Department of Electrical and Computer Engineering
North South University

Senior Design Project

Prompt Analysis For Stable Diffusion

Saifur Rahman	ID# 1931794642
Md Rezwanul Alam Sayem	ID#1931809642
Jahid Akand Nahid	ID# 1813142642

Faculty Advisor:

Dr. Mohammad Ashrafuzzaman Khan

Associate Professor

ECE Department

Spring, 2023

LETTER OF TRANSMITTAL

June, 2023

To

Dr. Rajesh Palit
Chairman,
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: **Submission of Capstone Project Report on "Prompt AnalysisFor Stable Diffusion"**

Dear Sir,
With due respect, we would like to submit our **Capstone Project Report** on **"Prompt AnalysisFor Stable Diffusion"** as a part of our BSc program. The report deals with prompting in diffusion models. This project helped us enhance our skills in prompting and aided in generating specific images.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report useful and informative to have an apparent perspective on the issue.

Sincerely Yours,

....

Saifur Rahman
ECE Department
North South University, Bangladesh

....

Md Rezwanul Alam Sayem
ECE Department
North South University, Bangladesh

....

Jahid Akand Nahid
ECE Department
North South University, Bangladesh

APPROVAL

Saifur Rahman (ID # 1931794642), Md Rezwanul Alam Sayem (ID # 1931809642) and Jahid Akand Nahid (ID # 1813142642) from Electrical and Computer Engineering Department of North South University, have worked on the Senior Design Project titled “Prompt Analysis For Stable Diffusion” under the supervision of DR. MOHAMMAD ASHRAFUZZAMAN KHAN partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

Supervisor’s Signature

……………………………………

Dr. Mohammad Ashrafuzzaman Khan

Associate Professor

Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh.

Chairman’s Signature

……………………………………

Dr. Rajesh Palit

Professor

Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh.

DECLARATION

This is to declare that this project is our original work. No part of this work has been submitted elsewhere partially or fully for the award of any other degree or diploma. All project related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been properly acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students’ names & Signatures

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

1. Saifur Rahman

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

2. Md Rezwanul Alam Sayem

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

3. Jahid Akand Nahid

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project and research supervisor, Dr. Mohammad Ashrafuzzaman Khan, Associate Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance and advice pertaining to the experiments, research and theoretical studies carried out during the course of the current project and also in the preparation of the current report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh for facilitating the research. The authors would also like to thank their loved ones for their countless sacrifices and continual support.

ABSTRACT

Prompt Analysis For Stable Diffusion

The purpose of employing diffusion models for generating art or photos is to leverage a unique form of computational creativity. Diffusion models, particularly those rooted in deep learning architectures such as diffusion probabilistic models (DPNs), aim to simulate and enhance the creative process by producing novel and diverse outputs based on existing data.

A persistent challenge is determining whether the model will yield desirable outputs in response to corresponding prompts or generate unintended results. To address this issue, refining and optimizing our prompts is crucial. This involves systematically experimenting with and analyzing prompts' structure and content to enhance the generated outputs' quality and relevance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background and Motivation

Artificial intelligence (AI) has emerged as a pivotal technological advancement in recent years, significantly impacting medical, social, and economic domains. The progression of AI has reached a point where it can generate complex images from a few lines of command. Our motivation is to investigate the underlying mechanisms of this prompting technology, as understanding and optimizing prompts can generate a wide range of desired images.

1.2 Purpose and Goal of the Project

In this research project, our team aims to develop and utilize specialized prompts to rigorously test the limitations of various diffusion models. This investigation is pivotal for advancing our understanding of these models' capabilities and constraints, thereby informing the future design and implementation of AI-generated artwork. By systematically evaluating the performance of diffusion models under different conditions, we seek to identify their strengths and weaknesses, contributing valuable insights to the field of computational creativity and machine learning.

1.3 Organization of the Report

Chapter 1 presents a concise overview of diffusion models, including their fundamental principles and applications. Additionally, we outline the problem statement, delineating the specific challenges this project addresses. We articulate our objectives, detailing our prompt analysis's aims and ultimate goals. This chapter sets the foundation for understanding the context and significance of our research endeavors.

In Chapter 2, we conduct a comprehensive literature review of existing research on related topics. Through this extensive analysis of the current literature, we establish a well-defined guide to our project. This review enables us to identify key insights and methodologies, ensuring a robust foundation for our study.

Chapter 3 explores the necessary tools and software for our analysis. We delineate the system requirements and provide a detailed account of the system design. We also discuss the rationale behind the chosen system architecture and its benefits. This chapter elucidates the technical foundation of our project and justifies our methodological choices.

Chapter 4 constitutes a primary focus of this project, presenting the results and analysis, and discussing the outcomes of various prompting scenarios. This chapter provides a comprehensive evaluation of the experimental findings, offering insights into the effectiveness of different prompts. Additionally, it includes a critical reflection on the project, suggesting potential improvements and alternative approaches for future research.

Chapter 5 examines the project's impacts on social, health, and safety aspects. It addresses the legal implications of diffusion models and explores the associated moral challenges. Additionally, this chapter provides an overview of how these models influence environmental sustainability, highlighting their potential benefits and risks.

Chapter 6 comprehensively analyzes the time allocation for each project component, providing a detailed weekly roadmap of our activities. In addition, this chapter presents an overview of the budget and total expenditures associated with the project, if applicable. This thorough accounting facilitates understanding the project's resource management and financial planning.

Chapter 7 addresses the complex engineering challenges and activities encountered during the project.

Chapter 8 presents the conclusions drawn from our comprehensive analysis of diffusion models. This chapter offers a detailed overview of potential future improvements for both the models and the analytical methods employed, providing a roadmap for ongoing and future research endeavors.

Chapter 2 Research Literature Review

2.1 Existing Research and Limitations

Investigating Prompt Engineering in Diffusion Models: This paper describes techniques for determining the effect of specific words and phrases in prompts. They accomplish this by dividing the prompt into two sections: the physical and factual part and the style or how they want to display the physical segment. The words and phrases they used will fall into one of two categories. The experiment discovered that nouns cause significant shifts in images, whereas Simple adjectives have a minor impact on the image generated.

Optimal Prompts for Text-to-Image Generation: This paper describes a general framework for automatically adapting original user input to model-preferred prompts. Instead of manually prompt engineering, this will convert any user-provided prompt into a model-specific prompt for improved image generation. They began by performing supervised fine-tuning with a pre-trained language model on a small set of manually engineered prompts. They then moved on to reinforcement learning to investigate better prompts, in which a reward function is used to generate more aesthetically pleasing images while retaining the original user intent. Their experiment demonstrates that this model outperforms manual prompts in terms of automatic metrics and human preference ratings.

PromptMagician: Interactive Prompt Engineering for Text-to-Image Creation: This paper facilitates the exploration of image results and the refinement of input prompts, using a multi-level visualization for cross-modal embedding of images. The proposed system includes four key views: Model Input View, Image Browser View, Image Evaluation View, and Local Exploration View. The backbone of their system is a prompt recommendation model that takes user prompts as input, retrieves similar prompt-image pairs from DiffusionDB, and identifies special (important and relevant) prompt keywords.Their work differs from prior work by combining database retrieval and adhoc generation, enabling users to explore the vast artistic search space to identify effective prompt keywords for personalized creation and iteratively refine the prompts. According to the results, PromptMagician offers a useful framework for interactive prompt engineering in text-to-image generation, presenting a viable model for user and generative model cooperation.

A Taxonomy of Prompt Modifiers for Text-To-Image Generation: Based on a three-month ethnographic investigation, the paper proposes a novel taxonomy of six categories of prompt modifiers employed by practitioners in the online text-to-image community. The taxonomy includes Subject terms, picture cues, style modifiers, quality enhancers, recurring terms, and magic terms.The practice of prompt engineering is essential for controlling the generative model’s output, and the taxonomy provides a conceptual starting point for investigating text-to-image generation. It also outlines the practice of "prompt engineering" in

Human-Computer Interaction (HCI) and Human-AI Interaction (HAI), and discusses broader implications of prompt engineering in the field. This paper also discusses the social aspects of prompt engineering within the art community , highlighting the shared practices of practitioners and their learning from the community provided resources.

Generative Adversarial Text-to-Image Synthesis (GAN-INT-CLS): This paper introduces a novel approach for generating images from textual descriptions using Generative Adversarial Networks (GANs). The proposed model consists of a text encoder to map textual descriptions to a semantic embedding space and a GAN-based image generator conditioned on these embeddings. Notably, the discriminator is extended to also consider the encoded text embeddings, enhancing the coherence between generated images and textual descriptions. The model is trained in an adversarial manner, where the generator aims to produce realistic images that are indistinguishable from real images, while the discriminator aims to distinguish between real and generated images. Experimental results demonstrate the effectiveness of the proposed approach in generating high-quality images conditioned on textual descriptions.

AttnGAN:Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks:

This paper presents AttnGAN, an attention-based generative adversarial network for fine-grained text-to-image generation. AttnGAN consists of two stages: a conditional text generation stage and an image generation stage. In the conditional text generation stage, a hierarchical attention mechanism is employed to generate informative and diverse textual descriptions based on a given input text. In the image generation stage, a novel attentional generative network is proposed to synthesize high-resolution images conditioned on the generated textual descriptions. The attention mechanism allows the model to focus on relevant regions of the image while generating details corresponding to different parts of the input text. Experimental results demonstrate that AttnGAN outperforms existing methods in generating visually appealing and diverse images from textual descriptions.

Chapter 3 Methodology

3.1 System Design

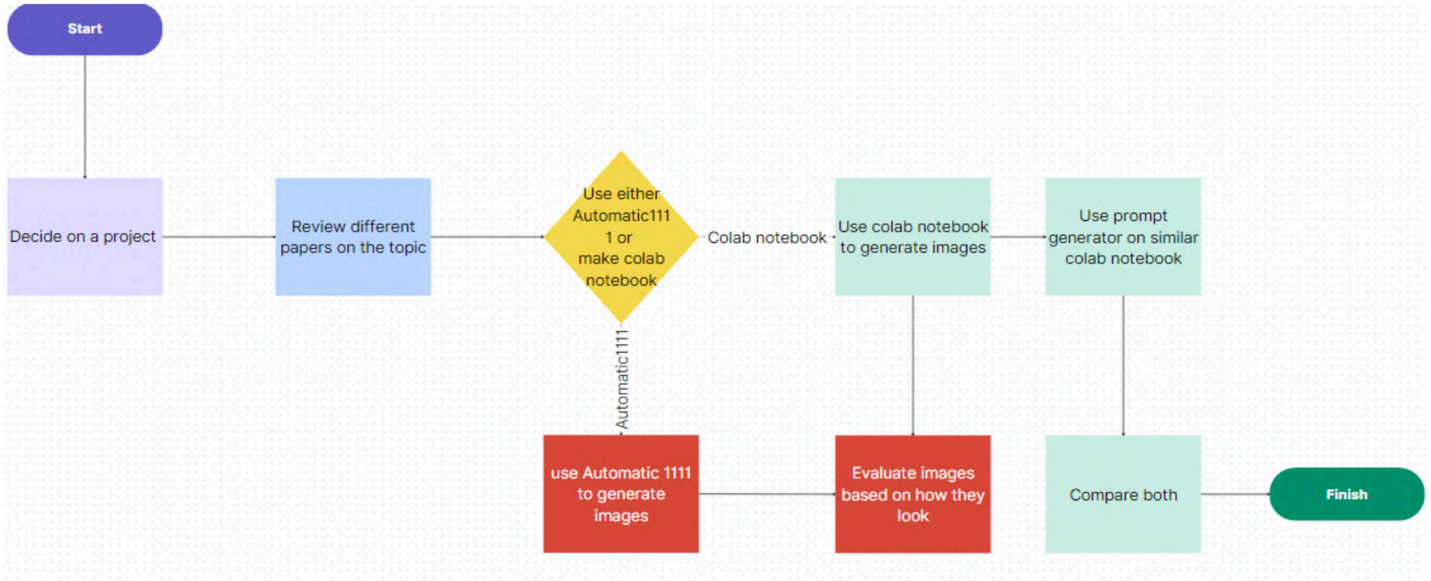

Start
Decide on a project
Review different papers on the topic
Use either Automatic111 1 or make colab notebook
Colab notebook
Use colab notebook to generate images
Use prompt generator on similar colab notebook
Automatic1111
use Automatic 1111 to generate images
Evaluate images based on how they look
Compare both
Finish

Figure:1 Block diagram of Prompt Analysis

We began by selecting "Prompt Analysis for Stable Diffusion" as our primary research topic and reviewed existing system designs for similar projects. We identified an existing system, Automatic 1111, which operates locally. Consequently, we developed our Google Colab notebook to generate images based on textual prompts. This approach enables us to systematically track all prompts and settings for generating specific images.

Subsequently, we evaluated the generated images using metrics such as the CLIP score, utilizing OpenAI's 'clip-vit-base' model. Following this, we manually compared the images we generated and those produced using prompts from a prompt generator. This comparison aimed to determine the efficacy and quality of the prompts, identifying the best-performing prompts and understanding the reasons for their success.

3.2 Hardware and/or Software Components

The entirety of our project was coded in Python. Google Colab served as the initial platform, providing free GPU access for image generation. Upon reaching the free usage limit, we transitioned to a local machine utilizing Automatic 1111 or leveraged Kaggle for additional free GPU resources. Visual Studio Code facilitated local execution of the code and intermittently ran the prompt generator. Additionally, we utilized GitHub to maintain backup copies of our code, ensuring continuity and version control in the event of any issues.

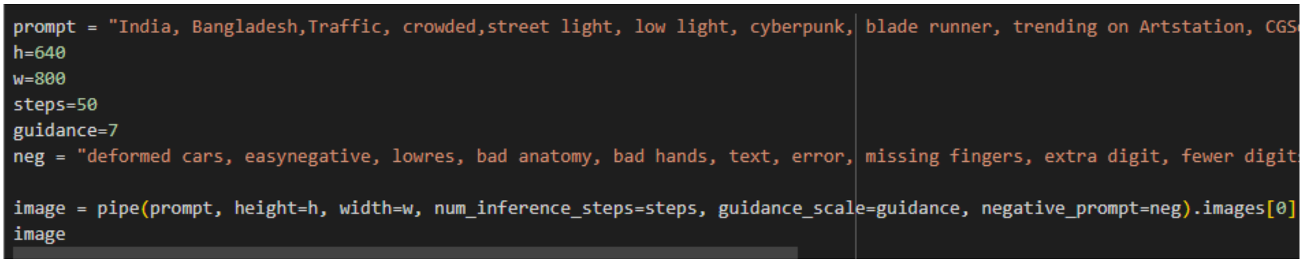
Table I. List of Software/Hardware Tools

Tool	**Functions**	**Other similar Tools (if any)**	**Why selected this tool**
Python 3	A high-level programming language used for developing the core functionality and scripts for the project.	R, MATLAB, Julia	Widely used in AI/ML research, extensive libraries, and ease of use.
Google colab	An online platform providing free GPU access for executing Python code and running experiments in a cloud environment.	Jupyter Notebooks, Kaggle	Provides free GPU access, facilitates collaboration, easy integration with Python.
diffusers pipeline	A library from Hugging Face for implementing diffusion models, enabling the generation of images from text prompts.	NA	Specialized for diffusion models,
Vs code	A source-code editor used for writing, debugging, and running the project's code.	PyCharm, Atom	Highly extensible, integrated Git support, and great for Python development.
Prompt generator	This tool is used to generate textual prompts for creating images, allowing comparison and analysis of different prompts.	NA	Enables generation of varied prompts, facilitating comparative analysis.

