Empowering Local Image Generation: Harnessing Stable Diffusion for Machine Learning and AI

Ahmed Imran KABIR, Limon MAHOMUD, Abdullah Al FAHAD, Ridwan AHMED
School of Business and Economics, United International University, Bangladesh
ahmedimran@bus.uiu.ac.bd, lmahomud192070@bba.uiu.ac.bd, aaf5623@gmail.com,
ridofficial2020@gmail.com

This paper examines the ability to use Stable Diffusion's diffusion models to get state-of-the-art synthesis results on image data and other types of data. Also, a guiding interface can be used to control the process of making images by converting text to images and image to image. But because these models usually work directly in pixel space, optimizing strong DMs often needs more GPU VRAM to run. Using Stable Diffusion and diffusion models on local hardware like this lets more information and depth be added while generating images, which greatly improves the quality detail of the image. By combining diffusion models to model architecture, I have made diffusion models into powerful and flexible producers for general conditioning inputs, such as when using XL-XDXL 1.0 and LoRA models. Overall, the paper highlights how a normal person can run their own Midjourney like AI image generation with the help of machine learning and generative AI.

Keywords: *Stable Diffusion, Machine Learning, Image generation, generative AI, VRAM, GPU, Diffusion Models, Prompt*
DOI: 10.24818/issn14531305/28.1.2024.03

1 Introduction

The operating concept of Stable Diffusion occurs by using the capabilities of a deep learning model to generate images based on textual descriptions. The primary mechanism employed is a diffusion process, wherein an image undergoes a transformation from a state of random noise to a state of coherence through a sequence of processing stages. The model undergoes training in order to effectively direct and control each stage of the process, ultimately overseeing the complete progression from initiation to conclusion, in accordance with the given textual prompt. The fundamental concept underlying Stable Diffusion involves the transformation of randomly generated elements, specifically noise, into a visual representation. The model initiates the procedure by introducing a substantial amount of random noise, akin to the colorized rendition of the white noise observed on a screen lacking signal. This noise is thereafter iteratively corrected, under the impact of the provided text prompt, resulting in the emergence of a recognizable image. The process of refining is conducted in a systematic manner, gradually reducing the presence of noise and enhancing the level of detail until a result of higher resolution is obtained. The starting point of the diffusion process is important in determining the primary elements that constitute the image, while further modifications to keywords have a limited impact on smaller sections. This underscores the importance of meticulous consideration of keyword weighting and timing in order to get the ideal result.

2 Background

The background of the project is understanding deep machine learning text-to-image or image-to-image generation using an open-source AI model and generating image local hardware for deployment ready for personal or commercial use. The main project focus is on running on a local personal machine to generate images with the help of machine learning and the AI tool Stable Diffusion. By using the hardware power of local use cases, greatly increases as more control is there during the generative process. The whole process required a high-end GPU for the generative machine deep learning process.

The scope of the project is to generate a Midjourney level of image generation process on own personal hardware for learning how machine learning and AI in general work during their generative processing. In the future, I believe it could be deployed as a service similar to Midjourney.

The objectives of the project are:

- stable Diffusion and its use cases
- setup Stable Diffusion in a private secure environment (Personal Computer)
- learn and implement Text to Image and Image to Image.
- use Diffusion models to generate images.

3 Literature Review

In the realm of empowering local image generation, the integration of stable diffusion techniques holds significant promise for advancing machine learning and artificial intelligence (AI) capabilities. Ahamed et al. [1] underscore the importance of cultivating proactive attitudes towards cybersecurity, emphasizing the need for robust cybersecurity measures in the era of the 4.0 Industrial Revolution. This emphasis on a proactive mindset aligns with the proactive approach required in developing stable diffusion methods, which are essential for generating high-quality images at the local level. Furthermore, Karim, Vyas, and Kabir [2] shed light on the legal challenges in regulating emerging technologies such as digital twins, highlighting the necessity for legal frameworks to adapt to technological advancements. Similarly, in the realm of local image generation, legal considerations regarding data usage and intellectual property rights are crucial for ensuring ethical and legal compliance. Additionally, Polas et al. [3]explore the behaviors of rural entrepreneurs towards green innovation, indicating the importance of community engagement and empowerment in driving sustainable development initiatives. This community-centric approach resonates with the concept of empowering local image generation, as it emphasizes the involvement of local communities in shaping and utilizing AI technologies for their benefit. By leveraging insights from Mia, Kabir, and Abdullah [4] regarding the psychological impact of global crises on digital behaviors, such as increased reliance on social media, AI researchers can better understand the evolving needs and preferences of local communities when developing image generation solutions.

In sum, by integrating insights from cybersecurity awareness, legal challenges, rural entrepreneurship, and psychological impacts, AI researchers can harness stable diffusion techniques to empower local image generation effectively, thereby advancing the capabilities of machine learning and AI in addressing local needs and challenges.

In addition to the aforementioned contributions, Kabir et al. [5] delve into the development of a network design tailored for smart airports utilizing Cisco Packet Tracer. This study underscores the practical application of technological solutions in enhancing infrastructure for efficient operations, a facet crucial for local image generation initiatives. Moreover, Polas et al. [6] explore blockchain technology as a transformative force for green innovation, particularly in fostering green entrepreneurship and economic sustainability. The integration of blockchain in local image generation endeavors could potentially enhance transparency and security in image data management, aligning with the overarching goal of sustainable development. Furthermore, Kabir, Akter, and Mitra [7] investigate methods for detecting students' engagement in online learning environments during the COVID-19 pandemic, highlighting the relevance of remote education strategies in shaping future workforce capabilities, including those involved in image generation technologies. Similarly, Kabir et al. [8] and Kabir, Ahmed, and Karim [9] contribute insights into the application of AI and data analytics techniques, such as face-mask detection and sentiment analysis, which could be leveraged in refining image generation algorithms and enhancing user experiences. Additionally, the study by Ahmed Imran Kabir et al. [10] emphasizes the power of social media analytics, aligning with the broader theme of understanding digital behaviors and preferences, which could inform strategies for disseminating locally generated images effectively. By integrating findings from these diverse studies, researchers can adopt a holistic approach towards empowering local image generation, leveraging technological innovations, legal frameworks, educational strategies, and community engagement initiatives to address local needs and challenges effectively.

