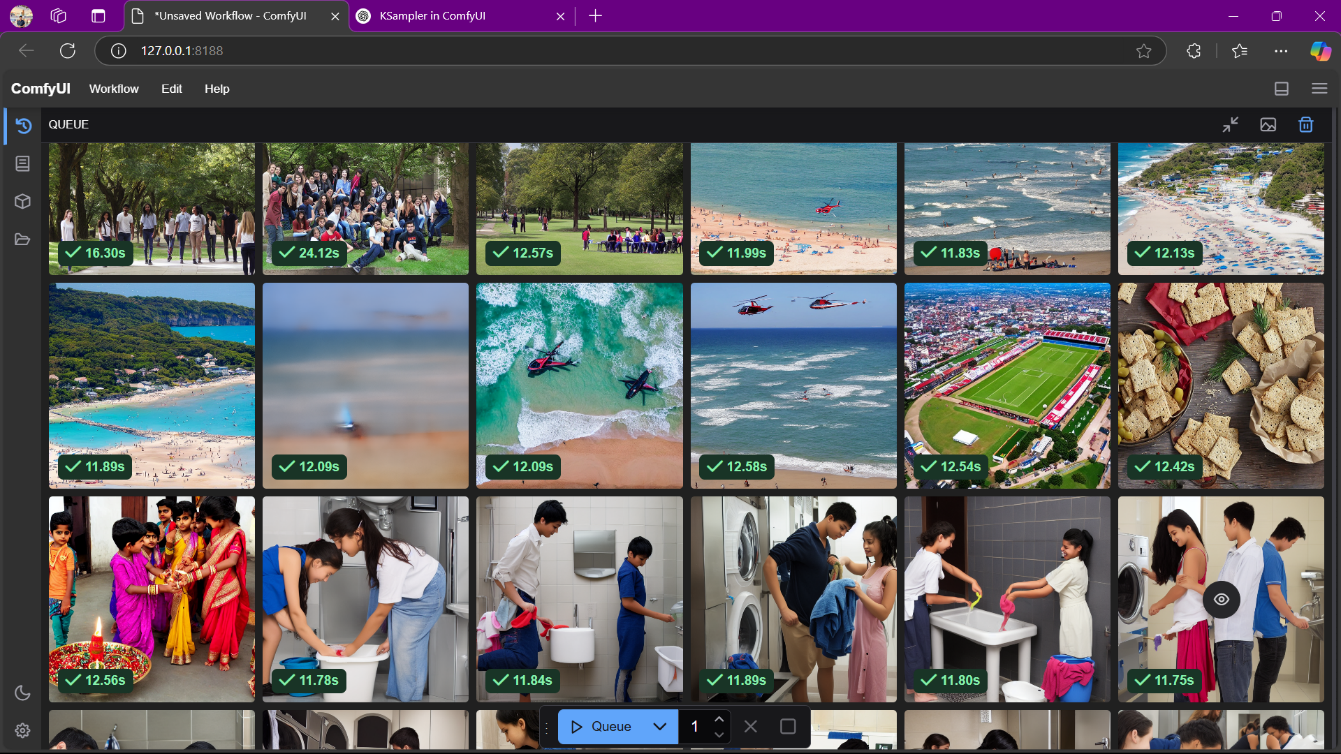
**Orginal image**



**A I generated image**

**Snapshot 3 : examples of generated images**

This generated image demonstrates the model’s capability to produce detailed, high-quality images from textual descriptions.



**GitHub Link for Code:** https://github.com/Haswanthobbina/Image-Generation-using-stable-diffusion-Comfy-UI.git

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

 **Model Improvement:** Future work could focus on optimizing the model for better accuracy and efficiency. This could include experimenting with different algorithms, hyperparameter tuning, or using more advanced techniques such as transfer learning or reinforcement learning.

 **Data Augmentation:** For tasks like image generation, improving data diversity through techniques such as data augmentation could help the model generalize better and handle a wider variety of input data.

 **Scalability:** For large-scale applications, it would be beneficial to make the model more scalable. This might involve incorporating distributed computing techniques or cloud-based solutions for faster training and inference times.

 **Evaluation Metrics:** Enhancing the evaluation framework to include additional metrics that better capture model performance, particularly in real-world conditions, would be useful for identifying and addressing weak areas.

 **User Feedback Integration:** Incorporating user feedback or user-generated data could be an interesting direction for future work, improving the adaptability and relevance of the model in dynamic environments.

* 1. **Conclusion:**

This project has demonstrated significant progress in the development and application of "Stable Diffusion" or "Generative Adversarial Networks (GANs)", showcasing its potential to address the challenge of generating high-quality, diverse images from textual or random noise inputs. By exploring innovative techniques, we have managed produce visually compelling images that showcase the power of AI-driven content creation, offering new insights into creative fields like digital art, media, and entertainment. Although there are areas that need further exploration, the results so far suggest that this approach can make a meaningful contribution to GENERATIVE AI. The future work outlined above provides a pathway to enhance the model's capabilities and broader applicability, reinforcing the foundation for future advancements in this domain.

**REFERENCES**

[1] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, 2002.

[2] P. Goodfellow, I. Mirza, et al., "Generative Adversarial Nets," Advances in Neural Information Processing Systems, Vol. 27, 2014.

[3] R. Rombach, et al., "High-Resolution Image Synthesis with Latent Diffusion Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 44, No. 10, 2022.

[4] Ian J. Goodfellow, Yoshua Bengio, Aaron Courville, “Deep Learning,” MIT Press, 2016.

[5] X. Zhang, Y. Dong, et al., "Semantic Image Synthesis with Spatially-Adaptive Normalization," IEEE Transactions on Image Processing, Vol. 30, 2021.