Github	Github is used to import different projects from their repositories. Furthermore it is also used to keep track of the different versions of the same code for better debugging.		To keep track of all the codes created all throughout the project.

3.3 Hardware and/or Software Implementation

Our project operates without a graphical user interface (GUI). Instead, we have chosen to utilize the 'diffusers' library from 'Hugging Face.' This approach allows us to correct any errors encountered during the image generation process and facilitates on-the-fly troubleshooting. Additionally, it enables us to deepen our understanding of the underlying mechanics. In the subsequent slide, we will present an example demonstrating the operational workflow of our project.



```
prompt = "India, Bangladesh,Traffic, crowded,street light, low light, cyberpunk, blade runner, trending on Artstation, CGS
h=640
w=800
steps=50
guidance=7
neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digit

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
image
```

Figure 2: Code block for generating images

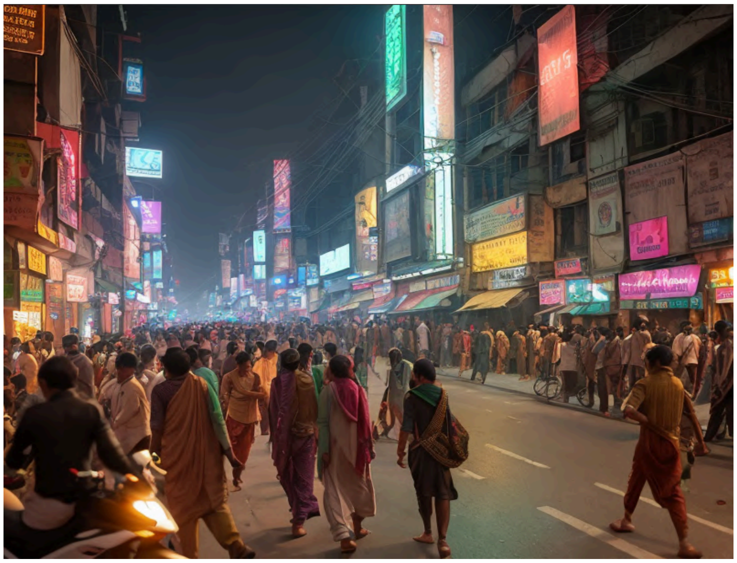


Figure 3: example of generated image

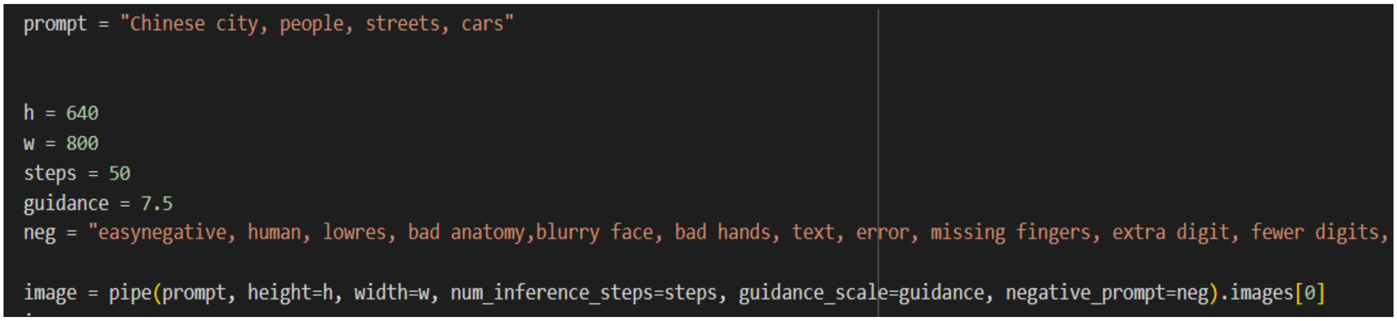
The image above is a direct result of executing the aforementioned code block. This method allows us to systematically track which images were generated by specific prompts, ensuring a clear and organized correlation between prompts and their corresponding outputs.

Chapter 4 Investigation/Experiment, Result, Analysis and Discussion

Our experiment successfully generated images that closely matched the given prompts, showcasing the capabilities of different diffusion models and how they respond to different types of prompts. Each image adhered to the described themes and settings, demonstrating each model's proficiency in rendering complex and detailed scenes through precise prompting.

Negative prompts helped refine the images by excluding undesirable elements, leading to cleaner and more accurate results. The guidance scale and number of steps played crucial roles in enhancing the quality and adherence of the generated images to the prompts.

We noticed while giving various prompts to different diffusion models to generate images that using one long sentence tends to produce lower-quality images compared to using shorter, comma-separated descriptions.



```
prompt = "Chinese city, people, streets, cars"


h = 640
w = 800
steps = 50
guidance = 7.5
neg = "easynegative, human, lowres, bad anatomy,blurry face, bad hands, text, error, missing fingers, extra digit, fewer digits,

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
```

Figure 4: Example of using not many words in the prompt (1)

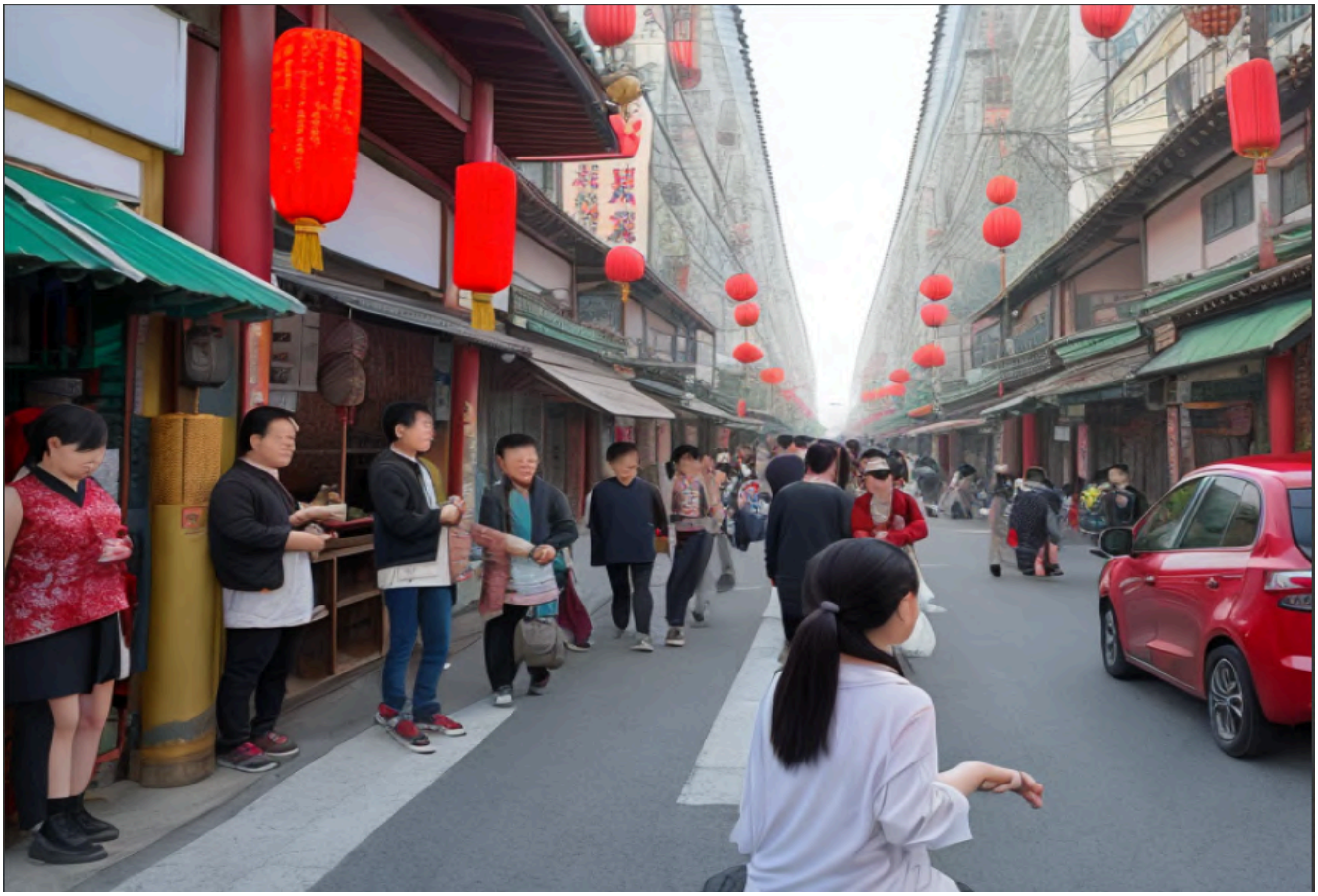
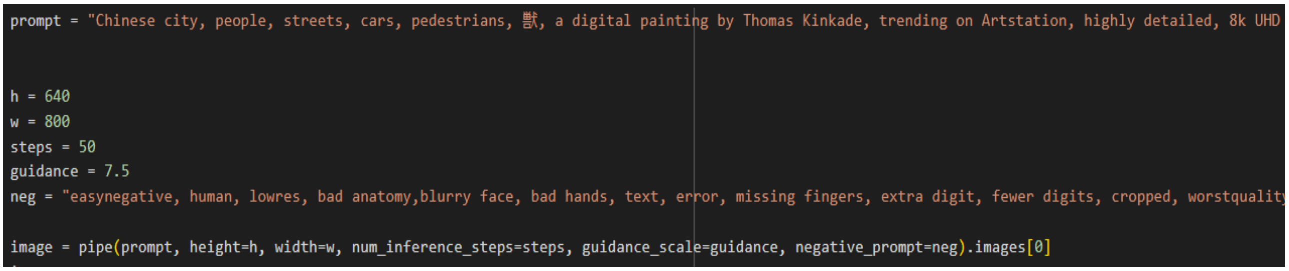


Figure 5: Example of using not many words in the prompt (2)



```
prompt = "Chinese city, people, streets, cars, pedestrians, 獸, a digital painting by Thomas Kinkade, trending on Artstation, highly detailed, 8k UHD


h = 640
w = 800
steps = 50
guidance = 7.5
neg = "easynegative, human, lowres, bad anatomy,blurry face, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
```

Figure 6: Example of using more words in the prompt separated by commas

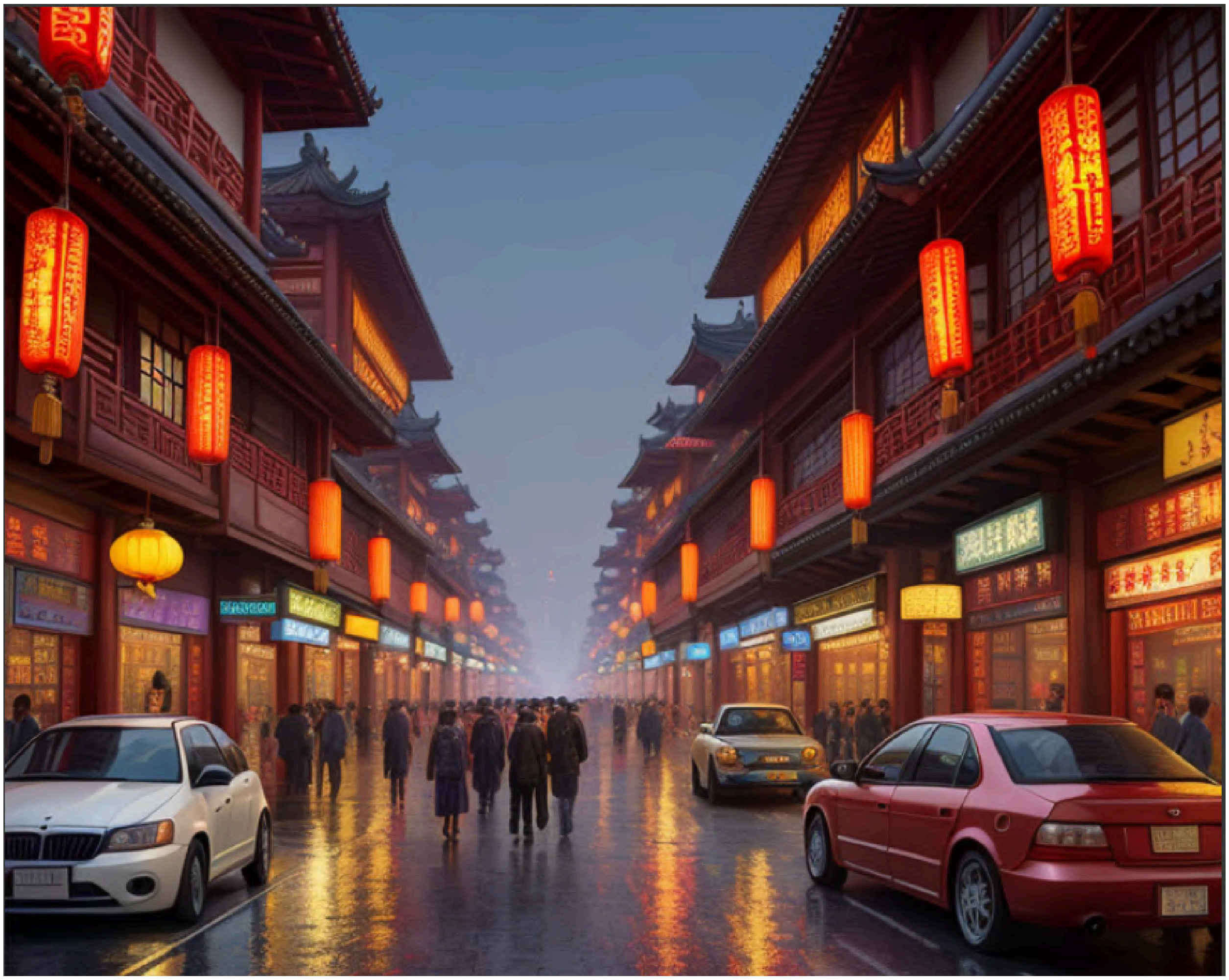
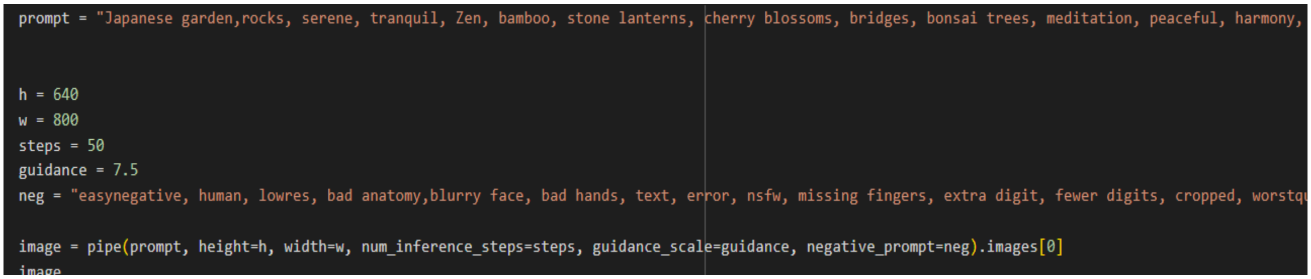