Stable Diffusion is a text-to-image and image to image model that use deep learning techniques in the year 2022. It is built upon the foundations of diffusion methods. The main application of this technology is to produce intricate visual representations based on textual descriptions. However, it can also be utilized for other purposes like as filling in missing parts of an image, creating new content beyond the original image boundaries, and generating image transformations based on a given text instruction [11]. The computational resources necessary for the project were generously provided by Stability AI, while the training data utilized in the system's training process was sourced from various non-profit organizations.

Stable Diffusion latent diffusion model is a deep generative artificial neural network. The code and model weights of the system have been made publicly available. It can run with personal hardware that is equipped with a NVIDIA GPU, with a minimum of 8 GB VRAM. This development represents a significant deviation from prior proprietary text-to-image models like DALL-E and Midjourney, which were exclusively accessible through cloud-based services [12].

Image Generation

The initial efforts to employ Artificial Intelligence for generating comprehensible and innovative content based on human cues may be historically attributed to the 1950s. During this period, a group of researchers at the Artificial Intelligence Laboratory of Massachusetts Institute of Technology developed a program known as ELIZA. ELIZA demonstrated the capability to produce rudimentary replies to textual input by employing pattern matching and techniques of natural language processing. Although not traditionally classified as art, ELIZA served as an early instance of Text-to-Text technology, wherein software was capable of producing novel textual output with the purpose of being comprehended by human users. A significant work of art generated by artificial intelligence emerged in the 1970s with the introduction of AARON, a program created by artist Harold Cohen. AARON was an advanced computer program with the ability to generate intricate drawings and paintings. AARON employed a prescribed set of rules and limitations to generate its artistic creations and illustrated the capacity to enhance its performance by self-learning from its previous outputs.

During the 1980s and 1990s, there was a notable progression in AI technology, which therefore led to the emergence of increasingly intricate and refined AI-generated art. An example of this is the work conducted by Karl Sims, who employed evolutionary algorithms to produce distinctive three-dimensional visuals and animations. In recent times, the emergence of deep learning has resulted in the production of increasingly lifelike outcomes. Consequently, there has been a growing interest in AI-generated art from both the art community and the general populace. In 2015, A research team at Google employed deep learning methodologies to train a neural network using a dataset including more than 10,000 paintings. The primary objective of this endeavor was to generate novel pieces of artwork based on input images. The program that emerged, referred to as DeepDream, demonstrated the capability to generate aesthetically captivating and strange visuals based on input photos (Image-to-Image). An additional noteworthy instance pertains to the artistic endeavors of a Parisian art collective known as "Obvious," wherein they produced a portrait through the use of software. This particular artwork achieved a substantial sale price of more than $432,000 during a Christie's auction in the year 2018 [13].

Stable Diffusion & OpenAI DALL-E

The year 2020 witnessed a significant advancement in Text-to-Text capabilities through the introduction of the third iteration of the Generative Pretrained Transformer (GPT-3) by OpenAI, a privately-owned research organization. The GPT-3 model represents a significant progression in the realm of Text-to-Text models, demonstrating enhanced versatility and the ability to produce highly coherent text in response to a wide range of prompts in natural language. The achievement was facilitated by the substantial scale of the model, with 175 billion parameters, surpassing the magnitude of any previously developed comparable model. The extensive range of parameters enabled GPT-3 to effectively understand and perform language-related activities that were not specifically included in its training, thereby marking the advent of the Large Language Models era. These models experience the capacity to produce text of superior quality that closely resembles human language. This attribute renders them suitable for a wide range of applications, such as machine translation, text summarization, and creative writing. The success of GPT-3 has resulted in the emergence of CLIP, a further pioneering model developed by OpenAI, with the specific objective of establishing connections between textual and visual content. The CLIP (Contrastive Language–Image Pretraining) model is a versatile image-text model that has been trained on a vast dataset of 400 million text-image pairs sourced from the internet. This extensive training enables CLIP to effectively classify images based on any label provided by the user. Additionally, it has the capability to produce textual descriptions that precisely depict any given input image, a process commonly referred to as Image-to-Text. OpenAI has launched DALL-E as a result of these advancements, enabling the generation of visually compelling images based on textual descriptions (Text-to-Image). Although DALL-E continues to be a proprietary and closed-source software, the code for CLIP has been made available as open-source. The development and training of Stable Diffusion, an open-source Text-to-Image model with performance comparable to DALL-E, was made possible by artificial intelligence business Stability AI. The software known as Stable Diffusion was made available to the public under a permissive license that grants permission for both commercial and non-commercial utilization.

Despite its significance as a notable technological advancement, both CLIP and the Text-to-Image systems that rely on it give rise to significant ethical and societal considerations. The utilization of large-scale, unselective internet data in the training of CLIP has led to its inclination to perpetuate biased and unjust preconceptions that exist throughout culture and society. Furthermore, legal professionals have raised concerns regarding the potential infringement of protected works by CLIP. These technologies also possess the capacity to be utilized for malicious intentions, such as fabricating false news or disseminating inaccurate information.

Diffusion models have had a notable increase in adoption in recent years due to their inherent generative aspect. This trend can be attributed to their numerous advantageous characteristics. Several influential articles published in the 2020s have demonstrated the outstanding capabilities of Diffusion models, incorporating their ability to compete with Generative Adversarial Networks (GANs) in the field of image generation. Recently, Diffusion Models have been discovered being utilized in DALL-E 2, a model developed by Open AI for generating images.

4 Research Methodology

This research focuses on the generative modeling of latent representations. By using skilled perceptual compression models, which are made up of an encoder (E) and a decoder (D), it was able to access a latent space that was effective and had fewer dimensions. In this latent space, intricate details that are invisible to human senses, particularly those pertaining to high frequencies, are condensed and removed (Figure 1). In contrast to the high-dimensional pixel space, this alternative space is better suited for likelihood-based generative models. This is because these models can now prioritize the significant semantic aspects of the data and train in a lower-dimensional space, which is computationally more efficient.

In contrast to autoregressive and attention-based transformer models inside a significantly compressed and discontinuous latent space, this model may exploit the image-specific inductive biases it possesses.

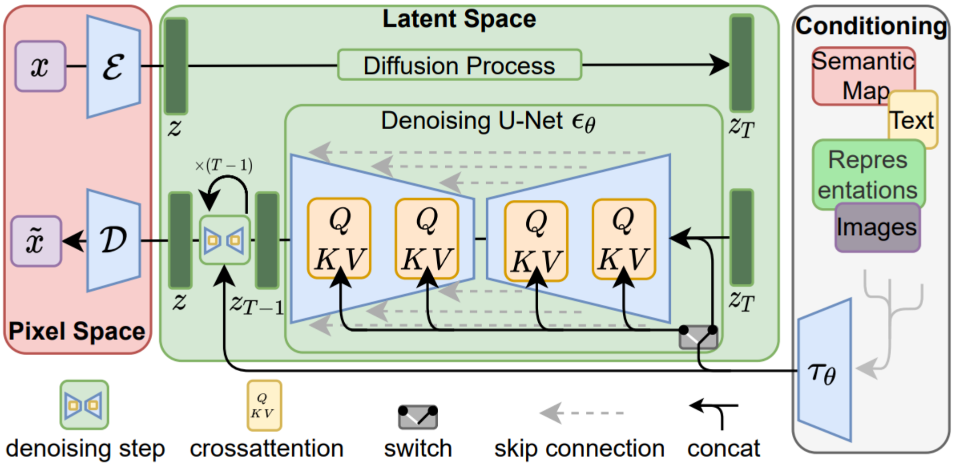


Fig. 1. Latent space [12]

This encompasses the capability to construct the foundational UNet architecture primarily utilizing 2D convolutional layers. Additionally, it involves emphasizing the most perceptually significant components of the objective through the utilization of the reweighted bound. The neural architecture, denoted in the model, is implemented as a time conditional UNet. Given that the forward process is predetermined, the value of zt may be effectively derived from E during the training phase. Additionally, samples from the distribution p(z) can be decoded into image space by making a single pass-through D.

Diffusion Model Image Generation

Every image generated through the use of Stable Diffusion displays a unique and distinctive characteristic. However, the differences observed in every image can vary considerably, ranging from highly dramatic to barely noticeable.

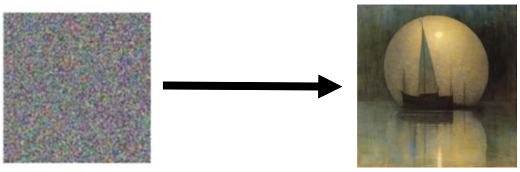


Fig. 2. From Diffusion Model image generation from noise.

The generations viewed from above look distinctive in some areas of the rendering but are mostly identical (Figure 2). If one continues to regard them as distinct, then each Stable Diffusion AI-generated image will likely be found unique. The probability of producing the same AI-generated image in Stable Diffusion by a different individual is significantly minimal. The framework behind this characteristic can be linked to the presence of variables. In order to generate an AI image, there are several variables to select from for processing. A Diffusion Model latent variable model utilizes a fixed Markov chain to map the latent space. The purpose of this chain is to introduce noise to the data in a step-by-step manner, with the aim of approximating the posterior distribution q(x1:T|x0). Here, x1,…,xT represent the latent variables, which possess the same dimensionality as x0. The provided diagram from Figure 3 illustrates the representation of a Markov chain specifically designed for image data [14]. The image undergoes an asymptotic transformation, resulting in the manifestation of pure Gaussian noise. The objective of training a diffusion model is to acquire knowledge of the inverse procedure (Figure 4) [15].

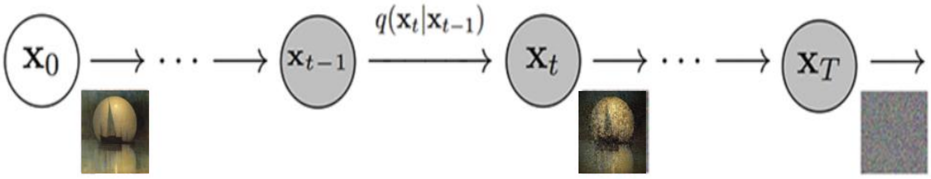


Fig. 3. From image generation to diffusion noise.

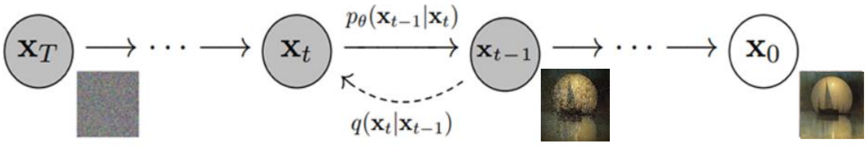

$\mathbf{x}_T$
$\mathbf{x}_t$
$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$
$\mathbf{x}_{t-1}$
$q(\mathbf{x}_t|\mathbf{x}_{t-1})$
$\mathbf{x}_0$

Fig. 4: From image generation to diffusion noise.