Figure 7: Image generated from figure 6’s prompt



```
prompt = "Japanese garden,rocks, serene, tranquil, Zen, bamboo, stone lanterns, cherry blossoms, bridges, bonsai trees, meditation, peaceful, harmony,


h = 640
w = 800
steps = 50
guidance = 7.5
neg = "easynegative, human, lowres, bad anatomy,blurry face, bad hands, text, error, nsfw, missing fingers, extra digit, fewer digits, cropped, worstqu

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
image
```

Figure 8: Example of a really long prompt separated by many commas

In figure 8 the prompt was "Japanese garden,rocks, serene, tranquil, Zen, bamboo, stone lanterns, cherry blossoms, bridges, bonsai trees, meditation, peaceful, harmony, water, wide angle, wide shot, sunrise, by gerald brom, by mikhail vrubel, by peter elson, muted colors, extreme detail, trending on artstation, 8 k"

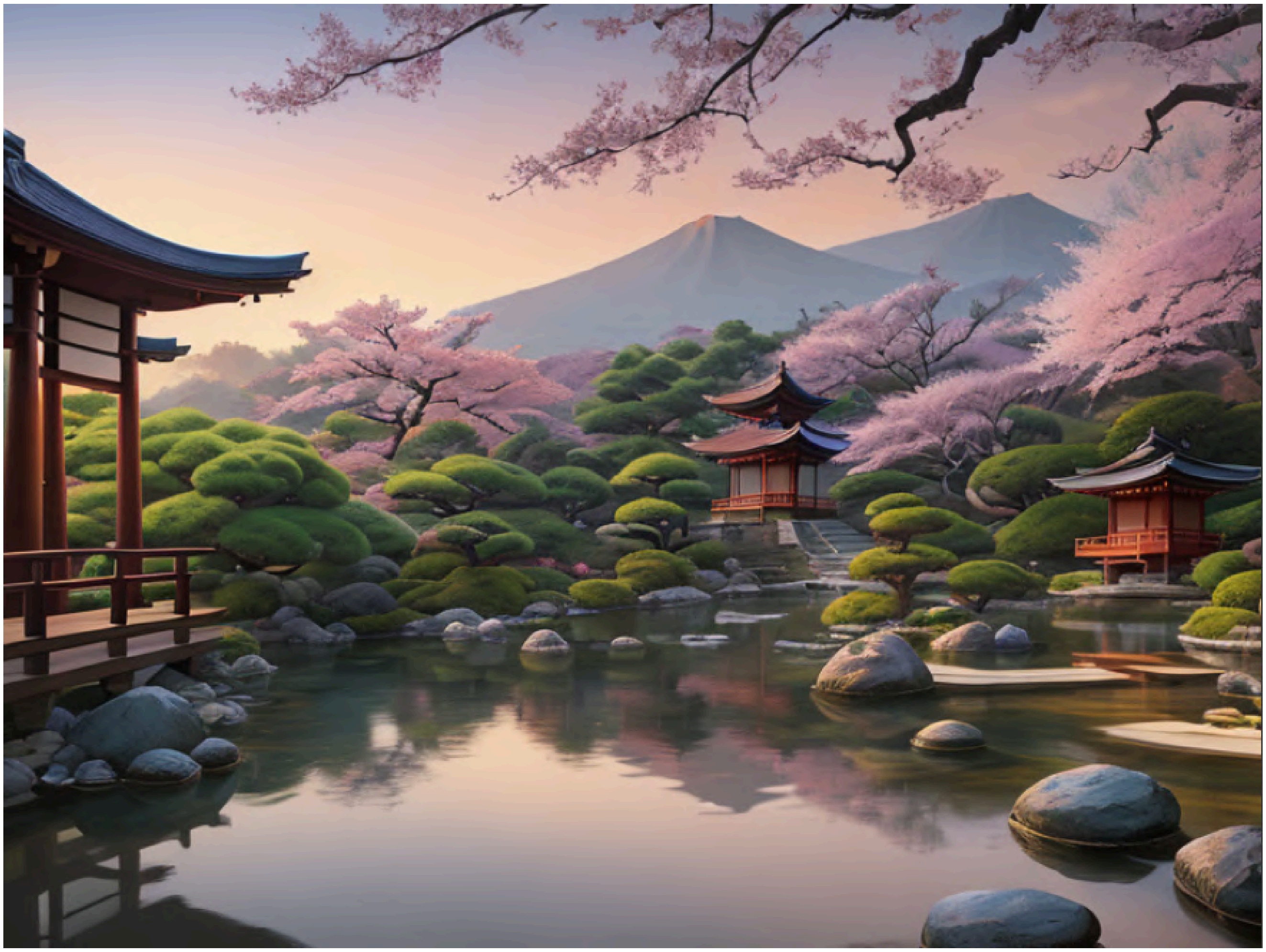
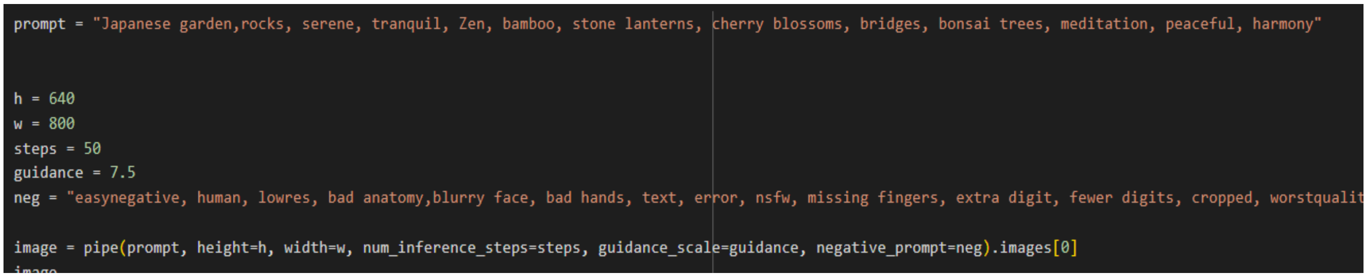


Figure 9: The image generated from figure 8's prompt



```
prompt = "Japanese garden,rocks, serene, tranquil, Zen, bamboo, stone lanterns, cherry blossoms, bridges, bonsai trees, meditation, peaceful, harmony"


h = 640
w = 800
steps = 50
guidance = 7.5
neg = "easynegative, human, lowres, bad anatomy,blurry face, bad hands, text, error, nsfw, missing fingers, extra digit, fewer digits, cropped, worstqualit

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
image
```

Figure 10: Example of a short description prompt

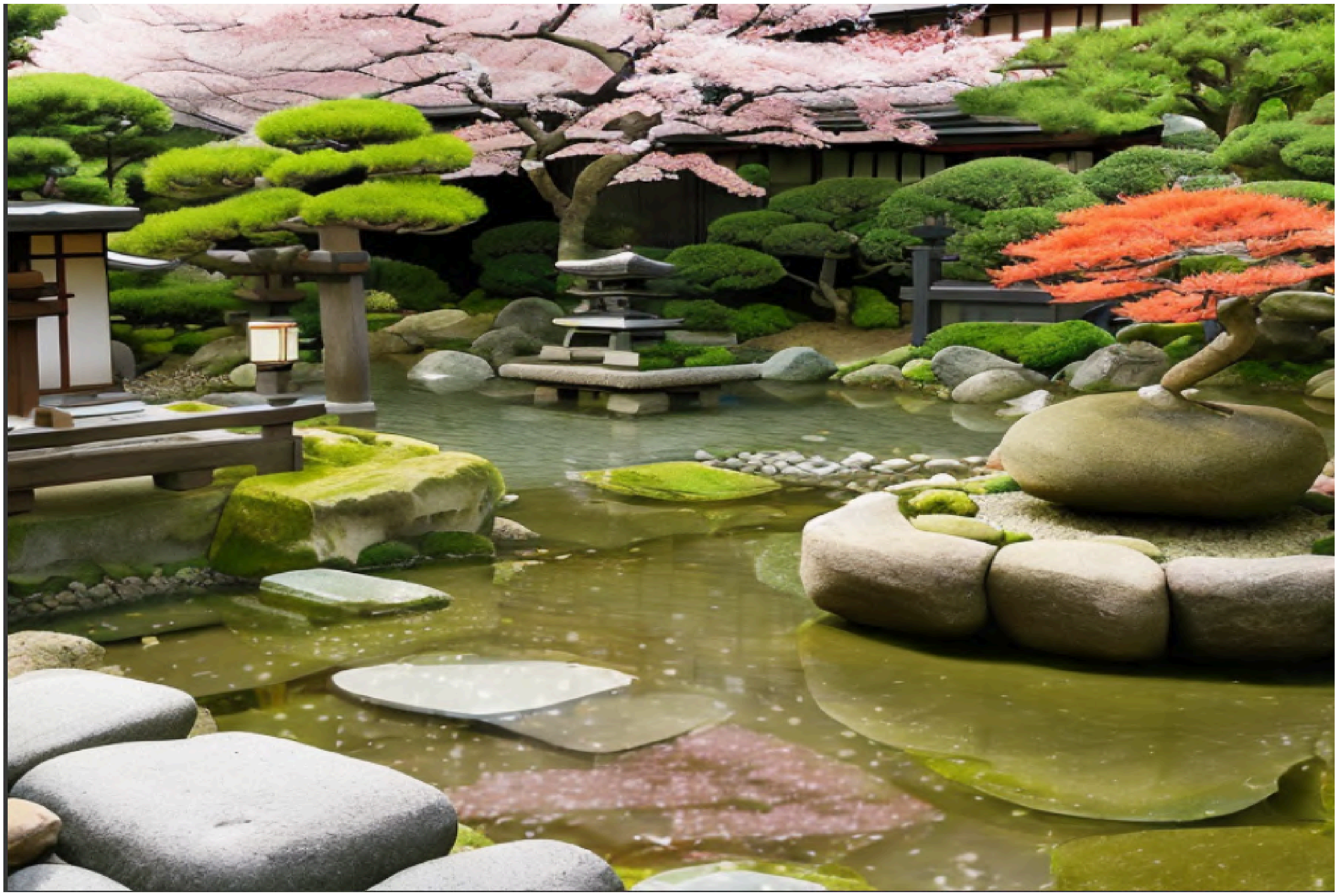
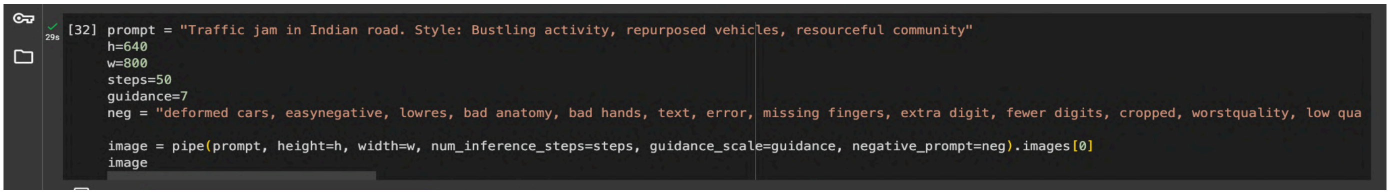


Figure 11:Image generated from figure 10's prompt

In order to produce realistic and aesthetically pleasing photographs of street crowds in Bangladesh and India utilizing diffusion models, the prompt's specificity and precision are essential. Certain places may yield superior results for models because of the variety and volume of training data that is available. However, we can enhance the quality of photos for Bangladeshi and Indian contexts by creating thorough and culturally relevant suggestions.



```
[32] prompt = "Traffic jam in Indian road. Style: Bustling activity, repurposed vehicles, resourceful community"
     h=640
     w=800
     steps=50
     guidance=7
     neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low qua

     image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
     image
```

Figure 12: Example of Indian crowds road.