Model Checkpoint

The Stable Diffusion checkpoint merger is a recently developed feature implemented in Stable Diffusion, which enables users to setup many combinations utilizing different modeling techniques with the objective to enhance the quality of their AI-generated images. This feature enables the merging of up to three models, including models that have been trained by the user. After combining our ideal scenario checkpoints, the output of the merger will be generated and stored in the allocated checkpoint directory.

Similar to the Stable Diffusion prompt matrix, the Stable Diffusion checkpoint merger facilitates the production of AI-generated images that possess a high degree of visual realism, catering to the specific artistic requirements of individuals. This is achieved through the combination of many checkpoints, enabling the creation of desired images with precision and accuracy. While Stable Diffusion models have been extensively trained on various aspects of image production, it is important to acknowledge that they possess certain limitations. This is the rationale for the necessity of integrating several models, including individuals who have trained models, for the purpose to produce the intended visuals.

One viable approach for combining two Stable Diffusion checkpoints and doing individual model training is to utilize one of the existing third-party interfaces. Stability AI does not officially approve or advocate for any of these interfaces, or the custom models derived from them. Therefore, the process of determining their efficacy will necessitate a trial-and-error approach. At present, AUTOMATIC1111 stands as one of the most often favored options.

Prompt Matrix

The Stable Diffusion Prompt Matrix is an additional feature incorporated into the Stable Diffusion technology. Its purpose is to enhance the generation of AI images by enabling users to combine many text prompts during the image creation process. For the process to effectively merge the prompts into one, it is necessary to inform Stable Diffusion of the specific prompts that you wish to utilize within the Prompt Matrix.

While both the Prompt Matrix and Stable Diffusion Checkpoint Merger serve the purpose of integrating essential components required for image production, it is important to note that they possess distinct characteristics. Hence, it is imperative to comprehend the distinctions among these functions and their potential synergistic utilization for enhancing the quality of AI-generated images. The utilization of the Prompt Matrix facilitates the simultaneous testing of diverse styles by enabling the combination of multiple prompts to augment the visual representation of images. By utilizing this method, users will be able to optimize their time distribution, allowing for increased productivity in generating images for both research purposes and business clients.

The concept of stable diffusion refers to the process of a substance or entity spreading or dispersing in a consistent and predictable manner throughout time. The Prompt Matrix facilitates the identification of discrepancies that arise from modifying strings, parameters, and spells across different styles. To achieve optimal stability in Diffusion, the user needs to experiment with different techniques with the Prompt Matrix until the appropriate mix of prompts is attained.

Given the multitude of variables containing multiple possibilities, the probability of an individual selecting similar options across every parameter is extremely low or nonexistent. However, when selecting to rely on a specifically trained checkpoint or model, it is common to observe the generation of comparable-style images and faces. Alternatively, by incorporating the work of a particular artist, for example, art by a well-known artist within a prompt, it is possible that the aesthetic impact will result in the AI-generated content bearing resemblance.

Image-to-Image Generation

Stable diffusion models and their various adaptations are highly effective in the generation of innovative visual representations. However, it is often the case that we have limited influence over the visuals that are produced. The Image-to-Image tool provides users with a certain degree of control over the style of the generated image. However, it is important to note that the pose and structural characteristics of objects in the resulting image may exhibit significant variations. In order to address this concern, a novel neural network for image production called ControlNet has been developed, which is based on Stable Diffusion.

ControlNet

The ControlNet neural network architecture has been specifically developed to effectively handle diffusion models by integrating supplementary conditions. The weights of neural network blocks are replicated into two separate copies, one being locked and the other being "trainable". The copy that is referred to as trainable acquires knowledge of the desired state through the learning process, whereas the copy that is referred to as locked maintains the original model without any modifications. This methodology guarantees that the utilization of limited datasets containing image pairings does not undermine the reliability and effectiveness of diffusion models that are suitable for deployment. The concept of "zero convolution" refers to a specific type of convolution operation that utilizes a 1×1 kernel. In this operation, both the weight and bias parameters are initialized at zero. Prior to undergoing training, it is seen that all zero convolutions result in a zero output, so effectively mitigating any potential distortion that may be induced by ControlNet. The training procedure does not involve training any layer from the start. Instead, it entails fine-tuning the existing model while ensuring the preservation of its original integrity. This approach facilitates training on devices with limited computational resources, including small-scale or personal devices. ControlNet is an innovative approach for conditioning input images and prompts in the context of image production. The utilization of diverse techniques such as pose estimation, edge detection, depth mapping, and others enables us to exert control over the process of generating the final image. In order to gain a comprehensive understanding of the efficacy of ControlNet, it is imperative to delve further into the construction and training methodologies employed in its development [15]. ControlNet initially generates two duplicate copies of a pre-trained large image diffusion model. The trainable copy acquires knowledge from task-specific datasets during the training process, so affording us enhanced control during the inference stage. ControlNet is trained by utilizing a pre-existing Stable Diffusion model that has been previously trained on a vast dataset comprising billions of images. Two copies are generated from the Stable Diffusion model: one of them has locked weights that remain unchanged, while the second copy has weights that can be trained that could be adjusted during training.