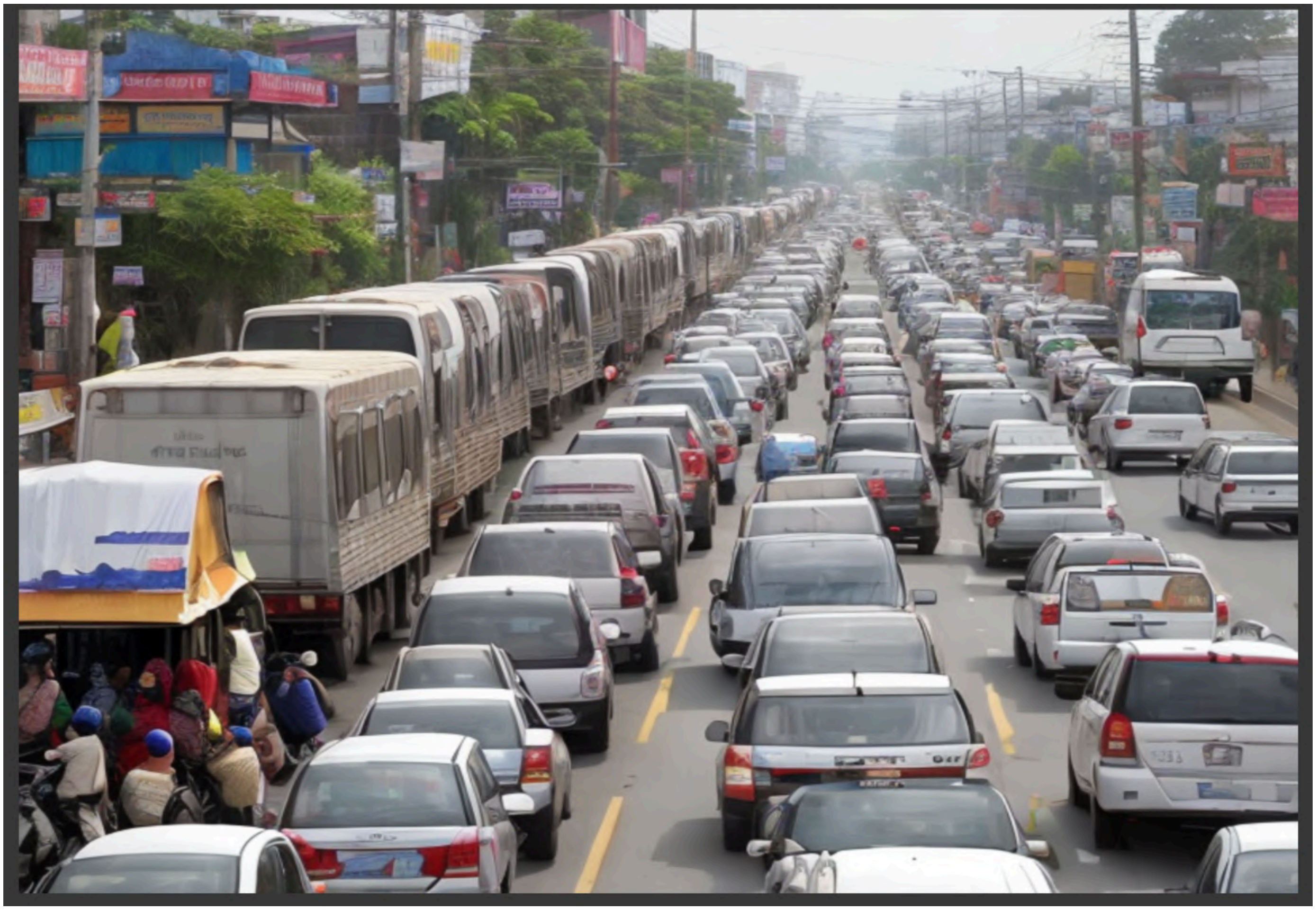
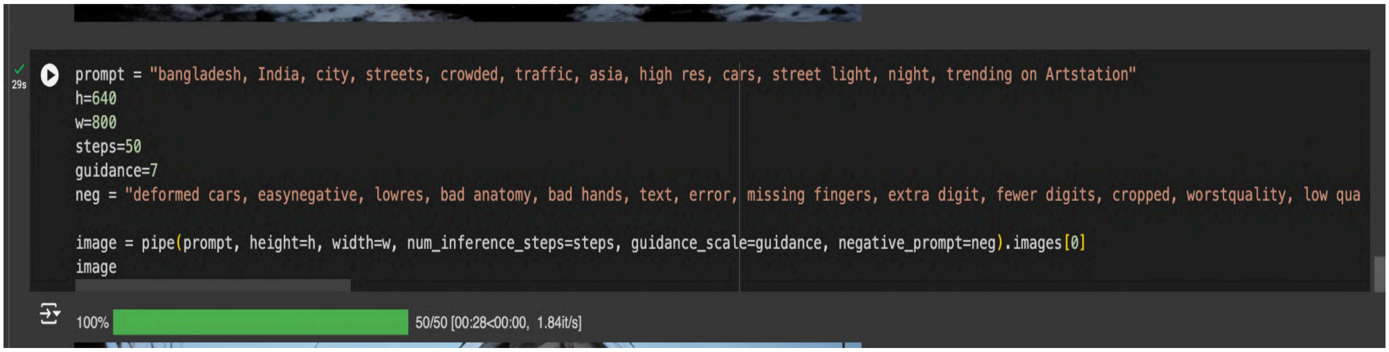


Figure 13:Image generated from figure 12’s prompt.



```
prompt = "bangladesh, India, city, streets, crowded, traffic, asia, high res, cars, street light, night, trending on Artstation"
h=640
w=800
steps=50
guidance=7
neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low qua

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
image
```

```
100% 50/50 [00:28<00:00,  1.84it/s]
```

Figure 14: Example of Bangladesh,India crowded streets.

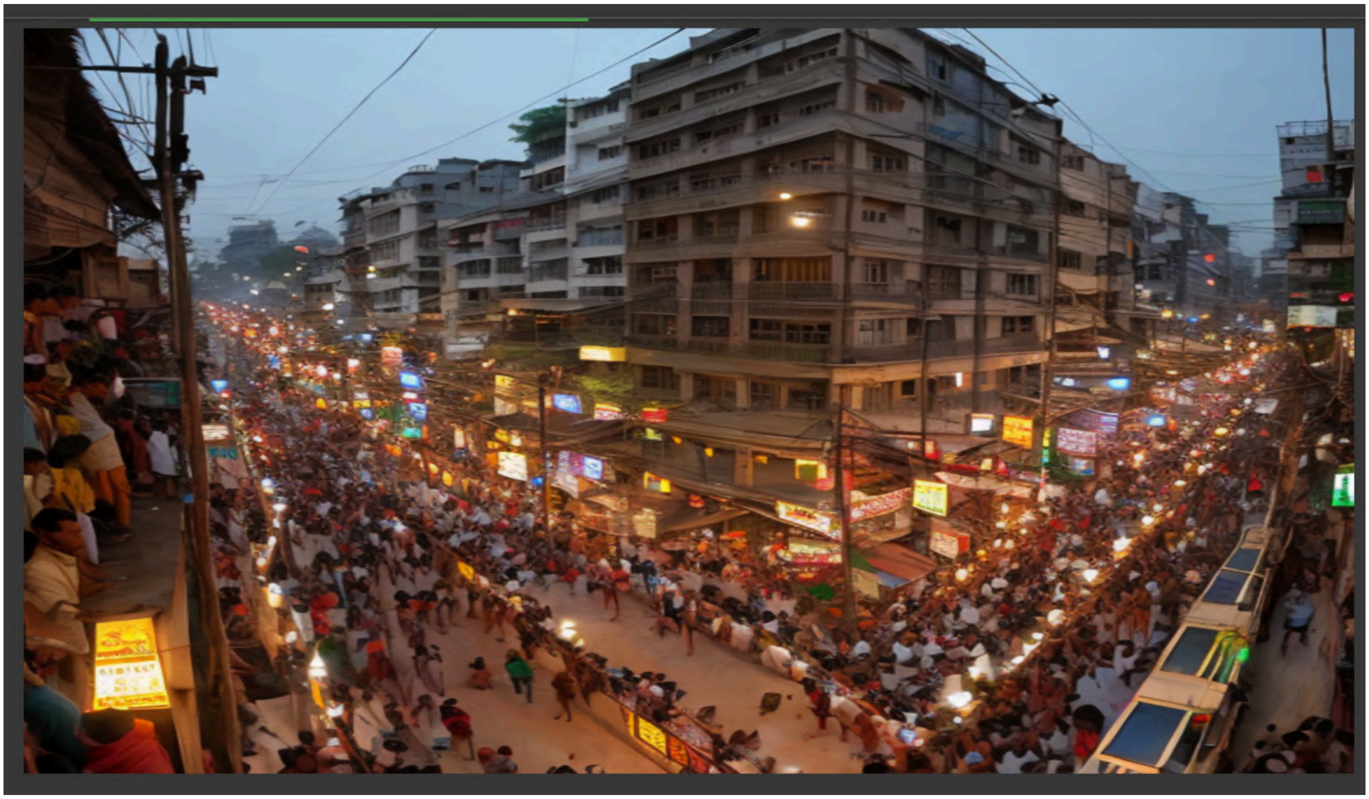
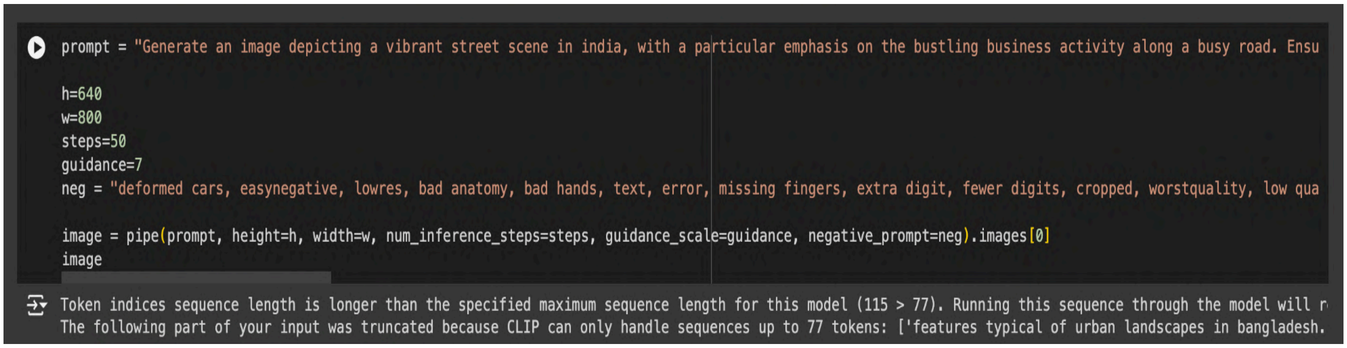


Figure 15:Image generated from figure 14’s prompt.

Using descriptive details and negative prompts effectively can significantly enhance the quality of face generation using diffusion models. By providing clear guidance on desired features and explicitly mentioning attributes to avoid, you can achieve more accurate and visually appealing results. Testing and refining these prompts based on generated outputs is crucial for obtaining the best possible images.



```
prompt = "Generate an image depicting a vibrant street scene in india, with a particular emphasis on the bustling business activity along a busy road. Ensu

h=640
w=800
steps=50
guidance=7
neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low qua

image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg).images[0]
image
```

```
Token indices sequence length is longer than the specified maximum sequence length for this model (115 > 77). Running this sequence through the model will r
The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['features typical of urban landscapes in bangladesh.
```

Figure 16: Example of descriptive prompt and descriptive negative prompt.

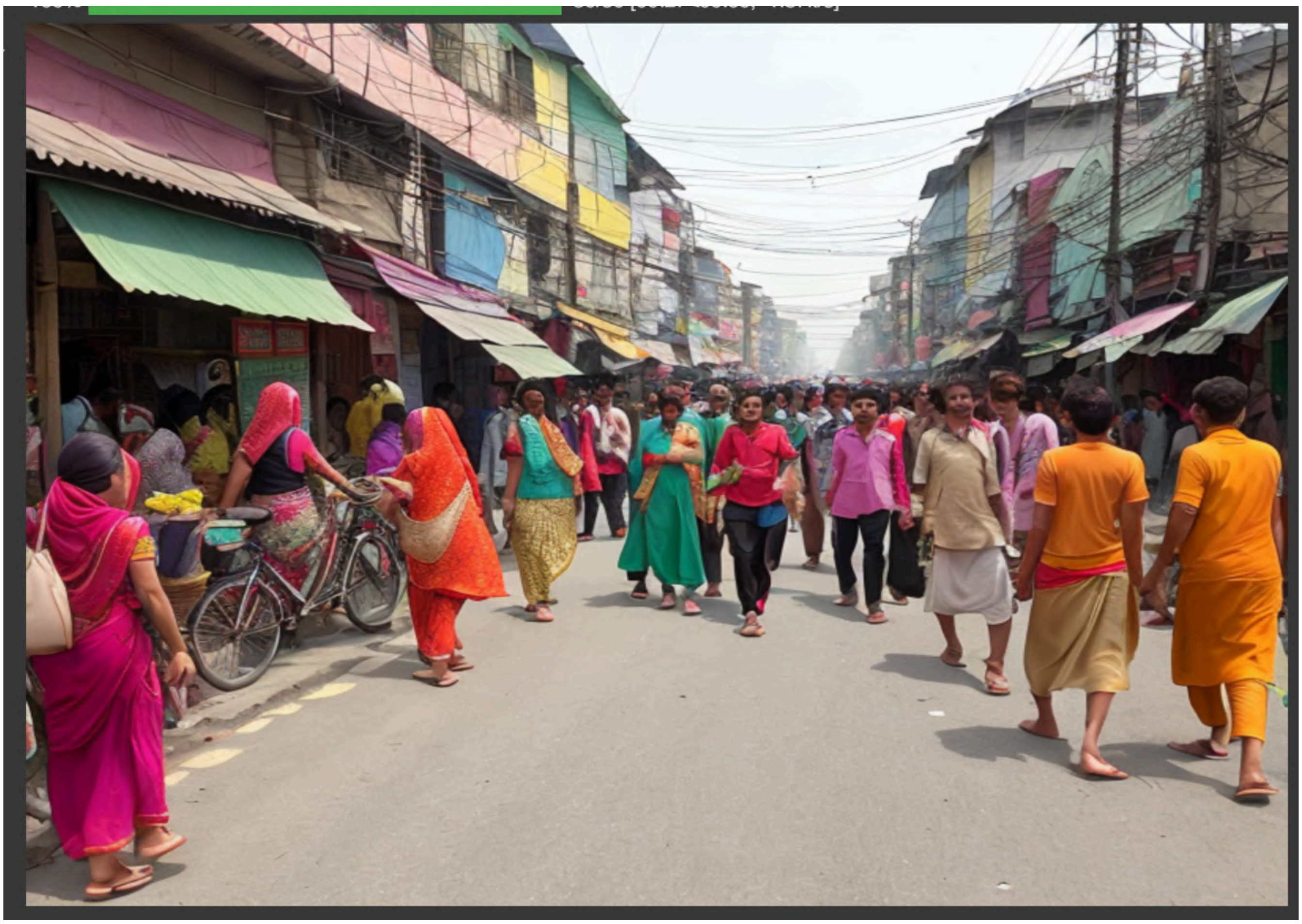
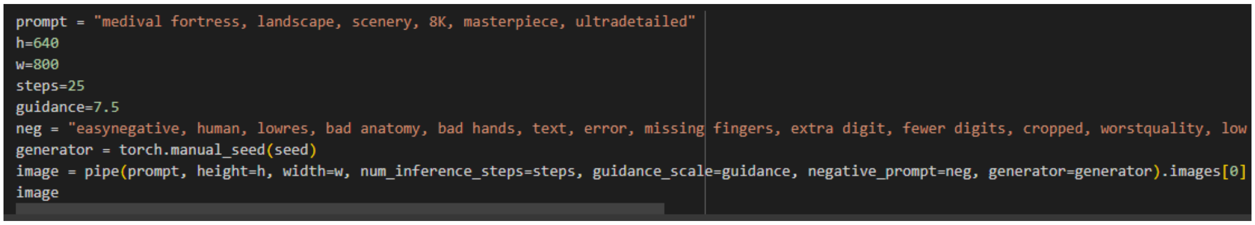


Figure 17:Image generated from figure 16's prompt.