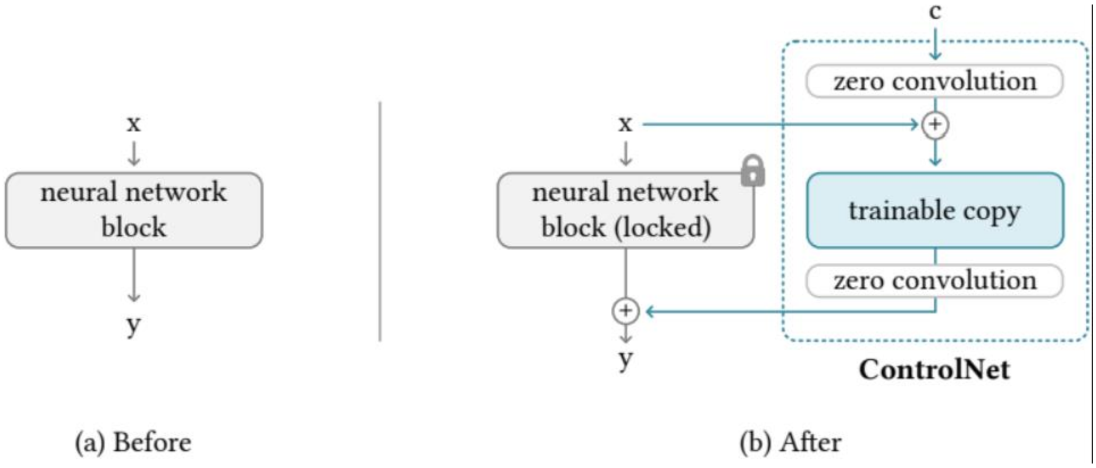


Fig. 5. ControlNet [15]

The trainable phase of the model undergoes training using external factors. The conditioning vector, denoted as c, plays a crucial role in enabling ControlNet to exert control over the global behavior of the neural network. During the training process, the settings of the locked copy remain unchanged. It is now recognized that ControlNet provides the ability to have control over our messages through the integration of task-specific training. To ensure effectiveness, ControlNet was trained to manage a comprehensive image diffusion model, enabling the gathering of task-specific learning from both the prompt and an input image.

5 Findings, Implementation, Experiments and Result

Findings

A viable approach to obtaining cutting-edge synthesis outcomes in a variety of data formats, especially picture data — is through the use of diffusion models. Use of the diffusion models provided by Stable Diffusion results in notable gains in image detail and quality. With the help of written descriptions, these models work by progressively bringing images from a condition of random noise to coherence. Diffusion models implemented on local hardware facilitate the addition of additional information and depth to images, therefore improving the synthesis process as a whole. Notably, diffusion models' adaptability and effectiveness in processing general conditioning inputs are improved when combined with model architecture, as in the case of LoRA and XL-XDXL 1.0 models. Therefore, the study emphasizes how machine learning and generative AI technologies can assist people in AI image generating projects.

Implementation

The implementation of Stable Diffusion's diffusion models requires careful consideration of hardware specifications. Running these models on local hardware allows for more control and flexibility, even if their direct operation in pixel space usually requires substantial GPU VRAM resources. Choosing the right software versions and obtaining models from specified repositories are necessary steps in the installation process. Moreover, the synthesis process is streamlined by the user interface's simple buttons for creating graphics from textual prompts.

Experiments

Diffusion model experiments demonstrate the effectiveness of these models in gradually learning data distributions via denoising procedures. To be more precise, these models use a series of denoising autoencoders to forecast denoised versions of noisy inputs, which leads to better synthesis results. To appreciate these models, one must comprehend that they are composed of a series of denoising autoencoders, each of which is trained to iteratively refine noisy inputs. The study emphasizes the potential of diffusion models for image synthesis through experimentation and emphasizes the significance of taking into account elements like model architecture and training procedures for the best outcomes.

Hardware Requirements and Stable Diffusion Installation

Stable Diffusion was successfully set up and installed on a local computer so that it could produce images in response to text requests. For this purpose, the hardware specs that followed worked well:

- GPU: The experiment employed an NVIDIA GeForce RTX 3070 with a minimum of 4GB VRAM and a preference for 8GB.
- CPU: 4-core (this experiment used an AMD Ryzen 5 5600 6-core processor).
- RAM: 8GB minimum (in this experiment, 16GB DDR4 was used).
- Installation requires at least 10GB of storage (2TB SSD utilized in this experiment).
- System software: Windows 10 or above

See the Stable Diffusion web UI repository (https://github.com/AUTOMATIC1111/stable-diffusion-webui) for comprehensive installation instructions. The suggested Python version, instructions for installing the program, and methods for creating directories are all described in this document.

Installing Models

Local execution is usually the best option for stable diffusion models, including well-known models like LORA and XL-SDXL. Repositories like GitHub and Hugging Face [16] offer these models for download. Once a Stable Diffusion model has been selected, it must be installed by downloading the required files to the correct directory on your computer from the specified source repository.

User Interface

By installing Stable Diffusion with Automatic1111, you may use text prompts to generate images, which is a valuable tool. Users get access to a web-based interface after completing the setup procedure, which may include typing commands in the Command Line Interface (CMD). This interface serves as an easy-to-use center for creating images. It offers several options for adjusting the generated image's creative style. Text prompts allow users to provide detailed descriptions of the desired material, including people, objects, and settings. The interface also provides a number of settings, such resolution and detail level, for adjusting the image. Users can experiment with these settings to produce one-of-a-kind, customized photos that realize their imaginative thoughts.