In order to see how we can better our images and also compare them side by side with an outside source we brought in a prompt generator to compare manual prompts with automated prompts.

The results are below:



```
prompt = "medival fortress, landscape, scenery, 8K, masterpiece, ultradetailed"
h=640
w=800
steps=25
guidance=7.5
neg = "easynegative, human, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low
generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 18: medieval fortress user prompt

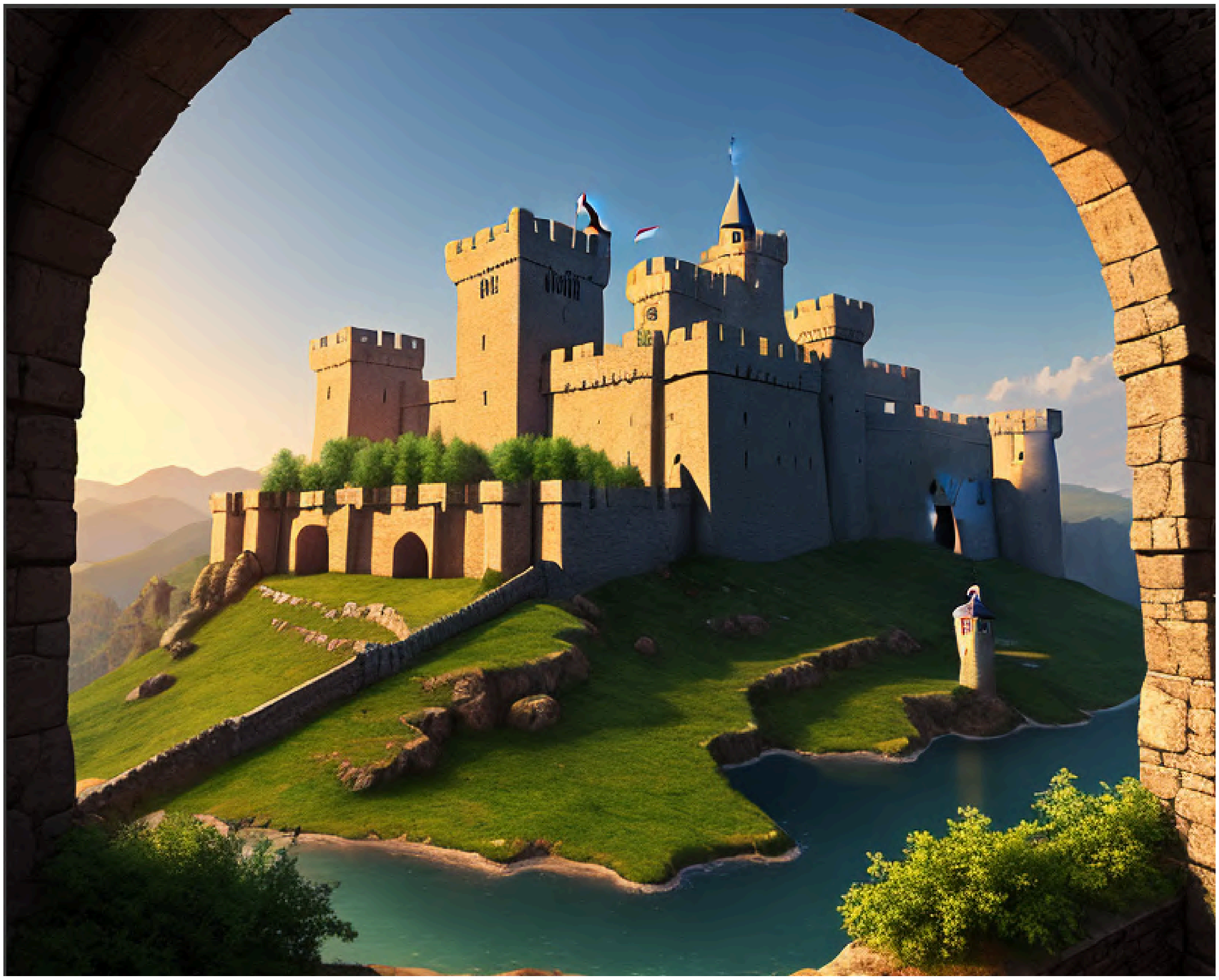
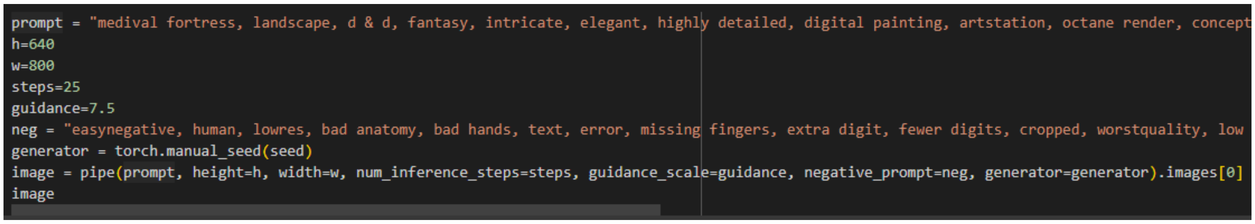


Figure 19: medieval fortress user prompt generated

Figure 19 adheres closely to its prompt, depicting a realistic and highly detailed medieval fortress set in a scenic landscape. The coherent image blends the fortress and landscape seamlessly, achieving an "8k, masterpiece, ultra-detailed" quality. The focus on realism and detailed architecture aligns perfectly with the given prompt.



```
prompt = "medival fortress, landscape, d & d, fantasy, intricate, elegant, highly detailed, digital painting, artstation, octane render, concept
h=640
w=800
steps=25
guidance=7.5
neg = "easynegative, human, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low
generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 20: medieval fortress prompt generator

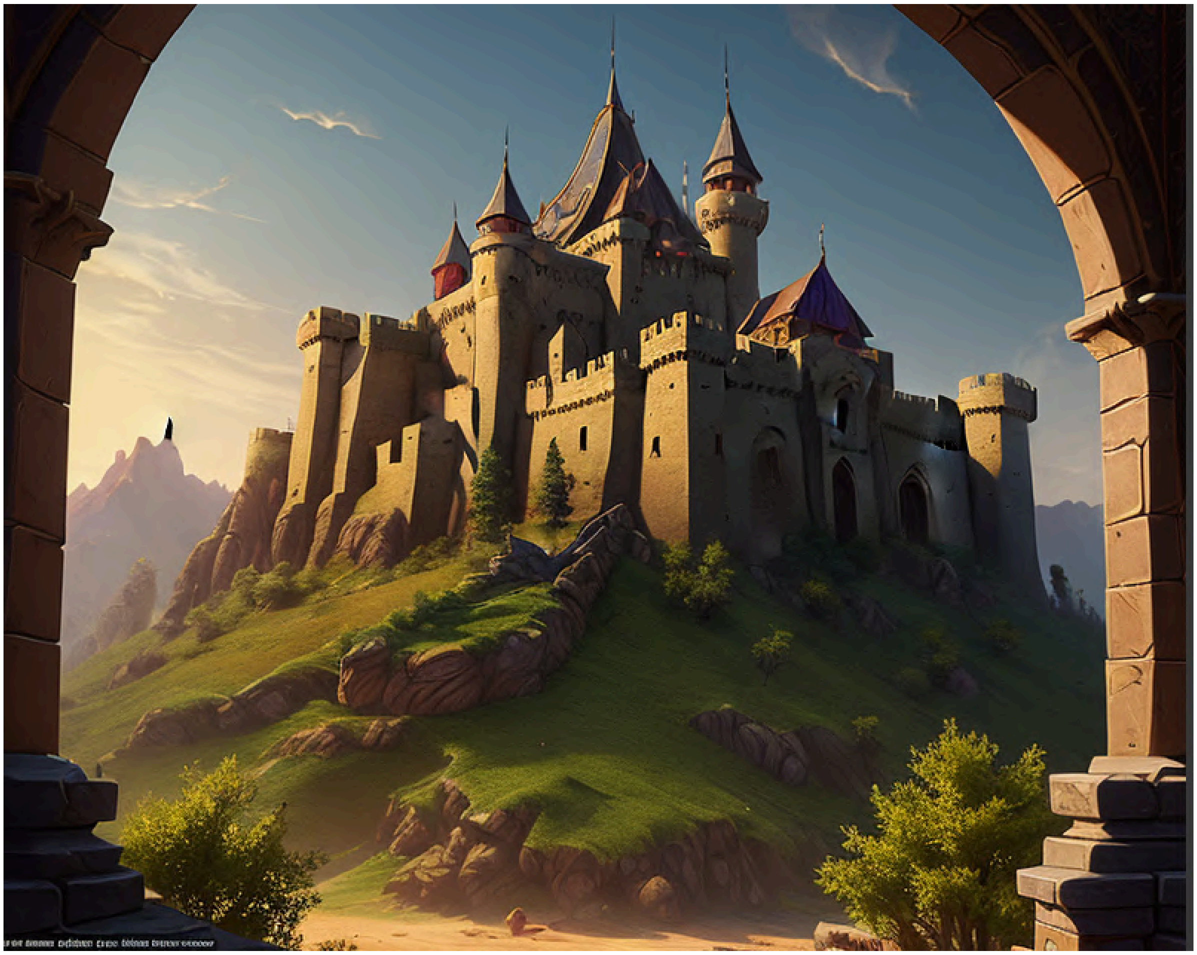
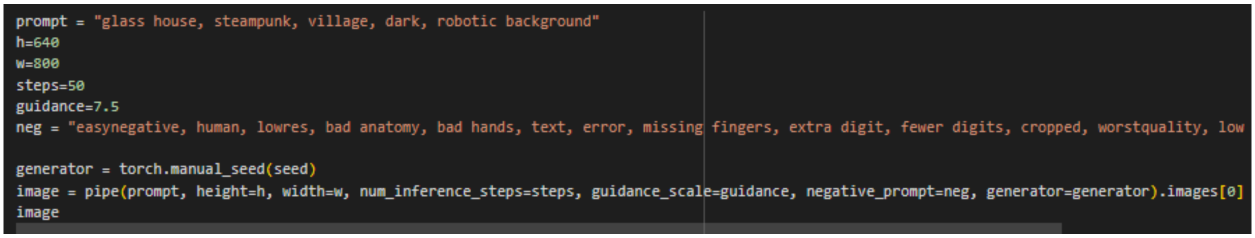


Figure 21: medieval fortress prompt generator generated

Figure 21 excels in creativity by incorporating fantasy elements and intricate architectural details. This image aligns well with its elaborate prompt, capturing the essence of "d & d, fantasy, intricate, elegant" themes. The artistic style reflects terms like "digital painting, artstation, octane render, concept art," blending fantasy with realism. Both images meet their prompts' criteria effectively, with Image 1 focusing on realistic depiction and Image 2 on a creative, fantasy-driven portrayal.



```
prompt = "glass house, steampunk, village, dark, robotic background"
h=640
w=800
steps=50
guidance=7.5
neg = "easynegative, human, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 22: glass house user prompt

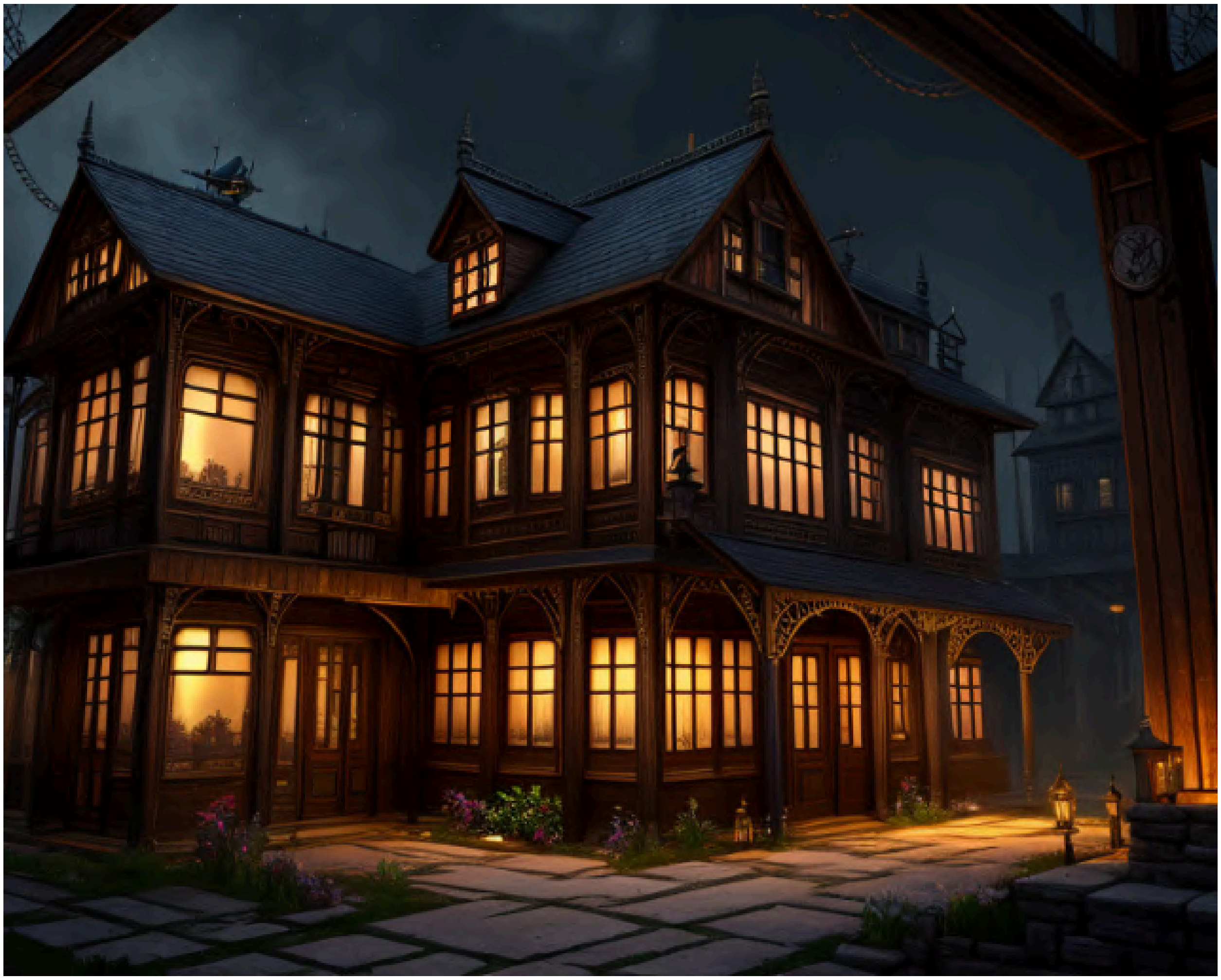
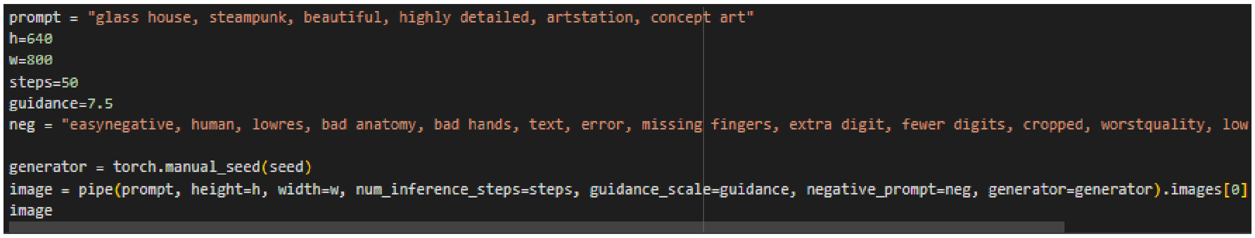


Figure 23: glass house user prompt generated

Figure 23 closely aligns with its prompt: "Glasshouse, steampunk, village, dark, robotic background." The image exhibits a high level of creativity and detail, surpassing what was achievable through the prompt generator's output. It performs exceptionally well in terms of coherence and creativity, effectively capturing the intricate and dark steampunk aesthetic described in the prompt.