Image Generation Parameters

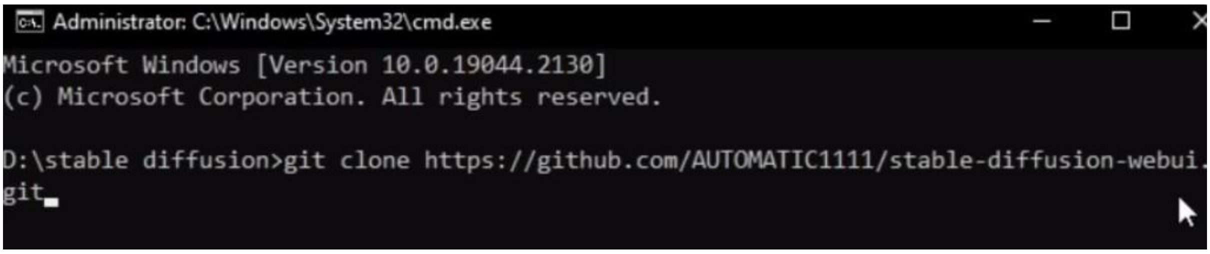
The process by which checkpoints and models are selected is crucial in deciding what the resulting image looks, as if it can accommodate different tastes from realistic to anime-style. Users can designate items to be included or deleted from the artwork and communicate their preferred visual representation by using the interface's places for positive and negative prompts.

Developing and employing stored styles, setting up sampling techniques, modifying batch size and count to manage processing resources, modifying CFG scale to modulate creativity, defining image dimensions, and controlling seed values for repeatability are more options. Together, these characteristics have an impact on the image-generating process and provide users with a variety of customizing options.

In addition, the platform allows for sophisticated features like automatic image storage, image-to-image conversions, and focused editing via the extras and imprint tabs. These characteristics improve Stable Diffusion's adaptability and usefulness for a range of picture synthesis applications.

Results

There will be a newly created folder with files and download may take some time (see Figure 6).



```
Administrator: C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.2130]
(c) Microsoft Corporation. All rights reserved.

D:\stable diffusion>git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui.
git_
```

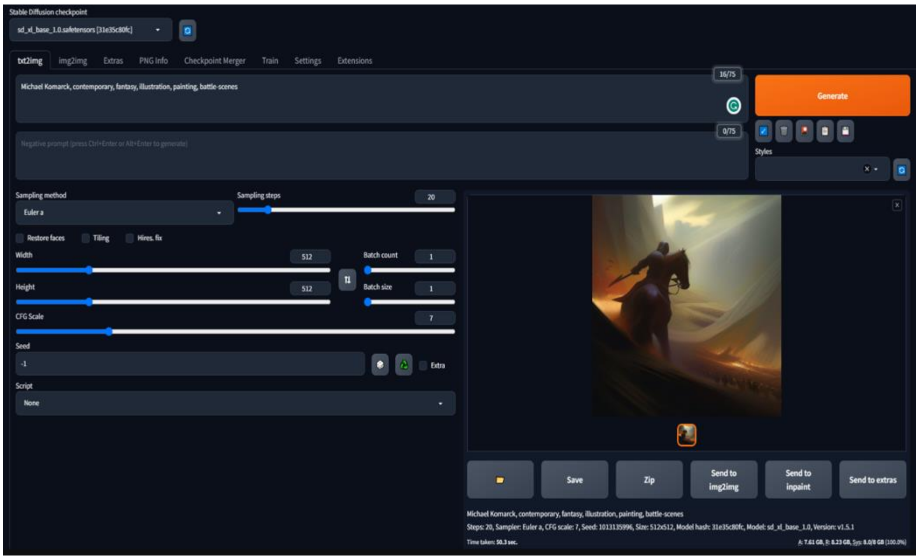
Fig. 6. Clone the repository

Experiments Diffusion Models

The process of installing models is relatively simple. Before starting the experiment process, it is important that you first find a model that fulfills the specific requirements and objectives available. I have used the SD XL 1.0 base model [17] [18] to conduct the experiment for results.

Generating Image using Prompt Text (txt2img)

Prompt: Michael Komarck, contemporary, fantasy, illustration, painting, battle-scene.



Stable Diffusion checkpoint: sd_xl_base_1.0.safetensors [31e35c80fc]

Michael Komarck, contemporary, fantasy, illustration, painting, battle-scenes

Negative prompt (press Ctrl+Enter or Alt+Enter to generate)

Sampling method: Euler a

Sampling steps: 20

□ Restore faces □ Tiling □ Hires. fix

Width: 512

Height: 512

Batch count: 1

Batch size: 1

CFG Scale: 7

Seed: -1

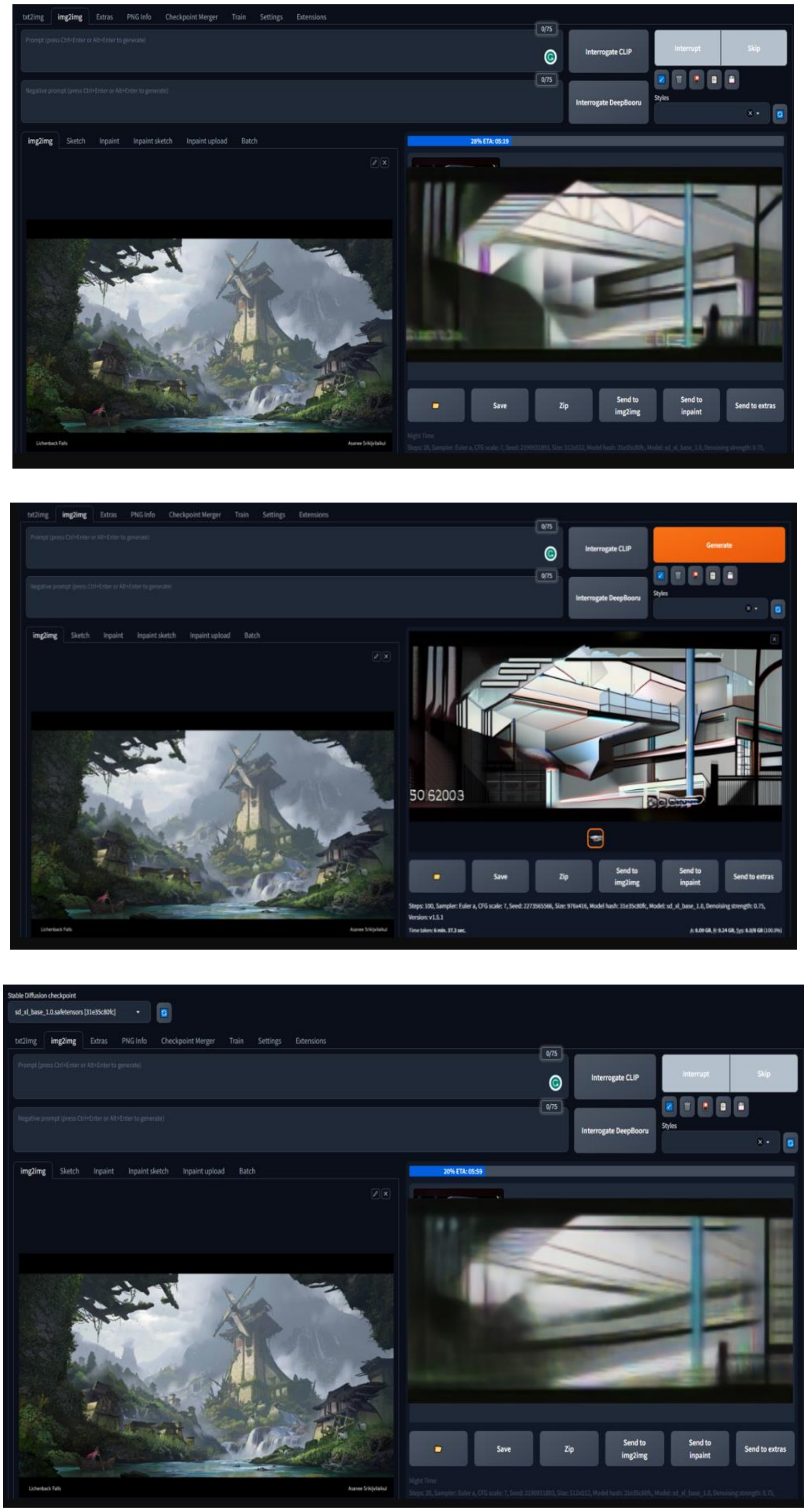
Script: None

Michael Komarck, contemporary, fantasy, illustration, painting, battle-scenes
Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 1013135996, Size: 512x512, Model hash: 31e35c80fc, Model: sd_xl_base_1.0, Version: v1.5.1
Time taken: 50.3 sec.
A: 7.61 GB, R: 8.23 GB, Sys: 8.0/8 GB (100.0%)

Fig. 7. Generating image using Prompt Text

Generating Image using Prompt Text (img2img)

The result did not come good enough for img2img generation as the source of images was less and prompt text addition with it could generate a better result.



Steps: 20, Sampler: Euler a, CFG scale: 7, Seed: 2193031893, Size: 512x512, Model hash: 31e35c80fc, Model: sd_xl_base_1.0, Denoising strength: 0.75,

Steps: 100, Sampler: Euler a, CFG scale: 7, Seed: 2273565566, Size: 976x416, Model hash: 31e35c80fc, Model: sd_xl_base_1.0, Denoising strength: 0.75,
Version: v1.5.1

Time taken: 6 min. 37.3 sec.

Fig. 8. Experimental samples
Image source: Asanee Srikijvilaikul (Artist)

Txt2img result of generation was very well and with a more refined prompt can easily generate outstanding images.

7 Conclusion

In conclusion, Stable Diffusion Automic1111 and models do help computer GPUs run machine deep learning and AI to generate images with the use of predictive text to identify user input requests for generative images. Stable Diffusion is open source and free to use. The entire project was to conduct the operation on a local machine, so an individual person doesn't really have to use a paid AI image generation application like Midjourney.

Key Understandings

SWOT analysis on Stable Diffusion for commercial generative AI image service.

Strengths

- Creative Potential: Generative AI image generation can create new and novel visuals that are impossible with traditional methods.
- Efficiency: Generative AI can swiftly create many photos, saving time and resources over manual creation or photo sessions.
- Customization: The technology can be tailored to design needs and user preferences. AI-generated images can keep a consistent style or theme, which can help branding and visual identity.
- Exploration: It allows for new styles, artistic combinations, and creative bounds.

Weaknesses

- Quality Control: Generated images may lack human-created details and aesthetic touch, causing quality issues.
- Unpredictability: The AI's outputs can vary, making it hard to ensure every image match requirement. The quality and diversity of the training data used to train the AI model strongly affects the quality of output photos.
- Ethics: Creating photos may involve offensive, inappropriate, or copyright-infringing content. Developing and training effective Generative AI models requires computational power, time, and experience.

Opportunities

- Personalized Art Generation: Generative AI can create personalized art to suit individual tastes and interior design trends.
- Automated Design: It lets designers focus on more complicated or creative design features by automating some elements.
- Content Generation: AI-generated visuals can enhance marketing, advertising, and social media content.
- Virtual Environments: Generative AI might create VR and AR backgrounds and environments. The technique could help human designers and AI models collaborate and leverage one other's skills.

Threats

- Job Displacement: AI may displace designers and artists as it advances.
- Overused Styles: Wide use of AI-generated styles may reduce visual diversity.
- Artificiality: AI-generated images may lack the originality and emotional impact of human art.
- Misuse: The technology could be used to create false or dangerous content. Legal and copyright difficulties may arise over ownership and usage rights when AI-generated content becomes more common.