```
prompt = "glass house, steampunk, beautiful, highly detailed, artstation, concept art"
h=640
w=800
steps=50
guidance=7.5
neg = "easynegative, human, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 24: glass house prompt generator

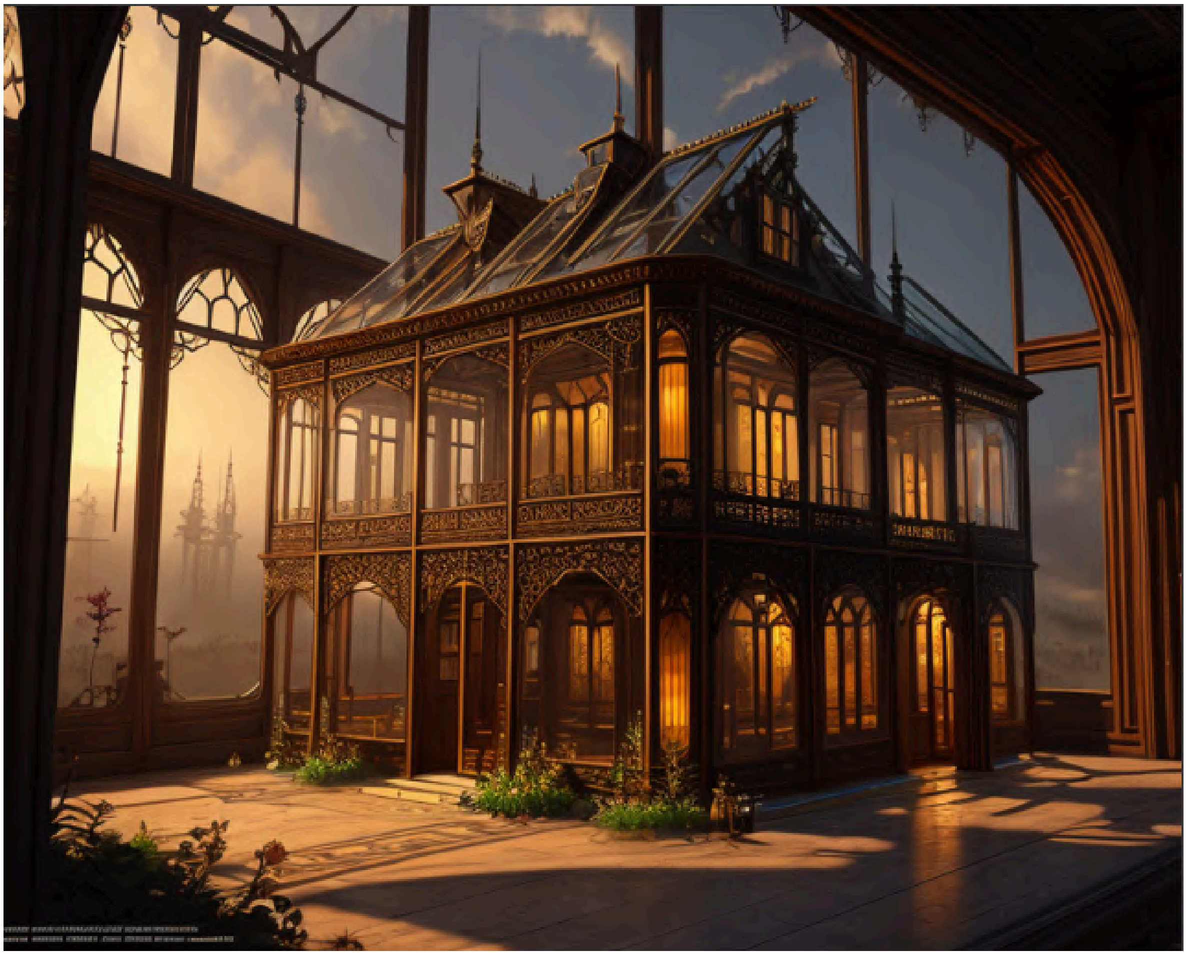
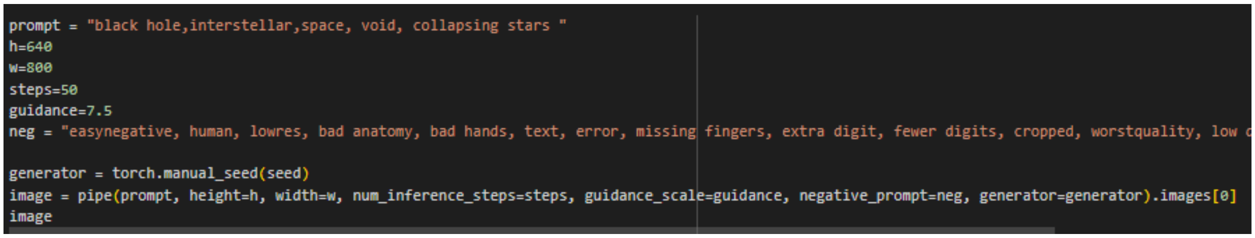


Figure 25: glass house prompt generator generated

Figure 25 excels in creativity, with the house and background blending seamlessly. However, it does have a slight drawback due to the watermark in the corner, which was fine in the user-generated prompt. The prompt, "glass house, steampunk, beautiful, highly detailed, artstation, concept art," strongly correlates with the image, emphasizing the environment and aesthetics over the fine details. This approach allows for greater creative expression, reflecting the flexibility and artistic focus of the generated image.



```
prompt = "black hole,interstellar,space, void, collapsing stars "
h=640
w=800
steps=50
guidance=7.5
neg = "easynegative, human, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 26: Black hole user prompt

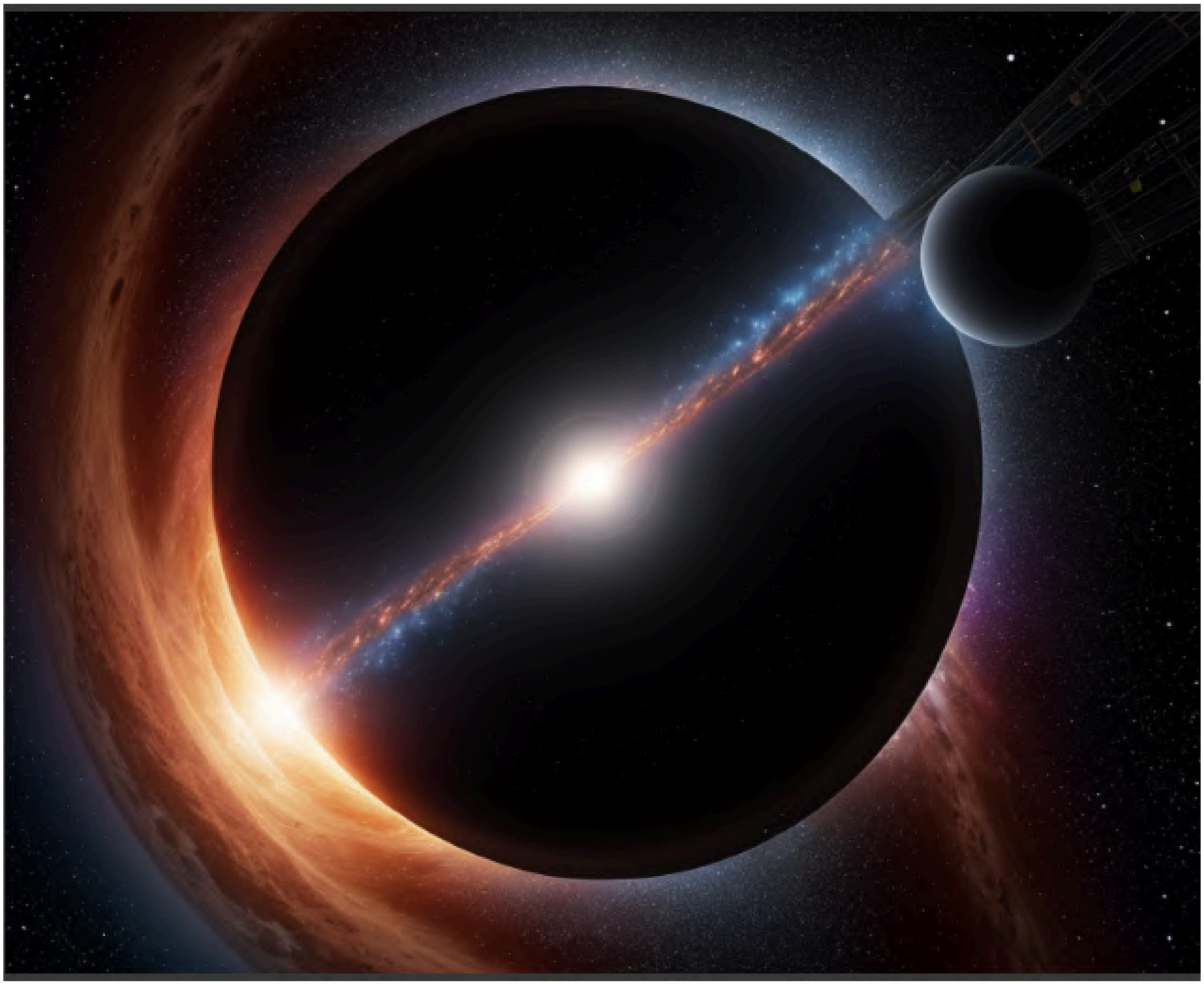
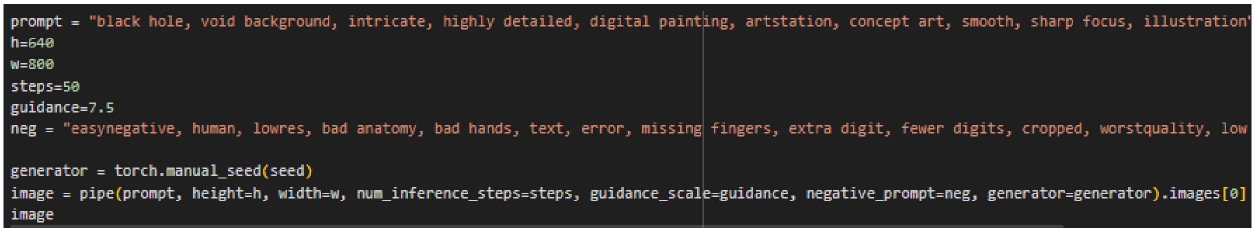


Figure 27: Black hole user prompt generated

Figure 27 presents a combination of space elements, including what appears to be an outline of a planet, diverging from the intended objective of depicting a black hole. The model did not accurately generate the specified prompt: "black hole, interstellar, space, void, collapsing stars." Despite this, the resulting image remains exciting and creative, showcasing a unique interpretation of space and void rather than a traditional black hole.



```
prompt = "black hole, void background, intricate, highly detailed, digital painting, artstation, concept art, smooth, sharp focus, illustration'
h=640
w=800
steps=50
guidance=7.5
neg = "easynegative, human, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality, low

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 28: Black hole prompt generator

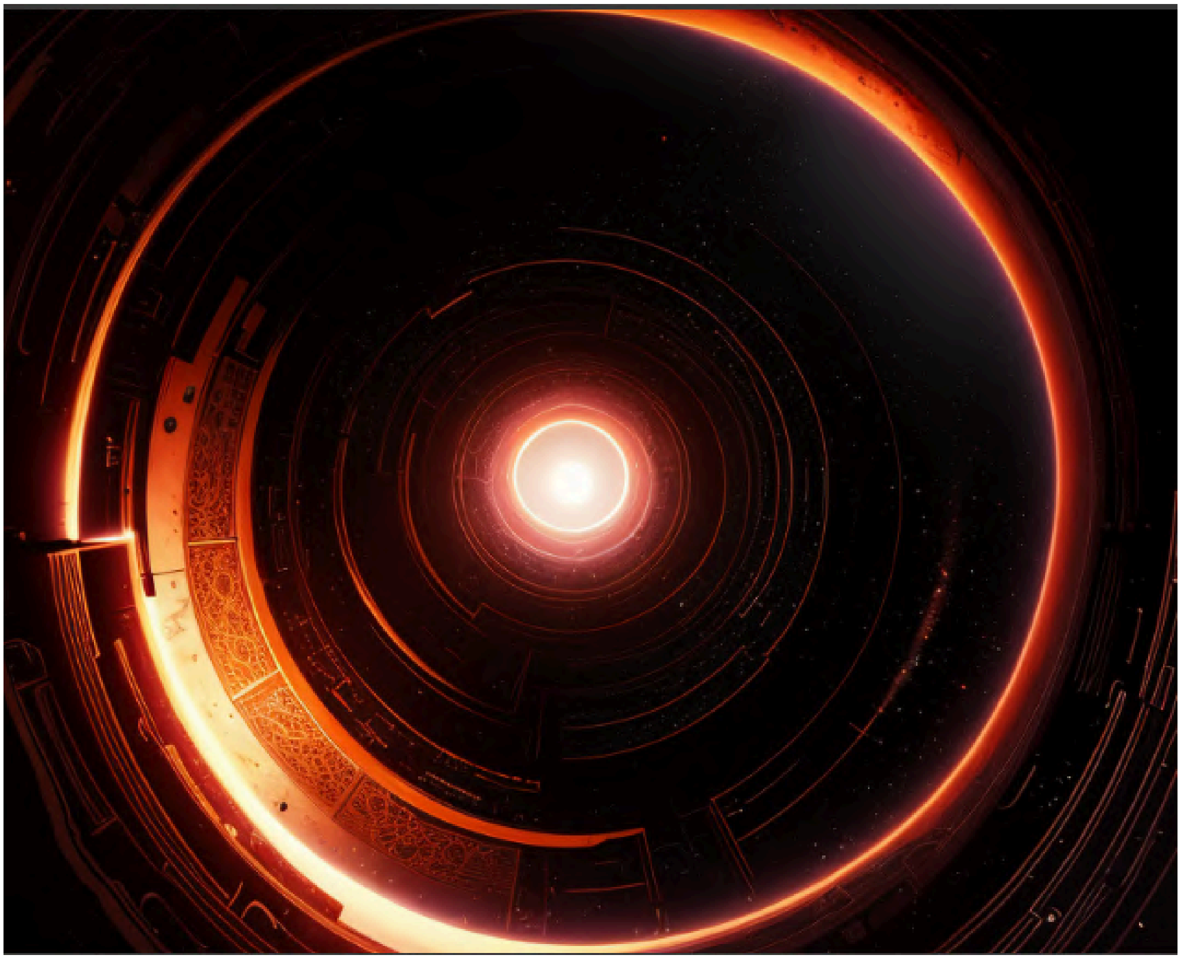
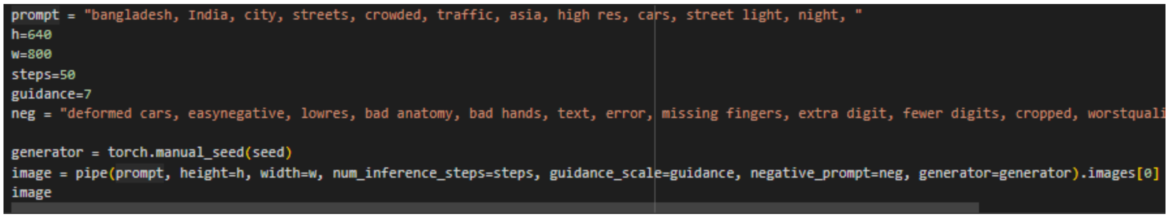


Figure 29: Black hole prompt generator generated

Figure 29 similarly does not meet the prompt's objective to create an image of a black hole from: "black hole, void background, intricate, highly detailed, digital painting, artstation, concept art, smooth, sharp focus, illustration." Instead, the model produced an aesthetically pleasing depiction of what appears to be a robotic black hole. Although it deviates from the prompt's specifics, the image stands out for its creativity and artistic appeal, reflecting a novel and intricate visualization.



```
prompt = "bangladesh, India, city, streets, crowded, traffic, asia, high res, cars, street light, night, "
h=640
w=800
steps=50
guidance=7
neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquali

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 30: Bangladesh crowded street user prompt

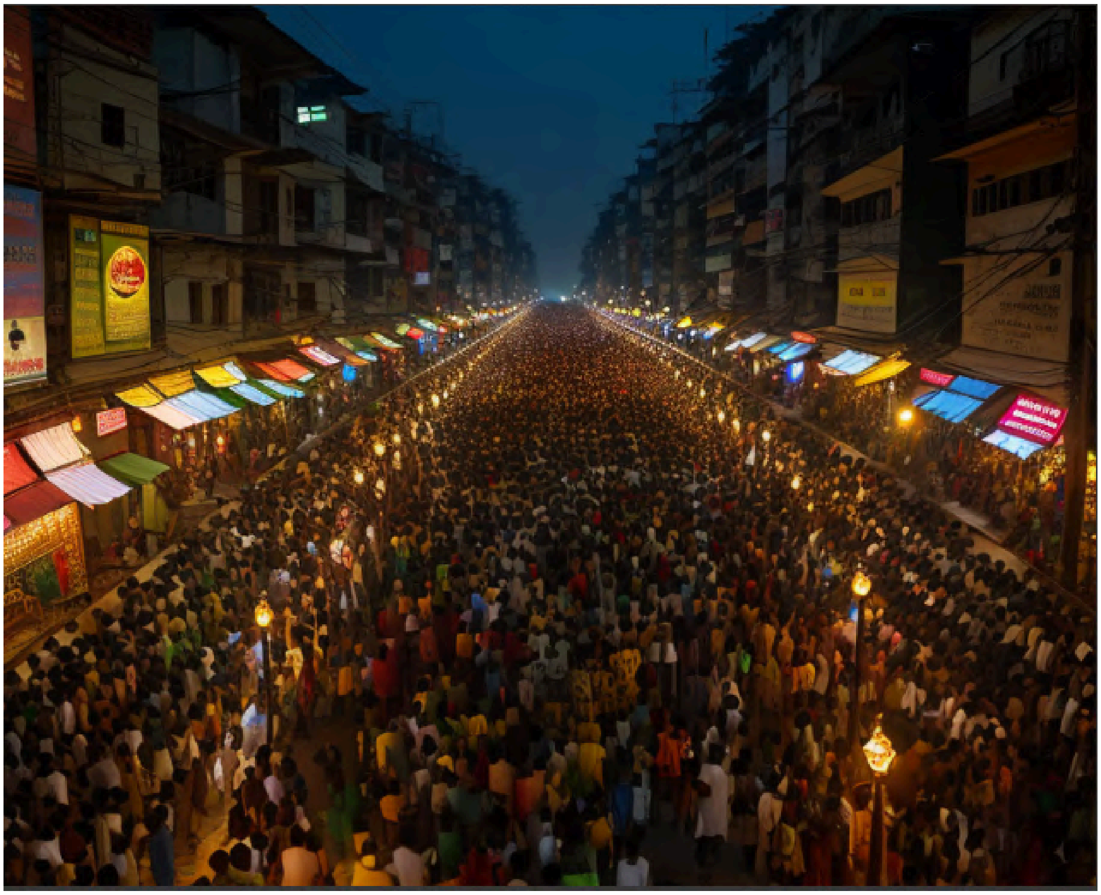
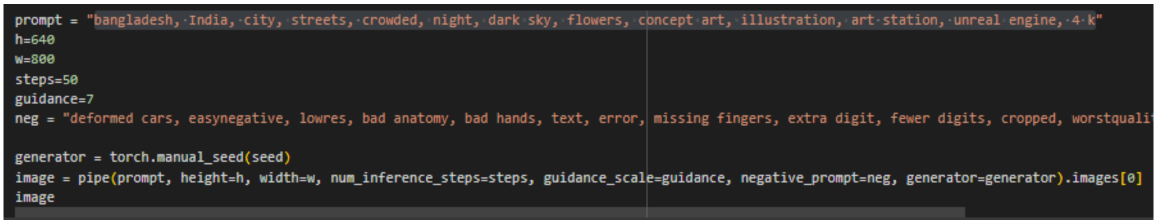


Figure 31: Bangladesh crowded street user prompt generated

Figure 31 provides a detailed depiction of the bustling streets of Bangladesh on a crowded evening. The image adheres closely to the prompt, striking an ideal balance between realism and artistic representation. The crowded streets, dense population, and vibrant city lights are rendered with remarkable accuracy, capturing the dynamic and lively atmosphere of an evening in a Bangladeshi city. The intricate details in the architecture, vehicles, and populace enhance the scene's authenticity, making it a compelling and well-executed visual representation of the prompt.



```
prompt = "bangladesh, India, city, streets, crowded, night, dark sky, flowers, concept art, illustration, art station, unreal engine, 4 k"
h=640
w=800
steps=50
guidance=7
neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquali

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 32: Bangladesh crowded street prompt generator

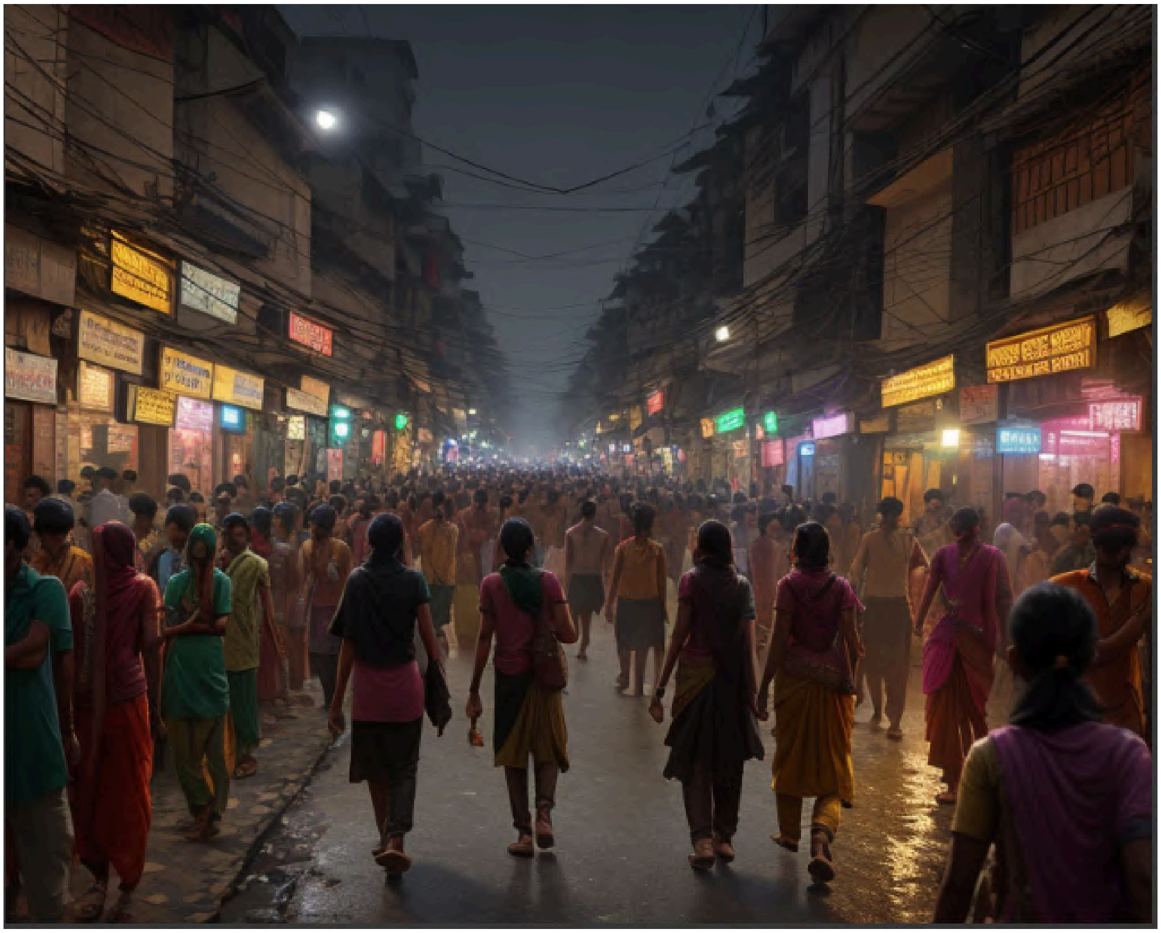
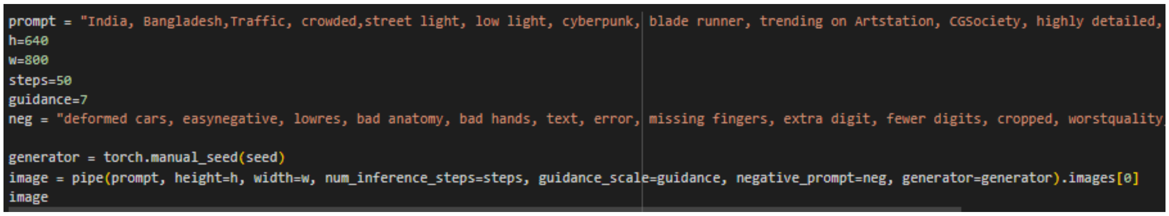


Figure 33: Bangladesh crowded street prompt generator generated

Figure 33 is generated using the prompt generator, resulting in a slightly futuristic depiction of Bangladesh while maintaining the same ethnic demographic. This image addresses its prompt: "Bangladesh, India, city, streets, crowded, night, dark sky, flowers, concept art, illustration, ArtStation, Unreal Engine, 4k." The primary distinction between this image and the user-generated prompt is its heightened level of creativity.



```
prompt = "India, Bangladesh,Traffic, crowded,street light, low light, cyberpunk, blade runner, trending on Artstation, CGSociety, highly detailed,
h=640
w=800
steps=50
guidance=7
neg = "deformed cars, easynegative, lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worstquality

generator = torch.manual_seed(seed)
image = pipe(prompt, height=h, width=w, num_inference_steps=steps, guidance_scale=guidance, negative_prompt=neg, generator=generator).images[0]
image
```

Figure 34: Bangladesh crowded street mixed prompt

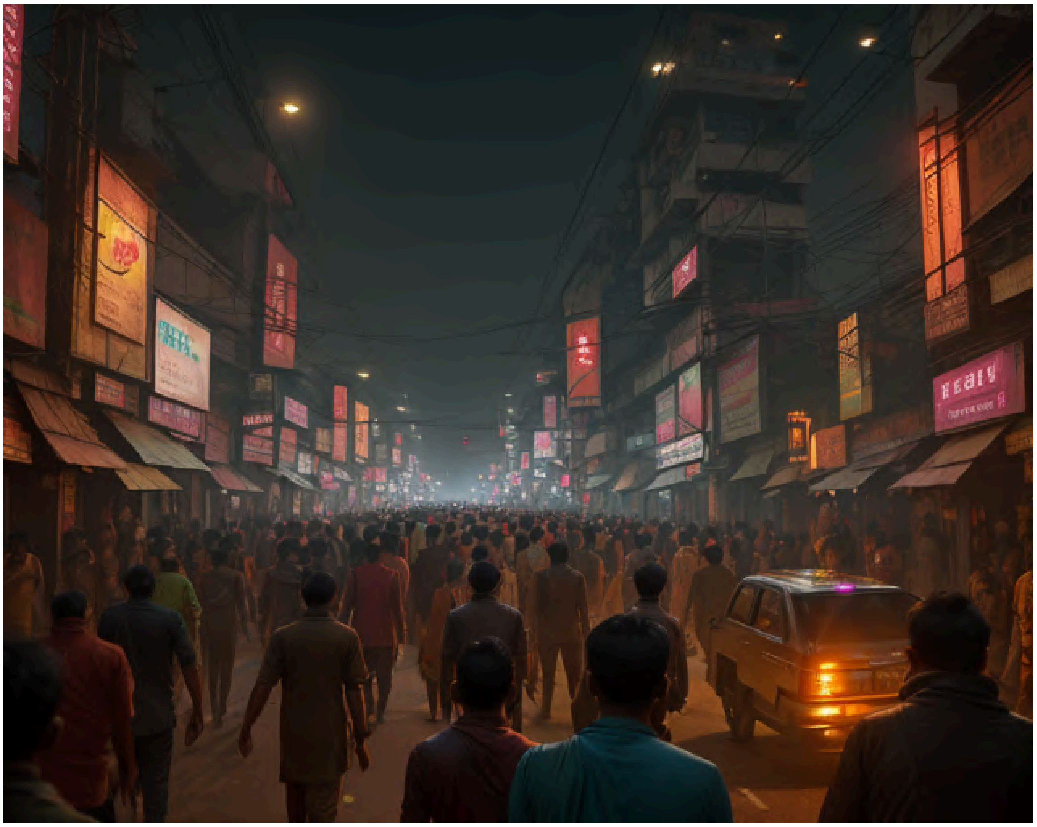


Figure 35: Bangladesh crowded street mixed prompt generated

Figure 35, the final image generated, exemplifies a blend of artistic creativity and a touch of realism. This endeavor presented a significant challenge, as rendering high-density crowds accurately is particularly difficult for diffusion models. The manual user prompting enhanced the realism of the image, and when combined with the prompt generator, this image achieves a harmonious integration of realism, creativity, and artistic vision. The image adheres closely to its prompt: "India, Bangladesh, Traffic, crowded, street light, low light, cyberpunk, blade runner, trending on Artstation, CGSociety, highly detailed, vibrant Bangladesh, Traffic, crowded, street light, low light, cinematic, medium shot, mid-shot, highly detailed, trending on Artstation, Unreal Engine 4k, cinematic wallpaper."