Recommendations

- Using a GPU that has more vRAM (DDR6)
- Look for models that can be combined with one another for better checkpoints.
- Keep the Stable Diffusion base files of Automic1111 updated

References

[1] B. Ahamed, M. R. H. Polas, A. I. Kabir, A. S. M. Sohel-Uz-Zaman, A. A. Fahad, S. Chowdhury, *et al.*, "Empowering Students for Cybersecurity Awareness Management in the Emerging Digital Era: The Role of Cybersecurity Attitude in the 4.0 Industrial Revolution Era," *SAGE Open*, vol. 14, p. 21582440241228920, 2024.

[2] R. Karim, S. Vyas, and A. I. Kabir, "Legal Challenges of Digital Twins in Smart

Manufacturing," in *The International Conference on Recent Innovations in Computing*, 2022, pp. 843-854.

[3] M. R. H. Polas, A. I. Kabir, A. A. Jahanshahi, A. S. M. Sohel-Uz-Zaman, R. Karim, and M. I. Tabash, "Rural entrepreneurs behaviors towards green innovation: Empirical evidence from Bangladesh," *Journal of Open Innovation: Technology, Market, and Complexity,* vol. 9, p. 100020, 2023.

[4] M. R. Mia and A. I. Kabir, "Post-Covid Psychological Impact on Social Media Users: A Study on Twitter Users," 2022.

[5] A. I. Kabir, S. Mitra, S. Akter, M. ISLAM, and S. S. Das, "Developing a Network Design for a Smart Airport Using Cisco Packet Tracer," *Informatica Economica,* vol. 26, 2022.

[6] M. R. H. Polas, A. I. Kabir, A. S. M. Sohel-Uz-Zaman, R. Karim, and M. I. Tabash, "Blockchain technology as a game changer for green innovation: Green entrepreneurship as a roadmap to green economic sustainability in Peru," *Journal of Open Innovation: Technology, Market, and Complexity,* vol. 8, p. 62, 2022.

[7] A. I. Kabir, S. Akter, and S. Mitra, "Students engagement detection in online learning during COVID-19 pandemic using r programming language," *Informatica Economica,* vol. 25, pp. 26-37, 2021.[8] A. I. Kabir, S. Mitra, and S. S. Das, "Development of a face-mask detection software using artificial intelligence (ai) in python for COVID-19 protection," *Journal of Management Information and Decision Sciences,* vol. 24, pp. 1-15, 2021.

[9] A. I. Kabir, K. Ahmed, and R. Karim, "Word cloud and sentiment analysis of Amazon earphones reviews with R programming language," *Informatica Economica,* vol. 24, pp. 55-71, 2020.

[10] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, "The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R," *Informatica Economica,* vol. 22, 2018.

[11] T. Smith. (2022, 21st March, 2024). *Leaked deck raises questions over Stability AI's Series A pitch to investors*. Available: https://sifted.eu/articles/stability-ai-fundraise-leak.

[12] R. Rombach. (2022, 21st March). *Stability-AI/stablediffusion*. Available: https://github.com/Stability-AI/stablediffusion

[13] N. Dehouche and K. Dehouche, "What’s in a text-to-image prompt? The potential of stable diffusion in visual arts education," *Heliyon,* vol. 9, 2023.

[14] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems,* vol. 33, pp. 6840-6851, 2020.

[15] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3836-3847.

[16] M. Woolf. (2022, 21st March). *sdxl-wrong-lora*. Available: https://huggingface.co/minimaxir/sdxl-wrong-lora

[17] S. AI. (2022, 21st March). *stable-diffusion-xl-base-1.0* Available: https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0

[18] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, *et al.*, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023.

Ahmed Imran KABIR is an experienced Faculty Member in the School of Business and Economics at United International University with working experience in the Management Information Systems and Business Analytics subjects. Strong educational background with a Master of Science in Business Analytics from Texas A&M University, United States. He has several research papers published in international and national journals and in ISI and Scopus- indexed journals Management Information System, Big Data Analytics, Blockchain Technology, and Multi-disciplinary studies.

Limon MAHOMUD is a graduate with an Honors Bachelor of Business Administration in Management Information System (MIS) in 2023 from the School of Business and Economics at United International University, Bangladesh. His academic years have helped him build a strong foundation in MIS principles, learning Python, Cisco Packet Tracer, HTML, and Microsoft Project.

Abdullah Al FAHAD is a Faculty Member of MIS in the School of Business and Economics at United International University. Abdullah completed his “BBA with a Finance concentration” from University of North Texas and “MS in Information Technology Management” from University of Dallas, Texas, in The United States. Abdullah’s corporate experience also speaks volume about his in-depth knowledge as he placed himself with prominent Fortune 500 giants such as J P Morgan Chase in Arlington, Texas, where he started a journey as a “Financial Analyst” and climbed the ladder to “Quality Control Analyst” and a “Subject Mater Expert”. After his Masters, Abdullah were positioned as a “Software QA Tester” at “CVS Health” in Richardson, Texas. Abdullah Also hold a directorship at Greenland Group, in Dhaka, Bangladesh. Abdullah’s objective now is to have a progressive academic career and contribute his knowledge in research and hence improve his efficiency as an educator. His research interest areas are – Management Information System, Data Analytics, Data Visualization, Python Programming etc.

Ridwan AHMED attained his Honours Bachelor of Business Administration with a concentration in Management Information System (MIS) in 2024 from the School of Business and Economics at United International University. Throughout his academic journey, he acquired proficiency in various technical tools and languages including Python, Cisco Packet Tracer, HTML, MySQL Workbench, and the data visualization tool Tableau. These skills have had a major role in helping him build a solid foundation in Management Information System (MIS) principles.