The CLIP score evaluates the compatibility between image-caption pairs, with higher CLIP scores indicating greater compatibility. This metric provides a quantitative assessment of the qualitative concept of "compatibility," which can also be understood as the semantic similarity between an image and its corresponding caption. Research has demonstrated that the CLIP score exhibits a strong correlation with human judgment, underscoring its reliability as a measure of image-caption alignment.

The following table shows the CLIP score of the generated images:

Table II. CLIP score

Name of picture	**user input prompts**	**prompt generator prompt**
Medieval fortress	31.2103	26.1724
Glass house	25.8556	33.8838
Black hole	30.0182	29.0773
Bangladesh city streets	28.7054	28.6212
Bangladesh city streets mixed	40.3734	

The CLIP score analysis reveals that both human-generated prompts and those generated by prompt generators yield comparable results. Combining elements from both prompts to create an image results in a significantly higher CLIP score. Our manual analysis of Figure 35 corroborates this finding, which supports the claim that a hybrid approach enhances image-caption compatibility.

Chapter 5 Impacts of the Project

5.1 Impact of this project on societal, health, safety, legal and cultural issues

There are benefits and drawbacks to the creation of diffusion models for text-to-image generation. In terms of society, it democratizes creativity and encourages invention, but if datasets aren't diverse, there's a risk of oversaturation with information and bias reinforcement. These models can be therapeutic and offer fresh avenues for artistic expression, but they also carry the risk of spreading false information, encouraging cyberbullying, and maintaining unattainable ideals of beauty. Two safety issues are the possibility of deep fakes and the challenge of monitoring massive volumes of created information. Legally, ownership and liability issues for AI-generated works are challenging existing laws, even though new intellectual property frameworks may be developed. These models can help preserve the cultural legacy and produce new art forms, but they also risk causing cultural appropriation and muddying the distinction between human and machine.

5.2 Impact of this project on environment and sustainability

Sustainability and the environment are significantly impacted by the creation and application of text-to-image generation utilizing diffusion models. Carbon footprints are increased as a result of the energy-intensive training and operation of these models, particularly if non-renewable energy sources are used. Further contributing factors to e-waste and resource depletion include the resource demands for large-scale databases and computational infrastructure. But these negative effects can be lessened by developments in optimisation and green computing. Regarding sustainability, these models can contribute to public involvement and the promotion of sustainable behaviors by creating visual information that draws attention to concerns like pollution and climate change. Moreover, they lessen material waste by using virtual prototyping to reduce waste and assist in the design and promotion of environmentally responsible actions and goods.The development and use of diffusion models in text-to-image generation has a substantial impact on sustainability and the environment. The energy-intensive training and operation of these models increases carbon footprints, especially when non-renewable energy sources are used. The need for massive databases and computing infrastructure in terms of resources are additional elements that lead to e-waste and resource depletion. But advances in green computing and efficiency can mitigate these adverse impacts. In terms of sustainability, these models can help raise awareness of issues like pollution and climate change through the creation of visual information that encourages people to engage in sustainable activities. Additionally, they use virtual prototyping to minimize material waste and help with the creation and promotion of ecologically friendly acts.

Chapter 6 Project Planning and Budget

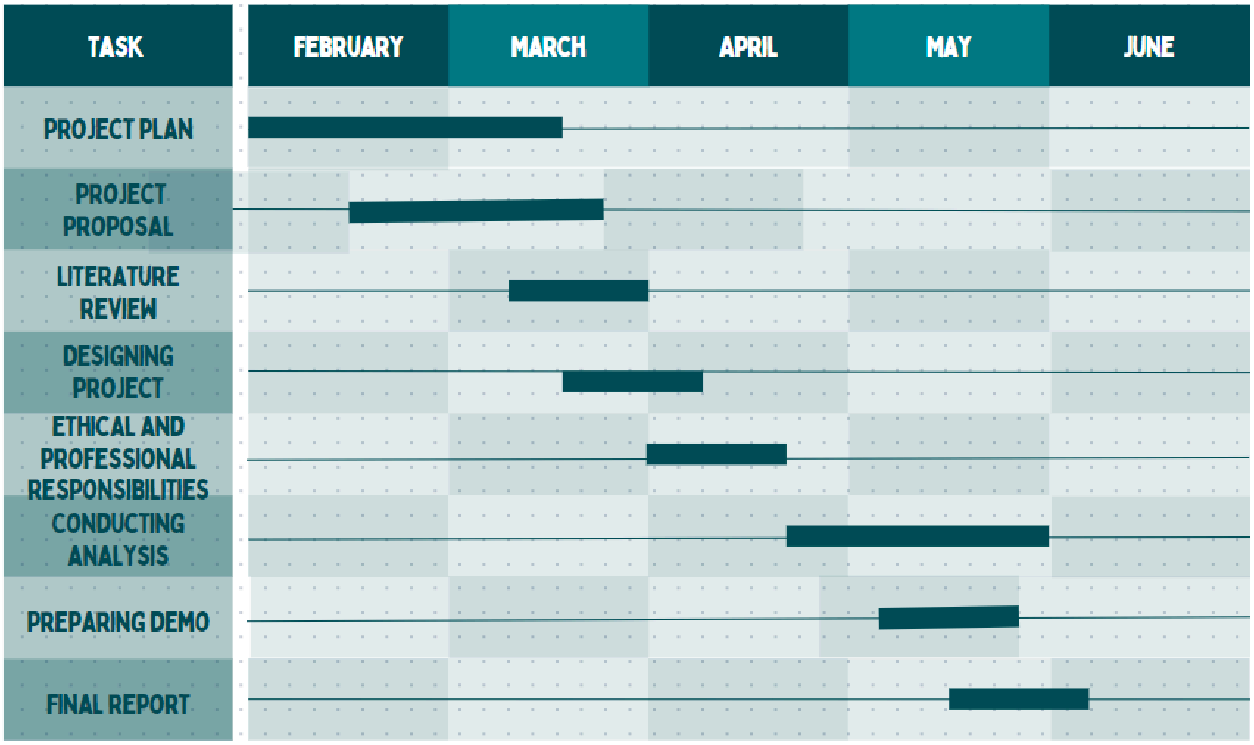


Figure 36: Gantt chart.

The Gantt chart above illustrates the project timeline, providing an approximate overview despite its lack of absolute precision. We deliberated on potential project topics during the initial weeks, finalizing our choice by March. Concurrently, we prepared a project proposal once our focus was determined.

We conducted a comprehensive literature review and conceptualized the project design. After formulating a preliminary design, we drafted reports on the ethical and professional responsibilities associated with our work.

Subsequently, we conducted an in-depth analysis using the selected diffusion models, constituting the project's critical phase. Upon completing the analysis, we prepared a demonstration of our findings and then compiled the final report.

Chapter 7 Complex Engineering Problems and Activities

(Since there is an instruction to discuss this part with our supervisor we decided to leave this part as it is for now)

7.1 Complex Engineering Problems (CEP)

[Describe the Complex Engineering Problems (CEP) attributes related to this project. Discuss with your capstone project supervisor regarding the table. A sample table is given below.]

Table III. Complex Engineering Problem Attributes

Attributes		**Addressing the complex engineering problems (P) in the project**
P1	Depth of knowledge required (K3-K8)	The project requires knowledge of Electrical Circuits, Electronics (K3), Wireless Communication, Embedded System, Sensors and Instrumentations (K4), Designing and Simulation (K5), Engineering & IT (Circuit Design/Smartphone Application) Tools (K6), Involve Environmental Effects (K7), Scientific Research Papers (WK8).
P2	Range of conflicting requirements	In the prototype, the strength of the structure (mass) and capability of weightlifting (# of sensors) is directly related to the capacity of the motors.
P3	Depth of analysis required	No unique way to design. Depth of analysis needed to select a specific solution from many alternatives. (Static/mobile/drone. Various microcontrollers. Various sensors)
P4	Familiarity of issues	Various air quality sensors, Raspberry Pi/Arduino Mega/Nano/Uno/NodeMCU Microcontroller.
P5	Extent of applicable codes	There is no existing code or standard for this project.
P6	Extent of stakeholder involvement	There are several stakeholders needs to be involved including the owner of the device, installing places, Ministry of Environment, etc.
P7	Interdependence	Project involves a number of interdependent sub-systems such as microcontrollers, sensors, wireless communication system, circuit designing tools, mobile apps.

Table I demonstrates a sample complex engineering problem attribute.

7.2 Complex Engineering Activities (CEA)

[Describe the Complex Engineering Activities (CEA) related to this project. Discuss with your capstone project supervisor regarding the table. A sample table is given below.]

Table IV. Complex Engineering Problem Activities

Attributes		**Addressing the complex engineering activities (A) in the project**
A1	Range of resources	This project involves human resource, money, modern tools (simulation software/mobile APP), hardware components, etc.
A2	Level of interactions	Involves interactions between different stakeholders including group members to design the device, installing places, Ministry of Environment to collect data, etc.
A3	Innovation	Employs innovative skills of engineering by introducing technology in a different manner in the environment and IoT sector
A4	Consequences to society / Environment	Impact in our environment since it helps to monitors the air quality data and measure AQI
A5	Familiarity	Needs to be familiar with the various sensors, microcontrollers, wireless communication system, circuit designing tools, mobile apps. UN SDG #04: Quality education; UN SDG #10: Reduce inequality

Chapter 8 Conclusions

8.1 Summary

In this project our team experimented with using the Stable Diffusion Pipeline to generate images based on specific textual prompts. We Utilized different models and generated images depicting various themes, such as medieval fortresses, steampunk villages, black holes, and urban traffic scenes. The results showcased the model's ability to produce high-quality, detailed images that closely matched the given prompts. However, it was noted that long, uninterrupted sentences in prompts often resulted in lower-quality images compared to shorter, more detailed descriptions which are comma-separated . The experiments highlighted the importance of prompt engineering and the effective use of model parameters to achieve the desired visual output.

8.2 Limitations

1. **Prompt Sensitivity**: The quality of generated images was susceptible to the phrasing and structure of the prompts. Long, complex sentences tended to produce less accurate results.
2. **Negative Prompts**: While negative prompts helped refine images by excluding unwanted elements, their effectiveness varied and sometimes led to the exclusion of desirable details.
3. **Model Constraints**: Using pre-trained models limited our ability to customize the output fully. Specific artistic styles or detailed requirements sometimes fell below expectations.
4. **Resource Intensive**: The image generation process was computationally intensive, requiring high-performance hardware, such as GPUs, which may not be accessible to all users.

8.3 Future Improvement

1. **Prompt Refinement**: Develop a more systematic approach to crafting prompts, possibly incorporating machine learning techniques to optimize prompt structures for better image quality.
2. **Object Addition**: Explore methods for adding new objects to already generated images using subsequent prompts, enhancing and enriching the initial outputs without starting from scratch.
3. **Custom Stable Diffusion Model**: Create a custom stable diffusion model trained on a specific dataset tailored to our needs, allowing for greater control over the stylistic and thematic elements of the generated images.
4. **Interactive Generation**: Implement an interactive interface that allows users to iteratively refine images by adjusting prompts and parameters in real time, providing greater creative control.
5. **Efficiency Improvements**: Optimize the computational efficiency of the pipeline, potentially through model pruning or utilizing more efficient algorithms, to make the technology more accessible.
6. **Enhanced Negative Prompt Handling**: Develop advanced techniques for more effective use of negative prompts, ensuring that unwanted elements are excluded without compromising desired details.
7. By addressing these limitations and implementing the proposed improvements, we aim to enhance the capabilities and accessibility of diffusion models for image generation, making them more versatile and user-friendly for a broader range of applications.

References

1. Witteveen, S., & Andrews, M. (2022, November 21). *Investigating Prompt Engineering in Diffusion Models*. arXiv.org. https://arxiv.org/abs/2211.15462
2. Hao, Y., Chi, Z., Dong, L., & Wei, F. (2022, December 19). *Optimizing Prompts for Text-to-Image Generation*. arXiv.org. https://arxiv.org/abs/2212.09611
3. Feng, Y., Wang, X., Wong, K. K., Wang, S., Lu, Y., Zhu, M., Wang, B., & Chen, W. (2023). PromptMagician: Interactive Prompt Engineering for Text-to-Image Creation. *IEEE Transactions on Visualization and Computer Graphics*, 1–11. https://doi.org/10.1109/tvcg.2023.3327168
4. Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, 1–14. https://doi.org/10.1080/0144929x.2023.2286532
5. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016, May 17). *Generative Adversarial Text to Image Synthesis*. arXiv.org. https://arxiv.org/abs/1605.05396
6. Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2017, November 28). *AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks*. arXiv.org. https://arxiv.org/abs/1711.10485